

**ФГОУ ВПО "КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ"
ФАКУЛЬТЕТ ПРИКЛАДНОЙ ИНФОРМАТИКИ
КАФЕДРА КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ И СИСТЕМ**

Е.В. ЛУЦЕНКО

ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

**Допущено
Министерством сельского хозяйства Российской Федерации
в качестве учебного пособия
для бакалавриата высших учебных заведений,
обучающихся по специальности
"Прикладная информатика"**

**Издание 2-е
дополненное и исправленное**

КРАСНОДАР – 2013

УДК 303.732.4

ББК 65.05

Л 86

Рецензенты:

Барановская Т.П., профессор, доктор экономических наук, заведующая кафедрой системного анализа и обработки информации, Кубанский государственный аграрный университет (г. Краснодар)

Ключко В. И., профессор, доктор технических наук, заведующий кафедрой вычислительной техники и автоматизированных систем управления, Кубанский государственный технологический университет (г.Краснодар)

Луценко Е. В.

Л 86 Интеллектуальные информационные системы: Учебное пособие для бакалавриата специальности: "Прикладная информатика". – Краснодар: КубГАУ. 2013. – 645 с.

ISBN 5-94672-060-0

Учебное пособие состоит из трех частей: курса лекций; практикума и программы самостоятельной работы студентов.

Курс лекций включает 16 лекций, сгруппированных в 4 раздела: введение в интеллектуальные информационные системы; теоретические основы и эксплуатация универсальной когнитивной аналитической системы "Эйдос"; принципы построения интеллектуальных информационных систем; применение и перспективы систем искусственного интеллекта.

Практикум базируется на универсальной когнитивной аналитической системе "Эйдос", разработанной автором данного учебного пособия, и включает 10 лабораторных работ.

Программа самостоятельной работы студентов по дисциплине включает теоретические вопросы и практические задания, выносящиеся на экзамен по дисциплине и государственный экзамен, а также список основной и дополнительной литературы, включая Internet-сайты по проблематике искусственного интеллекта.

Для студентов очной и заочной форм обучения, аспирантов, преподавателей и научных работников, интересующихся проблематикой систем искусственного интеллекта.

Ил. 216. Табл. 91. Формул 94. Библиогр.: 275 назв.

ISBN 5-94672-060-0 © Е.В. Луценко, 2013 г.
© КубГАУ 2013 г.

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	5
ЧАСТЬ I. КУРС ЛЕКЦИЙ.....	8
РАЗДЕЛ 1.1. ВВЕДЕНИЕ В ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ	9
1.1.1. ЛЕКЦИЯ-1. Интеллектуальные информационные системы, как закономерный и неизбежный этап развития средств труда	9
1.1.2. ЛЕКЦИЯ-2. Определение и критерии идентификации систем искусственного интеллекта. Место СИИ в структуре обработки информации и управления	27
РАЗДЕЛ 1.2. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ И ПРИМЕНЕНИЕ УНИВЕРСАЛЬНОЙ КОГНИТИВНОЙ АНАЛИТИЧЕСКОЙ СИСТЕМЫ "ЭЙДОС"	60
1.2.1. ЛЕКЦИЯ-3. Теоретические основы системно-когнитивного анализа и место СИИ в структуре АСУ	60
1.2.2. ЛЕКЦИЯ-4. Системная теория информации и семантическая информационная модель	94
1.2.3. ЛЕКЦИЯ-5. Методика численных расчетов (алгоритмы и структуры данных)	202
1.2.4. ЛЕКЦИЯ-6. Технология синтеза и эксплуатации приложений в системе "Эйдос"	238
РАЗДЕЛ 1.3. ПРИНЦИПЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ.....	261
1.3.1. ЛЕКЦИЯ-7. Системы с интеллектуальной обратной связью и интеллектуальными интерфейсами	261
1.3.2. ЛЕКЦИЯ-8. Автоматизированные системы распознавания образов	277
1.3.3. ЛЕКЦИЯ-9. Математические методы и автоматизированные системы поддержки принятия решений	290
1.3.4. ЛЕКЦИЯ-10. Экспертные системы	307
1.3.5. ЛЕКЦИЯ-11. Нейронные сети	313
1.3.6. ЛЕКЦИЯ-12. Генетические алгоритмы и моделирование биологической эволюции	341
1.3.7. ЛЕКЦИЯ-13. Когнитивное моделирование	349
1.3.8. ЛЕКЦИЯ-14. Выявление знаний из опыта (эмпирических фактов) и интеллектуальный анализ данных (data mining)	357
РАЗДЕЛ 1.4. ПРИМЕНЕНИЕ И ПЕРСПЕКТИВЫ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА	366
1.4.1. ЛЕКЦИЯ-15. Области применения систем искусственного интеллекта	366
1.4.2. ЛЕКЦИЯ-16. Перспективы развития систем искусственного интеллекта, в т.ч. в Internet	397
ЧАСТЬ II. ЛАБОРАТОРНЫЙ ПРАКТИКУМ.....	415
2.1. СООТВЕТСТВИЕ ЛАБОРАТОРНЫХ РАБОТ РАБОЧИМ ПРОГРАММАМ ПО СПЕЦИАЛЬНОСТЯМ	415
2.2. СТРУКТУРА ЛАБОРАТОРНОЙ РАБОТЫ, ПОРЯДОК ЕЕ УСТАНОВКИ, ВЫПОЛНЕНИЯ И СДАЧИ	416
2.2.1. Структура лабораторной работы	416
2.2.2. Инструкция по установке лабораторной работы	416
2.2.3. Выполнение и сдача лабораторной работы	418

2.3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ПРОВЕДЕНИЯ ЛАБОРАТОРНЫХ РАБОТ В КОМПЬЮТЕРНОМ КЛАССЕ	418
2.3.1. Материально-техническое обеспечение	418
2.3.2. Общее программное обеспечение	418
2.3.3. Специальное программное обеспечение	419
2.4. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ЛАБОРАТОРНЫМ РАБОТАМ	419
ЛР-1: "Прогнозирование вероятных пунктов назначения железнодорожных составов"	419
ЛР-2: "Прогнозирование учебных достижений студентов на основе их имеджевых фотороботов"	437
ЛР-3: "Прогнозирование учебных достижений студентов на основе особенностей их почерка"	457
ЛР-4: "Прогнозирование учебных достижений студентов на основе информации об их социальном статусе"	462
ЛР-5: "Идентификация слов по входящим в них буквам"	486
ЛР-6: "Атрибуция анонимных и псевдонимных текстов"	518
ЛР-7: "Идентификация и классификация натуральных чисел по их свойствам"	538
ЛР-8: "Идентификация трехмерных тел по их ортогональным проекциям"	551
ЛР-9: "Прогнозирование количественных и качественных результатов выращивания зерновых колосовых и поддержка принятия решений по выбору агротехнологий"	561
ЛР-10: "Исследование случайной семантической информационной модели при различных объемах выборки"	568
ЧАСТЬ III. ЗАОЧНОЕ ОБУЧЕНИЕ, САМОСТОЯТЕЛЬНАЯ РАБОТА СТУДЕНТОВ И ПОДГОТОВКА К ЭКЗАМЕНУ	574
3.1. ИЗУЧЕНИЕ ДИСЦИПЛИНЫ ПРИ ЗАОЧНОЙ ФОРМЕ ОБУЧЕНИЯ	574
3.2. ПРОГРАММА САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТА ПО ДИСЦИПЛИНЕ	574
3.2.1. Общие положения	574
3.2.2. Наименования тем для самостоятельной работы студентов	575
3.2.3. Примерная структура реферата, курсовой или дипломной работы и элементы научного исследования в самостоятельной работе	577
3.2.4. Методика использования репозитория UCI для оценки качества математических моделей систем искусственного интеллекта	578
3.3. ТЕОРЕТИЧЕСКИЕ ВОПРОСЫ	602
3.3.1. Вопросы, выносящиеся на экзамен по дисциплине	602
3.3.2. Вопросы, относящиеся к дисциплине "ИИС", выносящиеся на государственный экзамен	614
3.4. ПРАКТИЧЕСКИЕ ЗАДАНИЯ	614
ЗАКЛЮЧЕНИЕ	617
ЛИТЕРАТУРА	619
ОСНОВНАЯ ЛИТЕРАТУРА	619
ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА	619
INTERNET-САЙТЫ	634
КРАТКИЙ СЛОВАРЬ ТЕРМИНОВ ПО СК-АНАЛИЗУ И СИСТЕМАМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА	636

ПРЕДИСЛОВИЕ

Вниманию читателей предлагается второе издание учебного пособия по дисциплине: "Интеллектуальные информационные системы", которая читается для студентов, обучающихся по специальности «Прикладная информатика».

Учебное пособие состоит из трех частей:

1. Курса лекций.
2. Лабораторного практикума.
3. Программы самостоятельной работы студентов.

Количество и объем лекций и лабораторных работ соответствует утвержденным рабочим программам, образовательному стандарту по данным специальностям и рекомендациям УМО. При этом на лекции отводится 32 часа (16 лекций), на лабораторные занятия – 32 часа (8 лабораторных работ по 4 часа), и на самостоятельную работу студентов – 38 часов. С этим в определенной мере связаны ограниченный объем учебного пособия и краткость изложения материала, что во многом предопределило его обзорный характер.

Вместе с тем, автор ставил перед собой задачу не только дать студентам основы теоретических представлений по изучаемой дисциплине, насколько это возможно в обзорном курсе, но и предоставить им возможность приобретения определенных практических умений и навыков использования интеллектуальных информационных технологий для решения различных учебных задач по их специальностям.

Курс лекций включает 16 лекций, сгруппированных в 4 раздела:

1. Введение в интеллектуальные информационные системы.
2. Теоретические основы и эксплуатация универсальной когнитивной аналитической системы "Эйдос".
3. Принципы построения интеллектуальных информационных систем.
4. Применение и перспективы систем искусственного интеллекта.

Практикум базируется на универсальной когнитивной аналитической системе "Эйдос", разработанной автором пособия, и включает 10 лабораторных работ. Именно этим обусловлено размещение второго раздела, посвященного системе "Эйдос", раньше раздела о принципах построения интеллектуальных информационных систем.

Лабораторные работы однотипны по технологии и иллюстрируют возможности применения систем искусственного интеллекта в различных предметных областях. При этом в каждой лабораторной работе полностью изучается весь технологический цикл, но при этом основное внимание акцентируется на каком-либо из основных элементов технологии, который изучается более углубленно. Таким образом, студенты, пропустившие одну или даже несколько лабораторных работ, что, к сожалению, приходится учитывать, могут в определенной мере компенсировать последствия пропусков. В целом такой подход обеспечи-

вает создание у студентов целостной и универсальной картины применения систем искусственного интеллекта.

Каждая лабораторная работа включает: краткую теорию; одно или несколько заданий; примеры решения; контрольные вопросы; литературу, рекомендуемую для изучения.

Количество лабораторных работ соответствует утвержденным рабочим программам, образовательному стандарту по данным специальностям и рекомендациям УМО. Таким образом, преподаватель имеет возможность осуществлять выбор любых 8 из 10 вариантов разработанных лабораторных работ по своему усмотрению с учетом требования специальности. В последующем предполагается существенно увеличить количество предлагаемых для выбора вариантов лабораторных работ.

Учебное пособие включает также программу самостоятельной работы студентов по дисциплине, теоретические вопросы и практические задания, выносящиеся на экзамен, а также список основной и дополнительной литературы по дисциплине, включая Internet-сайты по проблематике искусственного интеллекта.

Хотелось бы отметить также, что системы искусственного интеллекта относятся к наиболее бурно развивающемуся направлению, буквально пронизывающему и революционизирующему практически все остальные направления развития современных информационных технологий. Объем общедоступной информации по этой проблематике огромен и очень быстро возрастает.

Поэтому автор полностью осознает, что данное учебное пособие ни в коей мере не может претендовать на полноту изложения и является не более чем кратким введением в проблематику искусственного интеллекта, причем в авторской интерпретации. На это, в общем-то, и рассчитан обзорный курс, на который в учебном плане отведено лишь 68 часов.

При изложении материала не удалось избежать некоторых повторов, что, правда, может быть как-то оправдано с методической точки зрения ("Повторение – мать учения").

В то время необходимо отметить, что наука о системах искусственного интеллекта пока даже не имеет общепринятого названия, является одной из самых бурно развивающихся, новые результаты появляются в ней чуть ли не ежедневно, многие ее положения спорны и находятся в процессе обсуждения, и говорить о об этой науке, как об "устоявшейся" не приходится и еще, по-видимому, долго не придется. По мнению автора в этой ситуации полезнее для дела, т.е. для качества обучения, не загаживать проблемы науки, строя изложение так, как будто они все уже решены, а открыто показывать их, т.к. они являются "точками роста" науки. По этой же причине автор, сам являющийся активно работающим в области систем искусственного интеллекта исследователем и разработчиком, счел возможным в ряде случаев выразить в порядке научной дискуссии и свою точку зрения, даже если она ранее не публиковалась в научной

печати. Поэтому данное "учебное пособие" в какой-то мере является и "научной работой". Кроме того авторское восприятие проблематики довольно сильно сказалось как на выборе материала, так и на характере его изложения.

Особенно хотелось бы выразить благодарность авторам web-публикаций и сайтов, по проблематике искусственного интеллекта, материалы которых были использованы при подготовке данного учебного пособия. Везде, где эти материалы использовались, на них сделаны ссылки:

– Васильев Л.Г. *Три парадигмы понимания: анализ литературы вопроса*. <http://newasp.omskreg.ru/intellect/f54.htm>;

– Орлов А.И. *"Высокие статистические технологии"*: <http://antorlov.chat.ru>;

– Максимов В.И., Корноушенко Е.К., Гребенюк Е.А., Григорян А.К. Качаев С.В., Кулинич А.А., Райков А.Н., Макаренко Д.И., Коврига С.В., специалисты сектора-51 ИПУ РАН и сотрудничающие с ними по комплексной теме: *"Когнитивный анализ и моделирование ситуаций"*: <http://www.ipu.ru/labs/lab51/projects.htm>;

– Сотник С.Л. *Конспект лекций по курсу "Основы проектирования систем искусственного интеллекта"*: (1997-1998), <http://neuroschoo1.narod.ru/books/sotnik.html>;

– Терехов С.А. *Лекции по теории и приложениям искусственных нейронных сетей. Лаборатория Искусственных Нейронных Сетей НТО-2, ВНИИТФ, Снежинск* (http://alife.narod.ru/lectures/neural/Neu_index.htm);

– Кива В. *Данные, информация, знания*. <http://vlak.webzone.ru/rus/it/knowledge.html>

Автор благодарен доктору экономических наук профессору Барановской Т.П., доктору технических наук профессору Лойко В.И., доктору технических наук профессору Ключко В. И, кандидату экономических наук профессору Курносову С.А. и кандидату технических наук доценту Лаптеву В.Н. за поддержку и многочисленные конкретные советы и предложения, высказанные при написании данного учебного пособия.

Для студентов, аспирантов, преподавателей и научных работников, интересующихся проблематикой систем искусственного интеллекта.

ЧАСТЬ I. КУРС ЛЕКЦИЙ

Лекции имеют типовую структуру, которая может включать следующие разделы:

- учебные вопросы;
- изложение учебных вопросов;
- контрольные вопросы к лекции;
- рекомендуемую литературу.

В разделе "**Учебные вопросы**" перечисляются учебные вопросы, раскрываемые в данной лекции.

В разделе "**Изложение учебных вопросов**" излагается теоретический материал по учебным вопросам.

Контрольные вопросы к лекции служат для проверки качества усвоения и понимания материала и могут быть включены в экзаменационные билеты.

В разделе "**Рекомендуемая литература**" приводятся конкретные литературные источники и Internet-сайты, использованные при разработке данной лекции и рекомендуемые для более углубленного изучения вопросов, затрагиваемых в лекции. При ссылках на литературу в квадратных скобках указываются номера источников по списку "Дополнительная литература", приведенному в конце учебного пособия.

РАЗДЕЛ 1.1. ВВЕДЕНИЕ В ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

1.1.1. ЛЕКЦИЯ-1. Интеллектуальные информационные системы, как закономерный и неизбежный этап развития средств труда

"Системы искусственного интеллекта позволяют с успехом решать сложнейшие проблемы, которых не возникало до создания этих систем"

/Из компьютерного фольклора/

Учебные вопросы:

1. Основные положения информационно-функциональной теории развития техники
2. Информационная теория стоимости
3. Интеллектуализация – генеральное направление и развития информационных технологий

1.1.1.1. Основные положения информационно-функциональной теории развития техники

Теоретические основы информационно-функциональной теории развития техники были разработаны автором в 1979 – 1981 годах и опубликованы в специальных материалах, а также в сокращенном виде в работах [81, 92, 99, 125]. Основываясь на этой теории рассмотрим следующие вопросы:

1. Процесс труда, как информационный процесс.
2. Организм человека и средства труда как информационные системы.
3. Законы развития техники:
 - закон перераспределения функций между человеком и средствами труда;
 - закон повышения качества базиса.
4. Детерминация формы сознания человека функциональным уровнем средств труда.
5. Неизбежность возникновения компьютеров, информационных систем и систем искусственного интеллекта.

1.1.1.1.1. Процессы труда и познания, как информационные процессы снятия неопределенности

Рассмотрим систему: "человек – объект" в точке бифуркации, т.е. в точке, после прохождения которой снимается (уменьшается) неопределенность в поведении этой системы (рисунок 1). Известно, что информация есть количественная мера снятия неопределенности, поэтому рассмотрим два основных направления информационных потоков, которые возможны в этой системе:

1. От человека к объекту: "Труд" (управление).
2. От объекта к человеку: "Познание" (идентификация, обобщение, абстрагирование, сравнение и классификация).



Рисунок 1. Направления потоков информации и локализация снятия неопределенности в процессах труда и познания

Труд представляет собой управляющее, по существу информационное, воздействие на предмет труда, при этом в результате осуществления процесса труда снимается неопределенность состояния предмета труда, в результате чего он трансформируется в продукт труда.

В результате познания снимается неопределенность наших представлений об объекте познания, т.е. снимается неопределенность в состоянии человека, в результате чего он трансформируется из "незнающего" в "знающего".

Если абстрагироваться от направления потока информации и, соответственно, от того, неопределенность в состоянии какой системы снимается (объекта или человека), то, очевидно, что в обоих случаях *количество переданной информации является количественной мерой степени снятия неопределенности*.

С позиций информационно-функциональной теории развития техники труд представляет собой прежде всего информационный процесс, средства труда являются информационными системами, передающими и усиливающими информационные потоки между человеком и внешней средой.

Труд (процесс опредмечивания) предлагается рассматривать как процесс перезаписи информации из образа продукта труда в структуру физической сре-

ды (предмета труда). По мере осуществления этого процесса физическая форма продукта труда за счет записи в ней информации структурируется и выделяется из окружающей среды.

Тело человека и средства труда выступают при этом как канал передачи информации. При передаче по этому каналу информация неоднократно меняет свой носитель и языковую форму представления, т.е. транслируется.

Таким образом, сам процесс передачи информации по каналу связи и запись ее в носитель информации – это и есть тот процесс (труд), который преобразует носитель информации в заранее заданную форму, т.е. в продукт труда.

1.1.1.1.2. Организм человека и средства труда как информационные системы

Очевидно образ продукта труда и сам продукт труда относятся к качественно различным уровням Реальности на которых тождественная по содержанию (семантике) информация просто физически не может находиться в одной и той же языковой, синтаксической форме. Поэтому тело человека и его средства труда как информационный канал, соединяющий качественно различные уровни Реальности, не просто передает информацию с сохранением ее содержания с одного уровня на другой, но при этом с необходимостью преобразует и языковую форму представления информации, т.е. осуществляет ее компиляцию, которая фактически и представляет собой технологический процесс.

Информация образа продукта труда, проявляющаяся первоначально в форме целесообразной и целенаправленной трудовой деятельности, т.е. как свободная информация, преобразуется затем в форму связанной информации, выступающей как покоящиеся полезные свойства продукта труда, определяемые его физической формой и структурой.

Свободная информация, существующая в форме целесообразности процесса труда не имеет стоимости, но образует ее в той мере, в какой преобразуется в информацию, связанную в структуре физической формы продукта труда. При этом смысл (содержание, семантика, качество) информации связанной в продукте труда определяет его потребительскую стоимость, тогда как ее количество связано с абстрактной себестоимостью продукта.

Время, за которое перезаписывается определенное количество информации из образа продукта труда в его физическую структуру определяется информационной пропускной способностью организма человека и его средств труда как информационного канала. Чем это время меньше, т.е. чем выше информационная пропускная способность тела человека и его средств труда, тем выше уровень развития человека и технологии. Так как процесс труда это сознательный процесс, то и количественные и содержательные возможности человека как информационного канала определяются типом и состоянием его сознания. Процесс увеличения информационной пропускной способности сознания человека поддерживается (обеспечивается, сопровождается) соответствующими психофизиологическими изменениями в теле человека.

Роль человека и его средств труда в процессе перезаписи информации из образа продукта труда в структуру его физической формы и в создании потреби-

тельной стоимости и стоимости продукта труда различна. Это различие определяется тем, что в процессе труда человек выполняет лишь ту часть работы по созданию определенного продукта труда, которая заключается в выполнении функций еще не переданных его средствам труда. Та же часть работы, которая состоит в выполнении уже полностью переданных средствам труда функций выполняется ими автоматически, т.е. без участия человека (рисунок 2):

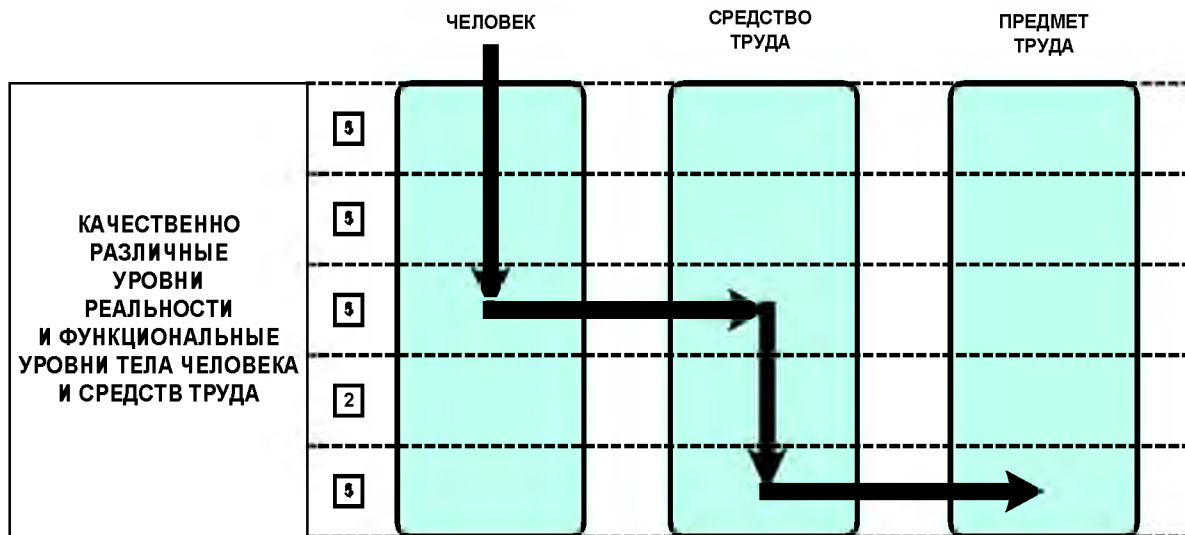


Рисунок 2. Упрощенная схема информационного канала для процессов труда с использованием средств труда

1.1.1.1.3. Законы развития техники

На любом этапе развития общества технологический базис общества основан на тех уровнях Реальности, которые осознаются как объективное при соответствующей данному этапу форме сознания. Технологический прогресс состоит в последовательной передаче трудовых функций организма человека средствам труда.

1.1.1.1.3.1. Закон перераспределения функций между человеком и средствами труда

Развитие средств труда происходит путем последовательной передачи им трудовых функций человеческого организма в результате чего они начинают выполняться средствами труда вне естественных ограничений организма человека, а человек выполняет оставшиеся функции вне ограничений, связанных с необходимостью выполнения переданных функций.

Физический организм выполняет следующие функции, трудовые функции, последовательно передаваемые средствам труда:

1. Функция контакта с физической средой.
2. Функция трансмиссии (передачи и перераспределения энергии).
3. Рабочая функция (преобразование простого движения в сложное и выполняющее работу).
4. Функция двигателя (преобразование формы энергии).
5. Функция преобразования формы информации.

Другие структурные уровни организма человека поддерживают еще ряд функций, связанных с чувственно-эмоциональной и интеллектуальной обработкой информации. Рассмотрение этих функций выходит за рамки данного учебного пособия.

Когда средствам передается очередная трудовая функция – происходит технологическая революция, которая с неизбежностью вызывает революцию экономическую и социальную, а значит переход к новой общественно-экономической формации и соответствующему состоянию сознания (этапу общественного познания). Так формулируется закон перераспределения трудовых функций в системе "человек–машина", т.е. закон количественного повышения базиса.

1.1.1.1.3.2. Закон повышения качества базиса

Формулировка закона повышения качества базиса. Развитие любой системы происходит путем разрешения противоречий между системой и средой в наинизшем качественном уровне системы, в котором они еще не разрешены. Этот уровень называется базисом (базисным). Разрешение противоречия в базисном уровне осуществляется поэтапно, путем перераспределения функций по преобразованию формы информации между внешним и внутренним.

Это перераспределение может осуществляться в двух формах:

- 1) в форме внешнего отчуждения (развитие средств труда и технологии);
- 2) путем внутреннего отчуждения (развитие сознания).

Причем развитие технологии детерминирует соответствующее развитие сознания, а уровень сознания определяет функциональный уровень технологии.

При отчуждении каждой очередной функции базисного уровня (передаче ее средствам труда или осознания ее как "не-Я") происходит количественное изменение системы. При отчуждении всех функций некоторого базисного уровня происходит качественное изменение системы и она переходит к развитию путем разрешения противоречий в следующем, более глубоком чем предыдущий уровне, который и становится базисным.

Когда средствам труда полностью и в массовом масштабе передается последняя функция некоторого относительно автономного уровня организации организма человека, то это вызывает переход к следующей группе общественно-экономических формаций и к следующему типу сознания. При этом человек как объективное начинает осознавать соответствующий качественно новый уровень Реальности и постепенно начинает действовать на нем как человек используя принцип свободы воли, в частности: сначала пользоваться тем, что "лежит на поверхности и ждет когда его возьмут", а затем трудиться и производить для потребления то, чего "на поверхности" не оказалось, и, наконец, производить средства производства. Таким образом при переходе к следующей группе формаций технологический базис общества повышается качественно, т.е. включает в себя средства труда, созданные на тех уровнях Реальности, которые ранее осознавались основной массой людей как субъективные. Так формулируется закон повышения качества базиса.

Таким образом, в процессе развития технологии создаются технические системы, в состав которых входят уровни Реальности, поддерживающие так называемые субъективные функции (чувственно–эмоциональное восприятие и формально–логическое мышление), которые на предыдущих этапах эволюции общества осознавались как субъективные и относились не к базису, а к надстройке. В результате этого *изменяется положение границы* между базисом и надстройкой и соответственно *изменяется содержание* этих понятий, хотя их соотношение остается тем же что и раньше. Конечно, в этой связи изменяется и содержание таких понятий, как "производственная сфера" и "непроизводственная сфера", под которыми ранее понимались соответственно "сфера материального производства" и сфера производства самого человека, т.е. в основном культура, наука, образование и медицина. Становится вполне очевидным, что главной производительной силой является сам человек.

1.1.1.1.4. Детерминация формы сознания человека функциональным уровнем средств труда

Взаимодействие человека со средствами труда приводит не только к созданию определенного материального продукта труда, но и к изменению самого человека. Уровень сознания человека во многом детерминируется функциональным уровнем средств труда, с помощью которых он трудится.

Труд не только создал человека, но через совершенствование форм и способов труда происходит развитие человека и в настоящее время.

А этот организм существует одновременно на многих уровнях Реальности и является значительно более сложным, чем обычно полагают. Функции этих тел также будут в будущем (некоторые в близком будущем) передаваться средствам труда, и в этом состоит блестящая перспектива развития техники, человека и общества.

Таким образом, при использовании средства труда определенного функционального уровня человек учится *не выполнять* функций, переданных этому средству труда, а оставшиеся функции выполняются человеком вне ограничений, связанных с необходимостью выполнения переданных функций. В результате человек частично высвобождается из процесса труда, отходит от него несколько в сторону и у него формируется новый адекватный этому "образ–Я" и сознание: они изменяются таким образом, что трудовые функции, переданные средству труда перестают осознаваться человеком как атрибут "образа–Я".

Здесь неявно предполагается, что если какая-либо функция может быть передана средству труда, то она не может быть атрибутом (неотъемлемой частью) "образа–Я".

Это значит, что происходит такое же изменение сознания и самосознания, как в йоге при (успешной) медитации над мантрами: "Я не это" и "Я есть то".

Этот принцип используется магами, а также почти осознанно применяется в тренажерах, основанных на методах "биологической обратной связи". Такие тренажеры обеспечивают за неделю овладение функциями своего физического тела в такой же степени, какой хатха-йоги добиваются за многие годы упорных тренировок.

В 1981 году Л.А. Бакурадзе и Е.В. Луценко были оформлены заявки на изобретение компьютерной системы, выполняющей все трудовые функции физического тела, обеспечивающую управление с использованием дистанционного мысленного воздействия, т.е. микротелекинеза. По мнению автора, телекинез представляет собой управление физическими объектами путем воздействия на них непосредственно с высших планов без использования физического тела, т.е. тем же способом, с помощью которого любой человек, осознает он это или нет, управляет своим физическим телом. Были предложены технические и программные решения и инженерно – психологические методики. Система предлагалась адаптивной, т.е. автоматически настраивающейся на индивидуальные особенности, "почерк" оператора и его состояние сознания, с плавным переключением на дистанционные каналы при повышении их надежности (которая измерялась автоматически) и могла одновременно с выполнением основной работы выступать в качестве тренажера. Человек, начиная работу с системой в обычной форме сознания с использованием традиционных каналов (интерфейса), имея мгновенную адекватную по форме и содержанию обратную связь об эффективности своего телекинетического воздействия, должен быстро переходить в одну из высших форм сознания, оптимальную для использования телекинеза в качестве управляющего воздействия.

1.1.1.1.5. Неизбежность возникновения компьютеров, информационных систем и систем искусственного интеллекта

Физический организм выполняет следующие трудовые функции:

1. Функция контакта с физической средой.
2. Функция трансмиссии (передачи и перераспределения энергии).
3. Рабочая функция (преобразование простого движения в сложное и выполняющее работу).
4. Функция двигателя (преобразование формы энергии).
5. Функция преобразования формы информации.

Другие структурные уровни организма человека поддерживают еще ряд функций, связанных с чувственно-эмоциональной и интеллектуальной обработкой информации. Рассмотрение этих функций выходит за рамки данного учебного пособия. Но именно с их передачей средствам труда будет связано создание компьютерных систем, не просто имитирующих некоторые стороны этих видов деятельности человека, а действительно реализующих их.

Развитие технологии связано с последовательной передачей всех этих функций средствам труда и, следовательно, настанет черед и функций, связанных с эмоциональной и интеллектуальной обработкой информации. Следовательно, создание систем искусственного интеллекта является столь же неизбежным и закономерным, как и создание рабочей машины или двигателя.

1.1.1.2. Информационная теория стоимости

Рассмотрим учебные вопросы:

1. Связь количества и качества информации с меновой и потребительной стоимостью.
2. Информация, как сырье и как товар: абсолютная, относительная и аналитическая информация. Данные, информация, знания.
3. Стоимость и амортизация систем искусственного интеллектуальных и баз знаний.
4. Источники экономической эффективности систем искусственного интеллекта и интеллектуальной обработки данных (Data mining) с позиций информационной теории стоимости.

1.1.1.2.1. Связь количества и качества информации с меновой и потребительной стоимостью

Информационная теория стоимости разработана автором в 1979 – 1981 годах и опубликована в специальных материалах, а также в сокращенном виде в работах [81, 92, 99, 125].

Эта теория базируется на двух положениях, имеющих очень высокую степень достоверности:

1. Стоимость товара определяется временем, необходимым в обществе на его производство.
2. Создание продукта труда есть информационный процесс воплощения информационного образа этого продукта в предмете труда.

Рассмотрим рисунок 3.



Рисунок 3. Схема образования потребительной стоимости и стоимости в процессе труда с позиций информационной теории стоимости

Информация записанная в структуре продукта труда непосредственно человеком создает и потребительную, и меновую стоимость. Информация же записанная в структуре продукта средствами труда, т.е. без участия человека, автоматически, не увеличивает стоимость этого продукта, хотя и создает его потребительную стоимость.

При этом совершенно неважно, каким образом записана эта информация в самих средствах труда: непосредственно человеком или также с помощью средств труда. Неважно также записана эта информация непосредственно в механической или другой консервативной структуре средств труда жестко один раз и навсегда, или в некотором мобильном устройстве памяти с возможностью его перепрограммирования (как в компьютерах, на гибких автоматизированных линиях и роботизированных комплексах).

Напротив информация стертая в средстве труда в процессе создания данного продукта (износ средства труда) переноситься на него и увеличивает его стоимость, хотя и не создает никакой потребительной стоимости. Но в процессе труда информация в средстве труда может не только стираться, но и накапливаться: это происходит, например, в интеллектуальных автоматизированных системах, как обучающихся с учителем, так и самообучающихся (поэтому их называют генераторами информации). В этом случае стоимость средств труда в процессе их использования не уменьшается, а возрастает, и стоимость продукта, созданного с их помощью соответственно уменьшается, а не увеличивается.

Итак, потребительная стоимость продукта труда определяется КАЧЕСТВОМ (смыслом, содержанием) связанной информации, записанной в физической форме и структуре того продукта непосредственно человеком или его средствами труда.

Абстрактная себестоимость продукта труда определяется алгебраической суммой КОЛИЧЕСТВА связанной информации, записанной в структуре физической формы продукта труда человеком и КОЛИЧЕСТВА связанной информации стертой или записанной в структуре физической формы средств труда в процессе производства данного продукта, причем последняя берется со знаком "+", если она стерта (износ средств труда), и со знаком "-", если она записана (генерация информации).

Производительность человеческого труда тем выше, чем большее количество функций тела человека передано его средствам труда, а также чем выше степень использования функциональных возможностей этих средств труда человеком. Чем выше производительность труда, тем большая доля информации записывается в продукте труда средствами труда автоматически, т.е. без участия человека. Таким образом, *в конечном счете производительность труда определяется прежде всего уровнем развития сознания человека.*

В отличие от производительности труда изменение его интенсивности не влияет на функциональный уровень технологии, а значит и на соотношение между количеством информации, записанной в продукт труда человеком и средствами труда. Поэтому только уменьшение рабочего времени, необходимого на производство данного продукта, достигнутое за счет увеличения производительности человеческого труда уменьшает абстрактную себестоимость этого продукта и может служить адекватной мерой изменения этой себестоимости. Так гени-

альные произведения искусства, содержащие колоссальную информацию, записанную в них непосредственно человеком-творцом практически без использования средств труда, всегда будут иметь наивысшую стоимость, значительно превосходящую стоимость самых качественных репродукций.

До сих пор мы использовали термин и понятие "время" без его специального анализа и определения в каком-то обыденно-экономическом значении. Теперь же основываясь на общности основных законов информационных взаимодействий проведем аналогию (а может быть и больше чем аналогию) между "временем физическим", "психофизиологическим" и "экономическим", естественно, насколько это возможно в рамках данной работы.

Из физики известно, что редукция виртуального объекта происходит при сообщении ему энергии, необходимой для образования его массы покоя. Очевидно, редуцируемый объект представляет собой канал взаимодействия классического и виртуального уровней Реальности и этот канал обеспечивает передачу энергии с первого уровня на второй. Однако для возникновения структуры редуцированной формы объекта одной энергии явно недостаточно: для этого необходима также и информация об этой структуре. Эта информация существовала еще до редукции на виртуальном уровне строения редуцируемого объекта и была передана по тому же каналу, но в направлении обратном энергетическому потоку.

Таким образом, в физике виртуальная сущность объекта выступает как источник информации, сам объект как информационно-энергетический канал взаимодействия виртуального и редуцированного уровней Реальности, а редуцированная форма объекта – как носитель информации, изменяющий свою структуру по мере записи соответствующей информации в структуре среды.

Чем выше уровень развития (сложность) объекта, тем более отдаленные друг от друга качественно различные уровни Реальности он соединяет как информационно-энергетический канал, тем выше пропускная способность (мощность) этого канала, тем большее разнообразие форм энергии и языковых форм представления информации он обеспечивает, и, наконец, *тем выше информационная емкость его формы*, т.е. тем большее количество информации может быть записано в структуре его формы до момента начала повышения ее энтропии.

Здесь уместным является пример с магнитофонной лентой на которую мы пытаемся записать как можно больше информации на единицу длины путем уменьшения скорости протяжки. Если при постоянном информационном потоке записи эту скорость уменьшать линейно, то первоначально плотность информации будет возрастать также практически линейно (а энтропия соответственно уменьшаться), однако скоро мы заметим, что плотность информации стала возрастать медленнее, т.к. возросли шумы (уменьшилось отношение сигнал/шум). Если продолжать и дальше уменьшать скорость протяжки, то конце концов это приведет к тому, что качественный записывающий сигнал будет восприниматься лентой практически как стирающий, т.е. на нее будет записываться один шум.

Таким образом, можно сделать по крайней мере следующие выводы:

Процесс труда можно рассматривать как процесс редукции образа продукта труда в структуре физической среды, что становится непосредственно очевидным при развитии интеллектуальных информационных технологий.

Человеческая Душа с ее неисчерпаемым творческим потенциалом является единственным источником всякой собственности и стоимости в этом мире. Поэтому даже в чисто экономическом, в общем-то достаточно "приземленном" смысле, нет ничего более ценного в мире, чем человеческая Душа.

Примечание: Согласно теории "Естественного права" (Сократ, Платон, Фома Аквинский) наиболее глубоким источником права является природа самого человека. Одним из основных правовых отношений является отношение собственности. На этом основании автор выдвигает (в качестве гипотезы, конечно) "Естественную теорию собственности":

1. Человеческая Душа является единственным и наиболее глубоким источником всех форм собственности и их фундаментом;

2. В зависимости от формы сознания человеческая душа отождествляет себя с различными "телами проявления" и, таким образом, возникает первая производная форма собственности: - собственность на свое тело и право на жизнь (при физической форме сознания – это собственность на физическое тело и физическую жизнь).

3. Из "Естественного права" собственности на свое тело возникает право собственности на все, что произведено непосредственно и исключительно с применением своего тела: прежде всего сам живой (собственный) труд, средства и продукты труда (физического и "умственного").

4. Право собственности на свой труд и средства труда приводят к праву собственности на продукты своего труда, произведенные с использованием собственных средств труда, а также к отсутствию права собственности на продукты труда, произведенные с использованием чужих средств труда (наемный труд).

Темп времени является величиной индивидуальной для каждого объекта и определяется мощностью информационно-энергетического канала, связывающего физическую форму объекта с его более глубокими структурными уровнями.

1.1.1.2.2. Информация, как сырье и как товар: абсолютная, относительная и аналитическая информация. Данные, информация, знания

Широко известны абсолютная и относительная формы информации. Абсолютная форма – это просто количество, частота. Относительная форма – это доли, проценты, относительные частоты и вероятности.

Менее знакомы специалисты с аналитической формой информации, примером которой является условные вероятности, стандартизированные статистические значения и **количество информации**.

Абсолютная информация – это информация содержащаяся в абсолютных числах, таких как количество чего-либо, взятого "само по себе", т.е. безотносительно к объему совокупности, к которой оно относится.

Относительная информация – это информация, содержащаяся в отношениях абсолютного количества к объему совокупности.

Относительная информация измеряется в частях, процентах, промиле, вероятностях и некоторых других подобных единицах. Очевидно, что и из относи-

тельной информации, взятой изолированно, вырванной из контекста, делать какие-либо обоснованные выводы не представляется возможным. Те, кто иногда делает это сознательно, просто вводит в заблуждение некомпетентных слушателей ("есть истина, есть заблуждение, а есть статистика").

Для того, чтобы о чем-то судить по процентам, нужен их *сопоставительный* анализ, т.е. анализ *всего* процентного распределения. Вариантов такого анализа может быть много, но суть не в этом, а в том, что такой анализ необходим. Рассмотрим один из возможных вариантов сопоставительного анализа процентных распределений на нашем примере. Этот вариант предполагает использование в качестве "базы оценки" среднего по всей совокупности (нормативный подход: норма – среднее).

Аналитическая (сопоставительная) информация – это информация, содержащаяся в *отношении* вероятности (или процента) к некоторой *базовой* величине, например к средней вероятности по всей выборке.

Аналитическими являются также стандартизированные величины в статистике и количество информации в теории информации.

Очевидно, именно аналитическая информация является наиболее **кондиционной** для употребления с той точки зрения, что позволяет **непосредственно** делать содержательные выводы об исследуемой предметной области (точнее будет сказать, что она сама и является выводом), тогда как для того, чтобы сделать аналогичные выводы на основе относительной, и особенно абсолютной информации требуется ее значительная предварительная обработка. Эта "предварительная обработка" и составляет значительную долю трудоемкости труда аналитиков и экспертов, которые полагаются во многом на чисто качественную (невербализуемую, интуитивную) оценку имеющейся у них сырой относительной информации, однако проводить необходимые для этого расчеты для реальных объемов данных вручную не представляется возможным.

Таким образом, есть все основания рассматривать абсолютную информацию как "**информационное сырье**", аналитическую – как "**информационный товар**". Относительная информация в этом смысле занимает промежуточное положение и может рассматриваться как "**информационный полуфабрикат**". Интеллектуальные информационные системы, преобразуют сырую информацию в кондиционный информационный продукт и, этим самым, многократно повышают ее потребительскую и меновую стоимость.

Для экономических исследований является естественным манипулировать понятиями "прибыль" – "убыток" или сходными понятиями теории игр: "выигрыш" – "проигрыш", измеряя их при этом, как правило, в денежных единицах.

Однако, на этом пути возникает ряд проблем:

1. Проблема выбора денежных единиц или сопоставимого во времени и пространстве способа их измерения.

2. Принципиальная проблема, состоящая в том, что не все явления, даже в экономике, уместно и целесообразно количественно оценивать (измерять) в денежных единицах.

В данной работе автор предлагает не решать эти проблемы, а обойти их, *выбрав в качестве количественной меры не "стоимость", а то, что лежит в основе стоимости*. Для этого предлагается раскрыть один из аспектов сущно-

сти понятия "стоимость" на основе применения аналитического понятия "информация", которое удовлетворяет всем сформулированным выше требованиям к количественной мере.

В эпоху господства капитала полагали, что "чистым товаром" является золото. Однако, если проанализировать те свойства золота, которые превратили его в основной эквивалент стоимости, то с очевидностью обнаружится, что это именно те свойства, из-за которых золото является идеальным носителем информации (информация легко "записывается" в золото, т.к. оно достаточно мягкое; информация долго сохраняется в золоте, т.к. оно не ржавеет, не является хрупким, не подвержено другим формам "порчи", т.е. стирания). Известно, что "стоимость" является не физическим свойством той или иной вещи, стоимость - это свойство вещи, которое приписывают ей люди, которые вступают посредством нее в определенные экономические отношения друг с другом. В действительности же все отношения людей друг с другом являются информационными. Золото, как эквивалент стоимости, также является лишь "информационным пакетом", несущим определенное количество информации, соответствующее его весу. Исходя из всех этих рассуждений и учитывая информационную теорию стоимости, основные положения которой были сформулированы выше, являясь свидетелями информационного общества, победившего в развитых странах, мы можем предположить, что единственным "чистым товаром" является (а по существу всегда и являлась) только информация.

Информация является квинтэссенцией стоимости.

Кратко рассмотрим вопрос о стоимости самой информации, имеющий самое непосредственное отношение к проблеме оценки экономической эффективности применения систем искусственного интеллекта и интеллектуальной обработки данных.

Мы знаем, как информация, в качестве рекламы, приносит деньги. Мы знаем также, что технология, "Ноу-хау", стоит значительно дороже, чем продукты ее применения.

Информация, как и нефть, может быть "сырой" или обработанной. Сырая информация может почти ничего не стоить по сравнению со стоимостью аналитически обработанной информации, т.е. информации, которая находится в форме пригодной для употребления, готовой немедленно принести громадные преимущества ее обладателю по сравнению с другими людьми, ею не обладающими.

1.1.1.2.3. Стоимость и амортизация систем искусственного интеллекта и баз знаний

Любая программная система представляет собой виртуальное средство труда работающее на базе универсального компьютера. Эти информационные средства труда, так же как и обычные "физические", могут быть предназначены либо для непосредственного потребления пользователем (группа "Б"), или для создания других подобных средств труда ("группа А").

Вопрос о стоимости программных систем – это вопрос о стоимости средств труда, начисто лишенных своего "физического тела". В их создание

вложен огромный высококвалифицированный труд, наукоемкие технологии, но тиражируются такие средства очень просто: путем перезаписи на магнитный носитель. Это своего рода "психосинтез", и если бы нечто аналогичное стало возможным с физическими объектами, то наступил бы настоящий "золотой век", по крайней мере в плане материальном (гибкие роботизированные комплексы уже приближают их к этому).

Следовательно, в соответствии с информационной теорией стоимости программные продукты имеют высокую потребительскую стоимость и практически никакой меновой стоимости.

Меновая стоимость программных продуктов определяется практически затратами на поиск места, где они уже есть, и на доставку потребителю. С появлением Internet практически решены вопросы и поиска, и доставки программного обеспечения и других "информационных товаров".

Чтобы повысить меновую стоимость программных систем их разработчики стараются затруднить их так называемое "свободное тиражирование" или попросту говоря – воровство, вводя необходимость инсталляции и "привязывая" систему к конкретному компьютеру (имеются также другие способы). Но, во-первых, против этого также есть свои средства, а во-вторых, такая привязка снижает функциональную ценность программной системы, т.к. делает ее более уязвимой при технических авариях, модернизации компьютера и т.п., что в общем делает ее просто менее удобной для пользователя.

Что касается износа программных систем, этих "нематериальных активов", то физический износ у них вообще отсутствует, а моральный может быть весьма значительным: после появления новой версии программного продукта старой уже никто не хочет пользоваться (хотя иногда появляется "сырая" и "не очень работающая" новая версия, а старая хорошо отработана и идеально выполняет свои функции). Итак, с появлением новой версии старая может "в один момент" потерять всю свою стоимость (в том числе и балансовую – в результате переоценки).

Иначе обстоит дело с базами данных и интеллектуальными системами, которые *накапливают и структурируют* информацию, обрабатывают ее по более или менее сложным алгоритмам, в результате чего *их стоимость непрерывно возрастает*. Стоимость баз данных, находящихся на банковском сервере, может в десятки миллионов раз превышать стоимость самого компьютера и в процессе работы это соотношение все больше увеличивается в пользу информации. Естественно, такую ценную информацию необходимо защищать, чтобы даже если сервер будет похищен, злоумышленники не смогли извлечь из него ни одно бита интересующей их информации.

1.1.1.2.4. Источники экономической эффективности систем искусственного интеллекта и интеллектуальной обработки данных с позиций информационной теории стоимости (повышение уровня системности и "охлаждение" объекта управления)

Информация, содержащаяся в системе, непосредственно связана с энтропией этой системы (обратно пропорционально), а та, в свою очередь, – с количеством энергии в системе (пропорционально). Поэтому *информация имеет энергетический эквивалент*, т.е. в принципе возможна формула, связывающая количество информации с количеством энергии, наподобие знаменитой формулы Альберта Эйнштейна $E=mc^2$, однако современной науке она неизвестна.

При сообщении некоторой системе определенного количества информации ее уровень системной организации возрастает и энтропия уменьшается, что приводит к выделению или экономии энергии (охлаждению системы).

Например, при внедрении системы оперативного управления процессом уборки зерновых в масштабах одного района Краснодарского края в 1983-1988 годах за счет *повышения уровня системной организации объекта управления* экономилось топлива на сумму около 400 тысяч рублей. Автором данной работы эти мысли высказывались в предложенной им информационной теории стоимости еще в 1979 году.

1.1.1.3. Интеллектуализация – генеральное направление и развития информационных технологий

1.1.1.3.1. От электронных вычислительных машин к компьютерам. Функциональное определение компьютера

На первых этапах развития информационных технологий считалось, что компьютеры предназначены в основном для проведения математических расчетов.

Это нашло отражение и в термине, первоначально введенному для обозначения компьютеров: "Электронная вычислительная машина" (ЭВМ). Этот термин включает общее понятие: "Машина", а также два специфических признака: "Электронная" и "Вычислительная". Первый из этих признаков является характеристикой элементной базы, как бы сейчас сказали Hard Ware, а второй является функциональным.

Давать определения средств труда, основываясь на их структуре, методологически неправильно с точки зрения информационно-функциональной теории развития техники, согласно которой определения должны быть функциональными.

Сегодня ясно, что у компьютеров может быть неэлектронная основа, и что их функции далеко выходят за пределы собственно проведения вычислений. По-

этому более правильным, по мнению автора, было бы такое определение компьютера: *"Компьютер – это средство труда для обработки информации (т.е. средство труда 5-го функционального уровня)"*.

1.1.1.3.2. Эволюция понятия: "Обработка информации" от информационного сырья к информационному продукту

Само понятие: "Обработка информации" также эволюционирует. Это происходит за счет изменения его содержания, которое постепенно меняется и сегодня включает в себя не только выполнение операций с уже *имеющейся* информацией, но также и очень важные операции *поиска* информации и *обеспечения удаленного доступа к информации*, в частности с применением средств глобальных компьютерных коммуникаций (например, Internet). Причем роль эффективного поиска и быстрого доступа к информации все больше и больше возрастает. Очевидно, что ценность и эффективность использования информации определяется прежде всего возможностью ее найти и получить к ней качественный доступ.

Операции по обработке уже имеющейся информации также становятся все более развитыми: они включают уже не только простейшие арифметические операции, но и операции базам данных (поиск, выборка, отбор, фильтрация, сортировка), статистические расчеты, численные методы, имитационное моделирование, методы распознавания образов и принятия решений, теории игр, экспертные системы, нейронные сети, генетические алгоритмы, когнитивное моделирование и др.

Развитие автоматизированных технологий обработки информации идет по пути все более глубокой переработки "информационного сырья", в результате чего "информационный продукт" все в большей степени отличается от исходного состояния информации.

Если на первых этапах эта переработка заключалась в основном в сортировке и выполнении арифметических операций, прежде всего суммирования, то в последующем она становится все более и более интеллектуальной. Все большее значение приобретают системы с элементами искусственного интеллекта: это и системы ввода текстов со сканера, и системы автоматизированного перевода, трехмерные графические системы.

1.1.1.3.3. Эволюция технологий создания и поддержки информационных систем: автоматизация функций посредников

С развитием информационных технологий радикально изменяется и подход к созданию программных систем. На начальных этапах ограниченная группа разработчиков, сосредоточенная в одном месте, проектировала и создавала программную систему *в целом*, после чего она внедрялась в готовом виде. В последующем могли создаваться новые версии системы, которые опять же внедрялись в готовом виде.

На пути создания таких систем возникли следующие основные сложно преодолимые трудности:

- систему нужно было создать очень быстро, т.к. при очень большом ее масштабе обычно система морально устаревала еще до начала ее внедрения;
- для создания таких систем необходимы огромный интеллектуальный потенциал и очень высокая скорость разработки, высокое качество сопровождения.

Опыт показывает, что *максимальная сложность систем, создаваемых централизованно по единому проекту, ограничена.*

Альтернативой централизованному проектированию и внедрению программных систем является *идеология открытых систем*, при которой над развитием системы одновременно и независимо друг от друга работают очень многие, иногда тысячи разработчиков, не придерживающихся какого-либо одного плана, но подчиняющихся общим некоторым общим правилам.

1.1.1.3.4. Перспективы информационных технологий: интеллектуализация, создание самообучающихся, саморазвивающихся (эволюционирующих) и самовоспроизводящихся систем

Сегодня мы все чаще видим как вместе или по отдельности реализуются и другие подходы:

1. Адаптивные и самообучающиеся системы настраиваются на решение тех или иных задач за счет учета априорной информации и информации, поступающей в систему в процессе ее эксплуатации. Можно считать, что подобные системы развиваются на основе опыта их эксплуатации и что *усвоение этого опыта есть один из технологических этапов создания таких систем.*

2. Саморазвивающиеся открытые системы очень большого масштаба, которые не спроектированы какой-либо одной группой разработчиков и развиваются не по какому-либо плану, созданному кем-либо заранее. Эти системы создают как бы информационную среду общего доступа в развитие которой могут вносить свой вклад разработчики и даже пользователи, независимо от своего места нахождения.

Таким образом перспектива информационных систем в создании технологий их создания и развития с учетом опыта эксплуатации, часто без заранее разработанного плана и в этом процессе могут принимать участие не только разработчики, но и пользователи, находящиеся где угодно в мире.

Такие системы будем называть открытыми распределенными адаптивными саморазвивающимися системами.

Контрольные вопросы

1. Основные положения информационно-функциональной теории развития техники.
2. Процессы труда и познания, как информационные процессы снятия неопределенности.
3. Организм человека и средства труда как информационные системы.
4. Законы развития техники.

5. Детерминация формы сознания человека функциональным уровнем средств труда.
6. Неизбежность возникновения компьютеров, информационных систем и систем искусственного интеллекта.
7. Информационная теория стоимости.
8. Связь количества и качества информации с меновой и потребительной стоимостью.
9. Информация, как сырье и как товар: абсолютная, относительная и аналитическая информация. Данные, информация, знания.
10. Стоимость и амортизация систем искусственного интеллекта и баз знаний.
11. Источники экономической эффективности систем искусственного интеллекта и интеллектуальной обработки данных с позиций информационной теории стоимости (повышение уровня системности и "охлаждение" объекта управления).
12. Интеллектуализация - генеральное направление и развития информационных технологий.
13. От электронных вычислительных машин к компьютерам. Функциональное определение компьютера.
14. Эволюция понятия: "Обработка информации" от информационного сырья к информационному продукту.
15. Эволюция технологий создания и поддержки информационных систем: автоматизация функций посредников.
16. Перспективы информационных технологий: интеллектуализация, создание самообучающихся, саморазвивающихся (эволюционирующих) и самовоспроизводящихся систем.

Рекомендуемая литература

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

1.1.2. ЛЕКЦИЯ-2.

Определение и критерии идентификации систем искусственного интеллекта. Место СИИ в структуре обработки информации и управления

Учебные вопросы

1. Данные, информация, знания. Системно-когнитивный анализ как развитие концепции смысла Шенка-Абельсона.
2. Понятие: "Система искусственного интеллекта", место СИИ в классификации информационных систем.
3. Определение и классификация систем искусственного интеллекта, цели и пути их создания.
4. Информационная модель деятельности специалиста и место систем искусственного интеллекта в этой деятельности.
5. Жизненный цикл системы искусственного интеллекта и критерии перехода между этапами этого цикла.

1.1.2.1. Данные, информация, знания. Системно-когнитивный анализ как развитие концепции смысла Шенка-Абельсона. Место СИИ в иерархической системе обработки информации

1.1.2.1.1. Постановка проблемы

При создании систем искусственного интеллекта разработчики оперируют такими основополагающими понятиями, как:

- данные, информация, знания;
- факт, смысл, мысль;
- мониторинг, анализ и управление.

От того, какое конкретное содержание вкладывается разработчиками в данные понятия, самым существенным образом зависят и подходы к созданию математических моделей, структур данных и алгоритмов функционирования СИИ.

Проблема состоит в том, что смысловое содержание этих понятий чаще всего не конкретизируется.

И это не случайно. Одной из основных причин этого положения дел, на наш взгляд, является то, что конкретизировать смысловое содержание данных понятий представляется возможным лишь на основе интуитивно-ясной и хорошо обоснованной концепции смысла.

Однако, как это ни удивительно и парадоксально, но реальные разработчики СИИ, обычно являющиеся математиками и программистами, чаще всего недостаточно знакомы с подобными концепциями.

Конечно, возникает вопрос о том, насколько вообще возможны, т.е. имеют смысл концепции смысла, не бессмысленны ли они? Может быть вопрос: "Какой

смысл имеют концепции смысла?" – является одним из вариантов логического парадокса Рассела? Хотя эти вопросы имеют "несерьезный" оттенок, по сути, они сводятся к очень серьезному вопросу о том, насколько или в какой степени интеллект может познать сам себя, т.е. о том, является интеллектуальная форма познания адекватным инструментом для познания интеллекта? В более общем теоретическом плане этот вопрос может быть сформулирован и так: "Может ли часть системы адекватно отразить (отобразить) систему в целом?", или, другими словами, "Может ли система в целом в определенном смысле включать себя как составную часть?"

От ответа на эти вопросы самым непосредственным образом зависит и ответ на ключевой вопрос о том, может ли человек создать искусственный интеллект по своему образу и подобию.

Мы отвечаем на эти вопросы утвердительно. Более того, системы, содержащие информацию о системе в целом в каждой своей части определенного уровня структурной иерархии, широко известны, это:

- биологические системы в каждой клетке которых (кроме половых) содержится полный геном;
- фрактальные системы;
- высокоорганизованные системы с большой взаимной информацией в своих частях, успешно противостоящие закону возрастания энтропии.

Выдающийся немецкий философ Георг Вильгельм Фридрих Гегель называл такие системы "Истинно бесконечными".

1.1.2.1.2. Традиционные пути решения проблемы

1.1.2.1.2.1. Классификация и характеристика концепций смысла

Наиболее полный и всесторонний на данный момент обзор различных подходов к автоматизации процессов понимания смысла дан в основополагающей работе Л.Г. Васильева [13].

Все теории понимания смысла классифицированы им на три группы: объектные; субъект-объектные; субъектные.

Объектные теории основаны на структурно-семантическом анализе, а понимание онтологизируется, т.е. приравнивается к самому объекту. Понять означает установить значение языкового знака.

Субъект-объектные теории синтактико-семантический анализ дополняют прагматическим описанием или описанием реальных психологических процессов при создании и/или анализе языкового знака.

Субъектные теории помимо собственно процедур понимания акцентируют внимание на оценке реципиентом **результатов** понимания (при помощи обратной связи в диалоге, рефлексии или монологе).

1.1.2.1.2.2. Суть концепции смысла Шенка-Абельсона

Согласно Л.Г. Васильеву [13] суть концепции смысла Шенка-Абельсона состоит в том, что факты рассматриваются как **причины** и их смысл считается известным, если известны **последствия** данного факта. Таким образом, понимание смысла определенных конкретных событий заключается в выявлении причинно-следственных взаимосвязей между этими событиями и другими.

По нашему мнению, данная концепция смысла является одной из наиболее интуитивно убедительных и хорошо обоснованных, поэтому она в целом принята нами за основу.

Естественно, в этой концепции одним из ключевых моментов является определение способа выявления силы и направленности влияния причинно-следственных взаимосвязей и их количественной оценки (меры).

На наш взгляд *слабым местом* концепции смысла Шенка-Абельсона является сложность корректного и обоснованного выбора количественной *меры* силы и направленности причинно-следственных связей, а также конкретного *способа* определения численной величины этой меры в каждом конкретном случае (т.е. для каждого факта), причем непосредственно на основе *эмпирических* данных.

Проблема в том, что в общественном сознании продолжает господствовать упрощенческая точка зрения, состоящая в том, что *корреляция является мерой причинно-следственных связей*. И это имеет место не смотря на многочисленные разъяснения в специальной литературе о том, что это не так, точнее не совсем так.

Поэтому одной из целей данной работы будет обоснована другой меры силы и направленности причинно-следственных взаимодействий.

1.1.2.1.3. Идея решения проблемы

1.1.2.1.3.1. Формулировка идеи

Предлагаемая идея создания искусственного интеллекта очень проста и состоит в том, что для этого предлагается:

- во-первых, выявить основные моменты играющие существенную роль при создании *естественного* интеллекта;
- во-вторых, попробовать реализовать эти моменты на базе современных компьютерных технологий.

Наблюдения за системами естественного интеллекта позволяют сформулировать следующую *гипотезу*.

1. *Естественный* интеллект реально существует.
2. *Естественный* интеллект создается не мгновенно, а в течение довольно длительного времени по вполне определенной сложной технологии, которая включает три основных этапа:

- создание материальной системы поддержки *естественного* интеллекта по сложной технологии в *изолированных* от среды условиях;
- создание активной информационной структуры, базирующейся на материальной системе поддержки, способной к развитию и саморазвитию в систему *естественного* интеллекта, т.е. создание системы потенциального естественного интеллекта (оболочки, инструментальной системы);
- формирование структуры и функций *естественного* интеллекта во *взаимодействии* системы его поддержки с другими подобными системами и с окружающей средой, как с природной, так и с "социальной", т.е. созданной другими подобными системами, в результате чего происходит трансформация системы *потенциального* естественного интеллекта в систему *реального* естественного интеллекта.

3. Системы *искусственного* интеллекта (СИИ) полностью функционально эквивалентные *естественному* интеллекту могут быть созданы на базе *другой* материальной структуры системы поддержки системы и *другой* системы потенциального искусственного интеллекта.

4. Создание СИИ должно включать три этапа:

- создание материальной системы поддержки (эта проблема в основном решена, т.к. СИИ могут создаваться даже на базе современных персональных компьютеров);

- создание системы потенциального искусственного интеллекта, т.е. программной оболочки, инструментальной системы (таких систем в настоящее время существует пока еще очень мало);

- обучение и самообучение системы потенциального искусственного интеллекта и преобразование ее в реальную СИИ.

5. Основополагающую роль в создании системы потенциального искусственного интеллекта играет разработка научной концепции и теории, адекватно отражающей способы реализации функций естественного интеллекта и пути его трансформации из потенциального в реальный.

1.1.2.1.3.2. Обоснование принципиальной возможности реализации идеи

Относительная независимость функций от поддерживающих их структур

При рассмотрении самой идеи реализации естественного интеллекта на иной материальной основе сразу возникает вопрос о том, на сколько это вообще в принципе возможно, т.е. вопрос о том, могут ли интеллектуальные *функции* быть реализованы на основе иной, чем мозг, материальной *структуры*. Это вопрос о соотношении функций и поддерживающих их структур.

Конечно, структура зависит от функций, которые она поддерживает, и в целом более сложные функции поддерживаются более сложными структурами. Однако к счастью эта связь "структура – функция" не однозначна, т.е. *одни и те же функции могут поддерживаться самыми разнообразными, т.е. альтернативными структурами*. Так, например, согласно современным представлениям (В.А. Драгавцев) определенные фенотипические признаки могут поддерживаться различными генными ансамблями, а не только конкретными генами, как в теории Менделя.

Если ген уподобить букве алфавита, а смысл фразы – фенотипическому признаку, то можно сказать, что возможно очень большое количество фраз с одним и тем же смысловым содержанием (тогда как в классической генетике считалось, что признак соответствует гену, хотя есть и такие). После расшифровки генома человека мы настолько же приблизились к его пониманию, как изучивший русскую или немецкую азбуку англичанин, не знающий этих языков, приблизился к чтению в оригинале и пониманию содержания "Войны и Мира" Льва Толстого или "Феноменологии Духа" Георга В.Ф.Гегеля.

При этом обычно функциональная универсальность достигается за счет определенной структурной избыточности, что неизбежно приводит к уменьшению эффективности поддержки конкретных функций, по сравнению с их специ-

ализированной реализацией. На этих закономерностях самым существенным образом основано все развитие нашей технократической цивилизации, т.к. прогресс техники представляет собой ни что иное, как процесс последовательной передачи трудовых функций человека средствам труда.

И системы искусственного интеллекта, с этой точки зрения, представляют собой совершенно закономерный, более того – даже неизбежный, этап развития информационных средств труда, т.е. современных информационных технологий.

Но функция (или определенное качество) вполне корректно рассматривать и как *свойство* соответствующей поддерживающей его *структуры*. Поэтому, для того, чтобы сформировать более углубленное представление о соотношении структуры и поддерживаемых ими функций кратко рассмотрим диалектику "Структура – свойство – отношение".

Диалектика: "Структура – свойство – отношение"

Рассмотрим простой пример. Два электрона определенным образом взаимодействуют друг с другом, находясь в определенных **отношениях**, а именно – отталкиваясь друг от друга с различной силой, зависящей от расстояния между ними. Но о каждом электроне можно сказать, что он обладает **свойством** отталкиваться от другого электрона. Чем поддерживается (обеспечивается) это свойство электронов, благодаря которому они могут быть в определенных **отношениях** друг с другом?

В науке настоящего времени считается, что существует соответствующая материальная **структура**, которая называется "электрическое поле" и является одной из форм электромагнитного поля. Возникает вопрос - а может ли эта структура существовать **независимо** от электрона?

Уже давно Герцем и Максвеллом получен однозначный положительный ответ на этот вопрос: "Да, может, и это – электромагнитные волны!!!"

*Таким образом, **свойство** может существовать как некая материальная **структура** отдельно и независимо от объекта, свойством которого оно являлось и благодаря которому этот объект вступал в определенные **отношения** с другими объектами, обладавшими тем же свойством.*

Так нам светят звезды, которых, давно уже нет. Мы восхищаемся произведениями художников, поэтов и музыкантов давно прошедших времен. Поэтому шутка Л.Кэррола об улыбке Чеширского кота, висевшей в воздухе, хотя сам кот и ушел, не так уж и далека от истины. Ряд исследований о квантовой природе сознания [27, 264] свидетельствует в пользу того, что определенные формы сознания может существовать независимо от физического организма.

Итак, *существуют различные точки зрения на одно и то же, которое, при различных условиях, может рассматриваться либо как определенные **отношения** объектов, либо как **свойства** этих объектов, на которых основаны их отношения, либо как самостоятельно существующая материальная **структура**.*

Каждая точка зрения имеет право на существование, но, по-видимому, каждая последующая из этих точек зрения является более глубокой, чем предыдущая.

1.1.2.1.4. Когнитивная концепция СК-анализа и синтез когнитивного конфигуратора

В качестве теории, позволяющей конкретизировать основополагающие понятия, связанные с СИИ, предлагается системно-когнитивный анализ (СК-анализ), представляющий собой системный анализ, как метод познания, т.е. структурированный по базовым когнитивным операциям.

Когнитивная концепция СК-анализа разработана с учетом двух основных требований:

1. Адекватное отражение в когнитивной концепции реальных процессов, реализуемых человеком в процессах познания.
2. Высокая степень приспособленности когнитивной концепции для формализации в виде достаточно простых математических и алгоритмических моделей, допускающих прозрачную программную реализацию в автоматизированной системе.

1.1.2.1.4.1. Понятие когнитивного конфигуратора и необходимость естественнонаучной (формализуемой) когнитивной концепции **Определение понятия конфигуратора**

Понятие конфигуратора, по-видимому, впервые предложено В.А. Лефевром [194], хотя безусловно это понятие использовалось и раньше, но, во-первых, оно не получало самостоятельного названия, а, во-вторых, использовалось в частных случаях и не получало обобщения. *Под конфигуратором В.А.Лефевр понимал минимальный полный набор понятийных шкал или структур, достаточный для адекватного описания предметной области.* Примеры конфигураторов приведены в [194].

Понятие когнитивного конфигуратора

В исследованиях по когнитивной психологии изучается значительное количество различных операций, связанных с процессом познания [81]. Однако, насколько известно из литературы, психологами не ставился вопрос о выделении из всего множества когнитивных операций такого минимального (базового) набора наиболее элементарных из них, из которых как составные могли бы строиться другие операции. Ясно, что для выделения таких базовых когнитивных операций (БКО) необходимо построить их иерархическую систему, в фундаменте которой будут находиться наиболее элементарные из них, на втором уровне – производные от них, обладающие более высоким уровнем интегративности, и т.д.

Таким образом, *под когнитивным конфигуратором будем понимать минимальный полный набор базовых когнитивных операций, достаточный для представления различных процессов познания.*

Когнитивные концепции и операции

Проведенный анализ когнитивных концепций показал, что они разрабатывались ведущими психологами (Пиаже, Солсо, Найсер) [81] без учета требований, связанных с их дальнейшей формализацией и автоматизацией. Поэтому имеющиеся концепции когнитивной психологии слабо подходят для этой цели; в

когнитивной психологии не ставилась и не решалась задача конструирования когнитивного конфигуратора и, соответственно, не сформулировано понятие базовой когнитивной операции.

1.1.2.1.4.2. Базовая когнитивная концепция

Автоматизировать процесс познания в целом безусловно значительно сложнее, чем отдельные операции процесса познания. Но для этого прежде всего необходимо:

- выявить эти операции;
- найти место каждой из них в системе или последовательности процесса познания.

Сделать это предлагается в форме когнитивной концепции, которая должна удовлетворять следующим требованиям:

- адекватность, т.е. точное отражение сущности процессов познания, характерных для человека, в частности описание процессов вербализации, семантической адаптации и семантического синтеза (уточнения смысла слов и понятий и включения в словарь новых слов и понятий);
- высокая степень детализации и структурированности до уровня достаточно простых базовых когнитивных операций;
- возможность математического описания, формализации и автоматизации.

Однако приходится констатировать, что даже концепции когнитивной психологии, значительно более конкретные, чем философские (гносеологические), разрабатывались без учета необходимости построения реализующих их математических и алгоритмических моделей и программных систем. Более того, в когнитивной психологии из всего многообразия различных исследуемых когнитивных операций не выделены базовые, к суперпозиции и различным вариантам сочетаний которых сводятся различные процессы познания. В этой науке в настоящее время господствует эмпирический подход, т.е. имеется тенденция не к теоретическим обобщениям, а тщательному изучению отдельных когнитивных операций. Поэтому *для достижения целей данного исследования* концепции когнитивной психологии мало применимы.

В связи с этим в данном исследовании предлагается когнитивная концепция, удовлетворяющая сформулированным выше требованиям. Эта концепция достаточно проста, иначе было бы невозможно ее формализовать, многие ее положения интуитивно очевидны или хорошо известны, тем ни менее в целостном виде она впервые сформулирована лишь в работах [81, 125, 149]. Положения когнитивной концепции приведены в определенном порядке, соответствующем реальному ходу процесса познания "от конкретных эмпирических исходных данных к содержательным информационным моделям, а затем к их верификации, адаптации и, в случае необходимости, к пересинтезу".

На базе выше сформулированных положений предложена целостная система взглядов на процесс познания, т.е. **формализуемая когнитивная концепция** [81], предназначенная для построения систем искусственного интеллекта.

Суть предложенной когнитивной концепции

Процесс познания рассматривается нами как многоуровневая иерархическая система обработки информации, в которой каждый последующий уровень является результатом интеграции элементов предыдущего уровня. На 1-м уровне этой системы находятся дискретные элементы потока чувственного восприятия, которые на 2-м уровне интегрируются в чувственный образ конкретного объекта. Те, в свою очередь, на 3-м уровне интегрируются в обобщенные образы классов и факторов, образующие на 4-м уровне кластеры, а на 5-м конструкторы. Система конструкторов на 6-м уровне образуют текущую парадигму реальности (т.е. человек познает мир путем синтеза и применения конструкторов). На 7-м же уровне обнаруживается, что текущая парадигма не является единственно возможной.

Данные, информация, знания

Существует неопределенность смыслового содержания "разночтения" терминов: "данные", "информация", "знания". Мы считаем целесообразным определить их следующим образом.

Данные представляют собой информацию, рассматриваемую в чисто синтаксическом аспекте, т.е. безотносительно к ее содержанию и использованию, т.е. семантике и телеологии (обычно на каком-либо носителе или в канале передачи).

Информация – это данные, *проинтерпретированные* с использованием *тезауруса*, т.е. осмысленные данные, рассматриваемые в единстве синтаксического и семантического аспектов.

Знания, есть *система* информации, обеспечивающая увеличение вероятности достижения какой-либо *цели*, т.е. по сути знания – это "Ноу-хау" или технологии.

Вышесказанное резюмируем в следующей форме:

$$\begin{aligned} \text{знание} &= \text{информация} + \text{цель} \\ \text{информация} &= \text{данные} + \text{смысл}; \\ \text{знания} &= \text{данные} + \text{смысл} + \text{цель}. \end{aligned}$$

При разработке предложенной интерпретации содержания данных понятий и формы его представления учтены и использованы разработки Лаптева В.Н., а также Владимира Кива, приведенные на его сайте: <http://vlak.webzone.ru/rus/it/knowledge.html>.

В этой связи необходимо отметить, что название теории информации Шеннона не вполне соответствует предлагаемому пониманию содержания понятий "данные – информация – знания": теорию Шеннона по сути дела более точно было бы называть "теорией передачи данных по каналам связи". Общеизвестная количественная мера Шеннона "для измерения количества информации" также в

этом смысле является лишь мерой количества данных, т.к. не содержит семантического аспекта.

Количественная теория автоматизированной интерпретации данных, основанная на теории информации, в настоящее время лишь только создается и получила название семантической теории информации. Предложены различные количественные меры для измерения смысла информации [81]. Из этих мер выделяется мера А.Харкевича, в основу которой положено понятие цели.

В литературе и самим А.Харкевичем это не отмечается, но, учитывая вышеизложенное, можно утверждать, что по сути дела им в 1962 году впервые предложена научно-обоснованная количественная мера знаний.

Однако предложенная А.Харкевичем мера не удовлетворяет *принципу соответствия*, обязательному для более общей теории, аналогично тому, как, например, мера Шеннона сводится к мере Хартли в случае равновероятных событий. Поэтому теория А.Харкевича как бы "стояла *особняком*" от теории информации Найквиста-Хартли-Больцмана и Шеннона и в этом плане она нуждалась в развитии, что и было осуществлено автором [1, 76, 78, 81, 87, 97, 98, 100, 103, 106, 107, 120, 122, 124, 126].

Известно, что *корреляция не является мерой причинно-следственных связей*. Если корреляция между действием некоторого фактора и переходом объекта управления в определенное состояние высока, то это еще не значит, что данный фактор является причиной этого перехода. Для того чтобы по корреляции можно было судить о наличии причинно-следственной связи необходимо сравнить исследуемую группу с *контрольной группой*, т.е. с группой, в которой данный фактор не действовал.

Также и высокая вероятность перехода объекта управления в определенное состояние в условиях действия некоторого фактора сама по себе не говорит о наличии причинно-следственной связи между ними, т.е. о том, что данный фактор обусловил переход объекта в это состояние. Это связано с тем, что вероятность перехода объекта в это состояние может быть и сама по себе очень высокой независимо от действия фактора. Поэтому в качестве *меры силы причинной обусловленности* определенного состояния объекта действием некоторого фактора Харкевич предложил логарифм *отношения* вероятностей перехода в объекта в это состояние в условиях действия фактора и при его отсутствии или *в среднем* (13). Таким образом, вероятность перехода объектов в некоторое состояние в исследуемой группе сравнивается с средней вероятностью перехода в это состояние по всем группам и это среднее выступает в качестве контрольной группы.

Таким образом семантическая мера информации Харкевича является мерой наличия причинно-следственных связей между факторами и состояниями объекта управления.

В связи с тем, что подход Харкевича является ярким примером нормативного подхода, будет уместно вспомнить о нормативном и критериальном подходах к оценке (идентификации) состояния объекта.

При критериальном подходе состояние объекта оценивают по жесткой системе критериев. Правда при этом остается открытым вопрос о том, каким образом формируется эта система критериев.

При нормативном подходе, который получил свое название от понятия "норма", сначала формируется эта норма, а затем состояние объектов оценивается относительно этой нормы. В частности, если это сформированное понятие нормы зафиксировать, стандартизировать, как и критерии его определения и сравнения с ним, то мы получим один из вариантов синтеза критериального подхода.

Мониторинг, анализ, прогнозирование, управление

Существует определенная иерархия задач обработки данных, информации и знаний:

Мониторинг – накопление данных по ряду показателей об объекте управления с привязкой ко времени.

Анализ – выявление смысла в данных, т.е. выявление в них причинно-следственных взаимосвязей, выявление информации.

Прогнозирование – использование смысла причинно-следственных зависимостей в предметной области, т.е. использование информации для прогнозирования и предсказания поведения объекта управления в условиях действия определенных факторов.

Управление – использование информации для достижения определенных *целей* управления, т.е. выявление и использование знаний для:

- сохранения стабильного функционирования объекта управления;
- перевода объекта управления в заранее заданное целевое состояние.

Таким образом, управление – это высшая форма обработки и использования информации.

Факт, смысл, мысль

Ключевым для когнитивной концепции является понятие факта.

Под фактом будем понимать соответствие дискретного и интегрального элементов познания, т.е. элементов разных уровней интеграции-иерархии процессов познания, обнаруженное на опыте.

Факт рассматривается как квант смысла. Это является основой для формализации смысла.

Смысл представляет собой "разность потенциалов" между смежными уровнями интеграции-иерархии в системе обработки информации в процессах познания.

Мысль является операцией выявления смысла из фактов.

Мышление есть процесс, состоящий из ряда взаимосвязанных по смыслу мыслей.

Но существуют различные формы мыслей, которые перед разработкой методов формализации и программной реализации необходимо классифицировать и выявить среди них основные, т.е. такие, к которым сводятся все остальные или по крайней мере большинство из них. Как уже отмечалось выше, сделать это предлагается на основе базовой когнитивной концепции.

Иерархическая структура (пирамида) обработки информации

Сказанное в разделах 4.2.2–4.2.4 можно резюмировать в графической диаграмме, в форме "Мексиканской" (ступенчатой) пирамиды (рисунок 4).



Рисунок 4. Иерархическая структура обработки информации в АСОИУ

При этом аббревиатуры расшифровываются следующим образом:

- СУБД - системы управления базами данных;
- АСОД - автоматизированные системы обработки данных;
- АСОИ - автоматизированные системы обработки информации;
- АСУ - автоматизированные системы управления.

Каждый последующих вид систем включает все предыдущие в качестве подсистем: АСОД включает БД и СУБД (БД - база данных), т.е. по сути является банком данных, АСОИ основывается на АСОД, АСУ - на АСОИ. Система, в которой в качестве различных иерархических уровней представлены все эти виды систем с полным основанием может быть названа автоматизированной системой обработки информации и управления (АСОИУ).

Базовая когнитивная концепция в свободном изложении

Исходные данные для процесса познания поставляются из нескольких *независимых* информационных источников, имеющих *качественно* различную природу, которые мы будем условно называть "*органы чувств*". Данные от органов чувств также имеют *качественно* различную природу, обусловленную конкретным видом информационного источника. Для обозначения этих исходных данных будем использовать термин "*атрибут*". В результате выполнения когнитивной операции "присвоение имен" атрибутам могут быть присвоены уникальные имена, т.е. они могут быть отнесены к некоторым градациям номинальных шкал. Получение информации о предметной области в атрибутивной форме осуществляется когнитивной операцией "восприятие".

Исходные данные содержат внутренние закономерности, объединяющие качественно разнородные исходные данные от различных информационных источников.

После восприятия предметной области может быть проведен ее первичный анализ путем выполнения когнитивной операции: "сопоставление опыта, воплощенного в модели, с общественным", т.е. с результатами восприятия той же предметной области другими. Это делается с целью исключения из дальнейшего анализа всех наиболее явных расхождений, как сомнительных.

Однако, закономерности в предметной области могут быть выявлены путем выполнения когнитивной операции "обобщение" только после накопления в результате мониторинга достаточно большого объема исходных данных в *памяти*.

Наличие этих закономерностей позволяют предположить, что:

- существуют некие интегративные структуры, не сводящиеся ни к одному из качественно-различных аспектов исходных данных и обладающие по отношению к ним системными, т.е. эмерджентными свойствами, которые не могут быть предметом прямого восприятия с помощью органов чувств, но могут являться предметом для других форм познания, например логической формы. Для обозначения этих структур будем использовать термин *"объект"*;

- *"объекты"* считаются причинами существования взаимосвязей между атрибутами.

Объектам приписывается объективное существование, в том смысле, что любой объект обнаруживается несколькими независимыми друг от друга способами с помощью различных органов чувств (этот критерий объективного существования в физике называется *"принцип наблюдаемости"*).

После обобщения возможны когнитивные операции: "определение значимости шкал и градаций атрибутов" и "определение степени сформированности шкал и градаций классов".

Путем выполнения когнитивной операции "присвоение имен" конкретным объектам могут быть присвоены уникальные имена, т.е. они могут быть отнесены к некоторым грациям номинальной шкалы, которые мы будем называть *"классами"*. В данном случае класс представляет собой отображение объекта шкалу, т.е. это своего рода целостный образ объекта. После этого возможно выполнение когнитивной операции "идентификация" объектов, т.е. их *"узнавание"*: при этом по атрибутам объекта определяется класс, к которому принадлежит объект. *При этом все атрибуты, независимо от их качественно различной природы, рассматриваются с одной-единственной точки зрения: "Какое количество информации они несут о принадлежности данного объекта к каждому из классов"*.

Кроме того возможно выполнение когнитивной операции: "дедукция и абдукция, обратная задача идентификации и прогнозирования", имеющей очень важное значение для управления, т.е. вывод всех атрибутов в порядке убывания содержащегося в них количества информации о принадлежности к данному классу.

Аналогично, может быть выполнена когнитивная операция: "семантический анализ атрибута", представляющий собой список классов, в по-

рядке убывания количества информации о принадлежности к ним, содержащейся в данном атрибуте.

Таким образом возможно два взаимно-дополнительных способа отображения объекта: в форме принадлежности к некоторому классу (целостное, интегральное, экстенциональное); в форме системы атрибутов (дискретное, интенциональное).

Дальнейшее изучение атрибутов позволяет ввести понятия "*порядковая шкала*" и "*градация*". Порядковая шкала представляет собой способ классификации атрибутов одного качества, обычно по степени выраженности (интенсивности). Градация – это конкретное положение или диапазон на шкале, которому ставится в соответствие конкретный атрибут, соответствующее определенной степени интенсивности. Каждому виду атрибутов, информация о которых получается с помощью определенного "*органа чувств*", ставится в соответствие одна шкала. Таким образом, если при анализе в номинальных шкалах, можно было в принципе ввести одну шкалу для всех атрибутов, то в порядковых шкалах каждому атрибуту будет соответствовать своя шкала.

После идентификации уникальных объектов с классами возможна их классификация и присвоение *обобщающих имен* группам похожих классов. Для обозначения группы похожих классов используем понятие "*кластер*". Формирование кластеров осуществляется с помощью когнитивной операции "классификация". Кластер представляет собой своего рода "объект, состоящий из объектов", т.е. объект 2-го порядка. Если объект выполняет интегративную функцию по отношению к атрибутам, то кластер – по отношению к объектам.

Необходимо подчеркнуть, что термин "класс" используется не только для обозначения образов уникальных объектов, но и для обозначения их кластеров, т.е. *классу может соответствовать не уникальное, а обобщающее имя, в этом случае мы имеем дело с обобщенным классом*. Да и кластеры могут быть не только кластерами уникальных объектов, но и обобщенных классов.

Если объективное существование уникальных объектов мало у кого вызывает сомнение, то вопрос об объективном существовании интегративных структур 2-го и более высоких порядков остается открытым. В некоторых философских системах подобным объектам приписывался даже более высокий статус существования, чем самим объектам, например обычные объекты рассматриваются лишь как "тени" "Эйдосов" (Платон). Известны и другие понятия для обозначения объектов высоких порядков, например "архетип" (Юм), "эгрегор" (Андреев) и др. Нельзя не отметить, что в современной физике (специальной и общей теории относительности) есть подобное понятие пространственно-временного интервала, который проявляется как *движение* объекта. По-видимому, статус существования структур реальности, отражаемых когнитивными структурами тем выше, чем выше интегративный уровень этих структур.

Являясь объектами 2-го порядка сами кластеры в результате выполнения когнитивной операции "генерация конструкторов" могут быть классифицированы по степени сходства друг с другом. Для обозначения системы двух противоположных кластеров, с "спектром" промежуточных кластеров между ними, будем использовать термин "*бинарный конструктор*", при этом сами противоположные

кластеры будем называть "*полюса бинарного конструкта*". Таким образом конструкт представляет собой объект 3-го порядка.

Словом "бинарный" определяется, что в данном случае полюсов у конструкта всего два, но этим самым подчеркивается, что в принципе их может быть 3, 4 и больше. Бинарный конструкт можно формально представить в виде порядковой шкалы или даже шкалы отношений, на которой градациям соответствуют кластеры, а значит и сами классы и соответствующие объекты. Конструкты с количеством полюсов больше 2 могут быть представлены графически в форме семантических сетей в которых полюса являются вершинами, а дуги имеют цвет и толщину, соответствующие степени сходства-различия этих вершин. Семантические сети можно считать также просто графической формой представления результатов кластерного анализа.

Аналогично кластерам и конструктам классов формируются кластеры и конструкты атрибутов. В кластеры объединяются атрибуты, имеющие наиболее сходный смысл. В качестве полюсов конструктов выступают кластеры атрибутов, противоположных по смыслу.

Бинарные конструкты классов и атрибутов представляет собой когнитивные структуры, играющие огромную роль в процессах познания. Не будет преувеличением сказать, что познание представляет собой процесс генерации, совершенствования и применения конструктов. Будем считать, что конструкт тем более совершенен и тем выше его качество, чем сильнее отличаются его полюса, т.е. чем больше диапазон его области значений.

В кластерном анализе определялась степень сходства или различия классов, а не то, чем конкретно сходны или отличаются. При выполнении когнитивной операции "содержательное сравнение" двух классов определяется вклад каждого атрибута в их сходство или различие. Результаты содержательного сравнения выводятся в наглядной графической форме когнитивных диаграмм, в которых изображаются информационные портреты классов с наиболее характерными и нехарактерными для них атрибутами и атрибуты разных классов соединяются линиями, цвет и толщина которых соответствуют величине и знаку вклада этих атрибутов в сходство или различие данных классов.

Результаты идентификации и прогнозирования, осуществленные с помощью модели, путем выполнения когнитивной операции "верификация" сопоставляются с опытом, после чего определяется выполнять ли когнитивную операцию "обучение", состоящую в том, что параметры модели могут изменяться количественно, и тогда мы имеем дело с адаптацией, или качественно, и тогда идет речь о перестроении модели.

Базовая когнитивная концепция в формальном изложении

1. Процесс познания начинается с чувственного восприятия. Различные органы восприятия дают качественно-различную чувственную информацию в форме дискретного потока *элементов восприятия*. Эти элементы формализуются с помощью описательных шкал и градаций.

2. В процессе накопления опыта выявляются взаимосвязи между элементами чувственного восприятия: одни элементы часто наблюдаются с другими (имеет место их пространственно-временная корреляция), другие же вместе встречаются достаточно редко. Существование устойчивых связей между элементами восприятия говорит о том, что они отражают некую реальность, *интегральную* по отношению к этим элементам. Эту реальность будем называть объектами восприятия. Рассматриваемые в *единстве* с объектами элементы восприятия будем называть *признаками* объектов. Таким образом, органы восприятия дают чувственную информацию о признаках наблюдаемых объектов, процессов и явлений окружающего мира (объектов). Чувственный образ конкретного объекта представляет собой систему, возникающую как результат *процесса* синтеза признаков этого объекта. В условиях усложненного восприятия синтез чувственного образа объекта может быть существенно замедленным и даже не завершаться в реальном времени.

3. Человек присваивает конкретным объектам названия (имена) и сравнивает объекты друг с другом. При сравнении выясняется, что одни объекты в различных степенях сходны по их признакам, а другие отличаются. Сходные объекты объединяются в обобщенные категории (классы), которым присваиваются имена, производные от имен входящих в категорию конкретных объектов. Классы формализуются с помощью классификационных шкал и градаций и обеспечивают интегральный способ описания действительности. Путем обобщения (синтеза, индукции) информации о признаках конкретных объектов, входящих в те или иные классы, формируются обобщенные образы классов. Накопление опыта и сравнение обобщенных образов классов друг с другом позволяет определить степень характерности признаков для классов, смысл признаков и ценность каждого признака для идентификации конкретных объектов с классами и сравнения классов, а также исключить наименее ценные признаки из дальнейшего анализа без существенного сокращения количества полезной информации о предметной области (абстрагирование). Абстрагирование позволяет существенно сократить затраты внутренних ресурсов системы на анализ информации. Идентификация представляет собой процесс узнавания, т.е. установление соответствия между чувственным описанием объекта, как совокупности дискретных признаков, и неделимым (целостным) именем класса, которое ассоциируется с местом и ролью воспринимаемого объекта в природе и обществе. Дискретное и целостное восприятие действительности поддерживаются, как правило, различными полушариями мозга: соответственно, правым и левым (доминантность полушарий). Таким образом, именно *системное* взаимодействие интегрального (целостного) и дискретного способов восприятия обеспечивает возможность установление *содержательного смысла* событий. При выполнении когнитивной операции "содержательное сравнение" двух классов определяется вклад каждого признака в их сходство или различие.

4. После идентификации уникальных объектов с классами возможна их классификация и присвоение *обобщающих имен* группам похожих классов. Для

обозначения группы похожих классов используем понятие "*кластер*". Но и сами кластеры в результате выполнения когнитивной операции "генерация конструктов" могут быть классифицированы по степени сходства друг с другом. Для обозначения системы двух противоположных кластеров, со "спектром" промежуточных кластеров между ними, будем использовать термин "*бинарный конструкт*", при этом сами противоположные кластеры будем называть "*полюса бинарного конструкта*". Бинарные конструкты классов и атрибутов, т.е. конструкты с двумя полюсами, наиболее типичны для человека и представляет собой когнитивные структуры, играющие огромную роль в процессах познания. Достаточно сказать, что *познание можно рассматривать как процесс генерации, совершенствования и применения конструктов*. Качество конструкта тем выше, чем сильнее отличаются его полюса, т.е. чем больше диапазон его смысла.

Результаты идентификации и прогнозирования, осуществленные с помощью модели, путем выполнения когнитивной операции "верификация" сопоставляются с опытом, после чего определяется целесообразность выполнения когнитивной операции "обучение". При этом может возникнуть три основных варианта:

1. Объект, входит в обучающую выборку и достоверно идентифицируется (внутренняя валидность, в адаптации нет необходимости).
2. Объект, не входит в обучающую выборку, но входит в исходную генеральную совокупность, по отношению к которой эта выборка репрезентативна, и достоверно идентифицируется (внешняя валидность, добавление объекта к обучающей выборке и адаптация модели приводит к количественному уточнению смысла признаков и образов классов).
3. Объект не входит в исходную генеральную совокупность и идентифицируется недостоверно (внешняя валидность, добавление объекта к обучающей выборке и синтез модели приводит к качественному уточнению смысла признаков и образов классов, исходная генеральная совокупность расширяется).

1.1.2.1.4.3. Когнитивный конфигуратор и базовые когнитивные операции системного анализа

Таким образом, из предложенной когнитивной концепции вытекает существование, по крайней мере, 10 базовых когнитивных операций системного анализа (БКОСА) (таблица 1):

**Таблица 1 – ОБОБЩЕННЫЙ СПИСОК БКОСА
(КОГНИТИВНЫЙ КОНФИГУРАТОР)**

№ п/п	Код БКОСА по схеме АСК-анализа	Полное наименование БКОСА
1.	1.	Присвоение имен классам и атрибутам (интенциональная и экстенциональная репрезентация)
2.	2.	Восприятие
3.	3.	Обобщение (синтез, индукция)
4.	4.	Абстрагирование классов и атрибутов
5.	5.	Оценка адекватности модели
6.	7.	Сравнение, идентификация и прогнозирование
7.	9.	Анализ (дедукция и абдукция) классов и атрибутов
8.	10.1, 10.2.	Классификация и генерация конструктов классов и атрибутов
9.	10.3, 10.4.	Содержательное сравнение классов и атрибутов
10.	11.	Планирование и принятие решений о применении системы управляющих факторов

Необходимо отметить, что классификация операций системного анализа по В.М.Казиеву ближе всего к позиции, излагаемой в данной работе, т.к. этим автором названы 6 из 10 базовых когнитивных операций системного анализа: формализация; синтез (индукция); абстрагирование; анализ (дедукция); распознавание, и идентификация образов; классификация. Вместе с тем им не приводятся математическая модель, алгоритмы и инструментарий реализации этих операций и не ставится задача их разработки, кроме того, некоторые из них приведены дважды под разными названиями, например: анализ и синтез это *тоже самое*, что дедукция и индукция.

Необходимо также отметить, что по-видимому, впервые идея сведения мышления и процессов познания к когнитивным операциям была четко и осознанно письменно сформулирована в V веке до н.э.: *"Сущность интеллекта проявляется в способностях обобщения, абстрагирования, сравнения и классификации"* (цит.по пам., Патанжали, Йога-Сутра). Более того, все перечисленные Патанжали операции вошли в предлагаемый когнитивный конфигуратор.

Познание предметной области с одной стороны безусловно является фундаментом, на котором строится все грандиозное здание системного анализа, а с другой стороны, процессы познания являются связующим звеном, органично объединяющим "блоки" принципов и методов системного анализа в стройное здание. Более того, процессы познания буквально пронизывают все методы и принципы системного анализа, входя в них как один из самых существенных элементов.

Однако, на этом основании неверным будет представлять, что когнитивные операции являются подмножеством понятия "системный анализ", скорее наоборот: *системный анализ представляет собой один из теоретических методов познания, представимый в форме определенной последовательности когнитивных операций, тогда как другие последовательности этих операций позволяют образовать другие формы теоретического познания.*

1.1.2.1.4.4. Задачи формализации базовых когнитивных операций системного анализа

Для решения задачи формализации БКОСА необходимо решить следующие задачи:

1. Выбор единой интерпретируемой численной меры для классов и атрибутов.
2. Выбор неметрической меры сходства объектов в семантических пространствах.
4. Определение идентификационной и прогностической ценности атрибутов.
5. Ортонормирование семантических пространств классов и атрибутов (Парето-оптимизация).

Выбор единой интерпретируемой численной меры для классов и атрибутов

При построении модели объекта управления одной из принципиальных проблем является выбор формализованного представления для индикаторов, критериев и факторов (далее: факторов). Эта проблема распадается на две подпроблемы:

1. Выбор и обоснование смысла выбранной численной меры.
2. Выбор математической формы и способа определения (процедуры, алгоритма) количественного выражения для значений, отражающих степень взаимосвязи факторов и будущих состояний АОУ.

Рассмотрим требования к численной мере, определяемые существом подпроблем. Эти требования вытекают из необходимости совершать с численными значениями факторов математические операции (сложение, вычитание, умножение и деление), что в свою очередь необходимо для построения полноценной математической модели.

Требование 1: из формулировки 1-й подпроблемы следует, что все факторы должны быть приведены к некоторой общей и универсальной для всех факторов единице измерения, имеющей какой-то смысл, причем смысл, поддающийся единой сопоставимой в пространстве и времени интерпретации.

Традиционно в специальной литературе [12] рассматриваются следующие смысловые значения для факторов: стоимость (выигрыш-проигрыш или прибыль-убытки); полезность; риск; корреляционная или причинно-следственная взаимосвязь. Иногда предлагается использовать безразмерные меры для факторов, например эластичность, однако, этот вариант не является вполне удовлетворительным, т.к. не позволяет придать факторам содержательный и сопоставимый смысл и получить содержательную интерпретацию выводов, полученных на основе математической модели.

Таким образом, возникает ключевая при выборе численной меры проблема выбора смысла, т.е. по сути единиц измерения, для индикаторов, критериев и факторов.

Требование 2: высокая степень адекватности предметной области.

Требование 3: высокая скорость сходимости при увеличении объема обучающей выборки.

Требование 4: высокая независимость от артефактов.

Что касается конкретной математической формы и процедуры определения числовых значений факторов в выбранных единицах измерения, то обычно применяется метод взвешивания экспертных оценок, при котором эксперты предлагают свои оценки, полученные как правило неформализованным путем.

При этом сами эксперты также обычно ранжированы по степени их компетентности. Фактически при таком подходе числовые значения факторов является не определяемой, искомой, а исходной величиной. Иначе обстоит дело в факторном анализе, но в этом методе, опять же на основе экспертных оценок важности факторов, требуется предварительно, т.е. перед проведением исследования, принять решение о том, какие факторы исследовать (из-за жестких ограничений на размерность задачи в факторном анализе). Таким образом оба эти подхода реализуемы при относительно небольших размерностях задачи, что с точки зрения достижения целей настоящего исследования, является недостатком этих подходов.

Поэтому самостоятельной и одной из ключевых проблем является обоснованный и удачный выбор математической формы для численной меры индикаторов и факторов.

Эта математическая форма с одной стороны должна удовлетворять предыдущим требованиям, прежде всего требованию 1, а также должна быть процедурно вычислимой, измеримой.

Выбор неметрической меры сходства объектов в семантических пространствах

Существует большое количество мер сходства, из которых можно было бы упомянуть скалярное произведение, ковариацию, корреляцию, евклидово расстояние, расстояние Махаланобиса и др.

Проблема выбора меры сходства состоит в том, что при выбранной численной мере для координат классов и факторов она должна удовлетворять определенным критериям:

1. Обладать высокой степенью адекватности предметной области, т.е. высокой валидностью, при различных объемах выборки, как при очень малых, так и при средних и очень больших.
2. Иметь обоснованную, четкую, ясную и интуитивно понятную интерпретацию.
3. Быть нетрудоемкой в вычислительном отношении.
4. Обеспечивать корректное вычисление меры сходства для пространств с неортонормированным базисом.
5. Обеспечивать высокую достоверность и устойчивость идентификации при неполных (фрагментарных) и зашумленных данных.

Определение идентификационной и прогностической ценности атрибутов

Не все факторы имеют одинаковую ценность для решения задач идентификации, прогнозирования и управления. Традиционно считается, что факторы имеют одинаковую ценность только в тех случаях (обычно в психологии), когда определить их действительную ценность не представляется возможным по каким-либо причинам.

Для достижения целей, поставленных в данном исследовании, необходимо решить проблему определения ценности факторов, т.е. разработать математическую модель и алгоритм, которые допускают программную реализацию и обеспечивают на практике определение идентификационной и прогностической ценности факторов.

Ортонормирование семантических пространств классов и атрибутов (Парето-оптимизация)

Если не все факторы имеют одинаковую ценность для решения задач идентификации, прогнозирования и управления, то возникает проблема исключения из системы факторов тех из них, которые не представляют особой ценности.

Удаление малоценных факторов вполне оправданно и целесообразно, т.к. сбор и обработка информации по ним в среднем связана с такими же затратами времени, вычислительных и информационных ресурсов, как и при обработке ценных факторов. В этом состоит идея Парето-оптимизации.

Однако это удаление должно осуществляться при вполне определенных граничных условиях, характеризующих результирующую систему:

- адекватность модели;
- количество признаков на класс;
- суммарное количество градаций признаков в описательных шкалах.

В противном случае удаление факторов может отрицательно сказываться на качестве решения задач. На практике проблема реализации Парето-оптимизации состоит в том, что факторы вообще говоря коррелируют друг с другом и поэтому их ценность может изменяться при удалении любого из них, в том числе и наименее ценного. Поэтому просто взять и удалить наименее ценные факторы не представляется возможным и необходимо разработать корректный итерационный вычислительный алгоритм обеспечивающий решение этой проблемы при заданных граничных условиях.

1.1.2.1.5. Мышление как вычисление смысла и реализация операций со смыслом в инструментарии СК-анализа – системе "Эйдос"

В связи с ограниченностью объема данной работы нет возможности подробно рассмотреть вопросы, вынесенные в заголовок данного раздела. Но в этом и нет особой необходимости, т.к. такие вопросы как:

- системное обобщение семантической меры целесообразности информации как количественная мера знаний;
- способ представления и обобщения фактов в модели СК-анализа;
- способ вычисления смысла фактов и неклассическая логика;
- структуры данных базовых когнитивных операций;
- алгоритмы базовых когнитивных операций;

и ряд других, раскрывающих содержание математической модели, методики численных расчетов и технологии СК-анализа, подробно описаны в работе [81], размещенной на сайте автора: <http://Lc.kubagro.ru> (кнопка: "АСК-анализ").

1.1.2.1.7. Выводы

Системы искусственного интеллекта (СИИ) реализуют все больше функций, ранее выполнявшиеся только человеком, например, таких как: получение новых знаний из фактов, выявление причинно-следственных взаимосвязей между факторами, действующим на объект, и переходом этого объекта в те или иные состояния. Но в основе любой математической модели, реализованной в СИИ,

всегда лежит некое представление о том, каким образом осуществляются аналогичные процессы человеком. Поэтому для разработчиков СИИ большой интерес представляет уточнение смыслового содержания и углубленный анализ таких базовых понятий, как *данные – информация – знания, факт – смысл – мысль, мониторинг – анализ – управление*. Один из возможных вариантов понимания этот терминов предлагается и обосновывается в данной работе в рамках предложенной и реализованной в СК-анализе когнитивной концепции.

1.1.2.2. Понятие: "Система искусственного интеллекта", место СИИ в классификации систем обработки информации и управления

Существует много различных подходов к классификации информационных систем. Сразу вполне закономерно возникают вопросы о том:

1. Чем обусловлено различие этих классификаций?
2. Какова классификация этих классификаций?
3. Каким образом выбрать ту классификацию, которая нам больше всего подходит в данном случае?

Попробуем ответить на эти вопросы.

Различия между этими классификациями определяются теми критериями, по которым производится классификация, например:

- по степени структурированности решаемых задач;
- по автоматизируемым функциям;
- по степени автоматизации реализуемых функций;
- по сфере применения и характеру использования информации, в частности, по уровням управления.

Известно, что при обучении людей существуют различные *уровни предметной обученности*: воспроизведение (память); решение стандартных задач (умения, навыки); решение нестандартных, творческих задач (знания, активное интеллектуальное понимание).

Интеллект может проявляться в различных областях, но мы рассмотрим его возможности в решении задач, т.к. эта область проявления является типичной для интеллекта. Задачи бывают стандартные и нестандартные. Для стандартных задач известны *алгоритмы* решения. Для нестандартных они неизвестны. Поэтому *решение нестандартной задачи представляет собой проблему*.

Само понятие "стандартности" задачи *относительно*, относительна сама "неизвестность": т.е. алгоритм может быть известен одним и неизвестен другим, или информация о нем может быть недоступной в определенный момент или период времени, и доступной – в другой. Поэтому для одних задач может быть стандартной, а для других нет. *Нахождение или разработка алгоритма решения переводит задачу из разряда нестандартных в стандартные*.

В математике и кибернетике задача считается **решенной**, если известен алгоритм ее решения. Тогда процесс ее фактического решения превращается в рутинную работу, которую могут в точности выполнить человек, вычислитель-

ная машина или робот, под управлением программы реализующей данный алгоритм, не имеющие ни малейшего представления о смысле самой задачи.

Разработка алгоритма решения задачи связано с тонкими и сложными рассуждениями, требующими изобретательности, опыта, высокой квалификации. Считается, что эта работа является творческой, существенно неформализуемой и требует участия человека с его "естественным" опытом и интеллектом.

Здесь необходимо отметить, что существует технология решения изобретательских задач (ТРИЗ), в которой сделана попытка, по мнению многих специалистов, довольно успешная, позволяющая в какой-то степени формализовать процедуру решения творческих задач.

Интеллектуальными считаются задачи, связанные с разработкой алгоритмов решения ранее нерешенных задач определенного типа.

Отличительной особенностью и одним из основных источников эффективности алгоритмов является то, что они сводят решение сложной задачи к определенной последовательности *достаточно простых* или даже элементарных для решения задач. В результате нерешаемая задача становится решаемой. Исходная информация поступает на вход алгоритма, на каждом шаге она преобразуется и в таком виде передается на следующий шаг, в результате чего на выходе алгоритма получается информация, представляющая собой решение задачи.

Алгоритм *может быть исполнен* такой системой, которая способна реализовать элементарные операции на различных шагах этого алгоритма.

Существует ряд задач, таких, как распознавание образов и идентификация, прогнозирование, принятие решений по управлению, для которых разбиение процесса поиска решения на отдельные элементарные шаги, а значит и разработка алгоритма, весьма затруднительны.

Из этих рассуждений вытекает следующее определение интеллекта: *интеллект представляет собой универсальный алгоритма, способный разрабатывать алгоритмы решения конкретных задач.*

С этой точки зрения профессия программиста является одной из самых творческих и интеллектуальных, т.к. продуктом деятельности программиста являются алгоритмы реализованные на некотором языке программирования (программы).

Исходя из вышесказанного можно сделать вывод о том, что в нашем случае наиболее подходит классификацией ИС, основанная на критерии, позволяющем оценить "степень интеллектуальности ИС", т.е. на критерии "степени структурированности решаемых задач" (рисунк 5).

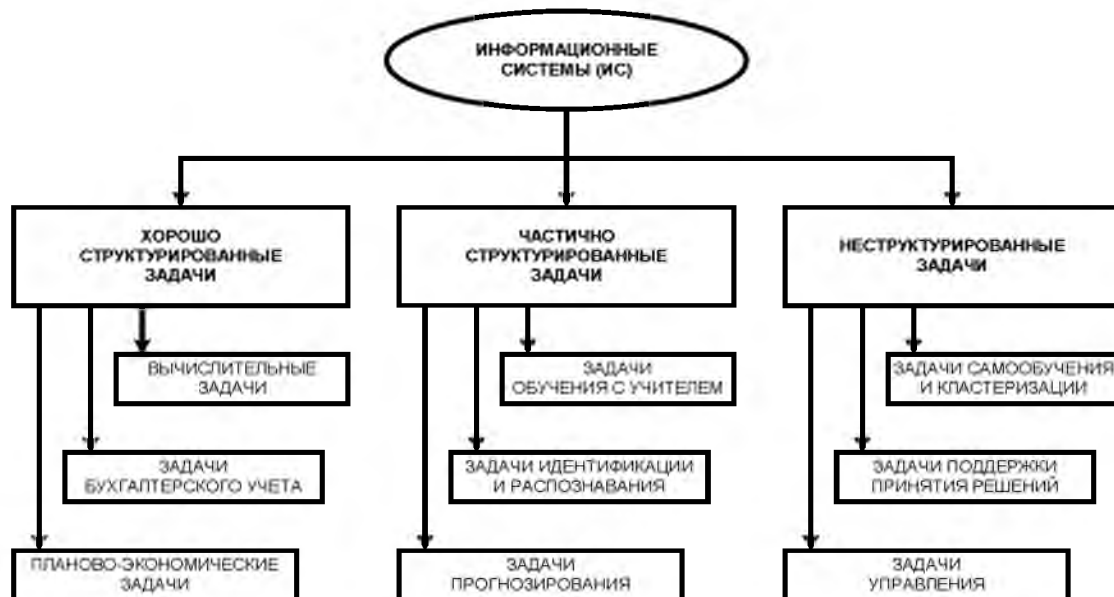


Рисунок 5. Классификация информационных систем по степени структурированности решаемых задач

Данная классификация не претендует на исчерпывающий характер.

Источники информации:

1. Сотник С.Л. Конспект лекций по курсу "Основы проектирования систем искусственного интеллекта": (1997-1998), <http://neuroscool.narod.ru/books/sotnik.html>.
2. Сайт: http://www.stu.ru/inform/glaves/glava3/gl_3_2.htm#klas_is.

1.1.2.3. Определение и классификация систем искусственного интеллекта, цели и пути их создания

1.1.2.3.1. Тест Тьюринга и критерии "интеллектуальности" информационных систем. Может ли машина мыслить? Может ли искусственный интеллект превзойти своего создателя?

"Одним из основных препятствий на пути внедрения систем искусственного интеллекта является недостаток интеллекта естественного"

/Из компьютерного фольклора/

В 1950 году в статье "Вычислительные машины и разум" (Computing machinery and intelligence) выдающийся английский математик и философ Алан

Тьюринг предложил тест, чтобы заменить бессмысленный, по его мнению, вопрос "может ли машина мыслить?" на более определённый.

Вместо того, чтобы отвлеченно спорить о критериях, позволяющих отличить живое мыслящее существо от машины, выглядящей как живая и мыслящая, он предложил реализуемый на практике способ установить это.

Судья-человек ограниченное время, например, 5 минут, переписывается в чате (в оригинале – по телеграфу) на естественном языке с двумя собеседниками, один из которых – человек, а другой – компьютер. Если судья за предоставленное время не сможет надёжно определить, кто есть кто, то компьютер прошёл тест.

Предполагается, что каждый из собеседников стремится, чтобы человеком признали его. С целью сделать тест простым и универсальным, переписка сводится к обмену текстовыми сообщениями.

Переписка должна производиться через контролируемые промежутки времени, чтобы судья не мог делать заключения исходя из скорости ответов. (Тьюринг ввел это правило потому, что в его времена компьютеры реагировали гораздо медленнее человека. Сегодня же это правило необходимо, наоборот, потому что они реагируют гораздо быстрее, чем человек).

Идею Тьюринга поддержал Джо Вайзенбаум, написавший в 1966 году первую "беседующую" программу "Элиза". Программа всего в 200 строк лишь повторяла фразы собеседника в форме вопросов и составляла новые фразы из уже использованных в беседе слов. Тем ни менее этого оказалось достаточно, чтобы поразить воображение тысяч людей.

А.Тьюринг считал, что компьютеры в конечном счёте пройдут его тест, т.е. на вопрос: "Может ли машина мыслить?" он отвечал утвердительно, но в будущем времени: "Да, смогут!"

Алан Тьюринг был не только выдающимся ученым, но и настоящим пророком компьютерной эры. Достаточно сказать, что в 1950 году (!!!), когда он писал, что к 2000 году, на столе у миллионов людей будут стоять компьютеры, имеющие оперативную память 1 миллиард бит (около 119 Мб) и оказался в этом абсолютно прав. Когда он писал это, все компьютеры мира вместе взятые едва ли имели такую память. Он также предсказал, что *обучение* будет играть важную роль в создании мощных интеллектуальных систем, что сегодня совершенно очевидно для всех специалистов по СИИ. Вот его слова: *"Пытаясь имитировать интеллект взрослого человека, мы вынуждены много размышлять о том процессе, в результате которого человеческий мозг достиг своего настоящего состояния... Почему бы нам вместо того, чтобы пытаться создать программу, имитирующую интеллект взрослого человека, не попытаться создать программу, которая имитировала бы интеллект ребенка? Ведь если интеллект ребенка получает соответствующее воспитание, он становится интеллектом взрослого человека... Наш расчет состоит в том, что устройство, ему подобное, может быть легко запрограммировано... Таким образом, мы расчленим нашу проблему на две части: на задачу построения "программы-ребенка" и задачу "воспитания" этой программы".*

Именно этот путь и используют практически все системы ИИ. Кроме того, именно на этом пути появляются и другие признаки интеллектуальной деятельности: накопление опыта, адаптация и т. д.

Против теста Тьюринга было выдвинуто несколько *возражений*.

1. Машина, прошедшая тест, может не быть разумной, а просто следовать какому-то хитроумному набору правил.

На что Тьюринг не без юмора отвечал: "А откуда мы знаем, что человек, который искренне считает, что он мыслит, на самом деле не следует какому-то хитроумному набору правил?"

2. Машина может быть разумной и не умея разговаривать, как человек, ведь и не все люди, которым мы не отказываем в разумности, умеют писать.

Могут быть разработаны варианты теста Тьюринга для неграмотных машин и судей.

3. Если тест Тьюринга и проверяет наличие разума, то он не проверяет сознание (consciousness) и свободу воли (intentionality), тем самым не улавливая весьма существенных различий между разумными людьми и разумными машинами.

Сегодня уже существуют многочисленные варианты интеллектуальных систем, которые не имеют цели, но имеют критерии поведения: генетические алгоритмы и имитационное моделирование эволюции. Поведение этих систем выглядит таким образом, как будто они имеют различные цели и добиваются их.

Ежегодно производится соревнование между разговаривающими программами, и наиболее человекоподобной, по мнению судей, присуждается приз Лебнера (Loebner).

Существует также приз для программы, которая, по мнению судей, пройдет тест Тьюринга. Этот приз ещё ни разу не присуждался.

В заключение отметим, что и сегодня тест Тьюринга не потерял своей фундаментальности и актуальности, более того – приобрел новое звучание в связи с возникновением Internet, общением людей в чатах и на форумах под условными никами и появлением почтовых и других программ-роботов, которые рассылают спам (некорректную навязчивую рекламу и другую не востребовавшую информацию), взламывают пароли систем и пытаются выступать от имени их зарегистрированных пользователей и совершают другие неправомерные действия.

Таким образом, возникает задачи:

- идентификации пола и других параметров собеседника (на эту возможность применения своего теста указывал и сам Тьюринг);

- выявления писем, написанных и посланных не людьми, а также такого автоматического написания писем, чтобы отличить их от написанных людьми было невозможно. Так что антиспамовый фильтр на электронной почте тоже представляет собой что-то вроде теста Тьюринга.

Не исключено, что скоро подобные проблемы (идентификации: человек или программа) могут возникнуть и в чатах. Что мешает сделать сетевых роботов типа программы "Элиза", но значительно более совершенных (все же сейчас не 1966, а 2004 год), которые будут сами регистрироваться в чатах и форумах

участвовать в них с использованием слов и модифицированных предложений других участников? Простейший вариант – дублирование тем с других форумов и перенос их с форума на форум без изменений, что мы уже иногда наблюдаем в Internet (например: сквозная тема про "Чакра-муни").

На практике чтобы на входе системы определить, кто в нее входит, человек или робот, достаточно при входе предъявить для решения простенькую *для человека*, но требующую огромных вычислительных ресурсов и системы типа неокогнитрона Фукушимы, задачу распознавания случайных наборов символов, представленных в нестандартных начертаниях, масштабах и поворотах на фоне шума (Vladimir Maximenko). Решил, – значит стучится человек-пользователь, не решил, – значит на входе робот, лазающий по мировой сети с неизвестными, чаще всего неблагоприятными целями.

Источники информации:

1. Свободная энциклопедия:

http://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D1%81%D1%82_%D0%A2%D1%8C%D1%8E%D1%80%D0%B8%D0%BD%D0%B3%D0%B0

2. Сотник С.Л. Конспект лекций по курсу "Основы проектирования систем искусственного интеллекта": (1997-1998),

<http://neuroscool.narod.ru/books/sotnik.html>.

3. Vladimir Maximenko. Реализация теста Тьюринга на Perl (ввод цифр изображенных на картинке) (perl image auth web cgi):
http://www.opennet.ru/base/dev/turing_test.txt.html

4. Captcha (<http://en.wikipedia.org/wiki/Captcha>) – полностью автоматизированные открытые тесты Тьюринга по разделению людей и машин (Completely Automated Public Turing tests to tell Computers and Humans Apart).

5. Сайт: <http://www.opennet.ru/opennews/art.shtml?num=4105>

1.1.2.3.2. Классификация систем искусственного интеллекта

В данном учебном пособии мы будем рассматривать следующие классы систем искусственного интеллекта:

1. Системы с интеллектуальной обратной связью и интеллектуальными интерфейсами.

2. Автоматизированные системы распознавания образов.

3. Автоматизированные системы поддержки принятия решений

4. Экспертные системы (ЭС).

5. Нейронные сети.

6. Генетические алгоритмы и моделирование эволюции.

7. Когнитивное моделирование.

8. Выявление знаний из опыта (эмпирических фактов) и интеллектуальный анализ данных (*data mining*).

Этими классами системы СИИ не исчерпываются, но мы вынуждены ограничиться ими, как основными, в связи с ограниченностью объема учебного пособия.

1.1.2.3.3. Особенности технологии создания систем искусственного интеллекта (обучение, "социализация", как технологический этап)

Система искусственного интеллекта в качестве существенной своей части включает базу знаний, которая является результатом обобщения опыта эксплуатации данной системы в определенных конкретных условиях. Это значит, что программистом может быть разработана только "пустая оболочка" системы искусственного интеллекта, которая превращается в работоспособную систему в результате процесса обучения, который, таким образом, является необходимым технологическим этапом создания подобных систем. Можно провести аналогию между такой системой и ребенком: ребенок не может идти работать, т.к. ему для этого предварительно требуется длительное обучение в школе, а затем часто и в вузе, чтобы он смог выполнять определенные виды работ.

1.1.2.4. Информационная модель деятельности специалиста и место систем искусственного интеллекта в этой деятельности

Информационная модель деятельности специалиста, представленная на рисунке 6, разработана на основе модели, впервые предложенной В.Н. Лаптевым (1984).

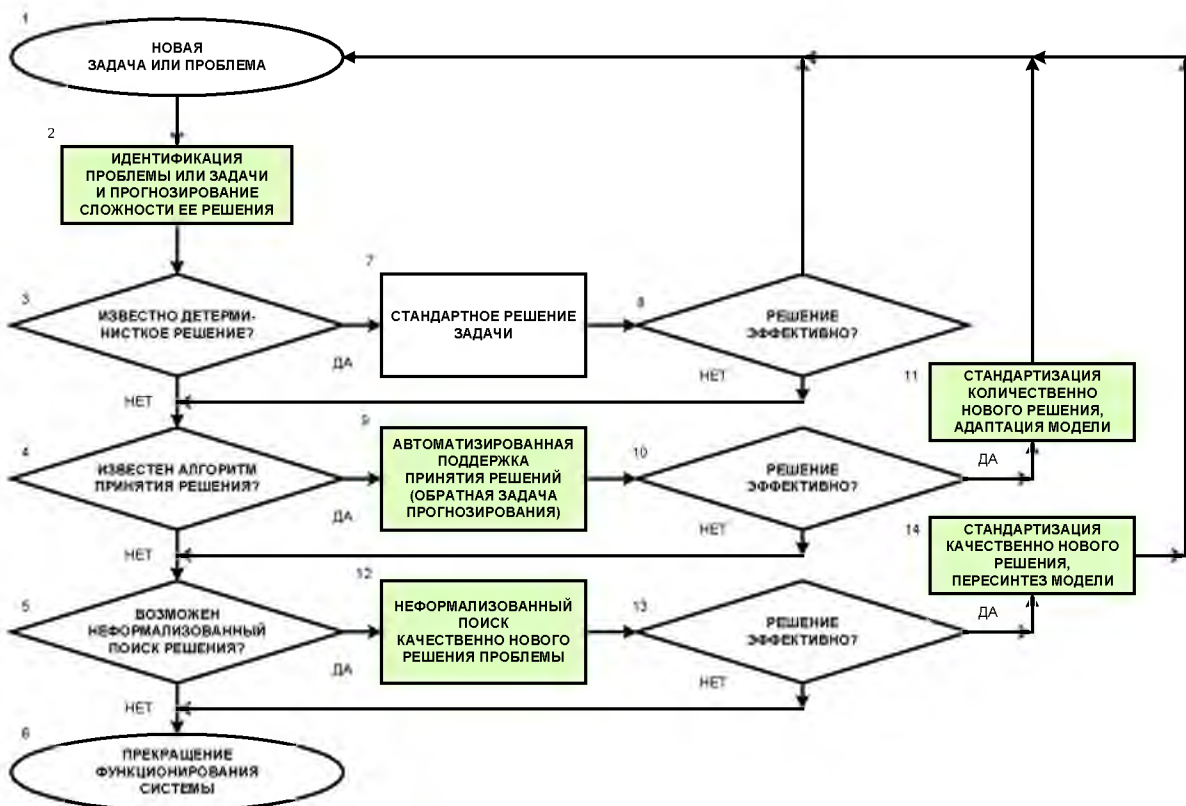


Рисунок 6. Информационная модель деятельности специалиста и место систем искусственного интеллекта в этой деятельности

На вход системы поступает задача или проблема. Толкование различия между ними также дано В.Н. Лаптевым и состоит в следующем.

*Ситуация, при которой фактическое состояние системы не совпадает с желаемым (целевым) называется **проблемной ситуацией** и представляет собой:*

- *задачу, если способ перевода системы из фактического состояния в желаемое точно известен, и необходимо лишь применить его;*
- *проблему, если способ перевода системы из фактического состояния в желаемое не известен, и необходимо сначала его разработать и лишь затем применить его.*

Таким образом, можно считать, что проблема – это задача, способ решения которой неизвестен. Это означает, что *если этот способ разработать, то этим самым проблема сводится к задаче, переводится в класс задач*. Проще говоря, проблема – это сложная задача, а задача – это простая проблема.

Но и проблемы различаются по уровню сложности:

- для решения одних достаточно *автоматизированной* системы поддержки принятия решений;
- для решения других – обязательным является творческое участие людей: специалистов, экспертов.

Рассмотрим *информационную модель деятельности специалиста*, представленную на рисунке 6.

Блок 1. На вход системы поступает задача или проблема. Что именно неясно, т.к. чтобы это выяснить необходимо идентифицировать ситуацию и обратиться к базе данных стандартных решений с запросом, существует ли стандартное решение для данной ситуации.

Блок 2. Далее осуществляется идентификация проблемы или задачи и прогнозирование сложности ее решения. На этом этапе применяется интеллектуальная система, относящаяся к классу систем распознавания образов, идентификации и прогнозирования или эта функция реализуется специалистом самостоятельно "вручную".

Блок 3. Если в результате идентификации задачи или проблемы по ее признакам установлено, что точно имеется стандартное решение, то это означает, что на вход системы поступила *точно такая же* задача, как уже когда-то ранее встречалась. Для установления этого достаточно информационно-поисковой системы, осуществляющей поиск по точному совпадению параметров запроса и в применении интеллектуальных систем нет необходимости. Тогда происходит переход на блок 7, а иначе на блок 4.

Блок 4. Если установлено, что точно такой задачи не встречалось, но встречались *сходные, аналогичные*, которые могут быть найдены в результате обобщенного (нечеткого) поиска системой распознавания образов, то решение может быть найдено с помощью автоматизированной системы поддержки принятия решений путем решения *обратной задачи прогнозирования*. Это значит, что на вход системы поступила не задача, а проблема, имеющая количественную

новизну по сравнению с решаемыми ранее (т.е. не очень сложная проблема). В этом случае осуществляется переход на блок 9, иначе – на блок 5.

Блок 5. Если установлено, что сходных проблем не встречалось, то необходимо качественно новое решение, поиск которого требует существенного творческого участия человека-эксперта. В этом случае происходит переход на блок 12, а иначе – на блок 6.

Блок 6. Переход на этот блок означает, что возможности поиска решения или выхода из проблемной ситуации системой исчерпаны и решения не найдено. В этом случае система обычно терпит ущерб целостности своей структуре и полноте функций, вплоть до разрушения и прекращения функционирования.

Блок 7. На этом этапе осуществляется реализация стандартного решения, соответствующего точно установленной задаче, а затем проверяется эффективность решения на блоке 8.

Блок 8. Если стандартное решение оказалось эффективным, это означает, что на этапах 2 и 3 идентификация задачи и способа решения осуществлены правильно и система может переходить к разрешению следующей *проблемной ситуации* (переход на блок 1). Если же стандартное решение оказалось неэффективным, то это означает, что проблемная ситуация идентифицирована как стандартная задача неверно и необходимо продолжить попытки ее разрешения с использованием более общих подходов, основанных на применении систем искусственного интеллекта (переход на блок 4), например, систем поддержки принятия решений.

Блок 9. Применяется автоматизированная система поддержки принятия решений, обеспечивающая решение обратной задачи прогнозирования. Отличие подобных систем от информационно-поисковых состоит в том, что они способны производить обобщение, выявлять силу и направление влияния различных факторов на поведение системы, и, на основе этого, по заданному целевому состоянию вырабатывать рекомендации по системе факторов, которые могли бы перевести систему в это состояние (обратная задача прогнозирования).

Блок 10. Если решение, полученное с помощью системы поддержки принятия решений, оказалось неэффективным, то это означает, что проблемная ситуация идентифицирована как аналогичная ранее встречавшимся неверно. Следовательно, что на вход системы поступила качественно новая, по сравнению с решаемыми ранее, т.е. сложная проблема. В этом случае необходимо продолжить попытки разрешения проблемы с использованием творческих неформализованных подходов с участием человека-эксперта и перейти на блок 5, иначе – на блок 11.

Блок 11. Информация об условиях и результатах решения проблемы заносится в базу знаний, т.е. стандартизируется. После чего база знаний количественно (не принципиально) изменяется, т.е. осуществляется ее адаптация. *В результате адаптации при встрече в будущем точно таких же проблемных ситуаций, как разрешенная, система уже будет разрешать ее не как проблему, а как стандартную задачу.*

Блок 12. На этом этапе с использованием неформализованных творческих подходов осуществляется поиск качественно нового решения проблемы, не встречавшейся ранее, после чего управление передается блоку 13.

Блок 13. Если решение, полученное экспертами с помощью неформализованных подходов, оказалось неэффективным, то это означает, что система терпит крах (осуществляется переход на блок 6). Если же адекватное решение найдено, то происходит переход на блок 14.

Блок 14. Стандартизация качественно нового решения, проблемы и пересинтез модели. Информация об условиях и результатах творческого решения проблемы заносится в базу знаний, т.е. стандартизируется. После этого база знаний качественно, принципиально изменяется, т.е. фактически осуществляется ее пересоздание (пересинтез). В результате пересинтеза базы знаний при встрече в будущем проблемных ситуаций, аналогичных разрешенной, система уже будет реагировать на них как проблемы, решаемые автоматизированными системами поддержки принятия решений.

Блоки, в которых используются системы искусственного интеллекта, на рисунке 6 показаны затемненными:

– **блоки 2 и 12:** система распознавания образов, идентификации и прогнозирования;

– **блоки 9, 11, 12 и 14:** автоматизированная система поддержки принятия решений.

В заключение раздела, с целью повышения настроения читателей-студентов, приведем шуточный алгоритм решения проблем (рисунок 7).

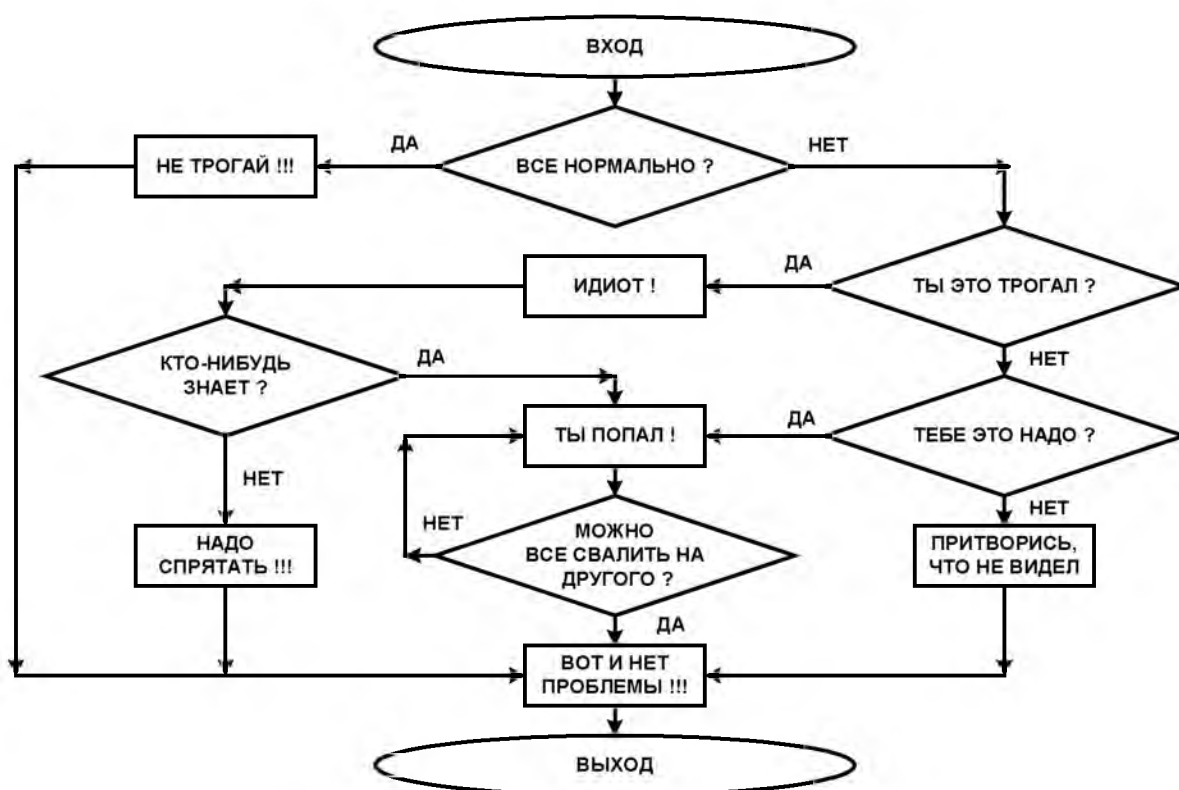


Рисунок 7. Шуточный алгоритм решения проблем (Internet-фольклор)

1.1.2.5. Жизненный цикл системы искусственного интеллекта и критерии перехода между этапами этого цикла

Жизненный цикл систем искусственного интеллекта сходен с жизненным циклом другого программного обеспечения и включает этапы и критерии перехода между ними, представленные в таблице 2.

Таблица 2 – ЭТАПЫ ЖИЗНЕННОГО ЦИКЛА СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И КРИТЕРИИ ПЕРЕХОДА МЕЖДУ НИМИ

№	Наименование этапа	Критерии перехода к следующему этапу
1	Разработка идеи и концепции системы	Появление (в результате проведения маркетинговых и рекламных мероприятий) заказчика или спонсора, заинтересовавшегося системой
2	Разработка теоретических основ системы	Обоснование выбора математической модели по критериям или обоснование необходимости разработки новой модели
3	Разработка математической модели системы	Детальная разработка математической модели
4	Разработка методики численных расчетов в системе:	
4.1	– разработка структур данных	детальная разработка структур входных, промежуточных и выходных данных
4.2	– разработка алгоритмов обработки данных	разработка обобщенных и детальных алгоритмов, реализующих на разработанных структурах данных математическую модель
5	Разработка структуры системы и экранных форм интерфейса	Разработка иерархической системы управления системой, структуры меню, экранных форм и средств управления на экранных формах
6	Разработка программной реализации системы	Разработка исходного текста программы системы, его компиляция и линковка. Исправление синтаксических ошибок в исходных текстах
7	Отладка системы	Поиск и исправление логических ошибок в исходных текстах на контрольных примерах. На контрольных примерах новые ошибки не обнаруживаются.
8	Экспериментальная эксплуатация	Поиск и исправление логических ошибок в исходных текстах на реальных данных без применения результатов работы системы на практике. На реальных данных новые ошибки практически не обнаруживаются, но считаются в принципе возможными.
9	Опытная эксплуатация	Поиск и исправление логических ошибок в исходных текстах на реальных данных с применением результатов работы системы на практике. На реальных данных новые ошибки не обнаруживаются и считаются недопустимыми.
10	Промышленная эксплуатация	Основной по длительности период, который продолжается до тех пор, пока система функционально устраивает Заказчика. У Заказчика появляется необходимость внесения количественных (косметических) изменений в систему на уровне п.5 (т.е. без изменения математической мо-

		дели, структур данных и алгоритмов)
11	Заказные модификации системы	У Заказчика формируется потребность внесения качественных (принципиальных) изменений в систему на уровне п.3 и п.4, т.е. с изменениями в математической модели, структурах данных и алгоритмах
12	Разработка новых версий системы	Выясняется техническая невозможность или финансовая нецелесообразность разработки новых версий системы
13	Снятие системы с эксплуатации	

Контрольные вопросы

1. Данные, информация, знания. Системно-когнитивный анализ как развитие концепции смысла Шенка-Абельсона.
2. Когнитивная концепция СК-анализа и синтез когнитивного конфигулятора.
3. Мышление как вычисление смысла и реализация операций со смыслом в инструментарии СК-анализа - системе "Эйдос".
4. Понятие: "Система искусственного интеллекта", место СИИ в классификации информационных систем.
5. Определение и классификация систем искусственного интеллекта, цели и пути их создания.
6. Тест Тьюринга и критерии "интеллектуальности" информационных систем. Может ли машина мыслить? Может ли искусственный интеллект превзойти своего создателя?
7. Классификация систем искусственного интеллекта.
8. Особенности технологии создания систем искусственного интеллекта (обучение, "социализация", как технологический этап).
9. Информационная модель деятельности специалиста и место систем искусственного интеллекта в этой деятельности.
10. Жизненный цикл системы искусственного интеллекта и критерии перехода между этапами этого цикла.

Рекомендуемая литература

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. – 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.
3. Кива Владимир, сайт: <http://vlak.webzone.ru/rus/it/knowledge.html>.

РАЗДЕЛ 1.2. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ И ПРИМЕНЕНИЕ УНИВЕРСАЛЬНОЙ КОГНИТИВНОЙ АНАЛИТИЧЕСКОЙ СИСТЕМЫ "ЭЙДОС"

1.2.1. ЛЕКЦИЯ-3.

Теоретические основы системно-когнитивного анализа и место СИИ в структуре АСУ

Учебные вопросы:

1. Системный анализ, как метод познания.
2. Когнитивная концепция и синтез когнитивного конфигулятора.
3. СК-анализ, как системный анализ, структурированный до уровня базовых когнитивных операций.
4. Место и роль СК-анализа в структуре управления.

1.2.1.1. Системный анализ, как метод познания

1.2.1.1.1. Принципы системного анализа

Анализ (дедукция) представляет собой метод познания "от общего к частному", "от целого к частям". Абдукция представляет собой обобщение дедукции на основе нечеткой логики. При анализе существует опасность за исследованием частей упустить из рассмотрения их взаимодействие, то общее, что их объединяет в целое (т.е. взаимодействие частей для достижения общей цели). Такой подход был характерен для метафизического (не диалектического) стиля мышления. Системный анализ лишен этого недостатка, поэтому многие совершенно справедливо считают системный анализ "современным воплощением прикладной диалектики" [194].

В этом контексте развиваемая в данной работе модель развития активных систем путем чередования детерминистских и бифуркационных состояний представляет собой ничто иное, как естественнонаучное трактовку закона диалектики "Перехода количественных изменений в качественные". Иначе говоря, детерминистские этапы – это этапы количественного, эволюционного изменения объекта управления, а бифуркационные – этапы его качественного, революционного преобразования. Поэтому *системный анализ рассматривается в данной работе как теоретический метод познания детерминистско-бифуркационной динамики систем*. Таким образом, *логически системный анализ можно считать результатом выполнения программы естественнонаучного развития диалектики, хотя исторически он и возник иначе*. Саму когнитивную психологию также в определенной мере можно рассматривать как результат выполнения программы естественнонаучного развития гносеологии.

"Системный анализ" – это такой анализ систем, при котором за исследованием частей не только теряется целое, но и весь процесс исследования струк-

туры системы и взаимосвязей ее элементов осуществляется под углом зрения целей и функций системы (авт.).

Система – это совокупность элементов (частей), взаимодействующих друг с другом для достижения некоторой общей цели. Система обеспечивает преимущество в достижении цели, т.е. достижение цели разрозненными элементами вне системы менее вероятно или вообще невозможно.

Система – это всегда нечто большее, чем просто сумма частей, т.е. она обладает качественно новыми (эмерджентными) свойствами, которые отсутствуют у ее частей. По мнению автора, *в конечном счете все свойства имеют эмерджентную природу, т.е. любое качество основано на уровне Реальности этим качеством не обладающим*. Термин "Реальность" включает и бытие, и небытие.

Например, качество "быть соленым" основано на свойствах Na и Cl, этим качеством ни в коей мере не обладающими. Движение с различными скоростями в метрическом пространстве основано на нелокальном уровне Реальности, в котором нет локализации объектов в физическом пространстве-времени. Об этом догадался еще Зенон и отразил логически в своих знаменитых апориях из которых следует не невозможность движения, как некоторые почему-то думают, а лишь невозможность адекватного отражения движения средствами формальной логики.

Системный анализ используется в тех случаях, когда стремятся исследовать объект с разных сторон, комплексно. Термин "системный анализ" впервые появился в 1948 г. в работах корпорации RAND в связи с задачами внешнего управления, а в отечественной литературе широкое распространение получил после перевода книги С. Оптнера. Дальнейшее развитие системный анализ получил в трудах зарубежных и отечественных ученых: Гэйна К., Сарсона Т., Клиланда Д., Кинга В., Перегудова Ф.И., Тарасенко Ф.П. [194], Юдина Б. Г. Валуева С.А., Губанова В.А., Захарова В.В., Коваленко А.Н., Кафарова В.В., Дорохова И.Н., Маркова Е.П., Мисюра Я.С., Купрюхина А.И., Дубенчака Г.И., Джагарова Ю.А. Дубенчака В.Е.

Во многих работах системный анализ развивается применительно к программно-целевому планированию и управлению. Однако, при этом получили развитие формализованные методики анализа систем (декомпозиции). В работах ведущих ученых по программированию урожая: Денисова Е.П., Ермохина Ю.И., Каюмова М. К., Мухортова С.Я., Неклюдова А.Ф., Филина В.И., Царева А.П., связанных с проблематикой данного исследования, в явной форме не используется автоматизированный системный анализ. Это, по-видимому, обусловлено тем, что формализованные средства системного анализа, обеспечивающие декомпозицию *с сохранением целостности* практически отсутствуют.

Системный анализ основывается на следующих принципах: *единства* – совместное рассмотрение системы как единого целого и как совокупности частей; *развития* – учет изменяемости системы, ее способности к развитию, накапливанию информации с учетом динамики окружающей среды; *глобальной цели* – ответственность за выбор глобальной цели (*оптимум для подсистем вообще говоря не является оптимумом для всей системы*); *функциональности* – совместное рассмотрение структуры системы и функций с приоритетом функций над структурой; *децентрализации* – сочетание децентрализации и централи-

зации; *иерархии* – учет соподчинения и ранжирования частей; *неопределенности* – учет вероятностного наступления события; *организованности* – степень выполнения решений и выводов.

Сущность системного подхода формулировалась многими авторами. В развернутом виде она сформулирована Афанасьевым В.Н., Колмановским В.Б. и Носовым В.Р., определившими ряд взаимосвязанных аспектов, которые в совокупности и единстве составляют системный подход: системно-элементный, отвечающий на вопрос, из чего (каких компонентов) образована система; системно-структурный, раскрывающий внутреннюю организацию системы, способ взаимодействия образующих ее компонентов; системно-функциональный, показывающий, какие функции выполняет система и образующие ее компоненты; системно-коммуникационный, раскрывающий взаимосвязь данной системы с другими как по горизонтали, так и по вертикали; системно-интегративный, показывающий механизмы, факторы сохранения, совершенствования и развития системы; системно-исторический, отвечающий на вопрос, как, каким образом возникла система, какие этапы в своем развитии проходила, каковы ее исторические перспективы.

Системный анализ используется для того, чтобы организовать процесс принятия решения в сложных проблемных ситуациях. При этом основным требованием системного анализа является полнота и всесторонность рассмотрения проблемы. Основной особенностью системного анализа является сочетание формальных методов и неформализованного (экспертного) знания. Последнее помогает неформализованным путем найти новые пути решения проблемы, не содержащиеся в формальной модели, а затем учесть последствия решений в модели, т.е. формализовать их, за счет чего непрерывно развивать модель и методы поддержки принятия решений.

С учетом вышесказанного в определении системного анализа нужно подчеркнуть, что системный анализ:

- применяется для решения таких проблем, которые не могут быть поставлены и решены отдельными методами математики, т.е. проблем с неопределенностью ситуации принятия решения, когда используют не только формальные методы, но и методы качественного анализа ("формализованный здравый смысл"), интуицию и опыт лиц, принимающих решения;
- объединяет разные методы на основе единой методик;
- опирается на научное мировоззрение;
- объединяет знания, суждения и интуицию специалистов различных областей знаний и обязывает их к определенной дисциплине мышления;
- уделяет основное внимание целям и целеобразованию.

В частности, основными специфическими особенностями системного анализа, отличающими его от других системных направлений, являются:

1. Наличие в системном анализе средств для организации процессов целеобразования, структуризации и анализа целей (другие системные направления ставят задачу достижения целей, разработки вариантов пути их достижения и выбора наилучшего из этих вариантов, а *системный анализ рассматривает объекты как активные системы, способные к целеобразованию и принятию*

решений, а затем уже и к достижению сформированных целей путем реализации принятых решений).

2. Разработка и использование методики, в которой определены этапы, подэтапы системного анализа и методы их выполнения, причем в методике сочетаются как формальные методы и модели, так и методы, основанные на интуиции специалистов, помогающие использовать их знания, что обуславливает особую привлекательность системного анализа для решения экономических проблем, в том числе в такой сложно формализуемой области как АПК.

1.2.1.1.2. Методы и этапы системного анализа

Основные методы, направленные на использование интуиции и опыта специалистов, а также методы формализованного представления систем, т.е. методы системного анализа, рассмотрены в работах [194]: метод "мозговой атаки"; метод экспертных оценок; метод "Дельфи"; метод "дерева целей"; морфологические методы.

Ведущие зарубежные Акофф Р., Бир С., Винер Р., Месарович М., Мако Д., Такахара И., Оптнер С.Л., Черчмен У., Эшби У.Р., Янг С., и отечественные ученые в области системного анализа Ф.И.Перегудов, Ф.П.Тарасенко [194], В.С.Симанков, Э.Х.Лийв [69], В.Н.Спицнадель, предлагают несколько отличающиеся друг от друга схемы основных этапов системного анализа.

Отечественные классики в области системного анализа Ф.И.Перегудов и Ф.П.Тарасенко считают [194], что системный анализ не может быть полностью формализован. Ими предложена следующая схема неформализованных этапов системного анализа (рисунок 8):

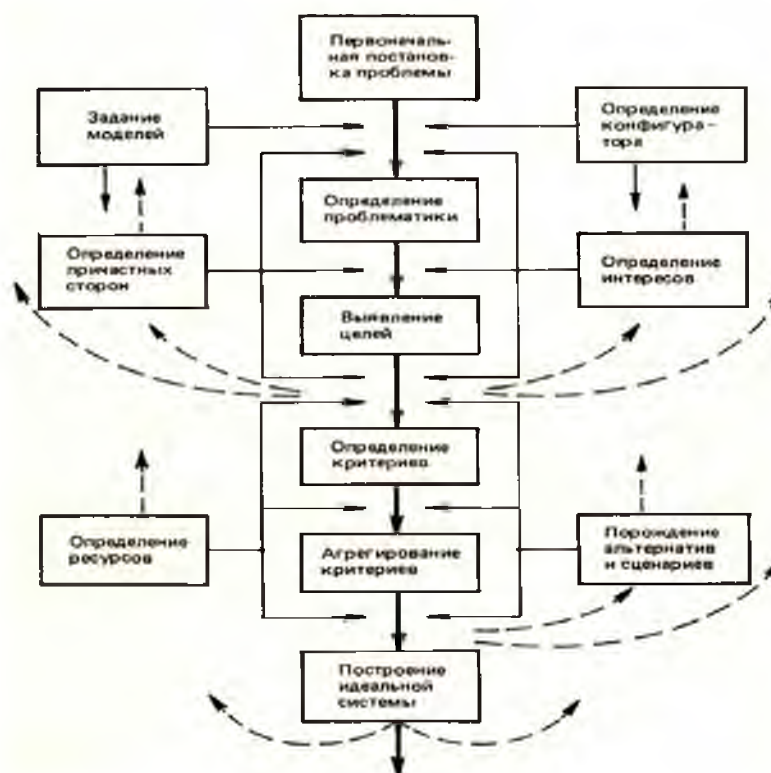


Рисунок 8. Неформализуемые этапы системного анализа по Ф.И.Перегудову и Ф.П.Тарасенко [194]

1. Определение конфигуратора.
2. Постановка проблемы – отправной момент исследования. В исследовании системы ему предшествует работа по структурированию проблемы.
3. Расширение проблемы до проблематики, т.е. нахождение системы проблем или задач, существенно связанных с исследуемой проблемой, без учета которых она не может быть решена.
4. Выявление целей: цели указывают направление, в котором надо двигаться, чтобы поэтапно решить проблему.
5. Формирование критериев. Критерий – это количественное отражение степени достижения системой поставленных перед ней целей. Критерий – это правило выбора предпочтительного варианта решения из ряда альтернативных. Критериев может быть несколько. Многокритериальность является способом повышения адекватности описания цели. Критерии должны описать по возможности все важные аспекты цели, но при этом необходимо минимизировать число необходимых критериев.
6. Агрегирование критериев. Выявленные критерии могут быть объединены либо в группы, либо заменены обобщающим критерием.
7. Генерирование альтернатив и выбор с использованием критериев наилучшей из них. Формирование множества альтернатив является творческим этапом системного анализа.
8. Исследование ресурсных возможностей, включая информационные потоки и ресурсы.
9. Выбор формализации (построение и использование моделей и ограничений) для решения проблемы.
10. Оптимизация (для простых систем).
11. Декомпозиция.
12. Наблюдение и эксперименты над исследуемой системой.
13. Построение системы.
14. Использование результатов проведенного системного исследования.

Однако в утверждении этих авторов есть некоторый смысловой парадокс, состоящий в том, что *предложенная ими схема сама может рассматриваться как первый шаг на пути формализации представленных на ней этапов системного анализа в форме алгоритма.*

Как уже отмечалось, специфической особенностью системного анализа является сочетание качественных и формальных методов. Такое сочетание составляет основу любой используемой методики. Различные схемы системного анализа, предлагаемые ведущими учеными в этой области (Оптнер С.Л., Янг С., Федоренко Н.П., Никаноров С.П., Черняк Ю.И., Перегудов Ф.И., Тарасенко Ф.П., Симанков В.С., Казиев В.М., Лийв Э.Х.) сведены в таблице 3.

Наиболее детализированная на данный момент многоуровневая иерархическая структуризация системного анализа в виде IDEF0-диаграмм, насколько известно автору, предложена в докторской диссертации В.С.Симанкова (в данной работе не приводится из-за ее ограниченного объема, но она приведена в работе [81]).

Таблица 3 – ЭТАПЫ СИСТЕМНОГО АНАЛИЗА СОГЛАСНО РАЗЛИЧНЫМ АВТОРАМ

№	Перегудов Ф.И. Тарасенко Ф.П.	Черняк Ю.И.	Оптнер С.Л.	Казиев В.М.	Никаноров С.П.	Янг С.	Федоренко Н.П.	Симанков В.С. ^а	Лийв Э.Х.
1.	Определение конфигурирования	Анализ проблемы	Идентификация симптомов	абстрагирование и конкретизация;	Обнаружение проблемы	Определение цели организации	Формулирование проблемы	Определение проблемы	Выделение объекта исследования от общей массы явлений. Очертание контур, пределов системы, его основных частей, элементов, связи с окружающей средой.
2.	Постановка проблемы	Определение системы	Определение актуальности проблемы	анализ и синтез;	Оценка актуальности проблемы	Выявление проблемы	Определение целей	Определение целей системы	Установление цели исследования: выяснение структуры или функций системы; изменение и преобразование её деятельности или наличие длительного механизма управления и функционирования.
3.	Расширение проблемы до проблематики	Анализ структуры системы	Определение целей	индукция и дедукция;	Анализ ограниченной проблемы	Диагноз	Сбор информации	Анализ системы	Система не обязательно является материальным объектом. Она может быть и воображаемым в мозгу сочетанием всех возможных структур для достижения определённой цели.
4.	Выявление целей	Формулирование общей цели и критерия	Определение структуры системы и её дефектов	формализация;	Определение критериев	Поиск решения	Разработка максимального количества альтернатив	Синтез системы	Выяснение основных критериев для обеспечения целесообразного или целенаправленного действия системы, а также основные ограничения и условия существования.
5.	Формирование критериев	Декомпозиция цели	Определений возможностей	структурирование;	Анализ существующей системы	Оценка и выбор альтернативы	Отбор альтернатив	Реализация системы	Определение альтернативных вариантов при выборе структур или элементов для достижения заданной цели. При этом необходимо учесть все факторы, влияющие на систему и все возможные варианты решения проблемы.
6.	Агрегирование критериев	Выявление потребностей в ресурсах, композиция целей	Нахождение альтернатив	макетирование;	Поиск возможностей (альтернатив)	Согласование решения	Построение модели в виде уравнений, программ или сценария	(*) Пояснение к приведенным этапам системного анализа приведено в тексте.	Составление модели функционирования системы, учитывая всех существенных факторов. Существенность факторов определяется по их влиянию на определяющие критерии цели
7.	Генерирование альтернатив и выбор	Прогноз и анализ будущих условий	Оценка альтернатив	алгоритмизация;	Выбор альтернативы	Утверждение решения	Оценка затрат		Оптимизация режима существования или работы системы. Градация решений по их оптимальному эффекту, по функционированию (достижению цели).
8.	Исследование ресурсных возможностей	Оценка целей и средств	Выработка решения	моделирование;	Обеспечение признания	Подготовка к вводу в действие	Испытание чувствительности решения (параметрическое исследование)		Проектирование оптимальных структур и функциональных действий системы. Определение оптимальной схемы их регулирования или управления.
9.	Выбор формализации (модели)	Отбор вариантов	Признание решения	программное управление;	Принятие решения (принятие формальной ответственности)	Управление применением решения			Контроль за работой системы в эксплуатации, определение ее надёжности и работоспособности. Установление надёжной обратной связи по результатам функционирования.
10.	Оптимизация модели	Диагноз существующей системы	Запуск процесса решения	распознавание, классификация и идентификация образов;	Реализация решения	Проверка эффективности			
11.	Декомпозиция	Построение комплексной программы развития	Управление процессом реализации решения	экспертное оценивание и тестирование и другие методы и процедуры	Определение результатов решения				
12.	Наблюдения и эксперименты над исследуемой системой	Проектирование организации для достижения цели	Оценка реализации и её последствий						
13.	Построение системы								

Работы по детализации системного анализа вдохновлялись надеждой на то, что более мелкие этапы легче автоматизировать. Этой надежде суждено было сбыться лишь частично. *Но парадокс этого пути автоматизации системного анализа, который оправданно было бы назвать путем "максимальной детализации" состоит в том, что на пути "максимальной детализации" сама автоматизация системного анализа велась не системно:* т.е. различные мелкие этапы СА автоматизировались различными не связанными друг с другом группами ученых и разработчиков, которые исходили при этом из своих целей, научных интересов и возможностей.

В результате на данный момент сложилась следующая картина:

- не все этапы системного анализа автоматизированы;
- для автоматизации различных этапов системного анализа применяются различные математические модели и теории;
- эти модели реализуются с применением различных программных систем, не связанных друг с другом и не образующие единого инструментального комплекса;
- эти программные системы созданы с использованием различных инструментальных средств, на различных платформах и языках программирования;
- как правило, эти программные системы имеют специализированный характер, т.е. автоматизируют отдельные этапы системного анализа не в универсальной форме, а лишь в одной конкретной предметной области.

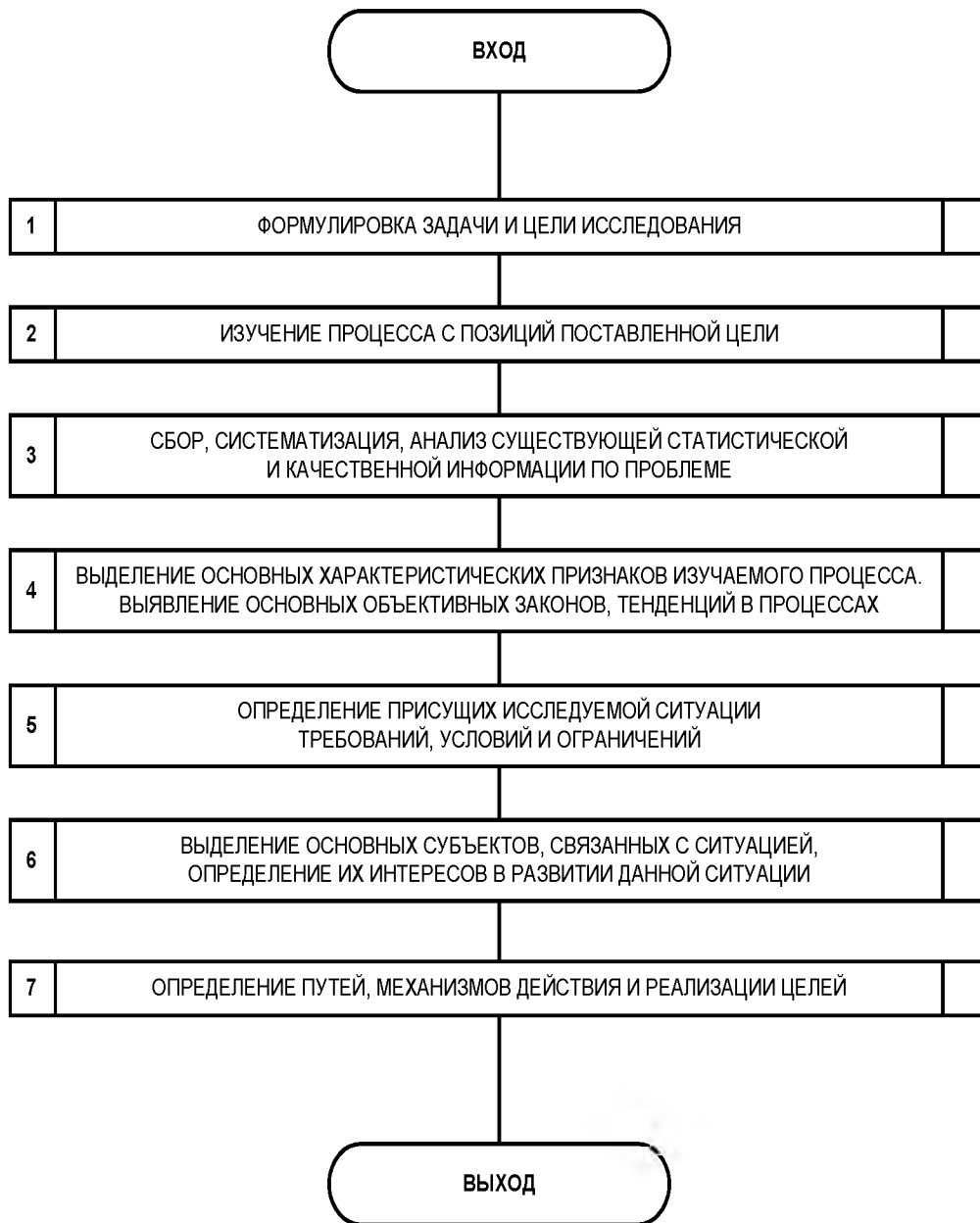
Поэтому автор считает, что "максимальная детализация системного анализа" – не самоцель, т.е. бессмысленна "детализация ради детализации". Безусловно, данное направление представляет интерес в научном плане, однако, по-видимому, оно не перспективно как путь автоматизации системного анализа, т.к. опыт показывает, что будучи изначально предназначено для облегчения процесса автоматизации на деле оно лишь фактически усложнило решение этой задачи.

Анализ приведенных детализированных схем этапов и процедур системного анализа показывает, что на всех этапах широко используются когнитивные операции, т.е. операции, связанные с познанием предметной области и объекта управления и с созданием их идеальной модели.

Поэтому в данной работе предлагается иной путь автоматизации системного анализа, основанный не на его максимальной детализации, а на интеграции с когнитивными технологиями путем структурирования по когнитивным операциям.

1.2.1.1.3. Этапы когнитивного анализа

Рассмотрим этапы когнитивного анализа в варианте, предлагаемом ведущими отечественными учеными в этой области Максимов В.И., Корноушенко Е.К., Гребенюк Е.А., Григорян А.К. (рисунок 9). В этой связи необходимо также отметить работы Казиева В.М., С.В.Качаева, А.А.Кулинич, А.Н.Райкова, Д.И.Макаренко, С.В.Ковриги, Е.А.Гребенюка, А.К.Григоряна в области когнитивного анализа [60, 160-164, 206-209].



**Рисунок 9. Этапы когнитивного анализа
по В.И.Максимову и Е.К.Корноушенко**

Если проанализировать перечисленные методы системного анализа, то можно сделать основополагающий для данного исследования вывод о том, что *все они самым существенным образом так или иначе основаны на процессах познания предметной области.*

Поэтому как одно из важных и перспективных направлений автоматизации системного анализа предлагается рассматривать автоматизацию когнитивных операций системного анализа. Чтобы выявить эти операции и определить их место и роль в процессах познания, рассмотрим базовую когнитивную концепцию.

1.2.1.1.4. Предлагаемая обобщенная схема системного анализа, ориентированного на интеграцию с когнитивными технологиями

Сопоставительный анализ приведенных в таблице 3 и рисунке 9 схем системного и когнитивного анализа, показывает, что они во многом взаимно дополняют друг друга. Это говорит о *возможности* объединения различных схем системного анализа и когнитивного анализа в одной схеме "системного анализа, ориентированного на интеграцию с когнитивными технологиями". Предполагается, что это *целесообразно*, т.к. полученная схема системного анализа более пригодна для формализации и автоматизации, чем приведенные схемы детализированного системного анализа. С учетом этого, а также модели реагирования открытых систем на вызовы среды, предложенной в 1984 В.Н. Лаптевым (рисунок 10), нами предложена схема системного анализа, ориентированного на интеграцию с когнитивными технологиями, представленная на рисунке 11.

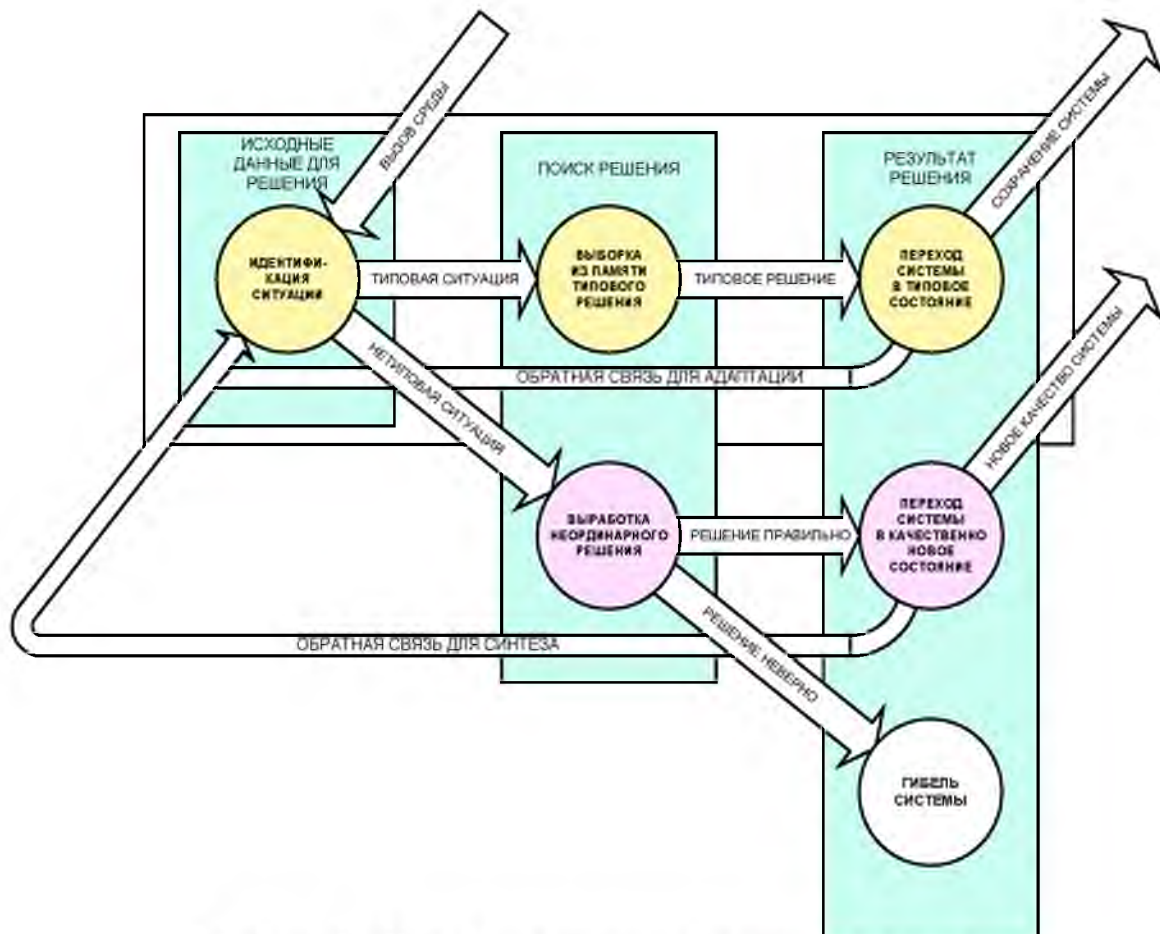


Рисунок 10. Схема реагирования открытой системы на вызовы среды по В.Н.Лаптеву

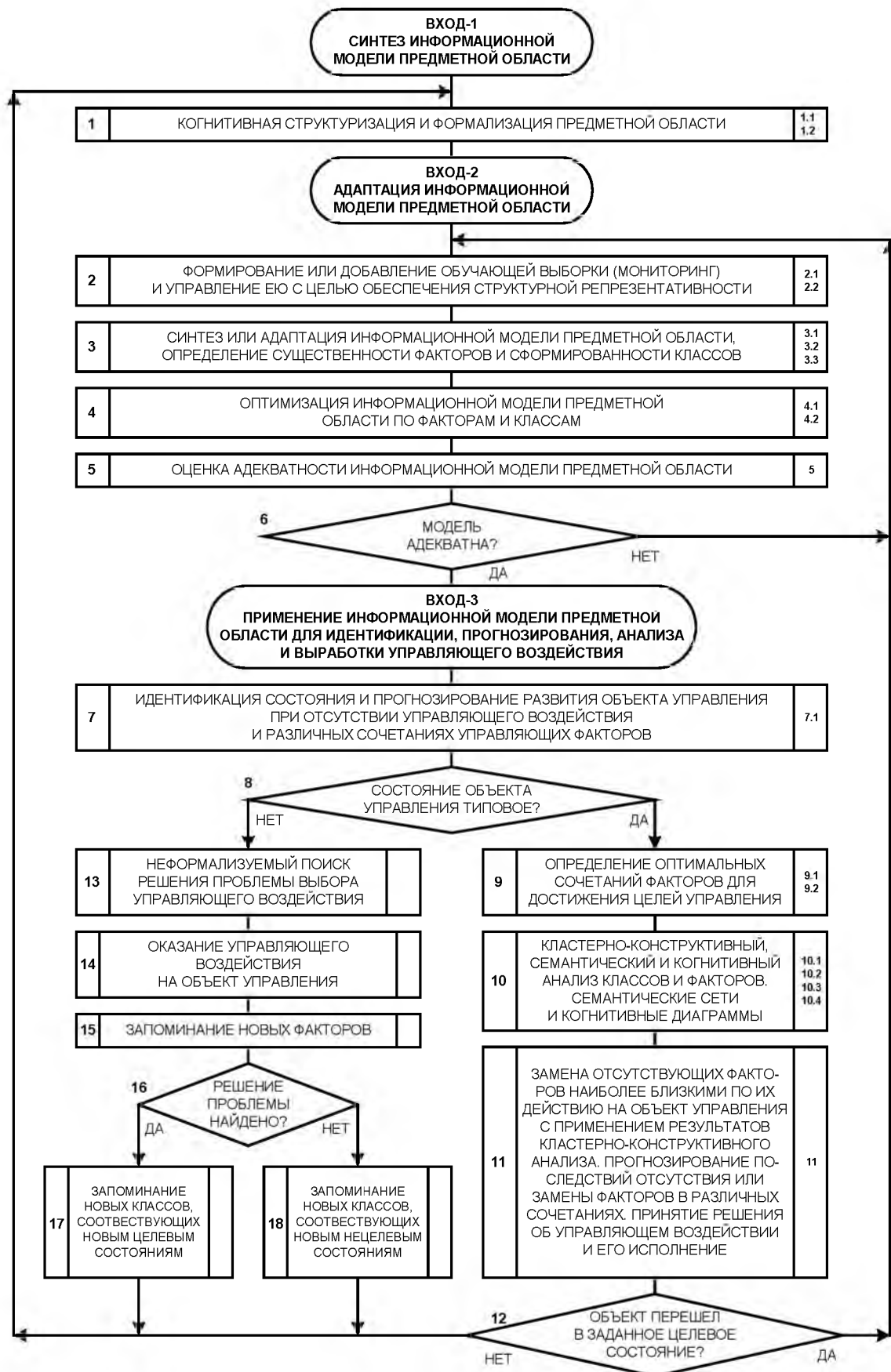


Рисунок 11. Схема системного анализа, ориентированного на интеграцию с когнитивными технологиями

1.2.1.2. Когнитивная концепция и синтез когнитивного конфигуратора

В данном разделе приводится когнитивная концепция, разработанная автором исследования в 1996 году [125], с учетом двух основных требований:

1. Адекватное отражение в когнитивной концепции реальных процессов, реализуемых человеком в процессах познания.
2. Высокая степень приспособленности когнитивной концепции для формализации в виде достаточно простых математических и алгоритмических моделей, допускающих прозрачную программную реализацию в автоматизированной системе.

1.2.1.2.1. Понятие когнитивного конфигуратора и необходимость естественнонаучной (формализуемой) когнитивной концепции

1.2.1.2.1.1. Определение понятия конфигуратора

Понятие конфигуратора, по-видимому, впервые предложено В.А.Лефевром [194], хотя безусловно это понятие использовалось и раньше, но, во-первых, оно не получало самостоятельного названия, а, во-вторых, использовалось в частных случаях и не получало обобщения. *Под конфигуратором В.А.Лефевр понимал минимальный полный набор понятийных шкал или конструктов, достаточный для адекватного описания предметной области.* Примеры конфигураторов приведены в [194].

1.2.1.2.1.2. Понятие когнитивного конфигуратора

В исследованиях по когнитивной психологии изучается значительное количество различных операций, связанных с процессом познания. Однако, насколько известно из литературы, психологами не ставился вопрос о выделении из всего множества когнитивных операций такого минимального (базового) набора наиболее элементарных из них, из которых как составные могли бы строиться другие операции. Ясно, что для выделения таких базовых когнитивных операций (БКО) необходимо построить их иерархическую систему, в фундаменте которой будут находиться наиболее элементарные из них, на втором уровне – производные от них, обладающие более высоким уровнем интегративности, и т.д.

Таким образом, *под когнитивным конфигуратором будем понимать минимальный полный набор базовых когнитивных операций, достаточный для представления различных процессов познания.*

1.2.1.2.1.3. Когнитивные концепции и операции

Проведенный анализ когнитивных концепций показал, что они разрабатывались ведущими психологами (Пиаже, Солсо, Найсер) без учета требований, связанных с их дальнейшей формализацией и автоматизацией. Поэтому имеющиеся концепции когнитивной психологии слабо подходят для этой цели; в ко-

гнитивной психологии не ставилась и не решалась задача конструирования когнитивного конфигулятора и, соответственно, не сформулировано понятие базовой когнитивной операции.

1.2.1.2.2. Предлагаемая когнитивная концепция

Автоматизировать процесс познания в целом безусловно значительно сложнее, чем отдельные операции процесса познания. Но для этого прежде всего необходимо выявить эти операции и найти место каждой из них в системе или последовательности процесса познания.

Сделать это предлагается в форме когнитивной концепции, которая должна удовлетворять следующим требованиям:

- адекватность, т.е. точное отражение сущности процессов познания, характерных для человека, в частности описание процессов вербализации, семантической адаптации и семантического синтеза (уточнения смысла слов и понятий и включения в словарь новых слов и понятий);
- высокая степень детализации и структурированности до уровня достаточно простых базовых когнитивных операций;
- возможность математического описания, формализации и автоматизации.

Однако приходится констатировать, что даже концепции когнитивной психологии, значительно более конкретные, чем гносеологические, разрабатывались без учета необходимости построения реализующих их математических и алгоритмических моделей и программных систем. Более того, в когнитивной психологии из всего многообразия различных исследуемых когнитивных операций не выделены базовые, к суперпозиции и различным вариантам сочетаний которых сводятся различные процессы познания. Поэтому *для достижения целей данного исследования* концепции когнитивной психологии мало применимы.

В связи с этим в данном исследовании предлагается когнитивная концепция, удовлетворяющая сформулированным выше требованиям. Эта концепция достаточно проста, иначе было бы невозможно ее формализовать, многие ее положения интуитивно очевидны или хорошо известны, тем ни менее в целостном виде она сформулирована лишь в работе [81]. Положения когнитивной концепции приведены в определенном порядке, соответствующем реальному ходу процесса познания "от конкретных эмпирических исходных данных к содержательным информационным моделям, а затем к их верификации, адаптации и, в случае необходимости, к пересинтезу".

На базе выше сформулированных положений автором предложена целостная система взглядов на процесс познания, т.е. **когнитивная концепция** [81] (рисунок 12).

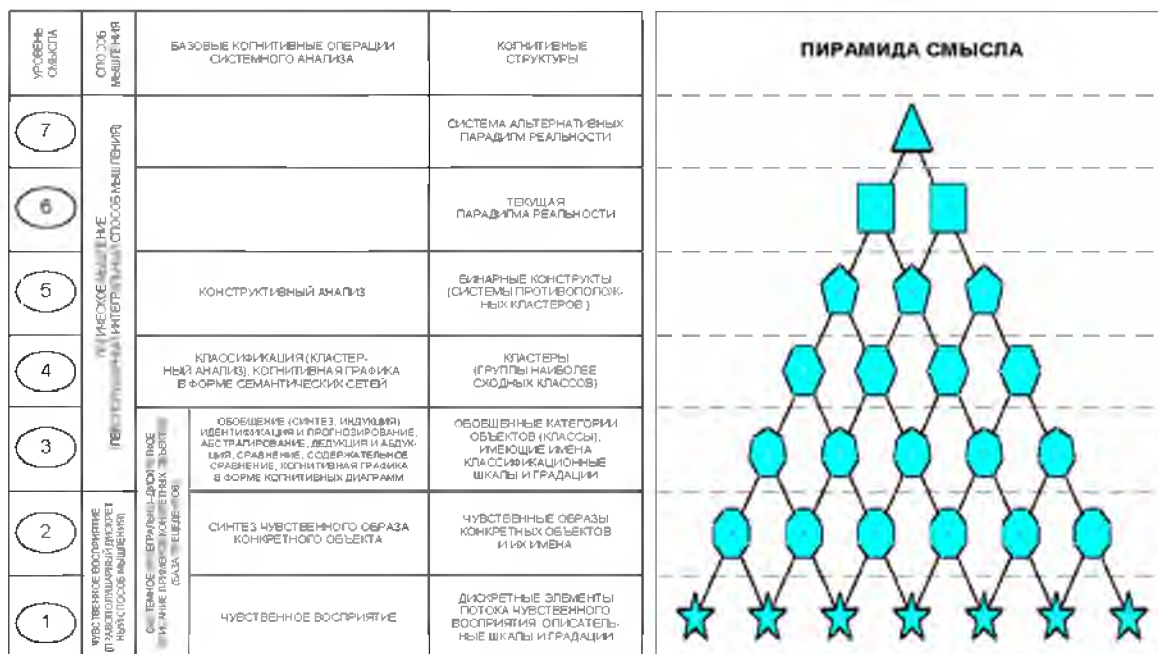


Рисунок 12. Обобщенная схема предлагаемой когнитивной концепции

Суть предложенной когнитивной концепции состоит в том, что процесс познания рассматривается как многоуровневая иерархическая система обработки информации, в которой каждый последующий уровень является результатом интеграции элементов предыдущего уровня. На 1-м уровне этой системы находятся дискретные элементы потока чувственного восприятия, которые на 2-м уровне интегрируются в чувственный образ конкретного объекта. Те, в свою очередь, на 3-м уровне интегрируются в обобщенные образы классов и факторов, образующие на 4-м уровне кластеры, а на 5-м конструкты. Система конструктов на 6-м уровне образуют текущую парадигму реальности (т.е. человек познает мир путем синтеза и применения конструктов). На 7-м же уровне обнаруживается, что текущая парадигма не является единственно-возможной.

Ключевым для когнитивной концепции является понятие **факта**, под которым понимается соответствие дискретного и интегрального элементов познания (т.е. элементов разных уровней интеграции-иерархии), обнаруженное на опыте. Факт рассматривается как квант смысла, что является основой для его формализации. Таким образом, происхождение смысла связывается со своего рода "*разностью потенциалов*", существующей между смежными уровнями интеграции-иерархии обработки информации в процессах познания.

1. Процесс познания начинается с чувственного восприятия. Различные органы восприятия дают качественно-различную чувственную информацию в форме дискретного потока *элементов восприятия*. Эти элементы формализуются с помощью описательных шкал и градаций.

2. В процессе накопления опыта выявляются взаимосвязи между элементами чувственного восприятия: одни элементы часто наблюдаются с другими (имеет место их пространственно-временная корреляция), другие же вместе встречаются достаточно редко. Существование устойчивых связей между элементами восприятия говорит о том, что они отражают некую реальность, *интегральную* по отношению к этим элементам. Эту реальность будем называть объ-

ектами восприятия. Рассматриваемые в *единстве* с объектами элементы восприятия будем называть *признаками* объектов. Таким образом, органы восприятия дают чувственную информацию о признаках наблюдаемых объектов, процессов и явлений окружающего мира (объектов). Чувственный образ конкретного объекта представляет собой систему, возникающую как результат *процесса* синтеза признаков этого объекта. В условиях усложненного восприятия синтез чувственного образа объекта может быть существенно замедленным и даже не завершаться в реальном времени.

3. Человек присваивает конкретным объектам названия (имена) и сравнивает объекты друг с другом. При сравнении выясняется, что одни объекты в различных степенях сходны по их признакам, а другие отличаются. Сходные объекты объединяются в обобщенные категории (классы), которым присваиваются имена, производные от имен входящих в категорию конкретных объектов. Классы формализуются с помощью классификационных шкал и градаций и обеспечивают интегральный способ описания действительности. Путем обобщения (синтеза, индукции) информации о признаках конкретных объектов, входящих в те или иные классы, формируются обобщенные образы классов. Накопление опыта и сравнение обобщенных образов классов друг с другом позволяет определить степень характерности признаков для классов, смысл признаков и ценность каждого признака для идентификации конкретных объектов с классами и сравнения классов, а также исключить наименее ценные признаки из дальнейшего анализа без существенного сокращения количества полезной информации о предметной области (абстрагирование). Абстрагирование позволяет существенно сократить затраты внутренних ресурсов системы на анализ информации. Идентификация представляет собой процесс узнавания, т.е. установление соответствия между чувственным описанием объекта, как совокупности дискретных признаков, и неделимым (целостным) именем класса, которое ассоциируется с местом и ролью воспринимаемого объекта в природе и обществе. Дискретное и целостное восприятие действительности поддерживаются как правило различными полушариями мозга: соответственно, правым и левым (доминантность полушарий). Таким образом именно *системное* взаимодействие интегрального (целостного) и дискретного способов восприятия обеспечивает возможность установление *содержательного смысла* событий. При выполнении когнитивной операции "содержательное сравнение" двух классов определяется вклад каждого признака в их сходство или различие.

4. После идентификации уникальных объектов с классами возможна их классификация и присвоение *обобщающих имен* группам похожих классов. Для обозначения группы похожих классов используем понятие "*кластер*". Но и сами кластеры в результате выполнения когнитивной операции "генерация конструктов" могут быть классифицированы по степени сходства друг с другом. Для обозначения системы двух противоположных кластеров, с "спектром" промежуточных кластеров между ними, будем использовать термин "*бинарный конструкт*", при этом сами противоположные кластеры будем называть "*полюса бинарного конструкта*". Бинарные конструкты классов и атрибутов, т.е. конструкты с двумя полюсами, наиболее типичны для человека и представляет собой когнитивные структуры, играющие огромную роль в процессах познания. Достаточно

сказать, что *познание можно рассматривать как процесс генерации, совершенствования и применения конструктов*. Качество конструкта тем выше, чем сильнее отличаются его полюса, т.е. чем больше диапазон его смысла.

Результаты идентификации и прогнозирования, осуществленные с помощью модели, путем выполнения когнитивной операции "верификация" сопоставляются с опытом, после чего определяется целесообразность выполнения когнитивной операции "обучение". При этом может возникнуть три основных варианта, которые на рисунке 13 обозначены цифрами:



Рисунок 13. К пояснению смысла понятий:

**"Адаптация и синтез когнитивной модели предметной области",
"Внутренняя и внешняя валидность информационной модели",**

1. Объект, входит в обучающую выборку и достоверно идентифицируется (внутренняя валидность, в адаптации нет необходимости).

2. Объект, не входит в обучающую выборку, но входит в исходную генеральную совокупность, по отношению к которой эта выборка репрезентативна, и достоверно идентифицируется (внешняя валидность, добавление объекта к обучающей выборке и адаптация модели приводит к количественному уточнению смысла признаков и образов классов).

3. Объект не входит в исходную генеральную совокупность и идентифицируется недостоверно (внешняя валидность, добавление объекта к обучающей выборке и синтез модели приводит к качественному уточнению смысла признаков и образов классов, исходная генеральная совокупность расширяется).

1.2.1.2.3. Когнитивный конфигуратор и базовые когнитивные операции системного анализа

Таким образом из предложенной когнитивной концепции вытекает существование по крайней мере 10 базовых когнитивных операций системного анализа (БКОСА) (таблица 4):

**Таблица 4 – ОБОБЩЕННЫЙ СПИСОК БКОСА
(КОГНИТИВНЫЙ КОНФИГУРАТОР)**

№ п.п.	Код БКОСА по схеме АСК-анализа	Полное наименование БКОСА
1.	1.	Присвоение имен классам и атрибутам (интенциональная и экстенциональная репрезентация)
2.	2.	Восприятие
3.	3.	Обобщение (синтез, индукция)
4.	4.	Абстрагирование классов и атрибутов
5.	5.	Оценка адекватности модели
6.	7.	Сравнение, идентификация и прогнозирование
7.	9.	Анализ (дедукция и абдукция) классов и атрибутов
8.	10.1, 10.2.	Классификация и генерация конструкторов классов и атрибутов
9.	10.3, 10.4	Содержательное сравнение классов и атрибутов
10.	11.	Планирование и принятие решений о применении системы управляющих факторов

Необходимо отметить, что классификация операций системного анализа по В.М.Казиеву [81] ближе всего к позиции, излагаемой в данной работе, т.к. этим автором названы 6 из 10 базовых когнитивных операций системного анализа: формализация; синтез (индукция); абстрагирование; анализ (дедукция); распознавание, и идентификация образов; классификация. Вместе с тем им не приводятся математическая модель, алгоритмы и инструментарий реализации этих операций и не ставится задача их разработки, кроме того некоторые из них приведены дважды под разными названиями, например: анализ и синтез это *тоже самое*, что дедукция и индукция (таблица 4).

Необходимо также отметить, что по-видимому, впервые идея сведения мышления и процессов познания к когнитивным операциям была четко и осознанно сформулирована в письменном виде в V веке до н.э.: *"Сущность интеллекта проявляется в способностях обобщения, абстрагирования, сравнения и классификации"* (цит.по пам., Патанжали, Йога-Сутра, авт.).

Познание предметной области с одной стороны безусловно является фундаментом, на котором строится все грандиозное здание системного анализа, а с другой стороны, процессы познания являются связующим звеном, органично объединяющим "блоки" принципов и методов системного анализа в стройное здание. Более того, процессы познания буквально пронизывают все методы и принципы системного анализа, входя в них как один из самых существенных элементов.

Однако, на этом основании неверным будет представлять, что когнитивные операции являются подмножеством понятия "системный анализ", скорее наоборот: *системный анализ представляет собой один из теоретических методов познания, представимый в форме определенной последовательности когнитивных операций, тогда как другие последовательности этих операций позволяют образовать другие формы теоретического познания.*

1.2.1.2.4. Задачи формализации базовых когнитивных операций системного анализа

Для решения задачи формализации БКОСА необходимо решить следующие задачи:

1. Выбор единой интерпретируемой численной меры для классов и атрибутов.

2. Выбор неметрической меры сходства объектов в семантических пространствах.

4. Определение идентификационной и прогностической ценности атрибутов.

5. Ортонормирование семантических пространств классов и атрибутов (Парето-оптимизация).

Выбор единой интерпретируемой численной меры для классов и атрибутов

При построении модели объекта управления одной из принципиальных проблем является выбор формализованного представления для индикаторов, критериев и факторов (далее: факторов). Эта проблема распадается на две подпроблемы:

1. Выбор и обоснование смысла выбранной численной меры.

2. Выбор математической формы и способа определения (процедуры, алгоритма) количественного выражения для значений, отражающих степень взаимосвязи факторов и будущих состояний АОУ.

Рассмотрим требования к численной мере, определяемые существом подпроблем. Эти требования вытекают из необходимости совершать с численными значениями факторов математические операции (сложение, вычитание, умножение и деление), что в свою очередь необходимо для построения полноценной математической модели.

Требование 1: из формулировки 1-й подпроблемы следует, что все факторы должны быть приведены к некоторой общей и универсальной для всех факторов единице измерения, имеющей какой-то смысл, причем смысл, поддающийся единой сопоставимой в пространстве и времени интерпретации.

Традиционно в специальной литературе [12] рассматриваются следующие смысловые значения для факторов: стоимость (выигрыш-проигрыш или прибыль-убытки); полезность; риск; корреляционная или причинно-следственная взаимосвязь. Иногда предлагается использовать безразмерные меры для факторов, например эластичность, однако, этот вариант не является вполне удовлетворительным, т.к. не позволяет придать факторам содержательный и сопоставимый смысл и получить *содержательную интерпретацию* выводов, полученных на основе математической модели.

Таким образом, возникает ключевая при выборе численной меры проблема выбора смысла, т.е. по сути единиц измерения, для индикаторов, критериев и факторов.

Требование 2: высокая степень адекватности предметной области.

Требование 3: высокая скорость сходимости при увеличении объема обучающей выборки.

Требование 4: высокая независимость от артефактов.

Что касается конкретной математической формы и процедуры определения числовых значений факторов в выбранных единицах измерения, то обычно применяется метод взвешивания экспертных оценок, при котором эксперты предлагают свои оценки, полученные как правило неформализованным путем. При этом сами эксперты также обычно ранжированы по степени их компетентности. Фактически при таком подходе числовые значения факторов является не

определяемой, искомой, а исходной величиной. Иначе обстоит дело в факторном анализе, но в этом методе, опять же на основе экспертных оценок важности факторов, требуется предварительно, т.е. перед проведением исследования, принять решение о том, какие факторы исследовать (из-за жестких ограничений на размерность задачи в факторном анализе). Таким образом оба эти подхода реализуемы при относительно небольших размерностях задачи, что с точки зрения достижения целей настоящего исследования, является недостатком этих подходов.

Поэтому самостоятельной и одной из ключевых проблем является обоснованный и удачный выбор математической формы для численной меры индикаторов и факторов.

Эта математическая форма с одной стороны должна удовлетворять предыдущим требованиям, прежде всего требованию 1, а также должна быть процедурно вычислимой, измеримой.

Выбор неметрической меры сходства объектов в семантических пространствах

Существует большое количество мер сходства, из которых можно было бы упомянуть скалярное произведение, ковариацию, корреляцию, евклидово расстояние, расстояние Махаланобиса и др.

Проблема выбора меры сходства состоит в том, что при выбранной численной мере для координат классов и факторов она должна удовлетворять определенным критериям:

1. Обладать высокой степенью адекватности предметной области, т.е. высокой валидностью, при различных объемах выборки, как при очень малых, так и при средних и очень больших.
2. Иметь обоснованную, четкую, ясную и интуитивно понятную интерпретацию.
3. Быть нетрудоемкой в вычислительном отношении.
4. Обеспечивать корректное вычисление меры сходства для пространств с неортонормированным базисом.

Определение идентификационной и прогностической ценности атрибутов

Не все факторы имеют одинаковую ценность для решения задач идентификации, прогнозирования и управления. Традиционно считается, что факторы имеют одинаковую ценность только в тех случаях (обычно в психологии), когда определить их действительную ценность не представляется возможным по каким-либо причинам.

Для достижения целей, поставленных в данном исследовании, необходимо решить проблему определения ценности факторов, т.е. разработать математическую модель и алгоритм, которые допускают программную реализацию и обеспечивают на практике определение идентификационной и прогностической ценности факторов.

Ортонормирование семантических пространств классов и атрибутов (Парето-оптимизация)

Если не все факторы имеют одинаковую ценность для решения задач идентификации, прогнозирования и управления, то возникает проблема исключения из системы факторов тех из них, которые не представляют особой ценно-

сти. Удаление малоценных факторов вполне оправданно и целесообразно, т.к. сбор и обработка информации по ним в среднем связана с такими же затратами времени, вычислительных и информационных ресурсов, как и при обработке ценных факторов. В этом состоит идея Парето-оптимизации. Однако это удаление должно осуществляться при вполне определенных граничных условиях, характеризующих результирующую систему: адекватность модели; количество признаков на класс; суммарное количество градаций признаков в описательных шкалах. В противном случае удаление факторов может отрицательно сказываться на качестве решения задач. На практике проблема реализации Парето-оптимизации состоит в том, что факторы вообще говоря коррелируют друг с другом и поэтому их ценность может изменяться при удалении любого из них, в том числе и наименее ценного. Поэтому просто взять и удалить наименее ценные факторы не представляется возможным и необходимо разработать корректный итерационный вычислительный алгоритм обеспечивающий решение этой проблемы при заданных граничных условиях.

1.2.1.3. СК-анализ, как системный анализ, структурированный до уровня базовых когнитивных операций

В предложенной схеме системного анализа (рисунок 11) наглядно прослеживается сходство с когнитивным анализом (рисунок 9). Это естественно, т.к. системный анализ рассматривается многими авторами, как одна из форм теоретического познания. Учитывая это и *с целью создания благоприятных условий для дальнейшей декомпозиции системного анализа до уровня, достаточного для разработки алгоритмов и программной реализации, предлагается структурировать системный анализ до уровня базовых когнитивных операций*. Предлагается рассматривать системный анализ, структурированный до уровня базовых когнитивных операций, как системно-когнитивный анализ (СК-анализ). Насколько известно впервые понятие "СК-анализ" предложено в 1995 году А.Е.Кибрик и Е.А.Богдановой. Однако этими авторами данное понятие было введено в другой предметной области, ими не ставилась и не решалась задача автоматизации СК-анализа.

В связи с тем, что СК-анализ структурируется нами до уровня БКОСА, его алгоритмизация и последующая автоматизация становится практически решаемой задачей, в отличие от автоматизация непосредственно системного анализа или детализированного системного анализа.

Отсюда органично вытекает возможность *структурирования системного анализа до уровня базовых когнитивных (познавательных) операций*.

Учитывая структуру когнитивного конфигуратора (таблица 4) конкретизируем обобщенную схему этапов системного анализа, ориентированного на интеграцию с когнитивными технологиями (рисунок 11), в результате чего получим обобщенную схему этапов СК-анализа (рисунок 14).

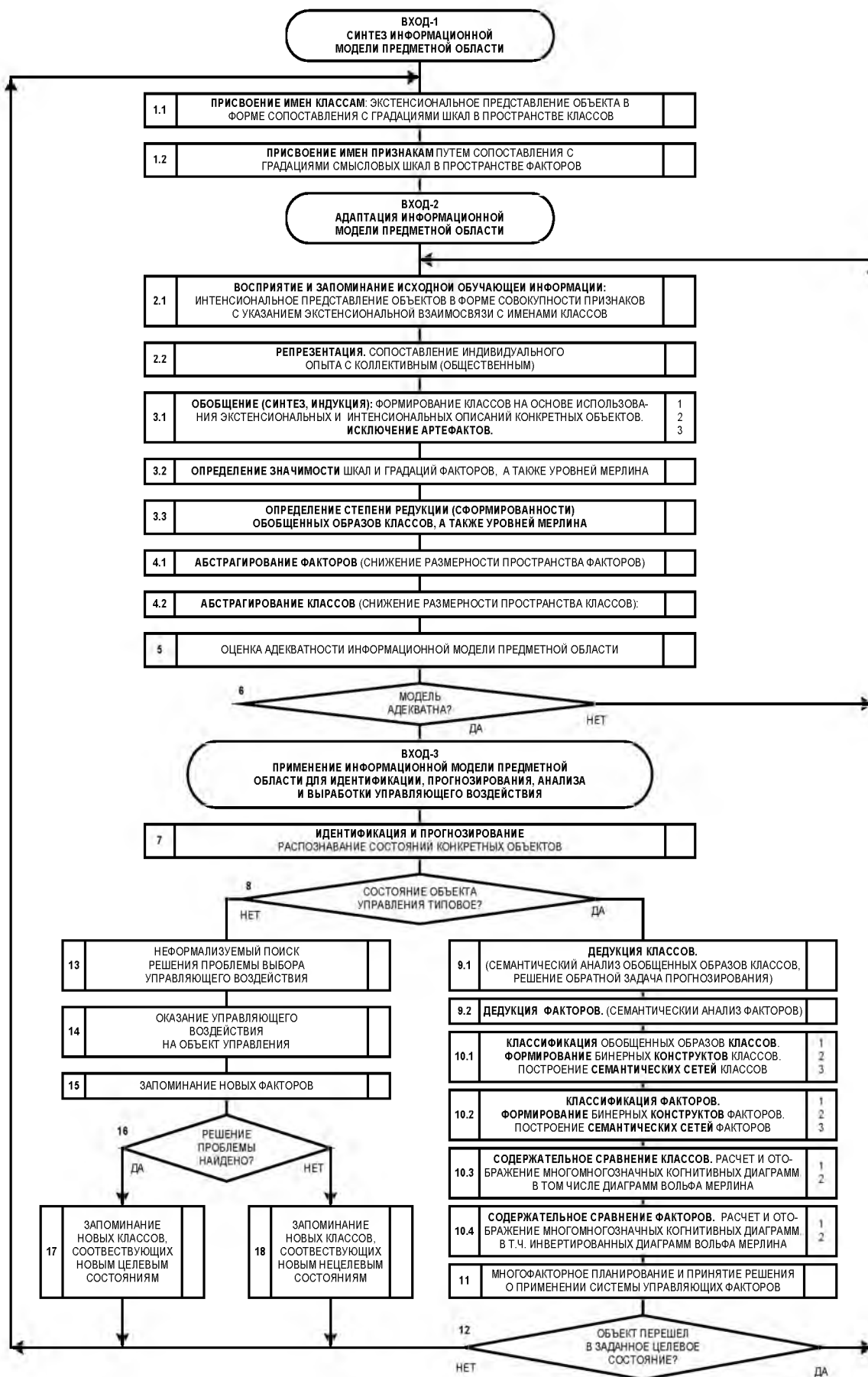


Рисунок 14. Обобщенная схема этапов СК-анализа

Предложенная схема представляет собой схему системного анализа, структурированного до уровня базовых когнитивных операций, который предлагается называть "Системно-когнитивным анализом (СК-анализ)". Нумерация блоков на рисунке 14 соответствует этапам СА на рисунке 11.

Схема, СК-анализа, представленная на рисунке 14, определяет место каждой базовой когнитивной операции в системном анализе.

1.2.1.4. Место и роль СК-анализа и систем искусственного интеллекта в структуре АСУ

Управление в АПК рассматривается в данной работе в контексте использования автоматизированных систем управления в этой области. Поэтому в данном разделе предложена классификация функционально-структурных типов АСУ и показано место адаптивных АСУ сложными системами и рефлексивных АСУ активными объектами в этой классификации; показаны место и роль СК-анализа в рефлексивных АСУ активными объектами [81].

1.2.1.4.1. Структура типовой АСУ

Цель применения АСУ обычно можно представить в виде некоторой суперпозиции трех подцелей:

- 1) стабилизация состояния объекта управления в динамичной или агрессивной внешней среде;
- 2) перевод объекта в некоторое конечное (целевое) состояние, в котором он приобретает определенные заранее заданные свойства;
- 3) повышение качества функционирования АСУ (синтез новых моделей и их адаптация).

Обычно АСУ рассматривается как система, состоящая из двух основных подсистем: управляющей и управляемой, т.е. из субъекта и объекта управления (рисунок 15).

Как правило, АСУ действует в определенной окружающей среде, которая является общей и для субъекта, и для объекта управления.

Граница между тем, что считается окружающей средой, и тем, что считается объектом управления относительно и определяется возможностью управляющей системы оказывать на них воздействие: на объект управления управляющее воздействие может быть оказано, а на среду нет.

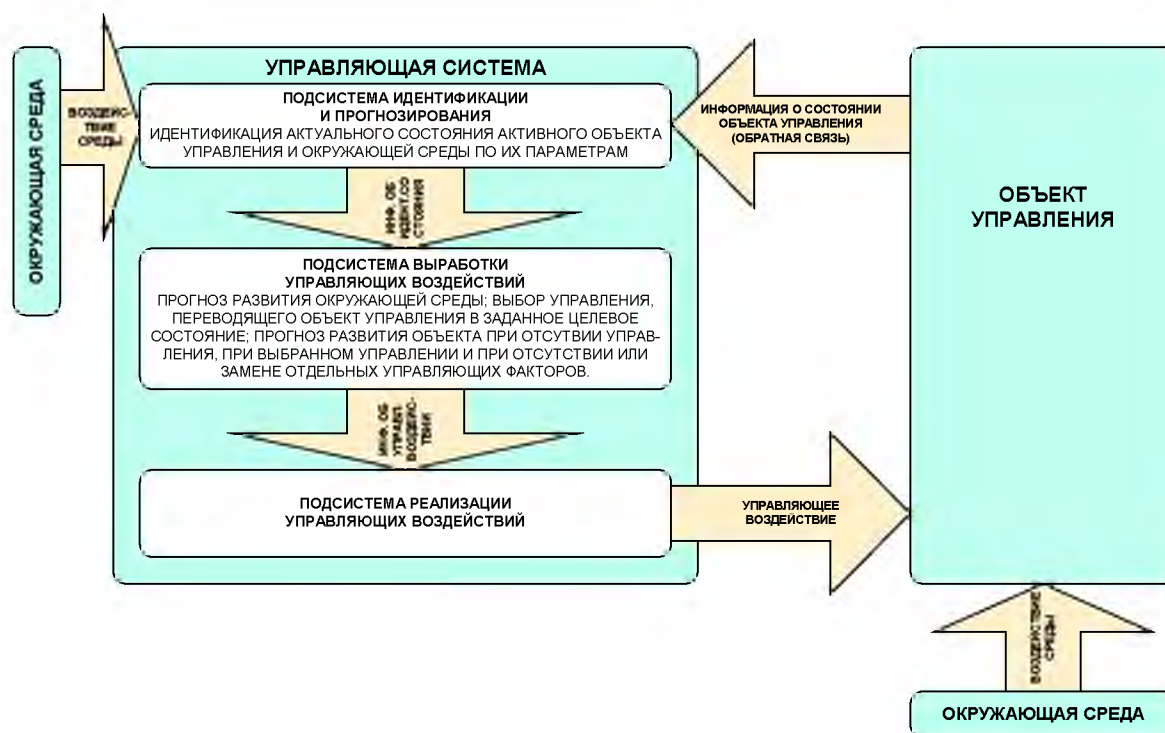


Рисунок 15. Структура типовой АСУ

1.2.1.4.2. Параметрическая модель адаптивной АСУ сложными системами

Конкретизируем типовую структуру АСУ (рисунок 15), используя классификацию входных и выходных параметров объекта управления. В результате получим параметрическую модель адаптивной АСУ сложными системами (рисунок 16).

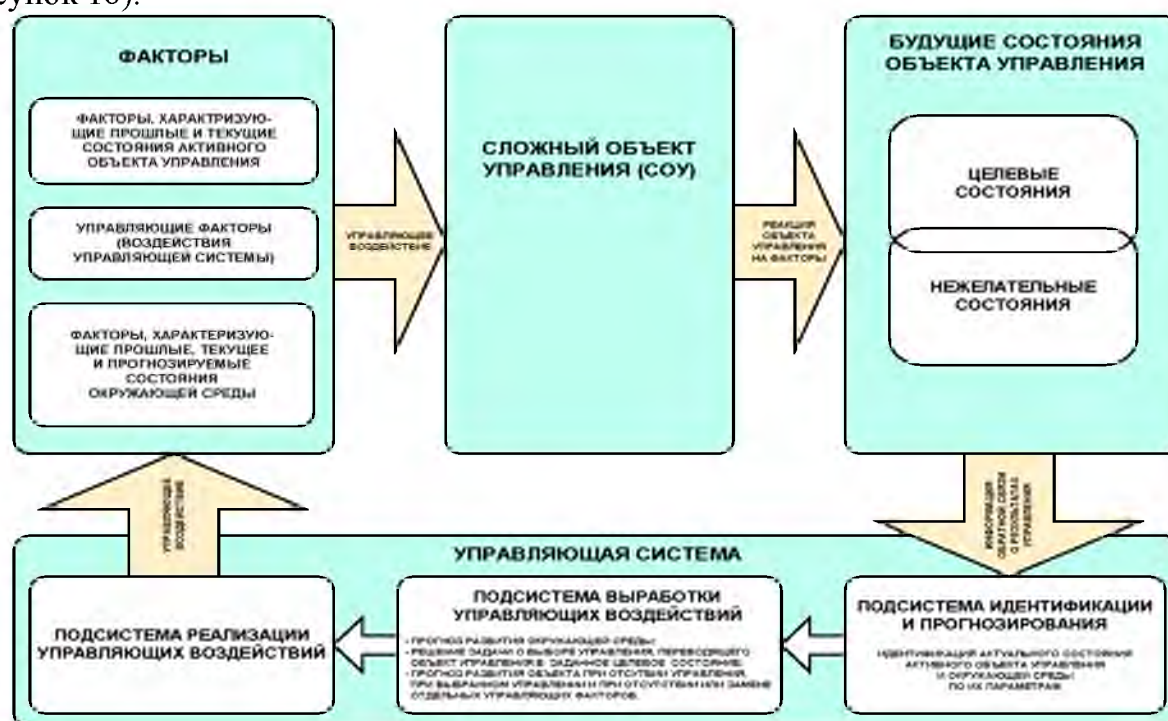


Рисунок 16. Параметрическая модель адаптивной АСУ сложными системами

Входные параметры (факторы) делятся на три группы: характеризующие предысторию и текущее состояние объекта управления, управляющие (технологические) факторы и факторы окружающей среды.

Выходные параметры – это свойства объекта управления, зависящие от входных параметров (в т.ч. параметров, характеризующих среду). В автоматизированных системах параметрического управления целью управления является получение определенных значений выходных параметров объекта управления, т.е. перевод объекта управления в заданное целевое состояние.

Однако, в случае сложного объекта управления (СОУ) его выходные параметры связаны с состоянием сложным и неоднозначным (нечетким) способом. Поэтому возможность параметрического управления сложными объектами является проблематичной и вводится более общее понятие "управление по состоянию СОУ".

Для ААСУ СС выполняется принцип соответствия, т.е. в предельном случае, когда связь выходных параметров и состояний объекта управления имеет однозначный и детерминистский характер, управление по состояниям сводится к управлению по параметрам и функции ААСУ СС сводится к их подмножеству: т.е. к функциям типовой АСУ. Однако, когда состояние объекта управления связано с его параметрами сложным и неоднозначным образом, возникает задача идентификации состояния СОУ по его выходным параметрам, которая решается подсистемой идентификации управляющей подсистемы, работающей на принципах адаптивного распознавания образов. При этом классами распознавания являются текущие состояния сложного объекта управления, а признаками – его выходные параметры.

Подсистема выработки управляющих воздействий, также основанная на алгоритмах распознавания образов, решает следующие задачи: прогноз развития окружающей среды; прогноз развития объекта управления в условиях отсутствия управляющих воздействий ("движение по инерции"); выбор управления, переводящего объект управления в целевое состояние.

Подсистема реализации управляющих воздействий осуществляет выбранное технологическое воздействие на объект управления.

1.2.1.4.3. Модель рефлексивной АСУ активными объектами и понятие мета-управления

АСУ активными объектами (объектами) (РАСУ АО), является обобщением ААСУ СС на случай, когда сложная система является активной, т.е. имеет собственные цели, которые в общем случае не совпадают с целями управляющей системы. Из этого обстоятельства следует, что активный объект управления (АОУ) имеет собственную модель себя и *своей* окружающей среды, включая и управляющую систему, как один из ее элементов.

Классификация различных уровней рефлексивности приведена в таблице 5.

Таблица 5 – УРОВНИ РЕФЛЕКСИВНОСТИ

Уровень рефлексии	Активный объект управления	Рефлексивная АСУ
0-й	ОУ не имеет модели окружающей среды и внешней системы управления	Нерефлексивная адаптивная модель ОУ
1-й	Модель окружающей среды, сформированная АОУ, включает модель АСУ, как адаптивной АСУ (ААСУ)	Модель АОУ, сформированная в РАСУ включает модель своего отражения в модели, сформированной АОУ
2-й	Модель РАСУ, сформированная АОУ, учитывает, что РАСУ обладает 1-м уровнем рефлексии	Модель АОУ, сформированная в РАСУ, учитывает, что АОУ обладает 2-м уровнем рефлексии
***	***	***
n-й	Модель РАСУ, сформированная АОУ, учитывает, что РАСУ обладает n-1 уровнем рефлексии	Модель АОУ, сформированная в РАСУ, учитывает, что АОУ обладает n-м уровнем рефлексии

Простейшая модель АОУ включает два уровня (рисунок 17) и предполагает возможность оказания управляющих воздействий на различных уровнях АОУ:

- уровне воздействия на систему поддержки системы управления;
- уровне системы управления.



Рисунок 17. Двухуровневая модель активной системы и различие в характере управляющего воздействия на АОУ в РАСУ ОУ и ААСУ СС

Различия между ААСУ СС и РАСУ АО приведены в таблице 6:

Таблица 6 – РАЗЛИЧИЯ МЕЖДУ ААСУ СС И РАСУ АО

	ААСУ СС	РАСУ АО
Модель объекта управления	Объект управления рассматривается как физическая система, пассивно воспринимающая управляющая воздействия	Объект управления рассматривается как субъект, имеющий системы: целеполагания; моделирования себя (рефлексивность) и окружающей среды (включая управляющую систему); принятия и реализации решений
Характер управляющего воздействия	Энергетическое (физическое) воздействие	Информационное воздействие, мета-управление

Конечно, РАСУ АО не исключает возможности энергетического воздействия на физическую структуру АОУ, как в ААСУ СС, но это также может осуществляться с учетом характеристик его интеллектуальной информационной системы.

Таким образом, в РАСУ АО управление АОУ осуществляется путем управления его системой управления, т.е. путем мета-управления: согласования целей системы управления и активного объекта управления; создания у активного объекта управления благоприятного для достижения целей управления и восприятия управляющих воздействий образа управляющей системы; создания у активного объекта управления мотивации, ориентирующей его на достижение целей управления. Таким образом, мета-управление представляет собой управление теми условиями, на основе которых активный объект управления формирует цели и принимает решения. Учитывая сказанное, получим структуру РАСУ АО как обобщение структуры ААСУ СС на случай активного объекта управления (рисунок 18).

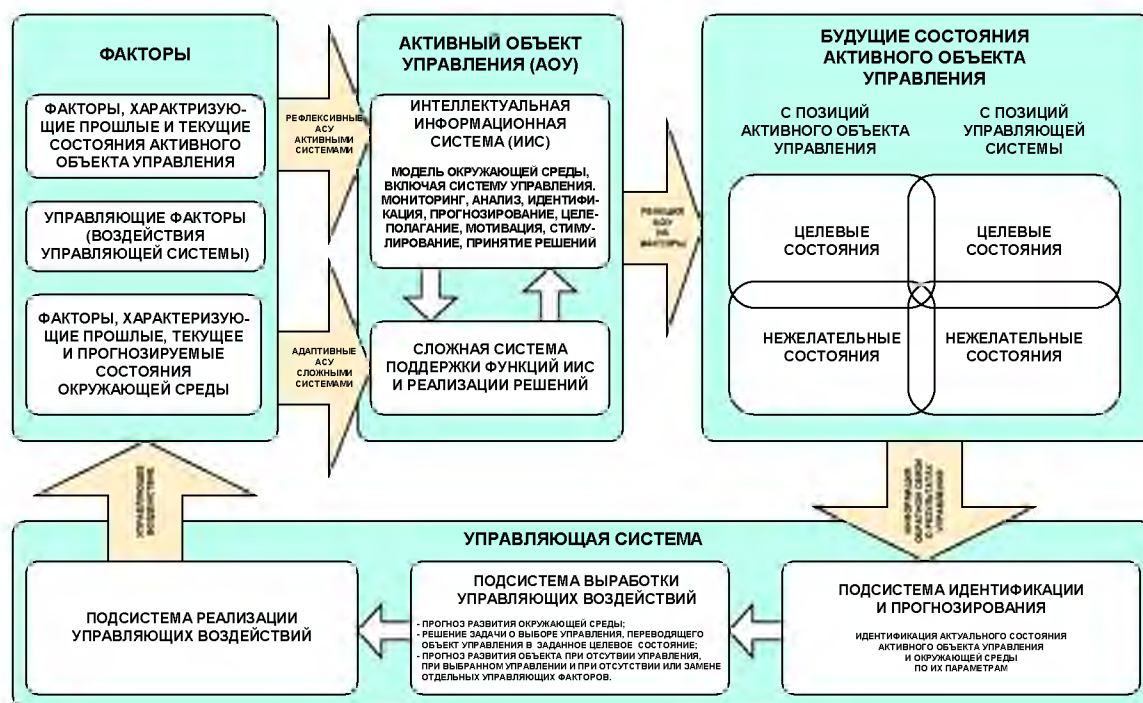


Рисунок 18. Модель рефлексивной АСУ активными объектами (системами)

1.2.1.4.4. Двухконтурная модель РАСУ в АПК

Концепция рефлексивной АСУ в АПК и технология QFD (технология развертывания функций качества)

Чтобы сформулировать концепцию управления В РАСУ АПК рассмотрим упрощенную формальную модель. Процесс управления состоит из последовательных циклов управления, каждый из которых включают следующие этапы: количественное сопоставимое измерение параметров и идентификация состояния объекта управления; оценка эффективности (качества) предыдущего управляющего воздействия; если предыдущее управляющее воздействие не обеспечило приближения цели, то выработка новых или корректировка (адаптация) имеющихся методов принятия решений; иначе – выработка нового управляющего воздействия на основе имеющихся методов принятия решений; реализация управляющего воздействия.

При этом объектами управления, в соответствии с технологией QFD (развертывания функций качества) на различных уровнях являются потребительские свойства продукта, свойства его компонент, технологический процесс и его элементы (операции) (рисунок 19) [81].

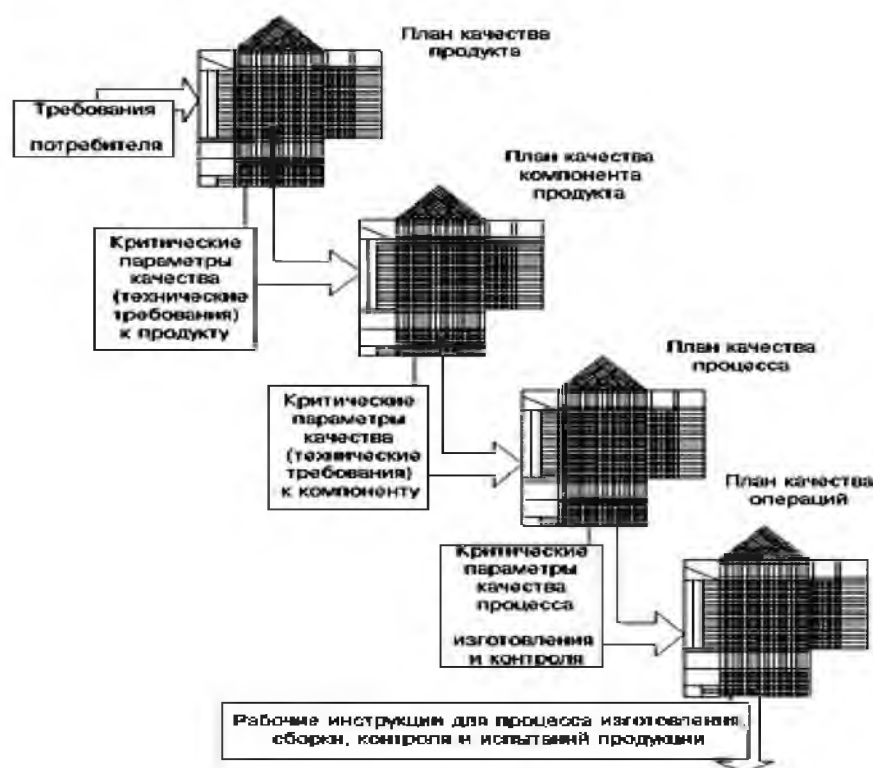


Рисунок 19. Обобщенная схема QFD-технологии (развертывание функций качества)

Рефлексивная АСУ АПК группы Б: 1-й контур: "Агротехнологии – конечный продукт"

Конкретизируем общие положения QFD-технологии (развертывание функций качества) для случая РАСУ АПК. Из этой технологии следует, что в этой РАСУ должно быть по крайней мере два уровня:

1-й уровень – управление производством конечной продукции:

2-й уровень – управление качеством технологии производства конечной продукции.

Такие АСУ, которые управляют производством конечного продукта, будем называть АСУ группы "Б" (АСУ средств потребления). Применительно к РАСУ АПК, АСУ группы "Б" – это АСУ управления производством сельхозпродукции с помощью агротехнологий (рисунок 20).

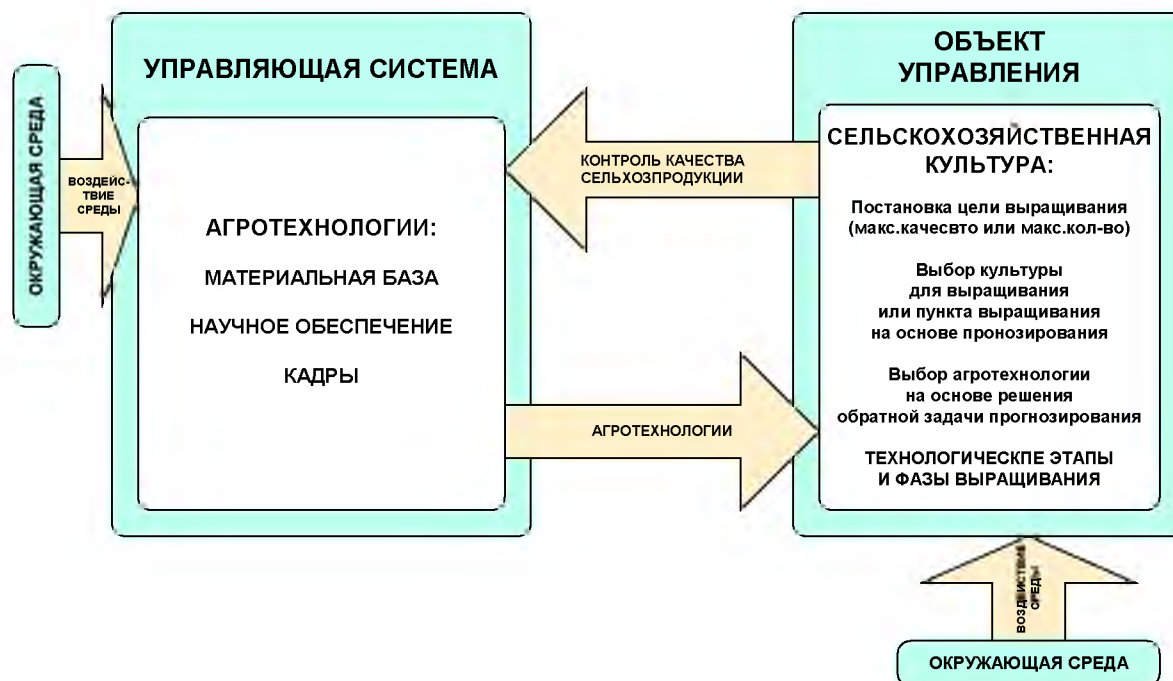


Рисунок 20. Обобщенная схема АСУ АПК группы "Б"

Обычно считается известным влияние тех или иных традиционных агротехнологий на потребительские свойства конечного продукта и его цену. Это положение не подвергается в данной работе сомнению, однако необходимо отметить, что само понятие "известно" существенно отличается в гуманитарной и технических областях, т.е. в этих областях приняты различные *критерии* для классификации исследуемых закономерностей на "известные" и "неизвестные". Это приводит к тому, что в ряде случаев то, что "гуманитарии" считают для себя известным не является таковым для "естественников", т.е. они, конечно, имеют эти знания, но они их не устраивают. Как правило гуманитариев устраивает **качественная** оценка связи, в результате они часто оперируют *нечеткими высказываниями* типа: "Бобовые предшественники приводят к повышению урожая зерновых колосовых". И это для них приемлемо. Однако для создания АСУ знаний выраженных в такой форме *недостаточно*, требуется **количественная** формулировка, значит специалист по созданию АСУ будет ставить вопрос о проведении специальных исследований для выявления и количественного измерения силы и направления влияния подобных связей.

Поэтому при создании РАСУ АПК возникают проблемы: количественного измерения различных параметров агротехнологических процессов и окружающей среды и выявления количественных зависимостей между этими параметра-

ми и количественными и качественными характеристиками конечной продукции. Причем характеристики конечной продукции могут быть выражены в интервальных величинах в натуральном или в ценовом выражении.

Во всех случаях внедрение АСУ означает прежде всего изменение (совершенствование) технологии воздействия на объект управления (рисунок 21).

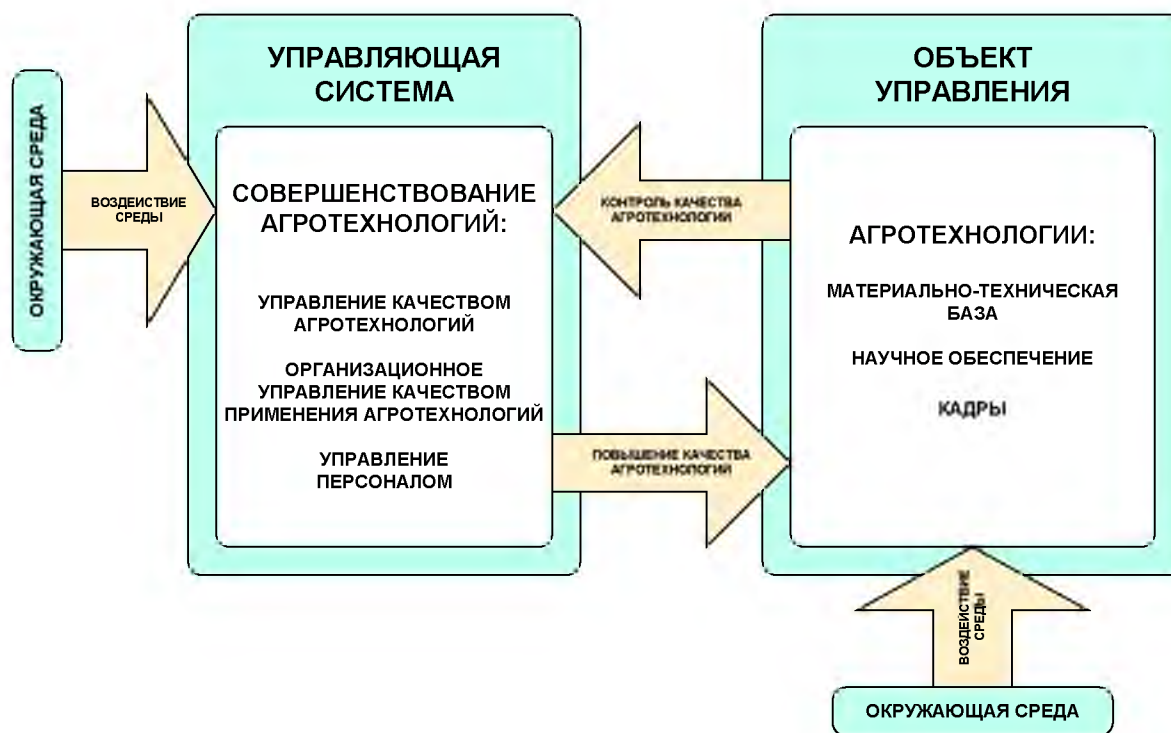


Рисунок 21. Обобщенная схема РАСУ АПК группы "А"

Таким образом, сам процесс внедрения АСУ можно рассматривать как процесс управления совершенствованием технологии производства конечного продукта.

Рефлексивная АСУ АПК группы А: 2-й контур: "Руководство – агротехнологический процесс"

АСУ, в которых сама агротехнология является объектом управления, мы отнесем к группе "А" (таблица 7):

Таблица 7 – КОМПОНЕНТЫ АСУ АГРОТЕХНОЛОГИЯМИ

№	Элементы АСУ	РАСУ АПК
1	Сырье	Агротехнологии и кадры до внедрения РАСУ АПК
2	Объект управления	Агротехнологический процесс и руководящие кадры
3	Управляющие факторы	Материально-техническое и научно-методическое обеспечение агротехнологического процесса, повышение квалификации руководящих кадров
4	Конечный продукт	Агротехнологии и руководящие кадры после внедрения РАСУ АПК
5	Потребитель	Производители сельскохозяйственной продукции
6	Окружающая среда	Рынок труда и агротехнологий

Технические АСУ группы "А" являются чем-то экзотическим, т.к. объект управления, как правило, представляет собой систему с медленноменяющимися параметрами. В этих областях АСУ после внедрения работают достаточно длительное время без существенных изменений.

В РАСУ АПК ситуация иная: и сам объект управления (сельхозкультуры и агротехнологии), и условия окружающей среды (природной, экономической, социальной), являются весьма динамичными, из чего с необходимостью следует и высокая динамичность агротехнологий. Следовательно РАСУ АПК группы "Б" фактически не только не может быть внедрена, но даже и разработана без одновременной разработки и внедрения РАСУ АПК группы "А", которая бы обеспечила ей высокий уровень адаптивности, достаточный для обеспечения поддержки адекватности модели как при количественных, так и при качественных изменениях предметной области. Обобщенная схема РАСУ АПК группы "А" приведена на рисунке 18.

Двухконтурная модель и обобщенная схема рефлексивной АСУ качеством подготовки специалистов

Объединение РАСУ АПК групп "А" и "Б" приводит к схеме двухуровневой РАСУ АПК, в которой первый контур управления включает управление сельхозкультурой, а второй контур управления обеспечивает управление самой агротехнологией. На уровне "А" РАСУ АПК осуществляется разработка и совершенствование агротехнологий, а на уровне "Б" – выбор и использование оптимальной агротехнологии для получения заданных количественных и качественных параметров конечного продукта.

Отметим, что в данной работе рассмотрение ведется на примере плодоводства и растениеводства, но это не является ограничением и легко обобщается на отрасли птицеводства, животноводства, рыбоводства и др.

Но и управление агротехнологиями будет беспредметным без обратной связи, содержащей информацию об эффективности как традиционных агротехнологических методов, так и инноваций, т.е. без учета их влияния на качество хозяйственных результатов.

Кроме того РАСУ АПК включает ряд обеспечивающих систем, работа которых направлена на создание наиболее благоприятных условий для выполнения основной функции РАСУ АПК, т.е. обеспечение максимальной прибыли путем производства и реализации заданного количества и качества наиболее рентабельной продукции. Это так называемые **обеспечивающие** подсистемы: стратегическое управление (включая совершенствование организационной структуры управления); управление инновационной деятельностью (НИР, ОКР, внедрение); управление информационными ресурсами (локальные и корпоративные сети, Internet); управление планово-экономической, финансовой и хозяйственной деятельностью, и др. Необходимо также отметить, что РАСУ АПК работает в определенной окружающей среде, которая, в частности, включает: социально-экономическую среду; рынок труда; рынок агротехнологий; рынок наукоемкой продукции.

Учитывая вышесказанное, предлагается следующая двухуровневая обобщенная модель РАСУ АПК, включающую в качестве базовых подсистем РАСУ АПК групп "А" и "Б", а также обеспечивающие подсистемы (рисунок 22).

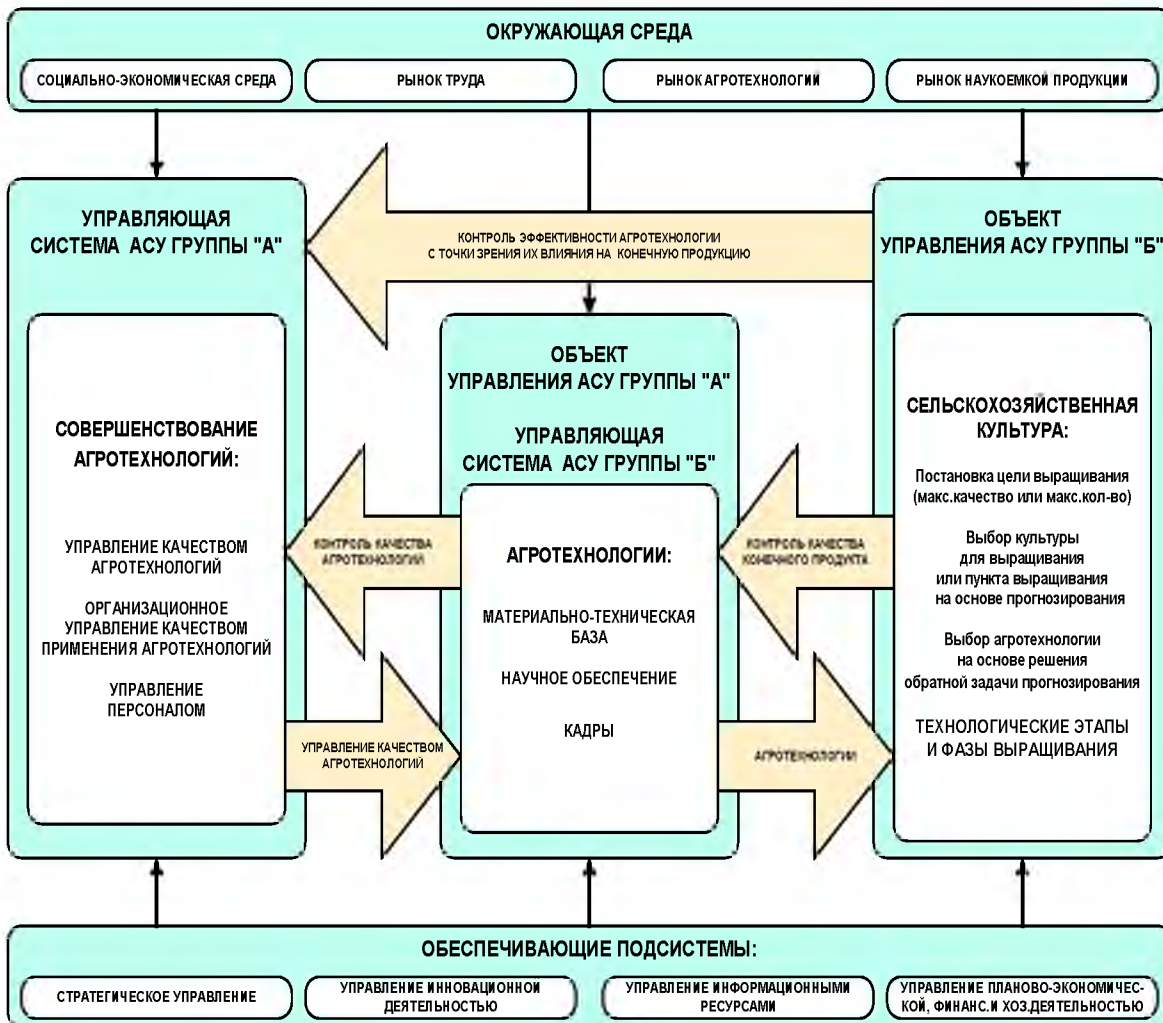
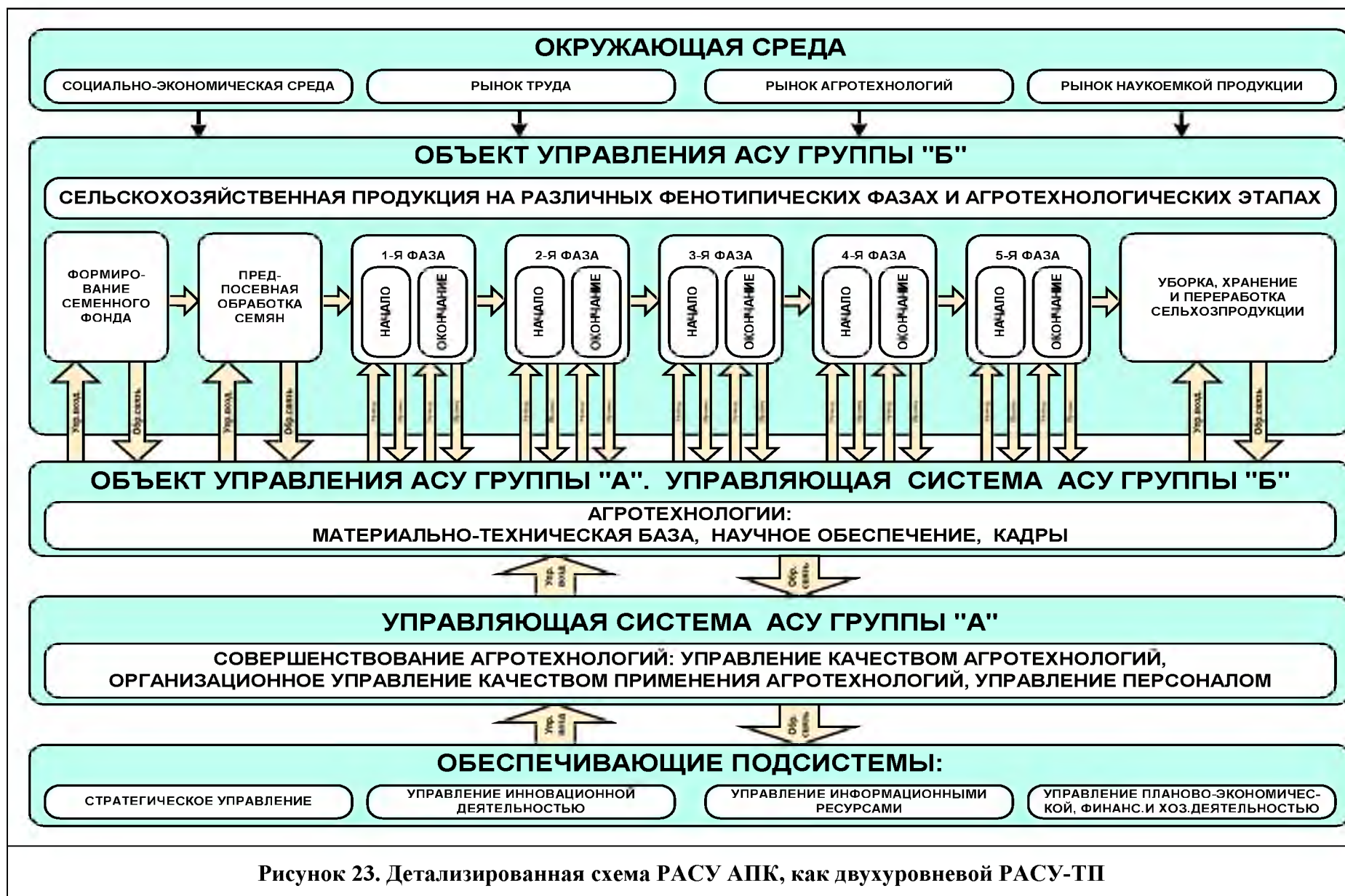


Рисунок 22. Обобщенная схема двухуровневой РАСУ АПК

Необходимо отметить, что двухуровневая схема АСУ АПК является обобщением структуры типовой АСУ, а не обобщением структуры РАСУ АО. Чтобы рассматривать ее именно как рефлексивную АСУ необходимо иметь в виду, что и агротехнологии, и объект управления в АПК, являются активными системами и управляющие воздействия на них имеют информационный характер, т.е. являются метауправляющими. Безусловно, что информационные потоки обуславливают соответствующие финансовые, энергетические и вещественные потоки, изучаемые логистически.

На рисунке 23 представлен вариант двухуровневой АСУ АПК, в котором показаны фазы развития сельскохозяйственной культуры и соответствующие агротехнологические этапы.



Резюме

1. С целью поиска путей автоматизации системного анализа проанализированы различные его варианты, предложенные ведущими учеными в этой области. Показана несостоятельность мнения о том, что автоматизацию системного анализа осуществить тем проще, чем более он детализирован. Отмечена не системность самой этой идеи, противоречащая духу системного анализа.

2. Предложена альтернативная идея поиска путей автоматизации системного анализа на пути его интеграции с когнитивными технологиями. В рамках этой идеи предложено структурировать системный анализ до уровня базовых когнитивных операций, достаточно элементарных, чтобы их было возможно автоматизировать при современном уровне развития науки и техники.

3. Для выявления базовых когнитивных операций разработана формализуемая когнитивная концепция. В ней процесс познания рассматривается как многоуровневая иерархическая система обработки информации, в которой когнитивные структуры каждого последующий уровня является результатом интеграции структур предыдущего уровня:

на 1-м уровне этой системы находятся дискретные элементы потока чувственного восприятия, которые получаются непосредственно от органов чувств и рассматриваются как исходная информация о реальности;

на 2-м уровне дискретные элементы потока чувственного восприятия интегрируются в чувственные образы конкретных объектов и факторов, которым присваиваются конкретные имена;

на 3-м уровне конкретные чувственные образы объектов и факторов интегрируются в обобщенные образы классов и факторов, которым присваиваются обобщенные и символические имена (обобщение и абстрагирование);

на 4-м уровне обобщенные образы классов и факторов сравниваются друг с другом и классифицируются в кластеры;

на 5-м уровне кластеры классов и факторов сравниваются друг с другом и образуют бинарные и многополюсные конструкторы;

на 6-м уровне конструкторы классов и факторов образуют текущую парадигму реальности, формулируется гипотеза о том, что человек познает мир путем синтеза и применения конструкторов;

на 7-м уровне сравниваются текущие парадигмы конкретных людей и их групп, в результате чего обнаруживается, что текущая парадигма не единственно-возможная.

Предложенная когнитивная концепция включает периодическое подтверждение на практике адекватности или неадекватности сформированной модели предметной области, а также ее количественное уточнение с учетом новых достоверных данных (адаптация) или ее качественное переформирование (синтез) в случае необходимости.

4. Понятие факта является ключевым для когнитивной концепции. Под фактом понимается соответствие дискретного и интегрального элементов познания (т.е. элементов разных, как правило смежных, уровней интеграции-иерархии), обнаруженное на опыте. Факт рассматривается как квант смысла, что является основой для математической модели смысла в предложенной семантической информационной модели. Таким образом, происхождение смысла связы-

вается со своего рода "разностью потенциалов", существующей между смежными уровнями интеграции-иерархии обработки информации в процессах познания.

5. В рамках когнитивной концепции сконструирован когнитивный конфигуратор (тер. авт.). Он представляющий собой минимальную полную систему когнитивных операций, названных "базовые когнитивные операции системного анализа" (БКОСА). Выявлено следующих десять БКОСА: 1) присвоение имен; 2) восприятие; 3) обобщение (синтез, индукция); 4) абстрагирование; 5) оценка адекватности модели; 6) сравнение, идентификация и прогнозирование; 7) дедукция и абдукция; 8) классификация и генерация конструкторов; 9) содержательное сравнение; 10) планирование и принятие решений об управлении.

Таким образом, на основе предложенной когнитивной концепции выводятся структура когнитивного конфигулятора, система базовых когнитивных операций и обобщенная схема автоматизированного СА, структурированного до уровня базовых когнитивных операций, получившего в данном исследовании название "СК-анализ".

6. Предложена структура рефлексивной АСУ активными объектами, включающая двухуровневую модель активного объекта управления, классификацию факторов и будущих состояний объекта управления.

Двухуровневая модель активного объекта управления предполагает два типа управляющих воздействий: информационное (мета-управляющее) воздействие на интеллектуальную информационную систему активного объекта управления; энергетическое (силовое) воздействие на сложную систему поддержки функций интеллектуальной информационной системы активного объекта управления.

Классификация факторов, включает: факторы, характеризующие активный объект управления в его прошлых и текущем состояниях, в том числе факторы, характеризующие его как активную, рефлексивную систему; управляющие факторы; факторы окружающей среды.

Будущие состояния активного объекта управления, классифицируются как целевые и нежелательные с позиций управляющей системы и самого активного объекта управления. В общем случае эти классификации не совпадают.

7. Разработаны классификация функционально-структурных типов АСУ и показано место адаптивных АСУ сложными системами и рефлексивных АСУ активными объектами в этой классификации; показаны роль и место СК-анализа в РАСУ АО.

Рассмотрена типовая структура АСУ, предложена параметрическая модель адаптивной АСУ сложными системами (ААСУ СС) и, на основе конкретизации технологии QFD (развертывания функций качества), предложена двухуровневая модель РАСУ АО, являющаяся обобщением ААСУ СС на случай активных объектов управления:

1-й уровень обеспечивает управление АОУ;

2-й уровень – управление технологией воздействия на АОУ.

Проведенное сравнение ААСУ СС и РАСУ АО АПК по способу управляющего воздействия на объект управления, степени управляемости на детерминистских и бифуркационных этапах развития объекта управления и уровню

адаптивности позволяет сделать вывод о предпочтительности эксплуатационных характеристик РАСУ АО при управлении активными объектами. Это обусловлено двумя основными обстоятельствами: в ААСУ СС обеспечивается лишь количественная адаптация модели АОУ, что не обеспечивает сохранение ее адекватности после прохождения объектом управления точки бифуркации, т.е. его качественной трансформации, тогда как в РАСУ АО в этом случае осуществляется повторный синтез модели АОУ; в ААСУ СС рефлексивность и активный характер объекта управления не учитываются и управляющее воздействие на него имеет энергетический (силовой) характер, тогда как в РАСУ АО – это прежде всего воздействие на информационный уровень объекта управления, т.е. мета-управление (коррекция его целей, модели себя и окружающей среды, мотиваций способов принятия и реализации решений в направлении их сближения и согласования с целями управляющей системы). Показано, что в рефлексивных АСУ активными объектами информационное мета-управление активным объектом состоит в коррекции его целей и мотиваций в направлении сближения его целей с целями управляющей системы.

Контрольные вопросы

1. Системный анализ, как метод познания.
2. Принципы системного анализа.
3. Методы и этапы системного анализа.
4. Этапы когнитивного анализа.
5. Обобщенная схема системного анализа, ориентированного на интеграцию с когнитивными технологиями.
6. Когнитивная концепция и синтез когнитивного конфигуратора.
7. Понятие когнитивного конфигуратора и необходимость естественно-научной (формализуемой) когнитивной концепции.
8. Формализуемая когнитивная концепция
9. Когнитивный конфигуратор и БКОСА.
10. Задачи формализации базовых когнитивных операций системного анализа.
11. СК-анализ, как системный анализ, структурированный до уровня базовых когнитивных операций.
12. Место и роль СК-анализа в структуре управления.
13. Структура типовой АСУ
14. Параметрическая модель адаптивной АСУ сложными системами.
15. Модель рефлексивной АСУ активными объектами и мета-управление.
16. Двухконтурная модель РАСУ в АПК.

Рекомендуемая литература

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

1.2.2. ЛЕКЦИЯ-4.

Системная теория информации и семантическая информационная модель

Учебные вопросы

1. Программа системного обобщения математики и предпосылки системной теории информации
2. Теоретические основы системной теории информации.
3. Семантическая информационная модель СК-анализа.
4. Некоторые свойства математической модели (сходимость, адекватность, устойчивость и др.).
5. Взаимосвязь математической модели СК-анализа с другими моделями.

1.2.2.1. Программа системного обобщения математики и предпосылки системной теории информации

Дадим определение понятия «система» классическим способом, т.е. путем его подведения под более общее понятие «множество» и выделение специфических признаков. Система представляет собой множество элементов, объединенных в целое за счет взаимодействия элементов друг с другом, т.е. за счет отношений между ними, и обеспечивающая преимущества в достижении целей. Преимущества в достижении целей обеспечиваются за счет системного эффекта. Системный эффект состоит в том, что свойства системы не сводятся к сумме свойств ее элементов, т.е. система как целое обладает рядом новых, т.е. эмерджентных свойств, которых не было у ее элементов. Уровень системности тем выше, чем выше интенсивность взаимодействия элементов системы друг с другом, чем сильнее отличаются свойства системы от свойств входящих в нее элементов, т.е. чем выше системный эффект, чем значительно отличается система от множества. Элементы взаимодействуют (вступают в отношения) друг с другом с помощью имеющихся у них общих свойств, а также свойств, которые коррелируют между собой. Таким образом, система обеспечивает тем большие преимущества в достижении целей, чем выше ее уровень системности. В частности, система с нулевым уровнем системности вообще ничем не отличается от множества образующих ее элементов, т.е. тождественна этому множеству и никаких преимуществ в достижении целей не обеспечивает. Этим самым обеспечивается выполнение принципа соответствия между понятиями системы и множества. Из соблюдения этого принципа для понятий множества и системы следует и его соблюдение для понятий, основанных на теории множеств и их системных обобщений.

На этой основе можно ввести и новое научное понятие: понятие «антисистемы», применение которого оправдано в случаях, когда централизация (моно-

полизация, интеграция) не только не дает положительного эффекта, но даже называется отрицательно. Антисистемой называется система с отрицательным уровнем системности, т.е. это такое объединение некоторого множества элементов за счет их взаимодействия в целое, которое препятствует достижению целей.

Фундаментом, находящимся в самом основании современной математики, является теория множеств. Эта теория лежит и в основе самого глубокого на сегодняшний день обоснования таких базовых математических понятий, как «число» и «функция». Определенный период этот фундамент казался незыблемым. Однако вскоре работы целой плеяды выдающихся ученых XX века, прежде всего Давида Гильберта, Бертрана Рассела и Курта Гёделя, со всей очевидностью обнажили фундаментальные логические и лингвистические проблемы, в частности проявляющиеся в форме парадоксов теории множеств, что в свою очередь привело к появлению ряда развернутых предложений по пересмотру самых глубоких оснований математики [20].

В задачи данной статьи не входит рассмотрение этой интереснейшей проблематики, а также истории возникновения и развития понятий числа и функции. Отметим лишь, что кроме рассмотренных в литературе вариантов существует возможность обобщения всех понятий математики, базирующихся на теории множеств, в частности теории информации, путем тотальной замены понятия множества на понятие системы и тщательного отслеживания всех последствий этой замены. Это утверждение будем называть «программной идеей системного обобщения понятий математики».

Строго говоря, реализация данной программной идеи потребует прежде всего системного обобщения самой теории множеств и преобразование ее в математическую теорию систем, которая будет плавно переходить в современную теорию множеств при уровне системности стремящемся к нулю. При этом необходимо заметить, что существующая в настоящее время наука под названием «Теория систем» ни в коей мере не является обобщением математической теории множеств и ее не следует путать с математической теорией систем. Вместе с тем, на наш взгляд, существуют некоторые возможности обобщения ряда понятий математики и без разработки математической теории систем. К таким понятиям относятся прежде всего понятия «информация» и «функция».

Системному обобщению понятия информации посвящены работы автора [5, 6, 9, 10, 11] и другие, поэтому в данной статье на этом вопросе мы останавливаться не будем. Отметим лишь, что на основе предложенной системной теории информации (СТИ) были разработаны математическая модель и методика численных расчетов (структуры данных и алгоритмы), а также специальный программный инструментарий (система «Эйдос») автоматизированного системно-когнитивного анализа (АСК-анализ), который представляет собой системный анализ, автоматизированный путем его рассмотрения как метода познания и структурирования по базовым когнитивным операциям.

В АСК-анализе теоретически обоснована и реализована на практике в форме конкретной информационной технологии процедура установления новой универсальной, сопоставимой в пространстве и времени, ранее не используемой

количественной, т.е. выражаемой числами, меры соответствия между событиями или явлениями любого рода, получившей название «системная мера целесообразности информации», которая по существу является количественной мерой знаний [10]. Это является достаточным основанием для того, чтобы назвать эти числа «когнитивными» от английского слова "cognition" – "познание".

В настоящее время функция понимается как соответствие друг другу нескольких множеств чисел. Поэтому виды функций можно классифицировать по крайней мере в зависимости от:

- природы этих чисел (натуральные, целые, дробные, действительные, комплексные и т.п.);
- количества и вида множеств чисел, связанных друг с другом в функции (функции одного, нескольких, многих, счетного или континуального количества аргументов, однозначные и многозначные функции, дискретные или континуальные функции) [10];
- степени жесткости и меры силы связи между множествами чисел (детерминистские функции, функции, в которых в качестве меры связи используется вероятность, корреляция и другие меры);
- степени расплывчатости чисел в множествах и самой формы функции (четкие и нечеткие функции, использование различных видов шкал, в частности интервальных оценок).

Так как функции, выявляемые модели предметной области методом АСК-анализа, связывают друг с другом множества когнитивных чисел, то предлагается называть их «когнитивными функциями». Учитывая перечисленные возможности классификации когнитивных функций, можно считать недетерминистскими многозначными функциями многих аргументов, в которых в качестве меры силы связи между множествами используется количественная мера знаний, т.е. системная мера целесообразности информации, основанная на интервальных оценках, номинальных и порядковых шкалах и шкалах отношений. Отметим, что детерминистские однозначные функции нескольких аргументов могут рассматриваться как частный случай когнитивных функций, к которому они сводятся при анализе жестко детерминированной предметной области, скажем явлений, описываемых классической физикой.

Итак, *предлагается программная идея системного обобщения понятий математики, в частности теории информации, основанных на теории множеств, путем замены понятия множества на более содержательное понятие системы*. Частично эта идея была реализована автором при разработке автоматизированного системно-когнитивного анализа (АСК-анализа), математическая модель которого основана на системном обобщении формул для количества информации Хартли и Харкевича. Реализация следующего шага: системного обобщения понятия функциональной зависимости рассматривается в работе [87], в ней же вводятся новые научные понятия и соответствующие термины: «когнитивные функции» и «когнитивные числа». На численных примерах показано, что АСК-анализ обеспечивает выявление когнитивных функциональных зависимостей в многомерных зашумленных фрагментированных данных.

1.2.2.2. Теоретические основы системной теории информации

1.2.2.2.1. Требования к математической модели и численной мере

Для практической реализации идеи решения проблемы необходимо сформулировать требования к математической модели и численной мере, вытекающие из когнитивной концепции и специфических свойств активного объекта управления в АПК (слабодетерминированность, многофакторность, активность).

Требования к математической модели предусматривают: содержательную интерпретируемость; эффективную вычислимость на основе непосредственно эмпирических данных (наличие эффективного численного метода); универсальность; адекватность; сходимость; семантическую устойчивость; сопоставимость результатов моделирования в пространстве и времени; непараметричность; формализацию базовых когнитивных операций системного анализа (прежде всего таких, как обобщение, абстрагирование, сравнение, классификация и др.); корректность работы на фрагментарных, неточных и зашумленных данных; возможность обработки данных очень больших размерностей (тысячи факторов и будущих состояний объекта управления); математическую и алгоритмическую ясность и простоту, эффективность программной реализации.

Требования к численной мере. Ключевым при построении математических моделей является выбор количественной меры, обеспечивающей учет степени причинно-следственной взаимосвязи исследуемых параметров. Эта мера должна удовлетворять следующим требованиям: 1) обеспечивать эффективную вычислимость на основе эмпирических данных, полученных непосредственно из опыта; 2) обладать универсальностью, т.е. независимостью от предметной области; 3) подчиняться единому для различных предметных областей принципу *содержательной интерпретации*; 4) количественно измеряться в единых единицах измерения а количественной шкале (шкала с естественным нулем, максимумом или минимумом); 5) учитывать понятия: "цели объекта управления", "цели управления"; "мощность множества будущих состояний объекта управления"; уровень системности объекта управления; степень детерминированности объекта управления; 6) обладать сопоставимостью в пространстве и во времени; 7) обеспечивать возможность введения метрики или неметрической функции принадлежности на базе выбранной количественной меры.

Для того, чтобы выбрать тип (класс) модели, удовлетворяющей сформулированным требованиям, необходимо решить на какой форме информации эта модель будет основана: абсолютной, относительной или аналитической.

1.2.2.2.2. Выбор базовой численной меры

Абсолютная, относительная и аналитическая информация.

Широко известны абсолютная и относительная формы информации. Абсолютная форма – это просто количество, частота. Относительная форма – это доли, проценты, относительные частоты и вероятности.

Менее знакомы специалисты с аналитической формой информации, примером которой являются условные вероятности, стандартизированные статистические значения, эластичность и количество информации.

Очевидно, что и из относительной информации, взятой изолированно, вырванной из контекста, делать какие-либо обоснованные выводы не представляется возможным. Для того, чтобы о чем-то судить по процентам, нужен их *сопоставительный* анализ, т.е. анализ *всего* процентного распределения. Обычно для используется "база оценки", в качестве которой используется среднего по всей совокупности или "скользящее среднее" (нормативный подход: норма – среднее).

Аналитическая (сопоставительная) информация – это информация, содержащаяся в *отношении* вероятности (или процента) к некоторой *базовой* величине, например к средней вероятности по всей выборке. Аналитическими являются также стандартизированные величины в статистике и количество информации в теории информации.

Очевидно, именно аналитическая информация является наиболее кондиционной для употребления с той точки зрения, что позволяет непосредственно делать содержательные выводы об исследуемой предметной области (точнее будет сказать, что она сама и является выводом), тогда как для того, чтобы сделать аналогичные выводы на основе относительной, и особенно абсолютной информации требуется ее значительная предварительная обработка.

Выбор в качестве базовой численной меры количества информации

Как было показано в лекции 2, системный анализ представляет собой теоретический метод познания, т.е. информационный процесс, в котором поток информации направлен от познаваемого объекта к познающему субъекту. Процесс труда, напротив, представляет собой процесс, в котором поток информации направлен от субъекта к объекту. При этом информация передается по каналу связи, представляющему собой средства труда, и записывается в носитель информации (предмет труда), который в ходе этого процесса преобразуется в заранее заданную форму, т.е. в продукт труда. Таким образом, процесс труда по сути дела представляет собой информационный процесс, обратный по направлению потока информации процессу познания. Управление представляет собой процесс, на различных этапах которого выполняются функции, сходные с процессами труда (управляющее воздействие) и познания (обратная связь). По мнению автора, информационный подход к управлению является наиболее общим. *Поэтому в качестве количественной меры взаимосвязи факторов и будущих состояний АОУ целесообразно использовать количество информации.* Более подробное обоснование целесообразности выбора в качестве численной меры количества информации приведено в работе автора [81].

Однако, известно много различных информационных мер и, следовательно, возникает задача выбора одной из них, оптимальной по выбранным критериям. Различные выражения классической теории информации для количества информации: Хартли, Шеннона, Харкевича и др., учитывают различные аспекты информационного моделирования объектов (таблице 8):

**Таблица 8 – СООТВЕТСТВИЕ ТРЕБОВАНИЯМ
ВЫРАЖЕНИЙ ДЛЯ КОЛИЧЕСТВА ИНФОРМАЦИИ**

	Хартли	Шеннон	Харкевич
Мощность множества состояний	+	–	–
Факторы	–	+	+
Цель	–	–	+
Уровень системности объекта	–	–	–
Степень детерминированности объекта	–	–	–
Рейтинг	1	1	2

– формула Хартли учитывает количество классов (мощность множества состояний объекта управления) но никак не учитывает их признаков или факторов, переводящих объект в эти состояния, т.е. содержит интегральное описание объектов;

– формула Шеннона основывается на учете признаков, т.е. основывается на дискретном описании объектов;

– формула Харкевича учитывает понятие цели и также как формула Шеннона основана на статистике признаков, но не учитывает мощности множества будущих состояний объекта управления, включающего целевые и другие будущие состояния объекта управления и также как формула Шеннона основывается на дискретном описании объектов.

Как видно из таблицы 8, классическая формула Харкевича по учитываемым критериям имеет преимущества перед классическими формулами Хартли и Шеннона, т.к. учитывает как факторы, так и *понятие цели, ключевое для системного анализа, теории и практики управления (в т.ч. АСУ)*. Поэтому именно выражение для семантической целесообразности информации Харкевича взято за основу при выводе обобщающего выражения, удовлетворяющего всем предъявляемым требованиям.

1.2.2.2.3. Конструирование системной численной меры на основе базовой

Системное обобщение формулы Хартли для количества информации

Классическая формула Хартли имеет вид:

$$I = \text{Log}_2 W \quad (3.1)$$

Будем искать ее системное обобщение в виде:

$$I = \text{Log}_2 W^\varphi \quad (3.2)$$

где:

W – количество чистых (классических) состояний системы.

φ – коэффициент эмерджентности Хартли (уровень системной организации объекта, имеющего W чистых состояний);

Учитывая, что возможны *смешанные состояния, являющиеся нелинейной суперпозицией или одновременной реализацией чистых (классических) состояний* "из W по m", всего возможно C_W^m состояний системы, являющихся сочетаниями классических состояний. Таким образом, *примем за аксиому*, что системное обобщение формулы Хартли имеет вид [81]:

$$I = \text{Log}_2 \sum_{m=1}^M C_W^m \quad (3.3)$$

где: W – количество элементов в системе альтернативных будущих состояний АОУ (количество чистых состояний); m – сложность смешанных состояний АОУ; M – максимальная сложность смешанных состояний АОУ.

Выражение (1) дает количество информации в активной системе, в которой чистые и смешанные состояния равновероятны. Смешанные состояния активных систем, возникающие под действием системы нелинейно-взаимодействующих факторов, считаются такими же измеримыми, как и чистые альтернативные состояния, возникающие под действием детерминистских факторов. Так как $C_W^1 = W$, то при $M=1$ выражение (3.3) приобретает вид (3.1), т.е. выполняется *принцип соответствия*, являющийся обязательным для более общей теории.

Рассмотрим подробнее смысл выражения (3.3), представив сумму в виде ряда слагаемых:

$$I = \text{Log}_2 (C_W^1 + C_W^2 + \dots + C_W^M) \quad (3.4)$$

Первое слагаемое в (3.4) дает количество информации по классической формуле Хартли, а остальные слагаемые – *дополнительное количество информации, получаемое за счет системного эффекта*, т.е. за счет наличия у системы иерархической структуры или смешанных состояний. По сути дела эта *дополнительная информация является информацией об иерархической структуре системы, как состоящей из ряда подсистем различных уровней сложности*.

Например, пусть система состоит из W пронумерованных элементов 1-го уровня иерархии. Тогда на 2-м уровне иерархии элементы соединены в подсистемы из 2 элементов 1-го уровня, на 3-м – из 3, и т.д. Если выборка любого элемента равновероятна, то из факта выбора n -го элемента по классической формуле Хартли мы получаем количество информации согласно (3.1). Если же при этом известно, что данный элемент входит в определенную подсистему 2-го уровня, то это дает дополнительное количество информации, за счет учета второго слагаемого, поэтому общее количество получаемой при этом информации будет определяться выражением (3.4) уже с двумя слагаемыми ($M=2$). Если элемент одновременно входит в M подсистем разных уровней, то количество информации, получаемое о системе и ее подсистемах при выборке этого элемента определяется выражением (3.4). Так, если мы вытаскиваем кирпич из неструктурированной кучи, состоящей из 32 кирпичей, то получаем 5 бит информации, если же из этих кирпичей сложен дом, то при аналогичном действии мы получаем дополнительное количество информации о том, из каких части дома (подсистем различного уровня иерархии) вытаскен этот кирпич. Действия каменщика, укладывающего кирпич на место, предусмотренное *проектом*, значительно выше по целесообразности, чем у грузчика, складывающего кирпичи в кучу. Учитывая, что при $M=W$:

$$\sum_{m=1}^M C_W^m = 2^W - 1 \quad (3.5)$$

в этом случае получаем:

$$I = \log_2(2^W - 1) \quad (3.6)$$

Выражение (3.5) дает оценку максимального количества информации, которое может содержаться в элементе системы с учетом его вхождения в различные подсистемы ее иерархической структуры.

Однако реально в любой системе осуществляются не все формально возможные сочетания элементов 1-го уровня иерархии, т.к. существуют различные **правила запрета**, различные для разных систем. Это означает, что возможно множество различных систем, состоящих из одинакового количества тождественных элементов, и отличающихся своей структурой, т.е. строением подсистем различных иерархических уровней. Эти различия систем как раз и возникают благодаря различию действующих для них этих правил запрета. По этой причине *систему правил запрета предлагается назвать информационным проектом системы*. Различные системы, состоящие из равного количества одинаковых элементов (например, дома, состоящие из 20000 кирпичей), отличаются друг от друга именно по причине различия своих информационных проектов.

Из выражения (3.5) очевидно, что I быстро стремится к W :

$$\begin{aligned} \text{при } W \rightarrow \infty \\ I \rightarrow W \end{aligned} \quad (3.7)$$

В действительности уже при $W > 4$ погрешность выражения (3.5) не превышает 1% (таблица 9):

**Таблица 9 – ЗАВИСИМОСТЬ ПОГРЕШНОСТИ
ВЫРАЖЕНИЯ (3.5) ОТ КОЛИЧЕСТВА КЛАССОВ W**

W	2	3	4	5	6	7	8	9	10
Погрешность в %	20,752	6,422	2,328	0,916	0,379	0,162	0,071	0,031	0,014

График зависимости погрешности выражения (3.5) от количества классов W приведен на рисунке 24.

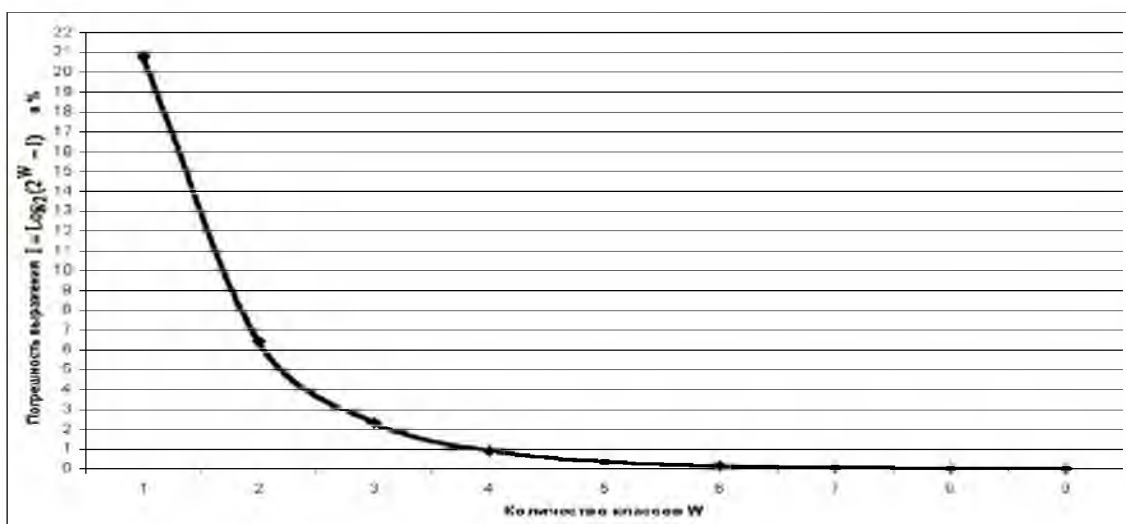


Рисунок 24. Зависимость погрешности приближенного выражения системного

обобщения формулы Хартли от количества классов W

Приравняв правые части выражений (3.2) и (3.3):

$$I = \text{Log}_2 W^\varphi = \text{Log}_2 \sum_{m=1}^M C_W^m \quad (3.8)$$

получим выражение для коэффициента эмерджентности Хартли (терм. авт.):

$$\varphi = \frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W} \quad (3.9)$$

Непосредственно из вида выражения для коэффициента эмерджентности Хартли (3.9) ясно, что он представляет собой относительное превышение количества информации о системе при учете системных эффектов (смешанных состояний, иерархической структуры ее подсистем и т.п.) над количеством информации без учета системности, т.е. этот коэффициент отражает уровень системности объекта.

С учетом выражения (3.9) выражение (3.2) примет вид:

$$I(W, M) = \text{Log}_2 W \frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W} \quad (3.10)$$

или при $M=W$ и больших W , учитывая (3.4 – 3.6):

$$I(W, M) = \text{Log}_2 W \frac{W}{\text{Log}_2 W} = W \quad (3.11)$$

Выражение (3.10) и представляет собой искомое системное обобщение классической формулы Хартли, а выражение (3.11) – его достаточно хорошее приближение при большом количестве элементов или состояний системы (W).

Коэффициент эмерджентности Хартли представляет собой относительное превышение количества информации о системе при учете системных эффектов (смешанных состояний, иерархической структуры ее подсистем и т.п.) над количеством информации без учета системности, т.е. этот коэффициент является аналитическим выражением для *уровня системности объекта*. Таким образом, **коэффициент эмерджентности Хартли отражает уровень системности объекта и изменяется от 1 (системность минимальна, т.е. отсутствует) до $W/\text{Log}_2 W$ (системность максимальна)**. Очевидно, для каждого количества элементов системы существует свой максимальный уровень системности, который никогда реально не достигается из-за действия **правил запрета** на реализацию в системе ряда подсистем различных уровней иерархии.

Например: из 32 букв русского алфавита может быть образовано не $C_{32}^6 = 906192$ осмысленных 6-буквенных слов, а значительно меньше. Если мы услышим одно из этих в принципе возможных слов, то получим не $5 \times 6 = 30$ информации, содержащейся непосредственно в буквах (в одной букве содержится $\text{Log}_2 32 = 5$ бит), а $30 + 19,79 = 49,79$ бит, т.е. в 1.66 раз больше. Это и есть *уровень*

системности иерархического уровня 6-буквенных слов русского языка. Уровень системности русского языка, как системы, состоящей из слов длиной от одной до 6 букв, согласно выражения (3.9) с учетом (3.5), равен примерно 6,4. Но при этом еще не была учтена информация, содержащаяся в последовательности слов, в последовательности предложений и т.д.

Итак, в предложении содержится значительно больше информации, чем в буквах, с помощью которых оно написано, т.к. кроме букв информацию содержат слова, сочетания слов, последовательность предложений и т.д.. Буквы образуют 1-й иерархический уровень языка, слова – 2-й, предложения – 3-й, абзацы – 4-й, параграфы – 5-й, главы – 6-й, произведения – 7-й. Теория Шеннона концентрирует основное внимание на рассмотрении 1-го уровня, т.е. рассматривает тексты, прежде всего, как последовательность символов. Но именно иерархическая организация, не учитываемая в теории Шеннона и отраженная в системной теории информации, обеспечивает языку его удивительную мощь, как средства отражения и моделирования реальности.

Аналогично и в генах, этих своеобразных "символах генома", содержится значительно больше информации о фенотипе, чем предполагается в классической генетике Менделя, т.к. гены образуют *ансамбли различных уровней иерархии* в зависимости от влияния среды и технологий управления (явление адаптивности системы "генотип-среда", Драгавцев В.А., 1993). Если ген уподобить букве алфавита, а смысл фразы – фенотипическому признаку, то можно сказать, что *возможно очень большое количество фраз с практически тождественным смысловым содержанием* (тогда как в классической генетике считалось, что признак соответствует гену, хотя есть и такие). После расшифровки генома человека мы настолько же приблизились к его пониманию, как изучивший русскую или немецкую азбуку англичанин, не знающий этих языков, приблизился к чтению в оригинале и пониманию содержания "Войны и Мира" Льва Толстого или "Феноменологии Духа" Георга В.Ф.Гегеля.

На уровне слов верхняя оценка уровня системности русского языка с учетом (3.5) составляет огромную величину: 2616,48 (предполагается, что в русском языке 40000 слов и предложения могут иметь любую длину). Необходимо отметить, что *правила запрета на порядок слов в русском языке значительно слабее, чем, например в английском, поэтому в русском языке возможно гораздо больше грамматически правильных и несущих различную информацию предложений из одних и тех же слов, чем в английском.* Это значит, что уровень системности русского языка на уровне предложений, по-видимому, значительно превосходит уровень системности английского языка. При длине предложения до 2-х слов системность русского языка на уровне предложений согласно (3.9) составляет: 52330916.

Анализ выражения (3.9) показывает, что при $M=1$ оно преобразуется в (3.1), т.е. *выполняется принцип соответствия*. При $M>1$ количество информации в соответствии с системной теорией информации (СТИ) (3.9) будет превосходить количество информации, рассчитанное по классической теории информации (КТИ) (3.1). Непосредственно из выражения (3.2) получаем:

$$I(W, M) = \log_2 W + \log_2 W^{\varphi-1} \quad (3.12)$$

Первое слагаемое в выражении (3.12) отражает количество информации, согласно КТИ, а второе – СТИ, т.е. доля системной информации.

Представляет несомненный интерес исследование закономерностей изменения доли системной информации в поведении элемента системы в зависимости от количества классов W и сложности смешанных состояний M .

В таблице 10 приведены результаты численных расчетов в соответствии с выражением (3.9). Сводные данные из таблицы 10 приведены в таблице 11, а в графическом виде они представлены на рисунке 25.

Таблица 10 – ЗАВИСИМОСТЬ $I(W, M)$ ОТ КОЛИЧЕСТВА КЛАССОВ W И СЛОЖНОСТИ СМЕШАННЫХ СОСТОЯНИЙ M

W=2				W=3				W=4			
M	Число сочетаний		I(W, M)	M	Число сочетаний		I(W, M)	M	Число сочетаний		I(W, M)
	Из W по m	Сумма по m от 1 до M			Из W по m	Сумма по m от 1 до M			Из W по m	Сумма по m от 1 до M	
1	2	2	1,00000	1	3	3	1,58496	1	4	4	2,00000
2	1	3	1,58496	2	3	6	2,58496	2	6	10	3,32193
3		4	2,00000	3	1	7	2,80735	3	4	14	3,80735
4				4		8	3,00000	4	1	15	3,90689
5				5				5		16	4,00000
W=5				W=6				W=7			
M	Число сочетаний		I(W, M)	M	Число сочетаний		I(W, M)	M	Число сочетаний		I(W, M)
	Из W по m	Сумма по m от 1 до M			Из W по m	Сумма по m от 1 до M			Из W по m	Сумма по m от 1 до M	
1	5	5	2,32193	1	6	6	2,58496	1	7	7	2,80735
2	10	15	3,90689	2	15	21	4,39232	2	21	28	4,80735
3	10	25	4,64386	3	20	41	5,35755	3	35	63	5,97728
4	5	30	4,90689	4	15	56	5,80735	4	35	98	6,61471
5	1	31	4,95420	5	6	62	5,95420	5	21	119	6,89482
6		32	5,00000	6	1	63	5,97728	6	7	126	6,97728
7				7		64	6,00000	7	1	127	6,98868
8				8				8		128	7,00000
W=8				W=9				W=10			
M	Число сочетаний		I(W, M)	M	Число сочетаний		I(W, M)	M	Число сочетаний		I(W, M)
	Из W по m	Сумма по m от 1 до M			Из W по m	Сумма по m от 1 до M			Из W по m	Сумма по m от 1 до M	
1	8	8	3,00000	1	9	9	3,16993	1	10	10	3,32193
2	28	36	5,16993	2	36	45	5,49185	2	45	55	5,78136
3	56	92	6,52356	3	84	129	7,01123	3	120	175	7,45121
4	70	162	7,33985	4	126	255	7,99435	4	210	385	8,58871
5	56	218	7,76818	5	126	381	8,57365	5	252	637	9,31515
6	28	246	7,94251	6	84	465	8,86109	6	210	847	9,72622
7	8	254	7,98868	7	36	501	8,96867	7	120	967	9,91737
8	1	255	7,99435	8	9	510	8,99435	8	45	1012	9,98299
9		256	8,00000	9	1	511	8,99718	9	10	1022	9,99718
10				10		512	9,00000	10	1	1023	9,99859

Таблица 11 – ЗАВИСИМОСТЬ КОЛИЧЕСТВА ИНФОРМАЦИИ $I(W, M)$ ОТ СЛОЖНОСТИ СМЕШАННЫХ СОСТОЯНИЙ M ДЛЯ РАЗЛИЧНОГО КОЛИЧЕСТВА КЛАССОВ W

M	W=2	W=3	W=4	W=5	W=6	W=7	W=8	W=9	W=10
	I(W, M)	I(W, M)	I(W, M)	I(W, M)	I(W, M)	I(W, M)	I(W, M)	I(W, M)	I(W, M)
1	1,00000	1,58496	2,00000	2,32193	2,58496	2,80735	3,00000	3,16993	3,32193
2	1,58496	2,58496	3,32193	3,90689	4,39232	4,80735	5,16993	5,49185	5,78136
3	2,00000	2,80735	3,80735	4,64386	5,35755	5,97728	6,52356	7,01123	7,45121
4	2,00000	3,00000	3,90689	4,90689	5,80735	6,61471	7,33985	7,99435	8,58871
5	2,00000	3,00000	4,00000	4,95420	5,95420	6,89482	7,76818	8,57365	9,31515
6	2,00000	3,00000	4,00000	5,00000	5,97728	6,97728	7,94251	8,86109	9,72622
7	2,00000	3,00000	4,00000	5,00000	6,00000	6,98868	7,98868	8,96867	9,91737
8	2,00000	3,00000	4,00000	5,00000	6,00000	7,00000	7,99435	8,99435	9,98299
9	2,00000	3,00000	4,00000	5,00000	6,00000	7,00000	8,00000	8,99718	9,99718
10	2,00000	3,00000	4,00000	5,00000	6,00000	7,00000	8,00000	9,00000	9,99859
(СТИ-КТИ)'	0,585	1,222	1,907	2,632	3,392	4,181	4,994	5,827	6,677
(СТИ-КТИ)''		0,637	0,684	0,725	0,760	0,789	0,813	0,833	0,849
(СТИ-КТИ)'''			0,047	0,041	0,035	0,029	0,024	0,020	0,017

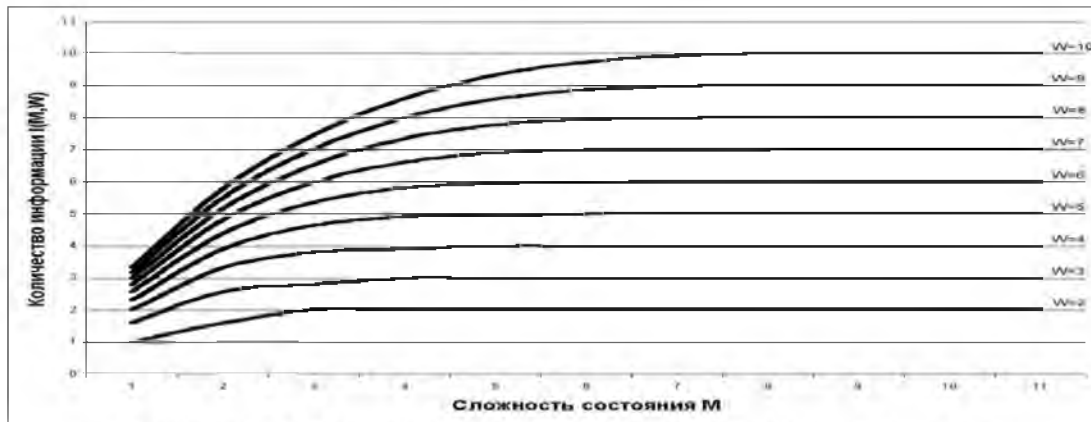


Рисунок 25. Зависимость количества информации $I(W, M)$ от сложности смешанных состояний M для разного количества классов W

Рост количества информации в СТИ по сравнению с КТИ обусловлен системным эффектом (эмерджентностью), который связан с учетом смешанных состояний, возникающих путем одновременной реализации (суперпозиции) нескольких чистых (классических) состояний под действием системы нелинейно-взаимодействующих недетерминистских факторов. Выражение (3.9) дает максимальную возможную оценку количества информации, т.к. могут существовать различные *правила запрета* на реализацию тех или иных смешанных состояний.

Фактически это означает, что в СТИ множество возможных состояний объекта рассматривается не как совокупность несвязанных друг с другом состояний, как в КТИ, а как *система*, уровень системности которой как раз и определяется коэффициентом эмерджентности Хартли ϕ (3.9), являющегося монотонно возрастающей функцией сложности смешанных состояний M . Следовательно, *дополнительная информация, которую мы получаем из поведения объекта в СТИ, по сути дела является информацией о системе всех возможных состояний объекта, элементом которой является объект в некотором данном состоянии.*

Гипотеза о законе возрастания эмерджентности и следствия из него

Численные расчеты и аналитические выкладки в соответствии с СТИ показывают, что при возрастании количества элементов в системе доля системной информации в поведении ее элементов возрастает. Это обнаруженное нами новое фундаментальное свойство систем предлагается назвать законом возрастания эмерджентности.

Закон возрастания эмерджентности: "Чем больше элементов в системе, тем большую долю содержащейся в ней информации составляет информация, содержащаяся во взаимосвязях ее элементов".

На рисунках 26 и 27 приведены графики скорости и ускорения возрастания эмерджентности в зависимости от количества элементов W в системе.



Рисунок 26. Возрастание доли системной информации в поведении элемента системы при увеличении количества элементов W

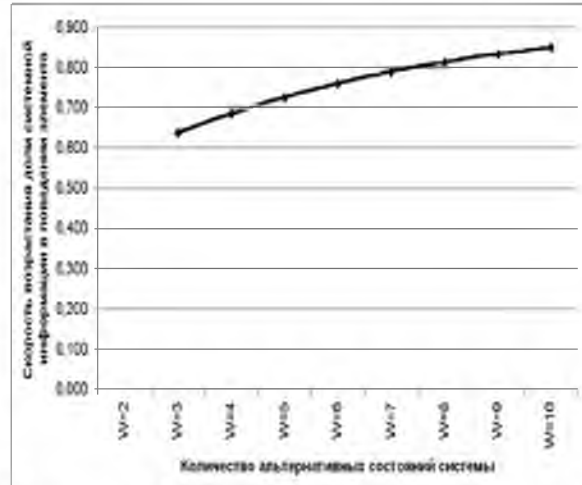


Рисунок 27. Ускорение возрастания доли системной информации в поведении элемента системы от количества элементов W

Более детальный анализ предполагаемого закона возрастания эмерджентности с использованием конечных разностей первого и второго порядка (таблица 11) показывает, что *при увеличении количества элементов в системе доля системной информации в ней возрастает с ускорением, которое постепенно уменьшается*. Это утверждение будем называть леммой 1.

Продолжим анализ закона возрастания эмерджентности. Учитывая, что:

$$C_W^m = \frac{W!}{m!(W-m)!}$$

выражение (3.3) принимает вид:

$$I = \log_2 \sum_{m=1}^M \frac{W!}{m!(W-m)!} \quad (3.13)$$

где: $1 \leq M \leq W$.

$$\begin{aligned} I &= \log_2 \sum_{m=1}^M \frac{W!}{m!(W-m)!} = \\ &= \log_2 \left(W! \cdot \sum_{m=1}^M \frac{1}{m!(W-m)!} \right) = \\ &= \log_2(W!) + \log_2 \sum_{m=1}^M \frac{1}{m!(W-m)!} = \\ &= \sum_{n=1}^W \log_2 n + \log_2 \sum_{m=1}^M \frac{1}{m!(W-m)!} \end{aligned}$$

и учитывая, что $\log_2 1 = 0$, выражение (3.13) приобретает вид:

$$I = \sum_{n=2}^W \log_2 n + \log_2 \sum_{m=1}^M \frac{1}{m!(W-m)!} = I(W) + I(W, M) \quad (3.14)$$

Где введены обозначения:

$$I(W) = \sum_{n=2}^W \text{Log}_2 n$$

$$I(W, M) = \text{Log}_2 \sum_{m=1}^M \frac{1}{m!(W-m)!}$$
(3. 15)

С учетом (3.14) выражение (3.9) для коэффициента эмерджентности Хартли приобретает вид:

$$\varphi = \frac{\sum_{n=2}^W \text{Log}_2 n}{\text{Log}_2 W} + \frac{\text{Log}_2 \sum_{m=1}^M \frac{1}{m!(W-m)!}}{\text{Log}_2 W}$$

Заменяя в (3.13) факториал на Гамма-функцию, получаем обобщение выражения (3.3) на непрерывный случай:

$$\begin{aligned} I &= \text{Log}_2 \int_{m=1}^M \left(\frac{G(W)}{G(m) \cdot G(W-m)} \right) dm = \\ &= \text{Log}_2 \left\{ G(W) \cdot \int_{m=1}^M \frac{dm}{G(m) \cdot G(W-m)} \right\} = \\ &= \text{Log}_2 G(W) + \text{Log}_2 \int_{m=1}^M \frac{dm}{G(m) \cdot G(W-m)} \end{aligned}$$

Или окончательно:

$$I = \text{Log}_2 G(W) + \text{Log}_2 \int_{m=1}^M \frac{dm}{G(m) \cdot G(W-m)} = I(W) + I(W, M)$$
(3. 16)

Для непрерывного случая обозначения (3.15) принимают вид:

$$I(W) = \text{Log}_2 G(W)$$

$$I(W, M) = \text{Log}_2 \int_{m=1}^M \frac{dm}{G(m) \cdot G(W-m)}$$
(3. 17)

Учитывая выражения (3.9) и (3.16) получим выражение для коэффициента эмерджентности Хартли для непрерывного случая:

$$\begin{aligned} \varphi &= \frac{\text{Log}_2 G(W) + \text{Log}_2 \int_{m=1}^M \frac{dm}{G(m) \cdot G(W-m)}}{\text{Log}_2 W} = \\ &= \frac{\text{Log}_2 G(W)}{\text{Log}_2 W} + \frac{\text{Log}_2 \int_{m=1}^M \frac{dm}{G(m) \cdot G(W-m)}}{\text{Log}_2 W} \end{aligned}$$

И окончательно для непрерывного случая:

$$\varphi = \frac{\text{Log}_2 G(W)}{\text{Log}_2 W} + \frac{\text{Log}_2 \int_{m=1}^M \frac{dm}{G(m) \cdot G(W-m)}}{\text{Log}_2 W}$$
(3. 18)

Анализируя выражения (3.14) и (3.16) видим, что количество информации, получаемое при выборке из системы некоторого ее элемента, состоит из двух слагаемых:

1) $I(W)$, зависящего только от количества элементов в системе W (первое слагаемое);

2) $I(W, M)$, зависящего как от количества элементов в системе W , так и от максимальной сложности, т.е. связности элементов подсистем M между собой (второе слагаемое).

Этот результат позволяет высказать гипотезы "О природе сложности системы" и "О видах системной информации":

– сложность системы определяется количеством содержащейся в ней информации;

– системная информация включает две составляющих: зависящее от количества элементов системы и зависящее также от характера взаимосвязей между элементами.

Изучим какой относительный вклад вносит каждое слагаемое в общее количество информации системы в зависимости от числа элементов в системе W и сложности подсистем M . Результаты численных расчетов показывают, что чем выше уровень системности, тем большая доля информации системы содержится во взаимосвязях ее элементов, и чем меньше элементов в системе, тем быстрее возрастает доля информации, содержащейся во взаимосвязях элементов при возрастании уровня системности. Эти утверждения будем рассматривать как леммы 2 и 3. Таким образом полная формулировка гипотезы о законе возрастания эмерджентности с гипотезой о видах информации в системе и тремя леммами приобретает вид:

ГИПОТЕЗА О ЗАКОНЕ ВОЗРАСТАНИЯ ЭМЕРДЖЕНТОСТИ: "Чем больше элементов в системе, тем большую долю содержащейся в ней информации составляет информация, содержащаяся во взаимосвязях ее элементов" (рисунок 28).

Гипотеза 1: "О природе сложности системы": сложность системы определяется количеством содержащейся в ней информации.

Гипотеза 2: "О видах системной информации": системная информация включает две составляющие:

– зависящую от количества элементов системы;

– зависящую как от количества элементов системы, так и от сложности взаимосвязей между ними.

Лемма-1: при увеличении количества элементов в системе доля системной информации в ней возрастает с ускорением, которое постепенно уменьшается.

Лемма-2: чем выше уровень системности, тем большая доля информации системы содержится во взаимосвязях ее элементов.

Лемма-3: чем меньше элементов в системе, тем быстрее возрастает доля информации, содержащейся во взаимосвязях элементов при возрастании уровня системности.

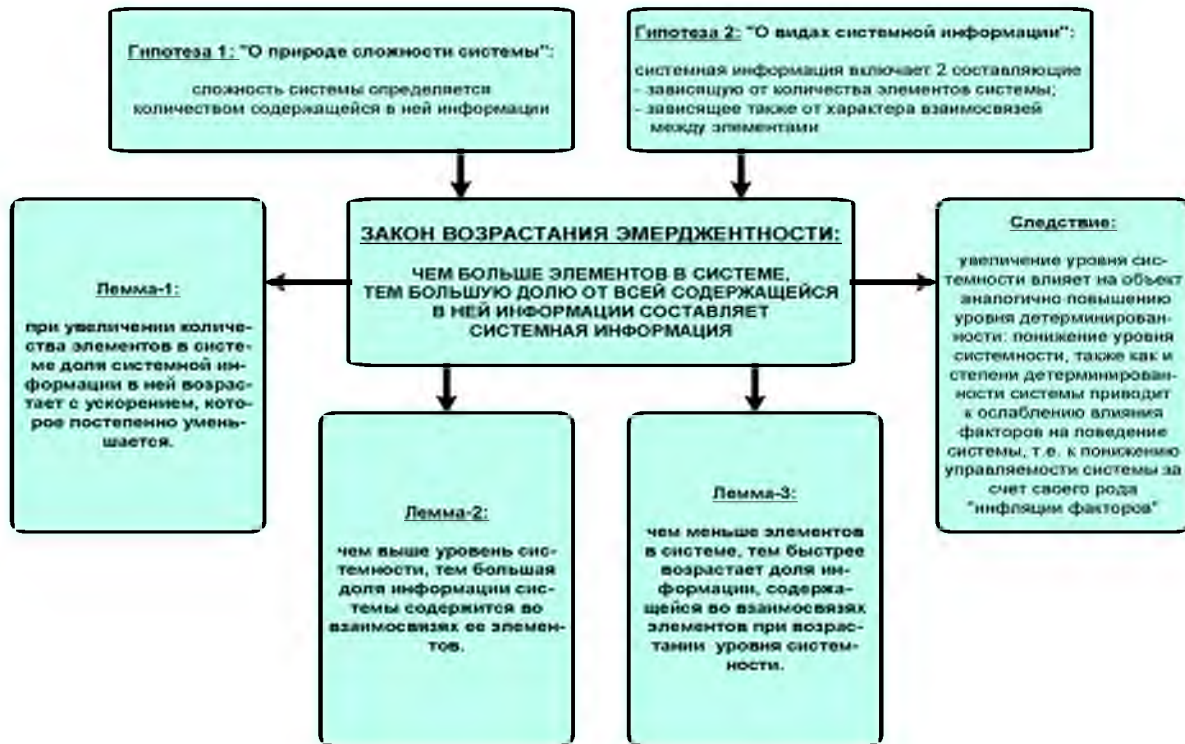


Рисунок 28. Закон возрастания эмерджентности

Системное обобщение классической формулы Харкевича, как количественная мера знаний

Это обобщение представляет большой интерес, в связи с тем, что А.Харкевич впервые ввел в теорию информации понятие цели. Он считал, что количество информации, сообщенное объекту, можно измерять по изменению вероятности достижения цели этим объектом за счет использования им этой информации.

Рассмотрим таблицу 12, в которой столбцы соответствуют будущим состояниям АОУ (целевым и нежелательным), а строки факторам, характеризующим объект управления, управляющую систему и окружающую среду.

Таблица 12 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ

Атрибуты		Классы - будущие состояния объекта управления					Сумма
		Целевые состояния		Нежелательные состояния			
		***	I	***	I	***	
Факторы, характеризующие текущее и прошлые состояния объекта управления, в т.ч. его рефлексивность	***						
	r		N_{rj}		N_{ri}		N_r

Управляющие факторы системы управления	i		N_{ij}		N_{ii}		N_i

Факторы, характеризующие прошлые, текущее и прогнозируемые состояния окружающей среды	k		N_{kj}		N_{ki}		N_k

Сумма			N_j		N_i		N

где:

 N_{ij} – количество встреч i-го признака у объектов j-го класса по данным обучающей выборки

Классическая формула А.Харкевича имеет вид:

$$I_{ij}(W, M) = \log_2 \frac{P_{ij}}{P_j} \quad (3.19)$$

где:

- W – количество классов (мощность множества будущих состояний объекта управления)
- M – максимальный уровень сложности смешанных состояний объекта управления;
- индекс i обозначает фактор: $1 \leq i \leq M$;
- индекс j обозначает класс: $1 \leq j \leq W$;
- P_{ij} – вероятность достижения объектом управления j -й цели при условии сообщения ему i -й информации;
- P_j – вероятность самопроизвольного достижения объектом управления j -й цели.

Ниже глобальные параметры модели W и M в выражениях для I опускаются, т.к. они являются константами для конкретной математической модели СК-анализа.

Однако: А.Харкевич в своем выражении для количества информации не ввел зависимости количества информации, *от мощности пространства будущих состояний объекта управления*, в т.ч. от количества его целевых состояний. Вместе с тем, один из возможных вариантов учета количества будущих состояний объекта управления обеспечивается классической и системной формулами Хартли (3.1) и (3.9); выражение (3.19) при подстановке в него реальных численных значений вероятностей P_{ij} и P_j не дает количества информации в битах; для выражения (3.19) не выполняется принцип соответствия, считающийся обязательным для обобщающих теорий. Возможно, в этом состоит одна из причин слабого взаимодействия между классической теорией информации Шеннона и семантической теорией информации.

Чтобы снять эти вопросы, приближенно выразим вероятности P_{ij} , P_i и P_j через частоты:

$$P_{ij} = \frac{N_{ij}}{N_i}; P_i = \frac{N_i}{N}; P_j = \frac{N_j}{N};$$

$$N_i = \sum_{j=1}^W N_{ij}; N_j = \sum_{i=1}^M N_{ij}; N = \sum_{i=1}^M \sum_{j=1}^W N_{ij} \quad (3.20)$$

В (3.20) использованы обозначения:

- N_{ij} – суммарное количество наблюдений факта: "действовал i -й фактор и объект перешел в j -е состояние";
- N_j – суммарное количество встреч различных факторов у объектов, перешедших в j -е состояние;
- N_i – суммарное количество встреч i -го фактора у всех объектов;
- N – суммарное количество встреч различных факторов у всех объектов.

Подставим в выражение (3.19) значения для P_{ij} и P_j из (3.20):

$$I_{ij} = \log_2 \frac{N_{ij} N}{N_i N_j} \quad (3.21)$$

Введем коэффициент эмерджентности Ψ в модифицированную формулу А.Харкевича:

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}N}{N_i N_j} \right)^{\Psi} \quad (3.22)$$

где: Ψ – коэффициент эмерджентности Харкевича (как будет показано выше, он определяет степень детерминированности объекта с уровнем системной организации φ , имеющего W чистых состояний, на переходы в которые оказывают влияние M факторов, о чем в модели накоплено N фактов).

Известно, что классическая формула Шеннона для количества информации для неравновероятных событий преобразуется в формулу Хартли при условии, что события равновероятны, т.е. удовлетворяет фундаментальному *принципу соответствия* [81].

Естественно потребовать, чтобы и обобщенная формула Харкевича также удовлетворяла аналогичному принципу соответствия, т.е. преобразовывалась в формулу Хартли в предельном случае, когда каждому классу (состоянию объекта) соответствует один признак (фактор), и каждому признаку – один класс, и эти классы (а, значит и признаки), равновероятны. Иначе говоря факторов столько же, сколько и будущих состояний объекта управления, все факторы детерминистские, а состояния объекта управления – альтернативные, т.е. каждый фактор однозначно определяет переход объекта управления в определенное состояние.

В этом предельном случае отпадает необходимость двухвекторного описания объектов, при котором 1-й вектор (классификационный) содержит интегральное описание объекта, как принадлежащего к определенным классам, а 2-й вектор (описательный) – дискретное его описание, как имеющего определенные атрибуты. Соответственно, двухвекторная модель, предложенная в данной работе, преобразуется в "вырожденный" частный случай – стандартную статистическую модель. В этом случае количество информации, содержащейся в признаке о принадлежности объекта к классу является *максимальным* и равным количеству информации, вычисляемому по системной формуле Хартли (3.9).

Таким образом при взаимно-однозначном соответствии классов и признаков:

$$\forall N_{ij} = N_i = N_j = 1 \quad (3.23)$$

формула А.Харкевича (3.13) приобретает вид:

$$I_{ij} = \text{Log}_2 N^{\Psi} = \text{Log}_2 W^{\varphi} \quad (3.24)$$

откуда:

$$\boxed{\Psi = \frac{\text{Log}_2 W^{\varphi}}{\text{Log}_2 N}} \quad (3.25)$$

или, учитывая выражение для коэффициента эмерджентности Хартли (3.8):

$$\Psi = \frac{\frac{\log_2 \sum_{m=1}^M C_W^m}{\log_2 W}}{\log_2 N} \quad (3.26)$$

Смысл коэффициента эмерджентности Харкевича (3.25) проясняется, если учесть, что при количестве состояний системы W равном количеству фактов N о действии на эту систему различных факторов он равен 1. В этом случае факторы *однозначно* определяют состояния объекта управления, т.е. являются детерминистскими. Если же количество фактов N о действии на эту систему различных факторов превосходит количество ее состояний W , что является гораздо более типичным случаем, то этот коэффициент меньше 1. По-видимому, это означает, что в этом случае факторы как правило не однозначно (и не так жестко как детерминистские) определяют поведение объекта управления, т.е. являются статистическими.

Таким образом, коэффициент эмерджентности Харкевича Ψ изменяется от 0 до 1 и определяет степень детерминированности системы:

- $\Psi=1$ соответствует полностью детерминированной системе, поведение которой однозначно определяется действием минимального количества факторов, которых столько же, сколько состояний системы;
- $\Psi=0$ соответствует полностью случайной системе, поведение которой никак не зависит действия факторов независимо от их количества;
- $0<\Psi<1$ соответствуют большинству реальных систем поведение которых зависит от многих факторов, число которых превосходит количество состояний системы, причем ни одно из состояний не определяется однозначно никакими сочетаниями действующих факторов (рисунок 29):

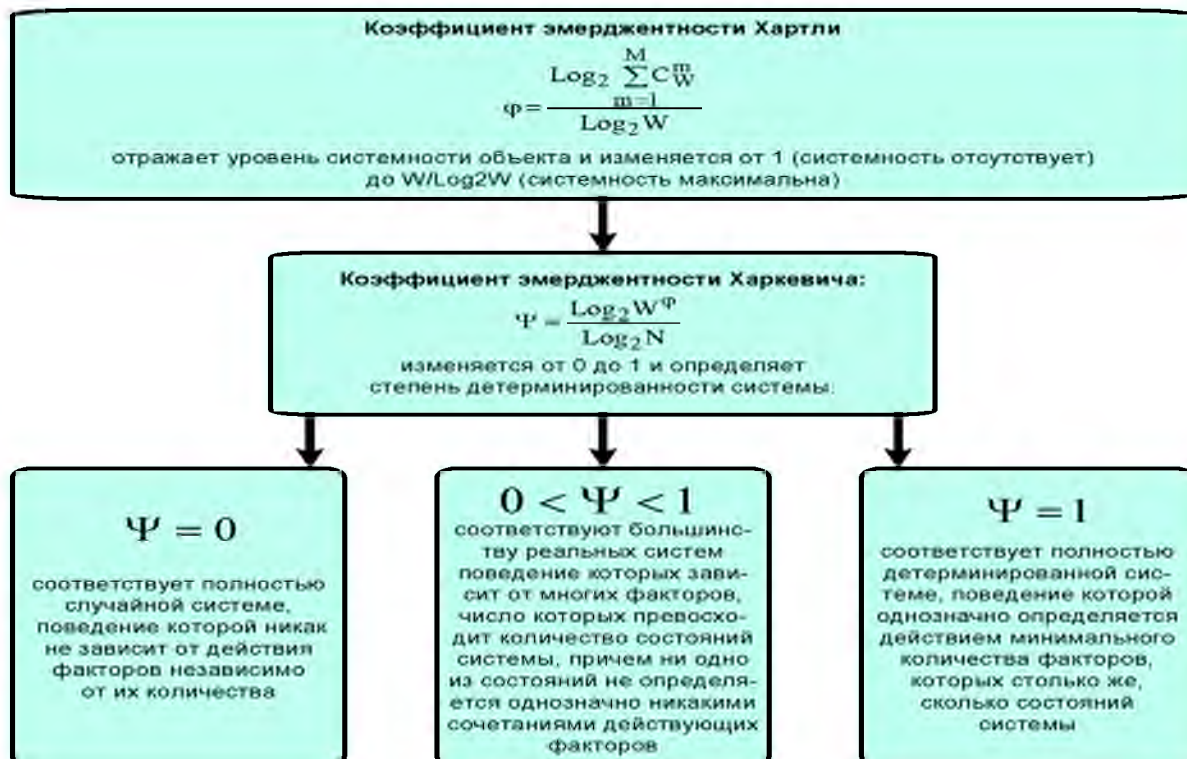


Рисунок 29. Интерпретация коэффициентов эмерджентности СТИ

Из выражения (3.25) видно, что в частном случае, когда реализуются только чистые состояния объекта управления, т.е. $M=1$, коэффициент эмерджентности А.Харкевича приобретает вид:

$$\Psi = \frac{\log_2 W}{\log_2 N} \quad (3.27)$$

Подставив коэффициент эмерджентности А.Харкевича (3.25) в выражение (3.22), получим:

$$\begin{aligned} I_{ij} &= \log_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^\Psi = \log_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^{\frac{\log_2 W^\varphi}{\log_2 N}} = \\ &= \frac{\log_2 W^\varphi}{\log_2 N} \left(\log_2 \left(\frac{N_{ij}}{N_i N_j} \right) + \log_2 N \right) = \\ &= \log_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\log_2 W^\varphi}{\log_2 N}} + \log_2 W^\varphi \end{aligned}$$

или окончательно:

$$I_{ij} = \log_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\log_2 W^\varphi}{\log_2 N}} + \log_2 W^\varphi \quad (3.28)$$

Из вида выражения (3.25) для Ψ очевидно, что увеличение уровня системности влияет на семантическую информационную модель (3.28) аналогично повышению уровня детерминированности системы: **понижение уровня системности, также как и степени детерминированности системы приводит к ослаблению влияния факторов на поведение системы, т.е. к понижению управляемости системы за счет своего рода "инфляции факторов".**

Например: управлять толпой из 1000 человек значительно сложнее, чем воздушно-десантным полком той же численности. Процесс превращения 1000 новобранцев в воздушно-десантный полк это и есть процесс повышения уровня системности и степени детерминированности системы. Этот процесс включает процесс иерархического структурирования (на отделения, взвода, роты, батальоны), а также процесс повышения степени детерминированности команд, путем повышения "степени беспрекословности" их исполнения. Оркестр, настраивающий инструменты, также весьма существенно отличается от оркестра, исполняющего произведение под управлением дирижера.

Необходимо отметить, что при повторном использовании *той же самой обучающей выборки* степень детерминированности модели уменьшается. Очевидно, с формальной математической точки зрения этого явления можно избежать, если перед расчетом информативностей признаков делить абсолютные частоты на количество объектов обучающей выборки.

С использованием выражения (3.28) непосредственно из матрицы абсолютных частот (таблица 12) рассчитывается матрица информативностей (таблица 13), содержащая связи между факторами и будущими состояниями АОУ и

имеющая много различных интерпретаций и играющая основополагающую роль в данном исследовании.

Таблица 13 – МАТРИЦА ИНФОРМАТИВНОСТЕЙ

Атрибуты		Классы - будущие состояния объекта управления					Средняя детерминирующая мощность фактора
		Целевые состояния			Нежелательные состояния		
		***	j	***	l	***	
Факторы, характеризующие текущее и прошлые состояния объекта управления, в т.ч. его рефлексивность	***						
	г	I_{gj}		I_{gl}		σ_g	

Управляющие факторы системы управления	i	I_{ij}		I_{il}		σ_i	

Факторы, характеризующие прошлые, текущее и прогнозируемые состояния окружающей среды	k	I_{kj}		I_{kl}		σ_k	

Средняя детерминированность будущих состояний АОУ			σ_j		σ_l		

Из рассмотрения основополагающего выражения (3.28) видно, что:

1. При выполнении условий взаимно-однозначного соответствия классов и признаков (3.23) первое слагаемое в выражении (3.28) обращается в ноль и при всех реальных значениях входящих в него переменных оно отрицательно.

2. Выражение (3.28) является нелинейной суперпозицией двух выражений: системного обобщения формулы Хартли (второе слагаемое), и первого слагаемого, которое имеет вид формулы Шеннона для плотности информации и отличается от него тем, что выражение под логарифмом находится в степени, которая совпадает с коэффициентом эмерджентности Харкевича, а также способом взаимосвязи входящих в него абсолютных частот с вероятностями.

Это дает основание предположить, что первое слагаемое в выражении (3.28) является одной из форм системного обобщения выражения Шеннона для плотности информации:

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\text{Log}_2 W^p}{\text{Log}_2 N}} \quad (3.29)$$

Поэтому вполне оправданным будет назвать степень в (3.29) коэффициентом эмерджентности Шеннона-Харкевича.

Генезис системной (эмерджентной) теории информации

Полученное системное обобщение формулы Харкевича (3.28) учитывает как взаимосвязь между признаками (факторами) и будущими, в т.ч. целевыми состояниями объекта управления, так и мощность множества будущих состояний объекта управления. Кроме того она объединяет возможности интегрального и дискретного описания объектов, учитывает уровень системности и степень детерминированности описываемой системы (таблица 14):

**Таблица 14 – СООТВЕТСТВИЕ ТРЕБОВАНИЯМ ФОРМУЛ
"КТИ / СТИ"**

	Хартли	Шеннон	Харкевич
Мощность множества состояний	+ / +	- / -	- / +
Факторы	- / -	+ / +	+ / +
Цель	- / -	- / -	+ / +
Уровень системности	- / +	- / +	- / +
Степень детерминированности	- / -	- / +	- / +
Рейтинг	1/2	1/3	2/5

При этом факторами являются управляющие факторы, т.е. управления со стороны системы управления, факторы окружающей среды, а также факторы, характеризующие текущее и прошлые состояния объекта управления. Все это делает полученное выражение (3.28) оптимальным по сформулированным критериям для целей построения содержательных информационных моделей активных объектов управления и для применения для синтеза адаптивных систем управления (см. диаграмму: "Генезис системного обобщения формулы Харкевича для количества информации", рисунок 30).

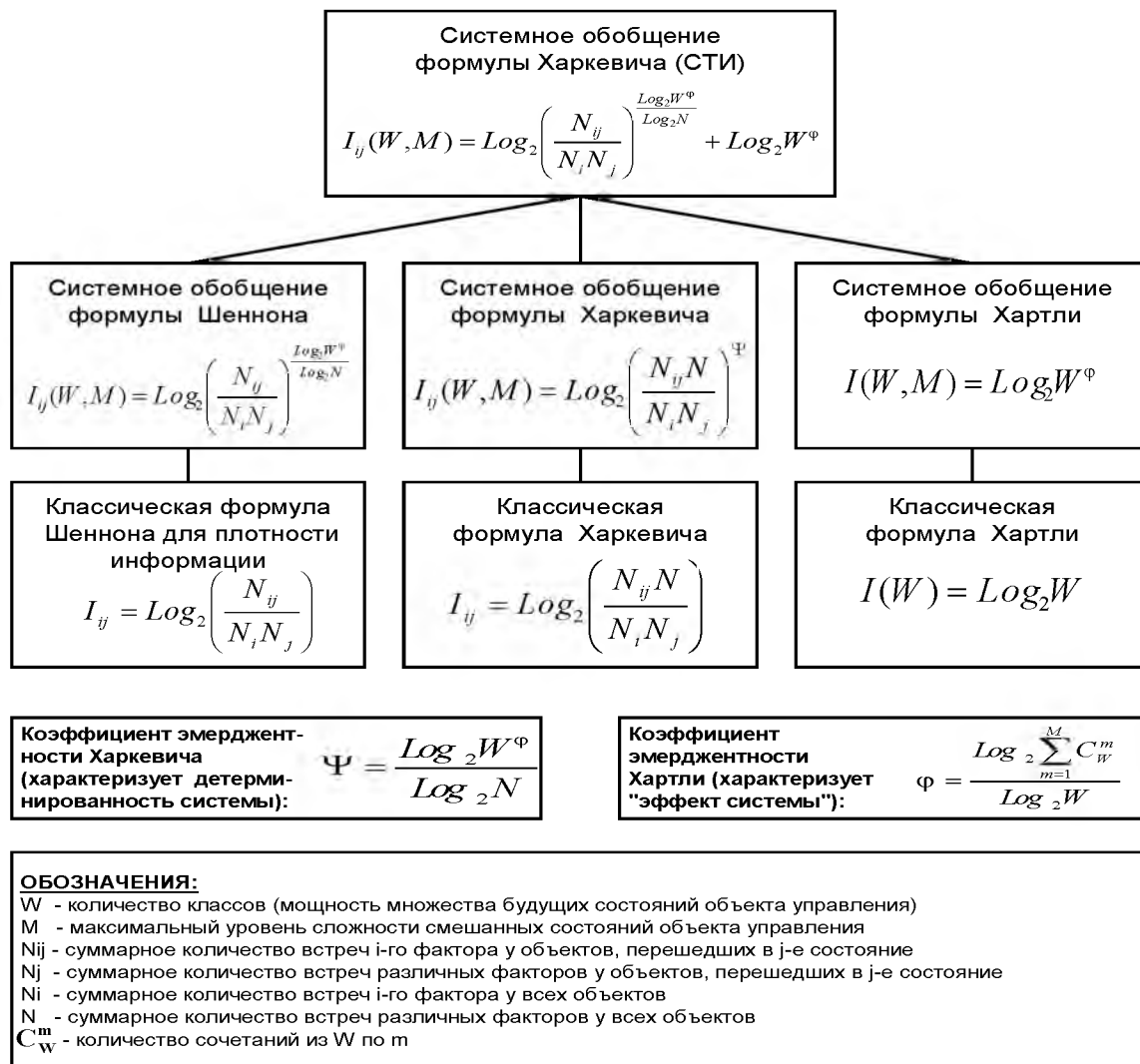


Рисунок 30. Генезис системной (эмерджентной) теории информации

Итак, различные выражения классической теории информации для количества информации: Хартли, Шеннона и Харкевича учитывают различные аспекты информационного моделирования объектов.

Полученное системное обобщение формулы А.Харкевича (3.28) учитывает как взаимосвязь между признаками (факторами) и будущими, в т.ч. целевыми состояниями объекта управления, так и мощность множества будущих состояний. Кроме того она объединяет возможности интегрального и дискретного описания объектов, учитывает уровень системности и степень детерминированности системы.

Различие между классическим понятием информации и его предложенным системным обобщением определяется различием между понятиями множества и системы, на основе которых они сформированы. Система при этом рассматривается как множество элементов, объединенных определенными видами взаимодействия ради достижения некоторой общей цели.

Все это делает полученное выражение (3.28) оптимальным по сформулированным критериям для целей построения содержательных информационных моделей активных объектов управления и для применения для синтеза рефлексивных АСУ активными объектами.

1.2.2.3. Семантическая информационная модель СК-анализа

Основная проблема, решаемая в аналитической модели: выбор способа вычисления весовых коэффициентов, отражающих степень и характер влияния факторов на переход активного объекта управления в различные состояния.

Основное отличие предлагаемого подхода от методов обобщения экспертных оценок состоит в том, что в предлагаемом подходе от экспертов требуется лишь само решение, а весовые коэффициенты автоматически подбираются в соответствии с моделью таким образом, что в сходных случаях будут приниматься решения, аналогичные предлагаемым экспертами. В традиционных подходах от экспертов требуют либо самих весовых коэффициентов, либо правил принятия решения (продукций).

1.2.2.3.1. Формализм динамики взаимодействующих семантических информационных пространств. Двухвекторное представление данных.

Не всегда и не все классы являются атрибутами, также не всегда и не все атрибуты являются классами по смыслу (в данной модели это может быть так в многослойной нейронной сети) Поэтому традиционное представление данных в форме одной матрицы с одинаковыми строками и столбцами представляется нецелесообразным и предлагается более общее – двухвекторное представление. В предлагаемой математической модели формальное описание объекта представляет собой совокупность его интенционального и экстенционального описаний.

Интенциональное (дискретное) описание – это последовательность информативностей (но не кодов) тех и только тех признаков, которые реально фактически встретились у данного конкретного объекта.

Экстенциональное (континуальное) описание состоит из информативностей (но не кодов) тех классов распознавания, для формирования образов которых по мнению экспертов целесообразно использовать интенциональное описание данного конкретного объекта.

Именно взаимодействие и взаимная дополнительность этих двух взаимоисключающих видов описания объектов формирует то, что психологи, логики и философы называют "смысл".

Таким образом, формальное описание объекта в предлагаемой модели состоит из двух векторов. Первый вектор описывает к каким обобщенным категориям (классам распознавания) относится объект с точки зрения экспертов (вектор субъективной, смысловой, человеческой оценки). Второй же вектор содержит информацию о том, какими признаками обладает данный объект (вектор объективных характеристик). Необходимо особо подчеркнуть, что связь этих двух векторов друг с другом имеет вообще говоря не детерминистский, а вероятностный, статистический характер.

Если объект описан обоими векторами, то это описание можно использовать для формирования обобщенных образов классов распознавания, а также для проверки степени успешности выполнения этой задачи.

Если объект описан только вторым вектором – вектором признаков, то его можно использовать только для решения задачи распознавания (идентификации), которую можно рассматривать как задачу восстановления вектора классов данного объекта по его известному вектору признаков.

Предлагаемая модель удовлетворяет принципу соответствия, т.е. в ней одновекторный вариант описания предметной области получается как некоторое подмножество из возможных в ней вариантов, определяемое двумя ограничениями:

- справочник классов распознавания тождественно совпадает со справочником признаков;
- наличие какого-либо признака у объекта обучающей выборки однозначно (детерминистским образом) определяет принадлежность этого объекта к соответствующему классу распознавания (взаимно-однозначное соответствие классов и признаков).

Очевидно, эти ограничения приводят и к соответствующим ограничениям, накладываемым в свою очередь на варианты обработки информации и анализа данных в подобных системах.

Если говорить конкретнее, такая модель данных стирает различие между атрибутами и классами и не позволяет решать ряд задач, в которых эта абстракция является недопустимым упрощением. Эти задачи будут подробнее рассмотрены ниже.

Семантические пространства классов и атрибутов

Наглядно модель данных целесообразно представить себе в виде двух взаимосвязанных фазовых (т.е. абстрактных) пространств, в первом из которых

осями координат служат шкалы атрибутов (пространство атрибутов), а во втором – шкалы классов (пространство классов).

В пространстве атрибутов векторами являются объекты обучающей выборки и обобщенные образы классов. Вектор класса представляет собой массив координат в фазовом пространстве, каждый элемент массива, т.е. координата, соответствует определенному атрибуту, а *значение этой координаты – весовому коэффициенту, отражающему количество информации, содержащееся в факте наблюдения данного атрибута у объекта о принадлежности этого объекта к данному классу.*

В пространстве классов векторами являются атрибуты. Вектор атрибута представляет собой массив координат в фазовом пространстве, каждый элемент массива, т.е. координата, соответствует определенному классу, а *значение этой координаты – весовому коэффициенту, отражающему количество информации, содержащееся в факте наблюдения объекта данного класса о том, что у этого объекта будет определенный атрибут.*

Таким образом, выбор смысла и математической формы значений весовых коэффициентов в виде количества информации вводит метрику в этих фазовых пространствах. *Поэтому данные пространства являются нелинейными самосогласованными пространствами.* Ясно, что линейная разделяющая функция в нелинейном пространстве является нелинейной функцией в линейном пространстве. Самосогласованность семантических пространств означает, что любое изменение одной координаты в общем случае связано с изменением всех остальных. Нелинейность и самосогласованность *самым существенным образом* отличает предложенные семантические информационные пространства классов и атрибутов от линейного семантического пространства, используемого в основном в психодиагностике [38], в котором осями являются признаки (шкалы), а значениями координат по осям являются непосредственно градации признаков.

Однако этого недостаточно. Чтобы над векторами в фазовых пространствах можно было корректно выполнять стандартные операции сложения, вычитания, скалярного и векторного умножения, выполнять преобразования системы координат, переход от одной системы координат к другой, и вообще применять аппарат линейной алгебры и аналитической геометрии, что представляет большой научный и практический интерес и является очень актуальным, необходимо корректно ввести в этих пространствах системы координат т.е. системы отсчета, удовлетворяющие определенным требованиям.

Требования к системам координат, свойства векторов классов и атрибутов, решение проблемы снижения размерности описания и ортонормирования

В качестве осей координат пространства атрибутов целесообразно выбрать вектора атрибутов, обладающие следующими свойствами:

1. Их должно быть минимальное количество, достаточное для полного описания предметной области.
2. Эти вектора должны пересекаться в одной точке.
3. Значения координат вектора должны измеряться в одной единице измерения, т.е. должны быть сопоставимы.

Для выполнения первого требования необходимо, чтобы математическая форма и смысл весовых коэффициентов были выбраны таким образом, чтобы модули векторов атрибутов в пространстве классов были пропорциональны их значимости для решения задач идентификации, прогнозирования и управления. Причем наиболее значимые вектора атрибутов не должны коррелировать друг с другом, т.е. должны быть ортонормированными. В этом случае при удалении векторов с минимальными модулями автоматически останутся наиболее значимые практически ортонормированные вектора, которые можно принять за базисные, т.е. в качестве осей системы координат.

Второе требование означает, что минимальное расстояние между этими векторами в пространстве классов должно быть равно нулю.

Третье требование предполагает соответствующий выбор математической формы для значений координат.

Эти идеальные требования практически никогда не будут соблюдаться на практике с абсолютной точностью. Однако этого и не требуется. Достаточно, чтобы реально выбранные в качестве базисных атрибуты отображались в пространстве классов векторами, для которых эти требования выполняются с точностью, достаточной для применения соответствующих математических моделей и математического аппарата на практике.

Аналогично обстоит дело и с минимизацией размерности пространства классов. В качестве базисных могут выбраны вектора классов, имеющие максимальную длину и взаимно (попарно) ортонормированные.

Очевидно, задача выбора базисных векторов имеет не единственное решение, т.е. может существовать несколько систем таких векторов, которые можно рассматривать как результат действия преобразований системы координат, состоящих из смещений и поворотов.

1.2.2.3.2. Применение классической теории информации К.Шеннона для расчета весовых коэффициентов и мер сходства

Формально, распознавание есть не что иное, как принятие решения о принадлежности распознаваемого объекта или его состояния к определенному классу (классам) [12, 125, 221]. Из этого следует внутренняя и органичная связь методов распознавания образов и принятия решений. Аналитический обзор позволяет сделать вывод, что наиболее глубокая основа этой связи состоит в том, что и распознавание образов, и принятие решений есть прежде всего снятие неопределенности. Распознавание снимает неопределенность в вопросе о том, к какому классу относится распознаваемый объект. Если до распознавания существовала неопределенность в вопросе о том, к какому классу относится распознаваемый объект или его состояние, то в результате распознавания эта неопределенность уменьшается, причем возможно и до нуля (когда объект идентифицируется однозначно). Принятие решения (выбор) также снимает неопределенность в вопросе о том, какое из возможных решений будет принято, если существовало несколько альтернативных вариантов решений и принимается одно из них.

Для строгого исследования процессов снятия неопределенности оптимальным является применение аппарата теории информации, которая как бы специально создана для этой цели. Из этого непосредственно следует возможность применения методов теории информации для решения задач распознавания и принятия решений в АСУ. Таким образом, теория информации может рассматриваться как единая основа методов распознавания образов и принятия решений.

Формальная постановка задачи

В рефлексивных АСУ активными объектами модели распознавания образов и принятия решений применимы в подсистемах идентификации состояния АОУ и выработки управляющего воздействия: идентификация состояния АОУ представляет собой принятие решения о принадлежности этого состояния к определенной классификационной категории (задача распознавания); выбор многофакторного управляющего воздействия из множества возможных вариантов представляет собой принятие решения (обратная задача распознавания).

Распознавание образов есть принятие решения о принадлежности объекта или его состояния к определенному классу. Если до распознавания существовала неопределенность в вопросе о том, к какому классу относится распознаваемый объект или его состояние, то в результате распознавания эта неопределенность уменьшается, в том числе может быть и до нуля (когда объект идентифицируется однозначно). Из данной постановки непосредственно следует возможность применения методов теории информации для решения задач распознавания образов и принятия решений в АСУ.

Информация как мера снятия неопределенности

Как было показано выше, теория информации применима в АСУ для решения задач идентификации состояния сложного объекта управления (задача распознавания) и принятия решения о выборе многофакторного управляющего воздействия (обратная задача распознавания).

Так в результате процесса познания уменьшается неопределенность в наших знаниях о состоянии объекта познания, а в результате процесса труда (по сути управления) – уменьшается неопределенность поведения продукта труда (или объекта управления). В любом случае количество переданной информации представляет собой количественную меру степени снятия неопределенности.

Процесс получения информации можно интерпретировать как изменение неопределенности в вопросе о том, от какого источника отправлено сообщение в результате приема сигнала по каналу связи. Подробно данная модель приведена в работе [81].

Количество информации в индивидуальных событиях и лемма Неймана–Пирсона

В классическом анализе Шеннона идет речь лишь о передаче символов по одному информационному каналу от одного источника к одному приемнику. Его интересует прежде всего передача самого сообщения.

В данном исследовании ставится другая задача: идентифицировать информационный источник по сообщению от него. Поэтому метод Шеннона был обобщен путем учета в математической модели возможности существования

многих источников информации, о которых к приемнику по зашумленному каналу связи приходят не отдельные символы–признаки, а сообщения, состоящие из последовательностей символов (признаков) любой длины.

Следовательно, ставится задача идентификации информационного источника по сообщению от него, полученному приемником по зашумленному каналу. Метод, являющийся обобщением метода К.Шеннона, позволяет применить классическую теорию информации для построения моделей систем распознавания образов и принятия решений, ориентированных на применение для синтеза адаптивных АСУ сложными объектами.

Для решения поставленной задачи необходимо вычислять не средние информационные характеристики, как в теории Шеннона, а количество информации, содержащееся в конкретном i -м признаке (символе) о том, что он пришел от данного j -го источника информации. Это позволит определить и суммарное количество информации в сообщении о каждом информационном источнике, что дает интегральный критерий для идентификации или прогнозирования состояния АОУ.

Логично предположить, что среднее количество информации, содержащейся в системе признаков о системе классов

$$I(Y, X) = \sum_{j=1}^W \sum_{i=1}^M p_{ij} \log_2 \frac{p_{ij}}{p_i p_j}, \quad (3.30)$$

является ничем иным, как усреднением (с учетом условной вероятности наблюдения) "индивидуальных количеств информации", которые содержатся в конкретных признаках о конкретных классах (источниках), т.е.:

$$i(x_j, y_i) = \log_2 \frac{p_{ij}}{p_i p_j}. \quad (3.31)$$

Это выражение определяет так называемую "плотность информации", т.е. количество информации, которое содержится в одном отдельно взятом факте наблюдения i -го символа (признака) на приемнике о том, что этот символ (признак) послан j -м источником.

Если в сообщении содержится M символов, то суммарное количество информации о принадлежности данного сообщения j -му информационному источнику (классу) составляет:

$$i(x_j) = \sum_{i=1}^M \log_2 \frac{p_{ij}}{p_i p_j}. \quad (3.32)$$

Необходимо отметить, что применение сложения в выражении (3.43) является вполне корректным и оправданным, так как информация с самого начала вводилась как аддитивная величина, для которой операция сложения является корректной.

Преобразуем выражение (3.50) к виду, более удобному для практического применения (численных расчетов). Для этого выразим вероятности встреч признаков через частоты их наблюдения:

$$P_{ij} = \frac{1}{N_{ij}}; P_i = \frac{1}{N_i}; P_j = \frac{1}{N_j}. \quad (3.33)$$

Подставив (3.44) в (3.25), получим:

$$i(x_j) = \sum_{i=1}^M \text{Log}_2 \frac{N_{ij}}{N_i N_j}. \quad (3.34)$$

Если ранжировать классы в порядке убывания суммарного количества информации о принадлежности к ним, содержащейся в данном сообщении (т.е. описании объекта), и выбирать первый из них, т.е. тот, о котором в сообщении содержится наибольшее количество информации, то мы получим обоснованную статистическую процедуру, основанную на классической теории информации, оптимальность которой доказывается в фундаментальной лемме Неймана–Пирсона [194].

Сравнивая выражения (3.34) и (3.28) видим, что в системное обобщенное формулы Харкевича входит слагаемое, *сходное* с выражением Шеннона для плотности информации. Различия состоят в том, что в выражении (3.28) это слагаемое возведено в степень, имеющую смысл коэффициента эмерджентности Харкевича. Необходимо отметить, что значения частот в этих формулах связаны с вероятностями несколько различным образом (выражения 3.20 и 3.33).

Если ранжировать классы в порядке убывания суммарного количества информации о принадлежности к ним, содержащейся в данном сообщении (т.е. описании объекта), и выбирать первый из них, т.е. тот, о котором в сообщении содержится наибольшее количество информации, то мы получим обоснованную статистическую процедуру, основанную на классической теории информации, оптимальность которой доказывается в фундаментальной лемме Неймана–Пирсона [194].

Таким образом, распознавание образов есть принятие решения о принадлежности объекта или его состояния к определенному классу. Если до распознавания существовала неопределенность в вопросе о том, к какому классу относится распознаваемый объект или его состояние, то в результате распознавания эта неопределенность уменьшается, в том числе может быть и до нуля. Понятие информации может быть определено следующим образом: "Информация есть количественная мера степени снятия неопределенности". Количество информации является мерой соответствия распознаваемого объекта (его состояния) обобщенному образу класса.

Количество информации имеет ряд вполне определенных свойств. Эти свойства позволяют ввести понятие "количество информации в индивидуальных событиях", которое является весьма перспективным для применения в системах распознавания образов и поддержки принятия решений.

1.2.2.3.3. Математическая модель метода распознавания образов и принятия решений, основанного на системной теории информации

Формальная постановка основной задачи рефлексивной АСУ активными объектами и ее декомпозиция

Рассмотрим некоторые основные понятия, необходимые для дальнейшего изложения. При этом будут использованы как литературные данные, так и результаты, полученные в предыдущих главах данной работы.

Принятие решения в АСУ – это выбор некоторого наиболее предпочтительного управляющего воздействия из исходного множества всех возможных управляющих воздействий, обеспечивающего наиболее эффективное достижение целей управления. В результате выбора неопределенность исходного множества уменьшается на величину информации, которая порождается самим актом выбора [81]. Следовательно, теория информации может быть применена как для идентификации состояний объекта управления, так и для принятия решений об управляющих воздействиях в АСУ.

Модель АСУ включает в себя: модель объекта управления, модель управляющей подсистемы, а также модель внешней среды. Управляющая подсистема реализует следующие функции: идентификация состояния объекта управления, выработка управляющего воздействия, реализация управляющего воздействия.

С позиций теории информации сложный объект управления (АОУ) может рассматриваться как шумящий (определенным образом) информационный канал, на вход которого подаются входные параметры \vec{X} , представляющие собой управляющие воздействия, а также факторы предыстории и среды, а на выходе фиксируются выходные параметры \vec{Y} , связанные как с входными параметрами, так и с целевыми и иными состояниями объекта управления.

Одной из основных задач АСУ является задача принятия решения о наиболее эффективном управляющем воздействии. В терминах теории информации эта задача формулируется следующим образом: *зная целевое состояние объекта управления, на основе его информационной модели определить такие входные параметры \vec{X} , которые с учетом предыстории и актуального состояния объекта управления, а также влияния среды с наибольшей эффективностью переведут его в целевое состояние, характеризующееся выходными параметрами \vec{Y} .*

С решением этой задачи тесно связана задача декодирования теории информации: *"По полученному в условиях помех сообщению определить, какое сообщение было передано"* [221]. Для решения данной задачи используются коды, корректирующие ошибки, а в более общем случае, – различные методы распознавания образов.

Учитывая вышесказанное, предлагается рассматривать принятие решения об управляющем воздействии в АСУ как решение обратной задачи декодирования, которая формулируется следующим образом: *"Какое сообщение необходимо подать на вход зашумленного канала связи, чтобы на его выходе получить заранее заданное сообщение"*. Данная задача решается на основе математической модели канала связи.

Декомпозиция основной задачи в ряд частных подзадач

Построение аналитической модели АОУ затруднено из-за отсутствия или недостатка априорной информации об объекте управления, а также из-за ограниченности и сложности используемого математического аппарата. В связи с этим предлагается путь решения данной проблемы, состоящий в поэтапном решении следующих задач:

1-я задача: разработать абстрактную модель более общего класса (содержательную информационную);

2-я задача: обучить абстрактную информационную модель путем учета информации о реальном поведении АОУ, поступающей в процессе экспериментальной эксплуатации АСУ; на этом этапе адаптируется и конкретизируется абстрактная модель АОУ, т.е. в ней все более точно отражаются взаимосвязи между входными параметрами и состояниями АОУ;

3-я задача: на основе конкретной содержательной информационной модели разработать алгоритмы решения следующих задач АСУ:

3.1. Расчет влияния факторов на переход АОУ в различные возможные состояния (обучение, адаптация).

3.2. Прогнозирование поведения АОУ при конкретном управляющем воздействии и выработка многофакторного управляющего воздействия (основная задача АСУ).

3.3. Выявление факторов, вносящих основной вклад в детерминацию состояния АОУ; контролируемое удаление второстепенных факторов с низкой дифференцирующей способностью, т.е. снижение размерности модели при заданных ограничениях.

3.4. Сравнение влияния факторов. Сравнение состояний АОУ.

Сформулируем предлагаемую абстрактную модель АОУ, опишем способ ее конкретизации и приведем алгоритмы решения задач адаптивных АСУ АОУ на основе данной модели.

Решение задачи 1: "Синтез семантической информационной модели активного объекта управления"

Исходные данные для выявления взаимосвязей между факторами и состояниями объекта управления предлагается представить в виде корреляционной матрицы – матрицы абсолютных частот (таблица 15):

В этой матрице в качестве классов (столбцов) приняты будущие состояния объекта управления, как целевые, так и нежелательные, а в качестве атрибутов (строк) – факторы, которые разделены на три основных группы, математически обрабатываемые единообразно: факторы, характеризующие текущее и прошлые состояния объекта управления; управляющие факторы системы управления; факторы, характеризующие прошлые, текущее и прогнозируемые состояния окружающей среды. Отметим, что форма таблицы 15 является универсальной формой представления и обобщения *фактов – эмпирических данных* в единстве их дискретного и интегрального представления (причины – следствия, факторы – результирующие состояния, признаки – обобщенные образы классов, обобщенное – логическое и т.п.).

Таблица 15 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ

Атрибуты		Классы - будущие состояния объекта управления					Сумма
		Целевые состояния		Нежелательные состояния			
		***	j	***	l	***	
Факторы, характеризующие текущее и прошлые состояния объекта управления, в т.ч. его рефлексивность	***						
	r		N_{rj}		N_{rl}		$N_r = \sum_{j=1}^w N_{rj}$

Управляющие факторы системы управления	i		N_{ij}		N_{il}		$N_i = \sum_{j=1}^w N_{ij}$

Факторы, характеризующие прошлые, текущее и прогнозируемые состояния окружающей среды	k		N_{kj}		N_{kl}		$N_k = \sum_{j=1}^w N_{kj}$

Сумма			$N_j = \sum_{i=1}^M N_{ij}$		$N_l = \sum_{i=1}^M N_{il}$		$N = \sum_{i=1}^M \sum_{j=1}^w N_{ij}$

где:

 N_{ij} – количество встреч i-го признака у объектов j-го класса по данным обучающей выборки

Управляющие факторы объединяются в группы, внутри каждой из которых они альтернативны (несовместны), а между которыми – нет (совместны). В этом случае внутри каждой группы выбирают одно из доступных управляющих воздействий с максимальным влиянием. Варианты содержательной информационной модели без учета прошлых состояний объекта управления и с их учетом, аналогичны, соответственно, простым и составным цепям Маркова, автоматам без памяти и с памятью.

В качестве количественной меры влияния факторов, предложено использовать обобщенную формулу А.Харкевича (3.28), полученную на основе предложенной эмерджентной теории информации. При этом по формуле (3.28) непосредственно из матрицы абсолютных частот (таблица 15) рассчитывается матрица информативностей (таблица 16), которая и представляет собой основу содержательной информационной модели предметной области.

Весовые коэффициенты таблицы 3.28 непосредственно определяют, какое количество информации I_{ij} система управления получает о наступлении события: "активный объект управления перейдет в j-е состояние", из сообщения: "на активный объект управления действует i-й фактор".

Принципиально важно, что эти весовые коэффициенты не определяются экспертами неформализуемым способом, а рассчитываются непосредственно на основе эмпирических данных и удовлетворяют всем ранее сформулированным требованиям, т.е. являются сопоставимыми, содержательно интерпретируемыми, отражают понятия "достижение цели управления" и "мощность множества будущих состояний объекта управления" и т.д.

Таблица 16 – МАТРИЦА ИНФОРМАТИВНОСТЕЙ

Атрибуты	Классы - будущие состояния объекта управления				Средняя детерминирующая мощность фактора
	Целевые состояния		Нежелательные состояния		
	***	j	***	l	***
Факторы, характеризующие текущее и прошлые состояния объекта управления, в т.ч. его рефлексивность	r	$I_{rj} = \Psi \cdot \text{Log}_2 \frac{N_{rj} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ij}}$		$I_{rl} = \Psi \cdot \text{Log}_2 \frac{N_{rl} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ij}}$	$\sigma_r = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{rj} - \bar{I}_r)^2}$

Управляющие факторы системы управления	i	$I_{ij} = \Psi \cdot \text{Log}_2 \frac{N_{ij} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ij}}$		$I_{il} = \Psi \cdot \text{Log}_2 \frac{N_{il} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ij}}$	$\sigma_i = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$

Факторы, характеризующие прошлые, текущее и прогнозируемые состояния окружающей среды	k	$I_{kj} = \Psi \cdot \text{Log}_2 \frac{N_{kj} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ij}}$		$I_{kl} = \Psi \cdot \text{Log}_2 \frac{N_{kl} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ij}}$	$\sigma_k = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{kj} - \bar{I}_k)^2}$

Средняя детерминированность будущих состояний АОУ		$\sigma_j = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (I_{ij} - \bar{I}_j)^2}$		$\sigma_l = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (I_{il} - \bar{I}_l)^2}$	$H = \sqrt{\frac{1}{(W \cdot M - 1) \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I}_j)^2}}$

$\bar{I}_j = \frac{1}{M} \sum_{i=1}^M I_{ij}$ – среднее значение координат вектора класса, M – количество факторов.

$\bar{I}_i = \frac{1}{W} \sum_{j=1}^W I_{ij}$ – среднее значение координат вектора фактора, W – количество классов (будущих состояний АОУ).

$\bar{I} = \frac{1}{W \cdot M} \sum_{j=1}^W \sum_{i=1}^M I_{ij}$ – средняя информативность признаков по матрице информативностей.

$\varphi = \frac{\log_2 \sum_{m=1}^M C_m}{\log_2 W}$ – коэффициент эмергентности Хартли.

$\Psi = \frac{\log_2 W^\varphi}{\log_2 N}$ – коэффициент эмергентности Харкевича,

H – мера уровня системности предметной области в рамках СИИ

В данной работе показано, что предложенная информационная мера обеспечивает сопоставимость индивидуальных количеств информации, содержащейся в факторах о классах, а также сопоставимость интегральных критериев, рассчитанных для одного объекта и разных классов, для разных объектов и разных классов.

Когда количество информации $I_{ij} > 0$ – i-й фактор способствует переходу объекта управления в j-е состояние, когда $I_{ij} < 0$ – препятствует этому переходу, когда же $I_{ij} = 0$ – никак не влияет на это. В векторе i-го фактора (строка матрицы информативностей) отображается, какое количество информации о переходе объекта управления в каждое из будущих состояний содержится в том факте, что данный фактор действует. В векторе j-го состояния класса (столбец матрицы информативностей) отображается, какое количество информации о переходе объекта управления в соответствующее состояние содержится в каждом из факторов.

Таким образом, матрица информативностей (таблица 16) является обобщенной таблицей решений, в которой входы (факторы) и выходы (будущие состояния АОУ) связаны друг с другом не с помощью классических (Аристотелевских) импликаций, принимающих только значения: "Истина" и "Ложь", а различными значениями истинности, выраженными в битах и принимающими значения от положительного теоретически-максимально-возможного ("Максимальная степень истинности"), до теоретически неограниченного отрицательного ("Степень ложности").

Фактически предложенная модель позволяет осуществить синтез обобщенных таблиц решений для различных предметных областей непосредственно

на основе эмпирических исходных данных и *продуцировать на их основе прямые и обратные правдоподобные (нечеткие) логические рассуждения по неклассическим схемам с различными расчетными значениями истинности*, являющимся обобщением классических импликаций (таблица 17).

Таблица 17 – ПРЯМЫЕ И ОБРАТНЫЕ ПРАВДОПОДОБНЫЕ ЛОГИЧЕСКИЕ ВЫСКАЗЫВАНИЯ С РАСЧЕТНОЙ (В СООТВЕТСТВИИ С СТИ) СТЕПЕНЬЮ ИСТИННОСТИ ИМПЛИКАЦИЙ

	Прямые высказывания:	Обратные высказывания
1	если A, то B (если действует фактор A , то мы предполагаем с степенью истинности I_{AB} , что АОУ перейдет в состояние B)	если B, то A (если АОУ перешел в состояние B , то мы предполагаем с степенью истинности I_{AB} , что действовал фактор A)
2	если A_1 и $A_2 \dots$ и A_m, то B (прогноз влияния системы факторов на поведение АОУ. Степень истинности обобщающей (итоговой) импликации равна алгебраической сумме истинностей составляющих ее элементарных импликаций вида: "если A то B ")	если B, то A_1 и $A_2 \dots$ и A_m (информационный портрет класса B , т.е. влияние различных факторов A_i на переход АОУ в будущее состояние B , решение обратной задачи прогнозирования, т.е. выработка управления)
3	если A, то B_1 или $B_2 \dots$ или B_w (семантический портрет фактора A , т.е. его влияние на переход АОУ в различные состояния)	
4	если A_1 и $A_2 \dots$ и A_m, то B_1 или $B_2 \dots$ или B_w (прогноз влияния системы факторов на переход АОУ в различные состояния)	

Приведем пример более сложного высказывания, которое может быть рассчитано непосредственно на основе матрицы информативностей – обобщенной таблицы решений (таблица 16):

Если A , со степенью истинности $\alpha(A,B)$ детерминирует B , и если C , со степенью истинности $\alpha(C,D)$ детерминирует D , и A совпадает по смыслу с C со степенью истинности $\alpha(A,C)$, то это вносит вклад в совпадение B с D , равный степени истинности $\alpha(B,D)$.

При этом в прямых рассуждениях как предпосылки рассматриваются факторы, а как заключение – будущие состояния АОУ, а в обратных – наоборот: как предпосылки – будущие состояния АОУ, а как заключение – факторы. Степень истинности i -й предпосылки – это просто количество информации I_{ij} , содержащейся в ней о наступлении j -го будущего состояния АОУ. Если предпосылок несколько, то степень истинности наступления j -го состояния АОУ равна суммарному количеству информации, содержащемуся в них об этом. Количество информации в i -м факторе о наступлении j -го состояния АОУ, рассчитывается в соответствии с выражением (3.28) СТИ.

Прямые правдоподобные логические рассуждения позволяют *прогнозировать* степень достоверности наступления события по действующим факторам, а обратные – по заданному состоянию восстановить степень необходимости и степень нежелательности каждого фактора для наступления этого состояния, т.е. принимать решение по выбору *управляющих воздействий* на АОУ, оптимальных для перевода его в заданное целевое состояние.

Необходимо отметить, что предложенная модель, основывающаяся на теории информации, обеспечивает *автоматизированное формирования системы нечетких правил по содержимому входных данных*, как и комбинация нечеткой логики Заде-Коско с нейронными сетями Кохонена. Принципиально важно, что качественное изменение модели путем добавления в нее новых классов не уменьшает достоверности распознавания уже сформированных классов. Кроме того, при сравнении распознаваемого объекта с каждым классом учитываются не только признаки, имеющиеся у объекта, но и отсутствующие у него, поэтому *предложенной моделью правильно идентифицируются объекты, признаки которых образуют множества, одно из которых является подмножеством другого* (как и в Неокогнитроне К.Фукушимы) [243].

Данная модель позволяет прогнозировать поведение АОУ при воздействии на него не только одного, но и целой системы факторов:

$$I_j = f(\vec{I}_{ij}). \quad (3.35)$$

В теории принятия решений скалярная функция I_j векторного аргумента называется интегральным критерием. Основная проблема состоит в выборе такого аналитического вида функции интегрального критерия, который обеспечил бы эффективное решение сформулированной выше задачи АСУ.

Учитывая, что частные критерии (3.28) имеют смысл количества информации, а информация по определению является аддитивной функцией, предлагается ввести интегральный критерий, как аддитивную функцию от частных критериев в виде:

$$I_j = (\vec{I}_{ij}, \vec{L}_i). \quad (3.36)$$

В выражении (3.54) круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^A I_{ij} L_i, \quad (3.37)$$

где:

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор j -го состояния объекта управления;

$\vec{L}_i = \{L_i\}$ – вектор состояния предметной области, включающий все виды факторов, характеризующих объект управления, возможные управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ \alpha_i, & \text{где } 0 < \alpha_i < 1, \text{ если } i - \text{й фактор действует с истинностью } \alpha_i; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В реализованной модели значения координат вектора состояния ПО принимались равными либо 1 (фактор действует), либо 0 (фактор не действует).

Таким образом, интегральный критерий представляет собой суммарное количество информации, содержащееся в системе факторов различной природы (т.е. факторах, характеризующих объект управления, управляющее воздей-

ствие и окружающую среду) о переходе активного объекта управления в будущее (в т.ч. целевое или нежелательное) состояние.

В многокритериальной постановке задача прогнозирования состояния объекта управления, при оказании на него заданного многофакторного управляющего воздействия I_j , сводится к максимизации интегрального критерия:

$$j^* = \arg \max_{j \in J} ((\vec{I}_{ij}, \vec{L}_i)), \quad (3.38)$$

т.е. к выбору такого состояния объекта управления, для которого интегральный критерий максимален.

Задача принятия решения о выборе наиболее эффективного управляющего воздействия является *обратной задачей* по отношению к задаче максимизации интегрального критерия (идентификации и прогнозирования), т.е. вместо того, чтобы по набору факторов прогнозировать будущее состояние АОУ, наоборот, по заданному (целевому) состоянию АОУ определяется такой набор факторов, который с наибольшей эффективностью перевел бы объект управления в это состояние.

Предлагается еще одно обобщение этой фундаментальной леммы, основанное на косвенном учете корреляций между информативностями в векторе состояний при использовании средних по векторам. Соответственно, вместо простой суммы количеств информации предлагается использовать корреляцию между векторами состояния и объекта управления, которая количественно измеряет степень сходства этих векторов:

$$I_j = \frac{1}{\sigma_j \sigma_l A} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}), \quad (3.39)$$

где:

\bar{I}_j – средняя информативность по вектору класса;

\bar{L} – среднее по вектору идентифицируемой ситуации (объекта).

σ_j – среднеквадратичное отклонение информативностей вектора класса;

σ_l – среднеквадратичное отклонение по вектору распознаваемого объекта.

Выражение (3.39) получается непосредственно из (3.37) после замены координат перемножаемых векторов их стандартизированными значениями:

$$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_l}.$$

Необходимо отметить, что выражение для интегрального критерия сходства (3.39) по своей математической форме является корреляцией двух векторов. Это означает, что если эти вектора являются суммой двух сигналов: полезного и белого шума, то при расчете интегрального критерия белый шум практически не будет играть никакой роли, т.е. его корреляция с самими собой равна нулю по определению. Это означает, что **выбранный интегральный критерий сходства является высокоэффективным средством подавления белого шума и выделе-**

ния полезной информации из шума, который неизбежно присутствует в эмпирических данных.

Результат прогнозирования поведения объекта управления, описанного данной системой факторов, представляет собой список его возможных будущих состояний, в котором они расположены в порядке убывания суммарного количества информации о переходе объекта управления в каждое из них.

Сравнения результатов идентификации и прогнозирования с опытными данными, с использованием выражений (3.37) и (3.39), показали, что при малых выборках они практически не отличаются, но при увеличении объема выборки до 400 и более (независимо от предметной области) выражение (3.39) дает погрешность идентификации (прогнозирования) на 5% – 7% меньше, чем (3.37). Поэтому в предлагаемой модели фактически используется не метрическая мера сходства (3.39).

В связи с тем, что в дальнейшем изложении широко применяются понятия теории АСУ, теории информации (связи), теории распознавания образов и методов принятия решений, приведем таблицу соответствия наиболее часто используемых нами терминов из этих научных направлений, имеющих сходный смысл (таблица 18):

**Таблица 18 – СООТВЕТСТВИЕ ТЕРМИНОВ
РАЗЛИЧНЫХ НАУЧНЫХ НАПРАВЛЕНИЙ**

Теория автоматизированного управления	Теория распознавания образов	Методы принятия решений	Теория информации (связи)
Фактор	Признак	Критерий	Входной сигнал
Состояние объекта управления	Класс распознавания	Альтернатива	Информационный источник
Идентификация состояния объекта управления	Распознавание	Принятие (синтез) решения	Декодирование
Выработка управляющего воздействия	Решение обратной задачи распознавания (прогнозирования)	Анализ решения	Решение обратной задачи декодирования

Вывод системного обобщения формулы Харкевича (3.28) приведен в разделе 3.1 данной работы. Чрезвычайно важное для данного исследования выражение (3.28) заслуживает специального комментария. Прежде всего нельзя не обратить внимания на то, что оно *по своей математической форме, т.е. формально, ничем не отличается от выражения для превышения сигнала над помехой* для информационного канала [242]. Из этого, на первый взгляд, внешнего совпадения следует интересная интерпретация выражения (3.28). А именно: можно считать, что обнаружив некоторый i -й признак у объекта, предъявленного на распознавание, мы тем самым получаем сигнал, содержащий некоторое количество информации

$$D_{ij} = \text{Log}_2 P_{ij}$$

о том, что этот объект принадлежит к j -му классу. По-видимому, это так и есть, однако чтобы оценить насколько много или мало этой информации нами получено, ее необходимо с чем-то сравнить, т.е. необходимо иметь точку отсчета или

базу для сравнения. В качестве такой базы естественно принять среднее по всем признакам количество информации, которое мы получаем, обнаружив этот j -й класс:

$$\bar{I}_j = \text{Log}_2 P_j.$$

Иначе говоря, если при предъявлении какого-либо объекта на распознавание у него обнаружен i -й признак, то для того, чтобы сделать из этого факта обоснованный вывод о принадлежности этого объекта к тому или иному классу, необходимо знать и учесть, насколько часто вообще (т.е. в среднем) обнаруживается этот признак при предъявлении объектов данного класса.

Фактически это среднее количество информации можно рассматривать как некоторый "информационный шум", который имеется в данном признаке и не несет никакой полезной информации о принадлежности объектов к тем или иным классам. Полезной же информацией является степень отличия от этого шума. Таким образом классическому выражению Харкевича (3.12) для семантической целесообразности информации может быть придан более привычный для теории связи вид:

$$I_{ij} = \text{Log} P_{ij} - \text{Log} P_j = D_{ij} - \bar{I}_j,$$

который интерпретируется как *вычитание шума из полезного сигнала*. Эта операция является совершенно стандартной в системах шумоподавления.

Если полезный сигнал выше уровня шума, то его обнаружение несет информацию в пользу принадлежности объекта к данному классу, если нет – то, наоборот, в пользу не принадлежности.

Возвращаясь к выражению (3.12), необходимо отметить, что сам А.А.Харкевич рассматривал P_{ij} как вероятность достижения цели, при условии, что система получила информацию I_{ij} , а P_j – как вероятность ее достижения при условии, что система этой информации не получала. Очевидно, что фактически P_j соответствует вероятности случайного угадывания объектом управления правильного пути к цели, или, что тоже самое, вероятности самопроизвольного, т.е. без оказания управляющих воздействий, достижения АОУ целевого заданного состояния.

Необходимо отметить также, что каждый признак объекта управления как канала связи может быть охарактеризован динамическим диапазоном, равным разности максимально возможного (допустимого) уровня сигнала в канале и уровня помех в логарифмическом масштабе:

$$I_{ij \max} = \text{Log} P_{ij \max} - \text{Log} P_j = \text{Log}(P_{ij \max} / P_j).$$

Максимальное количество информации, которое может содержаться в признаке, полностью определяется количеством классов распознавания W и равно количеству информации по Хартли: $I = \text{Log}_2 W$.

Динамический диапазон признака является количественной мерой его полезности (ценности) для распознавания, но все же предпочтительней для этой цели является *среднее количество полезной для классификации информации в*

признаке, т.е. исправленное выборочное среднеквадратичное отклонение информативностей:

$$\sigma_i = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2} \quad (3.40)$$

Очевидна близость этой меры к *длине вектора признака* в семантическом пространстве атрибутов:

$$d_i = \sqrt{\sum_{j=1}^W I_{ij}^2} \quad (3.41)$$

В сущности выражение (3.40) просто представляет собой *нормированный* вариант (3.41).

Решение задачи 2: "Адаптация модели объекта управления"

На основе обучающей выборки, содержащей информацию о том, какие факторы действовали, когда АОУ переходил в те или иные состояния, методом прямого счета формируется матрица абсолютных частот, имеющая вид, представленный в таблице 15. Необходимо отметить, что в случае АОУ в большинстве случаев нет возможности провести полный факторный эксперимент для заполнения матрицы абсолютных частот. В данной работе предполагается, что это и не обязательно, т.е. на практике достаточно воспользоваться *естественной вариативностью* факторов и состояний АОУ, представленных в обучающей выборке. *С увеличением объема обучающей выборки в ней со временем будут представлены все практически встречающиеся варианты сочетаний факторов и состояний АОУ.*

В соответствии с выражением (3.28), непосредственно на основе матрицы абсолютных частот $\|N_i^j\|$ (таблица 15) рассчитывается матрица информативностей факторов $\|I_i^j\|$ (таблица 16).

Количество информации в i-м факторе о наступлении j-го состояния АОУ является статистической мерой их связи и количественной мерой влияния данного фактора на переход АОУ в данное состояние.

Решение задачи 3: "Разработка алгоритмов решения основных задач АСУ"

Как было показано в разделе 3.2, решение задачи 3 предполагает решение следующих подзадач.

Решение подзадачи 3.1: "Расчет влияния факторов на переход объекта управления в различные состояния (обучение, адаптация)"

При изменении объема обучающей выборки или изменении экспертных оценок прежде всего пересчитывается матрица абсолютных частот, а затем, на ее основании и в соответствии с выражением (3.28), – матрица информативностей. Таким образом, *предложенная модель обеспечивает отображение динамических взаимосвязей, с одной стороны, между входными и выходными парамет-*

рами, а с другой, – между параметрами и состояниями объекта управления. Конкретно, это отображение осуществляется в форме так называемых векторов факторов и состояний.

В профиле (векторе) i -го фактора (строка матрицы информативностей) отображается, какое количество информации о переходе АОУ в каждое из возможных состояний содержится в том факте, что данный фактор действует.

В профиле (векторе) j -го состояния АОУ (столбец матрицы информативностей) отображается, какое количество информации о переходе АОУ в данное состояние содержится в каждом из факторов.

Решение подзадачи 3.2: "Прогнозирование поведения объекта управления при конкретном управляющем воздействии и выработка многофакторного управляющего воздействия (обратная задача прогнозирования)"

Прогнозирование состояния АОУ осуществляется следующим образом:

1. Собирается информация о действующих факторах, характеризующих состояние предметной области (активный объект управления описывается факторами, характеризующими его текущее и прошлые состояния; управляющая система характеризуется технологическими факторами, с помощью которых она оказывает управляющее воздействие на активный объект управления; окружающая среда характеризуется прошлыми, текущими и прогнозируемыми факторами, которые также оказывают воздействие на активный объект управления).

2. Для каждого возможного будущего состояния АОУ подсчитывается суммарное количество информации, содержащееся во всей системе факторов (согласно п.1), о наступлении этого состояния.

3. Все будущие состояния АОУ ранжируются в порядке убывания количества информации об их осуществлении.

Этот ранжированный список будущих состояний АОУ и представляет собой первичный результат прогнозирования.

Если задано некоторое определенное целевое состояние, то выбор *управляющих воздействий* для фактического применения производится из списка, в котором все возможные управляющие воздействия расположены в порядке убывания их влияния на перевод АОУ в данное целевое состояние. Такой список называется информационным портретом состояния АОУ [81].

Управляющие воздействия могут быть объединены в группы, внутри каждой из которых они альтернативны (несовместны), а между которыми – нет (совместны). В этом случае внутри каждой группы выбирают одно из фактически доступных управляющих воздействий с максимальным влиянием на достижение заданного целевого состояния АОУ.

Однако выбор многофакторного управляющего воздействия нельзя считать завершенным без прогнозирования результатов его применения. Описание АОУ в актуальном состоянии состоит из списка факторов окружающей среды, предыстории АОУ, описания его актуального (исходного) состояния, а также выбранных управляющих воздействий. Имея эту информацию по каждому из факторов в соответствии с выражением (3.39), нетрудно подсчитать, какое количество информации о переходе в каждое из состояний содержится суммарно во

всей системе факторов. Данный метод соответствует фундаментальной лемме Неймана–Пирсона, содержащей доказательство оптимальности метода выбора той из двух статистических гипотез, о которой в системе факторов содержится больше информации. В то же время он является обобщением леммы Неймана–Пирсона, так как вместо информационной меры Шеннона используется системное обобщение семантической меры целесообразности информации Харкевича.

Предлагается еще одно обобщение этой фундаментальной леммы, основанное на косвенном учете корреляций между информативностями в профиле состояния при использовании среднего по профилю. Соответственно, вместо простой суммы количеств информации предлагается использовать ковариацию между векторами состояния и АОУ, которая количественно измеряет степень сходства формы этих векторов.

Результат прогнозирования поведения АОУ, описанного данной системой факторов, представляет собой список состояний, в котором они расположены в порядке убывания суммарного количества информации о переходе АОУ в каждое из них.

Решение подзадачи 3.3: "Выявление факторов, вносящих основной вклад в детерминацию состояния АОУ; снижение размерности модели при заданных ограничениях"

Естественно считать, что некоторый фактор является тем более ценным, чем больше среднее количество информации, содержащееся в этом факторе о поведении АОУ [81]. Но так как в предложенной модели количество информации может быть и отрицательным (если фактор уменьшает вероятность перехода АОУ в некоторое состояние), то простое среднее арифметическое информативностей может быть близко к нулю. При этом среднее будет равно нулю и в случае, когда все информативности равны нулю, и тогда, когда они будут велики по модулю, но с разными знаками. Следовательно, более адекватной оценкой полезности фактора является среднее модулей или, что наиболее точно, исправленное (несмещенное) среднеквадратичное отклонение информативностей по профилю признака.

Ценность фактора по сути дела определяется его полезностью для различения состояний АОУ, т.е. является его дифференцирующей способностью или селективностью.

Необходимо также отметить, что различные состояния АОУ обладают различной степенью обусловленности, т.е. в различной степени детерминированы факторами: некоторые слабо зависят от учтенных факторов, тогда как другие определяются ими практически однозначно. *Количественно детерминируемость состояния АОУ предлагается оценивать стандартным отклонением информативностей вектора обобщенного образа данного состояния.*

Предложено и реализовано несколько итерационных алгоритмов корректного удаления малозначимых факторов и слабодетерминированных состояний АОУ при заданных граничных условиях [81]. *Решение задачи снижения размерности модели АОУ при заданных граничных условиях позволяет снизить эксплуатационные затраты и повысить эффективность РАСУ АО.*

Решение подзадачи 3.4: "Сравнение влияния факторов. Сравнение состояний объекта управления"

Факторы могут сравниваться друг с другом по тому влиянию, которое они оказывают на поведение АОУ. Сами состояния могут сравниваться друг с другом по тем факторам, которые способствуют или препятствуют переходу АОУ в эти состояния. Это сравнение может содержать лишь результат, т.е. различные степени сходства/различия (в кластерном анализе), или содержать также причины этого сходства/различия (в когнитивных диаграммах).

Эти задачи играют важную роль в теории и практике РАСУ АО при необходимости замены одних управляющих воздействий другими, но аналогичными по эффекту, а также при изучении вопросов семантической устойчивости управления (различимости состояний АОУ по детерминирующим их факторам).

Этот анализ проводится над классами распознавания и над признаками. Он включает: информационный (ранговый) анализ; кластерный и конструктивный анализ, семантические сети; содержательное сравнение информационных портретов, когнитивные диаграммы.

Семантический информационный анализ

Предложенная математическая модель позволяет сформировать информационные портреты обобщенных эталонных образов классов распознавания и признаков.

Портреты классов распознавания представляют собой списки признаков в порядке убывания содержащегося в них количества информации о принадлежности к этим классам.

Информационный портрет класса распознавания показывает нам, каков информационный вклад каждого признака в общий объем информации, содержащейся в обобщенном образе этого класса.

В подходе к решению задач рефлексивных АСУ АО, основанном на применении методов распознавания образов, классам распознавания соответствуют, во-первых, исходные, а во-вторых, результирующие, в том числе целевые состояния объекта управления. Это значит, что в первом случае портреты классов используются для идентификации исходного состояния АОУ, потому что именно с ними сравнивается состояние объекта управления, а во втором – для выработки управляющего воздействия, так как его выбирают в форме суперпозиции неальтернативных факторов из информационного портрета целевого состояния, оказывающих наибольшее влияние на перевод АОУ в это состояние.

Портреты признаков представляют собой списки классов распознавания в порядке убывания количества информации о них, которое содержит данный признак. По своей сути информационный портрет признака раскрывает нам смысл данного признака, т.е. его семантическую нагрузку. *В теории и практике рефлексивных АСУ АО информационный портрет фактора является развернутой количественной характеристикой, содержащей информацию о силе и характере его влияния на перевод АОУ в каждое из возможных результирующих состояний, в том числе в целевые.* Информационные портреты классов и признаков допускают наглядную графическую интерпретацию в виде двухмерных (2d) и трехмерных (3d) диаграмм.

Кластерно-конструктивный анализ и семантические сети

Кластеры представляют собой такие группы классов распознавания (или признаков), внутри которых эти классы наиболее схожи друг с другом, а между которыми наиболее различны [81]. В данной работе, в качестве классов распознавания рассматриваются как исходные, так и результирующие, в том числе целевые состояния объекта управления, а в качестве признаков – факторы, влияющие на переход АОУ в результирующие состояния.

Исходные состояния АОУ, объединенные в кластер, характеризуются общими или сходными методами перевода в целевые состояния. Результирующие состояния АОУ, объединенные в кластер, являются слабоблагодарными по факторам, детерминирующим перевод АОУ в эти состояния. Это означает, что одно и то же управляющее воздействие при одних и тех же предпосылках (исходном состоянии и предыстории объекта управления и среды) могут привести к переводу АОУ в одно из результирующих состояний, относящихся к одному кластеру. Поэтому кластерный анализ результирующих состояний АОУ является инструментом, позволяющим изучать вопросы устойчивости управления сложными объектами.

При выборе управляющего воздействия как суперпозиции неальтернативных факторов часто возникает вопрос о замене одних управляющих факторов другими, имеющими сходное влияние на перевод АОУ из данного текущего состояния в заданное целевое состояние. Кластерный анализ факторов как раз и позволяет решить эту задачу: при невозможности применить некоторый управляющий фактор его можно заменить другим фактором из того же кластера.

При формировании кластеров используются матрицы сходства объектов и признаков, формируемые на основе матрицы информативностей.

В соответствии с предлагаемой математической моделью могут быть сформированы кластеры для заданного диапазона кодов классов распознавания (признаков) или заданных диапазонов уровней системной организации с различными критериями включения объекта (признака) в кластер.

Эти критерии могут быть сформированы автоматически либо заданы непосредственно. В последнем уровне кластеризации, в частности при задании одного уровня, в кластеры включаются не только похожие, но и все непохожие объекты (признаки), и, таким образом, формируются конструкторы классов распознавания и признаков.

В данной работе под конструктором понимается система противоположных (наиболее сильно отличающихся) кластеров, которые называются "полюсами" конструктора, а также спектр промежуточных кластеров, к которым применима количественная шкала измерения степени их сходства или различия [81].

Понятия "кластер" и "конструктор" тесно взаимосвязаны:

– так как положительный и отрицательный полюса конструктора представляют собой кластеры, в наибольшей степени отличающиеся друг от друга, то конструкторы могут быть получены как результат кластерного анализа кластеров;

– конструкт может рассматриваться как кластер с нечеткими границами, включающий в различной степени, причем не только в положительной, но и отрицательной, все классы (признаки).

В теории рефлексивных АСУ АО, конструктивный анализ позволяет решить такие задачи, как:

1. Определение в принципе совместимых и в принципе несовместимых целевых состояний АОУ. Совместимыми называются целевые состояния, для достижения которых необходимы сходные предпосылки и управляющие воздействия, а несовместимыми – для которых они должны быть диаметрально противоположными. Например, обычно сложно совмещаются такие целевые состояния, как очень высокое качество продукции и очень большое ее количество.

2. Определение факторов, имеющих не только сходное (это возможно и на уровне кластерного анализа), но и совершенно противоположное влияние на поведение сложного объекта управления.

Современный интеллект имеет дуальную структуру и, по сути дела, мыслит в системе кластеров и конструктов. Поэтому инструмент автоматизированного кластерно–конструктивного анализа может быть успешно применен для рефлексивного управления активными объектами.

Необходимо отметить, что формирование кластеров затруднено из-за проблемы комбинаторного взрыва, так как требует полного перебора и проверки "из n по m " сочетаний элементов (классов или признаков) в кластеры. Конструкты же формируются непосредственно из матрицы сходства прямой выборкой и сортировкой, что значительно проще в вычислительном отношении, так как конструктов значительно меньше, чем кластеров (всего n^2). Поэтому учитывая, что *при формировании конструктов автоматически формируются и их полюса, т.е. кластеры, в предложенной математической модели реализован не кластерный анализ, а сразу конструктивный (как более простой в вычислительном отношении и более ценный по получаемым результатам).*

Диаграммы смыслового сходства–различия классов (признаков) соответствуют определению семантических сетей [81], т.е. представляют собой ориентированные графы, в которых признаки соединены линиями, соответствующими их смысловому сходству–различию.

Когнитивные диаграммы классов и признаков

В предложенной в настоящем исследовании математической модели в обобщенной постановке реализована возможность содержательного сравнения обобщенных образов классов распознавания и признаков, т.е. построения когнитивных диаграмм [81].

В информационных портретах классов распознавания мы видим, какое количество информации о принадлежности (или не принадлежности) к данному классу мы получаем, обнаружив у некоторого объекта признаки, содержащиеся в информационном портрете. В кластерно-конструктивном анализе мы получаем результаты сравнения классов распознавания друг с другом, т.е. мы видим, насколько они сходны и насколько отличаются. *Но мы не видим, какими призна-*

ками они похожи и какими отличаются, и какой вклад каждый признак вносит в сходство или различие некоторых двух классов.

Эту информацию мы могли бы получить, если бы проанализировали и сравнили два информационных портрета. Эту работу и осуществляет режим содержательного сравнения классов распознавания.

Аналогично, в информационных портретах признаков мы видим, какое количество информации о принадлежности (или не принадлежности) к различным классам распознавания мы получаем, обнаружив у некоторого объекта данный признак. В кластерно-конструктивном анализе мы получаем результаты сравнения признаков друг с другом, т.е. мы видим, насколько они сходны и насколько отличаются. *Но мы не видим, какими классами они похожи и какими отличаются, и какой вклад каждый класс вносит в смысловое сходство или различие некоторых двух признаков.*

Эту информацию мы могли бы получить, если бы проанализировали и сравнили информационные портреты двух признаков. Эту работу и осуществляет режим содержательного (смыслового) сравнения признаков.

Содержательное (смысловое) сравнение классов

Обобщим математическую модель, предложенную и развиваемую в данной главе, на случай содержательного сравнения двух классов распознавания: J-го и L-го.

Признаки, которые есть по крайней мере в одном из классов, будем называть связями, так как благодаря тому, что они либо тождественны друг другу, либо между ними имеется определенное сходство или различие по смыслу, они вносят определенный вклад в отношения сходства/различия между классами.

Список выявленных связей сортируется в порядке убывания модуля силы связи, причем учитывается не более заданного количества связей.

Пусть, например:

у J-го класса обнаружен i-й признак,

у L-го класса обнаружен k-й признак.

Используем те же обозначения, что и в разделе 3.1.

На основе обучающей выборки системой рассчитывается матрица абсолютных частот встреч признаков по классам (таблица 15).

В разделе 3.1. получено выражение (3.28) для расчета количества информации в i-м признаке о принадлежности некоторого конкретного объекта к j-му классу (плотность информации), которое имеет вид:

$$I_{ij} = \log_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\log_2 W^\varphi}{\log_2 N}} + \log_2 W^\varphi \quad (3.28)$$

Аналогично, формула для количества информации в k-м признаке о принадлежности к L-му классу имеет вид:

$$I_{kl} = \log_2 \left(\frac{N_{kl}}{N_k N_l} \right)^{\frac{\log_2 W^\varphi}{\log_2 N}} + \log_2 W^\varphi \quad (3.42)$$

Вклад некоторого признака i в сходство/различие двух классов j и l равен соответствующему слагаемому корреляции образов этих классов, т.е. просто произведению информативностей

$$R_{jl} = (I_{ij} - \bar{I}_j)(I_{il} - \bar{I}_l). \quad (3.43)$$

Классический коэффициент корреляции Пирсона, количественно определяющий степень сходства векторов двух классов: j и l , на основе учета вклада каждой связи, образованной i -м признаком, рассчитывается по формуле

$$K_{jl} = \frac{1}{MS_j S_l} \sum_{i=1}^M (I_{ij} - \bar{I}_j)(I_{il} - \bar{I}_l), \quad (3.44)$$

где:

$$\bar{I}_j = \frac{1}{M} \sum_{i=1}^M I_{ij} \quad - \text{средняя информативность признаков } j\text{-го класса};$$

$$\bar{I}_l = \frac{1}{M} \sum_{i=1}^M I_{il} \quad - \text{средняя информативность признаков } l\text{-го класса};$$

$$S_j = \sqrt{\frac{\sum_{i=1}^M (I_{ij} - \bar{I}_j)^2}{M-1}} \quad - \text{среднеквадратичное отклонение информативностей признаков } j\text{-го класса};$$

$$S_l = \sqrt{\frac{\sum_{i=1}^M (I_{il} - \bar{I}_l)^2}{M-1}} \quad - \text{среднеквадратичное отклонение информативностей признаков } l\text{-го класса}.$$

Проанализируем, насколько классический коэффициент корреляции Пирсона (3.62) пригоден для решения важных задач:

- содержательного сравнения классов;
- изучения внутренней многоуровневой структуры класса.

Упростим анализ, считая, что средние информативности признаков по обоим классам близки к нулю, что при достаточно больших выборках (более 400 примеров в обучающей выборке) практически близко к истине.

Каждое слагаемое (3.43) суммы (3.44) отражает связь между классами, образованную одним i -м признаком. I -я связь существует в том и только в том случае, если i -й признак есть у обоих классов. Поэтому эти связи уместно называть одно–однозначными.

Этот подход можно назвать классическим для когнитивного анализа. Рассмотрим когнитивную диаграмму, приведенную на стр. 222 работы основной работы классика когнитивной психологии Р.Солсо (Когнитивная психология. /Пер. с англ. - М.: Тривола, 1996. - 600с.) (рисунок 31).

Рис. 7.3. Атрибуты двух групп ("птица" и "малиновка") с высокой степенью пересечения.

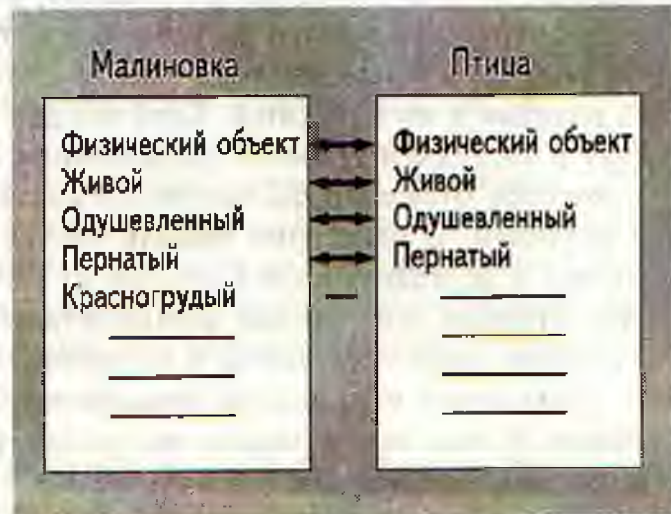


Рисунок 31. Когнитивная диаграмма из классической работы Роберта Солсо.

В приведенной когнитивной диаграмме наглядно в графической форме показано сравнение классов (обобщенных образов) "Малиновка" и "Птица" разных уровней общности по их атрибутам (признакам). Как видно из диаграммы, в ней:

1. Все атрибуты имеют одинаковый вес, т.е. не учитывается, что некоторые атрибуты более важны для идентификации класса, чем другие. Это соответствует предположению, что этот вес равен по модулю 1 для всех атрибутов.

2. Все признаки имеют одинаковый знак, т.е. они все характерны для классов и нет атрибутов нехарактерных. Это соответствует предположению, что вес всех признаков положительный, т.е. все признаки вносят вклад в сходство и нет признаков, вносящих вклад в различие.

3. Классы сравниваются только по тем атрибутам, которые есть одновременно у них обоих, т.е. признаки, имеющиеся у обоих классов вносят вклад в сходство классов, а признаки, которые есть только у одного из классов не вносят никакого вклада ни в сходство классов, ни в различие. Это соответствует предположению, что атрибуты ортонормированы, т.е. корреляция их друг с другом равна 0 (атрибуты семантически не связаны).

Каждое из этих трех допущений является довольно сильным и желательно их снять и, тем самым, обобщить принцип построения когнитивных диаграмм, приведенный в данном примере.

Но это означает, что данный подход не позволяет сравнивать классы, описанные различными, т.е. непересекающимися наборами признаков. Но даже если общие признаки и есть, то невозможность учета вклада остальных признаков является недостатком классического подхода, так как из содержательного анализа связей неконтролируемо исключается потенциально существенная информация. Таким образом, классический подход имеет ограниченную применимость при решении задачи №1. Для решения задачи №2 подход, основанный на формуле (3.44), вообще не применим, так как различные уровни системной организации классов образованы различными признаками и, следовательно, между уровнями не будет ни одной одно-однозначной связи.

Основываясь на этих соображениях, предлагается в общем случае учитывать вклад в сходство/различие двух классов, который вносят не только общие, но и остальные признаки. Логично предположить, что этот вклад (при прочих равных условиях) будет тем меньше, чем меньше корреляция между этими признаками.

Следовательно, для обобщения выражения для силы связи (3.43) необходимо умножить произведение информативностей признаков на коэффициент корреляции между ними, *отражающий степень сходства или различия признаков по смыслу*.

Таким образом, будем считать, что любые два признака (i,k) вносят определенный вклад в сходство/различие двух классов (j,l), определяемый сходством/различием признаков и количеством информации о принадлежности к этим классам, которое содержится в данных признаках:

$$R_{jl} = K_{ik} (I_{ij} - \bar{I}_j)(I_{kl} - \bar{I}_l), \quad (3.45)$$

где: K_{ik} – классический коэффициент корреляции Пирсона, количественно определяющий степень сходства по смыслу двух признаков: i и k, на основе учета вклада каждой связи, образованной содержащейся в них информацией о принадлежности к j-му классу

$$K_{ik} = \frac{1}{WS_i S_k} \sum_{j=1}^W (I_{ij} - \bar{I}_i)(I_{kj} - \bar{I}_k), \quad (3.46)$$

где

$$\bar{I}_i = \frac{1}{W} \sum_{j=1}^W I_{ij} \quad \text{– средняя информативность координат вектора i-го признака;}$$

$$\bar{I}_k = \frac{1}{W} \sum_{j=1}^W I_{kj} \quad \text{– средняя информативность координат вектора k-го признака;}$$

$$S_i = \sqrt{\frac{\sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}{W-1}} \quad \text{– среднее квадратичное отклонение координат вектора i-го признака;}$$

$$S_k = \sqrt{\frac{\sum_{j=1}^W (I_{kj} - \bar{I}_k)^2}{W-1}} \quad \text{– среднее квадратичное отклонение координат вектора k-го признака.}$$

Коэффициент корреляции между признаками (3.46) рассчитывается на основе *всей* обучающей выборки, а не только объектов двух сравниваемых классов. Так как коэффициент корреляции между признаками (3.46) практически всегда не равен нулю, то *каждый признак i образует связи со всеми признаками k*, где $k=\{1,...,A\}$, а каждый признак k в свою очередь связан со всеми остальными признаками. Это означает, что *выражение (3.45) является обобщением*

(3.43) с учетом много-многозначных связей.

На основе этих представлений сформулируем выражение для обобщенного коэффициента корреляции Пирсона между двумя классами: j и l , учитывающего вклад в их сходство/различие не только одно-однозначных, но и много-многозначных связей, образуемых коррелирующими признаками. Когнитивные диаграммы с много-многозначными связями предлагается называть обобщенными когнитивными диаграммами.

$$L_{jl} = \frac{1}{MS_j S_l} \sum_{i=1}^M \sum_{k=1}^M K_{ik} (I_{ij} - \bar{I}_j)(I_{kl} - \bar{I}_l) \quad (3.47)$$

где K_{ik} определяется выражением (3.46).

Сравним классический (3.44) и обобщенный (3.47) коэффициенты корреляции Пирсона друг с другом. Очевидно, при $i=k$ (3.47) преобразуется в (3.44), т.е. *соблюдается принцип соответствия*. Отметим, что модель позволяет задавать минимальный коэффициент корреляции (порог) между признаками, образующими учитываемые связи. При пороге 100% отображаются только одно-однозначные связи, учитываемые в классическом коэффициенте корреляции (3.44). Из выражений (3.47) и (3.44) видно, что

$$L_{jl} \geq K_{jl}, \quad (3.48)$$

так как в обобщенном коэффициенте корреляции учитываются связи между классами, образованные за счет учета корреляций между различными признаками. Ясно, что отношение

$$\frac{L_{jl}}{K_{jl}} \geq 1 \quad (3.49)$$

отражает *степень избыточности описания классов*. В модели имеется возможность исключения из системы признаков наименее ценных из них для идентификации классов. При этом в первую очередь удаляются сильно коррелирующие друг с другом признаки. В результате степень избыточности системы признаков уменьшается, и она становится ближе к ортонормированной.

Рассмотрим вопрос о единицах измерения, в которых количественно выражаются связи между классами.

Сходство двух признаков K_{ik} выражается величиной от -1 до $+1$.

Максимальная теоретически возможная информативность признака в *Bit* выражается формулой

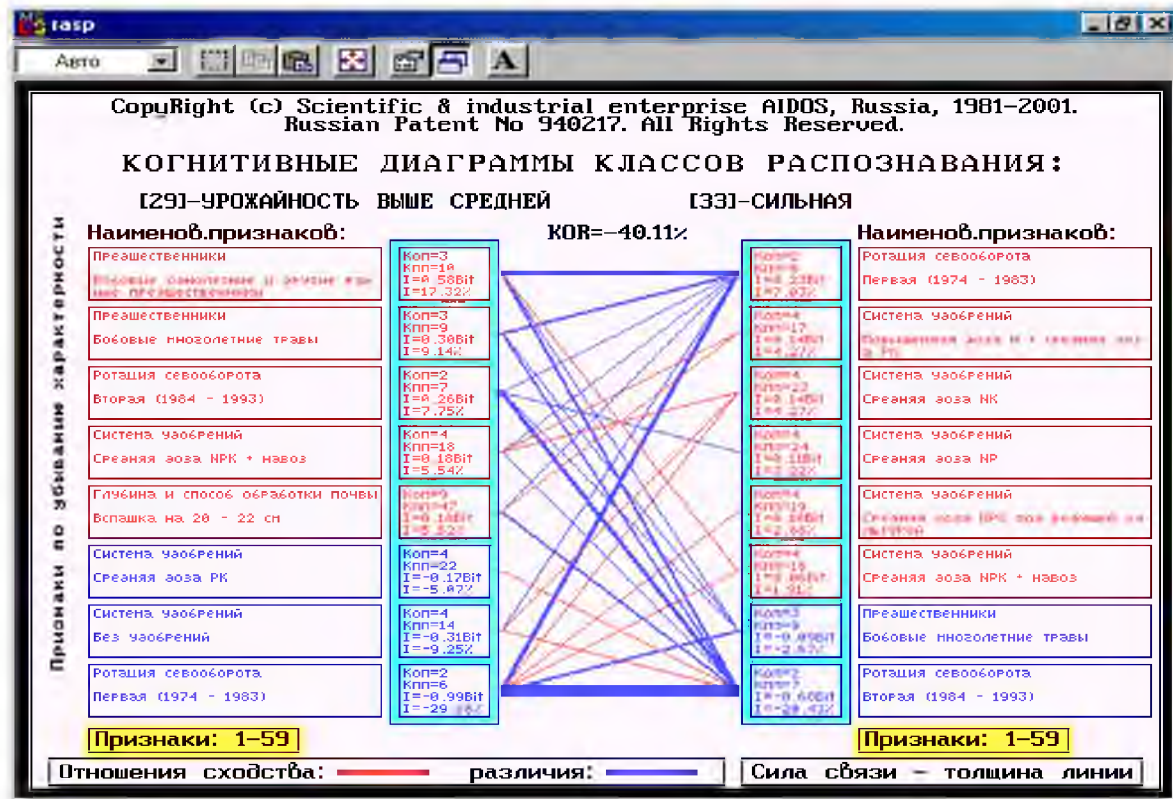
$$I_{\max} = \log_2(W), \quad (3.50)$$

Таким образом, учитывая выражения (3.45) и (3.50) получаем, что максимальная теоретически возможная сила связи R_{\max} равна

$$R_{\max} = I_{\max}^2. \quad (3.51)$$

В разработанном инструментарии СК-анализа, реализующем данную модель (описанном в лекции 6), реализован режим отображения когнитивной графики, где фактическая сила связи (3.45) в когнитивных диаграммах выражается

в процентах от максимальной теоретически возможной силы связи (3.50). На графической диаграмме (рисунок 32) отображается 8 наиболее сильных по модулю связей, рассчитанных согласно формулы (3.47), причем знак связи изображается цветом (красный +, синий –), а величина – толщиной линии.



**Рисунок 32. Когнитивная диаграмма
конструкта классов "Качество-количество"**

Имеется возможность выводить диаграммы только с положительными или только с отрицательными связями (для не цветных принтеров).

Частным случаем предложенных в данной работе обобщенных когнитивных диаграмм являются известные диаграммы В.С.Мерлина (Очерк интегрального исследования индивидуальности. - М., 1986. - 187с.). Эти диаграммы представляют обобщенные когнитивные диаграммы, формируемые в соответствии с предложенной моделью при следующих граничных условиях:

1. Класс сравнивается сам с собой.
2. Фильтрация левого и правого информационных портретов выбрана по уровням системной организации признаков (в данном случае – уровням Мерлина, терм. авт.).
3. Левый класс отображается с фильтрацией по одному уровню системной организации, а правый – по другому.
4. Диалог задания вида диаграмм предоставляет пользователю возможность задать следующие параметры:
 - способ нормирования толщины линий, отображающих связи: нормирование по текущей диаграмме или по всем диаграммам;

- способ фильтрации признаков в информационных портретах диаграммы: по диапазону признаков или по диапазону уровней системной организации (уровням Мерлина);

- сами диапазоны признаков или уровней для левого и правого информационных портретов;

- максимальное количество связей, отображаемых на диаграмме;

- уровень сходства признаков, образующих одну связь, отображаемую на диаграмме: от 0 до 100%. При уровне сходства 100% в диаграммах отображаются только связи, образованные теми признаками, которые есть в обоих портретах одновременно, т.е. взаимно-однозначные связи. При уровне сходства менее 100% вообще говоря связи становятся много-многозначными, так как каждый признак корреляционно связан со всеми остальными;

- уровень сходства классов, отображаемых на диаграмме.

Таким образом, в предлагаемой математической модели в общем виде реализована возможность содержательного сравнения обобщенных образов состояний АОУ и факторов, т.е. построения когнитивных диаграмм [81], веса атрибутов определяются автоматически на основе исходных данных в соответствии с математической моделью и могут принимать различные по величине положительные и отрицательные значения. Кроме того на основе кластерного анализа атрибутов определяются корреляции между ними, которые учитываются при определении вклада атрибутов в сходство или различие классов. Поэтому отношения между атрибутами разных классов в когнитивной диаграмме не "один к одному", как в диаграмме на рисунке 31, а "многие ко многим" (рисунок 32).

В информационном портрете состояния АОУ показано, какое количество информации о принадлежности (не принадлежности) АОУ к данному состоянию, а также о переходе (не переходе) АОУ в данное состояние содержится в том факте, что на АОУ действуют факторы, содержащиеся в данном информационном портрете.

Кластерно-конструктивный анализ дает результат сравнения состояний АОУ друг с другом, т.е. показывает, насколько эти состояния сходны друг с другом и насколько отличаются друг от друга. *Но он не показывает, какими факторами эти состояния АОУ похожи и какими отличаются, и какой вклад каждый фактор вносит в сходство или различие каждых двух состояний*. Чтобы получить эту информацию, необходимо проанализировать два информационных портрета, что и делается при содержательном сравнении состояний АОУ.

Смысл и значение диаграмм Мерлина применительно к проблематике АСУ состоит в том, что они наглядно представляют внутреннюю структуру детерминации состояний АОУ, т.е. показывают, каким образом связаны друг с другом факторы и будущие состояния АОУ.

Таким образом:

- для моделирования процессов принятия решений в рефлексивных АСУ активными системами целесообразно применение многокритериального подхода с аддитивным интегральным критерием, в котором в качестве частных критериев используется семантическая мера целесообразности информации (Харкевич, 1960);

– предложенная математическая модель обеспечивает эффективное решение следующих задач, возникающих при синтезе адаптивных АСУ АОУ: разработка абстрактной информационной модели АОУ; адаптация и конкретизация абстрактной модели на основе апостериорной информации о реальном поведении АОУ; расчет влияния факторов на переход АОУ в различные возможные состояния; прогнозирование поведения АОУ при конкретном управляющем воздействии и выработка многофакторного управляющего воздействия (основная задача АСУ); выявление факторов, вносящих основной вклад в детерминацию состояния АОУ; контролируемое удаление второстепенных факторов с низкой дифференцирующей способностью, т.е. снижение размерности модели при заданных ограничениях; сравнение влияния факторов, сравнение целевых и других состояний АОУ.

Предложенная методология, основанная на теории информации, обеспечивает эффективное моделирование задач принятия решений в адаптивных АСУ сложными системами.

Содержательное (смысловое) сравнение признаков

Предложенная математическая модель позволяет осуществить содержательное сравнение информационных портретов двух признаков.

Выявляются классы, которые есть по крайней мере в одном из векторов. Такие классы называются связями, так как благодаря тому, что они либо тождественны друг другу, либо между ними имеется определенное сходство или различие, они вносят определенный вклад в отношения сходства/различия между признаками по смыслу.

Все связи между признаками сортируются в порядке убывания модуля, в соответствии с определенными ограничениями, связанными с тем, что нет необходимости учитывать очень слабые связи.

Для каждого класса известно, какое количество информации о принадлежности к нему содержит данный признак – это информативность. Здесь необходимо уточнить, что информативность признака – это не только количество информации в признаке о принадлежности к данному классу, но и количество информации в классе о том, что при нем наблюдается данный признак, т.е. это взаимная информация класса и признака.

Если бы классы были тождественны друг другу, т.е. это был бы один класс, то его вклад в сходство/различие двух признаков был бы просто равен соответствующему данному классу слагаемому корреляции этих признаков, т.е. просто произведению информативностей.

Но поскольку это в общем случае это могут быть различные классы, то, очевидно, необходимо умножить произведение информативностей на коэффициент корреляции между классами.

Таким образом, будем считать, что любые два класса (j,l) вносят определенный вклад в сходство/различие двух признаков (i,k) , определяемый сходством/различием этих классов и количеством информации о принадлежности к ним, которое содержится в данных признаках

$$L_{ik} = \frac{1}{WS_i S_k} \sum_{j=1}^W \sum_{l=1}^W K_{jl} (I_{ij} - \bar{I}_i)(I_{kl} - \bar{I}_k). \quad (3.52)$$

Вывод формулы (3.52) обобщенного коэффициента корреляции Пирсона для двух признаков совершенно аналогичен выводу формулы (3.47), поэтому он здесь не приводится. Формулы для всех входящих в (3.52) величин приведены выше в предыдущем разделе.

Так же, как и в режиме содержательного сравнения классов, в данном режиме сила связи выражается в процентах от максимальной теоретически-возможной силы связи. На диаграммах отображается 16 наиболее значимых связей, рассчитанных согласно этой формуле, причем знак связи изображается цветом (красный +, синий –), а величина – толщиной линии. Имеется возможность вывода диаграмм только с положительными или только с отрицательными связями.

Математическая модель позволяет получить обобщенные инвертированные когнитивные диаграммы для любых двух заданных признаков, для пар наиболее похожих и непохожих признаков, для всех их возможных сочетаний, а также инвертированные диаграммы Мерлина.

Необходимо отметить, что понятия, соответствующие по смыслу терминам "обобщенная инвертированная когнитивная диаграмма" и "инвертированная диаграмма Мерлина" не упоминаются даже в фундаментальных руководствах по когнитивной психологии и впервые предложены автором в [125]. Эти диаграммы представляют собой частный случай обобщенных когнитивных диаграмм признаков, формируемых в соответствии с предложенной математической моделью при следующих ограничениях:

1. Признак сравнивается сам с собой.
2. Выбрана фильтрация левого и правого вектора по уровням системной организации классов (аналог уровней Мерлина для свойств).
3. Левый вектор отображается с фильтрацией по одному уровню системной организации классов, а правый – по другому.

Обоснование сопоставимости частных критериев I_{ij}

Применение этого метода корректно, если можно сравнивать суммарное количество информации о переходе АОУ в различные состояния, рассчитанное в соответствии с выражением (3.44), т.е. если они сопоставимы друг с другом.

Будем считать, что величины сопоставимы тогда и только тогда, когда одновременно выполняются следующие три условия:

1. Сопоставимы индивидуальные количества информации, содержащейся в признаках о принадлежности к классам.
2. Сопоставимы величины, рассчитанные для одного объекта и разных классов.
3. Сопоставимы величины, рассчитанные для разных объектов и разных классов.

Очевидно, для решения всех этих вопросов необходимо дать точное и полное определение самого термина "сопоставимость".

Считается, что величины сопоставимы, если существует некоторая количественная шкала для измерения этих величин.

Таким образом, в нашем случае сопоставимость обеспечивается, если на шкале определены направление и единица измерения, а также есть абсолютный минимум (ноль) или максимум.

Докажем теоремы о выполнении условий сопоставимости для упрощенной и полной информационных моделей объектов и классов распознавания. Для этого рассмотрим вышеперечисленные необходимые и достаточные условия сопоставимости для упрощенной и полной информационных моделей.

Теорема-1: Индивидуальные количества информации, содержащейся в признаках объекта о принадлежности к классам, сопоставимы между собой.

В упрощенной информационной модели класса и информационной модели объекта принято, что все признаки имеют одинаковый вес, который равен 1, если признак есть у класса, и 0, если его нет. Уже одним этим обеспечивается сопоставимость индивидуальных количеств информации в упрощенной модели.

В полной модели количество информации рассчитывается в соответствии с модифицированной формулой Харкевича (3.28). Таким образом, в полной информационной модели класса для каждого признака известно, какое количество информации о принадлежности к данному классу он содержит. Это количество информации может быть положительным, нулевым и отрицательным, но не может превосходить некоторой максимальной величины, определяемой количеством классов распознавания: $I = \log_2 W$ (мера Хартли), где W – количество классов распознавания. Следовательно, для полной информационной модели сопоставимость индивидуальных количеств информации также обеспечивается, так как для них применима шкала отношений.

Это означает, что индивидуальные количества информации можно суммировать и ввести интегральный критерий как аддитивную меру от индивидуальных количеств информации, что и требовалось доказать.

Теорема-2: Величины суммарной информации, рассчитанные для одного объекта и разных классов, сопоставимы друг с другом.

В упрощенной информационной модели вариант расстояния Хэмминга H^j , в котором учитываются только совпадения единиц (т.е. существующих признаков), для кодовых слов объекта и класса равно:

$$H^j = \sum_{i=1}^M I_i^j L_i, \quad (3.53)$$

где I_i^j – кодовое слово (профиль, массив–локатор) j -го класса;

$$I_i^j = \begin{cases} 1, & \text{если признак есть;} \\ 0, & \text{если признака нет;} \end{cases}$$

L_i – кодовое слово (профиль, массив–локатор) объекта.

$$L_i = \begin{cases} 1, & \text{если признак есть;} \\ 0, & \text{если признака нет.} \end{cases}$$

Пусть длина кодового слова (количество признаков) равна A . Длины кодовых слов объекта и классов одинаковы. Признаки могут принимать значения $\{0,1\}$. Тогда из этих условий и выражения (3.53) следует:

$$0 \leq H^j \leq M. \quad (3.54)$$

Но выражение (3.54) является математическим определением шкалы отношений, что означает полную сопоставимость предложенной меры сходства для упрощенной информационной модели одного объекта и многих классов. Для обобщенной информационной модели этот вывод сохраняет силу, т.к. в этой модели информация в соответствии с выражением (3.28) измеряется в единицах измерения – битах, определенных на шкале измерения информации, и на этой шкале имеется 0 и теоретический максимум, определяемый в соответствии с выражением Хартли. В полной информационной модели мера сходства объекта с классом I^j имеет вид, определяемый выражением (3.39).

Очевидно, величина I^j нормирована:

$$-1 \leq I^j \leq +1, \quad (3.55)$$

что и доказывает применимость шкалы отношений и полную сопоставимость меры сходства для полной информационной модели одного объекта и многих классов.

Это значит, что можно сравнивать меры сходства данного объекта с каждым из классов и ранжировать классы в порядке убывания сходства с данным объектом, что и требовалось доказать.

Теорема-3: Величины суммарной информации, рассчитанные для разных объектов и разных классов, а также классов и классов, признаков и признаков, взаимно-сопоставимы.

Очевидно, величина I^j , рассчитанная по формуле (3.39) для различных объектов и классов нормирована:

$$-1 \leq I^j \leq +1, \quad (3.56)$$

что и доказывает применимость шкалы отношений и полную сопоставимость мер сходства для полной информационной модели многих объектов и многих классов.

Это значит, что можно сравнивать меры сходства различных объектов с классами распознавания и делать выводы о том, что одни объекты распознаются лучше, а другие хуже на данном наборе классов и признаков, что и т.д.

Аналогичные рассуждения верны и для сравнения векторов классов друг с другом, а также векторов признаков друг с другом, что позволяет применить модели кластерно-конструктивного анализа и алгоритмы построения семантических сетей, что и требовалось доказать.

Теорема-4: Неметрический интегральный критерий сходства, основанный на модифицированной формуле А.Харкевича и обобщенной лемме Неймана-Пирсона, аддитивен.

Рассмотрим информационные модели распознаваемого объекта и классов распознавания, т.е. модели, основанные на теории кодирования – декодирования

и расстоянии Хэмминга (кодовое расстояние) в качестве критерия сходства. Эта модель является упрощенной, но достаточно адекватной для решения вопроса об аддитивности меры сходства объектов и классов.

Информационная модель распознаваемого объекта представляет собой двоичное слово, каждый разряд которого соответствует определенному признаку. Если признак есть у распознаваемого объекта, то соответствующий разряд имеет значение 1, если нет – то 0. Двоичное слово с установленными в 1 разрядами, соответствующими признакам распознаваемого объекта, называется его кодовым словом.

Упрощенная информационная модель класса распознавания есть двоичное слово, каждый разряд которого соответствует определенному признаку. Соответствие между двоичными разрядами и признаками для классов то же самое, что и для распознаваемых объектов. Если признак есть у класса, то соответствующий разряд имеет значение 1, если нет – то 0. Двоичное слово с установленными в 1 разрядами, соответствующими признакам класса, называется его кодовым словом.

Такая модель класса является упрощенной, так как в ней принято, что все признаки имеют одинаковый вес равный 1, если он есть у класса, и 0, если его нет, тогда как *в полной информационной модели класса для каждого признака известно, какое количество информации о принадлежности к данному классу он содержит*. Это количество информации может быть положительным, нулевым и отрицательным, но не может превосходить некоторой максимальной величины, определяемой количеством классов распознавания: $I = \log_2 W$ (мера Хартли), где W – количество классов.

Таким образом, в упрощенной информационной модели различные классы распознавания отличаются друг от друга только наборами признаков, которые им соответствуют.

При использовании этих упрощенных моделей задача распознавания объекта сводится к задаче декодирования, т.е. *кодовые слова объектов рассматриваются как искаженные зашумленным каналом связи кодовые слова классов*. Распознавание состоит в том, что по кодовому слову объекта определяется наиболее близкое ему в определенном смысле кодовое слово класса. При этом естественной и наиболее простой мерой сходства между распознаваемым объектом и классом является расстояние Хэмминга между их кодовыми словами, т.е. количество разрядов, которыми они отличаются друг от друга.

Рассмотрим теперь вопрос об *аддитивности количества информации как частного критерия в интегральном критерии*.

Известно [194], что существует всего два варианта формирования интегрального критерия из частных критериев: аддитивный и мультипликативный, поэтому задача сводится к выбору одного из этих вариантов. Следует отметить, что аддитивный критерий может рассматриваться как логарифм мультипликативного, а т.к. логарифм является монотонно-возрастающей функцией, то в принципе разница между этими критериями для принятия решений не так уж и велика, что подтверждается и в их сравнительных численных исследованиях.

Рассмотрим эти варианты. Пусть кодовое слово объекта состоит из N разрядов. Тогда добавление еще одного разряда, отображающего имеющийся (1)

или отсутствующий (0) признак, приведет к различным результатам в случаях, когда интегральный критерий есть аддитивная и мультипликативная функция индивидуальных количеств информации в признаках (таблица 19).

Таблица 19 – СРАВНЕНИЕ АДДИТИВНОГО И МУЛЬТИПЛИКАТИВНОГО ВАРИАНТОВ ИНТЕГРАЛЬНОГО КРИТЕРИЯ

Дополнительный признак	Аддитивная функция: $f(n) = f(n_1, n_2) = f(n_1) + f(n_2)$	Мультипликативная функция: $f(n) = f(n_1, n_2) = f(n_1) \times f(n_2)$
Есть (1)	$I + 1 = f(n_1, 1) = f(n_1) + f(1)$	$I \times 1 = f(n_1, 1) = f(n_1) \times f(1)$
Нет (0)	$I + 0 = f(n_1, 0) = f(n_1) + f(0)$	$I \times 0 = f(n_1, 0) = f(n_1) \times f(0)$

Здесь предполагается, что: $I=f(n)$, $f(1)=1$, $f(0)=0$.

Итак, если функция аддитивна – добавление еще одного разряда увеличит количество информации в кодовом слове на 1 бит, если соответствующий признак есть, и не изменит этого количества, если его нет; если же функция мультипликативна, – то это не изменит количества информации в кодовом слове, если соответствующий признак есть, и сделает его равным нулю, если его нет.

Очевидно, мультипликативный вариант интегрального критерия не соответствует классическим представлениям о природе информации, тогда как аддитивный вариант полностью им соответствует: требование аддитивности самой меры информации было впервые обосновано Хартли в 1928 году, подтверждено Шенноном в 1948 году, и в последующем развитии теории информации никогда не подвергалось сомнению. На аддитивности частных критериев, имеющих смысл количества информации, основана известная лемма Неймана-Пирсона [194, стр.152].

Пусть по выборке (т.е. совокупности факторов) $\{x=x_1, \dots, x_N\}$ требуется отдать предпочтение одной из конкурирующих гипотез (H_1 или H_0), т.е. определить в какое будущее состояние перейдет объект управления, если известны распределения наблюдений при каждой из них (по данным обучающей выборки), т.е. $p(x|H_0)$ и $p(x|H_1)$. Как обработать предпочтительную гипотезу? Из теории информации известно, что никакая обработка не может увеличить количества информации, содержащегося в выборке $\{x\}$. Следовательно, выборке $\{x\}$ нужно поставить в соответствие число, содержащее всю полезную информацию, т.е. обработать выборку без потерь. Возникает мысль о том, чтобы вычислить индивидуальные количества информации в выборке $\{x\}$ о каждой из гипотез и сравнить их:

$$\Delta i = i(x, H_1) - i(x, H_0) = \ln \frac{p(x|H_1)}{p(x)} - \ln \frac{p(x|H_0)}{p(x)} = \ln \frac{p(x|H_1)}{p(x|H_0)}$$

Какой из гипотез отдать предпочтение, зависит теперь от величины Δi и от того, какой порог сравнения мы назначим. Оптимальность данной статистической процедуры специально доказывается в математической статистике, – именно к этому сводится содержание фундаментальной Леммы Неймана-Пирсона, которая утверждает, что *предпочтение следует отдавать той статистической гипотезе, о которой в выборке содержится больше информации*.

Согласно описанной выше процедуре предполагается, что объект управления перейдет в то будущее состояние, о переходе в которое в системе факторов содержится большее суммарное количество информации.

Таким образом, аддитивность интегрального критерия, основанного на частных критериях, имеющих смысл количества информации, можно считать обоснованной, что и требовалось доказать.

Обобщение интегральной модели путем учета значений выходных параметров объекта управления

Выходные параметры \vec{Y} – это свойства объекта управления, зависят от входных параметров (в том числе параметров, характеризующих среду) и связаны с его целевым состоянием сложным и неоднозначным способом:

$$\vec{T} = T(\vec{Y}).$$

Задача идентификации состояния АОУ по его выходным параметрам решается подсистемой идентификации управляющей подсистемы, работающей на принципах системы распознавания образов. При этом классами распознавания являются выходные состояния АОУ, а признаками – его выходные параметры.

Подсистема выработки управляющих воздействий, также основанная на алгоритмах распознавания образов, обеспечивает выбор управления \vec{X} , переводящего объект управления в целевое состояние \vec{T} .

При этом последовательно решаются следующие две обратные задачи распознавания:

во-первых, по заданному целевому состоянию \vec{T} определяются наиболее характерные для данного состояния выходные параметры объекта управления:

$$\vec{Y} = T^T(\vec{T}),$$

во-вторых, по определенному на предыдущем шаге набору выходных параметров \vec{Y} определяются входные параметры \vec{X} , с наибольшей эффективностью переводящие объект управления в данное целевое состояние с этими выходными параметрами:

$$\vec{X} = Y^T(\vec{Y}).$$

1. Таким образом, определенная ограниченность подхода Шеннона, рассмотренная в данной главе, преодолевается в семантической информационной математической модели СК-анализа, основанной на СТИ. В рамках СТИ установлено, что одной из наиболее перспективных конкретизаций апостериорного подхода, является подход, предложенный в 1960 году А.А.Харкевичем [242]. Для моделирования процессов принятия решений в рефлексивных АСУ активными объектами предложено применить многокритериальный подхода с аддитивным интегральным критерием, в котором в качестве частных критериев используется системная мера семантической целесообразности информации. При этом количество информации оценивается косвенно: *по изменению степени це-*

лесообразности поведения системы, получившей эту информацию. В результате получения информации поведение системы улучшается (растет выигрыш), а в результате получения дезинформации – ухудшается (растет проигрыш). Известны и более развитые семантические меры информации [194], основанные на интересных и правдоподобных идеях, однако они наталкиваются на значительные математические трудности и сложности в программной реализации, поэтому их рассмотрение в данном исследовании признано нецелесообразным.

2. Предложенная математическая модель обеспечивает эффективное решение следующих задач, возникающих в рефлексивных АСУ АО:

- разработка абстрактной информационной модели АОУ;
- адаптация и конкретизация абстрактной модели на основе информации о реальном поведении АОУ;
- расчет влияния факторов на переход АОУ в различные возможные состояния;
- прогнозирование поведения АОУ при конкретном управляющем воздействии и выработка многофакторного управляющего воздействия (основная задача АСУ);
- выявление факторов, вносящих основной вклад в детерминацию состояния АОУ;
- корректное удаление второстепенных факторов с низкой дифференцирующей способностью, т.е. снижение размерности модели при заданных граничных условиях;
- сравнение влияния факторов, сравнение целевых и других состояний АОУ.

3. Показано, что предложенная методология, основанная на системном обобщении теории информации, обеспечивает эффективное моделирование задач принятия решений в РАСУ АОУ.

4. Доказана возможность сведения многокритериальной задачи принятия решений к однокритериальной, показана глубокая внутренняя взаимосвязь данной модели с математической моделью распознавания образов. На этой основе введено понятие "интегрального метода" распознавания и принятия решений и, после анализа и переосмысления основных понятий теории информации, предложена базовая математическая модель "интегрального метода", основанная на системной теории информации. Показано, что теория информации может рассматриваться как единая математическая и методологическая основа методов распознавания образов и теории принятия решений. При этом распознавание образов рассматривается как принятие решения о принадлежности объекта к определенному классу распознавания, прогнозирование – как распознавание будущих состояний, а принятие решения об управляющем воздействии на объект управления в АСУ как решение обратной задачи прогнозирования (распознавания).

5. Проведено исследование базовой математической модели на примере решения основной задачи АСУ – задачи принятия решения о наиболее эффек-

тивном управляющем воздействии. Осуществлена декомпозиция основной задачи в последовательность частных задач для каждой из которых найдено решение, показана взаимосвязь основной задачи АСУ с задачей декодирования теории информации.

1.2.2.4. Некоторые свойства математической модели (сходимость, адекватность, устойчивость и др.)

Под сходимостью семантической информационной модели в данной работе понимается:

- а) зависимость информативностей факторов (в матрице информативностей) от объема обучающей выборки;
- б) зависимость адекватности модели (интегральной и дифференциальной валидности) от объема обучающей выборки.

Для измерения сходимости в смыслах "а" и "б" в инструментарии СК-анализа – системе "Эйдос" реализован специальный исследовательский режим.

Под адекватностью модели понимается ее внутренняя и внешняя дифференциальная и интегральная валидность. Понятие валидности является уточнением понятия адекватности, для которого определены процедуры количественного измерения, т.е. *валидность – это количественная адекватность*. Это понятие количественно отражает способность модели давать правильные результаты идентификации, прогнозирования и способность вырабатывать правильные рекомендации по управлению. Под внутренней валидностью понимается валидность модели, измеренная после синтеза модели путем идентификации объектов обучающей выборки. Под внешней валидностью понимается валидность модели, измеренная после синтеза модели путем идентификации объектов, не входящих в обучающую выборку. Под дифференциальной валидностью модели понимается достоверность идентификации объектов в разрезе по классам. Под интегральной валидностью средневзвешенная дифференциальная валидность. Возможны все сочетания: внутренняя дифференциальная валидность, внешняя интегральная валидность и т.д. (таблица 20).

Таблица 20 – К ОПРЕДЕЛЕНИЮ ПОНЯТИЯ ВАЛИДНОСТИ

	Внутренняя валидность	Внешняя валидность
Дифференциальная валидность	Валидность модели, измеренная после синтеза модели путем идентификации объектов обучающей выборки в разрезе по классам	Валидность модели, измеренная после синтеза модели путем идентификации объектов, не входящих в обучающую выборку в разрезе по классам
Интегральная валидность	Средневзвешенная по всем классам достоверность идентификации объектов обучающей выборки	Средневзвешенная по всем классам достоверность идентификации объектов, не входящих в обучающую выборку

Под устойчивостью модели понимается ее способность давать незначительные различия в прогнозах и рекомендациях по управлению при незначительных различиях в исходных данных для решения этих задач.

1.2.2.4.1. Непараметричность модели. Робастные процедуры и фильтры для исключения артефактов

Предложенная семантическая информационная модель является непараметрической, т.к. не основана на предположениях о нормальности распределений исследуемой выборки. Под робастными понимаются процедуры, обеспечивающие устойчивую работу модели на исходных данных, зашумленных артефактами, т.е. данными, выпадающими из общих статистических закономерностей, которым подчиняется исследуемая выборка. Выявление артефактов возможно только при большой статистике, т.к. при малой статистике все частоты атрибутов малы и невозможно отличить артефакт от значимого атрибута. Критерий выявления артефактов основан на том, что при увеличении объема статистики частоты значимых атрибутов растут, как правило, пропорционально объему выборки, а частоты артефактов так и остаются чрезвычайно малыми, близкими к единице. В модели реализована такая процедура удаления наиболее вероятных артефактов, и она, как показывает опыт, существенно повышает качество (адекватность) модели.

1.2.2.4.2. Зависимость информативностей факторов от объема обучающей выборки

При учете в модели апостериорной информации, содержащейся в очередном объекте обучающей выборки, осуществляется перерасчет значений информативностей всех атрибутов. При этом изменяется количество информации, содержащейся в факте обнаружения у объекта данного атрибута о принадлежности объекта к определенному классу.

В этом процессе пересчета информативностей атрибута их значения "сходятся" к некоторому пределу в соответствии с двумя "сценариями":

- 1) процесс "последовательных приближений", напоминающего по своей форме "затухающие колебания" (рисунок 33);
- 2) относительно "плавное" возрастание или убывание с небольшими временными отклонениями от этой тенденции (рисунок 34).

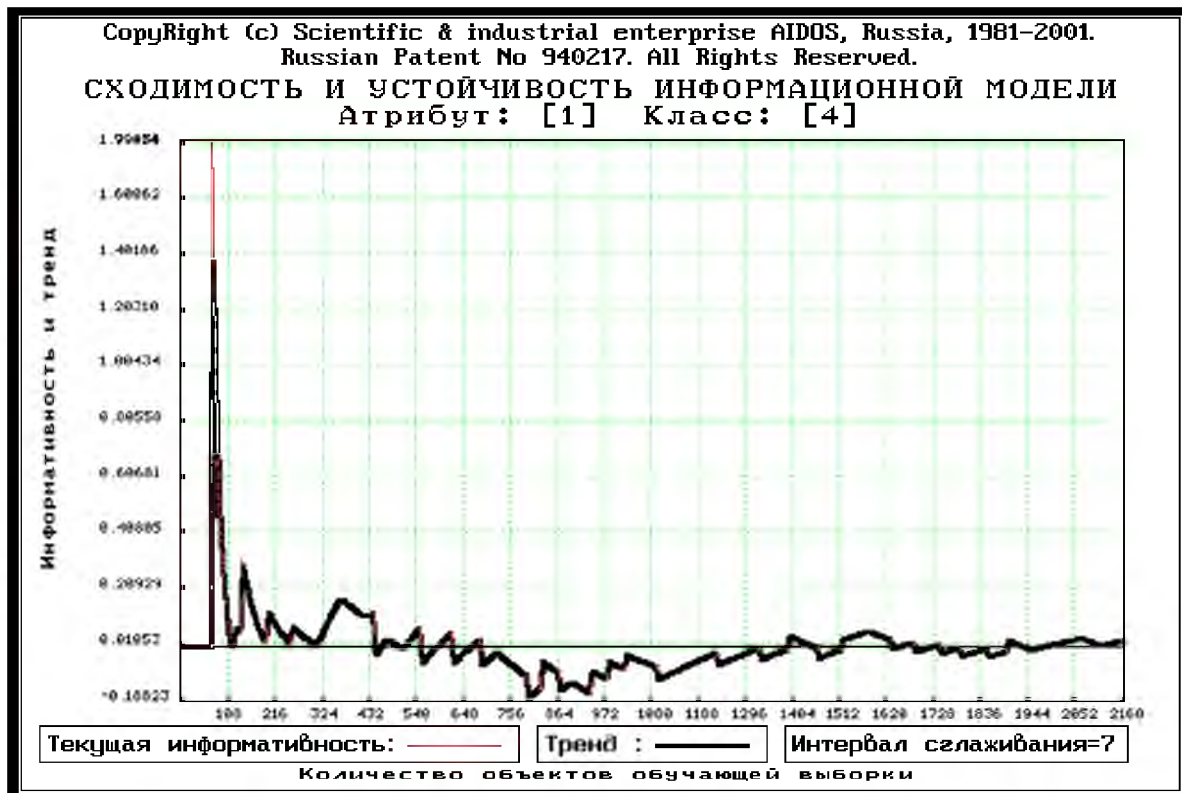


Рисунок 33. Зависимость количества информации, содержащегося в атрибуте №1 о принадлежности идентифицируемого объекта (обладающего этим атрибутом) к классу №4 от объема обучающей выборки

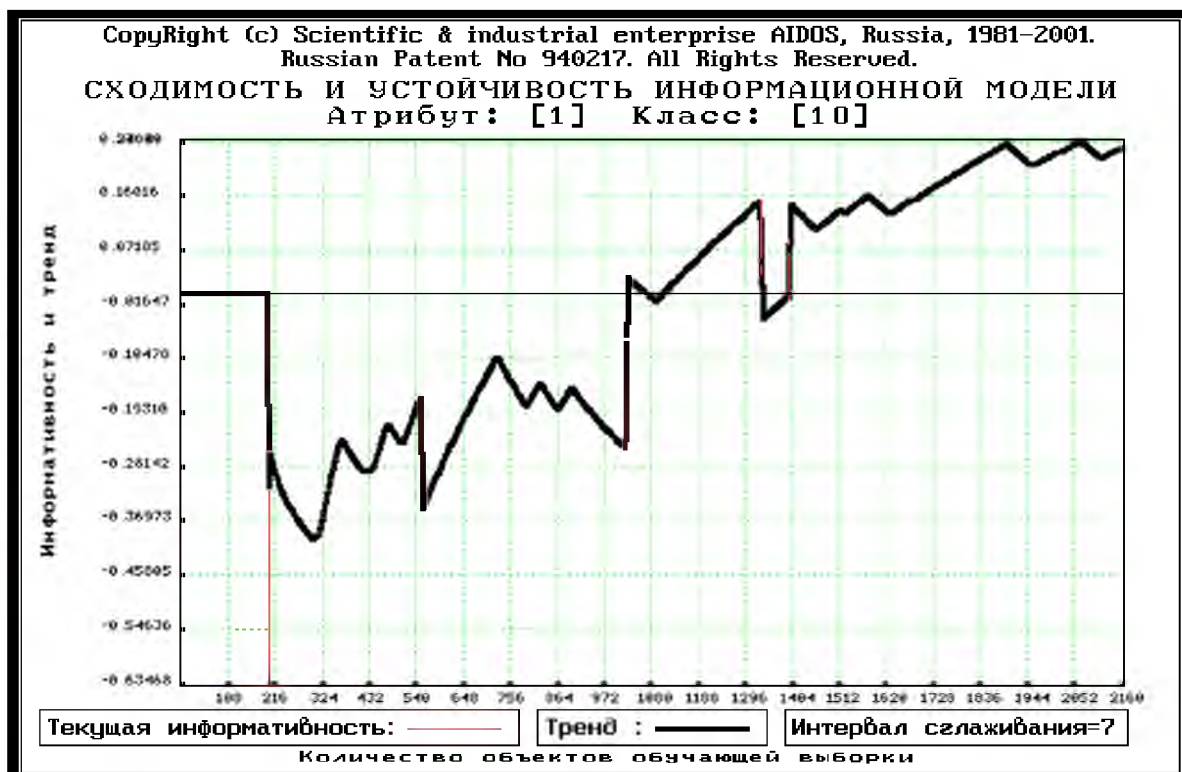


Рисунок 34. Зависимость количества информации, содержащегося в атрибуте №1 о принадлежности идентифицируемого объекта (обладающего этим атрибутом) к классу №10 от объема обучающей выборки

Как показали численные эксперименты и специально проведенные исследования, других сценариев на практике не наблюдается.

В любом случае при накоплении достаточно большой статистики и сохранении закономерностей предметной области, отражаемых обучающей выборкой, модель стабилизируется в том смысле, что значения информативностей атрибутов перестают существенно изменяться.

Это дает основание утверждать, что при достижении этого состояния добавление новых примеров из обучающей выборки не вносит в модель ничего нового в модель и процесс обучения продолжать нецелесообразно. Это и является одним из критериев для принятия решения об остановке процесса обучения.

1.2.2.4.3. Зависимость адекватности семантической информационно-математической модели от объема обучающей выборки (адекватность при малых и больших выборках)

При экспериментальном исследовании свойств предлагаемой математической модели было установлено следующее (рисунок 35).

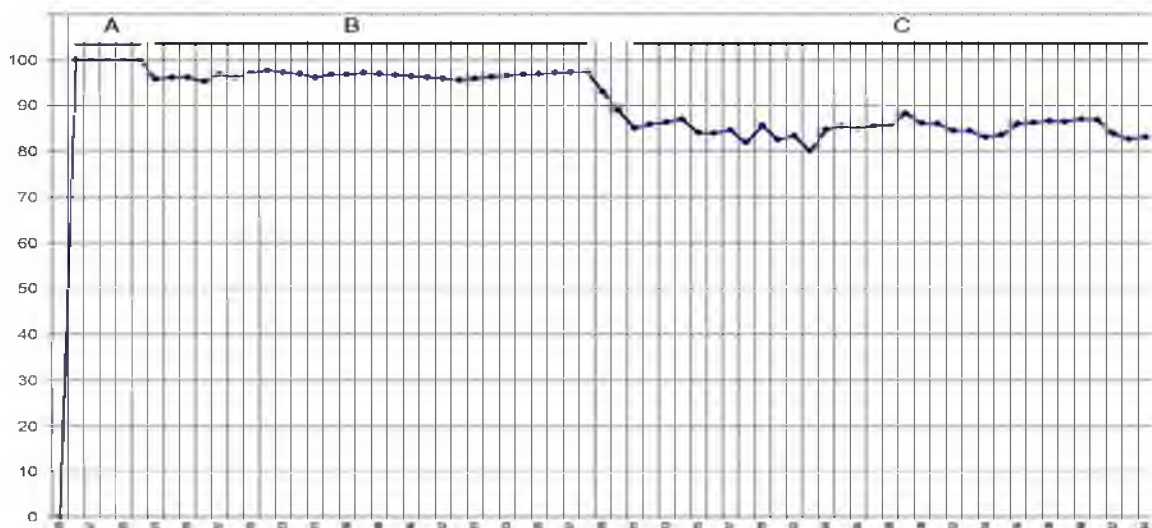


Рисунок 35. Зависимость адекватности модели от объема обучающей выборки

1. При малых выборках адекватность модели (внутренняя интегральная и дифференциальная валидность) равна 100% (рисунок 35, диапазон "А"). Это можно объяснить тем, что при малых объемах выборки все выявленные закономерности имеют детерминистский характер.

2. При увеличении объема исследуемой выборки происходит понижение адекватности модели (переход: А→В) и стабилизация ее адекватности на некотором уровне около 95-98% (рисунок 35, диапазон "В").

3. Учет в модели объектов обучающей выборки, отражающих закономерности, качественно отличающиеся от ранее выявленных, приводит к понижению адекватности модели (переход: В→С) и ее стабилизации на уровне от 80 до 90% (рисунок 35, диапазон "С").

4. Внутри диапазона "В" вариабельность объектов обучающей выборки по закономерностям "атрибут→класс" меньше, чем в диапазоне "С", т.е. объекты обучающей выборки диапазона "В" более однородны, чем "С".

Выявленные в модели причинно-следственные закономерности имеют силу для определенного подмножества обучающей выборки, например, отражающих определенный период времени, который соответствует детерминистскому периоду развития предметной области. При качественном изменении закономерностей устаревшие данные могут даже на некоторое время (пока модель не сойдется к новым закономерностям) нарушать ее адекватность.

В многочисленных проведенных практических исследованиях модель показала высокую скорость сходимости и высокую адекватность на малых выборках. На больших выборках (т.е. охватывающих несколько детерминистских и бифуркационных состояний предметной области) закономерности с коротким периодом "причина-следствие" переформируются заново, а с длительным (охватывающим несколько детерминистских и бифуркационных состояний) – автоматически становятся незначимыми и не ухудшают адекватность модели, если процесс аperiодический, или сохраняют силу, если они имеют фундаментальный характер.

Выявленные закономерности сходимости модели позволяют сформулировать следующий *критерий остановки процесса обучения*: если в модели ничего существенно не меняется при добавлении в обучающую выборку все новых и новых данных, то это означает, что модель адекватно отображает генеральную совокупность, к которой относятся эти данные, и продолжать процесс обучения нецелесообразно.

Здесь уместно рассмотреть ответ на следующий вопрос. Если для формирования образов классов распознавания предъявлено настолько малое количество обучающих объектов, что говорить об обобщении и статистике не приходится, то как это может повлиять на качество формирования модели и ее адекватность? При большой статистике, как показывает опыт, около 95% объектов, формирующих образ некоторого класса оказываются типичными для него, а остальные не типичными. Следовательно, если этот образ формируется на основе буквально одного - двух объектов, то вероятнее всего (т.е. с вероятностью около 95%) они являются типичными, и, следовательно, образ будет сформирован практически таким же, как и при большой статистике, т.е. правильным. При увеличении статистики в этом случае информативности признаков, составляющих образ практически не меняются). Но есть некоторая, сравнительно незначительная вероятность (около 5%), что попадется нетипичная анкета. Тогда при увеличении статистики образ быстро качественно изменится и "быстро сойдется" к адекватному, "нетипичная" анкета будет идентифицирована и ее данные либо будут удалены из модели, либо для нее специально будет создан свой класс.

При незначительной статистике относительный вклад каждого объекта в обобщенный образ некоторого класса, сформированный с его применением, будет достаточно велик. Поэтому в этом случае при распознавании модель уверенно относит объект к этому классу. При большой статистике модель также уверенно относит типичные объекты к классам, сформированным с их примени-

ем. Незначительное количество нетипичных объектов могут быть распознаны ошибочно, т.е. не отнесены моделью к тем классам, к которым их отнесли эксперты.

Наличие в системе очень сходных классов также может формально уменьшать валидность модели. Однако фактически эти очень сходные классы целесообразно объединить в один, т.к. по-видимому, их разделение объективно ничем не оправдано, т.е. не соответствует действительности. Для осуществления данной операции в математической модели целесообразно использовать режим: "Получение статистической характеристики обучающей выборки и объединение классов (ручной ремонт обучающей выборки)".

1.2.2.4.4. Семантическая устойчивость модели

Под семантической устойчивостью модели [81] нами понимается ее свойство давать малое различие в прогнозе при замене одних факторов, другими, мало отличающимися по смыслу (т.е. сходными по их влиянию на поведение АОУ). Проведенные автором исследования численные эксперименты в течение 1987 – 2005 годов показали, что разработанная математическая модель обладает очень высокой семантической устойчивостью.

1.2.2.4.5. Зависимость некоторых параметров модели от ее ортонормированности

Изучим зависимость уровня системности, степени детерминированности и адекватности модели от ее ортонормированности. В связи с тем, что соответствующий научно-исследовательский режим, позволяющий изучить эти зависимости методом численного эксперимента, на момент написания данной работы находится в стадии разработки, получим интересующие нас зависимости путем анализа выражений (3.9) и (3.25).

При этом будем различать ортонормированность модели по классам и ортонормированность по атрибутам.

Зависимость адекватности модели от ее ортонормированности

Модель изучалась методом численного эксперимента. При этом были получены следующие результаты.

На 1-м этапе ортонормирования адекватность модели (ее внутренняя дифференциальная и интегральная валидность) возрастает. Это можно объяснить тем, что, во-первых, уменьшается количество ошибок идентификации с близкими, т.е. коррелирующими классами, и, во-вторых, удаление из модели малоинформативных признаков по сути улучшает отношение "сигнал/шум" модели, т.е. качество идентификации.

На 2-м этапе ортонормирования адекватность модели стабилизируется и незначительно колеблется около максимума. Это объясняется тем, что атрибуты, удаляемые на этом этапе, не являются критическим для адекватности модели.

На 3-м этапе ортонормирования адекватность модели начинает уменьшаться, т.к. дальнейшее удаление атрибутов не позволяет адекватно описать предметную область.

При приближении процесса ортонормирования к 3-му этапу или его наступлении этот процесс должен быть остановлен.

Зависимость уровня системности модели от ее ортонормированности

Рассмотрим выражение (3.9):

$$\varphi = \frac{\log_2 \sum_{m=1}^M C_w^m}{\log_2 W} \quad (3.9)$$

При выполнении операции ортонормирования по классам из модели последовательно удаляются те из них, которые наиболее сильно корреляционно связаны друг с другом. В результате в модели остаются классы практически не коррелирующие, т.е. ортонормированные. Поэтому можно предположить, что *в результате ортонормирования правила запрета на образование подсистем классов становятся более жесткими, и уровень системности модели уменьшается.*

Зависимость степени детерминированности модели от ее ортонормированности

Рассмотрим выражение (3.25):

$$\Psi = \frac{\log_2 W^\varphi}{\log_2 N} \quad (3.25)$$

Так как каждый класс как правило описан более чем одним признаком, то *при ортонормировании классов и удалении некоторых из них из модели суммарное количество признаков N будет уменьшаться быстрее, чем количество классов W, поэтому степень детерминированности будет возрастать.*

При ортонормировании атрибутов числитель выражения (3.25) не изменяется, а знаменатель уменьшается, поэтому и в этом случае *степень детерминированности возрастает.*

Таким образом, ортонормирование модели приводит к увеличению степени ее детерминированности.

По этой причине предлагается считать "детерминированностью" и "системностью" модели не их значения в текущем состоянии модели, а тот предел, к которому стремятся эти величины при корректном ортонормировании модели при достижении ею точки максимума адекватности.

1.2.2.5. Взаимосвязь математической модели СК-анализа с другими моделями

1.2.2.5.1. Взаимосвязь системной меры целесообразности информации со статистикой χ^2 и новая мера уровня системности предметной области

Статистика χ^2 представляет собой сумму вероятностей совместного наблюдения признаков и объектов по всей корреляционной матрице или определенным ее подматрицам (т.е. сумму относительных отклонений частот совместного наблюдения признаков и объектов от среднего):

$$\chi^2 = \sum_{j=1}^W \sum_{i=1}^M \frac{(N_{ij} - t)^2}{t} \quad (3.57)$$

где:

– N_{ij} – фактическое количество встреч i -го признака у объектов j -го класса;

– t – ожидаемое количество встреч i -го признака у объектов j -го класса.

$$t = \frac{N_i N_j}{N} \quad (3.58)$$

Отметим, что статистика χ^2 математически связана с количеством информации в системе признаков о классе распознавания, в соответствии с системным обобщением формулы Харкевича для плотности информации (3.28)

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^{\Psi} \quad (3.59)$$

а именно из (3.58) и (3.59) получаем:

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{t} \right)^{\Psi} \quad (3.60)$$

Из (3.60) очевидно:

$$I_{ij} = \Psi (\text{Log}_2 N_{ij} - \text{Log}_2 t) \quad (3.61)$$

Сравнивая выражения (3.57) и (3.61), видим, что числитель в выражении (3.57) под знаком суммы отличается от выражения (3.61) только тем, что в выражении (3.61) вместо значений N_{ij} и t взяты их логарифмы. Так как логарифм является монотонно возрастающей функцией аргумента, то введение логарифма не меняет общего характера поведения функции.

Фактически это означает, что:

$$\begin{cases} \text{если } N_{ij} < t \text{ то } \chi_{ij} > 0, I_{ij} < 0 \\ \text{если } N_{ij} = t \text{ то } \chi_{ij} = 0, I_{ij} = 0 \\ \text{если } N_{ij} > t \text{ то } \chi_{ij} < 0, I_{ij} > 0 \end{cases} \quad (3.62)$$

Если фактическая вероятность наблюдения i -го признака при предъявлении объекта j -го класса равна ожидаемой (средней), то наблюдение этого признака **не несет никакой информации о принадлежности объекта к данному классу**. Если же она выше средней – то это говорит в пользу того, что предъявлен объект данного класса, если же ниже – то другого.

Поэтому наличие статистической связи (информации) между признаками и классами распознавания, т.е. отличие вероятностей их совместных наблюдений от предсказываемого в соответствии со случайным нормальным распределением, приводит к увеличению фактической статистики χ^2 по сравнению с теоретической величиной.

Из этого следует возможность использования в качестве количественной меры степени выраженности закономерностей в предметной области не матрицы абсолютных частот и меры χ^2 , а новой меры H , основанной на матрице информативностей и системном обобщении формулы Харкевича для количества информации:

$$H = \sqrt{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2} \quad (3.63)$$

где:

$$\bar{I} = \frac{1}{W \cdot M} \sum_{j=1}^W \sum_{i=1}^M I_{ij} \quad \text{– средняя информативность признаков по матрице информативностей.}$$

Меру H в выражении (3.63) предлагается назвать обобщенным критерием сформированности модели Харкевича.

Значение данной меры показывает среднее отличие количества информации в факторах о будущих состояниях активного объекта управления от среднего количества информации в факторе (которое при больших выборках близко к 0). По своей математической форме эта мера сходна с мерами для значимости факторов и степени сформированности образов классов и коррелирует с объемом пространства классов и пространства атрибутов.

Описанная выше математическая модель обеспечивает инвариантность результатов ее синтеза относительно следующих параметров обучающей выборки: *суммарное количество и порядок ввода анкет обучающей выборки*; количество анкет обучающей выборки по каждому классу распознавания; суммарное количество признаков во всех анкетах обучающей выборки; суммарное количество признаков по эталонным описаниям различных классов распознавания; количество признаков и их порядок в отдельных анкетах обучающей выборки.

Это обеспечивает высокое качество решения задач системой распознавания на неполных и разнородных (в вышеперечисленных аспектах) данных как обучающей, так и распознаваемой выборки, т.е. при таких статистических характеристиках потоков этих данных, которые чаще всего и встречается на практике и которыми невозможно или очень сложно управлять.

1.2.2.5.2. Сравнение, идентификация и прогнозирование как разложение векторов объектов в ряд по векторам классов (объектный анализ)

В разделе 3.2.3 были введены неметрические интегральные критерии сходства объекта, описанного массивом-локатором L_i с обобщенными образами классов I_{ij} (выражения 3.35 – 3.37)

$$I_j = (\vec{I}_{ij}, \vec{L}_i). \quad (3.64)$$

В выражении (3.64) круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^M I_{ij} L_i, \quad (3.65)$$

где:

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор j -го состояния объекта управления;

$\vec{L}_i = \{L_i\}$ – вектор состояния предметной области, включающий все виды факторов, характеризующих объект управления, возможные управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, \text{ если } i - \text{й фактор действует;} \\ \alpha_i, \text{ где } 0 < \alpha_i < 1, \text{ если } i - \text{й фактор действует с истинностью } \alpha_i; \\ 0, \text{ если } i - \text{й фактор не действует.} \end{cases}$$

Для непрерывного случая выражение (3.65) принимает вид:

$$I_j = \int_1^M L(i) I_j(i) di \quad (3.66)$$

Таким образом, выражение (3.66) представляет собой вариант выражения (3.65) интегрального критерия сходства объекта и класса для непрерывного случая в координатной форме.

Интересно и очень важно отметить, что коэффициенты ряда Фурье по своей математической форме и смыслу сходны с ненормированными коэффициентами корреляции, т.е. по сути скалярными произведениями для непрерывных функций в координатной форме: выражение (3.66), между разлагаемой в ряд кривой $f(x)$ и функциями Sin и Cos различных частот и амплитуд на отрезке $[-L, L]$ [81]:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \sin\left(\frac{n\pi x}{L}\right) \right)$$

где:

$$a_0 = \frac{1}{L} \int_{-L}^L f(x) dx \quad (3.67)$$

$$a_n = \frac{1}{L} \int_{-L}^L f(x) \cos\left(\frac{n\pi x}{L}\right) dx$$

$$b_n = \frac{1}{L} \int_{-L}^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx$$

где: $n = \{1, 2, 3, \dots\}$ – натуральное число.

Из сравнения выражений (3.66) и (3.67) следует вывод, что *процесс идентификации и прогнозирования (распознавания), реализованный в предложенной математической модели, может рассматриваться как разложение вектора-локатора распознаваемого объекта в ряд по векторам информативностей классов распознавания (которые представляют собой произвольные функции, сформированные при синтезе модели на основе эмпирических данных).*

Например, при результатах идентификации, представленных на рисунке 36.

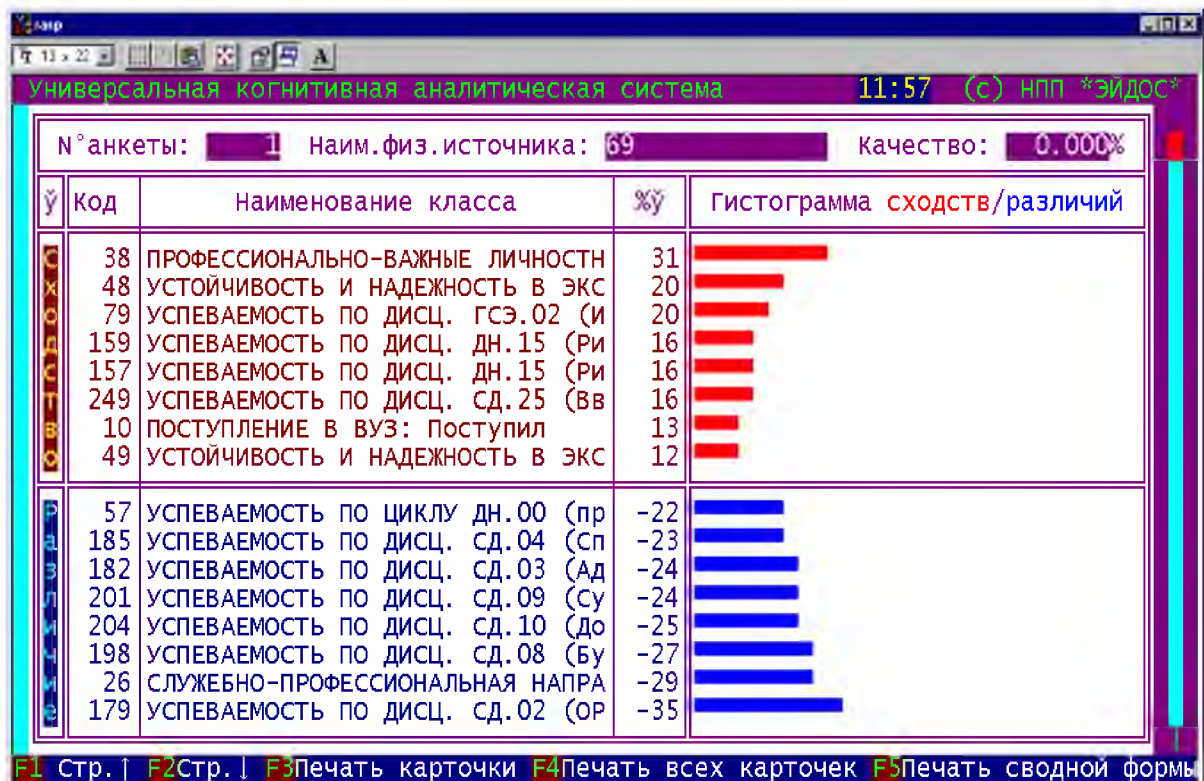


Рисунок 36. Пример разложения профиля курсанта усл. №69 в ряд по обобщенным образам классов

Продолжая развивать аналогию с разложением в ряд, данный результат идентификации можно представить в векторной аналитической форме:

$$\vec{K}_{\text{усл.№}} = 0,31 \cdot \bar{I}(38) + 0,20 \cdot \bar{I}(48) + 0,20 \cdot \bar{I}(79) + 0,16 \cdot \bar{I}(159) + 0,16 \cdot \bar{I}(157) + 0,16 \cdot \bar{I}(249) - 0,35 \cdot \bar{I}(179) - 0,29 \cdot \bar{I}(26) - 0,27 \cdot \bar{I}(198) - 0,25 \cdot \bar{I}(204) - 0,24 \cdot \bar{I}(201) - 0,24 \cdot \bar{I}(18$$

Или в координатной форме, более удобной для численных расчетов:

$$K(i) = \sum_{j=1}^W (I(j) \cdot I(i, j)) \quad (3.68)$$

где:

$I(j)$ – интегральный критерий сходства массива-локатора, описывающего состояние объекта, и j -го класса, рассчитываемый согласно выражения (3.39):

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}), \quad (3.39)$$

$I(i, j)$ – вектор обобщенного образа j -го класса, координаты которого рассчитываются в соответствии с системным обобщением формулы Харкевича (3.28):

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\text{Log}_2 W^\Phi}{\text{Log}_2 N}} + \text{Log}_2 W^\Phi \quad (3.28)$$

Примечание: обозначения $I(i, j)$ и I_{ij} , и т.п. эквивалентны. Смысл всех переменных, входящих в выражения (3.28) и (3.39) раскрыт в разделе 3.1.3 данной работы.

При дальнейшем развитии данной аналогии естественно возникают вопросы: о полноте, избыточности и ортонормированности системы векторов классов как функций, по которым будет вестись разложение вектора объекта; о сходимости, т.е. вообще возможности и корректности такого разложения.

В общем случае вектор объекта совершенно не обязательно должен разлагаться в ряд по векторам классов таким образом, что сумма ряда во всех точках точно совпадала со значениям исходной функции. Это означает, что система векторов классов может быть *неполна* по отношению к профилю распознаваемого объекта, и, тем более, всех возможных объектов.

Предлагается считать не разлагаемые в ряд, т.е. плохо распознаваемые объекты, суперпозицией хорошо распознаваемых объектов ("похожих" на те, которые использовались для формирования обобщенных образов классов), и объектов, которые и не должны распознаваться, так как объекты этого типа не встречались в обучающей выборке и не использовались для формирования обобщенных образов классов, а также не относятся к представляемой обучающей выборкой генеральной совокупности.

Нераспознаваемую компоненту можно рассматривать либо как шум, либо считать ее полезным сигналом, несущим ценную информацию о еще не исследованных объектах интересующей нас предметной области (в зависимости от целей и тезауруса исследователей). Первый вариант не приводит к осложнениям,

так как примененный в математической модели алгоритм сравнения векторов объектов и классов, основанный на вычислении нормированной корреляции Пирсона (сумма произведений), является *весьма устойчивым к наличию белого шума* в идентифицируемом сигнале. Во втором варианте необходимо дообучить систему распознаванию объектов, несущих такую компоненту (в этой возможности и заключается адаптивность модели). Технически этот вопрос решается просто копированием описаний плохо распознававшихся объектов из распознаваемой выборки в обучающую, их идентификацией экспертами и дообучением системы. Кроме того, может быть целесообразным расширить справочник классов распознавания новыми классами, соответствующими этим объектам.

Но на практике гораздо чаще наблюдается противоположная ситуация (можно даже сказать, что она типична), когда система векторов *избыточна*, т.е. в системе классов распознавания есть очень похожие классы (между которыми имеет место высокая корреляция, наблюдаемая в режиме: "кластерно-конструктивный анализ"). Практически это означает, что в системе сформировано несколько практически одинаковых образов с разными наименованиями. Для исследователя это само по себе является очень ценной информацией. Однако, если исходить только из потребности разложения распознаваемого объекта в ряд по векторам классов (чтобы определить суперпозицией каких образов он является, т.е. "разложить его на компоненты"), то наличие сильно коррелирующих друг с другом векторов представляется неоправданным, так как просто увеличивает размерности данных, внося в них мало нового по существу. Поэтому возникает задача *исключения избыточности системы классов распознавания*, т.е. выбора из всей системы классов распознавания такого минимального их набора, в котором профили классов минимально коррелируют друг с другом, *т.е. ортогональны в фазовом пространстве признаков*. Это условие в теории рядов называется "ортонормируемостью" системы базовых функций, а в факторном анализе связано с идеей выделения "главных компонент".

В предлагаемой математической модели релизованы два варианта выхода из данной ситуации:

- 1) исключение неформирующихся, расплывчатых классов;
- 2) объединение почти идентичных по содержанию (дублирующих друг друга) классов.

Но выбрать нужный вариант и реализовать его, используя соответствующие режимы, пользователь технологии АСК-анализа должен сам. Вся необходимая и достаточная информация для принятия соответствующих решений предоставляется пользователю инструментария АСК-анализа.

Если считать, что функции образов составляют формально-логическую систему, к которой применима теорема Геделя, то можно сформулировать эту теорему для данного случая следующим образом: "Для любой системы базисных функций в принципе всегда может существовать по крайней мере одна такая функция, что она не может быть разложена в ряд по данной системе базисных функций, т.е. функция, которая является ортонормированной ко всей системе базисных функций в целом".

Очевидно, не взаимосвязанными друг с другом могут быть только четко оформленные, детерминистские образы, т.е. образы с высокой степенью редук-

ции ("степень сформированности конструкта"). Поэтому в процессе выявления взаимно-ортогональных базисных образов в первую очередь будут выброшены аморфные "расплывчатые" образы, которые связаны практически со всеми остальными образами.

В некоторых случаях результат такого процесса представляет интерес и это делает оправданным его реализацию. Однако можно предположить, что и наличие расплывчатых образов в системе является оправданным, так как в этом случае система образов не будет формальной и подчиняющейся теореме Геделя, следовательно, система распознавания будет более полна в том смысле, что повысится вероятность идентификации *любого объекта*, предъявленного ей на распознавание. Конечно, уровень сходства с аморфным образом не может быть столь же высоким, как с четко оформленным, поэтому в этом случае может быть более уместно применить термин "ассоциация" или нечеткая, расплывчатая идентификация, чем "однозначная идентификация".

Итак, можно сделать следующий вывод: допустимость в математической модели СК-анализа не только четко оформленных (детерминистских) образов, но и образов аморфных, нечетких, расплывчатых является важным достоинством данной модели. Это обусловлено тем, что данная модель обеспечивает корректные результаты анализа, идентификации и прогнозирования даже в тех случаях, когда модели идентификации и информационно-поисковые системы детерминистского типа традиционных АСУ практически неработоспособны. В этих условиях данная модель СК-анализа работает как система *ассоциативной (нечеткой) идентификации*.

Таким образом, в предложенной семантической информационной модели при идентификации и прогнозировании по сути дела осуществляется разложение векторов идентифицируемых объектов по векторам классов распознавания, т.е. осуществляется "объектный анализ" (по аналогии с спектральным, гармоническим или Фурье-анализом), что позволяет рассматривать идентифицируемые объекты как суперпозицию обобщенных образов классов различного типа с различными амплитудами (3.68). При этом вектора обобщенных образов классов с математической точки зрения представляют собой произвольные функции, и не обязательно образуют полную и не избыточную (ортонормированную) систему функций.

Для любого объекта всегда существует такая система базисных функций, что вектор объекта может быть представлен в форме линейной суперпозиции (суммы) этих базисных функций с различными амплитудами. Это утверждение, по-видимому, является одним из следствий фундаментальной теоремы А.Н.Колмогорова, доказанной им в 1957 году (О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения // Докл. АН СССР, том 114, с. 953-956, 1957).

Теорема Колмогорова: Любая непрерывная функция от n переменных $F(x_1, x_2, \dots, x_n)$ может быть представлена в виде:

$$F(x_1, x_2, \dots, x_n) = \sum_{j=1}^{2n+1} \left(g_j \sum_{i=1}^n (h_{ij}(x_i)) \right)$$

где g_j и h_{ij} – непрерывные функции, причем h_{ij} не зависят от функции F .

Эта теорема означает, что для реализации функций многих переменных достаточно операций суммирования и композиции функций одной переменной. Удивительно, что в этом представлении лишь функции g_j зависят от представляемой функции F , а функции h_{ij} универсальны. *Необходимо отметить, что теорема Колмогорова является обобщением теоремы В.И. Арнольда (1957), которая дает решение 13-й проблемы Гильберта.*

К сожалению определение вида функций h_{ij} и g_j для данной функции F представляет собой математическую проблему, для которой пока не найдено общего строгого решения.

В данной работе предлагается рассматривать предлагаемую семантическую информационную модель как один из вариантов решения этой проблемы. В этом контексте функция F интерпретируется как образ идентифицируемого объекта, функция h_{ij} – как образ i -го класса, а функция g_j – как мера сходства образа объекта с образом класса.

1.2.2.5.3. Системно-когнитивный и факторный анализ. СК-анализ, как метод переменных контрольных групп

В науке широко известен "метод контрольных групп" (терм. авт.), позволяющий оценить влияние некоторого фактора на исследуемую группу по сравнению с контрольной, на которую он не влияет.

Обобщением метода контрольных групп является полный и дробный факторный анализ, при котором исследуется не одна контрольная группа, а столько, сколько факторов. При этом в каждой группе исследуется влияние одного фактора при остальных фиксированных. Таким образом факторный анализ можно было бы назвать "методом фиксированных контрольных групп". Факторный анализ требует проведения специально организованных экспериментов, что представляет собой проблему даже при нескольких факторах при большой длительности цикла управления (которая в АПК может составлять до десяти лет и более).

Например, для сбора исходных данных в факторном эксперименте при 3 факторах с 10 градациями каждый необходимо провести $10^3=1000$ экспериментов. На практике это редко осуществимо.

Поэтому перед проведением факторного эксперимента обычно выбирают небольшое количество наиболее значимых или интересных факторов для исследования. Вопрос о том, какие факторы исследовать, решается самим исследователем на основе неформальных методов.

СК-анализ является обобщением метода факторного анализа в том смысле, что контрольные группы отличаются не значениями одного фактора при остальных фиксированных, а в общем случае различными комбинациями значений действующих факторов. СК-анализ позволяет выявлять и корректно исследовать влияние *тысяч факторов* на объект управления на основе непосредственно эмпирических данных, причем неполных и неупорядоченных, как в факторном эксперименте. При этом определяется и значимость факторов, что позволяет обоснованно выбрать из них небольшое количество наиболее значимых для последующего более детального исследования методом факторного

анализа. Необходимо отметить, что СК-анализ является непараметрическим методом, в отличие от факторного анализа.

1.2.2.5.4. Семантическая мера целесообразности информации и эластичность

Эластичность в непрерывном случае

Рассмотрим связь эластичности и семантической меры целесообразности информации, опираясь на результаты работ автора [81, 91, 115, 135]. Пусть численное значение некоторого параметра экономической системы описывается переменной y , зависящей от фактора x и эта зависимость описывается функцией $y=f(x)$. Тогда степень и направление влияния фактора x на параметр y можно численно измерить производной (3.69), представляющей собой предел отношения абсолютных изменений величин y и x :

$$y_x = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \quad (3.69)$$

Однако применение производной не очень удобно, т.к. она зависит от размерности величин y и x и, по этой причине, обладает недостаточной сопоставимостью в пространстве и времени. Кроме того, сама по себе скорость абсолютного изменения некоторого параметра объекта безотносительно к средней величине этого параметра, содержит недостаточно информации об этом объекте. Например, если на очередных выборах за некоторого кандидата отдано на 500 голосов больше, чем на предыдущих, то важно знать, а на сколько это процентов больше. Поэтому в экономике введено понятие эластичности $E_x(y)$ функции $y=f(x)$, которое определяется как предел отношения не абсолютных, а относительных изменений значений переменных y и x :

$$E_x(y) = \lim_{\Delta x \rightarrow 0} \left(\frac{\frac{\Delta y}{y}}{\frac{\Delta x}{x}} \right) = \lim_{\Delta x \rightarrow 0} \left(\frac{\Delta y}{\Delta x} \cdot \frac{x}{y} \right) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \cdot \frac{x}{y} = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{f'(x)}{\frac{y}{x}} = \frac{f'(x)}{\bar{f}(x)}$$

$$E_x(y) = \frac{f'(x)}{\bar{f}(x)}, \text{ где: } f'(x) - \text{производная, а } \bar{f}(x) - \text{среднее значений функции } f(x)$$

Так как $d \ln y = dy/y$, и $d \ln x = dx/x$, то эластичность можно представить в виде логарифмической производной:

$$E_x(y) = \frac{d \ln(f(x))}{d \ln(x)} \quad (3.71)$$

Эластичность в дискретном случае

Для численных расчетов необходимо перейти к дискретному случаю, в частности для численного взятия производных используем метод конечных разностей. В конечных разностях выражение (3) принимает вид:

$$E_x(y) = \frac{\Delta \ln(f(x))}{\Delta \ln x} = \frac{\ln(f(x_2)) - \ln(f(x_1))}{\ln(x_2) - \ln(x_1)} = \frac{\ln(f(x_2)/f(x_1))}{\ln(x_2/x_1)} \quad (3.72)$$

Свойства эластичности

Рассмотрим некоторые свойства эластичности, которые, как мы заметили, *удивительным образом* полностью или частично совпадают со свойствами логарифма (таблица 21).

Таблица 21 – СВОЙСТВА ЭЛАСТИЧНОСТИ И ЛОГАРИФМА

№	ЭЛАСТИЧНОСТЬ	ЛОГАРИФМ	Примечание
1	Эластичность взаимно-обратной функции взаимно-обратна: $E_x(y) = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{1}{\frac{dx}{dy} \cdot \frac{y}{x}} = \frac{1}{E_y(x)}$	Логарифм взаимно-обратной функции равен той же функции с обратным знаком: $\ln y = -(\ln 1 - \ln y) = -\ln \frac{1}{y}$	Совпадает по модулю (с точностью до знака)
2	Эластичность произведения двух функций одного аргумента равна сумме эластичностей функций: $E_x(uv) = \frac{d(uv)}{dx} \cdot \frac{x}{uv} = \frac{v \left(\frac{du}{dx} \right) + u \left(\frac{dv}{dx} \right)}{uv} = \frac{du}{dx} \cdot \frac{x}{u} + \frac{dv}{dx} \cdot \frac{x}{v} = E_x(u) + E_x(v)$	Логарифм произведения двух функций одного аргумента равна сумме логарифмов функций: $\ln(uv) = \ln(u) + \ln(v)$	<i>Полностью совпадает</i>
3	Эластичность частного двух функций одного аргумента равна разности эластичностей функций: $E_x(u/v) = \frac{d(u/v)}{dx} \cdot \frac{x}{u/v} = \frac{v \cdot du - u \cdot dv}{v^2} \cdot \frac{xv}{u} = \frac{du}{dx} \cdot \frac{x}{u} - \frac{dv}{dx} \cdot \frac{x}{v} = E_x(u) - E_x(v)$	Логарифм частного двух функций одного аргумента равна разности логарифмов функций: $\ln(u/v) = \ln(u) - \ln(v)$	<i>Полностью совпадает</i>
4	Эластичность показательной функции $y = a^x$ пропорциональна показателю степени: $E_x(y) = \frac{da^x}{dx} \cdot \frac{x}{a^x} = a^x \cdot x \cdot \frac{\ln a}{a^x} = x \cdot \ln a$	Логарифм показательной функции $y = a^x$ пропорционален показателю степени: $\ln a^x = x \cdot \ln a$	<i>Полностью совпадает</i>
5	Область значений эластичности: $-\infty < E < +\infty$	Область значений логарифма: $-\infty < \ln < +\infty$	<i>Полностью совпадает</i>

Необходимо отметить, что ряд других свойств эластичности, таких как эластичность суммы функций, эластичность линейной функции и др., *не совпадают* со свойствами логарифма. Итак, учитывая свойства эластичности 2-5 (таблица 21) мы видим, что *большинство свойств эластичности совпадают со свойствами логарифмической функции*. Это позволяет высказать *гипотезу*, что свойства эластичности $E_x(y)$ схожи со свойствами количества информации I , т.к. во все выражения для количества информации Хартли-Найквиста-Больцмана, Шеннона и Харкевича входит логарифмическая функция.

Какая же из этих мер информации в наибольшей степени соответствует понятию эластичности? Ключевым в решении этого вопроса является свойство 5 (таблица 21):

- область значений мер Хартли-Найквиста-Больцмана и Шеннона изменяется от 0 до $+\infty$;
- область значений меры Харкевича, как и эластичности, изменяется от $-\infty$ до $+\infty$, как и эластичности.

Однако классическая мера семантической целесообразности информации мера Харкевича не удовлетворяет *принципу соответствия* с мерой Хартли в детерминистском случае, поэтому автором данной работы в [81] предложена *системная мера целесообразности информации (СМЦИ)* – $I_{ij}(W, M)$. В отличие от эластичности $E_x(y)$, которая определена для однозначной функции одного аргумента, $I_{ij}(W, M)$ определена для многозначной функции многих аргументов.

Таким образом, *системная мера целесообразности информации, предложенная в настоящем исследовании, имеет математические свойства сходные со свойствами эластичности многозначной функции многих аргументов.*

1.2.2.5.5. Связь семантической информационной модели с нейронными сетями

В 1943 году Дж. Маккалоки и У. Питт предложили формальную модель биологического нейрона как устройства, имеющего несколько входов (входные синапсы – дендриты), и один выход (выходной синапс – аксон). Дендриты получают информацию от источников информации (рецепторов) L_i , в качестве которых могут выступать и нейроны. Набор входных сигналов $\{L_i\}$ характеризует объект или ситуацию, обрабатываемую нейроном. Каждому i -му входу j -го нейрона ставится в соответствие некоторый весовой коэффициент I_{ij} , характеризующий степень влияния сигнала с этого входа на аргумент передаточной (активационной) функции, определяющей сигнал Y_j на выходе нейрона. В нейроне происходит взвешенное *суммирование* входных сигналов, и далее это значение используется как *аргумент* активационной (передаточной) функции нейрона. На рисунке 37 данная модель приведена в обозначениях, принятых в настоящей работе.

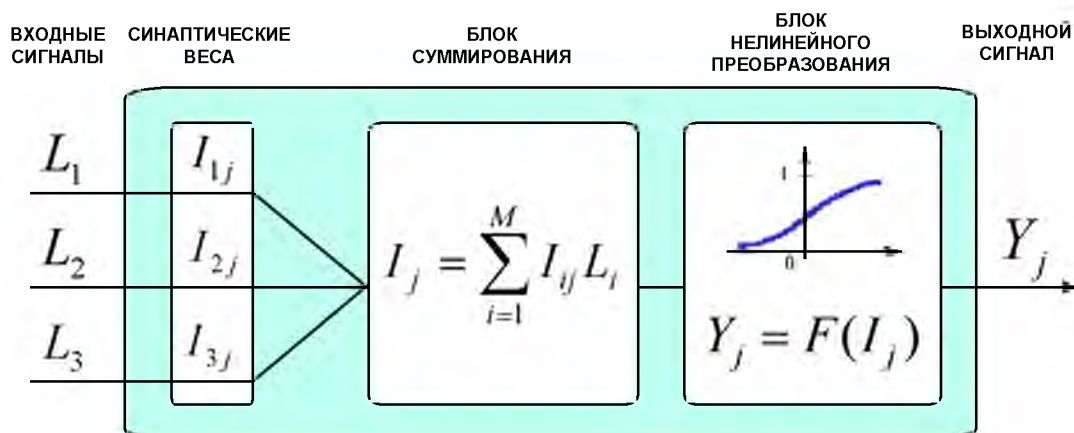


Рисунок 37. Классическая модель нейрона Дж. Маккалоки и У. Питта (1943) в обозначениях системной теории информации

Метафора нейросетевого представления семантической информационной модели

В данной работе предлагается представление, согласно которому каждый нейрон отражает определенное будущее состояние активного объекта управления, а нейронная сеть в целом – систему будущих состояний, как желательных (целевых), так и нежелательных. Весовые коэффициенты на дендридах нейронов имеют смысл силы и направления влияния факторов на переход активного объекта управления в то или иное будущее состояние. Таким образом, предложенная в данной работе семантическая информационная модель в принципе допускает представление в терминах и понятиях нейронных сетей. Однако при более детальном рассмотрении выясняется, что семантическая информационная модель является более общей, чем нейросетевая и для полного их соответствия необходимо внести в нейросетевую модель ряд дополнений.

Соответствие основных терминов и понятий

Предлагается следующая система соответствий, позволяющая рассматривать термины и понятия из теории нейронных сетей и предложенной семантической информационной модели практически как синонимы. Нейрон – вектор обобщенного образа класса в матрице информативностей. Входные сигналы – факторы (признаки). Весовой коэффициент – системная мера целесообразности информации. Обучение сети – *адаптация* модели, т.е. перерасчет значений весовых коэффициентов дендридов для каждого нейрона (матрицы информативностей) и изменение вида активационной функции. Самоорганизация сети – *синтез* модели, т.е. изменение количества нейронов и дендридов, изменение количества нейронных слоев и структуры связей между факторами и классами, а затем адаптация (перерасчет матрицы информативностей). Таким образом, адаптация – это обучение сети на уровне изменения информационных весовых коэффициентов и активационной функции, а синтез – на уровне изменения размерности и структуры связей нейронов сети. 1-й (входной) слой нейронной сети – формирование обобщенных образов классов. Сети Хопфилда и Хэмминга – обучение с учителем, сопоставление описательной и классификационной информации, идентификация и прогнозирование. 2-й слой, сети Хебба и Кохонена – самообучение, анализ структуры данных без априорной классификационной информации, формирование кластеров классов и факторов. 3-й слой – формирование конструкторов (в традиционных нейронных сетях не реализовано). Необходимо отметить, что любой слой нейронной сети является в предлагаемой модели не только обрабатывающим, но и выходным, т.е. с одной стороны дает результаты обработки информации, имеющие самостоятельное значение, а с другой – предоставляет информацию для последующих слоев нейронной сети, т.е. более высоких уровней иерархии информационной системы (в полном соответствии с формализуемой когнитивной концепцией).

Недостатки нейронных сетей и пути их преодоления в семантической информационной модели

К основным недостаткам нейронных сетей можно отнести:

1. Сложность содержательной интерпретации смысла интенсивности входных сигналов и весовых коэффициентов ("*проблема интерпретируемости весовых коэффициентов*").

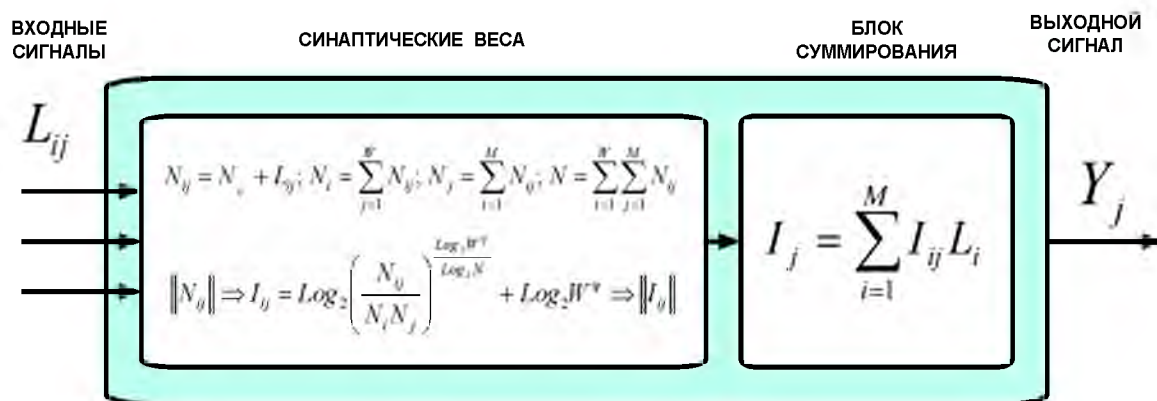
2. Сложность содержательной интерпретации и обоснования аддитивности аргумента и вида активационной (передаточной) функции нейрона ("проблема интерпретируемости передаточной функции").

3. "Комбинаторный взрыв", возникающий при определении структуры связей нейронов, подборе весовых коэффициентов и передаточных функций ("проблема размерности").

Проблемы интерпретируемости приводят к снижению ценности полученных результатов работы сети, а проблема размерности – к очень жестким ограничениям на количество выходных нейронов в сети, на количество рецепторов и на сложность структуры взаимосвязей нейронов с сети. Достаточно сказать, что количество выходных нейронов в реальных нейронных сетях, реализуемых на базе известных программных пакетов, обычно не превышает несколько сотен, а чаще всего составляет единицы и десятки.

Гипотеза о нелокальности нейрона и информационная нейросетевая парадигма

Модель нелокального нейрона: так как сигналы на дендридах различных нейронов вообще говоря коррелируют (или антикоррелируют) друг с другом, то, значения весовых коэффициентов, а значит и выходное значение на аксоне каждого конкретного нейрона вообще говоря не могут быть определены с использованием значений весовых коэффициентов на дендридах только данного конкретного нейрона, а должны учитывать интенсивности сигналов на всей системе дендритов *нейронной сети в целом* (рисунок 38).



**Рисунок 38. Модель нелокального нейрона
в обозначениях системной теории информации**

За счет учета корреляций входных сигналов (если они фактически присутствуют в структуре данных), т.е. наличия *общего самосогласованного* информационного поля исходных данных всей нейронной сети (информационное пространство), нелокальные нейроны ведут себя так, как будто связаны с другими нейронами, хотя могут быть и не связаны с ними синаптически по входу и выходу ни прямо, ни опосредованно. Самосогласованность семантического информационного пространства означает, что *учет любого одного нового факта в информационной модели вообще говоря приводит к изменению всех весовых коэффициентов всех нейронов, а не только тех, на рецепторе которых обнаружен*

этот факт и тех, которые непосредственно или опосредованно синаптически с ним связаны.

В традиционной (т.е. локальной) модели нейрона весовые коэффициенты на его дендридах однозначно определяются заданным выходом на его аксоне и никак не зависят от параметров других нейронов, с которыми с нет прямой или опосредованной синаптической связи. Это связано с тем, что в общепринятой энергетической парадигме Хопфилда весовые коэффициенты дендридов имеют смысл *интенсивностей* входных воздействий. В методе "обратного распространения ошибки" процесс переобучения, т.е. интерактивного перерасчета весовых коэффициентов, начинается с нейрона, состояние которого оказалось ошибочным и захватывает только нейроны, ведущие от рецепторов к данному нейрону. Корреляции между локальными нейронами обусловлены сочетанием трех основных причин:

- наличием в исходных данных определенной структуры: корреляцией входных сигналов;
- синаптической связью локальных нейронов;
- избыточностью (дублированием) нейронной сети.

Решение проблемы интерпретируемости весовых коэффициентов (семантическая мера целесообразности информации и закон Фехнера)

В данной работе предлагается использовать такие весовые коэффициенты дендридов, чтобы активационная функция была линейной, т.е. по сути была равна своему аргументу: сумме. Этому условию удовлетворяют весовые коэффициенты, рассчитываемые с применением системного обобщения формулы Харкевича (3.28).

Очень важно, что данная мера, удовлетворяет известному эмпирическому закону Г.Фехнера (1860), согласно которому существует *логарифмическая* зависимость между интенсивностью фактора и величиной отклика на него биологической системы (в частности, величина ощущения прямо пропорциональна логарифму интенсивности раздражителя).

Предлагается *информационный подход* к нейронным сетям, по аналогии с энергетическим подходом Хопфилда (1980).

Суть этого подхода состоит в том, что интенсивности входных сигналов рассматриваются не сами по себе и не с точки зрения только их *интенсивности*, а как сообщения, несущие определенное количество *информации* или *дезинформации* о переходе нейрона и моделируемого им активного объекта управления в некоторое будущее состояние.

Под интенсивностью входного сигнала на определенном дендриде мы будем понимать абсолютную частоту (количество) встреч фактора (признака), соответствующего данному дендриду, при предъявлении нейронной сети объекта, соответствующего определенному нейрону. Таким образом матрица абсолютных частот рассматривается как способ накопления и первичного обобщения эмпирической информации об интенсивностях входных сигналов на дендридах в разрезе по нейронам.

Весовые коэффициенты, отражающие влияние каждого входного сигнала на отклик каждого нейрона, т.е. величину его возбуждения или торможения, представляют собой элементы матрицы информативностей, получающиеся из

матрицы абсолютных частот методом прямого счета с использованием выражения для семантической меры целесообразности информации (3.28).

При этом предложенная мера семантической целесообразности информации, как переключается с нейронными сетями Кохонена, в которых также принято стандартизировать (нормализовать) входные сигналы, что позволяет в определенной мере уйти от многообразия передаточных функций.

Наличие ясной и обоснованной интерпретации весовых коэффициентов, как количества информации, позволяет предложить в качестве математической модели для их расчета системную теорию информации (СТИ).

Семантическая информационная модель, как нелокальная нейронная сеть

Учитывая большое количество содержательных параллелей между семантической информационной моделью и нейронными сетями предлагается рассматривать данную модель как нейросетевую модель, основанную на системной теории информации. В данной модели предлагается вариант решения важных нейросетевых проблем интерпретируемости и ограничения размерности за счет введения меры целесообразности информации (системное обобщение формулы Харкевича), обеспечивающей *прямой расчет* интерпретируемых весовых коэффициентов на основе непосредственно эмпирических данных. Итак, в данной работе предлагается новый класс нейронных сетей, основанных на семантической информационной модели и информационном подходе. Для этих сетей предлагается полное наименование: *"Нелокальные интерпретируемые нейронные сети прямого счета"* и сокращенное наименование: *"Нелокальные нейронные сети"*.

Нелокальная нейронная сеть является системой нелокальных нейронов, обладающей качественно новыми (системными, эмерджентными) свойствами, не сводящимися к сумме свойств нейронов. В такой сети поведение нейронов определяется как их собственными свойствами и поступающими на них входными сигналами, так и свойствами нейронной сети в целом, т.е. поведение нейронов в нелокальной нейронной сети согласовано друг с другом не только за счет их прямого и опосредованного синаптического взаимодействия (как в традиционных нейронных сетях), но за счет общего информационного поля весовых коэффициентов всех нейронов данной сети.

Гипотеза о физической природе нелокального взаимодействия нейронов в нелокальной нейронной сети

В данной работе предлагается математическая модель, численный метод и программный инструментальный нелокальных нейронных сетей (универсальная когнитивная аналитическая система "Эйдос"), успешно апробированные в ряде предметных областей. Данная система обеспечивает неограниченное количество слоев ННС при максимальном количестве весовых коэффициентов в слое до 16 миллионов и до 4000 выходных нейронов (в текущей версии 12.5 при решении ряда задач эти ограничения сняты). Но если рассматривать нелокальную нейронную сеть как модель реальных "биологических" нейронных сетей, то ясно, что формальной модели недостаточно и необходимо дополнить ее физической моделью о природе каналов нелокального взаимодействия нейронов в данной сети. По мнению автора данный механизм основан на парадоксе Эйнштей-

на-Подольского Розена (ЭПР) [10, 27, 264]. По мнению автора, физическая реализация нелокальных нейронов может быть осуществлена за счет соединения как минимум одного дендрита каждого нейрона с датчиком микротелекинетического воздействия, на который человек может оказывать влияние дистанционно. Некоторые из подобных датчиков описаны в работе [27] и более подробно на сайте автора <http://lc.kubagro.ru>. По мнению автора, квантовые компьютеры, основанные не на математических и программных моделях, а на физических нелокальных нейронах, могут оказаться во многих отношениях функционально эквивалентными физическому мозгу и организму.

Решение проблемы интерпретируемости передаточной функции

Вопрос об интерпретируемости передаточной функции нейрона включает два основных аспекта:

- об интерпретируемости *аргумента* передаточной функции;
- об интерпретируемости *вида* передаточной функции.

1. Возникает естественный вопрос о том, чем обосновано включение в состав модели нейрона Дж. Маккалоки и У. Питтом именно аддитивного элемента, суммирующего входные сигналы, а не скажем мультипликативного или в виде функции общего вида. По мнению автора *такой выбор обоснован и имеет явную и убедительную интерпретацию именно в том случае, когда весовые коэффициенты имеют смысл количества информации, т.к. в этом случае данная мера представляет собой неметрический критерий сходства (3.37), основанный на лемме Неймана-Пирсона. Сумма весовых коэффициентов, соответствующих набору действующих факторов (входных сигналов) дает величину выходного сигнала на аксоне каждого нейрона.*

2. Вид передаточной функции содержательно в теории нейронных сетей явно не обосновывается. Предлагается гипотеза, что на практике вид передаточной функции подбирается таким образом, чтобы соответствовать смыслу подобранных в данном конкретном случае весовых коэффициентов. Так как при применении в различных предметных областях смысл весовых коэффициентов в явном виде не контролируется и может отличаться, то выбор вида передаточной функции позволяет частично компенсировать эти различия.

Предлагаемый интерпретируемый вид весовых коэффициентов обеспечивает единую и стандартную интерпретацию аргумента и значения передаточной функции независимо от предметной области. Поэтому в нелокальной нейронной модели передаточная функция нейрона всегда линейна (аргумент равен функции). Следовательно в модели нелокального нейрона блок суммирования по сути дела объединен с блоком нелинейного преобразования (точнее, второй отсутствует, а его роль выполняет блок суммирования), в отличие от стандартных передаточных функций локальных нейронов: логистической, гиперболического тангенса, пороговой линейной, экспоненциально распределенной, полиномиальной и импульсно-кодовой.

Нелокальные нейроны как бы "резонируют" на ансамбли входных сигналов, причем *этот резонанс может быть обоснованно назван семантическим (смысловым)*, т.к. весовые коэффициенты рассчитаны на основе предложенной семантической меры целесообразности информации. Таким образом, разложение вектора идентифицируемого объекта в ряд по векторам обобщенных образов

классов осуществляется на основе семантического резонанса нейронов выходного слоя на ансамбль входных сигналов (признаков, факторов).

Решение проблемы размерности

Вместо итерационного подбора весовых коэффициентов путем полного перебора вариантов их значений при малых вариациях (методы обратного распространения ошибки и градиентного спуска к локальному экстремуму) предлагается *прямой расчет* этих коэффициентов на основе процедуры и выражений, обоснованных в предложенных системной теории информации и семантической информационной модели. Выигрыш во времени и используемых вычислительных ресурсах, получаемый за счет этого, быстро возрастает при увеличении размерности нейронной сети.

Моделирование причинно-следственных цепочек в нейронных сетях и семантической информационной модели

Факторы описывают причины, а классы – следствия. Но и следствия в свою очередь являются причинами более отдаленных последствий. Предлагаемая семантическая информационная модель позволяет рассматривать события, обнаружение которых осуществляется в режиме идентификации, как причины последующих событий, т.е. как факторы, их вызывающие. При этом факт наступления этих событий моделируется путем включения в модель факторов, соответствующих классам (событиям). В нейронных сетях этот процесс моделируется путем включения в сеть дополнительных нейронных слоев и создания обратных связей между слоями, обеспечивающих передачу в предыдущие слои результатов работы последующих слоев.

Моделирование иерархических структур обработки информации

Рассмотрим иерархическую структуру информации на примере использования психологического теста для оценки психологических качеств сотрудников и влияния этих качеств на эффективность работы фирмы. В нейронной сети иерархическим уровням обработки информации соответствуют слои, поэтому далее будем использовать термины "слой нейронной сети" и "иерархический уровень обработки информации" как синонимы. Рецепторы дают информацию по ответам сотрудника на опросник, нейроны 1-го слоя дают оценку психологических качеств и сигнал с их аксонов является входным для нейронов 2-го слоя, дающих оценку качества работы фирмы. В семантической информационной модели существует три варианта моделирования подобных иерархических структур обработки информации:

1. Заменить все слои одним слоем и выявлять зависимости непосредственно между исходными данными с первичных рецепторов и интересующими итоговыми оценками, например, ответами сотрудников на вопросы и результатами работы фирмы. Этот подход эффективен с прагматической точки зрения, но дает мало информации для теоретических обобщений.

2. Каждый слой моделируется отдельной семантической информационной моделью, включающей свои классификационные и описательные шкалы и градации, обучающую выборку, матрицы абсолютных частот и информативностей. Вся система иерархической обработки информации моделируется системой этих моделей, взаимосвязанных друг с другом по входу-выходу: результаты классификации объектов обучающей выборки 1-й моделью рассматриваются как свой-

ства этих объектов во 2-й модели, в которой они используются для классификации 2-го уровня. Например, психологические качества сотрудников, установленные в результате психологического тестирования, рассматриваются как свойства сотрудников, влияющие на эффективность работы фирмы. Данный подход эффективен и с прагматической, и с теоретической точек зрения, но является громоздким в программной реализации.

3. Моделирование каждого слоя соответствующими подматрицами матриц абсолютных частот и информативностей (таблица 22).

Таблица 22 – ЛОГИЧЕСКАЯ МОДЕЛЬ СТРУКТУРЫ ДАННЫХ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИОННОЙ МОДЕЛИ, СООТВЕТСТВУЮЩАЯ ТРЕХСЛОЙНОЙ НЕЛОКАЛЬНОЙ НЕЙРОННОЙ СЕТИ

Рецепторы – факторы, влияющие на поведение объекта управления	Нейроны - будущие состояния объекта управления			Дифференцирующая способность входного сигнала
	Нейроны 1-го слоя: психологические качества сотрудников	Нейроны 2-го слоя: успешность деятельности сотрудника	Нейроны 3-го слоя: успешность деятельности фирмы	
Рецепторы 1-го слоя: ответы сотрудников на вопросы анкеты	Весовые коэффициенты 1-го слоя	---	---	
Рецепторы 2-го слоя: психологические качества сотрудников	---	Весовые коэффициенты 2-го слоя	---	
Рецепторы 3-го слоя: успешность деятельности сотрудника	---	---	Весовые коэффициенты 3-го слоя	
Степень обученности нейрона				Степень обученности нейронной сети

Этот вариант обладает преимуществами первых двух и преодолевает их недостатки. В нем применяется следующий итерационный алгоритм послойного расчета, где $n=\{1, 2, \dots, N\}$, N – количество слоев нейронной сети:

Шаг n : расчет весовых коэффициентов n -го слоя, идентификация объектов обучающей выборки в нейронах n -го слоя, если слой $(n+1)$ существует, то занесение в обучающую выборку в качестве свойств объектов $(n+1)$ -го слоя результатов их идентификации в нейронах n -го слоя.

Примечание: в таблице 22 представлена именно логическая структура данных, т.е. в реальных базах данных нет записей, содержащих информацию о влиянии рецепторов n -го слоя на нейроны слоев, номера которых не равны n .

Нейронные сети и СК-анализ

Известные в литературе нейронные сети, в отличие от предлагаемой семантической информационной модели и нелокальных нейронных сетей, не обеспечивают реализацию всех базовых когнитивных операций, входящих в когнитивный конфигуратор. В частности, традиционные нейронные сети решают лишь задачу идентификации (прогнозирования) и не обеспечивают решение обратной задачи (дедукции), необходимой для принятия решения о выборе многофакторного управляющего воздействия. Кроме того не решается вопрос об уменьшении размерности нейронной сети без ущерба для ее адекватности (абстрагирование).

Результаты численного моделирования и исследования свойств нейронных сетей этого класса при управлении в АПК и других предметных областях позволяют предположить, в качестве модели реальных когнитивных процессов они обладает более высокой адекватностью, чем нейронные сети других типов.

Графическое отображение нейронов и Паррето-подмножеств нелокальной нейронной сети

Ниже приводятся примеры графического отображения нелокального нейрона и Паррето-подмножества (нейронов с наиболее значимыми связями) нелокальной нейронной сети в системе "Эйдос" (рисунки 39 и 40).

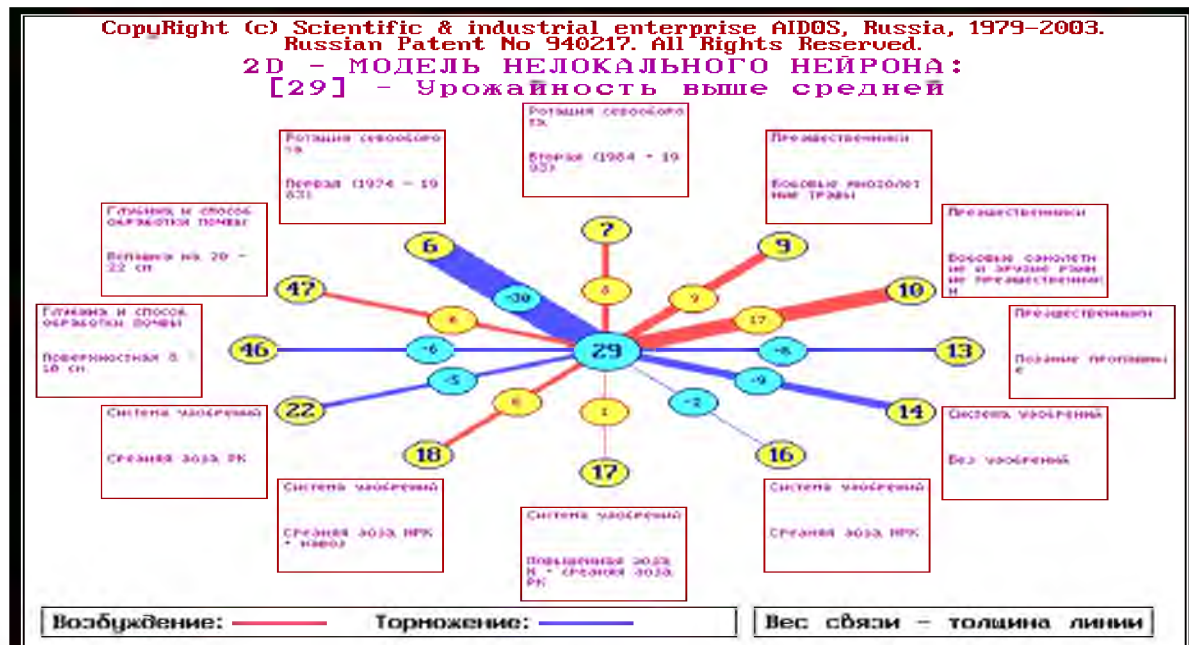


Рисунок 39. Графическое отображение нелокального нейрона в системе "Эйдос"

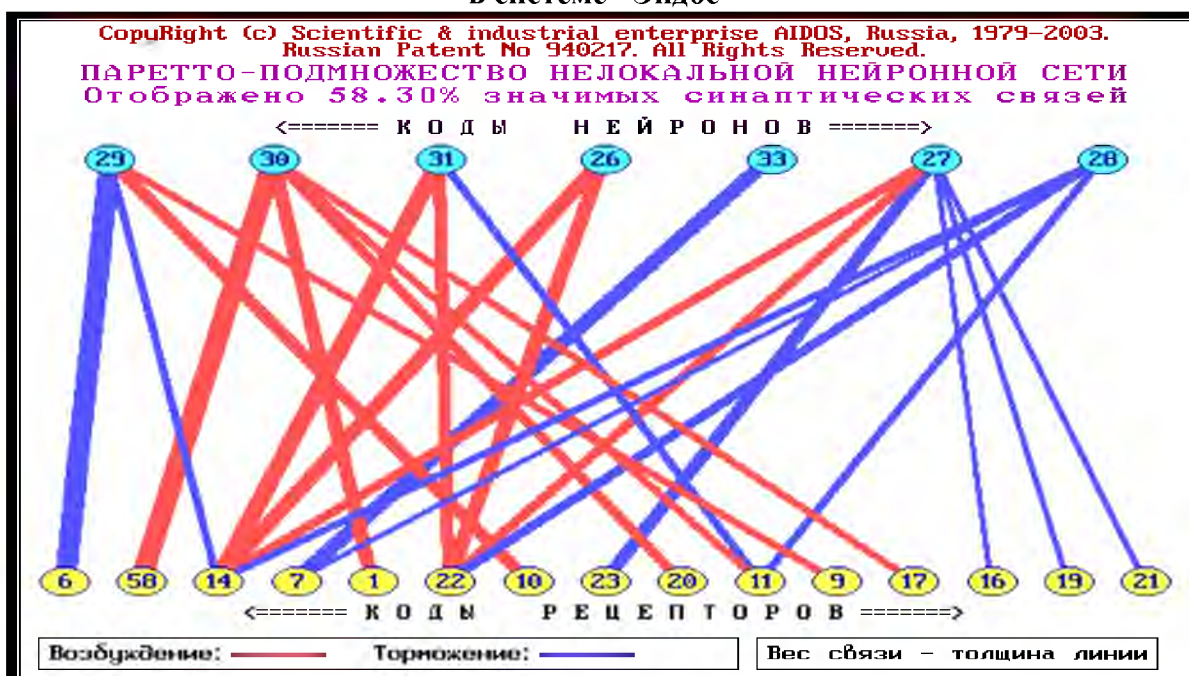


Рисунок 40. Графическое отображение нелокальной нейронной сети в системе "Эйдос"

1.2.2.5.6. Математический метод СК-анализа в свете идей интервальной бутстрепной робастной статистики объектов нечисловой природы

Постановка проблемы

Современный этап развития информационных технологий характеризуется быстрым ростом производительности компьютеров облегчением доступа к ним. С этим связан возрастающий интерес к использованию компьютерных технологий для организации мониторинга различных объектов, анализа данных, прогнозирования и управления в различных предметных областях. И у исследователей, и у руководителей, имеются определенные ожидания и надежды на повышение эффективности применения компьютерных технологий.

Однако на пути реализации этих ожиданий имеются определенные сложности, связанные с относительным отставанием в развитии математических методов и реализующего их программного инструментария.

И анализ, и прогнозирование, и управление самым непосредственным образом основываются на математическом моделировании объектов. Математическое моделирование в свою очередь предполагают возможность выполнения всех арифметических операций (сложение, вычитание, умножение и деление) над отображениями объектов в моделях и над их элементами.

В практике интеллектуального анализа данных в экономике, социологии, психологии, педагогике и других предметных областях все чаще встречаются ситуации, когда необходимо в рамках единой математической модели *совместно* обрабатывать числовые и нечисловые данные.

В свою очередь числовые данные могут быть различной природы и, соответственно, измеряться в самых различных единицах измерения. Ясно, что арифметические операции можно выполнять только над числовыми данными, измеряемыми в одних единицах измерения.

Данные нечисловой природы, т.е. различные факты и события, характеризуются тем, что с ними вообще нельзя выполнять арифметические операции.

Соответственно, *возникает потребность в математических методах и программном инструментарии, обеспечивающих совместную сопоставимую обработку разнородных числовых данных и данных нечисловой природы.*

Традиционные пути решения проблемы

Традиционно при необходимости проведения подобных исследований реализуется один из двух вариантов, т.е. либо изучается подмножество однородных по своей природе данных, измеряемых в одних единицах измерения; либо перед исследованием данные приводятся к сопоставимому виду, например, широко используются процентные или другие относительные величины, реже – стандартизированные значения.

Ясно, что первый вариант является не решением проблемы, а лишь ее вынужденным обходом, обусловленным ограничениями реально имеющегося в распоряжении исследователей инструментария.

Второй вариант лишь частично решает проблему, т.к. хотя и снимает различие в единицах измерения, но не преодолевает принципиального различия

между количественными и качественными (нечисловыми) величинами и не позволяет обрабатывать их совместно в рамках единой модели.

В последние годы развивается ряд новых методов статистики, полный обзор которых дан в работах А. И. Орлова [<http://antorlov.chat.ru>]. Прежде всего, это интервальная статистика, статистика объектов нечисловой природы, робастные, бутстрепные и непараметрические методы.

В частности методы интервальной статистики, позволяют сводить числовые величины к фактам попадания их значений в определенные интервалы, т.е. к событиям. При этом преодолевается проблема различия в размерности числовых величин. Это обеспечивает также обработку числовых величин, как событий совместно с информацией о других событиях, связанных с объектами нечисловой природы. Таким образом, *интервальные методы сводят обработку числовых величин к методам обработки нечисловой информации* и позволяет обрабатывать их *единообразно по одной методике*. И это является очень важным достижением.

Идея решения проблемы

Это, в общем-то, вполне очевидный и естественный ход. Однако достигается этот результат *дорогой ценой*, т.е. путем сведения числовых величин к нечисловым, т.е. путем сведения их к "низменному типу", что приводит к утрате ряда возможностей обработки. Это происходит потому, что для числовых величин существует гораздо больше методов и возможностей обработки, чем для нечисловых.

По нашему мнению более предпочтительным является противоположный подход, основанный на введении некоторой количественной меры, позволяющей единым и сопоставимым образом описывать как числовые данные различной природы, так и нечисловые величины с использованием всего арсенала возможностей, имеющегося при обработке числовых данных.

Аналогично, если у нас есть документы стандартов "Документ Word" и "Текст-DOS" и мы хотели бы обрабатывать их все в одном редакторе, то это можно сделать либо преобразовав все документы Word в "низменный стандарт" "Текст-DOS", либо наоборот, преобразовав "досовские" документы в формат Word.

В 1979 году автором разработана, а в 1981 году впервые применена математическая модель, обеспечивающая реализацию этой идеи. В последующем этот математический аппарат был развит в ряде работ, основной из которых является [81], был разработана соответствующая ему методика численных расчетов, включающая структуры данных и алгоритмы базовых когнитивных операций, а также создана программная система "Эйдос", реализующая математическую модель и методику численных расчетов [76-155, 185-192].

Предложенный метод получил название "Системно-когнитивный анализ" (СК-анализ) [81]. В СК-анализе нечисловым величинам тем же методом, что и числовым, *приписываются сопоставимые в пространстве и времени, а также между собой, количественные значения, позволяющие обрабатывать их как числовые.*

СК-анализ включает следующие этапы:

1. Когнитивная структуризация, а затем и формализация предметной области.
2. Ввод данных мониторинга в базу прецедентов за период, в течение которого имеется необходимая информация в электронной форме.
3. Синтез семантической информационной модели (СИМ).
4. Оптимизация СИМ.
5. Проверка адекватности СИМ (измерение внутренней и внешней, дифференциальной и интегральной валидности).
6. Анализ СИМ.
7. Решение задач идентификации состояний объекта управления, прогнозирование и поддержка принятия управленческих решений по управлению с применением СИМ.

На первых двух этапах СК-анализа, детально рассмотренных в работе [81], числовые величины сводятся к интервальным оценкам, как и информация об объектах нечисловой природы (фактах, событиях). Этот этап реализуется и в методах интервальной статистики.

На третьем этапе СК-анализа всем этим величинам по единой методике, основанной на системном обобщении семантической теории информации А.Харкевича, сопоставляются количественные величины, с которыми в дальнейшем и производятся все операции моделирования.

Математическая модель СК-анализа

Системное обобщение формулы Хартли

В выражении (3) приведено системное обобщение формулы Хартли для равновероятных состояний объекта управления.

$$I = \log_2 W \quad (1) \quad I = \log_2 (C_W^1 + C_W^2 + \dots + C_W^M) \quad (4)$$

$$I = \log_2 W^\varphi \quad (2) \quad \text{при } M = W : \sum_{m=1}^M C_W^m = 2^W - 1 \quad (5)$$

$$\boxed{I = \log_2 \sum_{m=1}^M C_W^m} \quad (3) \quad I = \log_2 (2^W - 1) \approx W \quad (6)$$

при $W \gg 1$; $I \approx W$ с очень малой и быстро уменьшающейся погрешностью

W – количество чистых (классических) состояний системы.

φ – коэффициент эмерджентности Хартли (уровень системной организации объекта, имеющего W чистых состояний).

Гипотеза о Законе возрастания эмерджентности

Исследование математических выражений системной теории информации (7 – 12) позволило сформулировать гипотезу о существовании "Закона возрастания эмерджентности". Суть этой гипотезы в том, что в самих элементах системы содержится сравнительно небольшая доля всей содержащейся в ней информации, а основной ее объем составляет системная информация, содержащаяся в подсистемах различного уровня иерархии.

Различие между классическим и предложенным системным понятиями информации соответствует различию между понятиями МНОЖЕСТВА И СИСТЕМЫ, на основе которых они сформированы.

$$I = \text{Log}_2 W^\varphi = \text{Log}_2 \sum_{m=1}^M C_W^m \quad (7)$$

$$\varphi = \frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W} \quad (8)$$

$$I(W, M) = \text{Log}_2 W \frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W} \quad (9) \quad I(W, M) \approx \text{Log}_2 W \frac{W}{\text{Log}_2 W} = W \quad (10)$$

$$I_{\text{системная}} \approx W - \text{Log}_2 W \quad (11) \quad I(W, M) = \text{Log}_2 W + \text{Log}_2 W^{\varphi-1} \quad (12)$$

Математическая формулировка:

$$\varphi = \frac{\text{Log}_2 \sum_{m=1}^M C_W^m}{\text{Log}_2 W} \approx \frac{W}{\text{Log}_2 W}$$

$$I_{\text{системная}} \approx W - \text{Log}_2 W$$

Интерпретация



Системное обобщение формулы Харкевича

Ниже приведен вывод системного обобщения формулы Харкевича, а именно:

– классическая формула Харкевича через вероятности перехода системы в целевое состояние при условии сообщения ей определенной информации и самопроизвольно (13);

– выражение классической формулы Харкевича через частоты (14, 15);

- вывод коэффициента эмерджентности Харкевича на основе принципа соответствия с выражением Хартли в детерминистском случае (16 – 19);
- вывод системного обобщения формулы Харкевича;
- окончательное выражение для системного обобщения формулы Харкевича (21).

Классическая формула Харкевича

$$I_{ij} = \text{Log}_2 \frac{P_{ij}}{P_j} \quad (13)$$

P_{ij} – вероятность перехода объекта управления в j -е состояние в условиях действия i -го фактора;

P_j – вероятность самопроизвольного перехода объекта управления в j -е состояние, т.е. в условиях отсутствия действия i -го фактора или в среднем.

Известно, что *корреляция не является мерой причинно-следственных связей*. Если корреляция между действием некоторого фактора и переходом объекта управления в определенное состояние высока, то это еще не значит, что данный фактор является причиной этого перехода. Для того чтобы по корреляции можно было судить о наличии причинно-следственной связи необходимо сравнить исследуемую группу с *контрольной группой*, т.е. с группой, в которой данный фактор не действовал.

Также и высокая вероятность перехода объекта управления в определенное состояние в условиях действия некоторого фактора сама по себе не говорит о наличии причинно-следственной связи между ними, т.е. о том, что данный фактор обусловил переход объекта в это состояние. Это связано с тем, что вероятность перехода объекта в это состояние может быть вообще очень высокой независимо от действия фактора. Поэтому в качестве меры силы причинной обусловленности определенного состояния объекта действием некоторого фактора Харкевич предложил логарифм *отношения* вероятностей перехода в объекта в это состояние в условиях действия фактора и при его отсутствии или в среднем (13).

Таким образом семантическая мера информации Харкевича является мерой наличия причинно-следственных связей между факторами и состояниями объекта управления.

Выражение классической формулы Харкевича через частоты фактов

$$P_{ij} = \frac{N_{ij}}{N_i}; P_i = \frac{N_i}{N}; P_j = \frac{N_j}{N}; \quad (14)$$

$$\text{где: } N_i = \sum_{j=1}^W N_{ij}; N_j = \sum_{i=1}^M N_{ij}; N = \sum_{i=1}^W \sum_{j=1}^M N_{ij}$$

$$I_{ij} = \text{Log}_2 \frac{N_{ij} N}{N_i N_j} \quad (15)$$

Вывод коэффициента эмерджентности Харкевича на основе принципа соответствия с выражением Хартли в детерминистском случае

Однако мера Харкевича (13) не удовлетворяет принципу соответствия мерой Хартли как мера Шеннона, т.е. не переходит в меру Хартли в детерминистском случае, т.е. когда каждому будущему состоянию объекта управления соответствует единственный уникальный фактор и между факторами и состояниями имеется взаимно однозначное соответствие (17).

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^\Psi \quad (16) \quad \forall N_{ij} = N_i = N_j = 1 \quad (17)$$

Откуда:

$$I_{ij} = \text{Log}_2 N^\Psi = \text{Log}_2 W^\varphi \quad (18) \quad \boxed{\Psi = \frac{\text{Log}_2 W^\varphi}{\text{Log}_2 N}} \quad (19)$$

Вывод системного обобщения формулы Харкевича

$$\begin{aligned} I_{ij} &= \text{Log}_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^\Psi = \text{Log}_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^{\frac{\text{Log}_2 W^\varphi}{\text{Log}_2 N}} = \\ \Psi &= \frac{\text{Log}_2 W^{\frac{\sum_{m=1}^M C_W^m}{\text{Log}_2 W}}}{\text{Log}_2 N} \quad (20) \quad = \frac{\text{Log}_2 W^\varphi}{\text{Log}_2 N} \left(\text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right) + \text{Log}_2 N \right) = \\ &= \text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\text{Log}_2 W^\varphi}{\text{Log}_2 N}} + \text{Log}_2 W^\varphi \end{aligned}$$

Окончательное выражение для системного обобщения формулы Харкевича

$$\boxed{I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\text{Log}_2 W^\varphi}{\text{Log}_2 N}} + \text{Log}_2 W^\varphi} \quad (21)$$

Связь системной теории информации (СТИ) с теорией Хартли-Найквиста-Больцмана и теорией Шеннона

Связь между выражениями для плотности информации в теориях Хартли, Шеннона и СТИ приведена на рисунке 41.

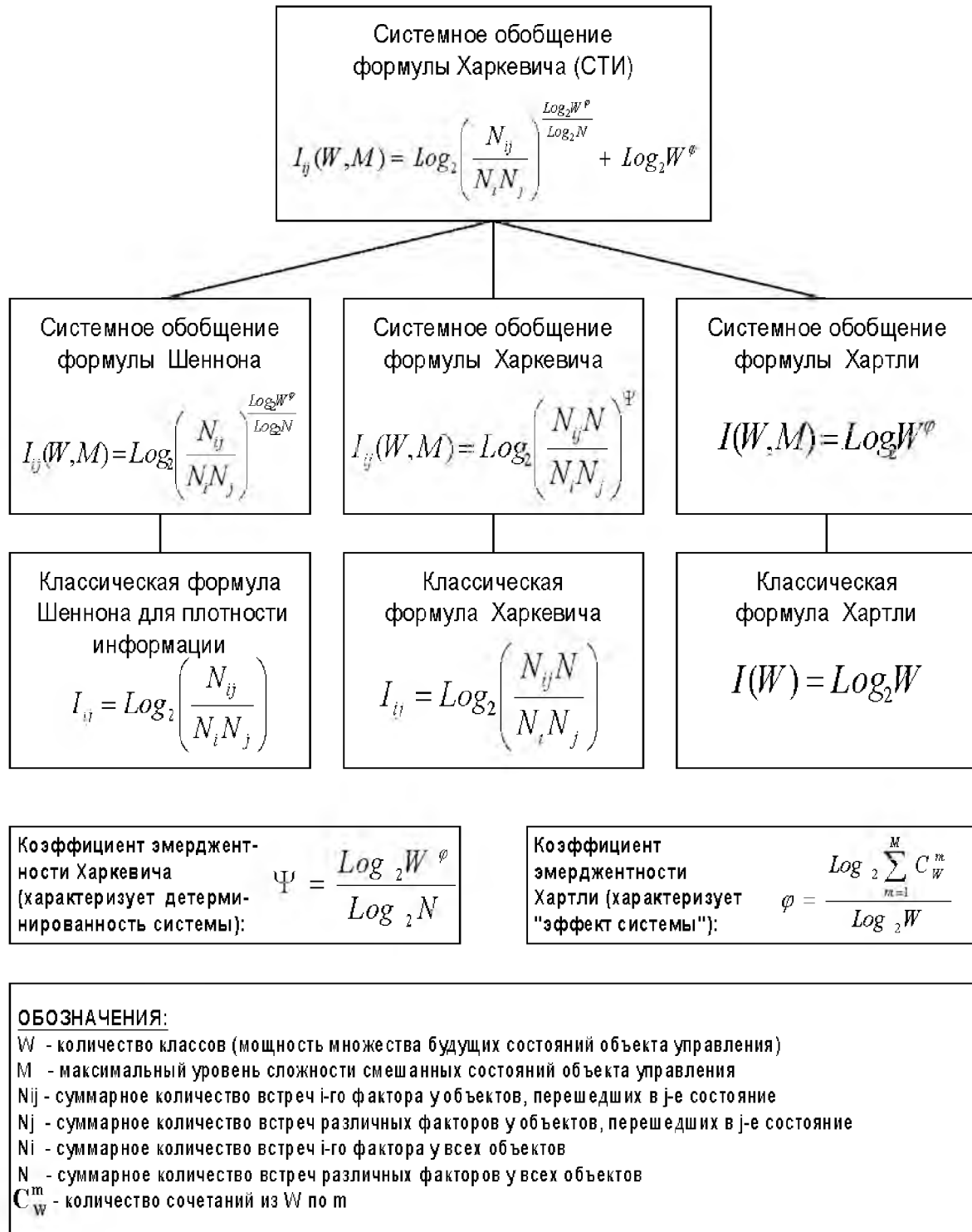


Рисунок 41. Связь между выражениями для плотности информации в теориях Хартли, Шеннона и СТИ

Интерпретация коэффициентов эмерджентности СТИ

Интерпретация коэффициентов эмерджентности, предложенных в рамках системной теории информации, приведена на рисунке 42.

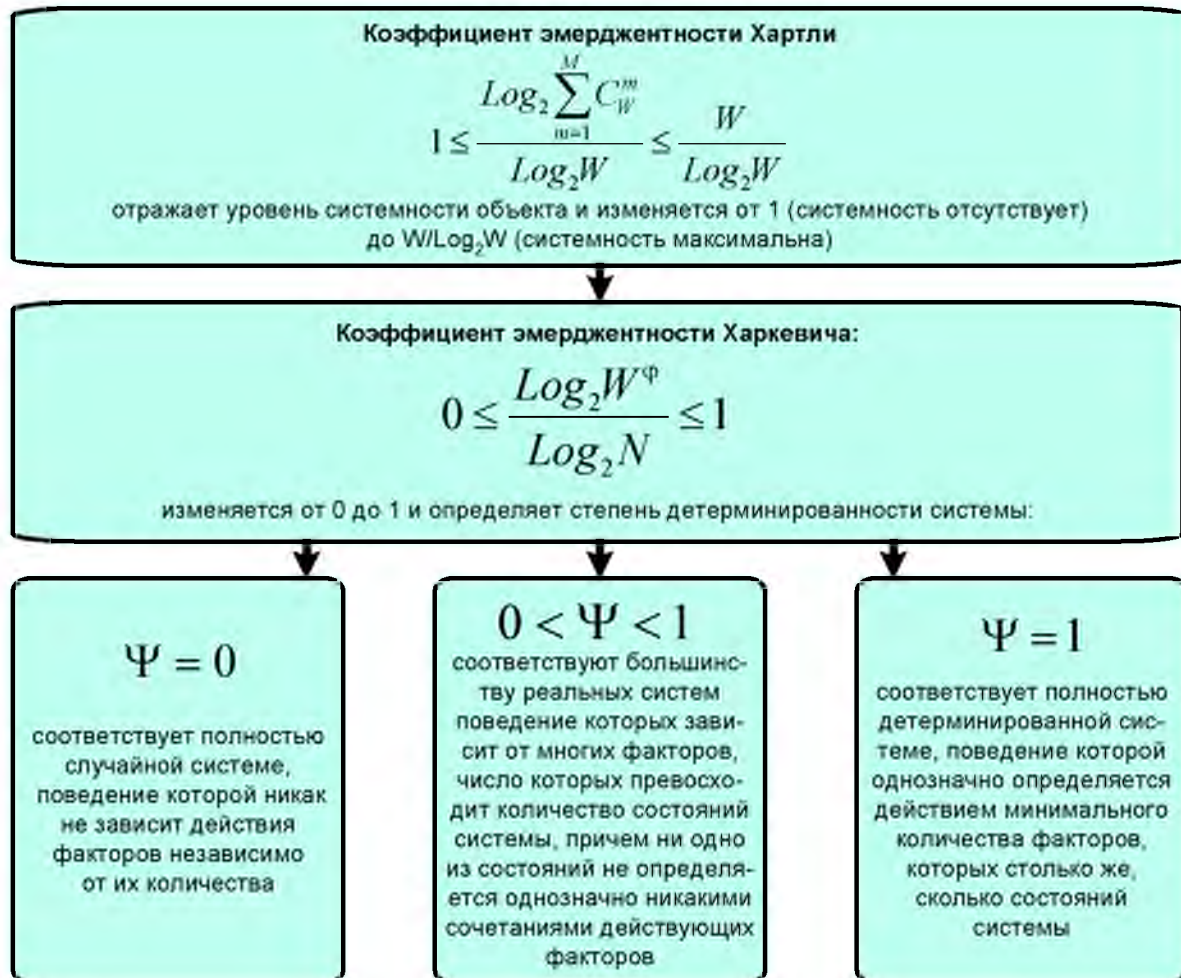


Рисунок 42. Интерпретация коэффициентов эмерджентности СТИ

Коэффициент эмерджентности Хартли ϕ (4) представляет собой относительное превышение количества информации о системе при учете системных эффектов (смешанных состояний, иерархической структуры ее подсистем и т.п.) над количеством информации без учета системности, т.е. этот коэффициент является аналитическим выражением для уровня системности объекта.

Коэффициент эмерджентности Харкевича Ψ , изменяется от 0 до 1 и определяет степень детерминированности системы.

Таким образом, в предложенном системном обобщении формулы Харкевича (21) впервые непосредственно в аналитическом выражении для самого понятия "Информация" отражены такие фундаментальные свойства систем, как "Уровень системности" и "Степень детерминированности" системы.

Матрица абсолютных частот

Основной формой первичного обобщения эмпирической информации в модели является матрица абсолютных частот (таблица 23). В этой матрице строки соответствуют факторам, столбцы – будущим целевым и нежелательным состояниям объекта управления, а на их пересечении приведено количество наблюдения фактов (по данным обучающей выборки), когда действовал некоторый i -й фактор и объект управления перешел в некоторое j -е состояние.

Таблица 23 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ

Атрибуты		Классы - будущие состояния объекта управления					Сумма
		Целевые состояния		Нежелательные состояния			
		***	J	***	I	***	
Факторы, характеризующие текущее и прошлые состояния объекта управления, в т.ч. его рефлексивность	***						
	r		N_{rj}		N_{ri}		$N_r = \sum_{j=1}^W N_{rj}$

Управляющие факторы системы управления	i		N_{ij}		N_{ii}		$N_i = \sum_{j=1}^W N_{ij}$

Факторы, характеризующие прошлые, текущее и прогнозируемые состояния окружающей среды	k		N_{kj}		N_{ki}		$N_k = \sum_{j=1}^W N_{kj}$

Сумма			$N_j = \sum_{i=1}^M N_{ij}$		$N_i = \sum_{i=1}^M N_{ii}$		$N = \sum_{i=1}^M \sum_{j=1}^W N_{ij}$

где:

 N_{ij} – количество встреч i-го признака у объектов j-го класса по данным обучающей выборки**Матрица информативностей**

Непосредственно на основе матрицы абсолютных частот с использованием системного обобщения формулы Харкевича (21) рассчитывается матрица информативностей (таблица 24).

Таблица 24 – МАТРИЦА ИНФОРМАТИВНОСТЕЙ

Атрибуты		Классы - будущие состояния объекта управления				Средняя детерминирующая мощность фактора
		Целевые состояния		Нежелательные состояния		
		***	j	***	I	
Факторы, характеризующие текущее и прошлые состояния объекта управления, в т.ч. его рефлексивность	***					
	r		$I_{rj} = \Psi \cdot \text{Log}_2 \frac{N_{rj} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ri}}$		$I_{ri} = \Psi \cdot \text{Log}_2 \frac{N_{ri} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ri}}$	$\sigma_r = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{rj} - \bar{I}_r)^2}$
Управляющие факторы системы управления	***					
	i		$I_{ij} = \Psi \cdot \text{Log}_2 \frac{N_{ij} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ii}}$		$I_{ii} = \Psi \cdot \text{Log}_2 \frac{N_{ii} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ii}}$	$\sigma_i = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
Факторы, характеризующие прошлые, текущее и прогнозируемые состояния окружающей среды	***					
	k		$I_{kj} = \Psi \cdot \text{Log}_2 \frac{N_{kj} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ki}}$		$I_{ki} = \Psi \cdot \text{Log}_2 \frac{N_{ki} \cdot \sum_{i=1}^M \sum_{j=1}^W N_{ij}}{\sum_{i=1}^M N_{ij} \cdot \sum_{j=1}^W N_{ki}}$	$\sigma_k = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{kj} - \bar{I}_k)^2}$
Средняя детерминированность будущих состояний АОУ	***					
			$\sigma_j = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (I_{ij} - \bar{I}_j)^2}$		$\sigma_i = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (I_{ii} - \bar{I}_i)^2}$	$H = \sqrt{\frac{1}{(W \cdot M - 1) \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}}$

 $\bar{I}_j = \frac{1}{M} \sum_{i=1}^M I_{ij}$ – среднее значение координат вектора класса, M – количество факторов.

 $\bar{I}_i = \frac{1}{W} \sum_{j=1}^W I_{ii}$
 $\bar{I} = \frac{1}{W \cdot M} \sum_{j=1}^W \sum_{i=1}^M I_{ij}$ – среднее значение координат вектора фактора, W – количество классов (будущих состояний АОУ).

 $\bar{I} = \frac{1}{W \cdot M} \sum_{j=1}^W \sum_{i=1}^M I_{ij}$ – средняя информативность признаков по матрице информативностей.

 $\Psi = \frac{\text{Log}_2 \sum_{i=1}^M C_{ij}^m}{\text{Log}_2 W}$ – коэффициент эмерджентности Хартели.

 $\Psi = \frac{\text{Log}_2 W^m}{\text{Log}_2 N}$
 Ψ – коэффициент эмерджентности Харкевича.

 H – мера уровня системности предметной области в рамках СИ

Матрица информативностей является универсальной формой представления **смысла** эмпирических данных в единстве их дискретного и интегрального представления (причины – последствия, факторы – результирующие состояния, признаки – обобщенные образы классов, образное – логическое, дискретное – интегральное).

Весовые коэффициенты матрицы информативностей непосредственно определяют, какое количество информации I_{ij} система управления получает о наступлении события: "объект управления перейдет в j -е состояние", из сообщения: "на объект управления действует i -й фактор".

Когда количество информации $I_{ij} > 0$ – i -й фактор способствует переходу объекта управления в j -е состояние, когда $I_{ij} < 0$ – препятствует этому переходу, когда же $I_{ij} = 0$ – никак не влияет на это.

Таким образом, предлагаемая семантическая информационная модель позволяет непосредственно на основе эмпирических данных и независимо от предметной области рассчитать, какие количество информации содержится в любом событии о любом другом событии.

Этот вывод является ключевым для данной работы, т.к. конкретно показывает возможность числовой обработки в СК-анализе как числовой, так и нечисловой информации.

Матрица информативностей является также обобщенной (неклассической) таблицей решений, в которой входы (факторы) и выходы (будущие состояния объекта управления) связаны друг с другом не с помощью классических (Аристотелевских) импликаций, принимающих только значения: "Истина" и "Ложь", а различными значениями истинности, выраженными в битах и принимающими значения от положительного теоретически-максимально-возможного, до теоретически неограниченного отрицательного. Некоторые неклассические высказывания, генерируемые на основе матрицы информативности, приведены на плакате.

Неметрический интегральный критерий сходства, основанный на лемме Неймана-Пирсона

В выражениях (22 – 24) приведен неметрический интегральный критерий сходства, основанный на фундаментальной лемме Неймана-Пирсона, обеспечивающий идентификацию и прогнозирование в предложенных **неортонормированных** семантических пространствах с **финитной метрикой**, *в которых в качестве координат векторов будущих состояний объекта управления и факторов выступает количество информации, рассчитанное в соответствии с системной теорией информации (21)*, а не Булевы координаты или частоты, как обычно.

$$I_j = f(\vec{I}_{ij}). \quad (22)$$

$$I_j = (\vec{I}_{ij}, \vec{L}_i). \quad (23)$$

Или в координатной форме:

$$I_j = \sum_{i=1}^M I_{ij} L_i, \quad (24)$$

$$j^* = \arg \max_{j \in J} ((\vec{I}_{ij}, \vec{L}_i)), \quad (25)$$

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор j -го состояния объекта управления;

$\vec{L}_i = \{L_i\}$ – вектор состояния предметной области, включающий все виды факторов, характеризующих объект управления, возможные управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, \text{ если } i - \text{й фактор действует;} \\ \alpha_i, \text{ где } 0 < \alpha_i < 1, \text{ если } i - \text{й фактор действует с истинностью } \alpha_i; \\ 0, \text{ если } i - \text{й фактор не действует.} \end{cases}$$

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}), \quad (26) \quad I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_l} \quad (27)$$

\bar{I}_j – средняя информативность по вектору класса;

\bar{L} – среднее по вектору идентифицируемой ситуации (объекта).

σ_j – среднеквадратичное отклонение информативностей вектора класса;

σ_l – среднеквадратичное отклонение по вектору распознаваемого объекта.

Связь системной меры целесообразности информации с критерием χ^2

В (28 – 33) показана связь системной меры целесообразности информации с известным критерием χ^2 , а также предложен новый критерий уровня системности предметной области, являющийся нормированным объемом семантического пространства (34, 35).

$$\chi^2 = \sum_{j=1}^W \sum_{i=1}^M \frac{(N_{ij} - t)^2}{t} \quad (28)$$

$$t = \frac{N_i N_j}{N} \quad (29)$$

– N_{ij} – фактическое количество встреч i -го признака у объектов j -го класса;

– t – ожидаемое количество встреч i -го признака у объектов j -го класса.

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij} N}{N_i N_j} \right)^\Psi \quad (30)$$

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{t} \right)^\Psi \quad (31)$$

$$I_{ij} = \Psi(\text{Log}_2 N_{ij} - \text{Log}_2 t) \quad (32) \quad \begin{cases} \text{если } N_{ij} < t \text{ то } \chi_{ij} > 0, & I_{ij} < 0 \\ \text{если } N_{ij} = t \text{ то } \chi_{ij} = 0, & I_{ij} = 0 \\ \text{если } N_{ij} > t \text{ то } \chi_{ij} > 0, & I_{ij} > 0 \end{cases} \quad (33)$$

$$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2} \quad (34) \quad \bar{I} = \frac{1}{W \cdot M} \sum_{j=1}^W \sum_{i=1}^M I_{ij} \quad (35)$$

Предлагается более точный критерий уровня системности модели является объем неортонормированного семантического пространства, рассчитанный как объем многомерного параллелепипеда, ребрами которого являются оси семантического пространства. Однако для этой меры сложнее в общем виде записать аналитическое выражение и для ее вычисления могут быть использованы численные методы с использованием многомерного обобщения смешанного произведения векторов.

Абстрагирование (ортонормирование) существенно уменьшает размерность семантического пространства без существенного уменьшения его объема.

Оценка адекватности семантической информационной модели в СК-анализе и бутстрепные методы

Под адекватностью модели СК-анализа понимается ее внутренняя и внешняя дифференциальная и интегральная валидность. Понятие валидности является уточнением понятия адекватности, для которого определены процедуры количественного измерения, т.е. валидность – это количественная адекватность. Это понятие количественно отражает способность модели давать правильные результаты идентификации, прогнозирования и способность вырабатывать правильные рекомендации по управлению.

Под внутренней валидностью понимается валидность модели, измеренная после синтеза модели путем идентификации объектов обучающей выборки.

Под внешней валидностью понимается валидность модели, измеренная после синтеза модели путем идентификации объектов, не входящих в обучающую выборку.

Под дифференциальной валидностью модели понимается достоверность идентификации объектов в разрезе по классам.

Под интегральной валидностью средневзвешенная дифференциальная валидность.

Возможны все сочетания: внутренняя дифференциальная валидность, внешняя интегральная валидность и т.д.

Основная идея бутстреп по Б.Эфрону [107] состоит в том, что методом Монте-Карло (статистических испытаний) многократно извлекаются выборки из эмпирического распределения. Эти выборки, естественно, являются вариантами исходной, напоминают ее.

Эта идея позволяет сконструировать алгоритм измерения адекватности модели, состоящий из двух этапов:

1. Синтез модели на одном случайном подмножестве обучающей выборки.

2. Измерение валидности модели на оставшемся подмножестве обучающей выборки, не использованном для синтеза модели.

Поскольку оба случайных подмножества имеют переменный состав по объектам обучающей выборки, то подобная процедура должна повторяться много раз, после чего могут быть рассчитаны статистические характеристики адекватности модели, например, такие как:

– средняя внешняя валидность;

– среднеквадратичное отклонение текущей внешней валидности от средней и другие.

Достоинство бутстрепного подхода к оценке адекватности модели состоит в том, что он позволяет измерить внешнюю валидность на уже имеющейся выборке и изучить статистические характеристики, характеризующие адекватность модели при изменении объема и состава выборки.

Непараметричность модели. Робастные процедуры и фильтры для исключения артефактов

Предложенная семантическая информационная модель является *непараметрической*, т.к. базируется на системной теории информации [81], которая никоим образом не основана на предположениях о нормальности распределений исследуемой выборки.

Под робастными понимаются процедуры, обеспечивающие устойчивую работу модели на исходных данных, зашумленных артефактами, т.е. данными, выпадающими из общих статистических закономерностей, которым подчиняется исследуемая выборка.

Критерий выявления артефактов, реализованный в СК-анализе, основан на том, что при увеличении объема статистики частоты значимых атрибутов растут, как правило, пропорционально объему выборки, а частоты артефактов так и остаются чрезвычайно малыми, близкими к единице. Таким образом, выявление артефактов возможно только при достаточно большой статистике, т.к. в противном случае недостаточно информации о поведении частот атрибутов с увеличением объема выборки.

В модели реализована такая процедура удаления наиболее вероятных артефактов, и она, как показывает опыт, существенно повышает качество (адекватность) модели.

Методика численных расчетов СК-анализа

Детальный список БКОСА и их алгоритмов

В таблице 25 приведен детальный список базовых когнитивных операций системного анализа, которым соответствует 24 алгоритма, которые здесь привести нет возможности из-за их объемности. Но они все приведены в полном виде в работе [81].

**Таблица 25 – ДЕТАЛЬНЫЙ СПИСОК
БАЗОВЫХ КОГНИТИВНЫХ ОПЕРАЦИЙ СИСТЕМНОГО АНАЛИЗА (БКОСА)**

№ алгоритма	Код БКОСА по схеме СК-анализа	№ БКОСА	Наименование БКОСА	Полное наименование базовых когнитивных операций системного анализа (БКОСА)
	1.1	1	Присвоение имен	Присвоение имен классам (интенциональная, интегральная репрезентация)
	1.2			Присвоение имен атрибутам (экстенциональная, дискретная репрезентация)
1	2.1.	2	Восприятие	Восприятие и запоминание исходной обучающей информации
2	2.2.			Репрезентация. Сопоставление индивидуального опыта с коллективным (общественным)

Продолжение таблицы 25

№ алгоритма	Код БКОСА по схеме СК-анализа	№ БКОСА	Наименование БКОСА	Полное наименование базовых когнитивных операций системного анализа (БКОСА)
3	3.1.1.	3	Обобщение (синтез, индукция).	Накопление первичных данных
4	3.1.2.			Исключение артефактов
5	3.1.3.			Расчет истинности смысловых связей между предпосылками и результатами (обобщенных таблиц решений)
6	3.2.			Определение значимости шкал и градаций факторов, уровней Мерлина
7	3.3.			Определение значимости шкал и градаций классов, уровней Мерлина
8	4.1.	4	Абстрагирование	Абстрагирование факторов (снижение размерности семантического пространства факторов)
9	4.2.			Абстрагирование классов (снижение размерности семантического пространства классов)
10	5.	5	Оценка адекватности	Оценка адекватности информационной модели предметной области
11	7.	6	Сравнение, идентификация и прогнозирование	Сравнение, идентификация и прогнозирование. Распознавание состояний конкретных объектов (объектный анализ)
12	9.1.	7	Анализ, дедукция и абдукция	Анализ, дедукция и абдукция классов (семантический анализ обобщенных образов классов, решение обратной задачи прогнозирования)
13	9.2.			Анализ, дедукция и абдукция факторов (семантический анализ факторов)
14	10.1.1.	8	Классификация и генерация конструктов	Классификация обобщенных образов классов
15	10.1.2.			Формирование бинарных конструктов классов
16	10.1.3.			Визуализация семантических сетей классов
17	10.2.1.			Классификация факторов
18	10.2.2.			Формирование бинарных конструктов факторов
19	10.2.3.	9	Содержательное сравнение	Визуализация семантических сетей факторов
20	10.3.1.			Содержательное сравнение классов
21	10.3.2.			Расчет и отображение многозначных когнитивных диаграмм, в т.ч. диаграмм Мерлина
22	10.4.1.			Содержательное сравнение факторов
23	10.4.2.			Расчет и отображение многозначных когнитивных диаграмм, в т.ч. инвертированных диаграмм Мерлина
24	11.	10	Планирование и управление	Многовариантное планирование и принятие решения о применении системы управляющих факторов

Иерархическая структура данных семантической информационной модели СК-анализа

На рисунке 43 приведена в обобщенном виде иерархическая структура баз данных семантической информационной модели системно-когнитивного анализа. На этой схеме базы данных обозначены **прямоугольниками**, а базовые когнитивные операции системного анализа, преобразующие одну базу в другую – **стрелками** с надписями. Имеются также базовые когнитивные операции, формирующие выходные графические формы. Из этой схемы видно, что одни базовые когнитивные операции готовят данные для других операций, относящихся к более высоким уровням иерархии системы процессов познания. Этим определяется возможная последовательность выполнения базовых когнитивных операций.

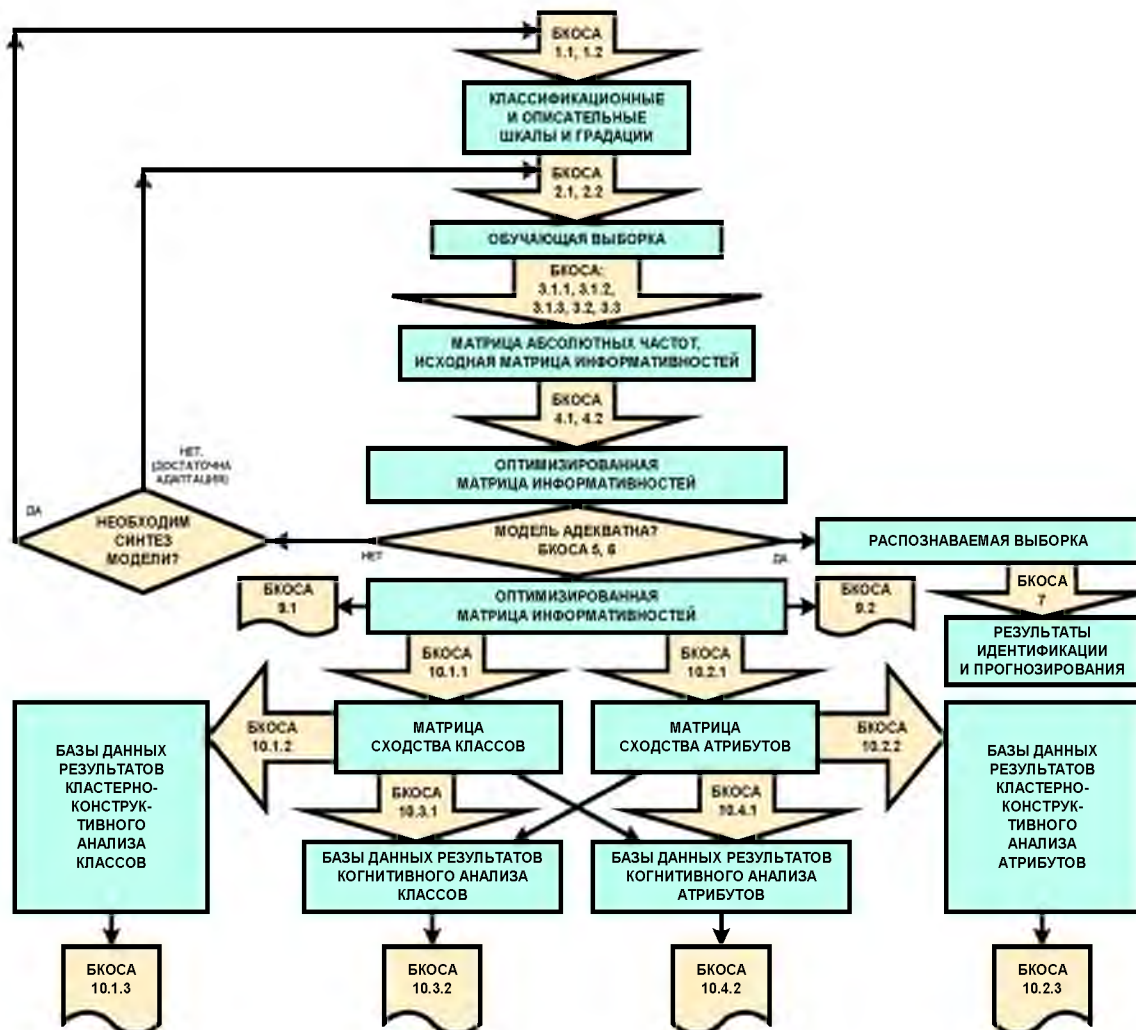


Рисунок 43. Иерархическая структура баз данных семантической информационной модели системно-когнитивного анализа

Специальный программный инструментарий СК-анализа – система "Эйдос"

На таблице 26 показана обобщенная схема когнитивной аналитической системы "Эйдос", которая реализует математическую модель и численный метод системно-когнитивного анализа и, таким образом, является его инструментари-ем.

В состав данной системы входит 7 подсистем.

Первые 3 подсистемы являются **инструментальными**, т.е. позволяют осуществлять **синтез и адаптацию** модели.

Остальные 4 подсистемы обеспечивают идентификацию, прогнозирование и кластерно-конструктивный анализ модели, в т.ч. верификацию модели и выработку управляющих воздействий.

Система "Эйдос" является довольно большой системой: распечатка ее исходных текстов 6-м шрифтом составляет около 800 листов, она генерирует 54 графических формы (двумерные и трехмерные) и 50 текстовых форм. На данную систему и системы окружения получено 8 свидетельств РосПатента РФ.

**Таблица 26 – ОБОБЩЕННАЯ СТРУКТУРА УНИВЕРСАЛЬНОЙ
КОГНИТИВНОЙ АНАЛИТИЧЕСКОЙ СИСТЕМЫ "ЭЙДОС"**

Подсистема	Режим	Функция	Операция	
1. Словари	1. Классификационные шкалы и градации			
	2. Описательные шкалы (и градации)			
	3. Градации описательных шкал (признаки)			
	4. Иерархические уровни систем	1. Уровни классов		
		2. Уровни признаков		
	5. Программные интерфейсы для импорта данных	1. Импорт данных из TXT-фалов стандарта DOS-текст		
		2. Импорт данных из DBF-файлов стандарта проф. А.Н.Лебедева		
		3. Импорт из транспонированных DBF-файлов проф. А.Н.Лебедева		
4. Генерация шкал и обучающей выборки RND-модели				
5. Генерация шкал и обучающей выборки для исследования чисел				
6. Почтовая служба по НСИ	6. Транспонирование DBF-матриц исходных данных			
	1. Обмен по классам			
	2. Обмен по обобщенным признакам			
3. Обмен по первичным признакам				
7. Печать анкеты				
2. Обучение	1. Ввод–корректировка обучающей выборки			
	2. Управление обучающей выборкой	1. Параметрическое задание объектов для обработки		
		2. Статистическая характеристика, ручной ремонт		
		3. Автоматический ремонт обучающей выборки		
	3. Пакетное обучение системы распознавания	1. Накопление абсолютных частот		
		2. Исключение артефактов (робастная процедура)		
		3. Расчет информативностей признаков		
		4. Расчет условных процентных распределений		
		5. Автоматическое выполнение режимов 1–2–3–4		
		6. Измерение схожести и устойчивости модели	1. Сходимость и устойчивость СИМ	
2. Зависимость валидности модели от объема обучающей выборки				
4. Почтовая служба по обучающей информации				
3. Оптимизация	1. Формирование ортонормированного базиса классов			
	2. Исключение признаков с низкой селективной силой			
	3. Удаление классов и признаков, по которым недостаточно данных			
4. Распознавание	1. Ввод–корректировка распознаваемой выборки			
	2. Пакетное распознавание			
	3. Вывод результатов распознавания	1. Разрез: один объект – много классов		
		2. Разрез: один класс – много объектов		
	4. Почтовая служба по распознаваемой выборке			
5. Типология	1. Типологический анализ классов распознавания	1. Информационные (ранговые) портреты (классов)		
		2. Кластерный и конструктивный анализ классов	1. Расчет матрицы сходства образов классов	
			2. Генерация кластеров и конструктов классов	
			3. Просмотр и печать кластеров и конструктов	
			4. Автоматическое выполнение режимов: 1,2,3	
			5. Вывод 2d семантических сетей классов	
	3. Когнитивные диаграммы классов			
	2. Типологический анализ первичных признаков	1. Информационные (ранговые) портреты признаков		
		2. Кластерный и конструктивный анализ признаков	1. Расчет матрицы сходства образов признаков	
			2. Генерация кластеров и конструктов признаков	
			3. Просмотр и печать кластеров и конструктов	
			4. Автоматическое выполнение режимов: 1,2,3	
			5. Вывод 2d семантических сетей признаков	
3. Когнитивные диаграммы признаков				
6. Анализ	1. Оценка достоверности заполнения объектов			
	2. Измерение адекватности семантической информационной модели			
	3. Измерение независимости классов и признаков			
	4. Просмотр профилей классов и признаков			
	5. Графическое отображение нелокальных нейронов			
	6. Отображение Паретто-подмножеств нейронной сети			
	7. Классические и интегральные когнитивные карты			

Продолжение таблицы 26

Подсистема	Режим	Функция	Операция
7. Сервис	1. Генерация (сброс) БД	1. Все базы данных	
		2. НСИ	1. Всех баз данных
			2. БД классов
			3. БД первичных признаков
			4. БД обобщенных признаков
		3. Обучающая выборка	
		4. Распознаваемая выборка	
		5. Базы данных статистики	
		2. Переиндексация всех баз данных	
		3. Печать БД абсолютных частот	
		4. Печать БД условных процентных распределений	
		5. Печать БД информативностей	
		6. Интеллектуальная дескрипторная информационно-поисковая система	

Выводы

Интервальные оценки сводят анализ чисел к анализу фактов и позволяют обрабатывать количественные величины как нечисловые данные. Это ограничивает возможности обработки количественных величин методами обработки нечисловых данных. В математической модели СК-анализа, основанной на системной теории информации, наоборот, качественным, нечисловым данным, сопоставляются количественные величины. Это позволяет использовать все возможности количественных методов для исследования нечисловых данных.

Таким образом, в СК-анализе числовые и нечисловые данные обрабатываются единообразно на основе единой математической модели как числовые данные.

Рассматривается связь метода измерения адекватности модели в СК-анализе с бутстрепными методами.

Описывается робастная процедура выявления и устранения артефактов в СК-анализе.

Резюме

1. Сформулированы требования к математической модели и к численной мере. Затем на их основе обоснован выбор базовой численной меры. Для этого рассмотрены три вида информации: абсолютная, относительная и аналитическая информация. Предпочтение отдано аналитической форме информации, к которой принадлежат относительные вероятности, относительные проценты и количество информации. Вместо традиционных мер, основанных на понятии "стоимости" и "полезности" предложено использовать информационную меру. Рассмотрены различные аспекты применения теории информации для анализа процесса труда и средств труда как информационных систем. Показано, что принятие решения об управляющем воздействии есть обратный процесс по отношению к идентификации и прогнозированию, т.е. познанию. Установлена связь количества (синтаксиса) и качества (содержания, семантики) информации, записываемой в структуре предмета труда, с меновой и потребительной стоимостью. Сформулирована информационная теория стоимости, в которой информация рассматривается как сущность стоимости и как "первичный" и по сути единственный товар. Рассмотрены вопросы определения стоимости и амортизация интеллектуальных систем и баз знаний. Показано, что их стоимость как генераторов информации возрастает в процессе эксплуатации. С позиций теории ин-

формации раскрыт фундаментальный источник экономической эффективности АСУ и систем интеллектуальной обработки данных: понижение энтропии объекта управления как приемника сообщений в результате получения управляющей информации. Сделан вывод о целесообразности выбора в качестве базовой численной меры количества информации. Поставлена задача выбора или конструирования конкретной численной меры, основанной на понятии информации.

2. В классической теории информации Шеннона, созданной на основе обобщения результатов Больцмана, Найквиста и Хартли, само понятие информации определяется на основе теоретико-множественных и комбинаторных представлений путем анализа поведения классического макрообъекта, который может переходить только в четко фиксированные альтернативные редуцированные состояния. Однако, квантовые объекты и сложные активные рефлексивные системы могут оказываться одновременно в двух и более альтернативных для классических объектов состояниях. Такие состояния названы смешанными. Таким образом в реальности наблюдается картина последствий, не сводящаяся к простой сумме последствий альтернативных вариантов. Она больше напоминает квантовое физическое явление, которое называется интерференцией плотности вероятности. Это явление, безусловно имеющее системный характер, предлагается называть "интерференция последствий выбора".

3. Предлагается обобщение классической теории информации Хартли-Шеннона путем рассмотрения активных объектов в качестве объектов, на основе анализа которых формируется само основополагающее понятие информации. Обобщенную таким образом теорию информации предлагается называть системной или эмерджентной теорией информации (СТИ). Основным отличием эмерджентной теории информации от классической является учет свойства системности, как фундаментального и универсального свойства всех объектов, на уровне самого понятия информации, а не только в последующем изложении, как в классической теории.

4. Предложена системная модификация формулы Хартли для количества информации:

$$I = \log_2 \sum_{m=1}^M C_W^m$$

где:

W – количество чистых (классических) состояний системы;

C_W^m – сочетания "по m " классических состояний.

Так как $C_W^1 = W$, то при $M=1$ выполняется принцип соответствия, являющийся обязательным для более общей теории. Данная формула дает верхнюю оценку возможного количества информации состояния системы, т.к. возможны различные правила запрета и реальное количество возможных состояний системы будет меньшим, чем C_W^m .

Предложено приближенное выражение для системной модификации формулы Хартли (при $M=W$):

$$I = \log_2 (2^W - 1)$$

При $W > 4$ погрешность данного выражения не превышает 1%.

Дополнительная информация, которую мы получаем из поведения объекта в СТИ, по сути дела является информацией о множестве всех его возможных состояний, как системы, элементом которой является объект в некотором данном состоянии.

5. Численные расчеты и аналитические выкладки согласно СТИ показывают, что при возрастании количества элементов в системе доля системной информации в поведении ее элементов возрастает, причем возрастает ускоренно. Это установленное нами свойство систем названо "Законом возрастания эмерджентности".

6. Предложена системная модификация классической формулы А.Харкевича, являющееся нелинейной суперпозицией классических выражений для плотности информации Шеннона и количества информации Хартли.

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{N_i N_j} \right)^{\frac{\text{Log}_2 W^\varphi}{\text{Log}_2 N}} + \text{Log}_2 W^\varphi$$

где: φ – коэффициент эмерджентности Хартли (уровень системной организации объекта, имеющего W чистых состояний):

$$\varphi = \left(\text{Log}_2 \sum_{m=1}^M C_W^m \right) / \text{Log}_2 W$$

Установлено, что полученное выражение учитывает как взаимосвязь между признаками (факторами) и будущими, в т.ч. целевыми состояниями объекта управления, так и мощность множества будущих состояний. Эта мера отражает уровень системности и степень детерминированности объекта, объединяет возможности их интегрального и дискретного описания, что является основой формализации смысла, а также удовлетворяет принципу соответствия, т.е. преобразуется в формулу Хартли в предельном случае, когда каждому классу (состоянию объекта) соответствует один признак (фактор), и каждому признаку – один класс, и эти классы, а значит и признаки, равновероятны.

7. Все это делает семантическую меру целесообразности информации оптимальной по сформулированным критериям для целей построения семантической информационной модели активных объектов управления и для применения при синтезе рефлексивных АСУ активными системами.

8. В разработанной семантической информационной модели генерируется обобщенная таблица решений, в которой входы (факторы) и выходы будущие состояния активного объекта управления (АОУ) связаны друг с другом не с помощью классических (Аристотелевских) импликаций, принимающих только значения: "Истина" и "Ложь", а различными значениями истинности, выраженными в битах и принимающими значения от положительного теоретически-максимально-возможного ("Максимальная степень истинности"), до теоретически неограниченного отрицательного ("Степень ложности"). Синтез обобщенных таблиц решений для различных предметных областей осуществляется непосредственно на основе эмпирических исходных данных. На основе этих таблиц продуцируются прямые и обратные правдоподобные (нечеткие) логические рас-

суждения по неклассическим схемам с различными расчетными значениями истинности, являющимся обобщением классических импликаций. При этом в прямых рассуждениях как предпосылки рассматриваются факторы, а как заключения – будущие состояния АОУ, а в обратных – наоборот. Степень истинности i -й предпосылки – это количество информации I_{ij} , содержащейся в предпосылке о наступлении j -го будущего состояния активного объекта управления.

9. В качестве меры сходства объекта с классом, класса с классом и атрибута с атрибутом предложено использовать *нечеткий* интегральный критерий, основанный на лемме Неймана-Пирсона, – суммарное количество информации. Если предпосылок несколько, то степень истинности наступления j -го состояния АОУ равна суммарному количеству информации, содержащемуся в них об этом:

$$I_j = (\bar{I}_{ij}, \bar{L}_i).$$

Или в координатной форме:

$$I_j = \sum_{i=1}^A I_{ij} L_i,$$

где: $\bar{I}_{ij} = \{I_{ij}\}$ – вектор j -го состояния объекта управления, координаты которого в информационном семантическом пространстве рассчитываются согласно системного обобщения формулы Харкевича, приведенной в п.6;

$\bar{L}_i = \{L_i\}$ – булев вектор состояния предметной области, включающий все виды факторов, характеризующих объект управления, возможные управляющие воздействия и окружающую среду (массив-локатор). Обоснована замена значений координат этих векторов их стандартизированными значениями.

10. Предложенная семантическая информационная модель позволяет решать задачи идентификации и прогнозирования развития активных систем (разложение вектора объекта по векторам классов – "Объектный анализ"), а также вырабатывать эффективные управляющие воздействия путем решения обратной задачи прогнозирования и применения элементов нетрадиционной логики и правдоподобных (нечетких) рассуждений. В ней объединены преимущества содержательных и статистических моделей, созданы предпосылки для реализации СК-анализа.

11. Исследована взаимосвязь примененной в модели семантической меры Харкевича со статистикой χ^2 , и, на этой основе, предложена новая статистическая мера наличия причинно-следственных связей в предметной области H , основанная на модифицированной формуле Харкевича:

$$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$$

где:

$$\bar{I} = \frac{1}{W \cdot M} \sum_{j=1}^W \sum_{i=1}^M I_{ij} \quad \text{– средняя информативность признаков по матрице информативностей.}$$

Обоснована устойчивость модели при малых и больших выборках, дано обоснование сопоставимости частных критериев, разработана интерпретация распознавания как объектного анализа (разложение вектора объекта в ряд по векторам классов), предложены робастные процедуры, а также процедуры обеспечения структурной репрезентативности выборки.

12. Предлагается модель нелокального нейрона, являющаяся обобщением классической модели Дж. Маккалоки и У. Питта. Суть нелокальности данной модели состоит в том, что весовые коэффициенты каждого нейрона зависят не только от нейронов, прямо или косвенно соединенных с ним синаптически, но и от всех остальных нейронов сети, не затрагиваемых при обратном распространении ошибки от данного нейрона. Предлагается новый класс нейронных сетей: "Нелокальные интерпретируемые нейронные сети прямого счета" (нелокальные нейронные сети – ННС). Организация ННС обеспечивает один из вариантов решения традиционных для нейронных сетей проблем: содержательной интерпретации смысла интенсивности входных сигналов и весовых коэффициентов (*"проблема интерпретируемости весовых коэффициентов"*); содержательной интерпретации и обоснования аддитивности аргумента и вида активационной (передаточной) функции нейрона (*"проблема интерпретируемости передаточной функции"*); "Комбинаторного взрыва" при определении структуры связей нейронов, подборе весовых коэффициентов и передаточных функций (*"проблема размерности"*). Математическая модель ННС основана на предложенной автором системной теории информации и семантической информационной модели автоматизированного системно-когнитивного анализа (АСК-анализ), и в отличие от известных нейронных сетей, обеспечивают автоматизацию всех 10 базовых когнитивных операций, образующих "когнитивный конфигуратор". Предложены не только математическая модель, но также и соответствующий численный метод (включая алгоритмы и структуры данных), а также программный инструментарий нелокальных нейронных сетей (универсальная когнитивная аналитическая система "Эйдос" версии 12.5), успешно апробированные в ряде предметных областей. Данная система обеспечивает практически неограниченное количество слоев ННС, рецепторов, выходных нейронов и связывающих их весовых коэффициентов (десятки миллионов), а также автоматическую визуализацию и запись в виде графических файлов сформированных моделей нелокальных нейронов и Паретто-подмножеств нелокальной нейронной сети.

13. Введено в научный оборот новое понятие: коэффициент эмерджентности Хартли ϕ , который представляет собой относительное превышение количества информации о системе при учете системных эффектов (смешанных состояний, иерархической структуры ее подсистем и т.п.) над количеством информации без учета системности, т.е. этот коэффициент отражает уровень системности объекта. Этот уровень системности объекта изменяется от 1 (системность минимальна, т.е. отсутствует) до $W/\log_2 W$ (системность максимальна). Для каждого количества элементов системы существует свой максимальный уровень системности, который никогда реально не достигается из-за действия правил запрета на реализацию в системе ряда подсистем различных уровней иерархии.

Введено в научный оборот новое понятие: коэффициент эмерджентности Харкевича Ψ , который изменяется от 0 до 1 и определяет степень детерми-

нированности системы: $\Psi=1$ соответствует полностью детерминированной системе, поведение которой однозначно определяется действием минимальным количеством факторов, которых столько же, сколько состояний системы; $\Psi=0$ соответствует полностью случайной системе, поведение которой никак не зависит действия факторов независимо от их количества; $0<\Psi<1$ соответствуют большинству реальных систем, поведение которых зависит от многих факторов, число которых превосходит количество состояний системы, причем ни одно из состояний не определяется однозначно никакими сочетаниями действующих факторов. Увеличение уровня системности влияет на семантическую информационную модель аналогично повышению уровня детерминированности. Понижение уровня системности, также как и степени детерминированности системы приводит к ослаблению влияния факторов на поведение системы, т.е. к своего рода "инфляции факторов".

Основной вывод:

В предложенном системном обобщении формулы Харкевича впервые непосредственно в аналитическом выражении для самого понятия "Информация" отражены такие фундаментальные свойства систем, как "Уровень системности" и "Степень детерминированности" системы, кроме того это выражение (как и формула Шеннона) удовлетворяет принципу соответствия с выражением Хартли в детерминистском случае, учитывает понятие цели и мощность множества будущих состояний объекта управления, объединяет возможности интегрального и дискретного описания объектов. По этим причинам полученное выражение является оптимальным и его целесообразно использовать в качестве основы для построения математической модели рефлексивных АСУ активными объектами.

Контрольные вопросы

1. Предпосылки и теоретические основы системной теории информации.
2. Требования к математической модели и численной мере СТИ.
3. Выбор базовой численной меры СТИ.
4. Конструирование системной численной меры на основе базовой в СТИ.
5. Семантическая информационная модель СК-анализа.
6. Формализм динамики взаимодействующих семантических информационных пространств. Двухвекторное представление данных.
7. Применение классической теории информации К.Шеннона для расчета весовых коэффициентов и мер сходства.
8. Математическая модель метода распознавания образов и принятия решений, основанного на системной теории информации.
9. Некоторые свойства математической модели СК-анализа (сходимость, адекватность, устойчивость и др.).
10. Непараметричность модели. Робастные процедуры и фильтры для исключения артефактов в математической модели СК-анализа.
11. Зависимость информативностей факторов от объема обучающей выборки.
12. Зависимость адекватности семантической информационной модели от объема обучающей выборки (адекватность при малых и больших выборках).
13. Семантическая устойчивость модели СК-анализа.
14. Зависимость параметров модели СК-анализа от ее ортонормированности.
15. Взаимосвязь математической модели СК-анализа с другими моделями.

16. Взаимосвязь системной меры целесообразности информации со статистической χ^2 и новая мера уровня системности предметной области.

17. Сравнение, идентификация и прогнозирование как разложение векторов объектов в ряд по векторам классов (объектный анализ).

18. Системно-когнитивный и факторный анализ. СК-анализ, как метод переменных контрольных групп.

19. Семантическая мера целесообразности информации и эластичность.

20. Связь семантической информационной модели с нейронными сетями.

21. Математический метод СК-анализа в свете идей интервальной бутстрепной робастной статистики объектов нечисловой природы.

Рекомендуемая литература

1. Орлов А.И. "Высокие статистические технологии":

<http://antorlov.chat.ru>.

2. Луценко Е.В. Автоматизированная система распознавания образов: математическая модель и опыт применения. //В сб.: "В.И. Вернадский и современность (к 130-летию со дня рождения)". - Краснодар: КНА, 1993. - С.37-42.

3. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). – Краснодар: КЮИ МВД РФ, 1996. – 280с.

4. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. –605 с.

5. Пат. № 940217. РФ. Универсальная автоматизированная система распознавания образов "ЭЙДОС". /Е.В.Луценко (Россия); Заяв. № 940103. Опубл. 11.05.94. – 50с.

6. Пат. № 2003610986 РФ. Универсальная когнитивная аналитическая система "ЭЙДОС" / Е.В.Луценко (Россия); Заяв. № 2003610510 РФ. Опубл. от 22.04.2003. – 50с.

7. Луценко Е.В. Типовая методика и инструментарий когнитивной структуризации и формализации задач в СК-анализе. // Научный журнал КубГАУ. – 2004.– № 1 (3). –18 с. <http://ej.kubagro.ru>

8. Эфрон Б. Нетрадиционные методы многомерного статистического анализа. - М.: Финансы и статистика, 1988. – 263 с.

1.2.3. ЛЕКЦИЯ-5.

Методика численных расчетов (алгоритмы и структуры данных)

Учебные вопросы

1. Принципы формализации предметной области и подготовки эмпирических данных.
2. Иерархическая структура данных и последовательность численных расчетов в СК-анализе.
3. Обобщенное описание алгоритмов СК-анализа.
4. Детальные алгоритмы СК-анализа.

1.2.3.1. Принципы формализации предметной области и подготовки эмпирических данных

Понятие шкалы и градации. Типы шкал

Формализация предметной области это процесс, состоящий из двух основных этапов:

1. Конструирование шкал и градаций для описания и кодирования состояний объекта управления и факторов, влияющих на его поведение.
2. Отнесение состояний объекта управления и факторов к определенным градациям соответствующих шкал.

В данной работе предлагается следующие определения.

Шкала – это способ классификации объектов по наименованиям или степени выраженности некоторого свойства.

Градация – это положение на шкале (или интервал, диапазон), соответствующее наименованию или определенной степени выраженности свойства.

Понятие шкалы тесно связано с ключевым понятием когнитивной психологии: понятием конструкта, более того, практически является синонимом или формальным аналогом этого понятия.

Конструктом называется понятие, имеющее полюса, противоположные по смыслу, и ряд промежуточных градаций, отражающих различную степень выраженности некоторого качества. Познание состоит в создании (генерировании) новых конструктов и их использовании для ориентации в предметной области. Таким образом, формализация предметной области по сути дела представляет собой ее познание, т.е. когнитивную структуризацию. В приведенной таблице 27 дана характеристика измерительных шкал согласно [81]. Конечно, наименования могут быть присвоены градациям всех видов измерительных шкал.

Таблица 27 – ХАРАКТЕРИСТИКА И ПРИМЕРЫ ИЗМЕРИТЕЛЬНЫХ ШКАЛ

Тип шкалы	Характеристики	Примеры
Номинальная (наименований)	Объекты классифицированы, классам присвоены словесные наименования или условные номера – коды. То, что номер одного класса больше или меньше другого, еще ничего не говорит о свойствах объектов, относящихся к этим классам, за исключением того, что они различаются	Раса, Национальность, цвет глаз, номера на футболках, пол, клинические диагнозы, автомобильные номера, номера страховых.
Порядковая	Объекты классифицированы, а классы обозначены номерами (закодированы). Значения чисел, присваиваемые классам, качественно отражают степень выраженности определенных свойств предметов, принадлежащих этим классам. То есть большим значениям кодов классов соответствует и большая степень выраженности измеряемого свойства, на основании чего классы можно ранжировать.	Твердость минералов, награды за заслуги, ранжирование по индивидуальным чертам личности, военные и гражданские ранги, должности и звания.
Интервальная	Существует единица измерения, при помощи которой классы можно не только упорядочить, но и приписать им числа так, чтобы равные разности чисел присвоенных классам, отражали равные различия в количествах измеряемых свойств. Нулевая точка интервальной шкалы произвольна (условна) и не указывает на отсутствие свойства.	Календарное время, шкалы температур по Фаренгейту и Цельсию.
Отношений	Числа, присвоенные классам, обладают всеми свойствами интервальной шкалы, но помимо этого на шкале существует абсолютный ноль, соответствующий полному отсутствию измеряемого свойства. Отношения чисел, присвоенных классам или объектам при измерении, отражают количественные отношения измеряемого свойства.	Рост, вес, время, цена, количество информации, температура по Кельвину (есть абсолютный ноль).

Шкалы классов (классификационные шкалы)

Плодотворным является представление классов, как некоторых областей в фазовом пространстве, в котором в качестве осей координат выступают некоторые шкалы классов меньшего уровня общности или признаков. Классы распознавания могут рассматриваться, также, как градации (конкретные значения, заданные с некоторой точностью, или диапазоны – зоны), заданные на этих шкалах. Количество шкал, тип шкал и количество градаций на них в предлагаемой модели задает сам пользователь.

Если представить эти шкалы как оси координат, то, очевидно, наиболее обобщенным классам распознавания соответствуют зоны на самих осях. Кроме того возможны варианты сочетаний по 2 оси, соответствующие областям на координатных плоскостях. Существуют также области в фазовом пространстве, образованные сочетаниями градаций сразу n -го количества шкал, где $n \leq N$, где N - размерность фазового пространства. Естественно, пользователь может исследовать только те классы, которые его интересуют, сознательно принимая решение не рассматривать остальных. Но он должен знать, что и остальные классы также могут быть сформированы и исследованы, а для этого нужно иметь их классификацию, принцип разработки которой мы только что рассмотрели.

Конкретными реализациями обобщенных категорий могут быть объекты, их состояния или ситуации (но применять мы, как правило, будем термин "объекты", всегда имея в виду и остальные возможные варианты). Синонимами понятия "класс" являются применяющиеся в специальной литературе термины "объекты", "категории", "образы", "эталоны", "типы", "профили", "вектора". В данной работе объекты рассматриваются как конкретные реализации классов, а классы – как обобщенные образы объектов определенной категории.

Когда классы распознавания сформированы с ними могут осуществляться три основные операции: сравнение конкретных объектов, их состояний или ситуаций с классами; сравнение классов друг с другом; вывод информации о со-

держании обобщенного образа класса в форме таблиц или графических диаграмм.

Шкалы атрибутов (описательные шкалы)

Конкретные объекты, предъявляемые на входе модели в качестве примеров или реализаций некоторых обобщенных классов (прецедентов), описываются на языке атрибутов, т.е. признаков.

Признаки могут иметь любую *природу*, в частности: объективную - физическую, химическую и др. (вес, температура, рост); социально-экономическую (меновую и потребительную стоимость, степень амортизации, процент дивидендов); эмоционально-психологическую (привлекательный, предупредительный, исполнительный, конфликтный и т.п.).

Система признаков двухуровневая, что позволяет формализовать (шкалировать) не только качественные (да/нет), но и количественные (числовые) признаки, а также позволяет обрабатывать вопросы со многими, в том числе и неальтернативными вариантами ответов. Вопрос с вариантами ответов можно рассматривать как шкалу с градациями. Такое понимание позволяет "ввести в оборот" хорошо разработанную теорию шкалирования, что является весьма ценным. В предлагаемой модели нет ограничений на тип и количество шкал, а также на количество градаций в них (за исключением суммарного общего количества градаций. Нет в предлагаемой модели и таких искусственных ограничений, как, например, необходимость одинакового количества градаций во всех шкалах, или необходимость использовать только шкалы только одного какого-либо типа, и т.п., которые, как правило, встречаются в других системах.

В принципе могут быть сконструированы системы признаков, представляемые деревьями трех и более уровней, однако программно реализовывать их нецелесообразно, т.к. они все сводятся к двухуровневым деревьям (вопросы с вариантами ответов).

1.2.3.2. Иерархическая структура данных и последовательность численных расчетов в СК-анализе

Рассмотрим 6 уровней базовых когнитивных операций системного анализа и 5-ти уровневую иерархическую структуру данных (рисунок 44), на базе которой и реализуются эти операции.

На 1-м уровне непосредственно на основе исходной информации, путем применения БКОСА 2.1 и 2.2 формируется матрица абсолютных частот.

На 2-м уровне на основе матрицы абсолютных частот путем применения БКОСА 3.1.1, 3.1.2, 3.1.3, 3.2, 3.3 формируется матрица информативностей, являющаяся основой для выполнения последующих БКОСА и обеспечивающая независимость времени их выполнения от объема обучающей выборки.

На 3-м уровне путем выполнения БКОСА 4.1 и 4.2 формируется оптимизированная матрица информативностей. Оптимизация обеспечивает экономию труда, времени и других затрат на эксплуатацию содержательной информационной модели.

На 4-м уровне с использованием оптимизированной матрицы информативностей выполняются БКОСА 9.1, 9.2, а также 10.1.1 и 10.2.1. Две последние

операции обеспечивают (соответственно) создание матриц сходства классов и атрибутов, являющихся, в свою очередь, основой для реализации последующих БКОСА.

На 5-м уровне на основе матриц сходства путем выполнения БКОСА 10.1.2, 10.2.2, 10.3.1 и 10.4.1 рассчитываются базы данных, когнитивного и кластерно-конструктивного анализа.

На 6-м уровне, с использованием баз данных, созданных на 5-м уровне, реализуются БКОСА 10.1.3, 10.3.2, 10.4.2 и 10.2.3.

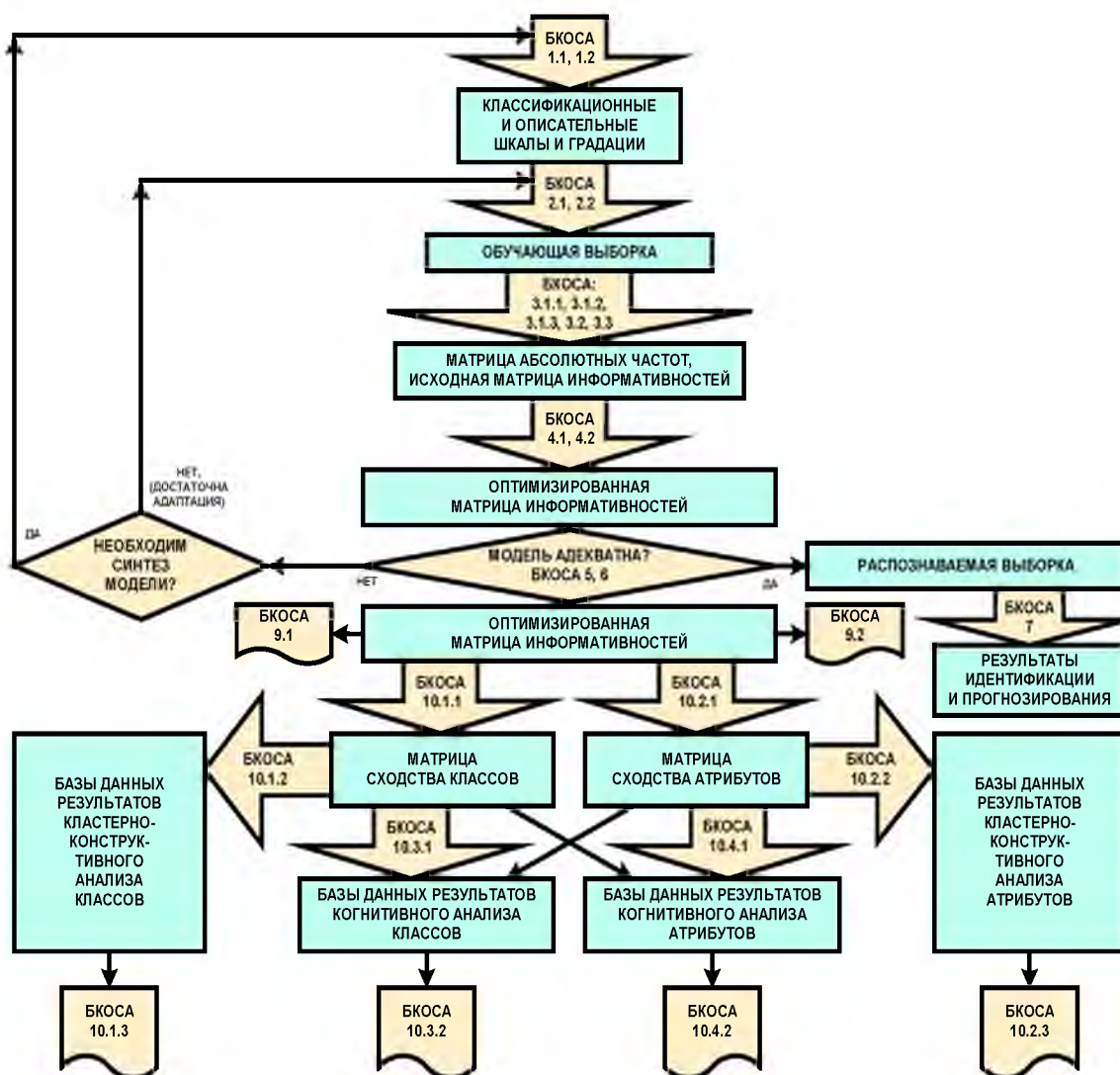


Рисунок 44. Иерархическая структура данных семантической информационной модели СК-анализа

1.2.3.3. Обобщенное описание алгоритмов СК-анализа

В данном разделе приведены 24 детальных алгоритма всех 10 базовых когнитивных операций системного анализа (таблица 28), коды которых полностью соответствуют обобщенной схеме СК-анализа (рисунок 44).

**Таблица 28 – БАЗОВЫЕ КОГНИТИВНЫЕ ОПЕРАЦИИ
СИСТЕМНОГО АНАЛИЗА (БКОСА)**

№ п/п	Код БКОСА по схеме АСК-анализа	Полное наименование БКОСА
1.	1.	Присвоение имен классам и атрибутам (интенциональная и экстенциональная репрезентация)
2.	2.	Восприятие
3.	3.	Обобщение (синтез, индукция)
4.	4.	Абстрагирование классов и атрибутов
5.	5.	Оценка адекватности модели
6.	7.	Сравнение, идентификация и прогнозирование
7.	9.	Анализ (дедукция и абдукция) классов и атрибутов
8.	10.1, 10.2.	Классификация и генерация конструкторов классов и атрибутов
9.	10.3, 10.4.	Содержательное сравнение классов и атрибутов
10.	11.	Планирование и принятие решений о применении системы управляющих факторов

В таблице 29 приведена структура *каждой* базовой когнитивной операции, дана их нумерация в соответствии с обобщенной схемой СК-анализа и нумерация реализующих их алгоритмов.

Таблица 29 – ДЕТАЛЬНЫЙ СПИСОК БКОСА И ИХ АЛГОРИТМОВ

№ алгоритма	Код БКОСА по схеме АСК-анализа	№ БКОСА	Наименование БКОСА	Полное наименование базовых когнитивных операций системного анализа (БКОСА)
	1.1	1	Присвоение имен	Присвоение имен классам (интенциональная, интегральная репрезентация)
	1.2			Присвоение имен атрибутам (экстенциональная, дискретная репрезентация)
1	2.1.	2	Восприятие	Восприятие и запоминание исходной обучающей информации
2	2.2.			Репрезентация. Сопоставление индивидуального опыта с коллективным (общественным)
3	3.1.1.	3	Обобщение (синтез, индукция)	Накопление первичных данных
4	3.1.2.			Исключение артефактов
5	3.1.3.			Расчет истинности смысловых связей между предпосылками и результатами (обобщенных таблиц решений)
6	3.2.			Определение значимости шкал и градаций факторов, уровней Мерлина
7	3.3.			Определение значимости шкал и градаций классов, уровней Мерлина
8	4.1.	4	Абстрагирование	Абстрагирование факторов (снижение размерности семантического пространства факторов)
9	4.2.			Абстрагирование классов (снижение размерности семантического пространства классов)
10	5.	5	Оценка адекватности	Оценка адекватности информационной модели предметной области
11	7.	6	Сравнение, идентификация и прогнозирование	Сравнение, идентификация и прогнозирование. Распознавание состояний конкретных объектов (объектный анализ)
12	9.1.	7	Анализ, дедукция и абдукция	Анализ, дедукция и абдукция классов (семантический анализ обобщенных образов классов, решение обратной задачи прогнозирования)
13	9.2.			Анализ, дедукция и абдукция факторов (семантический анализ факторов)
14	10.1.1.	8	Классификация и генерация конструкторов	Классификация обобщенных образов классов
15	10.1.2.			Формирование бинарных конструкторов классов
16	10.1.3.			Визуализация семантических сетей классов
17	10.2.1.			Классификация факторов
18	10.2.2.			Формирование бинарных конструкторов факторов
19	10.2.3.			Визуализация семантических сетей факторов
20	10.3.1.	9	Содержательное сравнение	Содержательное сравнение классов
21	10.3.2.			Расчет и отображение многозначных когнитивных диаграмм, в т.ч. диаграмм Мерлина
22	10.4.1.			Содержательное сравнение факторов
23	10.4.2.			Расчет и отображение многозначных когнитивных диаграмм, в т.ч. инвертированных диаграмм Мерлина
24	11.	10	Планирование и управление	Многовариантное планирование и принятие решения о применении системы управляющих факторов

Описания базовых когнитивных операций системного анализа и их реальные детализированные алгоритмы приведены ниже (рисунки 45 – 68).

БКОСА-2.1. "Восприятие и запоминание исходной обучающей информации"

В базы данных вводятся двухвекторные (дискретно-интегральные) описания объектов, включающие как их описание на языке признаков, так и принадлежность к определенным классификационным категориям – классам.

БКОСА-2.2. "Репрезентация. Сопоставление индивидуального опыта с коллективным (общественным)"

В ряде случаев, особенно при проведении политологических исследований, необходимо, чтобы исследуемая выборка корректно представляла генеральную совокупность не только в смысле традиционно понимаемой репрезентативности, но и по распределению респондентов по категориям (т.е. структурно) соответствовала ей. Добиться этого путем подбора объектов для исследования затруднительно, т.к. каждый объект может относиться одновременно ко многим классификационным категориям. Данный алгоритм обеспечивает выборку из исследуемого множества объектов последовательных подмножеств, наиболее близких по частотному распределению объектов по категориям к заданному распределению. Данная операция называется также "взвешивание или ремонт данных".

БКОСА-3.1.1. "Обобщение (синтез, индукция). Накопление первичных данных"

На основе анализа обучающей выборки обеспечивается накопление в базах данных первичных элементов смысла, т.е. фактов, состоящих в том, что определенный признак встретился у объекта определенного класса.

БКОСА-3.1.2. "Обобщение (синтез, индукция). Исключение артефактов"

При отсутствии статистики невозможно отличить закономерные факты от не вписывающихся в общую складывающуюся картину и искажающих ее, т.е. артефактов. При накоплении же достаточной статистики это возможно и данный алгоритм позволяет выявить и исключить из дальнейшего анализа артефакты. Необходимо отметить, что в результате действия данного алгоритма существенно повышается качество содержательной модели предметной области, в частности ее валидность.

БКОСА-3.1.3. "Обобщение (синтез, индукция). Расчет степени истинности содержательных смысловых связей между предпосылками и результатами (обобщенных таблиц решений)"

Непосредственно на основе матрицы абсолютных частот позволяет вычислить количество информации, содержащейся в факте наблюдения у некоторого объекта определенного признака о том, что данный объект принадлежит к определенной классификационной категории.

БКОСА-3.2. "Определение значимости шкал и градаций факторов, уровней Мерлина"

Рассчитывается среднее количество информации, которое система управления получает о поведении АОУ из фактов о действии тех или иных факторов и их значений. Кроме того, если факторы классифицированы независимым способом по уровням Мерлина, то определяется и значимость этих уровней.

БКОСА-3.3. "Определение значимости шкал и градаций классов, уровней Мерлина"

Рассчитывается среднее количество информации, которое система управления получает из одного признака, если известен класс. Если классы относятся к уровням Мерлина, то определяется и их значимость.

БКОСА-4.1. "Абстрагирование факторов (снижение размерности семантического пространства факторов)"

С помощью метода последовательных приближений (итерационный алгоритм) при заданных граничных условиях снижается размерность пространства атрибутов без существенного уменьшения его объема и адекватности модели. Критерий остановки итерационного процесса – достижение одного из граничных условий.

БКОСА-4.2. "Абстрагирование классов (снижение размерности семантического пространства классов)"

С помощью метода последовательных приближений (итерационный алгоритм) при заданных граничных условиях снижается размерность пространства классов без существенного уменьшения его и адекватности объема. Критерий остановки итерационного процесса – достижение одного из граничных условий.

БКОСА-5. "Оценка адекватности информационной модели предметной области"

Осуществляется идентификация объектов обучающей выборки (классификационный вектор которых уже известен) и затем рассчитывается средневзвешенная погрешность идентификации (интегральная валидность), а также погрешность идентификации с каждым классом (дифференциальная валидность). Если модель имеет приемлемый уровень адекватности, то принимается решение о возможности ее использования в адаптивном режиме на объектах, не входящих в обучающую выборку, но относящихся к генеральной совокупности, по отношению к которой эта выборка репрезентативна. Если же модель недостаточно адекватна, то продолжают работы по синтезу адекватной модели путем увеличения количества классов и факторов, а также корректировки описаний объектов обучающей выборки и увеличения их количества.

БКОСА-7. "Сравнение, идентификация и прогнозирование. Распознавание состояний конкретных объектов (объектный анализ)"

Рассчитывается количество информации, содержащееся в описании идентифицируемого объекта о его принадлежности к каждому из классов. Все классы ранжируются в порядке убывания количества информации о принадлежности к

ним в описании данного объекта. Таким образом, вектор объекта разлагается в ряд по векторам классов. Кроме того, все объекты ранжируются в порядке убывания сходства с каждым классом. Таким образом, вектор класса разлагается в ряд по векторам объектов.

БКОСА-9.1. "Дедукция и абдукция классов (семантический анализ обобщенных образов классов, решение обратной задачи прогнозирования)"

Координаты вектора класса (т.е. факторы) ранжируются в порядке убывания их значений. Таким образом, в начале списка оказываются факторы, оказывающие наиболее сильное влияние на переход АОУ в состояние, соответствующее данному классу, а в конце списка – препятствующие этому. Это позволяет выбрать факторы для управляющего воздействия, целью которого является перевод АОУ в состояние, соответствующее данному классу. Механизм фильтрации позволяет "изолированно" рассматривать влияние различных групп факторов: например, факторов, характеризующих объект управления, управляющую систему или окружающую среду. Абдукция представляет собой обобщение дедукции на основе нечеткой логики. В данном случае это означает, что фактор связан с классом не детерминистским образом, а через количество информации, которое в нем содержится о данном классе.

БКОСА-9.2. "Дедукция и абдукция факторов (семантический анализ факторов)"

Классы ранжируются в порядке убывания влияния данного фактора на переход АОУ в состояния, соответствующие этим классам. В начале списка оказываются состояния, на переход в которые данный фактор оказывает наибольшее влияние, а в конце – на переход в которые данный фактор препятствует. Этот список является развернутой характеристикой смысла фактора.

БКОСА-10.1.1. "Классификация обобщенных образов классов"

Сравниваются вектора классов и формируется диагональная матрица сходства классов, в которой по обоим осям расположены коды классов а в клетках находятся нормированные коэффициенты, численно отражающие степень сходства или различия векторов соответствующих классов.

БКОСА-10.1.2. "Формирование бинарных конструкторов классов"

На основе матрицы сходства классов для каждого из них формируется ранжированный список остальных, в котором они расположены в порядке убывания сходства с данным классом. Такие списки представляют собой бинарные конструкторы, а их полюса соответствуют кластерам.

БКОСА-10.1.3. "Визуализация семантических сетей классов"

На основе матрицы сходства классов визуализируются ориентированные графы, вершинам которых соответствуют классы, а ребрам – степени их сходства или различия. Знак связи обозначается цветом: красный цвет – сходство,

синий – различие, толщина линии соответствует модулю (силе) связи. Необходимо отметить, что для подобных графов в литературе пока нет устоявшегося общепринятого названия: в данном исследовании, как и в предшествующих работах автора, они называются семантическими сетями, в литературе по когнитивному анализу их называют когнитивными картами, а в литературе по когнитивному анализу – когнитивными картами или схемами [160 – 164].

БКОСА-10.2.1. "Классификация факторов"

Сравниваются вектора факторов и формируется диагональная матрица сходства факторов, в которой по обоим осям расположены коды факторов, а в клетках находятся нормированные коэффициенты, численно отражающие степень сходства или различия векторов соответствующих факторов.

БКОСА-10.2.2. "Формирование бинарных конструкторов факторов"

На основе матрицы сходства факторов для каждого из них формируется ранжированный список остальных, в котором они расположены в порядке убывания сходства с данным фактором. Такие списки представляют собой бинарные конструкторы, а их полюса соответствуют кластерам.

БКОСА-10.2.3. "Визуализация семантических сетей факторов"

На основе матрицы сходства факторов визуализируются ориентированные графы, вершинам которых соответствуют заданные факторы, а ребрам – степени их сходства или различия. Знак связи обозначается цветом: красный цвет – сходство, синий – различие, толщина линии соответствует модулю (силе) связи.

БКОСА-10.3.1. "Содержательное сравнение классов"

Каждая связь между классами в семантической сети, отражающая степень их сходства или различия, имеет определенную структуру, описанную в разделе 3.2.3 исследования. Эта структура включает ряд элементов, каждый из которых соответствует одному слагаемому обобщенной меры сходства векторов классов.

БКОСА-10.3.2. "Расчет и отображение многозначных когнитивных диаграмм, в т.ч. диаграмм Вольфа Мерлина"

Из всех составляющих связи между классами выбираются 8 наиболее сильных и отображаются в форме линий, цвет которых означает знак, а толщина – модуль силы связи. Классы изображаются в форме наиболее значимых фрагментов их информационных портретов. При этом учитываются корреляции между факторами.

БКОСА-10.4.1. "Содержательное сравнение факторов"

Каждая связь между факторами в семантической сети, отражающая степень их сходства или различия, имеет определенную структуру, описанную в разделе 3.2.3 исследования. Эта структура включает ряд элементов, каждый из

которых соответствует одному слагаемому обобщенной меры сходства векторов факторов.

БКОСА-10.4.2. "Расчет и отображение многозначных когнитивных диаграмм, в т.ч. инвертированных диаграмм Мерлина"

Из всех составляющих связи между факторами выбираются 16 наиболее сильных и отображаются в форме линий, цвет которых означает знак, а толщина – модуль силы связи. Факторы отображаются в форме наиболее значимых фрагментов их семантических портретов. При этом учитываются корреляции между классами.

БКОСА-11. "Многовариантное планирование и принятие решения о применении системы управляющих факторов"

Выполняется в несколько этапов:

- 1) выполняется прогноз развития АОУ в условиях отсутствия управляющих воздействий, т.е. реализуется БКОСА-7 ("движение по инерции");
- 2) если в соответствии с прогнозом по п.1 АОУ достигает заданного целевого состояния (т.е. прогноз "удовлетворительный"), то планирование прекращается (переход на п.6); иначе – выполняется п.3;
- 3) путем решения обратной задачи прогнозирования (БКОСА-9.1) определяется набор факторов, оптимальный для перевода АОУ в заданное целевое состояние;
- 4) если все эти факторы есть возможность использовать для управления, то на этом планирование прекращается (переход п.6); иначе переход на п.5;
- 5) используя результаты кластерно-конструктивного анализа факторов (БКОСА 10.2.1, 10.2.2, 10.2.3) последовательно ищется замена для факторов, которые нет возможности использовать и после каждой замены выполняется прогнозирование (БКОСА-7); если результаты прогнозирования удовлетворительные – окончание планирования (переход на п.6); иначе принятие решения о невозможности выработки корректного управляющего воздействия;
- 6) окончание планирования.

Информационный портрет представляет собой детализацию вершин семантической сети. Когнитивные диаграммы детально раскрывают структуру связи между двумя вершинами семантической сети, представленными в форме информационных портретов. Поэтому для расшифровки структуры вершин семантической сети и связей между ними, предлагается ввести новое понятие "Семантическая когнитивная сеть", которая представляет собой систему когнитивных диаграмм, объединенных в макроструктуру, соответствующую структуре семантической сети.

1.2.3.4. Детальные алгоритмы СК-анализа

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 2.1:
"ВОСПРИЯТИЕ И ЗАПОМИНАНИЕ ОБУЧАЮЩЕЙ ИНФОРМАЦИИ"

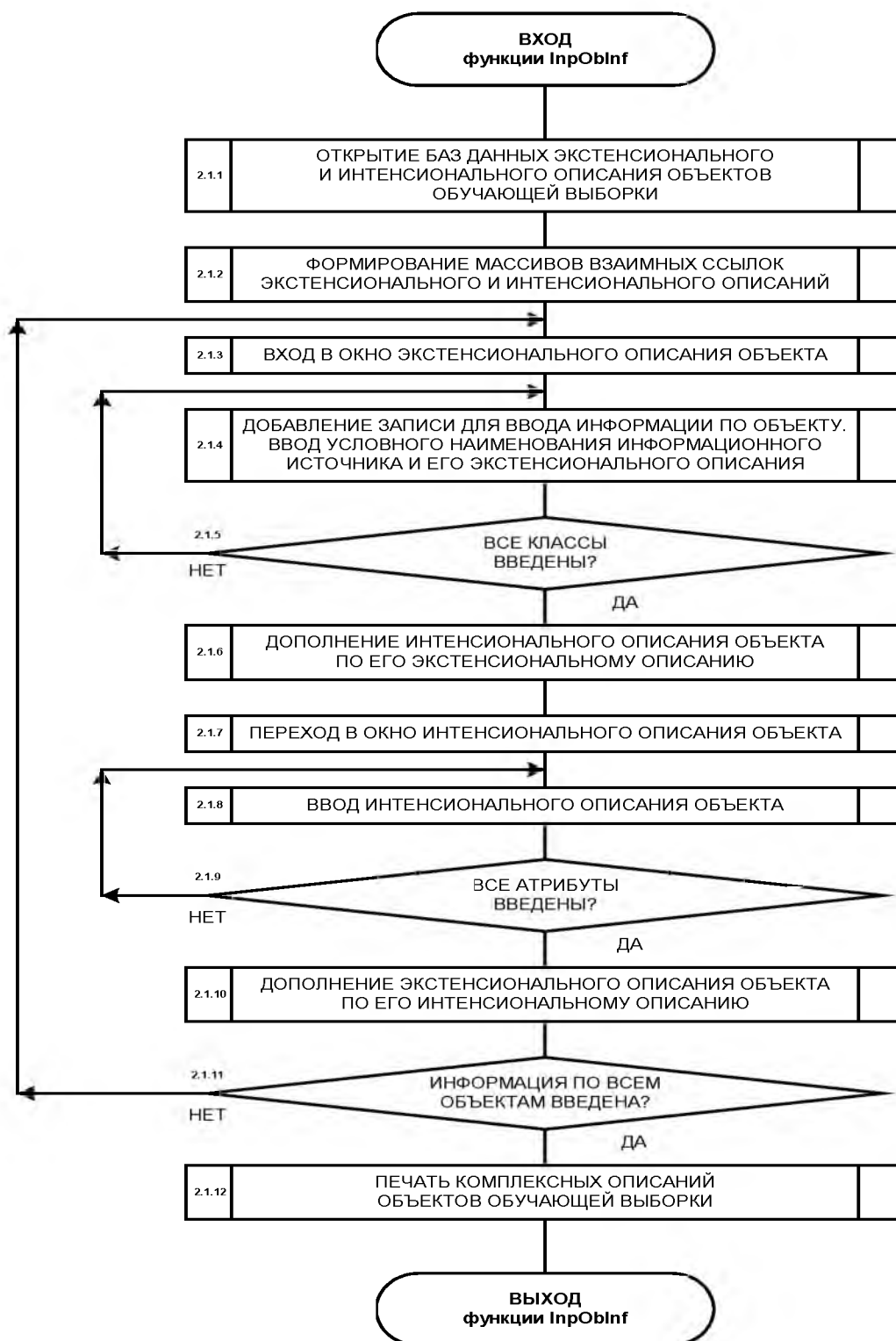
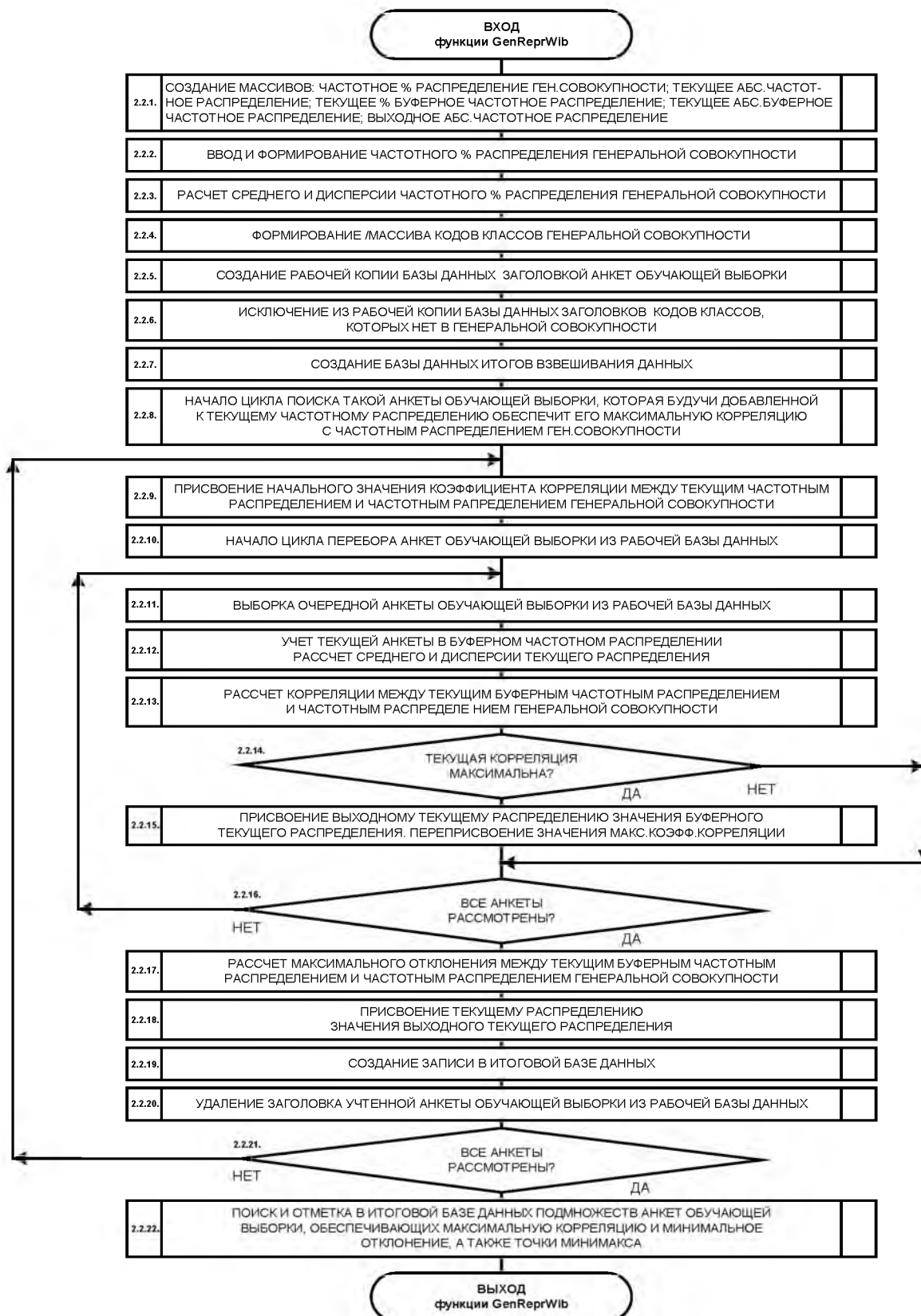


Рисунок 45. Алгоритм БКОСА-2.1. "Восприятие и запоминание исходной обучающей информации"

**АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 2.2:
"РЕПРЕЗЕНТАЦИЯ, СОПОСТАВЛЕНИЕ ИНДИВИДУАЛЬНОГО ОПЫТА С КОЛЛЕКТИВНЫМ"**



**Рисунок 46. Алгоритм БКОСА-2.2. "Репрезентация.
Сопоставление индивидуального опыта с коллективным"**

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 3.1.1:
 "ОБОБЩЕНИЕ (СИНТЕЗ, ИНДУКЦИЯ) -
 - РАСЧЕТ МАТРИЦЫ АБСОЛЮТНЫХ ЧАСТОТ"

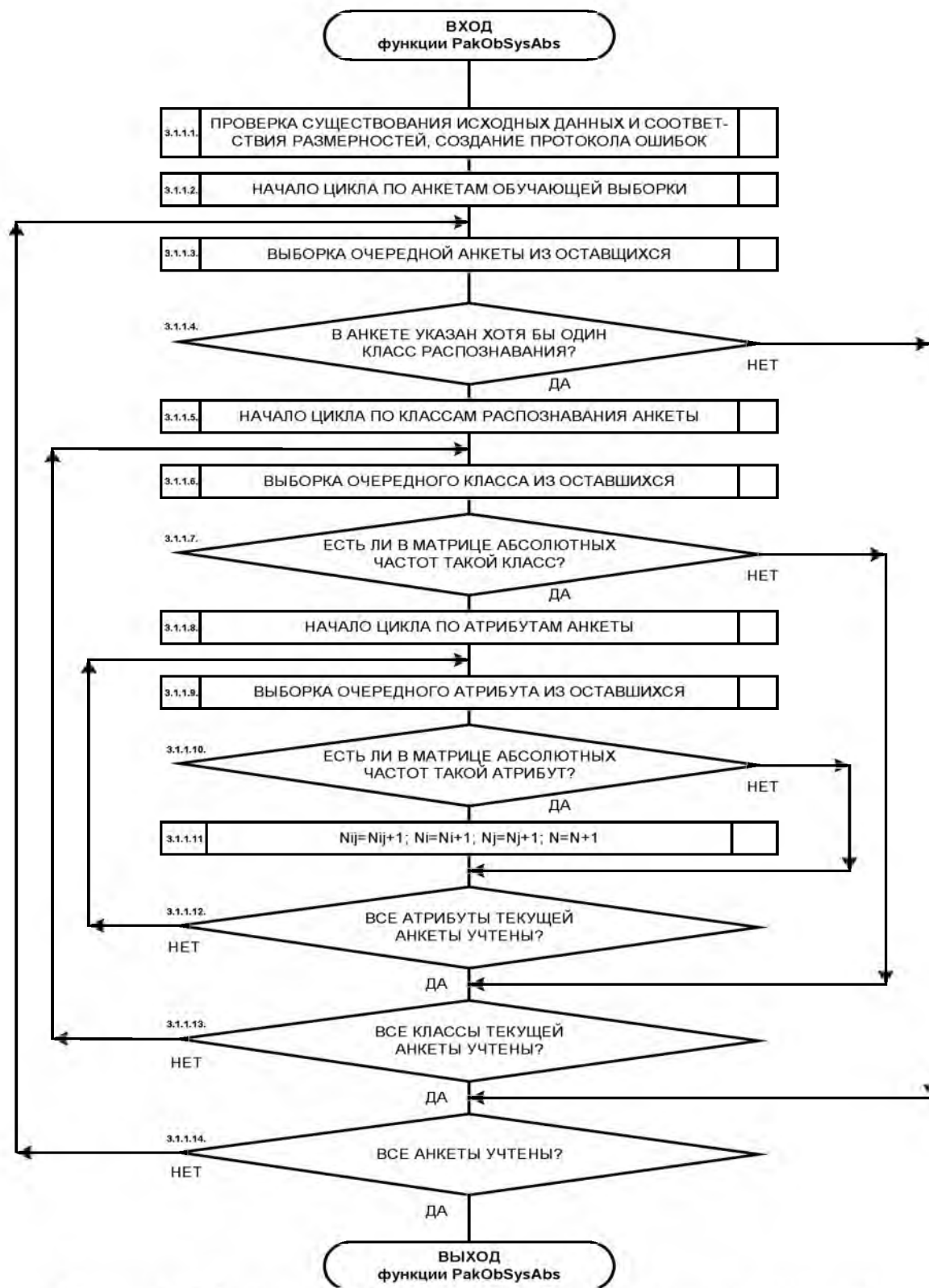


Рисунок 47. Алгоритм БКОСА-3.1.1. "Обобщение (синтез, индукция).
 Накопление первичных данных"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 3.1.2:
"ИСКЛЮЧЕНИЕ АРТЕФАКТОВ"

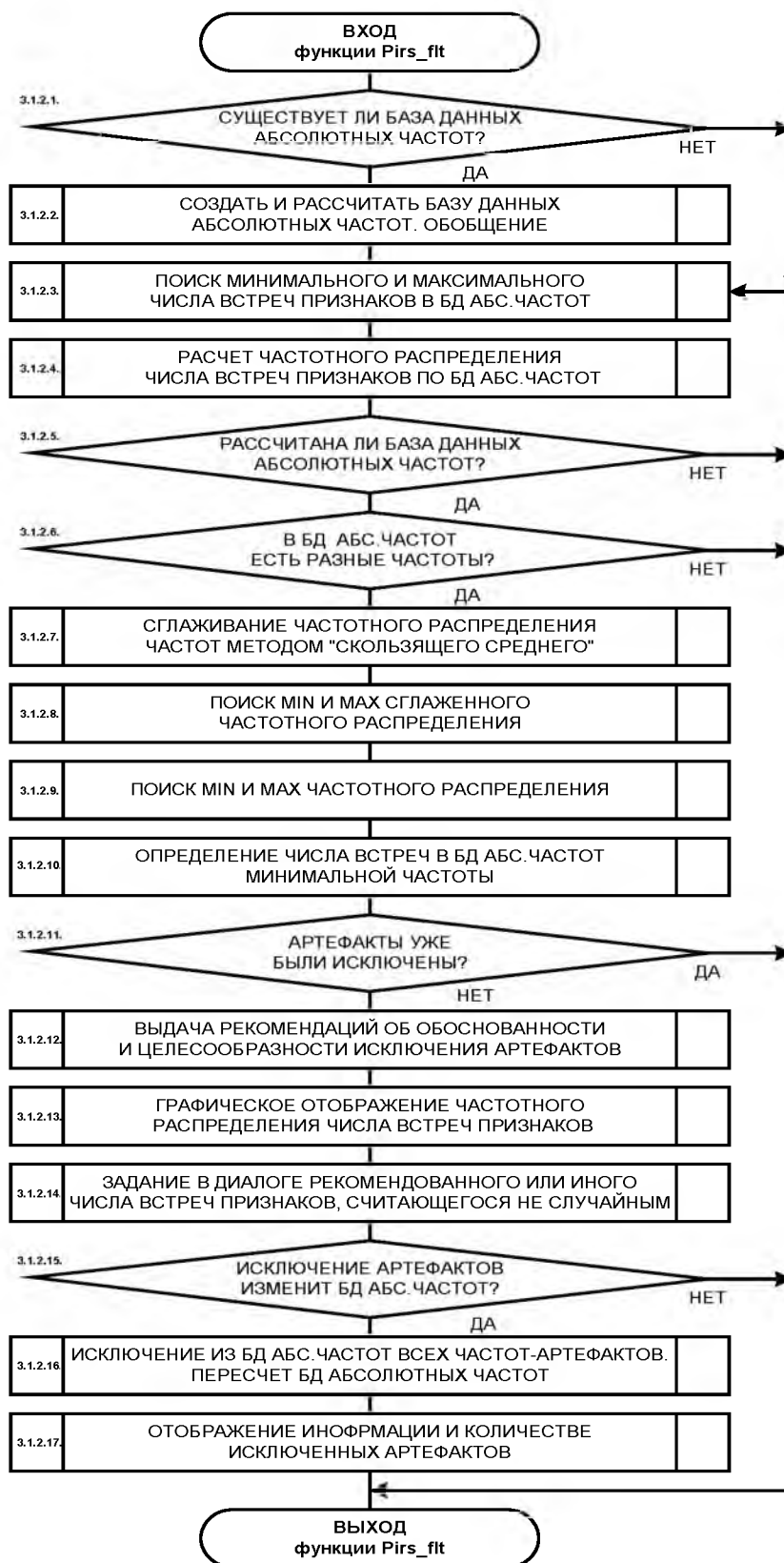


Рисунок 48. Алгоритм БКОСА-3.1.2. "Обобщение (синтез, индукция).
Исключение артефактов"

**АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 3.1.3:
"ОБОБЩЕНИЕ (СИНТЕЗ, ИНДУКЦИЯ) -
- РАСЧЕТ МАТРИЦЫ ИНФОРМАТИВНОСТЕЙ"**

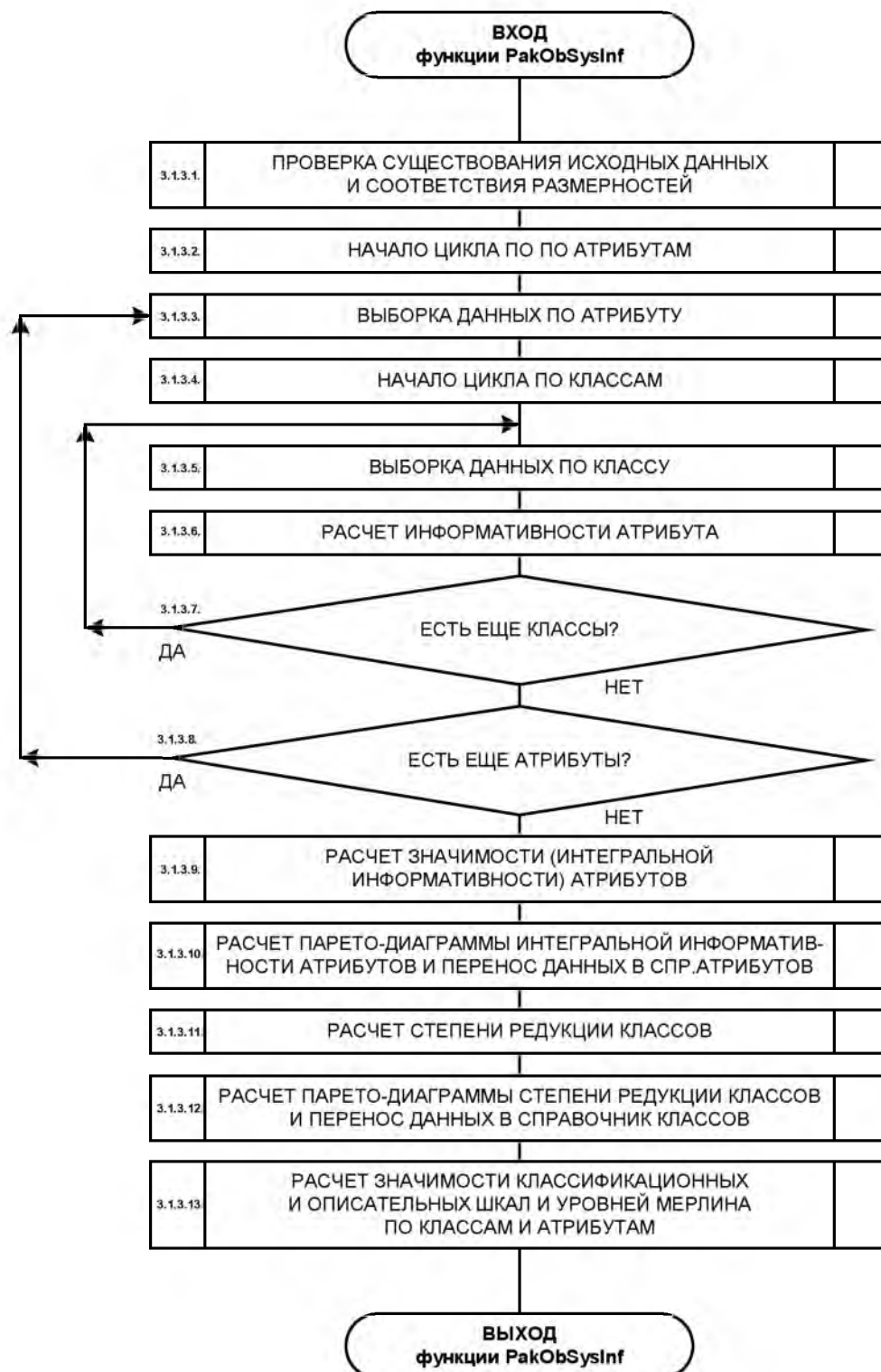


Рисунок 49. Алгоритм БКОСА-3.1.3. "Обобщение (синтез, индукция). Расчет степени истинности содержательных смысловых связей между предпосылками и результатами (обобщенных таблиц решений)"

**АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 3.2:
"ОПРЕДЕЛЕНИЕ ЗНАЧИМОСТИ ШКАЛ И ГРАДАЦИЙ ФАКТОРОВ,
А ТАКЖЕ УРОВНЕЙ МЕРЛИНА"**

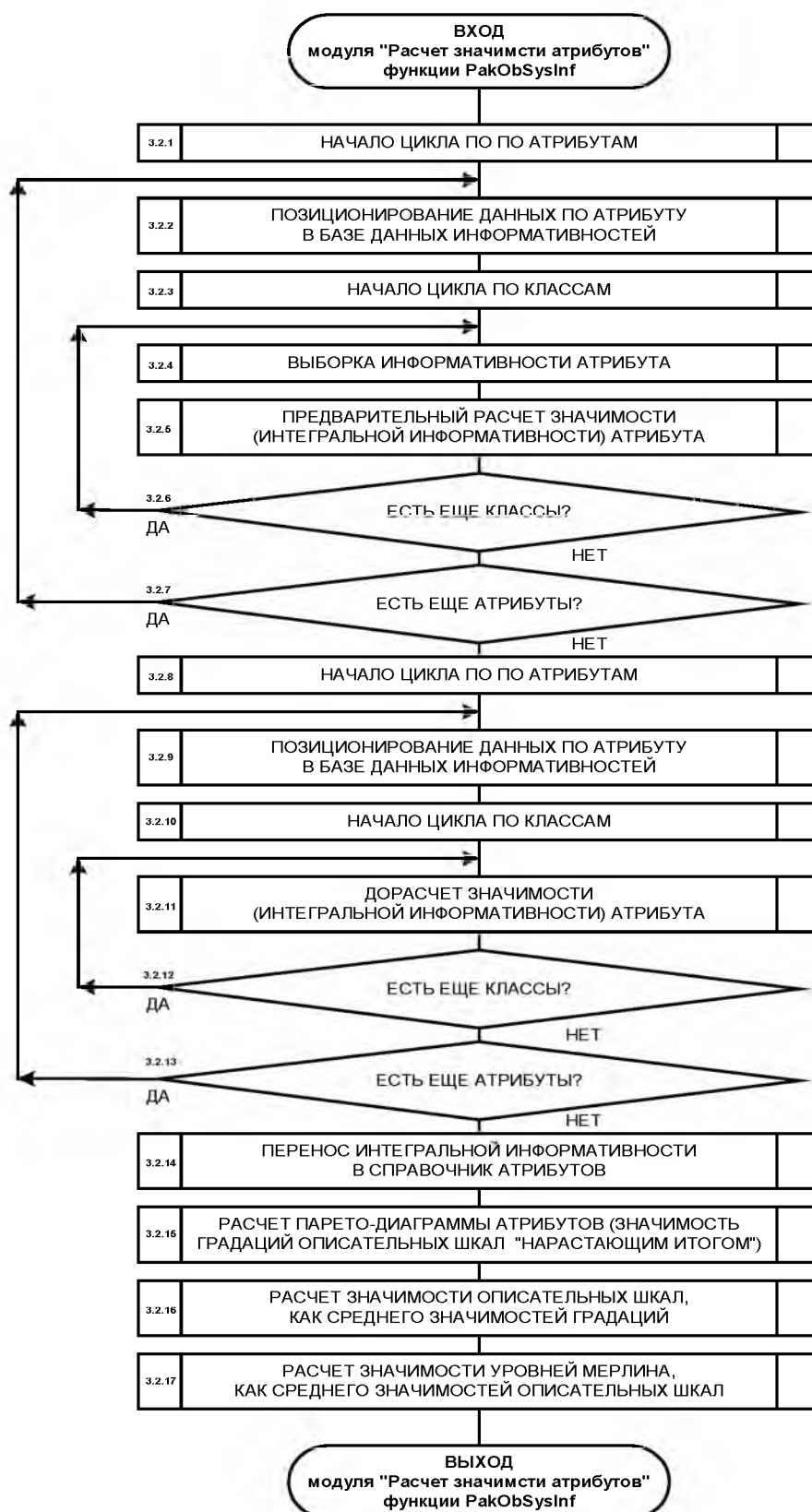


Рисунок 50. Алгоритм БКОСА-3.2. "Определение значимости шкал и градаций факторов, уровней Мерлина"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 3.3:
"РЕДУКЦИЯ КЛАССОВ"

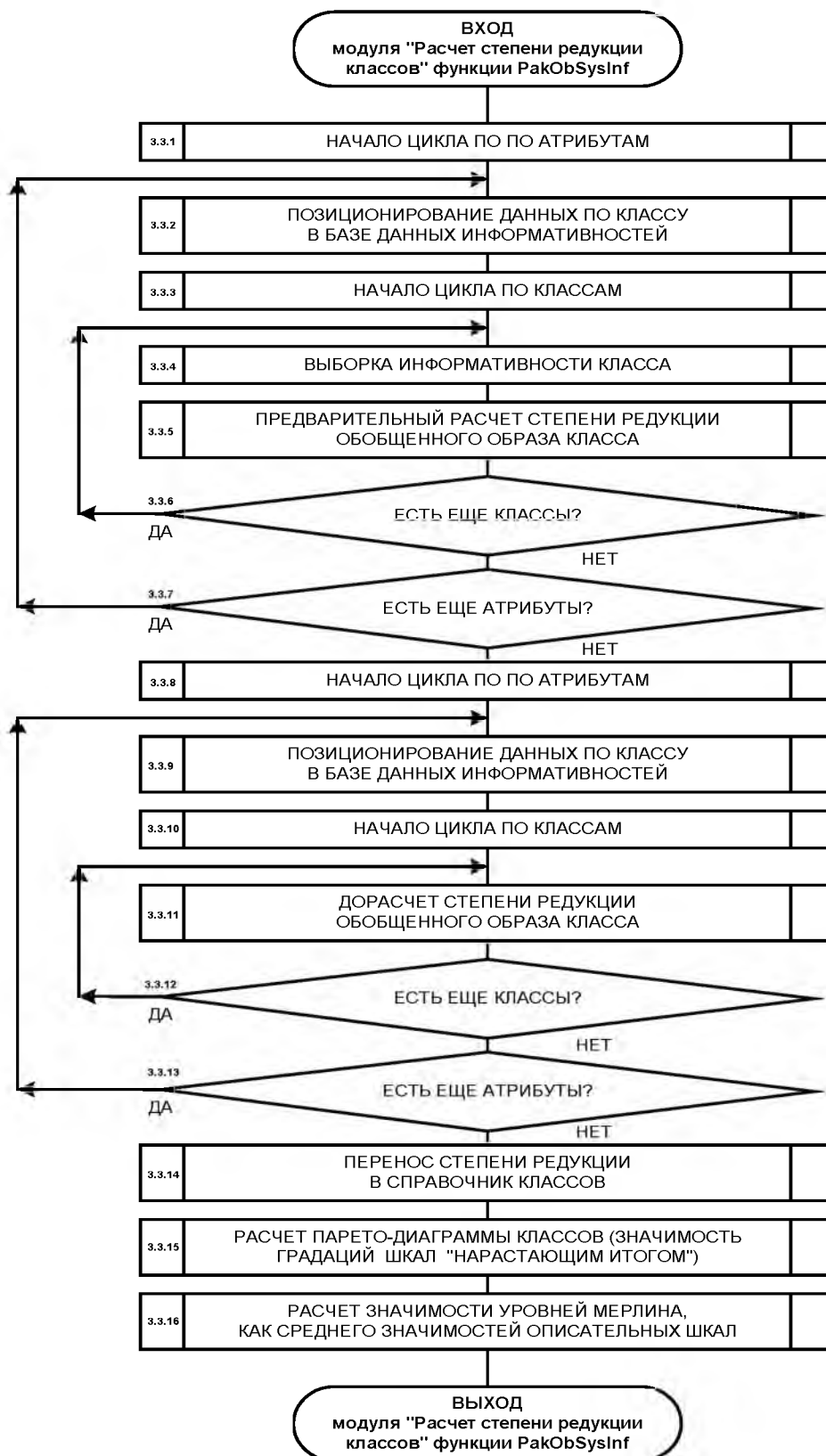


Рисунок 51. Алгоритм БКОСА-3.3. "Определение значимости шкал и градаций классов, уровней Мерлина"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 4.1: "АБСТРАГИРОВАНИЕ АТРИБУТОВ"

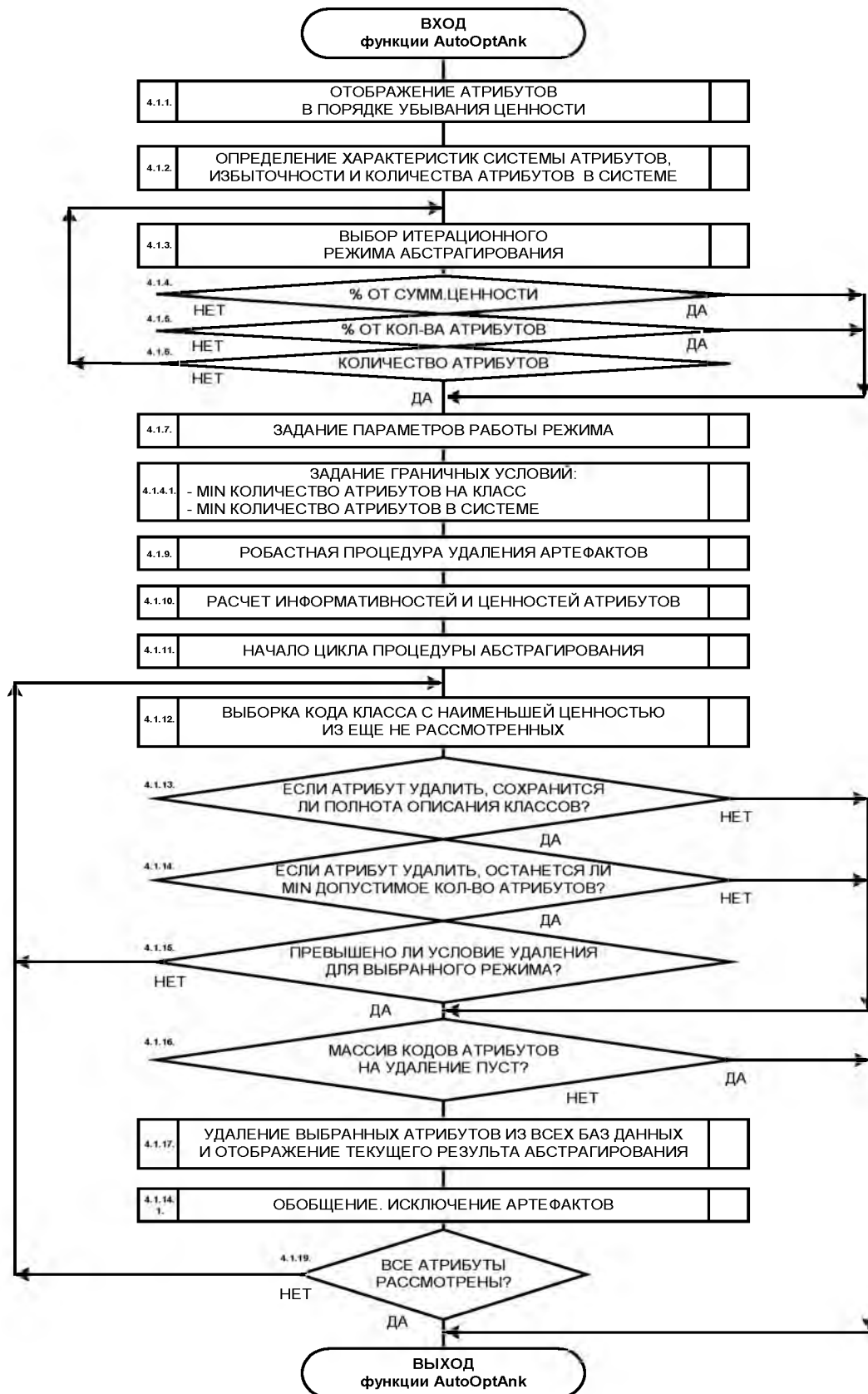


Рисунок 52. Алгоритм БКОСА-4.1. "Абстрагирование факторов (снижение размерности семантического пространства факторов)"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 4.2: "АБСТРАГИРОВАНИЕ КЛАССОВ"

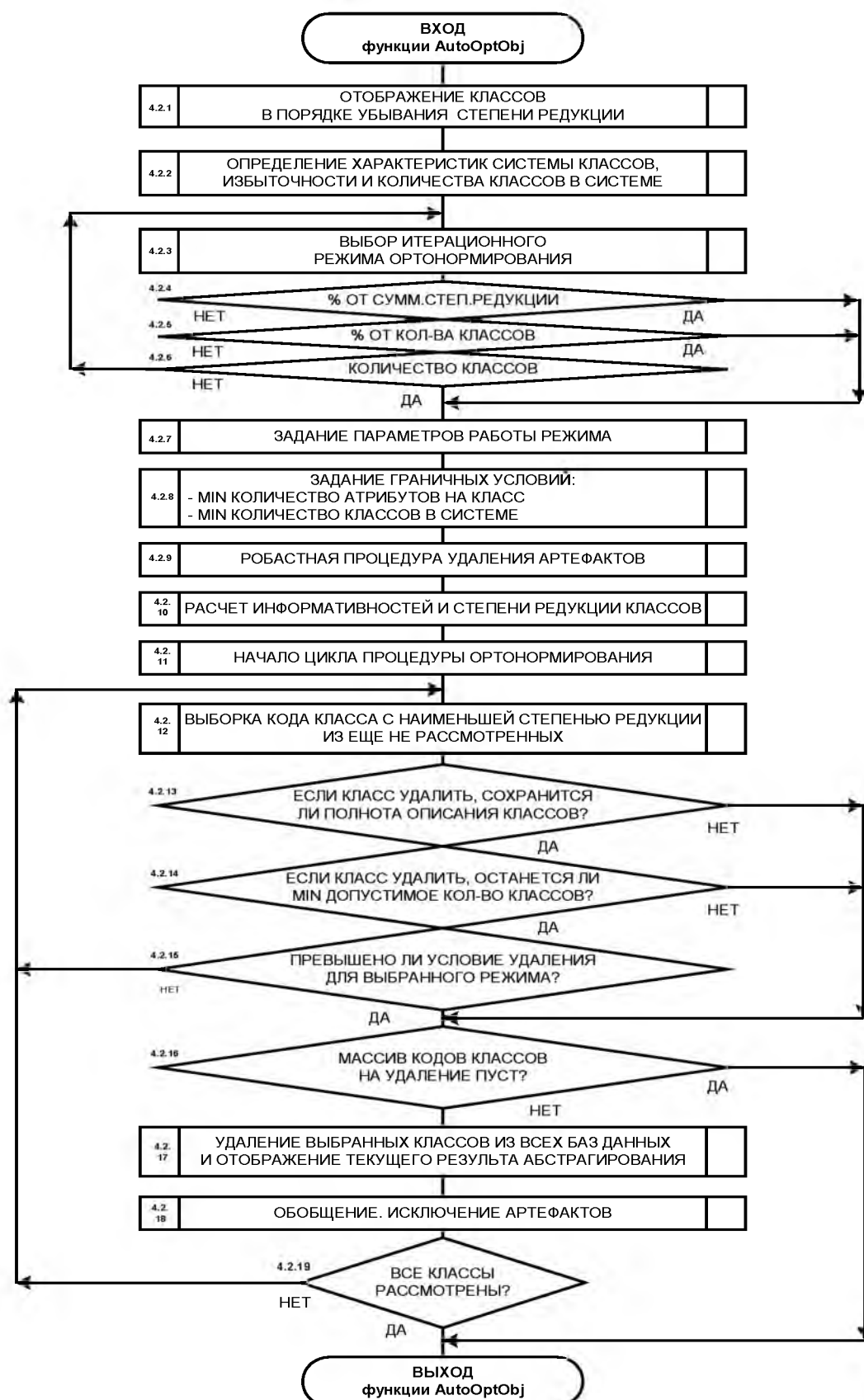


Рисунок 53. Алгоритм БКОСА-4.2. "Абстрагирование классов (снижение размерности семантического пространства классов)"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 5:
"ОЦЕНКА АДЕКВАТНОСТИ МОДЕЛИ (ВАЛИДНОСТЬ)"

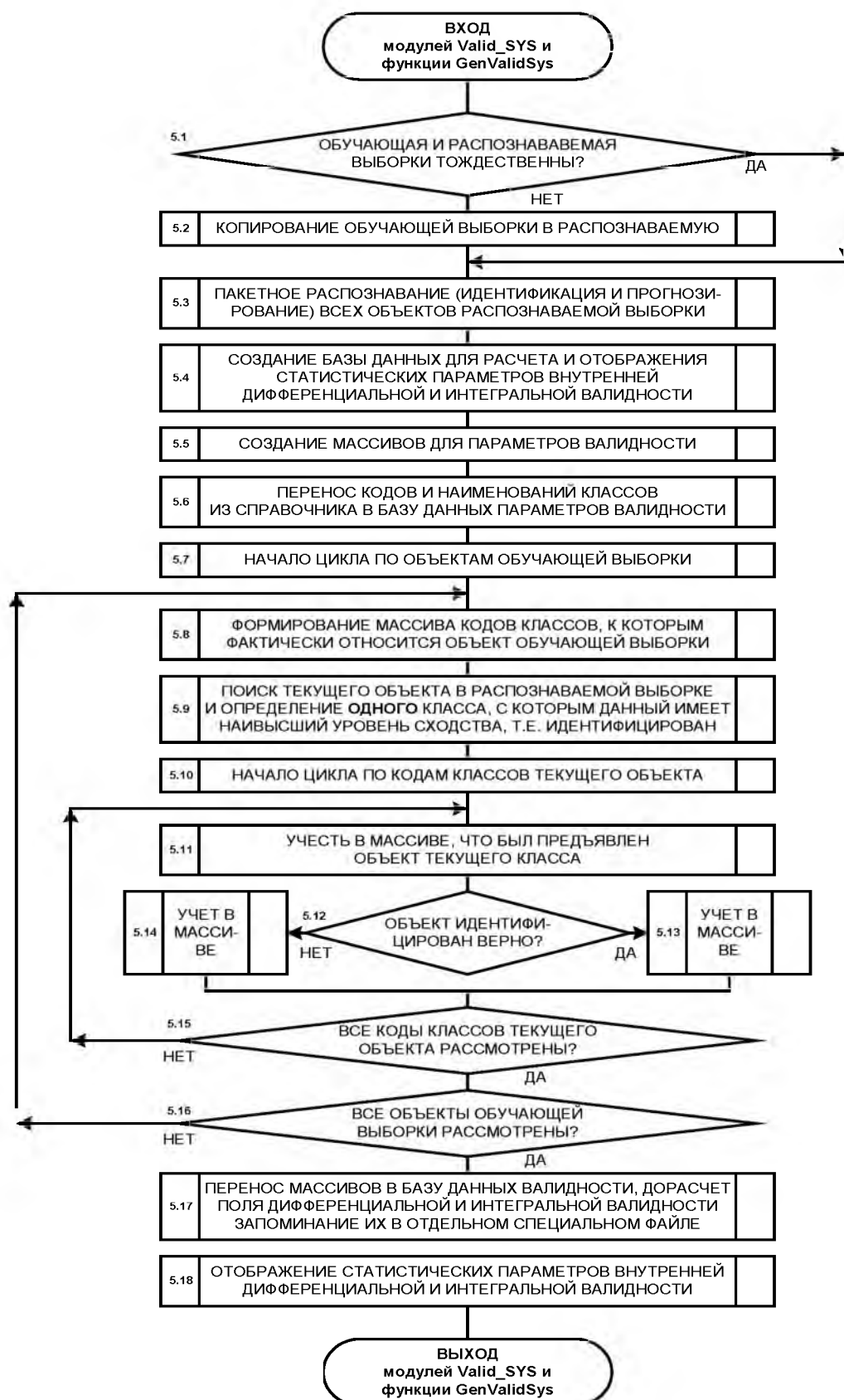


Рисунок 54. Алгоритм БКОСА-5. "Оценка адекватности семантической инфор-
мационной модели предметной области"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 7:
"ИДЕНТИФИКАЦИЯ И ПРОГНОЗИРОВАНИЕ"

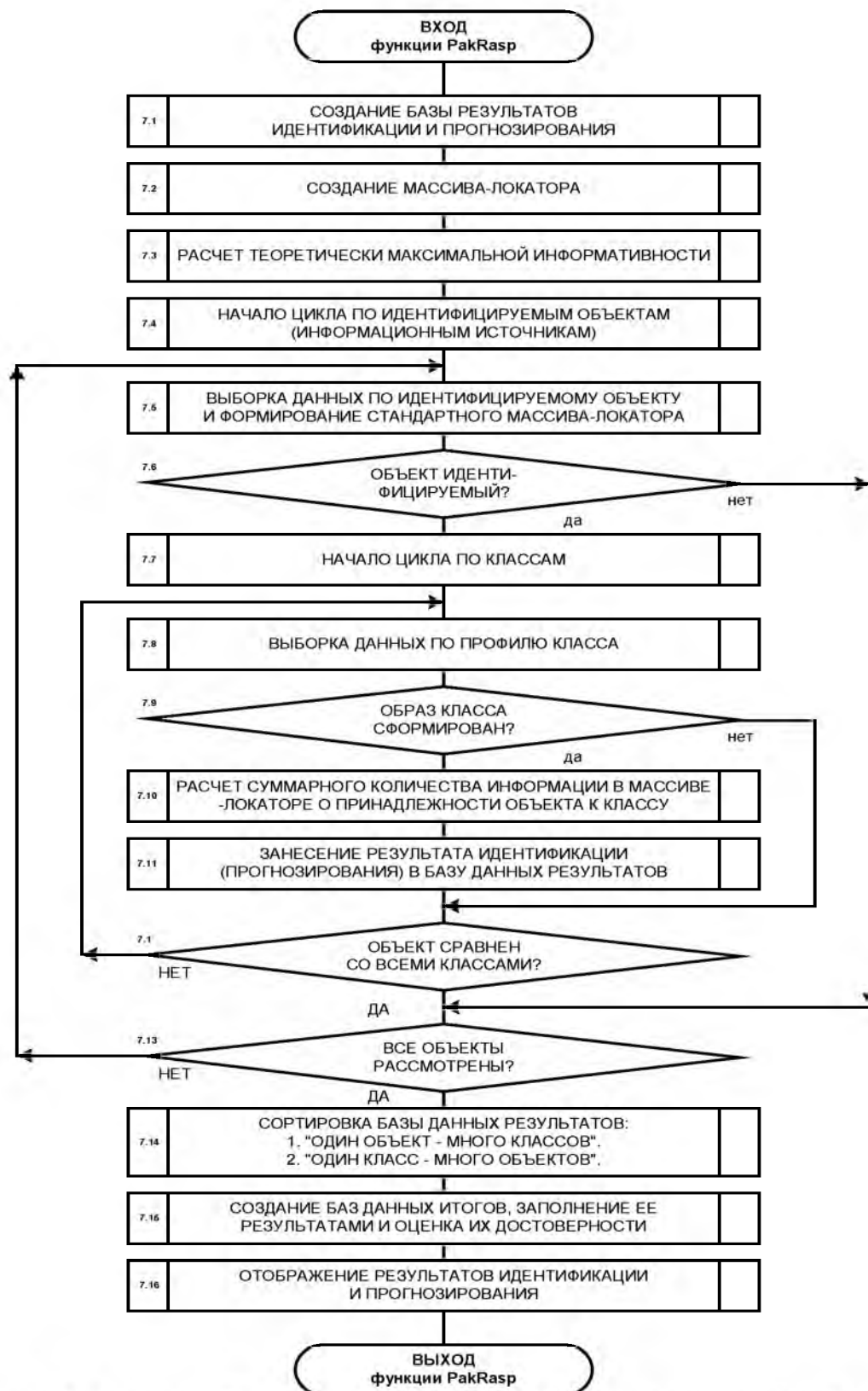


Рисунок 55. Алгоритм БКОСА-7. "Идентификация и прогнозирование. Распознавание состояний конкретных объектов (объектный анализ)"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 9.1:
"ДЕДУКЦИЯ КЛАССОВ"

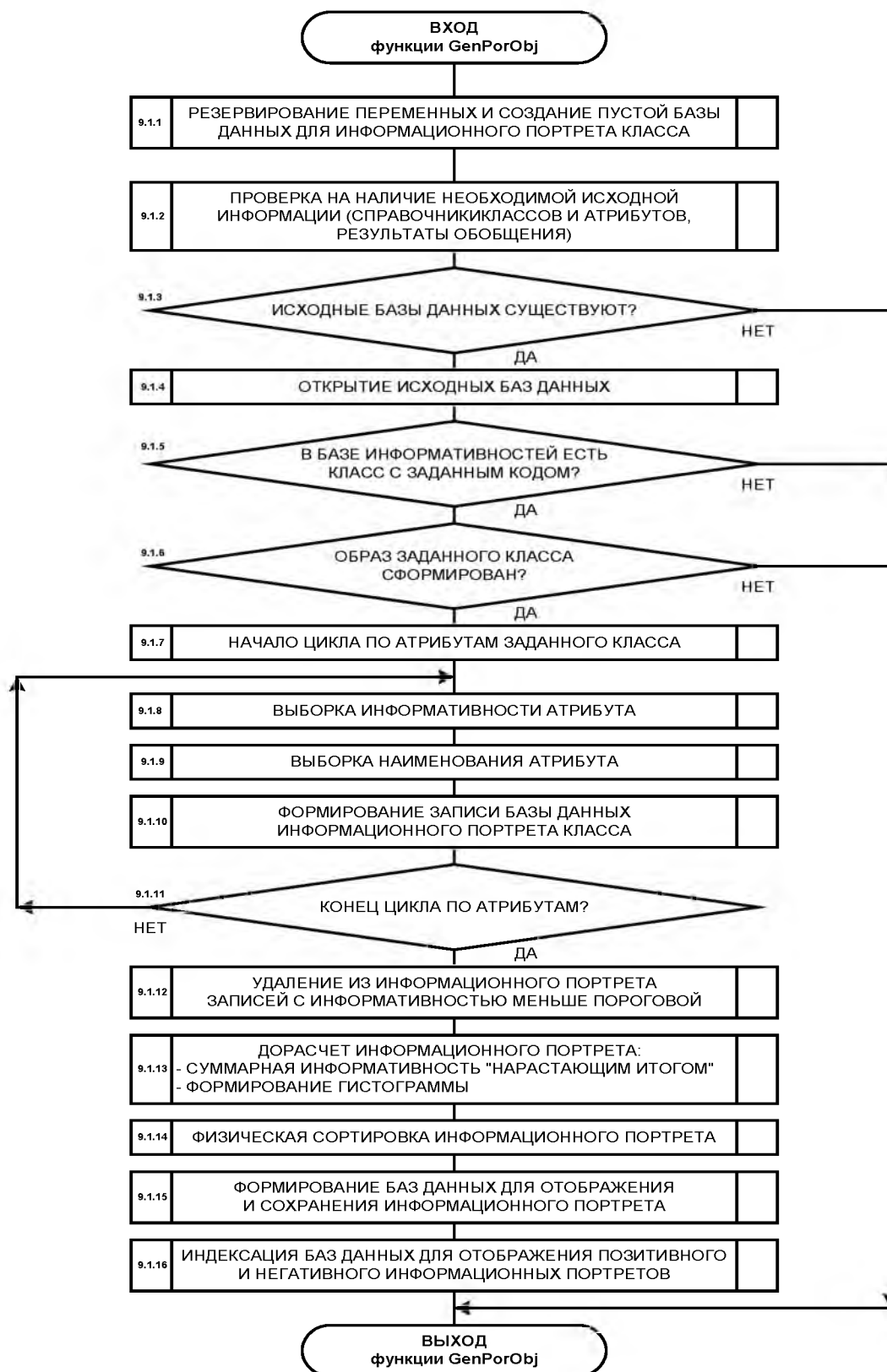


Рисунок 56. Алгоритм БКОСА-9.1. "Дедукция и абдукция классов (семантический анализ обобщенных образов классов, решение обратной задачи прогнозирования)"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 9.2:
"ДЕДУКЦИЯ АТРИБУТОВ"

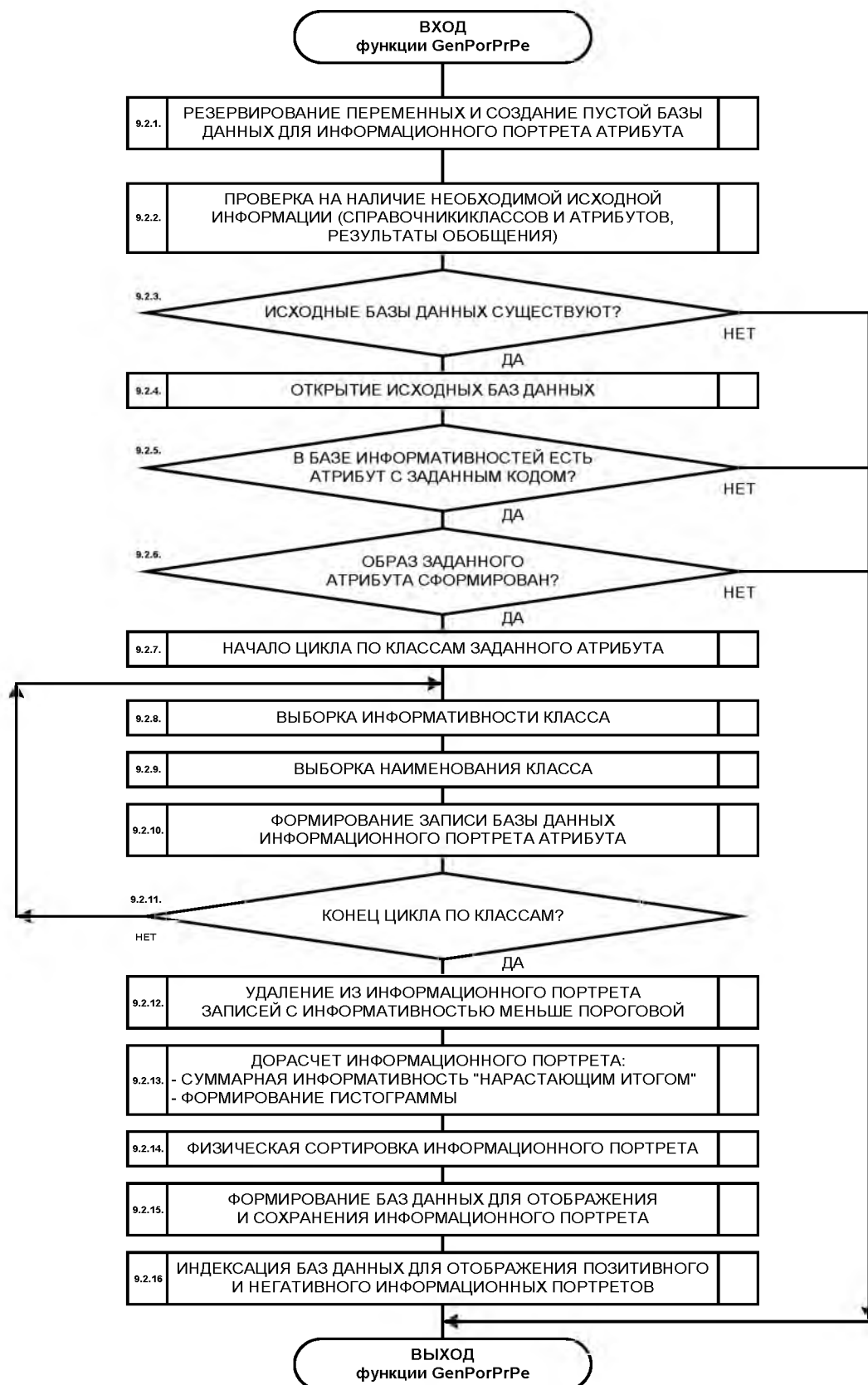


Рисунок 57. Алгоритм БКОСА-9.2. "Дедукция и абдукция факторов
(семантический анализ факторов)"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.1.1:
"РАСЧЕТ МАТРИЦЫ СХОДСТВА КЛАССОВ"

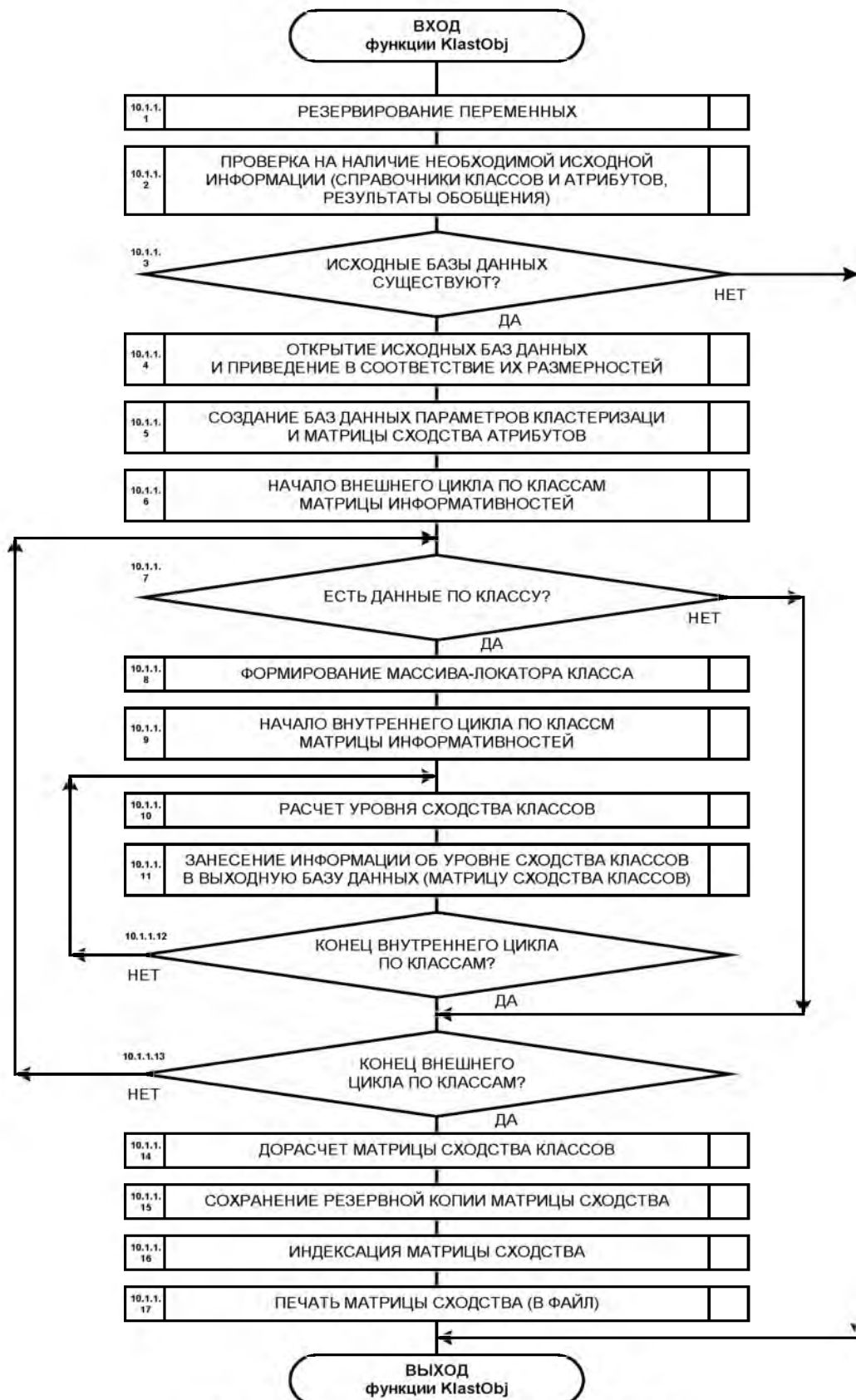


Рисунок 58. Алгоритм БКОСА-10.1.1.
"Классификация обобщенных образов классов"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.1.2: "РАСЧЕТ КЛАСТЕРОВ И КОНСТРУКТОВ КЛАССОВ"

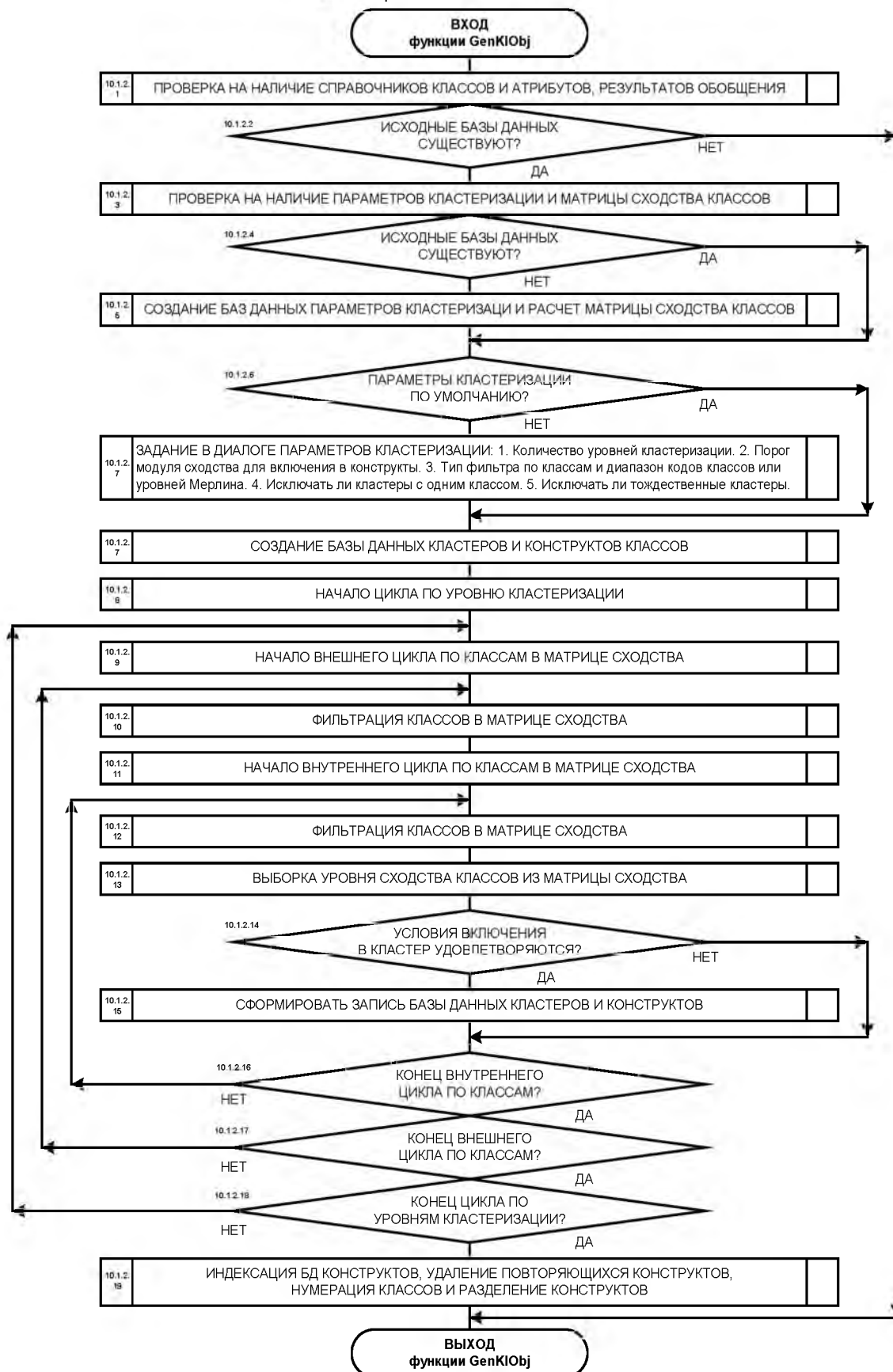


Рисунок 59. Алгоритм БКОСА-10.1.2. "Формирование бинарных конструктов классов"

**АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.1.3:
"ВИЗУАЛИЗАЦИЯ СЕМАНТИЧЕСКОЙ СЕТИ КЛАССОВ"**

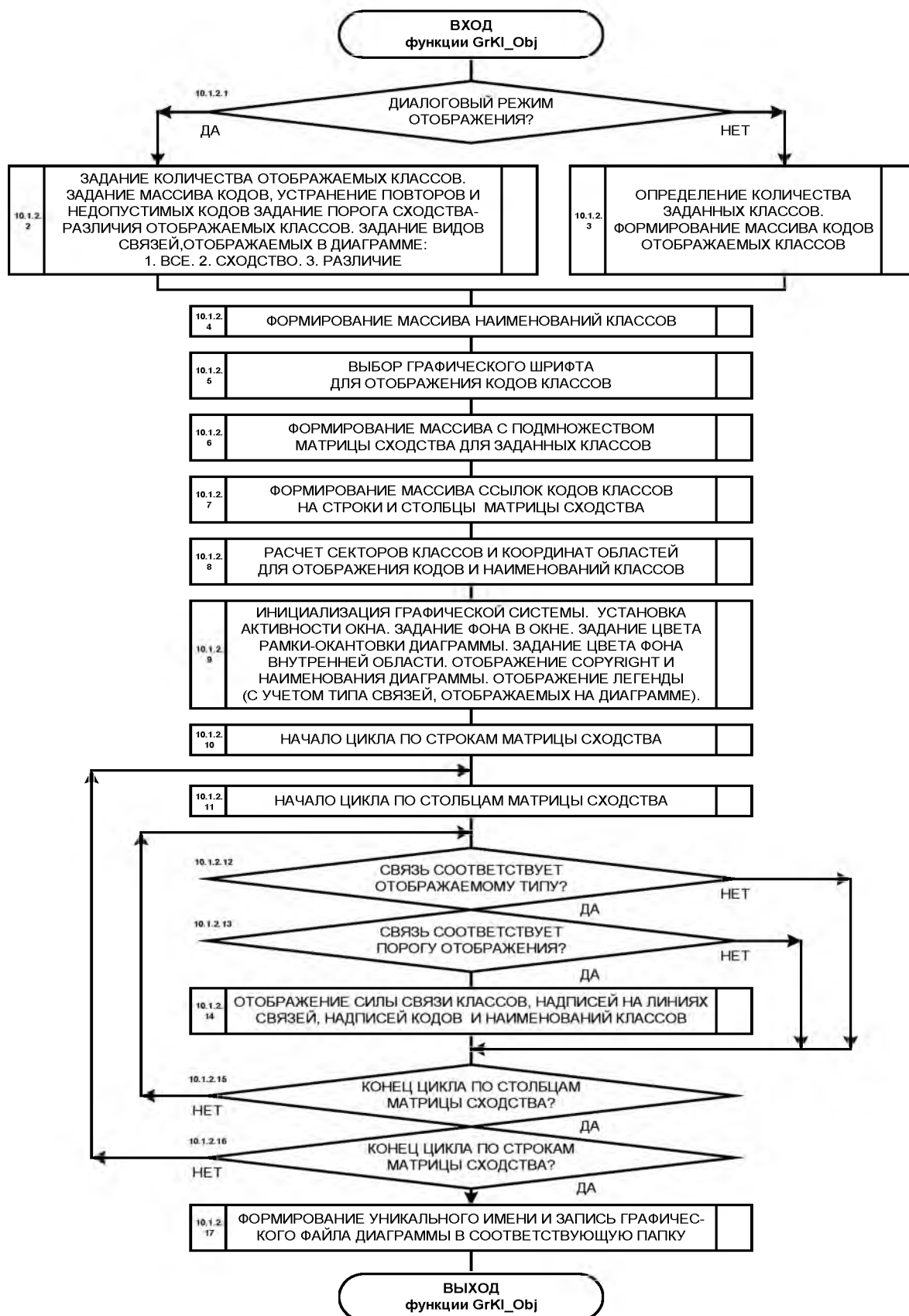


Рисунок 60. Алгоритм БКОСА-10.1.3. "Визуализация семантических сетей классов" (когнитивная графика)

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.2.1:
"РАСЧЕТ МАТРИЦЫ СХОДСТВА АТРИБУТОВ"

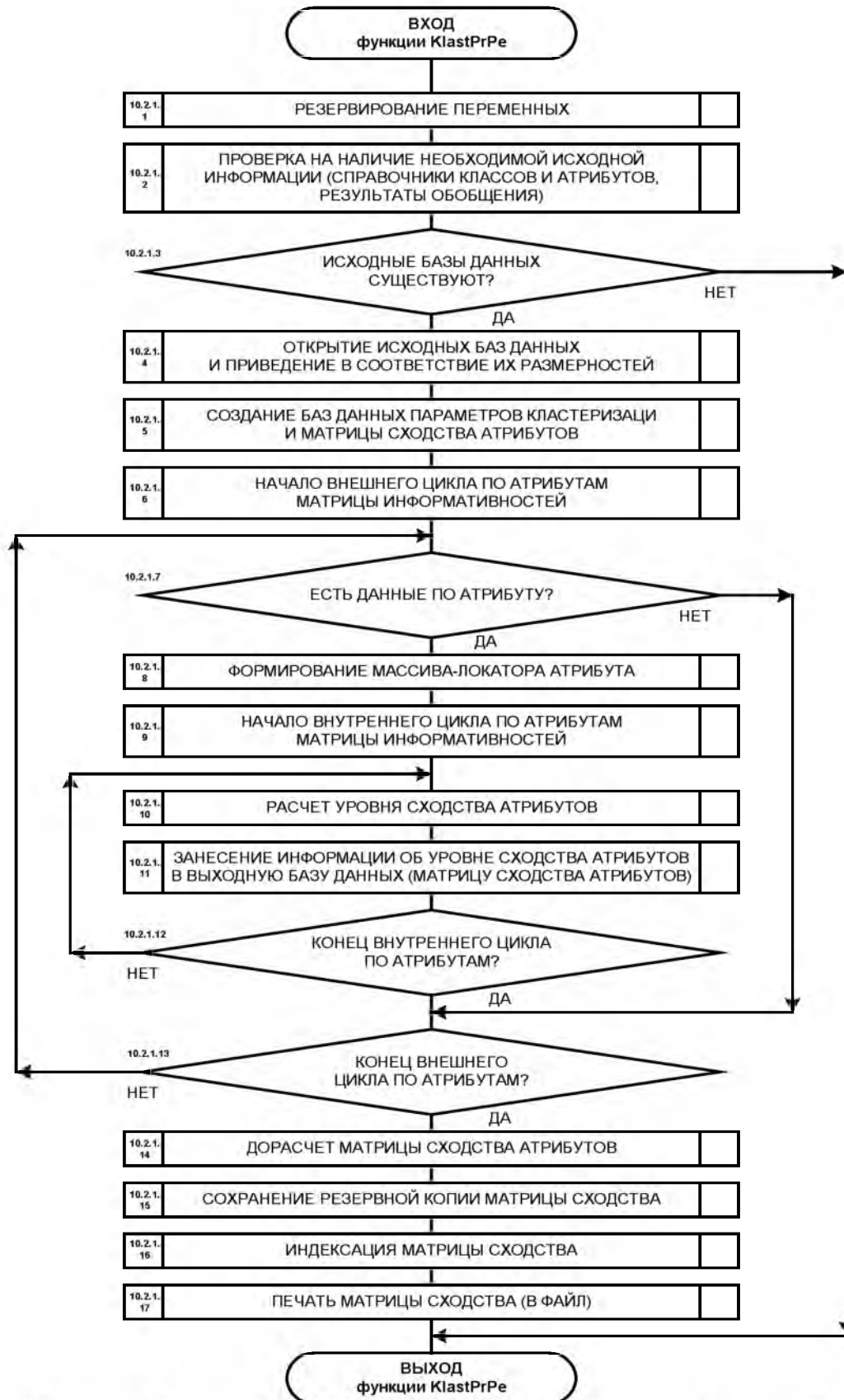


Рисунок 61. Алгоритм БКОСА10.2.1. "Классификация факторов"

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.2.2: "РАСЧЕТ КЛАСТЕРОВ И КОНСТРУКТОВ АТРИБУТОВ"

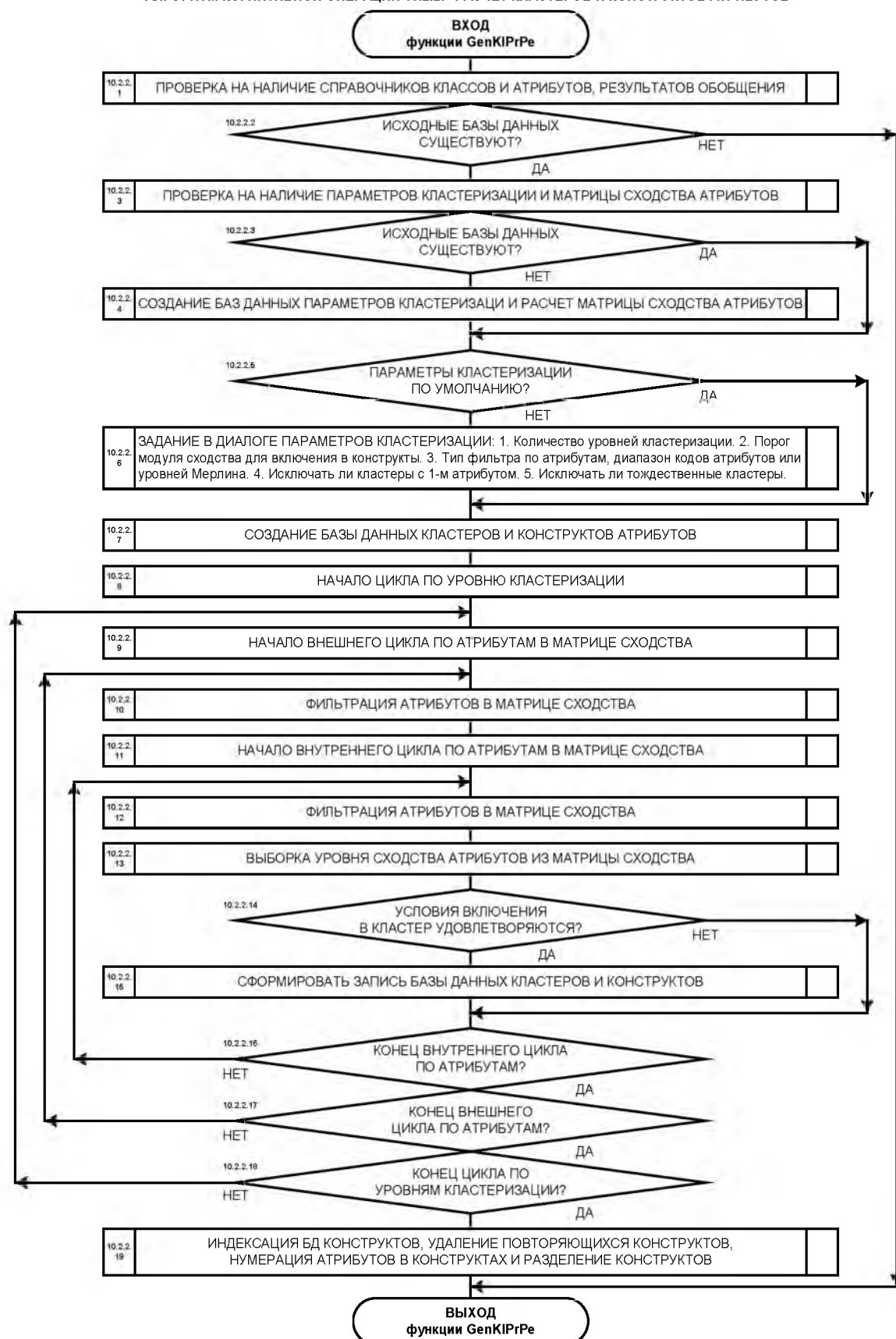


Рисунок 62. Алгоритм БКОСА-10.2.2. "Формирование бинарных конструктов факторов"

**АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.2.3:
"ВИЗУАЛИЗАЦИЯ СЕМАНТИЧЕСКОЙ СЕТИ АТРИБУТОВ"**

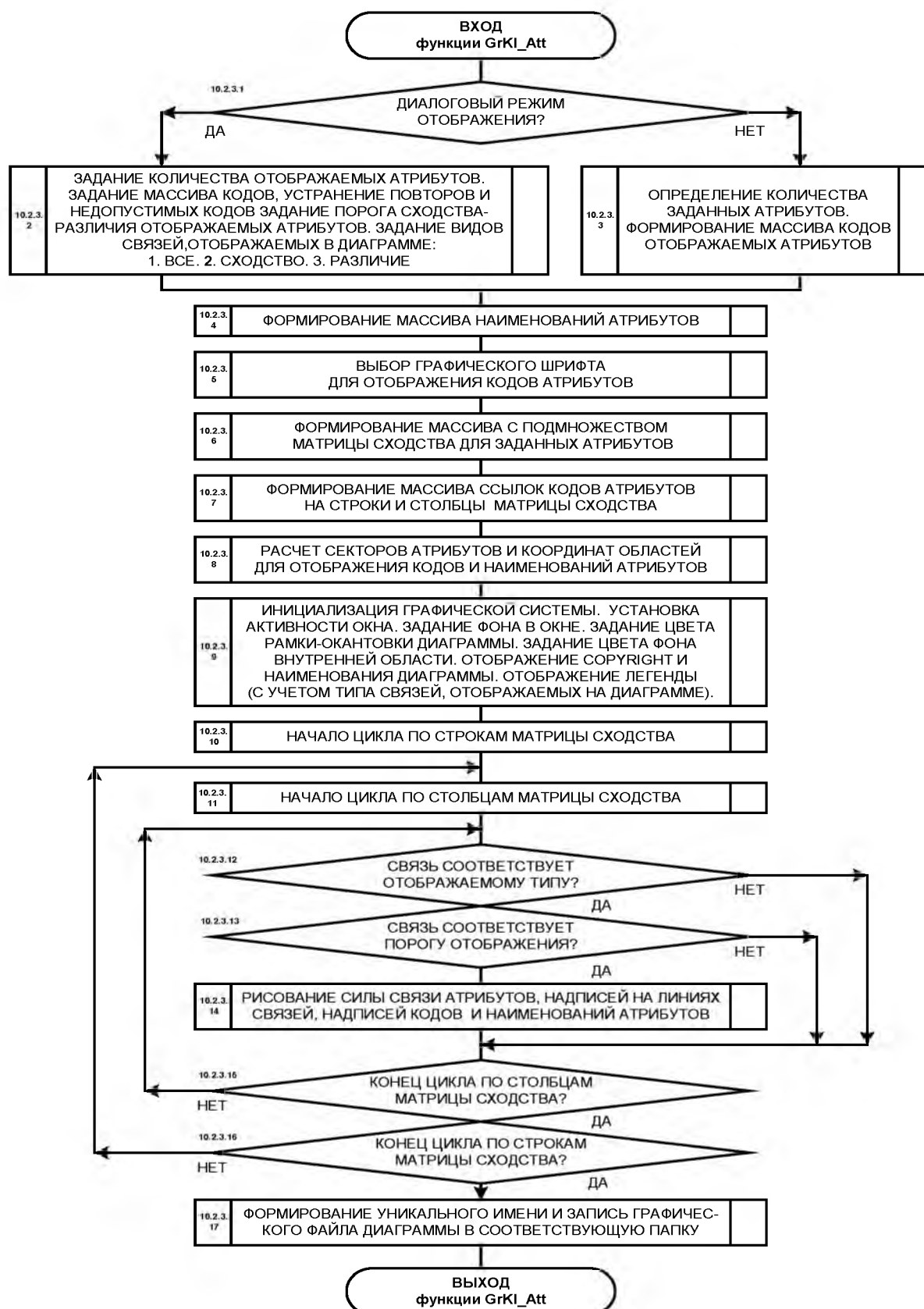


Рисунок 63. Алгоритм БКОСА-10.2.3. "Визуализация семантических сетей факторов" (когнитивная графика)

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.3.1:
"РАСЧЕТ КОГНИТИВНЫХ ДИАГРАММ КЛАССОВ"

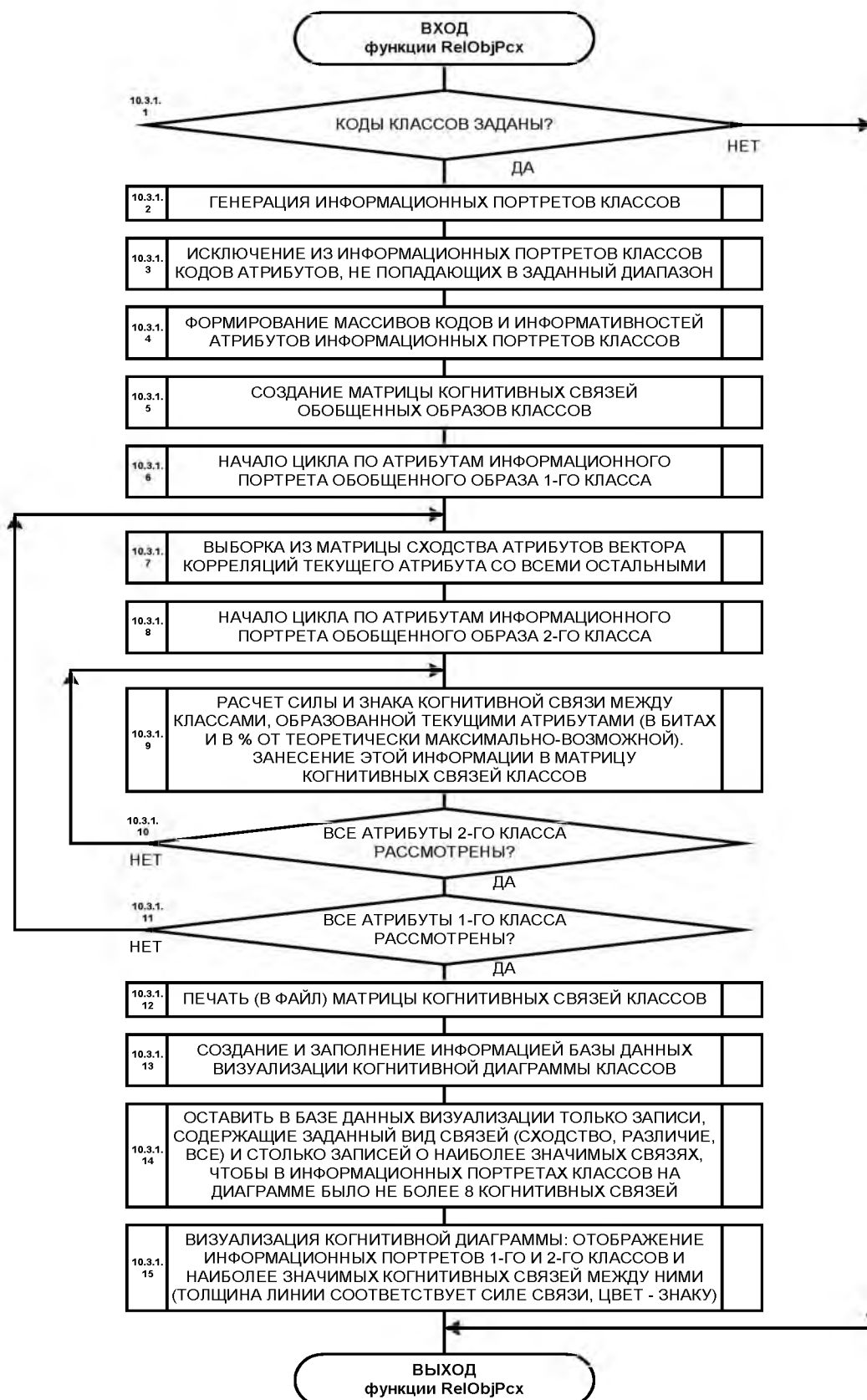


Рисунок 64. Алгоритм БКОСА-10.3.1. "Содержательное сравнение классов"

**АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.3.2:
"ВИЗУАЛИЗАЦИЯ КОГНИТИВНЫХ ДИАГРАММ КЛАССОВ"**

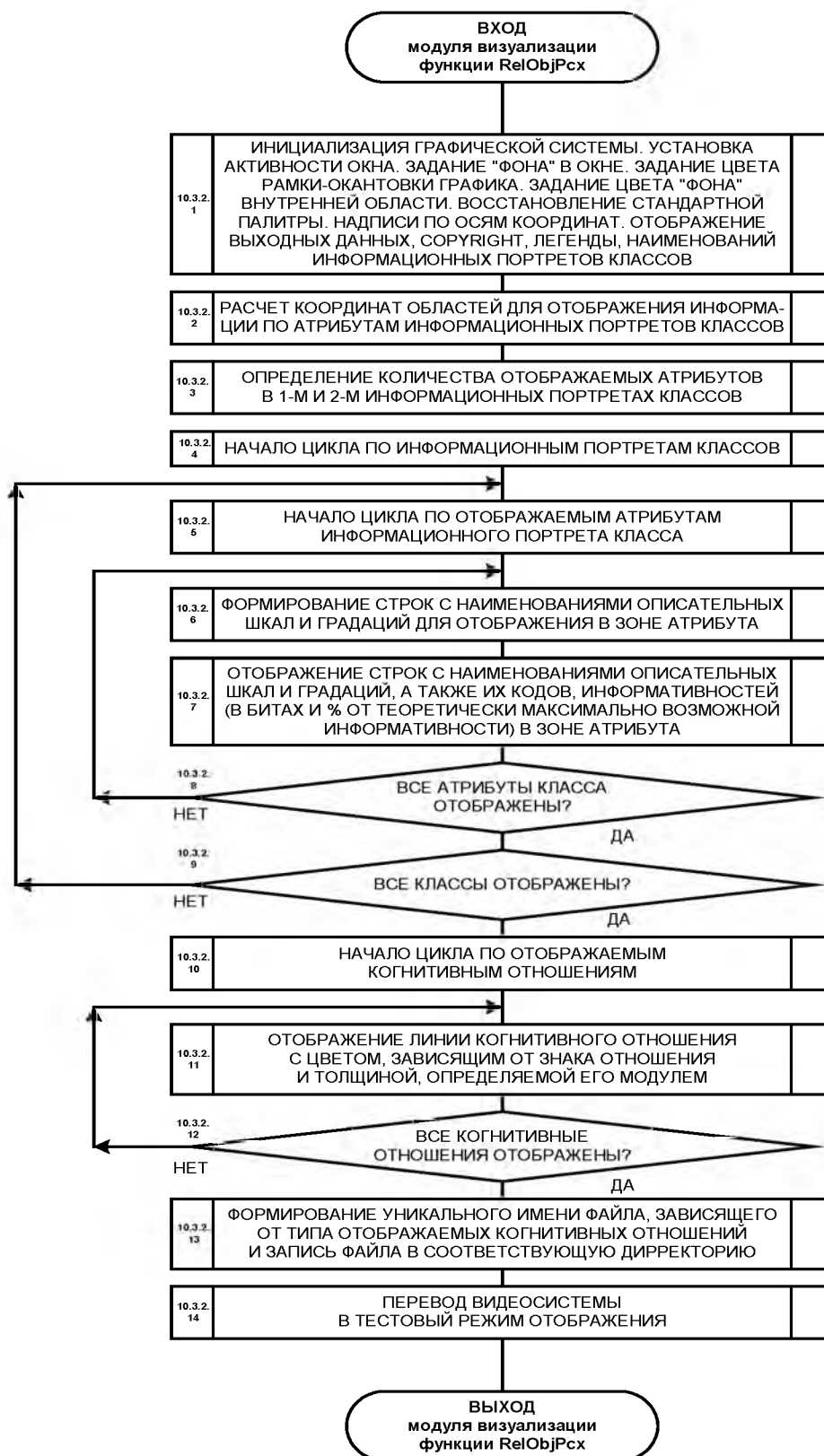


Рисунок 65. Алгоритм БКОСА-10.3.2. "Расчет и отображение много-многозначных когнитивных диаграмм, в т.ч. диаграмм Мерлина" (когнитивная графика)

**АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.4.1:
"РАСЧЕТ КОГНИТИВНЫХ ДИАГРАММ АТРИБУТОВ"**

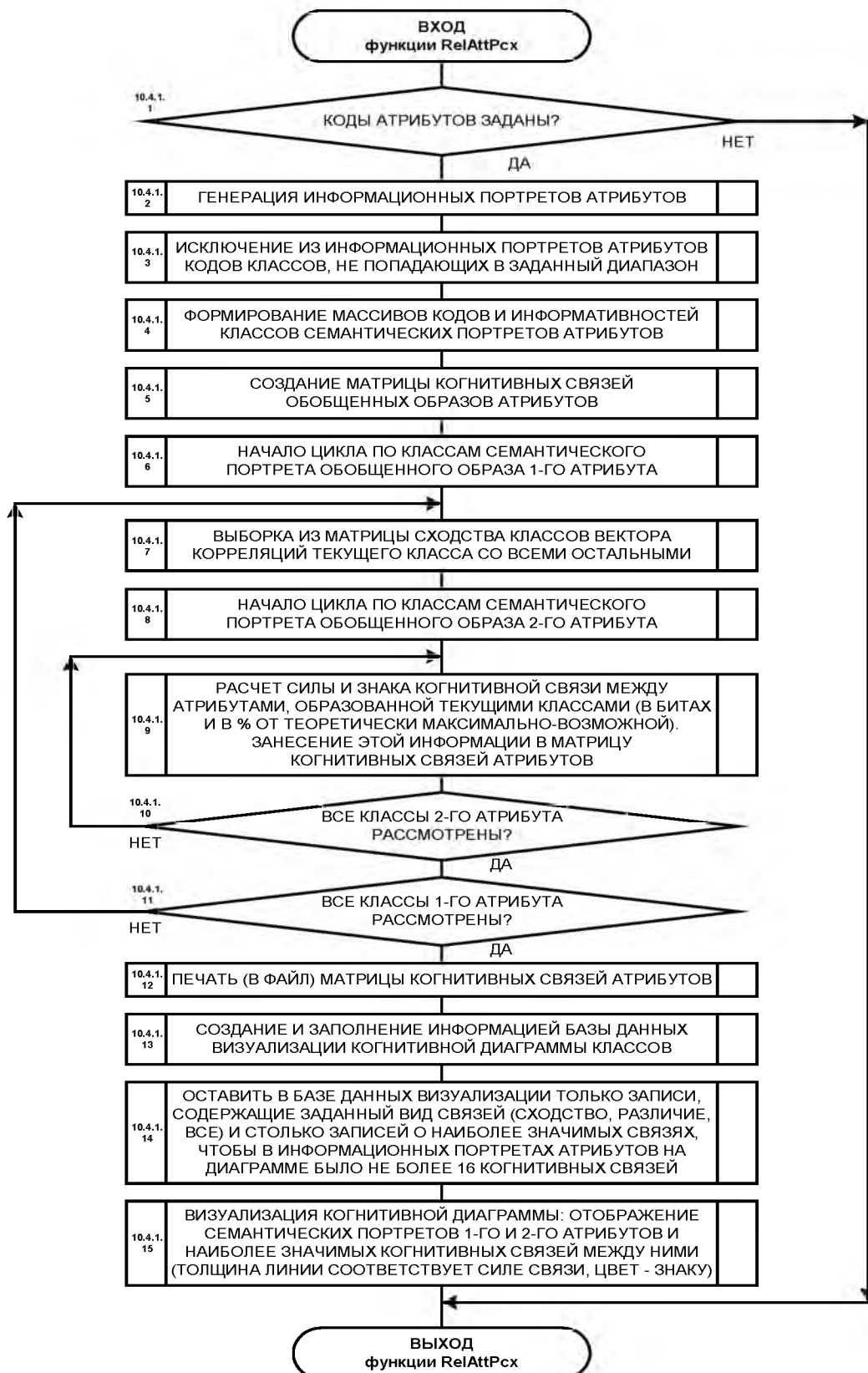
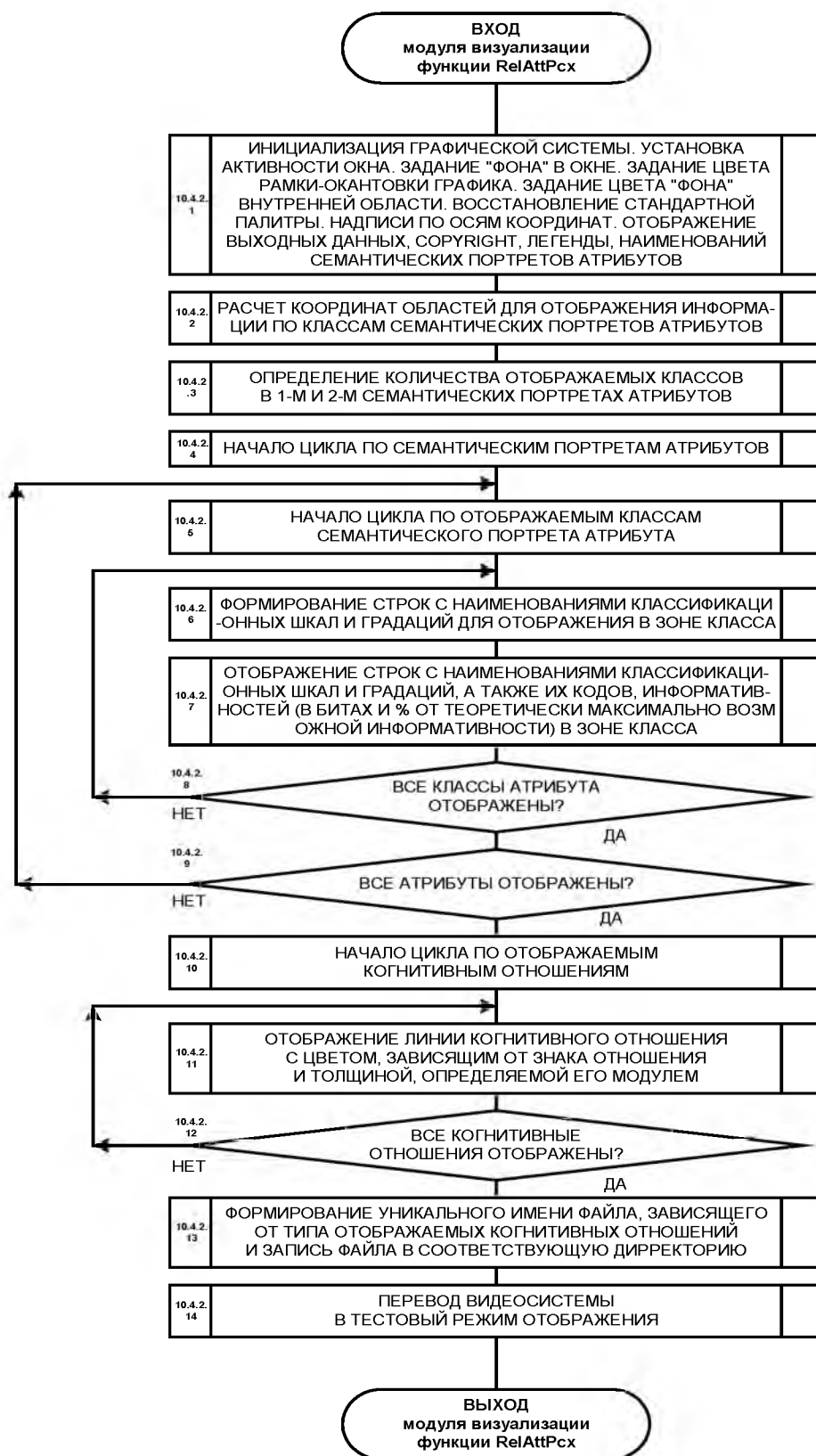


Рисунок 66. Алгоритм БКОСА-10.4.1. "Содержательное сравнение факторов"

**АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 10.4.2:
"ВИЗУАЛИЗАЦИЯ КОГНИТИВНЫХ ДИАГРАММ АТРИБУТОВ"**



**Рисунок 67. Алгоритм БКОСА-10.4.2. "Расчет и отображение много-
многозначных когнитивных диаграмм, в т.ч. инвертированных
диаграмм Мерлина" (когнитивная графика)**

АЛГОРИТМ КОГНИТИВНОЙ ОПЕРАЦИИ 11:
"МНОГОФАКТОРНОЕ ПЛАНИРОВАНИЕ"

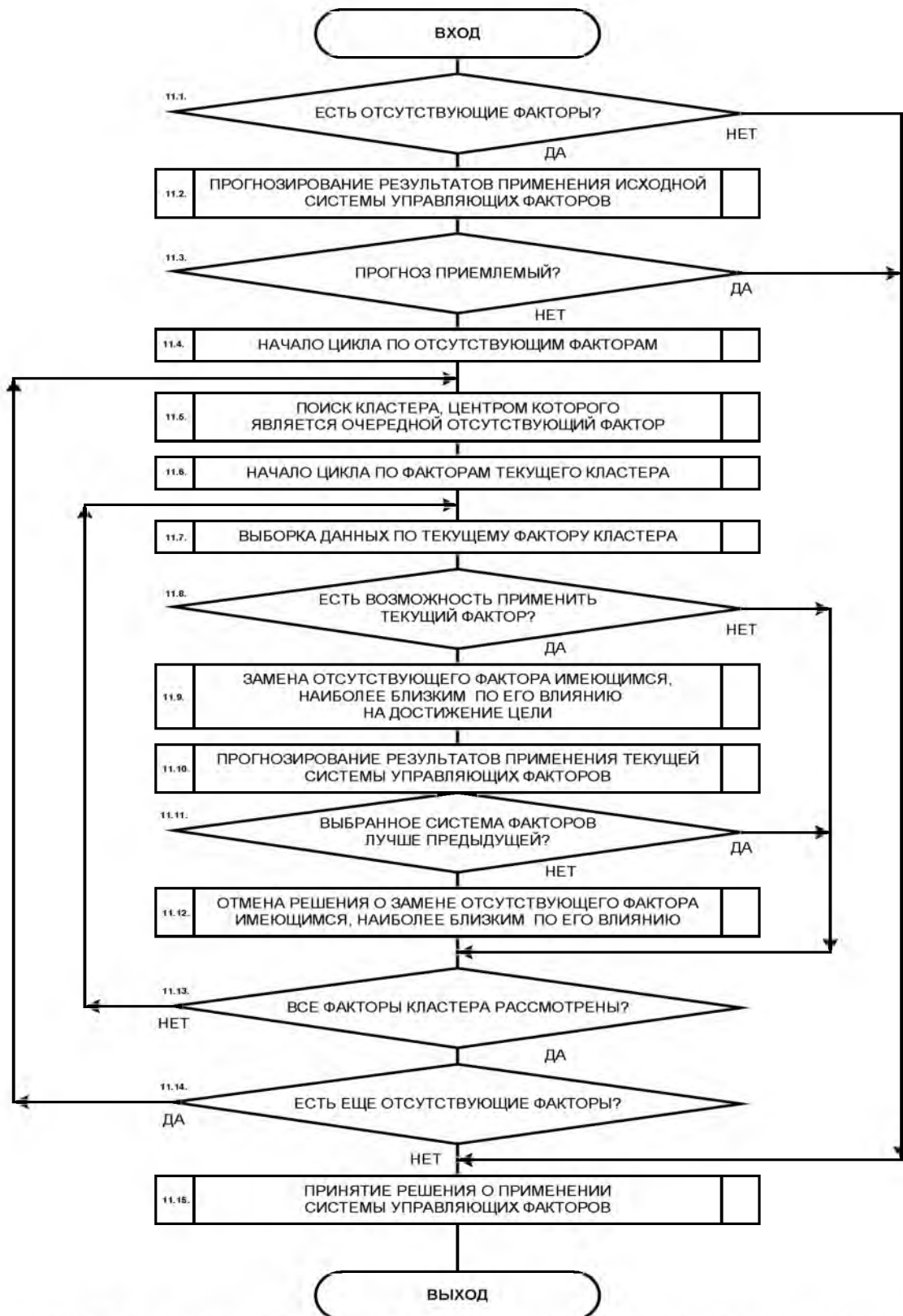


Рисунок 68. Алгоритм БКОСА-11. "Многовариантное планирование и принятие решения о применении системы управляющих факторов"

Резюме

1. Разработан **численный метод СК-анализа**, включающий:
 - иерархическую структуру данных семантической информационной модели;
 - 24 детальных алгоритма 10 базовых когнитивных операций системного анализа, алгоритмы кластерно-конструктивного и когнитивного анализа, нечеткой логики и когнитивной графики, обеспечивающие оригинальную визуализацию результатов интеллектуального анализа данных (нечеткие графы).
2. Предложенный численный метод СК-анализа обеспечил конкретизацию моделей БКОСА, достаточную для их реализации в одной программной системе.

Контрольные вопросы

1. Принципы формализации предметной области и подготовки эмпирических данных.
2. Иерархическая структура данных и последовательность численных расчетов в СК-анализе
3. Обобщенное описание алгоритмов СК-анализа
4. БКОСА-2.1. "Восприятие и запоминание исходной обучающей информации".
5. БКОСА-2.2. "Репрезентация. Сопоставление индивидуального опыта с коллективным (общественным)".
6. БКОСА-3.1.1. "Обобщение (синтез, индукция). Накопление первичных данных".
7. БКОСА-3.1.2. "Обобщение (синтез, индукция). Исключение артефактов".
8. БКОСА-3.1.3. "Обобщение (синтез, индукция). Расчет степени истинности содержательных смысловых связей между предпосылками и результатами (обобщенных таблиц решений)".
9. БКОСА-3.2. "Определение значимости шкал и градаций факторов, уровней Мерлина".
10. БКОСА-3.3. "Определение значимости шкал и градаций классов, уровней Мерлина".
11. БКОСА-4.1. "Абстрагирование факторов (снижение размерности семантического пространства факторов)".
12. БКОСА-4.2. "Абстрагирование классов (снижение размерности семантического пространства классов)".
13. БКОСА-5. "Оценка адекватности информационной модели предметной области".
14. БКОСА-7. "Сравнение, идентификация и прогнозирование. Распознавание состояний конкретных объектов (объектный анализ)".
15. БКОСА-9.1. "Дедукция и абдукция классов (семантический анализ обобщенных образов классов, решение обратной задачи прогнозирования)".

16. БКОСА-9.2. "Дедукция и абдукция факторов (семантический анализ факторов)".
17. БКОСА-10.1.1. "Классификация обобщенных образов классов".
18. БКОСА-10.1.2. "Формирование бинарных конструкторов классов".
19. БКОСА-10.1.3. "Визуализация семантических сетей классов".
20. БКОСА-10.2.1. "Классификация факторов".
21. БКОСА-10.2.2. "Формирование бинарных конструкторов факторов".
22. БКОСА-10.2.3. "Визуализация семантических сетей факторов".
23. БКОСА-10.3.1. "Содержательное сравнение классов".
25. БКОСА-10.3.2. "Расчет и отображение многозначных когнитивных диаграмм, в т.ч. диаграмм Вольфа Мерлина".
26. БКОСА-10.4.1. "Содержательное сравнение факторов".
27. БКОСА-10.4.2. "Расчет и отображение многозначных когнитивных диаграмм, в т.ч. инвертированных диаграмм Мерлина".
28. БКОСА-11. "Многовариантное планирование и принятие решения о применении системы управляющих факторов".
29. Детальные алгоритмы СК-анализа.

Рекомендуемая литература

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

1.2.4. ЛЕКЦИЯ-6.

Технология синтеза и эксплуатации приложений в системе "Эйдос"

Учебные вопросы

1. Назначение и состав системы "ЭЙДОС".
2. Пользовательский интерфейс, технология разработки и эксплуатации приложений в системе "ЭЙДОС".
3. Технические характеристики и обеспечение эксплуатации системы "ЭЙДОС" (версии 12.5).
4. АСК-анализ, как технология синтеза и эксплуатации рефлексивных АСУ активными объектами.

В данной лекции рассматривается инструментарий автоматизации СК-анализа в качестве которого выступает универсальная когнитивная аналитическая система "Эйдос". Данная система является *одним из вариантов* программной реализации предложенной математической модели и численного метода СК-анализа. Наличие данного инструментария, автоматизирующего СК-анализ, позволяет ввести в новый термин: *автоматизированный системно-когнитивный анализ (АСК-анализ)*, под которым понимается СК-анализ, оснащенный математическим методом, методикой численных расчетов и реализующим их программным инструментарием.

1.2.4.1. Назначение и состав системы "ЭЙДОС"

1.2.4.1.1. Цели и основные функции системы "Эйдос"

Универсальная когнитивная аналитическая система "Эйдос" является отечественным лицензионным программным продуктом [185-192], созданным с использованием официально приобретенного лицензионного программного обеспечения. По системе "Эйдос" и различным аспектам ее практического применения имеется более 100 публикаций автора с соавторами [76-155, 214-217, 219-231, 247-251, 269-271], в т.ч. 5 монографий [81, 125, 132, 221, 231] и одно учебное пособие [97]. Система "Эйдос" является программным инструментарием, реализующим математическую модель и методику численных расчетов СК-анализа. Она обеспечивает реализацию следующих функций:

1. Синтез и адаптация семантической информационной модели предметной области, включая активный объект управления и окружающую среду.
2. Идентификация и прогнозирование состояния активного объекта управления, а также разработка управляющих воздействий для его перевода в заданные целевые состояния.
3. Углубленный анализ семантической информационной модели предметной области.

Система "Эйдос" является специальным программным инструментарием, реализующим предложенные математическую модель и численный метод (структуры данных и алгоритмы) и решающим проблему данной работы.

Синтез содержательной информационной модели предметной области

Синтез модели в СК-анализе осуществляется с применением подсистем: "Словари", "Обучение", "Оптимизация", "Распознавание" и "Анализ". Он включает следующие этапы:

- 1) формализация (когнитивная структуризация предметной области);
- 2) формирование исследуемой выборки и управление ею;
- 3) синтез или адаптация модели;
- 4) оптимизация модели;
- 5) измерение адекватности модели (внутренней и внешней, интегральной и дифференциальной валидности), ее скорости сходимости и семантической устойчивости.

Идентификация и прогнозирование состояния объекта управления, выработка управляющих воздействий

Данный вид работ осуществляется с помощью подсистем "Распознавание" и "Анализ". Эти подсистемы обеспечивают: ввод распознаваемой выборки; пакетное распознавание; вывод результатов распознавания и их оценку, в т.ч. с использованием данных по дифференциальной валидности модели.

Углубленный анализ содержательной информационной модели предметной области

Этот анализ выполняется в подсистеме "Типология", которая включает:

1. Информационный и семантический анализ классов и признаков.
2. Кластерно-конструктивный анализ классов распознавания и признаков, включая визуализацию результатов анализа в оригинальной графической форме когнитивной графики (семантические сети классов и признаков).
3. Когнитивный анализ классов и признаков (когнитивные диаграммы и диаграммы Вольфа Мерлина).

1.2.4.1.2. Обобщенная структура системы "Эйдос"

Система "Эйдос" включает *семь* подсистем: "Словари", "Обучение", "Оптимизация", "Распознавание", "Типология", "Анализ", "Сервис" (таблица 30).

Таблица 30 – ОБОБЩЕННАЯ СТРУКТУРА СИСТЕМЫ "ЭЙДОС" (версии 12.5)

Подсистема	Режим	Функция	Операция
1. Словари	1. Классификационные шкалы и градации		
	2. Описательные шкалы (и градации)		
	3. Градации описательных шкал (признаки)		
	4. Иерархические уровни систем	1. Уровни классов	
		2. Уровни признаков	
	5. Программные интерфейсы для импорта данных	1. Импорт данных из TXT-файлов стандарта DOS-текст	
		2. Импорт данных из DBF-файлов стандарта проф. А.Н.Лебедева	
		3. Импорт из транспонированных DBF-файлов проф. А.Н.Лебедева	
		4. Генерация шкал и обучающей выборки RND-модели	
		5. Генерация шкал и обучающей выборки для исследования чисел	
		6. Транспонирование DBF-матриц исходных данных	
	6. Почтовая служба по НСИ	1. Обмен по классам	
2. Обмен по обобщенным признакам			
3. Обмен по первичным признакам			
7. Печать анкеты			

Подсистема	Режим	Функция	Операция
2. Обучение	1. Ввод–корректировка обучающей выборки	1. Параметрическое задание объектов для обработки 2. Статистическая характеристика, ручной ремонт 3. Автоматический ремонт обучающей выборки	
	3. Пакетное обучение системы распознавания	1. Накопление абсолютных частот	
		2. Исключение артефактов (робастная процедура)	
		3. Расчет информативностей признаков	
		4. Расчет условных процентных распределений	
5. Автоматическое выполнение режимов 1–2–3–4			
	6. Измерение схожести и устойчивости модели	1. Сходимость и устойчивость СИМ 2. Зависимость валидности модели от объема обучающей выборки	
	4. Почтовая служба по обучающей информации		
3. Оптимизация	1. Формирование ортонормированного базиса классов		
	2. Исключение признаков с низкой селективной силой		
	3. Удаление классов и признаков, по которым недостаточно данных		
4. Распознавание	1. Ввод–корректировка распознаваемой выборки		
	2. Пакетное распознавание		
	3. Вывод результатов распознавания	1. Разрез: один объект – много классов	
		2. Разрез: один класс – много объектов	
		4. Почтовая служба по распознаваемой выборке	
5. Типология	1. Типологический анализ классов распознавания	1. Информационные (ранговые) портреты (классов)	
		2. Кластерный и конструктивный анализ классов	1. Расчет матрицы сходства образов классов
			2. Генерация кластеров и конструктов классов
			3. Просмотр и печать кластеров и конструктов
			4. Автоматическое выполнение режимов: 1,2,3
		5. Вывод 2d семантических сетей классов	
		3. Когнитивные диаграммы классов	
	2. Типологический анализ первичных признаков	1. Информационные (ранговые) портреты признаков	
		2. Кластерный и конструктивный анализ признаков	1. Расчет матрицы сходства образов признаков
			2. Генерация кластеров и конструктов признаков
			3. Просмотр и печать кластеров и конструктов
			4. Автоматическое выполнение режимов: 1,2,3
5. Вывод 2d семантических сетей признаков			
	3. Когнитивные диаграммы признаков		
6. Анализ	1. Оценка достоверности заполнения объектов		
	2. Измерение адекватности семантической информационной модели		
	3. Измерение независимости классов и признаков		
	4. Просмотр профилей классов и признаков		
	5. Графическое отображение нелокальных нейронов		
	6. Отображение Паретто-подмножеств нейронной сети		
	7. Классические и интегральные когнитивные карты		
7. Сервис	1. Генерация (сброс) БД	1. Все базы данных	
		2. НСИ	1. Всех баз данных
			2. БД классов
			3. БД первичных признаков
			4. БД обобщенных признаков
		3. Обучающая выборка	
		4. Распознаваемая выборка	
		5. Базы данных статистики	
	2. Переиндексация всех баз данных		
	3. Печать БД абсолютных частот		
	4. Печать БД условных процентных распределений		
	5. Печать БД информативностей		
	6. Интеллектуальная дескрипторная информационно–поисковая система		

Структура и взаимодействие этих подсистем позволяют полностью реализовать все аспекты СК-анализа в удобной для пользователя форме. Обобщенной структуре соответствуют и структура управления и дерево диалога системы. По-

дробнее подсистемы, режимы, функции и операции, реализуемые системой "Эйдос", описаны в работах [81, 97, 125].

1.2.4.2. Пользовательский интерфейс, технология разработки и эксплуатации приложений в системе "ЭЙДОС"

Не смотря на то, что данный раздел посвящен интерфейсу системы "Эйдос", видеогаммы и экранные формы в нем не приводятся, т.к. они есть в описаниях лабораторных работ. В наименованиях разделов с описаниями подсистем и режимов системы "Эйдос" указаны коды реализуемых ими базовых когнитивных операций системного анализа в соответствии с обобщенной схемой СК-анализа (рисунок 14).

1.2.4.2.1. Начальный этап синтеза модели: когнитивная структуризация и формализация предметной области, подготовка исходных данных (подсистема "Словари") (БКОСА-1, БКОСА-2)

Подсистема "Словари" обеспечивает формализацию предметной области. Она реализует следующие режимы: классификационные шкалы и градации; описательные шкалы и градации; градации описательных шкал; иерархические уровни организации систем; автоматический ввод первичных признаков из текстовых файлов; почтовая служба по нормативно-справочной информации; печать анкеты.

Классификационные шкалы и градации (БКОСА-1.1)

Классификационные шкалы и градации предназначены для ввода справочника будущих состояний активного объекта управления – классов. Режим: "Классификационные шкалы и градации" обеспечивает ведение базы данных классификационных шкал и градаций классов: ввод; корректировку; удаление; распечатку (в текстовый файл); сортировку; поиск по базе данных.

Описательные шкалы и градации (БКОСА-1.2)

Описательные шкалы и градации предназначены для ввода справочников факторов, влияющих на поведение активного объекта управления – признаков. В этом режиме обеспечивается ввод, удаление, корректировка, копирование наименований описательных шкал и связанных с ними градаций. Характерной особенностью системы "Эйдос" является возможность использования неальтернативных градаций, которых может быть различное количество по различным шкалам (в широких пределах). Справочник позволяет работать непосредственно с градациями (с учетом связей со шкалами), видеть их общее количество, а также просматривать и распечатывать процентное распределение ответов респондентов по.

Уровни организации систем (уровни Вольфа Мерлина) являются независимым способом классификации классов и факторов, что позволяет легко создавать и анализировать различные их подмножества как сами по себе, так и в со-

поставлении друг с другом. В.С.Мерлин предложил интегральную концепцию индивидуальности, в которой рассматривал взаимодействие и взаимообусловленность различных уровней свойств личности: от генетически предопределенных, до социально-обусловленных и отражающих сиюминутное состояние. В системе "Эйдос" предусмотрен аппарат, позволяющий классифицировать факторы таким образом, что становится возможным исследовать различные уровни их организации и взаимообусловленности. Уровни организации классов предназначены для классификации будущих состояний активного объекта управления, как целевых и нежелательных с точки зрения самого объекта управления и управляющей системы, а также различных вариантов сочетаний этих вариантов. Возможны и другие виды классификации.

Система "Эйдос" обеспечивает решение задач атрибуции анонимных и псевдонимных текстов (установления вероятного авторства), датировки текстов, определения их принадлежности к определенным традициям, школам или течениям мысли [81, 97, 125]. При этом различные структуры, из которых состоят тексты, рассматриваются как их атрибуты. В системе "Эйдос" реализован специальный режим, обеспечивающий автоматическое выявление и ввод этих атрибутов текстов непосредственно из текстовых файлов.

Технология работы в системе "Эйдос" не предусматривает одновременной работы многих пользователей с одними и теми же базами данных в режиме корректировки записей. Поэтому возможна эффективная организация распределенной работы по многомашинной технологии без использования ЛВС. Для обеспечения необходимой тождественности справочников на различных компьютерах служит режим "Почтовая служба по НСИ".

Классификационные шкалы и градации в экономических, социально-психологических и политологических исследованиях часто представляют собой опросники (анкеты). Для их распечатки в файл (в поддиректорию "ТХТ") служит режим: "Печать анкеты". В системе "Эйдос" все текстовые и графические входные и выходные формы автоматически сохраняются в виде файлов, удобных для использования в различных приложениях под Windows.

Ввод-корректировка обучающей информации (БКОСА-2.1)

Данная подсистема обеспечивает ввод и корректировку обучающей выборки, управление ею, синтез и адаптацию модели на основе данных обучающей выборки, экспорт и импорт данных с других компьютеров.

Для ввода-корректировки обучающей выборки служит соответствующий режим, имеющий двухоконный интерфейс, позволяющий ввести в обучающую выборку двухвекторные описания объектов. Левое окно служит для ввода классификационной характеристики объекта. В этом окне каждому объекту соответствует одна строка с прокруткой. В правом окне вводится описательная характеристика объекта на языке признаков. Каждому объекту соответствует окно с прокруткой. Переход между окнами осуществляется по нажатию клавиши "ТАВ". Количество объектов в обучающей выборке не ограничено. Имеется практический опыт проведения расчетов с объемами обучающей выборки до

7000 объектов, суммарным количеством градаций описательных шкал до 3900 и количеством классов до 1500. Реализована также возможность автоматического формирования объектов обучающей выборки путем кодирования текстовых файлов.

В системе реализован ряд программных интерфейсов, обеспечивающих автоматическое формирование классификационных и описательных шкал и градаций, а также обучающей выборки:

- импорт данных из файлов стандарта "Текст DOS";
- импорт данных из DBF-файлов, стандарта проф. А.Н.Лебедева;
- импорт данных из транспонированных DBF-файлов, стандарта профессора А.Н.Лебедева;
- генерация случайной модели;
- генерация учебной модели для исследования свойств натуральных чисел.

Управление составом обучающей информации (БКОСА-2.2)

Данный режим предназначен для управления обучающей выборкой путем параметрического задания подмножеств анкет для обработки, объединения классов, автоматического ремонта обучающей выборки ("ремонт или взвешивание данных"). Параметрическое выделение подмножества анкет для обработки может осуществляться логически и физически (рекомендуется 2-й вариант), это осуществляется путем сравнения с анкетой-маской. В ней задаются коды тех классов и признаков, которые обязательно должны присутствовать во всех анкетах обрабатываемого подмножества. Режим: "Статистическая характеристика обучающей выборки. Ручной ремонт" предназначен для выявления слабо представленных классов (по которым недостаточно данных) и объединения нескольких классов в один. При этом производится переформирование справочника классов и автоматическое перекодирование анкет обучающей выборки. *В режиме "Автоматический ремонт обучающей выборки (ремонт или взвешивание данных)" реализуется БКОСА-2.2: задается частотное распределение объектов по категориям, характерное для генеральной совокупности (или другое), затем автоматически осуществляется формирование последовательных подмножеств анкет обучающей выборки (с увеличивающимся числом анкет), на каждом этапе максимально соответствующих заданному частотному распределению генеральной совокупности.* При этом используется метод последовательных приближений по минимаксному критерию: максимизация корреляции и минимизация максимального отклонения. Соответствующие графики представлены на рисунке 69.

Система рекомендует оптимальное (по этим двум критериям) подмножество и позволяет исключить остальные анкеты из рассмотрения. На рисунке 70 приведены графики частотных распределений объектов генеральной совокупности и выбранного подмножества обучающей выборки по категориям (классам), а также отклонение между этими распределениями.

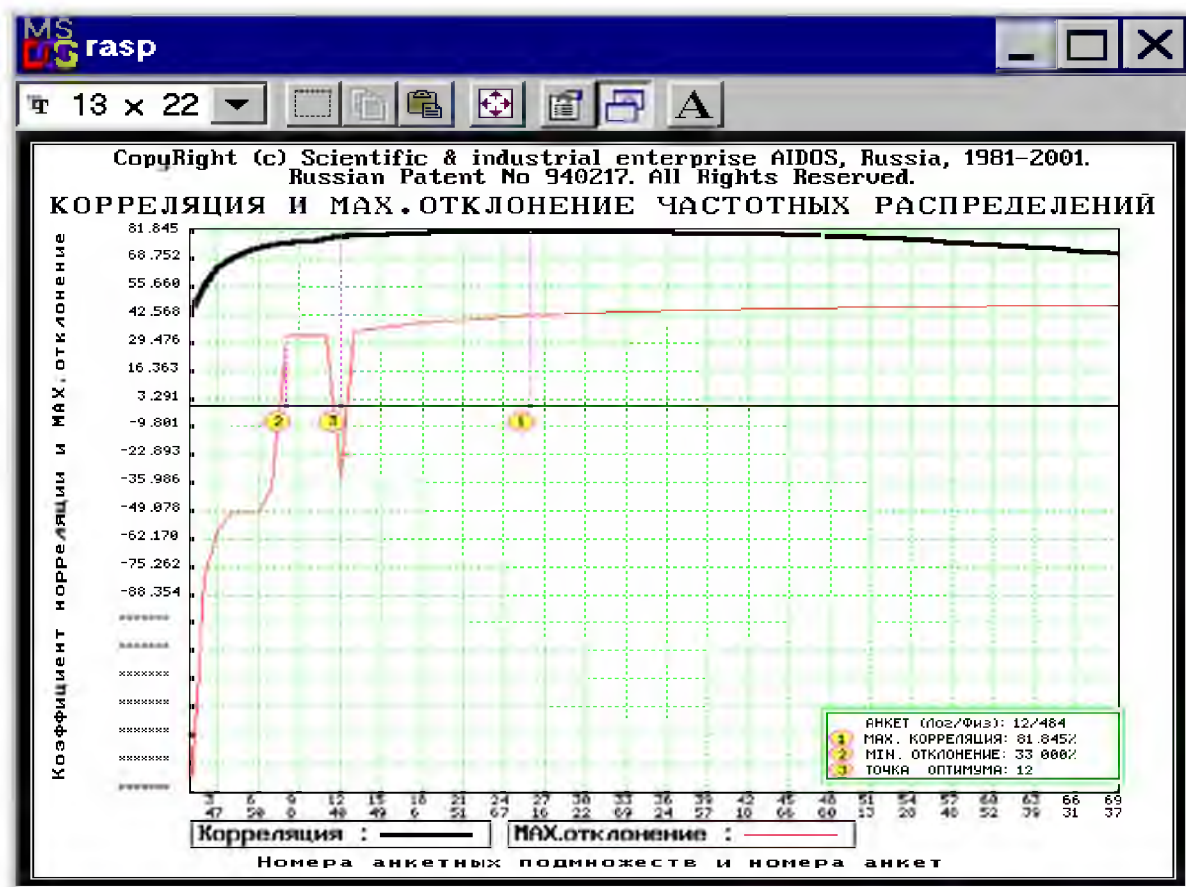


Рисунок 69. Автоматический ремонт обучающей выборки (диагр.1) (БКОСА-2.2)

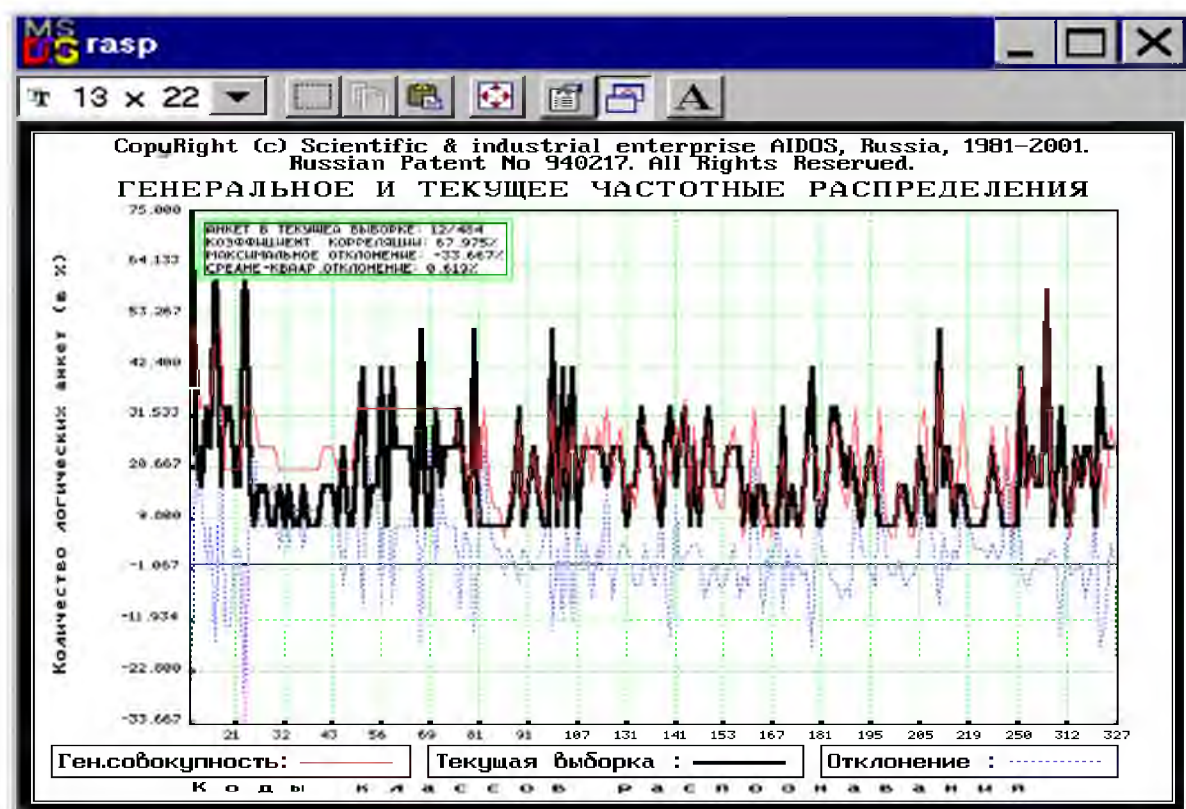


Рисунок 70. Автоматический ремонт обучающей выборки (диагр.2) (БКОСА-2.2)

При достижении минимакса можно говорить об обеспечении структурной репрезентативности [81].

1.2.4.2.2. Синтез модели: пакетное обучение системы распознавания (подсистема "Обучение") (БКОСА-3)

Данный режим обеспечивает: расчет матрицы абсолютных частот, поиск и исключение из дальнейшего анализа артефактов, расчет матрицы информативностей, расчет матрицы условных процентных распределений, пакетный режим автоматического выполнения вышеперечисленных 4-х режимов, а также исследовательский режим, обеспечивающий измерение скорости сходимости и семантической устойчивости сформированной содержательной информационной модели.

Расчет матрицы абсолютных частот (БКОСА-3.1.1)

В данном режиме осуществляется последовательное считывание всех анкет обучающей выборки и использование описаний объектов для формирования статистики встреч признаков в разрезе по классам. На экране в наглядной форме отображается стадия этого процесса, который может занимать значительное время при больших размерностях задачи и объеме обучающей выборки. Кроме того на качественном уровне красным отображается заполнение матрицы абсолютных частот данными: классы соответствуют столбцам, а признаки – строкам. Поэтому значительная фрагментарность данных легко обнаруживается еще на этой стадии. Данный режим обеспечивает полную "развязку по данным" и *независимость времени исполнения процессов синтеза модели и ее анализа от объема обучающей выборки*. Кроме того в данном режиме выявляются 4 типа формально-обнаружимых ошибок в исходных данных и по ним формируется файл отчета.

Исключение артефактов (робастная процедура) (БКОСА-3.1.2)

В данном режиме на основе исследования частотного распределения частот встреч признаков в матрице абсолютных частот, делаются выводы:

- об отсутствии статистики и невозможности обнаружения и исключения артефактов;
- о наличии статистики и возможности выявления артефактов (если частоты встреч признаков растут пропорционально объему обучающей выборки, то это нормально, артефактами считаются признаки, по которым эта закономерность нарушается).

На основе этих выводов рекомендуется частота, которая признается незначимой и характерной для артефактов и осуществляется переформирование баз данных с исключенными артефактами.

Расчет матриц информативностей (БКОСА-3.1.3, 3.2, 3.3)

В этом режиме непосредственно на основе матрицы абсолютных частот с применением системного обобщения формулы Харкевича, предложенного автором в рамках СТИ (3.28), рассчитывается матрица информативностей, определяются значимость признаков, степень сформированности обобщенных образов классов, а также обобщенный критерий сформированности модели Харкевича (3.63) для всей матрицы информативностей в целом. На экране монитора наглядно отображается стадия выполнения процесса и структура заполнения

матрицы информативностей значимыми данными (на качественном уровне). На основе матрицы абсолютных частот рассчитывается и матрица условных процентных распределений.

Автоматическое выполнение режимов 1-2-3-4. В данном пакетном режиме последовательно выполняются ранее перечисленные режимы обучения системы (кроме режима исключения артефактов).

Измерение сходимости и устойчивости модели

Для измерения сходимости и устойчивости модели СК-анализа задаются параметры, определяющие исследование скорости сходимости:

- порядок выборки анкет (физический, случайный, в порядке возрастания соответствия генеральной совокупности, в порядке убывания степени многообразия, вносимого анкетой в модель);
- количество и коды признаков, по которым исследуется сходимость модели;
- интервал сглаживания для расчета скользящей погрешности.

В данном режиме организован цикл по объектам обучающей выборки, в котором после учета каждой анкеты в матрице абсолютных частот пересчитывается матрица информативностей и в отдельной базе данных запоминаются информативности для заданных признаков. Это позволяет измерять и графически отображать скорость сходимости и семантическую устойчивость модели. В работах [81, 97], на примере прогнозирования фондового рынка, подробно рассматриваются вопросы сходимости и семантической устойчивости содержательной информационной модели.

Почтовая служба по обучающей информации обеспечивает экспорт и импорт баз данных обучающей выборки при решении задач в системе "Эйдос" по многомашинной технологии.

1.2.4.2.3. Оптимизация модели (подсистема "Оптимизация") (БКОСА-4)

В данной подсистеме различными способами корректно реализуется контролируемое существенное снижение размерности семантических пространств классов и атрибутов при несущественном уменьшении их объема.

Формирование ортонормированного базиса классов (БКОСА-4.2)

Формирование ортонормированного базиса классов реализуется с применением одного из трех итерационных алгоритмов оптимизации, относящиеся к методу последовательных приближений:

- 1) исключение из модели заданного количества наименее сформированных классов;
- 2) исключение заданного процента количества классов от оставшихся (адаптивный шаг);
- 3) исключение классов, вносящих заданный процент степени сформированности от оставшегося суммарного (адаптивный шаг).

Критерий остановки процесса последовательных приближений – срабатывание хотя бы одного из заданных ограничений:

- а) достигнуто заданное минимальное количество классов в модели;
- б) достигнута заданная полнота описания признака.

Прокрутка окна вправо позволяет просмотреть дополнительные характеристики, позволяющие оценить степень сформированности образов классов и ортонормированность пространства классов.

Исключение признаков с низкой селективной силой (БКОСА-4.1)

С этой целью реализовано три итерационных алгоритма оптимизации, относящиеся к методу последовательных приближений:

- путем исключения из модели заданного количества наименее значимых признаков;
- путем исключения заданного процента количества признаков от оставшихся (адаптивный шаг);
- путем исключения признаков, вносящих заданный процент значимости от оставшейся суммарной (адаптивный шаг).

Критерий остановки процесса исключения признаков с низкой селективной силой – срабатывание одного из заданных ограничений:

- а) достигнуто заданное минимальное количество признаков в модели;
- б) достигнуто заданное минимальное количество признаков на класс (полнота описания класса).

Удаление классов и признаков, по которым недостаточно данных

В данном режиме реализована возможность удаления из модели всех классов и признаков, по которым или вообще нет данных, или их недостаточно в соответствии с заданным критерием. Этот режим сходен с режимом выявления и исключения артефактов.

1.2.4.2.4. Верификация модели (оценка ее адекватности) (БКОСА-5)

Данный режим выполняется после синтеза модели. Верификация модели осуществляется путем копирования обучающей выборки в распознаваемую, пакетного распознавания и последующего анализа в режиме "Измерение валидности системы распознавания" подсистемы "Анализ". Он показывает средневзвешенную погрешность идентификации (интегральная валидность) и погрешность идентификации в разрезе по классам. При этом объект считается отнесенным к классу, с которым у него наибольшее сходство. Необходимо отметить, что остальные классы, находящиеся по уровню сходства на второй и последующих позициях *не учитываются*. Это обусловлено тем, что их учет привел бы к *завышению* оценки валидности модели.

Классы, по которым дифференциальная валидность неприемлемо низка считаются не сформированными. Причиной этого может быть очень высокая вариабельность объектов, отнесенных к данным классам (тогда имеет смысл разделить их на несколько), а также недостаток достоверной классификационной и описательной информации по этим классам (некорректная работа экспертов).

1.2.4.2.5. Эксплуатация приложения в режиме адаптации и периодического синтеза модели (БКОСА-7, БКОСА-9, БКОСА-10)

Идентификация и прогнозирование (подсистема "Распознавание") (БКОСА-7)

Данная подсистема реализует режимы ввода и корректировки распознаваемой выборки; пакетного распознавания; вывода результатов и межмашинного обмена данными. Ввод-корректировка распознаваемых анкет осуществляется в двухоконном интерфейсе: в левом окне показаны заголовки идентифицируемых объектов, в которых отображаются их коды и условные наименования, а в правом окне – описания объектов на языке признаков. В левом окне каждому объекту соответствует строка, а в правом – окно с прокруткой. Переход между окнами происходит по нажатию клавиши "ТАВ". В данном режиме каждая анкета распознаваемой выборки последовательно идентифицируется с каждым классом. Вывод результатов распознавания (идентификации и прогнозирования) возможен в двух разрезах:

- а) информация о сходстве каждого объекта со всеми классами;
- б) информация о сходстве каждого класса со всеми объектами.

Система генерирует обобщающий отчет по итогам идентификации, в котором в каждой строке дана информация о классе, с которым распознаваемый объект имеет наивысший уровень сходства (в процентах). Качество результата идентификации – это эвристическая оценка качества, учитывающая максимальную величину сходства, различие между первым и вторым классами по уровню сходства и в (меньшей степени) общий вид распределения классов по уровням сходства с данным объектом. Каждой строке обобщающего отчета соответствует карточка результатов идентификации (прогнозирования), которая по сути дела представляет собой результат разложения вектора объекта в ряд по векторам классов. Эти карточки распечатываются в файл с полными наименованиями классов и содержат классы, с уровнем сходства выше заданного.

Почтовая служба по распознаваемым анкетам обеспечивает запись на дискету распознаваемой выборки и считывание распознаваемой выборки с дискеты с добавлением к имеющейся на текущем компьютере. Этот режим служит для объединения информации по идентифицируемым объектам, введенной на различных компьютерах.

Подсистема "Типология" обеспечивает типологический анализ классов и признаков.

Типологический анализ классов включает: информационные (ранговые) портреты; кластерно-конструктивный и когнитивный анализ классов.

Информационные портреты классов (БКОСА-9.1)

Информационный портрет класса представляет собой список признаков в порядке убывания количества информации о принадлежности к данному классу. Такой список представляет собой результат решения обратной задачи идентификации (прогнозирования). Фильтрация (F6) позволяет выделить из информационного портрета класса диапазон признаков (по кодам или уровням Мерли-

на) и, таким образом, исследовать влияние заданных признаков на переход активного объекта управления в состояние, соответствующее данному классу.

Кластерный и конструктивный анализ классов обеспечивает: расчет матрицы сходства классов; генерацию кластеров и конструктов; просмотр и печать кластеров и конструктов; пакетный режим, обеспечивающий автоматическое выполнение первых трех режимов при установках параметров "по умолчанию"; визуализацию результатов кластерно-конструктивного анализа в форме семантических сетей и когнитивных диаграмм.

Расчет матрицы сходства эталонов классов (БКОСА-10.1.1)

В данном режиме непосредственно на основе оптимизированной матрицы информативностей рассчитывается матрица сходства классов. На экране в наглядной форме отображается информация о текущей стадии выполнения этого процесса.

Генерация кластеров и конструктов классов (БКОСА-10.1.2)

В данном режиме пользователем задаются параметры для генерации кластеров и конструктов классов, позволяющие исключить из форм центральную часть конструктов (оставить только полюса), а также сформировать кластеры и конструкты для заданных (кодами или уровнями Мерлина) подматриц. В данном режиме обеспечивается отображение отчета по конструктам и вывод его в виде текстового файла. Реализован режим быстрого поиска заданного конструкта и быстрый выход на него по заданному классу.

Автоматическое выполнение режимов 1-2-3

В данном пакетном режиме автоматически выполняются вышеперечисленные 3 режима с параметрами "по умолчанию". Выполнение пакетного режима целесообразно в самом начале проведения типологического анализа для общей оценки его результатов. Более детальные результаты получаются при выполнении отдельных режимов с конкретными значениями параметров.

Вывод 2d-семантических сетей классов (БКОСА-10.1.3)

В данном режиме пользователем в диалоге с системой "Эйдос" задаются коды от 3 до 12 классов (ограничение связано с тем, что больше классов не помещается на мониторе при используемом разрешении), а затем на основе данных матрицы сходства классов отображается ориентированный граф, в вершинах которого находятся классы, а ребра соответствуют знаку (красный – "+", синий – "-") и величине (толщина линии) сходства/различия между ними. Посередине каждой линии уровень сходства/различия соответствующих классов отображается в числовой форме (в процентах). Такие графы в данной работе называются 2d-семантическими сетями классов (2d означает "двухмерные").

Когнитивные диаграммы классов (БКОСА-10.3.1, 10.3.2)

В системе "Эйдос" реализован двухоконный интерфейс ввода задания на формирование когнитивных диаграмм и пример такой диаграммы. Переход между окнами осуществляется по клавише "TAB", выбор класса для когнитивной диаграммы – по нажатию клавиши "Enter". В верхней левой части верхнего окна отображаются коды выбранных классов. Генерация и вывод когнитивной диаграммы для заданных классов выполняется по нажатию клавиши F5. Отображаемые диаграммы всегда записываются в виде графических файлов в соответствующие поддиректории. Имеются также пакетные режимы генерации диа-

грамм: генерацию когнитивных диаграмм для полюсов конструкторов (F6), генерация всех возможных когнитивных диаграмм (F7), а также генерация диаграмм Вольфа Мерлина (F8). При задании всех этих режимов имеется возможность задания большого количества параметров, определяющих вид диаграмм и содержание отображаемой на них информации.

Типологический анализ атрибутов обеспечивает: формирование и отображение семантических портретов атрибутов (признаков), а также кластерно-конструктивный и когнитивный анализ атрибутов.

Семантические портреты атрибутов (БКОСА-9.2)

В данном режиме обеспечивается формирование семантического портрета заданного признака и его отображение в текстовой и графической формах. Окно для просмотра текстового отчета имеет прокрутку вправо, что позволяет отобразить количественные характеристики. Графическая диаграмма выводится по нажатию клавиши F5, и может быть непосредственно распечатана или записана в виде графического файла в соответствующую поддиректорию.

Кластерный и конструктивный анализ атрибутов обеспечивает: расчет матрицы сходства признаков; генерация кластеров и конструкторов признаков; просмотр и печать результатов кластерно-конструктивного анализа; автоматическое выполнение перечисленных режимов; отображение результатов кластерно-конструктивного анализа в форме семантических сетей и когнитивных диаграмм.

Расчет матрицы сходства атрибутов (БКОСА-10.2.1)

Стадия выполнения расчета матрицы сходства признаков наглядно отображается на мониторе.

Генерация кластеров и конструкторов атрибутов (БКОСА-10.2.2)

В данном режиме имеется возможность задания ряда параметров, детально определяющих обрабатываемые данные и форму вывода результатов анализа и отображаются результаты кластерно-конструктивного анализа. Имеются также многочисленные возможности манипулирования данными (различные варианты поиска, сортировки и фильтрации).

Автоматическое выполнение режимов 1-2-3. Автоматически реализуются три вышеперечисленные режима.

Вывод 2d-семантических сетей атрибутов (БКОСА-10.2.3)

Результаты кластерно-конструктивного анализа признаков отображаются для заданных признаков в наглядной графической форме семантических сетей.

Когнитивные диаграммы атрибутов (БКОСА-10.4.1, 10.4.2)

Это новый вид когнитивных диаграмм, не встречающийся в литературе. Частным случаем этих диаграмм являются инвертированные диаграммы Вольфа Мерлина (терм. авт.). При их генерации имеется возможность задания ряда параметров, определяющих обрабатываемые данные и форму отображения результатов.

В подсистеме "Анализ" реализованы режимы:

- оценки анкет по шкале лживости;
- измерения внутренней интегральной и дифференциальной валидности модели;
- измерения независимости классов и признаков (стандартный анализ χ^2);

– генерации большого количества разнообразных 2d & 3d графических форм на основе данных матриц абсолютных частот, условных процентных распределений и информативностей (2d & 3d означает: "двухмерные и трехмерные");

– генерации и графического отображения нелокальных нейронов, нейронных сетей, классических и интегральных когнитивных карт.

Оценка достоверности заполнения анкет

В данном режиме исследуются корреляции между ответами в каждой анкете, эти корреляции сравниваются с выявленными на основе всей обучающей выборки и все анкеты ранжируются в порядке уменьшения типичности обнаруженных в них корреляций. Считается, что если корреляции в анкете соответствуют "среднестатистическим", которые принимаются за "норму", то анкета отражает обнаруженные макрозакономерности, если же нет, то возникает подозрение в том, что она заполнена некорректно.

В режиме "Измерение независимости объектов и признаков" реализован стандартный анализ χ^2 , а также рассчитываются коэффициенты Пирсона, Чупрова и Крамера, популярные в экономических, социологических и политологических исследованиях. В системе задание на расчет матриц сопряженности вводится в специальный бланк, который служит также для отображения обобщающих результатов расчетов. На основе этого задания рассчитываются и записываются в форме текстовых файлов одномерные и двумерные матрицы сопряженности для заданных подматриц.

В отличие от матриц сопряженности, выводимых в известной системе SPSS, здесь они выводятся *с текстовыми пояснениями на том языке, на котором сформированы классификационные и описательные шкалы, с констатацией того, обнаружена ли статистически-значимая связь на заданном уровне значимости*. Необходимо также отметить, что в системе "Эйдос" не используются табулированные теоретические значения критерия χ^2 для различных степеней свободы, а необходимые теоретические значения *непосредственно рассчитываются системой, причем со значительно большей точностью, чем они приведены в таблицах* (при этом численно берется обратный интеграл вероятностей).

Режим "Просмотр профилей классов и признаков". Система "Эйдос" текущей версии 12.5 позволяет генерировать и выводить более 54 различных *видов* 2d & 3d графических форм, каждая из которых выводится в форме, определяемой задаваемыми в диалоге параметрами.

Подсистема "Сервис". Реальная эксплуатация ни одной программной системы невозможна либо без тщательного сопровождения эксплуатации и без наличия в системе *развитых средств обеспечения надежности эксплуатации*. В системе "Эйдос" для этого служит подсистема "Сервис" в которой:

- автоматически ведется архивирование баз данных;
- создаются отсутствующие базы данных и индексные массивы;
- распечатываются в текстовые файлы служебные формы, являющиеся основой содержательной информационной модели (базы абсолютных частот, условных процентных распределений и информативностей).

В подсистему "Сервис" входит также интеллектуальная дескрипторная информационно-поисковая система, автоматически генерирующая нечеткие дескрипторы и имеющая интерфейс нечетких запросов на любом естественном языке, использующем кириллицу или латиницу (т.е. не только русском). Отчет по результатам запроса содержит информационные объекты базы данных системы, ранжированные в порядке уменьшения степени соответствия запросу.

1.2.4.3. Технические характеристики и обеспечение эксплуатации системы "ЭЙДОС" (версии 12.5)

1.2.4.3.1. Состав системы "Эйдос": базовая система, системы окружения и программные интерфейсы импорта данных

Система "Эйдос" (текущей версии 12.5) включает базовую систему (система "Эйдос" в узком смысле слова), а также две системы окружения:

- систему комплексного психологического тестирования "Эйдос-Ψ", разработанную совместно с С.Д.Некрасовым [154, 188];
- систему анализа и прогнозирования ситуации на фондовом рынке "Эйдос-фонд", разработанную совместно с Б.Х.Шульман [192].

Данные системы окружения представляют собой программные интерфейсы базовой системы "Эйдос" с базами данных психологических тестов и биржевыми базами данных соответственно, а также выполняют ряд самостоятельных функций по предварительной обработке информации.

Кроме того в саму базовую систему "Эйдос" включены программные интерфейсы с двумя универсальными стандартами баз исходных данных.

1.2.4.3.2. Отличия системы "Эйдос" от аналогов: экспертных и статистических систем

От экспертных систем система "Эйдос" отличается тем, что для ее обучения от экспертов требуется лишь само их решение о принадлежности того или иного объекта или его состояния к определенному классу, а не формулирование правил (продукций) или весовых коэффициентов, позволяющих прийти к такому решению (система генерирует их сама, т.е. автоматически). Дело в том, что часто эксперт не может или не хочет вербализовать, тем более формализовать свои способы принятия решений. Система "Эйдос" генерирует обобщенную таблицу решений непосредственно на основе эмпирических данных и их *оценки* экспертами.

От систем статистической обработки информации система "Эйдос" отличается прежде всего своими целями, которые состоят в следующем: формирование обобщенных образов исследуемых классов распознавания и признаков по

данным обучающей выборки (т.е. обучение); исключение из системы признаков тех из них, которые оказались наименее ценными для решения задач системы; вывод информации по обобщенным образам классов распознавания и признаков в удобной для восприятия и анализа текстовой и графической форме (информационные или ранговые портреты); сравнение распознаваемых формальных описаний объектов с обобщенными образами классов распознавания (распознавание); сравнение обобщенных образов классов распознавания и признаков друг с другом (кластерно-конструктивный анализ); расчет частотных распределений классов распознавания и признаков, а также двумерных матриц сопряженности на основе критерия χ^2 и коэффициентов Пирсона, Чупрова и Крамера; результаты кластерно-конструктивного и информационного анализа выводятся в форме семантических сетей и когнитивных диаграмм. Система "Эйдос" в универсальной форме автоматизирует базовые когнитивные операции системного анализа, т.е. является инструментарием СК-анализа. Таким образом, система "Эйдос" выполняет за исследователя-аналитика ту работу, которую при использовании систем статистической обработки ему приходится выполнять вручную, что чаще всего просто невозможно при реальных размерностях данных. Поэтому система "Эйдос" и называется универсальной когнитивной аналитической системой.

1.2.4.3.3. Некоторые количественные характеристики системы "Эйдос"

Система "Эйдос" обеспечивает генерацию и запись в виде файлов более 54 видов 2d & 3d графических форм и 50 видов текстовых форм, перечень которых приведен в таблице 31.

При применении системы в самых различных предметных областях обеспечивается достоверность распознавания обучающей выборки: на уровне 90% (интегральная валидность), которая существенно повышается после Парето-оптимизации системы признаков (т.е. после исключения признаков с низкой селективной силой), удаления из модели артефактов, а также классов и признаков, по которым недостаточно данных. Система "Эйдос" версии 12.5 обеспечивает синтез модели, включающей десятки тысяч классов и признаков при неограниченном объеме обучающей выборки, причем признаки могут быть не только качественные (да/нет), но и количественные, т.е. числовые. В некоторых режимах анализа модели имеются ограничения на ее размерность, которые на данном этапе преодолеваются путем оптимизации модели. Реализована возможность разработки супертестов, в том числе интеграции стандартных тестов в свою среду, (при этом не играют роли известны ли методики интерпретации, т.е. "ключи" этих тестов). В системе имеется научная графика, обеспечивающая высокую степень наглядности, а также естественный словесный интерфейс при обучении Системы и запросах на распознавание.

Исходные тексты системы "Эйдос" и систем окружения "Эйдос-У" и "Эйдос-фонд" в формате "Текст-DOS" имеют объем около 2.5 Мб; их распечатка 6-м шрифтом составляет около 800 страниц.

Таблица 31 – ПЕРЕЧЕНЬ ТЕКСТОВЫХ ФОРМ СИСТЕМЫ "ЭЙДОС"

ПАРАМЕТРЫ ФОРМАТИРОВАНИЯ ВЫХОДНЫХ ФОРМ ДЛЯ РЕДАКТОРА WINWORD

03-07-02 15:26:14

г.Краснодар

п/п	Имя файла выходной формы	Ко-во строк	Courier New Cyr	Т и п листа	Н а и м е н о в а н и е ф о р м ы
1	Abs.txt	65	6	АЛЬБОМНЫЙ=>	Матрица абсолютных частот
2	Anketa.txt	59	10	Книжный	Анкета
3	AnkObKod.txt	0	10	Книжный	Анкеты обучающей выборки с кодами признаков
4	AnkObNam.txt	0	10	Книжный	Анкеты обучающей выборки с наим. признаков
5	AnkObZer.txt	0	10	Книжный	Анкеты обучающей выборки - пустографики
6	Ank.txt	0	9	Книжный	Распределение анк.об.выб.по классам распознавания
7	Arep_it.txt	65	9	Книжный	Итог ремонта обучающей выборки (взвешивания данных)
8	Arep_zad.txt	53	8	АЛЬБОМНЫЙ	Задание на ремонт обучающей выборки - част.распр.ГС
9	Dost_ank.txt	53	8	АЛЬБОМНЫЙ	Оценка анкет об.выборки по =шкале лживости=
10	Inf.txt	65	6	АЛЬБОМНЫЙ=>	Матрица информативностей
11	Ipo_####.txt	65	6	АЛЬБОМНЫЙ	Информационный портрет объекта-класса с кодом ####
12	Ipo_blf.txt	59	10	Книжный	Задание на формирование инф.портретов классов
13	Ipo_blz.txt	59	10	Книжный	Бланк задания на формирование инф.портретов классов
14	Ipp_####.txt	65	6	АЛЬБОМНЫЙ	Информационный портрет признака с кодом ####
15	Ips_obj.txt	59	10	Книжный	Обучающая база данных дискрипторной ИПС
16	Ips_rsi.txt	59	10	Книжный	Распознаваемая база данных дискрипторной ИПС
17	Klas_obj.txt	74	8	Книжный	Кластеры и конструкты объектов (классов)
18	Klas_prp.txt	74	8	Книжный	Кластеры и конструкты признаков
19	Matr_sf.txt	43	10	АЛЬБОМНЫЙ	Задание и сводная форма матриц сопряженности X^2
20	Matr_spl.txt	48	9	АЛЬБОМНЫЙ	Одномерные частотные распределения X^2
21	Matr_sp2.txt	0	6	АЛЬБОМНЫЙ	Двухмерные частотные распределения X^2
22	Matr_sz.txt	0	10	Книжный	Бланк задания на формирование матриц сопряженности
23	Object.txt	84	7	Книжный	Справочник классов распознавания
24	Objs_int.txt	59	10	Книжный	Интерпретация инф.портретов классов (объектов)
25	Opt_ank.txt	65	6	АЛЬБОМНЫЙ	Вопросы в порядке убывания селективной силы
26	Opt_obj.txt	65	6	АЛЬБОМНЫЙ	Классы в порядке убывания сформированности
27	Perc.txt	65	6	АЛЬБОМНЫЙ=>	Матрица условных процентных распределений
28	Port_obj.txt	65	6	АЛЬБОМНЫЙ	Информационные портреты классов (объектов)
29	Port_prp.txt	65	6	АЛЬБОМНЫЙ	Информационные портреты признаков
30	Priz_int.txt	59	10	Книжный	Справочник интерпретации первичных признаков
31	Priz_ob.txt	59	7	АЛЬБОМНЫЙ	Справочник обобщенных признаков (шкал, вопросов)
32	Priz_per.txt	84	7	Книжный	Справочник первичных признаков (градации, ответов)
33	Rasp_it1.txt	59	7	АЛЬБОМНЫЙ	Итоговая форма по результатам распознавания №1
34	Rasp_it2.txt	59	7	АЛЬБОМНЫЙ	Итоговая форма по результатам распознавания №2
35	RspKart1.txt	0	10	АЛЬБОМНЫЙ	Карта результатов идентификации инф.источника
36	RspKart2.txt	0	9	Книжный	Карта результатов идентификации инф.источников
37	RsAnkKod.txt	0	10	Книжный	Анкеты распознаваемой выборки с кодами ответов
38	RsAnkNam.txt	0	10	Книжный	Анкеты расп.выборки с наименованиями ответов
39	RsAnkZer.txt	0	10	Книжный	Анкеты-пустографики расп.выборки для заполнения
40	Stat_obj.txt	65	6	АЛЬБОМНЫЙ	Статистическая характеристика обучающей выборки
41	ValidSys.txt	65	6	АЛЬБОМНЫЙ	Отчет по интегральной и дифференциальной валидности
42	Err_obj.txt	65	9	Книжный	Протокол ошибок обучающей выборки
43	Merl_att.txt	50	14	Книжный	Наименования уровней иерархии признаков
44	Merl_obj.txt	50	14	Книжный	Наименования уровней иерархии классов
45	Ipp_blf.txt	59	10	Книжный	Задание на формирование инф.портретов признаков
46	Ipp_blz.txt	59	10	Книжный	Бланк задания на формиров-е инф.портретов признаков
47	Korr_obj.txt	59	10	АЛЬБОМНЫЙ	Корреляционная матрица классов
48	Korr_att.txt	59	10	АЛЬБОМНЫЙ	Корреляционная матрица признаков
49	Rel_obj.txt	59	10	АЛЬБОМНЫЙ	Матрица связей информационных портретов классов
50	Rel_att.txt	59	10	АЛЬБОМНЫЙ	Матрица связей информационных портретов признаков

1.2.4.3.4. Обеспечение эксплуатации системы "Эйдос"

Универсальная когнитивная аналитическая система "Эйдос" представляет собой программную систему, и для ее эксплуатации, как и для эксплуатации любой программной системы, необходима определенная инфраструктура. Без инфраструктуры эксплуатации любая программная система остается лишь файлом, записанным на винчестере. В зависимости от масштабности решаемых задач управления и специфики предметной области данная структура может быть как довольно малочисленной, так и более развитой. Однако в любом случае ее основные функциональные и структурные характеристики остаются примерно одними и теми же. Кратко рассмотрим эту инфраструктуру на примере гипотетической организации, производящей определенные виды продукции.

Основная цель: обеспечивать информационную и аналитическую поддержку деятельности организации, направленную на производство запланированного объема продукции заданного качества, достижение высокой эффективности управления и устойчивого поступательного развития.

Данная основная цель предполагает выполнение информационных и аналитических работ с различными объектами деятельности, находящимися на различных структурных уровнях как самой организации, так и ее окружения: персональный уровень; уровень коллективов (подразделений); уровень организации в целом; окружающая среда (непосредственное, региональное, международное окружение). Для достижения основной цели для каждого класса объектов должны регулярно выполняться следующие работы: оценка (идентификация) текущего состояния с накоплением данных (мониторинг); прогнозирование развития (оперативное, тактическое и стратегическое); выработка рекомендаций по управлению. *Необходимо особо подчеркнуть, что основная цель может быть достигнута только при условии соблюдения вполне определенной наукоемкой технологии, основы которой изложены в данном исследовании.*

Задачи, решаемые для достижения цели работы:

1. Мониторинг: оценка и идентификация текущего (фактического, актуального) состояния объекта управления; накопление данных идентификации в базах данных в течение длительного времени.

2. Анализ: выявление причинно-следственных зависимостей путем анализа данных мониторинга.

3. Прогнозирование: оперативное, тактическое и стратегическое прогнозирование развития объекта управления и окружающей среды путем использования закономерностей, выявленных на этапе анализа данных мониторинга.

4. Управление: анализ взаимодействия объекта управления с окружающей средой и выработка рекомендаций по управлению.

Таким образом, по мнению автора, *управление является высшей, существующей на данный момент формой обработки информации.*

Для достижения основной цели и решения задач управления необходимо выполнять работы по следующим направлениям: регулярное получение исходной информации о состоянии объекта управления; обработка исходной информации на компьютерах; анализ обработанной информации, прогнозирование развития объекта управления, выработка рекомендаций по оказанию управляю-

щих воздействий на объект управления; разработка и применение (или предоставление рекомендаций заказчикам) различных методов оказания управляющих воздействий на объект управления.

Для этого необходима определенная организационная структура: *научно-методический отдел* включает: научно-методический сектор; сектор разработки программного обеспечения; сектор внедрения и сопровождения программного обеспечения; сектор организационного и юридического обеспечения; *отдел мониторинга*: сектор исследования объекта управления; сектор по работе с независимыми экспертами; сектор по взаимодействию с поставляющими информацию организациями; сектор по анализу информации общего пользования; *отдел обработки информации*: сектор ввода исходной информации (операторы); сектор сетевых технологий и Internet; сектор внедрения, эксплуатации и сопровождения программных систем; сектор технического обслуживания компьютерной техники; сектор ведения архивов баз данных по проведенным исследованиям; *аналитический отдел* имеет структуру, обеспечивающую компетентный профессиональный анализ результатов обработки данных мониторинга по объектам, которые приняты для контроля и управления.

Для выполнения работ по этим направлениям необходимо определенное обеспечение деятельности: техническое, программное, информационное, организационное, юридическое и кадровое. Детально подобная структура и виды обеспечения ее деятельности описаны в работе [125].

1.2.4.4. АСК-анализ, как технология синтеза и эксплуатации рефлексивных АСУ активными объектами

Применение АСК-анализа обеспечивает выявление информационных зависимостей между факторами различной природы и будущими состояниями объекта управления, т.е. позволяет осуществить синтез содержательной информационной модели, а фактически – осуществить синтез АСУ. Применение АСК-анализа в составе АСУ обеспечивает ее эксплуатацию в режиме непрерывной адаптации модели (на детерминистских этапах), а когда это необходимо (т.е. после прохождения точек бифуркации) – и ее нового синтеза.

Ниже приведена технология системы "Эйдос" как инструментария АСК-анализа:

Шаг 1-й: формализация предметной области (БКОСА-1): разработка описательных и классификационных шкал и градаций, необходимых для формализованного описания предметной области. Описательные шкалы описывают факторы различной природы, влияющие на поведение АОУ, а классификационные – все его будущие состояния, в том числе целевые.

Шаг 2-й: формирование обучающей выборки (БКОСА-2): информация о состоянии среды и объекта управления, а также вариантах управляющих воздействий поступает на вход системы. Работа по преобразованию этой информации в формализованный вид (т.е. кодирование) осуществляется специалистами, обслуживающими систему с использованием описательных и классификационных шкал. Вся эта информация представляется в виде специальных кодированных

бланков, используемых также для ввода информации в компьютер. В результате ее формируется так называемая "обучающая выборка".

Шаг 3-й: обучение (БКОСА-3): обучающая выборка обрабатывается обучающим алгоритмом, на основе чего им формируются решающие правила (обобщенные образы состояний АОУ, отражающие весь спектр будущих возможных состояний объекта управления) и определяется ценность факторов для решения задач подсистем идентификации, мониторинга, прогнозирования и выработки управляющих воздействий.

Шаг 4-й: оптимизация (БКОСА-4): факторы, не имеющие особой прогностической ценности, корректным способом удаляются из системы. Данный процесс осуществляется с помощью итерационных алгоритмов, при этом обеспечивается выполнение ряда ограничений, таких как результирующая размерность пространства факторов, его информационная избыточность и т.п.

Шаг 5-й: верификация модели (БКОСА-5): выполняется после каждой адаптации или пересинтеза модели. На этом шаге обучающая выборка копируется в распознаваемую и осуществляется ее автоматическая классификация (в режиме распознавания). Затем рассчитываются так называемые внутренняя дифференциальная и интегральная валидности, характеризующие качество решающих правил.

Шаг-6: принятие решения об эксплуатации модели или ее пересинтезе. Если результаты верификации модели удовлетворяют разработчиков РАСУ АО, то она переводится из пилотного (экспериментального) режима, при котором управляющие решения генерировались, но не исполнялись, в режим экспериментальной эксплуатации, а затем и опытно-производственной эксплуатации, когда они реально начинают использоваться для управления. Иначе, т.е. если же модель признана недостаточно адекватной, то необходимо осуществить ее пересинтез, начиная с шага 1. При этом используются следующие приемы: расширение набора факторов, т.к. значимые факторы могли не войти в модель; увеличение объема обучающей выборки, т.к. существенные примеры могли не войти в обучающую выборку; исключение артефактов, т.к. в модель могли вкрасться существенно искажающие ее не подтверждающиеся данные; пересмотр экспертных оценок и, если необходимость этого возникает систематически, то и перестроение экспертного совета, т.к. причиной этого могла быть некомпетентность экспертов; объединение некоторых классов, т.к. по ним недостаточно данных; разделение некоторых классов, т.к. по ним слишком высокая вариабельность объектов по признакам, и т.д.

Шаг 7-й: идентификация и прогнозирование состояния АОУ (БКОСА-7).

Шаг 8-й: оценка качества идентификации состояния АОУ. Если качество идентификации высокое, то состояние АОУ рассматривается как типовое, а значит причинно-следственные взаимосвязи между факторами и будущими состояниями данного объекта управления считаются адекватно отраженными в модели и известными (т.е. если качество идентификации высокое, то считается, что объект относится к генеральной совокупности, по отношению к которой обучающая выборка репрезентативна). Поэтому в этом случае осуществляется переход на Шаг-9 (выработка управляющего воздействия и последующий анализ). Иначе – считается, что на вход системы идентификации попал объект, не относящийся к

генеральной совокупности, адекватно представленной обучающей выборкой. Поэтому в этом случае информация о нем поступает на Шаг-13, начиная с которого запускается процедура пересинтеза модели, что приводит к расширению генеральной совокупности, представленной обучающей выборкой.

Шаг 9-й: выработка решения об управляющем воздействии (БКОСА-9) путем решения обратной задачи прогнозирования [81].

Шаг 10-й: типологический анализ классов и факторов (БКОСА-10): кластерно-конструктивный и когнитивный анализ, семантические сети, когнитивные диаграммы состояний АОУ и факторов [81].

Шаг 11-й: многофакторное планирование и принятие решения о применении системы управляющих факторов (БКОСА-11).

Шаг 12-й: оценка адекватности принятого решения об управляющих воздействиях: если АОУ перешел в заданное целевое состояние, то осуществляется переход на вход адаптации содержательной информационной модели (Шаг- 2): в подсистеме идентификации предусмотрен режим дополнения распознаваемой выборки к обучающей, чтобы в последующем, когда станут известны результаты управления, этой верифицированной (т.е. достоверной) оценочной информацией дополнить обучающую выборку и переформировать решающие правила (обучающая обратная связь). Иначе, т.е. если АОУ не перешел в заданное целевое состояние, переход на вход пересинтеза модели (Шаг-1), при этом могут быть изменены и описательные, и классификационные (оценочные) шкалы, что позволяет качественно расширить сферу адекватного функционирования РАСУ АО.

Шаг 13-й (неформализованный поиск нетипового решения об управляющем воздействии и подготовка данных для пересинтеза модели, как в случае, если решения оказалось удачным, так и в противном случае).

Таким образом, предложена технология применения системы "Эйдос" как инструментария применения АСК-анализа, основанного на системной теории информации, ориентированной на синтез рефлексивных АСУ АО. В процессе эксплуатации системы "Эйдос" успешно решаются все задачи АСК-анализа: формирование обобщенных образов состояний АОУ на основе обучающей выборки (обучение); идентификация состояний АОУ на основе его параметров (распознавание); определение влияния входных параметров на перевод АОУ в различные будущие состояния (обратная задача прогнозирования); прогнозирование поведения АОУ в условиях полного отсутствия управляющих воздействий; прогнозирование поведения АОУ при различных вариантах многофакторных управляющих воздействий.

Кроме того, выявленные в результате работы рефлексивной АСУ причинно-следственные зависимости между факторами различной природы и будущими состояниями объекта управления позволяют, при условии неизменности этих закономерностей в течение достаточно длительного времени, построить АСУ с постоянной моделью классического типа.

Резюме

1. Создан программный и методический инструментарий СК-анализа – Универсальная когнитивная аналитическая система "Эйдос", защищенная 8 Свидетельствами РосПатента РФ [185-192].

2. Продemonстрировано, что система "Эйдос" на базе формализуемой когнитивной концепции успешно реализует предложенную семантическую информационную модель и алгоритмы базовых когнитивных операций системного анализа, и, таким образом, является специальным программным инструментарием для решения проблемы и задач, поставленных в данном исследовании. Показано, что *система "Эйдос" является эффективным инструментарием СК-анализа*. В функциях и структуре системы "Эйдос" нашли воплощение фундаментальные закономерности познания, связанные с функциональной асимметрией мозга и знаковых систем.

3. Подробно описана технология синтеза и эксплуатации приложений системы "Эйдос", видеогаммы ее пользовательского интерфейса и технические характеристики (текущая версия системы "Эйдос-12.5" позволяет обрабатывать десятки тысяч будущих состояний АОУ и градаций факторов).

4. Технология разработки приложения в системе "Эйдос" включает: когнитивную структуризацию и формализацию предметной области (подсистема "Словари"); синтез модели (подсистема "Обучение"); оптимизацию модели (подсистема "Оптимизация"); верификацию модели (оценка степени адекватности, скорости сходимости и семантической устойчивости модели).

5. Технология эксплуатации приложения системы "Эйдос" в режиме адаптации и синтеза модели включает: идентификацию и прогнозирование (подсистема "Распознавание"); кластерно-конструктивный, семантический и когнитивный анализ (подсистема "Типология"); анализ достоверности, валидности, независимости (подсистема "Анализ"); средства и инструменты обеспечения надежности эксплуатации (подсистема "Сервис").

5. Система "Эйдос" является большой системой. Листинг ее исходных текстов (вместе с системами окружения "Эйдос-фонд" и "Эйдос-Ψ") 6-м шрифтом составляет около 800 страниц, в процессе работы система оперирует десятками баз данных (около 70).

6. Разработана инфраструктура применения системы "Эйдос", детализированы организационные, юридические, экономические, технические и другие аспекты информационной технологии применения данной системы для решения задач синтеза рефлексивных АСУ активными объектами и эксплуатации этих АСУ в режиме адаптации и периодического синтеза модели.

7. Разработана технология синтеза рефлексивных АСУ активными объектами и методики ее применения в конкретных предметных областях. Предложено рассматривать алгоритм СК-анализа, как алгоритм синтеза рефлексивных АСУ активными. Предложено рассматривать СК-анализ, как технологию синтеза РАСУ АО и их эксплуатации в режиме адаптации и периодического синтеза модели:

шаг 1-й: формализация предметной области (БКОСА-1);

шаг 2-й: формирование обучающей выборки (БКОСА-2);

шаг 3-й: обучение (БКОСА-3);

шаг 4-й: оптимизация (БКОСА-4);

шаг 5-й: верификация модели (БКОСА-5);

шаг 6-й: принятие решения об эксплуатации модели или ее пересинтезе;

шаг 7-й: идентификация и прогнозирование состояния АОУ (БКОСА-7);

- шаг 8-й: оценка качества идентификации состояния АОУ;
- шаг 9-й: выработка решения об управляющем воздействии (БКОСА-9);
- шаг 10-й: типологический анализ классов и факторов (БКОСА-10);
- шаг 11-й: многофакторное планирование и принятие решения о применении системы управляющих факторов (БКОСА-11);
- шаг 12-й: оценка адекватности принятого решения об управляющих воздействиях;
- шаг 13-й: (неформализованный поиск нетипового решения об управляющем воздействии и подготовка данных для пересинтеза модели, причем как в случае, если решение оказалось удачным, так и в противном случае).

Предложенные концептуальные подходы, технология и методики их использования могут быть эффективно применены при проектировании и эксплуатации рефлексивных АСУ в АПК, а также в других предметных областях.

Контрольные вопросы

1. Назначение и состав системы "Эйдос".
2. Цели и основные функции системы "Эйдос".
3. Обобщенная структура системы "Эйдос".
4. Пользовательский интерфейс, технология разработки и эксплуатации приложений в системе "ЭЙДОС".
5. Начальный этап синтеза модели: когнитивная структуризация и формализация предметной области, подготовка исходных данных (подсистема "Словари") (БКОСА-1, БКОСА-2).
6. Синтез модели: пакетное обучение системы распознавания (подсистема "Обучение") (БКОСА-3).
7. Оптимизация модели (подсистема "Оптимизация") (БКОСА-4).
8. Верификация модели (оценка ее адекватности) (БКОСА-5).
9. Эксплуатация приложения в режиме адаптации и периодического синтеза модели (БКОСА-7, БКОСА-9, БКОСА-10).
10. Технические характеристики и обеспечение эксплуатации системы "ЭЙДОС" (версии 12.5).
11. Состав системы "Эйдос": Базовая система, системы окружения и программные интерфейсы импорта данных.
12. Отличия системы "Эйдос" от аналогов: экспертных и статистических систем.
13. Некоторые количественные характеристики системы "Эйдос".
14. Обеспечение эксплуатации системы "Эйдос".
15. АСК-анализ, как технология синтеза и эксплуатации рефлексивных АСУ активными объектами.

Рекомендуемая литература

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

РАЗДЕЛ 1.3. ПРИНЦИПЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

1.3.1. ЛЕКЦИЯ-7. Системы с интеллектуальной обратной связью и интеллектуальными интерфейсами

Учебные вопросы

1. *Интеллектуальные интерфейсы. Использование биометрической информации о пользователе в управлении системами.*
2. *Системы с биологической обратной связью.*
3. *Системы с семантическим резонансом. Компьютерные (Ψ-технологии и интеллектуальный подсознательный интерфейс.*
4. *Системы виртуальной реальности и критерии реальности. Эффекты присутствия, деперсонализации и модификация сознания пользователя.*
5. *Системы с дистанционным телекинетическим интерфейсом.*

1.3.1.1. Интеллектуальные интерфейсы. Использование биометрической информации о пользователе в управлении системами

1.3.1.1.1. Идентификация и аутентификация личности по почерку. Понятие клавиатурного почерка

Рассмотрим, в чем заключается различие между двумя формами представления одного и того же текста: рукописной и печатной. При этом могут исследоваться и сравниваться как сам процесс формирования текста, так и его результаты, т.е. уже сформированные тексты.

При исследовании уже сформированных текстов обнаруживается, что главное отличие рукописного текста от печатного состоит в значительно большей степени вариабельности начертаний одной и той же буквы разными людьми и одним и тем же человеком в различных состояниях, чем при воспроизведении тех же букв на различных пишущих машинках и принтерах.

Почерком будем называть систему индивидуальных особенностей начертания и динамики воспроизведения букв, слов и предложений вручную различными людьми или на различных устройствах печати.

В рукописной форме начертание букв является индивидуальным для каждого человека и зависит также от его состояния, хотя, конечно, в начертаниях каждой конкретной буквы всеми людьми безусловно есть и нечто общее, что и позволяет идентифицировать ее именно как данную букву при чтении.

К индивидуальным особенностям рукописного начертания букв в работе [171] отнесено 13 шкал с десятками градаций в каждой.

В печатной форме вариабельность начертания букв значительно меньше, чем в рукописной, но все же присутствует, особенно на печатных машинках, барабанных, знаковосинтезирующих и литерных принтерах.

В СССР печатные машинки при продаже регистрировались и образец печати всех символов вместе с паспортными данными покупателя направлялся в "компетентные" органы. Это позволяло установить на какой машинке и кем напечатан тот или иной материал. Считается, что принтер тем лучше, чем меньше у него индивидуальных особенностей, т.е. чем ближе реально распечатываемые им тексты к некоторому идеалу – стандарту. Современные лазерные и струйные принтеры в исправном состоянии (новый барабан и картридж) практически не имеют индивидуальных особенностей.

На современных компьютерах основным устройством ввода текстовой информации является клавиатура. Результат ввода текста в компьютер с точки зрения начертания букв, слов и предложений не имеет особых индивидуальных особенностей (если не считать частот использования различных шрифтов, кеглей, жирностей, подчеркиваний и других эффектов, изменяющих вид текста). Поэтому необходимо ввести понятие клавиатурного почерка, под которым будем понимать систему индивидуальных особенностей начертаний и динамики воспроизведения букв, слов и предложений на клавиатуре.

1.3.1.1.2. Соотношение психографологии и атрибуции текстов

Таким образом, *любой текст содержит не только ту информацию, для передачи которой его собственно и создавали, но и информацию о самом авторе этого текста и о технических средствах и технологии его создания.*

Существует целая наука – "*Психографология*", которая ставит своей задачей получение максимального количества информации об авторах текстов на основе изучения индивидуальных особенностей их почерка [171].

В настоящее время в России действует институт графологии. На сайте этого института <http://graphology.boom.ru> можно познакомиться с тем, что такое графология, с ее историей и задачами, которые она позволяет решать сегодня. Графологическое исследование имеет значительное преимущество перед простым тестированием или собеседованием, поскольку нет необходимости информировать человека, чей почерк подвергается изучению о производимых исследованиях.

Но текст представляет собой не просто совокупность букв, а сложную иерархическую структуру, в которой буквы образуют лишь фундамент пирамиды, а на более высоких ее уровнях находятся слова, предложения, и другие части текстов различных размеров, обладающие относительной целостностью и самостоятельностью (абзацы, параграфы, главы, части, книги).

Понятие почерка акцентирует внимание именно на начертании и динамике воспроизведения *букв и слов*. При этом в понятие почерка не входят индивидуальные особенности текстов, обнаруживаемые на более высоких уровнях иерархической организации текстов, например: частоты употребления тех или иных слов и словосочетаний, средние длины предложений и абзацев, и т.п. Но именно эти индивидуальные особенности текстов исследуются и используются

при *атрибуции* анонимных и псевдонимных текстов (определении их вероятного авторства) и датировки.

Соответственно и текст может представлять для читателя интерес по крайней мере с трех точек зрения:

1. Как источник информации о том, о чем говорит автор, т.е. о предмете изложения.

2. Как источник информации о самом авторе.

3. Как источник информации о предмете изложения и об авторе.

В этом смысле читать Пушкина в рукописи может быть значительно интереснее, чем в взяв томик с полки. Это объясняется просто: в томике есть лишь сам *результат* работы поэта и выхолощена вся информация о *процессе*, т.е. о самом поэте, содержащаяся в почерке, способе размещения текста на листе, порядке и динамике его формирования, различных вариантах и ассоциациях, возникавших в процессе создания произведения.

Таким образом, система, оснащенная интеллектуальным интерфейсом, может вести по-разному в зависимости от результатов идентификации пользователя, его профессионального уровня и текущего психофизиологического состояния.

1.3.1.1.3. Идентификация и аутентификация личности пользователя компьютера по клавиатурному почерку

Рассмотрим подробнее некоторые вопросы идентификации пользователей по клавиатурному почерку. При этом мы будем *самым существенным образом* основываться на работе: Завгородний В.В. и Мельников Ю.Н. Идентификация по клавиатурному почерку, "Банковские Технологии" №9, 1998 [43].

Проблемы идентификации и аутентификации пользователей компьютеров являются актуальными в связи с все большим распространением компьютерных преступлений. Использование для идентификации клавиатурного почерка является одним из направлений биометрических методов идентификации личности.

Подобные системы не обеспечивают такую же точность распознавания, как системы идентификации по отпечаткам пальцев или по рисунку радужной оболочки глаз, но имеют то преимущество, что система может быть полностью скрыта от пользователя, т. е. он может даже не подозревать о наличии такой системы контроля доступа.

Варианты постановки задачи распознавания клавиатурного почерка

Аутентификация – это проверка, действительно ли пользователь является тем, за кого себя выдает. При этом пользователь должен предварительно сообщить о себе идентификационную информацию: свое имя и пароль, соответствующий названному имени.

Идентификация – это установление его личности.

И идентификация, и аутентификация являются типичными задачами распознавания образов, которое может проводиться по заранее определенной или произвольной последовательности нажатий клавиш.

Характеристики клавиатурного почерка

При вводе информации пользователь последовательно нажимает и отпускает клавиши, соответствующие вводимому тексту. При этом для каждой нажимаемой клавиши можно фиксировать моменты нажатия и отпускания.

На IBM-совместимых персональных компьютерах на следующую клавишу можно нажимать до отпускания предыдущих, т.е. символ помещается в буфер клавиатуры только по нажатию клавиши, тогда как аппаратные прерывания от клавиатуры возникают и при нажатии, и при отпускании клавиши.

Основной характеристикой клавиатурного почерка следует считать временные интервалы между различными моментами ввода текста:

- между нажатиями клавиш;
- между отпусканиями клавиш;
- между нажатием и отпусканием одной клавиши;
- между отпусканием предыдущей и нажатием следующей клавиши.

Кроме того, могут учитываться производные от временных интервалов *вторичные показатели*, например такие как скорость и ускорение ввода.

Математические методы распознавания клавиатурного почерка

В литературе описано четыре математических подхода к решению задачи распознавания клавиатурного почерка пользователя ЭВМ:

- статистический;
- вероятностно-статистический;
- на базе теории распознавания образов и нечеткой логики;
- на основе нейросетевых алгоритмов.

За более подробным анализом подходов к использованию компьютерного почерка для идентификации и аутентификации пользователей компьютеров отсылаем к в вышеназванной статье.

1.3.1.1.4. Прогнозирование ошибок оператора по изменениям в его электроэнцефалограмме

В 2004 году Институтом психологии РАН, Институтом Высшей Нервной Деятельности и Нейрофизиологии РАН, Высшей Школой Экономики и Кубанским государственным аграрным университетом (Щукин Т.Н., Дорохов В. Б., Лебедев А.Н., Луценко Е.В.) проведены исследования, продемонстрировавшие принципиальную возможность прогнозирования ошибок оператора при работе с клавиатурой, типа "ошибочное нажатие клавиши", "ошибочное ненажатие клавиши" и т.п. по изменениям в его электроэнцефалограмме (ЭЭГ). При этом для обработки информации успешно была применена система "Эйдос" [152, 269-271].

Эти работы в перспективе позволяют создать интеллектуальные *высоко-надежные* интерфейсы, обеспечивающие решение этих и ряда других задач идентификации и прогнозирования состояния оператора *в режиме реального времени, т.е. непосредственно в процессе его работы с системой*. При этом система в своей работе будет гибко учитывать текущее и прогнозируемое состояние оператора, что может проявляться в адаптации как алгоритмов работы, так и вида и содержания интерфейса.

Эти работы дополняют возможности *заблаговременного* отбора операторов, обладающих свойствами, необходимыми для высоко ответственных работ в экстремальных ситуациях [77-79, 81, 87, 93, 97, 101, 102, 116-118, 125, 143, 148, 150, 151, 152, 154, 155, 187, 188, 219, 222, 224, 227, 252, 269-271].

1.3.1.2. Системы с биологической обратной связью

1.3.1.2.1. Общие положения

Системами с биологической обратной связью (БОС) будем называть системы, поведение которых зависит от психофизиологического (биологического) состояния пользователя.

Это означает, что в состав систем с БОС в качестве подсистем входят информационно-измерительные системы и системы искусственного интеллекта.

Съем информации о состоянии пользователя осуществляется с помощью контактных и/или дистанционных датчиков в режиме реального времени с применением транспьютерных или обычных карт (плат) с аналого-цифровыми преобразователями (АЦП).

*При этом информация может сниматься по большому количеству каналов – показателей (количество которых обычно кратно степеням двойки), подавляющее большинство которых обычно являются *несознаваемыми* для пользователя. Это является весьма существенным обстоятельством, т.к. означает, что системы БОС позволяют вывести на уровень сознания обычно ранее не осознаваемую информацию о состоянии своего организма, т.е. расширить область осознаваемого. А это значит, что у человека появляются условия, обеспечивающие возможность сознательного управления своими состояниями, ранее не управляемыми на сознательном уровне, что является важным эволюционным достижением технократической цивилизации.*

Передача информации от блока съема информации к АЦП-карте может также осуществляться либо по проводной связи, либо дистанционно с использованием каналов инфракрасной или радиосвязи.

Приведем три примера применения подобных систем:

1. Мониторинг состояния сотрудников на конвейере с целью обеспечения высокого качества продукции.
2. Компьютерные тренажеры, основанные на БОС, для обучения больных с функциональными нарушениями управлению своим состоянием.
3. Компьютерные игры с БОС.

1.3.1.2.1. Мониторинг состояния сотрудников сборочного конвейера с целью обеспечения высокого качества продукции

Известно, что одной из основных причин производственного брака является ухудшение состояния сотрудников. Но сотрудники не всегда могут вовремя заметить это ухудшение, т.к. самооценка (самочувствие) обычно запаздывает по времени за моментом объективного ухудшения состояния. Поэтому является актуальным своевременное обнаружение объективного ухудшения параметров и адекватное реагирование на него.

С помощью систем БОС это достигается тем, что:

1. Каждому сотруднику одевается на руку браслет с компактным устройством диагностики ряда параметров, например таких, как:

- частота и наполнение пульса;
- кожно-гальваническая реакция;
- температура;
- давление;
- пототделение.

2. Это же устройство и периодически передает значения данных параметров на компьютер по радиоканалу.

3. Параметры от каждого сотрудника накапливаются в базе данных системы мониторинга на сервере, а также анализируются в режиме реального времени с учетом текущего состояния и динамики, в т.ч. вторичных (расчетных) показателей.

4. Когда параметры выходят за пределы коридора "нормы" или по их совокупности может быть поставлен диагноз, – сотрудник оперативно снимается с рабочего места и заменяется другим из резерва, а затем, при наличии показаний, направляется на лечение.

1.3.1.2.2. Компьютерные тренажеры, основанные на БОС, для обучения больных навыкам управления своим состоянием

Некоторыми процессами в своем организме мы не можем управлять не потому, что у нас нет рычагов управления, а лишь потому, что мы их не знаем, не имеем навыков их использования и *не знаем результатов их применения*. Но ключевой проблемой, без решения которой невозможно управление, является отсутствие быстрого и надежного, адекватного по содержанию канала обратной связи.

Все эти проблемы снимаются системами БОС:

- на экран компьютера в наглядной и легко интерпретируемой форме в режиме реального времени выводится информация о состоянии какой-либо подсистемы организма, например, об уровне pH (кислотности) в желудке;
- в качестве рычагов управления пациенту предлагается применить метод визуализации тех или иных образов, которые сообщаются врачом;
- когда пациент ярко зрительно представляет заданные образы, то при этом он обнаруживает, что кривая кислотности на экране начинает ползти вверх или вниз в прямом соответствии с тем, что именно он себе представляет.

Через пару недель подобных тренировок, проводимых по 15-20 минут через день пациент приобретает такой уровень навыков управления ранее не осознаваемыми процессами в своем организме, которых Хатха-йоги добиваются за многие годы упорных тренировок под руководством профессиональных опытных и ответственных наставников (Гуру). Причем скоро пациент начинает понимать, когда необходимо повысить или понизить кислотность и без компьютера с системой БОС и может делать это прямо в той обстановке, в которой возникла такая необходимость. Столь высокая эффективность метода БОС объясняется высокой скоростью, наглядностью и адекватностью обратной связи, что яв-

ляется одним из основных факторов, влияющих на эффективность формирования навыков управления своим состоянием.

Имеется информация, что такими методами могут лечиться или облегчаться многие заболевания, вплоть до диабета, причем не только на стадии функциональных нарушений, но даже и при наличии органических изменений.

1.3.1.2.3. Компьютерные игры с БОС

В последнее время появляется все больше компьютерных игр, включающих элементы БОС. При этом от психофизиологического состояния игрока может зависеть, например, и развитие сценария, и точность прицеливания при использовании оптического прицела.

В этих играх часто создаются ситуации, в которых человеку нужно быстро принимать и реализовать решения, при этом цена ошибки, а значит и психическая напряженность, и волнение игрока, постоянно увеличиваются. Этим самым создается экстремальная ситуация, напряженность которой все больше возрастает. В этих условиях лучших результатов достигает тот, у кого "крепче нервы", кто лучше может управлять собой в экстремальных ситуациях.

Поэтому игры с элементами БОС можно считать своего рода тренажерами по формированию и совершенствованию навыков адекватного поведения в экстремальных ситуациях.

Здесь необходимо отметить один очень существенный момент. В обычной реальности развитие событий зависит не непосредственно от нашего психофизиологического состояния, а лишь от того, как оно проявляется в наших *действиях*. В случае же виртуальной реальности развитие сценария игры может зависеть *непосредственно* от состояния игрока. Таким образом, *в виртуальной реальности само сознательное (произвольное) или несознательное (непроизвольное) изменение нашего состояния по сути дела является действием*. Аналогичная ситуация в обычной реальности может иметь место при высших формах сознания и проявлении сверхспособностей.

1.3.1.3. Системы с семантическим резонансом. Компьютерные Ψ-технологии и интеллектуальный подсознательный интерфейс

Системами с семантическим резонансом будем называть системы, поведение которых зависит от состояния сознания пользователя и его психологической реакции на смысловые стимулы.

Это означает, что в состав систем с семантическим резонансом, также как и систем с БОС, в качестве подсистем входят информационно-измерительные системы и системы искусственного интеллекта, аналогично осуществляется и съём информации о состоянии пользователя.

Различие между системами с БОС и с семантическим резонансом состоит в том, что в первом случае набор снимаемых параметров и методы их математической обработки определяются необходимостью идентификации биологическо-

го состояния пользователя, тогда как во втором – его реакции на смысловые стимулы (раздражители).

В частности, имеется возможность по наличию в электроэнцефалограмме так называемых вызванных потенциалов [266] установить реакцию человека на стимул: заинтересовался он или нет.

Здесь принципиально важно, что *вызванные потенциалы после предъявления стимула по времени возникают гораздо раньше, чем его осознание.*

Из этого следует ряд важных выводов:

1. Если это осознание не наступает по каким-либо причинам, то вызванные потенциалы все равно с очень высокой достоверностью позволяют прогнозировать ту реакцию, которая была бы у человека, если бы информация о стимуле проникла в его сознание (причинами, по которым зрительный образ стимула может не успеть сформироваться и проникнуть в сознание пользователя, могут быть, например, его очень сильную зашумленность, фрагментарность или слишком короткое время его предъявления).

2. Реакция на стимул на уровне вызванных потенциалов не подвергается критическому анализу и корректировке на уровне сознания, т.е. является гораздо более "искренней" и "откровенной", адекватной и достоверной, чем сознательные ответы на опросник с тем же самым стимульным материалом (сознательные ответы зависят от мотивации, конъюнктуры и массы других обстоятельств).

3. Для получения информации о подсознательной реакции пользователя на стимульный материал он может предъявляться в значительно более высоком темпе, чем при сознательном тестировании.

4. При подсознательном тестировании пользователь может даже не знать о том, что оно проводится.

Все это в совокупности означает, что системы с семантическим резонансом позволяют получить и вывести на уровень сознания обычно ранее не осознаваемую адекватную информацию о состоянии своего сознания, систем мотивации, целеполагания, ценностей и т.д., т.е. расширить область осознаваемого. Это позволяет создать качественно более благоприятные условия для управления состоянием сознания, чем ранее, что является важным эволюционным достижением технократической цивилизации.

Системы с семантическим резонансом могут эффективно использоваться в ряде направлений:

- психологическое и профессиональное тестирование, подбор персонала, в т.ч. для действий в специальных условиях и в измененных формах сознания;
- модификация сознания, систем мотиваций, целеполагания, ценностей и др. (компьютерное нейролингвистическое программирование: "компьютерные НЛП-технологии");
- компьютерные игры с системами семантической обратной связи.

1.3.1.4. Системы виртуальной реальности и критерии реальности. Эффекты присутствия, деперсонализации и модификация сознания пользователя

1.3.1.4.1. Классическое определение системы виртуальной реальности

ВИРТУАЛЬНАЯ РЕАЛЬНОСТЬ (ВР) – модельная трехмерная (3D) окружающая среда, создаваемая компьютерными средствами и реалистично реагирующая на взаимодействие с пользователями (<http://dlc.miem.edu.ru/newsite.nsf/docs/CSD309>).

Технической базой систем виртуальной реальности являются современные мощные персональные компьютеры и программное обеспечение высококачественной трехмерной визуализации и анимации. В качестве устройств ввода-вывода информации в системах ВР применяются виртуальные шлемы с дисплеями (HMD), в частности шлемы со стереоскопическими очками, и устройства 3D-ввода, например, мышь с пространственно управляемым курсором или "цифровые перчатки", которые обеспечивают тактильную обратную связь с пользователем.

Совершенствование систем виртуальной реальности приводит ко все большей изоляции пользователя от обычной реальности, т.к. все больше каналов взаимодействия пользователя с окружающей средой замыкаются не на обычную, а на виртуальную среду – виртуальную реальность, которая, при этом, становится все более и более функционально-замкнутой и самодостаточной.

Создание систем ВР является закономерным следствием процесса совершенствования компьютерных систем отображения информации и интерфейса управления.

При обычной работе на компьютере монитор занимает не более 20% поля зрения пользователя. Системы ВР перекрывают *все* поле зрения.

Обычные мониторы не являются стереоскопическими, т.е. не создают объемного изображения. Правда, в последнее время появились разработки, которые, позволяют преодолеть это ограничение (достаточно сделать поиск в yandex.ru по запросу "Стереоскопический монитор"). Системы ВР изначально были стереоскопическими.

Звуковое сопровождение, в том числе со стерео и quadro-звуком, сегодня уже стали стандартом. В системах ВР человек не слышит ничего, кроме звуков этой виртуальной реальности.

В некоторых моделях систем виртуальной реальности пользователи имеют возможность восприятия изменяющейся перспективы и видят объекты с разных точек наблюдения, как если бы они сами находились и перемещались внутри модели.

Если пользователь располагает более развитыми (*погруженными*) устройствами ввода, например, такими, как цифровые перчатки и виртуальные шлемы, то модель может даже надлежащим образом реагировать на такие *действия* пользователя, как поворот головы или движение глаз.

Необходимо отметить, что в настоящее время системы виртуальной реальности развиваются очень быстрыми темпами и явно выражена тенденция проникновения технологий виртуальной реальности в стандартные компьютерные технологии широкого применения.

Развитие этих и других подобных средств привело к появлению качественно новых эффектов, которые ранее не наблюдались или наблюдались в очень малой степени:

- эффект присутствия пользователя в виртуальной реальности;
- эффект деперсонализации и модификации самосознания и сознания пользователя в виртуальной реальности.

1.3.1.4.2. "Эффект присутствия" в виртуальной реальности

Эффект присутствия – это создаваемая для пользователя иллюзия его присутствия в смоделированной компьютером среде, при этом создается полное впечатление "присутствия" в виртуальной среде, очень сходное с ощущением присутствия в обычном "реальном" мире.

При этом виртуальная среда начинает осознаваться как реальная, а о реальной среде пользователь на время как бы совершенно или почти полностью "забывает". При этом технические особенности интерфейса также вытесняются из сознания, т.е. мы не замечаем этот интерфейс примерно так же, как собственное физическое тело или глаза, когда смотрим на захватывающий сюжет. Таким образом, реальная среда *замещается* виртуальной средой.

Исследования показывают, что для возникновения и силы эффекта присутствия определяющую роль играет *реалистичность движения* различных объектов в виртуальной реальности, а также убедительность реагирования объектов виртуальной реальности при *взаимодействии* с ними виртуального тела пользователя или других виртуальных объектов. В то же время, как это ни странно, естественность вида объектов виртуальной среды играет сравнительно меньшую роль.

1.3.1.4.3. Применения систем виртуальной реальности

Системы виртуальной реальности уже в настоящее время широко применяются во многих сферах жизни.

Одними из первых технологии виртуальной реальности были применены НАСА США для тренировки пилотов космических челноков и военных самолетов, при отработке приемов посадки, дозаправки в воздухе и т.п.

Самолет-невидимка "Стелс" вообще управляется пилотом, практически находящемся в виртуальной реальности.

Из виртуальной реальности человек управляет роботом, выполняющим опасную или тонкую работу.

Технология Motion Capture, позволяет дистанционно "снять" движения с человека и присвоить их его трехмерной модели, что широко применяется для создания компьютерных игр и анимации рисованных персонажей в фильмах.

Особенно эффективно применение виртуальной реальности в рекламе, особенно в Интернет-рекламе на стадии информирования и убеждения.

С использованием виртуальной реальности можно показывать различные помещения, например, совершить виртуальную экскурсию по музею, учебному заведению, дому, коттеджу или местности (прогулка по Парижу от туристической фирмы).

Во всех этих приложениях важно, что в отличие от трехмерной графики, виртуальная реальность обеспечивает *эффект присутствия и личного участия пользователя в наблюдаемых им событиях*.

1.3.1.4.4. Модификация сознания и самосознания пользователя в виртуальной реальности

Сегодня уже для всех вполне очевидно, что виртуальная реальность может с успехом использоваться для развлечений, ведь *она помогает представить себя в другой роли и в другом облики*. Однако в действительности этот эффект связан с модификацией "Образа Я", т.е. сознания и самосознания пользователя. Это значит, что последствия этого в действительности значительно серьезнее, чем обычно представляют, и далеко выходит за рамки собственно развлечений.

Как показано автором в ряде работ, приведенных на сайте <http://Lc.kubagro.ru>, форма сознания и самосознания человека определяются тем, как он осознает себя и окружающее, т.е. тем:

- что он осознает, как объективное, субъективное и несуществующее;
- с чем он отождествляет себя и что осознает как объекты окружающей среды.

Очевидно, что разработчики новейших компьютерных технологий совершенно неожиданно вторглись в абсолютно новую для себя сферу исследования *измененных форм сознания*, и далеко идущие системные последствия этого ими, как и вообще научным сообществом, пока еще очень мало осознаны.

Еще в 1979-1981 годах автором и Л.А.Бакурадзе были оформлены заявки на изобретение компьютерной системы, выполняющей все трудовые функции физического тела, обеспечивающую управление с использованием дистанционного мысленного воздействия, т.е. микротелекинеза. По мнению автора телекинез представляет собой управление физическими объектами путем воздействия на них непосредственно с высших планов без использования физического тела, т.е. *тем же способом, с помощью которого любой человек, осознает он это или нет, управляет своим физическим телом*. Были предложены технические и программные решения и инженерно – психологические методики. Система предлагалась адаптивной, т.е. автоматически настраивающейся на индивидуальные особенности, "почерк" оператора и его состояние сознания, с плавным переключением на дистанционные каналы при повышении их надежности (которая измерялась автоматически) и могла одновременно с выполнением основной работы выступать в качестве тренажера. Человек, начиная работу с системой в обычной форме сознания с использованием традиционных каналов (интерфейса), имея мгновенную адекватную по форме и содержанию *обратную связь* об эффективности своего телекинетического воздействия, должен быстро переходить в форму сознания, оптимальную для использования телекинеза в качестве управляющего воздействия.

1.3.1.4.5. Авторское определение системы виртуальной реальности

С учетом вышесказанного, предлагается следующее определение виртуальной реальности.

Система виртуальной реальности (ВР) – это система, обеспечивающая:

1. **Генерацию полиперцептивной модели реальности** в соответствии с математической моделью этой реальности, реализованной в программной системе.

2. **Погружение пользователя в модель реальности** путем подачи на все или основные его перцептивные каналы – органы восприятия, программно-управляемых по величине и содержанию воздействий: зрительного, слухового, тактильного, термического, вкусового и обонятельного и других.

3. **Управление системой** путем использования виртуального "образа Я" пользователя и виртуальных органов управления системой (интерфейса), на которые он воздействует, представляющие собой зависящую от пользователя часть модели реальности.

4. **Реалистичную реакцию** моделируемой реальности на виртуальное воздействие и управление со стороны пользователя.

5. Разрыв отождествления пользователя со своим "Образом Я" из обычной реальности (деперсонализация), и отождествление себя с "виртуальным образом Я", генерируемым системой виртуальной реальности (**модификация сознания и самосознания пользователя**).

6. **Эффект присутствия** пользователя в моделируемой реальности в своем "виртуальном образе Я", т.е. эффект личного участия пользователя в наблюдаемых виртуальных событиях.

7. **Положительные результаты применения критериев реальности**, т.е. функциональную замкнутость и самодостаточность виртуальной реальности, вследствие чего никакими действиями внутри виртуальной реальности, осуществляемыми над ее объектами, в т.ч. объектами виртуального интерфейса, с помощью своего виртуального тела, невозможно установить, "истинная" эта реальность или виртуальная.

1.3.1.4.6. Критерии реальности при различных формах сознания и их применение в виртуальной реальности

В этой связи вспоминается ставший уже классическим первый фильм "Матрица", в котором Морфей, обращаясь к Нео, произносит свою знаменитую фразу: *"Сейчас я покажу тебе, как выглядит окончательная истинная реальность"*. Эта фраза сразу вызвала у меня массу ассоциаций и вопросов, в частности:

1. А каковы критерии реальности?

2. А вдруг и эта реальность, которую Морфей назвал окончательной, истинной, в действительности является не более, чем симулятором следующего иерархического уровня, так сказать более фундаментальным симулятором?

Здесь возникает сложный мировоззренческий вопрос о том, возможно ли *хотя бы в принципе* находясь в виртуальной реальности не выходя за ее преде-

лы установить, что ты находишься именно в виртуальной, а не истинной реальности, или это возможно сделать только *задним числом*, после выхода из виртуальной реальности и перехода в истинную реальность?

Итак, каковы же критерии реальности?

По нашему мнению, прежде всего это самосогласованность реальности, т.е. получение одной и той же информации качественно различными способами и по различным каналам связи (принцип наблюдаемости):

- согласованность реальности самой с собой во **времени**;
- согласованность и взаимное подтверждение информации от различных органов восприятия, которые обычно реагируют на различные формы материи и часто являются парными (зрение, слух, обоняние) и расположенными в различных точках **пространства**.

Например, мы не только что-то видим, но и слышим, и осязаем, и можем попробовать его на вкус и ощутить запах и все эти восприятия ОТ РАЗЛИЧНЫХ ОРГАНОВ ЧУВСТВ соответствуют друг другу и означают, что перед нами некий определенный объект, а не галлюцинация или визуализация. Согласованная и взаимно подтверждающая информация с различных органов чувств, в соответствии с принципом наблюдаемости, также может рассматриваться как повышающая достоверность и адекватность восприятия.

В современных компьютерных играх мы не только видим довольно качественную визуализацию, но и соответствующее реалистичное звуковое сопровождение. А в системах виртуальной реальности – визуализация стереоскопическая (то, что мы видим РАЗНЫМИ глазами как бы с разных точек в ПРОСТРАНСТВЕ также взаимно подтверждается), а также появляется тактильный канал с обратной связью, который позволяет ощутить даже твердость, вес и температуру моделируемого в виртуальной реальности объекта. Все это вместе уже создает на столько высокую степень реалистичности, что может возникнуть эффект присутствия в виртуальной реальности, деперсонализация и отождествление с измененным образом Я, моделируемым в виртуальной реальности (переход в измененную форму сознания).

Представим, что эти сформулированные критерии реальности не выполняются, т.е. нарушается ее самосогласованность. Как и в чем это может проявляться?

По-видимому, как своего рода "сбои" и различные "**нарушения физических законов**" и несогласованности в виртуальной реальности:

- "заикливание" событий, как на заезженной пластинке, т.е. их многократное повторное осуществление без каких-либо изменений (пример: повторный проход черной кошки, с характерной остановкой и поворотом головы, в дверном проеме в "Матрице");
- прохождение сквозь стены;
- полеты и очень длинные прыжки, а также телепортация в своем "реальном" теле;
- действия в другом темпе времени, т.е. эффект замедления внешнего времени, соответствующий аналогичному ускорению внутреннего времени;
- действия в другом масштабе пространства, "увеличение" и "уменьшение" размеров, наблюдение мега и микроструктуры материи;

- видение сквозь стены, видение на больших расстояниях (в т.ч. с увеличением "как в телескоп"), видение прошлого и будущего;
- телекинез, пирокинез, психосинтез, левитация и т.п.;
- одновременное нахождение в нескольких местах.

Нетрудно заметить, что все эти проявления весьма напоминают так называемые *"паранормальные явления"*, которые традиционно связывают с *сверхвозможностями* человека, т.е. с его возможностями при высших формах сознания.

Эти явления хотя и редко, но все же наблюдаются в нашем мире, что может указывать на то, что наша "истинная реальность" в определенной мере возможно является виртуальной, по крайней мере в большей степени, чем ранее предполагалось.

1.3.1.4.7. Принципы эквивалентности (относительности) Галилея и Эйнштейна и критерии виртуальной реальности

Вспомним известные в физике принципы относительности Галилея и Эйнштейна:

1. Никакими экспериментами внутри замкнутой системы невозможно отличить состояние покоя от состояния равномерного и прямолинейного движения (Галилей). Следовательно, покоящаяся система отсчета ***физически эквивалентна*** системе отсчета, движущейся равномерно и прямолинейно под действием сил инерции.

2. Никакими экспериментами внутри ограниченной по размерам замкнутой системы невозможно установить, движется она под действием сил гравитации или по инерции (Эйнштейн). Следовательно, система отсчета, движущаяся в поле сил тяготения ***физически эквивалентна*** системе отсчета, движущейся под действием сил инерции.

Легко заметить, что формулировка 7-го пункта в определении системы виртуальной реальности весьма сходна с формулировками принципов относительности Галилея и Эйнштейна: *никакими действиями внутри виртуальной реальности, осуществляемыми над ее объектами, в т.ч. объектами виртуального интерфейса, с помощью своего виртуального тела, невозможно установить, "истинная" эта реальность или виртуальная.*

Следовательно, виртуальная система отсчета, локализованная в ***полнофункциональной виртуальной реальности полностью физически эквивалентна*** физической системе отсчета, локализованной в "истинной реальности". Учитывая эту аналогию, принцип, предложенный автором, назовем принципом относительности или *принципом эквивалентности виртуальной и истинной реальности*.

1.3.1.5. Системы с дистанционным телекинетическим интерфейсом

В 1981 году Л.А.Бакурадзе и Е.В.Луценко были оформлены заявки на изобретение компьютерной системы, выполняющей все трудовые функции физического тела, обеспечивающую управление с использованием дистанционного мысленного воздействия, т.е. микротелекинеза.

По мнению автора телекинез представляет собой управление физическими объектами путем воздействия на них непосредственно с высших планов без использования физического тела, т.е. тем же способом, с помощью которого любой человек, осознает он это или нет, управляет своим физическим телом.

Были предложены технические и программные решения и инженерно – психологические методики. Система предлагалась адаптивной, т.е. автоматически настраивающейся на индивидуальные особенности, "почерк" оператора и его состояние сознания, *с плавным переключением на дистанционные каналы при повышении их надежности* (которая измерялась автоматически) и могла одновременно с выполнением основной работы выступать в качестве *тренажера для овладения высшими формами сознания*.

Человек, начиная работу с системой в обычной форме сознания с использованием традиционных каналов (интерфейса), имея мгновенную адекватную по форме и содержанию обратную связь об эффективности своего телекинетического воздействия, должен быстро переходить в одну из высших форм сознания, оптимальную для использования телекинеза в качестве управляющего воздействия.

Контрольные вопросы

1. *Интеллектуальные интерфейсы. Использование биометрической информации о пользователе в управлении системами.*
2. *Идентификация и аутентификация личности по почерку. Понятие клавиатурного почерка.*
3. *Соотношение психофизиологии и атрибуции текстов.*
4. *Идентификация и аутентификация личности пользователя компьютера по клавиатурному почерку.*
5. *Прогнозирование ошибок оператора по изменениям в его электроэнцефалограмме.*
6. *Системы с биологической обратной связью (БОС).*
7. *Мониторинг состояния сотрудников сборочного конвейера с целью обеспечения высокого качества продукции.*
8. *Компьютерные тренажеры, основанные на БОС, для обучения больных навыкам управления своим состоянием.*
9. *Компьютерные игры с БОС.*
10. *Системы с семантическим резонансом. Компьютерные (Ψ-технологии и интеллектуальный подсознательный интерфейс).*
11. *Системы виртуальной реальности и критерии реальности. Эффекты присутствия, деперсонализации и модификация сознания пользователя.*
12. *Классическое определение системы виртуальной реальности.*
13. *"Эффект присутствия" в виртуальной реальности.*
14. *Применения систем виртуальной реальности.*

15. Модификация сознания и самосознания пользователя в виртуальной реальности.
16. Авторское определение системы виртуальной реальности.
17. Критерии реальности при различных формах сознания и их применение в виртуальной реальности.
18. Принципы эквивалентности (относительности) Галилея и Эйнштейна и критерии виртуальной реальности.
19. Системы с дистанционным телекинетическим интерфейсом.

Рекомендуемая литература

1. Завгородний В.В., Мельников Ю.Н., Идентификация по клавиатурному почерку. "Банковские Технологии", №9, 1998.
2. Иванов А.И. Биометрическая идентификация личности по динамике подсознательных движений. Пенза. Издательство Пензенского государственного университета –2000, –188 с.
3. Луценко Е.В., Лаптев В.Н., Третьяк В.Г. Прогнозирование качества специальной деятельности методом подсознательного (подпорогового) тестирования на основе семантического резонанса. //В сб.: "Материалы II межвузовской научно-технической конференции". – Краснодар: КВИ, 2001. – С.127-128.
4. Луценко Е.В., Лебедев А.Н. Диагностика и прогнозирование профессиональных и творческих способностей методом АСК-анализа электроэнцефалограмм в системе "Эйдос". // Межвузовский сборник научных трудов, том 1. – Краснодар: КВИ. 2003.–С. 227-229.
5. Щукин Т.Н., Дорохов В. Б., Лебедев А.Н., Луценко Е.В. ЭЭГ прогноз успешности выполнения психомоторного теста при снижении уровня бодрствования: постановка задачи, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(6). – Режим доступа: <http://ej.kubagro.ru/2004/04/20/p20.asp>
6. Щукин Т.Н., Дорохов В. Б., Лебедев А.Н., Луценко Е.В. ЭЭГ прогноз успешности выполнения психомоторного теста при снижении уровня бодрствования: описание эксперимента, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(6). – Режим доступа: <http://ej.kubagro.ru/2004/04/21/p21.asp>
7. Щукин Т.Н., Дорохов В. Б., Лебедев А.Н., Луценко Е.В. ЭЭГ прогноз успешности выполнения психомоторного теста при снижении уровня бодрствования: анализ результатов исследования, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(6). – Режим доступа: <http://ej.kubagro.ru/2004/04/22/p22.asp>
8. Смирнов И., Безносюк Е., Журавлёв А. Психотехнологии: Компьютерный психосемантический анализ и психокоррекция на неосознаваемом уровне. - М.: Изд. группа Прогресс-Культура, 1995. - 416с. (легко найти в Internet)
9. Шагас Ч. Вызванные потенциалы мозга в норме и патологии. –М.: Мир, 1975. –314 с.
10. Сайт Луценко Е.В. <http://Lc.kubagro.ru>.

1.3.2. ЛЕКЦИЯ-8.

Автоматизированные системы распознавания образов

Учебные вопросы

1. Основные понятия и определения, связанные с системами распознавания образов.
2. Проблема распознавания образов.
3. Классификация методов распознавания образов.
4. Применение распознавания образов для идентификации и прогнозирования. Сходство и различие в содержании понятий "идентификация" и "прогнозирование".
5. Роль и место распознавания образов в автоматизации управления сложными системами.
6. Методы кластерного анализа.

1.3.2.1. Основные понятия

Системой распознавания образов будем называть класс систем искусственного интеллекта, обеспечивающих:

- формирование конкретных образов объектов и обобщенных образов классов;
- обучение, т.е. формирование обобщенных образов классов на основе ряда примеров объектов, классифицированных (т.е. отнесенных к тем или иным категориям – классам) учителем и составляющих обучающую выборку;
- самообучение, т.е. формирование кластеров объектов на основе анализа неклассифицированной обучающей выборки;
- распознавание, т.е. идентификацию (и прогнозирование) состояний объектов, описанных признаками, друг с другом и с обобщенными образами классов;
- измерение степени адекватности модели;
- решение обратной задачи идентификации и прогнозирования (обеспечивается не всеми моделями).

1.3.2.1.1. Признаки и образы конкретных объектов, метафора фазового пространства

Признаками объектов будем называть конкретные результаты измерения значений их свойств.

Свойства объектов отличаются своим качеством и измеряются с помощью различных органов восприятия или измерительных приборов в различных единицах измерения.

Результатом измерения является снижение неопределенности в наших знаниях о значении свойств объекта. Значения свойств конкретизируются путем

их сопоставления определенным градациям соответствующих измерительных шкал: номинальных, порядковых или отношений.

В номинальных шкалах отсутствуют отношения порядка, начало отсчета и единица измерения.

На порядковых шкалах определены отношения "больше – меньше", но отсутствуют начало отсчета и единица измерения.

На шкалах отношений определены отношения порядка, все арифметические операции, есть начало отсчета и единица измерения.

Можно представить себе, что шкалы образуют оси координат некоторого абстрактного многомерного пространства, которое будем называть "фазовым пространством".

В этом фазовом пространстве каждый конкретный объект представляется определенной точкой, имеющей координаты, соответствующие значениям его свойств по осям координат, т.е. градациям описательных шкал.

Оси координат фазового пространства в общем случае не являются взаимно-перпендикулярными шкалами отношений, т.е. в общем случае это пространство неортонормированное, более того – неметрическое. Следовательно, в нем в общем случае не применима Евклидова мера расстояний, т.е. не действует Евклидова метрика. Применение этой меры расстояний корректно, если одновременно выполняются два условия:

1. Все оси координат фазового пространства являются шкалами отношений.
2. Все оси координат взаимно-перпендикулярны или очень близки к этому.

1.3.2.1.2. Признаки и обобщенные образы классов

Обобщенный образ класса формируется из нескольких образов конкретных объектов, относящихся к данному классу, т.е. одной градации некоторой классификационной шкалы.

Обобщенные образы классов формализуются (кодируются) путем использования классификационных шкал и градаций, которые могут быть тех же типов, что и описательные, т.е. номинальные, порядковые и отношений.

Сама принадлежность конкретных объектов к данному классу определяется либо человеком-учителем, после чего фиксируется в обучающей выборке, либо самой системой автоматически на основе кластерного анализа конкретных объектов.

1.3.2.1.3. Обучающая выборка и ее репрезентативность по отношению к генеральной совокупности. Ремонт (взвешивание) данных

Обучающая выборка является некоторым подмножеством исследуемой совокупности, которая называется "генеральная совокупность". На основе изучения обучающей выборки мы хотели бы сделать выводы о генеральной совокупности, причем важно знать степень достоверности этих выводов.

Рассмотрим, как зависит степень достоверности выводов о генеральной совокупности от объема обучающей выборки.

Если обучающая выборка включает все объекты генеральной совокупности, т.е. они совпадают, то достоверность выводов будет наиболее высокой (при всех прочих равных условиях).

Если же обучающая выборка очень мала, то вряд ли на ее основе могут быть сделаны достоверные выводы о генеральной совокупности, т.к. в этом случае в обучающую выборку могут даже не входить примеры объектов всех или подавляющего большинства классов.

Под репрезентативностью обучающей выборки будем понимать ее способность адекватно представлять генеральную совокупность, так что изучение самой генеральной совокупности можно корректно заменить исследованием обучающей выборки.

Но репрезентативность зависит не только от объема, но и от структуры обучающей выборки, т.е. от того, насколько полно представлены все категории объектов генеральной совокупности (классы) и от того, насколько полно они описаны признаками.

Взвешивание данных или ремонт обучающей выборки – это операция, в результате которой частное распределение объектов по классам в обучающей выборке максимально, на сколько это возможно, приближается либо к частотному распределению генеральной совокупности (если оно известно из независимых источников), либо к равномерному.

В системе "Эйдос" режим взвешивания данных реализован.

1.3.2.1.4. Основные операции: обобщение и распознавание

Обобщение – это операция формирования обобщенных образов классов на основе описаний конкретных объектов, входящих в обучающую выборку.

Сразу необходимо отметить, что операция обобщения реализуется далеко не во всех моделях систем распознавания (например, в методе k-ближайших соседей), а в тех, в которых оно реализуется, – это делается по-разному.

Обычно, пока не реализовано обобщение нет возможности определить ценность признаков для решения задачи идентификации.

Например, если у нас есть 10 конкретных мячей разного размера и цвета, состоящих из разных материалов и предназначенных для разных игр, и мы рассматриваем их как совершенно независимые друг от друга объекты, наряду с другими, то у нас нет возможности определить, какие признаки являются наиболее характерными для мячей и наиболее сильно отличают их от этих других объектов. Но как только мы сформируем обобщенные образы "мяч", "стул", и т.д., сразу выяснится, что цвет мяча и материал, из которого он сделан, не является жестко связанными с обобщенным образом класса "мяч", а наиболее существенно то, что он круглый и его можно бросать или бить во время игры.

Распознавание – это операция сравнения и определения степени сходства образа данного конкретного объекта с образами других конкретных объектов или с обобщенными образами классов, в результате которой формируется рейтинг объектов или классов по убыванию сходства с распознаваемым объектом.

Ключевым моментом при реализации операции распознавания в математической модели является выбор вида *интегрального критерия или меры сходства*, который бы на основе знания о признаках конкретного объекта позволил бы количественно определить степень его сходства с другими объектами или обобщенными образами классов.

В ортонормированном пространстве, осями которого являются шкалы отношений, вполне естественным является использовать в качестве такой меры сходства Евклидово расстояние. Однако, такие пространства на практике встречаются скорее как исключение из правила, а операция ортонормирования является довольно трудоемкой в вычислительном отношении и приводит к обеднению модели, а значит ее не всегда удобно и целесообразно осуществлять.

Поэтому актуальной является задача выбора или конструирования интегрального критерия сходства, применение которого было бы корректно и в ортонормированных пространствах. Кроме того, этот интегральный критерий должен быть устойчив к наличию *шума, т.е. к неполноте и искажению* как в исходных данных, так и самой численной модели.

Требование устойчивости к наличию шума математически означает, что результат применения интегрального критерия к сигналу, состоящему только из белого шума, должен быть равным нулю. Это значит, что в качестве интегрального критерия может быть применена функция, используемая при определении самого понятия "белый шум", т.е. свертка, скалярное произведение, корреляция.

Такой интегральный критерий предложен в математической модели системно-когнитивного анализа и реализован в системе "Эйдос".

1.3.2.1.5. Обучение с учителем (экспертом) и самообучение (кластерный анализ)

Обучение с учителем – это процесс формирования обобщенных образов классов, на основе обучающей выборки, содержащей характеристики конкретных объектов как в описательных, так и в классификационных шкалах и градациях.

Причем, если описательные характеристики могут формироваться с помощью информационно-измерительной системы автоматически, то классификационные – представляют собой результат вообще говоря неформализуемого процесса оценки степени принадлежности данных объектов к различным классам, который осуществляется человеком-экспертом или, как традиционно говорят специалисты по распознаванию образов, "учителем". В этом случае не возникает вопроса о том, для формирования обобщенного образа каких классов использовать описание данного конкретного объекта.

Обучение без учителя или самообучение – это процесс формирования обобщенных образов классов, на основе обучающей выборки, содержащей характеристики конкретных объектов, причем только в описательных шкалах и градациях.

Поэтому этот процесс реализуется в три этапа:

1. Кластерный анализ объектов обучающей выборки, в результате которого определяются группы наиболее сходных из них по их признакам (кластеры).

2. Присвоение кластерам статуса обобщенных классов, для формирования обобщенных образов которых используются конкретные объекты, входящие именно в эти кластеры.

3. Формирование обобщенных образов классов, аналогично тому, как это делалось при обучении с учителем.

1.3.2.1.6. Верификация, адаптация и синтез модели

Как только произнесено или написано слово "модель", сразу неизбежно возникает вопрос о степени ее адекватности.

Верификация модели – это операция установления степени ее адекватности (валидности) путем сравнения результатов идентификации конкретных объектов с их фактической принадлежностью к обобщенным образам классов.

Различают внутреннюю и внешнюю, интегральную и дифференциальную валидность.

Внутренняя валидность – это способность модели верно идентифицировать объекты обучающей выборки.

Если модель имеет низкую внутреннюю валидность, то модель нельзя считать удачно сформированной.

Внешняя валидность – это способность модели верно идентифицировать объекты, не входящие в обучающую выборку.

Интегральная валидность – это средневзвешенная достоверность идентификации по всем классам и распознаваемым объектам.

Дифференциальная валидность – это способность модели верно идентифицировать объекты в разрезе по классам.

Адаптация модели – это учет в модели объектов, не входящих в обучающую выборку, но входящих в генеральную совокупность, по отношению к которой данная обучающая выборка репрезентативна.

Если моделью верно идентифицируются объекты, не входящие в обучающую выборку, то это означает, что эти объекты входят в генеральную совокупность, по отношению к которой данная обучающая выборка репрезентативна. Следовательно, на основе обучающей выборки удалось выявить закономерности взаимосвязей между признаками и принадлежностью объектов к классам, которые действуют не только в обучающей выборке, но имеют силу и для генеральной совокупности.

Адаптация модели не требует изменения классификационных и описательных шкал и градаций, а лишь объема обучающей выборки, и *приводит к количественному изменению модели.*

Синтез (или повторный синтез – пересинтез) модели – это учет в модели объектов, не входящих ни в обучающую выборку, ни в генеральную совокупность, по отношению к которой данная обучающая выборка репрезентативна.

Это объекты с новыми, ранее неизвестными закономерностями взаимосвязей признаков с принадлежностью этих объектов к тем или иным классам. Причем и признаки, и классы, могут быть как те, которые уже были отражены в модели ранее, так и новые. *Пересинтез модели приводит к ее качественному изменению.*

1.3.2.2. Проблема распознавания образов

Проблема распознавания образов сводится к двум задачам: обучения и распознавания. Поэтому, прежде чем сформулировать задачу обучения распознаванию образов уточним, в чем смысл их распознавания.

Простейшим вариантом распознавания является *строгий запрос* на поиск объекта в базе данных по его признакам, который реализуется в информационно-поисковых системах. При этом каждому полю соответствует признак (описательная шкала), а значению поля – значение признака (градация описательной шкалы). Если в базе данных есть записи, *все значения заданных полей которых точно совпадают* со значениями, заданными в запросе на поиск, то эти записи извлекаются в отчет, иначе запись не извлекается.

Более сложными вариантами распознавания является *нечеткий запрос с неполной информацией*, когда не все признаки искомых объектов задаются в запросе на поиск, т.к. не все они известны, и *нечеткий запрос с шумом*, когда не все признаки объекта известны, а некоторые считаются известными ошибочно. В этих случаях из базы данных извлекаются все объекты, у которых совпадает хотя бы один признак и в отчете объекты сортируются (ранжируются) в порядке убывания количества совпавших признаков. При этом при определении ранга объекта в отсортированном списке все признаки считаются имеющими одинаковый "вес" и учитывается только их количество.

Однако:

- во-первых, на самом деле признаки имеют разный вес, т.е. *один и тот же признак в разной степени характерен для различных объектов*;
- во-вторых, нас могут интересовать не столько сами объекты, извлекаемые из базы данных прецедентов по запросам, сколько *классификация самого запроса*, т.е. отнесение его к определенной категории, т.е. к тому или иному *обобщенному* образу класса.

Если реализация строгих и даже нечетких запросов не вызывает особых сложностей, то распознавание как идентификация с *обобщенными* образами классов, причем с учетом *различия весов признаков* представляет собой определенную проблему.

Обучение осуществляется путем предъявления системе отдельных объектов, описанных на языке признаков, с указанием их принадлежности тому или другому классу. При этом сама принадлежность к классам сообщается системе человеком – Учителем (экспертом).

В результате обучения распознающая система должна приобрести способность:

1. Относить объекты к классам, к которым они принадлежат (идентифицировать объекты верно).
2. Не относить объекты к классам, к которым они не принадлежат (не идентифицировать объекты ошибочно).

Эта и есть проблема обучения распознаванию образов, и состоит она в следующем:

1. В разработке математической модели, обеспечивающей: обобщение образов конкретных объектов и формирование обобщенных образов классов; расчет весов признаков; определение степени сходства конкретных объектов с

классами и ранжирование классов по степени сходства с конкретным объектом, включая и положительное, и отрицательное сходство.

2. В наполнении этой модели конкретной информацией, характеризующей определенную предметную область.

1.3.2.3. Классификация методов распознавания образов

Распознаванием образов называются задачи установления отношений эквивалентности между конкретными и обобщенными образами-моделями объектов реального или идеального мира.

Отношения эквивалентности выражают принадлежность оцениваемых объектов к каким-либо классам, рассматриваемым как самостоятельные семантические единицы.

При построении алгоритмов распознавания классы эквивалентности могут задаваться исследователем, который пользуется собственными содержательными представлениями или использует внешнюю дополнительную информацию о сходстве и различии объектов в контексте решаемой задачи. Тогда говорят о "распознавании с учителем". В противном случае, т.е. когда автоматизированная система решает задачу классификации без привлечения внешней обучающей информации, говорят об автоматической классификации или "распознавании без учителя".

Большинство алгоритмов распознавания образов требует привлечения весьма значительных вычислительных мощностей, которые могут быть обеспечены только высокопроизводительной компьютерной техникой.

Различные авторы (Ю.Л. Барабаш, В.И. Васильев, А.Л. Горелик, В.А. Скрипкин, Р. Дуда, П. Харт, Л.Т.Кузин, Ф.И. Перегудов, Ф.П. Тарасенко, Темников Ф.Е., Афонин В.А., Дмитриев В.И., Дж. Ту, Р. Гонсалес, П. Уинстон, К. Фу, Я.З. Цыпкин и др.) дают различную типологию методов распознавания образов. Одни авторы различают параметрические, непараметрические и эвристические методы, другие – выделяют группы методов, исходя из исторически сложившихся школ и направлений в данной области.

Например, в работах В.А. Дюка [38, 39], в которых дан академический обзор методов распознавания, используется следующая типология методов распознавания образов:

- методы, основанные на принципе разделения;
- статистические методы;
- методы, построенные на основе "потенциальных функций";
- методы вычисления оценок (голосования);
- методы, основанные на исчислении высказываний, в частности на аппарате алгебры логики.

В основе данной классификации лежит различие в формальных методах распознавания образов и поэтому опущено рассмотрение эвристического подхода к распознаванию, получившего полное и адекватное развитие в экспертных системах.

Эвристический подход основан на трудно формализуемых знаниях и интуиции исследователя. При этом исследователь сам определяет, какую инфор-

мацию и каким образом система должна использовать для достижения требуемого эффекта распознавания.

Подобная типология методов распознавания с той или иной степенью детализации встречается во многих работах по распознаванию. В то же время известные типологии не учитывают одну очень существенную характеристику, которая отражает специфику способа представления знаний о предметной области с помощью какого-либо формального алгоритма распознавания образов.

В.А. Дюка [38, 39] выделяет два основных способа представления знаний:

- интенциональное, в виде схемы связей между атрибутами (признаками).
- экстенциональное, с помощью конкретных фактов (объекты, примеры).

Интенциональное представление фиксируют закономерности и связи, которыми объясняется структура данных. Применительно к диагностическим задачам такая фиксация заключается в определении операций над атрибутами (признаками) объектов, приводящих к требуемому диагностическому результату. Интенциональные представления реализуются посредством операций над значениями атрибутов и не предполагают произведения операций над конкретными информационными фактами (объектами).

В свою очередь, экстенциональные представления знаний связаны с описанием и фиксацией конкретных объектов из предметной области и реализуются в операциях, элементами которых служат объекты как целостные системы.

На наш взгляд, можно провести глубокую и далеко идущую аналогию между интенциональными и экстенциональными представлениями знаний и механизмами, лежащими в основе деятельности левого и правого полушарий головного мозга человека. Если для правого полушария характерна целостная прототипная репрезентация окружающего мира, то левое полушарие оперирует закономерностями, отражающими связи атрибутов этого мира.

Описанные выше два фундаментальных способа представления знаний позволяют предложить следующую классификацию методов распознавания образов:

- интенциональные методы, основанные на операциях с признаками.
- экстенциональные методы, основанные на операциях с объектами.

Необходимо особо подчеркнуть, что существование именно этих двух (и только двух) групп методов распознавания: оперирующих с признаками, и оперирующих с объектами, на наш взгляд, глубоко закономерно. С этой точки зрения ни один из этих методов, взятый отдельно от другого, не позволяет сформировать адекватное отражение предметной области. Между этими методами существует отношение дополнительности в смысле Н.Бора, поэтому перспективные системы распознавания должны обеспечивать реализацию обоих этих методов, а не только какого-либо одного из них.

Таким образом, в основу классификации методов распознавания, предложенной В.А. Дюка [38, 39], положены фундаментальные закономерности, лежащие в основе человеческого способа познания вообще, что ставит ее в совершенно особое (привилегированное) положение по сравнению с другими классификациями, которые на этом фоне выглядят более легковесными и искусственными.

1.3.2.4. Применение распознавания образов для идентификации и прогнозирования. Сходство и различие в содержании понятий "идентификация" и "прогнозирование"

Термины "Распознавание образов" и "Идентификация" являются синонимами.

Идентификация и прогнозирование часто практически ничем друг от друга не отличаются по математическим моделям и алгоритмам. Основное различие между ними состоит в том, что при идентификации признаки и состояния объекта относятся к одному времени, тогда как при прогнозировании признаки (факторы) относятся к прошлому, а состояния объекта – к будущему.

Это означает, что системы распознавания образов с успехом могут применяться не только для решения задач идентификации, но и прогнозирования.

1.3.2.5. Роль и место распознавания образов в автоматизации управления сложными системами

1.3.2.5.1. Обобщенная структура системы управления

Автоматизированная система управления состоит из двух основных частей: объекта управления и управляющей системы (рисунок 71).

Управляющая система осуществляет следующие функции:

- идентификация состояния объекта управления;
- выработка управляющего воздействия исходя из целей управления с учетом состояния объекта управления и окружающей среды;
- оказание управляющего воздействия на объект управления.

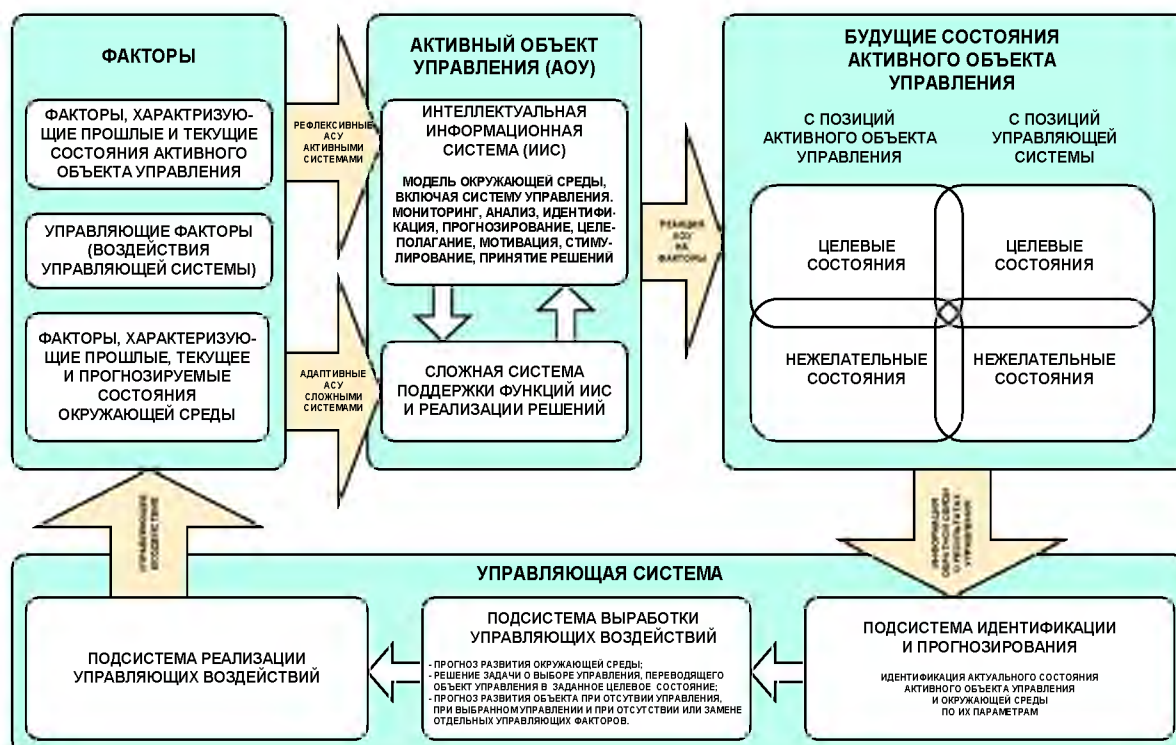


Рисунок 71. Обобщенная схема рефлексивной системы управления активными объектами

1.3.2.5.2. Место системы идентификации в системе управления

Распознавание образов есть не что иное, как идентификация состояния некоторого объекта. Автоматизированная система управления (АСУ), построенная на традиционных принципах, может работать только на основе параметров, закономерности связей которых уже известны, изучены и отражены в математической модели. В итоге АСУ, основанные на традиционном подходе, практически не эффективны с активными многопараметрическими слабодетерминированными объектами управления, такими, например, как макро– и микро– социально-экономические системы в условиях динамичной экономики "переходного периода", иерархические элитные и этнические группы, социум и электорат, физиология и психика человека, природные и искусственные экосистемы и многие другие.

Поэтому, в состав перспективных АСУ, обеспечивающих устойчивое управление активными объектами в качестве существенных функциональных звеньев должны войти подсистемы идентификации и прогнозирования состояний среды и объекта управления, основанные на методах искусственного интеллекта (прежде всего распознавания образов), методах поддержки принятия решений и теории информации.

1.3.2.5.3. Управление как задача, обратная идентификации и прогнозированию

Кратко рассмотрим вопрос о применении систем распознавания образов для принятия решений об управляющем воздействии. Очевидно, что применение систем распознавания для прогнозирования результатов управления при различных сочетаниях управляющих факторов позволяет рассмотреть и сравнить различные варианты управления и выбрать наилучшие из них по определенным критериям. Однако, этот подход на практике малоэффективен, особенно если факторов много, т.к. в этом случае количество сочетаний их значений может быть чрезвычайно большим.

Если в качестве классов распознавания взять целевые и иные будущие состояния объекта управления, а в качестве признаков – факторы, влияющие на него, то в модели распознавания образов может быть сформирована количественная мера причинно-следственной связи факторов и состояний.

Это позволяет по заданному целевому состоянию объекта управления получить информацию о силе и направлении влияния факторов, способствующих или препятствующих переходу объекта в это состояние, и, на этой основе, выработать решение об управляющем воздействии.

Задача выбора факторов по состоянию является обратной задачей прогнозирования, т.к. при прогнозировании, наоборот, определяется состояние по факторам.

Факторы могут быть разделены на следующие группы:

– характеризующие предысторию объекта управления и его актуальное состояние управления;

- технологические (управляющие) факторы;
- факторы окружающей среды;

Таким образом, системы распознавания образов могут быть применены в составе АСУ в подсистемах:

- идентификации состояния объекта управления;
- выработки управляющих воздействий.

Это целесообразно в случае, когда объект управления представляет собой сложную или активную систему.

1.3.2.6. Методы кластерного анализа

Термин "Кластерный анализ" впервые ввел Тьюн в 1939.

Кластеризация – это операция автоматической классификации, в ходе которой объекты объединяются в группы (кластеры) таким образом, что внутри групп различия между объектами минимальны, а между группами – максимальны. При этом в ходе кластеризации не только определяется состав кластеров, но и сам их набор и границы.

Поэтому вполне обоснованно считается, что методы кластерного анализа используются в большинстве случаев тогда, когда нет каких-либо априорных гипотез относительно классов, т.е. исследование находится на первой эмпирической стадии: описательной.

Существует большое количество различных алгоритмов кластеризации, которые обычно связаны с полным перебором объектов и весьма трудоемки в вычислительном отношении, здесь же мы упомянем лишь о трех из них:

- объединение (древовидная кластеризация);
- двухходовое объединение;
- метод К средних.

Рассмотрим кратко эти алгоритмы (описание взято с сайта <http://StatSoft.ru>).

1.3.2.6.1. Древовидная кластеризация

Древовидная диаграмма (*диаграмму* (рисунок 72) начинается с конкретных объектов (в левой части диаграммы). Теперь представим себе, что постепенно (очень малыми шагами) вы "ослабляете" ваш критерий о том, какие объекты являются уникальными, а какие нет. Другими словами, вы понижаете порог, относящийся к решению об объединении двух или более объектов в один кластер.

В результате, вы *связываете* вместе всё большее и большее число объектов и агрегируете (*объединяете*) все больше и больше кластеров, состоящих из все сильнее различающихся элементов. Окончательно, на последнем шаге все объекты объединяются вместе.

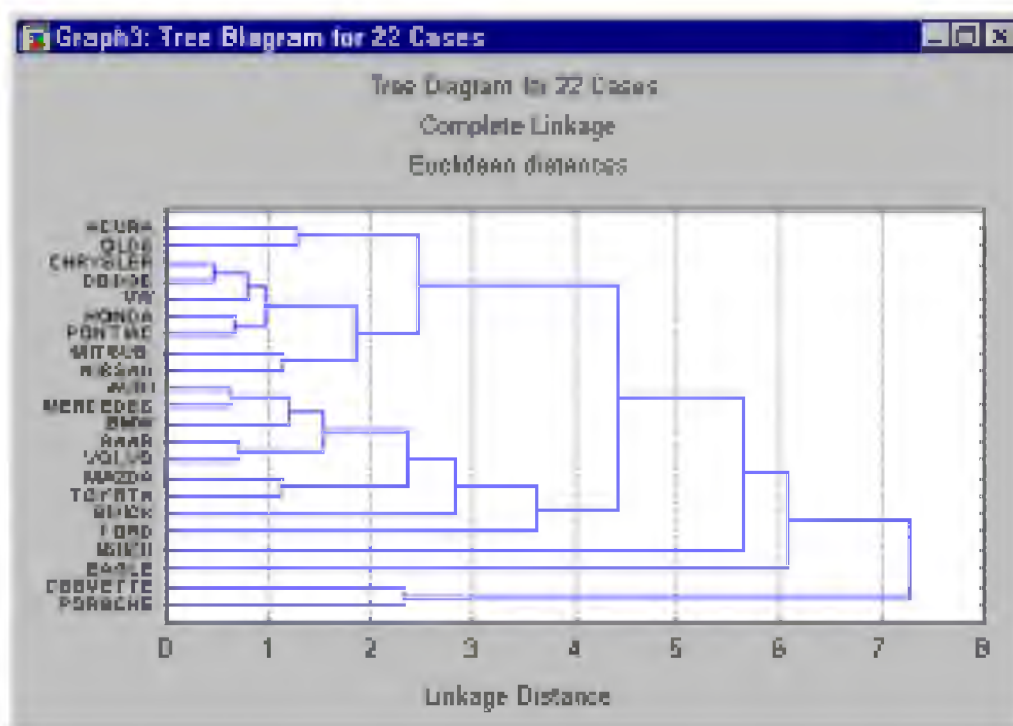


Рисунок 72. Древовидная диаграмма последовательной кластеризации

1.3.2.6.2. Двухходовое объединение

Исследователь может кластеризовать конкретные образы наблюдаемых объектов для определения кластеров объектов со сходными признаками.

Он может также кластеризовать признаки для определения кластеров признаков, которые связаны со сходными конкретными объектами.

В двухходовом алгоритме эти процессы осуществляются одновременно.

1.3.2.6.3. Метод *K* средних

В этом методе принадлежность объектов к кластерам определяется таким образом, чтобы:

- минимизировать изменчивость (различия) объектов внутри кластеров;
- максимизировать изменчивость объектов между кластерами.

Контрольные вопросы

1. Основные понятия и определения, связанные с системами распознавания образов.
2. Признаки и образы конкретных объектов, метафора фазового пространства.
3. Признаки и обобщенные образы классов.
4. Обучающая выборка и ее репрезентативность по отношению к генеральной совокупности. Ремонт (взвешивание) данных.

5. Основные операции: обобщение и распознавание.
 6. Обучение с учителем (экспертом) и самообучение (кластерный анализ).
 7. Верификация, адаптация и синтез модели.
 8. Проблема распознавания образов.
 9. Классификация методов распознавания образов.
 10. Применение распознавания образов для идентификации и прогнозирования.
- Сходство и различие в содержании понятий "идентификация" и "прогнозирование".*
11. Роль и место распознавания образов в автоматизации управления сложными системами.
 12. Обобщенная структура системы управления.
 13. Место системы идентификации в системе управления.
 14. Управление как задача, обратная идентификации и прогнозированию.
 15. Методы кластерного анализа.
 16. Метод кластеризации: "Древовидная кластеризация".
 17. Метод кластеризации: "Двууходовое объединение".
 18. Метод кластеризации: "Метод К средних".

Рекомендуемая литература

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

1.3.3. ЛЕКЦИЯ-9.

Математические методы и автоматизированные системы поддержки принятия решений

Учебные вопросы

1. Многообразие задач принятия решений.
2. Языки описания методов принятия решений.
3. Выбор в условиях неопределенности.
4. Решение как компромисс и баланс различных интересов. О некоторых ограничениях оптимизационного подхода.
5. Экспертные методы выбора.
6. Юридическая ответственность за решения, принятые с применением систем поддержки принятия решений.
7. Условия корректности использования систем поддержки принятия решений.
8. Хранилища данных для принятия решений.

1.3.3.1. Многообразие задач принятия решений

1.3.3.1.1. Принятие решений, как реализация цели

Определение: принятие решения есть действие над множеством альтернатив, в результате которого исходное множество альтернатив сужается. Это действие называется "**выбор**".

Выбор является действием, придающим всей деятельности целенаправленность. Именно через акты выбора реализуется подчиненность всей деятельности определенной цели или совокупности взаимосвязанных целей.

Таким образом, для того, чтобы стал возможен акт выбора, необходимо следующее:

Порождение или обнаружение множества альтернатив, на котором предстоит совершить выбор.

Определение целей, ради достижения которых осуществляется выбор.

Разработка и применение способа сравнения альтернатив между собой, т.е. определение рейтинга предпочтения для каждой альтернативы, согласно определенным критериям, позволяющим косвенно оценивать, насколько каждая альтернатива соответствует цели.

Современные работы в области поддержки принятия решений выявили характерную ситуацию, которая состоит в том, что *полная формализация нахождения наилучшего (в определенном смысле) решения возможна только для хорошо изученных, относительно простых задач, тогда как на практике чаще встречаются слабо структурированные задачи для которых полностью формализованных алгоритмов не разработано* (если не считать полного перебора и метода проб и ошибок). Вместе с тем, опытные, компетентные и способные специалисты, часто делают выбор, который оказывается достаточно хо-

рошим. Поэтому современная тенденция практики принятия решений в естественных ситуациях состоит в сочетании способности человека решать неформализованные задачи с возможностями формальных методов и компьютерного моделирования: диалоговые системы поддержки принятия решений, экспертные системы, адаптивные человеко-машинные автоматизированные системы управления, нейронные сети и когнитивные системы.

1.3.3.1.2. Принятие решений, как снятие неопределенности (информационный подход)

Процесс получения информации можно рассматривать как уменьшение неопределенности в результате приема сигнала, а количество информации, как количественную меру степени снятия неопределенности.

Но в результате выбора некоторого подмножества альтернатив из множества, т.е. в результате принятия решения, происходит тоже самое (уменьшение неопределенности).

Это значит, что каждый выбор, каждое решение порождает определенное количество информации, а значит может быть описано в терминах теории информации.

Простейшее понятие об информации (подход Хартли).

Будем считать, что если существует множество элементов и осуществляется выбор одного из них, то этим самым сообщается или генерируется определенное количество информации. Эта информация состоит в том, что если до выбора не было известно, какой элемент будет выбран, то после выбора это становится известным.

Найдем вид функции, связывающей количество информации, получаемой при выборе некоторого элемента из множества, с количеством элементов в этом множестве, т.е. с его мощностью.

Если множество элементов, из которых осуществляется выбор, состоит из одного-единственного элемента, то ясно, что его выбор предопределен, т.е. никакой неопределенности выбора нет. Таким образом, если мы узнаем, что выбран этот единственный элемент, то, очевидно, при этом мы не получаем никакой новой информации, т.е. получаем нулевое количество информации.

Если множество состоит из двух элементов, то неопределенность выбора минимальна. В этом случае минимально и количество информации, которое мы получаем, узнав, что совершен выбор одного из элементов. Минимальное количество информации получается при выборе одного из двух равновероятных вариантов. Это количество информации принято за единицу измерения и называется "бит".

Чем больше элементов в множестве, тем больше неопределенность выбора, тем больше информации мы получаем, узнав о том, какой выбран элемент.

Рассмотрим множество, состоящее из чисел в двоичной системе счисления длиной i двоичных разрядов. При этом каждый из разрядов может принимать значения только 0 и 1 (таблица 32).

**Таблица 32 – К ЭВРИСТИЧЕСКОМУ ВЫВОДУ ФОРМУЛЫ
КОЛИЧЕСТВА ИНФОРМАЦИИ ПО ХАРТЛИ**

Кол-во двоичных разрядов (<i>i</i>)	Кол-во состояний N, которое можно пронумеровать <i>i</i> -разрядными двоичными числами	Основание системы счисления		
		10	16	2
1	2	0 1	0 1	0 1
2	4	0 1 2 3	0 1 2 3	00 01 10 11
3	8	0 1 2 3 4 5 6 7	0 1 2 3 4 5 6 7	000 001 010 011 100 101 110 111
4	16	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	0 1 2 3 4 5 6 7 8 9 A B C D E F	0000 0001 0010 0011 0100 0101 0110 0111 1000 1001 1010 1011 1100 1101 1110 1111
***	***			
<i>i</i>	$N=2^i$			

Из таблицы 32 очевидно, что количество этих чисел (элементов) в множестве равно:

$$N = 2^i$$

Рассмотрим процесс выбора чисел из рассмотренного множества. До выбора вероятность выбрать любое число одинакова. Существует объективная неопределенность в вопросе о том, какое число будет выбрано. Эта неопределенность тем больше, чем больше N – количество чисел в множестве, а чисел тем больше – чем больше разрядность i этих чисел.

Примем, что выбор одного числа дает нам следующее количество информации:

$$i = \text{Log}_2 N$$

Таким образом, количество информации, содержащейся в двоичном числе, равно количеству двоичных разрядов в этом числе. Это количество информации i мы получаем, когда случайным равновероятным образом выпадает одно из двоичных чисел, записанных i разрядами, или из некоторого множества выбирается объект произвольной природы, пронумерованный этим числом (предполагается, что остальные объекты этого множества пронумерованы остальными числами и этим они и отличаются).

Это выражение и представляет собой формулу Хартли для количества информации. Отметим, что оно полностью совпадает с выражением для энтропии (по Эшби), которая рассматривалась им как количественная мера степени неопределенности состояния системы.

Сам Хартли, возможно, пришел к своей мере на основе эвристических соображений, подобных только что изложенным, но в настоящее время строго доказано, что логарифмическая мера для количества информации однозначно следует из этих двух постулированных им условий.

Таким образом, информация по своей сущности теснейшим и органичным образом связана с выбором и принятием решений.

Отсюда следует простейшее на первый взгляд заключение: *"Для принятия решений нужна информация, без информации принятие решений невозможно, значение информации для принятия решений является определяющим, процесс принятия решений генерирует информацию".*

Мера Шеннона, как обобщение меры Хартли для неравновероятных событий.

Представим себе, что имеются объекты различных видов, причем:

- всего имеется M видов объектов;
- объектов каждого i -го вида имеется N_i .

Тогда по Хартли, если мы извлекаем один из объектов i -го вида, то получаем I_i бит информации

$$I_i = \log_2 N_i$$

В среднем по \bar{I}_i на один объект i -го вида.

$$\bar{I}_i = \frac{I_i}{N_i} = \frac{\log_2 N_i}{N_i}$$

Сумма этих средних будет равна:

$$\bar{I}_i = \sum_{i=1}^M \frac{\log_2 N_i}{N_i} = - \sum_{i=1}^M \frac{1}{N_i} \log_2 \frac{1}{N_i} = - \sum_{i=1}^M p_i \log_2 p_i$$

(где: $p_i = 1/N_i$, – вероятность встречи объектов i -го вида).

Последнее выражение – это и есть *формула Шеннона*, которая, таким образом, позволяет рассчитать средневзвешенное количество информации, приходящееся на один объект, получаемое при предъявлении объектов различных видов.

Отметим, что идентификация объектов, как относящихся к тому или иному виду (i-му виду) осуществляется на основе *признаков* этих объектов. В простейшем варианте это может быть и один признак, например номер вида на бильярдном шаре, но в реальных случаях признаков может быть очень много и их различные наборы сложным и неоднозначным образом могут быть связаны с принадлежностью объектов к тем или иным классам.

Но главный вывод от этого не изменяется: *формула Шеннона дает средневзвешенное количество информации, приходящееся на один объект, получаемое при предъявлении объектов различных видов (классов), отличающихся своими наборами признаков. Мера Шеннона является обобщением меры Хартли для неравновероятных событий.*

1.3.3.1.3. Связь принятия решений и распознавания образов

Распознавание образов есть решение задачи идентификации состояния объекта или системы, отнесение его к той или иной классификационной категории.

Распознавание образов и принятие решений тесно взаимосвязаны:

- само распознавание безусловно представляет собой принятие решения об отнесении состояния системы к той или иной категории;
- от того, к какой категории отнесено состояние системы самым непосредственным образом зависят и решения по управлению этой системой;
- математический аппарат распознавания образов и принятия решений имеет очень много аналогий, а иногда и просто является тождественным (например, сведение многокритериальной задачи к однокритериальной).

1.3.3.1.4. Классификация задач принятия решений

Множественность задач принятия решений связана с тем, что каждая компонента ситуации, в которой осуществляется принятие решений может реализовываться в качественно различных вариантах.

Вот только некоторые из этих вариантов:

Множество альтернатив с одной стороны может быть конечным, счетным или континуальным, а с другой – закрытым (т.е. известным полностью), или открытым (включающим неизвестные элементы).

Оценка альтернатив может осуществляться по одному или нескольким критериям, которые, в свою очередь, могут иметь количественный или качественный характер.

Режим выбора может быть однократным (разовым), или многократным, повторяющимся, включающим обратную связь по результатам выбора, т.е. допускающим обучение алгоритмов принятия решений с учетом последствий предыдущих выборов.

Последствия выбора каждой альтернативы могут быть точно известны заранее (выбор в условиях определенности), иметь вероятностный характер, когда известны вероятности возможных исходов после сделанного выбора (выбор в

условиях риска), или иметь неоднозначный исход с неизвестными вероятностями (выбор в условиях неопределенности).

Ответственность за выбор может отсутствовать, быть индивидуальной или групповой.

Степень согласованности целей при групповом выборе может варьироваться от полного совпадения интересов сторон (кооперативный выбор), до их противоположности (выбор в конфликтной ситуации). Возможны также промежуточные варианты: компромисс, коалиция, нарастающий или затухающий конфликт.

Различные сочетания перечисленных вариантов и приводят к многочисленным задачам принятия решений, которые изучены в различной степени.

1.3.3.2. Языки описания методов принятия решений

Об одном и том же явлении можно говорить на различных языках различной степени общности адекватности. К настоящему времени сложилось три основных языка описания выбора.

Самым простым и наиболее развитым и наиболее популярным является критериальный язык.

1.3.3.2.1. Критериальный язык

Название этого языка связано с основным предположением, состоящим в том, что каждую отдельно взятую альтернативу можно оценить некоторым конкретным (одним) числом, после чего сравнение альтернатив сводится к сравнению соответствующих им чисел.

Пусть, например, $\{X\}$ - множество альтернатив, а x – некоторая определенная альтернатива, принадлежащая этому множеству: $x \in X$. Тогда считается, что для всех x может быть задана функция: $q(x)$, которая называется критерием (критерием качества, целевой функцией, функцией предпочтения, функцией полезности и т.п.), обладающая тем свойством, что если альтернатива x_1 предпочтительнее x_2 : (обозначается: $x_1 > x_2$),

то:

$$q(x_1) > q(x_2).$$

При этом выбор сводится к отысканию альтернативы с наибольшим значением критериальной функции.

Однако, на практике использование лишь одного критерия для сравнения степени предпочтительности альтернатив оказывается неоправданным упрощением, т.к. более подробное рассмотрение альтернатив приводит к необходимости оценивать их не по одному, а по многим критериям, которые могут иметь различную природу и качественно отличаться друг от друга.

Например, при выборе наиболее приемлемого для пассажиров и эксплуатирующей организации типа самолета на определенных видах трасс сравнение идет одновременно по многим группам критериев: техническим, технологическим, экономическим, социальным, эргономическим и др.

Многокритериальные задачи не имеют однозначного общего решения. Поэтому предлагается много способов придать многокритериальной задаче

частный вид, допускающий единственное общее решение. Естественно, что для разных способов эти решения являются в общем случае различными. Поэтому едва ли не главное в решении многокритериальной задачи - обоснование данного вида ее постановки.

Используются различные варианты упрощения многокритериальной задачи выбора. Перечислим некоторые из них.

1. Условная максимизация (находится не глобальный экстремум суперкритерия, или, как его еще называют, *интегрального* критерия, а локальный экстремум основного критерия).

2. Поиск альтернативы с заданными свойствами.

3. Нахождение множества Парето.

4. Сведение многокритериальной задачи к однокритериальной, путем ввода суперкритерия.

Рассмотрим подробнее формальную постановку метода сведения многокритериальной задачи к однокритериальной.

Введем суперкритерий $q_0(x)$, как скалярную функцию векторного аргумента:

$$q_0(x) = q_0((q_1(x), q_2(x), \dots, q_n(x))).$$

Суперкритерий позволяет упорядочить альтернативы по величине q_0 , выделяя тем самым наилучшую (в смысле этого критерия). Вид функции q_0 определяется тем, как конкретно мы представляем себе вклад каждого критерия в суперкритерий. Обычно используют аддитивные и мультипликативные функции:

Коэффициенты s_i обеспечивают:

1. Безразмерность или единую размерность числа $a_i q_i / s_i$ (различные частные критерии могут иметь разную размерность, и тогда над ними нельзя производить арифметических операций и свести их в суперкритерий).

2. Нормировку, т.е. обеспечение условия: $b_i q_i / s_i < 1$.

Коэффициенты a_i и b_i отражают относительный вклад частных критериев q_i в суперкритерий.

Итак, в многокритериальной постановке задача принятия решения о выборе одной из альтернатив сводится к максимизации суперкритерия.

Основная проблема в многокритериальной постановке задачи принятия решений состоит в том, что необходимо найти такой аналитический вид коэффициентов a_i и b_i , который бы обеспечил следующие свойства модели:

1. Высокую степень адекватности предметной области и точке зрения экспертов.

2. Минимальные вычислительные трудности максимизации суперкритерия, т.е. Его расчета для разных альтернатив.

3. Устойчивость результатов максимизации суперкритерия от малых возмущений исходных данных.

Устойчивость решения означает, что малое изменение исходных данных должно приводить к малому изменению величины суперкритерия, и, соответственно, к малому изменению принимаемого решения. То есть практически на тех же исходных данных должно приниматься или тоже самое, или очень близкое решение.

1.3.3.2.2. Язык последовательного бинарного выбора

Язык бинарных отношений является обобщением многокритериального языка и основан на учете того факта, что когда мы даем оценку некоторой альтернативе, то эта оценка всегда является относительной, т.е. явно или чаще неявно в качестве базы или системы отсчета для сравнения используются другие альтернативы из исследуемого множества или из генеральной совокупности. Мышление человека основано на поиске и анализе противоположностей (конструктов), поэтому, нам всегда проще выбрать один из двух противоположных вариантов, чем один вариант из большого и никак неупорядоченного их множества.

Таким образом, основные предположения этого языка сводятся к следующему:

1. Отдельная альтернатива не оценивается, т.е. критериальная функция не вводится.

2. Для каждой пары альтернатив некоторым образом можно установить, что одна из них предпочтительнее другой, либо они равноценны или несравнимы.

3. Отношение предпочтения в любой паре альтернатив не зависит от остальных альтернатив, предъявленных к выбору.

Существуют различные способы задания бинарных отношений: непосредственный, матричный, с использованием графов предпочтений, метод сечений и др.

Отношения между альтернативами одной пары выражают через понятия эквивалентности, порядка и доминирования.

1.3.3.2.3. Обобщенный язык функций выбора

Язык функций выбора основан на теории множеств и позволяет оперировать с отображениями множеств на свои подмножества, соответствующие различным вариантам выбора, без необходимости перечисления элементов. Этот язык является весьма общим и потенциально позволяет описывать любой выбор. Однако, математический аппарат обобщенных функций выбора в настоящее время еще только разрабатывается и проверяется в основном на задачах, которые уже решены с помощью критериального или бинарного подходов.

1.3.3.2.4. Групповой выбор

Пусть имеется группа лиц, имеющих право принимать участие в коллективном принятии решений. Предположим, что эта группа рассматривает некоторый набор альтернатив, и каждый член группы осуществляет свой выбор. Ставится задача о выработке решения, которое определенным образом согласует индивидуальные выборы и в каком-то смысле выражает "общее мнение" группы, т.е. принимается за групповой выбор.

Естественно, различным принципам согласования индивидуальных решений будут соответствовать различные групповые решения.

Правила согласования индивидуальных решений при групповом выборе называются правилами голосования. Наиболее распространенным является

"правило большинства", при котором за групповое решение принимается альтернатива, получившая наибольшее число голосов.

Необходимо понимать, что такое решение отражает лишь **распространенность** различных точек зрения в группе, а не действительно оптимальный вариант, за который вообще никто может и не проголосовать. **"Истина не определяется путем голосования"**, самой распространенной точкой зрения может быть и заблуждение.

Кроме того, существуют так называемые "парадоксы голосования", наиболее известный из которых парадокс Эрроу.

Эти парадоксы могут привести, и иногда действительно приводят, к очень неприятным особенностям процедуры голосования: например бывают случаи, когда группа вообще не может принять единственного решения (нет кворума или каждый голосует за свой уникальный вариант, и т.д.), а иногда (при многоступенчатом голосовании) меньшинство может навязать свою волю большинству, как это было на президентских выборах в США "Буш – Гор".

1.3.3.3. Выбор в условиях неопределенности

Выбор в условиях определенности – это частный случай выбора в условиях неопределенности (когда неопределенность близка к нулю).

Но неопределенность чего конкретно имеется в виду?

В современной теории выбора считается, что в задачах принятия решений существует три основных вида неопределенности:

1. Информационная (статистическая) неопределенность исходных данных для принятия решений.
 2. Неопределенность последствий принятия решений (выбора).
 3. Расплывчатость в описании компонент процесса принятия решений.
- Рассмотрим их по порядку.

1.3.3.3.1. Информационная (статистическая) неопределенность в исходных данных

Данные, полученные о предметной области, не могут рассматриваться как абсолютно точные. Кроме того, очевидно, эти данные нас интересуют не сами по себе, а лишь в качестве сигналов, которые, возможно, несут определенную информацию о том, что нас в действительности интересует.

То есть, реалистичнее считать, что мы имеем дело с данными, не только зашумленными и неточными, но еще и косвенными, а возможно и не полными.

Кроме того эти данные касаются не всей исследуемой (генеральной) совокупности, а лишь определенного ее подмножества, о котором мы смогли фактически собрать данные, однако при этом мы хотим сделать выводы о всей совокупности, причем хотим еще и знать достоверность этих выводов.

В этих условиях используется теория статистических решений.

В этой теории существует два основных источника неопределенности. Во-первых, неизвестно, какому распределению подчиняются исходные данные. Во-вторых, неизвестно, какое распределение имеет то множество (генеральная совокупность), о котором мы хотим сделать выводы по его подмножеству, образующему исходные данные.

Статистические процедуры это и есть процедуры принятия решений, снимающих оба эти виды неопределенности.

Необходимо отметить, что существует ряд причин, которые приводят к некорректному применению статистических методов:

1. Статистические выводы, как и любые другие, всегда имеют некоторую определенную надежность или достоверность. Но, в отличие от многих других случаев, достоверность статистических выводов известна и определяется в ходе статистического исследования.

2. Качество решения, полученного в результате применения статистической процедуры зависит, от качества исходных данных.

3. Не следует подвергать статистической обработке данные, не имеющие статистической природы.

4. Необходимо использовать статистические процедуры, соответствующие уровню априорной информации об исследуемой совокупности (например, не следует применять методы дисперсионного анализа к негауссовым данным). Если распределение исходных данных неизвестно, то надо либо его установить, либо использовать несколько различных методов и сравнить результаты. Если они сильно отличаются - это говорит о неприменимости некоторых из использованных процедур.

1.3.3.3.2. Неопределенность последствий

Когда последствия выбора той или иной альтернативы однозначно определяются самой альтернативой, тогда можно не различать альтернативу и ее последствия, считая само собой разумеющимся, что выбирая альтернативу мы в действительности выбираем ее последствия.

Однако, в реальной практике нередко приходится иметь дело с более сложной ситуацией, когда выбор той или иной альтернативы неоднозначно определяет последствия сделанного выбора.

В случае дискретного набора альтернатив и исходов их выбора, при условии, что сам набор возможных исходов общий для всех альтернатив, можно считать, что различные альтернативы отличаются друг от друга распределением вероятностей исходов. Эти распределения вероятностей вообще говоря могут зависеть от результатов выбора альтернатив и реально наступивших в результате этого исходов. В простейшем случае исходы равновероятны. Сами исходы обычно имеют смысл выигрышей или потерь и выражаются количественно.

Если исходы равны для всех альтернатив, то выбирать нечего. Если же они различны, то можно сравнивать альтернативы, вводя для них те или иные количественные оценки. Разнообразие задач теории игр связано с различным выбором числовых характеристик потерь и выигрышей в результате выбора альтернатив, различными степенями конфликтности между сторонами, выбирающими альтернативы и т.д.

1.3.3.3.3. Расплывчатая неопределенность

Любая задача выбора является задачей целевого сужения множества альтернатив. Как формальное описание альтернатив (сам их перечень, перечень их признаков или параметров), так и описание правил их сравнения (критериев, от-

ношений) всегда даются в терминах той или иной измерительной шкалы (даже тогда, когда тот, кто это делает, не знает об этом).

Известно, все шкалы размыты, но в разной степени. Под термином "размытие" понимается свойство шкал, состоящее в том, что всегда можно предъявить такие две альтернативы, которые различимы, т.е. различны в одной шкале и неразличимы, т.е. тождественны в другой - более размытой. Чем меньше градаций в некоторой шкале, тем более она размыта.

Таким образом, мы можем четко видеть альтернативы, и одновременно нечетко их классифицировать, т.е. иметь неопределенность в вопросе о том, к каким классам они относятся.

Уже в первой работе по принятию решений в расплывчатой ситуации Беллман и Заде выдвинули идею, состоящую в том, что и цели, и ограничения должны представляться как размытые (нечеткие) множества на множестве альтернатив.

1.3.3.4. Решение как компромисс и баланс различных интересов. О некоторых ограничениях оптимизационного подхода

Во всех рассмотренных выше задачах выбора и методах принятия решений проблема состояла в том, чтобы в исходном множестве найти наилучшие в заданных условиях, т.е. оптимальные в определенном смысле альтернативы.

Идея оптимальности является центральной идеей кибернетики и прочно вошла в практику проектирования и эксплуатации технических систем. Вместе с тем эта идея требует осторожного к себе отношения, когда мы пытаемся перенести ее в область управления сложными, большими и слабо детерминированными системами, такими, например, как социально-экономические системы.

Для этого заключения имеются достаточно веские основания. Рассмотрим некоторые из них.

1. Оптимальное решение нередко оказывается неустойчивым: т.е. незначительные изменения в условиях задачи, исходных данных или ограничениях могут привести к выбору существенно отличающихся альтернатив.

2. Оптимизационные модели разработаны лишь для узких классов достаточно простых задач, которые не всегда адекватно и системно отражают реальные объекты управления. Чаще всего оптимизационные методы позволяют оптимизировать лишь достаточно простые и хорошо формально описанные подсистемы некоторых больших и сложных систем, т.е. позволяют осуществить лишь локальную оптимизацию. Однако, если каждая подсистема некоторой большой системы будет работать оптимально, то это еще совершенно не означает, что оптимально будет работать и система в целом. То есть оптимизация подсистемы совсем не обязательно приводит к такому ее поведению, которое от нее требуется при оптимизации системы в целом. Более того, иногда локальная оптимизация может привести к негативным последствиям для системы в целом.

3. Часто максимизация критерия оптимизации согласно некоторой математической модели считается целью оптимизации, однако в действительности целью является оптимизация объекта управления. Критерии оптимизации и ма-

тематические модели всегда связаны с целью лишь косвенно, т.е. более или менее адекватно, но всегда приближенно.

Итак, идею оптимальности, чрезвычайно плодотворную для систем, подающихся адекватной математической формализации, нельзя перенести на сложные системы. Конечно, математические модели, которые удается иногда предложить для таких систем, можно оптимизировать. Однако всегда следует учитывать сильную упрощенность этих моделей, а также то, что степень их адекватности фактически неизвестна. Поэтому не известно, какое чисто практическое значение имеет эта оптимизация. Высокая практичность оптимизации в технических системах не должна порождать иллюзий, что она будет настолько же эффективна при оптимизации сложных систем. Содержательное математическое моделирование сложных систем является весьма затруднительным, приближительным и неточным. Чем сложнее система, тем осторожнее следует относиться к идее ее оптимизации.

Поэтому, при разработке методов управления сложными, большими слабо детерминированными системами, основным является не оптимальность выбранного подхода с формальной математической точки зрения, а его адекватность поставленной цели и самому характеру объекта управления.

1.3.3.5. Экспертные методы выбора

При исследовании сложных систем часто возникают проблемы, которые по различным причинам не могут быть строго поставлены и решены с применением разработанного в настоящее время математического аппарата. В этих случаях прибегают к услугам экспертов (системных аналитиков), чей опыт и интуиция помогают уменьшить сложность проблемы.

Однако, необходимо учитывать, что эксперты сами представляют собой сверхсложные системы, и их деятельность сама зависит от многих внешних и внутренних условий. Поэтому в методиках организации экспертных оценок большое внимание уделяется созданию благоприятных внешних и психологических условий для работы экспертов.

На работу эксперта оказывают влияние следующие факторы:

- ответственность за использование результатов экспертизы;
- знание того, что привлекаются и другие эксперты;
- наличие информационного контакта между экспертами;
- межличностные отношения экспертов (если между ними есть информационный контакт);
- личная заинтересованность эксперта в результатах оценки;
- личностные качества экспертов (самолюбие, конформизм, воля и др.)

Взаимодействие между экспертами может как стимулировать, так и подавлять их деятельность. Поэтому в разных случаях используют различные методы экспертизы, отличающиеся характером взаимодействия экспертов друг с другом: анонимные и открытые опросы и анкетирования, совещания, дискуссии, деловые игры, мозговой штурм и т.д.

Существуют различные методы математической обработки мнений экспертов. Экспертам предлагают оценить различные альтернативы либо одним, либо системой показателей. Кроме того им предлагают оценить степень важно-

сти каждого показателя (его "вес" или "вклад"). Самим экспертам также приписывается уровень компетентности, соответствующий его вкладу в результирующее мнение группы.

Развитой методикой работы с экспертами является метод "Дельфи". Основная идея этого метода состоит в том, что критика и аргументация благотворно влияет на эксперта, если при этом не задевается его самолюбие и обеспечиваются условия, исключающие персональную конфронтацию.

Необходимо особо подчеркнуть, что существует принципиальное различие в характере использования экспертных методов в экспертных системах и в поддержке принятия решений. Если в первом случае от экспертов требуется формализация способов принятия решений, то во втором, лишь само решение, как таковое.

Поскольку эксперты привлекаются для реализации именно тех функций, которые в настоящее время или вообще не обеспечиваются автоматизированными системами, или выполняются ими хуже, чем человеком, то перспективным направлением развития автоматизированных систем является максимальная автоматизация этих функций.

1.3.3.6. Юридическая ответственность за решения, принятые с применением систем поддержки принятия решений

Необходимо отметить, что система поддержки принятия решений (СППР) не является физическим или юридическим лицом и не может нести ответственность за те или иные решения, которые принимаются с ее использованием.

Согласно действующему сейчас и в обозримой перспективе законодательству автоматизированную систему невозможно привлечь к административной или уголовной ответственности. Но это не означает, что в будущем подобные вопросы не возникнут по поводу созданных методами генной инженерии *искусственных разумных биологических систем*, которые автором совместно с Л.А.Бакурадзе в 1980 году были названы "квазибиологические роботы", или по поводу клонов, не ассоциированных в общечеловеческую социальную среду.

Ответственность за принятое решение всегда несет специалист, или, как говорят, "лицо, принимающее решения" (ЛПР).

Вместе с тем при выработке решения это лицо может обращаться к различной справочной информации, а также консультироваться со специалистами в различных предметных областях. При этом справочная информация может извлекаться по запросам из *баз данных*, а консультации могут происходить не с самими специалистами лично, а с автоматизированными системами поддержки принятия решений, в которых в формализованном и обобщенном виде содержатся знания многих специалистов.

Но в какой форме содержаться консультирующая информация и знания – это сути дела не меняет, т.е. *ответственность за принятое решение все равно всегда лежит на лице, его принявшем, а не на консультирующих или поддерживающих решения системах*. Поэтому эти системы и называются не "системы принятия решений", а лишь "системы **поддержки принятия решений**".

Особым случаем являются автоматизированные системы управления оружием, например, создаваемые в США в рамках программы "Стратегической оборонной инициативы" (СОИ). В этих системах человек может участвовать в процессе принятия решений лишь на этапах их создания, т.к. в боевом дежурстве систем для участия человека просто не будет времени. И тем ни менее и эти системы принимают решения не самостоятельно, а на основе анализа фактической информации по тем критериям и алгоритмам, которые заложил в них человек. Таким образом, в подобных системах человек становится заложником своих же решений, принятых им ранее.

1.3.3.7. Условия корректности использования систем поддержки принятия решений

Этих условий три:

1. Само решение о выборе той или иной конкретной системы поддержки принятия решений должно приниматься лицом, принимающим решения, который и будет пользоваться ее рекомендациями, либо подчиненными ему компетентными специалистами по его личному поручению.

2. Сам выбор системы поддержки принятия решений должен осуществляться, как правило, не по их специальным, или, тем более, рекламным описаниям и литературным данным, а по результатам сравнительных испытаний на реальных примерах из практики работы организации.

3. Выбранная система должна быть официально принятой для решения тех задач, для решения которых она будет использоваться, т.е. должна быть *сертифицирована*.

Под сертификацией понимается:

– апробация системы поддержки принятия решений и оценка адекватности рекомендуемых ей решений на ряде реальных задач, которые возникали в прошлом или возникают в течение определенного периода экспериментальной эксплуатации;

– придание системе поддержки принятия решений *юридического статуса*, который предписывает и лицам принимающим решения "принимать к сведению" рекомендации системы.

Для сертификации системы создается соответствующая компетентная и полномочная комиссия, которая в течение определенного времени изучает систему и выдаваемые ею рекомендации на предмет адекватности складывающимся ситуациям. Результаты работы комиссии оформляются в форме юридического

документа, имеющего силу в ведомстве или организации, в которых будет применяться система.

В этом документе описывается круг задач, для решения которых применяется система, инфраструктура и технология ее применения, а также определяются обязанности и персональная ответственность специалистов по эксплуатации системы.

Если система не сертифицирована, то ее применение некорректно с юридической точки зрения. На практике это происходит гораздо чаще, чем многие представляют, например, с системами бухгалтерского учета или психологическими тестами, которые распространяются на пиратских компакт-дисках, не локализованы для регионов и организаций применения, не адаптированы для решения задач, для решения которых они фактически применяются.

1.3.3.8. Хранилища данных для принятия решений

В разделе 4.2.5. данной работы мы рассматривали иерархическую систему обработки информации в которой на различных уровнях производятся различные операции по обработке данных, информации и знаний:

- на 1-м уровне накапливаются *данные* мониторинга;
- на 2-м уровне осуществляется анализ данных мониторинга с целью выявления в них зависимостей, что позволяет содержательно *интерпретировать* данные, т.е. *генерировать информацию путем анализа данных*;
- на 3-м уровне знание зависимостей в данных мониторинга используется для прогнозирования;
- на 4-м уровне возможности многовариантного прогнозирования и решения обратной задачи прогнозирования позволяют вырабатывать рекомендации и решения по достижению поставленных *целей*, т.е. *генерировать и использовать знания путем системной обработки информации*.

Выполнение операций каждого последующего уровня возможно только построения предыдущего уровня. Здесь уместно провести аналогию со строительством здания: пока не выполнен фундамент – не возводят стены, пока не возведены стены – не делают крышу, пока нет крыши – не проводят отделку и т.д. Аналогично, чтобы вытащить внутреннюю матрешку сначала надо раскрыть внешнюю. Могут существовать и более сложные алгоритмы, определяющие последовательность, например, типа используемых в игре "Ханойская башня".

Таким образом, фундаментом для генерации информации и знаний являются данные мониторинга.

Хранилище Данных (ХД или Data warehouses) – это база данных, хранящая данные, агрегированные по многим измерениям. Данные из ХД никогда не удаляются. Пополнение ХД происходит на периодической основе. При этом автоматически формируются новые агрегаты данных, зависящие от старых. Доступ к ХД организован особым образом на основе модели многомерного куба.

Итак, Хранилище Данных – это не автоматизированная система принятия решений, не экспертная система, не система логического вывода, а "всего лишь" оптимально организованная база данных, обеспечивающая максимально быстрый и комфортный доступ к информации, необходимой при принятии решений.

Принять любое управленческое решение, невозможно не обладая необходимой для этого информацией, обычно количественной. Для этого необходимо создание хранилищ данных (Data warehouses), то есть процесс сбора, отсеивания и предварительной обработки данных с целью предоставления результирующей информации пользователям для статистического анализа (а нередко и создания аналитических отчетов). Ральф Кимбалл (Ralph Kimball), один из авторов концепции хранилищ данных сформулировал основные требования к ним:

- поддержка высокой скорости получения данных из хранилища;
- поддержка внутренней непротиворечивости данных;
- возможность получения и сравнения так называемых срезов данных (slice and dice);
- наличие удобных утилит просмотра данных в хранилище;
- полнота и достоверность хранимых данных;
- поддержка качественного процесса пополнения данных.

Типичное хранилище данных, как правило, отличается от обычной реляционной базы данных.

Во-первых, обычные базы предназначены для того, чтобы помочь пользователям выполнять повседневную работу, тогда как хранилища данных предназначены для принятия решений. Например, продажа товара и выписка счета производятся с использованием базы данных, предназначенной для обработки транзакций, а анализ динамики продаж за несколько лет, позволяющий спланировать работу с поставщиками, - с помощью хранилища данных.

Во-вторых, обычные базы данных подвержены постоянным изменениям в процессе работы пользователей, а хранилище данных относительно стабильно: сведения в нем обычно обновляются согласно расписанию (например, еженедельно, ежедневно или ежечасно – в зависимости от потребностей). В идеале процесс пополнения представляет собой просто добавление новых данных за определенный период времени без изменения прежней информации, уже находящейся в хранилище.

В-третьих, обычные базы данных чаще всего являются источником данных, попадающих в хранилище. Кроме того, хранилище может пополняться за счет внешних источников, например статистических отчетов.

Для более полного ознакомления с *концепцией хранилищ данных* рекомендуется обратиться к Internet-источникам.

Контрольные вопросы

1. Многообразие задач принятия решений.
2. Принятие решений, как реализация цели.
3. Принятие решений, как снятие неопределенности (информационный подход).

4. Связь принятия решений и распознавания образов.
5. Классификация задач принятия решений.
6. Языки описания методов принятия решений.
7. Критериальный язык.
8. Язык последовательного бинарного выбора.
9. Обобщенный язык функций выбора.
10. Групповой выбор.
11. Выбор в условиях неопределенности.
12. Информационная (статистическая) неопределенность в исходных данных.
13. Неопределенность последствий.
14. Расплывчатая неопределенность.
15. Решение как компромисс и баланс различных интересов. О некоторых ограничениях оптимизационного подхода.
16. Экспертные методы выбора.
17. Юридическая ответственность за решения, принятые с применением систем поддержки принятия решений.
18. Условия корректности использования систем поддержки принятия решений.
19. Хранилища данных для принятия решений.

Рекомендуемая литература

1. Перегудов Ф.И., Тарасенко Ф.П. Введение в системный анализ: Учебное пособие. – М.: Высшая школа, 1997. – 389с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.
3. Бирюков А. Системы принятия решений и Хранилища Данных.
//Системы управления базами данных #04/97.
<http://www.isuct.ru/~ivt/books/DBMS/DBMS7/dbms/1997/04/37.htm>
4. Львов В. Создание систем поддержки принятия решений на основе хранилищ данных. Ж-л "Системы управления базами данных", #03, 1997 г.//Издательство "Открытые системы" (www.osp.ru). Адрес статьи: <http://www.osp.ru/dbms/1997/03/30.htm>

1.3.4. ЛЕКЦИЯ-10.

Экспертные системы

Учебные вопросы

1. Экспертные системы, базовые понятия.
2. Экспертные системы, методика построения.
3. Этап-1 синтеза ЭС: "Идентификация".
4. Этап-2 синтеза ЭС: "Концептуализация".
5. Этап-3 синтеза ЭС: "Формализация".
6. Этап-4 синтеза ЭС: "Разработка прототипа".
7. Этап-5 синтеза ЭС: "Экспериментальная эксплуатация".
8. Этап-6 синтеза ЭС: "Разработка продукта".
9. Этап-7 синтеза ЭС: "Промышленная эксплуатация".

Данный раздел основан на конспекте лекций по курсу "Основы проектирования систем искусственного интеллекта": Сотник С.Л. (1997-1998), который несложно найти Internet, например, по адресу:

<http://neuroschool.narod.ru/books/sotnik.html> [237].

1.3.4.1. Экспертные системы, базовые понятия

Экспертная система (ЭС) – это программа, которая в определенных отношениях заменяет эксперта или группу экспертов в той или иной предметной области.

ЭС предназначены для решения практических задач, возникающих в слабо структурированных и трудно формализуемых предметных областях.

Исторически, ЭС были первыми системами искусственного интеллекта, которые привлекли внимание потребителей.

С ЭС связаны некоторые распространенные заблуждения.

Заблуждение первое: ЭС могут делать не более, а скорее даже менее того, чем эксперт, создавший данную систему.

Во-первых, существуют технологии синтеза самообучающихся ЭС, которые могут быть применены в предметной области, в которой вообще нет экспертов.

Во-вторых, технология ЭС позволяет **объединить** в одной системе знания **нескольких** экспертов, и, таким образом, в результате получить систему, которая может то, чего ни один из ее создателей не может.

Заблуждение второе: ЭС никогда не заменит человека-эксперта.

На практике часто ЭС могут создаваться и применяться для решения задач, в решении которых эксперты по ряду причин физического, юридического, финансового и организационного характера не могут принять **личного** участия, т.е. в точках, весьма удаленных от экспертов как в пространстве, так и во времени:

- знания могут извлекаться из научных работ или фактических данных, доступ к которым может обеспечиваться через Internet;
- доступ к ЭС и ее базе знаний также может быть получен через Internet.

1.3.4.2. Экспертные системы, методика построения

В настоящее время сложилась определенная технология разработки ЭС, которая включает следующие шесть этапов:

1. Идентификация.
2. Концептуализация.
3. Формализация.
4. Разработка прототипа.
5. Экспериментальная эксплуатация.
6. Разработка продукта.
7. Промышленная эксплуатация.

1.3.4.3. Этап-1: "Идентификация"

На этапе идентификации производится:

- неформальное осмысление задач, которые должна решать создаваемая ЭС;
- формирование требований к ЭС;
- определение ресурсов, необходимых для создания ЭС.

В результате идентификации *функционально* определяется что должна делать ЭС и что необходимо для ее создания.

Идентификация задачи заключается в составлении неформального (вербального, т.е. словесного) описания, в котором указываются:

- общие характеристики задачи;
- подзадачи, выделяемые внутри данной задачи;
- ключевые понятия (объекты), их входные и выходные данные;
- предположительный вид решения;
- знания, относящиеся к решаемой задаче.

В процессе идентификации задачи инженер по знаниям и эксперт работают в тесном контакте.

Начальное неформальное описание задачи, данное экспертом, затем используется инженером знаний для уточнения терминов и ключевых понятий.

Эксперт корректирует описание задачи, объясняет, как решать ее и какие рассуждения лежат в основе того или иного решения.

После нескольких циклов, уточняющих описание, эксперт и инженер по знаниям получают окончательное неформальное описание задачи.

При создании ЭС основными видами ресурсов являются:

- источники знаний (эксперты);
- инженеры знаний и программисты;
- инструментальные программные средства (экспертные оболочки);
- вычислительные средства;
- время разработки;
- объем финансирования.

1.3.4.4. Этап-2: "Концептуализация"

На данном этапе проводится *содержательный* анализ проблемной области, выявляются используемые понятия и их взаимосвязи, определяются методы решения задач.

Этот этап завершается созданием *модели* предметной области, включающей основные концепты и отношения между ними. На этапе концептуализации определяются следующие особенности задачи:

- типы доступных данных;
- исходные и выводимые данные,
- подзадачи общей задачи;
- используемые стратегии и гипотезы;
- виды взаимосвязей между объектами ПО, типы используемых отношений (иерархия, причина – следствие, часть – целое и т.п.);
- процессы, используемые в ходе решения;
- состав знаний, используемых при решении задачи;
- типы ограничений, накладываемых на процессы, используемые в ходе решения;
- состав знаний, используемых для обоснования решений.

Существует два подхода к процессу построения модели предметной области:

1. Атрибутивный подход (атрибутами называют существенные признаки) предполагает наличие полученной от экспертов информации в виде цепочек: "*Класс (градация классификационной шкалы) – объект обучающей выборки – атрибут (описательная шкала) – значение атрибута (градация описательной шкалы)*". Этот подход развивается в рамках направления, получившего название формирование знаний или "машинное обучение" (*machine learning*).

2. Структурный или когнитивный подход, основан на выделении элементов предметной области, их взаимосвязей и семантических (смысловых) отношений.

Атрибутивный подход требует полной информации о предметной области: об объектах, их атрибутах и о значениях атрибутов, а также *дополнительной* обучающей информации о принадлежности конкретных объектов к обобщенным классам, задаваемой экспертом. Отметим, что атрибутивный подход в экспертных системах имеет очень много общего с методами, применяемыми в распознавании образов.

Структурный подход к построению модели предметной области предполагает выделение следующих когнитивных элементов знаний:

1. Понятия.
2. Взаимосвязи.
3. Метапонятия.
4. Семантические отношения.

Выделяемые понятия предметной области должны образовывать систему, под которой понимается совокупность понятий, обладающая следующими свойствами:

- минимальностью (уникальностью, отсутствием избыточности);

- полнотой (достаточно полным описанием различных процессов, фактов, явлений предметной области);
- достоверностью (адекватностью, валидностью – соответствием выделенных единиц смысловой информации их реальным наименованиям).

Существует ряд методов выявления иерархической системы понятий и метапонятий (включая отношения между ними), позволяющей адекватно отразить предметную область:

1. Метод локального представления.
2. Метод вычисления коэффициента использования.
3. Метод формирования перечня понятий.
4. Ролевой метод.
5. Метода составления списка элементарных действий.
6. Метод составления оглавления учебника.
7. Текстологический метод.
8. Метод свободных ассоциаций для определения "смыслового расстояния" между понятиями.
9. Метод "сортировки карточек".
10. Метод обнаружения регулярностей.
11. Методы семантического дифференциала и репертуарных решеток.

Перечисленные методы применяются на этапе концептуализации при построении модели предметной области. Подробнее они описаны в упомянутой выше работе С.Л. Сотника.

1.3.4.5. Этап-3: "Формализация"

Этап формализации необходим для преобразования декларативных и процедурных знаний о предметной области, полученных на этапе концептуализации, в форму, пригодную для их обработки на компьютере.

На данном этапе:

- выбирается или разрабатывается формальный язык, обеспечивающий представление знаний и манипулирование ими;
- осуществляется *формализация* знаний, т.е. они преобразуются в форму, пригодную для обработки на компьютере.

Способы представления знаний: фреймы, сценарии, семантические сети, продукции.

Способы манипулирования знаниями: логический вывод, аналитическая модель, статистическая модель.

1.3.4.6. Этап-4: "Разработка прототипа"

Разработка прототипа включает три основных этапа:

1. Программная реализация системы, призванной обеспечить реальное решение поставленных задачи.
2. Наполнение базы знаний.
3. Тестирование (исследование) прототипа.

Исследование прототипа позволяет:

- оценить насколько реализованные в нем идеи, методы и способы представления знаний пригодны для решения поставленных задач;

– продемонстрировать тенденцию к повышению качества и эффективности решений для всех задач предметной области по мере увеличения объема знаний.

Положительные результаты тестирования прототипа являются основанием для его передачи в экспериментальную эксплуатацию.

1.3.4.7. Этап-5: "Экспериментальная эксплуатация"

На этапе экспериментальной эксплуатации прототипа экспертной системы с помощью нее решаются реальные задачи, однако, на практике результаты их решения не используются. При этом экспертной системой управляют представители разработчика и квалифицированные пользователи, прошедшие обучение у разработчика.

При этом круг предлагаемых для решения задач естественно расширяется, исправляются ошибки, собираются пожелания и замечания экспертов и пользователей, которые должны быть учтены в очередной версии экспертной системы.

В частности на этом этапе детализируются направления будущего развития экспертной системы путем добавления в нее:

- "дружественного" пользовательского интерфейса с системой контекстно-зависимых подсказок;
- развитых средств исследования и графического представления базы знаний и цепочек выводов, генерируемых системой;
- средств обеспечения адаптивности базы знаний;
- диспетчера задач, решаемых в системе и средств поддержки и использования архива уже созданных приложений.

1.3.4.8. Этап-6: "Разработка продукта"

На этом этапе программная реализация прототипа экспертной системы доводится до уровня программного продукта, который может успешно использоваться заказчиком *без прямой помощи разработчиков*.

При этом разрабатываются программные модули, поддерживающие возможности экспертной системы, определенные на предыдущем этапе, а также проводится:

- анализ функционирования системы при значительном расширении базы знаний;
- исследование возможностей системы в решении более широкого круга задач и принятие мер для обеспечения таких возможностей;
- анализ мнений пользователей о функционировании ЭС;
- разработка системы ввода-вывода, осуществляющей анализ или синтез предложений ограниченного естественного языка.

Пригодность ЭС для пользователя определяется удобством работы с ней и ее полезностью.

Полезность ЭС – это ее способность эффективно решать поставленные пользователем задачи.

Удобство работы с ЭС включает:

- естественность интерфейса (общение в привычном, не утомляющем пользователя виде);
- гибкость ЭС (способность системы настраиваться на различные задачи, а также учитывать изменения в квалификации пользователей);
- устойчивость системы к ошибкам (способность не выходить из строя при попадании неадекватных, т.е. неполных и зашумленных знаний в базу знаний, а также при ошибочных действиях пользователей).

1.3.4.9. Этап-7: "Промышленная эксплуатация"

После доведения экспертной системы до уровня программного продукта она передается в промышленную эксплуатацию. При этом задачи не только решаются экспертной системой, но и результаты решения используются на практике. Для этого при передаче системы в промышленную эксплуатацию она тестируется и сертифицируется заказчиком, как инструмент, пригодный для решения поставленных задач, т.е. экспертной системе и результатам ее работы присваивается соответствующий юридический статус. Экспертной системой управляют рядовые пользователи, прошедшие обучение у квалифицированных пользователей. При необходимости квалифицированным пользователям могут оказывать консультативную помощь представители разработчика.

Контрольные вопросы

1. Экспертные системы, базовые понятия.
2. Экспертные системы, методика построения.
3. Этап-1 синтеза ЭС: "Идентификация".
4. Этап-2 синтеза ЭС: "Концептуализация".
5. Этап-3 синтеза ЭС: "Формализация".
6. Этап-4 синтеза ЭС: "Разработка прототипа".
7. Этап-5 синтеза ЭС: "Экспериментальная эксплуатация".
8. Этап-6 синтеза ЭС: "Разработка продукта".
9. Этап-7 синтеза ЭС: "Промышленная эксплуатация".

Рекомендуемая литература

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

1.3.5. ЛЕКЦИЯ-11. Нейронные сети

Учебные вопросы

1. Биологический нейрон и формальная модель нейрона Маккалоки и Питтса.
2. Возможность решения простых задач классификации непосредственно одним нейроном.
3. Однослойная нейронная сеть и персептрон Розенблата.
4. Линейная разделимость и персептронная представляемость.
5. Многослойные нейронные сети.
6. Проблемы и перспективы нейронных сетей.
7. Модель нелокального нейрона и нелокальные интерпретируемые нейронные сети прямого счета.

В разделах 1.3.5.1 и 1.3.5.3 широко использованы материалы лекций по теории и приложениям искусственных нейронных сетей, размещенные в Internet Сергеем А. Тереховым, Лаборатория Искусственных Нейронных Сетей НТО-2, ВНИИТФ, Снежинск (http://alife.narod.ru/lectures/neural/Neu_index.htm) [237].

1.3.5.1. Биологический нейрон и формальная модель нейрона Маккалоки и Питтса

Биологический нейрон имеет вид, представленный на рисунке 73:

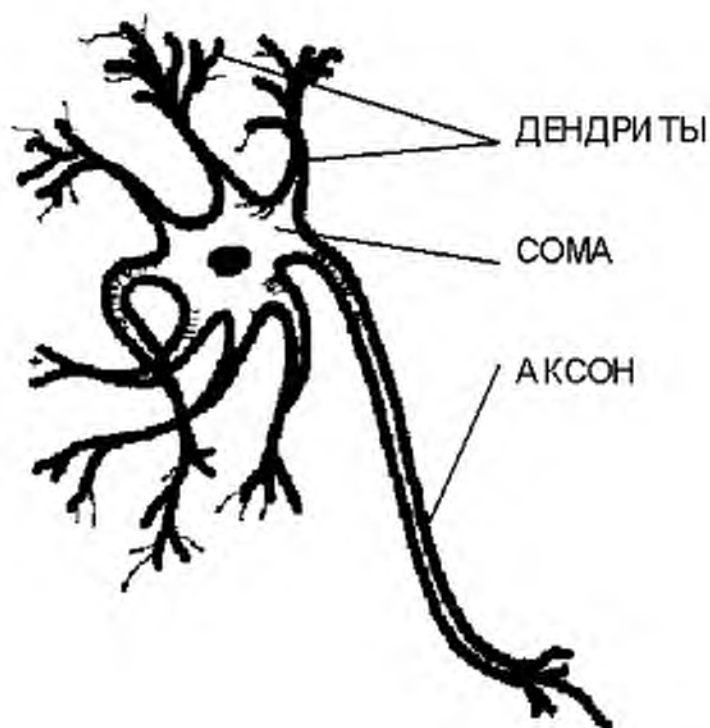


Рисунок 73. Структура биологического нейрона

В 1943 году Дж. Маккалоки и У. Питт предложили формальную модель биологического нейрона как устройства, имеющего несколько входов (входные синапсы – дендриты), и один выход (выходной синапс – аксон) (рисунок 74).

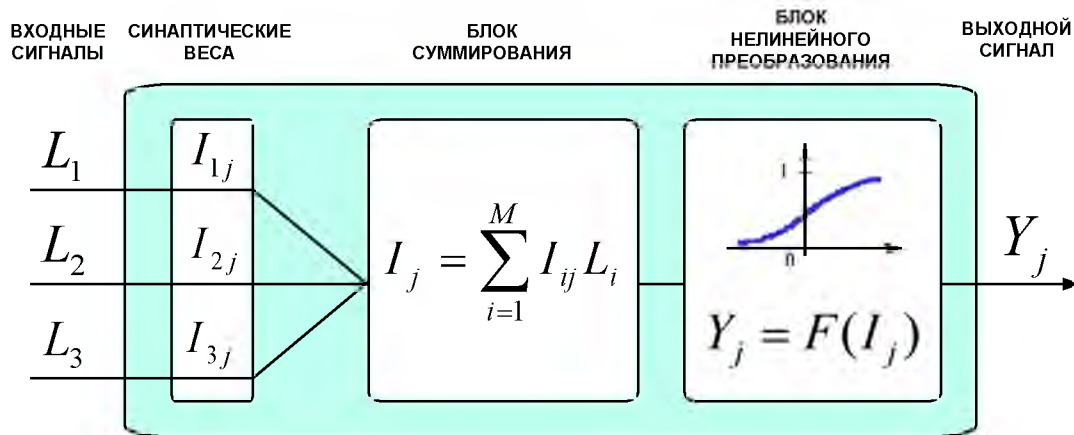


Рисунок 74. Классическая модель нейрона Дж. Маккалоки и У. Питта (1943) в обозначениях системной теории информации

Дендриты получают информацию от источников информации (рецепторов) L_i , в качестве которых могут выступать и нейроны. Набор входных сигналов $\{L_i\}$ характеризует объект, его состояние или ситуацию, обрабатываемую нейроном.

Каждому i -му входу j -го нейрона ставится в соответствие некоторый весовой коэффициент I_{ij} , характеризующий *степень влияния сигнала с этого входа на аргумент передаточной (активационной) функции, определяющей сигнал Y_j на выходе нейрона*. В нейроне происходит взвешенное суммирование входных сигналов, и далее это значение используется как *аргумент активационной (передаточной) функции нейрона*. На рисунке 74 данная модель приведена в обозначениях, принятых в настоящей работе.

1.3.5.2. Возможность решения простых задач классификации непосредственно одним нейроном

Представим себе, что необходимо решать задачу *определения пола студентов по их внешне наблюдаемым признакам*.

Есть, конечно, и более надежные способы, но мы их рассматривать не будем, т.к. они требуют дополнительных затрат для получения исходной информации и превращают задачу в тривиальную.

Поэтому будем рассматривать такие описательные шкалы и градации:

1. *Длина волос*: длинные, средние, короткие.
2. *Наличие брюк*: да, нет.
3. *Использование духов или одеколона*: да, нет.

Составим таблицу для определения весовых коэффициентов (таблица 33). Пусть столбцы этой таблицы соответствуют состояниям нейрона, а строки – дендритам, соединенным с соответствующими органами восприятия, которые способны устанавливать наличие или отсутствие соответствующего признака.

Тогда один из простейших способов определить значения весовых коэффициентов на дендритах будет заключаться в том, чтобы на пересечениях строк и столбцов просто проставить суммарное количество студентов в обучающей выборке, обладающих данным признаком.

**Таблица 33 – ОПРЕДЕЛЕНИЕ ВЕСОВЫХ
КОЭФФИЦИЕНТОВ НЕЙРОНОВ НЕПОСРЕДСТВЕННО
НА ОСНОВЕ ЭМПИРИЧЕСКИХ ДАННЫХ**

Описательные шкалы и градации	Классификационные шкалы и градации	
	Юноши	Девушки
<i>Длина волос:</i>		
– длинные	5	15
– средние	10	10
– короткие	15	5
<i>Наличие брюк:</i>		
– да;	30	10
– нет	0	20
<i>Использование духов или одеколона:</i>		
– да;	5	20
– нет	25	10

Если нейрон должен выдавать высокий выходной сигнал, когда на входе ему предъявляется юноша и низкий – когда девушка, то весовые коэффициенты на дендритах берутся из столбца: "Юноши". И наоборот, если нейрон должен выдавать высокий выходной сигнал, когда на входе ему предъявляется девушка и низкий – когда юноша, то весовые коэффициенты на дендритах берутся из столбца: "Девушки".

Можно представить себе сеть из двух нейронов, в которой весовые коэффициенты на дендритах взяты из столбцов: "Юноши" и "Девушки".

Большее количество нейронов для решения данной задачи будет избыточным. Его имеет смысл использовать в том случае, когда мы хотим повысить надежность идентификации объектов нейронной сетью и различные сходные по смыслу нейроны будут использовать *независимые* друг от друга рецепторы.

Например, если мы не только видим идентифицируемый объект, но можем его и обонять, и ощупывать, то это повышает надежность его идентификации. В этом состоит общепринятый в физике критерий реальности – *принцип наблюдаемости*, согласно которому объективное существование установлено для тех объектов и явлений, существование которых установлено несколькими, по крайней мере, двумя, независимыми способами.

В общем случае в нейронной сети каждому классу (градации классификационной шкалы) будет соответствовать один нейрон и объект, признаки которого будут измерены рецепторами на входе нейронной сети, будет идентифицирован сетью как класс, соответствующий нейрону с максимальным уровнем сигнала на выходе.

Психологические тесты обычно позволяют тестировать респондента сразу по нескольким шкалам. Очевидно, *нейронные сети, реализующие эти тесты,*

будут иметь как минимум столько нейронов, сколько шкал в психологическом тесте.

1.3.5.3. Однослойная нейронная сеть и персептрон Розенблатта

Исторически первой искусственной нейронной сетью, способной к перцепции (восприятию) и формированию реакции на воспринятый стимул, явился *Perceptron* Розенблатта (F.Rosenblatt, 1957). Термин "*Perceptron*" происходит от латинского *perceptio*, что означает восприятие, познание. Русским аналогом этого термина является "Персептрон". Его автором персептрон рассматривался не как конкретное техническое вычислительное устройство, а как модель работы мозга. Современные работы по искусственным нейронным сетям редко преследуют такую цель.

Простейший классический персептрон содержит элементы трех типов (рисунок 75), назначение которых в целом соответствует нейрону рефлекторной нейронной сети, рассмотренному выше.

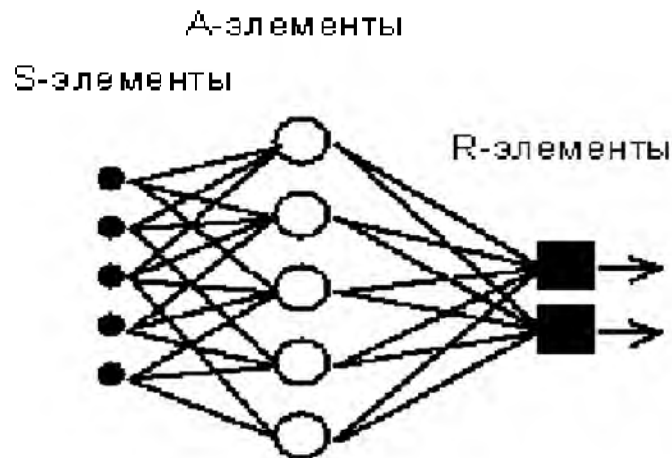


Рисунок 75. Элементарный персептрон Розенблатта

S-элементы – это сенсоры или рецепторы, принимающие *двоичные* сигналы от внешнего мира. Каждому S-элементу соответствует определенная градация некоторой описательной шкалы.

Далее сигналы поступают в слой ассоциативных или A-элементов (показана часть связей от S к A-элементам). Только ассоциативные элементы, представляющие собой формальные нейроны, выполняют *совместную аддитивную* обработку информации, поступающей от ряда S-элементов с учетом *изменяемых* весов связей (рисунок 75). Каждому A-элементу соответствует определенная градация некоторой классификационной шкалы.

R-элементы с фиксированными весами формируют сигнал реакции персептрона на входной стимул. R-элементы *обобщают* информацию о реакциях нейронов на входной объект, например могут выдавать сигнал об идентификации данного объекта, как относящегося к некоторому классу только в том случае, если *все* нейроны, соответствующие этому классу выдадут результат именно о такой идентификации объекта. Это означает, что в R-элементах может исполь-

зоваться *мультипликативная* функция от выходных сигналов нейронов. R-элементы также, как и A-элементы, соответствует определенным градациям классификационных шкал.

Розенблатт считал такую нейронную сеть трехслойной, однако по современной терминологии, представленная сеть является однослойной, так как имеет только один слой *нейропроцессорных* элементов.

Если бы R-элементы были тождественными по функциям A-элементам, то нейронная сеть классического персептрона была бы двухслойной. Тогда бы A-элементы выступали для R-элементов в роли S-элементов.

Однослойный персептрон характеризуется матрицей синаптических связей $\|W\|$ от S- к A-элементам. Элемент матрицы отвечает связи, ведущей от i -го S-элемента (*строки*) к j -му A-элементу (*столбцы*). Эта матрица очень напоминает матрицы абсолютных частот и информативностей, формируемые в семантической информационной модели, основанной на системной теории информации.

С точки зрения современной нейроинформатики однослойный персептрон представляет в основном чисто исторический интерес, вместе с тем на его примере могут быть изучены основные понятия и простые алгоритмы обучения нейронных сетей.

Обучение классической нейронной сети состоит в *подстройке* весовых коэффициентов каждого нейрона.

Пусть имеется набор *пар* векторов $\{x^\alpha, y^\alpha\}$, $\alpha = 1..p$, называемый *обучающей выборкой*, состоящей из p объектов.

Вектор $\{x^\alpha\}$ характеризует систему признаков конкретного объекта α обучающей выборки, зафиксированную S-элементами.

Вектор $\{y^\alpha\}$ характеризует картину возбуждения нейронов при предъявлении нейронной сети конкретного объекта α обучающей выборки:

$$x_i^\alpha = \begin{cases} 1, & \text{если у объекта } \alpha \text{ наблюдается } i - \text{й признак,} \\ 0, & \text{если у объекта } \alpha \text{ } i - \text{й признак не наблюдается;} \end{cases}$$

$$y_j^\alpha = \begin{cases} 1, & \text{если при предъявлении объекта } \alpha \text{ активизируется } j - \text{й нейрон,} \\ 0, & \text{если при предъявлении объекта } \alpha \text{ } j - \text{й нейрон не активизируется;} \end{cases}$$

Будем называть нейронную сеть *обученной* на данной обучающей выборке, если при подаче на вход сети вектора $\{x^\alpha\}$ на выходе всегда получается соответствующий вектор $\{y^\alpha\}$, т.е. каждому набору признаков соответствуют определенные классы.

Ф.Розенблаттом предложен итерационный алгоритм обучения из 4-х шагов, который состоит в подстройке матрицы весов, последовательно уменьшающей ошибку в выходных векторах:

Шаг 0: Начальные значения весов всех нейронов полагаются *случайными*.

Шаг 1: Сети предъявляется входной образ x^α , в результате формируется выходной образ.

Шаг 2: Вычисляется вектор ошибки, делаемой сетью на выходе.

- Шаг 3:** Вектора весовых коэффициентов корректируются таким образом, что величина корректировки пропорциональна ошибке на выходе и равна нулю если ошибка равна нулю:
- модифицируются только компоненты матрицы весов, отвечающие ненулевым значениям входов;
 - знак приращения веса соответствует знаку ошибки, т.е. положительная ошибка (значение выхода меньше требуемого) проводит к усилению связи;
 - обучение каждого нейрона происходит *независимо* от обучения остальных нейронов, что соответствует важному с биологической точки зрения, принципу *локальности* обучения.
- Шаг 4:** Шаги 1-3 повторяются для всех обучающих векторов. Один цикл последовательного предъявления всей выборки называется эпохой. Обучение завершается по истечении нескольких эпох, если выполняется по крайней мере одно из условий:
- когда итерации сойдутся, т.е. вектор весов перестает изменяться;
 - когда полная просуммированная по всем векторам абсолютная ошибка станет меньше некоторого малого значения.

Данный метод обучения был назван Ф.Розенблаттом "*методом коррекции с обратной передачей сигнала ошибки*". Имеется в виду передача сигнала ошибка от выхода сети на ее вход, где и определяются, и используются весовые коэффициенты. Позднее этот алгоритм назвали "*δ-правилом*".

Данный алгоритм относится к широкому классу алгоритмов *обучения с учителем*, т.к. в нем считаются известными не только входные вектора, но и значения выходных векторов, т.е. имеется учитель, способный оценить правильность ответа ученика, причем в качестве последнего выступает нейронная сеть.

Розенблаттом доказана "Теорема о сходимости обучения" по *δ-правилу*. Эта теорема говорит о том, что персептрон способен обучиться любому обучающему набору, который он *способен* представить. Но она ничего не говорит о том, какие именно обучающие наборы он способен представить. Ответ на этот вопрос мы получим в следующем разделе.

1.3.5.4. Линейная разделимость и персептронная представляемость

При *прямоугольной передаточной функции* (1) каждый нейрон представляет собой *пороговый* элемент, который может находиться только в одном из двух состояний:

- *возбужденном (активном)*, если взвешенная сумма входных сигналов больше некоторого порогового значения;
- *заторможенном (пассивном)*, если взвешенная сумма входных сигналов меньше некоторого порогового значения.

$$Y_j = \begin{cases} 1, & \sum_{i=1}^M W_{ij} x_i > \theta \\ 0, & \sum_{i=1}^M W_{ij} x_i \leq \theta \end{cases} \quad (1)$$

Следовательно, при заданных значениях весов и порогов, каждый нейрон имеет *единственное* определенное значение выходной активности для каждого возможного вектора входов. При этих условиях множество входных векторов, при которых нейрон активен ($Y=1$), отделено от множества векторов, на которых нейрон пассивен ($Y=0$) *гиперплоскостью* (2).

$$\sum_{i=1}^M W_{ij} x_i - \theta = 0 \quad (2)$$

Следовательно, нейрон способен *отделить* только такие два множества векторов входов, для которых существует гиперплоскость, отделяющая одно множество от другого. Такие множества называют *линейно разделимыми*.

Необходимо отметить, что линейно-разделимые множества являются составляют лишь очень незначительную часть всех множеств. Поэтому данное ограничение персептрона является принципиальным. Оно было преодолено лишь в 80-х годах путем введения *нескольких слоев нейронов* в сетях Хопфилда и неокогнитроне Фукушимы.

В завершении остановимся на некоторых *проблемах*, которые остались нерешенными после работ Ф.Розенблатта:

1. Возможно ли обнаружить линейную разделимость классов до обучения сети?
2. Как определить скорость обучения, т.е. количество итераций, необходимых для достижения заданного качества обучения?
3. Как влияют на результаты обучения *последовательность* предъявления образов и их *количество*?
4. Имеет ли алгоритм обратного распространения ошибки преимущества перед простым перебором весов?
5. Каким будет качество обучения, если обучающая выборка содержит *не все возможные на практике* пары векторов и какими будут ответы персептрона на *новые* вектора, отсутствующие в обучающей выборке?

Особенно важным представляется последний вопрос, т.к. индивидуальный опыт принципиально всегда не является полным.

1.3.5.5. Многослойные нейронные сети

1.3.5.5.1. Многослойный персептрон

Каким же образом в многослойных (иерархических) нейронных сетях преодолевается принципиальное ограничение однослойных нейронных сетей, связанное с требованием линейной разделимости классов?

Часто то, что не удастся сделать сразу, вполне возможно сделать по частям. Для этого изменяются задачи, решаемые слоями нейронной сети. Оказывается в 1-м слое не следует пытаться на основе *первичных признаков*, фиксируемых рецепторами, сразу идентифицировать классы, а нужно лишь сформировать линейно-разделимую систему *вторичных признаков*, которую уже во 2-м слое связать с классами (рисунок 76).

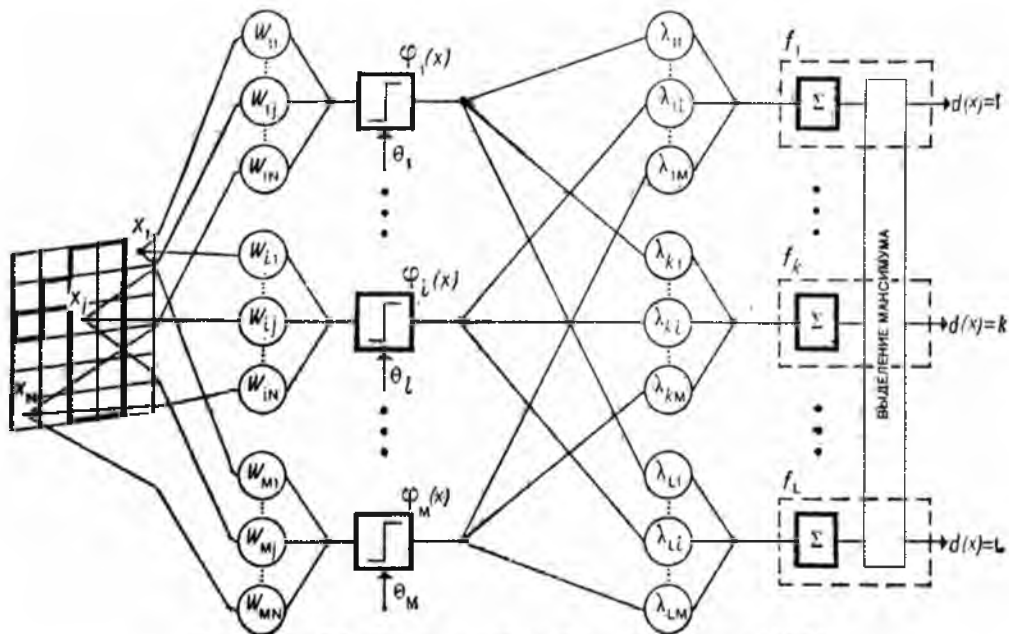


Рисунок 76. Двух-слойный перцептрон [233]

В многослойной сети выходные сигналы нейронов предыдущего слоя играют роль входных сигналов для нейронов последующего слоя, т.е. *нейроны предыдущего слоя выступают в качестве рецепторов* для нейронов последующего слоя.

Связи между смежными слоями нейронов будем называть непосредственными, а связи между слоями, разделенными N промежуточных слоев, будем называть связями N -го уровня опосредованности. Непосредственные связи – это связи 0-го уровня опосредованности. Промежуточные слои нейронов в многослойных сетях называют скрытыми.

Перцептрон переводит входной образ, определяющий степени возбуждения рецепторов, в выходной образ, определяемый нейронами самого верхнего уровня, которых обычно, не очень много. Состояния возбуждения нейронов на верхнем уровне иерархии сети характеризуют принадлежность входного образа к тем или иным классам.

Таким образом, **многослойный перцептрон** – это обучаемая *распознающая система*, реализующая корректируемое в процессе обучения линейное решающее правило в пространстве *вторичных признаков*, которые обычно являются фиксированными случайно выбранными линейными пороговыми функциями от *первичных признаков*.

При обучении на вход перцептрона поочередно подаются сигналы из *обучающей выборки*, а также указания о классе, к которому следует отнести данный сигнал. Обучение перцептрона заключается в коррекции весов при каждой

ошибке распознавания, т. е. при каждом случае несовпадения решения, выдаваемого персептроном, и истинного класса. Если персептрон ошибочно отнес сигнал, к некоторому классу, то веса функции, истинного класса увеличиваются, а ошибочного уменьшаются. В случае правильного решения все веса остаются неизменными.

Этот чрезвычайно простой алгоритм обучения обладает замечательным свойством: если существуют значения весов, при которых выборка может быть разделена безошибочно, то при определенных, легко выполнимых условиях эти значения будут найдены за конечное количество итераций.

При идентификации, распознавании, прогнозировании на вход многослойного персептрона поступает сигнал, представляющий собой набор первичных признаков, которые и фиксируются рецепторами. Сначала вычисляются вторичные признаки. Каждому такому вторичному признаку соответствует линейная от первичных признаков. Вторичный признак принимает значение 1, если соответствующая линейная функция превышает порог. В противном случае она принимает значение 0. Затем для каждого из классов вычисляется функция, линейная относительно вторичных признаков. Персептрон вырабатывает решение о принадлежности входного сигнала к тому классу, которому соответствует функция от вторичных параметров, имеющая наибольшее значение.

Показано, что для представления произвольного нелинейного функционального отображения, задаваемого обучающей выборкой, достаточно всего *двух слоев* нейронов. Однако на практике, в случае сложных функций, использование более чем одного скрытого слоя может давать экономию полного числа нейронов.

1.3.5.5.2. Модель Хопфилда

В модели Хопфилда (J.J.Hopfield, 1982) впервые удалось установить связь между нелинейными динамическими системами и нейронными сетями.

Модель Хопфилда является обобщением модели многослойного персептрона путем добавления в нее следующих двух новых свойств:

1. В нейронной сети все нейроны непосредственно связаны друг с другом: силу связи i -го нейрона с j -м обозначим как W_{ij} .

2. Связи между нейронами симметричны: $W_{ij}=W_{ji}$, сам с собой нейрон не связан $W_{ii}=0$.

Каждый нейрон может принимать лишь два состояния, которые определяются по классической формуле (1). Изменение состояний возбуждения всех нейронов может происходить либо последовательно, либо одновременно (параллельно), но свойства сети Хопфилда не зависят от типа динамики.

Сеть Хопфилда способна распознавать объекты при неполных и зашумленных исходных данных, однако не может этого сделать, если изображение смещено или повернуто относительно его исходного состояния, представленного в обучающей выборке.

1.3.5.5.3. Когнитрон и неокогнитрон Фукушимы

В целом *когнитрон* (К. Fukushima, 1975) представляет собой иерархию слоев, последовательно связанных друг с другом, как в персептроне. Однако, при этом есть два существенных отличия:

1. Нейроны образуют не одномерную цепочку, а покрывают *плоскость*, аналогично слоистому строению зрительной коры человека.
2. Когнитрон состоит из иерархически связанных слоев нейронов двух типов – тормозящих и возбуждающих.

В когнитроне каждый слой реализует свой уровень обобщения информации:

- *входные слои* чувствительны к отдельным элементарным структурам, например, линиям определенной ориентации или цвета;
- *последующие слои* реагируют уже на более сложные обобщенные образы;
- *в слое наивысшего уровня иерархии* активные нейроны определяют результат работы сети – узнавание определенного образа, при этом результатам распознавания соответствуют те нейроны, активность которых оказалась максимальной.

Однако добиться независимости (инвариантности) результатов распознавания от размеров и ориентации изображений удалось лишь в *неокогнитроне*, который был разработан Фукушимой в 1980 году и представляет собой как бы суперпозицию когнитронов, обученных распознаванию объектов различных типов, размеров и ориентации.

1.3.5.6. Проблемы и перспективы нейронных сетей

На наш взгляд к основным проблемам нейронных сетей можно отнести:

1. Сложность содержательной интерпретации смысла интенсивности входных сигналов и весовых коэффициентов (*"проблема интерпретируемости весовых коэффициентов"*).
2. Сложность содержательной интерпретации и обоснования аддитивности аргумента и вида активационной (передаточной) функции нейрона (*"проблема интерпретируемости передаточной функции"*).
3. "Комбинаторный взрыв", возникающий при определении структуры связей нейронов, подборе весовых коэффициентов и передаточных функций (*"проблема размерности"*).
4. *"Проблема линейной делимости"*, возникающая потому, что возбуждение нейронов принимают лишь булевы значения 0 или 1.

Проблемы интерпретируемости приводят к снижению ценности полученных результатов работы сети, а проблема размерности – к очень жестким ограничениям на количество выходных нейронов в сети, на количество рецепторов и на сложность структуры взаимосвязей нейронов с сети. Достаточно сказать, что количество выходных нейронов в реальных нейронных сетях, реализуемых на базе известных программных пакетов, обычно не превышает несколько сотен, а чаще всего составляет единицы и десятки.

Проблема линейной разделимости приводит к необходимости применения многослойных нейронных сетей для реализации тех приложений, которые вполне могли бы поддерживаться сетями с значительно меньшим количеством слоев (вплоть до однослойных), если бы значения возбуждения нейронов были не дискретными булевыми значениями, а континуальными значениями, нормированными в определенном диапазоне.

Перечисленные проблемы предлагается решить путем использования модели нелокального нейрона, обеспечивающего построение нейронных сетей прямого счета.

1.3.5.7. Модель нелокального нейрона и нелокальные интерпретируемые нейронные сети прямого счета

1.3.5.7.1. Метафора нейросетевого представления семантической информационной модели

В данной работе предлагается представление, согласно которому каждый нейрон отражает определенное будущее состояние активного объекта управления, а нейронная сеть в целом – систему будущих состояний, как желательных (целевых), так и нежелательных. Весовые коэффициенты на дендритах нейронов имеют смысл силы и направления влияния факторов на переход активного объекта управления в то или иное будущее состояние. Таким образом, предложенная в данной работе семантическая информационная модель в принципе допускает представление в терминах и понятиях нейронных сетей. Однако при более детальном рассмотрении выясняется, что семантическая информационная модель является более общей, чем нейросетевая и для полного их соответствия необходимо внести в нейросетевую модель ряд дополнений.

1.3.5.7.2. Соответствие основных терминов и понятий

Предлагается следующая система соответствий, позволяющая рассматривать термины и понятия из теории нейронных сетей и предложенной семантической информационной модели практически как синонимы. Нейрон – вектор обобщенного образа класса в матрице информативностей. Входные сигналы – факторы (признаки). Весовой коэффициент – системная мера целесообразности информации. Обучение сети – *адаптация* модели, т.е. перерасчет значений весовых коэффициентов дендритов для каждого нейрона (матрицы информативностей) и изменение вида активационной функции. Самоорганизация сети – *синтез* модели, т.е. изменение количества нейронов и дендритов, изменение количества нейронных слоев и структуры связей между факторами и классами, а затем адаптация (перерасчет матрицы информативностей). Таким образом, адаптация – это обучение сети на уровне изменения информационных весовых коэффициентов и активационной функции, а синтез – на уровне изменения размерности и структуры связей нейронов сети. 1-й (входной) слой нейронной сети – формирование обобщенных образов классов. Сети Хопфилда и Хэмминга – обучение с учителем, сопоставление описательной и классификационной информации,

идентификация и прогнозирование. 2-й слой, сети Хебба и Кохонена – самообучение, анализ структуры данных без априорной классификационной информации, формирование кластеров классов и факторов. 3-й слой – формирование конструкторов (в традиционных нейронных сетях не реализовано). Необходимо отметить, что любой слой нейронной сети является в предлагаемой модели не только обрабатывающим, но и выходным, т.е. с одной стороны дает результаты обработки информации, имеющие самостоятельное значение, а с другой – предоставляет информацию для последующих слоев нейронной сети, т.е. более высоких уровней иерархии информационной системы (в полном соответствии с формализуемой когнитивной концепцией).

1.3.5.7.3. Гипотеза о нелокальности нейрона и информационная нейросетевая парадигма

Модель нелокального нейрона: так как сигналы на дендритах различных нейронов вообще говоря коррелируют (или антикоррелируют) друг с другом, то, значения весовых коэффициентов, а значит и выходное значение на аксоне каждого конкретного нейрона вообще говоря не могут быть определены с использованием значений весовых коэффициентов на дендритах только данного конкретного нейрона, а должны учитывать интенсивности сигналов на всей системе дендритов нейронной сети в целом (рисунок 77).

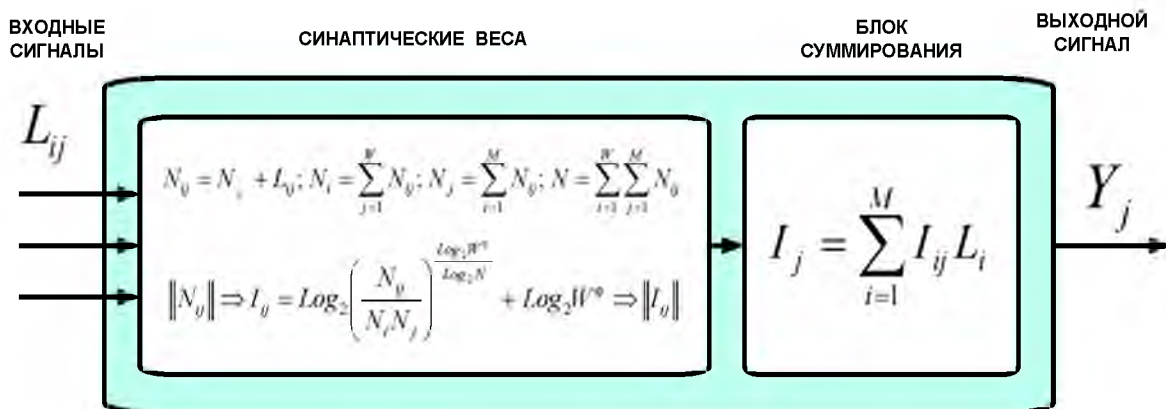


Рисунок 77. Модель нелокального нейрона
в обозначениях системной теории информации

За счет учета корреляций входных сигналов (если они фактически присутствуют в структуре данных), т.е. наличия *общего самосогласованного* информационного поля исходных данных всей нейронной сети (информационное пространство), нелокальные нейроны ведут себя так, как будто связаны с другими нейронами, хотя могут быть и не связаны с ними синаптически по входу и выходу ни прямо, ни опосредованно. Самосогласованность семантического информационного пространства [81, 109, 120] означает, что *учет любого одного нового факта в информационной модели вообще говоря приводит к изменению всех весовых коэффициентов всех нейронов, а не только тех, на рецепторе которых обнаружен этот факт и тех, которые непосредственно или опосредованно синаптически с ним связаны.*

В традиционной (т.е. локальной) модели нейрона весовые коэффициенты на его дендритах однозначно определяются заданным выходом на его аксоне и никак не зависят от параметров других нейронов, с которыми с ней нет прямой или опосредованной синаптической связи. Это связано с тем, что в общепринятой энергетической парадигме Хопфилда весовые коэффициенты дендритов имеют смысл *интенсивностей* входных воздействий. В методе "обратного распространения ошибки" процесс переобучения, т.е. интерактивного перерасчета весовых коэффициентов, начинается с нейрона, состояние которого оказалось ошибочным и захватывает только нейроны, ведущие от рецепторов к данному нейрону. Корреляции между локальными нейронами обусловлены сочетанием трех основных причин:

- наличием в исходных данных определенной структуры: корреляцией входных сигналов;
- синаптической связью локальных нейронов;
- избыточностью (дублированием) нейронной сети.

1.3.5.7.4. Решение проблемы интерпретируемости весовых коэффициентов (семантическая мера целесообразности информации и закон Фехнера)

В данной работе предлагается использовать такие весовые коэффициенты дендритов, чтобы активационная функция была линейной, т.е. по сути была равна своему аргументу: сумме. Этому условию удовлетворяют весовые коэффициенты, рассчитываемые с применением системного обобщения формулы Харкевича [81].

Очень важно, что данная мера, удовлетворяет известному эмпирическому закону Г.Фехнера (1860), согласно которому существует *логарифмическая* зависимость между интенсивностью фактора и величиной отклика на него биологической системы (в частности, величина ощущения прямо пропорциональна логарифму интенсивности раздражителя).

Предлагается *информационный подход* к нейронным сетям, по аналогии с энергетическим подходом Хопфилда (1980).

Суть этого подхода состоит в том, что интенсивности входных сигналов рассматриваются не сами по себе и не с точки зрения только их *интенсивности*, а как сообщения, несущие определенное количество *информации* или *дезинформации* о переходе нейрона и моделируемого им активного объекта управления в некоторое будущее состояние.

Под интенсивностью входного сигнала на определенном дендрите мы будем понимать абсолютную частоту (количество) встреч фактора (признака), соответствующего данному дендриту, при предъявлении нейронной сети объекта, соответствующего определенному нейрону. Таким образом матрица абсолютных частот рассматривается как способ накопления и первичного обобщения эмпирической информации об интенсивностях входных сигналов на дендритах в разрезе по нейронам.

Весовые коэффициенты, отражающие влияние каждого входного сигнала на отклик каждого нейрона, т.е. величину его возбуждения или торможения, представляют собой элементы матрицы информативностей, получающиеся из

матрицы абсолютных частот методом прямого счета с использованием выражения для семантической меры целесообразности информации [81].

При этом предложенная мера семантической целесообразности информации, как переключается с нейронными сетями Кохонена, в которых также принято стандартизировать (нормализовать) входные сигналы, что позволяет в определенной мере уйти от многообразия передаточных функций.

Наличие ясной и обоснованной интерпретации весовых коэффициентов, как количества информации, позволяет предложить в качестве математической модели для их расчета системную теорию информации (СТИ).

1.3.5.7.5. Семантическая информационная модель, как нелокальная нейронная сеть

Учитывая большое количество содержательных параллелей между семантической информационной моделью и нейронными сетями предлагается рассматривать данную модель как нейросетевую модель, основанную на системной теории информации. В данной модели предлагается вариант решения важных нейросетевых проблем интерпретируемости и ограничения размерности за счет введения меры целесообразности информации (системное обобщение формулы Харкевича), обеспечивающей *прямой расчет* интерпретируемых весовых коэффициентов на основе непосредственно эмпирических данных. Итак, в данной работе предлагается новый класс нейронных сетей, основанных на семантической информационной модели и информационном подходе. Для этих сетей предлагается полное наименование: ***"Нелокальные интерпретируемые нейронные сети прямого счета"*** и сокращенное наименование: ***"Нелокальные нейронные сети"***.

Нелокальная нейронная сеть является системой нелокальных нейронов, обладающей качественно новыми (системными, эмерджентными) свойствами, не сводящимися к сумме свойств нейронов. В такой сети поведение нейронов определяется как их собственными свойствами и поступающими на них входными сигналами, так и свойствами нейронной сети в целом, т.е. поведение нейронов в нелокальной нейронной сети согласовано друг с другом не только за счет их прямого и опосредованного синаптического взаимодействия (как в традиционных нейронных сетях), но за счет общего информационного поля весовых коэффициентов всех нейронов данной сети.

1.3.5.7.6. Гипотеза о физической природе нелокального взаимодействия нейронов в нелокальной нейронной сети

В данной работе предлагается математическая модель, численный метод и программный инструментальный нелокальных нейронных сетей (универсальная когнитивная аналитическая система "Эйдос"), успешно апробированные в ряде предметных областей. Данная система обеспечивает неограниченное количество слоев ННС при максимальном количестве весовых коэффициентов в слое до 16 миллионов в версии 9.0 и до 4000 выходных нейронов, а в текущей версии 12.5 эти количества ограничены практически только емкостью диска. Но если рассматривать нелокальную нейронную сеть как модель реальных "биологических"

нейронных сетей, то ясно, что формальной модели недостаточно и необходимо дополнить ее *физической моделью о природе каналов нелокального взаимодействия нейронов* в данной сети.

По мнению автора данный механизм основан на парадоксе Эйнштейна-Подольского-Розена (ЭПР) [10, 109, 264]. По мнению автора, физическая реализация нелокальных нейронов может быть осуществлена за счет соединения как минимум одного дендрита каждого нейрона с датчиком микротелекинетического воздействия, на который человек может оказывать влияние дистанционно. Некоторые из подобных датчиков описаны в работе [27] и на сайте автора по адресу: http://lc.kubagro.ru/History/vega-72/page_01.htm. По мнению автора, *мозг может рассматриваться как оптимальная среда для редукции мыслей, в этом смысле квантовые компьютеры, основанные не на математических и программных моделях, а на физических нелокальных нейронах, могут оказаться во многих отношениях функционально эквивалентными физическому организму.*

1.3.5.7.7. Решение проблемы интерпретируемости передаточной функции

Вопрос об интерпретируемости передаточной функции нейрона включает два основных аспекта:

- 1) об интерпретируемости *аргумента* передаточной функции;
- 2) об интерпретируемости *вида* передаточной функции.

1. Возникает естественный вопрос о том, чем обосновано включение в состав модели нейрона Дж. Маккалоки и У. Питтом именно *аддитивного* элемента, суммирующего входные сигналы, а не скажем мультипликативного или в виде какой-либо другой функции общего вида. По мнению автора *такой выбор обоснован и имеет явную и убедительную интерпретацию именно в том случае, когда весовые коэффициенты имеют смысл количества информации, т.к. в этом случае данная мера представляет собой неметрический критерий сходства, основанный на лемме Неймана-Пирсона [194]. Сумма весовых коэффициентов, соответствующих набору действующих факторов (входных сигналов) дает величину выходного сигнала на аксоне каждого нейрона.*

2. Вид передаточной функции содержательно в теории нейронных сетей явно не обосновывается. Предлагается гипотеза, что на практике вид передаточной функции подбирается таким образом, чтобы соответствовать смыслу подобранных в данном конкретном случае весовых коэффициентов. Так как при применении в различных предметных областях смысл весовых коэффициентов в явном виде не контролируется и может отличаться, то выбор вида передаточной функции позволяет частично компенсировать эти различия.

Предлагаемый интерпретируемый вид весовых коэффициентов обеспечивает единую и стандартную интерпретацию аргумента и значения передаточной функции независимо от предметной области. Поэтому в нелокальной нейронной модели передаточная функция нейрона всегда линейна (аргумент равен функции). Следовательно в модели нелокального нейрона блок суммирования по сути дела объединен с блоком нелинейного преобразования (точнее, второй отсутствует, а его роль выполняет блок суммирования), в отличие от стандартных передаточных функций локальных нейронов: логистической, гиперболического

тангенса, пороговой линейной, экспоненциально распределенной, полиномиальной и импульсно-кодовой.

Нелокальные нейроны как бы "резонируют" на ансамбли входных сигналов, причем *этот резонанс может быть обоснованно назван семантическим (смысловым)*, т.к. весовые коэффициенты рассчитаны на основе предложенной семантической меры целесообразности информации. Таким образом, разложение вектора идентифицируемого объекта в ряд по векторам обобщенных образов классов осуществляется на основе семантического резонанса нейронов выходного слоя на ансамбль входных сигналов (признаков, факторов).

1.3.5.7.8. Решение проблемы размерности

Вместо итерационного подбора весовых коэффициентов путем полного перебора вариантов их значений при малых вариациях (методы обратного распространения ошибки и градиентного спуска к локальному экстремуму) предлагается *прямой расчет* этих коэффициентов на основе процедуры и выражений, обоснованных в предложенных системной теории информации и семантической информационной модели. Выигрыш во времени и используемых вычислительных ресурсах, получаемый за счет этого, быстро возрастает при увеличении размерности нейронной сети.

1.3.5.7.9. Решение проблемы линейной разделимости

Вводятся промежуточные линейно-разделимые классы распознавания, которые рассматриваются как вторичные признаки при идентификации объектов с ранее не разделимыми классами. Это решение соответствует введению дополнительных слоев нейронной сети.

В системе "Эйдос" функция представления нейронов предыдущего слоя в качестве рецепторов последующего слоя автоматизирована, что в случае необходимости позволяет в полуавтоматическом режиме преобразовать однослойную сеть с линейно-неразделимыми классами в иерархическую нейронную сеть в которой эти классы линейно-разделены относительно вторичных признаков в слоях более высоких уровней иерархии.

1.3.5.7.10. Моделирование причинно-следственных цепочек в нейронных сетях и семантической информационной модели

Факторы описывают причины, а классы – следствия. Но и следствия в свою очередь являются причинами более отдаленных последствий. Предлагаемая семантическая информационная модель позволяет рассматривать события, обнаружение которых осуществляется в режиме идентификации, как причины последующих событий, т.е. как факторы, их вызывающие. При этом факт наступления этих событий моделируется путем включения в модель факторов, соответствующих классам (событиям). В нейронных сетях этот процесс моделируется путем включения в сеть дополнительных нейронных слоев и создания обратных связей между слоями, обеспечивающих передачу в предыдущие слои результатов работы последующих слоев.

1.3.5.7.11. Моделирование иерархических структур обработки информации

Рассмотрим иерархическую структуру информации на примере использования психологического теста для оценки психологических качеств сотрудников и влияния этих качеств на эффективность работы фирмы. В нейронной сети иерархическим уровням обработки информации соответствуют слои, поэтому далее будем использовать термины "слой нейронной сети" и "иерархический уровень обработки информации" как синонимы. Рецепторы дают информацию по ответам сотрудника на опросник, нейроны 1-го слоя дают оценку психологических качеств и сигнал с их аксонов является входным для нейронов 2-го слоя, дающих оценку качества работы фирмы. В семантической информационной модели существует три варианта моделирования подобных иерархических структур обработки информации:

1. Заменить все слои одним слоем и выявлять зависимости непосредственно между исходными данными с первичных рецепторов и интересующими итоговыми оценками, например, ответами сотрудников на вопросы и результатами работы фирмы. Этот подход эффективен с прагматической точки зрения, но дает мало информации для теоретических обобщений.

2. Каждый слой моделируется отдельной семантической информационной моделью, включающей свои классификационные и описательные шкалы и градации, обучающую выборку, матрицы абсолютных частот и информативностей. Вся система иерархической обработки информации моделируется системой этих моделей, взаимосвязанных друг с другом по входу-выходу: результаты классификации объектов обучающей выборки 1-й моделью рассматриваются как свойства этих объектов во 2-й модели, в которой они используются для классификации 2-го уровня. Например, психологические качества сотрудников, установленные в результате психологического тестирования, рассматриваются как свойства сотрудников, влияющие на эффективность работы фирмы. Данный подход эффективен и с прагматической, и с теоретической точек зрения, но является громоздким в программной реализации.

3. Моделирование каждого слоя соответствующими подматрицами матриц абсолютных частот и информативностей (таблица 34).

Примечание: в таблице 34 представлена именно *логическая* структура данных, т.е. в реальных базах данных нет записей, содержащих информацию о влиянии рецепторов n -го слоя на нейроны слоев, номера которых не равны n .

Этот вариант обладает преимуществами первых двух и преодолевает их недостатки. В нем применяется следующий итерационный алгоритм послойного расчета, где $n = \{1, 2, \dots, N\}$, N – количество слоев нейронной сети:

Шаг n : расчет весовых коэффициентов n -го слоя, идентификация объектов обучающей выборки в нейронах n -го слоя, если слой $(n+1)$ существует, то занесение в обучающую выборку в качестве свойств объектов $(n+1)$ -го слоя результатов их идентификации в нейронах n -го слоя.

Таблица 34 – ЛОГИЧЕСКАЯ СТРУКТУРА ДАННЫХ, СООТВЕТСТВУЮЩАЯ ТРЕХСЛОЙНОЙ НЕЛОКАЛЬНОЙ НЕЙРОННОЙ СЕТИ

Рецепторы – факторы, влияющие на поведение объекта управления	Нейроны - будущие состояния объекта управления			Дифференцирующая способность входного сигнала
	Нейроны 1-го слоя: психологические качества сотрудников	Нейроны 2-го слоя: успешность деятельности сотрудника	Нейроны 3-го слоя: успешность деятельности фирмы	
Рецепторы 1-го слоя: ответы сотрудников на вопросы анкеты	Весовые коэффициенты 1-го слоя	---	---	
Рецепторы 2-го слоя: психологические качества сотрудников	---	Весовые коэффициенты 2-го слоя	---	
Рецепторы 3-го слоя: успешность деятельности сотрудника	---	---	Весовые коэффициенты 3-го слоя	
Степень обученности нейрона				Степень обученности нейронной сети

1.3.5.7.12. Нейронные сети и СК-анализ

Известные в литературе нейронные сети, в отличие от предлагаемой семантической информационной модели и нелокальных нейронных сетей, не обеспечивают реализацию всех базовых когнитивных операций, входящих в когнитивный конфигуратор. В частности, традиционные нейронные сети решают лишь задачу идентификации (прогнозирования) и не обеспечивают решение обратной задачи (дедукции), необходимой для принятия решения о выборе многофакторного управляющего воздействия. Кроме того не решается вопрос об уменьшении размерности нейронной сети без ущерба для ее адекватности (абстрагирование).

Результаты численного моделирования и исследования свойств нейронных сетей этого класса при управлении в АПК и других предметных областях позволяют предположить, в качестве модели реальных когнитивных процессов они обладает более высокой адекватностью, чем нейронные сети других типов.

1.3.5.7.13. Графическое отображение нейронов, Парето-подмножеств нелокальной нейронной сети, семантических сетей, когнитивных карт и диаграмм в системе "Эйдос"

Для каждого технологического фактора в соответствии с предложенной моделью определяется величина и направление его влияния на осуществление всех желаемых и не желаемых хозяйственных ситуаций. Для каждой ситуации эта информация отображается в различных текстовых и графических формах, в частности в форме нелокального нейрона (рисунок 78).

На данной и последующих графических диаграммах цвет линии означает знак связи (красный – положительная, синий – отрицательная), а толщина – ее модуль.

Паретто-подмножеством нелокальной нейронной сети будем называть ее подмножество, включающее наиболее значимые связи. Пример графического отображения такого подмножества приведен в таблицах 35 – 37.

Факторы (сигналы с рецепторов) в модели нелокального нейрона взаимосвязаны друг с другом. Эти связи графически отображаются в форме семантической сети (рисунок 78).

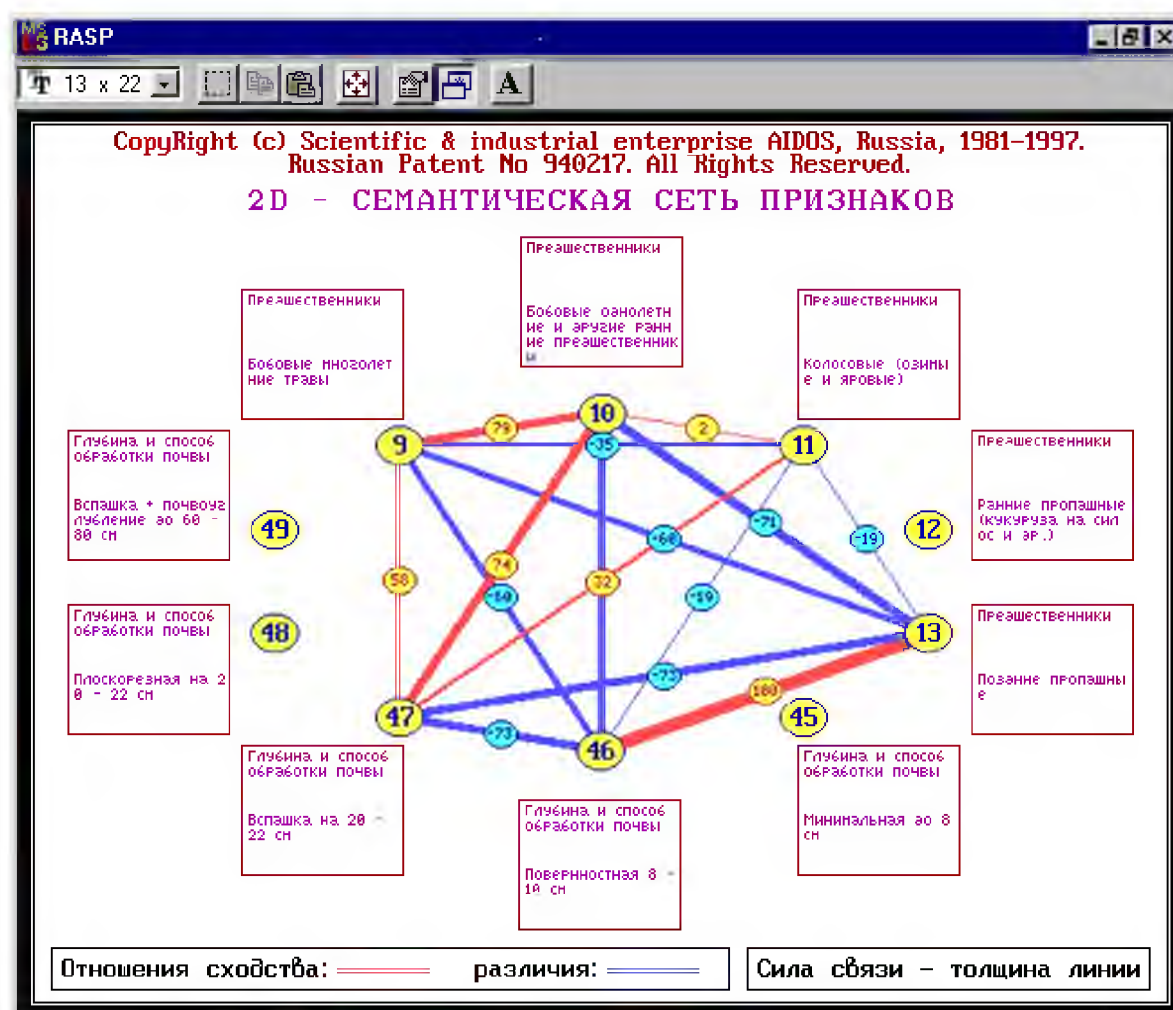


Рисунок 78. Отображение результатов кластерно-конструктивного анализа факторов в форме семантической сети (когнитивной карты) в системе "Эйдос"

Структуру любой линии связи семантической сети можно детально изучить в когнитивной диаграмме (рисунок 79).

Дополнение модели нейрона связями факторов позволяет построить классическую когнитивную карту ситуации (будущего состояния АОУ). Необходимо отметить, что все указанные графические формы генерируются системой "Эйдос" автоматически в соответствии с созданной моделью.

Нелокальный нейрон представляет собой будущее состояние объекта управления с изображением наиболее сильно влияющих на него факторов с указанием силы и направления (способствует-препятствует) их влияния (рисунок 80).

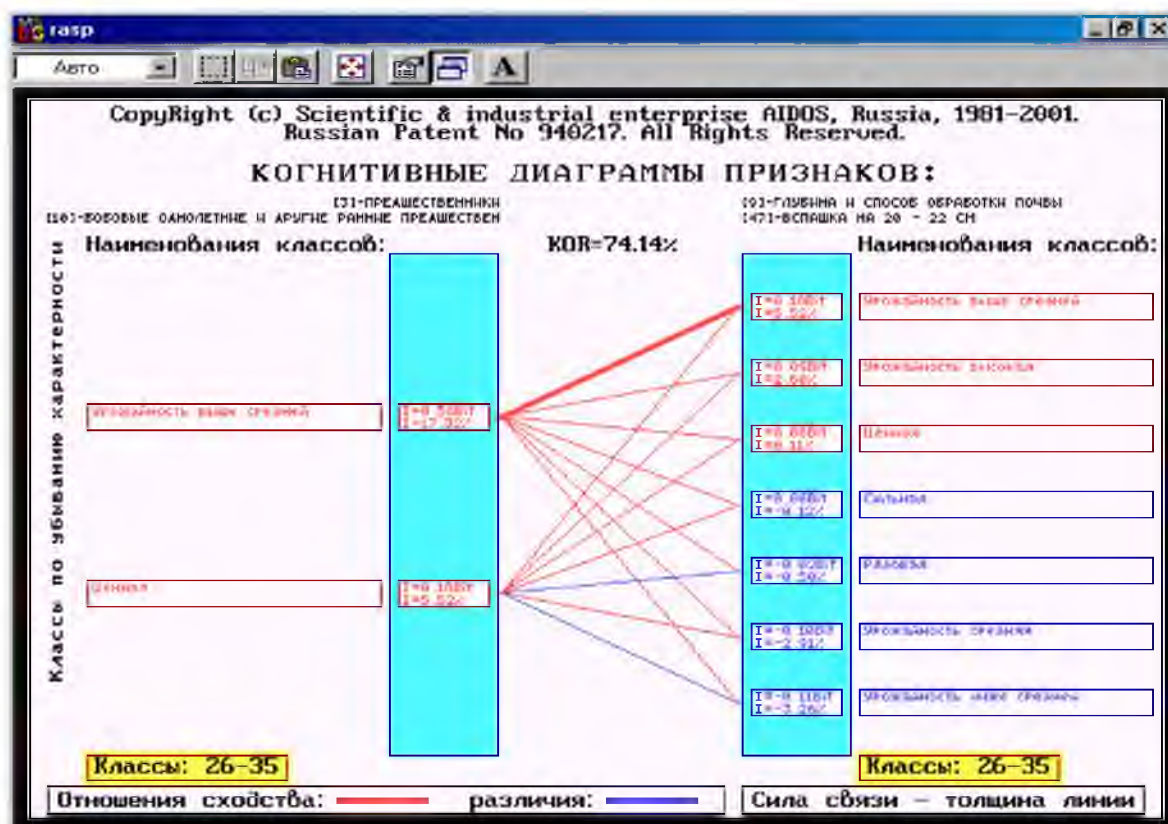


Рисунок 79. Когнитивная диаграмма кластера факторов: "Глубина и способ обработки почвы: вспашка на 20-22 см – Предшественники: Бобовые однолетние и другие ранние предшественники", генерируемая системой "Эйдос"

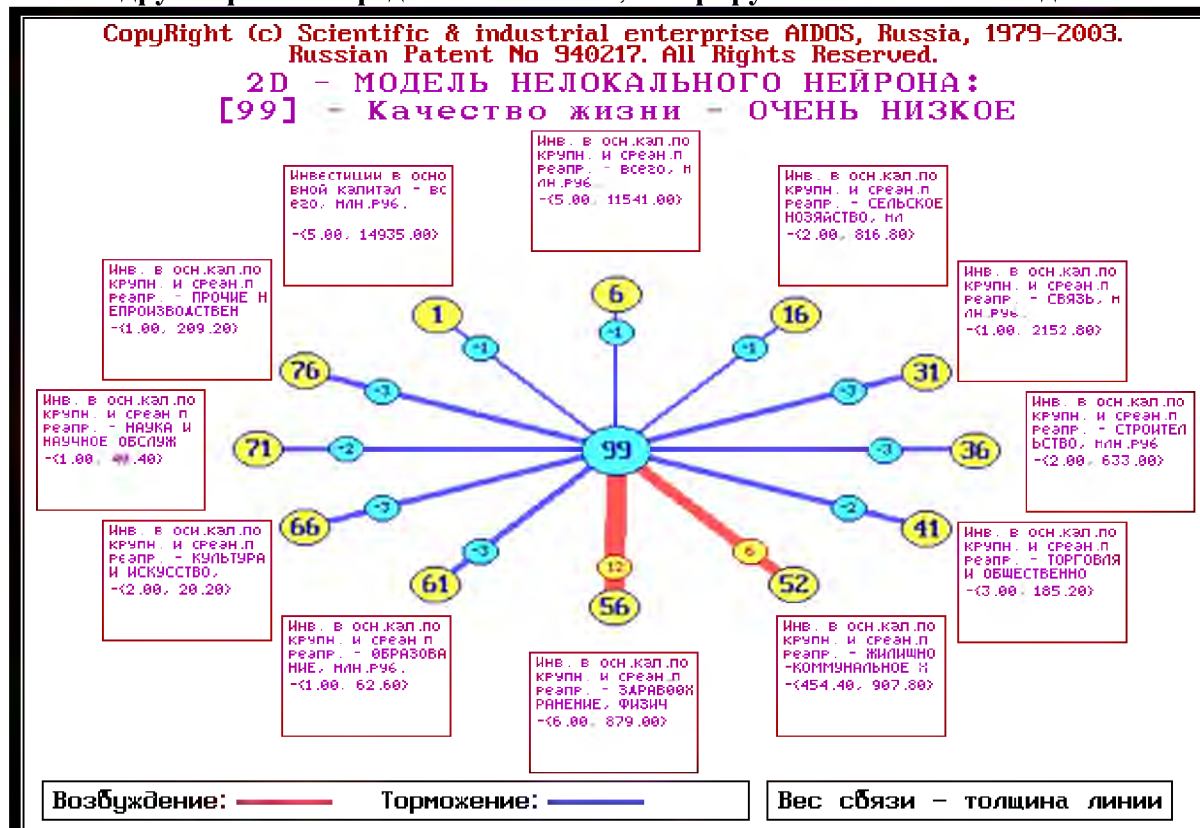


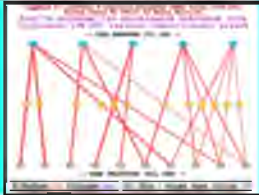
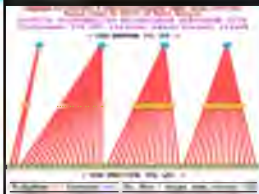
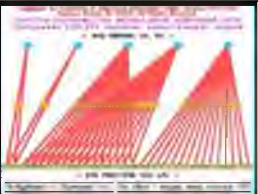
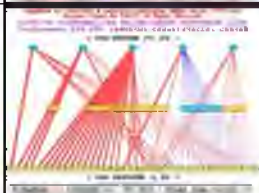
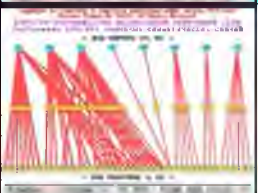

Рисунок 80. Примеры нелокальных нейронов, отражающих влияние инвестиций на уровень качества жизни в регионе (система "Эйдос")

Нейронная сеть представляет собой совокупность взаимосвязанных нейронов. В классических нейронных сетях связь между нейронами осуществляется по входным и выходным сигналам, а в нелокальных нейронных сетях [109] – на основе общего информационного поля.

В таблице 35 приводится классификация причинно-следственных связей по рангам. Эта классификация использована для отображения параметров заданий на генерацию и соответствующих фрагментов нейронных сетей (таблица 36). Фрагменты нейронной сети со связями 0-го уровня опосредованности, т.е. соответствующие *смежным* слоям многослойной сети, показаны на голубом фоне. В пустых клетках таблицы 35 могут быть отображены фрагменты нейронной сети, аналогичные показанным. Однако новой информации, по сравнению с уже показанными, они не содержат, т.к. практически они образуются из них путем перемены местами нейронов и рецепторов (инвертирования – отражения относительно горизонтальной оси). Сгенерированные по этим заданиям фрагменты нейронной сети приведены в форме, позволяющей составить из них многослойную нейронную сеть (таблицы 36 и 37). Эти таблицы взяты из работы [248].

Система "Эйдос" обеспечивает построение **любого** подмножества многослойной нейронной сети с заданными или выбираемыми по заданным критериям рецепторами и нейронами, связанными друг с другом связями любого уровня опосредованности.

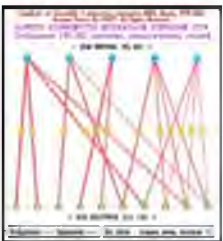
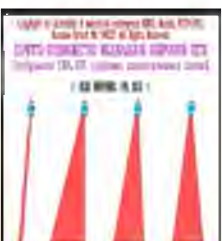
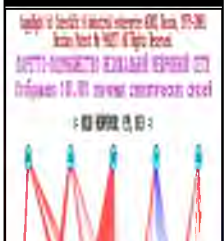
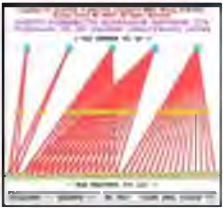
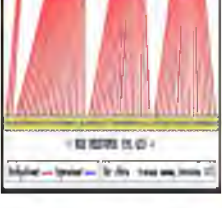




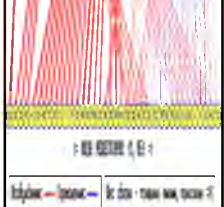
Таблица 35 – ВИДЫ КАУЗАЛЬНЫХ СВЯЗЕЙ МЕЖДУ ОБЪЕКТАМИ РАЗЛИЧНЫХ УРОВНЕЙ ИЕРАРХИЧЕСКОЙ МОДЕЛИ И СООТВЕТСТВУЮЩИЕ ФРАГМЕНТЫ НЕЙРОННОЙ СЕТИ

Факторы (наименования, коды)		Классы (наименования, коды)		
		Уровень качества жизни	Годы	Частные критерии уровня качества жизни
Наименования	Коды	99-103	86-98	1-85
Годы	626-638			
Частные критерии уровня качества жизни	541-625			
Первичные факторы (инвестиции)	1-85			

**Таблица 36 – ФРАГМЕНТЫ МНОГОУРОВНЕВОЙ СЕМАНТИЧЕСКОЙ
ИНФОРМАЦИОННОЙ МОДЕЛИ И НЕЙРОННОЙ СЕТИ
С НЕПОСРЕДСТВЕННЫМИ СВЯЗЯМИ**

№	Уровень Наименование	Нейронная сеть
4	Уровни качества Жизни (значения Интегрального критерия уровня качества жизни)	<p>Copyright (c) Scientific & Industrial enterprise AIDOS, Russia, 1979-2003. Russian Patent No 940217. All Rights Reserved. ПАРЕТТО-ПОДМНОЖЕСТВО НЕЛОКАЛЬНОЙ НЕЙРОННОЙ СЕТИ Отображено 100.00% значимых синаптических связей <= КОДЫ НЕЙРОНОВ: (99, 103) => <= КОДЫ РЕЦЕПТОРОВ: (626, 630) => Возбуждение: — Торможение: — Вес связи — толщина линии, показаны =>1</p>
3	Годы	<p>Copyright (c) Scientific & Industrial enterprise AIDOS, Russia, 1979-2003. Russian Patent No 940217. All Rights Reserved. ПАРЕТТО-ПОДМНОЖЕСТВО НЕЛОКАЛЬНОЙ НЕЙРОННОЙ СЕТИ Отображено 100.00% значимых синаптических связей <= КОДЫ НЕЙРОНОВ: (86, 98) => <= КОДЫ РЕЦЕПТОРОВ: (541, 625) => Возбуждение: — Торможение: — Вес связи — толщина линии, показаны =>17</p>
2	Вторичные факторы (частные критерии уровня качества жизни)	<p>Copyright (c) Scientific & Industrial enterprise AIDOS, Russia, 1979-2003. Russian Patent No 940217. All Rights Reserved. ПАРЕТТО-ПОДМНОЖЕСТВО НЕЛОКАЛЬНОЙ НЕЙРОННОЙ СЕТИ Отображено 100.00% значимых синаптических связей <= КОДЫ НЕЙРОНОВ: (1, 85) => <= КОДЫ РЕЦЕПТОРОВ: (1, 85) => Возбуждение: — Торможение: — Вес связи — толщина линии, показаны =>17</p>
1	Первичные факторы	

**Таблица 37 – ФРАГМЕНТЫ МНОГОУРОВНЕВОЙ СЕМАНТИЧЕСКОЙ
ИНФОРМАЦИОННОЙ МОДЕЛИ И МНОГОСЛОЙНОЙ НЕЙРОННОЙ СЕТИ
СО СВЯЗЯМИ РАЗЛИЧНОЙ СТЕПЕНИ ОПОСРЕДОВАННОСТИ**

Уровень		Нейронная сеть			
№	Наименование	Слои со связями 0-го уровня опосредованности	Слои со связями 1-го уровня опосредованности	Слои со связями 1-го уровня опосредованности	Слои со связями 2-го уровня опосредованности
4	Уровни качества Жизни (значения Интегрального критерия уровня качества жизни)				
3	Годы				
3	Годы				
2	Вторичные факторы (частные критерии уровня качества жизни)				
2	Вторичные факторы (частные критерии уровня качества жизни)				
1	Первичные факторы				

Обучение нелокальной нейронной сети

В классических нейронных сетях обучение состоит в таком подборе весовых коэффициентов на дендритах, что минимизируется ошибка выходного сигнала сети по сравнению с эталонным. Этот вид обучения аналогичен обучению человека с уже сформированным мозгом, т.е. в достаточно зрелом возрасте, когда структура нейронной сети уже сформирована и фиксирована.

Однако в раннем возрасте и до наступления зрелости может быть основную роль играет другой вид обучения, который состоит в том, что формируется и перестраивается сама структура нейронной сети. Этот процесс включает:

- формирование новых связей между уже существующими нейронами;
- формирование новых нейронов;
- формирование связей между уже существующими и новыми нейронами;
- формирование связей между новыми нейронами;
- формирование новых нейронных систем (ансамблей) и слоев.

В классической нейронной сети обучение в каждом конкретном случае затрагивает лишь те нейроны, которые физически связаны с входными рецепторами и конкретными выходными аксонами.

В нелокальной нейронной сети могут возникать и новые связи между рецепторами и аксонами (*адаптивность* СИМ и ННС), а также добавляться новые рецепторы и нейроны (*пересинтез* СИМ и ННС).

Классические когнитивные карты

Классическая когнитивная карта представляет собой нейрон, соответствующий некоторому состоянию объекта управления с рецепторами, каждый из которых соответствует фактору в определенной степени способствующему или препятствующему переходу объекта в это состояние. Рецепторы соединены связями как с нейроном, так и друг с другом. Связи рецепторов с нейроном отражают силу и направление влияния факторов, а связи рецепторов друг с другом, отображаемые в форме семантической сети факторов, – сходство и различие между рецепторами по характеру их влияния на объект управления. Таким образом, классическая когнитивная карта представляет собой нейрон с семантической сетью факторов, изображенные на одной диаграмме (рисунок 81).

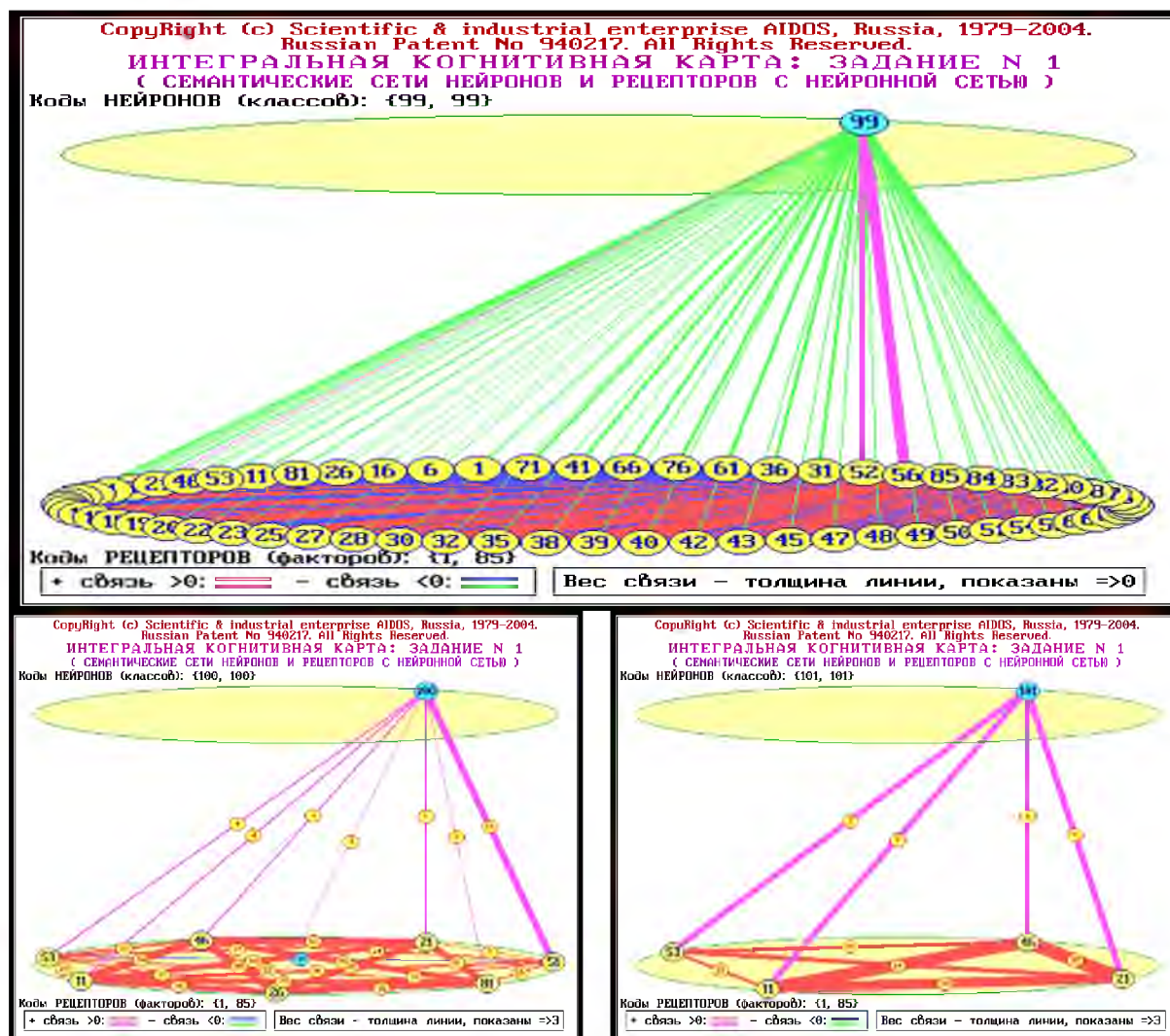


Рисунок 81. Примеры классических когнитивных карт, отражающих влияние структуры инвестиций на уровень качества жизни

Обобщенные когнитивные карты

Если объединить несколько классических когнитивных карт на одной диаграмме и изобразить на ней также связи между нейронами в форме семантической сети классов, то получим обобщенную (интегральную) когнитивную карту. На рисунке 82 приведена обобщенная когнитивная карта, отражающая резуль-

таты идентификации лет с помощью интегрального критерия уровня качества жизни, на рисунке 83 – влияние инвестиций на уровень качества жизни.

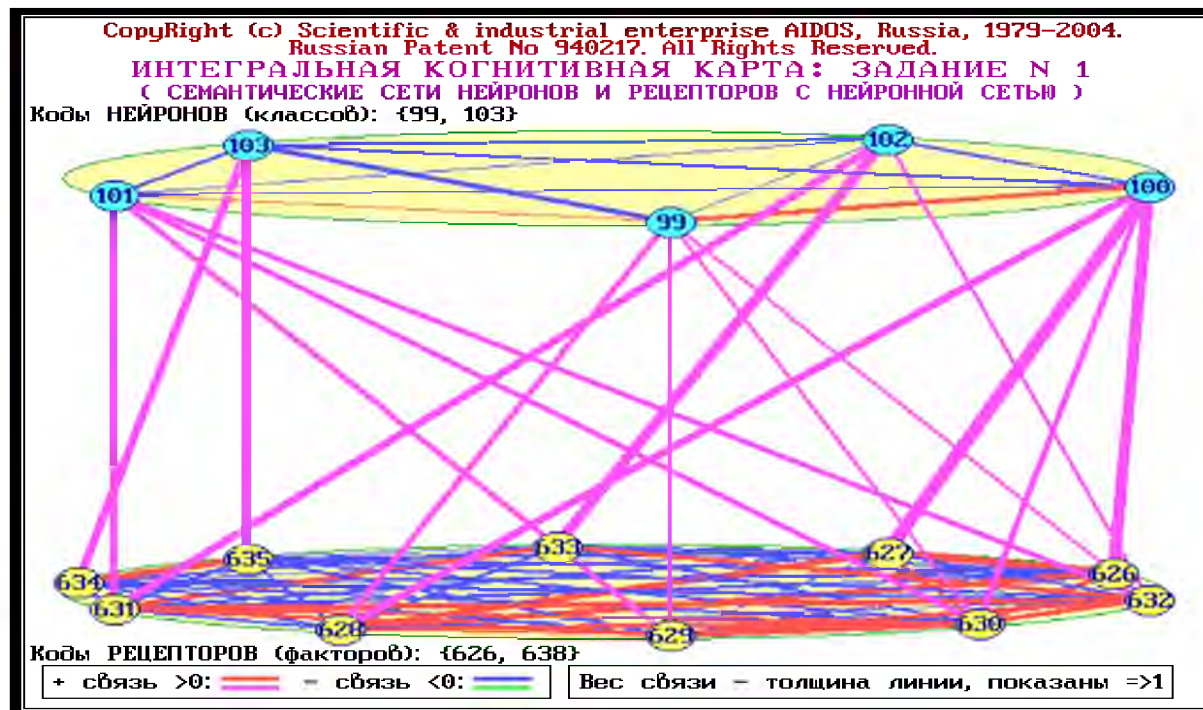


Рисунок 82. Результаты оценки лет с 1991 по 2003
с помощью интегрального критерия уровня качества жизни

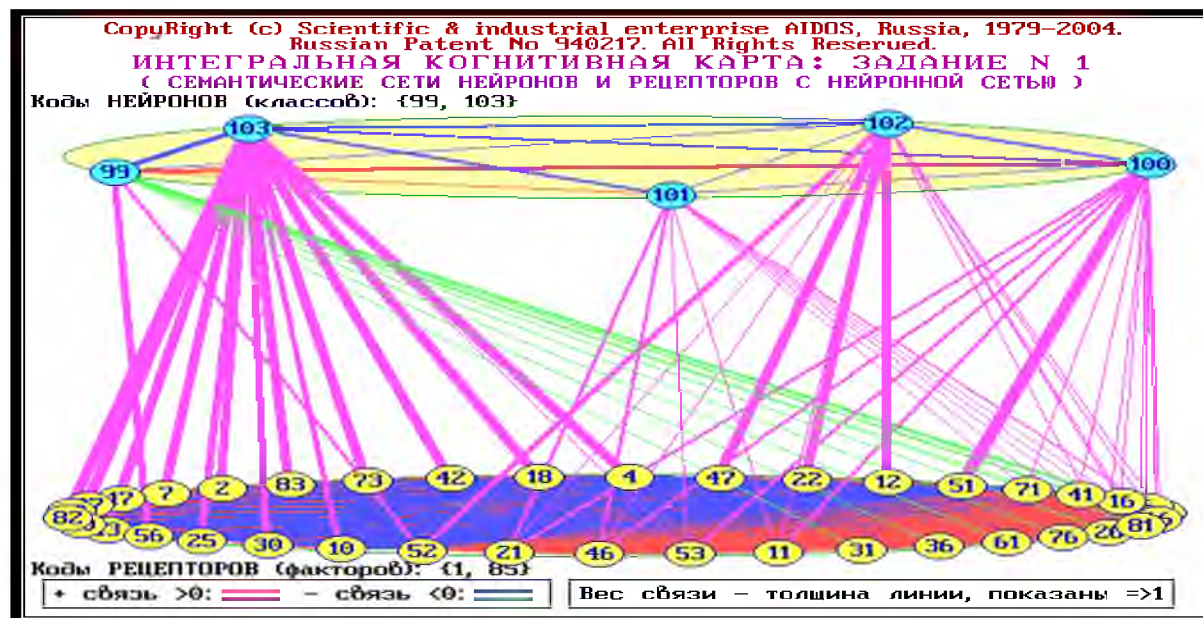


Рисунок 83. Обобщенная (интегральная) когнитивная карта,
визуализирующая связи 2-й степени опосредованности МСИМ
между структурой инвестиций и уровнем качества жизни

Система "Эйдос" обеспечивает построение любого подмножества многоуровневой семантической информационной модели с заданными или выбираемыми по заданным критериям рецепторами и нейронами, связанными друг с

другом связями любого уровня опосредованности в форме классических и обобщенных когнитивных карт. В частности, в системе полуавтоматически формируется задание на генерацию подмножеств обобщенной когнитивной карты, показанных на рисунках 84.

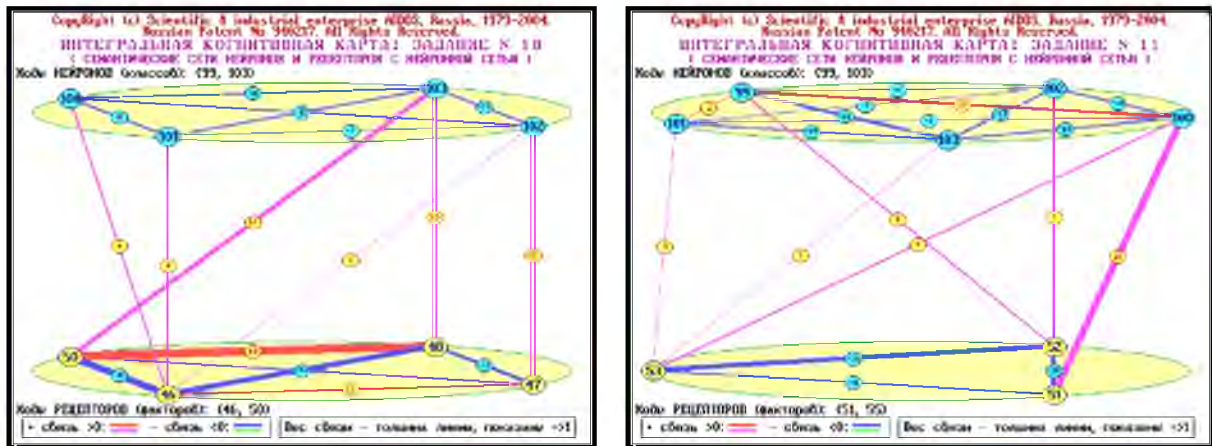


Рисунок 84. Примеры подмножеств интегральной когнитивной карты, отражающие влияние объемов инвестиций по конкретным отраслям на уровень качества жизни населения региона (система "Эйдос")

Резюме

Предлагается модель нелокального нейрона, являющаяся обобщением классической модели Дж. Маккалоки и У. Питтса. Суть нелокальности данной модели состоит в том, что весовые коэффициенты каждого нейрона зависят не только от нейронов, прямо или косвенно соединенных с ним синаптически, но и от всех остальных нейронов сети, не затрагиваемых при обратном распространении ошибки от данного нейрона. Предлагается новый класс нейронных сетей: "Нелокальные интерпретируемые нейронные сети прямого счета" (нелокальные нейронные сети – ННС). Организация ННС обеспечивает один из вариантов решения традиционных для нейронных сетей проблем:

- содержательной интерпретации смысла интенсивности входных сигналов и весовых коэффициентов (*"проблема интерпретируемости весовых коэффициентов"*);
- содержательной интерпретации и обоснования аддитивности аргумента и вида активационной (передаточной) функции нейрона (*"проблема интерпретируемости передаточной функции"*);
- "Комбинаторного взрыва" при определении структуры связей нейронов, подборе весовых коэффициентов и передаточных функций (*"проблема размерности"*);
- *"проблема линейной разделимости классов"* в случае отсутствия вариативности весовых коэффициентов нейронов, соответствующих тем или иным классам.

Математическая модель ННС основана на предложенной автором системной теории информации и семантической информационной модели автоматизированного системно-когнитивного анализа (АСК-анализ), и в отличие от известных нейронных сетей, обеспечивают автоматизацию всех 10 базовых когнитив-

ных операций, образующих "когнитивный конфигуратор". Предложены не только математическая модель, но также и соответствующий численный метод (включая алгоритмы и структуры данных), а также программный инструментальный нелокальных нейронных сетей (универсальная когнитивная аналитическая система "Эйдос" версии 12.5), успешно апробированные в ряде предметных областей. Данная система обеспечивает неограниченное количество слоев ННС при максимальном количестве весовых коэффициентов в слое до 16 миллионов (в текущей версии), до 4000 выходных нейронов, а также автоматическую визуализацию и запись в виде графических файлов сформированных моделей нелокальных нейронов и Паретто-подмножеств нелокальной нейронной сети.

Контрольные вопросы

1. Биологический нейрон и формальная модель нейрона Маккалоки и Питтса.
2. Возможность решения простых задач классификации непосредственно одним нейроном.
3. Однослойная нейронная сеть и персептрон Розенблата.
4. Линейная разделимость и персептронная представляемость.
5. Многослойные нейронные сети.
6. Многослойный персептрон.
7. Модель Хопфилда.
8. Когнитрон и неокогнитрон Фукушимы.
9. Проблемы и перспективы нейронных сетей.
10. Модель нелокального нейрона и нелокальные интерпретируемые нейронные сети прямого счета.
11. Метафора нейросетевого представления семантической информационной модели.
12. Соответствие основных терминов и понятий.
13. Гипотеза о нелокальности нейрона и информационная нейросетевая парадигма.
14. Решение проблемы интерпретируемости весовых коэффициентов (семантическая мера целесообразности информации и закон Фехнера).
15. Семантическая информационная модель, как нелокальная нейронная сеть.
16. Гипотеза о физической природе нелокального взаимодействия нейронов в нелокальной нейронной сети.
17. Решение проблемы интерпретируемости передаточной функции.
18. Решение проблемы размерности.
19. Решение проблемы линейной разделимости.
20. Моделирование причинно-следственных цепочек в нейронных сетях и семантической информационной модели.
21. Моделирование иерархических структур обработки информации.
22. Нейронные сети и СК-анализ.
23. Графическое отображение нейронов, Паретто-подмножеств нелокальной нейронной сети, семантических сетей, когнитивных карт и диаграмм в системе "Эйдос".

Рекомендуемая литература

1. Lutsenko E.V. Conceptual principles of the system (emergent) information theory & its application for the cognitive modelling of the active objects (entities) //2002 IEEE International Conference on Artificial Intelligence System (ICAIS 2002).

–Computer society, IEEE, Los Alamos, California, Washington-Brussels-Tokyo, p. 268-269.

2. Бранский В.П. Философские основания проблемы синтеза релятивистских и квантовых принципов. –Л: ЛГУ, 1973. –175с.

3. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

4. Луценко Е.В. Интерференция последствий выбора в результате одновременного выбора альтернатив и необходимость разработки эмерджентной теории информации. //В сб.: "Материалы III всероссийской межвузовской научно-технической конференции". – Краснодар: КВИ, 2002. – С.24-30.

5. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.

6. Роберт Г.Джан, Бренда Дж.Данн. Границы реальности. (Роль сознания в физическом мире). /Пер. с англ. - М.: Объединенный институт высоких температур РАН, 1995. - 287с.

7. Цехмистро И.З. Поиски квантовой концепции физических оснований сознания. –Харьков: ХГУ, 1981. - 275с.

8. Терехов С.А. Лекции по теории и приложениям искусственных нейронных сетей. Лаборатория Искусственных Нейронных Сетей НТО-2, ВНИИТФ, Снежинск, http://alife.narod.ru/lectures/neural/Neu_index.htm.

9. Ткачев А.Н., Луценко Е.В. Формальная постановка задачи и синтез многоуровневой семантической информационной модели влияния инвестиций на уровень качества жизни населения региона // Научный журнал КубГАУ. – 2004.– №4(6). –22 с. <http://ej.kubagro.ru>

10. Ткачев А.Н., Луценко Е.В. Исследование многоуровневой семантической информационной модели влияния инвестиций на уровень качества жизни населения региона // Научный журнал КубГАУ. – 2004.– №4(6). –28 с. <http://ej.kubagro.ru>

11. Сайт "Курс статистики", раздел "Нейронные сети": <http://www.statsoft.ru/home/textbook/modules/stneunet.html>.

1.3.6. ЛЕКЦИЯ-12.

Генетические алгоритмы и моделирование биологической эволюции

Учебные вопросы

1. Основные понятия, принципы и предпосылки генетических алгоритмов.
2. Пример работы простого генетического алгоритма.
3. Достоинства и недостатки генетических алгоритмов.
4. Примеры применения генетических алгоритмов.

1.1.6.1. Основные понятия, принципы и предпосылки генетических алгоритмов

Генетические Алгоритмы (ГА) – это адаптивные методы функциональной оптимизации, основанные на компьютерном **имитационном моделировании биологической эволюции**. Основные принципы ГА были сформулированы Голландом (Holland, 1975), и хорошо описаны во многих работах и на ряде сайтов в Internet.

В настоящее время существует ряд теорий биологической эволюции (Ж.-Б.Ламарка, П.Тейяра де Шардена, К.Э.Бэра, Л.С.Берга, А.А.Любищева, С.В.Мейена и др.), однако, ни одна из них не считается общепризнанной. Наиболее известной и популярной, конечно, является теория Чарльза Дарвина, которую он представил в работе "Происхождение Видов" в 1859 году.

Эта теория, как и другие, содержит довольно много *нерешенных проблем*, глубокое рассмотрение которых далеко выходит за рамки данной работы. Здесь мы можем отметить лишь некоторые наиболее известные из них. Как это ни парадоксально, но несмотря на то, что сам Чарльз Дарвин назвал свою работу "Происхождение Видов" но как раз именно *происхождения видов* она и не объясняет. Дело в том, что возникновение нового вида "по алгоритму Дарвина" является крайне маловероятным событием, т.к. для этого требуется случайное возникновение в одной точке пространства и времени сразу не менее 100 особей нового вида, т.е. особей, которые могли бы иметь плодовитое потомство. При меньшем количестве особей вид обречен на вымирание. Поэтому процесс видообразования на основе *случайных* мутаций должен был бы занять несуразно много времени (по некоторым оценкам даже в намного раз больше, чем время существования Вселенной). Кроме того, "алгоритм Дарвина" не объясняет явной *системности* в многообразии возникающих форм, типа *закона гомологичных рядов* Н.И. Вавилова. Поэтому Л.С. Берг предложил очень интересную концепцию *ногогенеза* – *закономерной* или *направленной* эволюции живого. В этой концепции предполагается, что филогенез имеет определенное направление и *смена форма является не случайной, а задается некоторым вектором*, природа которого не ясна. Идеи ногогенеза глубоко разработал и развил А.А. Любищев, высказавший гипотезу о математических закономерностях, которые определяют

многообразие живых форм. Кроме того, Дарвин не смог показать *механизм наследования*, при котором поддерживается и закрепляется изменчивость. Это было на пятьдесят лет до того, как генетическая теория наследственности начала распространяться по миру, и за тридцать лет до того, как "эволюционный синтез" укрепил связь между теорией эволюции и молодой генетикой.

Тем ни менее и не смотря на свои недостатки, *именно теория Дарвина традиционно и моделируется в ГА*, хотя, конечно, это не исключает возможности моделирования и других теорий эволюции в ГА. *Более того, возможно именно такое компьютерное моделирование и сравнение его результатов с картиной реальной эволюции жизни на Земле может быть и сыграет положительную роль в дальнейшей разработке наиболее адекватной теории биологической эволюции.*

Теория Дарвина применима не к отдельным особям, а к *популяциям* — большому количеству особей *одного вида*, т.е. способных давать плодовитое потомство, находящейся в определенной статичной или динамичной внешней среде.

В основе модели эволюции Дарвина лежат *случайные изменения* отдельных материальных элементов живого организма при переходе от поколения к поколению. Целесообразные изменения, которые облегчают выживание и производство потомков в данной конкретной внешней среде, сохраняются и передаются потомству, т.е. *наследуются*. Особи, не имеющие соответствующих приспособлений, погибают, не оставив потомства или оставив его меньше, чем приспособленные (считается, что количество потомства пропорционально степени приспособленности). Поэтому в результате *естественного отбора* возникает популяция из наиболее приспособленных особей, которая может стать основой нового вида.

Естественный отбор происходит в условиях *конкуренции* особей популяции, а иногда и различных видов, друг с другом за различные *ресурсы*, такие, например, как пища или вода. Кроме того, члены популяции одного вида часто конкурируют за привлечение брачного партнера. Те особи, которые наиболее приспособлены к окружающим условиям, будут иметь относительно больше шансов воспроизвести потомков. Слабо приспособленные особи либо совсем не произведут потомства, либо их потомство будет очень немногочисленным. Это означает, что гены от высоко адаптированных или приспособленных особей будут распространяться в увеличивающемся количестве потомков на каждом последующем поколении.

Таким образом, по сути дела каждый конкретный *генетический алгоритм представляют имитационную модель некоторой определенной теории биологической эволюции или ее варианта*. Вместе с тем необходимо отметить, что сами исследователи биологической эволюции пока еще не до конца определились с критериями и методами определения степени существенности для поддерживаемой ими теории эволюции тех или иных биологических процессов, которые собственно и моделируются в генетических алгоритмах.

1.1.6.2. Пример работы простого генетического алгоритма

На рисунке 85 приведен пример простого генетического алгоритма.

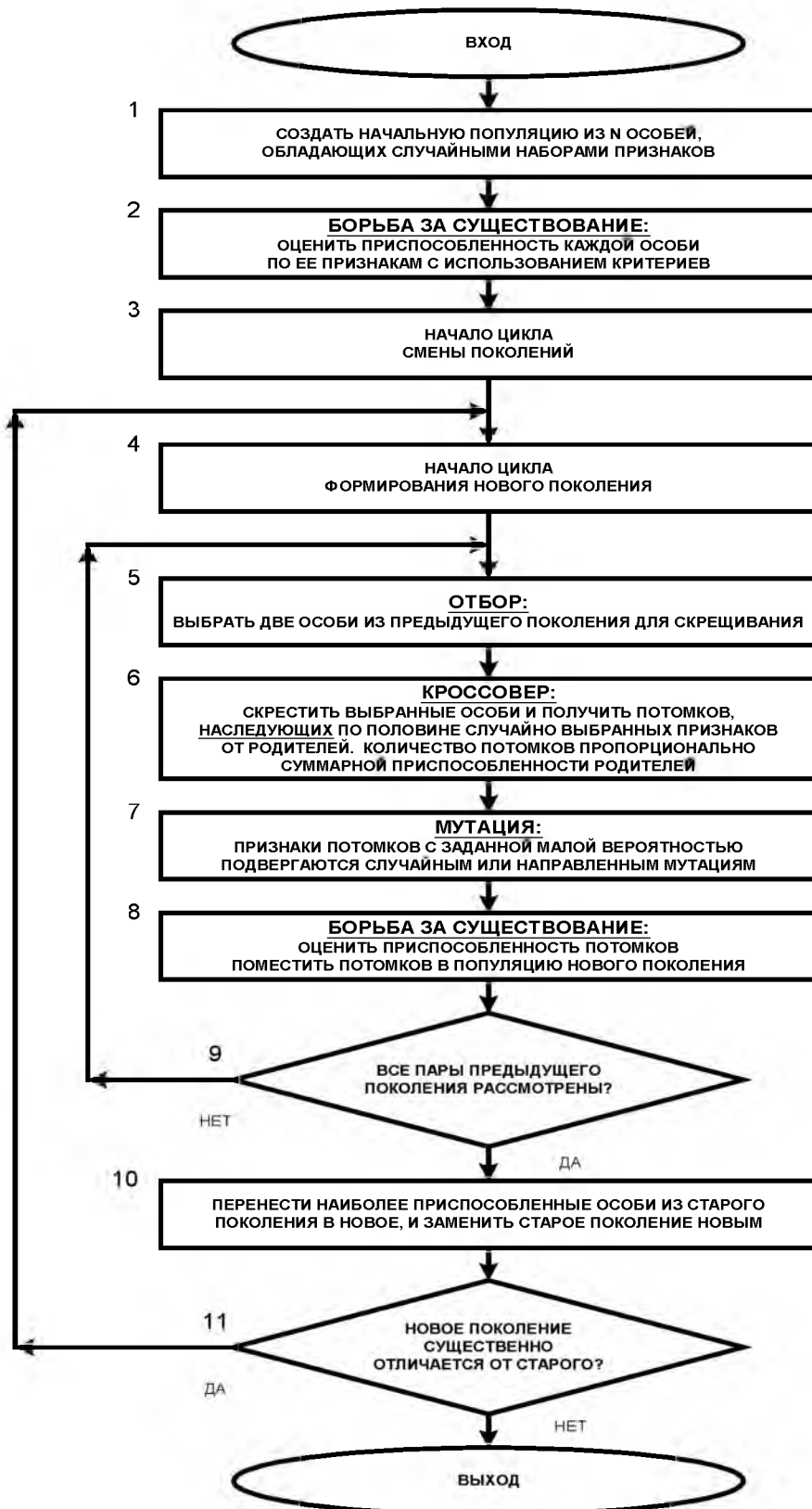


Рисунок 85. Простой генетический алгоритм

Работа ГА представляет собой итерационный процесс, который продолжается до тех пор, пока поколения не перестанут существенно отличаться друг от друга, или не пройдет заданное количество поколений или заданное время. Для каждого поколения реализуются отбор, кроссовер (скрещивание) и мутация. Рассмотрим этот алгоритм.

Шаг 1: генерируется начальная популяция, состоящая из N особей со случайными наборами признаков.

Шаг 2 (борьба за существование): вычисляется абсолютная приспособленность каждой особи популяции к условиям среды $f(i)$ и суммарная приспособленность особей популяции, характеризующая приспособленность всей популяции. Затем при *пропорциональном отборе* для каждой особи вычисляется ее *относительный вклад в суммарную приспособленность популяции* $P_s(i)$, т.е. отношение ее абсолютной приспособленности $f(i)$ к суммарной приспособленности всех особей популяции (3):

$$P_s(i) = \frac{f(i)}{\sum_{i=1}^N f(i)} \quad (3)$$

В выражении (3) сразу обращает на себя внимание возможность сравнения абсолютной приспособленности i -й особи $f(i)$ не с суммарной приспособленностью всех особей популяции, а со средней абсолютной приспособленностью особи популяции (4):

$$\bar{f} = \frac{1}{N} \sum_{i=1}^N f(i) \quad (4)$$

Тогда получим (5):

$$P(i) = \frac{f(i)}{\bar{f}} = \frac{f(i)}{\frac{1}{N} \sum_{i=1}^N f(i)} \quad (5)$$

Если взять логарифм по основанию 2 от выражения (5), то получим *количество информации, содержащееся в признаках особи о том, что она выживает и даст потомство* (6).

$$I(i) = \log_2 \frac{f(i)}{\bar{f}} \quad (6)$$

Необходимо отметить, что эта формула совпадает с формулой для семантического количества информации Харкевича, если целью считать *индивидуальное выживание и продолжение рода*. Это значит, что даже чисто формально *приспособленность особи представляет собой количество информации, содержащееся в ее фенотипе о продолжении ее генотипа в последующих поколениях*.

Поскольку количество потомства особи пропорционально ее приспособленности, то естественно считать, что *если это количество информации:*

- *положительно*, то данная особь выживает и дает потомство, численность которого пропорциональна этому количеству информации;
- *равно нулю*, то особь доживает до половозрелого возраста, но потомства не дает (его численность равна нулю);
- *меньше нуля*, то особь погибает до достижения половозрелого возраста.

Таким образом, можно сделать фундаментальный вывод, имеющий даже мировоззренческое звучание, о том, что *естественный отбор представляет собой процесс генерации и накопления информации о выживании и продолжении рода в ряде поколений популяции, как системы.*

Это накопление информации происходит на различных уровнях иерархии *популяции, как системы*, включающей:

- элементы системы: отдельные особи;
- взаимосвязи между элементами: отношения между особями в популяции, обеспечивающие передачу последующим поколениям максимального количества информации об их выживании и продолжении рода (путем скрещивания наиболее приспособленных особей и наследования рациональных приобретений);
- цель системы: сохранение и развитие популяции, реализуется через цели особей: индивидуальное выживание и продолжение рода.

Фенотип соответствует генотипу и представляет собой его внешнее проявление в признаках особи. Особь взаимодействует с окружающей средой и другими особями в соответствии со своим фенотипом. В случае, если это взаимодействие удачно, то особь передает генетическую информацию, определяющую фенотип, последующим поколениям.

Шаг 3: начало цикла смены поколений.

Шаг 4: начало цикла формирования нового поколения.

Шаг 5 (отбор): осуществляется *пропорциональный отбор* особей, которые могут участвовать в продолжении рода. Отбираются только те особи популяции, у которых количество информации в фенотипе и генотипе о выживании и продолжении рода положительно, причем вероятность выбора пропорциональна этому количеству информации.

Шаг 6 (кроссовер): отобранные для продолжения рода на предыдущем шаге особи с заданной вероятностью P_c подвергаются *скрещиванию* или *кроссоверу (рекомбинации)*.

Если кроссовер происходит, то потомки получают по половине случайным образом определенных признаков от каждого из родителей. Численность потомства пропорциональна суммарной приспособленности родителей. В некоторых вариантах ГА потомки после своего появления заменяют собой родителей и переходят к мутации.

Если кроссовер не происходит, то исходные особи – несостоявшиеся родители, переходят на стадию мутации.

Шаг 7 (мутация): выполняются операторы *мутации*. При этом признаки потомков с вероятностью P_m случайным образом изменяются на другие. Отметим, что использование механизма случайных мутаций роднит генетические алгоритмы с таким широко известным методом имитационного моделирования, как *метод Монте-Карло*.

Шаг 8 (борьба за существование): оценивается приспособленность потомков (по тому же алгоритму, что и на шаге 2).

Шаг 9: проверяется, все ли отобранные особи дали потомство.

Если нет, то происходит переход на шаг 5 и продолжается формирование нового поколения, иначе – переход на следующий шаг 10.

Шаг 10: происходит смена поколений:

- потомки помещаются в новое поколение;
- наиболее приспособленные особи из старого поколения переносятся в новое, причем для каждой из них это возможно не более заданного количества раз;
- полученная новая популяция замещает собой старую.

Шаг 11: проверяется выполнение условия останова генетического алгоритма. **Выход** из генетического алгоритма происходит либо тогда, когда новые поколения перестают существенно отличаться от предыдущих, т.е., как говорят, "алгоритм сходится", либо когда пройдено заданное количество поколений или заданное время работы алгоритма (чтобы не было "зацикливания" и динамического зависания в случае, когда решение не может быть найдено в заданное время).

Если ГА сошелся, то это означает, что решение найдено, т.е. получено поколение, идеально приспособленное к условиям данной фиксированной среды обитания.

Иначе – переход на шаг 4 – начало формирования нового поколения.

В реальной биологической эволюции этим дело не ограничивается, т.к. любая популяция кроме освоения некоторой экологической ниши пытается также выйти за ее пределы освоить и другие ниши, как правило "смежные". Именно за счет этих процессов жизнь вышла из моря на сушу, проникла в воздушное пространство и поверхностный слой почвы, а сейчас осваивает космическое пространство.

Конечно, реальные генетические алгоритмы, на которых проводятся научные исследования, чаще всего мало похожи на приведенный пример. Исследователи экспериментируют с различными параметрами генетических алгоритмов, например: способами отбора особей для скрещивания; критериями приспособленности и жесткостью влияния факторов среды; способами выбора признаков, передающихся от родителей потомкам (рецессивные и не рецессивные гены и т.д.); интенсивностью, видом случайного распределения и направленностью мутаций; различными подходами к воспроизводству и отбору.

Поэтому под термином "генетические алгоритмы" по сути дела надо понимать не одну модель, а довольно широкий класс алгоритмов, подчас мало похожих друг на друга.

В настоящее время рассматривается много различных операторов отбора, кроссовера и мутации: **турнирный отбор** (Brindle, 1981; Goldberg и Deb, 1991), реализует n турниров, чтобы выбрать n особей, при этом каждый турнир построен на выборке k элементов из популяции, и выбора лучшей особи среди них (наиболее распространен турнирный отбор с $k=2$); **элитный отбор** (De Jong, 1975) гарантируют, что при отборе обязательно будут выживать лучший или лучшие члены популяции совокупности (наиболее распространена процедура обязательного сохранения только одной лучшей особи, если она не прошла как другие через процесс отбора, кроссовера и мутации); **двухточечный кроссовер** (Cavichio, 1970; Goldberg, 1989с) и равномерный кроссовер (Syswerda, 1989) отличаются способами наследования потомками признаков родителей.

Не смотря на то, что модели биологической эволюции, реализуемые в ГА, обычно сильно упрощены по сравнению с природным оригиналом, тем ни менее ГА являются мощным средством, которое может с успехом применяться для решения широкого класса прикладных задач, включая те, которые трудно, а иногда и вовсе невозможно, решить другими методами.

1.1.6.3. Достоинства и недостатки генетических алгоритмов

Однако, ГА не гарантирует обнаружения глобального решения за приемлемое время. ГА не гарантируют и того, что найденное решение будет оптимальным решением. Тем ни менее они применимы для *поиска* "достаточно хорошего" решения задачи за "достаточно короткое время". ГА представляют собой разновидность алгоритмов поиска и имеют преимущества перед другими алгоритмами при очень больших размерностях задач и отсутствия упорядоченности в исходных данных, когда альтернативой им является метод полного перебора вариантов.

В случаях, когда задача может быть решена специально разработанным для нее методом, практически всегда такие методы будут эффективнее ГА как по быстрдействию, так и по точности найденных решений.

Главным же достоинством ГА является то, что они могут применяться для решения сложных неформализованных задач, для которых не разработано специальных методов, т.е. ГА обеспечивают решение проблем. Но даже в тех случаях, для которых хорошо работают существующие методики, можно достигнуть интересных результатов сочетая их с ГА.

1.1.6.4. Примеры применения генетических алгоритмов

Данный раздел основан на статье Алексея Андреева "Электродарвин" (<http://www.fuga.ru/articles/2004/03/genetic-pro.htm>), к которой мы и отсылаем за очень интересными подробностями.

В 1994 году Эндрю Кин из университета Саутгемптона использовал генетический алгоритм в дизайне космических кораблей. За основу была взята модель опоры космической станции, спроектированной в NASA из которой после смены 15 поколений, включавших 4.500 вариантов дизайна, получилась модель, превосходящая по тестам тот вариант, что разработали люди.

Аналогичный генетический алгоритм был использован NASA при разработке антенны для спутника.

Джон Коца из Стэнфорда разработал технологию генетического программирования, в которой результатом эволюции становятся не отдельные числовые параметры "особей", а целые имитационные программы, которые являются виртуальными аналогами реальных устройств. Эта технология позволила компании Genetic Programming повторить 15 человеческих изобретений, 6 из которых были запатентованы после 2000 года, то есть представляют собой самые передовые достижения, а один из контроллеров, "выведенных" в GP, даже превосходит аналогичную человеческую разработку.

Сейчас плоды электронной эволюции можно найти в самых разных сферах: от двигателя самолета Boeing 777 до новых антибиотиков.

Генетические алгоритмы представляют собой компьютерное моделирование эволюции. Материальное воплощение сконструированных таким образом систем до сих пор была невозможна без участия человека. Однако интенсивно ведутся работы, результатом которых является уменьшение зависимости машинной эволюции от человека. Эти работы ведутся по двум основным направлениям:

1. Естественный отбор, моделируемый ГА, переносится из виртуального мира в реальный, например, проводятся эксперименты по реальным битвам роботов на выживание.

2. Интеллектуальные системы, основанные на ГА, конструируют роботов, которые в принципе могут быть изготовлены на автоматизированных заводах без участия человека.

Пример воплощения ГА в реальной битве роботов на выживание: в 2002 году в британском центре Magna открылся павильон Live Robots, где боролись за выживание 12 роботов двух видов: "гелиофаги", способные добывать электроэнергию с использованием солнечных батарей; "хищники", которые могли получать электроэнергию только от гелиофагов. Выжившие роботы загружали свои "гены" в погибших и, таким образом, образовывали новые поколения. Те хищники, которые забирали всю энергию у гелиофагов, теряли источник питания и погибали, не передавая свою тактику потомкам, поступавшие же "более разумно" продолжили свой род. В результате возникла равновесная сбалансированная искусственная экосистема с двумя популяциями.

Пример конструирования роботов роботами: в Brandeis University была создана программа Golem, которая сама конструировала роботов. В программу была база деталей, а также механизм мутаций и функция пригодности для "отсеивания" неудачников – тех, кто не научился двигаться. После 600 поколений за несколько дней программа получила модели трех ползающих роботов. Показательно, что роботы оказались симметричными, хотя симметрия никак не была явно прописана в правилах эволюции и исходных данных. Это означает, что она появилась в ходе моделирования машинной эволюции как полезная черта, позволяющая двигаться прямолинейно.

Контрольные вопросы

1. Основные понятия, принципы и предпосылки генетических алгоритмов.
2. Пример работы простого генетического алгоритма.
3. Достоинства и недостатки генетических алгоритмов.
4. Примеры применения генетических алгоритмов.

Рекомендуемая литература

1. Исаев С. Популярно о генетических алгоритмах.
<http://home.od.ua/~relayer/algo/neuro/ga-pop/>
2. Алексей Андреев. Электродарвин.
<http://www.fuga.ru/articles/2004/03/genetic-pro.htm>
3. Сотник С.Л. Конспект лекций по курсу "Основы проектирования систем искусственного интеллекта": (1997-1998),
<http://neuroschool.narod.ru/books/sotnik.html>.

1.3.7. ЛЕКЦИЯ-13.

Когнитивное моделирование

Учебные вопросы:

1. Определение основных понятий: "Когнитивное моделирование" и "Классическая когнитивная карта", их связь с когнитивной психологией и гносеологией.
2. Когнитивная (познавательная-целевая) структуризация знаний об исследуемом объекте и внешней для него среды на основе PEST-анализа и SWOT-анализа.
3. Разработка программы реализации стратегии развития исследуемого объекта на основе динамического имитационного моделирования (при поддержке программного пакета Ithink).

1.3.7.1. Определение основных понятий: "Когнитивное моделирование" и "Классическая когнитивная карта", их связь с когнитивной психологией и гносеологией

Термин: "Когнитивный" происходит от "*cognition*" – "познание" (англ.) и используется для обозначения нового перспективного направления психологии (когнитивная психология), а также направления развития систем искусственного интеллекта (когнитивное моделирование и системно-когнитивный анализ), в которых ставится и решается задача автоматизации некоторых функций, реализуемых человеком, в процессе познания.

Исторически процессы познания первоначально изучались философами. В философии теория познания (сознания) называется *гносеологией*, от греч. *gnosis*, – знание, учение, познание, в отличие от *онтологии* – учения о бытие.

Однако, философский анализ процессов познания не касается исследования естественно-научными методами *конкретных форм сознания и характерных для них методов познания, а также конкретных способов их достижения и реализации.*

Когнитивная психология – это область психологии, непосредственно теоретически и экспериментально изучающая процессы познания у конкретных людей, различного пола, возраста, социального статуса и т.д.

Когнитивное моделирование – это способ анализа, обеспечивающий определение силы и направления влияния факторов на перевод объекта управления в целевое состояние с учетом сходства и различия в влиянии различных факторов на объект управления.

Классическая когнитивная карта – это ориентированный граф, в котором привилегированной вершиной является некоторое будущее (как правило, целевое) состояние объекта управления, остальные вершины соответствуют факторам, дуги, соединяющие факторы с вершиной состояния имеют толщину и знак, соответствующий силе и направлению влияния данного фактора на переход объекта управления в данное состояние, а дуги, соединяющие факторы

показывают сходство и различие в влиянии этих факторов на объект управления.

Ведущей научной организацией России, занимающейся разработкой и применением технологии когнитивного анализа, является Институт проблем управления РАН, подразделение: Сектор-51, ученые Максимов В.И., Корноушенко Е.К., Качаев С.В., Григорян А.К. и другие. На их научных трудах в области когнитивного анализа и основывается данная лекция.

В основе технологии когнитивного анализа и моделирования (рисунок 86) лежит когнитивная (познавательльно-целевая) структуризация знаний об объекте и внешней для него среды.

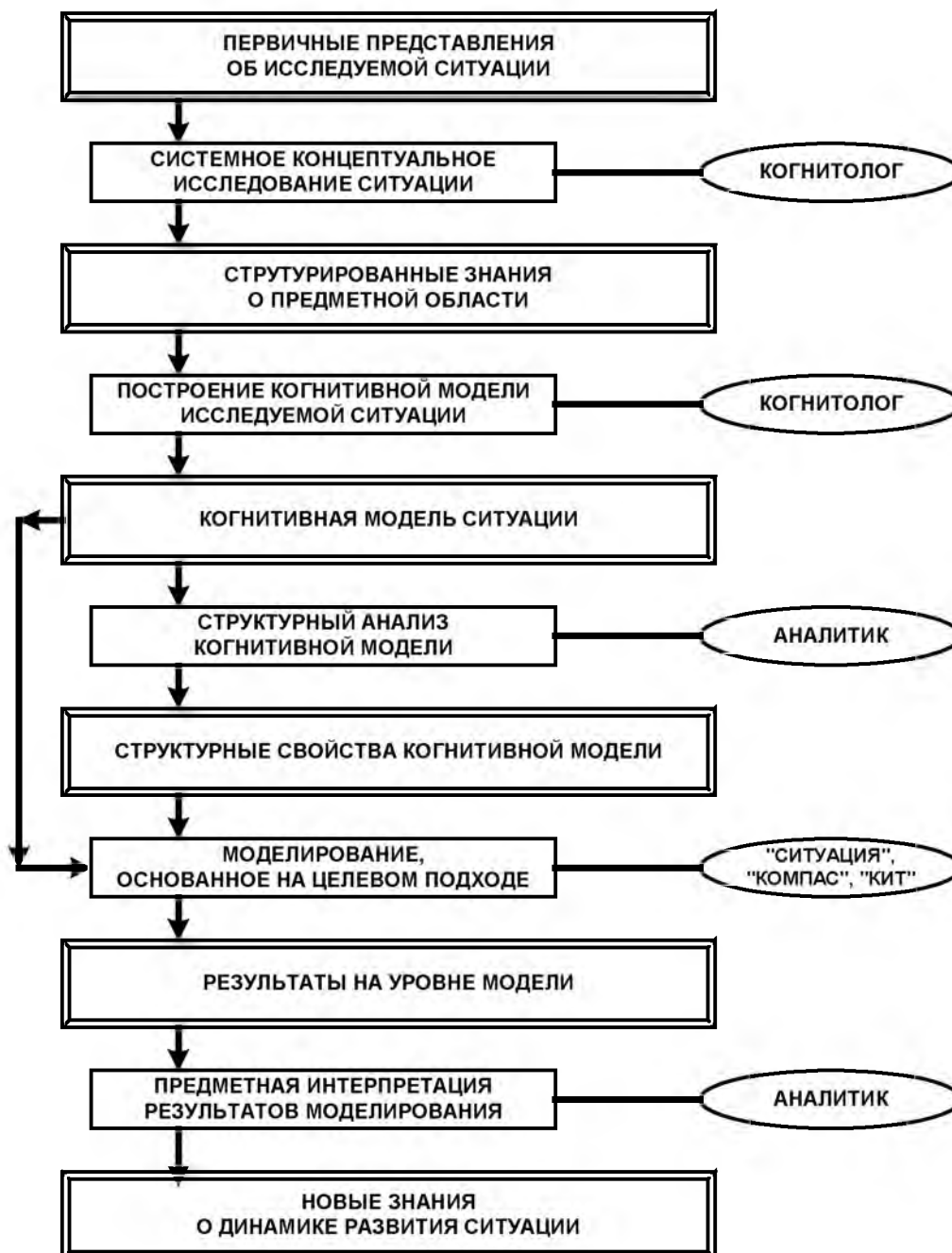


Рисунок 86. Технология когнитивного анализа и моделирования

Когнитивная структуризация предметной области – это выявление будущих целевых и нежелательных состояний объекта управления и наиболее существенных (базисных) факторов управления и внешней среды, влияющих на переход объекта в эти состояния, а также установление на качественном уровне причинно-следственных связей между ними, с учетом взаимовлияния факторов друг на друга.

Результаты когнитивной структуризации отображаются с помощью когнитивной карты (модели).

1.3.7.2. Когнитивная (познавательного-целевая) структуризация знаний об исследуемом объекте и внешней для него среды на основе PEST-анализа и SWOT-анализа

Отбор базисных факторов проводится путем применения *PEST-анализа*, выделяющего четыре основные группы факторов (аспекта), определяющих поведение исследуемого объекта (рисунок 87):

- Policy – политика;
- Economy – экономика;
- Society – общество (социокультурный аспект);
- Technology – технология.

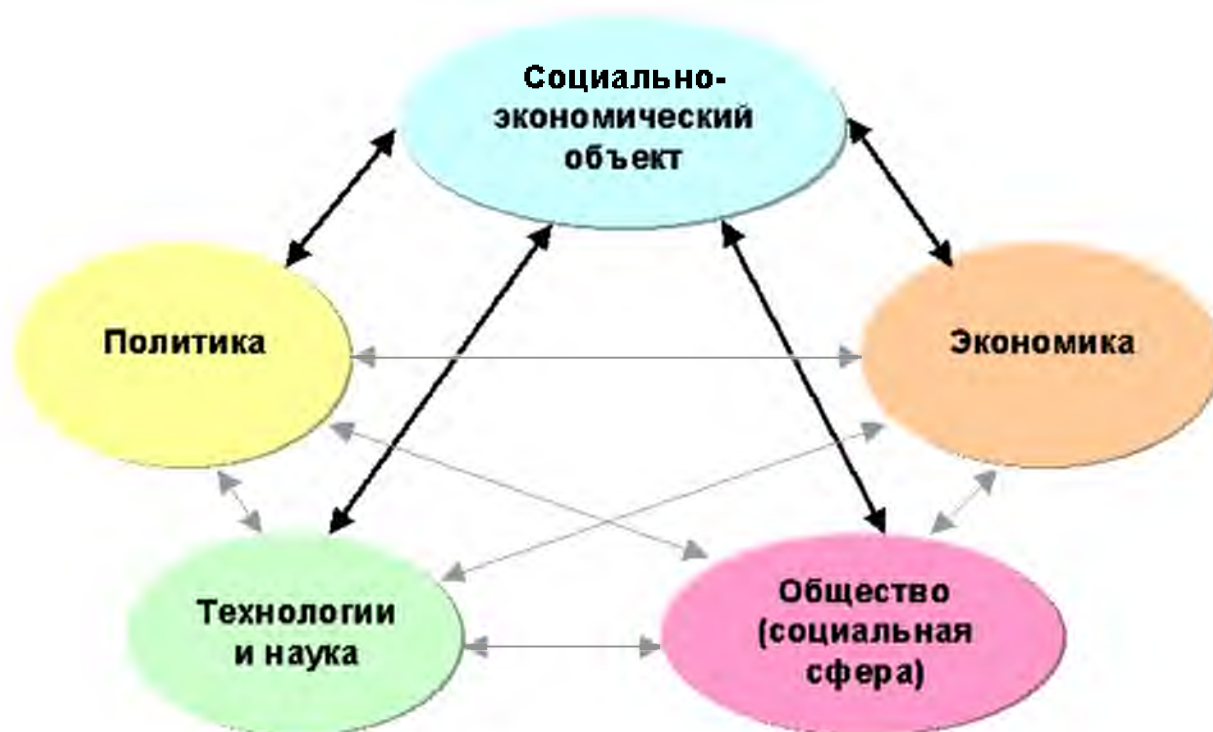


Рисунок 87. Факторы PEST-анализа

Для каждого конкретного сложного объекта существует свой особый набор наиболее существенных факторов, определяющих его поведение и развитие.

PEST-анализ можно рассматривать как вариант системного анализа, т.к. факторы, относящиеся к перечисленным четырём аспектам, в общем случае тес-

но взаимосвязаны и характеризуют различные иерархические уровни общества, как системы.

В этой системе есть детерминирующие связи, направленные с нижних уровней иерархии системы к верхним (наука и технология влияет на экономику, экономика влияет на политику), а также обратные и межуровневые связи. Изменение любого из факторов через эту систему связей может влиять на все остальные.

Эти изменения могут представлять угрозу развитию объекта, или, наоборот, предоставлять новые возможности для его успешного развития.

Следующий шаг – ситуационный анализ проблем, *SWOT-анализ* (рисунок 88):

- **Strengths** – сильные стороны;
- **Weaknesses** – недостатки, слабые стороны;
- **Opportunities** – возможности;
- **Threats** – угрозы.

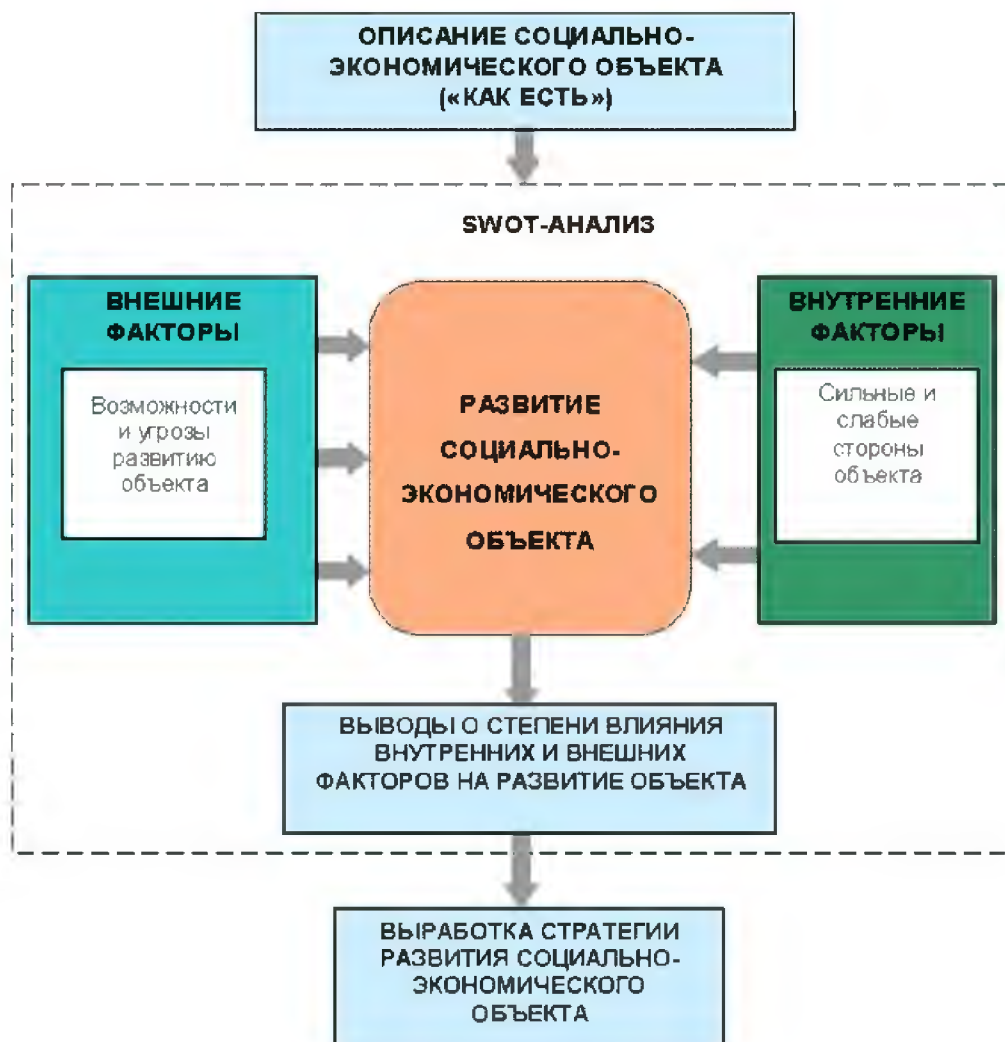


Рисунок 88. Факторы SWOT-анализа

Он включает анализ сильных и слабых сторон развития исследуемого объекта в их взаимодействии с угрозами и возможностями и позволяет опреде-

лить актуальные проблемные области, узкие места, шансы и опасности, с учетом факторов внешней среды.

Возможности определяются как обстоятельства, способствующее благоприятному развитию объекта.

Угрозы – это ситуации, в которых может быть нанесен ущерб объекту, например может быть нарушено его функционирование или он может лишиться имеющихся преимуществ.

На основании анализа различных возможных сочетаний сильных и слабых сторон с угрозами и возможностями формируется *проблемное поле* исследуемого объекта.

Проблемное поле – это совокупность проблем, существующих в моделируемом объекте и окружающей среде, в их взаимосвязи друг с другом.

Наличие такой информации – основа для определения целей (направлений) развития и путей их достижения, выработки стратегии развития.

Когнитивное моделирование на основе проведенного ситуационного анализа позволяет подготовить альтернативные варианты решений по снижению степени риска в выделенных проблемных зонах, прогнозировать возможные события, которые могут тяжелее всего отразиться на положении моделируемого объекта.

Этапы когнитивной технологии и их результаты, представленные на рисунке 86, конкретизированы в таблице 38:

**Таблица 38 – ЭТАПЫ КОГНИТИВНОЙ ТЕХНОЛОГИИ
И РЕЗУЛЬТАТЫ ЕЕ ПРИМЕНЕНИЯ**

Наименование этапа	Форма представления результата
1. Когнитивная (познавательная-целевая) структуризация знаний об исследуемом объекте и внешней для него среды на основе PEST-анализа и SWOT-анализа: Анализ исходной ситуации вокруг исследуемого объекта с выделением базисных факторов, характеризующих экономические, политические и др. процессы, протекающие в объекте и в его макро-окружении и влияющих на развитие объекта. 1.1. Выявление факторов, характеризующих сильные и слабые стороны исследуемого объекта 1.2. Выявление факторов, характеризующих возможности и угрозы со стороны внешней среды объекта 1.3. Построение проблемного поля исследуемого объекта	Отчет о системном концептуальном исследовании объекта и его проблемной области
2. Построение когнитивной модели развития объекта – формализация знаний, полученных на этапе когнитивной структуризации 2.1. Выделение и обоснование факторов 2.2. Установление и обоснование взаимосвязей между факторами 2.3. Построение графовой модели	Компьютерная когнитивная модель объекта в виде ориентированного графа (и матрицы взаимосвязей факторов)

Наименование этапа	Форма представления результата
3. Сценарное исследование тенденций развития ситуации вокруг исследуемого объекта (при поддержке программных комплексов "СИТУАЦИЯ", "КОМПАС", "КИТ") 3.1. Определение цели исследования 3.2. Задание сценариев исследования и их моделирование 3.3. Выявление тенденций развития объекта в его макроокружении 3.4. Интерпретация результатов сценарного исследования	Отчет о сценарном исследовании ситуации, с интерпретацией и выводами
4. Разработка стратегий управления ситуацией вокруг исследуемого объекта 4.1. Определение и обоснование цели управления 4.2. Решение обратной задачи 4.3. Выбор стратегий управления и упорядочивание их по критериям: возможности достижения цели; риска потери управления ситуацией; риска возникновения чрезвычайных ситуаций	Отчет о разработке стратегий управления с обоснованием стратегий по разным критериям качества управления
5. Поиск и обоснование стратегий достижения цели в стабильных или изменяющихся ситуациях <u>Для стабильных ситуаций:</u> а) выбор и обоснование цели управления ; б) выбор мероприятий (управлений) для достижения цели; в) анализ принципиальной возможности достижения цели из текущего состояния ситуации с использованием выбранных мероприятий; г) анализ реальных ограничений на реализацию выбранных мероприятий; д) анализ и обоснование реальной возможности достижения цели; е) выработка и сравнение стратегий достижения цели по: близости результатов управления к намеченной цели; затратам (финансовым, физическим и т.п.); по характеру последствий (обратимые, необратимые) от реализации этих стратегий в реальной ситуации; по риску возникновения чрезвычайных ситуаций <u>Для изменяющихся ситуаций:</u> а) выбор и обоснование текущей цели управления; б) по отношению к текущей цели справедливы предыдущие п.п. б-е; в) анализ изменений, происходящих в ситуации, и их отображение в графовой модели ситуации. Переход к п. а.	Отчет о разработке стратегий достижения цели в стабильных или изменяющихся ситуациях
6. Разработка программы реализации стратегии развития исследуемого объекта на основе динамического имитационного моделирования (при поддержке программного пакета Ithink) 6.1. Распределение ресурсов по направлениям и во времени 6.2. Координация 6.3. Контроль за исполнением	Программа реализации стратегии развития объекта. Компьютерная имитационная модель развития объекта

1.3.7.3. Разработка программы реализации стратегии развития исследуемого объекта на основе динамического имитационного моделирования (при поддержке программного пакета Ithink)

Технология когнитивного анализа и моделирования поддерживается программными комплексами "Ситуация", "Компас", "КИТ" (рисунок 86), созданными в ИПУ РАН, которые позволяют в сложных и неопределенных ситуациях быстро, комплексно и системно охарактеризовать и обосновать сложившуюся ситуацию и на качественном уровне предложить пути решения проблемы в этой ситуации с учетом факторов внешней среды.

Применение когнитивных технологий открывает новые возможности прогнозирования и управления в различных областях:

- в экономической сфере это позволяет в сжатые сроки разработать и обосновать стратегию экономического развития предприятия, банка, региона или даже целого государства с учетом влияния изменений во внешней среде;
- в сфере финансов и фондового рынка – учесть ожидания участников рынка;
- в военной области и области информационной безопасности – противостоять стратегическому информационному оружию, заблаговременно распознавая конфликтные структуры и вырабатывая адекватные меры реагирования на угрозы.

Когнитивные технологии автоматизируют часть функций процессов познания, поэтому они с успехом могут применяться во всех областях, в которых востребовано само познание. Вот лишь некоторые из этих областей:

1. Модели и методы интеллектуальных информационных технологий и систем для создания геополитических, национальных и региональных стратегий социально-экономического развития.
2. Модели выживания "мягких" систем в изменяющихся средах при дефиците ресурсов.
3. Ситуационный анализ и управление развитием событий в кризисных средах и ситуациях.
4. Информационный мониторинг социально-политических, социально-экономических и военно-политических ситуаций.
5. Разработка принципов и методологии проведения компьютерного анализа проблемных ситуаций.
6. Выработка аналитических сценариев развития проблемных ситуаций и управления ими.
7. Подготовка рекомендаций по решению первоочередных стратегических проблем на основе компьютерной системы анализа проблемных ситуаций.
8. Мониторинг проблем в социально-экономическом развитии корпорации, региона, города, государства.
9. Технология когнитивного моделирования целенаправленного развития региона РФ.
10. Анализ развития региона и мониторинг проблемных ситуаций при целенаправленном развитии региона.

11. Модели для формирования государственного регулирования и саморегулирования потребительского рынка.

12. Анализ и управление развитием ситуации на потребительском рынке.

Технология когнитивного моделирования может быть широко использована для уникальных проектов развития регионов, банков, корпораций (и др. объектов) в кризисных условиях после соответствующего обучения.

Контрольные вопросы

1. *Определение основных понятий: "Когнитивное моделирование" и "Классическая когнитивная карта", их связь с когнитивной психологией и гносеологией.*

2. *Когнитивная (познавательнo-целевая) структуризация знаний об исследуемом объекте и внешней для него среды на основе PEST-анализа и SWOT-анализа.*

3. *Разработка программы реализации стратегии развития исследуемого объекта на основе динамического имитационного моделирования (при поддержке программного пакета Ithink).*

Рекомендуемая литература

1. Сайт: ИПУ РАН, Сектор-51 сектор "Когнитивный анализ и моделирование ситуаций": <http://www.ipu.ru/labs/lab51/projects.htm>.

2. Максимов В.И., Корноушенко Е.К. Знание – основа анализа. Банковские технологии, № 4, 1997.

3. Корноушенко Е.К., Максимов В.И. Управление процессами в слабоформализованных средах при стабилизации графовых моделей среды. Труды ИПУ, вып.2, 1998.

4. Максимов В.И., Корноушенко Е.К. Аналитические основы применения когнитивного подхода при решении слабоструктурированных задач. Труды ИПУ, вып.2, 1998.

5. Максимов В.И., Качаев С.В., Корноушенко Е.К. Концептуальное моделирование и мониторинг проблемных и конфликтных ситуаций при целенаправленном развитии региона. В сб. "Современные технологии управления для администраций городов и регионов". Фонд "Проблемы управления", М. 1998.

6. Максимов В.И., Корноушенко Е.К., Качаев С.В. Анализ ситуации и компенсация теневых аспектов в свободной торговле. В сб. "Современные технологии управления для администраций городов и регионов". Фонд "Проблемы управления", М. 1998.

7. Максимов В.И., Корноушенко Е.К., Качаев С.В., Григорян А.К. Когнитивный подход к анализу проблемы демонополизации в транспортном комплексе. Труды ИПУ, вып.2, 1998.

8. Райков А.Н. Аналитическим службам - информационные технологии. /Ваш выбор. 1994. № 4. - С.28-29.

9. Райков А.Н. Гносеологическая декомпозиция процессов рефлексивного управления. /"Рефлексивное управление". Тезисы международного симпозиума (17-19.10.2000). – М.: Ин-т психол. РАН, 2000. – С.89-90.

10. Райков А.Н. Интеллектуальные информационные технологии и системы. В 2-х частях. – М.: МИРЭА, 1998. – 213с.

11. Райков А.Н. Интеллектуальные информационные технологии: Учебное пособие. – М.: МГИРЭА(ТУ), 2000. - 96с.

1.3.8. ЛЕКЦИЯ-14.

Выявление знаний из опыта (эмпирических фактов) и интеллектуальный анализ данных (*data mining*)

Учебные вопросы

1. Интеллектуальный анализ данных (*data mining*)
2. Типы выявляемых закономерностей *data mining*.
3. Математический аппарат *data mining*.
4. Области применения технологий интеллектуального анализа данных.
5. Автоматизированные системы для интеллектуального анализа данных.

Данная лекция основана на работе В. Дюк, А. Самойленко. *Data Mining: учебный курс (+ CD-ROM)*. 2001 г. Издательство: Питер. Серия: Учебный курс. – 368 с., а также материалах сайта компании BI Partner: <http://www.bipartner.ru/services/dm.html>.

1.3.8.1. Интеллектуальный анализ данных

Интеллектуальный анализ данных (ИАД или *data mining*) – это процесс обнаружения в "сырых" данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Обзор материалов Internet, посвященных ИАД показывает, что данное определение является классическим. Тем не менее, на наш взгляд, оно содержит несколько неточностей:

- в определение этого понятия входят слова "*анализ*" и "*знания*", тогда как *знания* нужны для управления, т.е. достижения цели;
- для *интерпретации* доступны не только знания, но уже и *информация* (см. раздел 1.1.2.1. данной работы);
- термин "ИАД" не подразумевает какого-либо одного метода анализа данных, но является *собирательным* и объединяет многие направления исследований и разработок в области СИИ.

Поэтому предлагается другое, более точное определение понятия.

Интеллектуальный анализ данных (ИАД или data mining) – это совокупность математических моделей, численных методов, программных средств и информационных технологий, обеспечивающих обнаружение в эмпирических данных доступной для интерпретации *информации* и синтез на основе этой информации ранее неизвестных, нетривиальных и практически полезных для достижения определенных *целей знаний*.

Технологии *data mining* являются наиболее совершенным инструментом для решения сложных аналитических задач. Они основаны на мощном математическом и статистическом аппарате, корректное применение которого позволяет достичь высоких результатов.

Все большим количеством компаний предлагаются услуги в области интеллектуального анализа данных, что предполагает проведение следующих работ:

- проведение исследования накопленной статистики;
- выявление закономерностей;
- создание моделей данных;
- верификация и апробация моделей данных;
- внедрение модели в практику.

1.3.8.2. Типы выявляемых закономерностей

Успех применения систем *data mining* основан на том, что эти технологии обеспечивают исследование эмпирических данных и выявление в них скрытых закономерностей различных видов.

Ассоциация (идентификация). Если некоторый факт-1 является частью определенного события, то с расчетной вероятностью и другой факт-2, связанный с первым, будет частью того же события.

Последовательность (прогнозирование). Если свершилось некоторое событие-1, то с расчетной вероятностью через определенный период времени свершится другое событие-2, связанное с первым.

Классификация. На основании информации о свойствах объекта ему присваивается определенное дискретное значение показателя, по которому проводится классификация (идентификатор).

Кластеризация. Наиболее сходные по своим признакам объекты объединяются в группы (кластеры) таким образом, что в разных кластерах оказываются наиболее сильно отличающиеся друг от друга объекты. Кластеризация аналогична классификации, но в отличие от последней классы – кластеры объектов заранее не известны, а формируются в процессе кластеризации.

Прогнозирование. Прошлые фактические значения величин используются для прогнозирования будущих значений тех же или других величин на основании на основании знания зависимостей между ними, трендов и статистики.

1.3.8.2.1. Ассоциация

Ассоциация используется для определения закономерностей в событиях или процессах. Ассоциации связывают различные факты одного события. В качестве примера может служить выявленная закономерность, что мужчины, предпочитающие элитные сорта кофе в три раза чаще покупают импортные сигареты, чем мужчины, покупающие обычный кофе.

Результатом ассоциативного анализа являются правила вида: Если факт А является частью события, то с вероятностью X факт В будет частью того же события.

Например:

Если покупатель берет чипсы, то существует 85-ти процентная вероятность, что он приобретет еще легкие алкогольные напитки или пиво.

Если человек едет в отпуск и покупает авиабилеты на всю семью, то с вероятностью 95 процентов они снимут машину в месте отдыха на весь период.

Имея историю продаж в розничном магазине, можно разработать шаблоны покупок (стандартные наборы), например для продуктового отдела. Из-за

широкого применения в торговле ассоциативный анализ часто называют анализом рыночной корзины.

1.3.8.2.2. Последовательность

Последовательные шаблоны аналогичны ассоциациям с той лишь разницей, что связывают события, разнесенные во времени. Например, последовательный шаблон может предсказывать, что человек, купивший посудомоечную машину с вероятностью 0.7 купит сушилку для одежды в течение следующих шести месяцев. Для увеличения этой вероятности магазин может предложить ему скидку в 10% на покупку сушильного аппарата в течение трех или четырех месяцев после покупки посудомоечной машины.

1.3.8.2.3. Классификация

Классификация – наиболее часто используемый метод интеллектуального анализа. Данный метод сосредотачивает внимание на поведении и атрибутах уже существующих групп. Группы могут включать людей, которые часто летают, много тратят, лояльных клиентов, людей, которые откликаются на компании прямой рассылки и т.д. С помощью алгоритмов классификации можно классифицировать объекты по заранее известным характеристикам. В классификации часто используются индукционные алгоритмы, которые позволяют по небольшому массиву данных, отнесенных к известным классам, определять дополнительные классы. Примером использования классификации может служить определение характеристик клиентов, которые вероятно захотят/не захотят приобрести определенный продукт или услугу. Имея такую группировку клиентов, мы можем значительно сократить расходы на акциях продвижения товара или прямой рассылки.

1.3.8.2.4. Кластеризация

Кластеризация используется для обнаружения классов схожих объектов в имеющемся наборе данных. Кластеризация аналогична классификации, но в отличие от последней классы объектов заранее не известны. Методы кластеризации широко используют алгоритмы нейронных сетей и статистическую обработку. В процессе кластеризации средство интеллектуального анализа определяет схожие характеристики объектов и на их основе объединяет объекты в классы. Качество процесса кластеризации определяется схожестью объектов внутри класса и степенью отличия разных классов между собой. Кластеризация обычно применяется для решения таких задач, как определение производственного брака или выявления групп услуг по кредитным карточкам предпочтительных для разных групп клиентов.

1.3.8.2.5. Прогнозирование

Регрессия – один из двух методов прогнозирования. Данный метод использует имеющиеся фактические значения величин для прогнозирования будущих на основании трендов и имеющейся статистики. Например, объем продаж аксессуаров для спортивных машин можно спрогнозировать по количеству проданных спортивных машин в прошлом месяце.

Различие между регрессией и временными рядами состоит в том, что временные ряды предсказывают значения переменных, зависящих от времени. Например, с их помощью можно прогнозировать количество несчастных случаев

во время каникул на основе аналогичных данных за прошлый период. Время в данном случае может содержать иерархии (рабочая неделя, календарная неделя, период) праздники, сезоны, интервалы дат.

1.3.8.3. Математический аппарат

Основой систем *data mining* является выявление различных закономерностей в данных. При этом применяются следующие методы:

- деревья решений;
- алгоритмы кластеризации;
- регрессионный анализ;
- нейронные сети;
- временные ряды.

1.3.8.3.1. Деревья решений

Деревья решения являются одним из наиболее популярных подходов к решению задач *data mining*. Они создают иерархическую структуру классифицирующих правил типа «ЕСЛИ... ТО...», имеющую вид дерева. Для того чтобы решить, к какому классу отнести некоторый объект или ситуацию, требуется ответить на вопросы, стоящие в узлах этого дерева, начиная с его корня. Вопросы имеют вид «значение параметра А больше х». Если ответ положительный, осуществляется переход к правому узлу следующего уровня, если отрицательный – то к левому узлу; затем снова следует вопрос, связанный с соответствующим узлом.

Популярность подхода связана с наглядностью и понятностью. Но очень остро для деревьев решений стоит проблема значимости. Дело в том, что отдельным узлам на каждом новом построенном уровне дерева соответствует все меньшее и меньшее число записей данных – дерево дробит данные на большое количество частных случаев. Чем больше этих частных случаев, чем меньше обучающих примеров попадает в каждый такой частный случай, тем менее уверенной становится их классификация. Если построенное дерево слишком «кустистое» – состоит из неоправданно большого числа мелких веточек – оно не будет давать статистически обоснованных ответов. Как показывает практика, в большинстве систем, использующих деревья решений, эта проблема не находит удовлетворительного решения. Кроме того, общеизвестно, и это легко показать, что деревья решений дают полезные результаты только в случае независимых признаков. В противном случае они лишь создают иллюзию логического вывода.

Область применения деревьев решений в настоящее время широка, но все задачи, решаемые этим аппаратом, могут быть объединены в следующие три класса:

Описание данных. Деревья решений позволяют хранить информацию о данных в компактной форме, вместо них мы можем хранить дерево решений, которое содержит точное описание объектов.

Классификация. Деревья решений отлично справляются с задачами классификации, т.е. отнесения объектов к одному из заранее известных классов. Целевая переменная должна иметь дискретные значения.

Регрессия. Если целевая переменная имеет непрерывные значения, деревья решений позволяют установить зависимость целевой переменной от независимых (входных) переменных. Например, к этому классу относятся задачи численного прогнозирования (предсказания значений целевой переменной).

1.3.8.3.2. Регрессионный анализ

Регрессионный анализ позволяет исследовать формы связи, устанавливающие количественные соотношения между случайными величинами изучаемого процесса.

Регрессия наиболее часто используется для построения прогнозных моделей.

1.3.8.3.3. Нейронные сети

Нейронные представляют собой большой класс систем, архитектура которых пытается имитировать построение нервной ткани из нейронов. В одной из наиболее распространенных архитектур, многослойном персептроне с обратным распространением ошибки, эмулируется работа нейронов в составе иерархической сети, где каждый нейрон более высокого уровня соединен своими входами с выходами нейронов нижележащего слоя. На нейроны самого нижнего слоя подаются значения входных параметров, на основе которых нужно принимать какие-то решения, прогнозировать развитие ситуации и т. д.

Эти значения рассматриваются как сигналы, передающиеся в вышележащий слой, ослабляясь или усиливаясь в зависимости от числовых значений (весов), приписываемых межнейронным связям. В результате на выходе нейрона самого верхнего слоя вырабатывается некоторое значение, которое рассматривается как ответ, реакция всей сети на введенные значения входных параметров. Для того чтобы сеть можно было применять в дальнейшем, ее прежде надо «натренировать» на полученных ранее данных, для которых известны и значения входных параметров, и правильные ответы на них. Эта тренировка состоит в подборе весов межнейронных связей, обеспечивающих наибольшую близость ответов сети к известным правильным ответам.

1.3.8.3.4. Временные ряды

Временной ряд – это расположение во времени статистических показателей, которые в своих последовательных изменениях отражают ход развития изучаемых процессов.

Временные ряды исследуются с различными целями. В одном ряде случаях бывает достаточно получить описание характерных особенностей ряда, а в другом ряде случаев требуется не только предсказывать будущие значения временного ряда, но и управлять его поведением. Метод анализа временного ряда определяется, с одной стороны, целями анализа, а с другой стороны, вероятностной природой формирования его значений.

Спектральный анализ. Позволяет находить периодические составляющие временного ряда

Корреляционный анализ. Позволяет находить существенные периодические зависимости и соответствующие им задержки (лаги) как внутри одного ряда (автокорреляция), так и между несколькими рядами. (кросскорреляция)

Модели авторегрессии и скользящего среднего. Модели ориентированы на описание процессов, проявляющих однородные колебания, возбуждаемые случайными воздействиями. Позволяют предсказывать будущие значения ряда.

1.3.8.4. Области применения технологий интеллектуального анализа данных

Системы, основанные на технологиях интеллектуального анализа данных, используются в компаниях различного профиля. Однако существует целый ряд областей, для которых накоплен богатый и очень успешный опыт применения подобных систем.

Торговля. Анализ потребительской корзины, исследование временных шаблонов, создание прогнозирующих моделей, оптимизация складских запасов.

Банковское дело. Сегментация клиентов, выявление мошенничества с кредитными картами, прогнозирование изменения клиентуры, анализ финансовых рисков.

Страховой бизнес. Сегментация клиентов, выявление фактов мошенничества, анализ страховых рисков, разработка новых продуктов, расчет страховых премий.

Телекоммуникации. Анализ лояльности клиентов, сегментирование клиентской базы и услуг, анализ внешних факторов на отказы оборудования, выявление случаев несанкционированного доступа к сети.

Производственные предприятия. Оптимизация закупок, диагностика брака на ранних стадиях, диагностика оборудования, маркетинг.

Нефтегазовая отрасль. Диагностика оборудования и нефтегазопроводов, прогнозирование цен, разведка месторождений, анализ влияния внешних и внутренних факторов на объемы продаж.

1.3.8.4.1. Розничная торговля

Предприятия розничной торговли сегодня собирают подробную информацию о каждой отдельной покупке, используя кредитные карточки с маркой магазина и компьютеризованные системы контроля. Вот типичные задачи, которые можно решать с помощью технологий data mining в сфере розничной торговли:

Анализ покупательской корзины предназначен для выявления товаров, которые покупатели стремятся приобретать вместе. Знание покупательской корзины необходимо для улучшения рекламы, выработки стратегии создания запасов товаров и способов их раскладки в торговых залах.

Исследование временных шаблонов помогает торговым предприятиям принимать решения о создании товарных запасов. Оно дает ответы на вопросы типа «Если сегодня покупатель приобрел видеокамеру, то, через какое время он вероятнее всего купит новые батарейки и пленку?»

Создание прогнозирующих моделей дает возможность торговым предприятиям узнавать характер потребностей различных категорий клиентов с определенным поведением, например, покупающих товары известных дизайнеров или посещающих распродажи. Эти знания нужны для разработки точно направленных, экономических мероприятий по продвижению товаров.

1.3.8.4.2. Банковская деятельность

Достижения технологии data mining используются в банковском деле для решения проблем Телекоммуникации.

В области телекоммуникаций характерен растущий уровень конкуренции. Здесь методы data mining помогают компаниям более энергично продвигать свои программы маркетинга и ценообразования, чтобы удержать существующих клиентов и привлечь новых. В число типичных мероприятий входят:

- анализ записей о подробных характеристиках вызовов;
- выявление степени лояльности клиентов.

Анализ записей о подробных характеристиках вызовов. Назначение такого анализа – выявление категорий клиентов с похожими стереотипами пользования их услугами и разработка привлекательных наборов цен и услуг.

Выявление степени лояльности клиентов. Некоторые клиенты все время меняют провайдеров, пользуясь программами новых компаний, стимулирующими появление новых клиентов. Data mining можно использовать для определения характеристик клиентов, которые, один раз воспользовавшись услугами данной компании, с большой долей вероятности останутся ей верными. В итоге средства, выделяемые на маркетинг, можно тратить там, где отдача больше всего.

Технологии data mining активно применяются в центрах обработки вызовов телекоммуникационных компаний.

1.3.8.4.3. Страховой бизнес

Страховые компании накапливают значительные объемы подробнейшей информации о клиентах, используемых ими услугах, страховых премиях и выплатах. Технологии data mining позволяют использовать накопившиеся данные для решения следующих задач:

Классификация и кластеризация клиентов. Система интеллектуального анализа данных позволяет страховой компании проводить эффективную тарифную политику, основанную на индивидуальных предпочтениях различных категорий клиентов.

Разработка нового товара. Технологии data mining являются инструментом, с помощью которого можно спрогнозировать спрос на услугу, оценить страховые выплаты и сформировать политику в отношении взимаемых страховых премий.

1.3.8.4.4. Производство

Большинство производственных компаний используют системы интеллектуального анализа данных для решения следующих задач.

Оптимизации логистических цепочек. Data mining позволяет снизить затраты на логистику за счет эффективного прогнозирования продаж товаров и закупок сырья/комплектующих.

Проведение маркетинговых исследований. Накопленные данные о сбыте продукции могут быть использованы при разработке новых продуктов или для повышения эффективности рекламных кампаний.

Диагностика брака на ранних стадиях. Анализ зависимостей позволяет оценить степень риска изготовления бракованного изделия на ранних стадиях производства. Очевидно, что это позволяет сэкономить существенные средства.

1.3.8.5. Автоматизированные системы для интеллектуального анализа данных

Эта тема рекомендуется для самостоятельного изучения и написания реферата, с демонстрацией презентации или демо-версии.

Резюме

Интеллектуальный анализ данных

Интеллектуальный анализ данных (data mining) – это процесс обнаружения в "сырых" данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Технологии интеллектуального анализа данных на сегодняшний день являются наиболее совершенным инструментом для решения сложных аналитических задач. Необходимо отметить, что *технологии data mining не имеют собственного уникального математического аппарата и программного инструментария, а объединяют различные математические методы и системы искусственного интеллекта.*

Системы интеллектуального анализа данных основаны на мощном математическом и статистическом аппарате, грамотное применение которого позволяет достичь высоких результатов в бизнесе.

Компания BI Partner предлагает реализацию полного спектра услуг в области интеллектуального анализа данных, что подразумевает проведение исследования накопленной статистики, выявление закономерностей, создание модели данных, ее апробация и внедрение в бизнес-процессы предприятия. ПО, которое мы используем в своих решениях, успешно используется во многих Российских компаниях.

Области применения технологий интеллектуального анализа данных.

Торговля. Анализ потребительской корзины, исследование временных шаблонов, создание прогнозирующих моделей, оптимизация складских запасов.

Банковское дело. Сегментация клиентов, выявление мошенничества с кредитными картами, прогнозирование изменения клиентуры, анализ финансовых рисков.

Страховой бизнес. Сегментация клиентов, выявление фактов мошенничества, анализ страховых рисков, разработка новых продуктов, расчет страховых премий.

Телекоммуникации. Анализ лояльности клиентов, сегментирование клиентской базы и услуг, анализ внешних факторов на отказы оборудования, выявление случаев несанкционированного доступа к сети.

Производственные предприятия. Оптимизация закупок, диагностика брака на ранних стадиях, диагностика оборудования, маркетинг.

Нефтегазовая отрасль. Диагностика оборудования и нефте/газопроводов, прогнозирование цен, разведка месторождений, анализ влияния внешних и внутренних факторов на объемы продаж.

Типы выявляемых закономерностей

Основой систем data mining является выявление различных закономерностей в данных.

Ассоциация. Если факт А является частью события, то с вероятностью Х% факт В будет частью того же события.

Последовательность. Если свершилось событие А, то с вероятностью Х% через период времени Т свершится событие В.

Классификация. На основании информации о свойствах объекта присвоение ему того или иного дискретного значения показателя, по которому проводится классификация.

Кластеризация. Кластеризация аналогична классификации, но в отличие от последней классы объектов заранее не известны.

Прогнозирование. Использование имеющихся фактических значений величин для прогнозирования будущих на основании трендов и имеющейся статистики.

Математический аппарат

Основой систем *data mining* является выявление различных закономерностей в данных: деревья решений; алгоритмы кластеризации; регрессионный анализ; нейронные сети; временные ряды.

Контрольные вопросы

1. *Интеллектуальный анализ данных (data mining).*
2. *Типы выявляемых закономерностей data mining: ассоциация, последовательность, классификация, кластеризация, прогнозирование.*
3. *Математический аппарат data mining: деревья решений, регрессионный анализ, нейронные сети, временные ряды.*
4. *Области применения технологий интеллектуального анализа данных: розничная торговля, банковская деятельность, страховой бизнес, производство, автоматизированные системы для интеллектуального анализа данных.*

Рекомендуемая литература

1. Дюк В., Самойленко А. Data Mining: учебный курс (+ CD-ROM). 2001 г. Издательство: Питер. Серия: Учебный курс. – 368 с.
2. Сайт компании BI Partner: <http://www.bipartner.ru/services/dm.html>.
3. Шапот М., Рощупкина В. Интеллектуальный анализ данных и управление процессами. // Открытые системы. – № 4-5, 1998. – С. 29.
4. Шапот М. Интеллектуальный анализ данных в системах поддержки принятия решений. Журнал "Открытые системы", #01, 1998 год // Издательство "Открытые системы" (www.osp.ru),
адрес статьи: <http://www.osp.ru/os/1998/01/30.htm>.

РАЗДЕЛ 1.4. ПРИМЕНЕНИЕ И ПЕРСПЕКТИВЫ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

1.4.1. ЛЕКЦИЯ-15. Области применения систем искусственного интеллекта

Учебные вопросы

1. *Обзор опыта применения АСК-анализа для управления и исследования социально-экономических систем.*
2. *Поддержка принятия решений по выбору агротехнологий, культур и пунктов выращивания.*
3. *Прогнозирование динамики сегмента рынка.*
4. *Анализ динамики макроэкономических состояний городов и районов на уровне субъекта Федерации в ходе экономической реформы (на примере Краснодарского края) и прогнозирование уровня безработицы (на примере Ярославской области).*

1.4.1.1. Обзор опыта применения АСК-анализа для управления и исследования социально-экономических систем

Впервые предложенная модель была практически применена автором в 1981 году для разработки оптимальной методики тестирования и диагностики специальных способностей. Имеются акты об успешном проведении исследований с помощью данной системы в области социологии, политологии и прикладной психологии по заказам НИИ культуры Российской Федерации и Краснодарского государственного института культуры (ноябрь 1987г.), Академии общественных наук под руководством доцента Б.И.Бессалаева (март 1987г.), Института социологических исследований АН СССР под руководством доктора социологии профессора А.А.Хагурова (май 1987г. и сентябрь 1987г.). В 1994 году система "Эйдос" была представлена в ВЦ РАН в присутствии иностранных специалистов (в основном из Японии) и получила **положительную оценку** председателя комитета по искусственному интеллекту Российской Академии наук академика Д.А.Поспелова (рекомендована для применения в социологических и психологических исследованиях), имеются положительные **отзывы** о Системе от Северо-Кавказской поисково-спасательной службы МЧС России (начальник В.М.Нархов) и Министерства Труда Российской Федерации (зам.нач. Управления, профессор Е.В.Белкин), прошла экспериментальную эксплуатацию в отделе информационного взаимодействия Аналитического центра при Президенте РФ (начальник отдела доктор технических наук, академик МАИ профессор

А.Н.Райков), а также **приобретена** Аналитическим центром администрации Ярославской области (руководитель А.В.Бушуев), Информационно - аналитическим центром администрации Краснодарского края (директор С.Б.Лисицын), АО "Новороссийское морское пароходство" (вице-президент, капитан 1-го ранга, к.т.н. доцент В.А.Бобыр), Краснодарским юридическим институтом МВД России (начальник генерал-майор, доктор философских наук, профессор Ю.А.Агафонов). Работы в МЧС и АО "Новошип" проводились совместно с Научно-производственной фирмой "ЭВРИТОП" (директор Б.В.Гауфман) под научным руководством кандидата технических наук, доктора биологических профессора О.А.Засухиной. Под ее же руководством Кубанским Агроуниверситетом совместно с НПП "Эйдос", на основе эмпирических исходных данных, представленных профессором Н.Г.Малюгой, была разработана методика прогнозирования урожайности и качества сельскохозяйственных культур в зависимости от вида почв, культуры-предшественника, применяемой агротехнологии (нормы посева, удобрения, вспашка и т.п.). Данная методика позволяет "просматривать" различные варианты технологии, прогнозировать последствия их применения, и, на этой основе, вырабатывать научно обоснованные рекомендации по выбору возделываемой культуры и оптимальной для поставленных целей агротехнологии. Это исследование показательно как само по себе, так и в качестве примера применения Системы для выявления взаимосвязей между различными технологиями и хозяйственными результатами. Материалы этих и других работ были представлены на международной конференции "Партнерство во имя прогресса", проходившей в конце мая 1994 года в г.Анапе. Совместно с канд.мед.наук. Г.А.Авакимяном на базе системы "Эйдос" была разработана методика диагностики избыточности и недостаточности меридианов, основанная на внешней симптоматике и анамнезе. Для строительной фирмы "ЭКСТЕРН" (директор В.Г.Очередько) была разработана методика прогнозирования профессиональной пригодности сотрудников для работы по должностям, имеющимся в фирме. Кроме того прогнозируется возможность криминогенных проявлений. *Данная методика является автоматизированным рабочим местом (АРМом) менеджера по персоналу и эксплуатировалась несколько лет в адаптивном режиме, обеспечивающем постоянное повышение качества прогнозирования на основе роста представительности обучающей выборки.*

Опыт применения АСК-анализа подтверждает эффективность его использования в следующих предметных областях: прогнозирование результатов применения агротехнологий и управление выращиванием сельскохозяйственных культур (Кубанский государственный аграрный университет); идентификация и анализ макроэкономических состояний городов и районов Краснодарского края и края в целом на основе оценки значимости индикаторов макроэкономики и применения методов статистической теории информации и искусственного интеллекта (Департамент экономики и прогнозирования администрации Краснодарского края); прогнозирование уровня безработицы (Аналитический центр администрации Ярославской области); социологические и политологические исследования, анализ общественного мнения (Краснодарский сектор Института социологических исследований АН СССР, Краснодарский краевой агропромышленный комитет, Академия государственной службы при Президенте РФ,

Краснодарский государственный институт культуры, Информационно-аналитический центр администрации Краснодарского края, Аналитический центр администрации Ярославской области и др.); решение ряда задач управления качеством подготовки специалистов (Краснодарский юридический институт МВД РФ); прогнозирование успешности будущей профессиональной деятельности в экстремальных ситуациях в условиях высокой ответственности и дефицита времени (Северокавказская региональная поисково-спасательная служба Государственного комитета РФ по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий МЧС, АО "Новошип" Новороссийское морское пароходство); решение задач Центра Правительственной связи в Краснодарском крае; проектирование адаптивных АСУ сложными техническими объектами (Государственное предприятие особое конструкторское бюро "Икар", Кубанский государственный технологический университет), исследование устойчивости управления АПК (профессор Крохмаль В.В., 2004, <http://ej.kubagro.ru/2004/01/15/> <http://ej.kubagro.ru/2004/01/09/>), исследование влияния объемов и направленности инвестиций на качество уровня жизни населения региона (Профессор Ткачев А.Н., 2004), управление качеством грунтовых вод на рисовых оросительных системах (Сафронова Т.И., 2004). Это подтверждается 17 актами внедрения предлагаемых технологий (<http://lc.kubagro.ru/aidos/aidos02/pr-4/pr-4..htm>) и рядом публикаций за период с 1987 по 2005 годы [33, 35, 36, 74, 77, 78, 79, 80, 81, 82, 86, 88, 89, 93, 96, 97, 101, 102, 113, 114, 116, 117, 118, 121, 122, 125, 130, 131, 132, 135, 143, 148, 150, 151, 152, 154, 155, 187, 188, 192, 214-217, 219, 221, 222, 223, 224, 227-231, 247-252, 268, 269-271].

1.4.1.1.1. Синтез и решение задач управления качеством подготовки специалистов

Данное исследование проведено совместно с В.Н.Лаптевым и В.Г.Третьяком [67, 81, 148-151, 155, 231, 252]. Наиболее подробно технология проведения работ и полученные результаты приведены в работе [81], в которой есть специальный большой раздел, посвященный этому исследованию, а также приложение 5, которое содержит видеогаммы интерфейса системы "Эйдос", полученные на основе данного приложения. На основе данных по абитуриентам и учащимся Краснодарского юридического института МВД РФ (КЮИ МВД РФ) за 1995 – 2002 годы (7-летний лонгитюд) с помощью системы "Эйдос" был осуществлен синтез семантической информационной модели, отражающей информационные взаимосвязи между индивидуальными личностными особенностями учащихся и их учебными и профессиональными достижениями. *Размерность модели составила: 318 прогнозируемых состояний учащихся и выпускников КЮИ МВД РФ, 129 градаций факторов, 69 прецедентов в обучающей выборке, 76128 фактов.* С помощью модели решены следующие задачи прогнозирования, входящие в состав рефлексивной АСУ качеством подготовки специалистов:

- 1) поступление в вуз;
- 2) успешность обучения в вузе по различным дисциплинам и циклам дисциплин;
- 3) успешность окончания вуза;

- 4) причины отчисления;
- 5) успешность профессиональной деятельности после окончания вуза;
- 6) продолжительность профессиональной деятельности по специальности, полученной в вузе;
- 7) причины ухода из ОВД (средневзвешенная достоверность прогнозирования составила 83%).

Система "Эйдос", как специальное программное средство (инструментарий интеллектуальной обработки информации), продемонстрировала возможность решения задач выбора оптимальной педагогической технологии для перевода учащегося из текущего состояния в заданное целевое состояние. Технология применения системы "Эйдос", разработанная автором для КЮИ МВД РФ, позволяет заблаговременно принимать обоснованные решения о целесообразности обучения конкретных курсантов по тем или иным специальностям. Перспективно применение данного инструментария и технологии в адаптивном режиме на систематической основе на выборках значительно большего и постоянно увеличивающегося объема по широкому спектру специальностей, в том числе и на межвузовском и межведомственном уровне. Развитие данного направления исследований и разработок целесообразно осуществить за счет включения в модель данных, характеризующих динамику личностных свойств учащихся, т.е. путем учета данных, полученных в процессе обучения, а не только перед его началом (как в проведенном исследовании), а также путем более детального учета влияния педагогических технологий, учебной, бытовой и профессиональной среды.

Описанная технология АСК-анализа может быть успешно применена для решения задач профотбора, подготовки и переподготовки кадров среднего и высшего звена управления АПК в вузах сельскохозяйственного профиля.

1.4.1.1.2. Экономические исследования

Все эти исследования проводятся примерно по одной технологии: разрабатывается и тиражируется опросный лист; проводится массовый опрос общественного мнения; данные опроса вводятся в компьютер и обрабатываются с помощью специальных программных систем; результаты компьютерной обработки анализируются, разрабатывается аналитический отчет, содержащий как описание ситуации, так и рекомендации по управлению.

Различие состоит в том, к какой сфере относится информация, собираемая с помощью опросного листа:

- в случае маркетинговых исследований – это потребительские предпочтения различных групп населения (имеется успешный опыт проведения подобных исследований по закрытой тематике);
- в рекламных исследованиях нас интересует, из каких источников информации различные группы населения получают информацию о приобретаемых ими товарах, а также то, каким образом они реагируют на различные виды и методы рекламы (имеется успешный опыт проведения подобных исследований по закрытой тематике);
- при социально–психологических исследованиях для нас важны психологические качества личности и точка зрения людей на те или иные преобразова-

ния (имеется успешный опыт проведения подобных исследований по закрытой тематике).

Различают разовые (точечные) и периодические исследования. Первые дают возможность выявить картину на момент проведения опроса, тогда как периодические – обеспечивают изучение динамики процессов. Вторые со временем дают возможность прогнозирования развития ситуации.

Исследования могут иметь также различный "территориальный охват", т.е. могут ограничиваться лишь одним населенным пунктом или включать сбор информации по району или краю. Во втором случае имеется возможность выявления территориальной специфики, а также возможность выявления регионов, сходных в том или ином отношении.

Использование информации, получаемой в результате маркетинговых, рекламных и социально–психологических исследований, позволяет принимать точные и эффективные решения в области планирования закупок и продвижения на рынок (адресной реализации) различных товаров, а также управлять кадрами, обеспечивая сплоченную, согласованную и надежную работу коллектива.

6.1.3.1. Маркетинговые исследования

Маркетинговые исследования проводятся в целях получения полной и достоверной информации, необходимой для определения объема и номенклатуры закупаемых товаров и организации их адресной реализации конкретным группам потребителей.

Опросный лист разрабатывается таким образом, чтобы с помощью него можно было получить необходимую информацию о потребителях. В опросный лист могут входить фирменные товарные знаки и каталоги товаров как изготовленные типографским способом, так и с изображения товаров, снятые с их фотографий при помощи сканера и отпечатанные на цветном принтере. Опросный лист может служить одновременно и бланком для заполнения, но могут использоваться и специальные бланки для ответов (что более рационально).

Вопросы делятся на несколько групп:

- социальный и гражданский статус потребителя (пол, возраст, образование, состоит ли в браке, сколько имеет детей, форма занятости, если работает, то руководитель он или исполнитель и т.п.);
- уровень достатка в семье потребителя;
- товары, реально приобретенные потребителем за последнее время;
- товары, которые потребитель приобрел бы, если бы у него была финансовая возможность;
- частота приобретения товаров по их видам;
- какие потребительские и стоимостные характеристики товаров привлекают потребителя и какие его не устраивают;
- где потребитель реально приобретает товары: у частных лиц, в магазинах (каких), на рынках (каких);
- какие места приобретения товаров потребитель знает: наименования магазинов и рынков;
- какую роль в выборе товара играет сервис и способ обслуживания в магазине или на рынке.

В результате проведения маркетингового исследования выявляются потребительские предпочтения различных групп населения в определенном секторе рынка. Выясняется, "что, где, когда", кем и по каким причинам приобретает-ся. Например, выясняется, что данная категория населения приобретает такие-то виды товаров в таких-то местах, а приобрела бы и такие, но если бы цена на них была более приемлемая. На основе этой информации может быть разработан план закупок или производства товаров, определена потенциальная емкость рынка на данный момент и перспективу, а также спрогнозирована его динамика, товар может поставляться адресно в те торговые точки, которые пользуются популярностью у конкретных категорий потребителей. Регулярное проведение маркетинговых исследований позволяет постоянно "держать руку на пульсе" и строить свою работу используя достоверное знание предпочтений возможностей и желаний своих потребителей. Имеется положительный опыт подобных исследований (по закрытой тематике).

6.1.3.2. Рекламные исследования

Рекламные исследования проводятся в целях получения полной и достоверной информации, необходимой для адресного управления формой и содержанием информации, на основе которой конкретными группами потребителей принимаются решения о приобретении тех или иных товаров. Подобные исследования проводились автором по закрытой тематике на основе применения технологии АСК-анализа.

Опросный лист разрабатывается таким образом, чтобы с помощью него можно было выяснить, какими источниками информации реально пользуются различные категории потребителей, а также какие формы подачи информации они предпочитают и лучше воспринимают. В опросный лист могут быть включены фирменные знаки, ключевые слова и рекламные фразы, а также наименования источников информации, выполненные их фирменными стилями (которые можно снять с помощью сканера и отпечатать на цветном принтере). Опросный лист может служить одновременно и бланком для заполнения, но могут использоваться и специальные бланки для ответов.

Вопросы делятся на несколько групп:

- социальный и гражданский статус потребителя (пол, возраст, образование, состоит ли в браке, сколько имеет детей, форма занятости, если работает, то руководитель он или исполнитель и т.п.);
- наиболее популярные телевизионные программы и передачи; наиболее популярные радиопрограммы и передачи;
- отношение к рекламе, включаемой в состав радио и телепередач; наиболее популярные журналы и рубрики;
- наиболее популярные газеты и рубрики;
- отношение к внешней рекламе;
- роль форм, методов и фирменных стилей рекламы;
- иные источники и формы информации, на основе которой потребителями принимаются решения.

В результате проведения рекламного исследования выясняется, какими источниками информации пользуются различные потребительские группы насе-

ления и какие формы подачи информации для них предпочтительны. Например, выясняется, что определенная группа населения в основном слушает "Радио Рокс" и именно определенную передачу, которая идет с 18 до 19. Кроме того, известно, что эту группу населения раздражает, когда любимая передача прерывается рекламой товаров, которые представителями данной группы в принципе не могут быть куплены. Известно, также, что эта группа предпочитает юмор политике. На основе этой информации может быть разработан план использования этой передачи "Радио Рокс" для адресного сообщения данной группе населения информации о "ее товарах", причем в такой форме, которая будет воспринята этой группой эмоционально положительно. Регулярное проведение рекламных исследований позволяет строить свою работу, используя достоверное знание источников и форм информации, которыми реально пользуются различные категории потребителей. Имеется положительный опыт подобных исследований (по закрытой тематике).

1.4.1.1.3. Социологические и политологические исследования

С применением предложенной методологии и технологии может быть выполнен ряд работ в области социологии и политологии:

- определение социальной базы (структуры электората) партий, их блоков и объединений, конкретных кандидатов в разрезах по краевому центру, городам и районам края (социальный, возрастной, профессиональный, национальный, образовательный, половой и т.п.;
- состав, поддерживающих и отвергающих программные лозунги предвыборной борьбы);
- изучение социальных запросов различных групп населения и формирование на этой основе предвыборной программы блока, партии, кандидата;
- выявление объективных союзников и оппозиции, как в среде партий, блоков и кандидатов, так и среди различных групп населения;
- изучение динамики и территориальных особенностей по всем этим аспектам.

Имеется положительный опыт исследований по данной проблематике, подтвержденный рядом актов (ДСП).

В настоящее время работы, проводящиеся различными группами и организациями по этим проблемам не основаны на использовании автоматизированных систем искусственного интеллекта, позволяющих разрабатывать и использовать оптимальные методики получения социологической информации и ее анализа. Обычно применяемое для компьютерной обработки результатов анкетирования программное обеспечение позволяет получить лишь простейшие характеристики исследуемой выборки, вроде процентного распределения голосов по районам или различным группам населения.

Рассмотрим подробнее вопросы применения АСК-анализа для выявления конфликтующих коалиций в сложных организационных системах.

В теории конфликтов традиционно считается, что конфликтующие стороны известны, т.е. заданы априорно, и, таким образом, вопрос об их обнаружении и выявлении в рамках этой теории не стоит.

Однако в ряде предметных областей, таких, например, как психология общения, социально-экономические и политологические системы, которые представляют собой сложные организационные системы, перед применением методов теории конфликтов часто бывает необходимо предварительно выявить сами конфликтующие стороны и их возможные коалиции, определить содержание и источники конфликтов, и уже только после этого конкретизировать типы конфликтов и применять стандартные методы их разрешения или компромиссного согласования интересов. Более того, именно в обнаружении конфликтующих сторон, а не в дальнейшем анализе, зачастую и состоит основной смысл интеллектуальной обработки данных в этих предметных областях.

Таким образом проблема выявления конфликтующих сторон и коалиций является актуальной научно-технической проблемой, относящейся к области, непосредственно смежной с теорией конфликтов и как бы подготавливающей исходные данные для ее применения.

Авторы предлагают применить для выявления конфликтующих коалиций в сложных организационных системах методы кластерно-конструктивного анализа и теории информации, реализованные в адаптивной системе анализа и прогнозирования состояний сложных систем "Эйдос".

Теория конструкторов возникла в 1955 году (Дж.Келли) в области когнитивной психологии. Конструктором называется понятие, имеющее семантические полюса и шкалу промежуточных смысловых значений (градаций). В частности, конструктором является и система из двух наиболее сильно отличающихся кластеров со спектром объектов, занимающих между этими полюсами промежуточные положения.

Более конкретно суть предлагаемой авторами технологии состоит в следующем.

Осуществляется формализация предметной области, которая состоит в том, что каждой относительно самостоятельной компоненте системы ставится в соответствие некоторая обобщенная категория и все компоненты описываются в одной системе свойств и качеств (атрибутов).

На основе описаний компонент формируются обобщенные образы категорий, которые соответствуют конфликтующим или вступающим в коалицию сторонам. Для этой цели применяются методы теории информации, в частности апостериорная семантическая мера информации А.А.Харкевича.

Каждый из обобщенных образов представляет собой список атрибутов, расположенных в порядке убывания их характеристики для данной категории. Кроме того, каждый атрибут описывается количеством информации, которое он содержит о принадлежности обладающего данным атрибутом компонента системы к каждой из категорий.

Обобщенные образы непосредственно сравниваются между собой методами корреляционного анализа, на основе чего формируется матрица сходства, которая является непосредственной основой для расчета конструкторов и кластеров.

Важно, что в предложенной математической модели при формировании конструкторов одновременно формируются и кластеры, причем если для формирования кластеров по классическим алгоритмам необходимо произвести число

операций, равное количеству сочетаний из "n по m", то для формирования конструкторов, всего n^2 , т.е. в предложенной модели в данном случае снимается проблема комбинаторного взрыва.

Результаты кластерно-конструктивного анализа представляются в наглядной графической форме в виде семантических сетей, которые представляют собой ориентированные графы, в вершинах которых находятся обобщенные категории, а ребрами являются вектора, величина и направление которых соответствуют величине сходства или различия каждой пары категорий.

Конкретное содержание конфликтов и основа для коалиций раскрывается в форме когнитивных диаграмм, представляющих собой по сути дела графическое изображение обобщенных образов двух категорий в форме списков наиболее характерных и нехарактерных для них атрибутов с указанием в форме векторов вклада каждой пары атрибутов в сходство или различие данных категорий.

Таким образом, предложенные математические модели, основанные на теории информации, и конкретная технология интеллектуальной обработки информации, реализованная на базе адаптивной системы анализа и прогнозирования состояний сложных систем "Эйдос", обеспечивают успешное выявление конфликтующих коалиций в сложных организационных системах, в частности коллективах, социально-экономических и политологических системах.

1.4.1.2. Поддержка принятия решений по выбору агротехнологий, культур и пунктов выращивания

1.4.1.2.1. Поддержка принятия решений по выбору агротехнологий

Исследование проведено совместно с О.А.Засухиной на базе Кубанского государственного аграрного университета в 1993-1996 годах [81] на основе эмпирических исходных данных, представленных Н.Г.Малюгой. Предложенные технологии АСК-анализа реализовывались на базе системы "Эйдос". С помощью сформированной содержательной информационной модели прогнозировались результаты выращивания сельскохозяйственных культур и вырабатывались научно-обоснованные рекомендации по управлению урожайностью и качеством сельскохозяйственной продукции.

Параметры семантической информационной модели

Созданная модель включает:

- активный объект управления (сельскохозяйственную культуру);
- классы (будущие состояния объекта управления, т.е. количественные и качественные результаты выращивания);
- факторы управляющей системы (агротехнологии, т.е. нормы посева, виды и нормы внесения удобрений, методы вспашки, и т.п.);
- факторы окружающей среды (вид почв, культуры-предшественники по предшествующим годам и др.).

Размерность модели составила: 35 прогнозируемых результатов выращивания, 188 градаций факторов, 217 прецедентов в обучающей выборке, 18594 факта. На основе предложенной технологии АСК-анализа разработано конкретное приложение, обеспечивающее поддержку принятия решений по выбору оп-

тимальной агротехнологии, обеспечивающей достижение поставленной цели выращивания с учетом параметров окружающей среды.

Методика проведения работ

1. Формулирование целей методики и в соответствии с ними разработка перечня прогнозируемых, в т.ч. целевых и нежелательных хозяйственных ситуаций, т.е. результатов выращивания (например, для классификации будущих состояний, в том числе целевых, могут быть использованы "шкала качества" и "шкала количества", а также разработка шкал и градаций факторов для описания причин, которые привели к этим результатам. Предложен формализованный паспорт результатов выращивания сельхозкультуры, состоящий из трех частей: первая из них содержит шкалы и градации, позволяющие классифицировать хозяйственный результат; вторая содержит шкалы и градации, описывающие не зависящие от воли человека факторы (факторы среды, в т.ч. технологическая предыстория); третья – технологические факторы, которые можно рассматривать как средство достижения желаемых результатов.

2. Использование бумажного архива по выращиванию сельхозкультуры для заполнения формализованных паспортов выращивания и *ввода* в программную инструментальную систему "Эйдос" в качестве примеров выращивания (*обучающей выборки*).

3. Синтез семантической информационной модели, т.е. выявление (на основе предъявленных реальных примеров выращивания сельхозкультур) причинно-следственных взаимосвязей между факторами окружающей среды, применяемыми технологиями и полученными результатами. Матрица абсолютных частот, является обобщением обучающей выборки, и матрица информативностей, представляет собой семантическую информационную модель предметной области.

4. Оценка достоверности модели, т.е. проверка способности созданного приложения правильно прогнозировать хозяйственные результаты на массиве уже введенных формализованных паспортов. Проведенная проверка показала, что валидность оказалась недостаточно высокой для практического применения: на уровне 58%. Причиной этого являются артефакты (неточные и зашумленные данные), из-за которых некоторые хозяйственные ситуации оказались слабо детерминированными. *Удаление артефактов (Парето-автоматизация) привело к повышению интегральной валидности до 80%, что достаточно для практического использования методики.*

5. Формирование информационных портретов по каждому возможному результату выращивания. Информационный портрет хозяйственной ситуации представляет собой перечень технологических факторов с количественным указанием того, какое влияние оказывает каждый из них на осуществление данной ситуации. По сути дела информационный портрет хозяйственной ситуации является системой научно-обоснованных рекомендаций по выбору культуры для выращивания и применению тех или иных технологий.

6. Каждый из технологических факторов на основе приведенных примеров характеризуется тем, какое влияние он оказывает на осуществление каждой (целевой или нежелательной) хозяйственной ситуации. Эта информация содержится в семантических портретах факторов.

7. Сравнение различных хозяйственных ситуаций и формирование групп наиболее сходных из них (кластеров), а также определение кластеров, наиболее сильно отличаются друг от друга (конструктов). При этом непосредственно на основе эмпирических данных путем их анализа был выявлен конструкт "качество–количество", означающий, что для получения высокого качества и большого количества продукции необходимы совершенно противоположные и несовместимые (т.е. невозможные одновременно) почвы, предшественники и агротехнологические приемы. Вместе с тем низкое качество и небольшое количество продукции вполне совместимы.

8. Группировка технологических факторов в кластеры и конструкты. Кластерно-конструктивный анализ факторов показал, что *некоторые различные по своей природе факторы имеют сходное влияние на хозяйственные результаты. Эти сходные по влиянию факторы предложено использовать для замены друг друга в случае необходимости.* Так, например, однолетние и многолетние бобовые и ранние предшественники оказывают практически такое же влияние на результаты выращивания пшеницы, как вспашка 20-22 см, а поздние пропашные предшественники как поверхностная вспашка 8-10 см.

Таким образом, с применением математического метода и инструментария АСК-анализа решены две основные задачи:

1. Прогнозирование того, какие хозяйственные результаты наиболее вероятны (а какие практически невозможны) на данном виде почв, с данными предшественниками и при условии применения имеющихся конкретных имеющихся в распоряжении агротехнологий. Указана мера сходства прогнозируемой ситуации с каждым будущим состоянием.

2. Разработка рекомендаций по выбору управляющих воздействий, т.е. консультирование по вопросам о том, какие виды почв, предшественники и *агротехнологии* должны быть, чтобы можно было рассчитывать с определенной уверенностью на желаемый хозяйственный результат. Необходимо отметить, что эта же задача позволяет также разрабатывать и совершенствовать агротехнологии.

Система "Эйдос" позволяет оценивать степень достоверности своих прогнозов и рекомендаций по управлению, т.е. она не просто дает рекомендацию, но и количественно оценивает степень ее надежности (как в самой карточке прогнозирования или информационном портрете, так в форме по валидности методики). Кроме того, система дает характеристику влияния каждого технологического приема и рекомендации по замене желательных, но очень дорогих или не имеющихся в наличии технологических приемов, другими, более дешевыми и доступными, и, при этом, имеющими сходное влияние на хозяйственные результаты. Таким образом, данная методика позволяет "просматривать" различные варианты технологии, прогнозировать последствия применения различных технологических приемов, и на этой основе вырабатывать научно обоснованные рекомендации по выбору возделываемой культуры и оптимальной для поставленных целей агротехнологии.

Необходимо отметить, что в данном исследовании *в количественной форме* были обнаружены как уже известные закономерности по влиянию предшественников, почв, удобрений, способов вспашки и т.д. на результаты выращивания сельхозкультур, так и новые, ранее неизвестные. В качестве второго можно

привести пример, условно названный автором "бройлерным эффектом для растений". Суть его состоит в том, что применение технологий, дающих особо большие урожаи, приводит к росту зеленой массы в большей степени, чем к росту массы зерна (плодов). В результате себестоимость зерна (плодов) при применении таких технологий резко возрастает и рентабельность производства снижается.

Ограничения и недостатки предложенных модели и инструментария и пути их преодоления

Однако, в приведенном исследовании были и определенные ограничения и недостатки:

- предложенный программный инструментарий системы "Эйдос" разработан еще *до создания в операционной системы MS Windows* и нуждается в приведении пользовательского интерфейса к современным стандартам;
- в описанном приложении, разработанном на данных Кубанского государственного аграрного университета совместно с О.А.Засухиной в 1993 – 1996 годах никак не учитывались метеоданные (по причине отсутствия соответствующей исходной информации);
- данное приложение было разработано на примере единственной агрокультуры (зерновые колосовые);
- использованные объемы данных: размерность модели составила: 35 прогнозируемых результатов выращивания, 188 градаций факторов, 217 прецедентов в обучающей выборке, 18594 факта для данной области являются достаточно скромными и *позволяют лишь продемонстрировать работоспособность предлагаемых технологий*, тогда как для обеспечения реального практического использования эти объемы должны быть увеличены на несколько порядков;
- не ставилась и не решалась задача районирования сельхозкультур, и не использовалась технология геоинформационных систем;
- не ставилась и не решалась задача создания многоагентной консалтинговой системы по выращиванию сельхозкультур в Internet.

Некоторые из указанных ограничений и недостатков были преодолены в следующей разработке автора совместно с Драгавцевой И.А. и Лопатиной Л.М.: [29-36, 40, 74, 153, 186, 189], кратко описанной ниже.

1.4.1.2.2. Поддержка принятия решений по выбору культур и пунктов выращивания

Данная работа проводилась в рамках гранта "ИННО-РФФИ" № 02-04-08080 "Модифицирование и коммерциализация методов компьютерной оценки ресурсного потенциала территорий и адаптивности генотипов для оптимизации размещения и зональных технологий с.х.культур" и гранта "Инициативный научный проект" № 02-05-64233 "Разработка теории многокритериальной оценки ландшафтных и метеорологических характеристик Юга России для увеличения продуктивности плодовых культур на основе создания системы банков данных и компьютерного моделирования" совместно с И.А.Драгавцевой и Л.М.Лопатиной [29-36, 40, 74, 153, 186, 189].

Цель работы

Исторически сложившиеся традиционные ареалы выращивания плодовых культур в общем случае могут не совпадать с фактически оптимальными. Само понятие "оптимальности" нуждается в изучении и определении. Прежде всего оно является многокритериальным, т.е. включает не только экономическую *необходимость* выращивания тех или иных культур, но и *возможность* этого, обусловленную всей совокупностью факторов, характеризующих природные и технологические условия выращивания в сопоставлении с требованиями, определяемыми биологическими свойствами сортов. В данном исследовании ставится цель создать технологию, обеспечивающую определение и картографическую визуализацию зон оптимального и рискованного выращивания сельскохозяйственных культур (на примере районирования плодовых косточковых в Южно-Российском регионе).

Проблема

Основная проблема, которую необходимо решить для достижения поставленной цели, состоит в отсутствии инструментария и технологий, обеспечивающих выявление и изучение факторов, влияющих на успешность выращивания сельскохозяйственных культур.

Это связано с тем, что огромные объемы информации по условиям и результатам выращивания, накопленной за десятки лет, находятся в бумажной форме, и не преобразованы в электронную форму, допускающую автоматизированную количественную обработку. Не ясно также какие математические модели могут быть использованы для обработки этой информации, т.к. она характеризуется огромной размерностью (десятки тысяч градаций факторов и результатов выращивания), непараметричностью, неполнотой (фрагментарностью), неточностью. Кроме того, отсутствующие данные не могут быть восполнены в результате специально организованного эксперимента. Эффективность математической модели должна быть подтверждена, в соответствии с принятыми в науке стандартами, на новых данных, что требует многолетних исследований с уже существующим инструментарием. Однако разработка эффективных и работоспособных алгоритмов и структур данных, а также программной реализации математической модели, весьма трудоемка и требует высокой квалификации разработчиков. Кроме того, должна быть детально разработана технология применения созданного специального инструментария.

Традиционные пути решения

Традиционно проблема районирования сортов решалась путем обобщения фактических результатов выращивания тех или иных сортов в различных регионах за длительный исторический период. В этом подходе, однако, есть следующие недостатки. Традиционные ареалы выращивания сортов обусловлены не только объективными ландшафтно-климатическими факторами, но и традициями этносов, а также путями и исторически сложившимися зонами их расселения. Это означает, что, с одной стороны, в ряде случаев сорта возделываются там, где это не вполне оправданно по объективным условиям, а с другой стороны, наоборот, они могут не возделываться там, где для этого существуют наилучшие природные условия. Традиционный подход не позволяет на количественном уровне оценить степень соответствия объективных условий в той или иной точке для

выращивания заданного сорта, и таким образом, не соответствует критериям строгости, принятым в современной науке.

Еще в 1935 г. Всесоюзной Академией сельскохозяйственных наук бывшего Советского Союза перед Государственной системой испытания сельскохозяйственных культур была поставлена задача проведения их районирования.

Однако, эта задача не решена в полной мере до сих пор, особенно по многолетним культурам. И это понятно, т.к. государственное испытание проводит изучение плодовых в *конкретных точках*, которые привязываются чаще всего к крупным плодовым хозяйствам иногда без возможности проводить испытание всего набора культур.

Например, в Краснодарском крае нет *ни одного* сортоучастка, где абрикос давал бы более 4-х урожаев за 10 лет. Госсортосеть рекомендует к закладке садов сорта по данным эмпирического испытания, без учета и анализа их адаптивного потенциала и природно-ресурсного потенциала конкретного пункта. Эти рекомендации с аналогичными данными НИИ используются при разработке проектов закладки садов проектными организациями, например, по Краснодарскому краю ОАО "Краснодарагроспецпроект".

Недостатки традиционных путей

Отсутствие системного подхода к оценке и использованию потенциала природных ресурсов и биологического потенциала сортов. Отсутствие анализируемого разработчиками проекта массива данных в системе взаимодействия "генотип-среда". Отсутствие информации по степени адаптивного потенциала конкретных сортов к конкретным условиям среды. Отсутствие прогноза вероятности проявления лимитирующих факторов (морозов, заморозков в конкретные фазы развития растений в зимне-весенний и поздне-осенний периоды, засухи, жары в летне-осенний период).

В связи с вышеизложенным неизбежен субъективизм и низкая адекватность решений по выбору культур и сортов для посадки со стороны руководителя хозяйства и специалистов проектных организаций.

Предлагаемая концепция решения проблемы

Предлагается путь решения поставленной проблемы путем поэтапной постановки и решения следующих задач:

1. **Формализация предметной области:** проектирование иерархической информационной модели предметной области, обеспечивающей оптимальность формализованного представления фактографической информации. Выбор исследуемой системы факторов на основе системного подхода (основное требование: полнота и всесторонность описания объекта исследования).

2. **Организация источников информации:** разработка формализованного паспорта для подготовки исходной информации, организация сбора и поступления информации для ввода в автоматизированную систему. Необходимо отметить, что получить исходную информацию из имеющихся в ведомственных организациях архивов практически не представляется возможным без поддержки данных исследований высших руководством этих организаций.

3. **Мониторинг:** накопление в электронной форме информации по условиям и результатам выращивания (прежде всего с использованием верифицированной ретроспективной информации).

4. **Анализ:** изучение силы и направленности влияния факторов; оптимизация, т.е. устранение избыточности системы факторов.

5. **Прогнозирование:**

– пространственная интерполяция метеопараметров для заданного пункта по их значениям в трех ближайших метеостанциях (с учетом высоты места и наклона поверхности);

– определение начала и окончания фаз для различных сортов в заданном пункте выращивания на основе ретроспективных данных;

– прогнозирование количественных и качественных результатов выращивания различных сортов в заданном пункте выращивания.

6. **Управление:** разработка научно-обоснованных рекомендаций по выбору:

– оптимальных пунктов для выращивания заданного сорта;

– оптимального сорта для выращивания в заданных пунктах.

7. **Картографическая визуализация** результатов прогнозирования и рекомендаций по управлению, получение отчетов и графических выходных форм.

При решении поставленных задач необходимо разработать математическую модель, алгоритмы и структуры данных, программный инструментарий и технологию его применения, обеспечивающие накопление баз данных о фактических результатах выращивания (мониторинг), анализ этих данных с целью выявления причинно-следственных зависимостей между свойствами сортов, условиями и результатами выращивания, прогнозирование результатов выращивания заданного сорта для любой точки на карте, картографическую визуализацию результатов прогнозирования.

Математическая модель

Математическая модель, обеспечивающая решение поставленной проблемы, относится к классу моделей распознавания образов и принятия решений, т.к. по системе факторов, характеризующих место выращивания (прежде всего характеристика почв и метеоусловий) определяются оптимальные сорта для выращивания или принимается решение о выборе оптимальных сортов для выращивания в данном пункте. (*Распознавание* есть не что иное, как *принятие решения* о принадлежности распознаваемого объекта или его состояния к определенному классу).

Планируется применить математические модели, основанные на методах распознавания образов и принятия решений, регрессионного и дискриминантного анализа, для решения следующих задач:

– оценка степени адекватности условий в нише или макронише выращивания для заданной номенклатуры плодовых культур;

– картографическая визуализация результатов решения поставленной задачи с применением технологий существующих геоинформационных систем;

– реализация этой технологии непосредственно в создаваемой системе.

Для решения указанных задач имеется определенный научный и практический задел, реализация которого позволит приблизиться к управлению продуктивностью выращиваемых культур на основе создаваемой компьютерно-информационной системы, решению задачи научно-обоснованного размещения и районирования плодовых культур и сортов.

В качестве математической модели планируется использовать непараметрическую математическую модель, предложенную в [81], основанную на системном обобщении семантической теории информации. В этой модели решено две основные научные проблемы:

1. Получено выражение для весовых коэффициентов, удовлетворяющее поставленным требованиям, основанная на системном обобщении семантической теории информации и учитывающее понятие цели, а также уровень системности и степень детерминированности системы.

2. Предложена неметрическая мера сходства объекта с классом, класса с классом, фактора с фактором, основанная на предложенном выражении для весового коэффициента и лемме Неймана-Пирсона, применение которой корректно в неортонормированном семантическом пространстве (в отличие от меры Евклида) и отражает сходство двух объектов в отличие от меры Махаланобиса.

Алгоритмы и структуры данных

Система включает две основные базы данных и большое количество вспомогательных. К основным относится база биологических данных, и банк метеоданных. Банк биологических данных состоит из двух взаимосвязанных баз данных: базы данных заголовков паспортов выращивания сельскохозяйственных культур (ссылающейся на справочники пунктов выращивания и сортов) и базы результатов выращивания сельскохозяйственных культур, состоящая из шести разделов: дифференциация, фенология, развитие, рост, адаптация, агротехника. Справочник пунктов выращивания ссылается на справочник типов почв и справочник районов, последний – на справочник регионов, а тот – стран. Справочник сортов ссылается на справочники подвоев, культур, а последний – на справочник типов культур. Справочник культур при вводе в паспорт используется для удобства с целью фильтрации справочника сортов.

Программный инструментарий

В 2001 автором совместно с И.А.Драгавцевой и Л.М.Лопатиной начата разработка и в настоящее время создана первая версия Автоматизированной системы мониторинга, анализа и прогнозирования развития сельхозкультур ("ПРОГНОЗ-АГРО"). Эта система содержит пять основных подсистем:

- 1) словари;
- 2) генерация метеобаз данных;
- 3) ввод-корректировка паспортов биологических баз данных;
- 4) расчеты выходных форм;
- 5) режим администратора системы;
- 6) информация о системе.

В свою очередь, подсистема "Словари" содержит 11 режимов, обеспечивающих ведение справочников: страны, регионы, районы, населенные пункты, метеостанции, типы почв, пункты выращивания, типы культур, культуры, подвои, сорта. Справочники взаимосвязаны друг с другом. Они организованы таким образом, чтобы минимизировать трудоемкость ввода информации и количество ошибок ввода.

Подсистема "Генерация метеобаз данных" преобразует метеобазы из одного стандарта в другой, удобный для пользователя и введения метеобаз данных. Это преобразование осуществляется в несколько этапов:

- 1) преобразование из исходного стандарта в текстовый файл;
- 2) преобразование из текстового файла в используемый стандарт, ведение метеобазы.

База метеоданных в настоящее время содержит более сотни тысяч записей о суточных метеоусловиях по десяткам метеопараметров в точках расположения метеостанций.

Подсистема "Ввод паспортов биологической базы данных" включает режим каталога паспортов и режим ввода-корректировки конкретного паспорта. Для ввода различных разделов паспорта реализованы вкладки:

- фазы дифференциации плодовых почек сортов,
- фенология, образование и развитие археспориальной ткани, ростовые характеристики;
- адаптивные свойства;
- агротехнические мероприятия.

Подсистема "Расчеты выходных форм" обеспечивает численные расчеты с использованием информации метео и биологических баз данных с привязкой к географическим координатам. Подсистема "Сервис" позволяет переиндексировать все базы данных и создать их архив. Режим администратора позволяет изменить права доступа пользователей к различным подсистемам и режимам системы.

Технология применения программного инструментария

Эта технология включает следующие этапы:

- 1) формализация предметной области;
- 2) сбор и ввод исходных данных в систему;
- 3) синтез информационной модели;
- 4) оптимизация информационной модели;
- 5) оценка адекватности информационной модели;
- 6) если модель адекватна – переход на п.7, иначе – на п.1.;
- 7) эксплуатация информационной модели для прогнозирования и управления в режиме адаптации и периодического синтеза модели.

Картографическое районирование косточковых плодовых в Южно-Российском регионе

В рамках гранта "Инициативный научный проект" № 02-05-64234 "Разработка теории многокритериальной оценки ландшафтных и метеорологических характеристик Юга России для увеличения продуктивности плодовых культур на основе создания системы банков данных и компьютерного моделирования" в Северо-Кавказском зональном НИИ садоводства и виноградарства в 1999-2003 годах совместно с И. А. Драгавцевой, И. Ю. Савиным были проведены работы по оценке потенциала выращивания различных пунктов по видам почв и метео-факторам с картографической визуализацией результатов прогнозирования в геоинформационной технологии. В результате были получены картографические отображения (с применением геоинформационной системы ILWIS) результатов оценки пригодности почв Краснодарского края для выращивания различных сортов и культур, пригодности климата, а также обобщающую двухфакторную оценку пригодности территории. При этом считалось, что если территория не подходит по почве или климату для выращивания конкретной культуры, то она

и вообще не подходит. Таким образом, использовалась не аддитивная, а мультипликативная оценка интегрального критерия, характерная для учета *лимитирующих* факторов.

Необходимо отметить, что технология АСК-анализа позволяет прогнозировать количественные и качественные результаты выращивания сельхозкультур для конкретных полей по значительно большему количеству факторов (почвы, предшественники за различные годы, метеоусловия, высота над уровнем моря, величина и ориентация наклон местности и т.п.) с использованием аддитивного интегрального критерия.

Таким образом, предлагаемые технологии АСК-анализа позволяют построить агрономическую модель Краснодарского края, в т.ч. с картографической визуализацией, в которой для каждой точки территории края будут определена степень адекватности ее условий для выращивания каждой культуры. На основе этой модели, с учетом имеющейся хозяйственной инфраструктуры или возможностей ее создания, а также прогнозируемой конъюнктуры рынка, могут быть выработаны научно-обоснованные рекомендации по объемам и номенклатуре выращиваемых сельскохозяйственных культур.

Следовательно, для принятия решений недостаточно оценки технологического потенциала пункта выращивания, а необходимо также иметь прогноз цен на продукцию. В этой связи рассмотрим применение технологии АСК-анализа для прогнозирования динамики одного из сегментов рынка.

1.4.1.3. Прогнозирование динамики сегмента рынка

Применение математического метода и инструментария АСК-анализа для прогнозирования динамики рынка продукции АПК рассмотрим на примере одного из сегментов фондового рынка Российской Федерации (рынок доллара США). Разработанные при этом технологии и подходы применимы и для других сегментов рынка при прогнозировании спроса и цен на различные виды продукции.

1.4.1.3.1. Предыстория исследования

В 1993-1994 годах, автором совместно с Б.Х.Шульман (США) были проведены исследования Российского фондового рынка [80, 81]. При этом были применены предложенные технологии и специальный программный инструментарий АСК-анализа – базовая система "Эйдос" [187] и специально созданная система окружения "Эйдос-фонд" [192]. Было создано несколько вариантов приложений для различных сегментов фондового рынка и с различной детализацией прогнозов: на каждый день на 30 дней вперед, и на каждый час рабочего дня. В дальнейшем данная разработка была усовершенствована (усилены графические возможности анализа данных), разработаны режимы исследования созданной модели и др.

1.4.1.3.2. Когнитивная структуризация и формализация предметной области

Одним из основных требований системного анализа является требование полноты и всесторонности рассмотрения, а также необходимость исключения из рассмотрения несущественной информации. При когнитивной структуризации предметной области были выявлены 300 ситуаций на фондовом рынке, которые описывались 1500 показателями. В соответствии с разработанной методикой причинно-следственные взаимосвязи между ситуациями на фондовом рынке выявлялись на основании анализа биржевых баз данных, содержащих сведения примерно за 1993-1994 годы. Позже эти базы были дополнены и в настоящее время содержат данные за 1393 дня с 1992 по 1998 годы. Таким образом, *размерность семантической информационной модели составила: 300 прогнозируемых ситуаций на фондовом рынке, 1500 градаций факторов, 1393 прецедента за 1992 по 1998 годы, 1765907 фактов.* В соответствии с разработанной методикой, путем анализа биржевых баз данных, содержащих сведения за 1393 дня с 1992 по 1998 годы, выявлялись причинно-следственные взаимосвязи между ситуациями на фондовом рынке. При этом были применены математические модели и технологии, а также специальный программный инструментарий СК-анализа: когнитивная аналитическая система "Эйдос" [187]. Реализованные в данной системе когнитивные технологии основаны на системной автоматизации 10 базовых когнитивных операций с применением системной меры семантической целесообразности информации, предложенной в работе [81].

1.4.1.3.3. Исходные данные для прогноза: биржевые базы данных

Система "Эйдос-фонд" относится к окружению системы Универсальной когнитивной аналитической системы "Эйдос" и представляет собой по сути дела программный интерфейс между биржевыми базами данных и базовой системой "Эйдос". Кроме того система "Эйдос-фонд" выполняет функции по визуализации результатов анализа. Система "Эйдос-фонд" обеспечивает прогнозирование динамики курсов валют и ценных бумаг на 1-й 2-й, 3-й, ... , 30-й день от текущего дня с достоверностью около 85%. Преобразование данных из стандартов биржевых баз данных в стандарт Системы "Эйдос" осуществляется автоматически с помощью специально для этого созданного автором программного интерфейса. При этом также осуществляется преобразование первичных параметров, т.е. чисел из временных рядов, характеризующих предметную область, во вторичные параметры – характеризующие наступление тех или иных экономических ситуаций (событий). Система "Эйдос" выявляет взаимосвязи между прошлыми и будущими событиями, и, на этой основе, позволяет осуществлять прогнозирование ситуаций. Технические решения, реализованные в данном программном интерфейсе, являются типовыми и могут быть использованы в других приложениях, где необходимо преобразование временных рядов, характеризующих динамику предметной области, в события, анализируемые системой "Эйдос". При использовании данного приложения были "вновь открыты" многие "сильнодействующие" закономерности валютного и фондового рынка, давно из-

вестные специалистам и вписывающиеся в так называемые "фундаментальные" (т.е. содержательные аналитические модели) и "технические" (т.е. феноменологические аналитические) модели. В то же время необходимо подчеркнуть, что было открыто много новых, как правило "более слабых" и специфических закономерностей валютного и фондового рынка, *характерных именно для ММВБ на момент проведения работ*. Была также обнаружена определенная динамика этих закономерностей. Подход, реализованный на базе системы "Эйдос" во многом является синтезом подходов фундаментальной и технической школ и имеет определенную новизну.

1.4.1.3.4. Формирование точечных прогнозов

На приведенных ниже диаграммах, которые выдает данная система, приводится фактический курс \$ США, а также множество точечных прогнозов и средневзвешенный прогноз динамики курса \$, рассчитанный с учетом достоверностей точечных прогнозов. Дело в том, в данном приложении на каждый конкретный день получается 30 прогнозов, первый из которых "с позиции во времени" на 30 дней назад, второй – на 29 дней назад, и т.д. Фрагмент карточки результатов прогнозирования, получающейся на основе данных, имеющихся на текущий день, приводится на рисунке 89.

РЕЗУЛЬТАТ ИДЕНТИФИКАЦИИ ИНФОРМАЦИОННОГО ИСТОЧНИКА С КЛАССАМИ РАСПОЗНАВАНИЯ				
Номер анкеты:	24	Наим.физ.источника:	11.01.94	Качество результата распознавания:
Код	Наименование класса распознавания			%у Гистограмма сходств/различий
242	На 2-й день:	РЕЗКОЕ Понижение курса \$ (3% - 5%)	53	
241	На 1-й день:	РЕЗКОЕ Понижение курса \$ (3% - 5%)	49	
276	На 6-й день:	ОЧЕНЬ РЕЗКОЕ Понижение курса \$ (более 5%)	34	
280	На 10-й день:	ОЧЕНЬ РЕЗКОЕ Понижение курса \$ (более 5%)	30	
277	На 7-й день:	ОЧЕНЬ РЕЗКОЕ Понижение курса \$ (более 5%)	23	
278	На 8-й день:	ОЧЕНЬ РЕЗКОЕ Понижение курса \$ (более 5%)	21	
283	На 13-й день:	ОЧЕНЬ РЕЗКОЕ Понижение курса \$ (более 5%)	20	
290	На 20-й день:	ОЧЕНЬ РЕЗКОЕ Понижение курса \$ (более 5%)	19	
287	На 17-й день:	ОЧЕНЬ РЕЗКОЕ Понижение курса \$ (более 5%)	18	
265	На 25-й день:	РЕЗКОЕ Понижение курса \$ (3% - 5%)	-10	
185	На 5-й день:	СЛАБОЕ Понижение курса \$ (0.5% - 1%)	-10	
298	На 28-й день:	ОЧЕНЬ РЕЗКОЕ Понижение курса \$ (более 5%)	-10	
184	На 4-й день:	СЛАБОЕ Понижение курса \$ (0.5% - 1%)	-11	
231	На 21-й день:	Понижение курса \$ (1% - 3%)	-11	
224	На 14-й день:	Понижение курса \$ (1% - 3%)	-11	
154	На 4-й день:	курс \$ БЕЗ ОСОБЫХ ИЗМЕНЕНИЙ (не более 0.5%)	-12	
230	На 20-й день:	Понижение курса \$ (1% - 3%)	-13	
221	На 11-й день:	Понижение курса \$ (1% - 3%)	-14	
228	На 18-й день:	Понижение курса \$ (1% - 3%)	-15	

Рисунок 89. Фрагмент карточки точечных прогнозов, получающихся на основе данных, имеющихся на текущий день

Любой прогноз основан на использовании ранее выявленных закономерностей в предметной области. Так как эти закономерности вообще говоря изменяются при прохождении активной системой точки бифуркации, то следует ожидать, что сразу после прохождения системой этой точки адекватность модели будет резко снижаться, а затем плавно возрастать со скоростью, которую называют "скоростью сходимости" за счет синтеза и адаптации новой модели

(фактически исследование зависимости адекватности модели от объема выборки показывает, что погрешность модели после прохождения системой точки бифуркации уменьшается по закону близкому по форме к "затухающим колебаниям"). Из этого следует, что выявление причинно-следственных связей между событиями, между которыми было несколько точек бифуркации, вряд ли имеет смысл. Поэтому увеличение объема фактографической базы для принятия решений само по себе еще не гарантирует повышения их качества. Более того, учет данных, подчиняющихся закономерностям уже потерявшим силу, вполне может и ухудшить характеристики модели. Свойство модели сохранять адекватность при прохождении точки бифуркации будем называть устойчивостью. Результаты взвешивания "точечных прогнозов" приведены на рисунке 90, зависимость погрешности средневзвешенного прогноза курса рубля от разброса точечных прогнозов на рисунке 91.

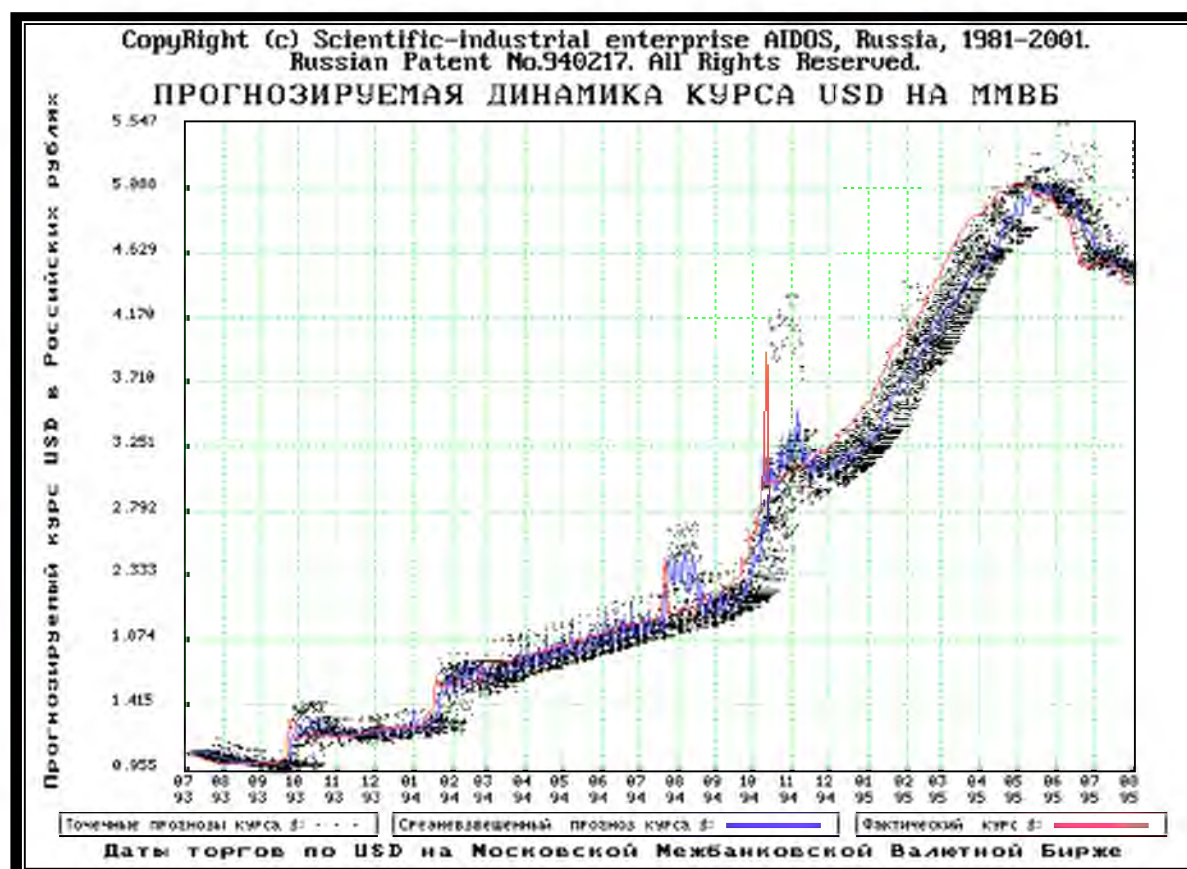


Рисунок 90. Точечные прогнозы курса Российского рубля к доллару США, средневзвешенный прогноз и фактический курс (1993-1995)

Из сравнения рисунков 90 и 91 видно, что после прохождения активной системой каждой точки бифуркации начинается колебательный процесс приближения атрибута к новому оптимальному значению, минимизирующему ошибку прогнозирования. После достижения этого состояния значение атрибута практически не изменяется до достижения системой новой точки бифуркации.

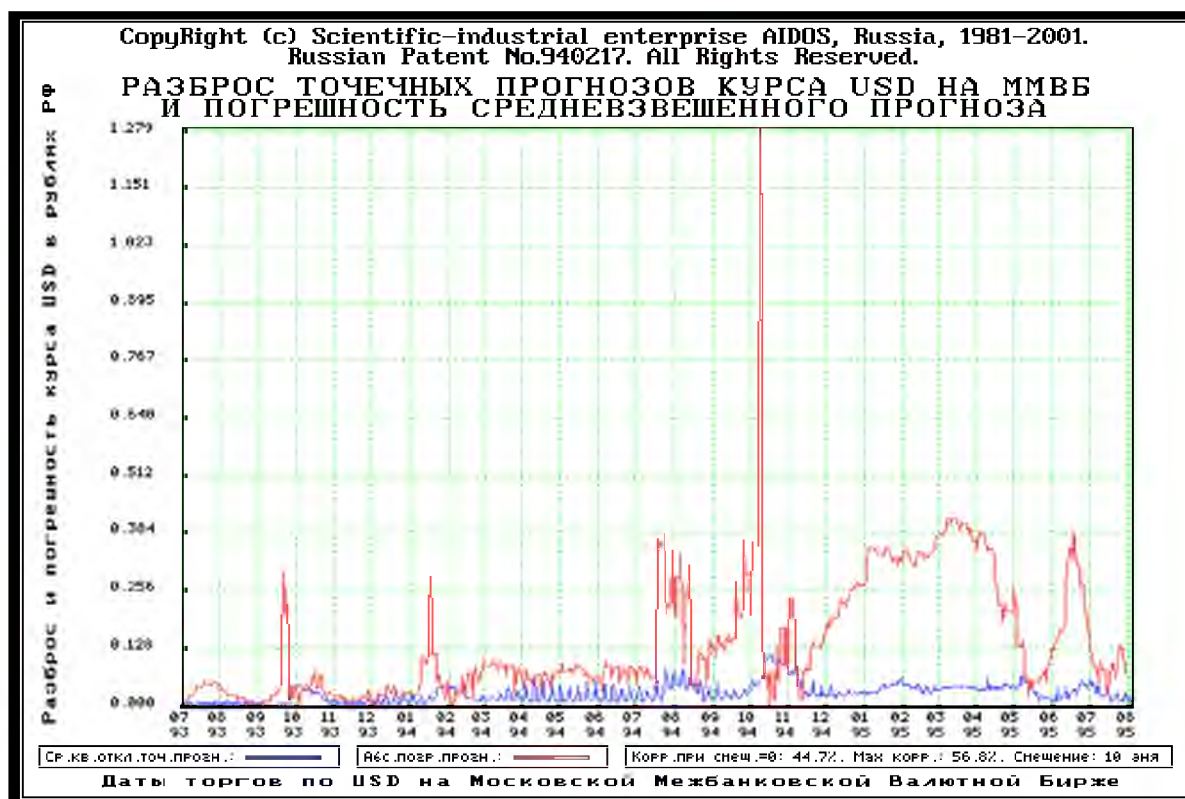


Рисунок 91. Зависимость погрешности средневзвешенного прогноза курса рубля от разброса точечных прогнозов (ММВБ, 1993-1995)

Когда разброс точечных прогнозов незначителен (рисунок 91), средневзвешенному прогнозу можно доверять, т.к. система находится на детерминистском участке своего развития, на котором ее поведение хорошо прогнозируется, т.к. закономерности, управляющие этим поведением на детерминистском этапе известны и не изменяются. Если же разброс точечных прогнозов велик, то средневзвешенному прогнозу доверять нельзя, т.к. система находится в бифуркационном состоянии, на котором ее дальнейшее поведение неопределенно, т.к. закономерности, управляющие этим будущим поведением только формируются, еще не определены и не отражены в модели.

1.4.1.3.5. Формирование средневзвешенного прогноза

Каждый точечный прогноз может быть представлен в форме вектора, параллельного оси Y (величина курса доллара США), имеющего определенную направленность, т.е. знак ("+" повышение, "-" понижение), а также величину, модуль, отражающую скорость изменения курса. Кроме того, каждый точечный прогноз имеет свою достоверность, нормированную от 0 до 1. Было принято, что вклад каждого точечного прогноза в средневзвешенный зависит не только от знака и модуля вектора, но от достоверности, которая просто умножается на модуль и уменьшает его пропорционально достоверности. Таким образом, *средневзвешенный прогноз является векторной суммой всех точечных прогнозов с учетом их достоверностей.* Так как вектора всех точечных прогнозов параллельны оси Y , то векторную сумму можно заменить их скалярной суммой в ко-

ординатной форме, в которой направление вектора учитывается просто его знаком:

$$P_D = \sum_{i=D-30}^{D-1} \sum_{j=1}^N a_{ij} \times P_{ij} \quad (6.1)$$

где:

- i – дата, с которой сделан прогноз.
- j – внутренний номер прогноза, сделанного с i -й даты.
- N – количество прогнозов, сделанных с i -й даты на дату D .
- P_D – средневзвешенный прогноз курса доллара США на дату D .
- P_{ij} – модуль и знак j -го точечного прогноза курса доллара США, сделанного с i -й даты на дату D .
- a_{ij} – оценка достоверности j -го точечного прогноза курса доллара США, сделанного с i -й даты на дату D (формируется системой "Эйдос" автоматически на основе данных, представленных системой окружения "Эйдос-фонд")

Необходимо отметить, что учитываются только те точечные прогнозы, которые имеют положительное сходство с ситуацией, реально сложившейся на фондовом рынке. В результате средневзвешенный прогноз показывает, что "скорее всего произойдет на фондовом рынке". Прогнозы, имеющие отрицательное сходство, также могут быть обобщены по аналогичной методике, но полученный средневзвешенный прогноз будет означать "чего скорее всего не произойдет на фондовом рынке".

1.4.1.3.6. Оценка адекватности модели. Зависимость достоверности прогнозирования от разброса точечных прогнозов

Адекватность модели определяется несколькими способами: путем численного эксперимента, т.е. ретроспективного прогнозирования по данным обучающей выборки (внутренняя валидность); путем экспериментально сопоставления прогнозируемого и фактического развития активного объекта (внешняя валидность). И внутренняя, и внешняя валидность может определяться в разрезе по классам (дифференциальная) или как средневзвешенная по всем классам (интегральная валидность).

При исследовании системно-когнитивной модели было обнаружено, что ошибка прогнозирования курса Российского рубля по отношению к доллару США зависит от прогнозируемой ситуации, т.е. дифференциальная валидность существенно отличается от интегральной.

Прохождение системой точек бифуркации изменяет значения атрибута, но не изменяет его *смысла* по отношению к рассмотренным классам, т.е. не изменяет порядка факторов в семантическом портрете атрибута, поэтому данный вид устойчивости предлагается называть "Семантическая устойчивость". Из рисунков 90 и 91 видно, что погрешность прогнозирования (т.е. отклонение фактического курса от средневзвешенного) и разброс точечных прогнозов резко возрастают синхронно с "обвалами" рубля. При искусственном (волевом) "удержании" курса рубля, которое обычно следует за периодами его "обвала", прогноз сильно

отличается от фактического курса. Это можно объяснить тем, что фактически прогнозируется рыночный, а не искусственно установленный курс.

Таким образом, *разброс точечных прогнозов является количественным измерителем степени неопределенности состояния системы и позволяет классифицировать это состояние как "детерминистское" или "бифуркационное" или оценить степень близости к этим состояниям.*

Когда разброс точечных прогнозов незначителен, средневзвешенному прогнозу можно доверять, т.к. система находится на детерминистском участке своего развития, на котором ее поведение хорошо прогнозируется, т.к. закономерности, управляющие этим поведением на детерминистском этапе известны и не изменяются. Если же разброс точечных прогнозов велик, то средневзвешенному прогнозу доверять нельзя, т.к. система находится в бифуркационном состоянии, на котором ее дальнейшее поведение неопределенно, т.к. закономерности, управляющие этим будущим поведением только формируются и еще не определены.

Из сравнения рисунков 90 и 91 видно, что погрешность прогнозирования и разброс точечных прогнозов сильно коррелируют. Таким образом, разброс точечных прогнозов может быть использован как количественный измеритель степени неопределенности состояния системы и позволяет оценить степень близости этого состояния к "детерминистскому" или "бифуркационному". А так как разброс точечных прогнозов может быть измерен за долго до наступления прогнозируемого состояния системы, то это позволяет прогнозировать переход системы в бифуркационное состояние.

Здесь необходимо особо отметить, что в предложенной математической модели и технологии АСК-анализа разброс точечных прогнозов не увеличивается пропорционально увеличению длительности периода прогнозирования, как доверительный интервал в статистических моделях, а является именно функцией степени объективной неопределенности состояния системы и изменяется сложным образом.

Кроме того, в системе "Эйдос" реализован режим автоматического удаления из модели классов, по которым в сформированной модели оказалась низкая достоверность идентификации и прогнозирования, после чего адекватность модели резко возрастают. Этот режим аналогичен использованию для этих целей доверительных интервалов в процедуре предсказания с помощью регрессионного анализа.

Таким образом, развитые методика, технология и программный инструментарий АСК-анализа позволяют либо надежно прогнозировать развитие активного объекта, либо надежно прогнозировать его переход в бифуркационное состояние, что само по себе также чрезвычайно ценно.

1.4.1.3.7. Детерминистские и бифуркационные участки траектории

При исследовании динамики среднеквадратичного отклонения точечных прогнозов от средневзвешенного (разброс точечных прогнозов от среднего) было обнаружено, что эта характеристика позволяет выделить два основных типа периодов, которые сменяют друг друга: относительно длительные периоды, характеризующиеся "высокой кучностью точечных прогнозов". Это детерминист-

ские периоды, для которых характерна высокая точность прогнозов, высокое совпадение средневзвешенного прогноза с фактическим курсом доллара США (на уровне 90-95%); относительно короткие периоды, характеризующиеся высокой степенью разброса точечных прогнозов. Это бифуркационные периоды, для которых характерна низкая точность прогнозов, относительно слабое совпадение средневзвешенного прогноза с фактическим курсом доллара США (на уровне 50-60%). Необходимо специально отметить, что сам факт наступления бифуркационного периода прогнозируется по величине разброса точечных прогнозов с очень высокой достоверностью (около 95%), а что произойдет конкретно в бифуркационный период с курсом доллара США и рубля спрогнозировать вообще не представляет проблемы, т.к. всегда происходило одно и то же: во все эти "черные вторники, четверги и пятницы" происходил катастрофический курса обвал рубля по отношению к доллару.

1.4.1.3.8. Прогнозирования времени перехода системы в бифуркационное состояние

Остановимся подробнее на прогнозе перехода активного объекта в бифуркационное состояние. Прежде всего, если говорить о Российском фондовом рынке, то многолетний опыт его исследования убедительно свидетельствует, что в точках бифуркации неизменно происходит "обвал" рубля. Но в возможности прогнозирования сроков перехода системы в состояние бифуркации содержится и значительно более глубокий смысл, состоящий в том, что имеется теоретическая и практическая возможность определения сроков окончания любого дела (т.е. перехода его в иное качество) на основании сроков его начала и данных о ходе реализации. Это и прогнозирование срока окончания ВУЗа по данным о сдаче сессий, прогнозирование сроков безотказной эксплуатации различных технических систем (от мобильной энергоустановки до сложной территориально распределенной энергосистемы), "сроков жизни" различных организмов, а также экономических, общественных, военных, политических и государственных организаций, и т.п. и т.д. Примерно также по изучению участка траектории снаряда специалисты по баллистике определяют точку его вылета и предполагаемую цель.

1.4.1.3.9. Исследование семантической информационной модели

Устойчивость модели, скорость ее сходимости и повышения степени адекватности при изменении объема обучающей выборки являются важнейшими характеристиками модели и определяются ее способностью к выявлению и учету новых закономерностей в предметной области, вступивших в действие после прохождения системой точки бифуркации. Кратко рассмотрим на примере исследования фондового рынка основные параметры семантической информационной модели предметной области: способ взвешивания точечных прогнозов; ослабление влияния факторов со временем; старение информации и периоды эргодичности процессов в предметной области; время реакции системы на изменение факторов (ригидность); автоколебания системы; детерминистские и бифуркационные участки траектории; управление фондовым рынком на детерминистских участках траектории и в точках бифуркации.

Ослабление влияния факторов со временем

Разработка методики прогнозирования основывалась на предположении о том, что факторы в среднем должны оказывать тем меньшее воздействие на текущую ситуацию на фондовом рынке, чем больше прошло времени со времени окончания их действия. Поэтому в реализованной методике учитывались факторы, действующие в течение прошедшего месяца. Специальное исследование, проведенное после разработки методики, показало, что факторы, действующие примерно 30 дней до возникновения текущей ситуации на фондовом рынке, незначительно отличаются по силе влияния на фондовый рынок от факторов, действующих 10 или 20 дней назад. Это означает, что в принципе оправдано было бы взять и значительно больший предстоящий период, например 6 месяцев, в течение которого учитывалось бы действие факторов на текущую ситуацию. Для определения длины предстоящего периода необходимо проведение специального исследования с очень большим предстоящим периодом, составляющим, например, несколько лет.

Старение информации и периоды эргодичности процессов в предметной области

В то же время при исследовании зависимости валидности методики от продолжительности предстоящего периода, в течение которого учитывается действие факторов, было обнаружено, что при увеличении этого периода валидность методики сначала возрастает, а затем начинает плавно, а иногда и скачком снижаться. Рост валидности объясняется увеличением статистики базы примеров, что повышает ее представительность и адекватность. Последующее уменьшение валидности может быть объяснено тем, что закономерности в предметной области изменяются с течением времени, и, поэтому, очень старые данные основаны на иных закономерностях, чем действующие в настоящее время, и значит они уже не повышают адекватность методики, а снижают ее, по сути искажают картину.

Периоды времени, в течение которых закономерности в предметной области существенно не меняются, называются периодами эргодичности. Именно на эргодичных периодах неадаптивные АСУ сохраняют свою адекватность. Эргодичность процессов нарушается либо в результате длительного действия эволюционных изменений в предметной области, которые в конце концов приводят к ее качественному изменению, или в результате действия кратковременных революционных (качественных, скачкообразных) изменений.

Для каждой методики должны быть определены периоды эргодичности, т.к. при выходе за эти периоды необходима адаптация старой или разработка новой методики.

Автоколебания фондового рынка

При исследовании чередования детерминистских и бифуркационных этапов в развитии активных объектов методами корреляционного анализа и конечных разностей были обнаружены автоколебания средневзвешенного прогноза, разброса точечных прогнозов и ошибки средневзвешенного прогноза на фондовом рынке с периодичностью около одной недели и одного месяца (рисунок 92).

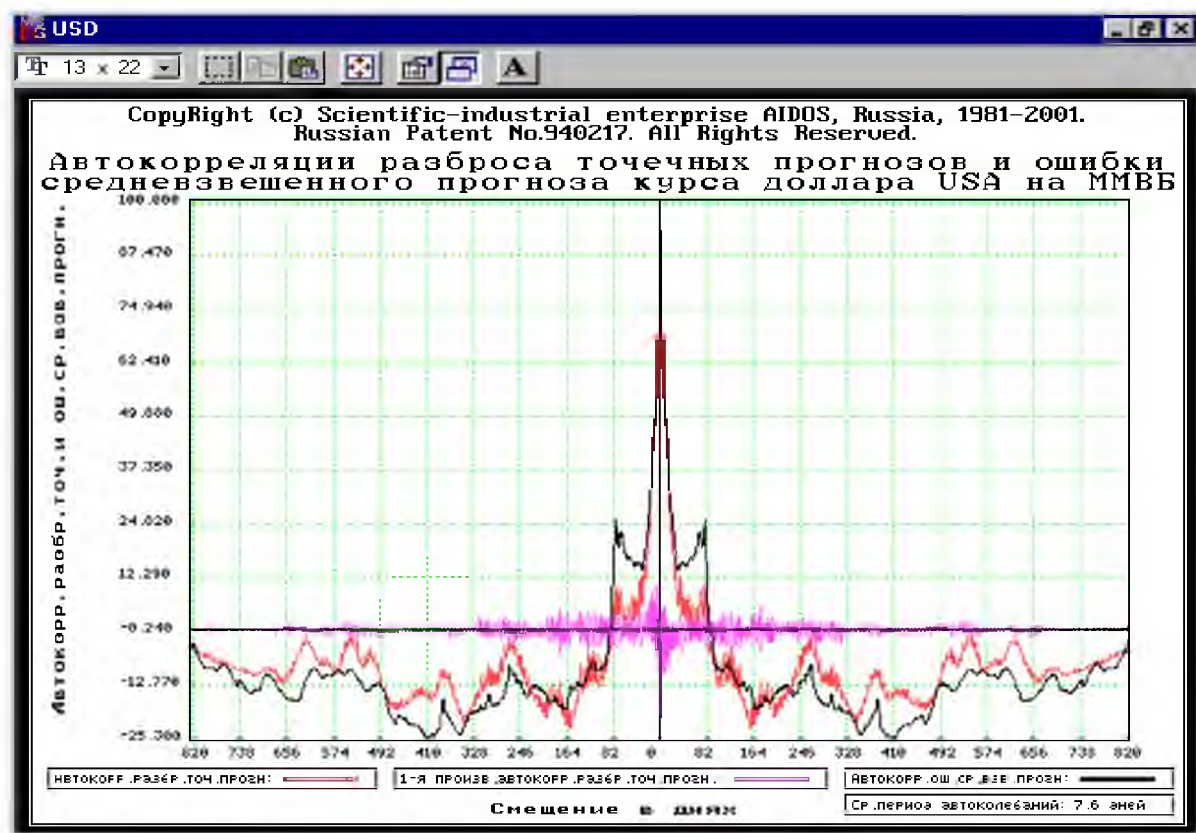


Рисунок 92. Автоколебания фондового рынка (ММВБ, 1993-1995)

В качестве варианта содержательного объяснения полученных результатов (т.е. их интерпретации) можно предположить, что именно с этой периодичностью на фондовом рынке в 1993-1995 годах осуществлялись принятие и реализация решений. Причем в субботу и воскресенье принимаются решения на следующую неделю, а в конце месяца – на следующий месяц. Таким образом, можно говорить об *уровнях бифуркаций*: для достижения цели месяца как закон реализуется некий план, для реализации которого необходимо каждую неделю принимать и реализовать решения так сказать 2-го уровня, подчиненные главному плану и цели месяца. Изучение графика средневзвешенного прогноза курса доллара США показало, что его автокорреляция имеет ярко выраженную недельную периодичность, т.е. ММВБ в 1995 году имела период собственных колебаний, составляющий одну неделю. Скорее всего это связано с отсутствием торгов в выходные дни.

Предложенные методология, технология, методика и специальный программный инструментарий обеспечивают системно-когнитивный анализ детерминистско-бифуркационной динамики активных объектов и в других предметных областях [81]. Однако необходимо подчеркнуть, что одной из предпосылок для формирования точечных прогнозов с различных точек во времени является *наличие данных мониторинга за достаточно длительный период времени*, т.е. отдельных точечных исследований для этого недостаточно. А для получения таких данных необходима соответствующая служба, использующая вполне определенные наукоемкие технологии.

Время реакции системы на изменение факторов (ригидность)

Исследование, проведенное после разработки методики, показало, что наиболее сильное влияние на текущую ситуацию оказывают факторы, действующие неделю назад, а более поздние факторы практически не оказывают на нее никакого влияния. Это означает, что в ММВБ в 1995 году имела высокую "инерционность", "время реакции", "ригидность", составляющую примерно одну неделю. По некоторым данным аналогичный параметр для Лондонской биржи в этот же период времени составлял около 32 секунд.

Управление фондовым рынком на детерминистских участках траектории

Из факторов, существенно влияющих на динамику курса доллара США лишь некоторые зависят от решений крупных финансовых негосударственных руководителей, да и то, если они будут действовать скоординировано. Прежде всего это количество банков, участвующих в торгах, а также объем первоначального спроса и предложения. На второй параметр может оказывать существенное влияние Центральный банк, путем крупных интервенций или закупок доллара США.

Управление фондовым рынком в точках бифуркации

Однако, в точках бифуркации обычные закономерности фондового рынка нарушаются или практически теряют силу, в игру вступают совсем другие факторы, которые имеют в основном не экономическую, а психологическую природу. Изучение этих факторов и разработка тактики оперативных действий требует прогнозирования динамики курса в течение суток буквально по часам. Такая работа была проведена совместно с Б.Х.Шульман (США). На первом этапе автором была разработана универсальная формальная классификация, включающая очень большое (заранее избыточное) количество вариантов суточной динамики курса доллара США, которая генерировалась автоматически. Затем было изучено влияние факторов, действующих на фондовом рынке в течение предшествующего месяца. Проведенная работа показала, что: не все теоретически-возможные варианты суточной динамики курса фактически реализуются; существует возможность надежного прогнозирования *суточной* динамики курса доллара не только в детерминистские периоды, но и в точках бифуркации. Это позволяет участникам рынка сознательно и спокойно принимать ответственные решения не только заблаговременно, но и точно привязываясь к времени в течение дня.

1.4.1.4. Анализ динамики макроэкономических состояний городов и районов на уровне субъекта Федерации в ходе экономической реформы (на примере Краснодарского края) и прогнозирование уровня безработицы (на примере Ярославской области)

В 1993 году в Департаменте экономики и прогнозирования администрации Краснодарского края по заказу заместителя Главы администрации Краснодарского края доктора экономических наук профессора Р.А.Попова с применением технологии АСК-анализа был проведен анализ макроэкономических состояний городов и районов Краснодарского края и всего края в целом в ходе экономической реформы на основе оценки значимости индикаторов макроэкономики и применения методов статистической теории информации и искусственного интеллекта. Данная работа проведена по закрытой теме и получила положительную оценку, что подтверждается актом внедрения (приложение 6).

В 1994 году в Аналитическом центре администрации Ярославской области было разработано приложение (ДСП), обеспечивающее: программный интерфейс между базами данных Аналитического центра, созданных в среде MS Works; анализ данных мониторинга, содержащих ежемесячную информацию по ряду социально-экономических показателей за 5 лет (5-летний лонгитюд) представленных Аналитическим центром, с целью выявления причинно-следственных зависимостей между ними и последующим уровнем безработицы; прогнозирование уровня безработицы на основе новых данных по социально-экономическим показателям и знания ранее выявленных причинно-следственных зависимостей. Разработанное приложение показало высокую достоверность и эффективность на ретроспективных данных и получило положительную оценку. Имеется акт внедрения (приложение 6).

Резюме

1. Продемонстрирована эффективность АСК-анализа для прогнозирования развития одного из сегментов фондового рынка РФ проведено в 1993-1998 (5-летний лонгитюд). При этом применены оригинальные авторские технологии и специальный программный инструментарий СК-анализа (система "Эйдос"), а также созданный программный интерфейс с биржевыми базами данных (система "Эйдос-фонд"). Размерность модели составила: 300 прогнозируемых ситуаций на фондовом рынке, 1500 градаций факторов, 1393 прецедента за 1992 по 1998 годы, 1765907 фактов. Обнаружено, что ошибка прогнозирования курса Российского рубля по отношению к доллару США является периодической величиной и сильно коррелирует с разбросом точечных прогнозов: чем больше разброс точечных прогнозов, тем больше ошибка средневзвешенного прогноза. Таким образом, было показано, что АСК-анализ позволяет прогнозировать как развитие активного объекта, так и его переход в бифуркационное состояние. Средневзвешенная достоверность прогнозирования составила около 87%.

2. Продемонстрирована эффективность АСК-анализа для прогнозирования результатов и управления выращиванием сельскохозяйственных культур.

Исследование проведено в 1993-1996 годах на базе Кубанского государственного аграрного университета. С помощью системы "Эйдос" была сформирована семантическая информационная модель, обеспечивающая прогнозирование результатов выращивания сельскохозяйственных культур и выработку научно-обоснованных рекомендаций по управлению урожайностью и качеством сельскохозяйственной продукции. Созданная модель включала: объект управления (сельскохозяйственную культуру); классы (будущие состояния объекта управления, т.е. количественные и качественные результаты выращивания); факторы управляющей системы (агротехнологии, т.е. нормы посева, виды и нормы внесения удобрений, методы вспашки, ротация севооборота и т.п.); факторы окружающей среды (вид почв, культуры–предшественники по предшествующим годам и др.). *Размерность модели составила: 35 прогнозируемых результатов выращивания, 188 градаций факторов, 217 прецедентов в обучающей выборке, 18594 факта.* Для каждого технологического фактора получена количественная информация о его влиянии на осуществление всех желаемых и не желаемых хозяйственных ситуаций. Кластерно-конструктивный анализ факторов показал, что *некоторые различные по своей природе факторы имеют сходное влияние на хозяйственные результаты.* Эти факторы предложено использовать для замены друг друга в случае необходимости.

Проведено количественное сравнение различных хозяйственных ситуаций и формирование групп наиболее сходных из них (кластеров), а также определены кластеры, наиболее сильно отличаются друг от друга (конструкты). *Получен конструкт "качество–количество" показывающий, что для получения высокого качества и большого количества продукции необходимы совершенно противоположные и практически несовместимые (т.е. одновременно неосуществимые) агротехнологии, почвы и предшественники.*

В данном приложении в количественной форме были обнаружены как уже известные закономерности по влиянию предшественников, почв, удобрений, способов вспашки и т.д. на результаты выращивания сельхозкультур, так и новые, ранее неизвестные. Необходимо отметить, что это сделано на современной информационно-технологической основе, что впервые позволяет количественно учесть в единой математической модели действие системы разнородных по своей природе факторов, ранее изучавшихся различными научными направлениями и дисциплинами. Система правильно оценивает степень достоверности своих прогнозов и рекомендаций по управлению. Внутренняя интегральная валидность модели составила около 83%, что достаточно для практического применения. В то же время были выявлены некоторые слабо детерминированные хозяйственные ситуации, которые наиболее сложно как вызывать, так и достоверно прогнозировать.

Были решены две основные задачи:

1) *прогнозирование* хозяйственных результатов с учетом видов почв, предшественников и применяемых агротехнологий;

2) *разработка научно-обоснованных рекомендаций по управлению выращиванием сельхозкультур*, т.е. консультирование аграриев по вопросам о том, какие виды почв, предшественники и *агротехнологии* должны быть использованы для достижения заранее заданного хозяйственного результата.

Приведенные примеры практического применения предложенной технологии и инструментария, а также успешный опыт их применения в других предметных областях, подтверждают правильность принятых научно-технических решений и их пригодность для решения проблемы, поставленной в данной работе.

5. Приведенные численные примеры успешного практического применения предложенной технологии АСК-анализа в различных предметных областях подтверждают правильность принятых научно-технических решений и их адекватность для решения проблемы, поставленной в данной работе, т.е. обеспечивают синтез семантической информационной модели и устойчивое управление активными объектами как на детерминистских, так и бифуркационных этапах их развития за счет адаптации и синтеза модели в режиме реального времени непосредственно в процессе эксплуатации РАСУ АО.

Контрольные вопросы

1. Обзор опыта применения АСК-анализа для управления и исследования социально-экономических систем.
2. Применение СИИ для синтеза и решения задач управления качеством подготовки специалистов.
3. Применение СИИ в экономических исследованиях.
4. Применение СИИ в социологических и политологических исследованиях.
5. Поддержка принятия решений по выбору агротехнологий, культур и пунктов выращивания.
6. Поддержка принятия решений по выбору агротехнологий.
7. Поддержка принятия решений по выбору культур и пунктов выращивания.
8. Применение СИИ для прогнозирования динамики сегмента рынка.
9. Когнитивная структуризация и формализация предметной области.
10. Формирование точечных и средневзвешенных прогнозов.
11. Оценка адекватности модели. Зависимость достоверности прогнозирования от разброса точечных прогнозов.
12. Детерминистские и бифуркационные участки траектории развития активных систем.
13. Прогнозирование времени перехода системы в бифуркационное состояние
14. Системно-когнитивный анализ семантической информационной модели.
15. Анализ динамики макроэкономических состояний городов и районов на уровне субъекта Федерации в ходе экономической реформы (на примере Краснодарского края) и прогнозирование уровня безработицы (на примере Ярославской области)

Рекомендуемая литература

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

1.4.2. ЛЕКЦИЯ-16.

Перспективы развития систем искусственного интеллекта, в т.ч. в Internet

Учебные вопросы

1. Ограничения АСК-анализа и обоснованное расширение области его применения на основе научной индукции.
2. Перспективы применения АСК-анализа в управлении.
3. Развитие АСК-анализа.
4. Другие перспективные области применения АСК-анализа и систем искусственного интеллекта.

1.4.2.1. Ограничения АСК-анализа и обоснованное расширение области его применения на основе научной индукции

Необходимо отметить, что ряд материалов данного раздела носят в определенной степени проблемно–дискуссионный характер.

Предложенные и развитые в данной работе математическая модель, методология и конкретная технология синтеза РАСУ АО основаны на фундаментальных положениях теории информации, распознавания образов и принятия решений, имеющих весьма общий характер. Они разрабатывались в общем виде без привязки к специфике какого–либо конкретного объекта управления. По этим причинам методы, развитые в данной работе и детализированные в главе 6 данной работы, обладают высокой универсальностью и применимы для синтеза РАСУ АО в самых различных предметных областях, в частности таких как:

- **техника** (управление сложными техническими системами); технология (управление технологиями в целях получения заданных хозяйственных и финансовых результатов);
- **психология** (разработка и применение профессиограмм, идентификация, мониторинг, прогнозирование и управление психологическими состояниями);
- **обучение** (прогнозирование успешности профессиональной деятельности, управление индивидуальным обучением, исследования влияния учебной активности на качество обучения и т.д.);
- другие применения (социология, политология, реклама, маркетинг, правоохранительная сфера и др.).

Рассмотрим перспективы применения методики и технологии АСК-анализа в некоторых из этих предметных областей более подробно. Многие из рассмотренных ниже применений технологии АСК-анализа описаны в работах автора с соавторами: [76-155, 214-217, 219-231, 247-251, 269-271], в т.ч. 5 монографиях [81, 125, 132, 221, 231] и одном учебном пособии [97].

Предложенная технология АСК-анализа продемонстрировала эффективность в различных предметных областях: это и синтез систем управления сложными техническими системами, и решение задач управления качеством подготовки специалистов; и прогнозирование ситуаций на фондовом рынке; и прогнозирование результатов выращивания сельскохозяйственных культур; и выбор оптимальных агротехнологий.

На первый взгляд эти области имеют мало общего между собой, между тем с математической точки зрения и с точки зрения технологии интеллектуальной обработки информации во всех этих областях решались одни и те же задачи. В этой связи возникает закономерный вопрос о существовании ограничений предложенной технологии и возможном корректном расширении области ее применения на другие предметные области.

Очевидно, ограничения АСК-анализа должны существовать, а значит они должны быть изучены, чтобы, с одной стороны, пользователь обоснованно пользовался этой технологией в областях ее оптимального, уверенного и рискованного применения, а с другой стороны, не возлагал необоснованных надежд на ее применение в тех предметных областях и для решения тех задач, для которых она не предназначена.

При решении вопроса о расширении области применения АСК-анализа целесообразно основываться на методе научной индукции.

В данном контексте этот метод научной индукции состоит в следующем:

1. Применить АСК-анализ для синтеза рефлексивной АСУ активными объектами на одном примере, в нашем случае – на примере управления выращиванием сельхозкультур.

2. Выявить факторы, которые являются причинами успешности, т.е. обусловили успех применения АСК-анализа в данном случае.

3. Выдвинуть гипотезу, что и в других случаях, когда сформулированные факторы (причины) успешности применения АСК-анализа действуют, т.е. имеют место, применение этой технологии будет успешным, и наоборот, при отсутствии хотя бы одного из этих факторов – на успех надеяться нет оснований.

4. Применить АСК-анализ в других предметных областях, обеспечив в обязательном порядке действие факторов успешности.

5. Проверить, успешным ли оказалось применение АСК-анализа в этих других случаях.

6. Если да, т.е. успешным, то считается, что такой результат получен за счет действия факторов успешности, и что на этом основании можно обоснованно утверждать, что и в других случаях, в которых будут присутствовать факторы успешности, можно надеяться на успешное применение АСК-анализа, и наоборот, в тех случаях, когда один или несколько факторов не соблюдаются, надеяться на успех нет оснований.

Примечание к п.6: если для какого-либо фактора успешности неизвестно, действует он или нет в данном конкретном случае, то результат применения АСК-анализа может быть положительным и отрицательным, в зависимости от того, действует ли данный фактор или нет и от того, является ли данный фактор детерминистским. В этом случае мы имеем дело с рискованным применением АСК-анализа.

Выполним эту программу.

1. АСК-анализ успешно применен для синтеза рефлексивной АСУ активными объектами: выращивание сельхозкультур.

2. Этот результат обусловлен следующими факторами:

- семантическая информационная модель адекватна предметной области;
- информация в обучающей выборке полна и достоверна:

1) достоверны тесты, применяемые для получения вторичных параметров (качественны и сами тесты и процедуры их использования)\$

2) достоверны и согласованны (объективны) независимые экспертные оценки объектов обучающей в выборки, (эксперты компетентны, процедуры сбора и обобщения экспертных оценок корректны и эффективны);

3) ввод информации в автоматизированную систему произведен качественно;

- в предметной области существуют реальные закономерности;
- обучающая выборка обладает общей и структурной репрезентативностью;
- методические и технологические этапы АСК-анализа реализованы корректно;
- специальный программный инструментарий выполняет возложенные на него функции автоматизации БКОСА.

3. Выдвигается гипотеза (т.е. предполагается), что когда обязательные факторы успешности применения АСК-анализа, приведенные в п.2 действуют, его применение будет успешным.

4. АСК-анализ реально эффективно применялся в десятках различных предметных областей с обязательным соблюдением факторов успешности.

5. Успешность применения АСК-анализа подтверждена актами внедрения (всего имеется 17 таких актов, подтверждающих успешность применения). О некоторых из этих применений информация приведена в данном разделе (ниже). При этом необходимо отметить, что примерно треть применений не документировались по ряду причин.

6. Это подтверждает гипотезу, сформулированную в п.3 и означает, что на основании принципа научной индукции *АСК-анализ должен сохранять адекватность и эффективность и в других новых предметных областях при условии действия факторов успешности.*

1.4.2.2. Перспективы применения АСК-анализа в управлении

1.4.2.2.1. Оценка рисков страхования и кредитования предприятий

Актуальность.

Основные цели организации мониторинга предприятий – прогнозирование целесообразности инвестирования в данное предприятие, оценка рисков его кредитования и страхования, выявление предприятий, попадающих в различные "зоны риска". Известны зарубежные методики оценки рисков страхования и кредитования предприятий (РСК). Однако эти методики основаны на предполо-

жении о существовании определенных экономических реалий (макро– и микро–экономических законов и закономерностей), которые в экономике переходного периода, каковой является экономика России, действуют слабо или вообще не имеют места. Поэтому зарубежные методики в наших условиях не эффективны.

Возможность адаптации зарубежных методик к нашим условиям также выглядит проблематичной, так как даже если ее и осуществить, то все равно из–за высокой динамичности нашей экономики и ее законодательно–правового аспекта эти адаптированные методики достаточно быстро потеряют свою адекватность. Возникает также вопрос о том, насколько для Южно–Российского региона, имеющего свою ярко выраженную специфику, годятся методики, адаптированные для Москвы или С.Петербурга и насколько они лучше, чем, например, Кембриджские или Стэндфордские.

Как правило, технология адаптации является весьма сложной, наукоемкой и интеллектуальной, и стоит на несколько порядков дороже самих методик или вообще не поставляется ее разработчиком.

Следовательно, весьма актуальной является разработка отечественных методик и технологий, обеспечивающих как оценку РСК, так и позволяющих разработать рекомендации по управлению предприятием, которые изменяли бы эти оценки в желательном направлении (санация).

Традиционные подходы к решению проблемы: фундаментальный и технический подход (содержательные и феноменологические модели)

Существуют два основных подхода к решению сформулированной выше проблемы:

- фундаментальный, основанный на выявлении взаимосвязей между внутренними и внешними по отношению к предприятию факторами и (событиями) и уровнем целесообразности инвестиций в данное предприятие;
- технический, основанный на анализе временных рядов различных параметров предприятия и его окружения средствами регрессионного анализа и математической статистики.

Фундаментальный подход оперирует средствами многомерного факторного анализа и содержательным аналитическим аппаратом математической экономики. Технический подход основан на статистических феноменологических моделях, отражающих внешнюю сторону явлений. Первое выглядит более обоснованным, однако наталкивается на технические трудности сбора и подготовки исходной информации. Технический подход более технологичен в плане сбора информации и ее обработки, но является более поверхностным в ее анализе и дает менее качественные прогнозы.

Предлагаемая методология и технология как синтез технического и фундаментального подхода. Общие принципы решения проблемы

В данной работе автор предлагает методологию и технологию, сочетающие глубину, содержательность и интерпретируемость фундаментального подхода с технологичностью технического подхода. Суть этой технологии состоит в том, что, *как в техническом подходе*, на основе анализа временных рядов опре-

деленных параметров, характеризующих как само предприятие, так и экономическую и правовую среду, в которой оно действует, выделяется информация о событиях в этих областях. После этого, *как в фундаментальном подходе*, анализируются причинно–следственные взаимосвязи между событиями, и на основе этого разрабатываются прогнозы и рекомендации по управлению.

В самом общем виде принцип решения сформулированной выше проблемы с помощью инструментальной программной системы состоит в следующем:

- классифицируются изучаемые типы предприятий (например, по четырем шкалам: форма собственности;
- направление деятельности; объем деятельности; оправданность инвестиций);
- классифицируются параметры как самих предприятий, так и окружения, в котором они действуют;
- выбираются и описываются предприятия, относящиеся к различным категориям;
- описанные предприятия представляются инструментальной системе в качестве образцов, т.е. конкретных реализаций, относящихся к тем или иным категориям;
- формируются обобщенные образы различных категорий предприятий и определяется ценность выбранных параметров описания для идентификации типа и прогнозирования развития предприятий и уровня РСК;
- исключаются наименее ценные параметры описания;
- выводятся в текстовом и графическом виде развернутые характеристики различных уровней РСК;
- осуществляется кластерно-конструктивный анализ сходства и различия предприятий с разными уровнями РСК;
- проводится содержательный структурно–функциональный анализ сходства и различия всевозможных предприятий с отличающимися уровнями РСК по параметрам: дается их смысловая характеристика, анализируется сходство и различие по смыслу, проводится их содержательное сравнение;
- создается инфраструктура разработки, применения и развития методик, основанных на предложенных интеллектуальных технологиях.

Подробнее эти этапы рассмотрены в работе [81].

1.4.2.2.2. Мониторинг, прогнозирование и управление в финансово-экономической сфере

Актуальность

Повторяющиеся с завидной периодичностью межбанковские кризисы заставили участников финансового рынка всерьез задуматься о необходимости серьезной организации социально–экономического мониторинга, а также аналитических исследований. Наиболее дальновидные банки и финансовые компании, в том числе и кубанские, начали создавать и оснащать необходимым инструментарием свои аналитические службы. Руководители давно поняли нехитрое правило: "Хочешь заработать сто миллионов – вложи в аналитика сто тысяч".

Предлагаемый подход

Рассмотрим службу мониторинга и прогнозирования, основная цель которой – прогнозирование и выработка рекомендаций по управлению на основе централизованного сбора, обработки и анализа разноплановой информации, необходимой для принятия управленческих решений. Для достижения этой цели служба мониторинга непрерывно ведет исследование следующих объектов управления:

- 1) ценообразование по всем секторам рынка;
- 2) ход экономической реформы по ее направлениям;
- 3) инвестиционные процессы в экономике;
- 4) кредитно-финансовая система;
- 5) предприятия различных форм собственности, направлений и объемов деятельности;
- 6) социально-психологический статус управленческого и вспомогательного персонала, а также населения. Рассмотрим эти направления подробнее.

Ценовой мониторинг

Ценовой мониторинг обеспечивает непрерывный сбор и накопление баз данных по ценам на основные товары всех секторов рынка, а также по основным финансово-экономическим показателям (факторам), влияющим на цены, что позволяет выявить и изучить зависимости между ценами на разные товары, между ценами и факторами, и на этой основе осуществлять краткосрочное и перспективное прогнозирование динамики цен.

Данная функция не дублирует ведомственный ценовой мониторинг, который уже организуется некоторыми департаментами, так как:

во-первых, акцентирует внимание на взаимосвязи динамических закономерностей цен на товары различных секторов рынка;

во-вторых, изучает влияние макроэкономических и других показателей на динамику цен;

в-третьих, держит в поле зрения лишь основные товары каждого сектора рынка, тогда как в ведомственных системах перечень товаров по соответствующему сектору рынка значительно шире;

в-четвертых, перед ведомственными службами ценового мониторинга не ставится задача сбора и обобщения информации по ценовому мониторингу различных секторов рынка в комплексе.

Мониторинг хода экономической реформы

Существует ежеквартальная отчетность регионов края по ходу экономической реформы. Однако эта отчетность слабо стандартизирована и формализована, не накапливается в компьютерных базах данных, содержит лишь абсолютные значения экономических показателей, а не информацию об экономических событиях и их динамике. В результате автоматизированный анализ этой информации с помощью интеллектуальных методов не проводится.

Предлагаемая методология и технология обеспечивают организацию этих работ, что позволяет в соответствии с установленным регламентом: получить аналитические отчеты о ходе реформы по ее направлениям (приватизация, фон-

довый рынок, земельная реформа, финансовая сфера, производство, услуги, и т.д.); выявить и изучить взаимосвязи темпа реформы в различных направлениях; выявить и изучить пути движения инвестируемых средств между различными предметными областями и направлениями реформы.

Мониторинг инвестиционных проектов

Конечно, в каждом инвестиционном проекте дается экономическое обоснование целесообразности инвестиций в выбранном направлении. Однако данное обоснование часто страдает некоторой односторонностью, так как обычно разработчики инвестиционной программы не в состоянии учесть обобщенный опыт инвестирования в ту или иную сферу в конкретных условиях рынка Кубани, так как просто не владеют необходимой информацией. Этот обобщенный опыт может быть формализован и накоплен лишь в специально созданной и уполномоченной на получение соответствующей информации структуре. Такая структура может дать независимую комплексную (а не только экономическую) оценку каждого инвестиционного проекта, т.е. сделать по нему заключение типа: "Опыт осуществления 25 подобных проектов показывает, что такого рода проекты, как правило, (в 85% случаев) в условиях Кубани приводят к тому-то и тому-то". Это заключение может быть отрицательным даже в том случае, если его экономическое обоснование является блестящим, но... проведенным по неадаптированным западным методикам, слабо учитывающим специфику России вообще и Кубани в частности.

Кредитно-финансовый мониторинг

На основе непрерывно ведущихся баз данных по динамике цен на ценные бумаги, ставкам банковских кредитов, курсам валют и т.д. позволяет выявлять и изучать закономерности кредитно-финансового и фондового рынка, и на этой основе прогнозировать его развитие, разрабатывать обоснованные рекомендации по управлению им.

Мониторинг предприятий

Основные цели организации мониторинга предприятий – прогнозирование целесообразности инвестирования в данное предприятие, оценка рисков его кредитования и страхования, выявление предприятий, попадающих в различные "зоны риска".

Для достижения этих целей необходимы следующие виды работ:

- 1) классификация предприятий, например, по трем шкалам: форма собственности; направление деятельности; объем деятельности;
- 2) организация ведения базы данных, которая содержит и накапливает в формализованном виде информацию по предприятиям, необходимую для достижения сформулированных целей;
- 3) формирование обобщенных информационных портретов предприятий различных категорий;
- 4) выдача характеристики на основании запроса в необходимых аспектах по любому предприятию, по которому имеется заполненный формализованный паспорт.

Социально–психологический мониторинг

Предлагаемая методология и технология обеспечивает сбор и обработку данных мониторинга по тем срезам социума, которые приняты как объекты постоянного контроля и управления: общественное и индивидуальное сознание; национально–этнические проблемы; культурно–религиозные проблемы; демографические проблемы; проблемы возрастных групп (молодежи, трудоспособного населения и пенсионеров); проблемы профессиональных групп; проблемы безработицы; проблемы групп населения с различным образовательным уровнем; классовые проблемы (профсоюзы, забастовки, приватизация и национализация, группы различного уровня достатка); политические ситуации; криминальные ситуации; экономические ситуации; рейтинг политических лидеров, партий и движений.

Кроме того, может быть дан психологический анализ различных типов управленческого и вспомогательного персонала, разработаны фотороботы этих типов и автоматизированные методики их идентификации.

Результаты и перспективы

Таким образом, предлагаемый подход позволяет решить следующие задачи:

- 1) разработка выводов о состоянии и динамике объекта управления на основе сбора и анализа информации;
- 2) разработка рекомендаций по способам и содержанию управляющих воздействий на объект управления;
- 3) подготовка регламентных тематических и сводных аналитических отчетов (ежедневных, недельных, месячных, квартальных и годовых отчетов), а также заказных отчетов по ранее проведенным исследованиям.

1.4.2.2.3. Некоторые перспективные области применения АСК-анализа

Технологии АСК-анализа могут быть успешно применены в ряде предметных областей:

- 1) экологический мониторинг, анализ и прогнозирование;
- 2) технология (РАСУ индивидуального управления лечением, управление неклассическими техническими системами и др.);
- 3) психология;
- 4) правоохранительная сфера.

Подробнее эти вопросы освещены в работе [81] и ряде других работ автора с соавторами: [76-155, 214-217, 219-231, 247-251, 269-271].

Некоторые перспективные направления применения АСК-анализа для управления в АПК на уровне хозяйств приведены в таблице 39.

**Таблица 39 – ПЕРСПЕКТИВНЫЕ ОБЛАСТИ ПРИМЕНЕНИЯ
АСК-АНАЛИЗА В АПК**

№	Наименование задачи	Имеющийся задел	Ожидаемые результаты
1	Прогнозирование количественных и качественных результатов выращивания сельскохозяйственных культур на основе данных о предшественниках, видах почв и применяемых агротехнологиях	Имеются апробированные математическая модель, численный метод и инструментарий. Разработана автоматизированная методика прогнозирования результатов выращивания по одной культуре (зерновые колосовые)	Повышение прибыли и рентабельности производства сельскохозяйственных культур на 3-5% за счет повышения адекватности принимаемых решений по выбору культур, полей и агротехнологий для выращивания
2	Выработка рекомендаций по выбору поля и агротехнологии для получения заданных количественных и качественных результатов выращивания сельскохозяйственных культур	Имеются апробированные математическая модель, численный метод и инструментарий. Разработана автоматизированная методика прогнозирования результатов выращивания по одной культуре (зерновые колосовые)	Повышение прибыли и рентабельности производства сельскохозяйственных культур на 5-7% за счет повышения адекватности принимаемых решений по выбору культур, полей и агротехнологий для выращивания
3	Оценка и прогнозирование уровня безработицы в хозяйстве	Имеются апробированные математическая модель, численный метод и инструментарий, опыт разработки методики прогнозирования уровня безработицы (для Ярославской области)	Принятие превентивных мер по созданию дополнительных рабочих мест
4	Управление персоналом (оценка и прогнозирование оптимальной профессиональной принадлежности)	Имеются апробированные математическая модель, численный метод и инструментарий, опыт разработки и эксплуатации ряда методик управления персоналом	Повышение эффективности использования кадров и отдачи от них, степени удовлетворенности работой, уменьшение текучести, профилактика криминогенных проявлений
5	Прогнозирование структуры и объема рынка сельскохозяйственной продукции	Имеются апробированные математическая модель, численный метод и инструментарий, опыт разработки методики прогнозирования фондового рынка	Выработка рекомендаций по реструктурированию производства с учетом прогнозируемой конъюнктуры рынка
6	Прогнозирование молочной и мясной продуктивности крупного рогатого скота	Имеются апробированные математическая модель, численный метод и инструментарий	Повышение удоев и приростов мяса за счет более рационального использования поголовья
7	Анализ качества жизни населения и выработка рекомендаций	Имеются апробированные математическая модель, численный метод и инструментарий, опыт разработки и эксплуатации подобных методик	Повышение удовлетворенности жизнью, уменьшение социальной напряженности
8	Анализ общественного мнения по актуальным вопросам жизни и развития хозяйства	Имеются апробированные математическая модель, численный метод и инструментарий, опыт разработки и эксплуатации ряда анализа общественного мнения	Знание мнения населения в разрезе по различным возрастным, половым, профессиональным и др. группам позволит руководству принимать более обоснованные решения по вопросам развития хозяйства и его социальной инфраструктуры, повысить популярность руководства

Работы, необходимые для внедрения: разработка методик по различным культурам, увеличение базы прецедентов: формализация исходных данных; ввод исходных данных в систему; синтез, оптимизация, проверка адекватности модели; эксплуатация методики в режиме адаптации и пересинтеза модели. **Основная проблема:** получение исходной информации.

1.4.2.3. Развитие АСК-анализа

1.4.2.3.1. Многоагентные распределенные системы обнаружения, накопления и использования знаний в Internet

Является перспективным создание на базе предложенной технологии и методологии АСК-анализа новой многозадачной версии системы "Эйдос", обеспечивающей (после регистрации пользователей на сайте при выполнении ими заданных условий) доступ пользователей Internet к базам данных системы как в режиме синтеза новых моделей, так и в режиме их эксплуатации в адаптивном режиме по направлениям, которые определяют сами пользователи и разработчики. Это позволит "интеллектуализировать" Internet, *создать в нем активные сервисы, способные стать со временем накопителями, хранилищами и анализаторами знаний и опыта по самым различным направлениям деятельности человека (например, автоматизированные консалтинговые службы).* Эти сервисы могли бы предоставлять услуги по интеллектуальной обработке данных пользователям независимо от их локализации, и, таким образом, стать реальной технической базой для интеграции знаний, что может стать еще одним шагом на пути развития ноосферы (В.И.Вернадский, Ле Руа, Тейяр де Шарден, 1923). В настоящее же время Internet используются, в основном, просто для получения или просмотра готовых, созданных кем-либо файлов, т.е. просто для обмена информацией.

1.4.2.3.2. Развитие АСК-анализа с применением теории нечетких множеств и неклассической логики

Весьма перспективным является развитие результатов, полученных в данном исследовании, с применением аппарата нечетких множеств Заде-Коско [81] и основанной на этом аппарате нечеткой логики (которую иногда более удачно называют непрерывной или континуальной, в отличие от дискретной бинарной Аристотелевской логики или дискретной многозначной логики).

Эта перспектива основана на том, что матрицу информативностей (таблица 16) вполне можно рассматривать как обобщенную (в смысле нечеткой логики) таблицу решений, в которой входы (факторы) и выходы (будущие состояния АОУ) связаны друг с другом не с помощью классических (Аристотелевских) импликаций, принимающих только значения: "Истина" и "Ложь", а **различными значениями истинности, выраженными в битах и принимающими значения от положительного теоретически-максимально-возможного ("Максимальная степень истинности"), до теоретически неограниченного отрицательного ("Степень ложности").**

Фактически это означает, что предложенная модель АСК-анализа позволяет осуществить синтез обобщенных таблиц решений для различных предмет-

ных областей непосредственно на основе эмпирических исходных данных и продуцировать на их основе *огромное количество* прямых и обратных правдоподобных (нечетких) логических рассуждений по неклассическим схемам с различными расчетными значениями истинности, являющимся обобщением классических импликаций.

При этом в прямых рассуждениях как предпосылки рассматриваются факторы, а как заключение – будущие состояния АОУ, а в обратных – наоборот: как предпосылки – будущие состояния АОУ, а как заключение – факторы. Степень истинности i -й предпосылки – это просто количество информации I_{ij} , содержащейся в ней о наступлении j -го будущего состояния АОУ. Если предпосылок несколько, то степень истинности наступления j -го состояния АОУ равна суммарному количеству информации, содержащемуся в них об этом. Количество информации в i -м факторе о наступлении j -го состояния АОУ, рассчитывается в соответствии с выражением (3.28) СТИ.

Прямые правдоподобные логические рассуждения позволяют прогнозировать степень достоверности наступления события по действующим факторам, а обратные – по заданному состоянию восстановить степень необходимости и степень нежелательности каждого фактора для наступления этого состояния, т.е. принимать решение по выбору управляющих воздействий на АОУ, оптимальных для перевода его в заданное целевое состояние.

Число вариантов подобных логических формул определяется по сути дела произведением числа сочетаний предпосылок на число сочетаний заключений. Однако, реально из этих формул имеет смысл использовать только *полные*, т.е. включающие все заданные предпосылки или все заданные заключения. В простейшем случае заданными могут считаться все предпосылки, или предпосылки, соответствующие факторам определенной группы, и т.д. Для развития этого направления, по-видимому, целесообразно задействовать логику предикатов.

Необходимо также отметить, что предложенная модель, основывающаяся на теории информации, обеспечивает **автоматизированное формирования системы нечетких правил по содержимому входных данных**, как и комбинация нечеткой логики Заде-Коско с нейронными сетями Кохонена. Принципиально важно, что качественное изменение модели путем добавления в нее новых классов не уменьшает достоверности распознавания уже сформированных классов. Кроме того, при сравнении распознаваемого объекта с каждым классом учитываются не только признаки, имеющиеся у объекта, но и отсутствующие у него, поэтому предложенной моделью правильно идентифицируются объекты, признаки которых образуют множества, одно из которых является подмножеством другого (как и в Неокогнитроне К.Фукушимы).

1.4.2.3.3. Динамика взаимодействующих семантических пространств и создание континуального АСК-анализа

Следующим чрезвычайно перспективным направлением развития полученных в данном исследовании результатов представляется развитие теории динамики взаимодействующих семантических пространств классов и атрибутов. С этой целью для семантических пространств и объектов в этих пространствах, т.е. векторов классов и факторов, на основе понятия "информация" предлагается

определить понятия, являющиеся аналогами понятий геометрии, кинематики и динамики: это метрика, система отсчета, скорость и ускорение, путь, масса, сила, импульс и энергия, закон (в частности законы сохранения), и т.д. Предлагается получить соответствующие аналитические выражения и дать их содержательную интерпретацию и способы численного расчета.

Высказывается гипотеза, состоящая в том, что *физическая картина мира (и не только физическая) является не более чем подмножеством некоторой информационной модели, в какой-то мере аналогичной той, которая предложена в данном исследовании*. Это позволяет подойти к исследованию природы реальности с пониманием того, что *в действительности мы исследуем не реальность, а лишь содержательные информационные модели этой реальности, основанные на нашем ограниченном опыте*. Например, если прогноз поведения системы, сделанный на основании некоторой модели, адекватен, то мы получаем информацию об адекватности модели, если же нет – то новую информацию о системе, не отраженную в модели. Вторая ситуация возможна при исследовании систем после прохождения ими точки бифуркации и систем, качественно отличающихся от описываемых моделью. Учет этой новой информации в модели повышает ее адекватность и качественно расширяет область адекватного применения модели. Таким образом, *количество информации о системе, полученное в процесс познания, можно измерять по степени модификации модели системы при учете в ней этой информации и приведении модели в адекватное состояние (аналогично, количество информации, записанное в структуре предмета труда, можно измерять по степени его модификации)*. При учете в модели незначительного количества информации происходит ее адаптация, а при учете значительного количества – синтез. Учет в теории информации эффектов системного взаимодействия альтернативных состояний (явление интерференции последствий выбора квантовых и активных объектов) приводит к обобщению понятия "информация" в рамках системной теории информации, один из вариантов которой предложен в данной работе. Подобные идеи в свое время привели к возникновению специальной теории относительности (учет свойств релятивистских свойств объектов), квантовой механики (учет квантовых свойств объектов) и нейролингвистического (НЛП) программирования. В принципе философы знали это всегда (правда не всегда признавали), но предложенный в данном исследовании подход позволяет перейти в этой области от умозрительных рассуждений к *строгим* расчетам, и это соответствует внутренней логике развития науки, которую здесь неуместно рассматривать подробнее.

Придание модели онтологического статуса (гипостазирование) – ошибка, которая часто совершается и в результате которой начинают считать, что "в действительности все устроено так, как в модели". Например, придание онтологического статуса объектам – дело решенное для материалистов, но для последователей остальных философских направлений этот вопрос остается открытым. Если объективное существование объектов мало у кого вызывает сомнение, то вопрос об объективном существовании интегративных структур 2-го и более высоких порядков остается открытым. Но что это за структуры? В предложенной когнитивной концепции, кластер, например, представляет собой своего рода "объект,

состоящий из объектов", т.е. объект 2-го порядка. Если объект выполняет интегративную функцию по отношению к атрибутам, то кластер – по отношению к объектам. Необходимо подчеркнуть, что термин "класс" используется не только для обозначения образов уникальных объектов, но и для обозначения их кластеров, т.е. *классу может соответствовать не уникальное, а обобщающее имя, в этом случае мы имеем дело с обобщенным классом*. Да и кластеры могут быть не только кластерами уникальных объектов, но и обобщенных классов.

Возникает вопрос об онтологическом статусе понятия "кластер". В некоторых философских системах подобным объектам приписывался даже более высокий статус существования, чем самим объектам, например обычные объекты рассматриваются лишь как "тени" "Эйдосов" (Платон). Известны и другие понятия для обозначения объектов высоких порядков, например "архетип" (Юм), "эгрегор" (Даниил Андреев) и др. Нельзя не отметить, что в современной физике (специальной и общей теории относительности) есть подобное понятие пространственно-временного интервала, который проявляется на практике как *движение* объекта. Являясь объектами 2-го порядка сами кластеры в результате выполнения когнитивной операции "генерация конструкторов" могут быть классифицированы по степени сходства друг с другом. Для обозначения системы двух противоположных кластеров, с "спектром" промежуточных кластеров между ними, будем использовать термин *"бинарный конструктор"*, при этом сами противоположные кластеры будем называть *"полюса бинарного конструктора"*. Таким образом конструктор представляет собой объект 3-го порядка. Словом "бинарный" определяется, что в данном случае полюсов у конструктора всего два, но этим самым подчеркивается, что в принципе их может быть 3, 4 и больше. Бинарный конструктор можно формально представить в виде порядковой шкалы или даже шкалы отношений, на которой градациям соответствуют кластеры, а значит и сами классы и соответствующие объекты. Конструкторы с количеством полюсов больше 2 могут быть представлены графически в форме семантических сетей в которых полюса являются вершинами, а дуги имеют цвет и толщину, соответствующие степени сходства-различия этих вершин. Семантические сети можно считать также просто графической формой представления результатов кластерного анализа. По-видимому, статус существования структуры реальности, отображаемой некоторой когнитивной структурой определенного уровня интегративности тем выше, чем выше этот уровень интегративности, т.е. наивысшим статусом существования обладает Мир-в-Целом.

Аналогично кластерам и конструкторам классов формируются кластеры и конструкторы атрибутов. В кластеры объединяются атрибуты, имеющие наиболее сходный смысл. В качестве полюсов конструкторов выступают кластеры атрибутов, противоположных по смыслу.

Бинарные конструкторы классов и атрибутов представляет собой когнитивные структуры, играющие огромную роль в процессах познания. Не будет преувеличением сказать, что познание представляет собой процесс генерации, совершенствования и применения конструкторов. Будем считать, что конструктор тем более совершенен и тем выше его качество, чем сильнее отличаются его полюса, т.е. чем больше диапазон его области значений.

Однако, возникает вполне закономерный вопрос о том, по какой причине в мышлении современного человека при обычных формах сознания используются лишь **бинарные** конструкты, причем, как правило, при их использовании редко кто вспоминает про нечеткую логику Заде, т.е. на практике промежуточные между полюсами значения вообще не рассматриваются. Представляется весьма перспективным исследовать и формализовать законы мышления, характерные для измененных (прежде всего, высших) форм сознания, разработать теорию конструктов с большим чем два дискретным количеством полюсов (многозначные конструкты), а также теорию конструктов с любым иррациональным количеством полюсов. Таким образом, будущим исследователям и разработчикам возможно предстоит обобщить предложенную в данном исследовании когнитивную концепцию и разработать на ее основе более общие математические модели "многозначного а затем и континуального АСК-анализа", в рамках которых предложенная модель получит статус частного случая – "бинарного АСК-анализа". Возможно в рамках континуального АСК-анализа удастся обнаружить новые базовые когнитивные операции и интегративные когнитивные структуры более высоких порядков, чем в предложенной когнитивной концепции и построить их математические модели.

1.4.2.4. Другие перспективные области применения АСК-анализа и систем искусственного интеллекта

В перспективе предложенные методология, технология и инструментальная программная система могут быть применены, в частности, в следующих областях:

В медицине, в том числе гомеопатии, иридодиагностике, рефлексотерапии: диагностики заболеваний по их симптоматике, в том числе при массовых профилактических обследованиях населения как с выездом в район обследования или на предприятие, так и без личного присутствия пациентов; сжатия диагностической информации, содержащейся в медицинской литературе и архивах историй болезни; проведения научных исследований по выявлению причинно-следственных зависимостей между клинической картиной, применяемыми лечебными средствами (например, лекарственными) и методами лечения, с одной стороны, и лечебным эффектом, с другой, мониторинга состояний пациентов.

В профессиональной диагностике и профориентации в учебных заведениях, на предприятиях в отделах кадров, в центрах занятости населения (на биржах труда и в службах трудоустройства). Для психосоциальной диагностики и анализа общественного мнения, компьютерной обработки и интерпретации результатов социологических исследований (public relations). Для тестирования предметной общности учащихся и определения решаемости контрольных заданий.

Для криминогенной профилактики физических и юридических лиц, выявления и прогнозирования "зон риска", в том числе в налоговых, финансовых и других контролирующих органах, для криминалистической и психофизиологической идентификации личности. Для классификации и типологизации преступлений, формализации фабулы (почерка), нахождения аналогов преступлений, а

также автоматизированного поиска ранее проведенных проверок, в которых были получены аналогичные результаты.

Для диагностики способностей, в том числе экстрасенсорных и других парапсихологических способностей, косвенного измерения уровня развития сознания, интеллектуальной и эмоциональной сферы, способностей к специальным видам деятельности, связанных с риском, высокой ответственностью, работой на высоте, под землей, под водой, с применением оружия, в экстремальных ситуациях и т.п. Для автоматизированного синтеза образа объекта по его фрагментам, полученным с помощью подсознательного информационного канала, в том числе при получении одной и той же информации многими людьми, ни один из которых в явной и целостной форме не осознает ее; для количественного сравнения и идентификации образов, полученных в результате дистанционной перцепции, ретрогнции и прекогнции, с образами – мишенями. Для синтеза образа объекта – мишени (и количественного его сравнения с оригиналом) при индивидуальном и коллективном восприятии по каналам телепатии, неклассической дистанционной перцепции, ретрогнции и прекогнции.

Для атрибуции анонимных и псевдонимных текстов, т.е. для установления вероятностного авторства текста или его тематической или иной принадлежности по незначительным, поврежденным, неполным и несвязанным фрагментам. В этом случае система распознавания работает как дескрипторная информационно-поисковая система с автоматическим формированием дескрипторов и поиску по нечеткому или некорректному запросу (на естественном языке, причем какой этот язык: русский, английский или какой-либо другой – роли не играет).

Для автоматической классификации химических веществ по их составу или внешним признакам. Для вероятностной идентификации элементов в смесях и при неполном или некачественном их анализе с помощью химических тестов, спектральных методов, ядерно-магнитного резонанса.

Для формирования обобщенных образов различных видов недвижимости и сопоставления конкретных объектов с этими образами в целях оценки недвижимости (развитие метода сравнительных продаж). Для разработки и применения стандарта земель в земельном кадастре.

Для автоматической классификации сортов растений и пород животных по их внешним, анатомическим, физиологическим и другим признакам, а также на основе измерения генетического расстояния.

Для прогнозирования месторождений полезных ископаемых по внешним сопутствующим признакам, в том числе и установленным с помощью биолокации.

Для синтеза образов подсознательно воспринимаемых объектов, признаки которых установлены (выведены на уровень сознания) с помощью биолокации.

Для долгосрочного и оперативного прогнозирования изменений погодных факторов, для регионов, для которых известны суточные значения этих факторов за достаточный период времени.

Для выявления влияния любых технологических приемов и условий на качество и количество хозяйственных результатов. Для анализа и прогнозирования ситуаций на сельскохозяйственном секторе натурального и фондового рынка: ценовой мониторинг и прогнозирование цен и объемов продаж продукции

растениеводства, птицеводства и животноводства, а также ценных бумаг; макроэкономический анализ.

Косвенная профессиональная идентификация, прогнозирование успешности работы и совместимости сотрудников службы безопасности, других служб банка (при этом выявляются сотрудники – "опора фирмы" и "зоны риска"). Прогнозирование остатков и движения средств на счетах филиалов банка и на счетах клиентов. Прогнозирование развития фондового рынка, других сегментов рынка. Косвенная оценка рисков страхования, кредитования, инвестирования, бизнеса, других форм взаимодействия юридических и физических лиц. Косвенная профессиональная идентификация, прогнозирование успешности работы и совместимости сотрудников различных служб фонда (при этом выявляются сотрудники – "опора фирмы", а также "зоны риска"). Косвенная профессиональная идентификация и прогнозирование успешности работы клиентов фонда по определяемым профессиям и специальностям.

Прогнозирование развития рынка труда и уровня безработицы, других сегментов рынка.

Для решения задач распознавания объектов и их состояний по признакам, в том числе и при неполном или искаженном описании. Для изучения динамики и территориальных зависимостей обобщенных образов классов распознавания. Для выявления и исследования причинно – следственных связей между событиями (признаками, технологией, составом) и их следствиями (объектами, состояниями, свойствами, эффективностью) и др.

Для аналитической обработки данных экологического мониторинга (на примере Черного моря): разработка оптимального формализованного паспорта для экологического мониторинга и экологического зонирования (районирования) побережья Черного моря; экологическая паспортизация и зонирование побережья Черного моря; разработка информационных портретов экологических зон побережья; их классификация, обобщение и сопоставительный анализ; комплексные и специальные, регламентные и инновационные аналитические исследования по данным экологического мониторинга побережья Черного моря; исследование влияния факторов различного происхождения (геофизических и биосферных; антропогенных; рекреационных и других экологических мер и технологий) на экологическое состояние побережья Черного моря; исследование динамики экологического состояния побережья Черного моря в связи с динамикой влияющих на него факторов. Оперативное и долгосрочное прогнозирование развития экологического состояния побережья Черного моря; разработка методик подбора персонала экологических служб: оценка индивидуальных качеств сотрудников; оптимальный подбор групп по совместимости; прогнозирование успешности профессиональной деятельности в различных областях; прогнозирование рисков нежелательных проявлений; косвенная оценка рисков инвестирования, страхования, кредитования, бизнеса, других форм деятельности и взаимодействия юридических и физических лиц, в том числе в области экологической деятельности; организация Информационно-аналитического центра Public Relations, для проведения систематических (регламентных) и заказных исследований по изучению и формированию мнения различных групп населения по вопросам экологии ("экологического сознания"), а также для участия в выполне-

нии работ по вышеупомянутым проектам; изучение аудитории и роли различных средств массовой информации в освещении экологической проблематики и формировании экологического сознания у различных групп населения.

Резюме

1. Предлагается рассматривать автоматизированный системный анализ, структурированный до уровня базовых когнитивных операций, как автоматизированный системно-когнитивный анализ (АСК-анализ). Приведены различные подходы к оценке экономической эффективности методологии, методов и методик АСК-анализа, при этом:

- АСК-анализ рассматривается как возможная системная основа конструирования геоинформационных технологий в АПК;

- программное обеспечение АСК-анализа рассматривается как инструментарий, обеспечивающий интеллектуальную подготовку данных для геоинформационных технологий;

- геоинформационные технологии, основанные на использовании данных АСК-анализа, позволяют повысить эффективность производства, снизить издержки и обеспечить совокупный эффект, выражающийся в повышении продуктивности и качества продукции и экономии агротехнологических ресурсов.

Таким образом, АСК-анализ представляет собой методологический базис, а реализованный в программном виде – и инструментарий геоинформационной технологии оптимального размещения породно-сортового состава плодовых культур в ландшафте с присущими ему почвенно-климатическими ограничениями, что в практике их реализации дает расчетный эффект.

2. На основе метода научной индукции обоснована возможность корректного расширения области применения АСК-анализа не только в АПК, но и в других предметных областях.

3. Описаны перспективы применения предложенных технологий АСК-анализа в следующих предметных областях: управление технологиями; экономика; психология; социология и политология; маркетинговые, рекламные, социально-психологические исследования; правоохранительная сфера; Internet (создание многоагентных распределенных систем обнаружения, накопления и использования знаний по различным предметным областям). Приведены другие перспективные области. В частности отмечаются перспективные применения для: атрибуции анонимных и псевдонимных текстов (реализовано в системе "Эйдос"); контент-анализа (реализовано в системе "Эйдос"); синтеза интеллектуальных дескрипторных информационно-поисковых систем с интерфейсом запросов на естественном языке (реализовано в системе "Эйдос"); идентификации и классификации изображений по их вербальным описаниям; формирования обобщенных фотороботов лиц по направлениям деятельности и идентификации с ними; применения в области психографологии и дерматоглифики, и т.д. Кроме того, рассматриваются ограничения предложенного варианта АСК-анализа и рассматриваются вопросы его развития (путем преодоления указанных ограничений) с применением теории нечетких множеств и неклассической логики. В частности *выдвигается идея обобщения понятия конструкта путем введения многополюсных конструктов.*

4. **Основной вывод данного раздела** состоит в том, что обоснованным выглядит предположение об универсальности предложенной методологии, специального программного инструментария, технологии и методики АСК-анализа и возможности их применения в различных предметных областях, в которых накапливаются факты и производятся их обобщение и анализ, а также выполняются другие операции, связанные с процессами познания и управления, получившие в данном исследовании аббревиатуру БКОСА.

Контрольные вопросы

1. *Ограничения АСК-анализа и обоснованное расширение области его применения на основе научной индукции.*
2. *Перспективы применения АСК-анализа в управлении.*
3. *Оценка рисков страхования и кредитования предприятий.*
4. *Мониторинг, прогнозирование и управление в финансово-экономической сфере.*
5. *Некоторые перспективные области применения АСК-анализа.*
6. *Развитие АСК-анализа.*
7. *Многоагентные распределенные системы обнаружения, накопления и использования знаний в Internet.*
8. *Развитие АСК-анализа с применением теории нечетких множеств и неклассической логики.*
9. *Динамика взаимодействующих семантических пространств и создание континуального АСК-анализа.*
10. *Перспективные области применения АСК-анализа и систем искусственного интеллекта.*

Рекомендуемая литература

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

ЧАСТЬ II. ЛАБОРАТОРНЫЙ ПРАКТИКУМ

2.1. СООТВЕТСТВИЕ ЛАБОРАТОРНЫХ РАБОТ РАБОЧИМ ПРОГРАММАМ ПО СПЕЦИАЛЬНОСТЯМ

В соответствии с рабочей программой при обучении по специальности: 220200 – автоматизированные системы обработки информации и управления, рекомендуется 2 варианта (набора) лабораторных работ, приведенные в таблицах 40 и 41.

**Таблица 40 – ЛАБОРАТОРНЫЕ РАБОТЫ
(вариант 1)**

№ п/п	Наименование	Ча- сов
1	ЛР-1: "Прогнозирование вероятных пунктов назначения железнодорожных составов"	8
2	ЛР-2: "Прогнозирование учебных достижений студентов на основе их имеджевых фотороботов"	4
3	ЛР-5: "Идентификация слов по входящим в них буквам"	4
4	ЛР-7: "Идентификация и классификация натуральных чисел по их свойствам"	4
5	ЛР-8: "Идентификация трехмерных тел по их ортогональным проекциям"	4
6	ЛР-9: "Прогнозирование количественных и качественных результатов выращивания зерновых колосовых и поддержка принятия решений по выбору агротехнологий"	4
7	ЛР-10: "Исследование случайной семантической информационной модели при различных объемах выборки"	4
	Всего:	32

**Таблица 41 – ЛАБОРАТОРНЫЕ РАБОТЫ
(вариант 2)**

№ п/п	Наименование	Ча- сов
1	ЛР-1: "Прогнозирование вероятных пунктов назначения железнодорожных составов"	8
2	ЛР-2: "Прогнозирование учебных достижений студентов на основе их имеджевых фотороботов"	4
3	ЛР-3: "Прогнозирование учебных достижений студентов на основе особенностей их почерка"	4
4	ЛР-4: "Прогнозирование учебных достижений студентов на основе информации об их социальном статусе"	4
5	ЛР-6: "Атрибуция анонимных и псевдонимных текстов"	4
6	ЛР-9: "Прогнозирование количественных и качественных результатов выращивания зерновых колосовых и поддержка принятия решений по выбору агротехнологий"	4
7	ЛР-10: "Исследование случайной семантической информационной модели при различных объемах выборки"	4
	Всего:	32

На первую лабораторную работу отводится 8 часов, т.к. на ней у учащихся будут формироваться первичные умения по применению универсальной когнитивной аналитической системы "Эйдос", которые на последующих лабораторных работах должны быть развиты и доведены до уровня навыков.

По технологии выполнения лабораторные работы имеют много общего, не смотря на различное содержание заданий. Это позволяет освоить эту технологию в общем виде на таком уровне, который в принципе позволяет учащимся самостоятельно осуществить формальную постановку и решение прикладной задачи применения систем искусственного интеллекта, независимо от предметной области. Кроме того, при такой структуре занятий материал лучше усваивается теми учащимися, которые пропускали занятия, что является реальностью, которую также необходимо учитывать.

2.2. СТРУКТУРА ЛАБОРАТОРНОЙ РАБОТЫ, ПОРЯДОК ЕЕ УСТАНОВКИ, ВЫПОЛНЕНИЯ И СДАЧИ

2.2.1. Структура лабораторной работы

Лабораторные работы имеют типовую структуру, включающую следующие разделы:

- краткая теория;
- одно или несколько заданий;
- пример решения;
- контрольные вопросы;
- литература по данной лабораторной работе.

В разделе "**Краткая теория**" излагается минимум теоретических понятий, необходимых для осмысленного выполнения студентом данной работы.

В разделе "**Задание**" ставится цель лабораторной работы и формулируются этапы ее достижения.

В разделе "**Пример решения**" приводятся примеры выполнения **некоторых** этапов сформулированных заданий. Примеры выполнения служат для пояснения наиболее сложных этапов выполнения работы и не должны тождественно повторяться студентами в своих лабораторных работах.

Контрольные вопросы по лабораторной работе служат для проверки качества усвоения и понимания материала, могут быть заданы преподавателем при защите лабораторной работы и включены в экзаменационные билеты.

В разделе "**Литература**" приводятся конкретные литературные источники, использованные при разработке данной лабораторной работы.

2.2.2. Инструкция по установке лабораторной работы

Предварительно дадим некоторые пояснения. На учебных компьютерах, на которых проводятся лабораторные работы, на запись доступна только директория **C:\WORK** (или **C:\STUDENT**, далее будем упоминать только директорию **C:\WORK**). Система "Эйдос" и базы данных, подготовленные для проведения

лабораторных работ, находятся не в этой директории, а в директории **C:\AIDOS** на диске **C**, которая доступна только на чтение, соответственно в поддиректориях **C:\AIDOS\SYSTEM** и **C:\AIDOS\LAB_RAB**.

Это сделано для того, чтобы студенты не могли нарушить исполнимый код системы "Эйдос", а также содержание учебного пособия и баз данных лабораторных работ (что, как показывает опыт преподавания, происходит незамедлительно, если директории, где находятся эти файлы, доступны на запись). Поэтому перед выполнением каждой лабораторной работы необходимо установить в директорию **C:\WORK** сначала саму систему "Эйдос", а затем в директорию, где она находится – базы данных выполняемой лабораторной работы. Это можно рассматривать как своеобразную плату за неуместную активность студентов.

Здесь необходимо отметить, что:

- текущая версия системы "Эйдос" (12.5) работает с базами данных, находящимися в той же директории, где установлена сама система;
- поскольку при работе системы ее базы данных модифицируются, то рабочая директория системы "Эйдос" должна быть доступна на запись.

ИНСТРУКЦИЯ

по установке системы "Эйдос" и баз данных лабораторной работы

1. Скопировать любым файл-менеджером *поддиректорию* **SYSTEM** (*целиком, а не файлы находящиеся в ней*) из директории **AIDOS** на диске **C** в поддиректорию **C:\WORK**.

2. Скопировать все файлы из указанной преподавателем поддиректории директории **C:\AIDOS\LAB_RAB** с подготовленными базами данных выполняемой лабораторной работы (или со своего носителя, в случае продолжения лабораторной работы) в поддиректорию **C:\WORK\SYSTEM**. При этом файлы копируются "поверх", т.е. с заменой имеющихся в этой директории.

3. Войти в директорию **C:\WORK\SYSTEM** и запустить на исполнение систему "Эйдос". Исполнимый модуль запуска системы называется **_AIDOS.EXE**.

4. Во 2-м режиме 7-й подсистемы системы "Эйдос" выполнить переиндексацию всех баз данных.

5. В 4-м подрежиме 1-го режима 7-й подсистемы системы "Эйдос" сгенерировать базы данных распознаваемых анкет.

6. В 3-м подрежиме 1-го режима 7-й подсистемы системы "Эйдос" сгенерировать базы данных обучающей информации (по указанию преподавателя).

7. Перед окончанием занятия, в случае если планируется продолжить данную работу на следующем занятии, скопировать следующие базы данных на дискету или flash-диск: **object.dbf**, **priz_per.dbf**, **priz_ob.dbf**, **obinfzag.dbf**, **obinfkpr.dbf**.

2.2.3. Выполнение и сдача лабораторной работы

Выполняются лабораторные работы, как правило, студентами индивидуально. В качестве исключения (при недостатке компьютеров в классе) допускается выполнение одной работы небольшими группами по 2-3 студента.

Сдаются лабораторные работы студентами только индивидуально в форме:

- демонстрации и объяснению преподавателю созданного ими приложения *на компьютере* непосредственно в среде используемой интеллектуальной информационной системы;
- предъявления и анализа сгенерированных в системе текстовых и графических выходных форм (в виде файлов);
- ответов на контрольные вопросы к данной лабораторной работе и на понимание базовых понятий.

2.3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ПРОВЕДЕНИЯ ЛАБОРАТОРНЫХ РАБОТ В КОМПЬЮТЕРНОМ КЛАССЕ

2.3.1. Материально-техническое обеспечение

Для проведения лабораторных работ по дисциплине "Интеллектуальные информационные системы" необходим компьютерный класс с учебной доской, оснащенный 30 рабочими местами в составе:

1. Стол.
2. Стул или кресло.
3. ИБМ-совместимый персональный компьютер класса Pentium IV с монитором не менее 17 дюймов, оперативной памятью не менее 256 Мб и внешней памятью не менее 80 Гб.

Все рабочие места должны быть объединены в локальную сеть не менее 100 Мбит/с с выходом в Internet.

2.3.2. Общее программное обеспечение

2.3.2.1. Операционные системы:

MS DOS

MS Windows.

2.3.2.2. Программы под MS Windows

MS Word – текстовый редактор;

MS Excel – табличный процессор;

PhotoShop – графический редактор;

Windows & Total Commander.

2.3.2.3. Программы под MS DOS

Norton Commander (NC) – файл-менеджер;

MultiEdit (ME) – текстовый редактор.

2.3.3. Специальное программное обеспечение

При выполнении лабораторных работ и заданий по самостоятельной работе студентов используется следующее специальное программное обеспечение:

– Универсальная когнитивная аналитическая система "Эйдос" версии 12.5 или выше.

– Нейросетевой пакет Neuro Office фирмы "АЛЬФА-СИСТЕМ" (С`Петербург).

2.4. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ЛАБОРАТОРНЫМ РАБОТАМ

ЛР-1:

"Прогнозирование вероятных пунктов назначения железнодорожных составов"

Краткая теория

Данная задача взята из книги Д.Мичи и Р.Джонстона "Компьютер – творец" (с.205-208), в которой она приводится в качестве примера задачи, решаемой методами искусственного интеллекта. Авторами этой задачи являются Рышард Михальски и Джеймс Ларсон.

Суть этой задачи сводится к тому, чтобы выработать правила, обеспечивающие идентификацию железнодорожных составов и прогнозирование направления их следования на основе их формализованных или вербальных описаний.

Выбор данной задачи не накладывает ограничений на выводы, полученные в результате ее исследования. Это обусловлено тем, что она имеет ряд характерных особенностей, наблюдающихся в подобных задачах в самых различных предметных областях. Поэтому ее с полным основанием можно рассматривать как типовую для широкого класса задач идентификации и прогнозирования.

Эти особенности состоят в следующем:

1. Рассматривается ряд объектов, имеющих сложную многоуровневую структуру признаков.

2. Для каждого из этих объектов известно, к каким обобщенным категориям (классам) он относится.

3. Необходимо сформировать модель, обеспечивающую как идентификацию объектов, так и определение их принадлежности к обобщенным классам.

Если признаки и классы относятся к одному времени, то имеет место задача идентификации (распознавания). Если же признаки (факторы, причины) относятся к прошлому, а классы, характеризующие состояния объектов, – к будущему, то это задача прогнозирования. Математически эти задачи не отличаются.

Существуют различные подходы к решению данной задачи, отличающиеся способами формализации предметной области, объектов обучающей выборки и синтеза математической модели.

В данной работе мы исследуется два основных подхода:

1. "Классический", основанный на изучении объектов предметной области экспертами (когнитивный анализ), выделении признаков объектов и формировании описательных шкал и градаций, которым соответствуют уникальные коды.

2. "Лингвистический", в котором вербальные описания объектов предметной области на естественном языке используются для автоматизированной формализации предметной области, формирования обучающей выборки и синтеза модели.

Кратко рассмотрим реализацию обоих этих подходов в интеллектуальной технологии "Эйдос" [81]. Исходные данные к задаче представлены в графической форме (рисунок 93).

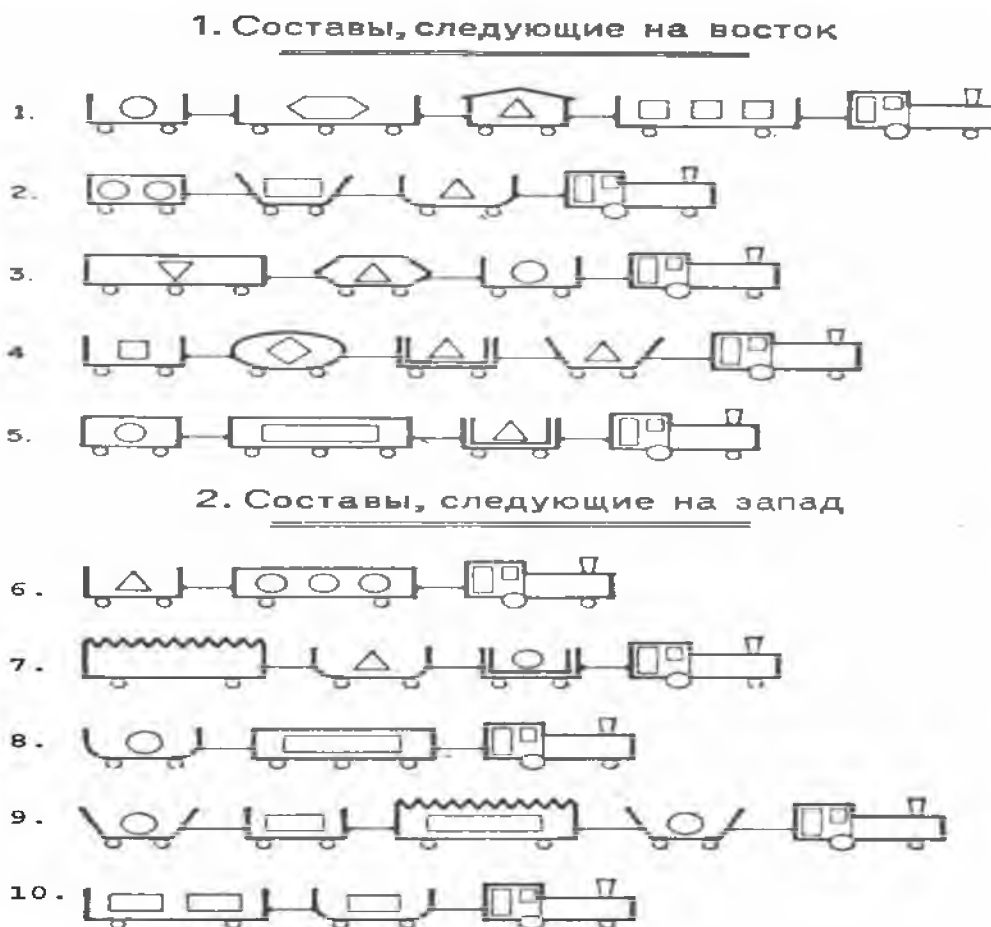


Рисунок 93. Примеры поездов, идущих на запад и на восток

Железнодорожный состав является сложным объектом, имеющим несколько иерархических уровней и допускающим, соответственно, несколько уровней описания. Некоторые из этих уровней приведены в таблице 42.

Можно, например, описывать составы с использованием шкал только 2-го или только 3-го уровней. Возможны и смешанные варианты.

1-й вариант соответствует представлению о том, что на запад или восток идут не составы, а отдельные вагоны (отличающиеся типом и грузом), а состав идет туда же, куда и большинство вагонов.

2-й вариант предполагает, что составы как бы не состоят из различных вагонов с различными грузами, а свойства вагонов и грузов являются свойствами непосредственно состава.

Таблица 42 – УРОВНИ ОПИСАНИЯ ЖЕЛЕЗНОДОРОЖНЫХ СОСТАВОВ

№	Уровень описания	Шкала	Градация
1.	На уровне направления движения состава	Направление движения состава	Запад
			Восток
2.	На уровне поезда	Количество вагонов в составе	2
			3
			4
		Типы вагонов с грузами	Закодировать все варианты
3.	На уровне признаков вагонов и грузов	Форма вагона	V-образная
			Прямоугольная
			Ромбовидная
			U-образная
			Эллипсоидная
		Длина вагона	Короткий
			Длинный
		Количество осей вагона	2
			3
		Вид стенок вагона	Одинарные
			Двойные
		Вид крыши	Отсутствует
			Гофрированная
			Двухскатная
			Прямая (эллипсоидная)
		Вид и количество груза	1 большой круг
			2 маленьких круга
			3 маленьких круга
			1 квадрат
			3 квадрата
			1 короткий прямоугольник
			2 коротких прямоугольника
			1 длинный прямоугольник
			1 треугольник
			1 перевернутый треугольник
			1 ромб
			1 шестиугольник
			Груз отсутствует

Необходимо отметить, что сравнительно небольшое количество признаков вагонов обеспечивает огромное количество различных типов вагонов с различными сочетаниями этих признаков, из которых реально в приведенных составах встречается лишь незначительная часть. Очевидно, существует еще большее число вариантов сочетаний различных типов вагонов с видами грузов, порядков следования вагонов и грузов друг за другом и т.п. Из этого следуют по крайней мере два основных вывода:

1. Составить исчерпывающий справочник для описания состава на 2-м уровне, в котором бы указывались все варианты сочетаний различных типов вагонов с различными грузами, на практике довольно трудоемко и вряд ли целесообразно (из-за его огромной размерности).

2. Реально встречающиеся в составах сочетания типов вагонов и видов грузов практически все будут являться уникальными, что обеспечит однозначную идентификацию составов, если их описывать только на 2-м уровне. Это превращает задачу в тривиальную. Поэтому будем рассматривать описание составов на 3-м уровне с элементами 2-го.

Вербальные описания железнодорожных составов практически на естественном языке являются их лингвистическими моделями, которые могут обрабатываться в системе "Эйдос". *При этом в справочники будут заноситься, причем автоматически, только реально встретившиеся признаки составов.*

Рассмотрим классический и лингвистический подходы на примере решения задач 1 и 2.

Задача 1

1. Формализовать задачу, создав классификационные и описательные шкалы (с использованием таблицы 42) и обучающую выборку на основе рисунка 93.

2. Осуществить синтез и верификацию модели.

3. Провести анализ модели, сформулировав правила для прогнозирования направления движения составов (в режиме: "Типология", "Информационные портреты классов").

4. Оценить ценность признаков для прогнозирования. Выделить признаки, наиболее существенные для решения поставленной задачи.

5. Сравнить составы по степени "типичности" для своих кластеров ("Идущие на запад", "Идущие на восток"). Вывести в графической форме семантические сети составов, построить классические когнитивные карты для составов, идущих на запад и на восток.

Задача 2

1. Создать стандартизированные (с использованием исходных данных, приведенных на рисунке 93) **текстовые** описания составов в виде отдельных файлов стандарта DOS-текст с концами строк, записать их в поддиректорию DOB в виде: #####-zap.txt и #####-vos.txt.

2. Сгенерировать классификационные и описательные шкалы в режиме: "Автоввод первичных признаков и ТХТ-файлов", "Признаки – слова".

3. Сгенерировать обучающую выборку с использованием режима: "Ввод – корректировка обучающей выборки", "F7 InpTХT", "F6 Ввод из всех файлов".

Дополнить анкеты, соответствующие составам, кодами принадлежности к обобщенным образам классов: "Идущие на запад", "Идущие на восток".

4. Осуществить синтез и верификацию семантической информационной модели.

5. Провести анализ модели, сформулировав правила для прогнозирования направления движения составов (в режиме: "Типология", "Информационные портеты классов").

6. Оценить ценность признаков для прогнозирования. Выделить признаки, наиболее существенные для решения поставленной задачи.

7. Сравнить составы по степени "типичности" для своих кластеров ("Идущие на запад", "Идущие на восток"). Отобразить в графической форме семантические сети составов, построить классические когнитивные карты для составов, идущих на запад и на восток.

Пример решения задачи 1

Пример решения задания 1.1: "Формализовать задачу, создав классификационные и описательные шкалы и обучающую выборку"

Для этих целей используем таблицу 43 и рисунок 93. В результате получим таблицы 43 и 44.

**Таблица 43 – КЛАССИФИКАЦИОННЫЕ
И ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ**

КЛАССЫ		ПРИЗНАКИ	
Код	Наименование	Код	Наименование
класса	класса распознавания		
1	Состав следует на ВОСТОК...	1	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ
2	Состав следует на ЗАПАД....	2	ФОРМА ВАГОНА:
3	Состав-01.....	3	V-образная
4	Состав-02.....	4	Прямоугольная
5	Состав-03.....	5	Ромбовидная
6	Состав-04.....	6	U-образная
7	Состав-05.....	7	Эллипсоидная
8	Состав-06.....	8	ДЛИНА ВАГОНА:
9	Состав-07.....	9	Короткий
10	Состав-08.....	10	Длинный
11	Состав-09.....	11	КОЛИЧЕСТВО ОСЕЙ ВАГОНА:
12	Состав-10.....	12	2
		13	3
		14	ВИД СТЕНОК ВАГОНА:
		15	Одинарные
		16	Двойные
		17	ВИД КРЫШИ ВАГОНА:
		18	Отсутствует
		19	Гофрированная
		20	Двухскатная
		21	Прямая (эллипсоидная)
		22	ГРУЗ (КОЛИЧЕСТВО И ВИД):
		23	1 большой круг
		24	2 маленьких круга
		25	3 маленьких круга
		26	1 квадрат
		27	3 квадрата
		28	1 короткий прямоугольник
		29	2 коротких прямоугольника
		30	1 длинный прямоугольник
		31	1 треугольник
			2 перевернутый треугольник
			3 ромб
			4 шестиугольник
			5 Груза нет

Таблица 44 – ОБУЧАЮЩАЯ ВЫБОРКА

№	Наименование состава	Коды классов		Коды признаков										
1	Состав-1	1	3	3	5	5	5	5	9	9	10	10	11	11
				11	12	13	13	13	13	15	15	15	17	19
				29	27	23								
2	Состав-2	1	4	2	5	4	7	9	9	9	2	2	2	13
				13	13	15	15	18	20	24	27			
3	Состав-3	1	5	2	5	6	5	10	9	9	12	11	11	13
				13	13	18	18	15	28	27	19			
4	Состав-4	1	6	3	5	8	5	4	9	9	9	9	13	13
				13	14	15	15	15	18	22	29	27	27	11
				11	11	11								
5	Состав-5	1	7	2	5	5	5	9	9	10	11	11	12	13
				13	14	18	18	15	19	26	27			
6	Состав-6	2	8	1	5	5	9	10	11	11	13	13	15	18
				27	21									
7	Состав-7	2	9	2	5	5	7	10	9	9	11	11	11	13
				13	14	16	15	15	31	27	19			
8	Состав-8	2	10	1	7	5	9	10	11	12	19	26	15	18
				13	13									
9	Состав-9	2	11	3	4	4	5	5	9	9	9	10	11	11
				11	11	13	13	13	13	15	15	15	16	19
				24	26	19								
10	Состав-10	2	12	1	5	7	9	10	11	11	13	13	15	15
				25	24									

Пример решения задачи 2

Пример решения задания 2.1: "Создать стандартизированные текстовые описания составов в виде отдельных файлов стандарта DOS-текст"

Создать стандартизированные (с использованием рисунка 1) текстовые описания составов в виде отдельных файлов стандарта DOS-текст с концами строк, записать их в поддиректорию DOB с именами вида: #####-zap.txt и #####-vos.txt, где ##### – номер анкеты (состава): 0001, 0002 и т.д., а остальные символы – произвольные, но выбираются таким образом, чтобы они отражали содержание анкеты.

0001-VOS.TXT

Кол-во вагонов=4
 форма вагона прямоугольная
 форма вагона прямоугольная
 форма вагона прямоугольная
 форма вагона прямоугольная
 длина вагона короткий
 длина вагона короткий
 длина вагона короткий
 длина вагона длинный
 длина вагона длинный
 N осей вагона=2
 N осей вагона=2
 N осей вагона=2

0002-VOS.TXT

Кол-во вагонов=3
 форма вагона прямоугольная
 форма вагона V-образная
 форма вагона U-образная
 длина вагона короткий
 длина вагона короткий
 длина вагона короткий
 N осей вагона=2
 N осей вагона=2
 N осей вагона=2
 стенки вагона одинарные
 стенки вагона одинарные

<p> N_осей_вагона=3 стенки_вагона_одинарные стенки_вагона_одинарные стенки_вагона_одинарные стенки_вагона_одинарные крыша_вагона_отсутствует крыша_вагона_отсутствует крыша_вагона_отсутствует крыша_вагона_двухскатная груз_1_большой_круг груз_1_шестиугольник груз_1_треугольник груз_3_квадрата </p>	<p> стенки_вагона_одинарные крыша_вагона_отсутствует крыша_вагона_отсутствует крыша_вагона_прямая груз_2_маленьких_круга груз_1_короткий_прямоугольник груз_1_треугольник </p>
<p> 0003-VOS.TXT Кол-во_вагонов=3 форма_вагона_прямоугольная форма_вагона_прямоугольная форма_вагона_ромбовидная длина_вагона_длинный длина_вагона_короткий длина_вагона_короткий N_осей_вагона=3 N_осей_вагона=2 N_осей_вагона=2 стенки_вагона_одинарные стенки_вагона_одинарные стенки_вагона_одинарные крыша_вагона_отсутствует крыша_вагона_прямая крыша_вагона_прямая груз_1_большой_круг груз_1_треугольник груз_1_перевернутый_треугольник </p>	<p> 0004-VOS.TXT Кол-во_вагонов=4 форма_вагона_прямоугольная форма_вагона_прямоугольная форма_вагона_эллипсоидная форма_вагона_V-образная длина_вагона_короткий длина_вагона_короткий длина_вагона_короткий длина_вагона_короткий длина_вагона_короткий N_осей_вагона=2 N_осей_вагона=2 N_осей_вагона=2 N_осей_вагона=2 стенки_вагона_одинарные стенки_вагона_одинарные стенки_вагона_одинарные стенки_вагона_двойные крыша_вагона_отсутствует крыша_вагона_отсутствует крыша_вагона_отсутствует крыша_вагона_прямая груз_1_квадрат груз_1_треугольник груз_1_треугольник груз_1_ромб </p>
<p> 0005-VOS.TXT Кол-во_вагонов=3 форма_вагона_прямоугольная форма_вагона_прямоугольная форма_вагона_прямоугольная длина_вагона_короткий длина_вагона_короткий длина_вагона_длинный N_осей_вагона=3 N_осей_вагона=2 N_осей_вагона=2 стенки_вагона_одинарные стенки_вагона_одинарные стенки_вагона_двойные крыша_вагона_отсутствует крыша_вагона_прямая крыша_вагона_прямая груз_1_большой_круг груз_1_треугольник груз_1_длинный_прямоугольник </p>	<p> 0006-ZAP.TXT Кол-во_вагонов=2 форма_вагона_прямоугольная форма_вагона_прямоугольная длина_вагона_короткий длина_вагона_длинный N_осей_вагона=2 N_осей_вагона=2 стенки_вагона_одинарные стенки_вагона_одинарные крыша_вагона_отсутствует крыша_вагона_прямая груз_3_маленьких_круга груз_1_треугольник </p>

<p>0007-ZAP.TXT Кол-во вагонов=3 форма вагона прямоугольная форма вагона прямоугольная форма вагона U-образная длина вагона короткий длина вагона короткий длина вагона длинный N осей вагона=2 N осей вагона=2 N осей вагона=2 стенки вагона одинарные стенки вагона одинарные стенки вагона двойные крыша вагона отсутствует крыша вагона отсутствует крыша вагона гофрированная груза нет груз 1 большой круг груз 1 треугольник</p>	<p>0008-ZAP.TXT Кол-во вагонов=2 форма вагона прямоугольная форма вагона U-образная длина вагона короткий длина вагона длинный N осей вагона=2 N осей вагона=3 стенки вагона одинарные стенки вагона одинарные крыша вагона отсутствует крыша вагона прямая груз 1 большой круг груз 1 длинный прямоугольник</p>
<p>0009-ZAP.TXT Кол-во вагонов=4 форма вагона прямоугольная форма вагона прямоугольная форма вагона V-образная форма вагона V-образная длина вагона короткий длина вагона короткий длина вагона короткий длина вагона длинный N осей вагона=2 N осей вагона=2 N осей вагона=2 N осей вагона=2 стенки вагона одинарные стенки вагона одинарные стенки вагона одинарные стенки вагона одинарные крыша вагона отсутствует крыша вагона отсутствует крыша вагона отсутствует крыша вагона гофрированная груз 1 большой круг груз 1 большой круг груз 1 длинный прямоугольник груз 1 короткий прямоугольник</p>	<p>0010-ZAP.TXT Кол-во вагонов=2 форма вагона прямоугольная форма вагона U-образная длина вагона короткий длина вагона длинный N осей вагона=2 N осей вагона=2 стенки вагона одинарные стенки вагона одинарные крыша вагона отсутствует крыша вагона отсутствует груз 1 короткий прямоугольник груз 2 коротких прямоугольника</p>

Пример решения задания 2.2: "Сгенерировать классификационные и описательные шкалы"

Для этого используем режим: "F1 Словари – Автовывод первичных признаков и TXT-файлов – F3 Признаки – слова".

Классы во втором задании те же самые, что и в первом. Признаки выглядят несколько иначе, т.к. формируются автоматически из текстовых описаний составов, но по смыслу они также совпадают (таблица 45).

**Таблица 45 – КЛАССИФИКАЦИОННЫЕ И ОПИСАТЕЛЬНЫЕ
ШКАЛЫ И ГРАДАЦИИ**

КЛАССЫ		ПРИЗНАКИ	
Код	Наименование		
класса	класса распознавания		
1	Состав следует на ВОСТОК	1	N_осей_вагона=2
2	Состав следует на ЗАПАД	2	N_осей_вагона=3
3	Состав-01	3	Кол-во_вагонов=2
4	Состав-02	4	Кол-во_вагонов=3
5	Состав-03	5	Кол-во_вагонов=4
6	Состав-04	6	груз_1_большой_круг
7	Состав-05	7	груз_1_длинный_прямоугольник
8	Состав-06	8	груз_1_квадрат
9	Состав-07	9	груз_1_короткий_прямоугольник
10	Состав-08	10	груз_1_перевернутый_треугольник
11	Состав-09	11	груз_1_ромб
12	Состав-10	12	груз_1_треугольник
		13	груз_1_шестиугольник
		14	груз_2_коротких_прямоугольника
		15	груз_2_маленьких_круга
		16	груз_3_квадрата
		17	груз_3_маленьких_круга
		18	груза_нет
		19	длина_вагона_длинный
		20	длина_вагона_короткий
		21	крыша_вагона_гофрированная
		22	крыша_вагона_двухскатная
		23	крыша_вагона_отсутствует
		24	крыша_вагона_прямая
		25	стенки_вагона_двойные
		26	стенки_вагона_одинарные
		27	форма_вагона_U-образная
		28	форма_вагона_V-образная
		29	форма_вагона_прямоугольная
		30	форма_вагона_ромбовидная
		31	форма_вагона_эллипсоидная

Пример решения задания 2.3: "Сгенерировать обучающую выборку"

Используем режим: "F2 Обучение – Ввод – корректировка обучающей выборки – F7 InpTXT – F6 Ввод из всех файлов". Затем необходимо дополнить анкеты, соответствующие составам, кодами принадлежности к обобщенным образам классов: "Идущие на запад", "Идущие на восток". Обучающая выборка будет иметь вид, приведенный в таблице 46.

Таблица 46 – ОБУЧАЮЩАЯ ВЫБОРКА

[illegible]

Этапы синтеза модели, ее оптимизации, проверки адекватности (которая в данном случае равна 100 %) и анализа подробно описаны в работах [81]. Поэтому в данной работе мы приведем лишь их результаты.

Пример решения задания 2.4: "Осуществить синтез и верификацию семантической информационной модели"

Основная матрица семантической информационной модели приведена в таблице 47.

Таблица 47 – МАТРИЦА ИНФОРМАТИВНОСТЕЙ

Атр	Коды классов												Сум.	Ср.	Отк.
	1	2	3	4	5	6	7	8	9	10	11	12			
1	-0,03	0,03	-0,08	0,09	-0,16	0,09	-0,16	0,07	0,09	-0,35	0,09	0,07	-0,23	-0,02	0,14
2	0,17	-0,34	0,39		0,55		0,55			0,78			2,11	0,18	0,32
3		0,5						0,96		0,96		0,96	3,37	0,28	0,43
4	0,17	-0,34		0,55	0,55		0,55		0,55				2,05	0,17	0,30
5	0,1	-0,16	0,56			0,56					0,56		1,62	0,14	0,26
6	-0,16	0,16	0,05		0,22		0,22		0,22	0,44	0,47		1,60	0,13	0,19
7	-0,32	0,26					0,73			0,96	0,56		2,18	0,18	0,37
8	0,35					1,22							1,57	0,13	0,36
9	-0,32	0,26		0,73							0,56	0,96	2,18	0,18	0,37
10	0,35				1,39								1,74	0,14	0,40
11	0,35				0	1,22							1,57	0,13	0,36
12	0,17	-0,34	-0,03	0,13	0,13	0,39	0,13	0,36	0,13				1,09	0,09	0,19
13	0,35		1,22										1,57	0,13	0,36
14		0,5										1,62	2,12	0,18	0,48
15	0,35			1,39									1,74	0,14	0,40
16	0,35		1,22										1,57	0,13	0,36
17		0,5						1,62					2,12	0,18	0,48
18		0,5						0	1,39				1,89	0,16	0,41
19	-0,14	0,15	0,32		0,06		0,06	0,29	0,06	0,29	-0,1	0,29	1,29	0,11	0,16
20	0,06	-0,08	-0,2	0,22	-0,03	0,22	-0,03	-0,22	-0,03	-0,22	0,05	-0,22	-0,48	-0,04	0,16
21		0,5							0,97		0,81		2,28	0,19	0,36
22	0,35		1,22										1,57	0,13	0,36
23	-0,04	0,05	0,11	0,03	-0,39	0,11	-0,39	-0,16	0,03	-0,16	0,11	0,26	-0,43	-0,04	0,20
24	0,17	-0,34		0,13	0,55	-0,03	0,55	0,36		0,36			1,77	0,15	0,27
25	0,1	-0,16				0,56	0,73		0,73				1,95	0,16	0,31
26	-0,01	0,01	0,07	0,06	0,06	-0,1	-0,18	0,05	-0,18	0,05	0,07	0,05	-0,05	0,00	0,10
27	-0,49	0,33		0,55					0,55	0,78		0,78	2,51	0,21	0,39
28	-0,07	0,08		0,55		0,39					0,81	0	1,76	0,15	0,28
29	0,04	-0,05	0,25	-0,42		-0,17	0,24	0,23		-0,19	-0,17	-0,19	-0,42	-0,03	0,21
30	0,35				1,39								1,74	0,14	0,40
31	0,35					1,22							1,57	0,13	0,36
Сум.	2,53	2,01	5,12	4,02	4,34	5,70	3,01	3,56	4,51	3,71	3,82	4,58	46,91		
Ср.	0,08	0,06	0,17	0,13	0,14	0,18	0,10	0,12	0,15	0,12	0,12	0,15		0,13	
Отк.	0,22	0,25	0,38	0,32	0,38	0,39	0,26	0,35	0,34	0,33	0,26	0,39			0,33

Пример решения задания 2.5 "Провести анализ модели, сформулировав правила для прогнозирования направления движения составов"

В подсистеме: "Типология", "Информационные портеты классов" системы "Эйдос" получаем следующие информационные портреты классов (таблицы 48 и 49).

ТАБЛИЦА 48 – ИНФОРМАЦИОННЫЙ ПОРТРЕТ КЛАССА РАСПОЗНАВАНИЯ:						
Код:		Наименование: Состав следует на ВОСТОК				
08-11-04 18:48:42		фильтр: All, Positive			г.Краснодар	
N	Код	Наименования признаков	Инфор- мат.	Инфор- мат.	Суммар инф-	
п/п	приз нака		т. Бит.	т. %	т. %%	
1	8	груз_1_квадрат.....	0.347	9.67	9.7	
2	10	груз_1_перевернутый_треугольник	0.347	9.67	19.3	
3	11	груз_1_ромб.....	0.347	9.67	29.0	
4	13	груз_1_шестиугольник.....	0.347	9.67	38.7	
5	15	груз_2_маленьких_круга.....	0.347	9.67	48.4	
6	16	груз_3_квадрата.....	0.347	9.67	58.0	
7	22	крыша_вагона_двухскатная.....	0.347	9.67	67.7	
8	30	форма_вагона_ромбовидная.....	0.347	9.67	77.4	
9	31	форма_вагона_эллипсоидная.....	0.347	9.67	87.0	
10	2	N_осей_вагона=3.....	0.173	4.82	91.8	
11	4	Кол-во_вагонов=3.....	0.173	4.82	96.7	
12	12	груз_1_треугольник.....	0.173	4.82	101.5	
13	24	крыша_вагона_прямая.....	0.173	4.82	106.3	
14	5	Кол-во_вагонов=4.....	0.102	2.84	109.2	
15	25	стенки_вагона_двойные.....	0.102	2.84	112.0	
16	20	длина_вагона_короткий.....	0.057	1.59	113.6	
17	29	форма_вагона_прямоугольная.....	0.038	1.07	114.7	
18	26	стенки_вагона_одинарные.....	-0.008	-0.23	114.9	
19	1	N_осей_вагона=2.....	-0.027	-0.75	115.6	
20	23	крыша_вагона_отсутствует.....	-0.041	-1.14	116.8	
21	28	форма_вагона_V-образная.....	-0.072	-2.00	118.8	
22	19	длина_вагона_длинный.....	-0.143	-3.99	122.8	
23	6	груз_1_большой_круг.....	-0.165	-4.60	127.4	
24	7	груз_1_длинный_прямоугольник..	-0.316	-8.83	136.2	
25	9	груз_1_короткий_прямоугольник..	-0.316	-8.83	145.0	
26	27	форма_вагона_U-образная.....	-0.490	-13.67	158.7	

Универсальная когнитивная аналитическая система НПП *Эйдос*

ТАБЛИЦА 49 – ВЛИЯНИЕ ПРИЗНАКОВ НА РЕЗУЛЬТАТ ПРОГНОЗИРОВАНИЯ: "НАПРАВЛЕНИЕ СЛЕДОВАНИЯ – НА ЗАПАД"					
ИНФОРМАЦИОННЫЙ ПОРТРЕТ КЛАССА РАСПОЗНАВАНИЯ:					
Код: 2		Наименование: Состав следует на ЗАПАД			
08-11-04 18:49:04		Фильтр: All, Positive		г.Краснодар	
N	Код	Наименования	Инфор-	Инфор-	Суммар
п/п	приз	признаков	мат-ть	мат-ть	инф-ть
	нака		Бит.	%	%%
1	3	Кол-во вагонов=2.....	0.500	13.94	13.9
2	14	груз_2 коротких прямоугольника	0.500	13.94	27.9
3	17	груз_3 маленьких круга.....	0.500	13.94	41.8
4	18	груза нет.....	0.500	13.94	55.8
5	21	крыша вагона гофрированная....	0.500	13.94	69.7
6	27	форма вагона U-образная.....	0.326	9.10	78.8
7	7	груз_1 длинный прямоугольник..	0.255	7.12	85.9
8	9	груз_1 короткий прямоугольник.	0.255	7.12	93.0
9	6	груз_1 большой круг.....	0.162	4.52	97.6
10	19	длина вагона длинный.....	0.145	4.05	101.6
11	28	форма вагона V-образная.....	0.082	2.27	103.9
12	23	крыша вагона отсутствует.....	0.049	1.36	105.2
13	1	N осей вагона=2.....	0.033	0.93	106.2
14	26	стенки вагона одинарные.....	0.010	0.29	106.5
15	29	форма вагона прямоугольная....	-0.053	-1.48	107.9
16	20	длина вагона короткий.....	-0.083	-2.30	110.2
17	5	Кол-во вагонов=4.....	-0.163	-4.55	114.8
18	25	стенки вагона двойные.....	-0.163	-4.55	119.3
19	2	N осей вагона=3.....	-0.337	-9.40	128.7
20	4	Кол-во вагонов=3.....	-0.337	-9.40	138.1
21	12	груз_1 треугольник.....	-0.337	-9.40	147.5
22	24	крыша вагона прямая.....	-0.337	-9.40	156.9
Универсальная когнитивная аналитическая система					
				НПП *ЭЙДОС*	

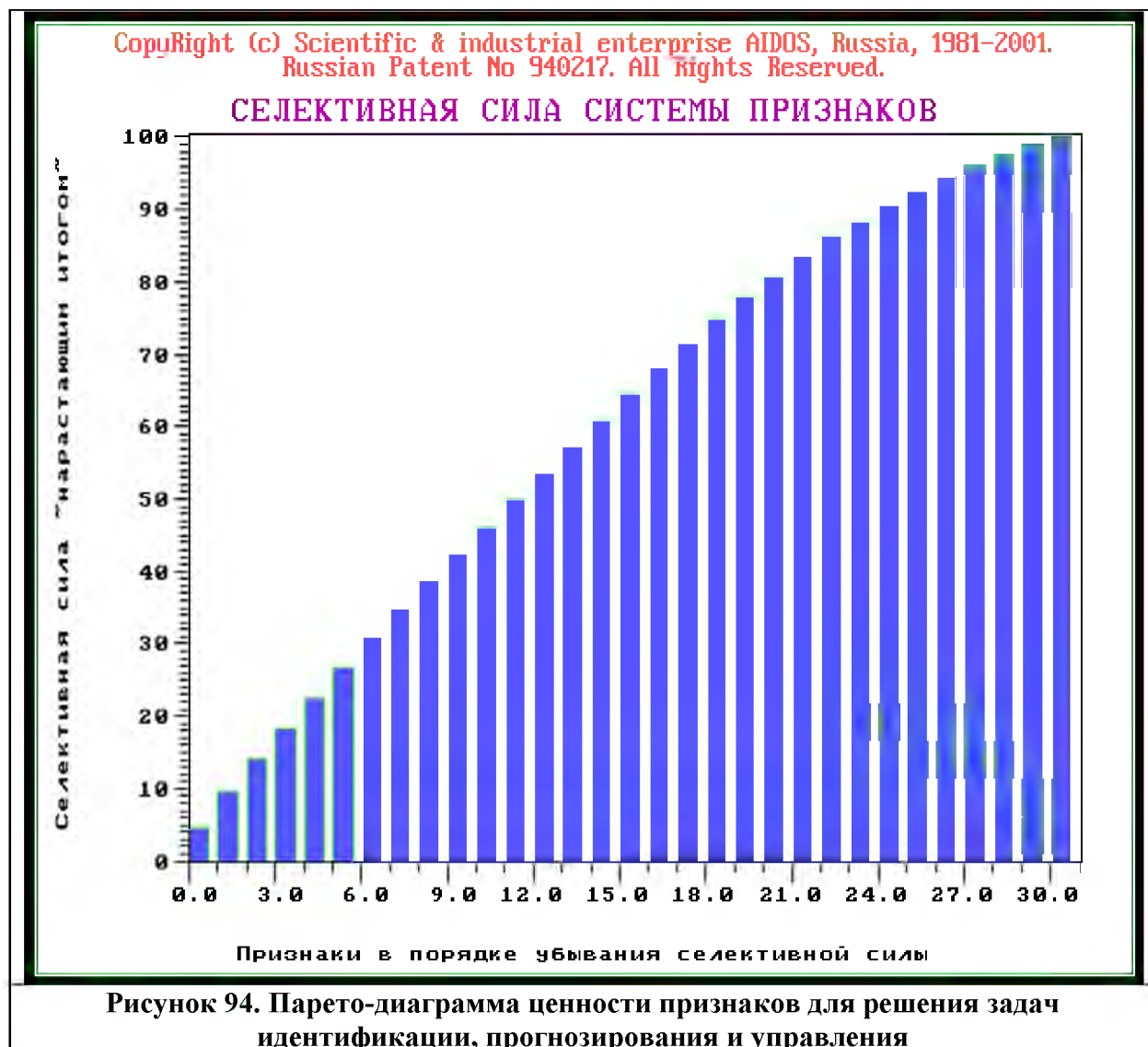
Пример решения задания 2.6: "Оценить ценность признаков для прогнозирования. Выделить признаки, наиболее существенные для решения поставленной задачи"

В подсистеме "Оптимизация" режиме "Исключение признаков с низкой селективной силой" получаем перечень признаков, ранжированных в порядке убывания среднего количества информации о направлении следования состава (таблица 50).

Таблица 50 – ПРИЗНАКИ В ПОРЯДКЕ УБЫВАНИЯ СЕЛЕКТИВНОЙ СИЛЫ

08-11-04		18:49:35		г. Краснодар		
=====						
N°	Код	Н а и м е н о в а н и е	Ценн.	Сумма	Сумма	
п/п	п.пр.	п р и з н а к о в	Бит	Бит	%	
=====						
1	14	груз_2_коротких_прямоугольника.	0.476	0.476	4.888	
2	17	груз_3_маленьких_круга.....	0.476	0.953	9.776	
3	3	Кол-во_вагонов=2.....	0.431	1.384	14.201	
4	18	груза_нет.....	0.414	1.798	18.448	
5	10	груз_1_перевернутый_треугольник	0.405	2.202	22.599	
6	15	груз_2_маленьких_круга.....	0.405	2.607	26.750	
7	30	форма_вагона_ромбовидная.....	0.405	3.011	30.901	
8	27	форма_вагона_U-образная.....	0.388	3.399	34.885	
9	7	груз_1_длинный_прямоугольник...	0.372	3.772	38.705	
10	9	груз_1_короткий_прямоугольник..	0.372	4.144	42.526	
11	8	груз_1_квадрат.....	0.358	4.502	46.203	
12	11	груз_1_ромб.....	0.358	4.861	49.881	
13	13	груз_1_шестиугольник.....	0.358	5.219	53.559	
14	16	груз_3_квадрата.....	0.358	5.577	57.236	
15	22	крыша_вагона_двухскатная.....	0.358	5.936	60.914	
16	31	форма_вагона_эллипсоидная.....	0.358	6.294	64.591	
17	21	крыша_вагона_гофрированная.....	0.358	6.652	68.266	
18	2	N_осей_вагона=3.....	0.323	6.975	71.577	
19	25	стенки_вагона_двойные.....	0.315	7.290	74.807	
20	4	Кол-во_вагонов=3.....	0.304	7.594	77.927	
21	28	форма_вагона_V-образная.....	0.280	7.873	80.797	
22	24	крыша_вагона_прямая.....	0.266	8.139	83.523	
23	5	Кол-во_вагонов=4.....	0.263	8.402	86.224	
24	29	форма_вагона_прямоугольная.....	0.206	8.608	88.340	
25	23	крыша_вагона_отсутствует.....	0.201	8.809	90.403	
26	12	груз_1_треугольник.....	0.191	9.000	92.358	
27	6	груз_1_большой_круг.....	0.190	9.190	94.307	
28	19	длина_вагона_длинный.....	0.160	9.350	95.954	
29	20	длина_вагона_короткий.....	0.158	9.508	97.573	
30	1	N_осей_вагона=2.....	0.141	9.649	99.015	
31	26	стенки_вагона_одинарные.....	0.096	9.745	100.000	
=====						
Универсальная когнитивная аналитическая система				НПП *ЭЙДОС*		

Накопительная диаграмма селективной силы (Парето-диаграмма) приведена на рисунке 94.



Пример решения задания 2.7: "Сравнить составы по степени "типичности" для своих кластеров ("Идущие на запад", "Идущие на восток"). Вывести в графической форме семантические сети составов. Построить классические когнитивные карты для составов, идущих на запад и на восток"

Сравним составы по степени "типичности" для своих кластеров ("Идущие на запад", "Идущие на восток"). В подсистеме "Типология" режиме "Типологический анализ классов распознавания – Кластерный и конструктивный анализ – просмотр и печать кластеров и конструктов" выводим конструкт: "Идущие на запад" и "Идущие на восток" (рисунки 95 и 96).

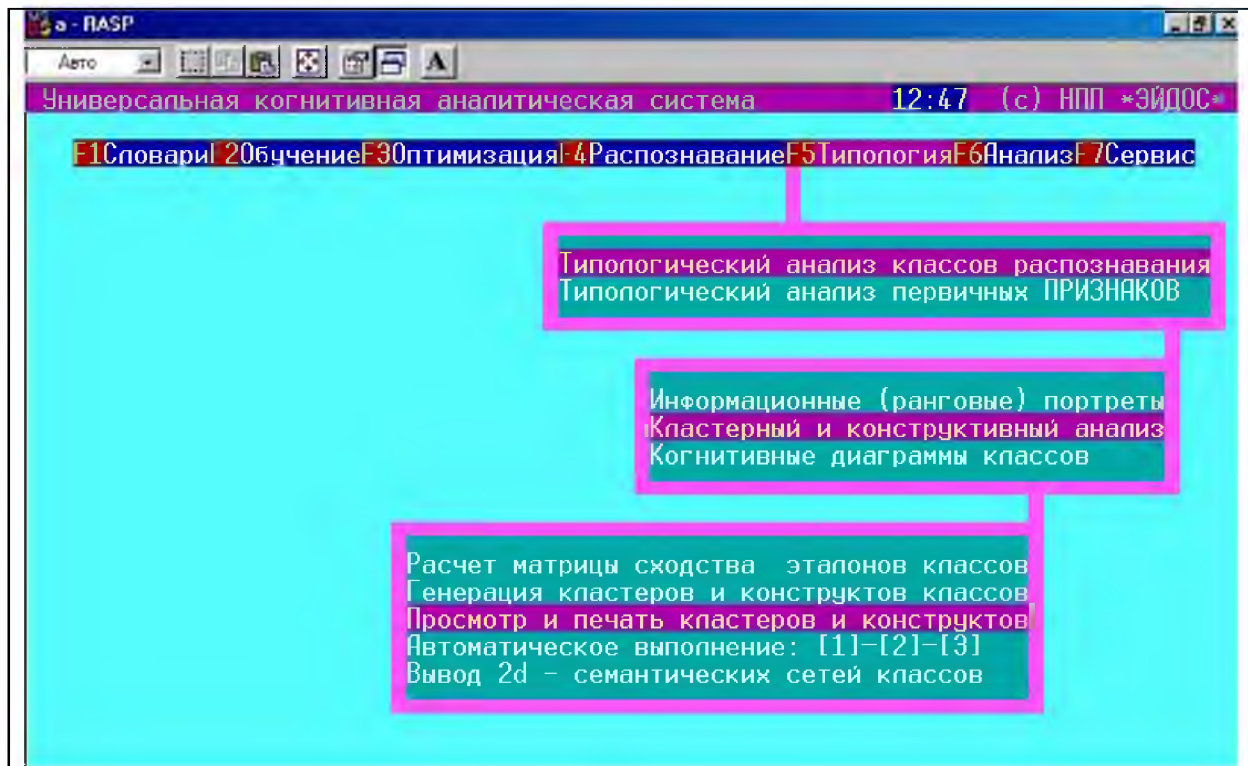


Рисунок 95. Подсистема "Типология", режим "Типологический анализ классов распознавания – Кластерный и конструктивный анализ – просмотр и печать кластеров и конструкторов"

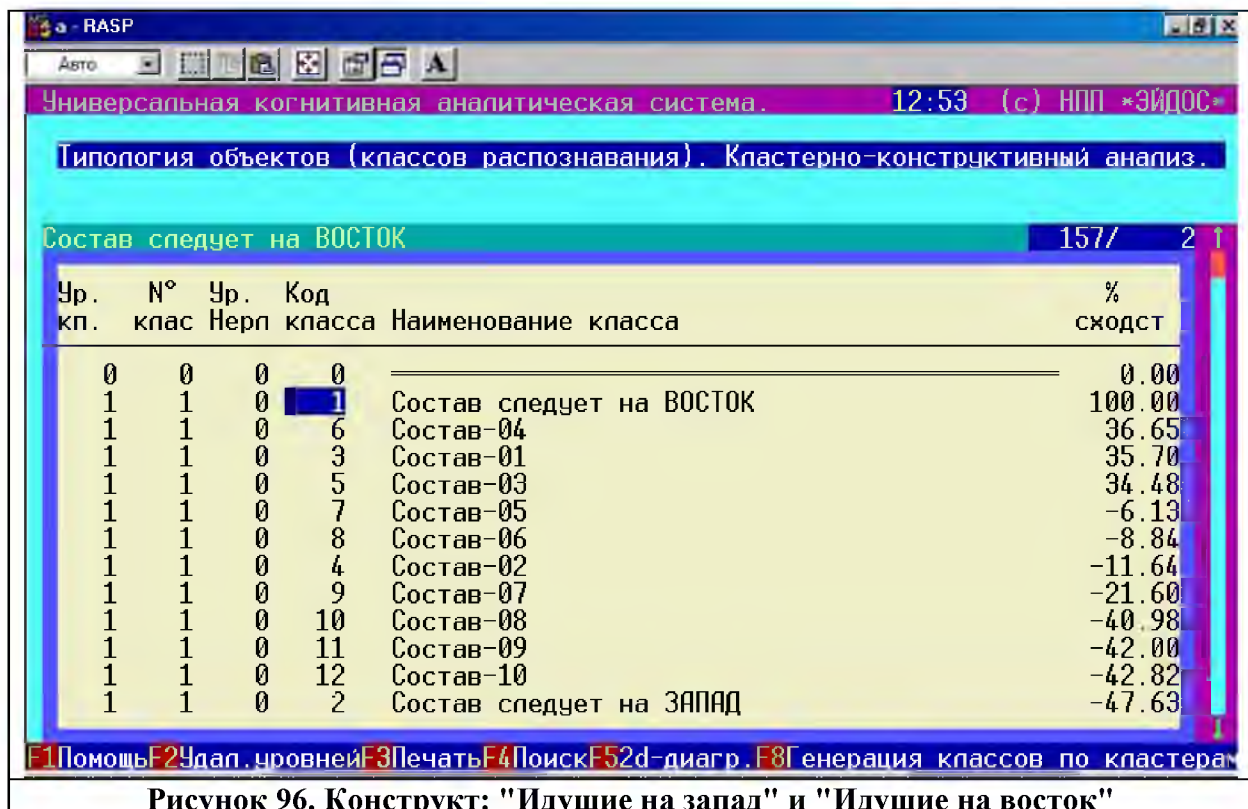


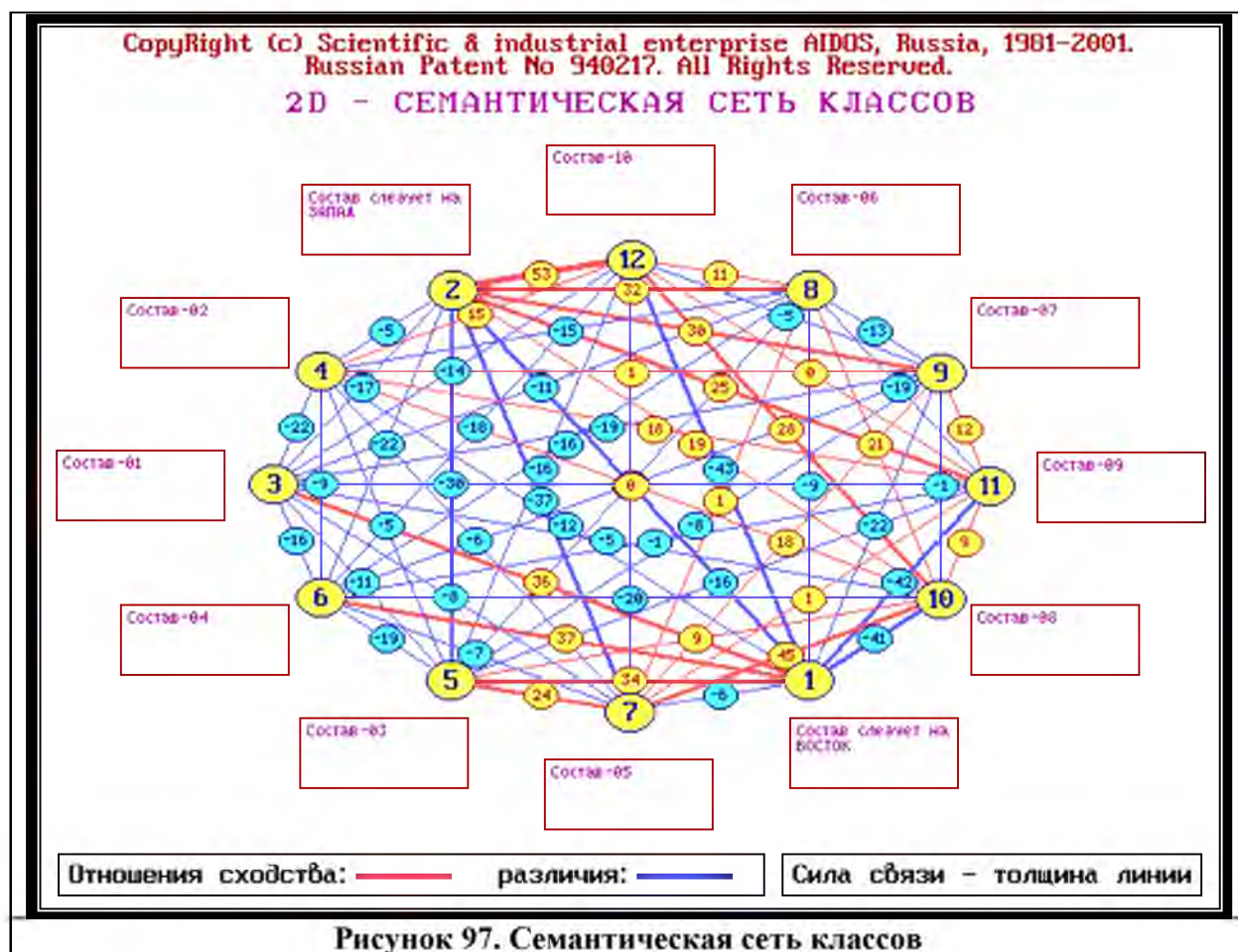
Рисунок 96. Конструкт: "Идущие на запад" и "Идущие на восток"

Из рисунка 4 видно, что:

- составы 4-й, 1-й и 3-й являются типичными для "Идущих на восток", а 5-й и, особенно 2-й, – нетипичными;
- составы 10-й, 9-й и 8-й являются типичными для "Идущих на запад", а 7-й и, особенно 6-й, – нетипичными.

Выведем в графической форме семантические сети составов.

Семантические сети классов отображают результаты кластерно-конструктивного анализа в графической форме. Для этого используется режим: "Вывод 2d-семантических сетей классов" (рисунок 95). Результат приведен на рисунке 97.



Построим классические когнитивные карты для составов, идущих на запад и на восток". В Системе "Эйдос" классическая когнитивная карта строится из двух графических диаграмм:

1. Неклассического нейрона (подсистема "Анализ", режим "Графическое отображение нелокальных нейронов" (рисунок 98)).
2. Семантической сети признаков (подсистема "Типология", режим "Кластерный и конструктивный анализ признаков – вывод 2d-семантических сетей признаков" (рисунок 99)).

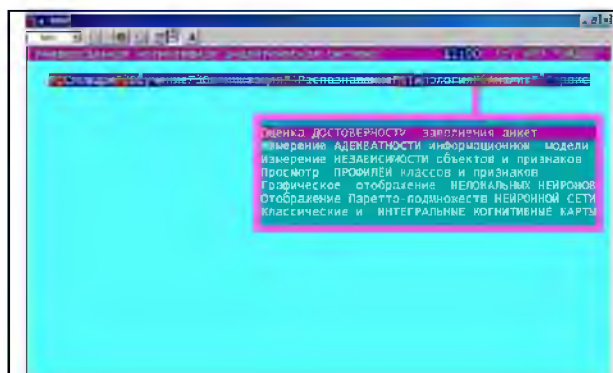


Рисунок 98. Задание режима отображения нелокальных нейронов

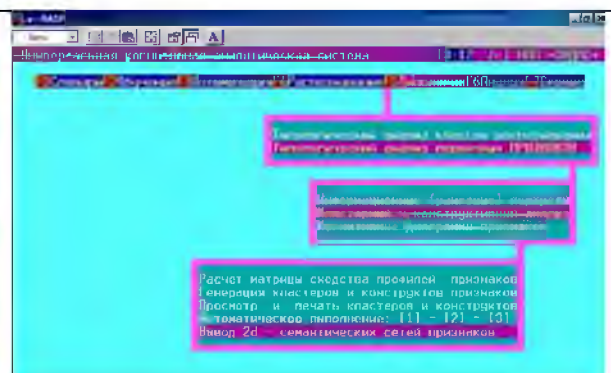


Рисунок 99. Задание режима отображения семантических сетей признаков

Результаты, т.е. когнитивные карты для составов, идущих на восток и запад, приведены на рисунках 100 и 101.

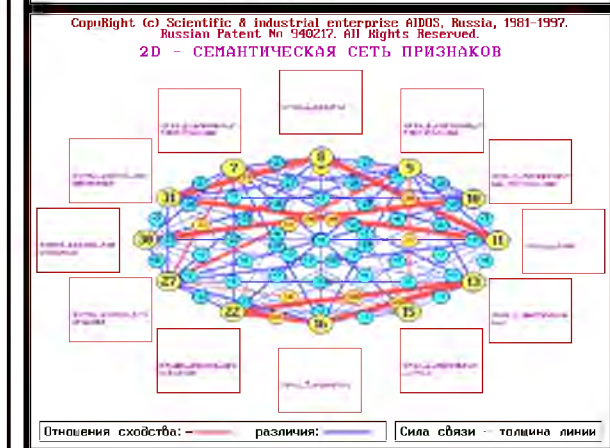
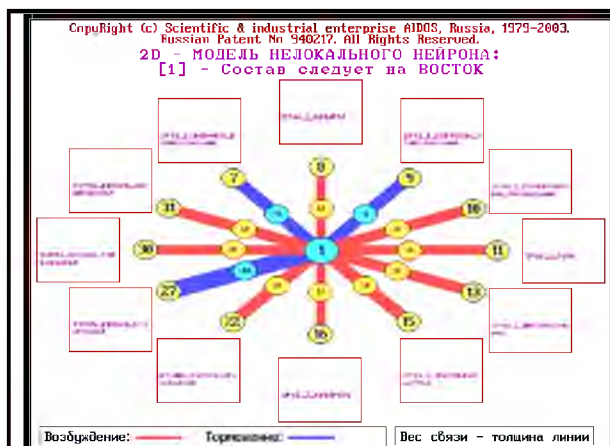


Рисунок 100. Когнитивная карта для составов, идущих на восток

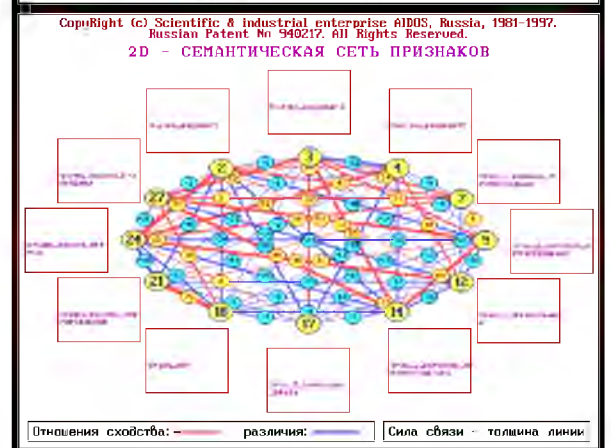
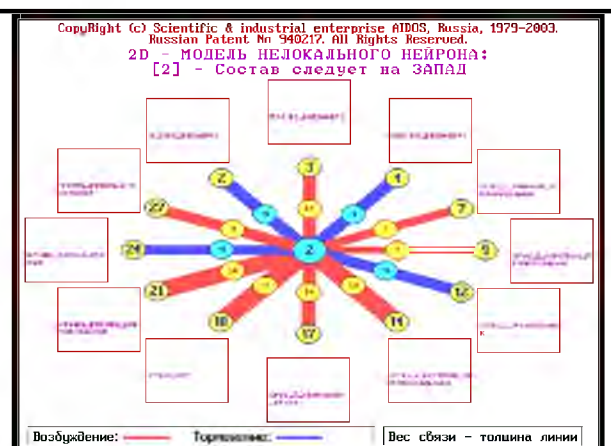


Рисунок 101. Когнитивная карта для составов, идущих на запад

Из этих рисунков видно, что классическая когнитивная карта может быть изображена в форме конуса, но для наглядности изображения большого объема информации его вершина и боковая поверхность показаны в форме нейрона, а основание – в форме семантической сети.

Выводы

Таким образом, вербальные описания объектов реальности на естественном языке с полным основанием могут рассматриваться как их иерархические лингвистические модели. Предложены методика и автоматизированная технология, основанные на универсальной когнитивной аналитической системе "Эйдос", которые обеспечивают:

- автоматизированную формализацию предметной области на основе вербального описания ее объектов;
- автоматизированное формирование описательных шкал и градаций;
- автоматизированную генерацию обучающей выборки;
- синтез семантической информационной модели, ее оптимизацию, проверку адекватности и анализ.

Предлагаемые технологии обеспечивают значительную экономию труда и времени по сравнению с традиционным подходом.

Необходимо отметить также, что в системе "Эйдос" реализована полнофункциональная интеллектуальная информационно-поисковая система, обеспечивающая работу по приведенной в работе технологии с объектами, описанными на естественном языке.

Контрольные вопросы

1. Что такое классификационные и описательные шкалы и градации?
2. Какие существуют виды шкал?
3. Как шкалы связаны с конструктами и с познанием?
4. Чем обусловлена возможность текстового описания объектов обучающей и распознаваемой выборки на естественном языке?

Литература по лабораторной работе

1. Мичи Д., Джонстон Р. Компьютер – творец. – М.: Мир, 1987. – 251 с.
2. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.
3. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). – Краснодар: КЮИ МВД РФ, 1996. – 280 с.

ЛР-2:**"Прогнозирование учебных достижений студентов на основе их имеджевых фотороботов"*****Краткая теория***

Задачами данной лабораторной работы являются:

1. Продемонстрировать студентам возможность выявления причинно-следственных связей между признаками внешнего вида (описательные шкалы и градации), и их полом, успеваемостью, тем, откуда они родом, обучением в той или иной группе (классификационные шкалы и градации).

2. Сформировать у студентов навыки формализации предметной области, подготовки и ввода обучающей выборки, синтеза информационной семантической модели и проверки ее адекватности, анализа модели (информационные портреты, кластерно-конструктивный анализ, семантические сети и когнитивные диаграммы, графическое отображение векторов классов и признаков).

Задание

1. Формализовать задачу:

- создать классификационные и описательные шкалы;
- собрать исходную фактографическую информацию и ввести в систему обучающую выборку.

2. Осуществить синтез и верификацию модели.

3. Оценить ценность признаков для прогнозирования. Выделить признаки, наиболее существенные для решения поставленной задачи.

4. Провести анализ модели, дав ответы на следующие вопросы:

- как посещаемость занятий по системам искусственного интеллекта влияет на успеваемость по этой дисциплине?
- как сказывается пол на посещаемости?
- как выглядят конструкторы "Пол", "Город-деревня", "Учебная группа", "Успеваемость", "Посещаемость"?
- какие студенты являются "типичными представителями" для своих учебных групп, а какие обладают своеобразием и выраженной индивидуальностью;

Результаты анализа отобразить в графической форме нелокальных нейронных и семантических сетей признаков. На их основе построить классические когнитивные карты для хорошо и плохо успевающих студентов.

Пример решения***Пример решения задания 1: Формализовать задачу:***

Пример решения задания 1.1: Формализовать задачу: создать классификационные и описательные шкалы

Один из вариантов классификационных шкал и градаций представлен в таблице 51, а описательных – в таблице 52:

Таблица 51 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ И ГРАДАЦИИ

Код	Наименование	Код	Наименование
1	ПОЛ - мужской	28	Нагапетян ПИ-51
2	ПОЛ - женский	29	Полонская ПИ-51
3	ОТКУДА РОДОМ - город-краевой центр	30	Трунина ПИ-51
4	ОТКУДА РОДОМ - город-районный центр	31	Черкашина ПИ-51
5	ОТКУДА РОДОМ - поселок городского типа	32	Чепурченко ПИ-51
6	ОТКУДА РОДОМ - село	33	Чушкин ПИ-51
7	УСПЕВАЕМОСТЬ - "5" более 75%	34	Шульгин ПИ-51
8	УСПЕВАЕМОСТЬ - "5" более 50% но меньше 75%	35	Арушанян ПИ-52
9	УСПЕВАЕМОСТЬ - "5" более 25% но меньше 50%	36	Быченко ПИ-52
10	УСПЕВАЕМОСТЬ - "5" менее 25%	37	Вережкина ПИ-52
11	ФОРМА ОБУЧЕНИЯ - бюджетная	38	Григорьева ПИ-52
12	ФОРМА ОБУЧЕНИЯ - платная	39	Давыдич ПИ-52
13	ГРУППА ПИ-51	40	Дронова ПИ-52
14	ГРУППА ПИ-52	41	Еременко ПИ-52
15	Бабенко ПИ-51	42	Жмурко ПИ-52
16	Воробьева ПИ-51	43	Иванова ПИ-52
17	Гура ПИ-51	44	Костенко ПИ-52
18	Головнев ПИ-51	45	Крейс ПИ-52
19	Дыбова ПИ-51	46	Куркина ПИ-52
20	Жеребятнев ПИ-51	47	Люлик ПИ-52
21	Заяц ПИ-51	48	Максимов ПИ-52
22	Иванова ПИ-51	49	Мануйлов ПИ-52
23	Котенко ПИ-51	50	Нарижний ПИ-52
24	Кузина О. ПИ-51	51	Ольховская ПИ-52
25	Кузина Я. ПИ-51	52	Паршакова ПИ-52
26	Лях ПИ-51	53	Силенко ПИ-52
27	Мясников ПИ-51	54	Соколова ПИ-52
		55	Турбин ПИ-52
		56	Цисарь ПИ-52

Таблица 52 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ

N	Код	Наименование			
	[1]	ДЛИНА ВОЛОС:		[5]	ОДЕЖДА:
1	1	Короткие.....	17	17	Пиджак.....
2	2	Средние.....	18	18	Брюки или джинсы.....
3	3	Длинные.....	19	19	Юбка.....
	[2]	ЦВЕТ ВОЛОС:		[6]	МАКИЯЖ, МАНИКЮР, ПЕДИКЮР:
4	4	Очень темные.....	20	20	Отсутствует.....
5	5	Русые.....	21	21	Незаметный.....
6	6	Каштановые.....	22	22	Заметный.....
7	7	Крашенные.....	23	23	Вызывающий.....
8	8	Очень светлые.....		[7]	ДОСТАТОК:
	[3]	ЦВЕТ ГЛАЗ:	24	24	Есть своя квартира.....
9	9	Серые.....	25	25	Есть автомобиль.....
10	10	Голубые.....	26	26	Есть компьютер.....
11	11	Карие.....	27	27	Есть мобильный телефон.....
12	12	Зеленые.....	28	28	Ничего нет.....
	[4]	НАЛИЧИЕ УКРАШЕНИЙ:		[8]	ПОСЕЩАЕМОСТЬ:
13	13	Цепочка.....	29	29	Очень плохая.....
14	14	Перстень.....	30	30	Плохая.....
15	15	Серьги.....	31	31	Средняя.....
16	16	Браслет.....	32	32	Хорошая.....
			33	33	Очень хорошая.....

Пример решения задания 1.2: Формализовать задачу: собрать исходную фактографическую информацию и ввести в систему обучающую выборку

С точки зрения методики организации занятия возможны различные варианты сбора и ввода информации обучающей выборки. Опыт проведения занятий по данной лабораторной работе показал, что с точки зрения *экономии времени* и обеспечения качества и единого вида обучающей выборки наиболее рациональным является вариант, приведенный ниже.

1. На доске чертится таблица вида 53. Это делается с таким расчетом, чтобы каждому присутствующему студенту группы соответствовала строка.

Таблица 53 – ШАБЛОН ДЛЯ ВВОДА ОБУЧАЮЩЕЙ ВЫБОРКИ

Код	Наименование	Классы						Признаки										
		1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11
1																		
2																		
3																		
4																		
5																		
6																		
7																		
8																		
9																		
10																		
11																		
12																		
13																		
14																		

2. Студентам дается задание с использованием классификационных и описательных шкал (таблицы 50 и 52) описать каждому *самого себя* и занести эту информацию в таблицу 53 на доске. При необходимости количество строк и столбцов в этой таблице можно увеличить.

В результате на доске появляется таблица вида 54. Видно, что по ряду студентов нет описательной информации, а классификационная – минимальна. Это связано с тем, что они *отсутствовали* на занятиях, когда проводилась данная лабораторная работа и не участвовали в формализации предметной области и подготовке обучающей выборки. У этих студентов данная работа не была зачтена. При этом для удобства кодирования информации о себе студентами с помощью Блокнота открываются файлы:

Object.txt и **Priz_per.txt** из поддиректории TXT, содержащие ту же информацию, что и таблицы 50 и 51.

Для правильного отображения этих файлов задается шрифт **Courier New** (при Windows-98) **Terminal** (при Windows-2000 и Windows^{XP})

3. Студентам дается задание переписать эту таблицу себе в тетради и затем ввести в систему "Эйдос" в подсистеме "F2 Обучение" в режиме "Ввод-корректировка обучающей выборки".

**Таблица 54 – ПРИМЕР ЗАПОЛНЕННОГО ШАБЛОНА
ДЛЯ ВВОДА ОБУЧАЮЩЕЙ ВЫБОРКИ**

Код	Наименование	Классы						Признаки										
		1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11
1	Бабенко ПИ-51	1	4	10	12	13	15	1	5	10	18	27	31					
2	Воробьева ПИ-51	2	5	10	12	13	16	3	7	12	13	18	20	25	26	27	32	
3	Гура ПИ-51	1	13	17														
4	Гловнев ПИ-51	1	13	18														
5	Дыбова ПИ-51	2	13	19														
6	Жеребятьев ПИ51	1	5	8	12	13	20	2	8	10	18	21	27	30				
7	Заяц ПИ-51	2	4	10	12	13	21	3	6	12	15	18	21	26	27	31		
8	Иванова ПИ-51	2	3	7	12	13	22	2	5	12	13	14	15	18	21	26	32	
9	Котенко ПИ-51	2	4	7	12	13	23	1	6	11	13	15	18	21	26	27	32	
10	Кузина О. ПИ-51	2	3	8	12	13	24	1	7	11	14	18	22	26	27	32		
11	Кузина Я. ПИ-51	2	3	8	12	13	25	1	7	14	17	22	32					
12	Лях ПИ-51	1	3	8	12	13	26	1	5	9	13	17	18	20	24	26	27	31
13	Мясников ПИ-51	1	3	9	12	13	27	1	5	10	13	16	18	20	24	27	32	0
14	Нагапетян ПИ-51	1	4	9	12	13	28	1	6	11	15	18	21	24	26	27	30	0
15	Полонская ПИ-51	2	13	29														
16	Трунина ПИ-51	2	13	30														
17	Черкашина ПИ-51	2	4	10	12	13	31	3	6	11	13	18	21	24	26	27	30	
18	Чепурченко ПИ51	1	13	32														
19	Чушкин ПИ-51	1	13	33														
20	Шульгин ПИ-51	1	5	8	12	13	34	1	5	9	17	20	26	30				
21	Арушунян ПИ-52	1	14	35														
22	Быченко ПИ-52	1	14	36														
23	Веревкина ПИ-52	2	3	9	12	14	37	1	8	9	15	18	21	28	33			
24	Григорьева ПИ52	2	4	8	12	14	38	2	5	9	13	18	22	28	33			
25	Давыдич ПИ-52	2	14	39														
26	Дронова ПИ-52	2	14	40														
27	Еременко ПИ-52	2	4	8	12	14	41	2	5	12	9	13	15	18	21	26	27	31
28	Жмурко ПИ-52	1	14	42														
29	Иванова ПИ-52	2	3	9	12	14	43	2	4	10	14	15	17	18	20	28	33	
30	Костенко ПИ-52	2	14	44														
31	Крейс ПИ-52	2	4	8	12	14	45	1	5	9	15	19	21	28	32			
32	Куркина ПИ-52	2	3	8	12	14	46	3	6	11	13	15	18	19	22	26	27	32
33	Люлик ПИ-52	2	5	8	12	14	47	2	7	12	15	18	22	24	26	32		
34	Максимов ПИ-52	1	14	48														
35	Мануйлов ПИ-52	1	3	7	12	14	49	1	5	9	14	18	20	24	25	26	27	30
36	Нарижный ПИ-52	1	3	7	12	14	50	1	5	9	14	18	20	24	25	26	27	31
37	Ольховская ПИ52	2	14	51														
38	Паршакова ПИ-52	2	6	8	12	14	52	2	6	11	13	19	22	26	32			
39	Силенко ПИ-52	1	3	7	12	14	53	1	6	12	13	18	20	28	32			
40	Соколова ПИ-52	2	4	8	12	14	54	1	5	12	17	18	22	24	26	27	32	
41	Турбин ПИ-52	1	14	55														
42	Цисарь ПИ-52	2	5	9	12	14	56	2	7	11	13	18	22	27	33			

Пример решения задания 2: Осуществить синтез и верификацию модели

Верификацию модели предлагается проверить путем расчета внутренней дифференциальной и интегральной валидности.

Для этого студентам необходимо выполнить следующие действия:

1. Скопировать обучающую выборку в распознаваемую в подсистеме: "F2 Обучение – ввод корректировка обучающей информации – F5 Об.инф.->Расп.анк. – F2 Перезапись БД распознаваемых анкет – F1 Копировать всю БД".
2. Выполнить пакетное распознавание в подсистеме: "F4 Распознавание – Пакетное распознавание – Критерий сходства [81] корреляция".
3. Измерить внутреннюю интегральную и дифференциальную валидность информационной модели в подсистеме: "F6 Анализ".

Пример решения задания 3: Оценить ценность признаков для прогнозирования. Выделить признаки, наиболее существенные для решения поставленной задачи.

В подсистеме: "F3 Оптимизация" выбрать режим: "Исключение признаков с низкой селективной силой". В результате появится экранная форма, представленная на рисунке 102:

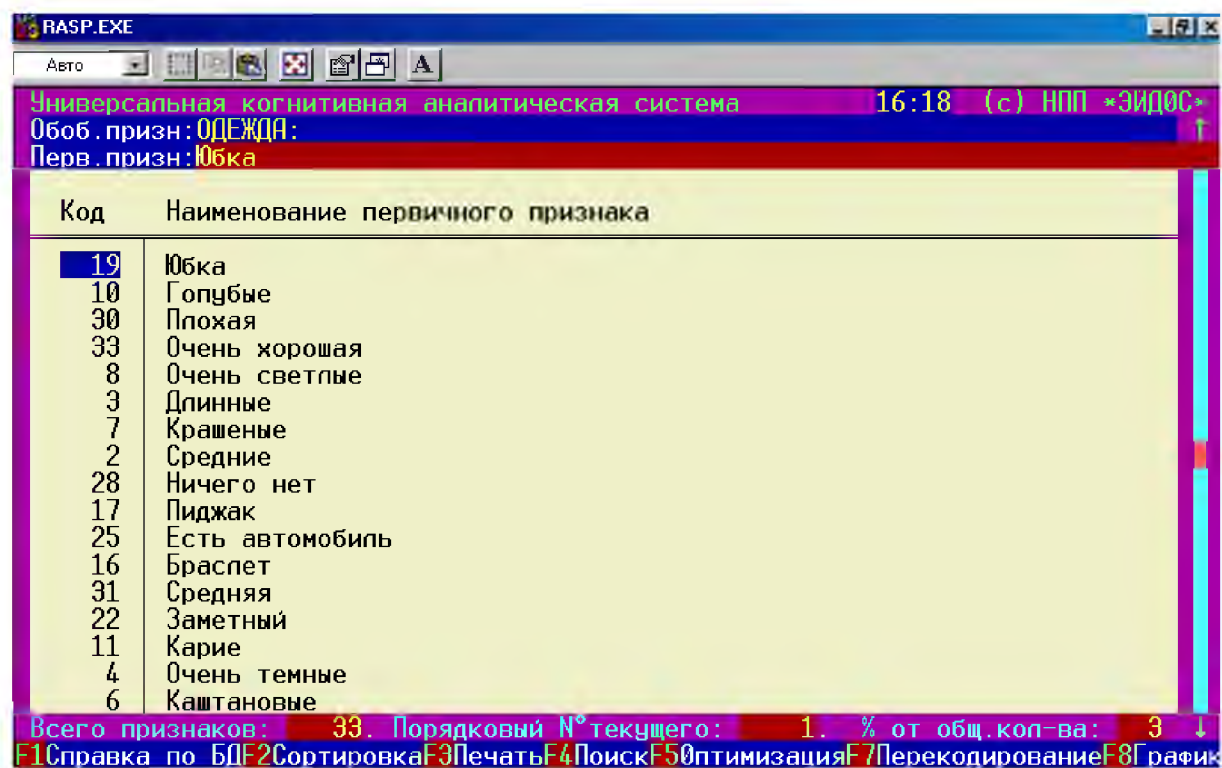


Рисунок 102. Экранная форма "Селективная сила признаков"

Количественные значения селективной силы можно просмотреть, прокручивая экранную форму по горизонтали.

Более подробно эта информация представлена в таблице 55:

**Таблица 55 – ПРИЗНАКИ (ГРАДАЦИИ ШКАЛ)
В ПОРЯДКЕ УБЫВАНИЯ СЕЛЕКТИВНОЙ СИЛЫ (ИНТ.ИНФ.)**

10-10-04

16:15:55

г. Краснодар

N°	Код	Код	Наименование	Наименование	Интегр.	Сумм. инт	Сумм. ин
	К.шк	град	классификационной шкалы	градации	инф-ть	инф. Бит	в %
1	5	19	ОДЕЖДА:.....	Юбка.....	0.470	0.470	4.147
2	3	10	ЦВЕТ ГЛАЗ:.....	Голубые.....	0.467	0.938	8.271
3	8	30	ПОСЕЩАЕМОСТЬ:.....	Плохая.....	0.456	1.393	12.290
4	8	33	ПОСЕЩАЕМОСТЬ:.....	Очень хорошая.....	0.438	1.831	16.150
5	2	8	ЦВЕТ ВОЛОС:.....	Очень светлые.....	0.437	2.268	20.002
6	1	3	ДЛИНА ВОЛОС:.....	Длинные.....	0.429	2.697	23.788
7	2	7	ЦВЕТ ВОЛОС:.....	Крашенные.....	0.421	3.118	27.502
8	1	2	ДЛИНА ВОЛОС:.....	Средние.....	0.418	3.536	31.192
9	7	28	ДОСТАТОК:.....	Ничего нет.....	0.413	3.949	34.836
10	5	17	ОДЕЖДА:.....	Пиджак.....	0.408	4.357	38.432
11	7	25	ДОСТАТОК:.....	Есть автомобиль.....	0.408	4.765	42.027
12	4	16	НАЛИЧИЕ УКРАШЕНИЙ:.....	Браслет.....	0.404	5.169	45.590
13	8	31	ПОСЕЩАЕМОСТЬ:.....	Средняя.....	0.400	5.569	49.122
14	6	22	МАКИЯЖ, МАНИКЮР, ПЕДИКЮР:	Заметный.....	0.399	5.968	52.638
15	3	11	ЦВЕТ ГЛАЗ:.....	Карие.....	0.398	6.365	56.147
16	2	4	ЦВЕТ ВОЛОС:.....	Очень темные.....	0.391	6.757	59.599
17	2	6	ЦВЕТ ВОЛОС:.....	Каштановые.....	0.389	7.146	63.032
18	4	14	НАЛИЧИЕ УКРАШЕНИЙ:.....	Перстень.....	0.381	7.527	66.395
19	3	12	ЦВЕТ ГЛАЗ:.....	Зеленые.....	0.367	7.894	69.629
20	6	20	МАКИЯЖ, МАНИКЮР, ПЕДИКЮР:	Отсутствует.....	0.361	8.255	72.817
21	3	9	ЦВЕТ ГЛАЗ:.....	Серые.....	0.356	8.611	75.958
22	6	21	МАКИЯЖ, МАНИКЮР, ПЕДИКЮР:	Незаметный.....	0.352	8.963	79.062
23	4	15	НАЛИЧИЕ УКРАШЕНИЙ:.....	Серьги.....	0.337	9.300	82.035
24	8	32	ПОСЕЩАЕМОСТЬ:.....	Хорошая.....	0.324	9.625	84.897
25	2	5	ЦВЕТ ВОЛОС:.....	Русые.....	0.322	9.947	87.739
26	7	24	ДОСТАТОК:.....	Есть своя квартира.....	0.318	10.265	90.545
27	1	1	ДЛИНА ВОЛОС:.....	Короткие.....	0.293	10.558	93.130
28	4	13	НАЛИЧИЕ УКРАШЕНИЙ:.....	Цепочка.....	0.264	10.822	95.458
29	7	26	ДОСТАТОК:.....	Есть компьютер.....	0.218	11.041	97.384
30	7	27	ДОСТАТОК:.....	Есть мобильный телефон.....	0.194	11.235	99.097
31	5	18	ОДЕЖДА:.....	Брюки или джинсы.....	0.102	11.337	100.000
32	6	23	МАКИЯЖ, МАНИКЮР, ПЕДИКЮР:	Вызывающий.....	0.000	11.337	100.000
33	8	29	ПОСЕЩАЕМОСТЬ:.....	Очень плохая.....	0.000	11.337	100.000

Универсальная когнитивная аналитическая система

НПП *Эйдос*

Соответствующую логистическую кривую (накопительную селективной силы) можно получить в графической форме нажав на клавишу "F8 График".

Пример решения задания 4: Провести анализ модели

Анализ модели позволяет дать обоснованные ответы, например, на следующие вопросы:

1. Как посещаемость занятий по системам искусственного интеллекта влияет на успеваемость по этой дисциплине?
2. Как сказывается пол на посещаемости?
3. Как выглядят конструкты "Пол", "Город-деревня", "Учебная группа", "Успеваемость", "Посещаемость"?
4. Какие студенты являются "типичными представителями" для своих учебных групп, а какие обладают своеобразием и выраженной индивидуальностью?

Результаты анализа отобразить в графической форме нелокальных нейрон и семантических сетей признаков. На их основе построить классические когнитивные карты для хорошо и плохо успевающих студентов.

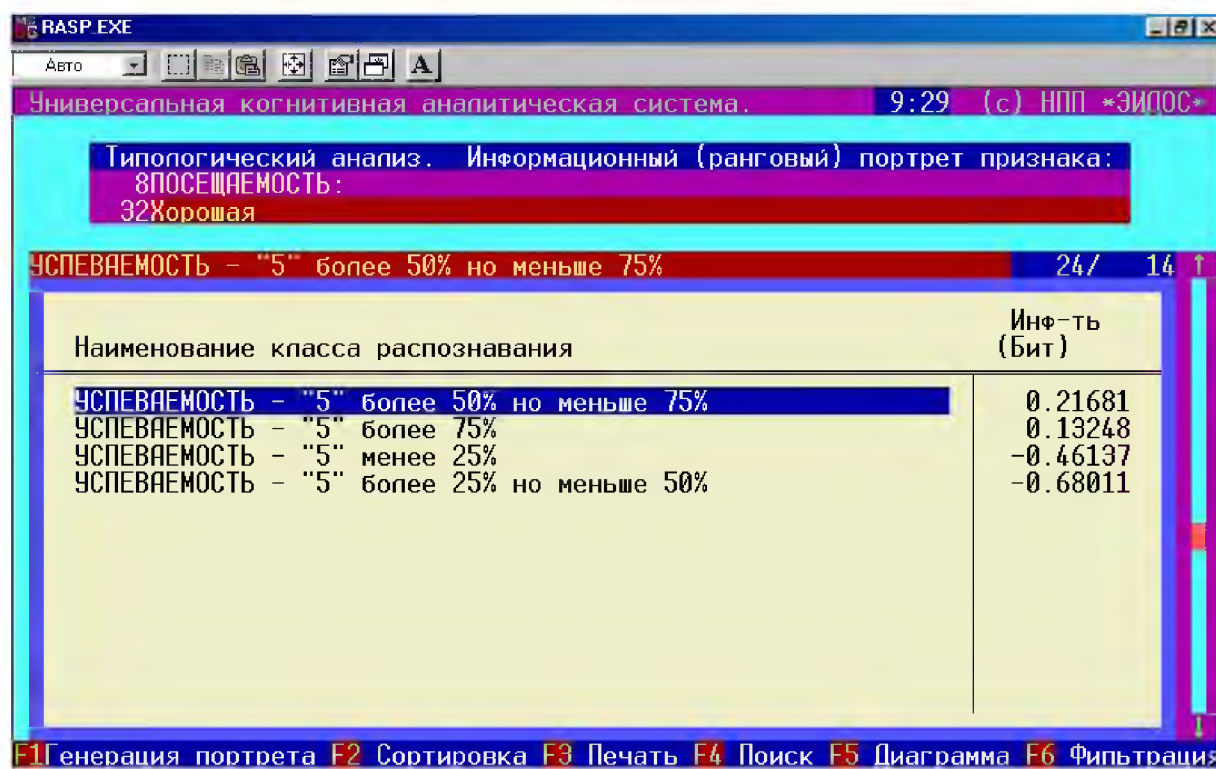
Пример решения задания 4.1: Как посещаемость занятий по системам искусственного интеллекта влияет на успеваемость по этой дисциплине?

Для ответа на этот вопрос войдем в подсистему: "Типология" режим: "Типологический анализ первичных признаков – Информационные (ранговые) портреты" и установив курсор на строке с кодом 30 "Посещаемость плохая" нажмем Enter. В результате в экранной форме отобразиться информационный портрет признака, т.е. состояния студентов, расположенные в порядке убывания значимости данного признака для их наступления. Однако в этой экранной форме приведены все состояния, а нас интересует только успеваемость. Успеваемость – это классификационная шкала, с кодами градаций от 7 до 10. Поэтому, чтобы отобразить только эти классы, нажмем клавиши: "F6 Фильтрация", затем "F1 Диапазон кодов" и зададим коды 7 – 10. В результате получим экранную форму, приведенную на рисунке 103:

Наименование класса распознавания	Инф-ть (Бит)
УСПЕВАЕМОСТЬ - "5" менее 25%	0.23936
УСПЕВАЕМОСТЬ - "5" более 25% но меньше 50%	0.02061
УСПЕВАЕМОСТЬ - "5" более 75%	-0.04613
УСПЕВАЕМОСТЬ - "5" более 50% но меньше 75%	-0.08518

Рисунок 103. Информационный портрет признака: 30 "Посещаемость плохая", фильтр по успеваемости.

Аналогично получим экранную форму для информационного портрета признака: 32 "Посещаемость хорошая", представленную на рисунке 104.



**Рисунок 104. Информационный портрет признака:
32 "Посещаемость хорошая", фильтр по успеваемости.**

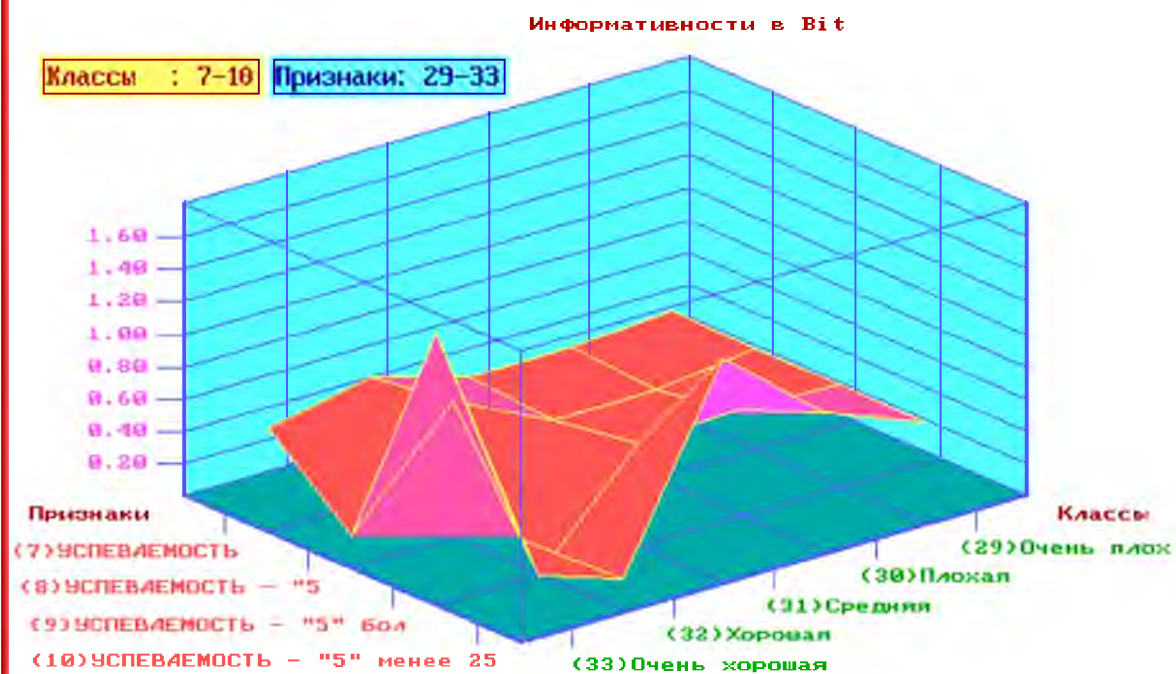
Из этих рисунков видно, что *плохой посещаемости соответствует низкая успеваемость, а хорошей – высокая, как и можно было ожидать*. НО теперь этот вывод получен непосредственно на основе анализа *эмпирических данных* и является *количественным*.

Эту же информацию можно получить и другим способом.

В подсистеме F6 "Анализ" запустить режим "Просмотр профилей классов и признаков" и нажать функциональную клавишу F2 "Выбор статистической матрицы" F3 "Загрузить матрицу информативностей". Затем нажать F5 "Профили 3d", F1 "Трехмерная поверхность" и задать диапазон кодов классов: 7-10 (классификационная шкала: "Успеваемость") и диапазон кодов признаков 29-33 (классификационная шкала: "Посещаемость"). После этого появится возможность настройки ориентации трехмерной системы координат, описанная в Help в системе. Выход из настройки по клавише Esc. После этого в экранных формах в виде трехмерных графиков последовательно отображаются информативности выбранной подматрицы, которые имеется возможность записать в форме РСХ-файлов, пропустить или распечатать. Некоторые из этих форм представлены на рисунках 105 – 106.

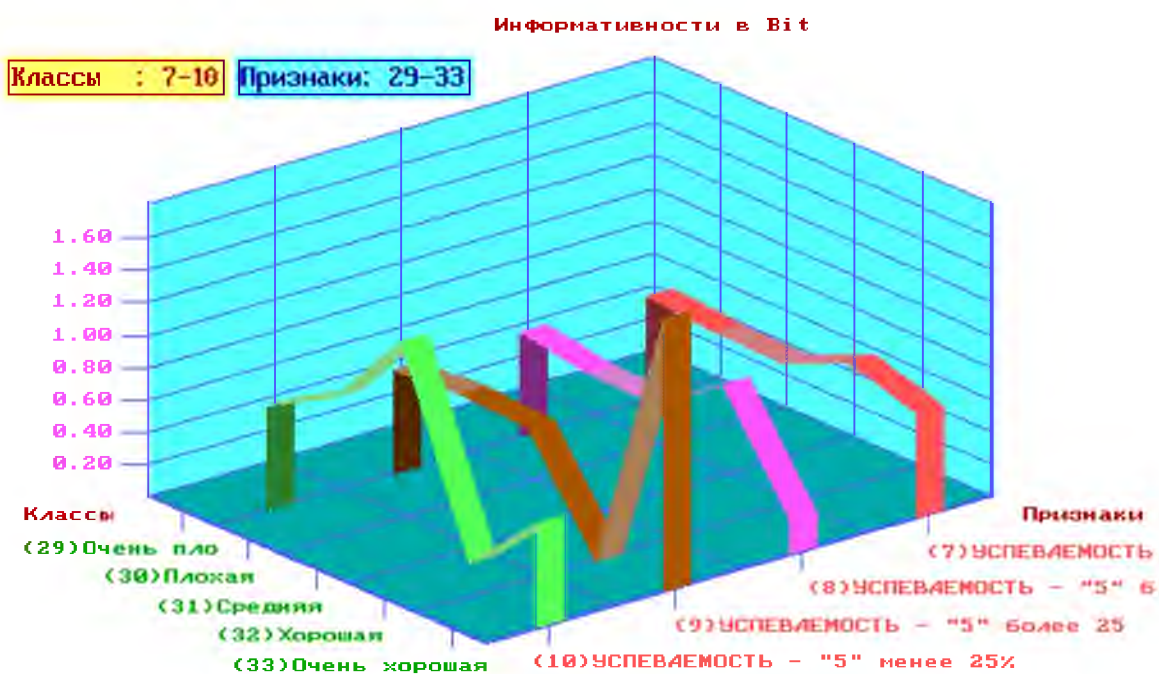
Copyright (c) Scientific & industrial enterprise AIDS, Russia, 1981-2001.
Russian Patent No 940217. All Rights Reserved.

ПРОФИЛИ КЛАССОВ РАСПОЗНАВАНИЯ И ПРИЗНАКОВ



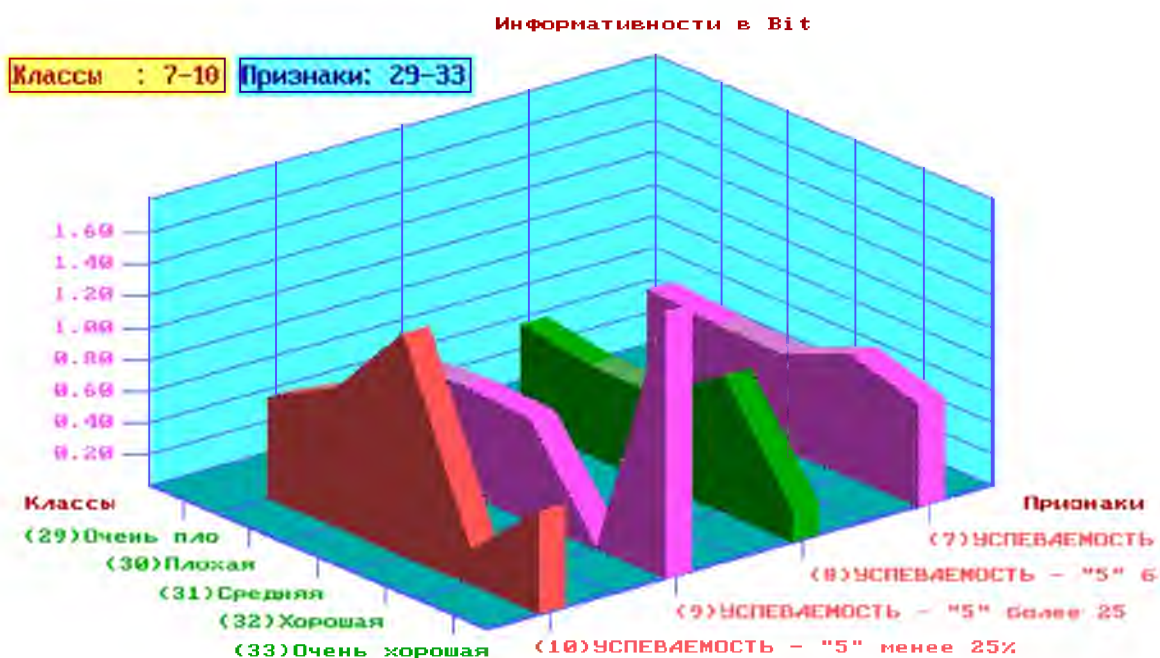
Copyright (c) Scientific & industrial enterprise AIDS, Russia, 1981-2001.
Russian Patent No 940217. All Rights Reserved.

ПРОФИЛИ КЛАССОВ РАСПОЗНАВАНИЯ И ПРИЗНАКОВ



CopyRight (c) Scientific & industrial enterprise AIDOS, Russia, 1981-2001.
Russian Patent No 940217. All Rights Reserved.

ПРОФИЛИ КЛАССОВ РАСПОЗНАВАНИЯ И ПРИЗНАКОВ



CopyRight (c) Scientific & industrial enterprise AIDOS, Russia, 1981-2001.
Russian Patent No 940217. All Rights Reserved.

ПРОФИЛИ КЛАССОВ РАСПОЗНАВАНИЯ И ПРИЗНАКОВ

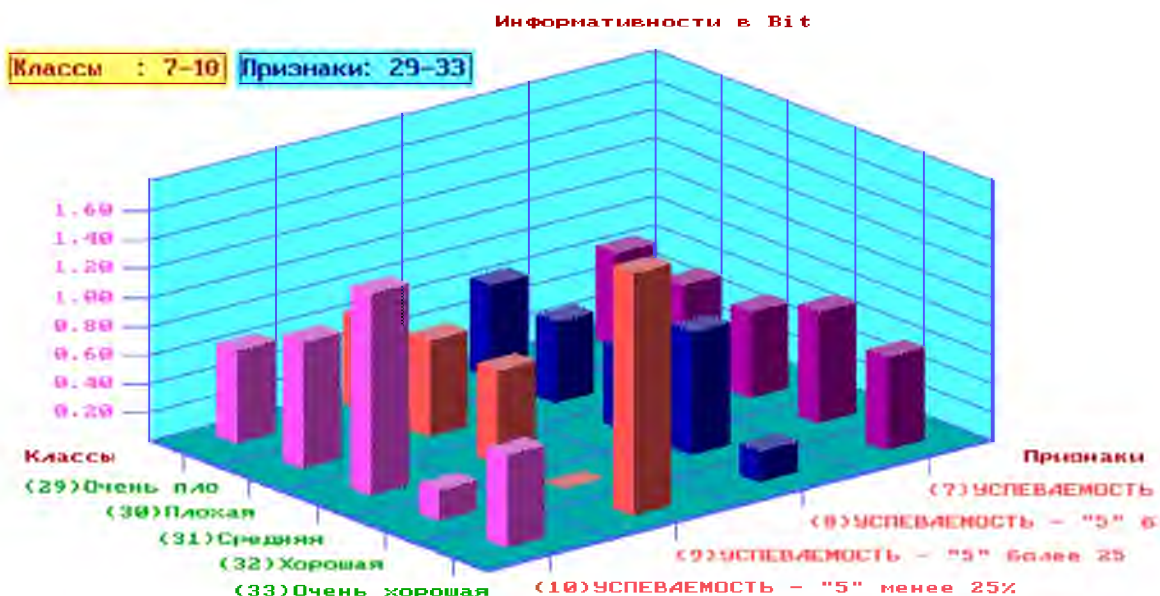


Рисунок 105. Некоторые варианты отображения подматрицы:
классификационная шкала "Успеваемость"
× описательная шкала "Посещаемость" в системе "Эйдос"

Из анализа этих трехмерных графиков видно, что очень хорошая успеваемость слабо зависит от посещаемости, а низкая – сильно. На основании этого можно сформулировать гипотезу, что хорошая посещаемость может гарантировать от низкой успеваемости, но сама по себе не гарантирует высокой, для достижения которой, видимо, играют важную роль также другие факторы, например, уровень достатка. Чтобы проверить эту гипотезу в подсистеме F6 "Анализ" запустим режим "Просмотр профилей классов и признаков" и поставим курсор на класс: код 7 "Успеваемость – 5 более 75%", а затем нажмем F4 "Профили 2d", F1 "Класс распознавания" и зададим диапазон признаков, градаций описательных шкал: 24-33 (шкалы: "Достаток" и "Посещаемость"). В результате получим графическую форму, представленная на рисунке 106:

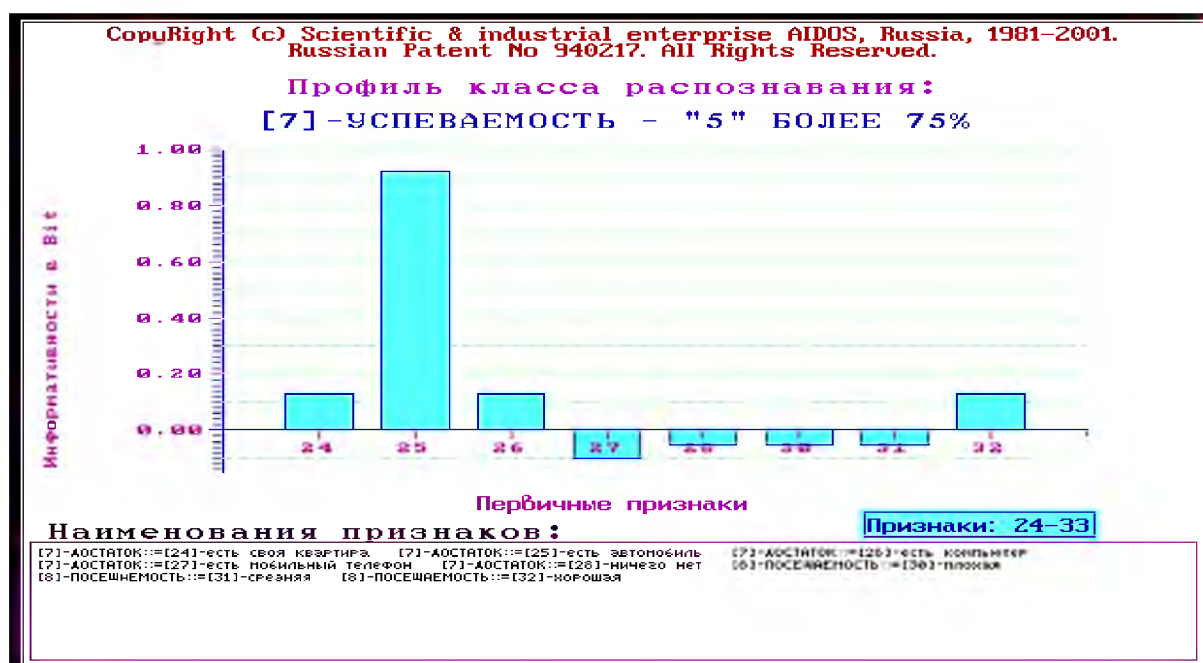


Рисунок 106. Профиль влияния описательных шкал: "Достаток" и "Посещаемость" на классификационную шкалу "Успеваемость"

Из этой формы видно, что хорошая посещаемость (код 32) положительно влияет на высокую успеваемость примерно также, как наличие квартиры (код 24) и компьютера (код 26), но все эти факторы вместе взятые уступают влиянию одного фактора: "Есть автомобиль" (код 25).

Пример решения задания 4.2: Как сказывается пол на посещаемости?

В подсистеме: "Типология" зайдём в режим: "Типологический анализ классов распознавания – Информационные (ранговые) портреты" и установив курсор на строке с кодом 1 "ПОЛ - мужской" нажмем Enter. В результате в экранной форме отобразится информационный портрет класса, т.е. признаки студентов, расположенные в порядке убывания характерности признаков для данного класса. Однако в этой экранной форме приведены все признаки, а нас интересует только посещаемость. Посещаемость – это описательная шкала, с кодами градаций от 29 до 33. Поэтому, чтобы отобразить только эти признаки, нажмем клавиши: "F6 Фильтрация", затем "F1 Диапазон кодов" и зададим коды 29 – 33. В результате получим экранную форму, приведенную на рисунке 107.

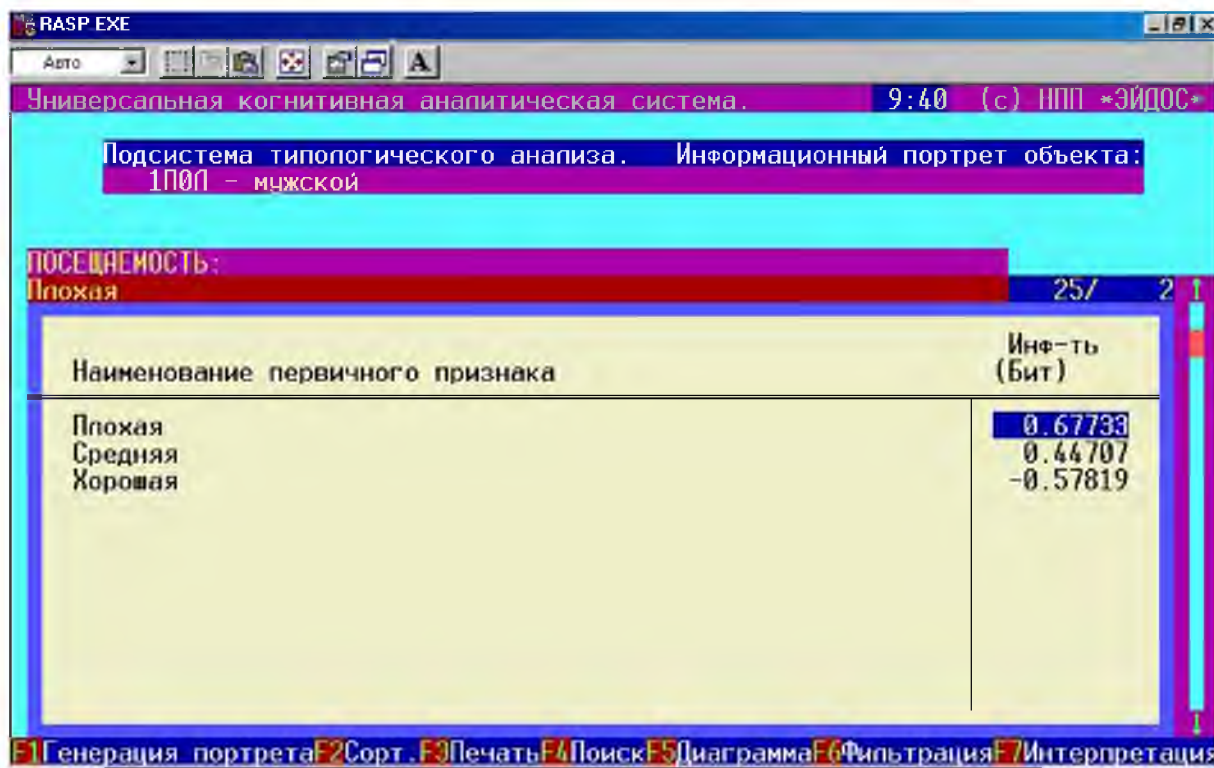


Рисунок 107. Информационный портрет класса:
1 "ПОЛ - мужской", фильтр по посещаемости.

Аналогично получим информационный портрет класса: код 2, "ПОЛ – женский". Результат представлен на рисунке 108.

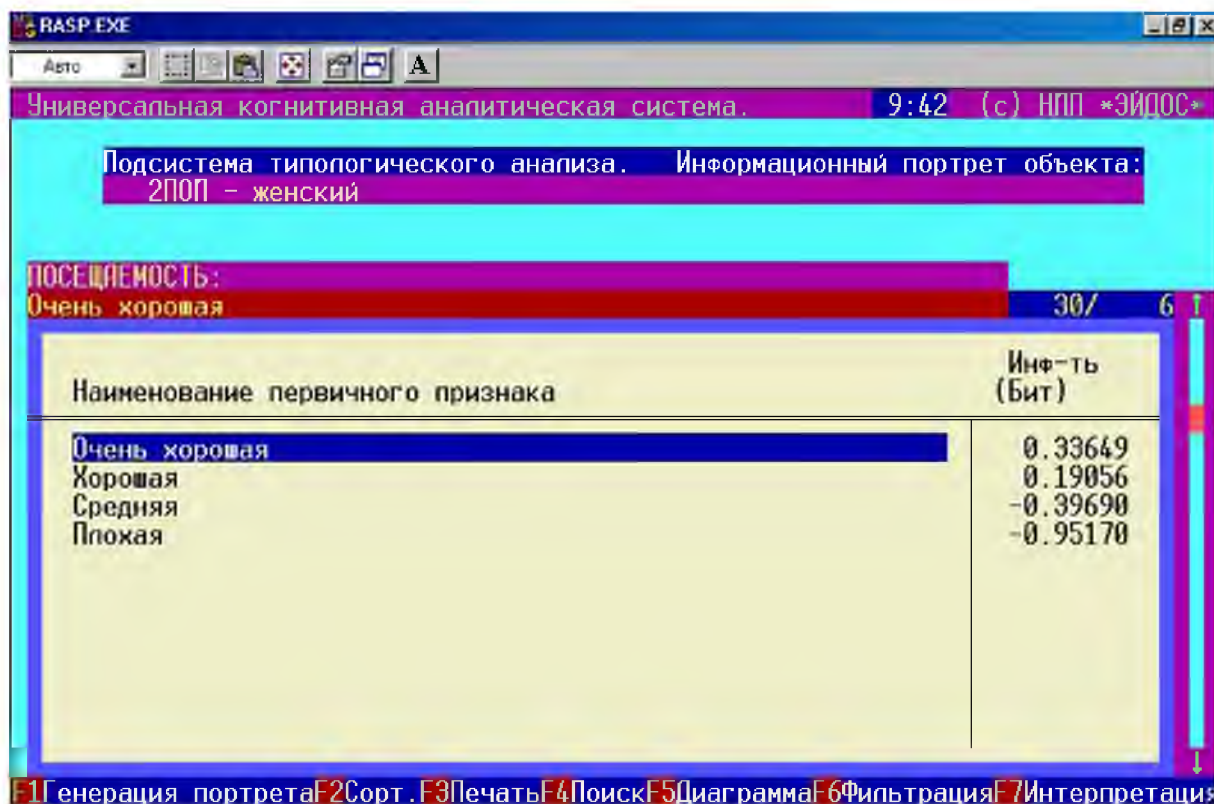


Рисунок 108. Информационный портрет класса:
код 2 "ПОЛ – женский", фильтр по посещаемости.

Из этих рисунков видно, что для ребят характерны плохая и средняя посещаемость, а для девушек – очень хорошая и хорошая. Как это сказывается на успеваемости мы уже знаем.

Пример решения задания 4.3: Как выглядят конструкторы "Пол", "Город-деревня", "Учебная группа", "Посещаемость"?

Ответ на первые три вопроса дает анализ, который можно провести в подсистеме F5 "Типология", "Типологический анализ классов", "Кластерный и конструктивный анализ", "Автоматическое выполнение: [1]-[2]-[3]".

После этого выполним режим: "Генерация кластеров и конструкторов классов", при следующих параметрах:

- количество уровней кластеризации – 1;
- порог модуля сходства – 0;
- диапазон кодов классов: 1-14;
- исключаем кластеры с одним классом;
- не исключаем кластеры с тождественным набором классов.

Результаты кластерно-конструктивного анализа классов можно просмотреть в режиме: "Просмотр и печать кластеров и конструкторов классов" (рисунок 109), а также получить в виде ТХТ-файла с именем **Klas_obj.txt** в поддиректории ТХТ (таблица 56):

Ур. кл.	№ клас	Ур. Мерп	Код класса	Наименование класса	% сходст
1	1	0	1	ПОП - мужской	100.00
1	1	0	3	ОТКУДА РОДОМ - город-краевой центр	38.64
1	1	0	13	ГРУППА ПИ-51	35.39
1	1	0	7	УСПЕВАЕМОСТЬ - "5" более 75%	33.46
1	1	0	9	УСПЕВАЕМОСТЬ - "5" более 25% но меньше 50%	16.24
1	1	0	10	УСПЕВАЕМОСТЬ - "5" менее 25%	10.76
1	1	0	5	ОТКУДА РОДОМ - поселок городского типа	9.69
1	1	0	4	ОТКУДА РОДОМ - город-районный центр	-8.19
1	1	0	8	УСПЕВАЕМОСТЬ - "5" более 50% но меньше 75%	-14.60
1	1	0	14	ГРУППА ПИ-52	-28.51
1	1	0	6	ОТКУДА РОДОМ - село	-38.49
1	1	0	2	ПОП - женский	-71.70
1	0	0	0		0.00

F1Помощь F2Удал. уровней F3Печать F4Поиск F52d-диагр. F8Генерация классов по кластерам

Рисунок 109. Конструктор: "Пол"

Таблица 56 – КОНСТРУКТЫ И КЛАСТЕРЫ КЛАССОВ РАСПОЗНАВАНИЯ

15-10-04 10:07:48

Фильтр по кодам классов: 1-14

г.Краснодар

№ класт	Код класса	Наименование класса распознавания	Сход- ство %
1	1	ПОЛ - мужской.....	100.00
	3	ОТКУДА РОДОМ - город-краевой центр.....	38.64
	13	ГРУППА ПИ-51.....	35.39
	7	УСПЕВАЕМОСТЬ - "5" более 75%.....	33.46
	9	УСПЕВАЕМОСТЬ - "5" более 25% но меньше 50%.	16.24
	10	УСПЕВАЕМОСТЬ - "5" менее 25%.....	10.76
	5	ОТКУДА РОДОМ - поселок городского типа.....	9.69
	4	ОТКУДА РОДОМ - город-районный центр.....	-8.19
	8	УСПЕВАЕМОСТЬ - "5" более 50% но меньше 75%.	-14.60
	14	ГРУППА ПИ-52.....	-28.51
	6	ОТКУДА РОДОМ - село.....	-38.49
	2	ПОЛ - женский.....	-71.70
3	3	ОТКУДА РОДОМ - город-краевой центр.....	100.00
	7	УСПЕВАЕМОСТЬ - "5" более 75%.....	45.92
	1	ПОЛ - мужской.....	38.64
	9	УСПЕВАЕМОСТЬ - "5" более 25% но меньше 50%.	38.46
	14	ГРУППА ПИ-52.....	38.29
	13	ГРУППА ПИ-51.....	0.35
	8	УСПЕВАЕМОСТЬ - "5" более 50% но меньше 75%.	-3.78
	2	ПОЛ - женский.....	-12.93
	5	ОТКУДА РОДОМ - поселок городского типа.....	-14.76
	4	ОТКУДА РОДОМ - город-районный центр.....	-36.40
	10	УСПЕВАЕМОСТЬ - "5" менее 25%.....	-36.68
	6	ОТКУДА РОДОМ - село.....	-36.95
13	13	ГРУППА ПИ-51.....	100.00
	1	ПОЛ - мужской.....	35.39
	9	УСПЕВАЕМОСТЬ - "5" более 25% но меньше 50%.	27.98
	4	ОТКУДА РОДОМ - город-районный центр.....	25.70
	10	УСПЕВАЕМОСТЬ - "5" менее 25%.....	25.08
	3	ОТКУДА РОДОМ - город-краевой центр.....	0.35
	5	ОТКУДА РОДОМ - поселок городского типа.....	-1.78
	7	УСПЕВАЕМОСТЬ - "5" более 75%.....	-10.67
	2	ПОЛ - женский.....	-21.13
	6	ОТКУДА РОДОМ - село.....	-26.13
	8	УСПЕВАЕМОСТЬ - "5" более 50% но меньше 75%.	-43.88
	14	ГРУППА ПИ-52.....	-64.51

Универсальная когнитивная аналитическая система

НПП *ЭЙДОС*

Конструкт: "Посещаемость" является конструктом градаций описательных шкал, т.е. признаков. Он генерируется в подсистеме F5 "Типология", "Типологический анализ признаков", "Кластерный и конструктивный анализ", "Автоматическое выполнение: [1]-[2]-[3]".

После этого выполним режим: "Генерация кластеров и конструктов признаков", при следующих параметрах:

- количество уровней кластеризации – 1;
- порог модуля сходства – 0;
- диапазон кодов классов: 1-33;
- исключаем кластеры с одним признаком;

- не исключаем кластеры с тождественным набором признаков.

Результаты кластерно-конструктивного анализа признаков можно получить в виде TXT-файла с именем **Klas_prp.txt** в поддиректории TXT (таблица 57):

Таблица 57 – КОНСТРУКТ ПРИЗНАКОВ: "ПОСЕЩАЕМОСТЬ"					
15-10-04 10:31:44 Фильтр по кодам: 1-33			17	Пиджак.....	16.77
			[2]	ЦВЕТ ВОЛОС:	
Коды	Наименования описательных признаков и градаций (признаков)	Сходство %	5	Русые.....	15.52
приз-ов	шكال и градаций (признаков)	ство %	[3]	ЦВЕТ ГЛАЗ:	
			11	Карие.....	12.19
[8]	ПОСЕЩАЕМОСТЬ:		[8]	ПОСЕЩАЕМОСТЬ:	
30	Плохая.....	100.00	31	Средняя.....	3.48
[7]	ДОСТАТОК:		[1]	ДЛИНА ВОЛОС:	
27	Есть мобильный телефон....	42.53	2	Средние.....	3.15
[7]	ДОСТАТОК:		[4]	НАЛИЧИЕ УКРАШЕНИЙ:	
24	Есть своя квартира.....	35.49	14	Перстень.....	-1.44
[6]	МАКИЯЖ, МАНИКЮР, ПЕДИКЮР:		[4]	НАЛИЧИЕ УКРАШЕНИЙ:	
21	Незаметный.....	33.92	16	Браслет.....	-2.47
[2]	ЦВЕТ ВОЛОС:		[2]	ЦВЕТ ВОЛОС:	
8	Очень светлые.....	33.72	7	Крашенные.....	-3.16
[3]	ЦВЕТ ГЛАЗ:		[4]	НАЛИЧИЕ УКРАШЕНИЙ:	
10	Голубые.....	32.63	13	Цепочка.....	-6.03
[6]	МАКИЯЖ, МАНИКЮР, ПЕДИКЮР:		[3]	ЦВЕТ ГЛАЗ:	
20	Отсутствует.....	31.99	12	Зеленые.....	-10.96
[7]	ДОСТАТОК:		[8]	ПОСЕЩАЕМОСТЬ:	
26	Есть компьютер.....	28.63	33	Очень хорошая.....	-13.16
[1]	ДЛИНА ВОЛОС:		[5]	ОДЕЖДА:	
3	Длинные.....	19.72	19	Юбка.....	-13.38
[2]	ЦВЕТ ВОЛОС:		[4]	НАЛИЧИЕ УКРАШЕНИЙ:	
6	Каштановые.....	19.15	15	Серьги.....	-13.91
[7]	ДОСТАТОК:		[6]	МАКИЯЖ, МАНИКЮР, ПЕДИКЮР:	
25	Есть автомобиль.....	18.43	22	Заметный.....	-16.05
[5]	ОДЕЖДА:		[2]	ЦВЕТ ВОЛОС:	
18	Брюки или джинсы.....	18.27	4	Очень темные.....	-17.52
[3]	ЦВЕТ ГЛАЗ:		[7]	ДОСТАТОК:	
9	Серые.....	17.53	28	Ничего нет.....	-18.06
[1]	ДЛИНА ВОЛОС:		[8]	ПОСЕЩАЕМОСТЬ:	
1	Короткие.....	17.11	32	Хорошая.....	-19.54
[5]	ОДЕЖДА:				

Пример решения задания 4.4: Какие студенты являются "типичными представителями" для своих учебных групп, а какие обладают своеобразием и выраженной индивидуальностью?

Чтобы ответить на этот вопрос необходимо провести в подсистеме: F5 "Типология", "Типологический анализ классов", "Кластерный и конструктивный анализ", "Автоматическое выполнение: [1]-[2]-[3]".

После этого выполним режим: "Генерация кластеров и конструкторов классов", при следующих параметрах:

- количество уровней кластеризации – 1;
- порог модуля сходства – 0;

- диапазон кодов классов: 13-56 (коды классов учебных групп и индивидуально студентов);
- исключаем кластеры с одним классом;
- не исключаем кластеры с тождественным набором классов.

Результаты кластерно-конструктивного анализа классов можно получить в виде TXT-файла с именем **Klas_obj.txt** в поддиректории TXT (таблица 58):

Таблица 58 – КОНСТРУКТ: "УЧЕБНАЯ ГРУППА"

15-10-04 10:53:27 Фильтр кодов: 13-56

Код класса	Наименование класса распознавания	Сход- ство %
13	ГРУППА ПИ-51.....	100.00
27	Мясников ПИ-51.....	43.60
31	Черкашина ПИ-51.....	40.11
28	Нагапетян ПИ-51.....	27.38
15	Бабенко ПИ-51.....	24.17
21	Заяц ПИ-51.....	21.70
20	Жеребятъев ПИ-51.....	16.21
23	Котенко ПИ-51.....	13.90
34	Шульгин ПИ-51.....	3.78
46	Куркина ПИ-52.....	0.98
16	Воробьева ПИ-51.....	-0.35
53	Силенко ПИ-52.....	-1.79
43	Иванова ПИ-52.....	-3.81
26	Лях ПИ-51.....	-3.83
24	Кузина О. ПИ-51.....	-4.22
25	Кузина Я. ПИ-51.....	-5.27
37	Веревкина ПИ-52.....	-13.84
54	Соколова ПИ-52.....	-14.71
49	Мануйлов ПИ-52.....	-16.14
45	Крейс ПИ-52.....	-17.43
56	Цисарь ПИ-52.....	-23.11
22	Иванова ПИ-51.....	-23.42
50	Нарижний ПИ-52.....	-24.32
52	Паршакова ПИ-52.....	-26.13
41	Еременко ПИ-52.....	-30.56
47	Люлик ПИ-52.....	-36.87
38	Григорьева ПИ-52.....	-52.30
14	ГРУППА ПИ-52.....	-64.51

Из таблицы 58 видно, что:

- студенты: Мясников, Черкашина, Нагапетян, Бабенко являются "типичными представителями" группы ПИ-51;
- студенты: Григорьева, Люлик, Еременко, Паршакова, Нарижний, Иванова, Цисарь являются "типичными представителями" группы ПИ-52;

– студенты: Шульгин, Куркина, Воробьева, Силенко, Иванова, Лях, Кузина О. и Кузина Я. обладают своеобразием и более ярко выраженной индивидуальностью, по сравнению с другими, т.к. они не похожи на типичных представителей ни одной группы (в рассматриваемой системе признаков).

Необходимо также отметить, что в учебной группе ПИ-51 степень разнообразия студентов в рассматриваемой системе признаков выше, чем в группе ПИ-52, что можно объяснить тем, что в 51 группе больше ребят, чем девушек и они в основном из краевого центра, а в 52 группе больше девушек и в основном они из села, как это видно из конструкта "Пол" в таблице 13.

Пример решения задания 4.5: Результаты анализа отобразить в графической форме семантических сетей классов, нелокальных нейронов и семантических сетей признаков. На их основе построить классические когнитивные карты для студентов с различной успеваемостью

Семантические сети классов отображаются на экране и записываются в форме РСХ-файла в поддиректории: **PCX/Klas-obj** в подсистеме F5 "Типология", "Типологический анализ классов", "Кластерный и конструктивный анализ", "Вывод 2d семантических сетей классов". На рисунке 110 приведена семантическая сеть классов с кодами от 1 до 14 (обобщенные классы).

Нелокальные нейроны отображаются в подсистеме F6 "Анализ", режиме: "Графическое отображение нелокальных нейронов". Они записываются в форме РСХ-файла в поддиректории: **PCX/NeuroNet**. На рисунке 111 приведен нелокальный нейрон: "Успеваемость: очень хорошая".

Семантическая сеть признаков отображается в подсистеме: F6 "Типология", в режиме: "Типологический анализ первичных признаков – Кластерный и конструктивный анализ – Вывод 2d семантических сетей признаков". Семантическая сеть отображается на экране и записывается в виде РСХ-файла в поддиректории: **PCX/Klas-att**.

На рисунке 112 приведена семантическая сеть для тех признаков, у которых оказались наибольшие весовые коэффициенты в приведенном на рисунке 111 нейроне (показаны только те связи между признаками, которые оказались больше 30%):

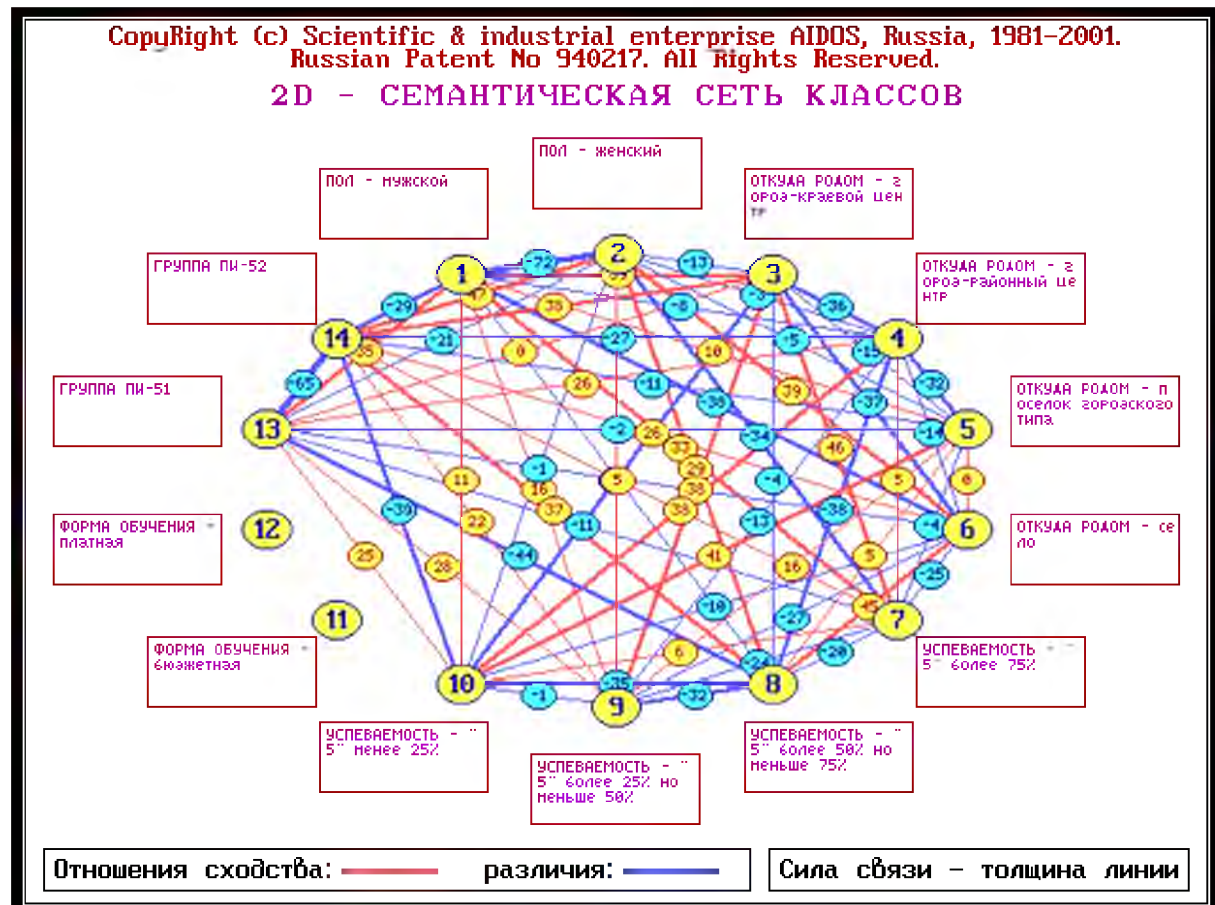


Рисунок 110. Семантическая сеть классов

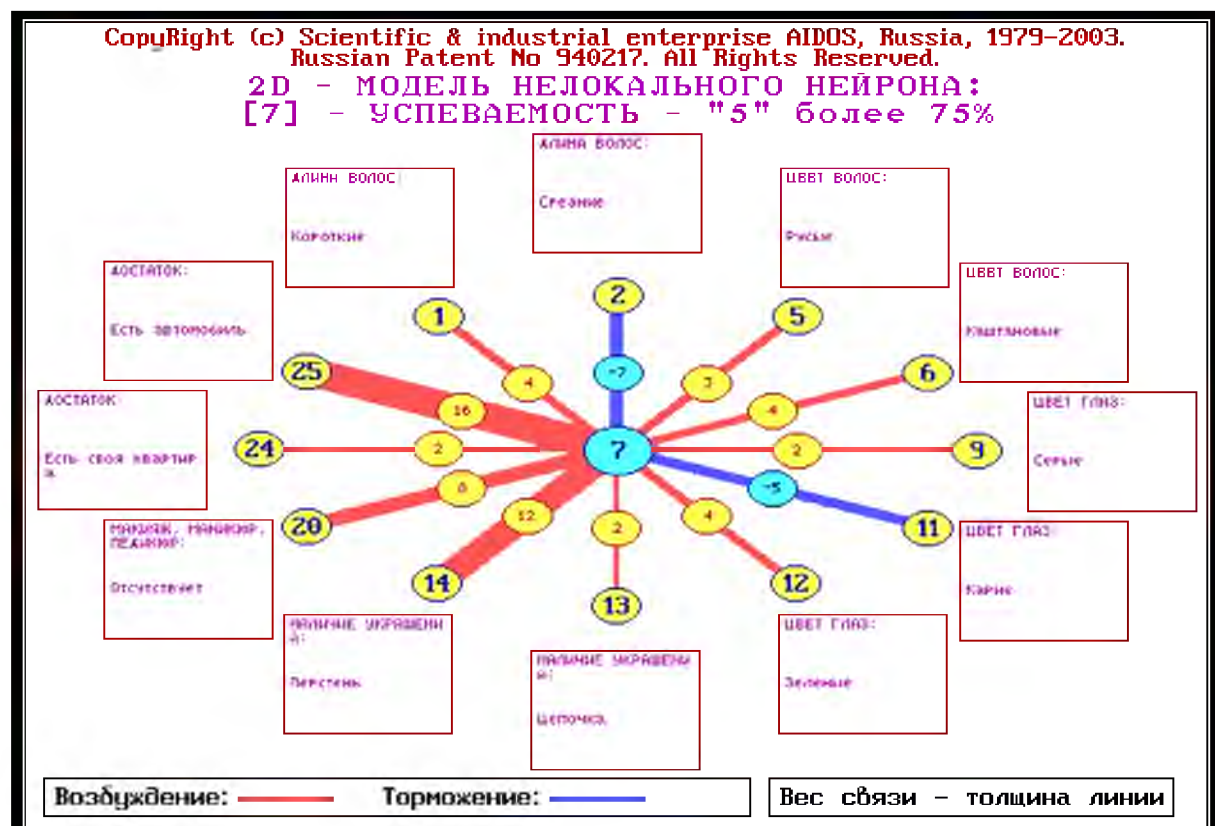


Рисунок 111. Нелокальный нейрон: "Успеваемость: очень хорошая"

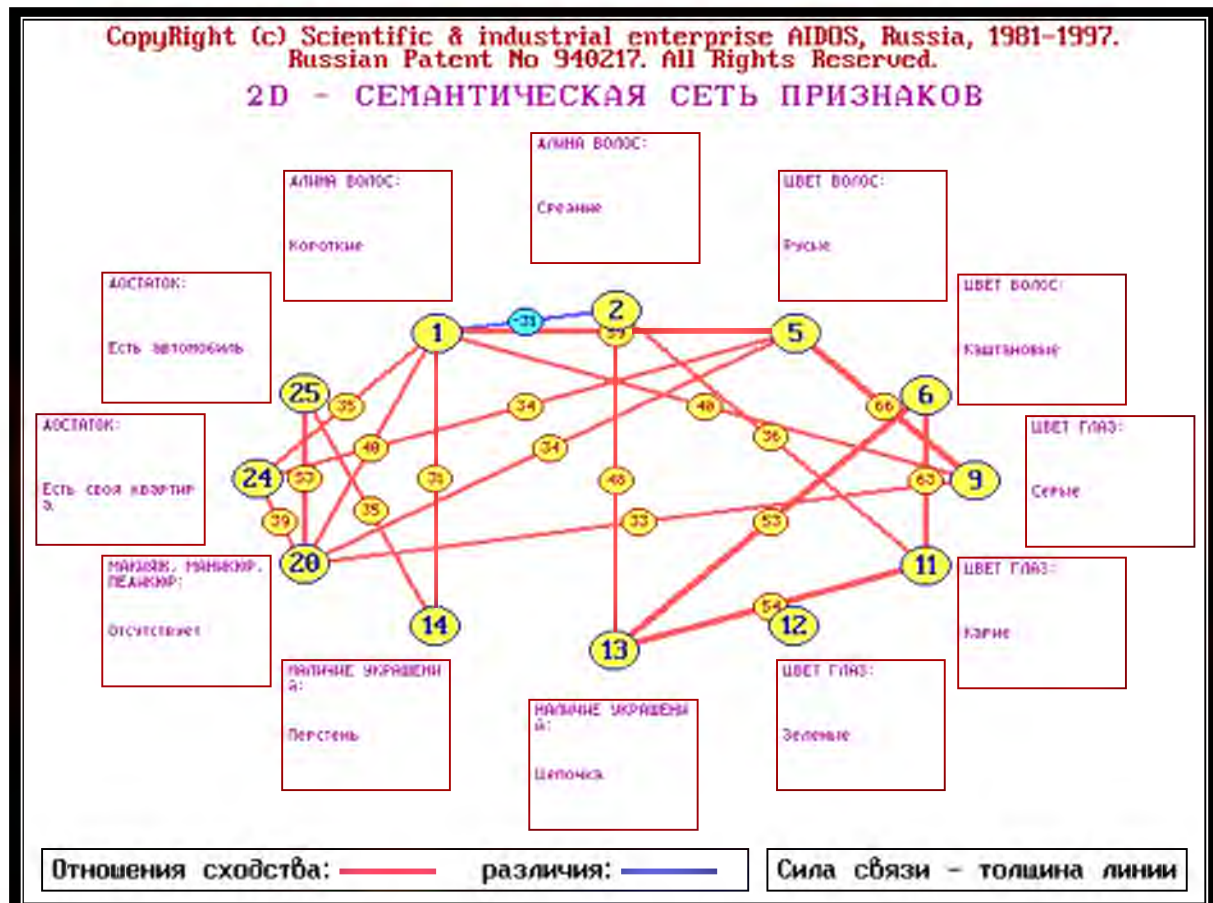


Рисунок 112. Семантическая сеть признаков, оказавшихся наиболее значимыми в нелокальном нейроне: "Успеваемость: очень хорошая"

Если дополнить графическое изображение нейрона на рисунке 111 связями между признаками, показанными в семантической сети на рисунке 112, то получим классическую когнитивную карту. Студентам это предлагается сделать самостоятельно в 7-м режиме 6-й подсистемы, также как и получить когнитивные карты для других уровней успеваемости.

Контрольные вопросы

1. Что понимается под формализацией задачи.
2. Как создать классификационные и описательные шкалы в системе "Эйдос"?
3. Как собрать исходную фактографическую информацию и ввести в систему обучающую выборку?
4. Как осуществить синтез и верификацию модели?
5. Как оценить ценность признаков для прогнозирования и выделить признаки, наиболее существенные для решения поставленной задачи?

6. Как провести анализ модели, чтобы ответить на следующие вопросы:

– как посещаемость занятий по системам искусственного интеллекта влияет на успеваемость по этой дисциплине?

– как сказывается пол на посещаемости?

– как выглядят конструкты "Пол", "Город-деревня", "Учебная группа", "Успеваемость", "Посещаемость"?

– какие студенты являются "типичными представителями" для своих учебных групп, а какие обладают своеобразием и выраженной индивидуальностью;

7. Как отобразить результаты анализа в графической форме нелокальных нейронов и семантических сетей признаков. На их основе построить классические когнитивные карты для хорошо и плохо успевающих студентов.

Литература по лабораторной работе

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.

2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

ЛР-3: "Прогнозирование учебных достижений студентов на основе особенностей их почерка"

Краткая теория

Известно, что подчерк каждого человека зависит от его темперамента и других генетически обусловленных, конституционных качеств личности, обладающих высокой стабильностью (некоторые из этих качеств могут быть измерены с применением 16PF-опросника Кеттелла), а также от социального статуса и текущего психического и физического состояния человека.

Учебные достижения по различным дисциплинам также зависят от конституционных качеств. Поэтому вполне корректно поставить задачу прогнозирования учебных достижений студентов на основе изучения особенностей их почерка.

Задание

1. Формализовать задачу.

1.1. Сконструировать классификационные шкалы и градации, выбрав в качестве классов – различные уровни учебных достижений по различным дисциплинам, перечень которых взять из зачетной книжки.

1.2. В качестве описательных шкал и градаций использовать характеристики почерка.

1.3. Обучающую выборку заполнить на основе данных по учащимся своей группы и дополнить данными параллельной группы.

2. Осуществить синтез и верификацию (измерение адекватности) семантической информационной модели.

3. Провести системно-когнитивный анализ модели:

3.1. Решить задачи идентификации и прогнозирования (для себя).

3.2. Сгенерировать информационные портреты классов и факторов, т.е. решить обратную задачу прогнозирования (результаты отобразить в графической форме двухмерных и трехмерных профилей классов и факторов).

3.3. Провести кластерно-конструктивный анализ классов и факторов (результаты отобразить в форме семантических сетей классов и факторов).

3.4. Осуществить содержательное сравнение классов и факторов (результаты отобразить в форме когнитивных диаграмм классов и факторов).

3.5. Построить нелокальные нейроны и интерпретируемые нейронные сети.

3.6. Построить классические когнитивные модели (отобразить в форме когнитивных карт).

3.7. Построить интегральные когнитивные модели (отобразить в форме интегральных когнитивных карт).

Пример решения

Задание 1. Формализовать задачу.

Задание 1.1. Сконструировать классификационные шкалы и градации, выбрав в качестве классов – различные уровни учебных достижений по различным дисциплинам, перечень которых взять из зачетной книжки.

Пример формирования классификационных шкалы и градаций приведен в таблице 59.

Таблица 59 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ И ГРАДАЦИИ

Код	Наименование класса	Код	Наименование класса
1	ПОЛ - мужской	29	Полонская ПИ-51
2	ПОЛ - женский	30	Трунина ПИ-51
3	ОТКУДА РОДОМ - город-краевой центр	31	Черкашина ПИ-51
4	ОТКУДА РОДОМ - город-районный центр	32	Чепурченко ПИ-51
5	ОТКУДА РОДОМ - поселок городского типа	33	Чушкин ПИ-51
6	ОТКУДА РОДОМ - село	34	Шульгин ПИ-51
7	УСПЕВАЕМОСТЬ - "5" более 75%	35	Арушанян ПИ-52
8	УСПЕВАЕМОСТЬ - "5" более 50% но меньше 75%	36	Быченков ПИ-52
9	УСПЕВАЕМОСТЬ - "5" более 25% но меньше 50%	37	Веревкина ПИ-52
10	УСПЕВАЕМОСТЬ - "5" менее 25%	38	Григорьева ПИ-52
11	ФОРМА ОБУЧЕНИЯ - бюджетная	39	Давыдич ПИ-52
12	ФОРМА ОБУЧЕНИЯ - платная	40	Дронова ПИ-52
13	ГРУППА ПИ-51	41	Еременко ПИ-52
14	ГРУППА ПИ-52	42	Жмурко ПИ-52
15	Бабенко ПИ-51	43	Иванова ПИ-52
16	Воробьева ПИ-51	44	Костенко ПИ-52
17	Гура ПИ-51	45	Крейс ПИ-52
18	Головнев ПИ-51	46	Куркина ПИ-52
19	Дыбова ПИ-51	47	Люлик ПИ-52
20	Жеребятёв ПИ-51	48	Максимов ПИ-52
21	Заяц ПИ-51	49	Мануйлов ПИ-52
22	Иванова ПИ-51	50	Нарижный ПИ-52
23	Котенко ПИ-51	51	Ольховская ПИ-52
24	Кузина О. ПИ-51	52	Паршакова ПИ-52
25	Кузина Я. ПИ-51	53	Силенко ПИ-52
26	Лях ПИ-51	54	Соколова ПИ-52
27	Мясников ПИ-51	55	Турбин ПИ-52
28	Нагапетян ПИ-51	56	Цисарь ПИ-52

Задание 1.2. В качестве описательных шкал и градаций использовать характеристики подчерка.

Существует много различных систем выявления признаков подчерка. Мы в учебных целях воспользуемся одной из самых простых из них, используемой на сайте: Альянс-медиа "Деловые тесты":

http://www.businessitest.ru/test.asp?test_id=155&topic_id=3

В таблице 60 эта система приведена в виде, преобразованном для удобства использования в системе "Эйдос".

Таблица 60 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ
СПРАВОЧНИК НАИМЕНОВАНИЙ ШКАЛ И ГРАДАЦИЙ ПРИЗНАКОВ

06-09-04 10:39:45

г. Краснодар

N	Код	Наименование
[1]	1]	РАЗМЕР БУКВ:
1	1	Очень мелкие.....
2	2	Мелкие.....
3	3	Средние.....
4	4	Крупные.....
[2]	2]	НАКЛОН БУКВ:
5	5	Влево.....
6	6	Легкий влево.....
7	7	Вправо.....
8	8	Резкий вправо.....
9	9	Прямое написание.....
[3]	3]	ФОРМА БУКВ:
10	10	Округлые.....
11	11	Бесформенные.....
12	12	Угловатые.....
[4]	4]	НАПРАВЛЕНИЕ ПОЧЕРКА:
13	13	Строчки "ползут" вверх.....
14	14	Строчки прямые.....
15	15	Строчки "сползают" вниз.....
[5]	5]	ИНТЕНСИВНОСТЬ ПОЧЕРКА И СИЛА НАЖИМА:
16	16	Легкая.....
17	17	Средняя.....
18	18	Сильная.....
[6]	6]	ХАРАКТЕР НАПИСАНИЯ СЛОВ:
19	19	Склонность к соединению букв.....
20	20	Склонность к отделению букв друг от друга.....
21	21	Смешанный стиль.....
[7]	7]	ОБЩАЯ ОЦЕНКА ПОЧЕРКА:
22	22	Почерк старательный, буквы аккуратные.....
23	23	Почерк неровный, одни слова видны, другие читаются с трудом.....
24	24	Буквы написаны кое-как, почерк неразборчивый и небрежный.....

Универсальная когнитивная аналитическая система

НПП *ЭЙДОС*

Задание 1.3. Обучающую выборку заполнить на основе данных по учащимся своей группы и дополнить данными параллельной группы.

Чтобы собрать информацию для обучающей выборки, студенты на доске рисуют и заполняют таблицу, аналогичную таблице 61. Каждый заполняет строку по себе, и все учащиеся переписывают таблицу целиком к себе в тетрадь. Затем, когда таблица в тетради заполнена – она заносится в систему "Эйдос" в 1-м режиме 2-й подсистемы.

Таблица 61 – ОБУЧАЮЩАЯ ВЫБОРКА

Код	Наименование	Коды классов						Коды признаков						
2	Воробьева ПИ-51	2	5	10	12	13	16	4	7	10	14	17	21	23
3	Гура ПИ-51	1	13	17	0	0	0	3	7	11	15	17	21	24
5	Дыбова ПИ-51	2	13	19	0	0	0							
6	Жеребятъев ПИ51	1	5	8	12	13	20							
8	Иванова ПИ-51	2	3	7	12	13	22	3	6	7	12	14	17	19
9	Котенко ПИ-51	2	4	7	12	13	23	3	7	10	14	17	21	23
10	Кузина О. ПИ-51	2	3	8	12	13	24	3	7	12	14	17	21	22
11	Кузина Я. ПИ-51	2	3	8	12	13	25	3	7	11	14	17	21	22
12	Лях ПИ-51	1	3	8	12	13	26	2	6	11	14	16	20	24
13	Мясников ПИ-51	1	3	9	12	13	27	3	7	10	14	17	21	22
14	Нагапетян ПИ-51	1	4	9	12	13	28	3	8	12	13	18	19	23
15	Полонская ПИ-51	2	13	29	0	0	0							
16	Трунина ПИ-51	2	13	30	0	0	0	3	7	12	13	18	19	23
17	Черкашина ПИ-51	2	4	10	12	13	31	3	7	12	14	17	19	23
18	Чепурченко ПИ51	1	13	32	0	0	0	3	8	12	13	16	19	23
19	Чушкин ПИ-51	1	13	33	0	0	0	3	6	11	14	17	20	23
20	Шульгин ПИ-51	1	5	8	12	13	34							
21	Арушунян ПИ-52	1	14	35	0	0	0							
22	Быченков ПИ-52	1	14	36	0	0	0							
23	Веревкина ПИ-52	2	3	9	12	14	37							
24	Григорьева ПИ52	2	4	8	12	14	38							
25	Давыдич ПИ-52	2	14	39	0	0	0							
26	Дронова ПИ-52	2	14	40	0	0	0	3	7	11	14	17	20	23
27	Еременко ПИ-52	2	4	8	12	14	41							
28	Жмурко ПИ-52	1	14	42	0	0	0	3	7	10	14	17	21	23
29	Иванова ПИ-52	2	3	9	12	14	43							
30	Костенко ПИ-52	2	14	44	0	0	0							
31	Крейс ПИ-52	2	4	8	12	14	45							
32	Куркина ПИ-52	2	3	8	12	14	46							
33	Люлик ПИ-52	2	5	8	12	14	47							
34	Максимов ПИ-52	1	14	48	0	0	0							
35	Мануйлов ПИ-52	1	3	7	12	14	49							
36	Нарижний ПИ-52	1	3	7	12	14	50							
37	Ольховская ПИ52	2	14	51	0	0	0							
38	Паршакова ПИ-52	2	6	8	12	14	52							
39	Силенко ПИ-52	1	3	7	12	14	53	3	5	11	14	17	21	23
40	Соколова ПИ-52	2	4	8	12	14	54							
41	Турбин ПИ-52	1	14	55	0	0	0	3	7	11	14	17	21	23
42	Цисарь ПИ-52	2	5	9	12	14	56	3	9	10	14	17	21	23
43	Бабенко ПИ-51	0	0	0	0	0	0	2	7	11	14	17	21	24

Так как по ряду студентов данных нет (что связано с низкой посещаемостью) и технология дальнейшей обработки информации уже освоена студентами в предыдущей лабораторной работе, то пример выполнения заданий 2 и 3 не приводится и они выполняются студентами **самостоятельно**.

Контрольные вопросы

1. Какие работы выполняются на этапе формализации задачи?
2. Как в системе "Эйдос" ввести классификационные шкалы и градации, выбрав в качестве классов – различные уровни учебных достижений по различным дисциплинам, перечень которых взять из зачетной книжки?
3. Как в системе "Эйдос" ввести описательные шкалы и градации, используя характеристики подчерка?
4. Каким образом подготовить и ввести в систему "Эйдос" обучающую выборку?
5. Как осуществить синтез и верификацию (измерение адекватности) семантической информационной модели в системе "Эйдос"?
6. Что включает системно-когнитивный анализ модели?
7. Как решаются задачи идентификации и прогнозирования в системе "Эйдос"?
8. В каких подсистемах и режимах системы "Эйдос" генерируются информационные портреты классов и факторов и отображаются в графической форме двумерных и трехмерных профилей классов и факторов?
9. Каким образом в системе "Эйдос" провести кластерно-конструктивный анализ классов и факторов и отобразить его в форме семантических сетей классов и факторов?
10. Как в системе "Эйдос" осуществить содержательное сравнение классов и факторов и отобразить результаты в форме когнитивных диаграмм классов и факторов?
11. Как в системе "Эйдос" построить нелокальные нейроны и интерпретируемые нейронные сети?
12. Как в системе "Эйдос" построить классические когнитивные модели и отобразить их в форме когнитивных карт?
13. Как в системе "Эйдос" построить интегральные когнитивные модели и отобразить в форме интегральных когнитивных карт?

Литература по лабораторной работе

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

ЛР-4:

"Прогнозирование учебных достижений студентов на основе информации об их социальном статусе"

Данная лабораторная работа основана на данных, полученных в Кубанском государственном технологическом университете (КубГТУ) при преподавании дисциплин кафедры Информатики: "Новые информационные технологии в учебном процессе" и "Комплексные технологии в науке и образовании" для специальности: 020300 – Социология (Социально-гуманитарный факультет), в 2002 году (акт внедрения имеется).

Краткая теория

При проведении в 2001 году одного из первых в г. Краснодаре экспериментов по централизованному компьютерному тестированию для поступления в вуз его организаторы предложили заполнить всем тестирующимся дополнительную социологическую анкету. Если бы удалось собрать данные с помощью этой анкеты по ряду регионов России и обработать их *совместно* с результатами тестирования по предметной обученности по ряду предметов, то в принципе оказалось бы возможным найти *зависимости между характеристиками социального статуса абитуриента и его учебными достижениями*. Безусловно, такие результаты представляли бы огромный научный интерес. Однако провести это дополнительное тестирование не удалось по ряду причин, среди которых основными, по-видимому, явились следующие:

- это тестирование не было бы анонимным;
- в анкету были включены некоторые вопросы, имеющие спорный характер;
- работа по организации проведению этого дополнительного тестирования потребовало бы от абитуриентов и организаторов проведения тестирования ранее незапланированных затрат времени и других усилий.

Но то, что в свое время не удалось сделать в масштабе России, вполне возможно осуществить в рамках одной лабораторной работы, в которой и предлагается использовать "ту самую социологическую анкету".

Задание

1. Формализовать задачу.

1.1. Сконструировать классификационные шкалы и градации, выбрав в качестве классов – различные уровни учебных достижений по различным дисциплинам, перечень которых взять из зачетной книжки.

1.2. В качестве описательных шкал и градаций использовать предлагаемую анкету.

1.3. Обучающую выборку заполнить на основе данных по учащимся своей группы и дополнить данными параллельной группы.

2. Осуществить синтез и верификацию (измерение адекватности) семантической информационной модели.

3. Провести системно-когнитивный анализ модели:

3.1. Решить задачи идентификации и прогнозирования (для себя).

3.2. Сгенерировать информационные портреты классов и факторов, т.е. решить обратную задачу прогнозирования (результаты отобразить в графической форме двухмерных и трехмерных профилей классов и факторов).

3.3. Провести кластерно-конструктивный анализ классов и факторов (результаты отобразить в форме семантических сетей классов и факторов).

3.4. Осуществить содержательное сравнение классов и факторов (результаты отобразить в форме когнитивных диаграмм классов и факторов).

3.5. Построить нелокальные нейроны и интерпретируемые нейронные сети.

3.6. Построить классические когнитивные модели (отобразить в форме когнитивных карт).

3.7. Построить интегральные когнитивные модели (отобразить в форме интегральных когнитивных карт).

Пример решения

Задание 1. Формализовать задачу.

Задание 1.1. Сконструировать классификационные шкалы и градации, выбрав в качестве классов – различные уровни учебных достижений по различным дисциплинам, перечень которых взять из образовательного стандарта или зачетной книжки.

Мы для выполнения этого задания используем данные, полученные в КубГТУ в 2002 году (таблица 62).

Таблица 62 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ И ГРАДАЦИИ

Код	Наименование класса
1	УСП.ПО ЦИКЛУ ДИСЦ.ГСД (гум.соц.-эк.дисц.) - высокая
2	УСП.ПО ЦИКЛУ ДИСЦ.ГСД (гум.соц.-эк.дисц.) - средняя
3	УСП.ПО ЦИКЛУ ДИСЦ.ГСД (гум.соц.-эк.дисц.) - низкая
4	УСП.ПО ЦИКЛУ ДИСЦ.ЕН (ест.научн.дисц.) - высокая
5	УСП.ПО ЦИКЛУ ДИСЦ.ЕН (ест.научн.дисц.) - средняя
6	УСП.ПО ЦИКЛУ ДИСЦ.ЕН (ест.научн.дисц.) - низкая
7	УСП.ПО ЦИКЛУ ДИСЦ.ОПД (общ.проф.дисц.) - высокая
8	УСП.ПО ЦИКЛУ ДИСЦ.ОПД (общ.проф.дисц.) - средняя
9	УСП.ПО ЦИКЛУ ДИСЦ.ОПД (общ.проф.дисц.) - низкая
10	УСП.ПО ЦИКЛУ ДИСЦ.СД (спец.дисц.) - высокая
11	УСП.ПО ЦИКЛУ ДИСЦ.СД (спец.дисц.) - средняя
12	УСП.ПО ЦИКЛУ ДИСЦ.СД (спец.дисц.) - низкая
13	УСП.ПО ДИСЦ.ГСД-01 (философия) - высокая
14	УСП.ПО ДИСЦ.ГСД-01 (философия) - средняя
15	УСП.ПО ДИСЦ.ГСД-01 (философия) - низкая
16	УСП.ПО ДИСЦ.ГСД-02 (иностр.язык) - высокая
17	УСП.ПО ДИСЦ.ГСД-02 (иностр.язык) - средняя
18	УСП.ПО ДИСЦ.ГСД-02 (иностр.язык) - низкая
19	УСП.ПО ДИСЦ.ГСД-03 (культурология) - высокая
20	УСП.ПО ДИСЦ.ГСД-03 (культурология) - средняя
21	УСП.ПО ДИСЦ.ГСД-03 (культурология) - низкая
22	УСП.ПО ДИСЦ.ГСД-04 (история) - высокая
23	УСП.ПО ДИСЦ.ГСД-04 (история) - средняя

Продолжение таблицы 62

Код	Наименование класса
24	УСП.ПО ДИСЦ.ГСД-04 (история) - низкая
25	УСП.ПО ДИСЦ.ГСД-05 (физ.культура) - высокая
26	УСП.ПО ДИСЦ.ГСД-05 (физ.культура) - средняя
27	УСП.ПО ДИСЦ.ГСД-05 (физ.культура) - низкая
28	УСП.ПО ДИСЦ.ГСД-06 (право) - высокая
29	УСП.ПО ДИСЦ.ГСД-06 (право) - средняя
30	УСП.ПО ДИСЦ.ГСД-06 (право) - низкая
31	УСП.ПО ДИСЦ.ГСД-07 (экономика) - высокая
32	УСП.ПО ДИСЦ.ГСД-07 (экономика) - средняя
33	УСП.ПО ДИСЦ.ГСД-07 (экономика) - низкая
34	УСП.ПО ДИСЦ.ГСД-09 (политология) - высокая
35	УСП.ПО ДИСЦ.ГСД-09 (политология) - средняя
36	УСП.ПО ДИСЦ.ГСД-09 (политология) - низкая
37	УСП.ПО ДИСЦ.ГСД-10 (психология и педагогика) - высокая
38	УСП.ПО ДИСЦ.ГСД-10 (психология и педагогика) - средняя
39	УСП.ПО ДИСЦ.ГСД-10 (психология и педагогика) - низкая
40	УСП.ПО ДИСЦ.ГСД-11 (дисц.и курсы по выбору) - высокая
41	УСП.ПО ДИСЦ.ГСД-11 (дисц.и курсы по выбору) - средняя
42	УСП.ПО ДИСЦ.ГСД-11 (дисц.и курсы по выбору) - низкая
43	УСП.ПО ДИСЦ.ЕН-01 (математика и информатика) - высокая
44	УСП.ПО ДИСЦ.ЕН-01 (математика и информатика) - средняя
45	УСП.ПО ДИСЦ.ЕН-01 (математика и информатика) - низкая
46	УСП.ПО ДИСЦ.ЕН-02 (конц.совр.естествознания) - высокая
47	УСП.ПО ДИСЦ.ЕН-02 (конц.совр.естествознания) - средняя
48	УСП.ПО ДИСЦ.ЕН-02 (конц.совр.естествознания) - низкая
49	УСП.ПО ДИСЦ.ЕН-03 (дисц.и курсы по выбору) - высокая
50	УСП.ПО ДИСЦ.ЕН-03 (дисц.и курсы по выбору) - средняя
51	УСП.ПО ДИСЦ.ЕН-03 (дисц.и курсы по выбору) - низкая
52	УСП.ПО ДИСЦ.ОПД-01 (история социологии) - высокая
53	УСП.ПО ДИСЦ.ОПД-01 (история социологии) - средняя
54	УСП.ПО ДИСЦ.ОПД-01 (история социологии) - низкая
55	УСП.ПО ДИСЦ.ОПД-02 (общая социология) - высокая
56	УСП.ПО ДИСЦ.ОПД-02 (общая социология) - средняя
57	УСП.ПО ДИСЦ.ОПД-02 (общая социология) - низкая
58	УСП.ПО ДИСЦ.ОПД-03 (мет.и техн.социолог.иссл.) - высокая
59	УСП.ПО ДИСЦ.ОПД-03 (мет.и техн.социолог.иссл.) - средняя
60	УСП.ПО ДИСЦ.ОПД-03 (мет.и техн.социолог.иссл.) - низкая
61	УСП.ПО ДИСЦ.ОПД-04 (демография) - высокая
62	УСП.ПО ДИСЦ.ОПД-04 (демография) - средняя
63	УСП.ПО ДИСЦ.ОПД-04 (демография) - низкая
64	УСП.ПО ДИСЦ.ОПД-05 (социальная статистика) - высокая
65	УСП.ПО ДИСЦ.ОПД-05 (социальная статистика) - средняя
66	УСП.ПО ДИСЦ.ОПД-05 (социальная статистика) - низкая
67	УСП.ПО ДИСЦ.ОПД-06 (социальная антропология) - высокая
68	УСП.ПО ДИСЦ.ОПД-06 (социальная антропология) - средняя
69	УСП.ПО ДИСЦ.ОПД-06 (социальная антропология) - низкая
70	УСП.ПО ДИСЦ.ОПД-07 (социальная психология) - высокая
71	УСП.ПО ДИСЦ.ОПД-07 (социальная психология) - средняя

Продолжение таблицы 62

Код	Наименование класса
72	УСП.ПО ДИСЦ.ОПД-07 (социальная психология) – низкая
73	УСП.ПО ДИСЦ.ОПД-08 (соц.труда и экон.соц.) - высокая
74	УСП.ПО ДИСЦ.ОПД-08 (соц.труда и экон.соц.) - средняя
75	УСП.ПО ДИСЦ.ОПД-08 (соц.труда и экон.соц.) - низкая
76	УСП.ПО ДИСЦ.ОПД-09 (социология организаций) - высокая
77	УСП.ПО ДИСЦ.ОПД-09 (социология организаций) - средняя
78	УСП.ПО ДИСЦ.ОПД-09 (социология организаций) - низкая
79	УСП.ПО ДИСЦ.ОПД-10 (социология коммуникаций) - высокая
80	УСП.ПО ДИСЦ.ОПД-10 (социология коммуникаций) - средняя
81	УСП.ПО ДИСЦ.ОПД-10 (социология коммуникаций) - низкая
82	УСП.ПО ДИСЦ.ОПД-11 (основы менеджмента) - высокая
83	УСП.ПО ДИСЦ.ОПД-11 (основы менеджмента) - средняя
84	УСП.ПО ДИСЦ.ОПД-11 (основы менеджмента) - низкая
85	УСП.ПО ДИСЦ.ОПД-12 (социальная экология) - высокая
86	УСП.ПО ДИСЦ.ОПД-12 (социальная экология) - средняя
87	УСП.ПО ДИСЦ.ОПД-12 (социальная экология) - низкая
88	УСП.ПО ДИСЦ.ОПД-13 (социология религий) - высокая
89	УСП.ПО ДИСЦ.ОПД-13 (социология религий) - средняя
90	УСП.ПО ДИСЦ.ОПД-13 (социология религий) - низкая
91	УСП.ПО ДИСЦ.ОПД-14 (дисц.и курсы по выбору) - высокая
92	УСП.ПО ДИСЦ.ОПД-14 (дисц.и курсы по выбору) - средняя
93	УСП.ПО ДИСЦ.ОПД-14 (дисц.и курсы по выбору) - низкая
94	УСП.ПО ДИСЦ.СД-01 (соц.прогнозирование и проект.) - высокая
95	УСП.ПО ДИСЦ.СД-01 (соц.прогнозирование и проект.) - средняя
96	УСП.ПО ДИСЦ.СД-01 (соц.прогнозирование и проект.) - низкая
97	УСП.ПО ДИСЦ.СД-02 (соц.проблемы изуч.общ.мнения) - высокая
98	УСП.ПО ДИСЦ.СД-02 (соц.проблемы изуч.общ.мнения) - средняя
99	УСП.ПО ДИСЦ.СД-02 (соц.проблемы изуч.общ.мнения) - низкая
100	УСП.ПО ДИСЦ.СД-03 (соц.полит.процессов) - высокая
101	УСП.ПО ДИСЦ.СД-03 (соц.полит.процессов) - средняя
102	УСП.ПО ДИСЦ.СД-03 (соц.полит.процессов) - низкая
103	УСП.ПО ДИСЦ.СД-04 (социология семьи) - высокая
104	УСП.ПО ДИСЦ.СД-04 (социология семьи) - средняя
105	УСП.ПО ДИСЦ.СД-04 (социология семьи) - низкая
106	УСП.ПО ДИСЦ.СД-05 (социология культуры) - высокая
107	УСП.ПО ДИСЦ.СД-05 (социология культуры) - средняя
108	УСП.ПО ДИСЦ.СД-05 (социология культуры) - низкая
109	УСП.ПО ДИСЦ.СД-06 (дисц.и курсы по выбору) - высокая
110	УСП.ПО ДИСЦ.СД-06 (дисц.и курсы по выбору) - средняя
111	УСП.ПО ДИСЦ.СД-06 (дисц.и курсы по выбору) - низкая
112	Группа-1
113	Группа-2

Задание 1.2. В качестве описательных шкал и градаций использовать предлагаемую анкету.

Анкета организаторов тестирования преобразована к виду, удобному для обработки в системе "Эйдос" (таблица 63).

Таблица 63 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ

Код	Наименования шкал и градаций признаков
[1]	ВАШ ВОЗРАСТ:
1	17
2	18
3	19
4	20
5	21
6	22
7	23
8	24
9	25 или более
[2]	ВЫ РОДИЛИСЬ В РОССИИ?
10	Да
11	Нет
[3]	ЕСЛИ ВЫ РОДИЛИСЬ НЕ В РОССИИ, ТО ВО СКОЛЬКО ЛЕТ ВЫ СЮДА ПРИЕХАЛИ?
12	До 7 лет
13	От 7 до 17 лет
14	Больше 17 лет
[4]	КОЛИЧЕСТВО ЖИЛЫХ КОМНАТ У ВАС ДОМА:
15	1
16	2
17	3
18	4
19	5 или более
[5]	СКОЛЬКО ВСЕГО ЧЕЛОВЕК ЖИВУТ У ВАС ДОМА?
20	1
21	2
22	3
23	4
24	5
25	6
26	7 и более
[6]	ДОХОД НА ЧЛЕНА СЕМЬИ (В МЕСЯЦ):
27	Низкий
28	Средний
29	Высокий
[7]	КАК ВЫ САМИ ОЦЕНИВАЕТЕ СВОИ ШКОЛЬНЫЕ ЗНАНИЯ ПО РУССКОМУ ЯЗЫКУ:
30	Отличные
31	Хорошие
32	Средние
33	Низкие
[8]	КАК ВЫ САМИ ОЦЕНИВАЕТЕ СВОИ ШКОЛЬНЫЕ ЗНАНИЯ ПО АЛГЕБРЕ:
34	Отличные
35	Хорошие
36	Средние
37	Низкие
[9]	КАК ВЫ САМИ ОЦЕНИВАЕТЕ СВОИ ШКОЛЬНЫЕ ЗНАНИЯ ПО ГЕОМЕТРИИ:
38	Отличные
39	Хорошие
40	Средние
41	Низкие
[10]	КАК ВЫ САМИ ОЦЕНИВАЕТЕ СВОИ ШКОЛЬНЫЕ ЗНАНИЯ ПО ФИЗИКЕ:
42	Отличные
43	Хорошие
44	Средние
45	Низкие

Продолжение таблицы 63

Код	Наименования шкал и градаций признаков
[11]	КАК ВЫ САМИ ОЦЕНИВАЕТЕ СВОИ ШКОЛЬНЫЕ ЗНАНИЯ ПО ХИМИИ:
46	Отличные
47	Хорошие
48	Средние
49	Низкие
[12]	КАК ВЫ САМИ ОЦЕНИВАЕТЕ СВОИ ШКОЛЬНЫЕ ЗНАНИЯ ПО ИНФОРМАТИКЕ:
50	Отличные
51	Хорошие
52	Средние
[13]	КАК ВЫ САМИ ОЦЕНИВАЕТЕ СВОИ ШКОЛЬНЫЕ ЗНАНИЯ ПО БИОЛОГИИ:
54	Отличные
55	Хорошие
56	Средние
57	Низкие
[14]	КАК ВЫ САМИ ОЦЕНИВАЕТЕ СВОИ ШКОЛЬНЫЕ ЗНАНИЯ ПО ИСТОРИИ:
58	Отличные
59	Хорошие
60	Средние
61	Низкие
[15]	КАК ВЫ САМИ ОЦЕНИВАЕТЕ СВОИ ШКОЛЬНЫЕ ЗНАНИЯ ПО ИНОСТРАННОМУ ЯЗЫКУ
62	Отличные
63	Хорошие
64	Средние
65	Низкие
[16]	КАК ВЫ ГОТОВИЛИСЬ К ВСТУПИТЕЛЬНОМУ ЭКЗАМЕНУ ПО РУССКОМУ ЯЗЫКУ:
66	Уроки в школе
67	Со знакомыми
68	С репетитором
69	Факультатив в школе
70	Курсы при ВУЗе
71	Заочно при ВУЗе
[17]	КАК ВЫ ГОТОВИЛИСЬ К ВСТУПИТЕЛЬНОМУ ЭКЗАМЕНУ ПО АЛГЕБРЕ:
72	Уроки в школе
73	Со знакомыми
74	С репетитором
75	Факультатив в школе
76	Курсы при ВУЗе
77	Заочно при ВУЗе
[18]	КАК ВЫ ГОТОВИЛИСЬ К ВСТУПИТЕЛЬНОМУ ЭКЗАМЕНУ ПО ГЕОМЕТРИИ:
78	Уроки в школе
79	Со знакомыми
80	С репетитором
81	Факультатив в школе
82	Курсы при ВУЗе
83	Заочно при ВУЗе
[19]	КАК ВЫ ГОТОВИЛИСЬ К ВСТУПИТЕЛЬНОМУ ЭКЗАМЕНУ ПО ФИЗИКЕ:
84	Уроки в школе
85	Со знакомыми
86	С репетитором
87	Факультатив в школе
88	Курсы при ВУЗе
89	Заочно при ВУЗе
[20]	КАК ВЫ ГОТОВИЛИСЬ К ВСТУПИТЕЛЬНОМУ ЭКЗАМЕНУ ПО ХИМИИ:
90	Уроки в школе
91	Со знакомыми
92	С репетитором
93	Факультатив в школе
94	Курсы при ВУЗе
95	Заочно при ВУЗе

Продолжение таблицы 63

Код	Наименования шкал и градаций признаков
[21]	КАК ВЫ ГОТОВИЛИСЬ К ВСТУПИТЕЛЬНОМУ ЭКЗАМЕНУ ПО ИНФОРМАТИКЕ:
96	Уроки в школе
97	Со знакомыми
98	С репетитором
99	Факультатив в школе
100	Курсы при ВУЗе
101	Заочно при ВУЗе
[22]	КАК ВЫ ГОТОВИЛИСЬ К ВСТУПИТЕЛЬНОМУ ЭКЗАМЕНУ ПО БИОЛОГИИ:
102	Уроки в школе
103	Со знакомыми
104	С репетитором
105	Факультатив в школе
106	Курсы при ВУЗе
107	Заочно при ВУЗе
[23]	КАК ВЫ ГОТОВИЛИСЬ К ВСТУПИТЕЛЬНОМУ ЭКЗАМЕНУ ПО ИСТОРИИ:
108	Уроки в школе
109	Со знакомыми
110	С репетитором
111	Факультатив в школе
112	Курсы при ВУЗе
113	Заочно при ВУЗе
[24]	КАК ВЫ ГОТОВИЛИСЬ К ВСТУПИТЕЛЬНОМУ ЭКЗАМЕНУ ПО ИНОСТРАННОМУ ЯЗЫКУ
114	Уроки в школе
115	Со знакомыми
116	С репетитором
117	Факультатив в школе
118	Курсы при ВУЗе
119	Заочно при ВУЗе
[25]	СКОЛЬКО ВРЕМЕНИ СУММАРНО УХОДИТ У ВАС НА ВЫПОЛНЕНИЕ ДОМАШНИХ ЗАДАНИЙ?
120	До 15 минут
121	Около 30 минут
122	1 - 2 часа
123	Более 2-х часов
[26]	УКАЖИТЕ УРОВЕНЬ ОБРАЗОВАНИЯ ОТЦА:
124	Неполное среднее
125	Среднее
126	Среднее техническое
127	Незаконченное высшее
128	Высшее
129	Кандидат наук
130	Доктор наук
[27]	УКАЖИТЕ УРОВЕНЬ ОБРАЗОВАНИЯ МАТЕРИ:
131	Неполное среднее
132	Среднее
133	Среднее техническое
134	Незаконченное высшее
135	Высшее
136	Кандидат наук
137	Доктор наук
[28]	УКАЖИТЕ НАЦИОНАЛЬНОСТЬ ОТЦА:
138	Русский
139	Украинец
140	Белорус
141	Татарин
142	Мордвин
143	Удмурт
144	Мари
145	Башкир
146	Чуваш

Продолжение таблицы 63

Код	Наименования шкал и градаций признаков
147	Еврей
148	Калмык
149	Бурят
150	Коми
151	Грузин
152	Армянин
153	Азербайджанец
154	Другая
[29]	УКАЖИТЕ НАЦИОНАЛЬНОСТЬ МАТЕРИ:
155	Русский
156	Украинец
157	Белорус
158	Татарин
159	Мордвин
160	Удмурт
161	Мари
162	Башкир
163	Чуваш
164	Еврей
165	Калмык
166	Бурят
167	Коми
168	Грузин
169	Армянин
170	Азербайджанец
171	Другая
[30]	ВЫ ПОСТОЯННО ЖИВЕТЕ В ДАННОМ МЕСТЕ ИЛИ ПЕРЕЕХАЛИ СЮДА ИЗ ДРУГОЙ МЕСТНОСТИ?
172	Родился и живу здесь постоянно
173	Приехал сюда из другого города России
174	Приехал из сельской местности России
175	Приехал из бывших республик СССР
[31]	ВЫ ПРОЖИВАЕТЕ:
176	Москве
177	Санкт-Петербурге
178	В областном (краевом) центре или столице республики
179	В крупном городе
180	В среднем городе
181	В малом городе
182	В поселке, деревне, селе, хуторе и т.п.
[32]	КАК ЧАСТО ВЫ ДОМА РАЗГОВАРИВАЕТЕ НА РУССКОМ ЯЗЫКЕ?
183	Всегда
184	Почти всегда
185	Иногда
186	Очень редко
187	Никогда
[33]	ВМЕСТЕ С ВАМИ ПОСТОЯННО ПРОЖИВАЮТ:
188	Мать
189	Отец
190	Отчим/мачеха
191	Братья/сестры
192	Бабушки/дедушки
193	Другие родственники
[34]	КОЛИЧЕСТВО ДЕТЕЙ В СЕМЬЕ, ВКЛЮЧАЯ ВАС:
194	Один
195	Двое
196	Трое
197	Четверо или более

Продолжение таблицы 63

Код	Наименования шкал и градаций признаков
[35]	ИМЕЕТ ЛИ ВАША СЕМЬЯ:
198	Видеомагнитофон
199	Видеокамеру
200	Несколько цветных телевизоров
201	Компьютер
202	Отечественную автомашину
203	Несколько отечественных автомашин или иномарку
[36]	ЕСЛИ ВЫ ИМЕЕТЕ КОМПЬЮТЕР ДОМА, ТО КАК ВЫ ЕГО ИСПОЛЬЗУЕТЕ?
204	Для игр
205	Для изучения учебных предметов (например, иностранных языков)
206	Для обмена информацией с помощью модема, в т.ч. в Internet
207	Для оплачиваемой работы
[37]	СКОЛЬКО ПРИМЕРНО КНИГ У ВАС В СЕМЬЕ (НЕ СЧИТАЯ УЧЕБНИКОВ И УЧЕБНЫХ ПОСОБИИ)?
208	Менее 50
209	Около 100
210	Несколько сотен
211	Более тысячи
[38]	ВАШЕ ОТНОШЕНИЕ К ПЕРИОДИЧЕСКОЙ ПЕЧАТИ:
212	Выписываем на дом
213	Покупаем ежедневно
214	Покупаем еженедельно
215	Почти не читаем
[39]	ЖИЛИЩНЫЕ УСЛОВИЯ ВАШЕЙ СЕМЬИ:
216	Комната в коммунальной квартире
217	1-комнатная квартира
218	2-комнатная квартира
219	Квартира из 3-х или более комнат
220	Обычный дом в сельской местности
221	Дом повышенной комфортности (особняк)
[40]	ПРИ СОСТАВЛЕНИИ ПЛАНОВ НА БУДУЩЕЕ ЧЬЕ МНЕНИЕ ВЫ УЧИТЫВАЕТЕ В ПЕРВУЮ ОЧЕРЕДЬ (ВЫБЕРИТЕ ТОЛЬКО ОДИН ОТВЕТ):
222	Родителей
223	Учителей
224	Сверстников (друзей)
225	Средств массовой информации
226	Свое собственное
[41]	ГДЕ ВЫ ОБЫЧНО ПРОВОДИТЕ ЕЖЕГОДНЫЙ ОТПУСК (КАНИКУЛЫ):
227	Дома
228	На собственной даче
229	В другом городе или на даче у родственников или знакомых
230	В туристическом походе
231	На Черном море в России
232	В бывших республиках СССР
233	За рубежом
234	Другое
[42]	ВЫ ЗАКОНЧИЛИ ВЫПУСКНОЙ КЛАСС:
235	Общеобразовательный
236	Профилированный на ВУЗ
237	Гимназический или лицейский
238	другой (экономический, юридический и др.)
[43]	ВЫ ЗАКОНЧИЛИ ВЫПУСКНОЙ КЛАСС С УГЛУБЛЕННЫМ ИЗУЧЕНИЕМ:
239	Математики
240	Физики
241	Биологии
242	Химии
243	Географии
244	Информатики
245	Истории, обществоведения
246	Русского языка, литературы

Продолжение таблицы 63

Код	Наименования шкал и градаций признаков
247	Иностранных языков
[44]	ВЫБЕРИТЕ ПРЕДМЕТЫ, КОТОРЫЕ ВАМ НРАВЯТСЯ:
248	Русский язык
249	Математика
250	Физика
251	Химия
252	Информатика
253	Биология
254	История
255	Иностранный язык
256	География
[45]	ЧЕМ ВАМ НРАВИТСЯ ШКОЛА, КОТОРУЮ ВЫ ЗАКОНЧИЛИ?
257	Хорошие отношения со сверстниками
258	Хорошие учителя
259	Хорошая подготовка к поступлению в ВУЗ
260	Расположена рядом с домом
261	Ничем не нравится
262	Другое
[46]	ОБЛАСТЬ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ ОТЦА:
263	Промышленность (тяжелая)
264	Промышленность (легкая)
265	Строительство
266	Сельское хозяйство
267	Транспорт
268	Коммерция, торговля, сфера бытового обслуживания
269	Экономика
270	Частный бизнес, предпринимательство
271	Государственная служба
272	Служба в силовых структурах
273	Выборные органы власти
274	Образование
275	Здравоохранение, медицина
276	Культура, искусство
277	Наука
278	Юриспруденция
279	Спорт
280	Религия
281	Другое
[47]	ОБЛАСТЬ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ МАТЕРИ:
282	Промышленность (тяжелая)
283	Промышленность (легкая)
284	Строительство
285	Сельское хозяйство
286	Транспорт
287	Коммерция, торговля, сфера бытового обслуживания
288	Экономика
289	Частный бизнес, предпринимательство
290	Государственная служба
291	Служба в силовых структурах
292	Выборные органы власти
293	Образование
294	Здравоохранение, медицина
295	Культура, искусство
296	Наука
297	Юриспруденция
298	Спорт
299	Религия
300	Другое

Продолжение таблицы 63

Код	Наименования шкал и градаций признаков
[48]	ОБЛАСТЬ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ, НАИБОЛЕЕ ПРИВЛЕКАЮЩАЯ МЕНЯ
301	Промышленность (тяжелая)
302	Промышленность (легкая)
303	Строительство
304	Сельское хозяйство
305	Транспорт
306	Коммерция, торговля, сфера бытового обслуживания
307	Экономика
308	Частный бизнес, предпринимательство
309	Государственная служба
310	Служба в силовых структурах
311	Выборные органы власти
312	Образование
313	Здравоохранение, медицина
314	Культура, искусство
315	Наука
316	Юриспруденция
317	Спорт
318	Религия
319	Другое
[49]	УКАЖИТЕ ФОРМЫ ДОСУГА, КОТОРЫЕ ВЫ ПРАКТИЧЕСКИ НЕ ИСПОЛЬЗУЕТЕ:
320	Просмотр видео и кинофильмов
321	Чтение художественной литературы
322	Спорт, туризм
323	Дискотеки и т.п.
324	Общение с родителями
325	Общение с любимым человеком
326	Работа на дому
327	Оплачиваемая работа
328	Дополнительное образование
329	Техническое творчество
330	Занятия искусством
331	Коллекционирование
332	Посещение музеев и выставок
333	Посещение оперы и балета
334	Посещение драматических театров
335	Посещение концертов классической музыки
336	Посещение концертов попмузыки
337	Слушание радио
338	Слушание аудио-музыки
339	Просмотр теленовостей
340	Просмотр развлекательных TV-программ
341	Просмотр спортивных TV-программ
342	Просмотр мультфильмов
343	Просмотр телесериалов
344	Просмотр познавательных передач
345	Просмотр религиозных программ
346	Посещение религиозных центров
[50]	УКАЖИТЕ ФОРМЫ ДОСУГА, КОТОРЫЕ ВЫ ИСПОЛЬЗУЕТЕ ПО ВОЗМОЖНОСТИ:
347	Просмотр видео и кинофильмов
348	Чтение художественной литературы
349	Спорт, туризм
350	Дискотеки и т.п.
351	Общение с родителями
352	Общение с любимым человеком
353	Работа на дому
354	Оплачиваемая работа
355	Дополнительное образование

Продолжение таблицы 63

Код	Наименования шкал и градаций признаков
356	Техническое творчество
357	Занятия искусством
358	Коллекционирование
359	Посещение музеев и выставок
360	Посещение оперы и балета
361	Посещение драматических театров
362	Посещение концертов классической музыки
363	Посещение концертов попмузыки
364	Слушание радио
365	Слушание аудио-музыки
366	Просмотр теленовостей
367	Просмотр развлекательных TV-программ
368	Просмотр спортивных TV-программ
369	Просмотр мультфильмов
370	Просмотр телесериалов
371	Просмотр познавательных передач
372	Просмотр религиозных программ
373	Посещение религиозных центров
[51]	УКАЖИТЕ ФОРМЫ ДОСУГА, КОТОРЫЕ ВЫ ЧАСТО ИСПОЛЬЗУЕТЕ:
374	Просмотр видео и кинофильмов
375	Чтение художественной литературы
376	Спорт, туризм
377	Дискотеки и т.п.
378	Общение с родителями
379	Общение с любимым человеком
380	Работа на дому
381	Оплачиваемая работа
382	Дополнительное образование
383	Техническое творчество
384	Занятия искусством
385	Коллекционирование
386	Посещение музеев и выставок
387	Посещение оперы и балета
388	Посещение драматических театров
389	Посещение концертов классической музыки
390	Посещение концертов поп-музыки
391	Слушание радио
392	Слушание аудио-музыки
393	Просмотр теленовостей
394	Просмотр развлекательных TV-программ
395	Просмотр спортивных TV-программ
396	Просмотр мультфильмов
397	Просмотр телесериалов
398	Просмотр познавательных передач
399	Просмотр религиозных программ
400	Посещение религиозных центров
[52]	УКАЖИТЕ ФОРМЫ ДОСУГА, КОТОРЫЕ ВЫ ИСПОЛЬЗУЕТЕ ПРАКТИЧЕСКИ ВСЕГДА:
401	Просмотр видео и кинофильмов
402	Чтение художественной литературы
403	Спорт, туризм
404	Дискотеки и т.п.
405	Общение с родителями
406	Общение с любимым человеком
407	Работа на дому
408	Оплачиваемая работа
409	Дополнительное образование
410	Техническое творчество
411	Занятия искусством
412	Коллекционирование
413	Посещение музеев и выставок

Продолжение таблицы 63

Код	Наименования шкал и градаций признаков
414	Посещение оперы и балета
415	Посещение драматических театров
416	Посещение концертов классической музыки
417	Посещение концертов поп-музыки
418	Слушание радио
419	Слушание аудио-музыки
420	Просмотр теленовостей
421	Просмотр развлекательных TV-программ
422	Просмотр спортивных TV-программ
423	Просмотр мультфильмов
424	Просмотр телесериалов
425	Просмотр познавательных передач
426	Просмотр религиозных программ
427	Посещение религиозных центров
[53]	УКАЖИТЕ СТОРОНЫ ЖИЗНИ, ИМЕЮЩИЕ ДЛЯ ВАС МАЛУЮ ЦЕННОСТЬ:
428	Общение с друзьями
429	Общение с родителями
430	Наличие вещей, имеющих продуктивную работу и учебу
431	Наличие престижных вещей
432	Наличие вещей, обеспечивающих комфортные условия жизни
433	Возможность влияния на ход важных событий в семье
434	Материальная независимость
435	Свобода высказываться и публиковаться
436	Свобода религии
437	Удовольствия, развлечения
438	Работа и учеба как творчество
439	Работа на благо общества
440	Работа как средство заработка
441	Счастливая семья
442	Любовь
443	Красота человека
444	Природа
445	Культура, искусство
[54]	УКАЖИТЕ СТОРОНЫ ЖИЗНИ, ИМЕЮЩИЕ ДЛЯ ВАС НЕКОТОРУЮ ЦЕННОСТЬ:
446	Общение с друзьями
447	Общение с родителями
448	Наличие вещей, имеющих продуктивную работу и учебу
449	Наличие престижных вещей
450	Наличие вещей, обеспечивающих комфортные условия жизни
451	Возможность влияния на ход важных событий в семье
452	Материальная независимость
453	Свобода высказываться и публиковаться
454	Свобода религии
455	Удовольствия, развлечения
456	Работа и учеба как творчество
457	Работа на благо общества
458	Работа как средство заработка
459	Счастливая семья
460	Любовь
461	Красота человека
462	Природа
463	Культура, искусство
[55]	УКАЖИТЕ СТОРОНЫ ЖИЗНИ, ИМЕЮЩИЕ ДЛЯ ВАС БОЛЬШУЮ ЦЕННОСТЬ:
464	Общение с друзьями
465	Общение с родителями
466	Наличие вещей, имеющих продуктивную работу и учебу
467	Наличие престижных вещей
468	Наличие вещей, обеспечивающих комфортные условия жизни
469	Возможность влияния на ход важных событий в семье

Продолжение таблицы 63

Код	Наименования шкал и градаций признаков
470	Материальная независимость
471	Свобода высказываться и публиковаться
472	Свобода религии
473	Удовольствия, развлечения
474	Работа и учеба как творчество
475	Работа на благо общества
476	Работа как средство заработка
477	Счастливая семья
478	Любовь
479	Красота человека
480	Природа
481	Культура, искусство

Задание 1.3. Обучающую выборку заполнить на основе данных по учащимся своей группы и дополнить данными параллельной группы.

Для этих целей каждый учащийся заполняет данными о себе "Карточку респондента", имеющую вид, представленный на таблице 64:

Таблица 64 – КАРТОЧКА РЕСПОНДЕНТА №

Учебная группа:							
Фамилия, имя, отчество:							
Коды классов, к которым он принадлежит:							
Коды признаков, которыми он характеризуется:							

При необходимости строки с кодами добавляются. На каждой карточке имеется номер студента по списку группы, который затем используется как номер анкеты в обучающей выборке. Необходимо сначала ввести пустые анкеты обучающей выборки по списку группы, а затем заполнить их, используя карточки. Учащиеся передают друг другу карточки передавать по компьютерному классу. В графе "Наим.физ.источника" указываются полностью: фамилия, и сокращенно: имя и отчество студента. Аналогично поступаем с карточками студентов 2-й группы, за исключением того, что для определения номера анкеты в

обучающей выборке к списочному номеру студента 2-й группы прибавляется номер последнего студента 1-й группы.

Задание 2. Осуществить синтез и верификацию (измерение адекватности) семантической информационной модели.

Синтез и верификация модели осуществляется в соответствующих подсистемах системы "Эйдос". Результаты верификации представлены в таблице 65.

Таблица 65 – РЕЗУЛЬТАТЫ ВЕРИФИКАЦИИ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИОННОЙ МОДЕЛИ

№	Код	Наименование класса	Ко-во лог. анк.	% верной иденти- фикации
ВЫСОКАЯ адекватность модели				
1	87	УСП.ПО ДИСЦ.ОПД-12 (социальная экология) - низкая	1	100,00
2	106	УСП.ПО ДИСЦ.СД-05 (социология культуры) - высокая	27	88,89
3	46	УСП.ПО ДИСЦ.ЕН-02 (конц.совр.естествознания) - высокая	35	88,57
4	55	УСП.ПО ДИСЦ.ОПД-02 (общая социология) - высокая	26	88,46
5	45	УСП.ПО ДИСЦ.ЕН-01 (математика и информатика) - низкая	8	87,50
6	85	УСП.ПО ДИСЦ.ОПД-12 (социальная экология) - высокая	24	87,50
7	13	УСП.ПО ДИСЦ.ГСД-01 (философия) - высокая	23	86,96
8	28	УСП.ПО ДИСЦ.ГСД-06 (право) - высокая	35	85,71
9	37	УСП.ПО ДИСЦ.ГСД-10 (психология и педагогика) - высокая	26	84,62
10	14	УСП.ПО ДИСЦ.ГСД-01 (философия) - средняя	6	83,33
11	16	УСП.ПО ДИСЦ.ГСД-02 (иностр.язык) - высокая	24	83,33
12	18	УСП.ПО ДИСЦ.ГСД-02 (иностр.язык) - низкая	6	83,33
СРЕДНЯЯ адекватность модели				
13	24	УСП.ПО ДИСЦ.ГСД-04 (история) - низкая	9	77,78
14	22	УСП.ПО ДИСЦ.ГСД-04 (история) - высокая	17	76,47
15	17	УСП.ПО ДИСЦ.ГСД-02 (иностр.язык) - средняя	12	75,00
16	56	УСП.ПО ДИСЦ.ОПД-02 (общая социология) - средняя	8	75,00
17	57	УСП.ПО ДИСЦ.ОПД-02 (общая социология) - низкая	8	75,00
18	43	УСП.ПО ДИСЦ.ЕН-01 (математика и информатика) - высокая	15	73,33
19	29	УСП.ПО ДИСЦ.ГСД-06 (право) - средняя	7	71,43
20	47	УСП.ПО ДИСЦ.ЕН-02 (конц.совр.естествознания) - средняя	7	71,43
НИЗКАЯ адекватность модели				
21	113	Группа-2	23	69,57
22	23	УСП.ПО ДИСЦ.ГСД-04 (история) - средняя	16	68,75
23	44	УСП.ПО ДИСЦ.ЕН-01 (математика и информатика) - средняя	19	68,42
24	39	УСП.ПО ДИСЦ.ГСД-10 (психология и педагогика) - низкая	3	66,67
25	112	Группа-1	19	63,16
26	15	УСП.ПО ДИСЦ.ГСД-01 (философия) - низкая	13	61,54
ОЧЕНЬ НИЗКАЯ адекватность модели				
27	107	УСП.ПО ДИСЦ.СД-05 (социология культуры) - средняя	11	54,55
28	38	УСП.ПО ДИСЦ.ГСД-10 (психология и педагогика) - средняя	13	53,85
29	86	УСП.ПО ДИСЦ.ОПД-12 (социальная экология) - средняя	17	52,94
30	108	УСП.ПО ДИСЦ.СД-05 (социология культуры) - низкая	4	50,00

Из таблицы 65 видно, что:

- из 113 классов сформированными оказались лишь 30, т.к. по остальным в обучающей выборке просто не было данных;
- если исключить из модели не сформированные классы, а также классы с дифференциальной валидностью (% верной идентификации) меньше 70% (строки с 21 по 30), то получившаяся в результате модель по 20 классам модель будет иметь среднюю адекватность не ниже 70%. Для *автоматизации* этой операции в системе "Эйдос" в подсистеме измерения адекватности имеется специальный режим. Необходимо отметить, что *этот прием аналогичен исключению из рассмотрения результатов обработки с низким доверительным интервалом, что широко используется в дискриминантном анализе.*

Задание 3. Провести системно-когнитивный анализ модели:

Задание 3.1. Решить задачи идентификации и прогнозирования (для себя).

Эти операции выполняются в 4-й подсистеме системы "Эйдос". Результаты выводятся в двух разрезах:

- индивидуальная универсальная характеристика конкретного респондента (рисунок 113);
- выборка респондентов, имеющих наивысшие сходство с заданным классом (рисунок 114).

Анализ карточек прогноза успеваемости показывает, что вариабельность внутри группы успевающих студентов гораздо выше, чем в группе неуспевающих. В результате в среднем уровень сходства конкретных респондентов с обобщенными образами классов, соответствующих высоким уровням успеваемости, гораздо ниже, чем с классами, соответствующими низким уровням успеваемости.

Это дает основание предложить *гипотезу*, что высокая успеваемость детерминируется менее жестко, чем низкая, т.е. существуют *конкретные факторы*, фактически *предопределяющие* низкую успеваемость, тогда как для высокой успеваемости можно говорить лишь о *системе факторов, способствующих* высокой успеваемости.

Конечно, чтобы подобные выводы имели достаточную научную достоверность необходима значительно большая статистика, чем использованная в данном учебном примере. Например, если бы подобное исследование было проведено хотя бы в масштабах КубГАУ (более 17000 студентов), то тогда уже с достаточной уверенностью можно было бы говорить о реально обнаруженных закономерностях.

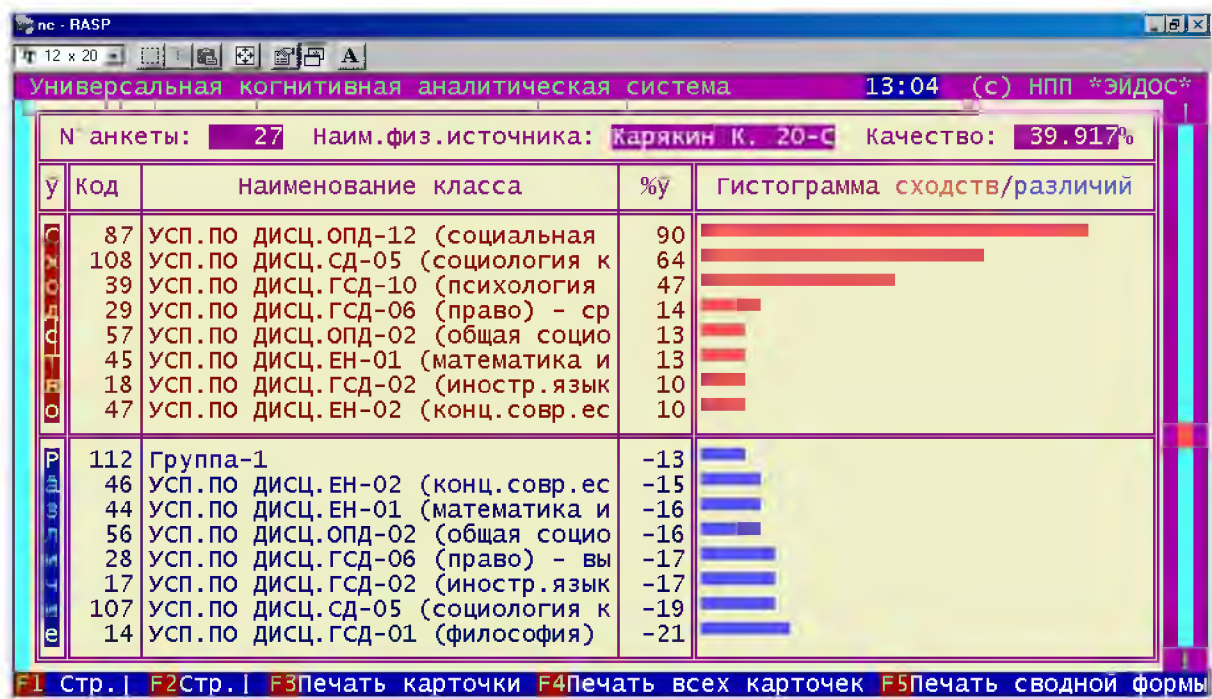


Рисунок 113. Прогноз успеваемости по различным предметам для конкретного студента



Рисунок 114. Список респондентов в порядке убывания сходства с заданным классом

Формы, подобные представленной на рисунке 113, могут интересовать потенциальных работодателей, а также приемную комиссию вуза в качестве дополнительной информации для принятия решения.

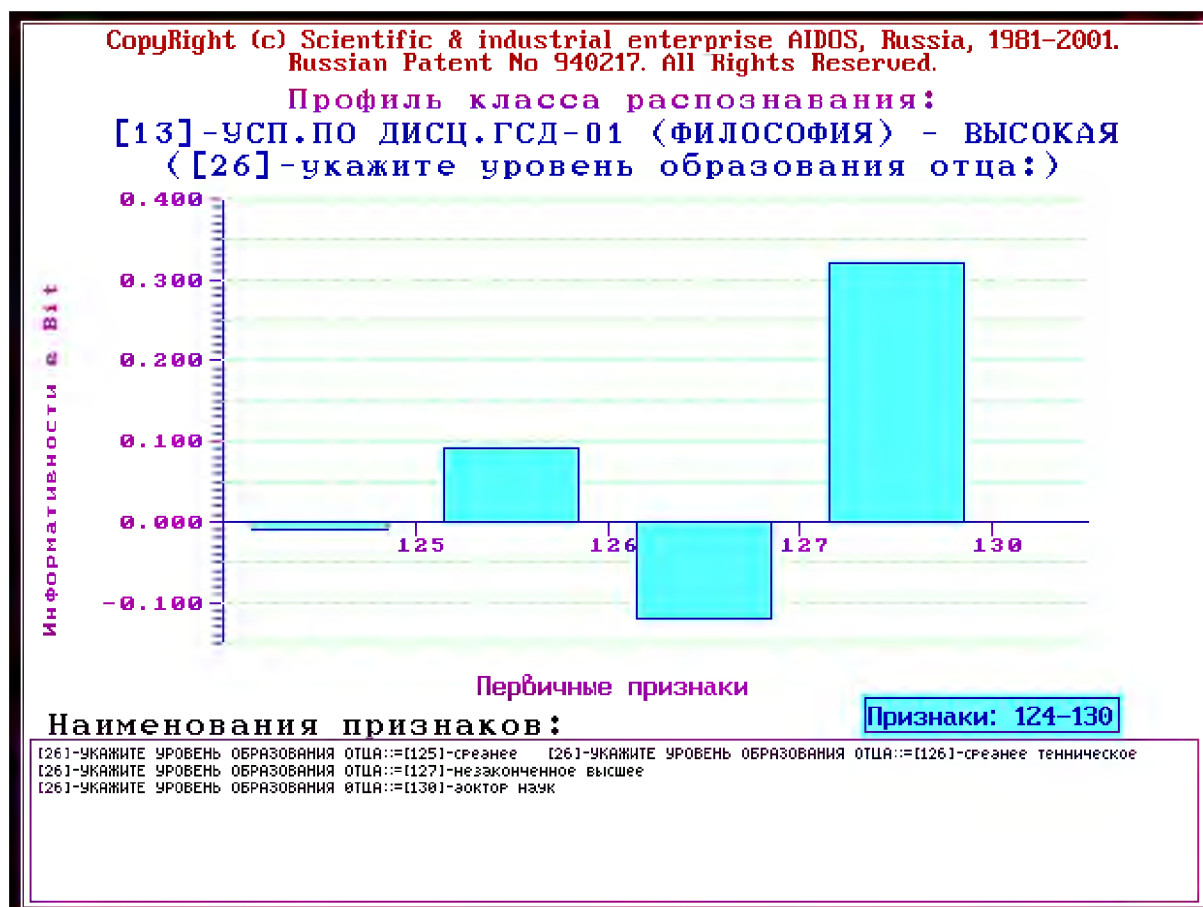
Задание 3.2. Сгенерировать информационные портреты классов и факторов, т.е. решить обратную задачу прогнозирования.

Определим, как влияют "ФОРМЫ ДОСУГА, КОТОРЫЕ ВЫ ИСПОЛЬЗУЕТЕ ПО ВОЗМОЖНОСТИ" на успеваемость по философии. Для этого в 1-м режиме 5-й подсистемы системы "Эйдос" получим информационный портрет класса с кодом 13: УСП.ПО ДИСЦ.ГСД-01 (философия) – высокая и нажав клавишу F6 зададим фильтр по диапазону градаций названной шкалы с кодом 50: {347, 373}. В результате получим информационный портрет данного класса, представленный в таблице 66.

Таблица 66 – ИНФОРМАЦИОННЫЙ ПОРТРЕТ КЛАССА, КОД: 13

Наименование класса: УСП.ПО ДИСЦ.ГСД-01 (философия) – высокая				
ФИЛЬТР по описательной шкале: коды градаций: 347-379, Positive				
Код шкалы : 50 Формы досуга, которые Вы используете по возможности				
15-08-04 11:16:45		г.Краснодар		
Код	Наименования	Инфор-	Инфор-	Суммар
приз	ОБОБЩЕННЫХ и первичных признаков	мат-ть	мат-ть	инф-ть
нака		Бит.	%	%%
=====				
362	Посещение концертов классической музыки..	0.320	4.69	164.2
372	Просмотр религиозных программ.....	0.320	4.69	168.8
373	Посещение религиозных центров.....	0.320	4.69	173.5
355	Дополнительное образование.....	0.203	2.98	251.5
360	Посещение оперы и балета.....	0.177	2.59	267.6
368	Просмотр спортивных ТВ-программ.....	0.159	2.33	290.0
364	Слушание радио.....	0.152	2.22	294.5
353	Работа на дому.....	0.135	1.98	316.5
357	Занятия искусством.....	0.135	1.98	318.5
359	Посещение музеев и выставок.....	0.135	1.98	320.5
361	Посещение драматических театров.....	0.135	1.98	322.4
351	Общение с родителями.....	0.096	1.41	356.8
348	Чтение художественной литературы.....	0.076	1.12	366.9
365	Слушание аудио-музыки.....	0.041	0.59	393.2
350	Дискотеки и т.п.....	0.025	0.37	398.0
352	Общение с любимым человеком.....	0.018	0.27	399.1
367	Просмотр развлекательных ТВ-программ.....	0.018	0.27	399.3
347	Просмотр видео и кинофильмов.....	0.012	0.18	400.4
~~~~~				
349	Спорт, туризм.....	-0.031	-0.45	404.3
371	Просмотр познавательных передач.....	-0.031	-0.45	404.8
366	Просмотр теленовостей.....	-0.049	-0.72	412.6
358	Коллекционирование.....	-0.125	-1.83	522.9
363	Посещение концертов попмузыки.....	-0.125	-1.83	524.7
369	Просмотр мультфильмов.....	-0.125	-1.83	526.6
370	Просмотр телесериалов.....	-0.125	-1.83	528.4
=====				
Универсальная когнитивная аналитическая система			НПП *Эйдос*	

Профиль класса в разрезе по диапазону факторов можно получить в 4-м режиме 6-й подсистемы системы "Эйдос" (рисунок 115).



**Рисунок 115. Профиль влияния образовательного уровня отца на отличную успеваемость студента по философии**

Из этого рисунка видно, что на отличную успеваемость студента по философии наиболее положительно сказывается если его отец – доктор наук.

В этой связи позволим себе немного повеселить читателей следующей историей. Профессор спрашивает абитуриента на вступительном экзамене:

– из каких соображений Вы выбрали для поступления именно наш вуз и именно эту специальность?

– не задавай глупых вопросов, папа!

отвечает абитуриент.

Оставшуюся часть задания студентам предлагается выполнить самостоятельно.

**Задание 3.3. Провести кластерно-конструктивный анализ классов и факторов (результаты отобразить в форме семантических сетей классов и факторов).**

В 5-й подсистеме системы "Эйдос" получены следующие семантические сети классов, отражающие степень их сходства по детерминирующим их факторам (рисунок 116), и факторов, отражающие, их сходство и различие по влиянию на уровень учебных достижений учащихся (рисунок 117).

CopyRight (c) Scientific & industrial enterprise AIDOS, Russia, 1981-2001.  
Russian Patent No 940217. All Rights Reserved.

## 2D - СЕМАНТИЧЕСКАЯ СЕТЬ КЛАССОВ

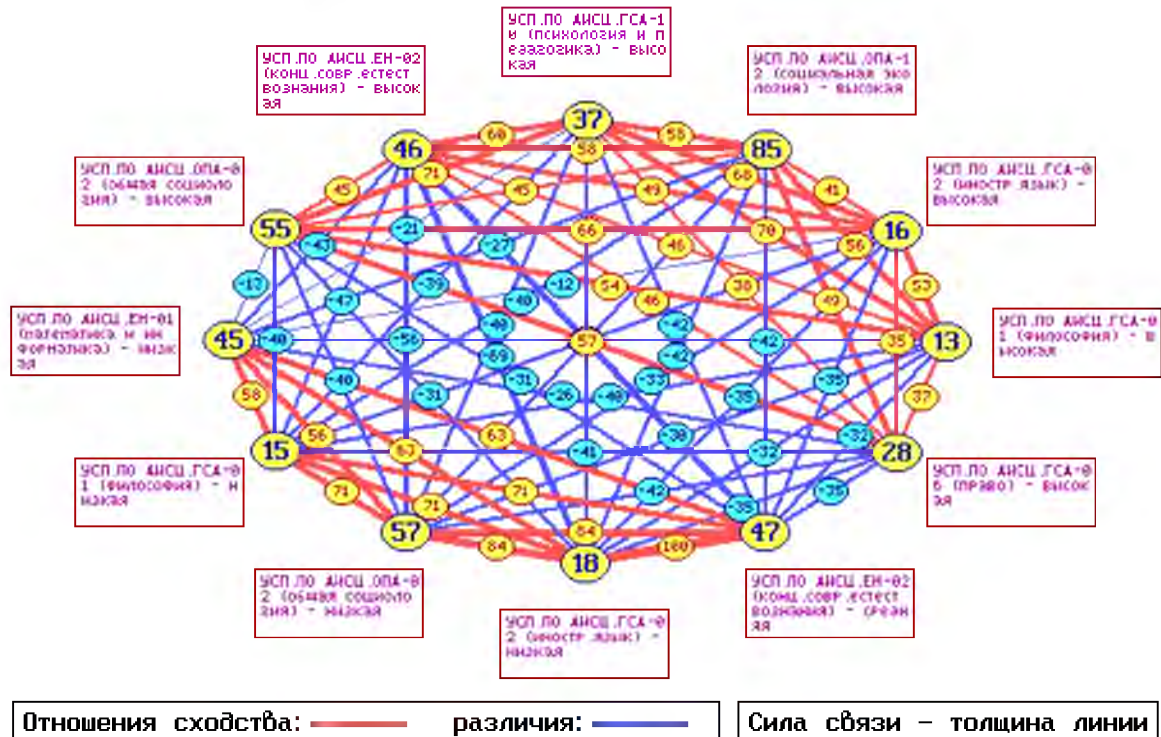


Рисунок 116. Пример семантической сети классов

CopyRight (c) Scientific & industrial enterprise AIDOS, Russia, 1981-1997.  
Russian Patent No 940217. All Rights Reserved.

## 2D - СЕМАНТИЧЕСКАЯ СЕТЬ ПРИЗНАКОВ

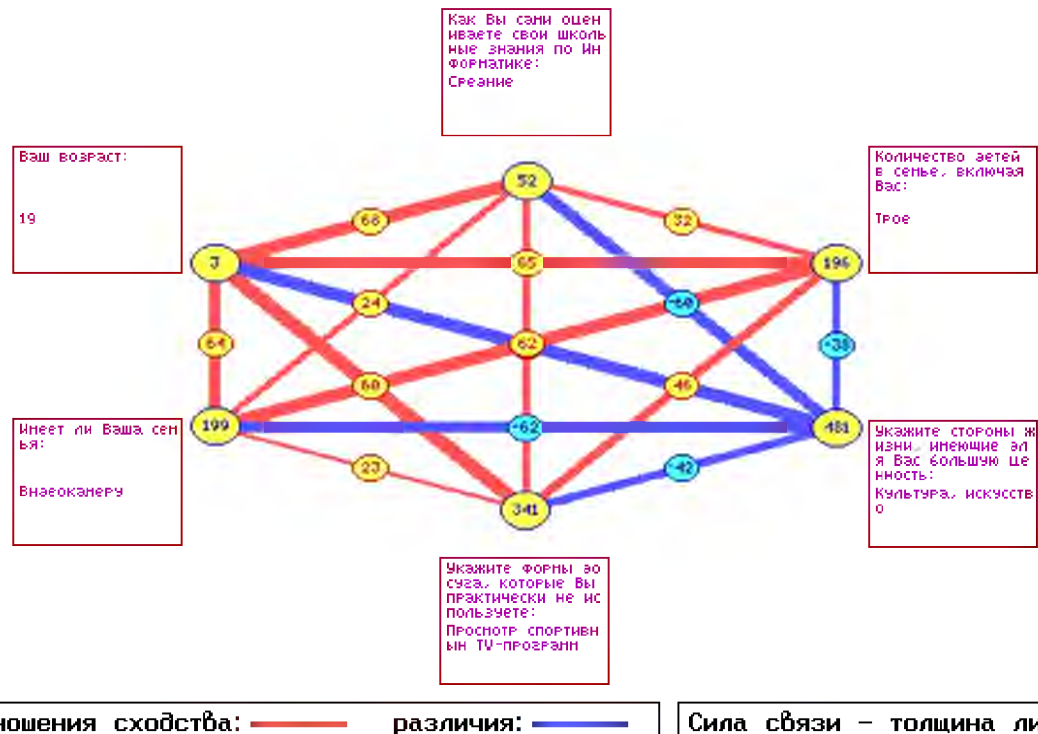
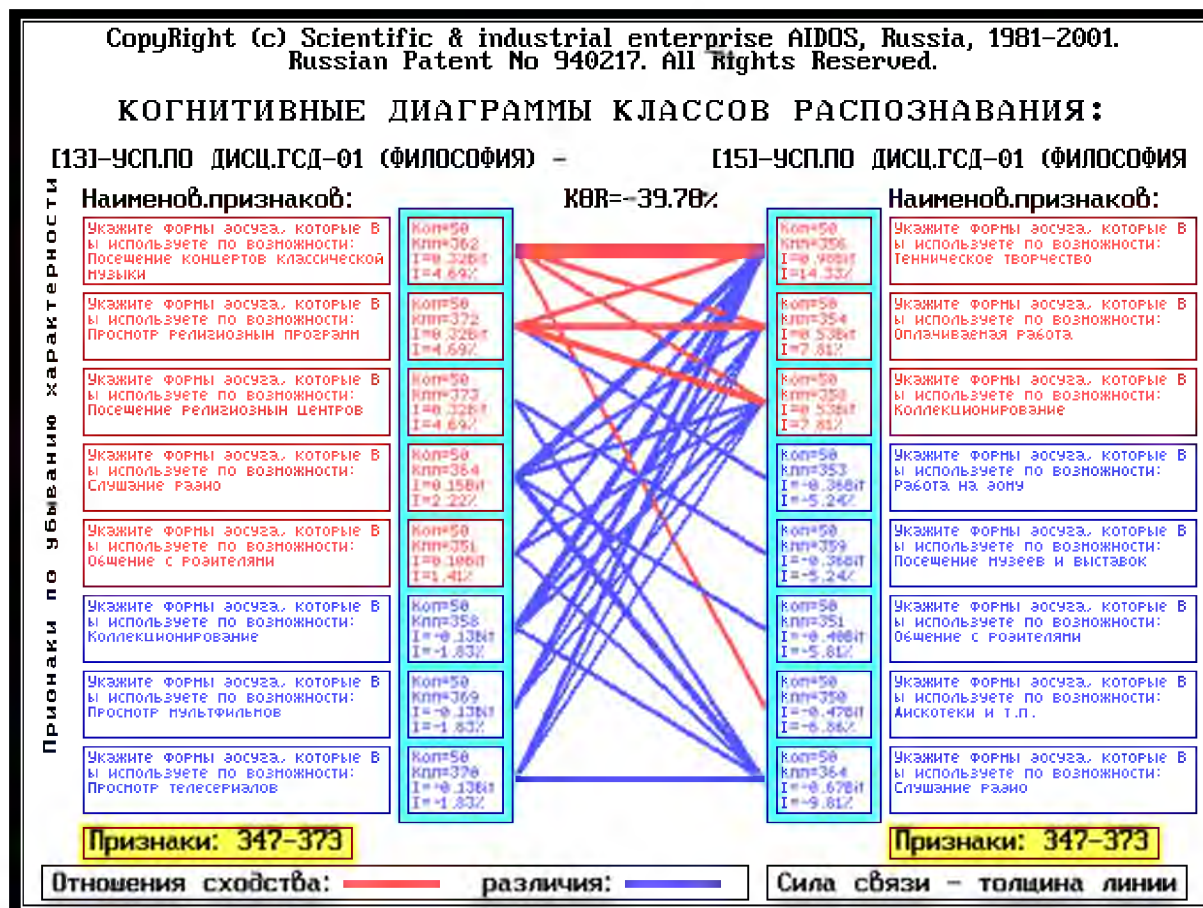


Рисунок 117. Пример семантической сети факторов



**Задание 3.4. Осуществить содержательное сравнение классов и факторов (результаты отобразить в форме когнитивных диаграмм классов и факторов).**

Для этого в 3-й функции 1-го режиме 5-й подсистемы системы "Эйдос" зададим коды классов 13 и 15 (высокая и низкая успеваемость по философии) и фильтр по факторам для обоих портретов от 347 до 373 (влияние образовательного уровня отца). В результате получим когнитивную диаграмму, представленную на рисунке 118.



**Рисунок 118. Когнитивная диаграмма, показывающая результаты содержательного сравнения двух классов по системам их детерминации**

Проинтерпретировать данную когнитивную диаграмму студенты должны самостоятельно.

**Задание 3.5. Построить нелокальные нейроны и интерпретируемые нейронные сети.**

Это задание предлагается выполнить самостоятельно. При этом использовать 6-й режим 6-й подсистемы системы "Эйдос".

**Задание 3.6. Построить классические когнитивные модели (отобразить в форме когнитивных карт).**

В 7-м режиме 6-й подсистемы системы "Эйдос" при задании кода класса 13 и диапазона кодов факторов 1 – 11 генерируется следующая классическая когнитивная карта (рисунок 119).

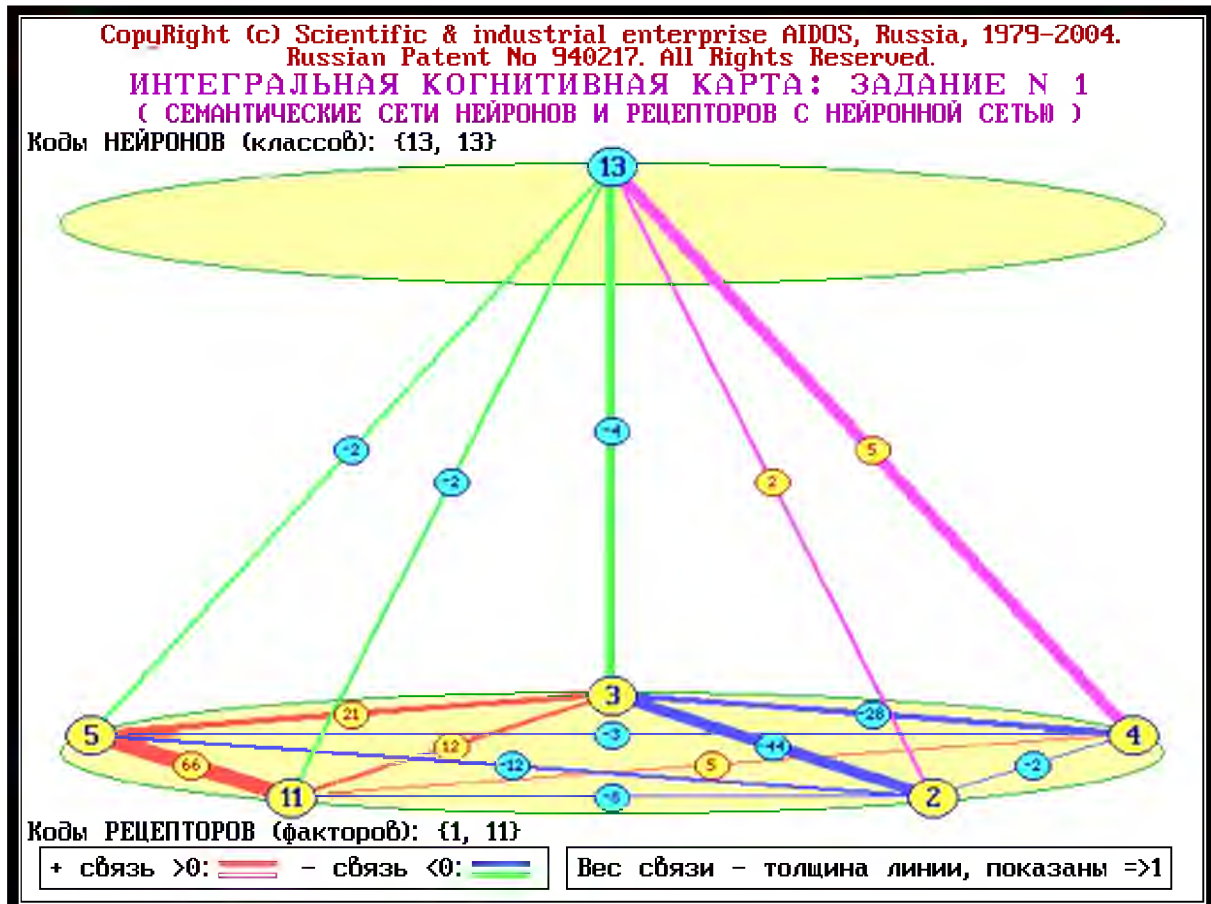


Рисунок 119. Пример классической когнитивной карты

**Задание 3.7. Построить интегральные когнитивные модели (отобразить в форме интегральных когнитивных карт).**

В 7-м режиме 6-й подсистемы системы "Эйдос" при задании диапазона кода классов {13, 15} и диапазона кодов факторов {1, 11} генерируется следующая классическая когнитивная карта (рисунок 120), которая, по сути, является суперпозицией нескольких классических когнитивных карт.

Из этого рисунка видно, что факторы с кодами 3, 5, 11 сходны по влиянию на учебные достижения по философии:

- они все препятствуют достижению результата с кодом 13;
- они все способствуют достижению результата с кодом 15;
- результаты с кодами 13 и 15 детерминируются несовместимыми системами факторов и одновременно недостижимы, на что указывает антикорреляция между ними в семантической сети классов;
- фактор 3 способствует результату 14 и препятствует результату 13, а фактор 2, наоборот, препятствует 14 и способствует 13, и факторы 2 и 3 имеют

различное влияние на поведение объекта управления, на что указывает антикорреляция между ними в семантической сети факторов.

Насколько известно, система "Эйдос" на данный момент является единственной системой, обеспечивающей автоматический синтез непосредственно на основе эмпирических данных и отображение в графической форме классических и интегральных когнитивных карт (интегральные когнитивные карты впервые предложены автором).

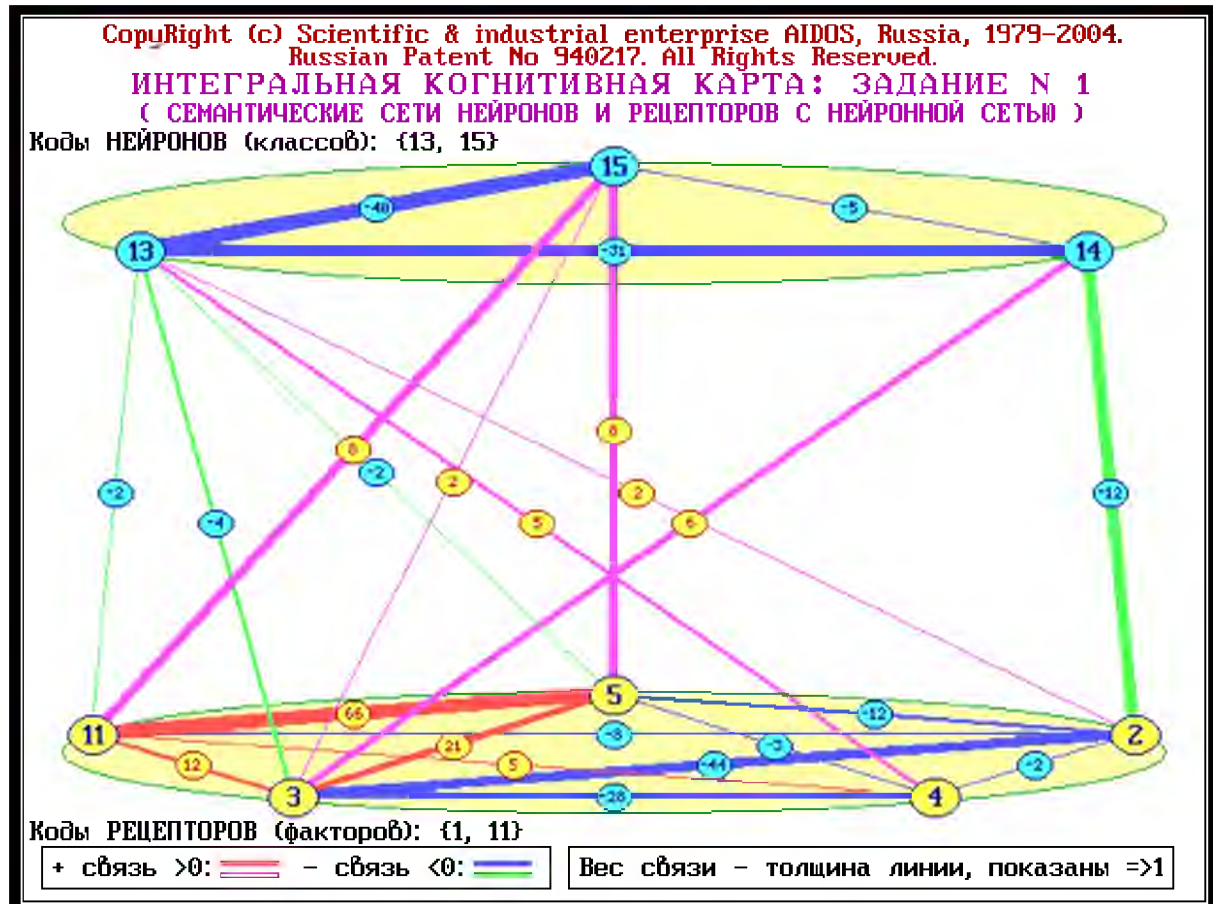


Рисунок 120. Пример интегральной когнитивной карты

### Контрольные вопросы

1. Какие работы выполняются на этапе формализации задачи?
2. Как в системе "Эйдос" ввести классификационные шкалы и градации, выбрав в качестве классов – различные уровни учебных достижений по различным дисциплинам, перечень которых взять из зачетной книжки?
3. Как в системе "Эйдос" ввести описательные шкалы и градации, используя характеристики подчерка?
4. Каким образом подготовить и ввести в систему "Эйдос" обучающую выборку?
5. Как осуществить синтез и верификацию (измерение адекватности) семантической информационной модели в системе "Эйдос"?
6. Что включает системно-когнитивный анализ модели?



7. Как решаются задачи идентификации и прогнозирования в системе "Эйдос"?

8. В каких подсистемах и режимах системы "Эйдос" генерируются информационные портреты классов и факторов и отображаются в графической форме двумерных и трехмерных профилей классов и факторов)?

9. Каким образом в системе "Эйдос" провести кластерно-конструктивный анализ классов и факторов и отобразить его в форме семантических сетей классов и факторов?

10. Как в системе "Эйдос" осуществить содержательное сравнение классов и факторов и отобразить результаты в форме когнитивных диаграмм классов и факторов?

11. Как в системе "Эйдос" построить нелокальные нейроны и интерпретируемые нейронные сети?

12. Как в системе "Эйдос" построить классические когнитивные модели и отобразить их в форме когнитивных карт?

13. Как в системе "Эйдос" построить интегральные когнитивные модели и отобразить в форме интегральных когнитивных карт?

14. Кто такой "респондент"?

15. Что понимается под "социальным статусом" респондента?

### ***Литература по лабораторной работе***

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.

2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

**ЛР-5:****"Идентификация слов по входящим в них буквам"**

Первую лабораторную работу рассмотрим более подробно, чтобы студенты могли ознакомиться с базовой системой "Эйдос", на которой основан курс лекций и лабораторных работ.

В данной работе исследуется возможность идентификации слов по входящим в них буквам. При этом слова рассматриваются как классы распознавания, а буквы – как признаки. Подробно описана технология синтеза и анализа модели с применением инструментария СК-анализа – системы "Эйдос", приведены видеограммы, табличные и графические выходные формы, причем впервые – классическая и интегральная когнитивные карты, генерируемые в версии 12.5. Продемонстрировано на численных примерах, что при идентификации слов главную роль играет информация о входящих в них буквах, а последовательность букв и их сочетаний не играет особой роли. Продемонстрирована устойчивость модели от неполноты и зашумленности информации.

***Краткая теория***

Язык является мощным средством моделирования реальности. Иерархическая структура языка, включающего в частности такие структуры, как символы, слова, предложения и тексты, является ярким примером иерархической структуры обработки информации, обеспечивающей адекватное отражение структуры реальности. Описание некоторых объектов на естественном языке является их моделированием и позволяет решать задачи идентификации, прогнозирования, сравнения и классификации этих объектов. Автоматизированный синтез моделей объектов, описанных на естественном языке, представляет большой интерес для теории и практики систем искусственного интеллекта.

Эффективность языка, как системы моделирования, удивительна: 32 буквы русского алфавита позволяют составить более 40 тысяч слов русского языка, из которых может быть образовано огромное количество осмысленных предложений.

Каждый символ содержит некоторое количество информации о словах, предложениях и текстах, в которые он входит, каждое слово – о предложениях и текстах, и каждое предложение – о текстах. Однако в буквах содержится больше информации о словах, в словах – о предложениях, в предложениях – о текстах. Поэтому на основе анализа букв целесообразно идентифицировать лишь слова, а на основе слов – предложения, на основе предложений – тексты.

*В данной работе исследуется возможность идентификации слов по входящим в них буквам. При этом слова рассматриваются как классы распознавания, а буквы – как признаки. Кроме того добавлен еще признак слов - количество букв в слове.*

Эта задача проста и наглядна. Поэтому она рекомендуется в качестве первой лабораторной работы для освоения инструментария системно-когнитивного анализа – универсальной когнитивной аналитической системы "Эйдос" [81, 125].

Подобные задачи имеют большое практическое значение и решаются в ряде систем, например, в редакторе Word – при проверке орфографии и подборе

рекомендуемых слов для замены, в системе FineReader – для поиска слов с неверно распознанными символами и др.

### **Задание**

1. Создать файл в стандарте DOS-текст с концами строк, записать его в поддиректорию DOB.
2. Сгенерировать классификационные и описательные шкалы и градации, а также обучающую выборку.
3. Осуществить синтез и верификацию модели.
4. Провести анализ устойчивости модели к неполноте информации и наличию шума.
5. Проверить способность модели правильно идентифицировать классы, один из которых является подмножеством другого.
6. Оценить ценность букв для идентификации слов. Сравнить суммарную ценность для этой цели гласных и согласных букв.
7. Выполнить кластерно-конструктивный анализ слов и букв, вывести информационные и семантические портреты слов и букв, построить их профили.
8. Вывести в графической форме семантические сети и когнитивные диаграммы слов и букв, а также классическую и интегральную когнитивные карты.

### **Пример решения**

#### ***Пример решения задания 1: "Создать файл в стандарте DOS-текст с концами строк, записать его в поддиректорию DOB"***

Текстовый файл создается в редакторе Word или MultiEdit. Если он создан в Word, то при сохранении выбирается режим: "Файл – Сохранить как – Тип файла: Текст DOS с разбиением на строки". Имя файла произвольное, но удовлетворяющее требованиям DOS. Этот файл каждый студент создает самостоятельно из двух – трех абзацев текста. Например, это может быть краткая биография студента или текст может быть взят из какого-либо файла, имеющегося на компьютере. Могут использоваться и русский, и латинский алфавиты, а также цифры. Регистр роли не играет.

Пример текста файла:

**Молоко око срок окорок огород В. Г. Белинский высоко и по достоинству оценил гениальное творение нашего национального поэта. Великий критик писал: "Пусть идет время и приводит с собой новые потребности, пусть растет русское общество и обгоняет "Онегина": как бы далеко оно ни ушло, всегда будет оно любить эту поэму, всегда будет останавливать на ней исполненный любви и благодарности взор...".**

Студентам рекомендуется выделить блоком текст в рамке, скопировать его в буфер обмена, создать новый файл, вставить в него этот текст из буфера обмена и записать как "Текст DOS" в поддиректорию DOB директории System, в которой находится система "Эйдос". Слова из этого файла будут использованы системой для выполнения следующего этапа работы.

**Пример решения задания 2: "Сгенерировать классификационные и описательные шкалы и градации, а также обучающую выборку"**

Генерация классификационных и описательных шкал и градаций осуществляется в соответствующих режимах системы "Эйдос": "Автовод первичных признаков и ТХТ-файлов", "F5 Признаки – Буквы". (рисунки 121 и 122):

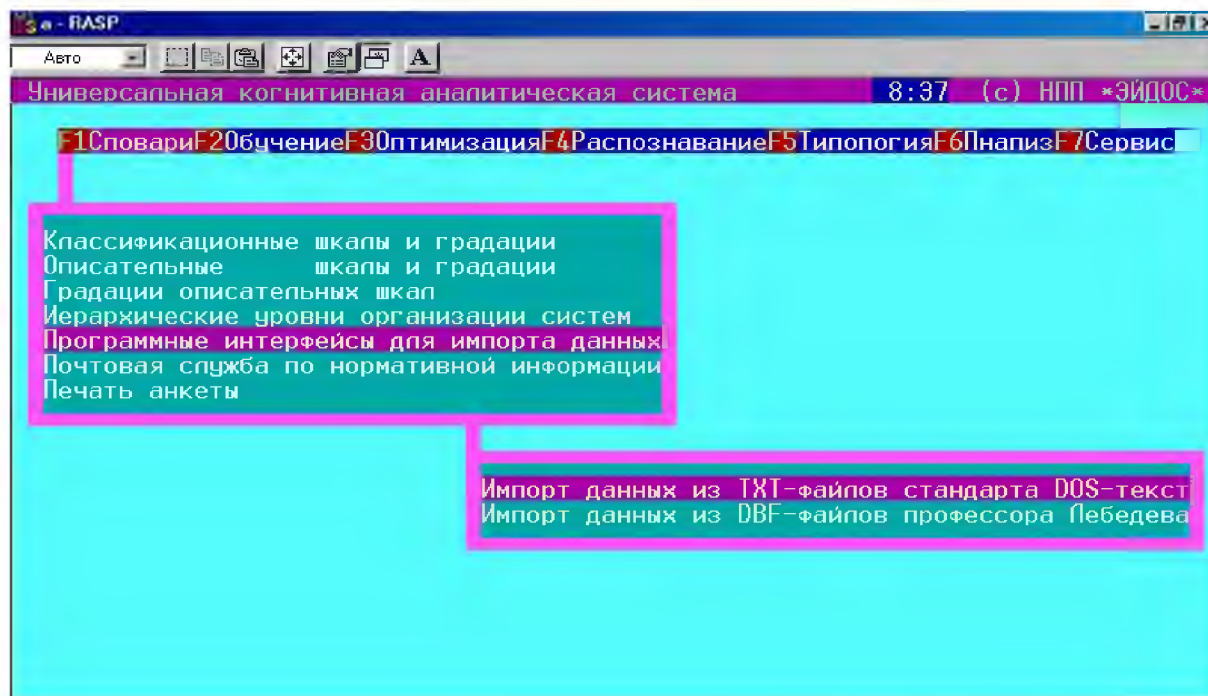


Рисунок 121. Выбор режима "Импорт данных из ТХТ-файлов стандарта DOS-текст"

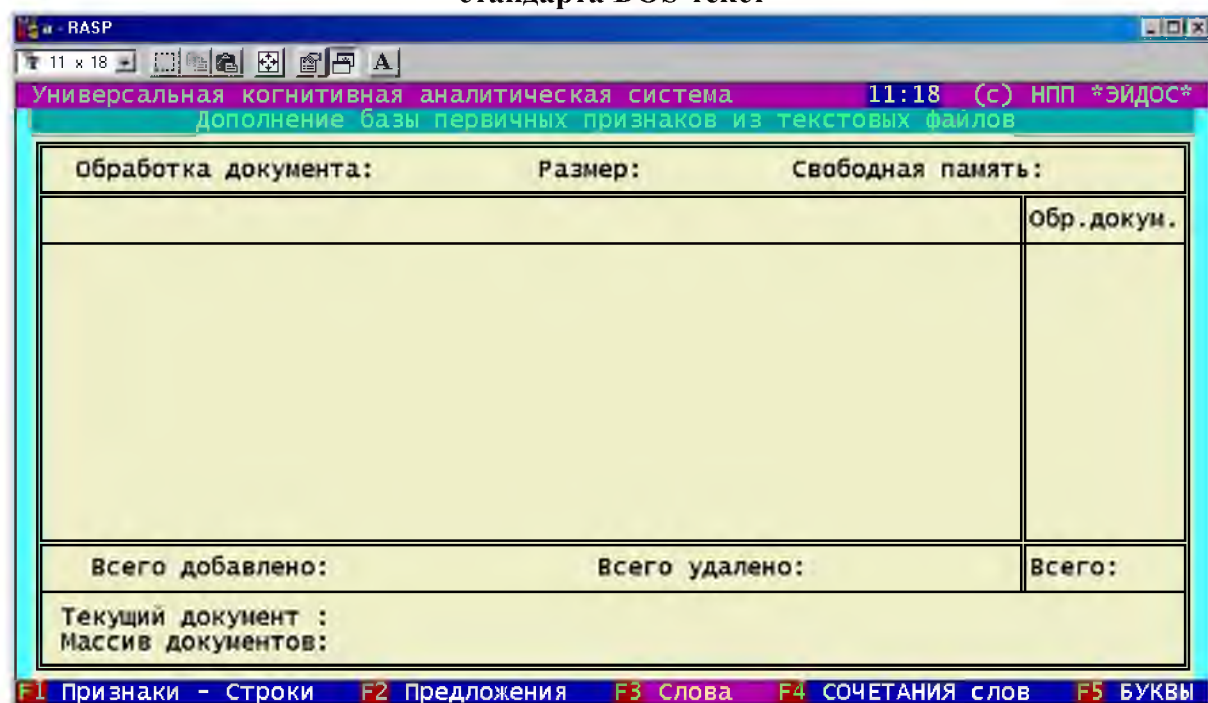


Рисунок 122. Выбор режима "Признаки – F5 Буквы"

В результате будут автоматически сгенерированы классификационные и описательные шкалы и градации, а также обучающая выборка (приведена полностью) (таблицы 67, 68, 69).

**Таблица 67 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ И ГРАДАЦИИ (СЛОВА)**

Код	Слово	Код	Слово	Код	Слово
1	БЕЛИНСКИЙ	16	ЛЮБВИ	31	ОЦЕНИЛ
2	БЛАГОДАРНОСТИ	17	ЛЮБИТЬ	32	ПИСАЛ
3	БУДЕТ	18	МОЛОКО	33	ПОТРЕБНОСТИ
4	ВЕЛИКИЙ	19	НАЦИОНАЛЬНОГО	34	ПОЭМУ
5	ВЗОР	20	НАШЕГО	35	ПОЭТА
6	ВРЕМЯ	21	НЕИ	36	ПРИВОДИТ
7	ВСЕГДА	22	НОВЫЕ	37	ПУСТЬ
8	ВЫСОКО	23	ОБГОНЯЕТ	38	РАСТЕТ
9	ГЕНИАЛЬНОЕ	24	ОБЩЕСТВО	39	РУССКОЕ
10	ДАЛЕКО	25	ОГОРОД	40	СОБОЙ
11	ДОСТОИНСТВУ	26	ОКО	41	СРОК
12	ИДЕТ	27	ОКОРОК	42	ТВОРЕНИЕ
13	ИСПОЛНЕННЫЙ	28	ОНЕГИНА	43	УШЛО
14	КАК	29	ОНО	44	ЭТУ
15	КРИТИК	30	ОСТАНАВЛИВАТЬ		

**Таблица 68 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ (БУКВЫ)**

Код	Буква	Код	Буква	Код	Буква	Код	Буква
1	0	18	Н	35	У	52	П
2	1	19	І	36	Z	53	Р
3	2	20	J	37	A	54	С
4	3	21	K	38	Б	55	Т
5	4	22	L	39	В	56	У
6	5	23	M	40	Г	57	Ф
7	6	24	N	41	Д	58	Х
8	7	25	O	42	Е	59	Ц
9	8	26	P	43	Ж	60	Ч
10	9	27	Q	44	З	61	Ш
11	A	28	R	45	И	62	Щ
12	B	29	S	46	Й	63	Ъ
13	C	30	T	47	К	64	Ы
14	D	31	U	48	Л	65	Ь
15	E	32	V	49	М	66	Э
16	F	33	W	50	Н	67	Ю
17	G	34	X	51	О	68	Я

Кроме того как признак используется количество букв в слове (таблица 68-1). Однако в данной работе студентам предлагается удалить из справочника

признаков (градации описательных шкал) признаки, приведенные в этой таблице, чтобы не учитывать их влияние и более явно проследить именно роль букв на идентификацию слов в различных условиях.

**Таблица 68-1 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ  
(КОЛ-ВО БУКВ В СЛОВЕ)**

Код	Наим.признака
69	В слове 1 симв.
70	В слове 2 симв.
71	В слове 3 симв.
72	В слове 4 симв.
73	В слове 5 симв.
74	В слове 6 симв.
75	В слове 7 симв.
76	В слове 8 симв.
77	В слове 9 симв.
78	В слове 10 симв.
79	В слове 11 симв.
80	В слове 12 симв.
81	В слове 13 симв.

Максимальное количество символов в слове определяется по словам обучающей выборки.

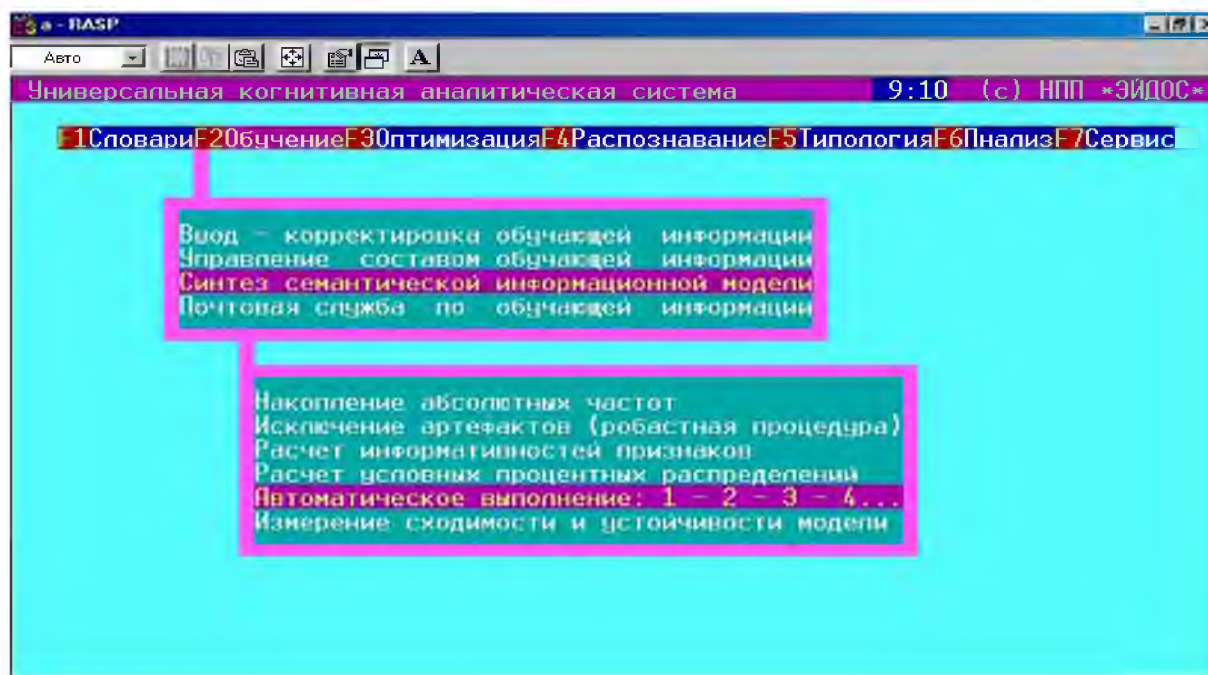
**Таблица 69 – ОБУЧАЮЩАЯ ВЫБОРКА (ФРАГМЕНТ)**

№	Класс	Коды признаков													
1	БЕЛИНСКИЙ	38	42	48	45	50	54	47	45	46	77				
2	БЛАГОДАРНОСТИ	38	48	37	40	51	41	37	53	50	51	54	55	45	81
3	БУДЕТ	38	56	41	42	55	73								
4	ВЕЛИКИЙ	39	42	48	45	47	45	46	75						
5	ВЗОР	39	44	51	53	72									
6	ВРЕМЯ	39	53	42	49	68	73								
7	ВСЕГДА	39	54	42	40	41	37	74							
8	ВЫСОКО	39	64	54	51	47	51	74							
9	ГЕНИАЛЬНОЕ	40	42	50	45	37	48	65	50	51	42	78			
10	ДАЛЕКО	41	37	48	42	47	51	74							
11	ДОСТОИНСТВУ	41	51	54	55	51	45	50	54	55	39	56	79		
12	ИДЕТ	45	41	42	55	72									
13	ИСПОЛНЕННЫЙ	45	54	52	51	48	50	42	50	50	64	46	79		
14	КАК	47	37	47	71										
15	КРИТИК	47	53	45	55	45	47	74							
16	ЛЮБВИ	48	67	38	39	45	73								
17	ЛЮБИТЬ	48	67	38	45	55	65	74							
18	МОЛОКО	49	51	48	51	47	51	74							
19	НАЦИОНАЛЬНОГО	50	37	59	45	51	50	37	48	65	50	51	40	51	81
20	НАШЕГО	50	37	61	42	40	51	74							
21	НЕЙ	50	42	46	71										
22	НОВЫЕ	50	51	39	64	42	73								
23	ОБГОНЯЕТ	51	38	40	51	50	68	42	55	76					
24	ОБЩЕСТВО	51	38	62	42	54	55	39	51	76					
25	ОГОРОД	51	40	51	53	51	41	74							

26	ОКО	51	47	51	71										
27	ОКОРОК	51	47	51	53	51	47	74							
28	ОНЕГИНА	51	50	42	40	45	50	37	75						
29	ОНО	51	50	51	71										
30	ОСТАНАВЛИВАТЬ	51	54	55	37	50	37	39	48	45	39	37	55	65	81
31	ОЦЕНИЛ	51	59	42	50	45	48	74							
32	ПИСАЛ	52	45	54	37	48	73								

### ***Пример решения задания 3: " Осуществить синтез и верификацию модели"***

Синтез модели осуществляется на основе обучающей выборки (таблица 69) в соответствующем режиме (рисунок 123) после формирования классификационных и описательных шкал и градаций и обучающей выборки.



**Рисунок 123. Режим "Синтез семантической информационной модели" системы "Эйдос"**

В результате синтеза семантической информационной модели рассчитываются две ее основные матрицы: матрица абсолютных частот (таблица 70) и матрица информативностей (таблица 71).

В матрице частот отражается суммарное количество встреч каждого признака у объектов каждого класса по данным обучающей выборки. Каждый элемент матрицы информативностей показывает, какое количество информации о переходе объекта управления в состояние, соответствующее классу, мы получаем, если узнаем, что действует некоторый признак.

**Таблица 70 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ (ФРАГМЕНТ)**

Коды букв	Коды классов (слов)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
37		2					1		1	1				1					2
38	1	1	1													1	1		
39				1	1	1	1	1			1					1			
40		1					1		1										1
41		1	1				1			1	1	1							
42	1		1	1		1	1		2	1		1	1						
43																			
44					1														
45	2	1		2					1		1	1	1		2	1	1		1
46	1			1									1						
47	1			1				1		1				2	2			1	
48	1	1		1					1	1			1			1	1	1	1
49						1												1	
50	1	1							2		1		3						3
51		2			1			2	1	1	2		1					3	3
52													1						
53		1			1	1									1				
54	1	1					1	1			2		1						
55		1	1								2	1			1		1		
56			1								1								

Таблица 71 – МАТРИЦА ИНФОРМАТИВНОСТЕЙ (ФРАГМЕНТ)

Ко- ды букв	Коды классов (слов)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
37		1,0					1,1		0,6	1,1				1,7					1,0
38	1,2	0,9	1,8													1,8	1,6		
39				1,1	1,6	1,4	1,3	1,3			0,7					1,4			
40		1,0					1,7		1,2										1,0
41		1,0	1,9				1,7			1,7	1,1	2,1							
42	0,4		0,9	0,6		0,9	0,7		0,9	0,7		1,1	0,2						
43																			
44					4,1														
45	1,0			1,3					0,3		0,2	1,1	0,2		1,4	0,9	0,7		
46	1,8			2,0									1,6						
47	0,8			1,0				1,2		1,2				2,5	1,9			1,2	
48	0,8	0,4		1,0					0,7	1,2			0,6			1,4	1,2	1,2	0,4
49						2,8												2,7	
50	0,4								0,9		0,2		1,2						1,1
51					0,4			0,7	-0,5		0,1		-0,6					1,1	0,3
52													1,3						
53		0,6			1,7	1,5									1,3				
54	0,6	0,2					1,0	1,0			1,1		0,4						
55		0,1	1,0								0,9	1,2			0,8		0,8		
56			2,0								1,3								



Верификация модели осуществляется путем копирования обучающей выборки в распознаваемую (рисунок 124), распознавания (рисунок 125, 126) и измерения дифференциальной и интегральной валидности (рисунок 127):

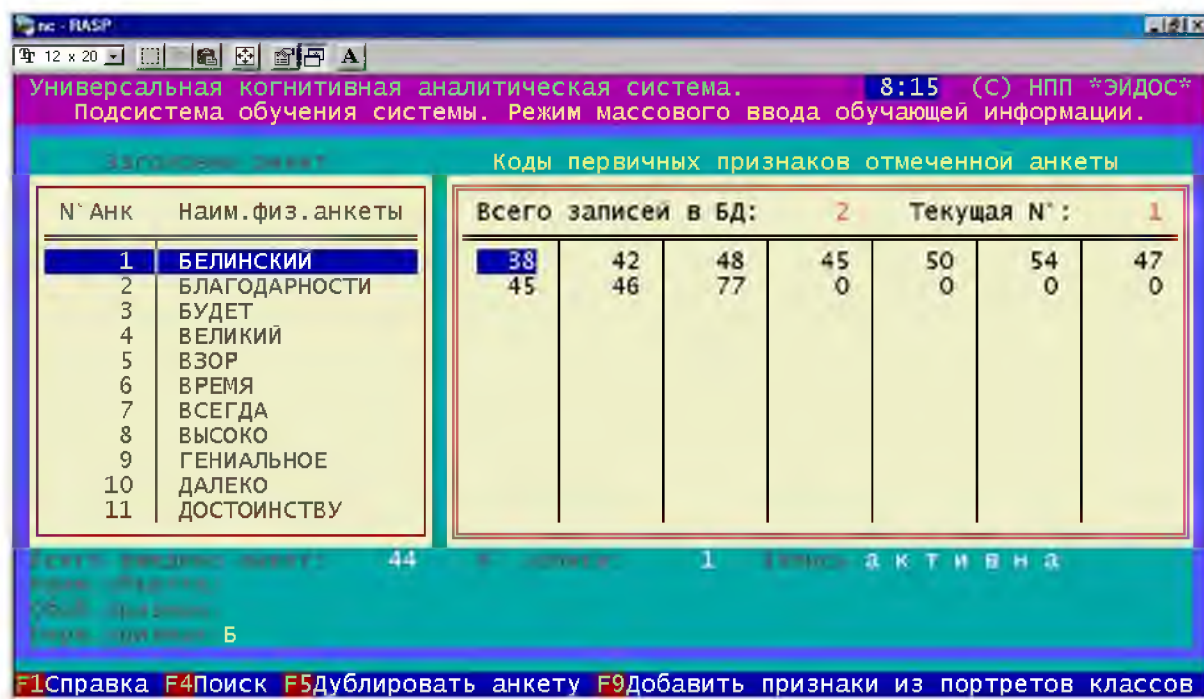


Рисунок 124. Режим "Ввод-корректировка обучающей выборки" системы "Эйдос" (копирование ее в распознаваемую – F5)

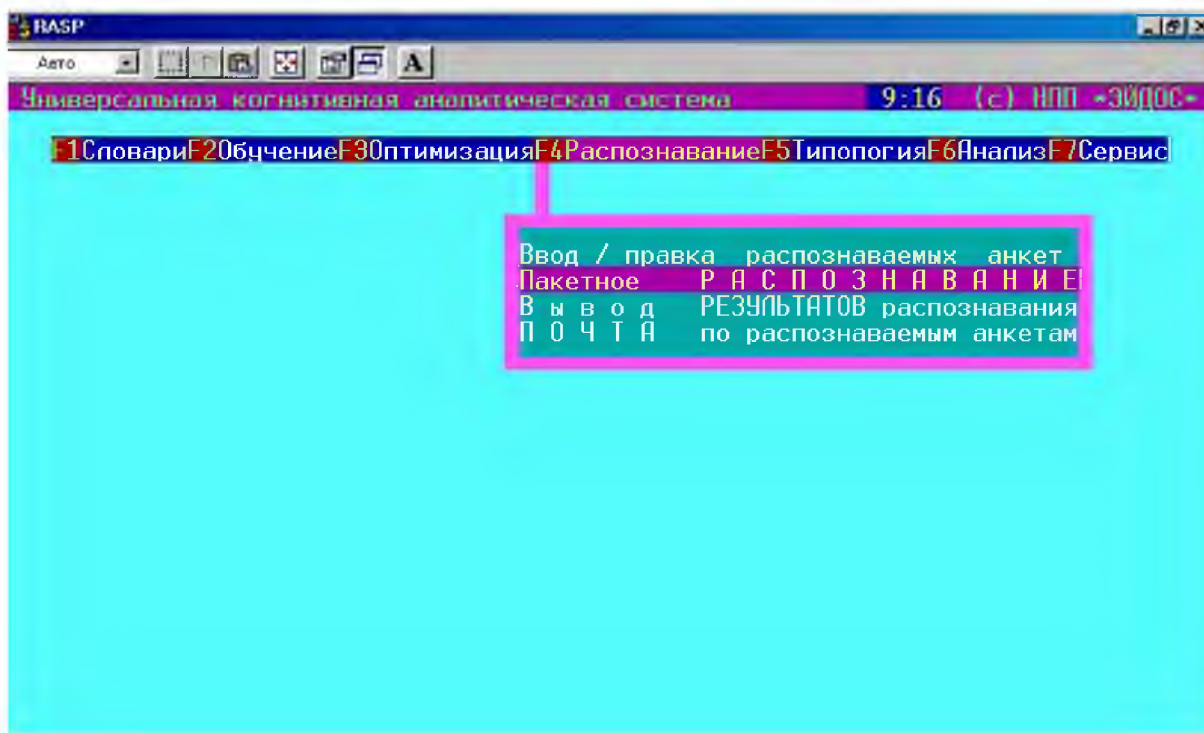


Рисунок 125. Выбор режима "Пакетное распознавание" системы "Эйдос"

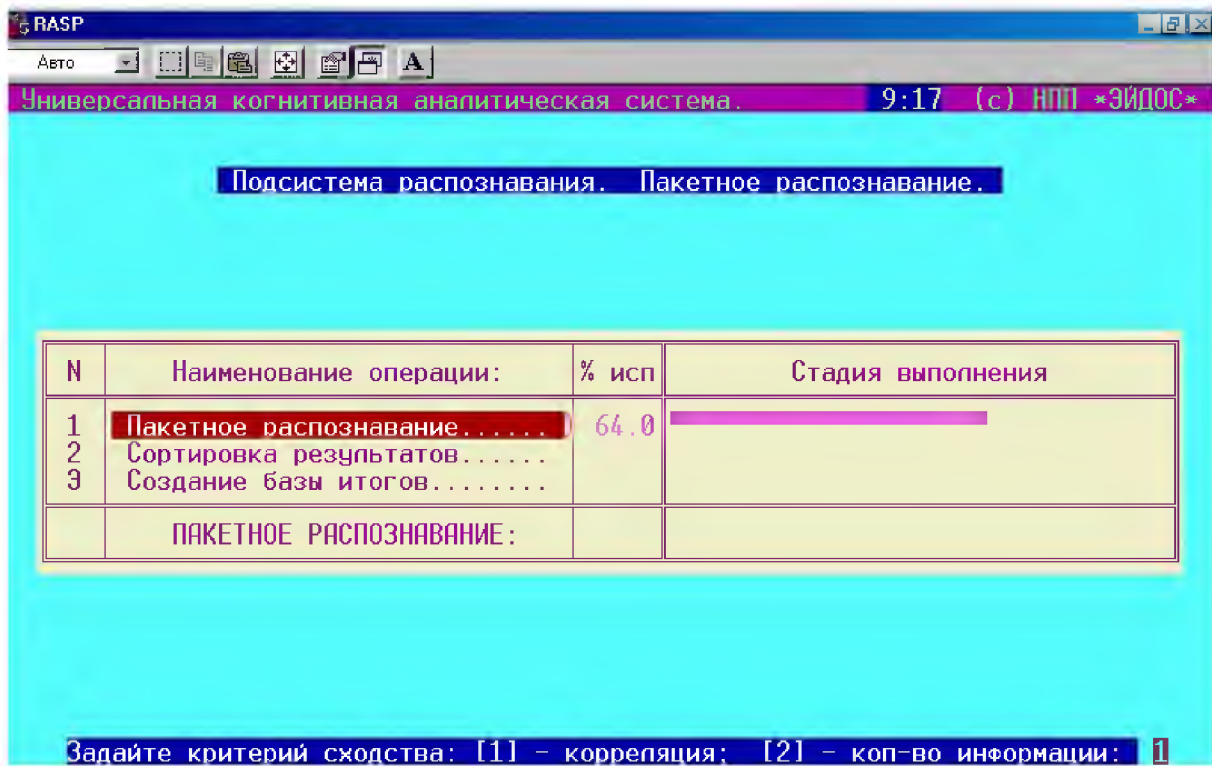


Рисунок 126. Выполнение режима "Пакетное распознавание" системы "Эйдос"

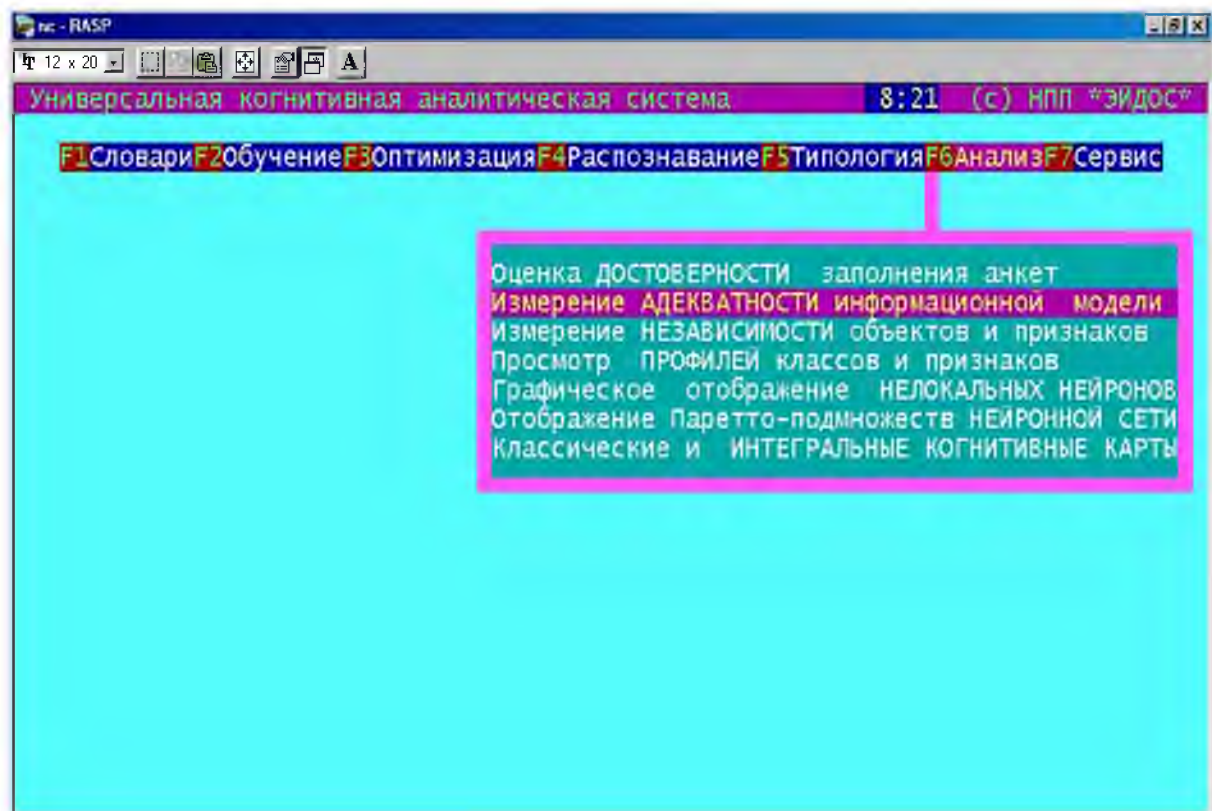


Рисунок 127. Выбор режима "Измерение адекватности модели" системы "Эйдос"

nc - RASP

универсальная когнитивная аналитическая система. 8:23 (с) НПП *Эйдос*

Подсистема анализа. Измерение адекватности информационной модели

Анкет физических: 44 логических (всего/факт): 44/ 44  
 Верная идентификация: 44 Ошибочная неидентификация: 0  
 Верная идентификация: 100.00% Ошибочная неидентификация: 0.00%

БЕЛИНСКИЙ 44/ 1

Код	Наименование класса	Анкет лог-х.	Идент. верно	Идент. ошиб.	Неидент. верно	Неидент. ошибоч.	ВЕРНАЯ. ИДЕНТ.%	Ошибочн. идентиф.%
1	БЕЛИНСКИЙ	1	1	37	6	0	100.00	86.05
2	БЛАГОДАРНОСТИ	1	1	32	11	0	100.00	74.42
3	БУДЕТ	1	1	32	11	0	100.00	74.42
4	ВЕЛИКИЙ	1	1	35	8	0	100.00	81.40
5	ВЗОР	1	1	21	22	0	100.00	48.84
6	ВРЕМЯ	1	1	28	15	0	100.00	65.12
7	ВСЕГДА	1	1	34	9	0	100.00	79.07
8	ВЫСОКО	1	1	36	7	0	100.00	83.72
9	ГЕНИАЛЬНОЕ	1	1	28	15	0	100.00	65.12
10	ДАЛЕКО	1	1	36	7	0	100.00	83.72

F1Генерация отчета F2Сортировка F3Печать F4Поиск F5Расч. внешней валид. F9Удал. классов

Рисунок 128. Режим "Измерение адекватности модели" системы "Эйдос"

Видим, что модель адекватна, т.к. интегральная валидность составляет 100%. Это означает, что *при идентификации слов на основе знания входящих в них букв системой не было допущено ни одной ошибки, причем необходимо специально отметить, что при этом в модели не учитывались последовательность букв и их сочетаний.*

#### ***Пример решения задания 4: "Проверить устойчивость модели к неполноте информации и наличию шума"***

В примере, исследуемом в данной лабораторной работе, неполнота информации – это пропуск букв, а наличие шума – замена верных букв неверными.

##### **Устойчивость модели к неполноте информации**

Подготовим распознаваемую выборку, состоящую из идентифицируемых слов с отсутствующими буквами.

Для этого выполним следующую последовательность шагов:

**Шаг 1.** Сбросим распознаваемую выборку в режиме "F7 Сервис – Генерация (сброс) баз данных – Распознаваемые анкеты" (рисунок 129):



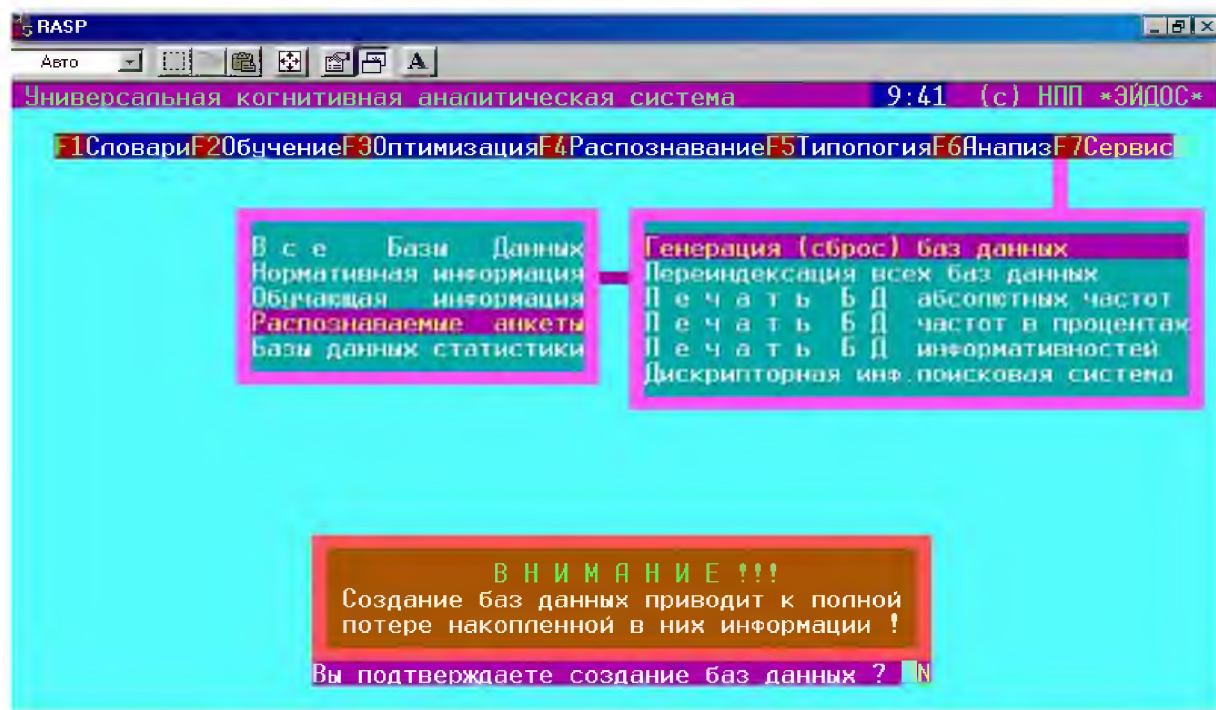


Рисунок 129. Режим "Сброс распознаваемой выборки" системы "Эйдос"

**Шаг 2.** Скопируем, например, первую анкету из обучающей выборки в распознаваемую, используя возможности режима "F2 Обучение – Ввод-корректировка обучающей выборки" (рисунок 124);

**Шаг 3.** Выберем режим "F4 Распознавание – Ввод-корректировка распознаваемой выборки" (рисунок 130):

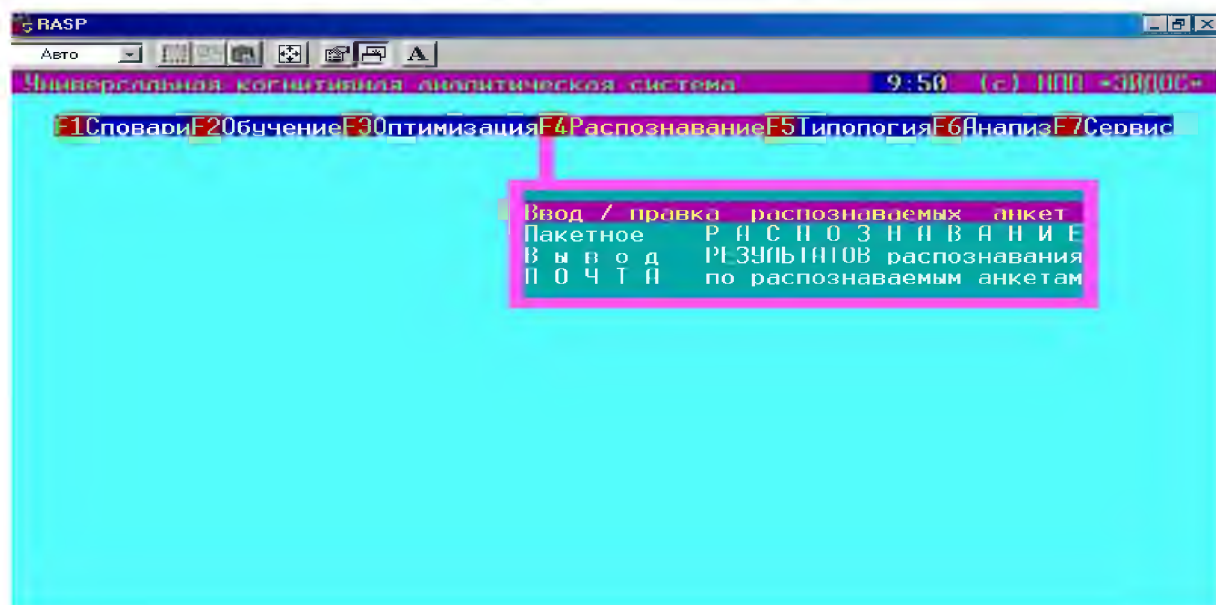


Рисунок 130. Выбор режима "Ввод-корректировка распознаваемой выборки" системы "Эйдос"

Выбор режима осуществляется нажатием клавиши Enter.

**Шаг 4.** Перейдем в правое окно, в котором задаются коды признаков, нажав клавишу "TAB".

**Шаг 5.** Удаляем последний код признака и дублируем анкету, нажав клавишу "F5 Дублирование анкеты".

Повторяем шаги 4 и 5 до тех пор, пока в описании слова останется одна буква. В результате получится видеограмма, представленная на рисунке 131.

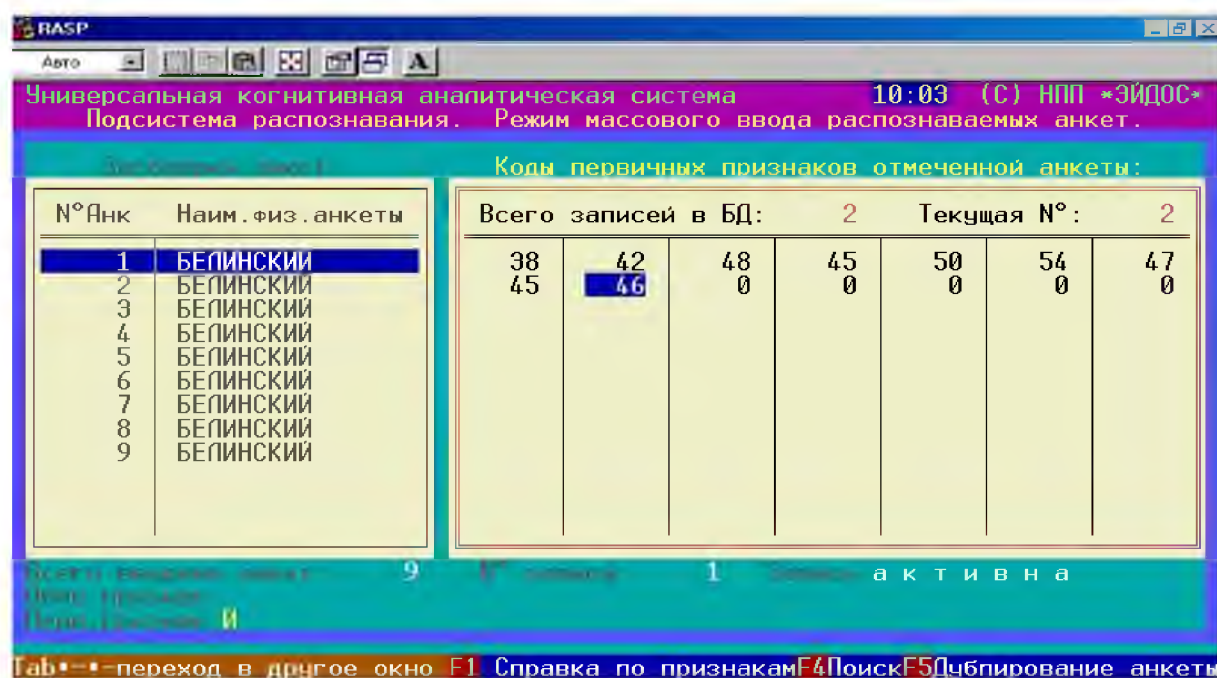
Студенты при выполнении этого этапа работы могут взять несколько анкет на выбор. При этом набор анкет должен отличаться у разных студентов.

Обучающая выборка в этом случае будет иметь вид, представленный на таблице 72:

**Таблица 72 – ВАРИАНТЫ КОДИРОВАНИЯ ОБЪЕКТА ОБУЧАЮЩЕЙ ВЫБОРКИ, ОТЛИЧАЮЩИЕСЯ СТЕПЕНЬЮ НЕПОЛНОТЫ ИНФОРМАЦИИ**

№	Класс	Коды признаков													
1	<b>БЕЛИНСКИЙ</b>	38	42	48	45	50	54	47	45	46					
2	<b>БЕЛИНСКИЙ</b>	38	42	48	45	50	54	47	45						
3	<b>БЕЛИНСКИЙ</b>	38	42	48	45	50	54	47							
4	<b>БЕЛИНСКИЙ</b>	38	42	48	45	50	54								
5	<b>БЕЛИНСКИЙ</b>	38	42	48	45	50									
6	<b>БЕЛИНСКИЙ</b>	38	42	48	45										
7	<b>БЕЛИНСКИЙ</b>	38	42	48											
8	<b>БЕЛИНСКИЙ</b>	38	42												
9	<b>БЕЛИНСКИЙ</b>	38													

Жирным шрифтом выделены символы, коды которых есть в анкете.



**Рисунок 131. Выполнение режима "Ввод-корректировка распознаваемой выборки" системы "Эйдос"**

**Шаг 6.** Выполним пакетное распознавание, выбрав и выполнив режим "F4 Распознавание – Пакетное распознавание", как показано на рисунках 125 и 126.

**Шаг 7.** Затем выберем и выполним режим "F4 Распознавание – Вывод результатов распознавания" (рисунок 132):

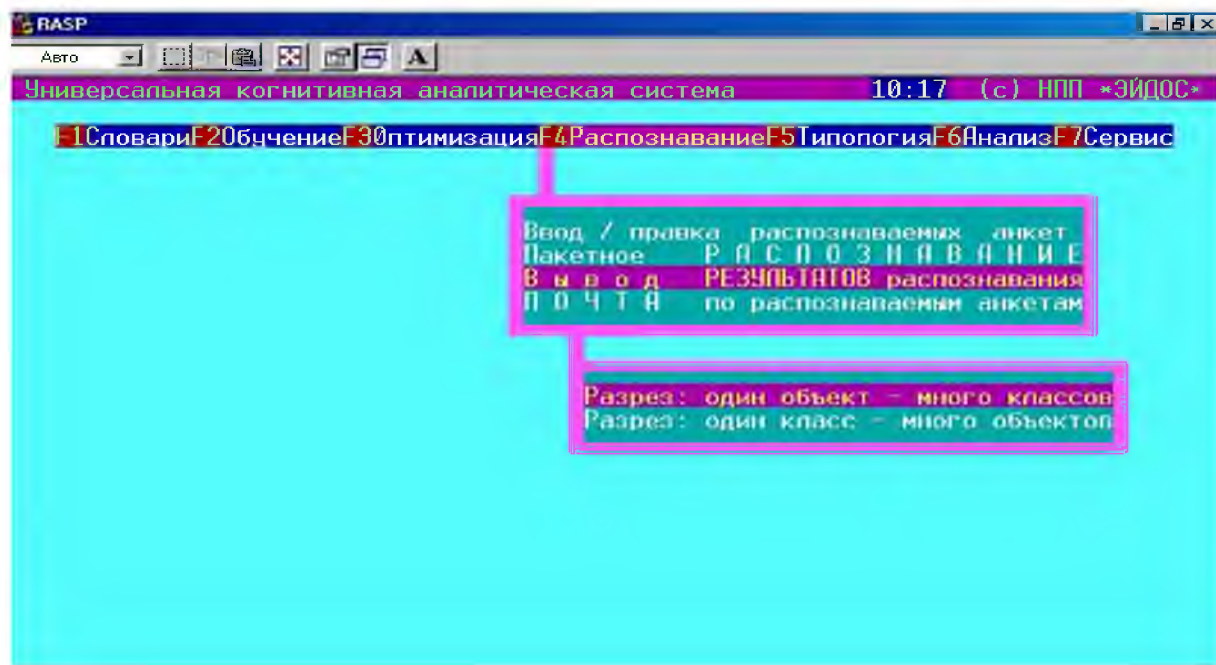


Рисунок 132. Выбор режима "Вывод результатов распознавания" системы "Эйдос"

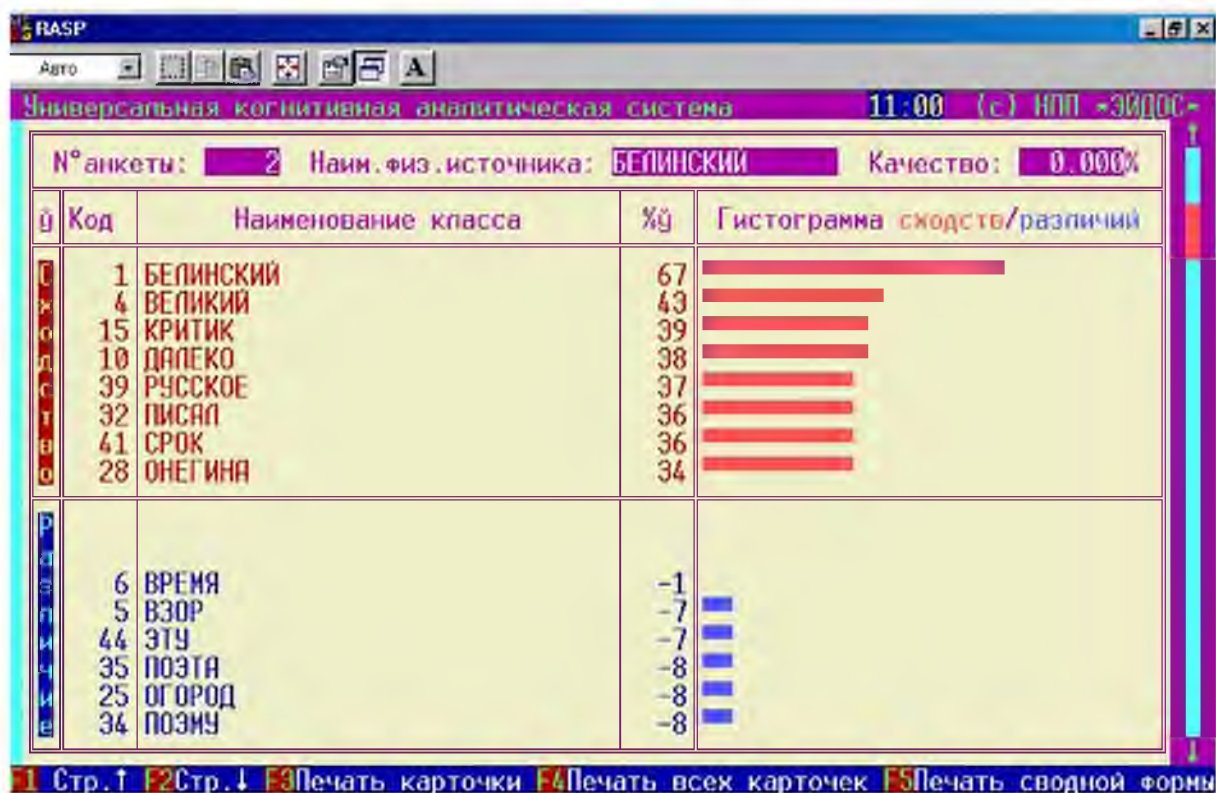
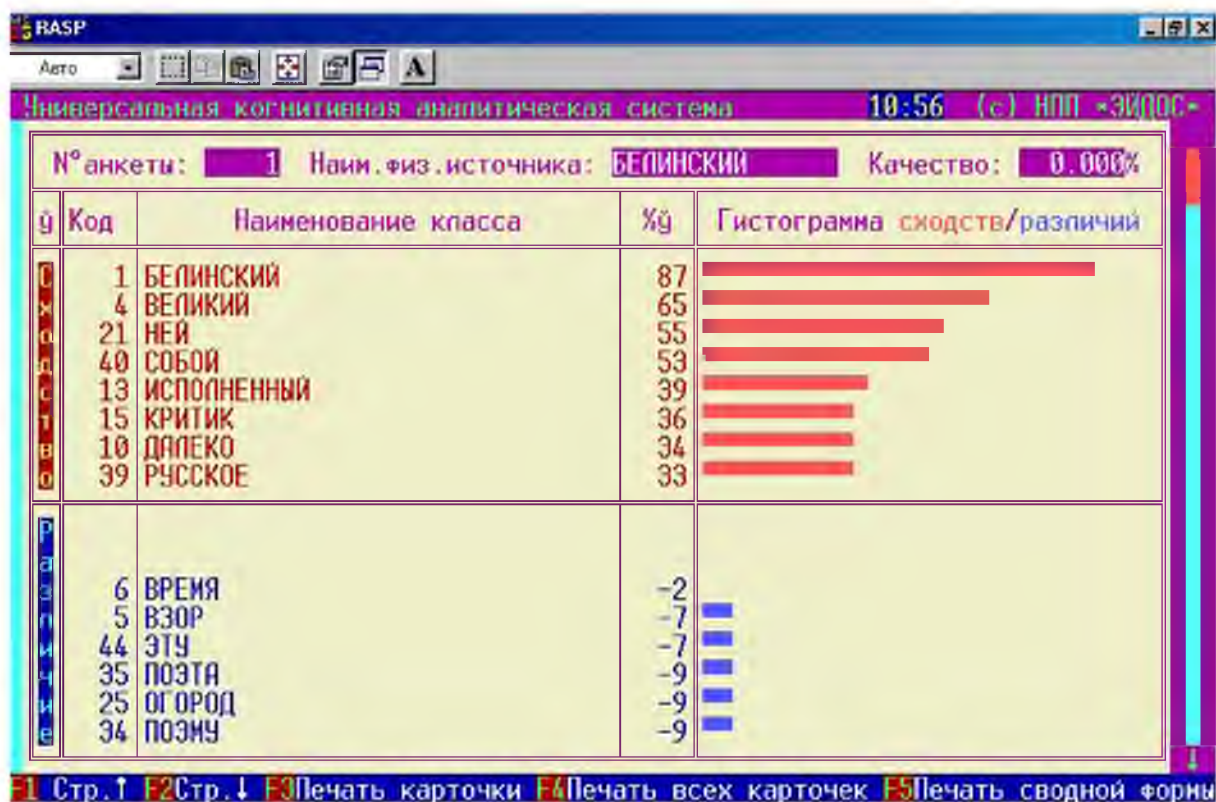
**Шаг 8.** Войдя в этот режим получим видеогармму, представленную на рисунке 133:

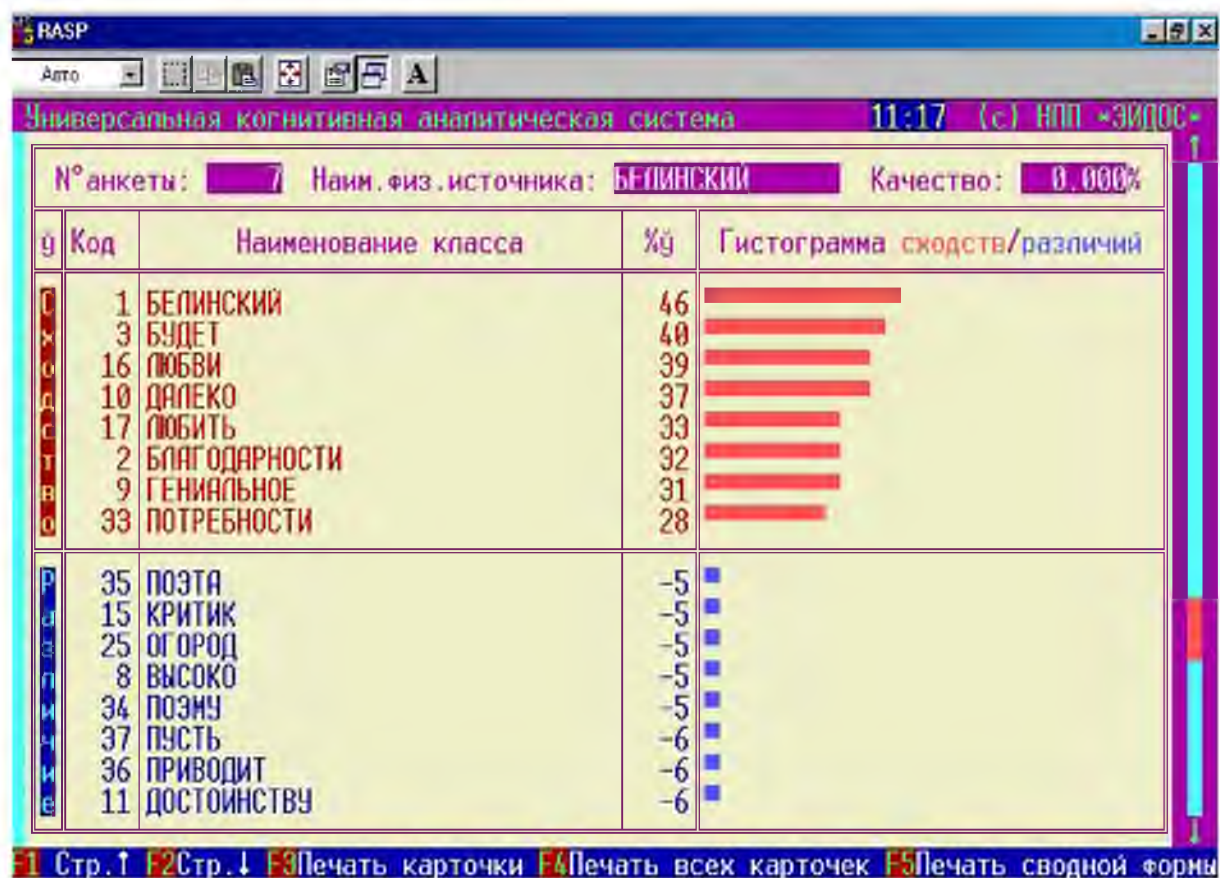
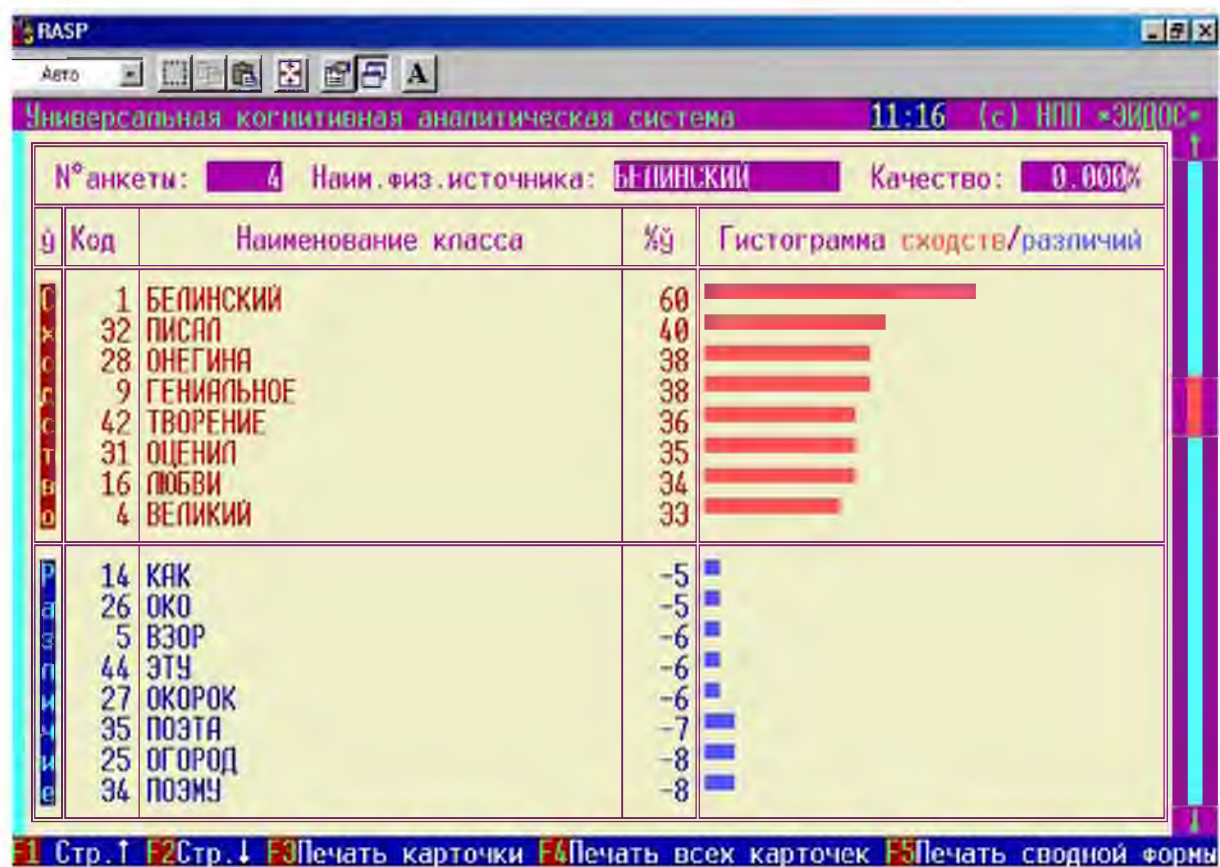
N°N° анкет	Наименование физ. источн. инф.	Код клас	Наименование класса распознавания	% Сход	Кач-во результат	?
1	БЕЯИАСКИЙ	1	БЕЯИАСКИЙ.....	87	0.000	◀
2	БЕЯИАСКИЙ	1	БЕЯИАСКИЙ.....	67	0.000	◀
3	БЕЛИНСКИЙ	1	БЕЛИНСКИЙ.....	67	0.000	◀
4	БЕЯИАСКИЙ	1	БЕЯИАСКИЙ.....	60	0.000	◀
5	БЕЯИНСКИЙ	1	БЕЯИАСКИЙ.....	57	0.000	◀
6	БЕЛИНСКИЙ	1	БЕЛИНСКИЙ.....	58	0.000	◀
7	БЕЯИАСКИЙ	1	БЕЯИАСКИЙ.....	46	0.000	◀
8	БЕЯИНСКИЙ	3	БУДЕТ.....	51	0.000	◀
9	БЕЛИНСКИЙ	40	СОБОЙ.....	53	0.000	◀

Рисунок 133. Обобщенная форма по результатам выполнения режима "Вывод результатов распознавания" системы "Эйдос"



**Шаг 9.** Нажав клавишу "F1 Карта идентификации объекта с классами" получим более подробные результаты идентификации, представленные в карточках распознавания на рисунке 134:







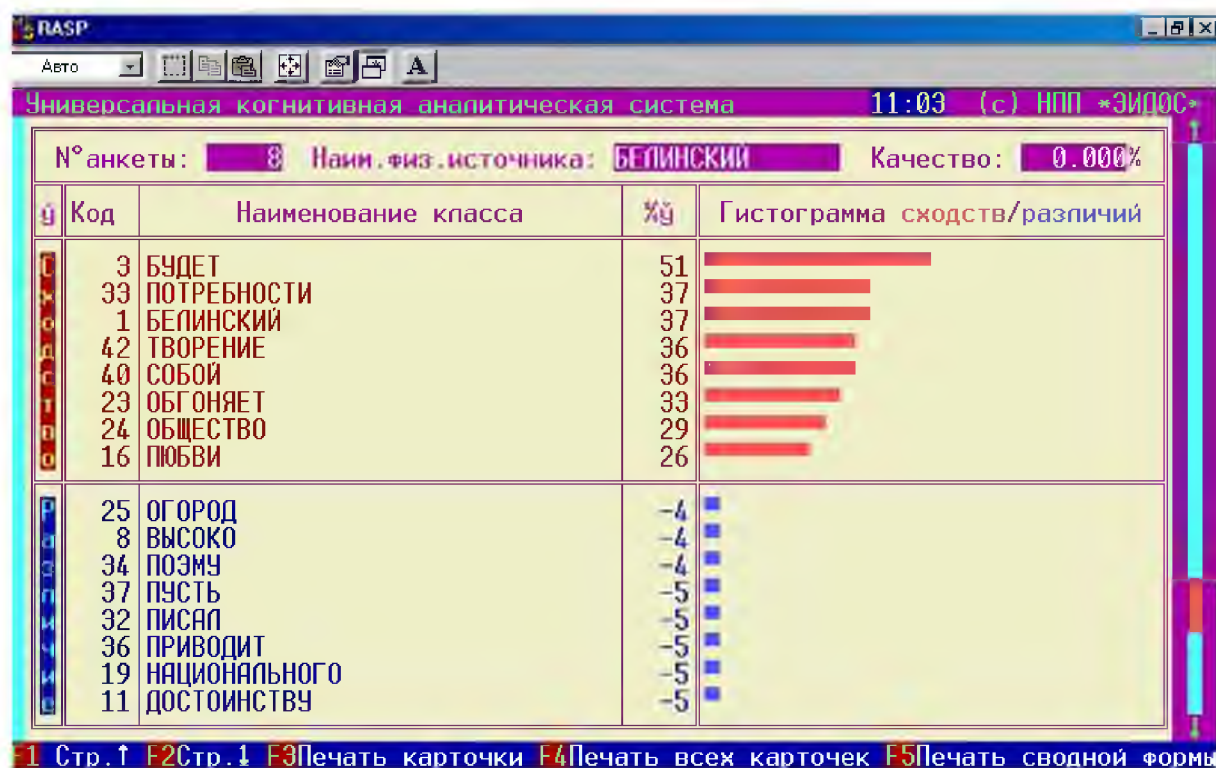


Рисунок 134. Идентификация в условиях неполноты информации в системе "Эйдос"

Из обобщенных и детальных выходных форм по результатам распознавания слова по его неполному описанию видно, что модель обладает очень высокой устойчивостью к неполноте информации в описании идентифицируемых объектов.

### Устойчивость модели к наличию шума

Шум можно рассматривать как сочетание неполноты информации (т.к. некоторые значащие символы исчезают из описаний объектов), и дезинформации (т.к. в описание включаются ложные символы).

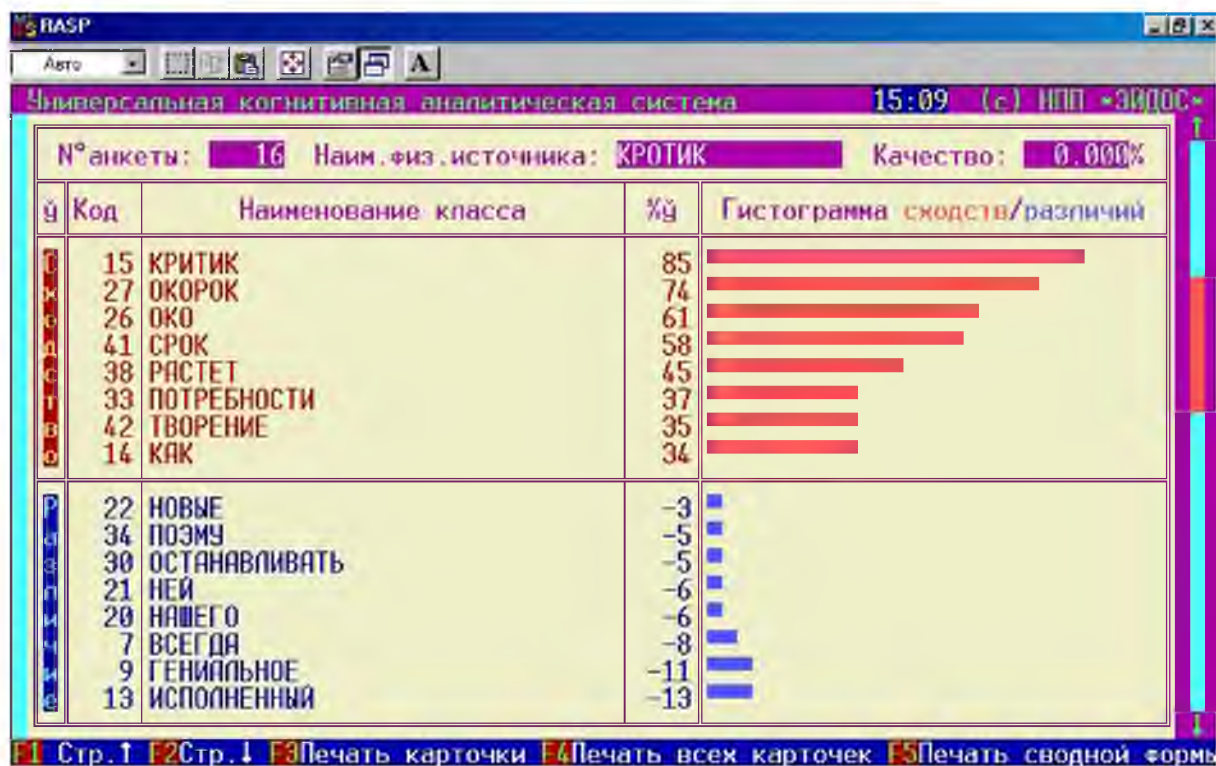
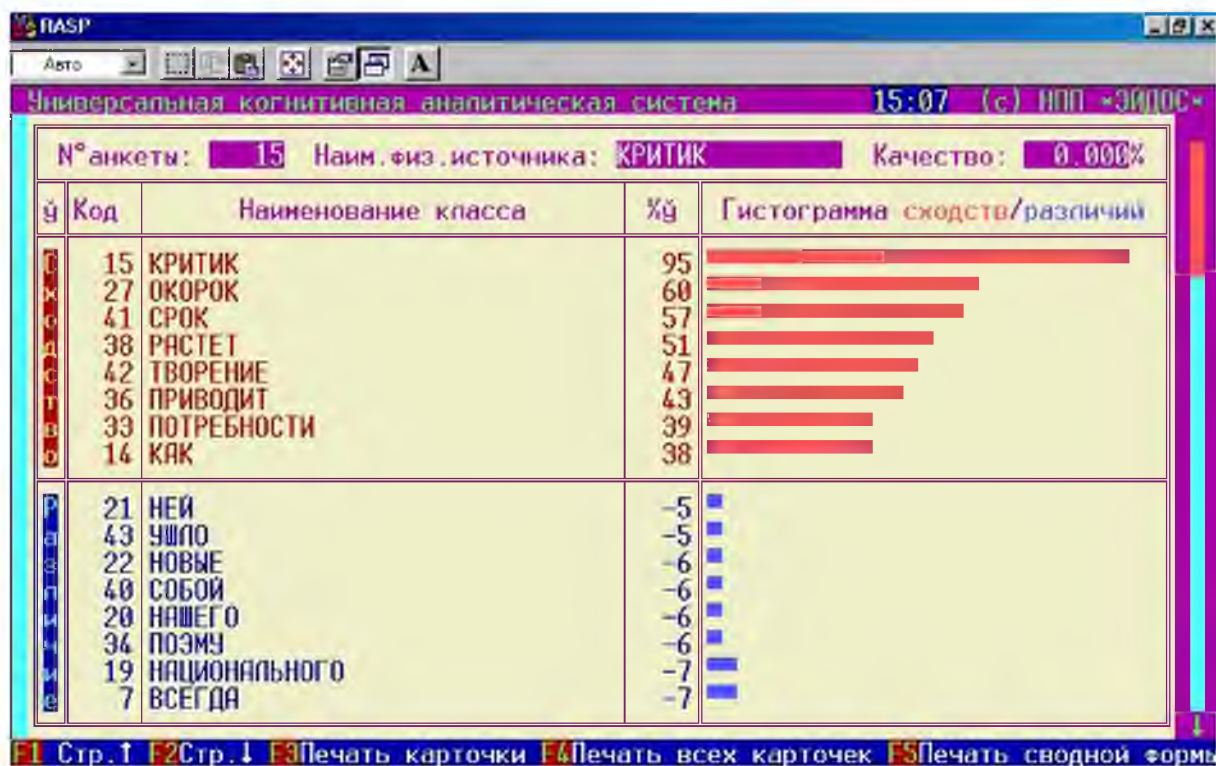
Поэтому замену символов в словах на символы, которые не встречаются по обучающей выборке можно считать неполнотой информации. Этот случай мы рассматривать не будем, т.к. по сути уже рассмотрели его в предыдущем пункте.

Рассмотрим пример, в котором одно слово заменой букв преобразуется в другое слово, например, слово "критик" преобразуется в слово "окорок". Каждой замене будет соответствовать одна анкета распознаваемой выборки (таблица 73):

Таблица 73 – ВАРИАНТЫ КОДИРОВАНИЯ ОБЪЕКТА ОБУЧАЮЩЕЙ ВЫБОРКИ, ОТЛИЧАЮЩИЕСЯ УРОВНЕМ ШУМА

№	Класс	Коды признаков											
1	КРИТИК	47	53	45	55	45	47						
2	КРОТИК	47	53	51	55	45	47						
3	КРОТОК	47	53	51	55	51	47						
4	ОКОРОК	51	47	51	53	51	47						

Результаты идентификации представлены на рисунке 135:



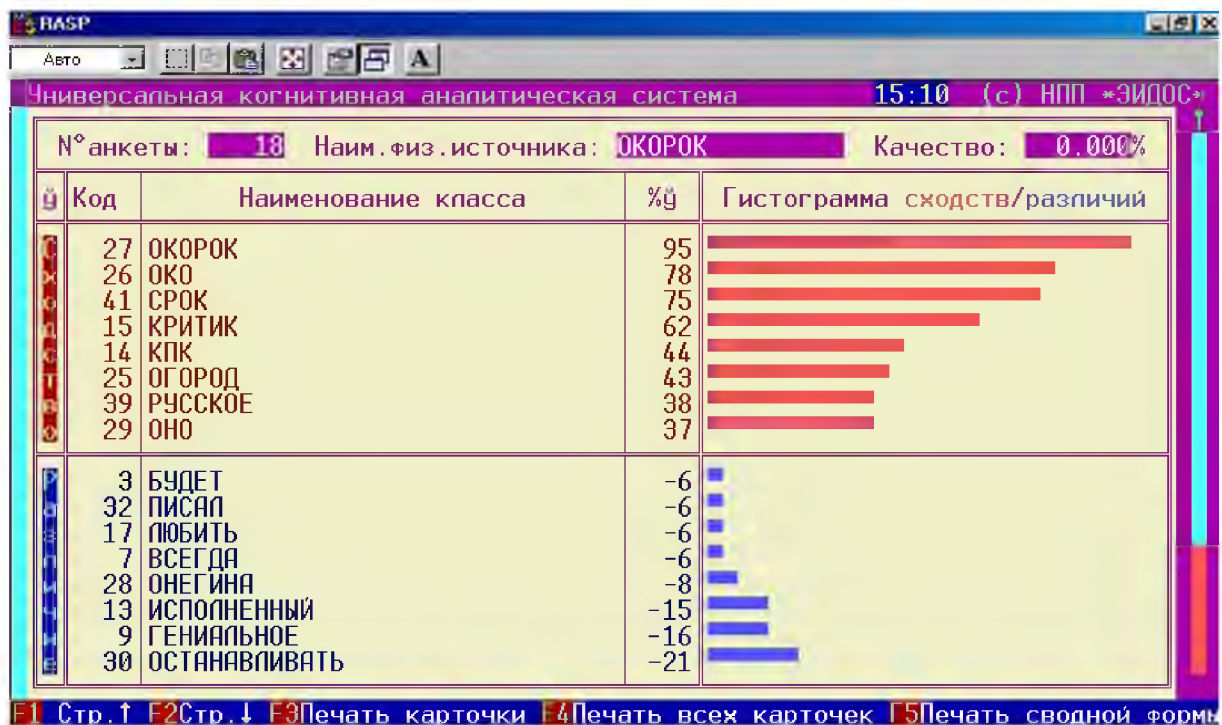
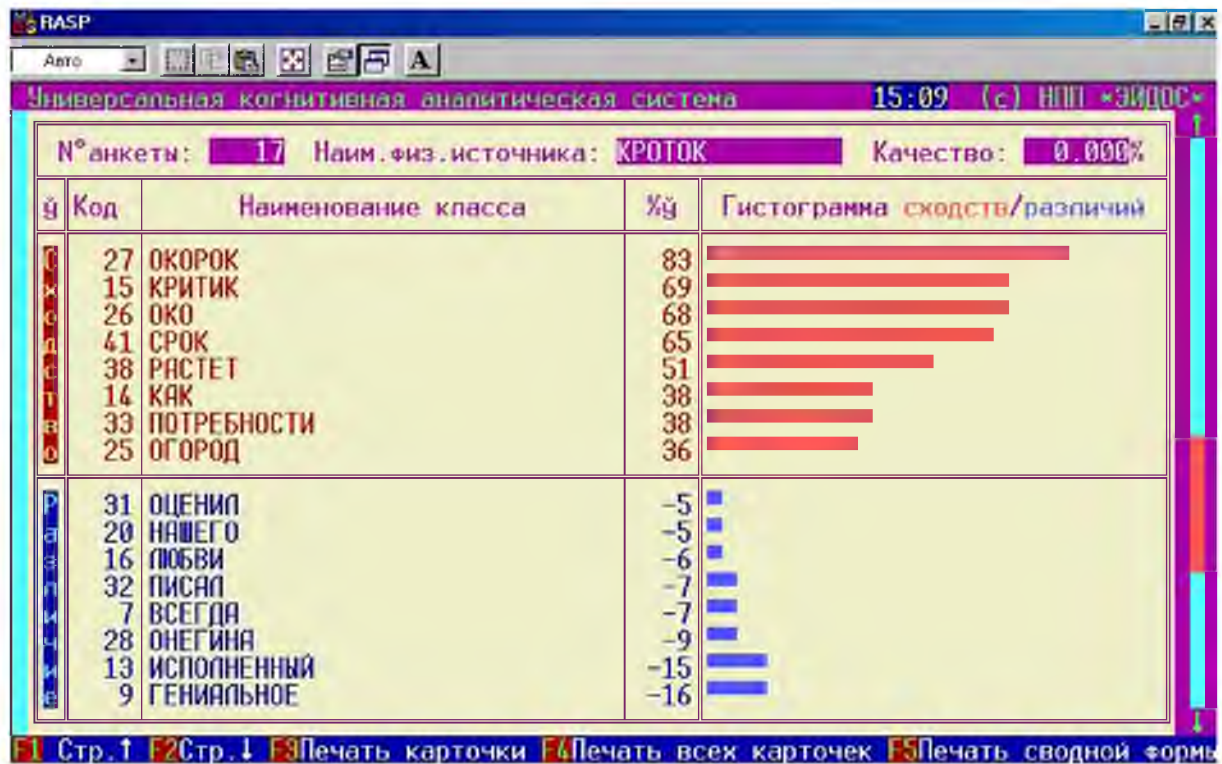


Рисунок 135. Результаты идентификации в условиях шума в системе "Эйдос"

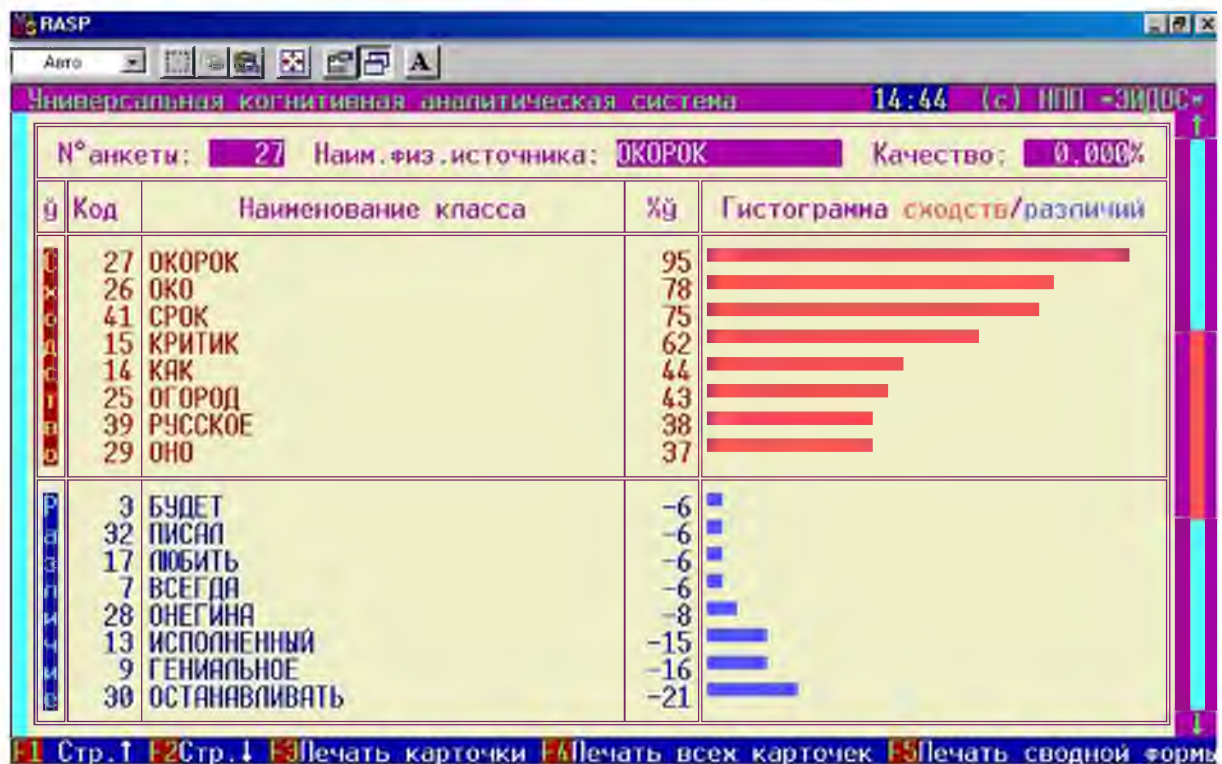
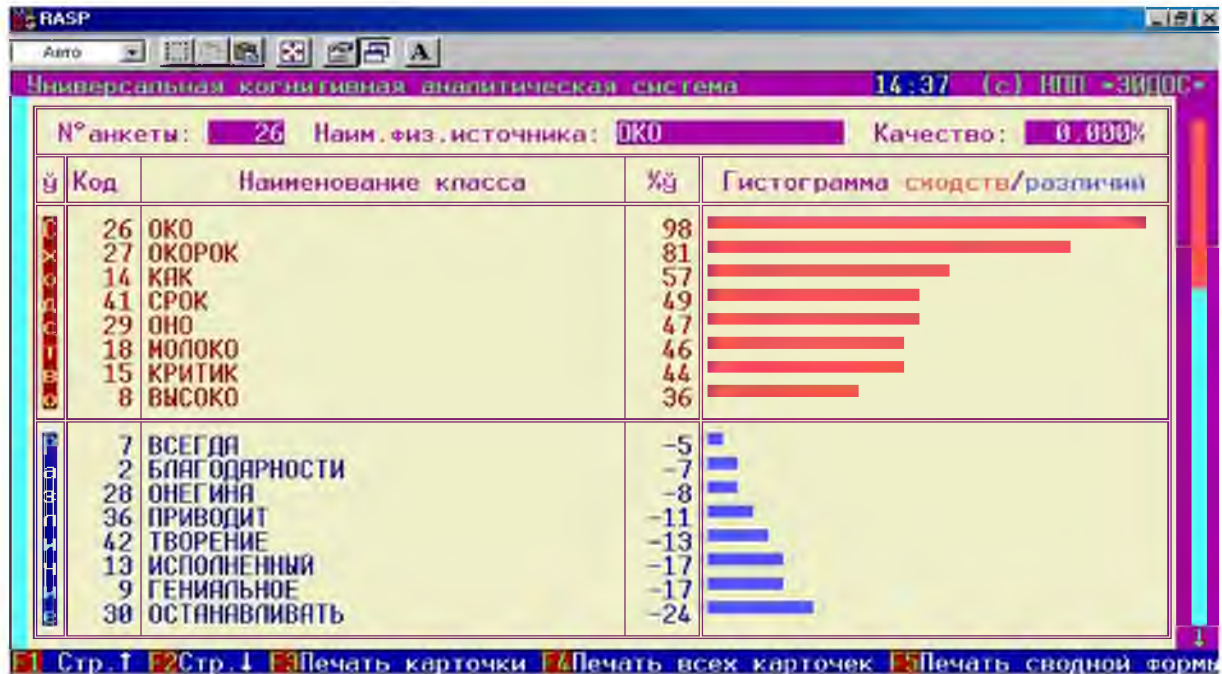
Видим, что модель обладает определенной устойчивостью и к шуму.

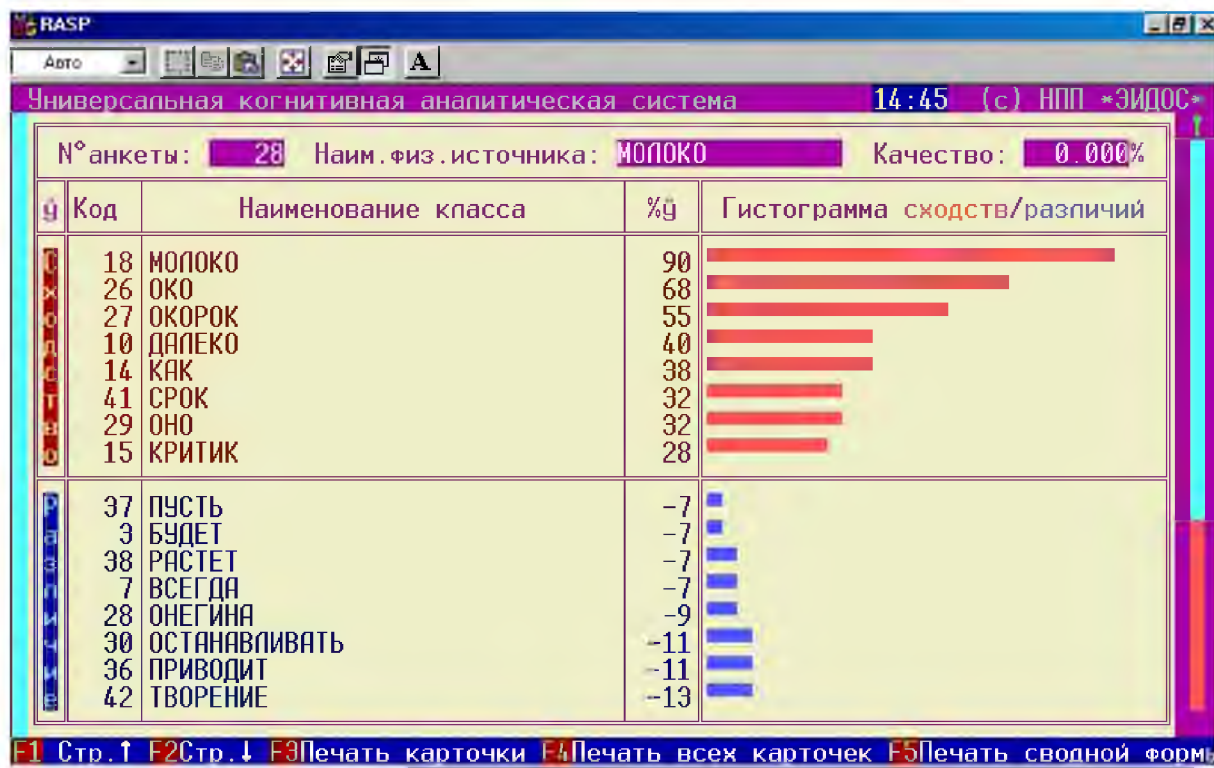


**Пример решения задания 5: "Проверить способность модели правильно идентифицировать классы, один из которых является подмножеством другого"**

С этой целью в текстовый файл специально включены такие слова, как: "око", "окорок", "молоко".

Результаты их идентификации приведены на рисунке 136:





**Рисунок 136. Карточка результатов распознавания системы "Эйдос" в случае идентификации классов, один из которых является подмножеством других**

Как видим, идентификация классов, один из которых является подмножеством других, осуществляется правильно. Это является важным достоинством семантической информационной модели системы "Эйдос", т.к. представляет собой проблему для многих типов нейронных сетей. Достаточно отметить, что возможность решения подобных задач считается одним из основных достоинств развитой модели нейронной сети, реализованной в неоконитроне Фукушимы.

### **Пример решения задания 6: Оценить ценность букв для идентификации слов. Сравнить суммарную ценность для этой цели гласных и согласных букв**

Для решения этой задачи запустим 2-й режим в 3-й подсистеме (рисунок 137). В этом режиме все признаки, которыми в данном примере являются буквы, выводятся системой "Эйдос" в порядке убывания среднего количества информации, которое в них содержится о принадлежности к словам. Если просуммировать ценность букв "нарастающим итогом" то получим накопительную кривую, представленную на рисунке 138. Эта кривая называется "Парето-диаграмма" по имени известного итальянского математика и экономиста XIX, Вильфредо Парето, впервые предложившего оценивать силу влияния факторов, исключать из рассмотрения незначимые факторы и впервые построившего подобные диаграммы.

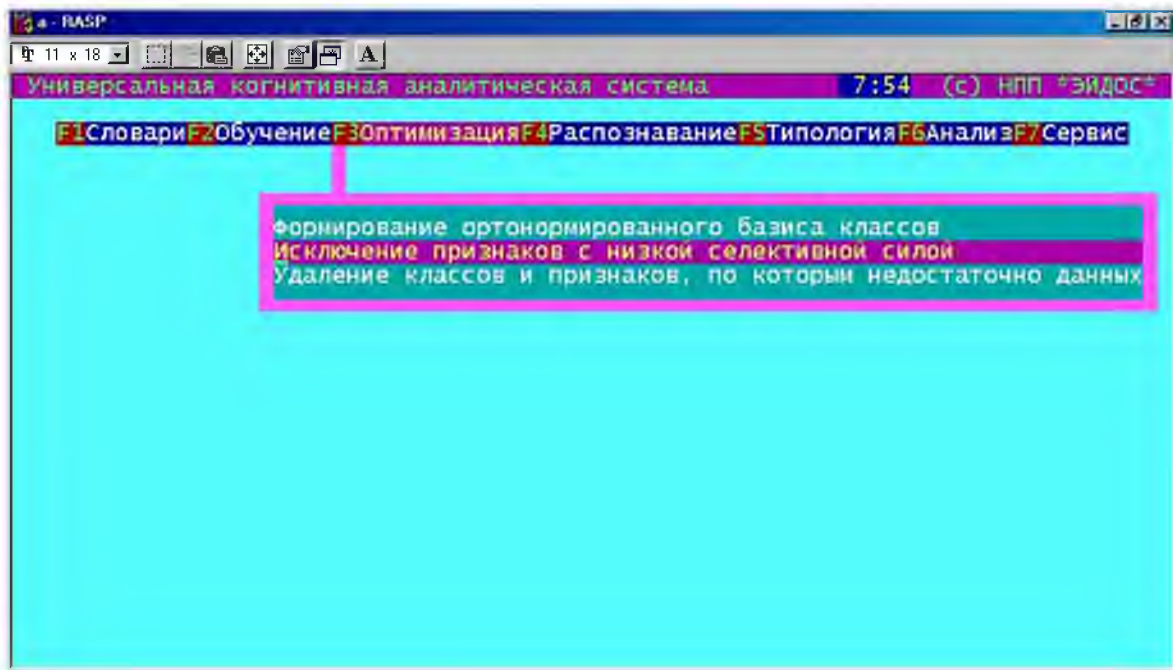
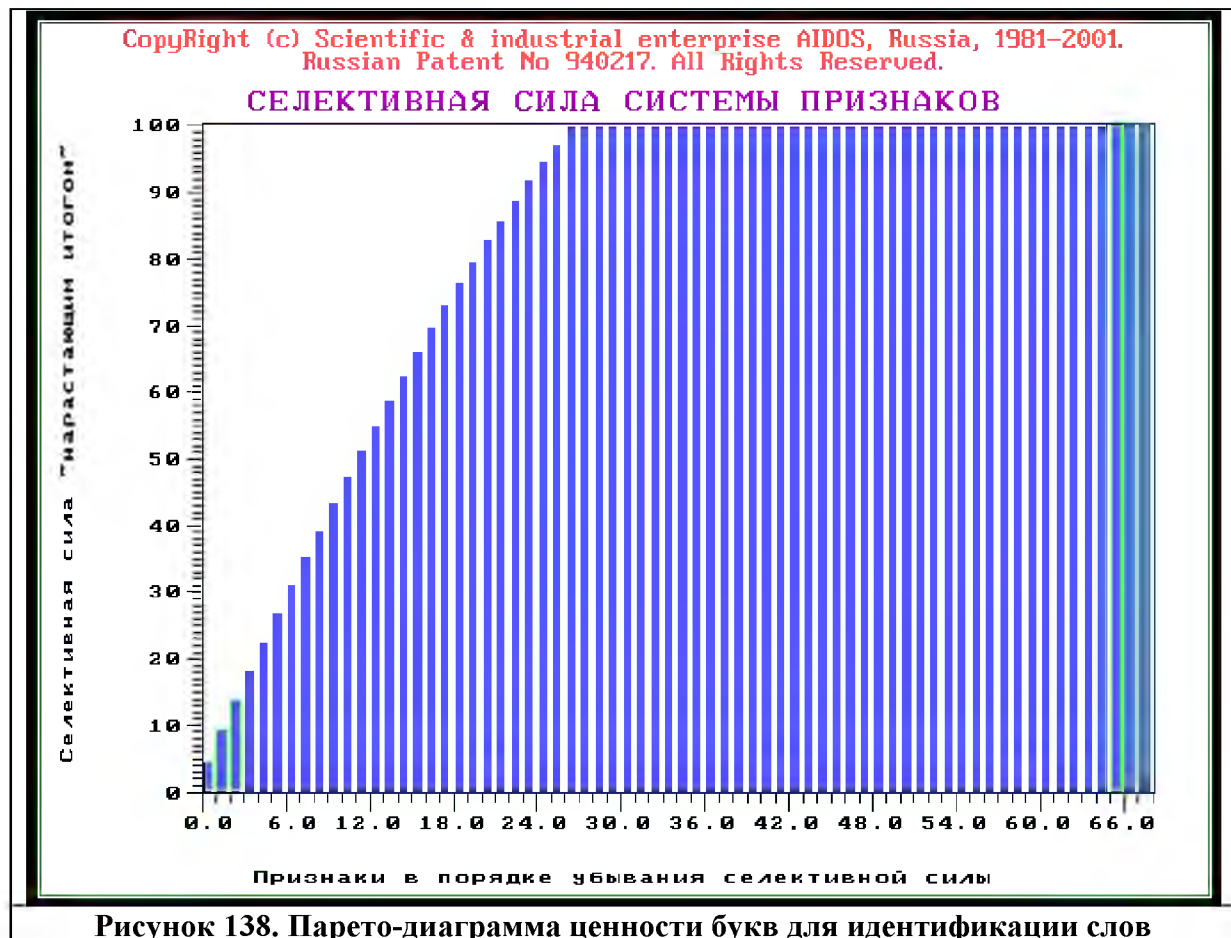


Рисунок 137. Запуск режима системы "Эйдос" измерения ценности признаков для решения задач идентификации, прогнозирования и управления



Характерная "полочка" на Парето-диаграмме соответствует цифрам и буквам латинского алфавита, которые не встретились в словах обучающей выбор-

ки. В таблице 74 приведены буквы, проранжированные в порядке убывания среднего количества информации в них, о принадлежности к словам.

**Таблица 74 – ЦЕННОСТЬ БУКВ ДЛЯ ИДЕНТИФИКАЦИИ СЛОВ**

№	Код	Буква	Ценность (бит)	Ценность (бит) "нарастающим итогом"	Ценность (%) "нарастающим итогом"
1	66	Э	0,76988	0,76988	4,841
2	56	У	0,74529	1,51517	9,526
3	49	М	0,71090	2,22607	13,996
4	46	Й	0,69728	2,92335	18,380
5	61	Ш	0,68748	3,61083	22,703
6	47	К	0,68569	4,29652	27,014
7	67	Ю	0,66376	4,96028	31,187
8	52	П	0,66014	5,62042	35,338
9	64	Ы	0,65157	6,27199	39,434
10	41	Д	0,64023	6,91222	43,460
11	68	Я	0,63612	7,54834	47,459
12	44	З	0,62131	8,16965	51,366
13	65	Ь	0,59697	8,76662	55,119
14	38	Б	0,59622	9,36284	58,868
15	53	Р	0,58610	9,94894	62,553
16	59	Ц	0,57201	10,52095	66,149
17	40	Г	0,56958	11,09053	69,730
18	39	В	0,55490	11,64543	73,219
19	62	Щ	0,52045	12,16588	76,492
20	37	А	0,51477	12,68065	79,728
21	48	Л	0,50010	13,18075	82,872
22	54	С	0,47977	13,66052	85,889
23	55	Т	0,47498	14,13550	88,875
24	51	О	0,46548	14,60098	91,802
25	50	Н	0,45089	15,05187	94,637
26	45	И	0,43046	15,48233	97,343
27	42	Е	0,42253	15,90486	100,000

***Пример решения задания 7: "Выполнить кластерно-конструктивный анализ слов и букв, вывести информационные и семантические портреты слов и букв, построить их профили"***

#### **Кластерно-конструктивный анализ слов**

Кластерно-конструктивный анализ выполняется в 5-й подсистеме "Типология" (рисунок 139). Сначала на основе матрицы информативностей рассчитывается матрица сходства классов (таблица 75), а затем на основе нее формируется таблица кластеров и конструкторов классов (таблица 76).



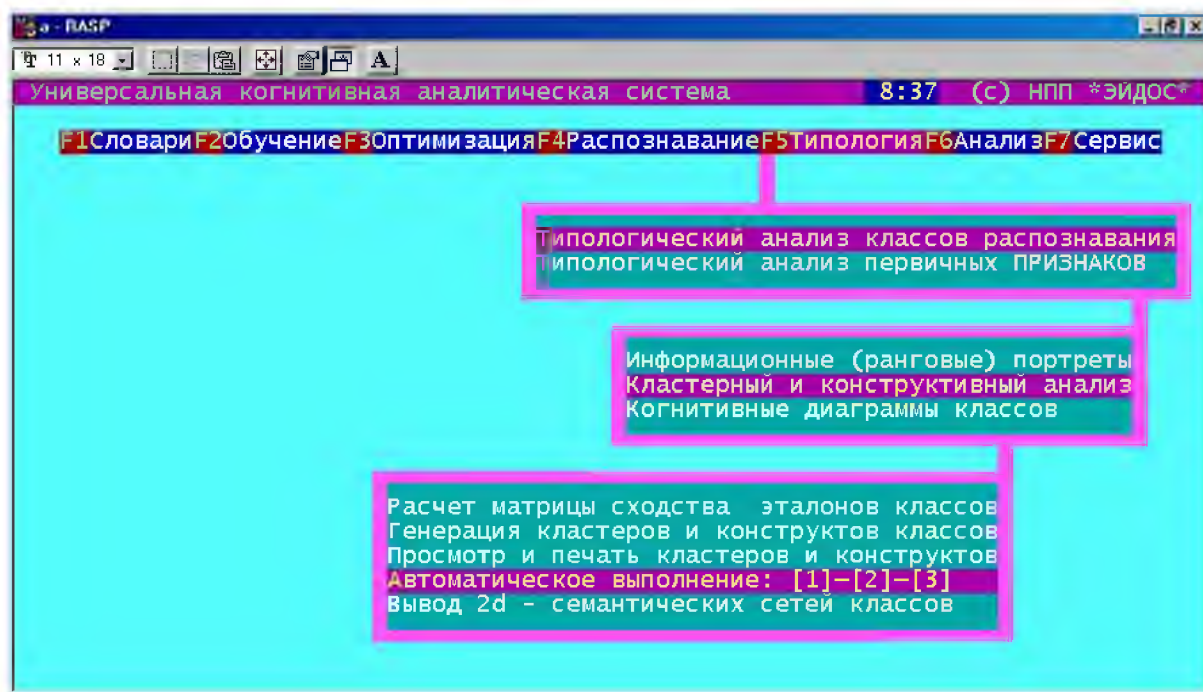


Рисунок 139. Запуск режима кластерно-конструктивного анализа классов (слов)

Таблица 75 – ФРАГМЕНТ МАТРИЦЫ СХОДСТВА КЛАССОВ (СЛОВ)

Код	Коды классов											
	1	2	3	4	5	6	7	8	9	10	11	12
1	100,0	20,0	19,0	78,1	-6,6	-5,5	1,3	9,7	12,9	22,2	5,3	13,3
2	20,0	100,0	43,1	-1,4	4,3	1,5	68,2	-4,6	32,4	54,0	23,9	31,5
3	19,0	43,1	100,0	-2,4	-5,5	-1,6	29,5	-6,7	2,2	36,3	65,5	57,8
4	78,1	-1,4	-2,4	100,0	7,2	7,9	11,7	19,5	12,9	29,3	5,6	18,3
5	-6,6	4,3	-5,5	7,2	100,0	17,2	8,3	10,2	-7,2	-5,5	4,9	-4,9
6	-5,5	1,5	-1,6	7,9	17,2	100,0	9,6	5,1	0,0	-1,4	1,4	1,7
7	1,3	68,2	29,5	11,7	8,3	9,6	100,0	18,0	35,1	49,5	47,9	44,5
8	9,7	-4,6	-6,7	19,5	10,2	5,1	18,0	100,0	-10,6	9,6	19,6	-6,0
9	12,9	32,4	2,2	12,9	-7,2	0,0	35,1	-10,6	100,0	23,3	-5,8	11,7
10	22,2	54,0	36,3	29,3	-5,5	-1,4	49,5	9,6	23,3	100,0	25,3	53,4
11	5,3	23,9	65,5	5,6	4,9	1,4	47,9	19,6	-5,8	25,3	100,0	51,4
12	13,3	31,5	57,8	18,3	-4,9	1,7	44,5	-6,0	11,7	53,4	51,4	100,0
13	42,0	-2,2	-5,7	36,4	-7,0	-5,8	-2,4	46,5	16,9	2,4	0,5	-2,3
14	19,8	23,2	-4,6	24,4	-3,5	-4,4	14,7	25,7	7,8	55,9	-5,1	-4,1
15	32,9	8,4	2,5	39,0	13,0	9,4	-7,2	18,2	-1,5	23,9	9,2	28,2
16	30,3	18,2	15,4	26,5	6,9	3,9	6,5	6,7	4,0	7,7	4,3	2,8
17	24,1	15,2	18,0	9,3	-5,6	-7,0	-8,2	-6,9	35,3	5,0	1,2	9,0
18	14,6	0,1	-6,1	18,9	-1,6	43,7	-6,8	14,5	-2,5	26,9	-5,1	-5,5
19	-0,6	24,6	-7,8	-3,6	-5,0	-7,4	19,5	-4,9	57,2	10,9	-5,7	-7,0
20	-2,3	29,7	-1,1	-2,9	-4,9	-2,2	33,7	-5,8	36,1	10,5	-5,7	2,3
21	61,5	-6,4	5,7	59,9	-4,1	3,0	4,1	-5,0	24,2	6,3	-3,1	11,8
22	0,1	-7,4	1,2	14,9	10,8	12,4	17,4	80,2	12,3	1,6	8,2	5,5
23	13,3	30,3	21,5	-4,9	-4,3	51,6	19,3	-4,0	20,5	-3,2	0,5	6,2
24	13,9	10,1	18,6	4,7	4,8	3,7	12,0	11,3	-4,4	-3,2	15,7	5,2



Таблица 76 – КОНСТРУКТ СЛОВ: "КРИТИК – ВСЕГДА"

№	Код слова	Слово	Уровень сходства	№	Код слова	Слово	Уровень сходства
1	15	<b>КРИТИК</b>	100,00	23	2	БЛАГОДАРНОСТИ	8,37
2	27	ОКОРОК	71,79	24	28	ОНЕГИНА	7,32
3	41	СРОК	65,59	25	31	ОЦЕНИЛ	5,70
4	14	КАК	52,61	26	44	ЭТУ	5,67
5	26	ОКО	51,07	27	16	ЛЮБВИ	5,34
6	42	ТВОРЕНИЕ	40,52	28	30	ОСТАНАВЛИВАТЬ	3,81
7	39	РУССКОЕ	40,23	29	35	ПОЭТА	2,59
8	36	ПРИВОДИТ	39,13	30	3	БУДЕТ	2,53
9	4	ВЕЛИКИЙ	38,98	31	37	ПУСТЬ	1,54
10	38	РАСТЕТ	37,88	32	23	ОБГОНЯЕТ	-1,37
11	1	БЕЛИНСКИЙ	32,92	33	9	ГЕНИАЛЬНОЕ	-1,50
12	33	ПОТРЕБНОСТИ	29,38	34	24	ОБЩЕСТВО	-1,55
13	12	ИДЕТ	28,16	35	13	ИСПОЛНЕННЫЙ	-3,75
14	10	ДАЛЕКО	23,87	36	29	ОНО	-4,20
15	18	МОЛОКО	19,13	37	21	НЕЙ	-4,87
16	8	ВЫСОКО	18,23	38	43	УШЛО	-5,20
17	25	ОГОРОД	16,45	39	22	НОВЫЕ	-5,51
18	5	ВЗОР	12,98	40	40	СОБОЙ	-5,64
19	6	ВРЕМЯ	9,39	41	20	НАШЕГО	-5,81
20	32	ПИСАЛ	9,36	42	34	ПОЭМУ	-6,07
21	11	ДОСТОИНСТВУ	9,19	43	19	НАЦИОНАЛЬНОГО	-7,01
22	17	ЛЮБИТЬ	8,54	44	7	<b>ВСЕГДА</b>	-7,17

### Кластерно-конструктивный анализ букв

Аналогично в режиме кластерно-конструктивного анализа признаков (рисунк 140) получаем матрицу сходства букв и конструкты букв (таблицы 77 и 78).

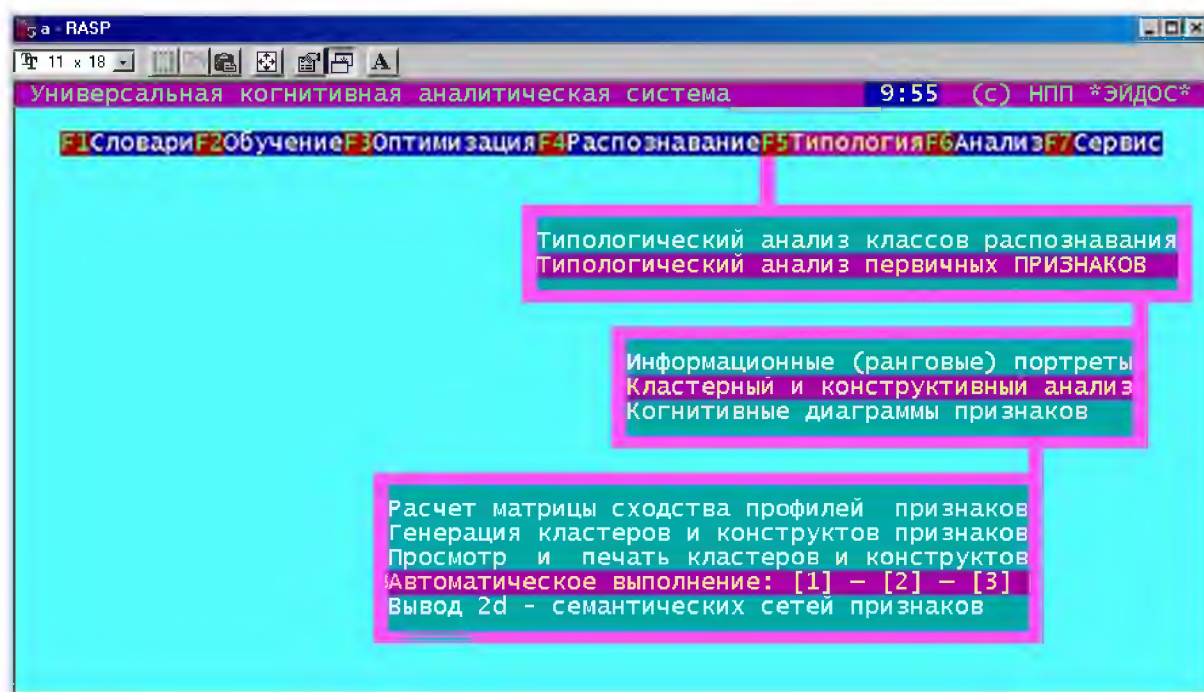


Рисунок 140. Запуск режима кластерно-конструктивного анализа признаков (букв)

Таблица 77 – ФРАГМЕНТ МАТРИЦЫ СХОДСТВА ПРИЗНАКОВ (БУКВ)

Код	Коды букв											
	37	38	39	40	41	42	43	44	45	46	47	48
37	100,0	-22,5	-11,1	33,7	5,4	-1,6	0,0	-9,0	-15,9	-20,6	5,1	6,5
38	-22,5	100,0	-2,1	-3,6	3,3	-6,5	0,0	-7,5	9,4	18,5	-20,9	13,9
39	-11,1	-2,1	100,0	-11,3	1,2	10,7	0,0	36,7	7,5	-7,3	-15,9	-6,5
40	33,7	-3,6	-11,3	100,0	22,9	15,2	0,0	-7,0	-15,4	-16,0	-24,7	-14,5
41	5,4	3,3	1,2	22,9	100,0	21,7	0,0	-7,0	8,5	-16,0	-13,8	-10,4
42	-1,6	-6,5	10,7	15,2	21,7	100,0	0,0	-12,0	6,1	20,8	-23,3	-11,4
43	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
44	-9,0	-7,5	36,7	-7,0	-7,0	-12,0	0,0	100,0	-9,1	-5,3	-8,2	-9,4
45	-15,9	9,4	7,5	-15,4	8,5	6,1	0,0	-9,1	100,0	15,3	5,4	35,3
46	-20,6	18,5	-7,3	-16,0	-16,0	20,8	0,0	-5,3	15,3	100,0	-1,8	8,1
47	5,1	-20,9	-15,9	-24,7	-13,8	-23,3	0,0	-8,2	5,4	-1,8	100,0	-2,8
48	6,5	13,9	-6,5	-14,5	-10,4	-11,4	0,0	-9,4	35,3	8,1	-2,8	100,0
49	-15,9	-13,3	7,9	-12,4	-12,4	-1,0	0,0	-4,1	-16,1	-9,5	0,5	3,9
50	0,9	-17,8	-16,8	29,0	-25,0	35,9	0,0	-8,8	-5,8	28,7	-29,0	-4,8
51	-31,8	2,6	-15,6	-0,9	-6,2	-36,3	0,0	7,6	-32,1	-6,3	30,3	-14,0
52	9,1	-13,2	-15,5	-19,5	-7,0	-30,2	0,0	-6,5	8,0	-4,7	-22,9	-1,6
53	-19,2	-20,2	15,7	-6,9	0,5	-4,8	0,0	36,4	-0,6	-20,4	19,8	-34,1
54	1,8	5,0	-0,5	-12,2	-4,6	-14,4	0,0	-9,6	-14,8	9,2	4,4	-14,3
55	-1,9	13,6	-12,1	-23,1	18,7	2,1	0,0	-10,3	10,2	-23,5	-25,1	-28,2
56	-25,1	-2,0	-20,5	-19,6	6,3	-12,3	0,0	-6,5	-23,8	-14,9	-14,9	-4,3
57	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
58	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
59	5,1	-10,6	-12,8	6,2	-9,9	4,9	0,0	-3,3	8,7	-7,6	-11,6	24,3
60	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Таблица 78 – КОНСТРУКТ БУКВ: "А – О"

№	Код буквы	Буква	Уровень сходства	№	Код буквы	Буква	Уровень сходства
1	37	<b>А</b>	100,00	13	42	Е	-1,57
2	40	Г	33,67	14	55	Т	-1,90
3	65	Ь	11,94	15	44	З	-8,98
4	52	П	9,08	16	62	Щ	-8,98
5	61	Ш	8,37	17	39	В	-11,08
6	48	Л	6,54	18	68	Я	-12,81
7	41	Д	5,41	19	67	Ю	-12,84
8	47	К	5,13	20	64	Ы	-15,79
9	59	Ц	5,13	21	45	И	-15,90
10	66	Э	4,66	22	49	М	-15,92
11	54	С	1,79	23	53	Р	-19,19
12	50	Н	0,90	24	46	Й	-20,58
				25	38	Б	-22,52
				26	56	У	-25,09
				27	51	<b>О</b>	-31,79

### Информационные портреты классов (слов)

Информационные портреты классов (слов) представляют собой списки признаков (букв), проранжированных в порядке убывания количества информации, содержащихся в них о принадлежности к данным классам.

Выход на режим генерации информационных портретов классов показан на рисунке 140. На рисунке 141 приведена круговая диаграмма информационного портрета класса (слова) "Достоинству". Обращает внимание, что в 4-х буквах из 8: "У", "Д", "С", "Т" содержится более 80% информации о принадлежности к данному слову.

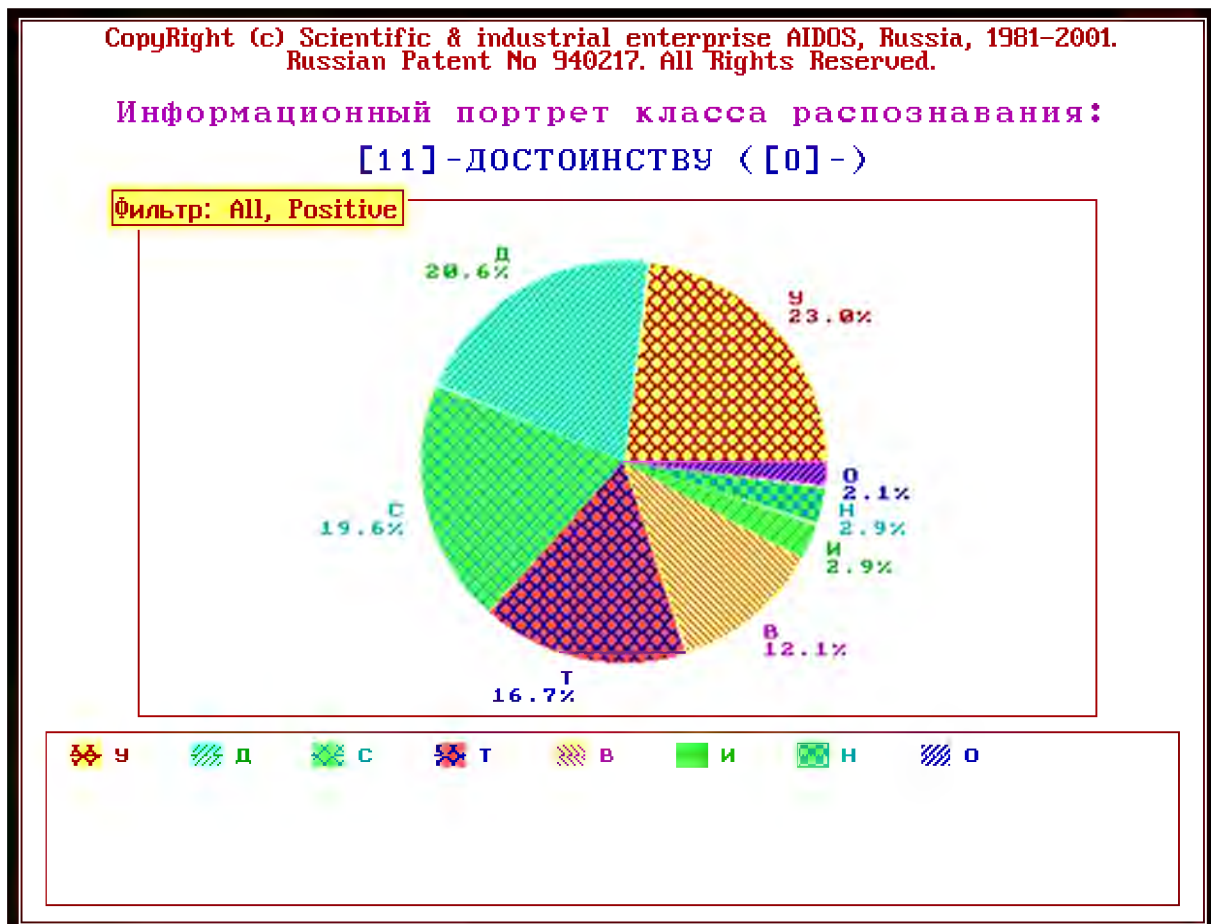


Рисунок 141. Информационный портрет слова "Достоинству"

### Профили слов

Профиль класса представляет собой гистограмму, в которой показан вклад каждого признака в общее количество информации, содержащееся в образе данного класса.

Профили классов и признаков отображаются в 4-м режиме 6-й подсистемы системы "Эйдос" (рисунок 142). Для примера на рисунке 143 показан профиль того же слова "Достоинству".

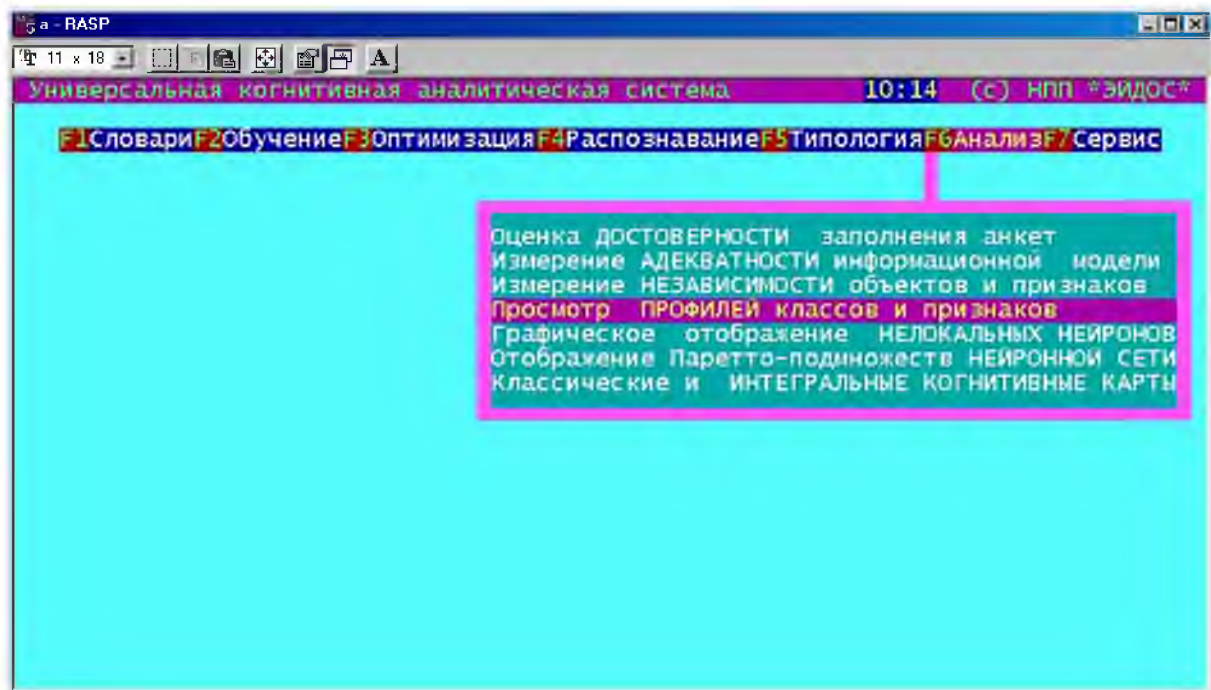


Рисунок 142. Выход на режим вывода профилей классов и признаков

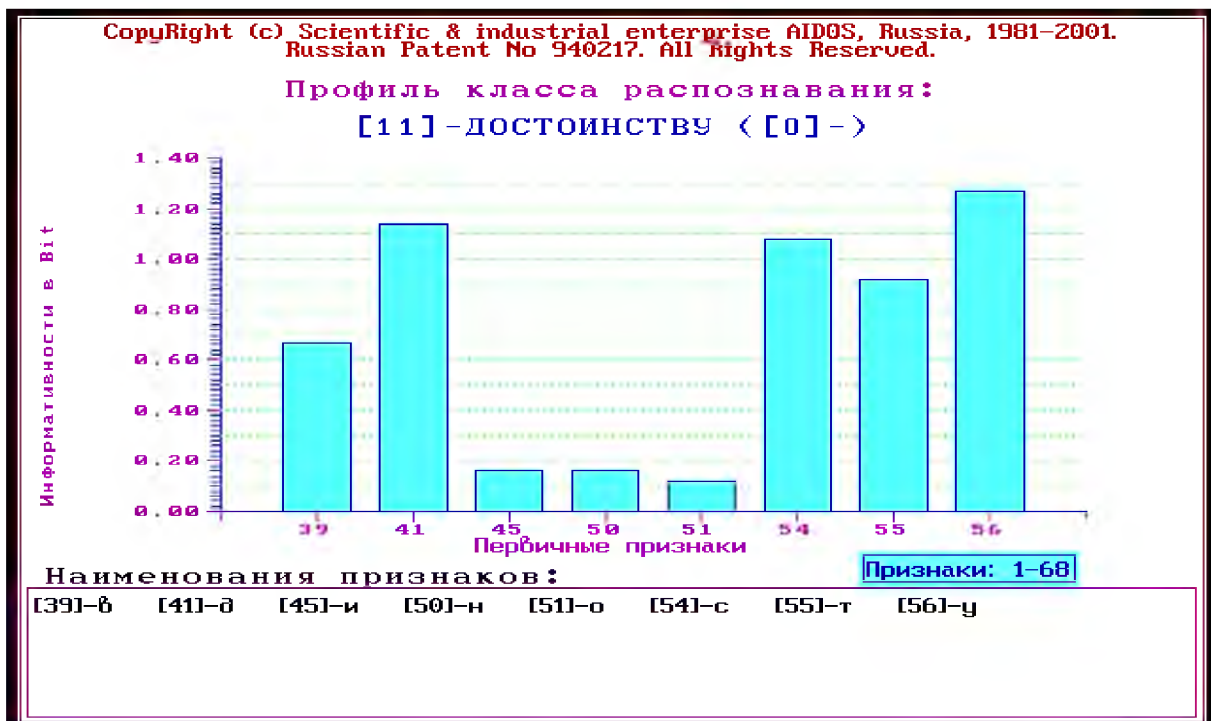


Рисунок 143. Профиль слова "Достоинству"

### Семантические портреты и профили букв

Выход на режим генерации семантических портретов признаков (букв) показан на рисунке 140. Один таких портретов, а именно портрет буквы "Й", приведен на рисунке 144, а ее профиль – на рисунке 145.

Copyright (c) Scientific & industrial enterprise AIDOS, Russia, 1981-2001.  
Russian Patent No 940217. All Rights Reserved.

Информационный портрет признака:  
[46]-Й

Фильтр: All, Positive

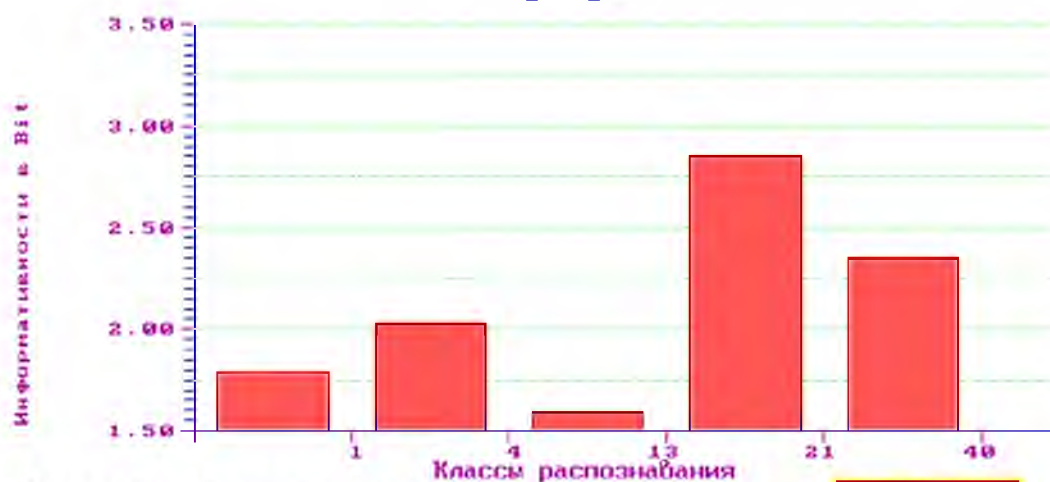


НЕИ СОБОИ ВЕЛИКИЙ БЕЛИНСКИЙ ИСПОЛНЕННЫЙ

Рисунок 144. Информационный портрет буквы "Й"

Copyright (c) Scientific & industrial enterprise AIDOS, Russia, 1981-2001.  
Russian Patent No 940217. All Rights Reserved.

ПРОФИЛЬ ПРИЗНАКА  
[46]-Й



Наименования классов:

Классы: 1-44

[1]-БЕЛИНСКИЙ [4]-ВЕЛИКИЙ [13]-ИСПОЛНЕННЫЙ [21]-НЕИ [40]-СОБОИ

Рисунок 145. Профиль буквы "Й"



**Пример решения задания 8: "Вывести в графической форме семантические сети и когнитивные диаграммы слов и букв, классическую и интегральную когнитивные карты"**

Эти возможности реализуются в режимах, выход на которые показан на рисунке 140. Семантическая сеть классов отображает в графической форме результаты кластерно-конструктивного анализа слов (рисунок 146), а признаков – букв (рисунок 147). Красный цвет линии связи в семантических сетях означает сходство, а синяя – различие, толщина линии соответствует степени выраженности этого свойства. Детальную структуру любой линии связи на семантических сетях можно исследовать на когнитивных диаграммах классов (рисунок 148) и признаков (рисунок 149). Классическая когнитивная карта для слова "Останавливать" приведена на рисунке 150. Классическая когнитивная карта представляет собой графическую диаграмму, объединяющую изображение нейрона и семантической сети его рецепторов. Классическая когнитивная карта представляет собой подмножество "Интегральной когнитивной карты", в которой объединены семантический сети нейронов и рецепторов и фрагмент нейронной сети, соединяющей рецепторы с нейронами. Режим системы "Эйдос" "Классические и интегральные когнитивные карты" позволяет при соответствующих параметрах, задаваемых в диалоге, генерировать и выводить в графической форме как интегральные, так и классические когнитивные карты, а также инвертированные когнитивные карты, представляющие собой семантическую сеть нейронов, соединенных фрагментов нейронной сети с одним рецептором.

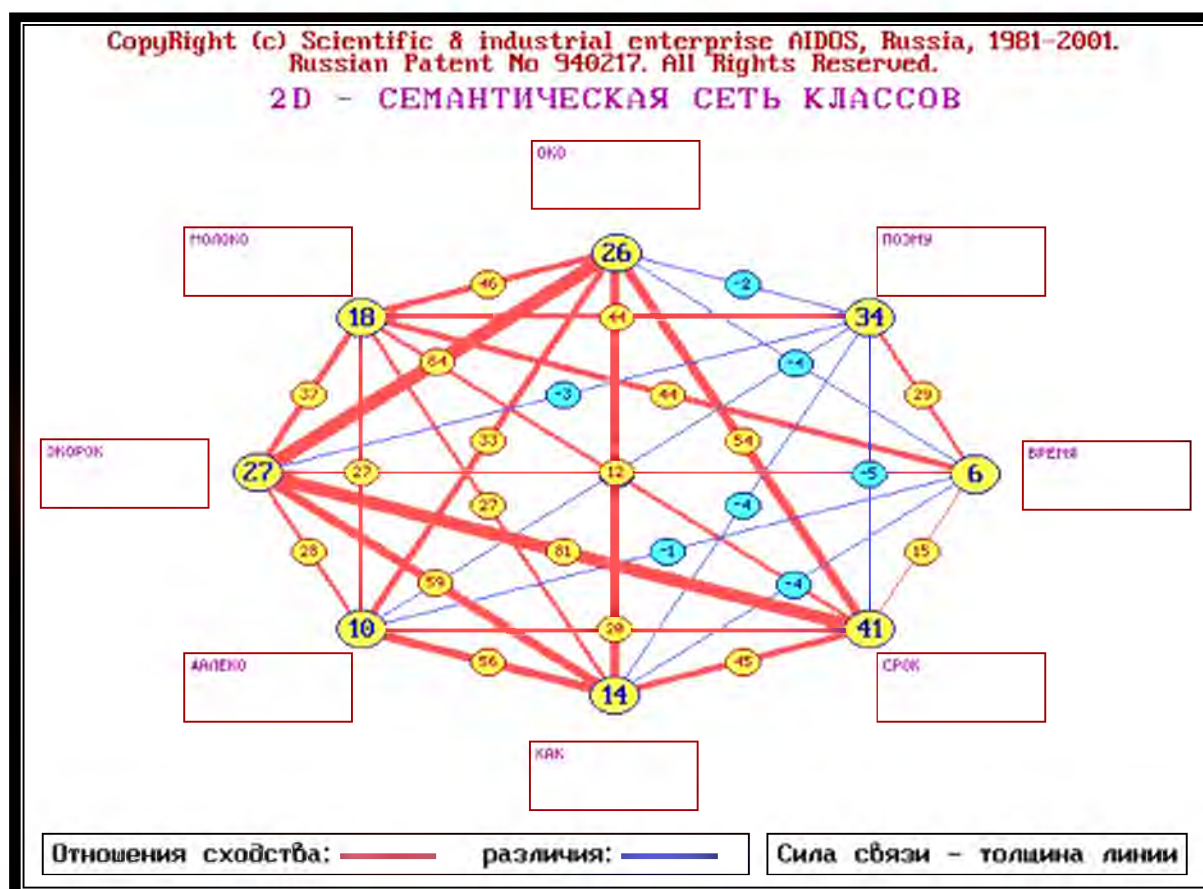


Рисунок 146. Семантическая сеть классов (слов)

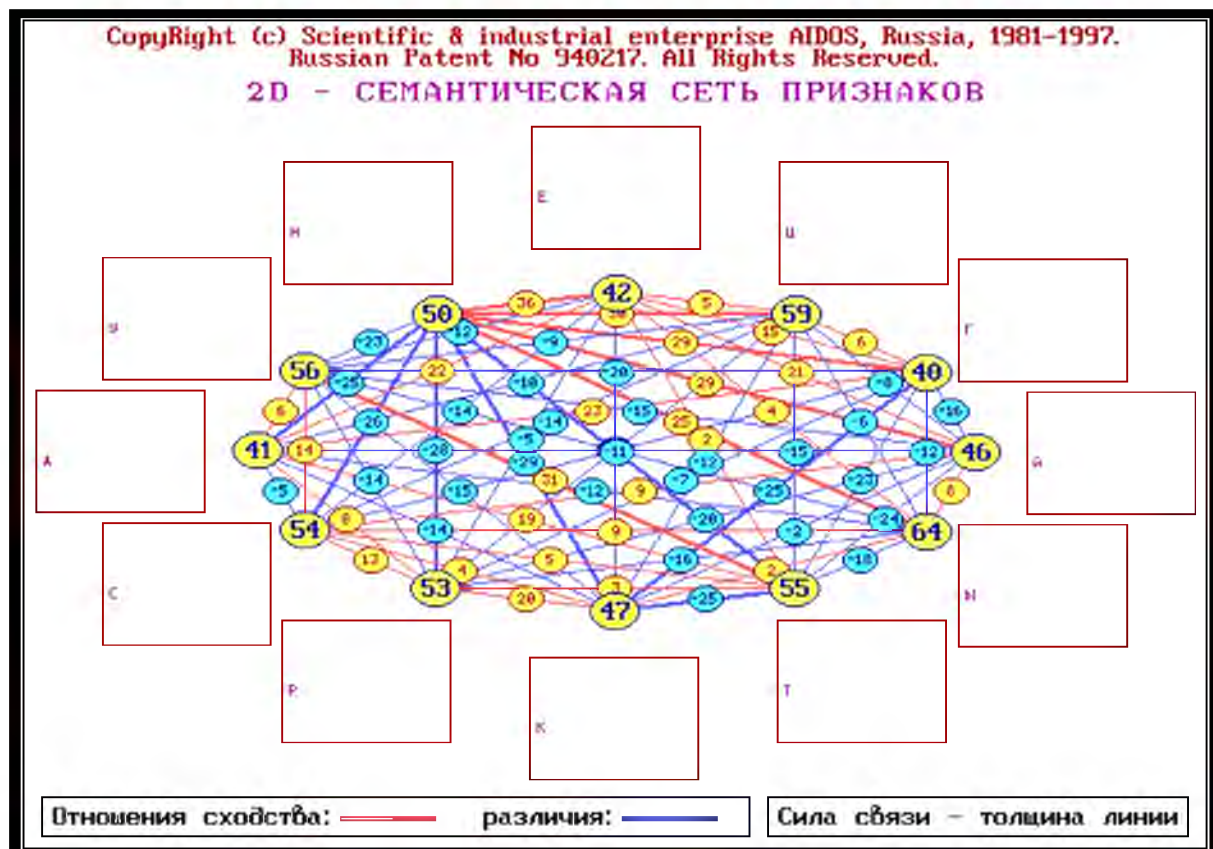


Рисунок 147. Семантическая сеть признаков (букв)

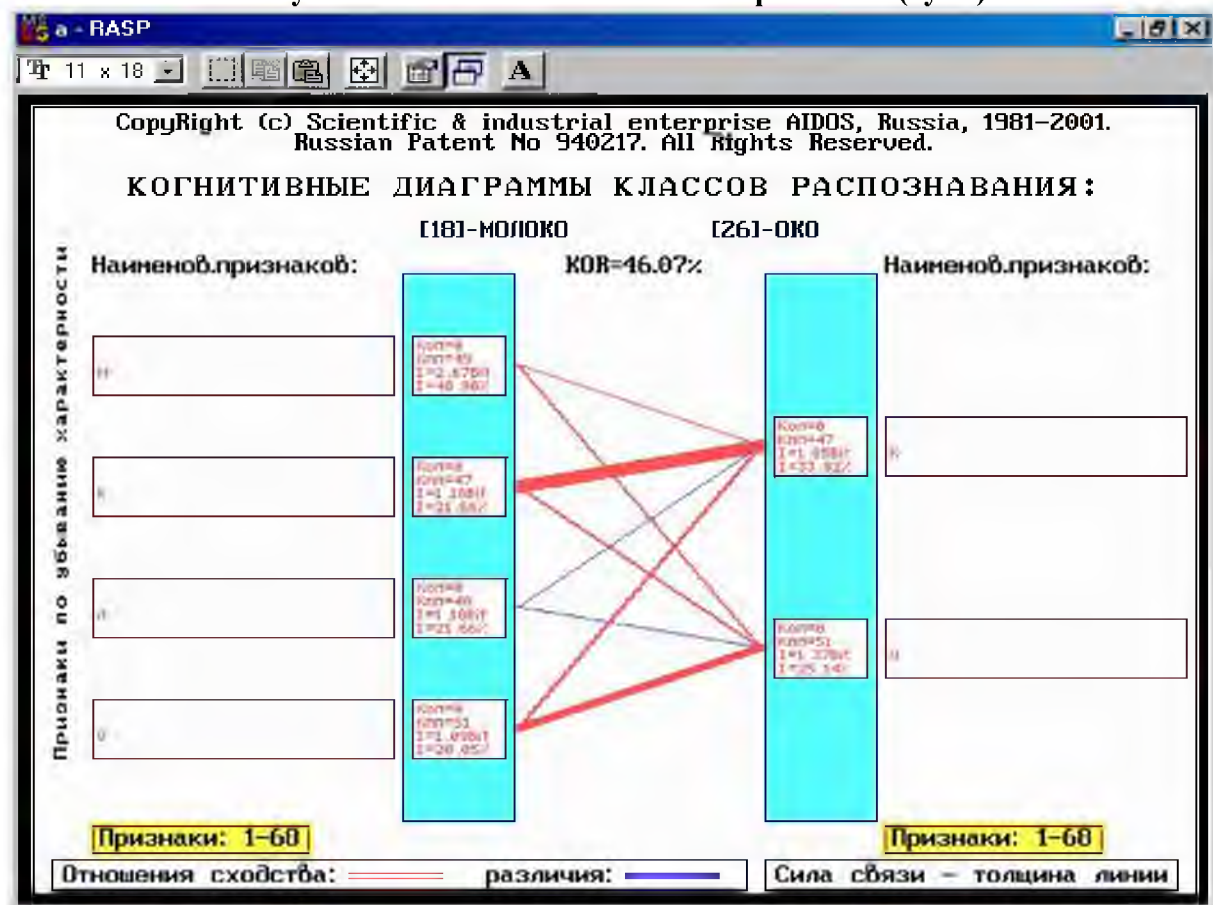
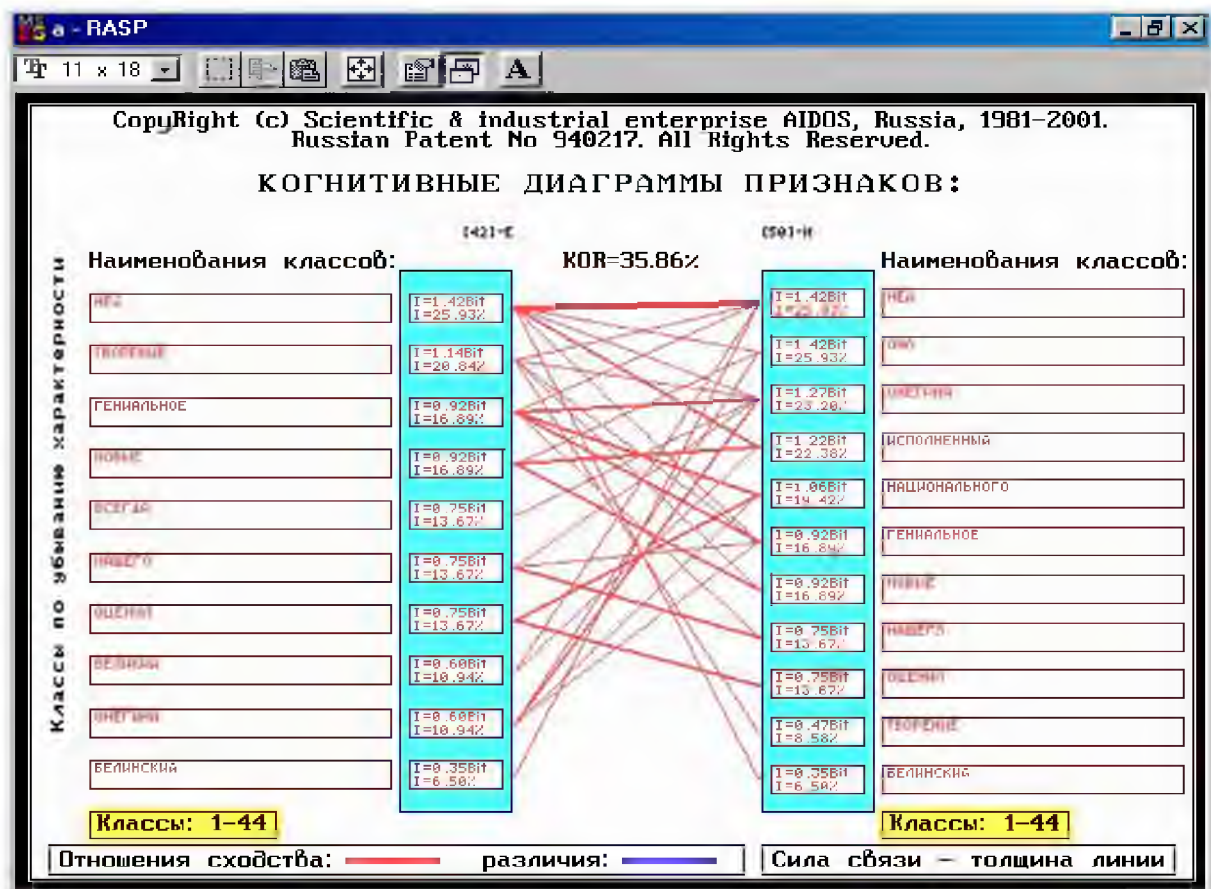


Рисунок 148. Детальная структура сходства слов "Молоко" и "Око"



**Рисунок 149. Детальная структура сходства букв "Е" и "Н"**

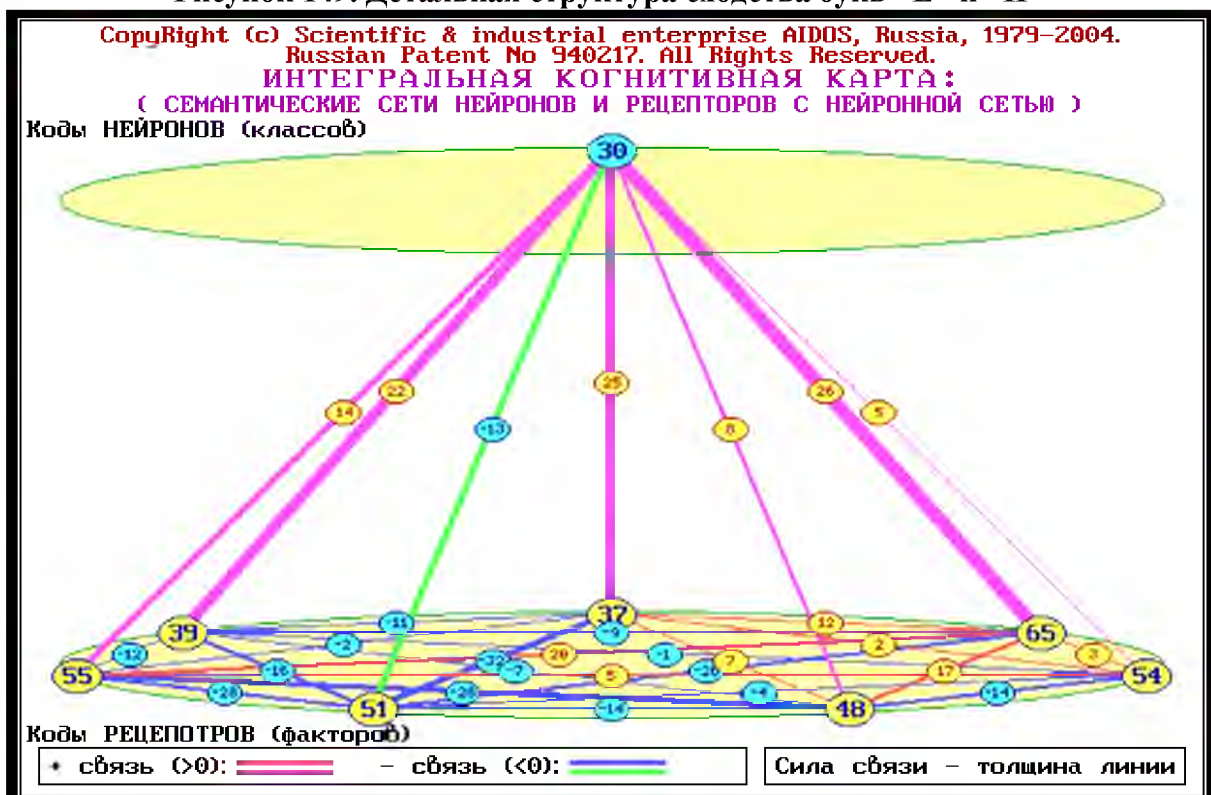


Рисунок 150. Классическая когнитивная карта для слова "Останавливать", генерируемая системой "Эйдос", начиная с версии 10.7



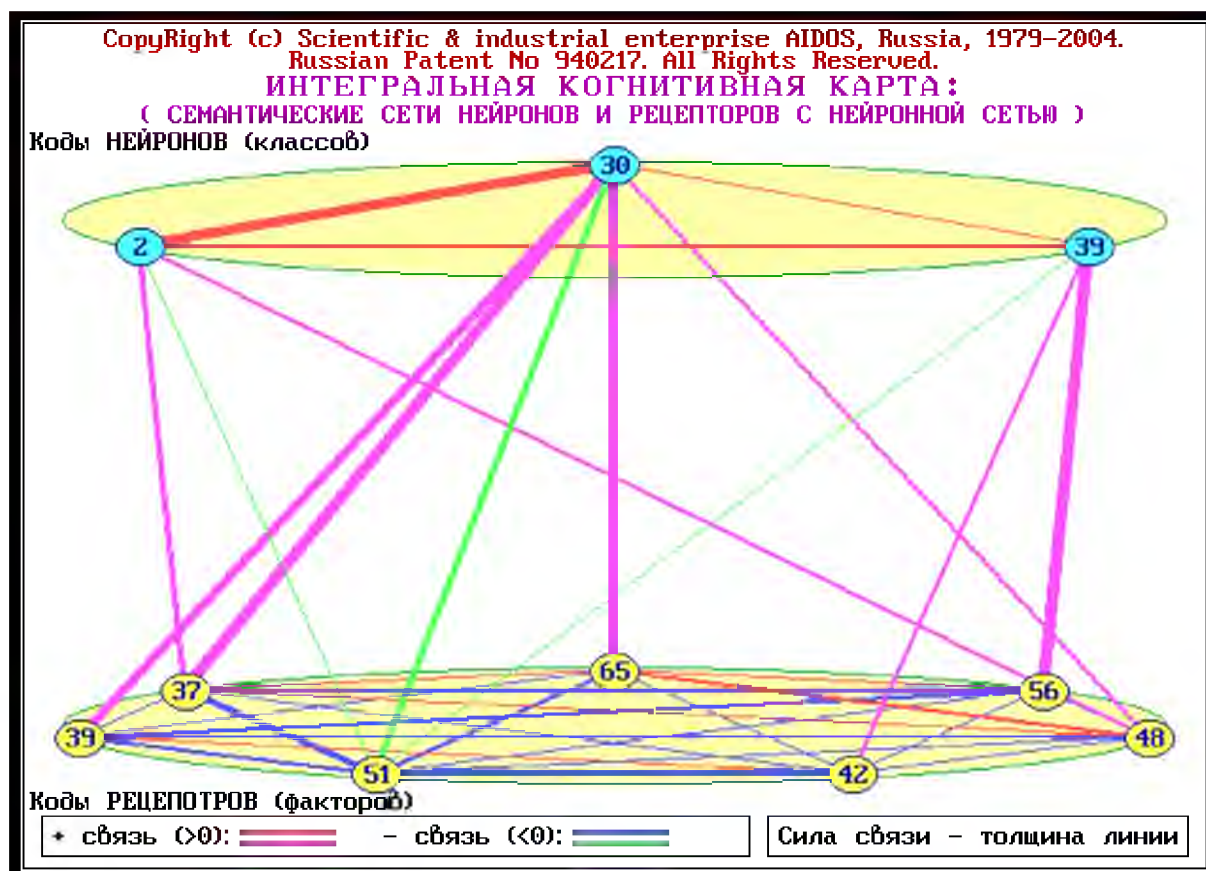


Рисунок 151. Пример интегральной когнитивной карты, генерируемой системой "Эйдос" версии 12.5

## Выводы

На основе вышеизложенного можно сделать вывод о том, что для надежной и достоверной идентификации слов (по крайней мере при сравнительно небольшом их количестве) вполне достаточно информации о входящих в них буквах, и для этого нет особой необходимости привлекать дополнительную информацию о последовательности букв и их сочетаний. Продемонстрирована устойчивость модели от неполноты и зашумленности информации. Приведено более 30 графических форм, генерируемых системой "Эйдос", в т.ч. новые формы – классическая и интегральная когнитивные карты.

## Контрольные вопросы

1. Каким образом провести анализ устойчивости модели к неполноте информации и наличию шума?
2. Каким образом можно проверить способность модели правильно идентифицировать классы, один из которых является подмножеством другого?
3. Как оценить ценность букв для идентификации слов. Сравнить суммарную ценность для этой цели гласных и согласных букв?

## Литература по лабораторной работе

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.

## ЛР-6: "Атрибуция анонимных и псевдонимных текстов"

### ***Краткая теория***

Данная лабораторная работа является продолжением предыдущей, поэтому внимательно прочитайте теорию по предыдущей лабораторной работе. В этой работе исследуется возможность атрибуции текстов с применением технологии и инструментария системно-когнитивного анализа. Приведен подробный численный пример реализации всех этапов СК-анализа при атрибуции текстов, т.е. когнитивной структуризации и формализации предметной области; формирования обучающей выборки; синтеза семантической информационной модели; ее оптимизации и измерения адекватности; адаптации и пересинтеза; а также типологического и кластерно-конструктивного анализа.

*Под атрибуцией анонимных и псевдонимных текстов понимается установление их вероятного авторства* ([88, 89] рекомендуемой литературы).

Анонимные тексты – это тексты вообще без подписи автора, а псевдонимные – подписанные не фамилией автора, а псевдонимом.

Задача идентификации текстов на основе анализа предложений является тривиальной из-за практически абсолютной уникальности предложений. Поэтому больший интерес представляет задача идентификация текстов на основе анализа слов, т.е. задача атрибуции текстов, имеющая очень большое научное и практическое значение. К этой задаче сводится определение вероятного авторства текстов в случае, когда автор не указан (анонимный текст) или указан его псевдоним (псевдонимный текст), а также датировка текста.

Но самое главное, что *к задаче атрибуции текстов сводятся задачи идентификации, прогнозирования, сравнения и классификации объектов, описанных на естественном языке (причем не важно, на каком именно).*

С ней связаны также задачи автоматического выделения дескрипторов и задачи нечеткого поиска и идентификации.

Все эти задачи имеют практическое значение для специалистов по прикладной информатике в экономике и юриспруденции, которых готовит Кубанский государственный аграрный университет.

Одному из вариантов решения этих задач с применением интеллектуальной технологии "Эйдос" и посвящена данная лабораторная работа.

### ***Задания***

Следуя логике Системно-когнитивного анализа выполнить следующие работы.

1. Осуществить когнитивную структуризацию предметной области.

2. Выполнить формализацию предметной области.
3. Сформировать обучающую выборку.
4. Осуществить синтез семантической информационной модели.
5. Оптимизировать семантическую информационную модель.
6. Проверить семантическую информационную модель на адекватность, измерить внутреннюю и внешнюю, дифференциальную и интегральную валидность.
7. Выполнить адаптацию модели и измерить, как изменилась ее адекватность.
8. Осуществить пересинтез модели и измерить, как изменилась ее адекватность.
9. Вывести информационные портреты текстов и дать их интерпретацию.
10. Выполнить кластерно-конструктивный анализ модели.

## **Пример решения**

### ***1. Осуществить когнитивную структуризацию предметной области.***

Под когнитивной структуризацией в СК-анализе понимается определение причин и следствий, факторов и состояний объекта управления, исходной информации и того, на что она влияет.

В данной лабораторной работе необходимо решить задачу идентификации текстов по входящим в них словам. Следовательно, необходимо будет сформировать обобщенные образы текстов, соответствующих определенной тематике или автору (будем считать, что сочинение принадлежит тому писателю, творчеству которого оно посвящено). Для этого в качестве объектов обучающей выборки использоваться фрагменты текстов школьных сочинений, взятые из Internet, а в качестве признаков текстов будут использоваться входящие в них слова.

Каждое сочинение разобьем случайным образом на примерно равные по размеру небольшие фрагменты, которые используем в качестве объектов обучающей выборки.

### ***2. Выполнить формализацию предметной области.***

Под формализацией предметной области понимается разработка классификационных и описательных шкал и градаций и ввод их в программную систему "Эйдос", являющуюся инструментарием СК-анализа.

#### ***2.1. Формирование классификационных шкал и градаций***

В подсистеме "Классификационные шкалы и градации" введем классы, соответствующие следующим писателям: Ф.М. Достоевский; Н.В. Гоголь; А.С. Грибоедов; М.Ю. Лермонтов; А.С. Пушкин; Л.Н. Толстой; И.С. Тургенев (рисунок 152).

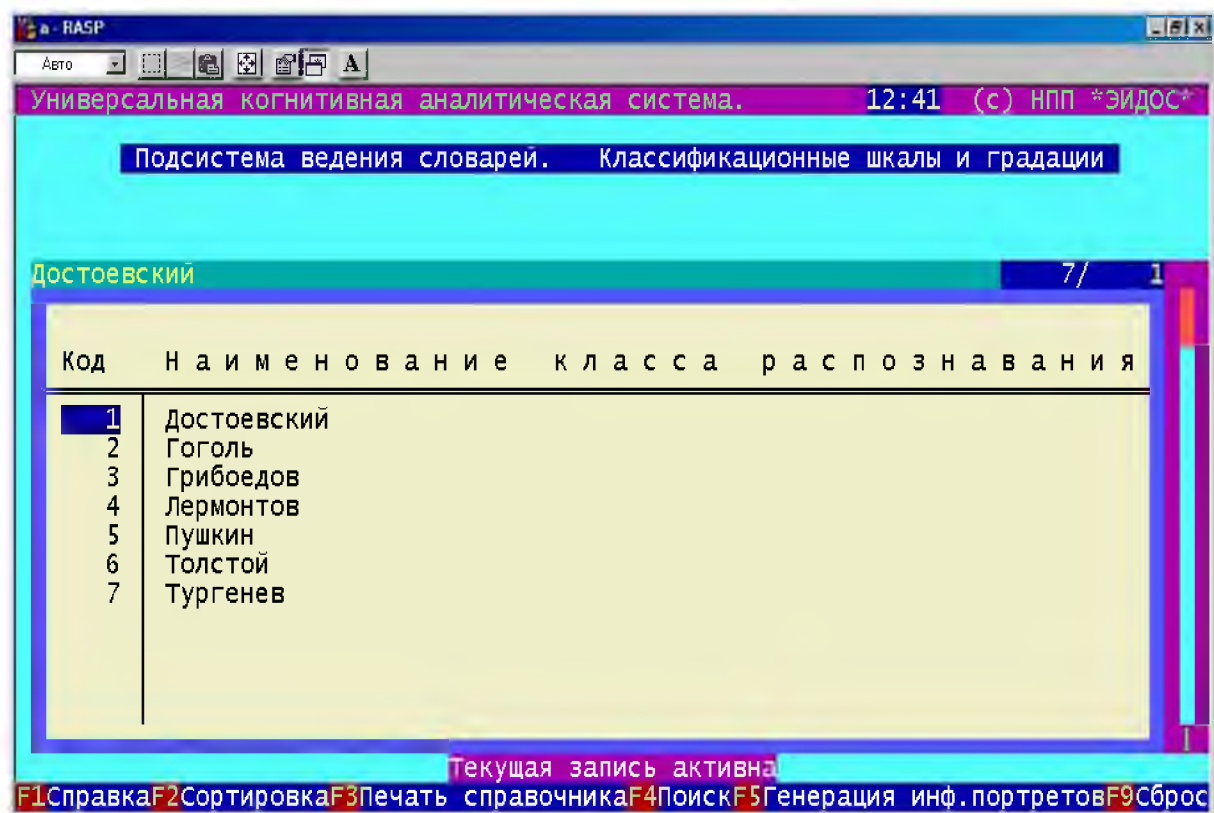


Рисунок 152. Ввод классов

### 2.1. Формирование описательных шкал и градаций

Для этого исходные файлы для формирования объекты обучающей выборки должны быть средствами Word представлены в виде текстовых файлов, стандарта "Текст DOS" (без разбиения на строки).

Затем каждый из этих файлов разбивается на столько файлов, сколько в нем строк, причем имена этих файлов должны иметь вид: #####SUBSTR(File_name,4).TXT, где ##### – сквозной номер файлов, соответствующий будущему номеру анкеты обучающей выборки, SUBSTR(File_name,4) – первые 4 символа имени исходного файла.

Полученные файлы должны быть помещены в поддиректорию DOB системы "Эйдос", а исходные – удалены из нее.

Это осуществляется одним из трех способов:

1. Вручную.
2. С использованием специальной программы, текст которой приводится ниже (язык программирования xBase).

3. В режиме: "Словари – Программные интерфейсы для импорта данных – Импорт данных из TXT-файлов стандарта "Текст DOS", формируем описательные шкалы и градации (рисунок 153), причем в качестве признаков выбираем слова.

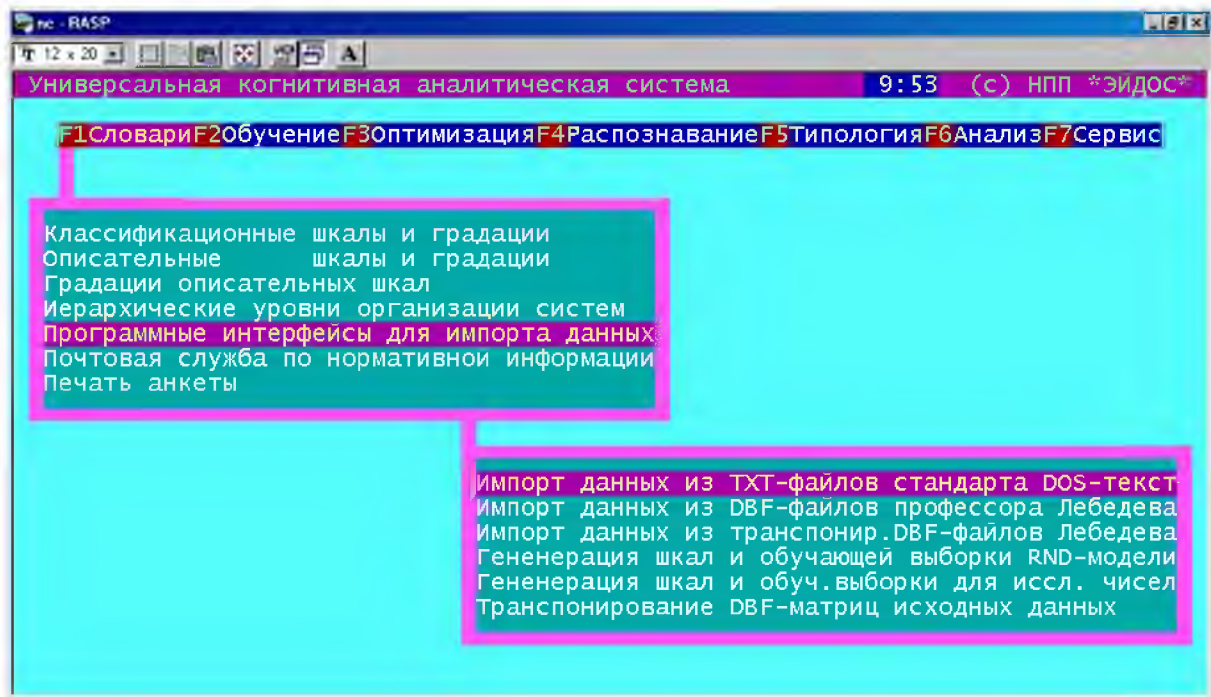
*Исходный текст программы записи TXT-файлов с данными по строкам*

```

*****
***** Разбиение текстовых файлов DOS на нумерованные файлы по строкам
***** Луценко Е.В., 03/31/04 04:24pm
*****

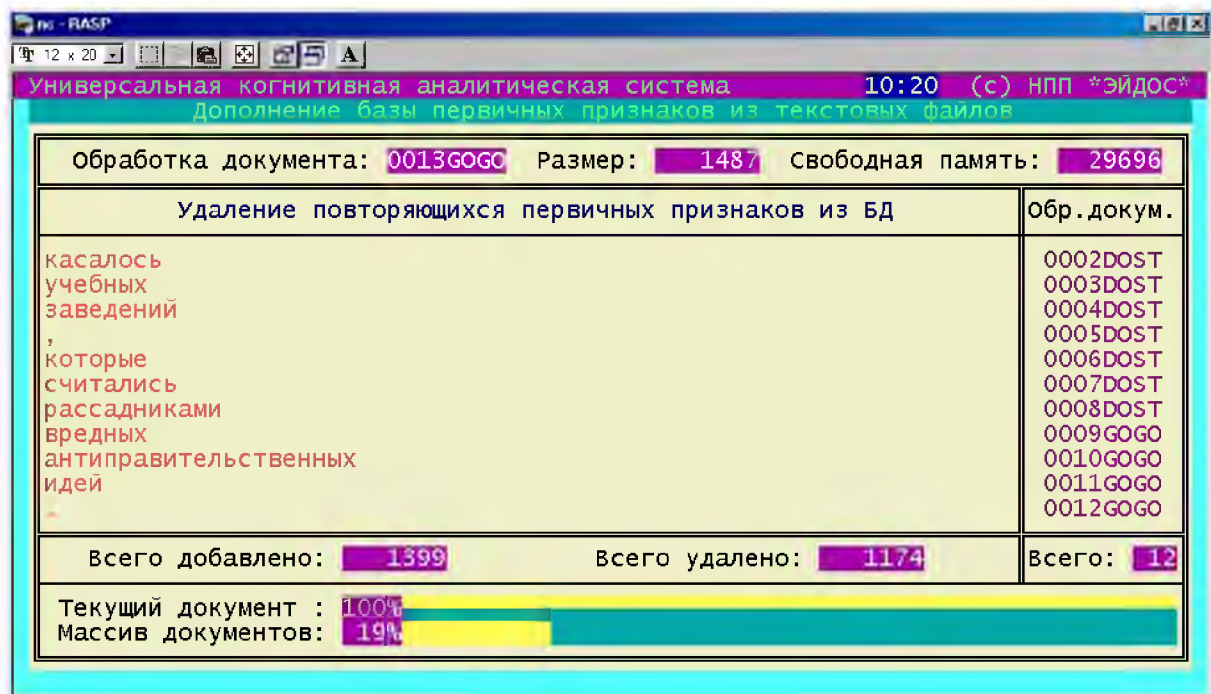
scr_start=SAVESCREEN(0,0,24,79)
SHOWTIME(0,58,.T.,"rb/n")
FOR j=0 TO 24
    @j,0 SAY SPACE(80) COLOR "n/n"
NEXT
***** Удаление TXT-файлов, имена которых начинаются на 0
FILEDELETE("0*.TXT")
***** РЕКОГНОСЦИРОВКА
Count = ADIR("*.TXT")          && Кол-во TXT-файлов
IF Count = 0
    Mess = "В текущей директории TXT-файлов не обнаружено !!!"
    @15,40-LEN(Mess)/2 SAY Mess COLOR "gr+/n"
    INKEY(0)
    RESTSCREEN(0,0,24,79,scr_start)
    SHOWTIME()
    QUIT
ENDIF
PRIVATE Name[Count],Size[Count]  && Имена и размеры файлов
Count = ADIR("*.txt",Name,Size)
SortData(Name,Size,LEN(Name),1)  && Сортировка файлов по алфавиту
CrLf = CHR(13)+CHR(10)           && Конец строки (абзаца) (CrLf)
*** Загрузка TXT-файлов
Num_pp = 0                       && Номера выходных файлов
FOR f = 1 TO Count               && Начало цикла по TXT-файлам
    ***** Загрузка файла
    Buffer = FILESTR(Name[f],.T.)
    Buffer = CHARONE(" ",Buffer)  && Удаление повторяющихся пробелов
    Buffer = Buffer + CrLf
    Len = AT(CrLf,Buffer)
    DO WHILE Len > 0 .AND. LASTKEY() <> 27  && Цикл по строкам
        Len = AT(CrLf,Buffer)
        IF Len > 0
            ***** Запись фрагмента файла
            Str_pr = ALLTRIM(SUBSTR(Buffer,1,Len-1))
            Fn_out = STRTRAN(STR(++Num_pp,4)," ","0")+SUBSTR(Name[f],1,4)+".TXT"
            STRFILE(Str_pr,Fn_out)
            ***** Исключение из буфера записанной строки
            Buffer = ALLTRIM(SUBSTR(Buffer,Len+1))
        ENDIF
    ENDDO
NEXT
*** Удаление исходных TXT-файлов
FOR f=1 TO Count
    FILEDELETE(Name[f])
NEXT
RESTSCREEN(0,0,24,79,scr_start)
SHOWTIME()
QUIT

```



**Рисунок 153. Выход на режим генерации справочников  
на основе текстовых файлов**

На рисунке 154 приведена экранная форма, отображающая ход процесса генерации описательных шкал и градаций и TXT-файлов, содержащих примеры текстов с разбиением по строкам.



**Рисунок 154. Генерация описательных шкал и градаций  
на основе TXT-файлов**

В результате получаем классификационные и описательные шкалы и градации, приведенные в таблицах 79 и 80.

**Таблица 79 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ И ГРАДАЦИИ**

Код	Наименование
1	Достоевский
2	Гоголь
3	Грибоедов
4	Лермонтов
5	Пушкин
6	Толстой
7	Тургенев

**Таблица 80 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ (фрагмент)**

Код	Наименование	Код	Наименование	Код	Наименование
1	!	41	Бедные	81	Все
2	(	42	Без	82	Вспомним
3	(основной	43	Бездушных	83	Встреча
4	)	44	Безумным	84	Всюду
5	,	45	Безумных	85	Вы
6	-	46	Безухов	86	Вызывают
7	.	47	Безухову	87	Высокие
8	1812	48	Белинский	88	Высокопарные
9	20-	49	Бессильной	89	Г
10	30-е	50	Бог	90	Герой
11	30-х	51	Боже	91	Главная
12	60-х	52	Болконский	92	Глухость
13	:	53	Болконскому	93	Говоря
14	;	54	Бордо	94	Гоголь
15	?	55	Борис	95	Гоголя
16	Встает	56	Бориса	96	Годунов
17	XIX	57	Бородинским	97	Горе
18	А	58	Бородинского	98	Гости
19	Автор	59	Буянов	99	Грибоедов
20	Авторский	60	Была	100	Грибоедова
21	Агрессивная	61	В	101	Гулливера
22	Адама	62	Ведь	102	Да
23	Александр	63	Везде	103	Даже
24	Александра	64	Век	104	Дворянин-аристократ
25	Алексевна	65	Великий	105	Действительно
26	Алексеевна	66	Великолепная	106	Дельвигу
27	Аммоса	67	Вернулся	107	Денисова
28	Андреевич	68	Взволнованный	108	Дидло
29	Андрей	69	Взгляды	109	Для
30	Андрею	70	Власы	110	Дмитриевна
31	Анной	71	Вместе	111	Добролюбова
32	Архивам	72	Внешней	112	Достоевского
33	Афанасьевича	73	Внешние	113	Драматична
34	Ах	74	Воды	114	Друбецкого
35	Базаров	75	Возникает	115	Другое
36	Базарова	76	Война	116	Думы
37	Базаровым	77	Вообще	117	Дуни
38	Балы	78	Вопрос	118	Дуня
39	Бегущим	79	Вот	119	Душа
40	Бедность	80	Время	120	Евгений



Приводится лишь фрагмент описательных шкал и градаций, т.к. размерность справочника составляет 3522 градации (т.е. слова).

### 3. Сформировать обучающую выборку

Обучающая выборка представляет собой фрагменты текстов различных авторов, используемые в качестве примеров для формирования семантической информационной модели. На основе анализа этих примеров выявляются взаимосвязи между теми или иными словами и принадлежностью текстов разным авторам.

Для генерации обучающей выборки используется 1-й режим 2-й подсистемы, функция F7InpTXT – F6Ввод из всех файлов. При этом в качестве признаков, также как при формировании описательных шкал и градаций, выбираются слова (рисунок 155).

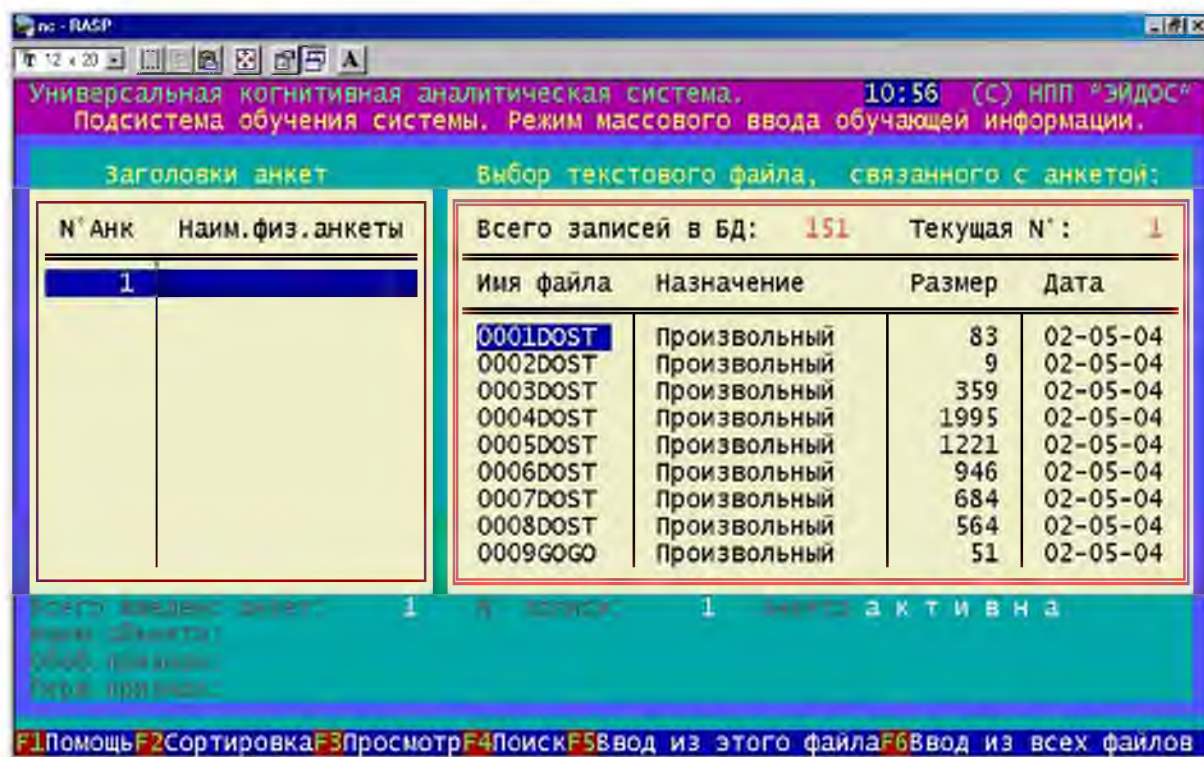


Рисунок 155. Генерация обучающей выборки из TXT-файлов

В результате формируется обучающая выборка, состоящая из 151 примера фрагментов текстов различных авторов. Остается лишь проставить в каждом примере (анкете) код писателя, о котором данный текст, т.е. код класса (в левом окне).

### 4. Осуществить синтез семантической информационной модели

Синтез модели осуществляется во 2-й подсистеме, 4-м режиме, 5-й функции (рисунок 156).



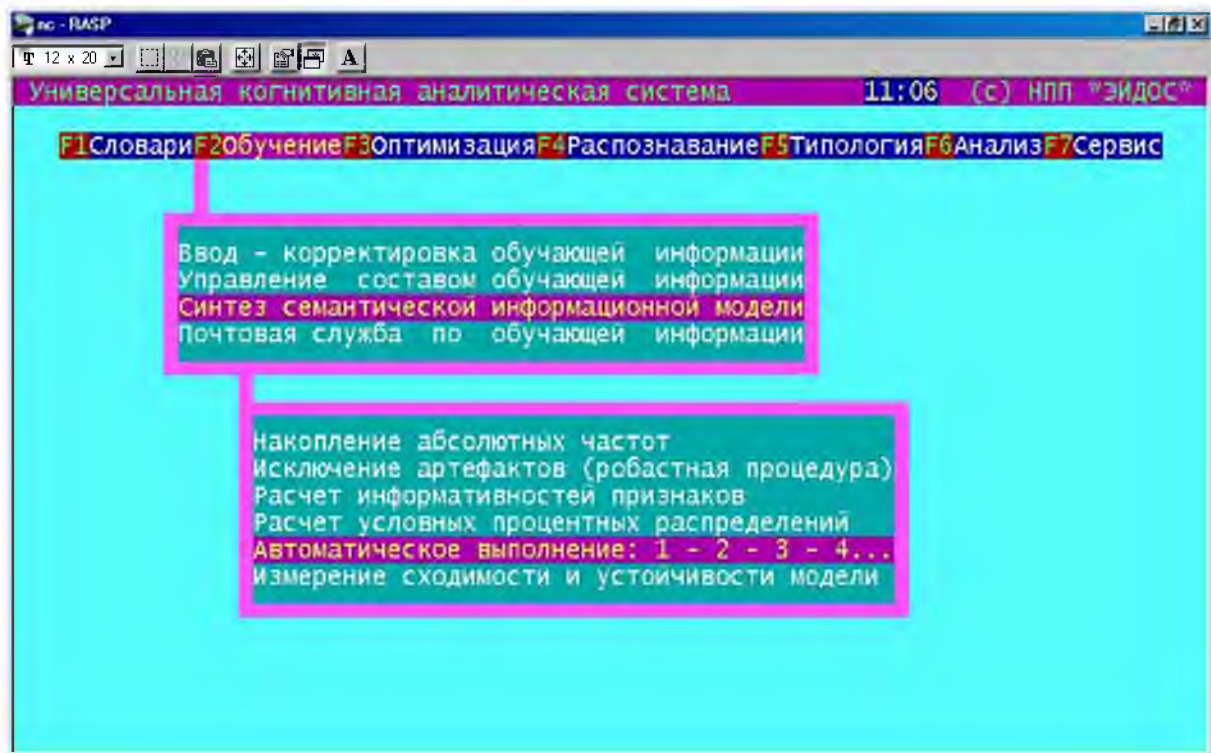


Рисунок 156. Запуск режима:  
"Синтез семантической информационной модели"

Стадия процесса синтеза отображается в ряде экранных форм, одна из которых приводится на рисунке 157.

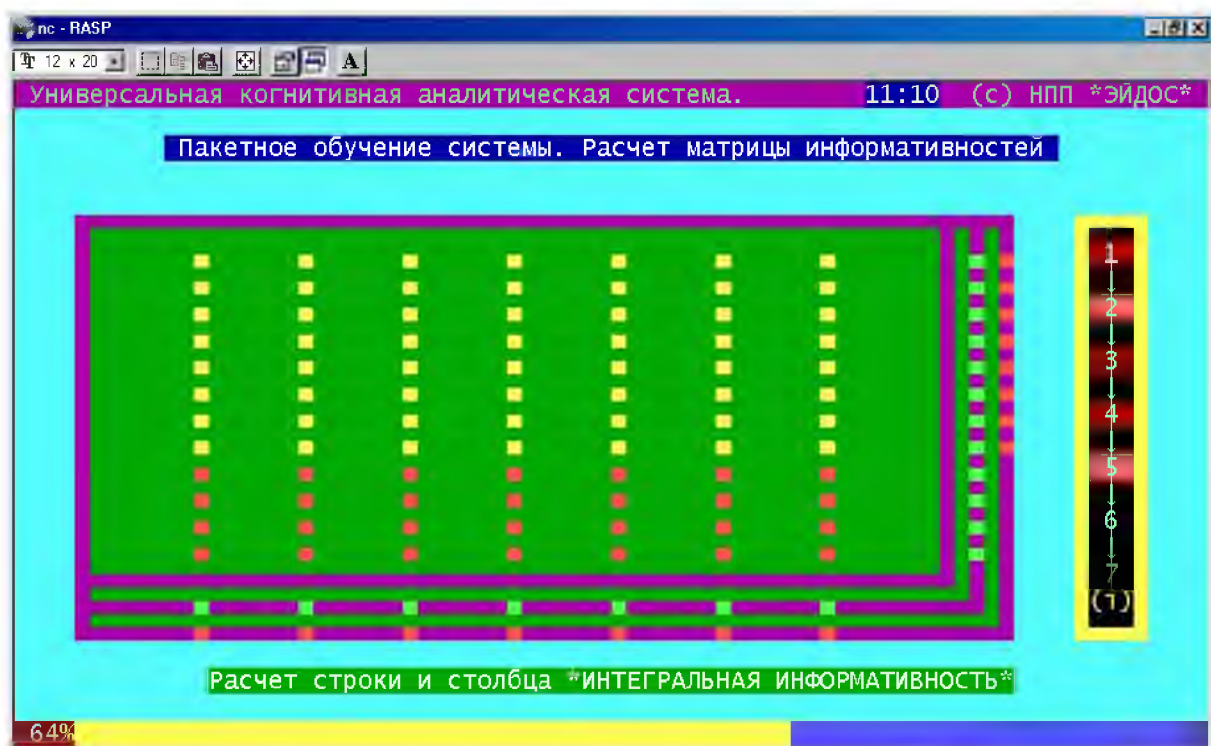


Рисунок 157. Экранная форма, отображающая одну из стадий процесса синтеза семантической информационной модели

### 5. Оптимизировать семантическую информационную модель

Оптимизация модели представляет собой исключение из нее малозначущих признаков без потери адекватности модели. Эта операция осуществляется во 2-м режиме 3-й подсистемы (рисунок 158).

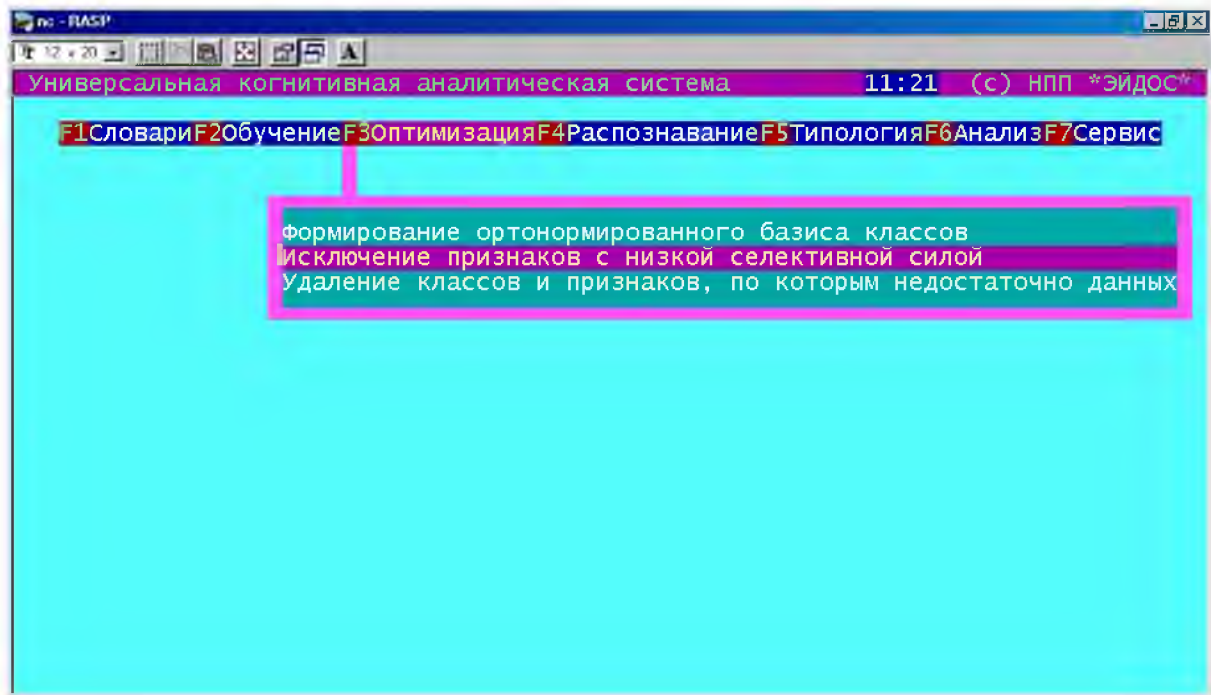


Рисунок 158. Выход на режимы оптимизации модели

При том имеется возможность вывести график ценности признаков "нарастающим итогом", т.е. Паретто-диаграмму признаков (рисунок 159).

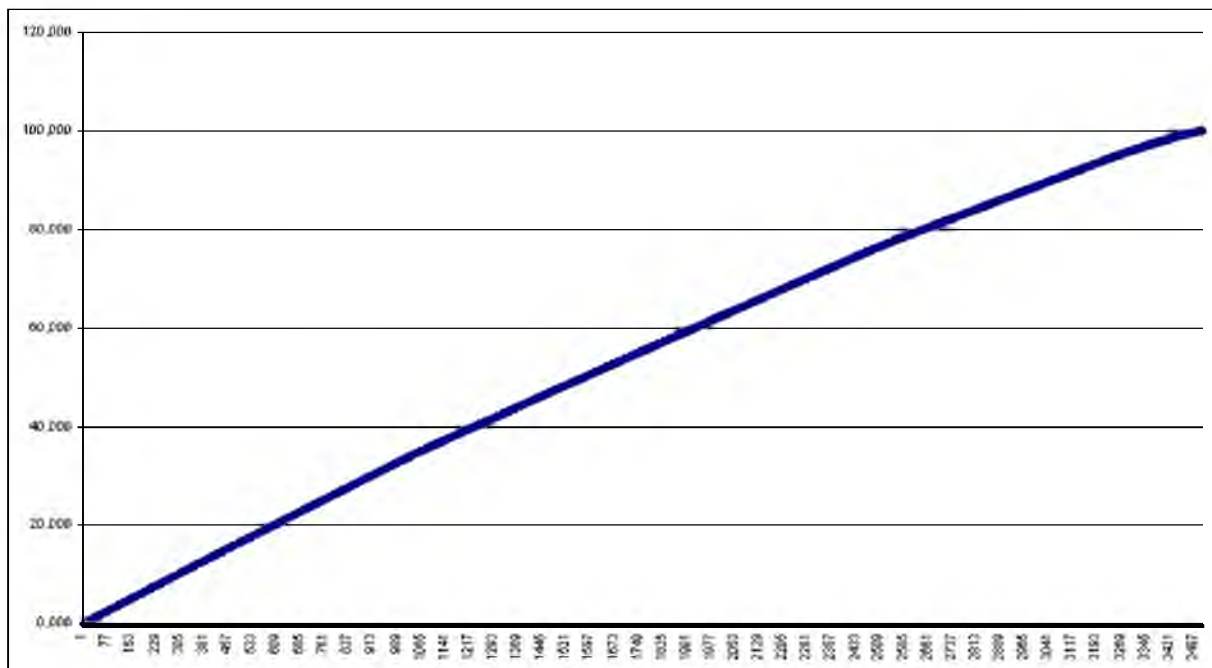


Рисунок 159. Паретто-диаграмма признаков

Видно, что в системе признаков нет имеющих очень малую или нулевую ценность. Это связано с тем, что все слова являются практически уникальными для фрагментов текстов, т.е. встречаются во всех текстах в основном от 1 до 5 раз (рисунок 160).

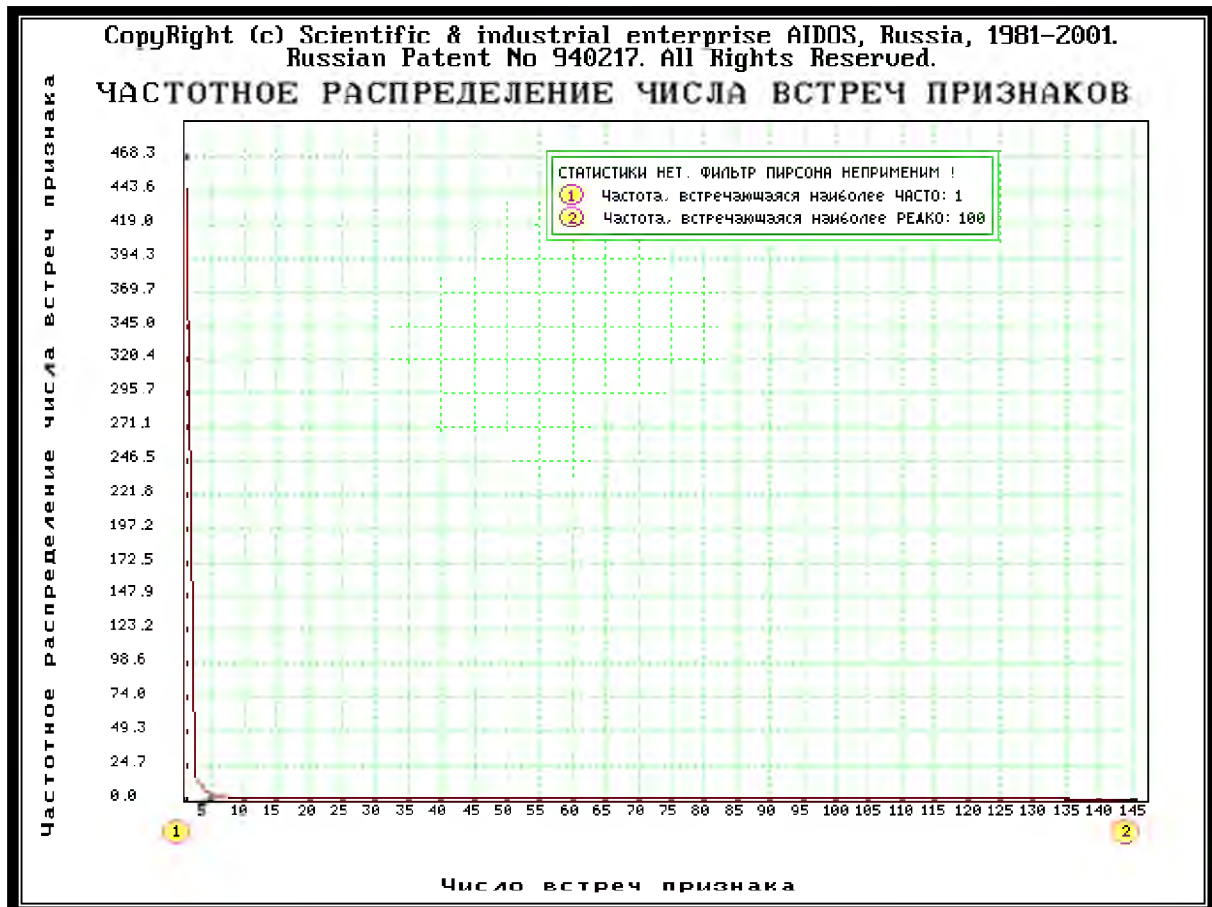


Рисунок 160. Частотное распределение частот признаков

**6. Проверить семантическую информационную модель на адекватность, измерить внутреннюю и внешнюю, дифференциальную и интегральную валидность**

#### **6.1. Внутренняя дифференциальная и интегральная валидность**

Под внутренней валидностью понимается способность модели верно идентифицировать объекты, входящие в обучающую выборку.

Для измерения адекватности модели необходимо выполнить следующие действия:

1. Скопировать обучающую выборку в распознаваемую (во 1-м режиме 2-й подсистемы нажав клавишу F5).
2. Выполнить пакетное распознавание (во 2-м режиме 4-й подсистемы, задав 1-й критерий сходства) (рисунок 161).
3. Измерить адекватность модели (во 2-м режиме 6-й подсистемы) (рисунки 162 и 163).

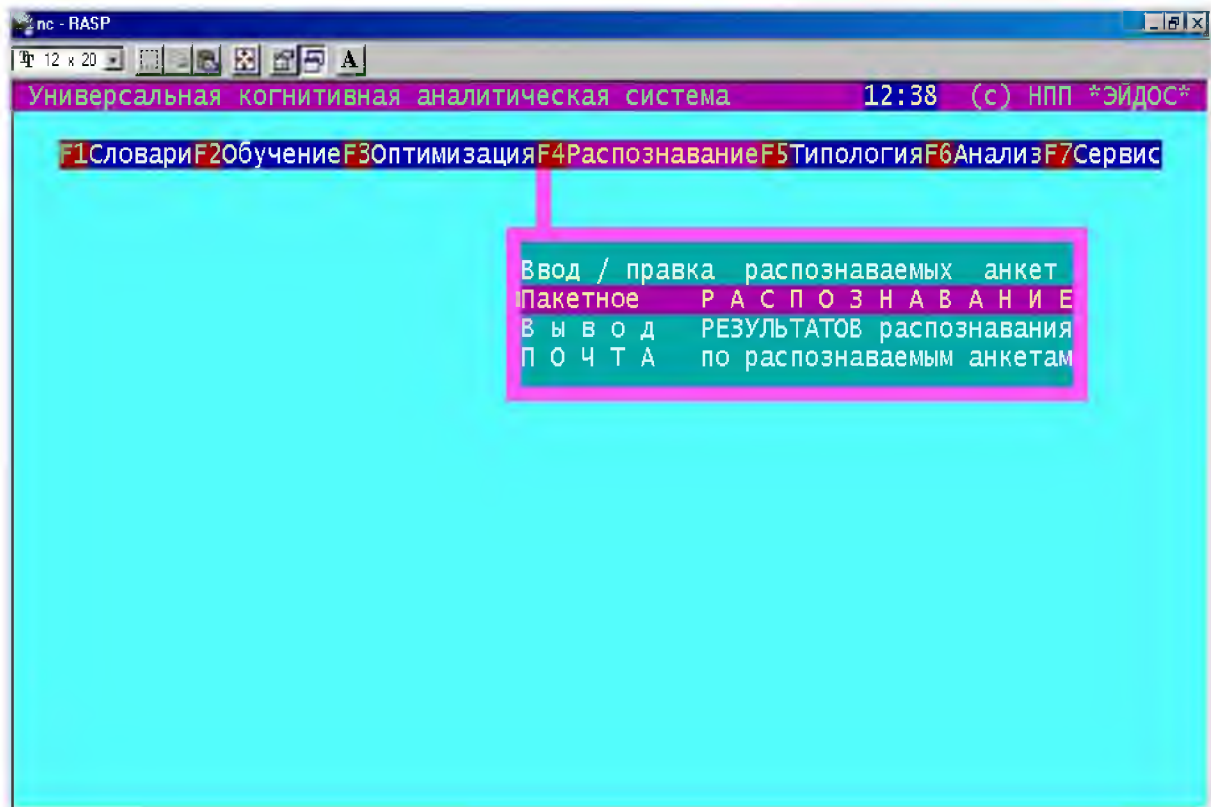


Рисунок 161. Выход на режим пакетного распознавания

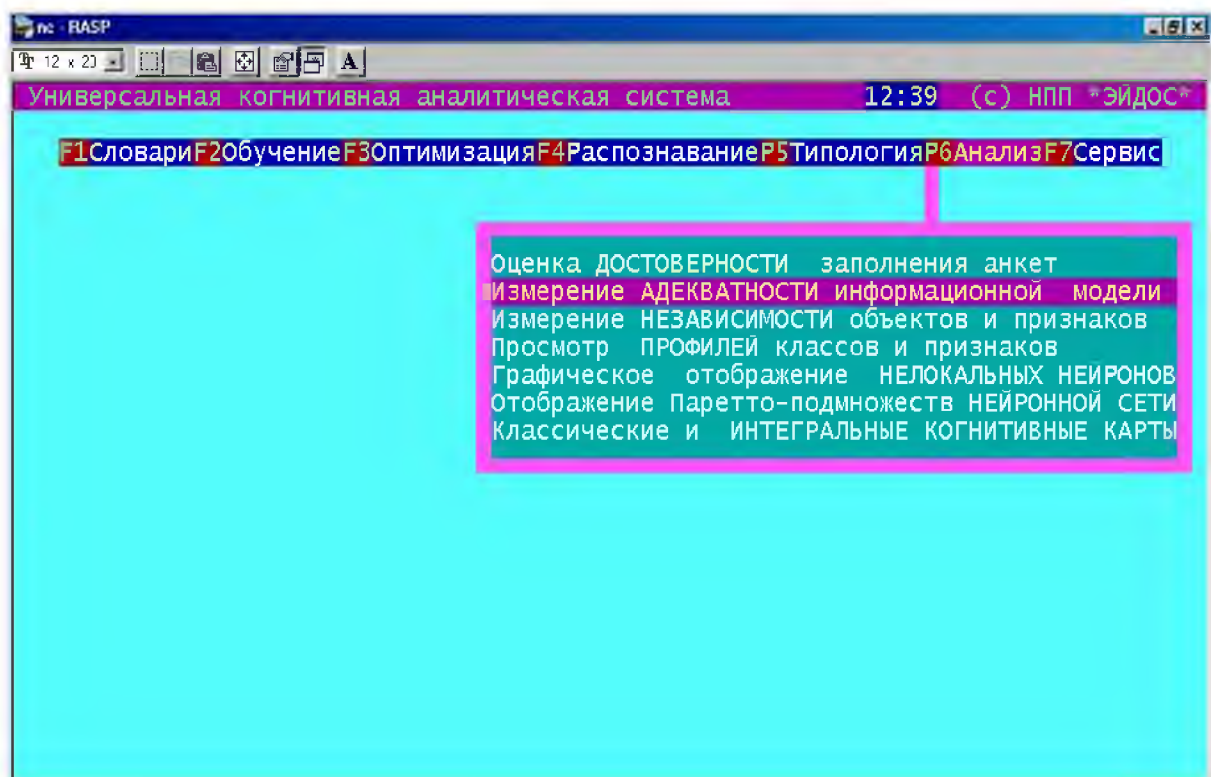


Рисунок 162. Выход на режим измерения адекватности модели



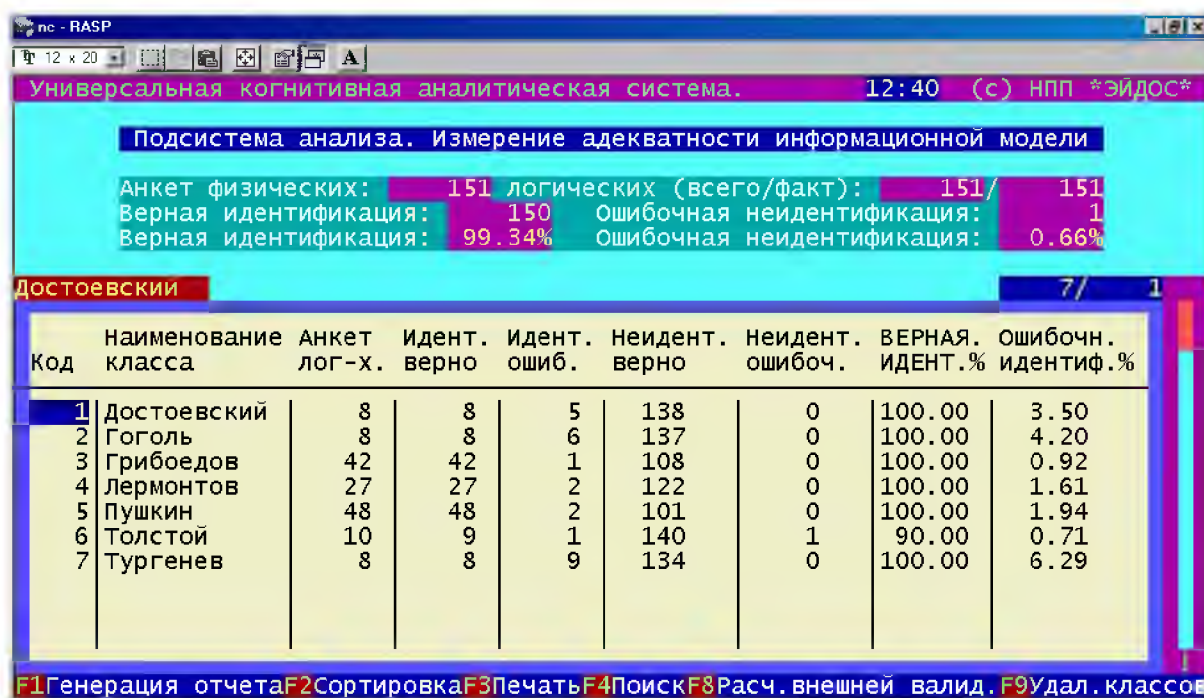


Рисунок 163. Экранная форма управления измерением адекватности модели и отображения результатов

Эта форма может прокручиваться вправо-влево. В верхней части формы приведены показатели интегральной валидности (средневзвешенные по всей обучающей выборке), а в самой таблице – дифференциальной валидности, т.е. в разрезе по классам.

Кроме того, результаты измерения адекватности модели выводятся в форме файлов с именами ValidSys.txt (рисунок 164) и ValAnkSt.txt (рисунок 165) стандарта "TXT-текст DOS" в поддиректории TXT. Первый файл имеет вид:

#### ИЗМЕРЕНИЕ АДЕКВАТНОСТИ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИОННОЙ МОДЕЛИ

Анкет физических: 151 логических (всего/факт): 151/ 151  
 Верная идентификация: 150 Ошибочная неидентификация: 1  
 Верная идентификация: 99.34% Ошибочная неидентификация: 0.66%  
 Минимальный уровень сходства: 0.0 Максимальное кол-во классов: 99999  
 02-05-04 12:40:09

г. Краснодар

N п/п	Код класса	Наименование класса	Всего логич анкет	ИДЕНТИФИЦИР		Неидентифиц		ИДЕНТИФИЦИРОВ		Неидентифицир	
				ВЕРНО	Ошиб.	Верно	Ошиб.	ВЕРНО%	Ошиб.%	Верно%	Ошиб.%
1	1	Достоевский	8	8	5	138	0	100.00	3.50	96.50	0.00
2	2	Гоголь	8	8	6	137	0	100.00	4.20	95.80	0.00
3	3	Грибоедов	42	42	1	108	0	100.00	0.92	99.08	0.00
4	4	Лермонтов	27	27	2	122	0	100.00	1.61	98.39	0.00
5	5	Пушкин	48	48	2	101	0	100.00	1.94	98.06	0.00
6	6	Толстой	10	9	1	140	1	90.00	0.71	99.29	10.00
7	7	Тургенев	8	8	9	134	0	100.00	6.29	93.71	0.00

Универсальная когнитивная аналитическая система

НПП *ЭЙДОС*

Рисунок 164. Выходная форма ValidSys.txt с результатами измерения адекватности модели и отображения результатов

Рассмотрим, что означают графы этой выходной формы.

"Всего логических анкет" – это количество анкет (примеров текстов) в обучающей выборке, на основе которых формировался образ данного класса.

**"Идентифицировано верно"** – это количество анкет обучающей выборки, которые идентифицированы как классы, к которым они действительно относятся.

**"Идентифицировано ошибочно"** – это количество анкет обучающей выборки, которые идентифицированы как классы, к которым они в действительности не относятся (ошибка идентификации).

**"Неидентифицировано верно"** – это количество анкет обучающей выборки, которые неидентифицированы как классы, к которым они действительно не относятся.

**"Неидентифицировано ошибочно"** – это количество анкет обучающей выборки, которые неидентифицированы как классы, к которым они в действительности относятся (ошибка неидентификации).

В правой части формы приведены те же показатели, но в процентном выражении:

– для анкет, идентифицированных верно и неидентифицированных ошибочно за 100% принимается количество логических анкет обучающей выборки по данному классу;

– для анкет, идентифицированных ошибочно и неидентифицированных верно за 100% принимается суммарное количество логических анкет обучающей выборки за вычетом логических анкет по данному классу.

А Н К Е Т Ы распознаваемой выборки  
Класс распознавания : 1 – ДОСТОЕВСКИЙ  
Результат идентификации : Верная идентификация  
Минимальный уровень сходства: 0.0 Максимальное кол-во классов: 99999  
02-05-04 12:40:09 г. Краснодар

К о д ы а н к е т р а с п о з н а в а е м о й в ы б о р к и							
2	3	4	5	6	7	8	

Универсальная когнитивная аналитическая система

НПП *ЭЙДОС*

**Рисунок 165. Фрагмент выходной формы ValAnkSt.txt с результатами измерения адекватности модели и отображения результатов**

В данной форме приведены коды анкет обучающей выборки, которые были учтены в каждой графе предыдущей формы по каждому классу.

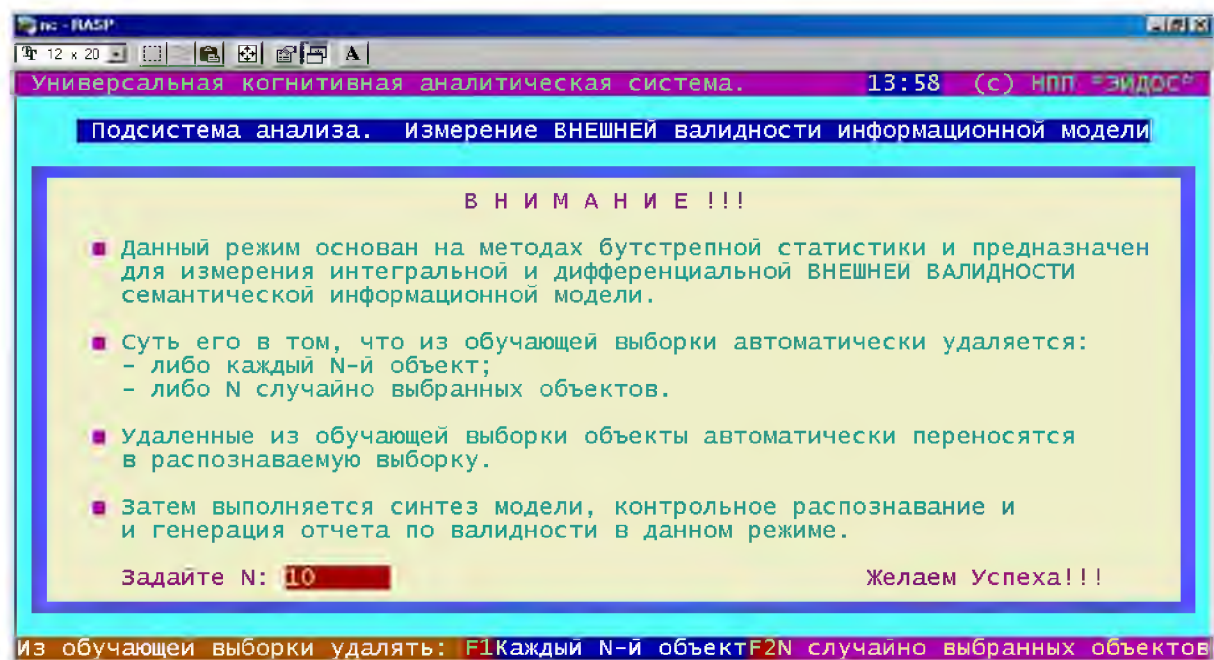
## 6.2. Внешняя дифференциальная и интегральная валидность

Под внешней валидностью понимается способность модели верно идентифицировать объекты, не входящие в обучающую выборку, но относящиеся к генеральной совокупности, по отношению к которой она репрезентативна.

Для измерения внешней валидности необходимо выполнить следующие действия:

1. В режиме измерения адекватности модели запустить режим измерения внешней валидности (нажав F8 Измерение внешней валидности) (рисунок 166).

2. Выбрать один из режимов удаления объектов обучающей выборки, приведенный на экранной форме (рисунок 167).



**Рисунок 166. Режим переноса анкет обучающей выборки в распознаваемую для измерения внешней валидности**

Результат выполнения всех указанных на рисунке 166 действий приведен на рисунке 167.

#### ИЗМЕРЕНИЕ АДЕКВАТНОСТИ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИОННОЙ МОДЕЛИ

Анкет физических: 75 логических (всего/факт): 75/ 75  
 Верная идентификация: 61 Ошибочная неидентификация: 14  
 Верная идентификация: 81.33% Ошибочная неидентификация: 18.67%  
 Минимальный уровень сходства: 0.0 Максимальное кол-во классов: 99999  
 09-05-04 08:20:09

г. Краснодар

N п/п	Код класса	Наименование класса	Всего логич анкет	ИДЕНТИФИЦИР		Неидентифиц		ИДЕНТИФИЦИРОВ		Неидентифицир	
				ВЕРНО	Ошиб.	Верно	Ошиб.	ВЕРНО%	Ошиб. %	Верно%	Ошиб. %
1	1	Достоевский	4	3	55	16	1	75.00	77.46	22.54	25.00
2	2	Гоголь	4	3	47	24	1	75.00	66.20	33.80	25.00
3	3	Грибоедов	21	19	43	11	2	90.48	79.63	20.37	9.52
4	4	Лермонтов	13	8	43	19	5	61.54	69.35	30.65	38.46
5	5	Пушкин	24	19	40	11	5	79.17	78.43	21.57	20.83
6	6	Толстой	5	5	45	25	0	100.00	64.29	35.71	0.00
7	7	Тургенев	4	4	53	18	0	100.00	74.65	25.35	0.00

Универсальная когнитивная аналитическая система

НПП *Эйдос*

**Рисунок 167. Выходная форма с результатами измерения внешней валидности методом бутстрепной статистики**

При этом исходная выборка была разделена на две:

- в обучающей выборке остались только нечетные анкеты;
- в распознаваемую выборку были включены только четные анкеты;

– при распознавании был использован 2-й интегральный критерий: сумма количества информации.

Анализ отчета по внешней валидности, приведенного на рисунке 167, позволяет сделать вывод о высокой степени адекватности семантической информационной модели. Это значит, что взаимосвязи между словами, использованными в текстах, и принадлежностью этих текстов различным авторам, выявленные по примерам обучающей выборки, оказались имеющими силу и для других фрагментов текстов, приведенных в распознаваемой выборке. Это означает, что они относятся к генеральной совокупности, по отношению к которой обучающая выборка репрезентативна.

### ***7. Выполнить адаптацию модели и измерить, как изменилась ее адекватность***

Под адаптацией модели понимается ее количественная модификация, осуществляемая путем включения в обучающую выборку дополнительных примеров реализации объектов, относящихся к тем же самым классам и описанным в той же системе признаков.

**На первом этапе**, для изучения адаптивности модели осуществим ее синтез на основе обучающей выборки, состоящей из нечетных анкет, которая использовалась в примере для измерения внешней валидности. Но в отличие от этого примера эту же выборку используем и как распознаваемую.

**На втором этапе** осуществим синтез модели на основе полной обучающей выборки, включающей как четные, так и нечетные анкеты.

Адаптация модели повышает точность идентификации объектов той же самой генеральной совокупности.

### ***8. Осуществить пересинтез модели и измерить, как изменилась ее адекватность***

Под повторным синтезом (пересинтезом) модели понимается ее качественная модификация, осуществляемая путем включения в модель новых дополнительных классификационных и описательных шкал и градаций, представленных примерами в обучающей выборке.

Пересинтез модели обеспечивает возможность ее применения для идентификации объектов расширенной или новой генеральной совокупности.

Приведем пример синтеза новой модели, обобщающей предыдущую.

В модель добавлены новые классы распознавания (таблица 81).

**Таблица 81 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ**

№	Наименования классов распознавания
1	Загадки о животных
2	А.П.Чехов "Вишневый сад"
3	Ф.М.Достоевский "Преступление и наказание"
4	Н.В.Гоголь "Ревизор"
5	А.С.Грибоедов "Горе от ума"
6	И.А.Крылов
7	М.Ю.Лермонтов "Мцыри"
8	Фольклорные загадки о природе
9	Некрасов "Кому на Руси жить хорошо"
10	Пословицы



## Продолжение таблицы 81

№	Наименования классов распознавания
11	А.С.Пушкин "Евгений Онегин"
12	Загадки о саде и огороде
13	В.Шекспир
14	М.А.Шолохов "Тихий Дон"
15	Скороговорки
16	Л.Н.Толстой "Война и мир"
17	И.С.Тургенев "Отцы и дети"

Описательные шкалы и градации не приводятся, т.к. размерность составляет 6974 градации. Необходимо отметить, что текущая версия 12.5 системы "Эйдос" не имеет принципиальных ограничений на суммарное количество градаций классификационных и описательных шкал при синтезе модели и решении задач идентификации и прогнозирования, а также количество объектов обучающей выборки. Реально решались задачи с объемом обучающей выборки до 25000 объектов с 1500 классами и 7000 признаками. При этом был осуществлен синтез и исследование моделей, содержащих до 25 миллионов фактов.

В программном интерфейсе импорта данных из 17 исходных текстовых файлов, посвященных различным темам (см. таблицу 81) было сформировано 592 фрагмента, которые стали основой обучающей выборки.

После синтеза модели измеряется ее адекватность. Для этого обучающая выборка копируется в распознаваемую, после чего проводится распознавание и измерение валидности (рисунок 168). Продемонстрирована очень высокая внутренняя валидность новой модели.

## ИЗМЕРЕНИЕ АДЕКВАТНОСТИ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИОННОЙ МОДЕЛИ

Анкет физических: 592 логических (всего/факт): 592/ 592  
 Верная идентификация: 591 Ошибочная неидентификация: 1  
 Верная идентификация: 99.83% Ошибочная неидентификация: 0.17%  
 Минимальный уровень сходства: 0.0 Максимальное кол-во классов: 99999  
 09-05-04 11:25:19

г. Краснодар

N п/п	Код класса	Наименование класса	Всего логич анкет	ИДЕНТИФИЦИР		Неидентифиц		ИДЕНТИФИЦИРОВ		Неидентифицир	
				ВЕРНО	Ошиб.	ВЕРНО	Ошиб.	ВЕРНО%	Ошиб. %	ВЕРНО%	Ошиб. %
1	1	Загадки о животных	66	66	214	312	0	100.00	40.68	59.32	0.00
2	2	А.П.Чехов "Вишневый сад"	10	10	24	558	0	100.00	4.12	95.88	0.00
3	3	Ф.М.Достоевский "Преступление и наказание"	8	8	38	546	0	100.00	6.51	93.49	0.00
4	4	Н.В.Гоголь "Ревизор"	8	8	26	558	0	100.00	4.45	95.55	0.00
5	5	А.С.Грибоедов "Горе от ума"	42	42	33	517	0	100.00	6.00	94.00	0.00
6	6	И.А.Крылов	35	35	12	545	0	100.00	2.15	97.85	0.00
7	7	М.Ю.Лермонтов "Мцыри"	27	27	34	531	0	100.00	6.02	93.98	0.00
8	8	Фольклорные загадки о природе	31	31	263	298	0	100.00	46.88	53.12	0.00
9	9	Некрасов "Кому на Руси жить хорошо"	55	55	17	520	0	100.00	3.17	96.83	0.00
10	10	Пословицы	43	43	213	336	0	100.00	38.80	61.20	0.00
11	11	А.С.Пушкин "Евгений Онегин"	48	48	48	496	0	100.00	8.82	91.18	0.00
12	12	Загадки о саде и огороде	33	33	288	271	0	100.00	51.52	48.48	0.00
13	13	В.Шекспир	59	58	27	506	1	98.31	5.07	94.93	1.69
14	14	М.П.Шолохов "Тихий Дон"	7	7	34	551	0	100.00	5.81	94.19	0.00
15	15	Скороговорки	102	102	120	370	0	100.00	24.49	75.51	0.00
16	16	Л.Н.Толстой "Война и мир"	10	10	20	562	0	100.00	3.44	96.56	0.00
17	17	И.С.Тургенев "Отцы и дети"	8	8	61	523	0	100.00	10.45	89.55	0.00

Универсальная когнитивная аналитическая система

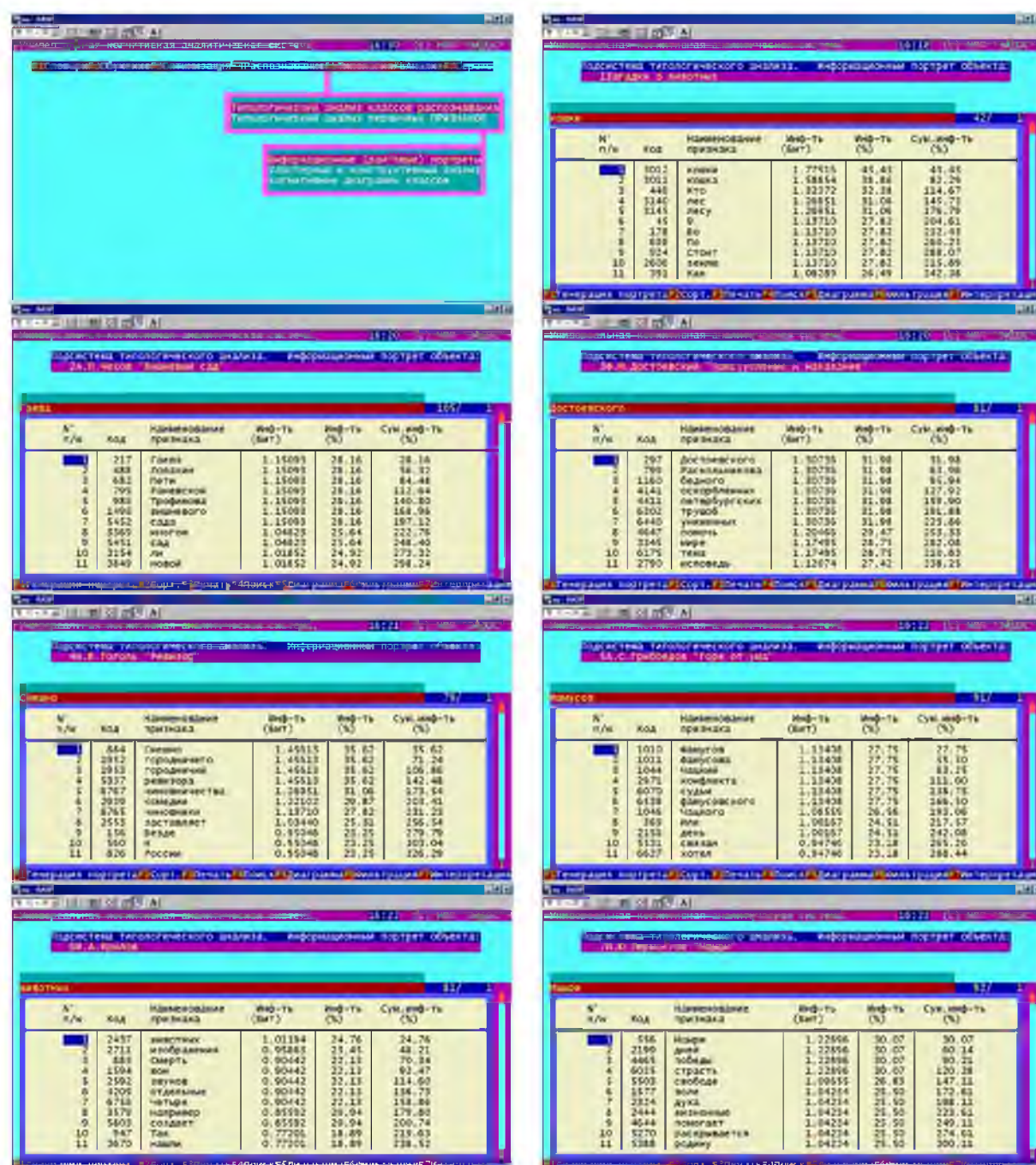
НПП «Эйдос»

**Рисунок 168. Выходная форма с результатами измерения внутренней валидности после пересинтеза модели**

## 9. Вывести информационные портреты текстов и дать их интерпретацию

Информационный портрет класса представляет собой список признаков в порядке убывания количества информации, содержащегося в этих признаках о принадлежности к данному классу.

Генерируются они 1-м режиме 5-й подсистемы "Типология" (рисунок 169). Информационные портреты классов отображаются системой "Эйдос" в виде экранных форм, круговых диаграмм и гистограмм, а также в распечатываемой форме таблиц в поддиректории ТХТ. Графические формы записываются в поддиректории РСХ.





Подсистема типологического анализа. Информационный портрет объекта: Беспокойные загадки в городе					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	102	Без	1.72610	42.30	42.30
2	3795	нету	1.62639	39.79	82.09
3	3354	миру	1.54248	37.74	119.83
4	45		1.41007	34.50	154.33
5	564	Над	1.41007	34.50	188.83
6	698	По	1.41007	34.50	223.33
7	2606	завыл	1.41007	34.50	257.83
8	4049	одним	1.41007	34.50	292.33
9	43		1.30736	31.98	324.31
10	44		1.30736	31.98	356.29
11	5490	свет	1.30736	31.98	388.27

Подсистема типологического анализа. Информационный портрет объекта: Снегидов "кому на Руси жить хорошо"					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	838	савелий	0.98474	24.08	24.08
2	2111	дед	0.95454	23.35	47.43
3	4765	поэмы	0.95454	23.35	70.78
4	287	Добролюбова	0.90032	22.03	92.81
5	643	от	0.90032	22.03	114.84
6	2481	забыть	0.90032	22.03	136.87
7	2456	какие	0.90032	22.03	158.90
8	3145	лесу	0.90032	22.03	180.93
9	3467	музыков	0.90032	22.03	202.96
10	3587	народного	0.90032	22.03	224.99
11	4554	поисках	0.90032	22.03	247.02

Подсистема типологического анализа. Информационный портрет объекта: Уроки Родины					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	4716	пословица	1.59669	39.06	39.06
2	4781	правду	1.59669	39.06	78.12
3	17		1.41007	34.50	112.62
4	34		1.41007	34.50	147.12
5	36		1.41007	34.50	181.62
6	739	правда	1.41007	34.50	216.12
7	96		1.38036	35.77	249.89
8	40		1.35912	32.76	282.65
9	41		1.35912	32.76	315.41
10	2045	да	1.32132	32.33	347.74
11	5571	семья	1.27766	31.26	379.00

Подсистема типологического анализа. Информационный портрет объекта: Цикл "Путь к Великой Победе"					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	313	Евгений	1.10169	29.15	29.15
2	628	Онегин	1.10169	29.15	58.30
3	956	Татьяны	1.10169	29.15	87.45
4	3647	нации	1.10169	29.15	116.60
5	6222	тогда	1.10169	29.15	145.75
6	777	Пушкин	1.02753	25.14	170.89
7	2097	дворянской	1.00507	24.59	195.48
8	3043	крепостного	1.00507	24.59	220.07
9	3063	критик	1.00507	24.59	244.66
10	3170	лица	1.00507	24.59	269.25
11	3383	могил	1.00507	24.59	293.84

Подсистема типологического анализа. Информационный портрет объекта: 128 загадки в саду и огороде					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	5987	стоит	1.66050	40.62	40.62
2	226	где	1.57658	38.57	79.19
3	4139	арбуз	1.57658	38.57	117.76
4	5453	саде	1.57658	38.57	156.33
5	2242	дом	1.50563	36.84	193.17
6	46		1.44417	35.33	228.50
7	43		1.34147	32.82	261.32
8	44		1.34147	32.82	294.14
9	4582	поле	1.25755	30.77	324.91
10	40		1.18660	29.03	353.94
11	41		1.18660	29.03	382.97

Подсистема типологического анализа. Информационный портрет объекта: Цикл "Путь к Великой Победе"					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	3226	любят	0.95726	23.42	23.42
2	3991	общественном	0.95726	23.42	46.84
3	4287	отца	0.95726	23.42	70.26
4	5904	восприимчивость	0.95726	23.42	93.68
5	1484	видеть	0.90305	22.09	115.77
6	1855	гибель	0.90305	22.09	137.86
7	2907	классицизма	0.90305	22.09	159.95
8	3351	мировой	0.90305	22.09	182.04
9	3590	народность	0.90305	22.09	204.13
10	3669	наших	0.90305	22.09	226.22
11	4002	обих	0.90305	22.09	248.31

Подсистема типологического анализа. Информационный портрет объекта: Цикл "Путь к Великой Победе"					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	296	дом	1.24782	30.53	30.53
2	520	Милослав	1.24782	30.53	61.06
3	1067	Волкова	1.24782	30.53	91.59
4	1206	Белый	1.24782	30.53	122.12
5	2441	экзотический	1.24782	30.53	152.65
6	2850	калечества	1.24782	30.53	183.18
7	2851	калечества	1.24782	30.53	213.71
8	507	Н	1.11541	27.29	241.00
9	2234	должен	1.11541	27.29	268.29
10	232	Григорий	1.09285	26.74	295.03
11	1316	бум	1.06123	25.96	320.99

Подсистема типологического анализа. Информационный портрет объекта: Цикл "Путь к Великой Победе"					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	2902	жилит	1.53164	37.47	37.47
2	2932	колокол	1.53164	37.47	74.94
3	2954	коллаж	1.53164	37.47	112.41
4	5846	сорос	1.53164	37.47	149.88
5	6261	траве	1.53164	37.47	187.35
6	6412	укуп	1.53164	37.47	224.82
7	2293	дрова	1.46070	35.74	260.56
8	2076	дес	1.34502	32.91	293.47
9	4968	про	1.34502	32.91	326.38
10	5575	сено	1.34502	32.91	359.29
11	6280	три	1.27407	31.17	390.46

Подсистема типологического анализа. Информационный портрет объекта: Цикл "Путь к Великой Победе"					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	781	Пьер	1.22896	30.07	30.07
2	782	Пьера	1.22896	30.07	60.14
3	973	Толстого	1.22896	30.07	90.21
4	974	Толстой	1.22896	30.07	120.28
5	1316	армя	1.22896	30.07	150.35
6	5821	судят	1.22896	30.07	180.42
7	1571	война	1.14504	28.01	208.43
8	5182	ради	1.09655	26.83	235.26
9	5478	смысл	1.09655	26.83	262.09
10	186	вина	1.04234	25.50	287.59
11	560	н	1.04234	25.50	313.09

Подсистема типологического анализа. Информационный портрет объекта: Цикл "Путь к Великой Победе"					
N п/к	Код	Наименование признака	Инф-ть (бит)	Инф-ть (%)	Сум. инф-ть (%)
1	90	Базаров	1.29440	31.67	31.67
2	91	Базарова	1.29440	31.67	63.34
3	993	Тургенев	1.29440	31.67	95.01
4	2175	дети	1.29440	31.67	126.68
5	384	Евгения	1.10777	27.10	153.78
6	1851	героя	1.10777	27.10	180.88
7	2558	заставляя	1.10777	27.10	207.98
8	4170	остается	1.10777	27.10	235.08
9	6965	всей	1.10777	27.10	262.18
10	273	действительн	0.97537	23.86	286.04
11	1470	взглядов	0.97537	23.86	309.90

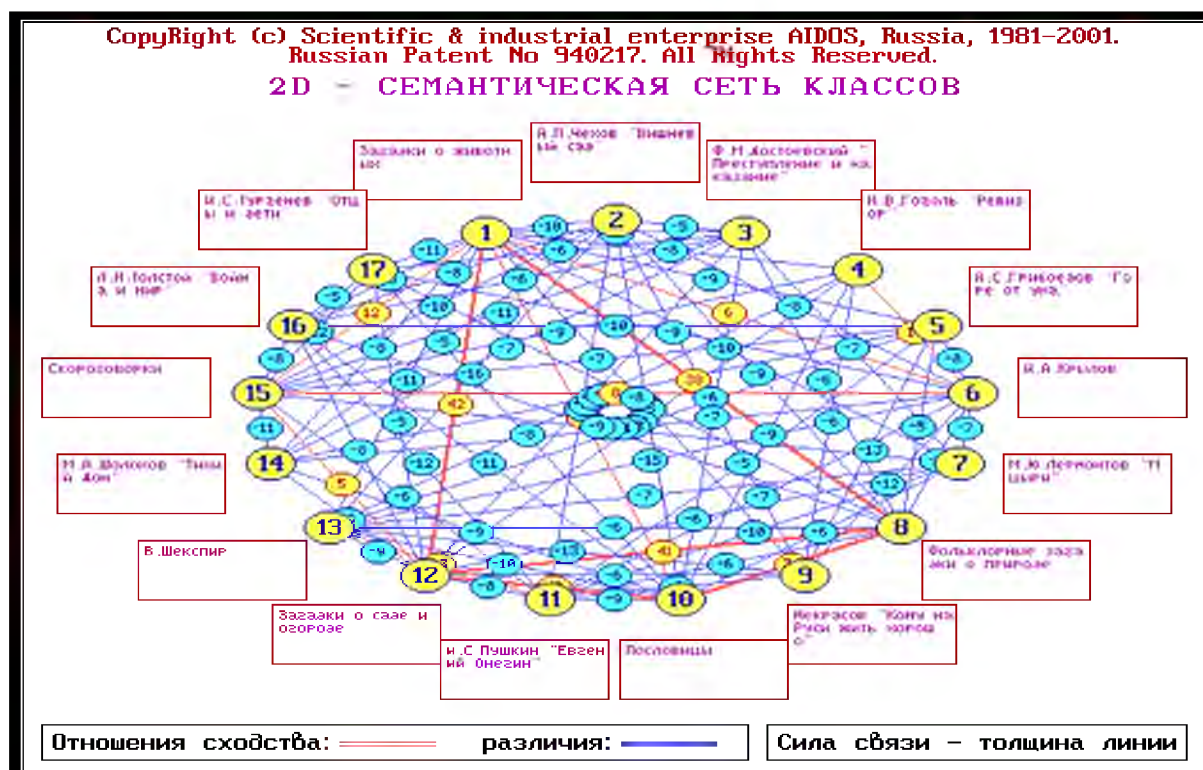
Рисунок 169. Информационные портреты классов

**10. Выполнить кластерно-конструктивный анализ модели**

Кластерно-конструктивный анализ классов и признаков реализуется в 5-й подсистеме "Типология". В результате рассчитываются матрицы сходства классов и признаков, на основе которых генерируется и выводится ряд текстовых и графических форм. В данной работе мы приведем для примера лишь матрицу сходства классов (таблица 82 и отображающую ее в графической форме семантическую сеть классов (рисунок 170)).

**Таблица 82 – МАТРИЦА СХОДСТВА КЛАССОВ**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	100,0	-9,55	-8,94	-8,16	-9,46	6,32	-10,08	38,11	-6,24	33,86	-4,34	41,96	-11,12	-8,17	11,70	-8,00	-11,05
2	-9,55	100,0	-5,11	-0,35	-2,47	-7,84	-2,95	-9,07	-5,50	-5,44	-6,47	1,20	-16,31	-5,43	-10,08	-8,26	2,04
3	-8,94	-5,11	100,0	-2,39	1,12	-3,97	-6,81	-5,61	-0,10	-6,56	-3,05	-10,13	-0,41	-7,36	-11,47	-6,04	-6,32
4	-8,16	-0,35	-2,39	100,0	2,52	9,73	1,33	-4,36	-6,23	-9,19	-4,94	-8,33	-9,55	-1,41	-9,14	-1,91	3,14
5	-9,46	-2,47	1,12	2,52	100,0	-8,31	-1,87	-5,24	-12,74	-1,25	-5,30	-4,55	-12,89	-8,85	-6,73	-9,59	-3,52
6	6,32	-7,84	-3,97	9,73	-8,31	100,0	-6,55	-5,05	-12,06	0,49	-7,34	-2,99	-15,19	-11,13	8,18	-3,10	-8,75
7	-10,08	-2,95	-6,81	1,33	-1,87	-6,55	100,0	-4,35	-1,04	-6,10	-10,14	-5,71	-7,09	-0,21	-9,40	-3,97	3,67
8	38,11	-9,07	-5,61	-4,36	-5,24	-5,05	-4,35	100,0	-2,38	34,04	-6,03	41,21	-6,48	-4,72	0,87	-8,50	-8,17
9	-6,24	-5,50	-0,10	-6,23	-12,74	-12,06	-1,04	-2,38	100,0	-1,85	-8,20	-6,28	-12,89	-1,18	-2,41	0,73	-3,53
10	33,86	-5,44	-6,56	-9,19	-1,25	0,49	-6,10	34,04	-1,85	100,0	-8,76	39,59	-9,83	-9,07	-1,63	-11,22	-7,73
11	-4,34	-6,47	-3,05	-4,94	-5,30	-7,34	-10,14	-6,03	-8,20	-8,76	100,0	-7,79	13,47	-3,96	-5,98	-11,77	-2,47
12	41,96	1,20	-10,13	-8,33	-4,55	-2,99	-5,71	41,21	-6,28	39,59	-7,79	100,0	-8,80	-8,13	5,09	-8,29	-5,24
13	-11,12	-16,31	-0,41	-9,55	-12,89	-15,19	-7,09	-6,48	-12,89	-9,83	13,47	-8,80	100,0	-3,67	-3,20	-1,92	1,77
14	-8,17	-5,43	-7,36	-1,41	-8,85	-11,13	-0,21	-4,72	-1,18	-9,07	-3,96	-8,13	-3,67	100,0	-11,07	-0,69	-3,25
15	11,70	-10,08	-11,47	-9,14	-6,73	8,18	-9,40	0,87	-2,41	-1,63	-5,98	5,09	-3,20	-11,07	100,0	-8,44	-12,23
16	-8,00	-8,26	-6,04	-1,91	-9,59	-3,10	-3,97	-8,50	0,73	-11,22	-11,77	-8,29	-1,92	-0,69	-8,44	100,0	-5,50
17	-11,05	2,04	-6,32	3,14	-3,52	-8,75	3,67	-8,17	-3,53	-7,73	-2,47	-5,24	1,77	-3,25	-12,23	-5,50	100,0



**Рисунок 170. Отображение матрицы сходства классов в графической форме семантической сети классов (отображены связи значимостью более 5%)**

## **Выводы**

Продемонстрирована возможность и эффективность применения технологии и инструментария системно-когнитивного анализа для решения ряда задач атрибуции текстов.

Приведен подробный численный пример (с большим количеством конкретных иллюстративных материалов) реализации всех этапов СК-анализа при атрибуции текстов: когнитивной структуризации и формализации предметной области; формирования обучающей выборки; синтеза семантической информационной модели; оптимизации и измерения адекватности модели; адаптации и пересинтеза модели; типологического и кластерно-конструктивного анализа модели.

## **Контрольные вопросы**

1. Что такое атрибуция текстов?
2. Каким образом выполняется когнитивная структуризация предметной области?
3. В чем состоит формализацию предметной области и как ее осуществить в системе "Эйдос"?
4. Какие средства формирования обучающей выборки используются в системе "Эйдос" при решении задач атрибуции текстов?
5. В какой подсистеме и в каком режиме системы "Эйдос" осуществляется синтез семантической информационной модели и в чем он состоит?
6. В чем заключается оптимизация семантической информационной модели и как она осуществляется в системе "Эйдос"?
7. Как семантическая информационная модель проверяется на адекватность?
8. Как связана адекватность модели с внутренней и внешней, дифференциальной и интегральной валидностью?
9. Каким образом можно в системе "Эйдос" выполнить адаптацию модели и измерить, как изменилась ее адекватность?
10. Как в системе "Эйдос" осуществить пересинтез модели и измерить, как изменилась ее адекватность?
11. Чем отличается адаптация модели от пересинтеза в системе "Эйдос"?
12. В каких подсистемах и режимах системы "Эйдос" можно вывести информационные портреты?
13. В каких подсистемах и режимах системы "Эйдос" можно выполнить кластерно-конструктивный анализ модели?

## **Литература по лабораторной работе**

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

## ЛР-7: "Идентификация и классификация натуральных чисел по их свойствам"

### ***Краткая теория***

Натуральные числа имеют ряд свойств, такие, например, как: делители, неделители, количество делителей и др. Существуют и более сложные свойства натуральных чисел, изучаемые в теории чисел. По этим свойствам числа классифицируются как простые, четные и т.д. Таким образом, исследование свойств чисел и является типичной задачей обучения распознаванию образов, идентификации и автоматической классификации. Эта задача довольно проста и наглядна, поэтому эта задача представляет для нас интерес в качестве учебной.

### ***Задание***

1. Формализовать задачу.
  - 1.1. Сконструировать классификационные шкалы и градации.
  - 1.2. Сконструировать описательные шкалы и градации.
  - 1.3. Сгенерировать обучающую выборку.
2. Осуществить синтез и верификацию (измерение адекватности) семантической информационной модели.
3. Провести системно-когнитивный анализ модели:
  - 3.1. Решить задачи идентификации и прогнозирования.
  - 3.2. Сгенерировать информационные портреты классов и факторов, т.е. решить обратную задачу прогнозирования (результаты отобразить в графической форме двухмерных и трехмерных профилей классов и факторов).
  - 3.3. Провести кластерно-конструктивный анализ классов и факторов (результаты отобразить в форме семантических сетей классов и факторов).
  - 3.4. Осуществить содержательное сравнение классов и факторов (результаты отобразить в форме когнитивных диаграмм классов и факторов).
  - 3.5. Построить нелокальные нейроны и интерпретируемые нейронные сети.
  - 3.6. Построить классические когнитивные модели (отобразить в форме когнитивных карт).
  - 3.7. Построить интегральные когнитивные модели (отобразить в форме интегральных когнитивных карт).

### ***Пример решения***

#### ***1. Формализовать задачу.***

Для этой цели используем 5-ю функцию 5-го режима 1-й подсистемы системы "Эйдос" (реальный исходный текст программы приводится ниже).



```

*****
*** Формирование модели для исследования свойств чисел *****
*** Луценко Е.В., 11/06/05 09:11am *****
*****

FUNCTION Div_chis()

scr23 = SAVESCREEN(0,0,24,79)

SHOWTIME(0,58,.T.,"rb/n")

Titul(.T.)

Mess = "=== ГЕНЕРАЦИЯ ИСХОДНЫХ ДАННЫХ ДЛЯ МОДЕЛИ ИССЛЕДОВАНИЯ СВОЙСТВ ЧИСЕЛ
=== "
@5,40-LEN(Mess)/2 SAY Mess COLOR "rg+/rb"
*
01234567890123456789012345678901234567890123456789012345678901234
5678
*      0          10          20          30          40          50          60
70
@10,24 SAY                      "Задайте максимальное число: #####" COLOR
"w+/rb"
N_Obj = 99
@10,52 GET N_Obj PICTURE "#####" COLOR "rg+/r"
SET CURSOR ON;READ;SET CURSOR OFF
IF LASTKEY()=27
    RESTSCREEN(0,0,24,79,scr23)
    SHOWTIME()
    RETURN
ENDIF

*** Формирование справочника первичных признаков

A_Pr := {}
FOR j=1 TO N_Obj
    AADD(A_Pr, "Делится на:" + ALLTRIM(STR(j,4)))          && 1*N_Obj
NEXT
FOR j=1 TO N_Obj
    AADD(A_Pr, "Не дел.на : " + ALLTRIM(STR(j,4)))          && 2*N_Obj
NEXT
FOR j=1 TO N_Obj
    AADD(A_Pr, "Делителей : " + ALLTRIM(STR(j,4)))          && N_Del+2*N_Obj
NEXT
FOR j=1 TO N_Obj
    AADD(A_Pr, "Степ.числа: " + ALLTRIM(STR(j,4)))          && 4*N_Obj
NEXT
FOR j=1 TO N_Obj
    AADD(A_Pr, "Чис.в степ: " + ALLTRIM(STR(j,4)))          && 5*N_Obj
NEXT

USE Priz_per EXCLUSIVE NEW
ZAP
FOR j=1 TO LEN(A_Pr)
    APPEND BLANK
    REPLACE Kod WITH j
    REPLACE Name WITH A_Pr[j]
NEXT
GenNtxPrp(.F.)
CLOSE ALL

```

```

*** Формирование справочника классов распознавания
USE Object EXCLUSIVE NEW
ZAP
FOR j=1 TO N_Obj
  APPEND BLANK
  REPLACE Kod WITH j
  REPLACE Name WITH ALLTRIM(STR(j,3))
NEXT
GenNtxObj(.F.)
CLOSE ALL

*** Формирование обучающей выборки

USE ObInfZag EXCLUSIVE NEW;ZAP
USE ObInfKpr EXCLUSIVE NEW;ZAP
@24,0 SAY REPLICATE("■",80) COLOR "rb/n"

FOR s=1 TO N_Obj

  ***** Подготовка записей
  SELECT ObInfZag
  APPEND BLANK
  REPLACE Kod_ist WITH s
  REPLACE Name_ist WITH ALLTRIM(STR(s,4))
  REPLACE Obj_1 WITH s
  SELECT ObInfKpr
  APPEND BLANK
  REPLACE Kod_ist WITH s
  p=0                                && Позиция для записи в БД
  N_Del=0                            && Кол-во делителей

  *** Проверка делимости
  FOR j=1 TO N_Obj
    IF s-j*INT(s/j) = 0
      Kod = j
      ++N_Del
      IF p+1 <= 11
        FIELDPUT(++p+1, Kod)
      ELSE
        APPEND BLANK
        REPLACE Kod_ist WITH s
        p=0
        FIELDPUT(++p+1, Kod)
      ENDIF
    ENDIF
  NEXT
  ***** Занесение количества делителей
  Kod = N_Del+2*N_Obj
  IF p+1 <= 11
    FIELDPUT(++p+1, Kod)
  ELSE
    APPEND BLANK
    REPLACE Kod_ist WITH s
    p=0
    FIELDPUT(++p+1, Kod)
  ENDIF

  *** Проверка не делимости
  FOR j=1 TO N_Obj
    IF s-j*INT(s/j) <> 0
      Kod = j+1*N_Obj

```



```

    IF p+1 <= 11
        FIELDPUT(++p+1, Kod)
    ELSE
        APPEND BLANK
        REPLACE Kod_ist WITH s
        p=0
        FIELDPUT(++p+1, Kod)
    ENDIF
ENDIF
NEXT

*** Проверка, является ли данное число s
*** целой степенью j некоторого натурального числа n
FOR j=1 TO N_Obj
    FOR n=1 TO N_Obj
        IF s=n^j
            *** Степень числа n
            Kod = n+3*N_Obj
            IF p+1 <= 11
                FIELDPUT(++p+1, Kod)
            ELSE
                APPEND BLANK
                REPLACE Kod_ist WITH s
                p=0
                FIELDPUT(++p+1, Kod)
            ENDIF
            *** Число в степени j
            Kod = j+4*N_Obj
            IF p+1 <= 11
                FIELDPUT(++p+1, Kod)
            ELSE
                APPEND BLANK
                REPLACE Kod_ist WITH s
                p=0
                FIELDPUT(++p+1, Kod)
            ENDIF
        ENDIF
    NEXT
NEXT

p=s/N_Obj*100;p=IF(p<=100,p,100)
@24,0 SAY STR(p,3)+"%" COLOR "rg+/r+"
@24,4 SAY REPLICATE("█",p*0.76) COLOR "rg+/g"

NEXT
GenNtxOin(.F.)

@24,0 SAY REPLICATE("█",80) COLOR "rb/n"
Mess = " ПРОЦЕСС ГЕНЕРАЦИИ ШКАЛ И ОБУЧАЮЩЕЙ ВЫБОРКИ ЗАВЕРШЕН УСПЕШНО !!! "
@24,40-LEN(Mess)/2 SAY Mess COLOR "rg+/rb"
nj=INKEY(0)

RESTSCREEN(0,0,24,79,scr23)
SHOWTIME()

RETURN(nj)

```

### 1.1. Сконструировать классификационные шкалы и градации.

Выберем в качестве классов натуральные числа от 1 до 30 (вместо 30 может быть взято другое число). В результате получим 30 классов с кодами от 1 до

30, наименования которых совпадают с их кодом. Приводить здесь эту элементарную таблицу не имеет смысла.

### 1.2. Сконструировать описательные шкалы и градации.

В качестве описательных шкал и градаций используем простейшие свойства натуральных чисел, такие как: делители, неделители, количество делителей (таблица 83). Могут быть использованы и более сложные свойства, например: какой степенью и степенью какого числа является данное число, а также другие свойства, изучаемые в теории чисел, но суть задачи от этого не изменится.

**Таблица 83 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ (ФРАГМЕНТ)**

Код	Наименование признака	Код	Наименование признака	Код	Наименование признака
1	Делится на 1	31	Не делится на 1	61	Делителей: 1
2	Делится на 2	32	Не делится на 2	62	Делителей: 2
3	Делится на 3	33	Не делится на 3	63	Делителей: 3
4	Делится на 4	34	Не делится на 4	64	Делителей: 4
5	Делится на 5	35	Не делится на 5	65	Делителей: 5
6	Делится на 6	36	Не делится на 6	66	Делителей: 6
7	Делится на 7	37	Не делится на 7	67	Делителей: 7
8	Делится на 8	38	Не делится на 8	68	Делителей: 8
9	Делится на 9	39	Не делится на 9	69	Делителей: 9
10	Делится на 10	40	Не делится на 10	70	Делителей: 10
---	---	---	---	---	---
28	Делится на 28	58	Не делится на 28	88	Делителей: 28
29	Делится на 29	59	Не делится на 29	89	Делителей: 29
30	Делится на 30	60	Не делится на 30	90	Делителей: 30

### 1.3. Сгенерируем обучающую выборку.

Обучающая выборка генерируется автоматически вместе с шкалами и градациями и здесь не приводится из-за ее большого объема.

## 2. Осуществить синтез и верификацию (измерение адекватности) семантической информационной модели.

Синтез модели осуществляется в 5-й функции 3-го режима 2-й подсистемы системы "Эйдос", а верификация (после копирования обучающей выборки в распознаваемую и распознавания со 2-м интегральным критерием сходства) – во 2-м режиме 6-й подсистемы. При этом получен результат, представленный на рисунке 171, который говорит о высокой степени адекватности модели и корректности выводов, полученных путем ее исследования.

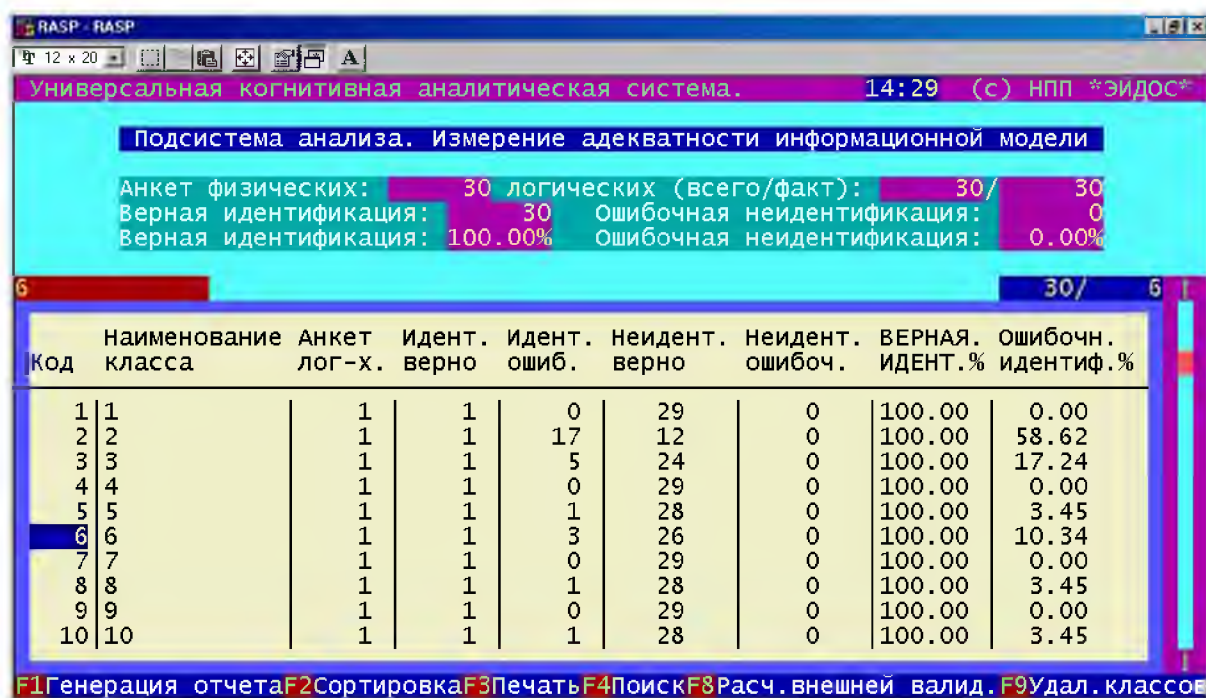


Рисунок 171. Экранная форма режима измерения адекватности СИМ

### 3. Провести системно-когнитивный анализ модели:

#### 3.1. Решить задачи идентификации и прогнозирования.

Идентификация проводится во 2-м режиме 4-й подсистемы системы "Эйдос". Результаты идентификации выводятся в форме карточек в 1-й и 2-й функциях 3-го режима 4-й подсистемы системы "Эйдос" (примеры карточек на рисунке 172).

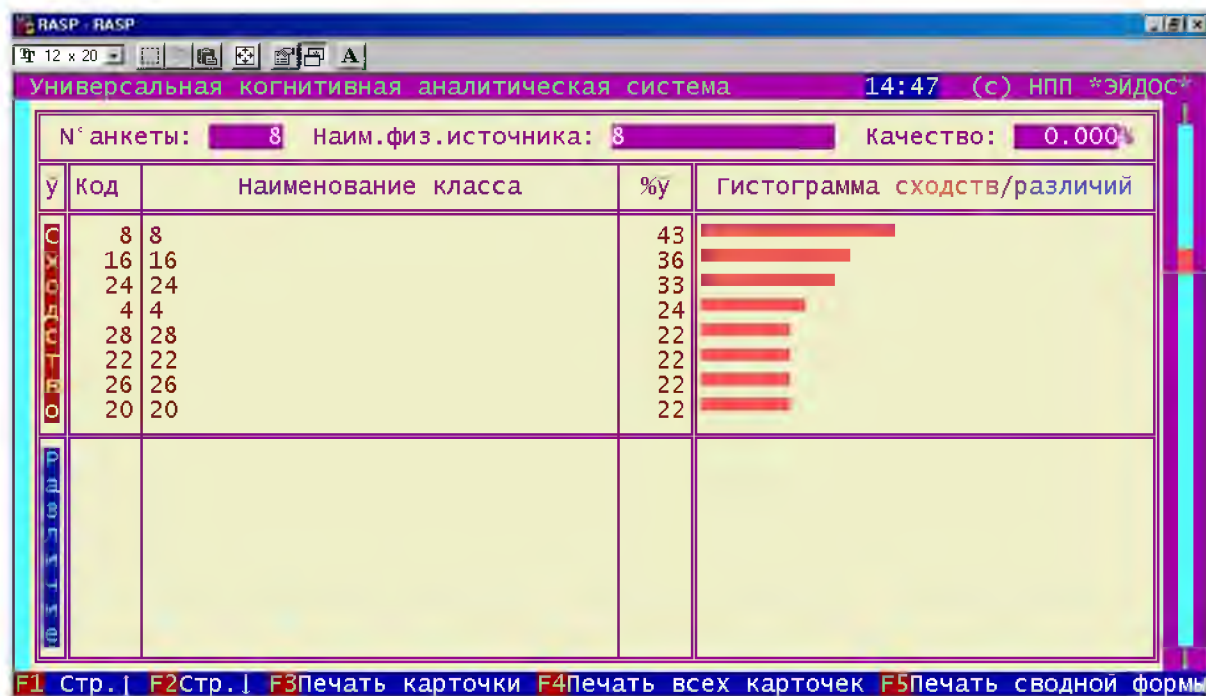


Рисунок 172. Пример карточки идентификации объекта с классами

3.2. Сгенерировать информационные портреты классов и факторов, т.е. решить обратную задачу прогнозирования (результаты отобразить в графической форме двухмерных и трехмерных профилей классов и факторов).

Информационные портреты классов генерируются и отображаются в 1-й функции 1-го режима 5-й подсистемы системы "Эйдос" (рисунок 173).

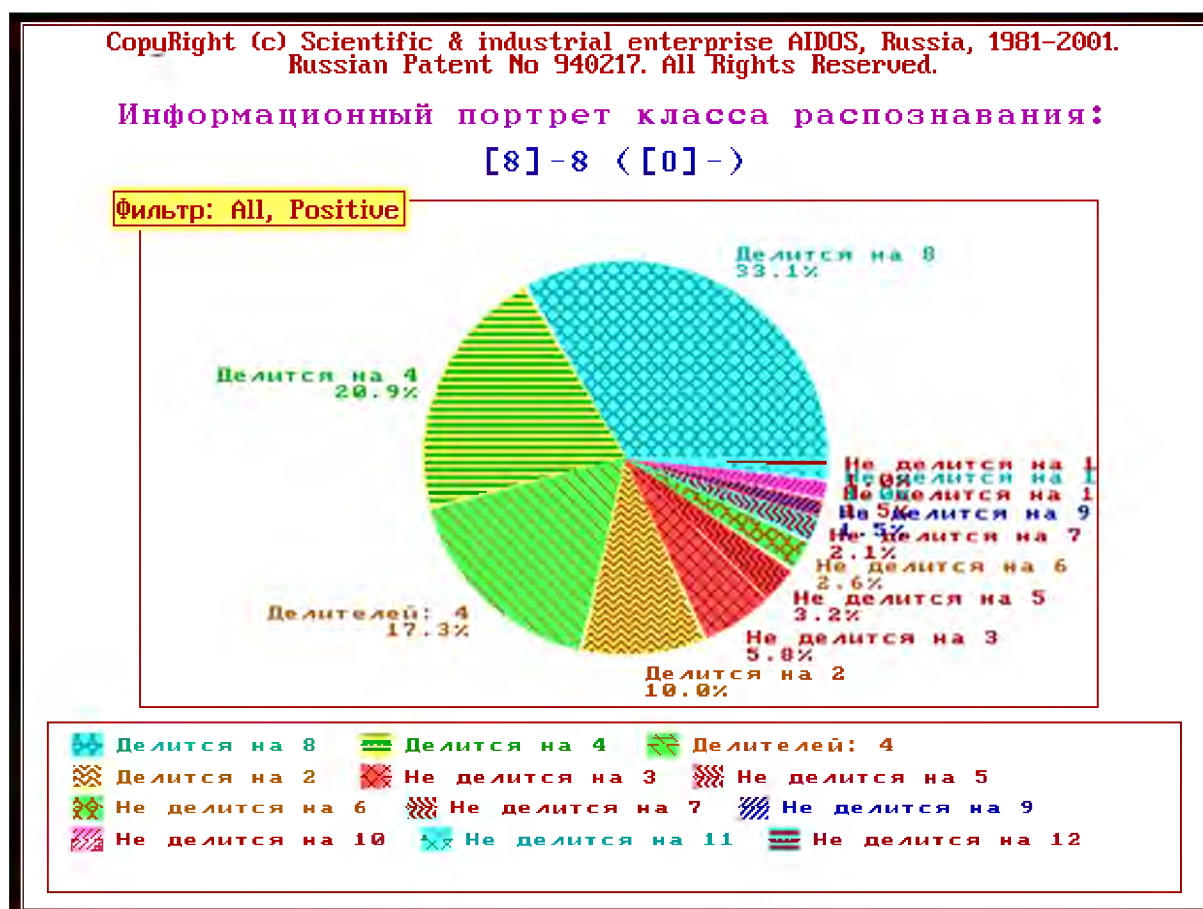


Рисунок 173. Пример информационного портрета класса в форме круговой диаграммы

Двухмерные и трехмерные профили классов и признаков генерируются и отображаются в 4-м режиме 6-й подсистемы системы "Эйдос" (рисунок 174).

Информационные портреты признаков (факторов) генерируются и отображаются в 1-й функции 2-го режима 5-й подсистемы системы "Эйдос" (рисунок 175). Размеры секторов в круговой диаграмме соответствуют относительному вкладу признаков в общее количество информации, содержащейся в информационном портрете.

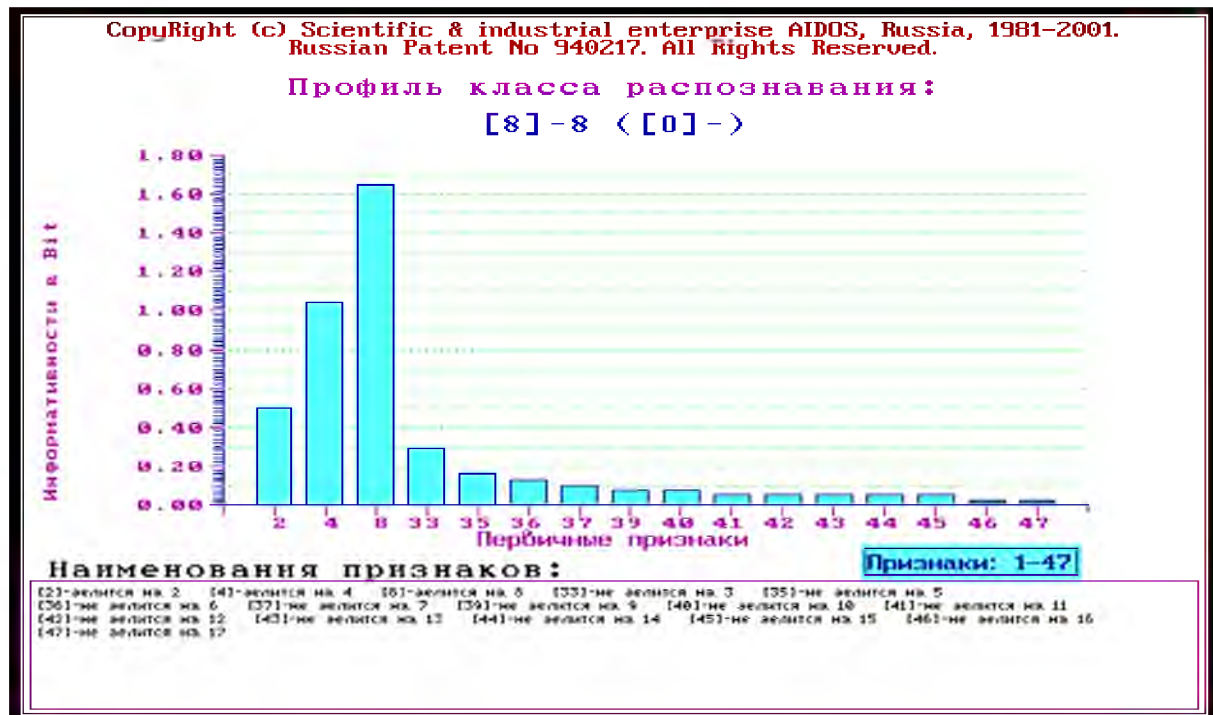


Рисунок 174. Пример профиля класса

НАСП - НАСП

Универсальная когнитивная аналитическая система. 15:26 (с) НПП "Эйдос"

Типологический анализ. Информационный (ранговый) портрет признака:  
62делителей: 2

N° n/w	Код	Наименование класса	Инф-ть (Бит)	Инф-ть (%)	Сум.инф-ть (%)
1	2	2	0.78868	16.07	16.07
2	3	3	0.78868	16.07	32.14
3	5	5	0.78868	16.07	48.21
4	7	7	0.78868	16.07	64.28
5	11	11	0.78868	16.07	80.35
6	13	13	0.78868	16.07	96.42
7	17	17	0.78868	16.07	112.49
8	19	19	0.78868	16.07	128.56
9	23	23	0.78868	16.07	144.63
10	29	29	0.78868	16.07	160.70

F1 Генерация портрета F2 Сортировка F3 Печать F4 Поиск F5 диаграмма F6 Фильтрация

Рисунок 175. Экранная форма информационного портрета фактора в форме таблицы

Из рисунка 148 видно, что система способна выявить простые числа по признаку: "Число делителей 2".

**3.3. Провести кластерно-конструктивный анализ классов и факторов (результаты отобразить в форме семантических сетей классов и факторов).**

Этот анализ проводится в во 2-й функции 1-го и 2-го режимов 5-й подсистемы системы "Эйдос" (рисунок 176).



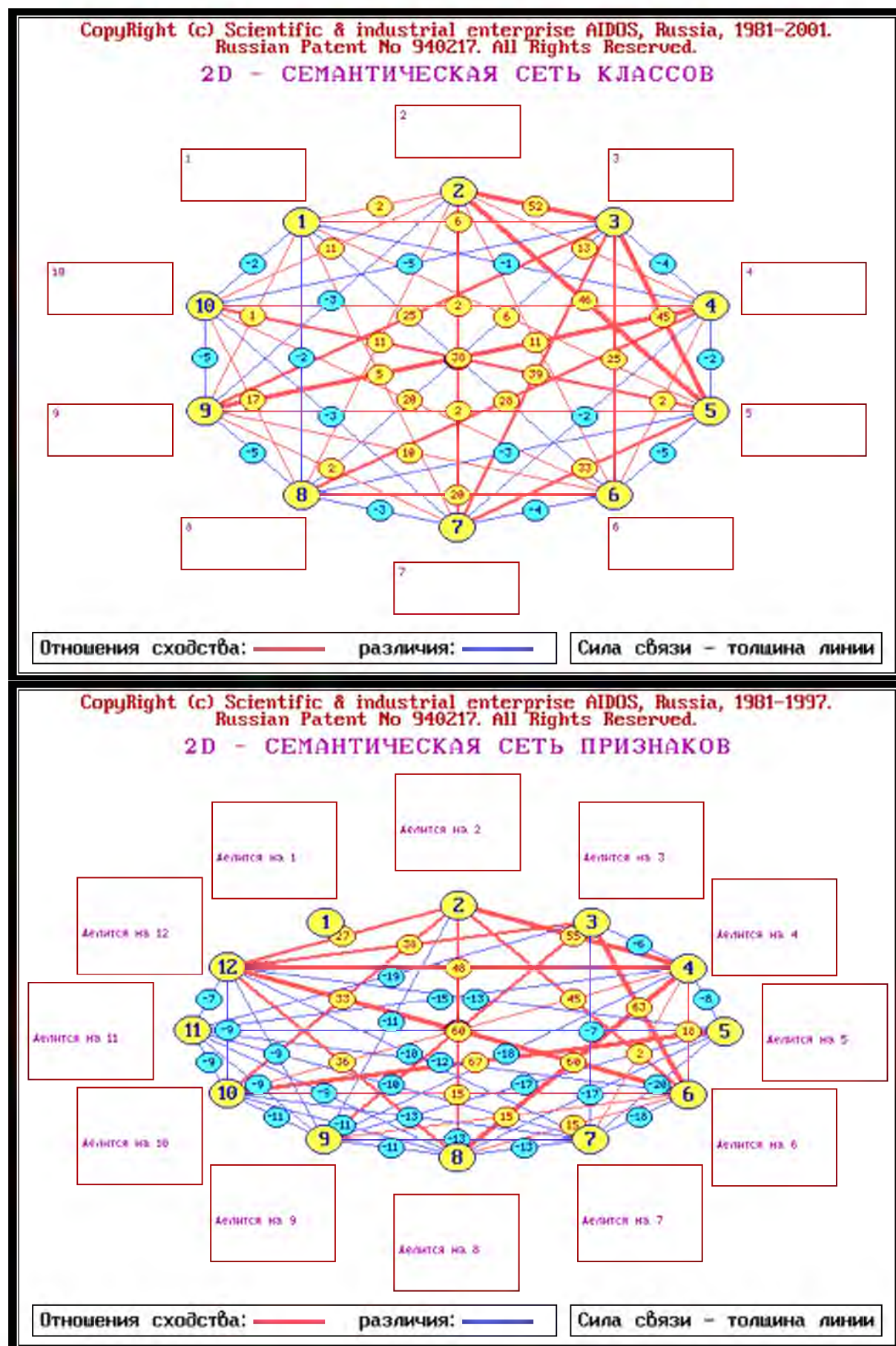
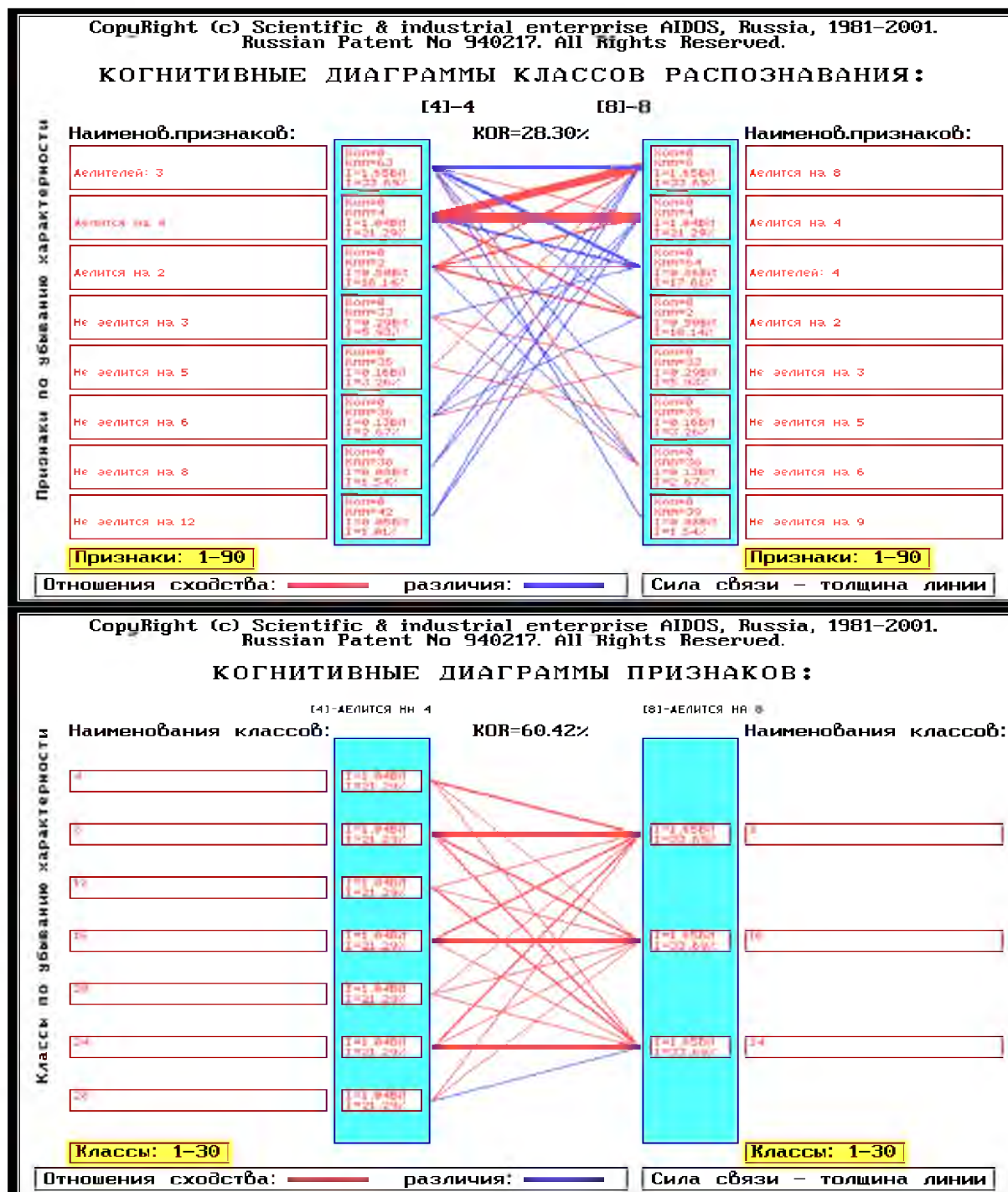


Рисунок 176. Примеры семантических сети классов и признаков

Семантические сети отражают сходство классов по характеризующим их признакам и сходство признаков по тем классам, о принадлежности к которым они несут информацию.

**3.4. Осуществить содержательное сравнение классов и факторов (результаты отобразить в форме когнитивных диаграмм классов и факторов).**

Структуру любой линии в семантических сетях классов или признаков можно детально увидеть в когнитивных диаграммах классов и признаков (рисунок 177).



**Рисунок 177. Примеры когнитивных диаграмм классов и признаков**



### 3.5. Построить нелокальные нейроны и интерпретируемые нейронные сети.

Нелокальные нейроны отражают систему детерминации состояний факторами и отображаются в 5-м режиме 6-й подсистемы системы "Эйдос", а нейронные сети представляют собой совокупность нейронов и отображаются в 6-м режиме той же подсистемы (рисунок 178).

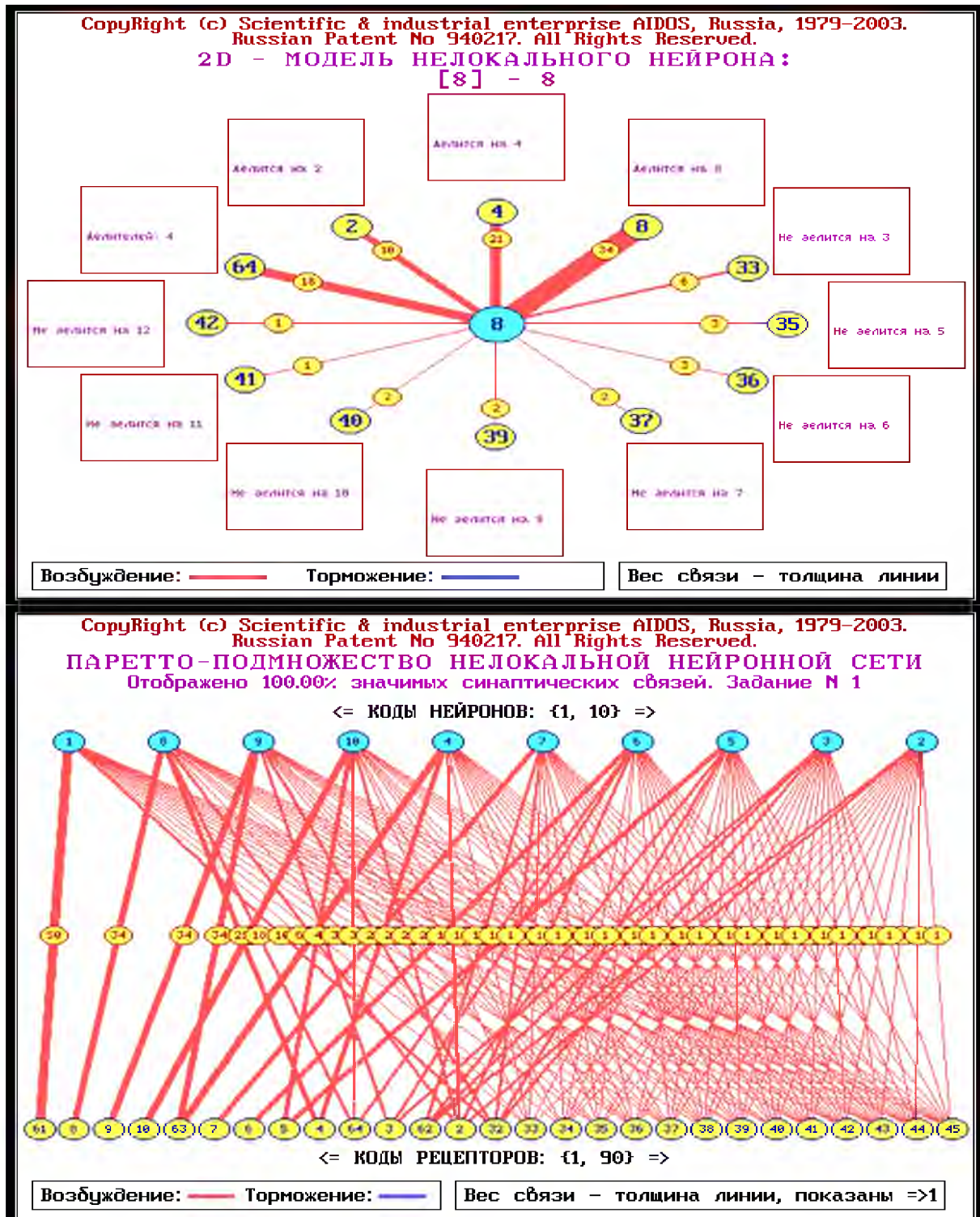


Рисунок 178. Примеры нелокального нейрона и Паретто-подмножества нейронной сети



### 3.6. Построить классические когнитивные модели (отобразить в форме когнитивных карт).

Классическая когнитивная модель отражает структуру детерминации состояния с указанием сходств и различий между факторами и отображается в 7-м режиме 6-й подсистемы системы "Эйдос" (рисунок 179).

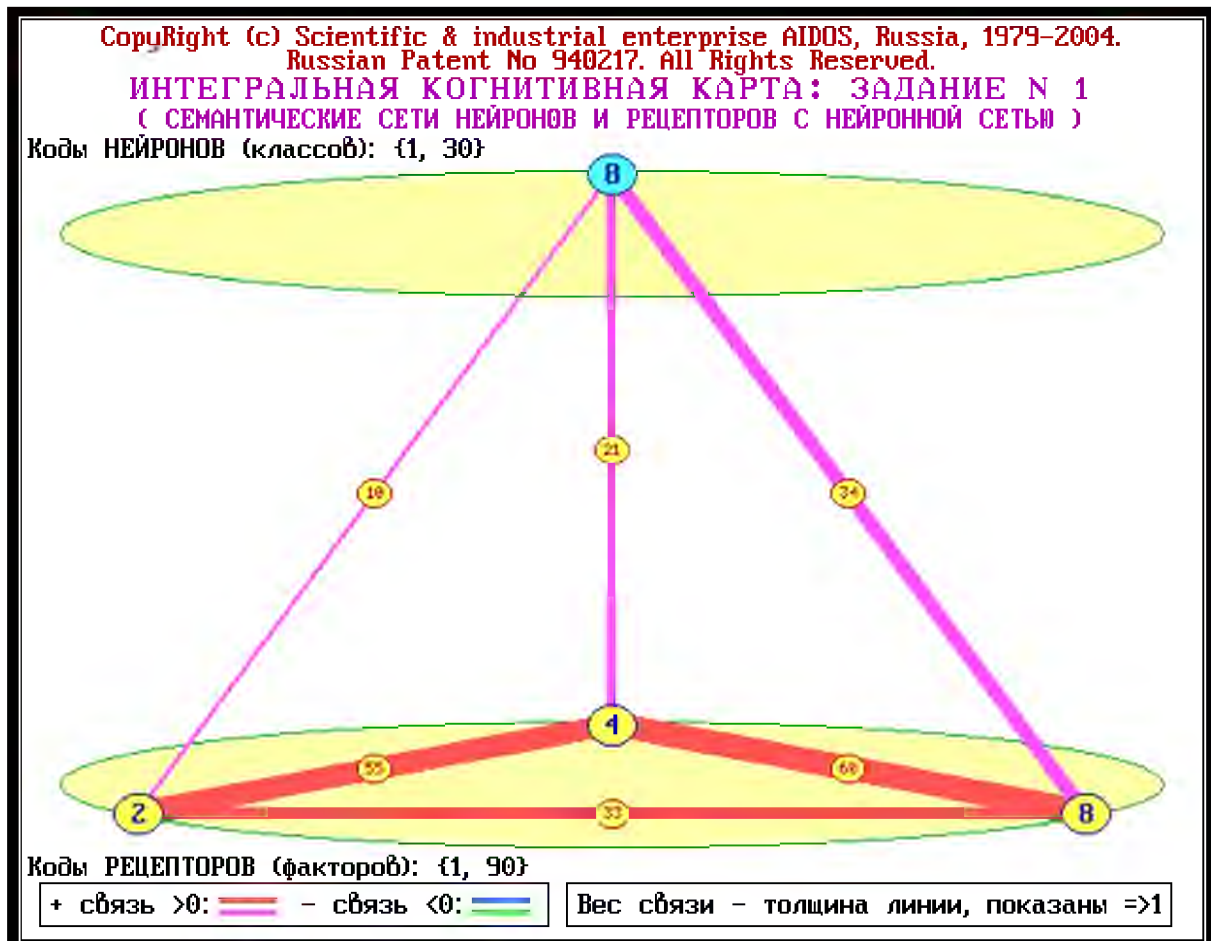


Рисунок 179. Пример классической когнитивной карты

### 3.7. Построить интегральные когнитивные модели (отобразить в форме интегральных когнитивных карт).

Интегральная когнитивная модель отражает структуру детерминации нескольких состояний с указанием сходств и различий между состояниями по их системам детерминации, между факторами, по влиянию, которое они оказывают на поведение объекта управления, отображаются в 7-м режиме 6-й подсистемы системы "Эйдос" (рисунок 180).

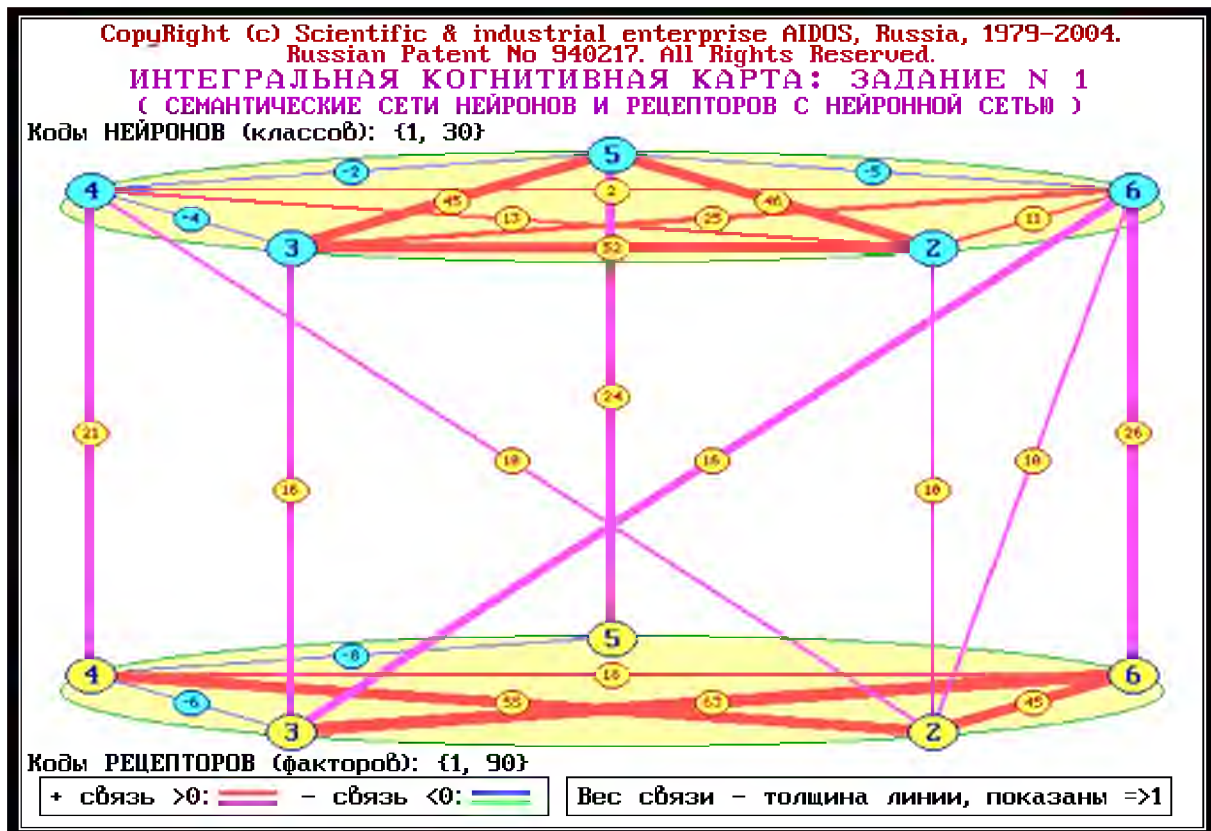


Рисунок 180. Пример интегральной когнитивной карты

### Контрольные вопросы

1. Какие свойства натуральных чисел мы рассматривали в качестве их признаков?
2. Каким образом выполняется когнитивная структуризация предметной области?
3. В чем состоит формализацию предметной области и как ее осуществить в системе "Эйдос" при изучении свойств натуральных чисел?
4. Какие средства формирования обучающей выборки используются в системе "Эйдос" при изучении свойств натуральных чисел?

### Литература по лабораторной работе

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

## ЛР-8: "Идентификация трехмерных тел по их ортогональным проекциям"

### ***Краткая теория***

Когда мы зрительно воспринимаем окружающую действительность, то постоянно решается задача идентификации трехмерных тел по их проекциям. При этом мы можем наблюдать одну, две или даже три проекции тела.

Одну проекцию мы наблюдаем при наблюдении тела на достаточно большом расстоянии, при котором бинокулярность зрения несущественна, или при монокулярном наблюдении за очень короткое время, за которое точка зрения на тело не меняется или оно не успевает повернуться.

Ясно, что наиболее сложными условиями для идентификации формы тела являются те, когда мы видим лишь одну его проекцию (большое расстояние и малое время наблюдения).

### ***Задание***

1. Формализовать задачу.
  - 1.1. Сконструировать классификационные шкалы и градации.
  - 1.2. Сконструировать описательные шкалы и градации.
  - 1.3. Сгенерировать обучающую выборку.
2. Осуществить синтез и верификацию (измерение адекватности) семантической информационной модели.
3. Провести системно-когнитивный анализ модели:
  - 3.1. Решить задачи идентификации и прогнозирования.
  - 3.2. Сгенерировать информационные портреты классов и факторов, т.е. решить обратную задачу прогнозирования (результаты отобразить в графической форме двухмерных и трехмерных профилей классов и факторов).
  - 3.3. Провести кластерно-конструктивный анализ классов и факторов (результаты отобразить в форме семантических сетей классов и факторов).
  - 3.4. Осуществить содержательное сравнение классов и факторов (результаты отобразить в форме когнитивных диаграмм классов и факторов).
  - 3.5. Построить нелокальные нейроны и интерпретируемые нейронные сети.
  - 3.6. Построить классические когнитивные модели (отобразить в форме когнитивных карт).
  - 3.7. Построить интегральные когнитивные модели (отобразить в форме интегральных когнитивных карт).

## Пример решения

### 1. Формализовать задачу.

#### 1.1. Сконструировать классификационные шкалы и градации.

Так как задачей является идентификация формы тела, то в качестве классов выберем для исследования простые и широко известные формы: шар, тетраэдр, куб, конус, пирамиду, призму и цилиндр, а также их наблюдения *по две* ортогональные проекции, которые будем обозначать числами 11, 12, 23, и одно наблюдение сразу трех проекций: 123 (таблица 84). Одновременно или последовательно наблюдаются проекции, в данном случае неважно.

**Таблица 84 – КЛАССИФИКАЦИОННЫЕ ШКАЛЫ И ГРАДАЦИИ**

Код	Наименование	Код	Наименование	Код	Наименование	Код	Наименование
1	Шар	11	Куб	21	Пирамида	31	Призма
2	Шар-11	12	Куб-11	22	Пирамида-11	32	Призма-11
3	Шар-12	13	Куб-12	23	Пирамида-12	33	Призма-12
4	Шар-22	14	Куб-23	24	Пирамида-23	34	Призма-23
5	Шар-123	15	Куб-123	25	Пирамида-123	35	Призма-123
6	Тетраэдр	16	Конус	26	Цилиндр		
7	Тетраэдр-11	17	Конус-11	27	Цилиндр-11		
8	Тетраэдр-12	18	Конус-12	28	Цилиндр-12		
9	Тетраэдр-23	19	Конус-23	29	Цилиндр-23		
10	Тетраэдр-123	20	Конус-123	30	Цилиндр-123		

#### 1.2. Сконструировать описательные шкалы и градации.

Проекциями перечисленных трехмерных тел на взаимно-ортогональные плоскости являются двухмерные фигуры: круг, квадрат и треугольник. Соответственно, сконструируем и описательные шкалы и градации, чтобы они позволяли отразить все варианты проекций трехмерных тел (таблица 85).

**Таблица 85 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ**

Код	Наименование
[1]	ВИД СПЕРЕДИ:
1	Круг
2	Квадрат
3	Треугольник
[2]	ВИД СБОКУ:
4	Круг
5	Квадрат
6	Треугольник
[3]	ВИД СВЕРХУ:
7	Круг
8	Квадрат
9	Треугольник

### 1.3. Сгенерировать обучающую выборку.

Чтобы сгенерировать обучающую выборку составим таблицу 86, в которой в наглядной форме изобразим проекции выбранных нами тел на ортогональные плоскости.

**Таблица 86 – КОДИРОВАНИЕ ПРОЕКЦИЙ ТЕЛ  
ДЛЯ ОБУЧАЮЩЕЙ ВЫБОРКИ**

Код	Тело	Проекция	1-я проекция X↔Y вращение вокруг Z 1			2-я проекция X↔Z вращение вокруг Y 2			3-я проекция Y↔Z вращение вокруг X 3		
			X	Y	Z	X	Y	Z	X	Y	Z
1	Шар	Вид	○	○	○	○	○	○	○	○	○
		Код	1	4	7	1	4	7	1	4	7
2	Тетраэдр	Вид	△	△	△	△	△	△	△	△	△
		Код	3	6	9	3	6	9	3	6	9
3	Куб	Вид	□	□	□	□	□	□	□	□	□
		Код	2	5	8	2	5	8	2	5	8
4	Конус	Вид	△	△	○	○	△	△	△	○	△
		Код	3	6	7	1	6	9	3	4	9
5	Пирамида	Вид	△	△	□	□	△	△	△	□	△
		Код	3	6	8	2	6	9	3	5	9
6	Цилиндр	Вид	□	□	○	○	□	□	□	○	□
		Код	2	5	7	1	5	8	2	4	8
7	Призма	Вид	□	□	△	△	□	□	□	△	□
		Код	2	5	9	3	5	8	2	6	8

С использованием таблицы 86 составим таблицу 87 с обучающей выборкой.

**Таблица 87 – ФОРМА ДЛЯ ВВОДА ОБУЧАЮЩЕЙ ВЫБОРКИ**

Код	Наименование	Классы			Признаки								
		1	2	3	Проекция-1			Проекция-2			Проекция-3		
					1	2	3	4	5	6	7	8	9
1	Шар	1			1	4	7						
2	Шар-11	2			1	4	7	1	4	7			
3	Шар-12	3			1	4	7				1	4	7
4	Шар-22	4						1	4	7	1	4	7
5	Шар-123	5			1	4	7	1	4	7	1	4	7
6	Тетраэдр	6			3	6	9						
7	Тетраэдр-11	7			3	6	9	3	6	9			
8	Тетраэдр-12	8			3	6	9				3	6	9
9	Тетраэдр-23	9						3	6	9	3	6	9
10	Тетраэдр-123	10			3	6	9	3	6	9	3	6	9

Продолжение таблицы 87

Код	Наименование	Классы			Признаки								
		1	2	3	Проекция-1			Проекция-2			Проекция-3		
					1	2	3	4	5	6	7	8	9
11	Куб	11			2	5	8						
12	Куб-11	12			2	5	8	2	5	8			
13	Куб-12	13			2	5	8				2	5	8
14	Куб-23	14						2	5	8	2	5	8
15	Куб-123	15			2	5	8	2	5	8	2	5	8
16	Конус	16			3	6	7						
17	Конус-11	17			3	6	7	1	6	9			
18	Конус-12	18			3	6	7				3	4	9
19	Конус-23	19						1	6	9	3	4	9
20	Конус-123	20			3	6	7	1	6	9	3	4	9
21	Пирамида	21			3	6	8						
22	Пирамида-11	22			3	6	8	2	6	9			
23	Пирамида-12	23			3	6	8				3	5	9
24	Пирамида-23	24						2	6	9	3	5	9
25	Пирамида-123	25			3	6	8	2	6	9	3	5	9
26	Цилиндр	26			2	5	7						
27	Цилиндр-11	27			2	5	7	1	5	8			
28	Цилиндр-12	28			2	5	7				2	4	8
29	Цилиндр-23	29						1	5	8	2	4	8
30	Цилиндр-123	30			2	5	7	1	5	8	2	4	8
31	Призма	31			2	5	9						
32	Призма-11	32			2	5	9	3	5	8			
33	Призма-12	33			2	5	9				2	6	8
34	Призма-23	34						3	5	8	2	6	8
35	Призма-123	35			2	5	9	3	5	8	2	6	8

## ***2. Осуществить синтез и верификацию (измерение адекватности) семантической информационной модели.***

Вводим классификационные и описательные шкалы и градации (в 1-м и 2-м режимах 1-й подсистемы), а также обучающую выборку (в 1-м режиме 2-й подсистемы) в систему "Эйдос" и осуществляем синтез модели (в 5-й функции 3-го режима 2-й подсистемы), а затем ее верификацию (во 2-м режиме 6-й подсистемы). В результате получаем семантическую информационную модель, обладающую достаточно высокой степенью адекватности: 94,29%, чтобы результаты ее исследования можно было считать исследованием самой моделируемой предметной области (рисунок 181).

RASP - RASP

Т 12 x 20

Универсальная когнитивная аналитическая система. 7:28 (с) НПП *Эйдос*

Подсистема анализа. Измерение адекватности информационной модели

Анкет физических: 35 логических (всего/факт): 35/ 35  
 Верная идентификация: 33 Ошибочная неидентификация: 2  
 Верная идентификация: 94.29% Ошибочная неидентификация: 5.71%

Шар 35/ 1

Код	Наименование класса	Анкет лог-х.	Идент. верно	Идент. ошиб.	Неидент. верно	Неидент. ошибоч.	ВЕРНАЯ. ИДЕНТ. %	Ошибочн. идентиф. %
1	Шар	1	1	4	30	0	100.00	11.76
2	Шар-11	1	1	4	30	0	100.00	11.76
3	Шар-12	1	1	4	30	0	100.00	11.76
4	Шар-23	1	1	4	30	0	100.00	11.76
5	Шар-123	1	1	4	30	0	100.00	11.76
6	Тетраэдр	1	1	12	22	0	100.00	35.29
7	Тетраэдр-11	1	1	10	24	0	100.00	29.41
8	Тетраэдр-12	1	1	7	27	0	100.00	20.59
9	Тетраэдр-23	1	1	7	27	0	100.00	20.59
10	Тетраэдр-123	1	1	4	30	0	100.00	11.76

F1Генерация отчета F2Сортировка F3Печать F4Поиск F8Расч. внешней валид. F9Удал. классов

Рисунок 181. Экранная форма по измерению адекватности СИМ

### 3. Провести системно-когнитивный анализ модели:

#### 3.1. Решить задачи идентификации и прогнозирования.

После выполнения 2-го режима 4-й подсистемы "Распознавание" в 1-й функции 3-го режима 4-й подсистемы получаем итоговую форму по результатам идентификации (таблица 88).

Таблица 88 – ИТОГОВЫЕ РЕЗУЛЬТАТЫ ИДЕНТИФИКАЦИИ

Объект		Класс		Сходство
Код	Наименован.	Код	Наименован.	
1	Шар	1	Шар	94,083
2	Шар-11	1	Шар	76,819
3	Шар-12	1	Шар	76,819
4	Шар-23	1	Шар	76,819
5	Шар-123	1	Шар	54,319
6	Тетраэдр	6	Тетраэдр	94,257
7	Тетраэдр-11	6	Тетраэдр	76,960
8	Тетраэдр-12	6	Тетраэдр	76,960
9	Тетраэдр-23	6	Тетраэдр	76,960
10	Тетраэдр-123	6	Тетраэдр	54,419
11	Куб	11	Куб	94,189
12	Куб-11	11	Куб	76,905
13	Куб-12	11	Куб	76,905
14	Куб-23	11	Куб	76,905
15	Куб-123	11	Куб	54,380
16	Конус	16	Конус	92,171
17	Конус-11	17	Конус-11	75,941
18	Конус-12	20	Конус-123	77,354
19	Конус-23	20	Конус-123	82,367
20	Конус-123	20	Конус-123	69,532



## Продолжение таблицы 88

Объект		Класс		Сходство
Код	Наименован.	Код	Наименован.	
21	Пирамида	21	Пирамида	94,257
22	Пирамида-11	22	Пирамида-11	66,725
23	Пирамида-12	23	Пирамида-12	64,571
24	Пирамида-23	24	Пирамида-23	62,412
25	Пирамида-123	32	Призма-11	44,934
26	Цилиндр	26	Цилиндр	92,128
27	Цилиндр-11	27	Цилиндр-11	77,694
28	Цилиндр-12	27	Цилиндр-11	77,694
29	Цилиндр-23	30	Цилиндр-123	82,136
30	Цилиндр-123	30	Цилиндр-123	69,328
31	Призма	31	Призма	94,189
32	Призма-11	32	Призма-11	67,933
33	Призма-12	33	Призма-12	63,420
34	Призма-23	34	Призма-23	62,412
35	Призма-123	32	Призма-11	44,934

Из таблицы 88 видно, что объект "Пирамида-123" неверно идентифицирован как класс "Призма-11". В остальных случаях тип объекта идентифицирован верно, что не исключает в ряде случаев неверной идентификации вида проекции (что, конечно, не имеет отношения к телам Платона – первым трем классам). Карточка, дающая расшифровку результатов идентификации 25-го объекта "Пирамида-123", представлена на рисунке 182.

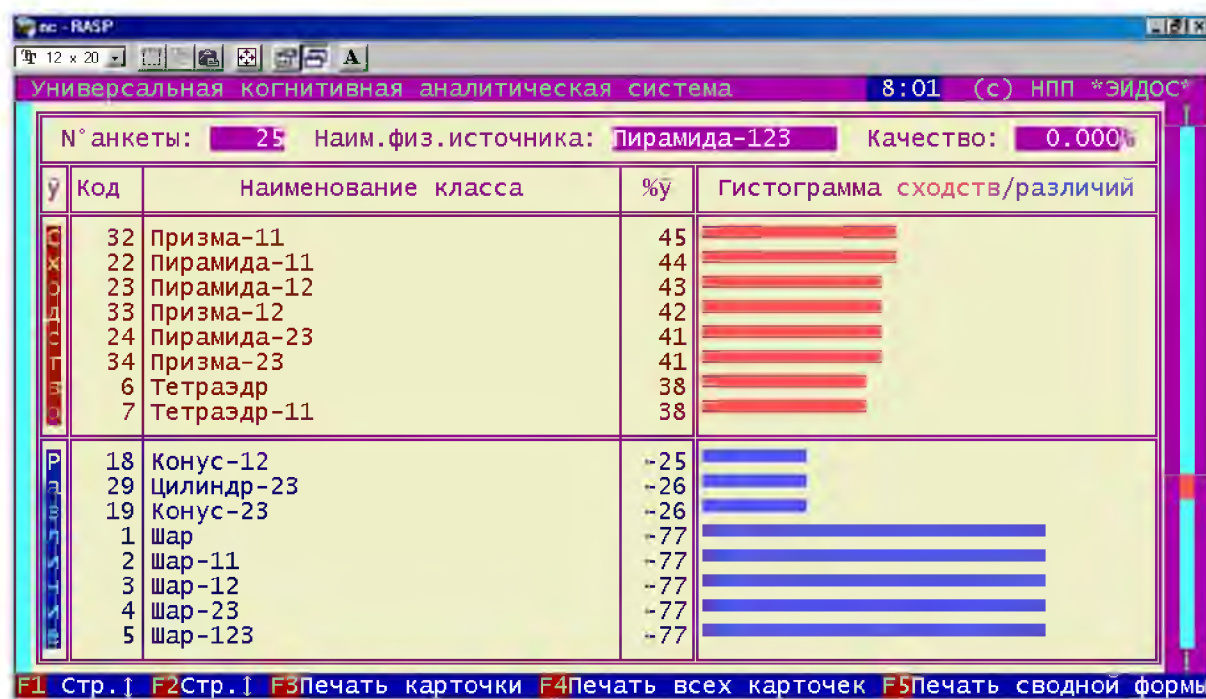


Рисунок 182. Карточка результатов идентификации объекта 25.

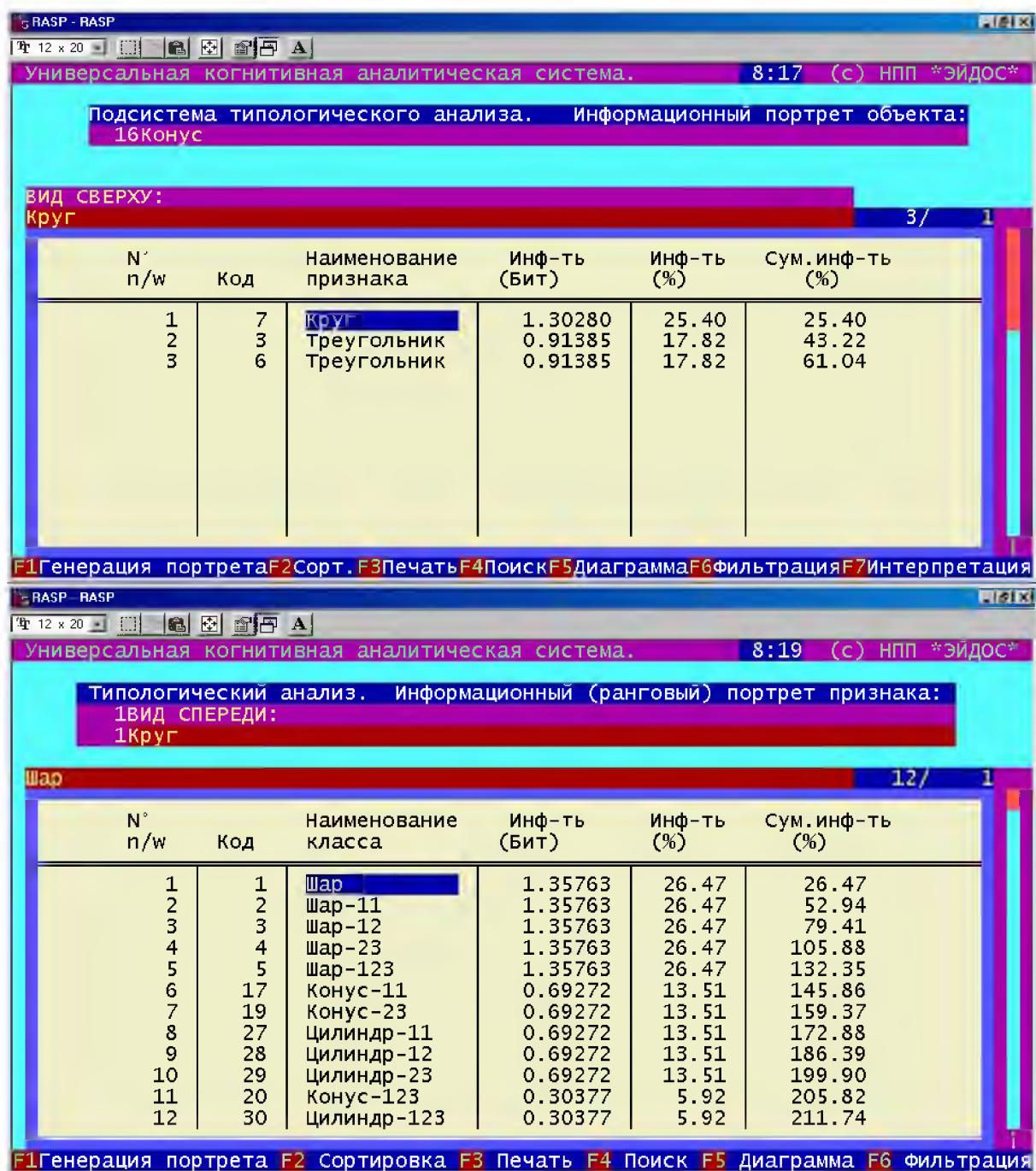
Если проанализировать коды признаков призмы и пирамиды, то можно увидеть, что они одни и те же, и отличие состоит лишь в количестве повторно-



стей признаков. Это и является причиной сложностей при дифференцировании этих объектов.

**3.2. Сгенерировать информационные портреты классов и факторов, т.е. решить обратную задачу прогнозирования (результаты отобразить в графической форме двухмерных и трехмерных профилей классов и факторов).**

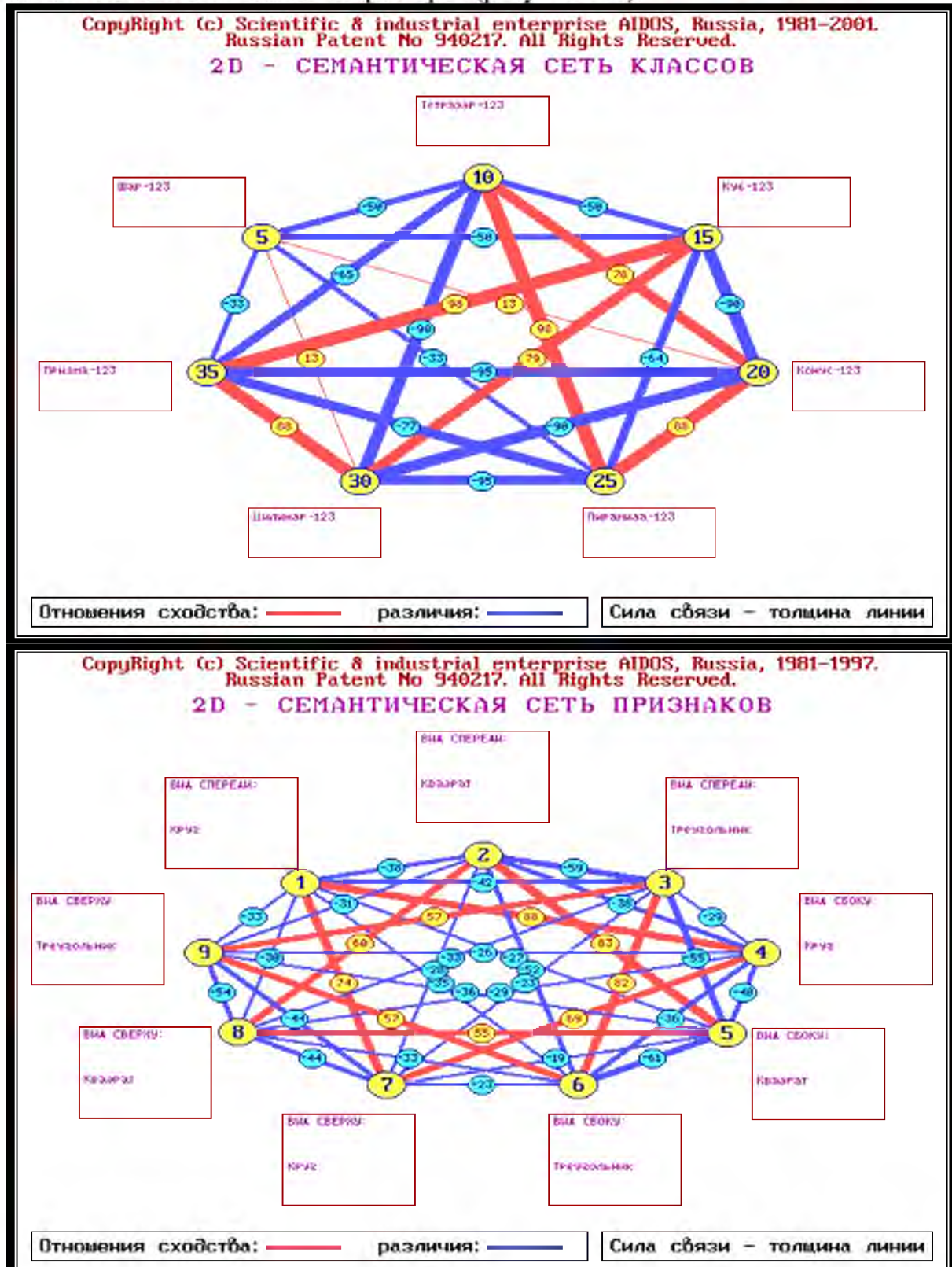
В 1-й функции 1-го режима 5-й подсистемы системы "Эйдос" получим информационный портрет класса, а в 1-й функции 2-го режима той же подсистемы – информационный портрет признака (рисунок 183).



**Рисунок 183. Примеры информационных портретов классов и признаков**

**3.3. Провести кластерно-конструктивный анализ классов и факторов (результаты отобразить в форме семантических сетей классов и факторов).**

В 5-й операции 2-й функции 1-го и 2-го режимов 5-й подсистемы получим семантические сети классов и факторов (рисунок 184).



**Рисунок 184. Примеры семантических сетей классов и факторов**

Интерпретацию семантических сетей дать самостоятельно.

### 3.4. Осуществить содержательное сравнение классов и факторов (результаты отобразить в форме когнитивных диаграмм классов и факторов).

В 3-й функции 1-го и 2-го режимов 5-й подсистемы получим когнитивные диаграммы классов и признаков (рисунок 185).

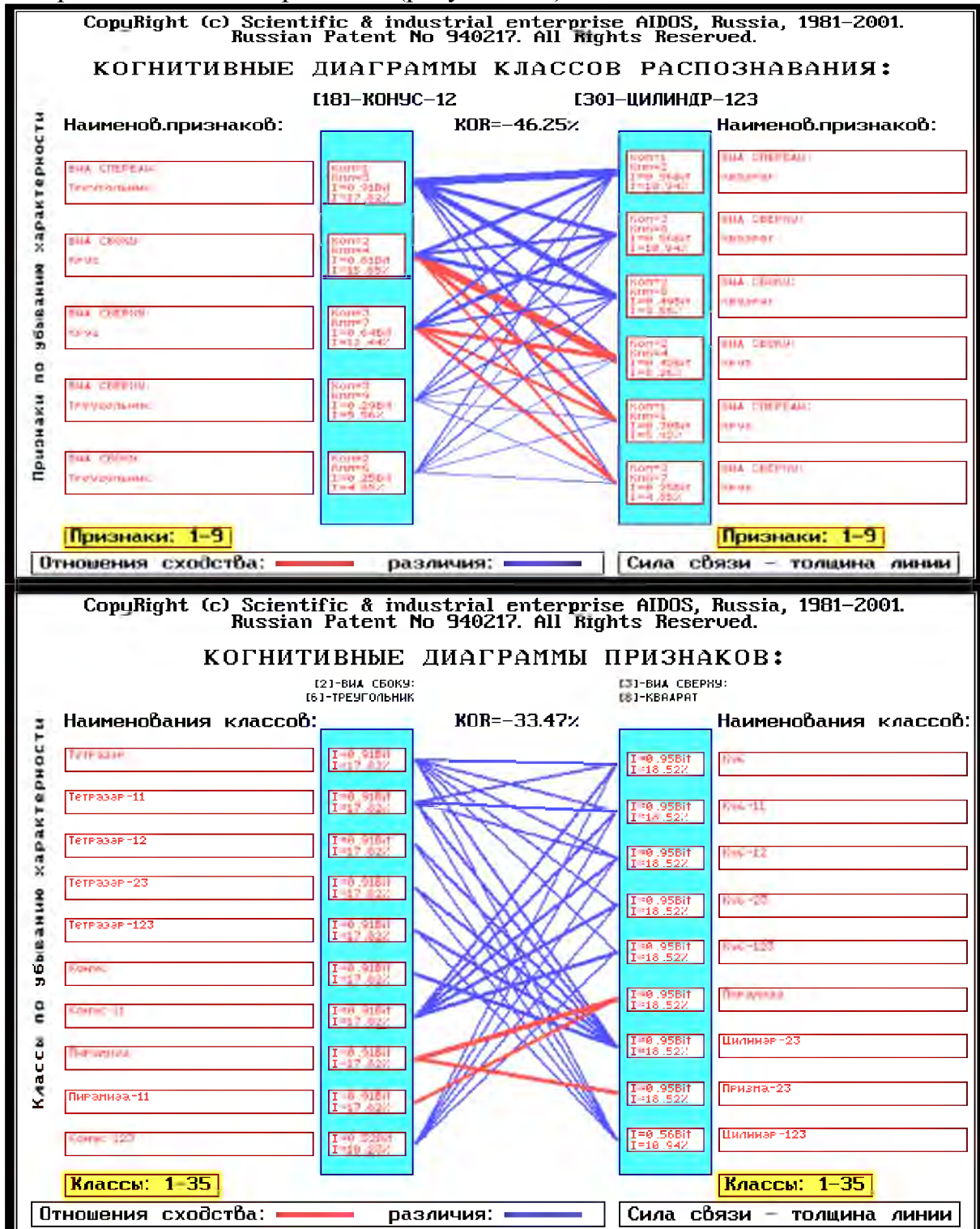


Рисунок 185. Пример когнитивных диаграмм, содержащих результаты содержательного сравнения классов и факторов



Из диаграммы классов мы видим, что основной вклад в сходство конуса и цилиндра вносит то, что одна из проекций у них одинаковая, т.е. круг, а основной вклад в различие – что остальные проекции у конуса – треугольник, а у цилиндра – квадрат.

Из диаграммы сравнения признаков мы видим, что такие проекции есть и у пирамиды, и у призмы, а в различие – то, что эти проекции характерны для различных тел.

Задания 3.5 – 3.7 выполнить самостоятельно.

### ***Контрольные опросы***

1. Что называется ортогональными проекциями тел?
2. В чем заключается сложность идентификации тел по их ортогональным проекциям?
3. За счет чего облегчается задача идентификации тел по их проекциям при наблюдении с нескольких точек или в движении?

### ***Литература по лабораторной работе***

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

**ЛР-9:**  
**"Прогнозирование количественных  
и качественных результатов выращивания зерновых  
колосовых и поддержка принятия решений по выбору  
агротехнологий"**

***Краткая теория***

Исследование проведено совместно с О.А.Засухиной на базе Кубанского государственного аграрного университета в 1993-1996 годах [81]. Предложенные технологии АСК-анализа реализовывались на базе системы "Эйдос". С помощью сформированной содержательной информационной модели прогнозировались количественные и качественные результаты выращивания культур и вырабатывались научно-обоснованные рекомендации по управлению продуктивностью сельхозкультур и качеством сельхозпродукции.

Созданная модель включала:

- объект управления (сельскохозяйственную культуру); классы (будущие состояния объекта управления, т.е. количественные и качественные результаты выращивания);
- факторы управляющей системы (агротехнологии, т.е. нормы посева, виды и нормы внесения удобрений, методы вспашки, ротация севооборота и т.п.);
- факторы окружающей среды (вид почв, культуры–предшественники по предшествующим годам и др.).

***Задание***

На основе предложенной технологии АСК-анализа разработать конкретное приложение системы "Эйдос", обеспечивающее управление урожайностью и качеством сельскохозяйственных культур путем выбора и применения оптимальной агротехнологии в зависимости от поставленной цели и вида почв, культуры–предшественника, а также ряда других параметров объекта управления и окружающей среды, например, таких как: нормы посева, виды и нормы внесения удобрений, методы вспашки, ротация и др.

***Пример решения***

При создании методики выполнялись следующие работы:

1. Формулировка целей методики и в соответствии с ними разработка перечня прогнозируемых хозяйственных ситуаций, т.е. результатов выращивания (например, для классификации будущих состояний, в том числе целевых, могут быть использованы "шкала качества" и "шкала количества", рисунок 186):

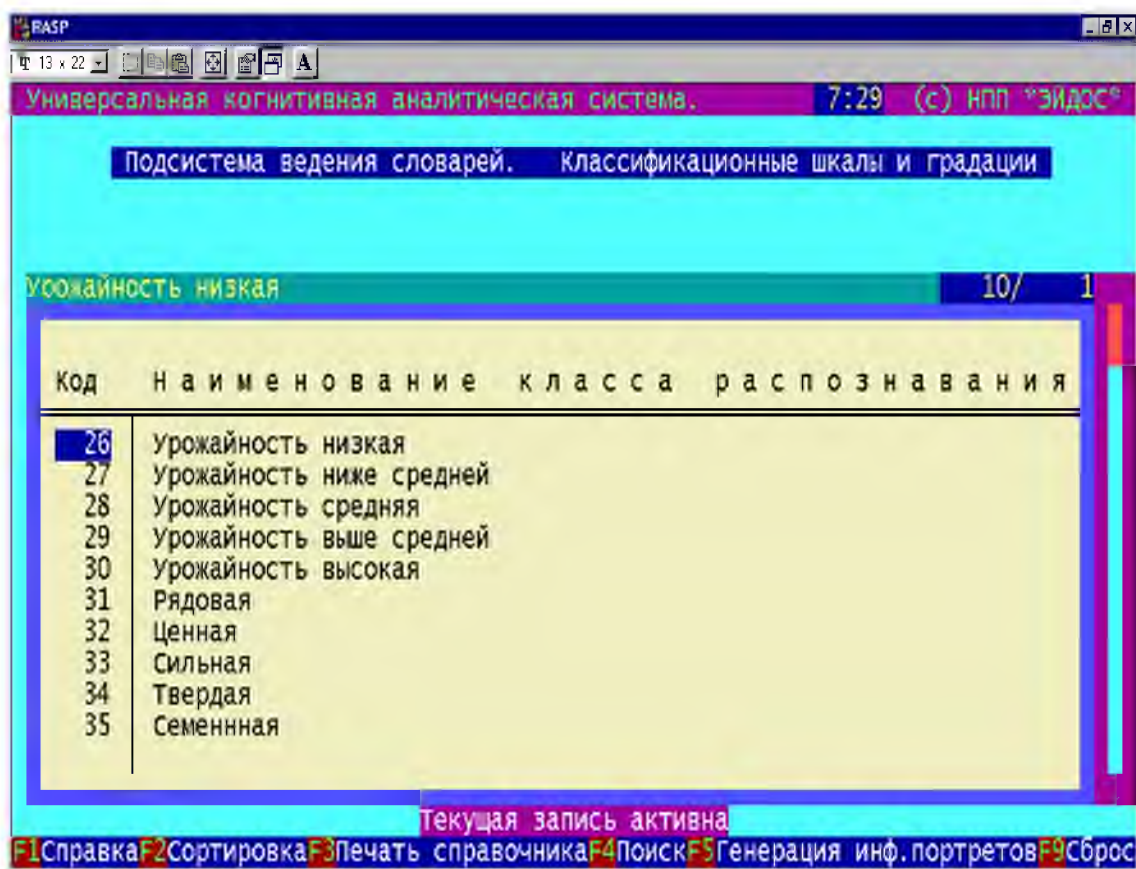


Рисунок 186. Будущие состояния объекта управления: количественные и качественные результаты выращивания сельхозкультуры (зерновые колосовые)

2. Разработка формализованного паспорта результатов выращивания сельхозкультур, позволяющего описать в пригодной для компьютерной обработки форме результаты выращивания конкретной сельхозкультуры на конкретном поле, по конкретной технологии при конкретных условиях окружающей среды.

Соответственно, формализованный паспорт результатов выращивания состоит из трех частей:

- первая содержит целевые и нежелательные количественные и качественные результаты выращивания (*классы*);
- вторая – *технологические факторы*, которые можно рассматривать как средство достижения желаемых хозяйственных результатов (рисунок 187).
- третья – описывает не зависящие от воли человека *факторы окружающей среды*, прежде всего метеорологические, а также виды и состояние почв;

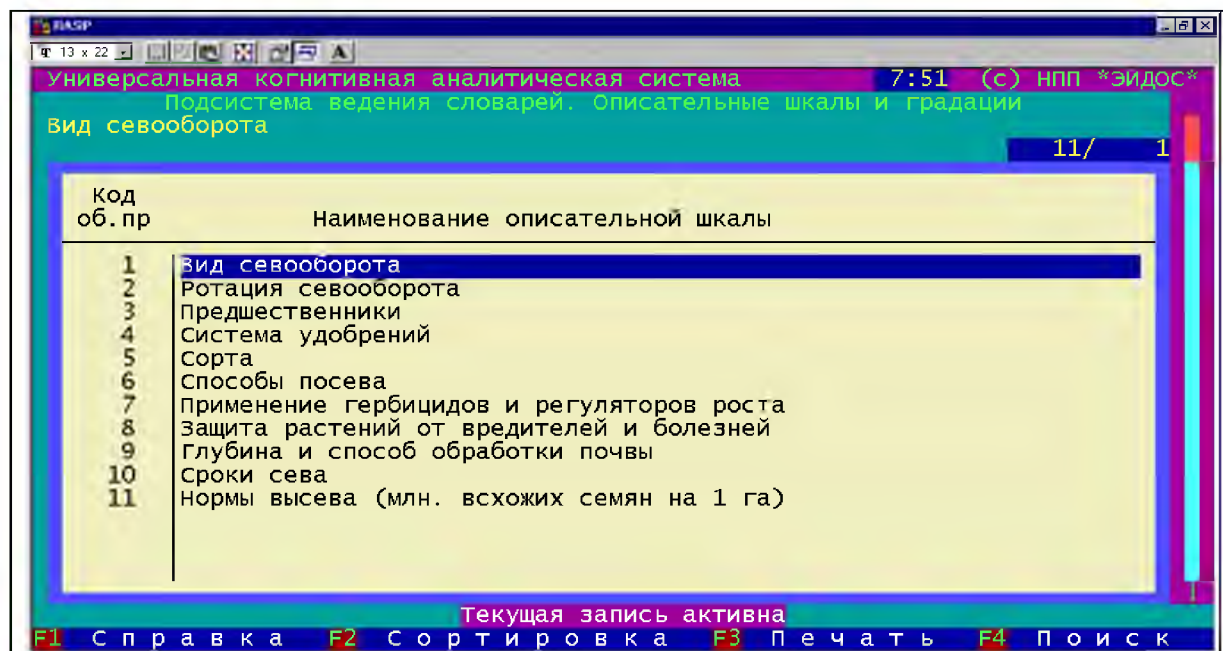


Рисунок 187. Видеограмма с фрагментом справочника описательных шкал (технологические факторы)

3. Использование бумажного архива по выращиванию сельхозкультур для заполнения формализованных паспортов и ввода в программную инструментальную систему в качестве примеров выращивания (обучающей выборки) (рисунок 188). Синтез семантической информационной модели. *Размерность модели составила: 35 прогнозируемых результатов выращивания, 188 градаций факторов, 217 прецедентов в обучающей выборке, 18594 факта.*

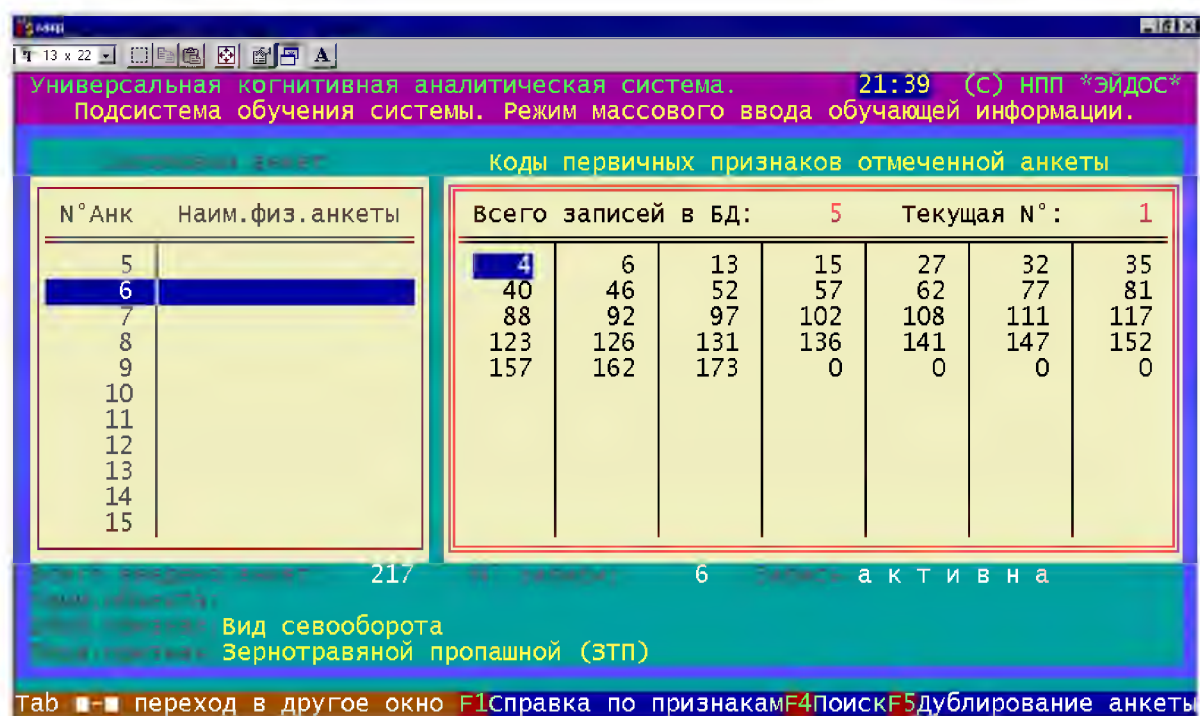
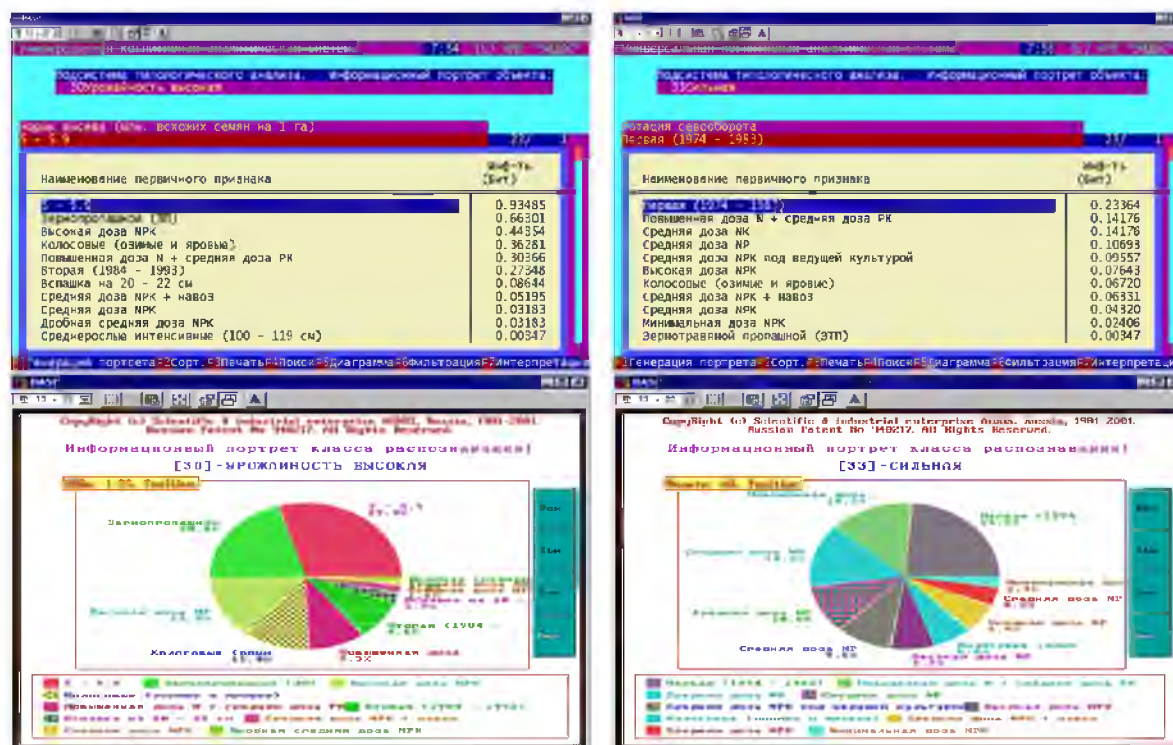


Рисунок 188. Интерфейс ввода обучающей выборки

4. Выявление (на основе предъявленных реальных примеров выращивания сельхозкультур) взаимосвязей между применяемыми технологиями и полученными

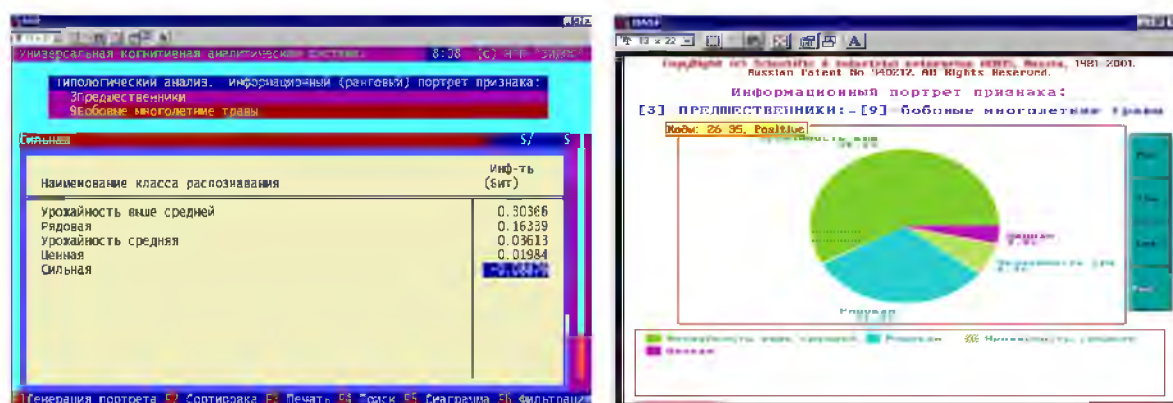


ми результатами и формирование информационных портретов по каждому возможному результату выращивания. Информационный портрет хозяйственной ситуации представляет собой перечень технологических факторов с количественным указанием того, какое влияние оказывает каждый из них на осуществление данной ситуации (рисунок 189):



**Рисунок 189. Примеры информационных портретов результатов выращивания "Высокое качество" и "Высокое количество"**

5. Каждый из технологических факторов на основе приведенных примеров характеризуется тем, какое влияние он оказывает на осуществление каждой (целевой или нежелательной) хозяйственной ситуации (рисунок 190):



**Рисунок 190. Семантический портрет признака: "Предшественники – бобовые многолетние травы"**

6. Сравнение различных хозяйственных ситуаций и формирование групп наиболее сходных из них (кластеров), а также определение кластеров, наиболее сильно отличаются друг от друга (конструктов). При этом на экспериментальной



базе данных был выявлен конструкт "качество–количество", означающий, что для получения высокого качества и большого количества необходимы совершенно противоположные и несовместимые (т.е. невозможные одновременно) почвы: предшественники и агротехнологические приемы (рисунок 191):

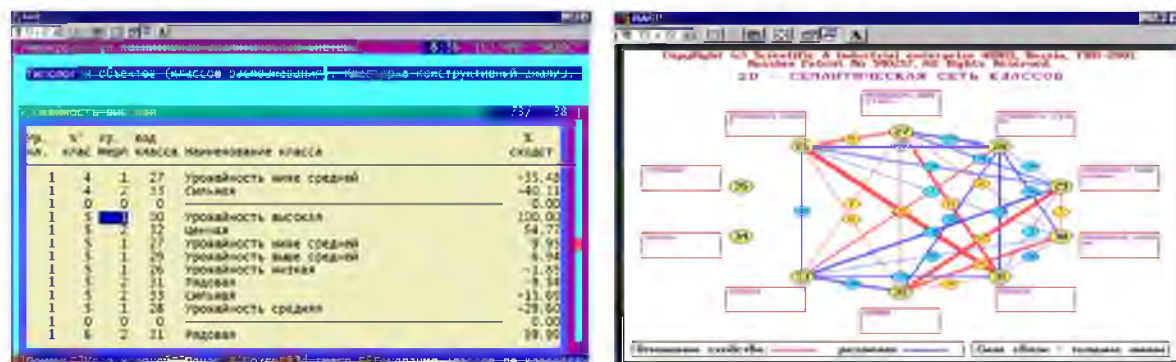


Рисунок 191. Конструкт классов: "Качество – количество" и семантическая сеть классов по шкалам: "Качество – количество"

7. Группировка технологических факторов в кластеры и конструкты. Кластерно-конструктивный анализ факторов показал, что *некоторые различные по своей природе факторы имеют сходное влияние на хозяйственные результаты*. Эти факторы предложено использовать для замены друг друга в случае необходимости (рисунок 192):

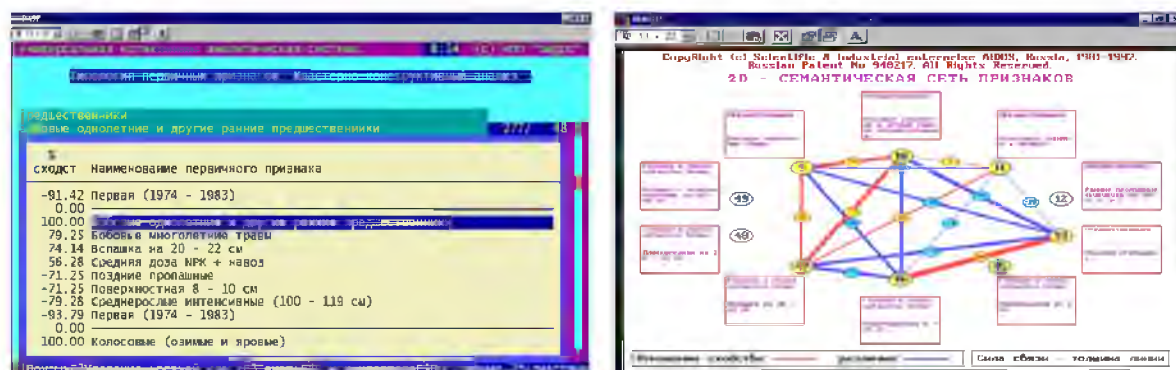


Рисунок 192. Конструкт факторов: "Предшественники бобовые ... – Ротация первая..." и семантической сети факторов: "Предшественники – Глубина обработки почвы"

8. Проверка способности созданного приложения правильно прогнозировать хозяйственные результаты на массиве уже введенных формализованных паспортов показала, что валидность оказалась недостаточно высокой для практического применения: на уровне 58%. Причиной этого являются артефакты, из-за которых некоторые хозяйственные ситуации оказались слабо детерминированными (рисунок 193). Удаление артефактов привело к повышению интегральной валидности до 80%, что достаточно для практического использования методики (рисунок 194)

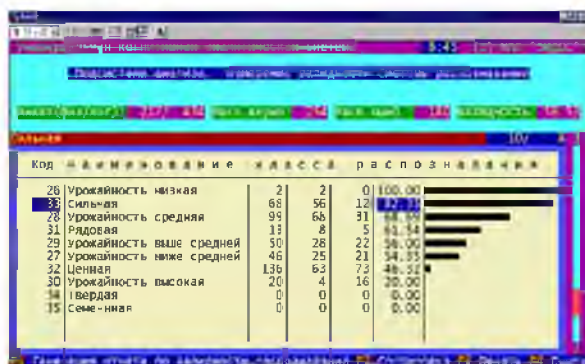


Рисунок 193. Интегральная и дифференциальная валидность методики до исключения артефактов

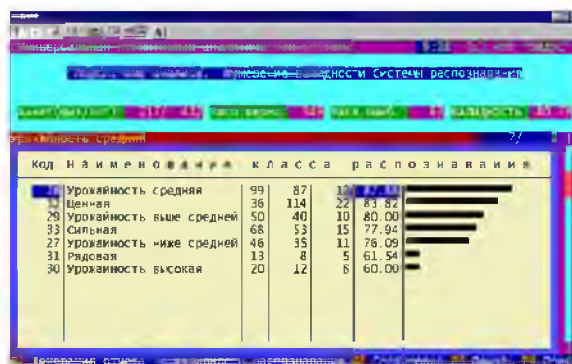


Рисунок 194. Интегральная и дифференциальная валидность методики после исключения артефактов

Таким образом, были решены две основные задачи:

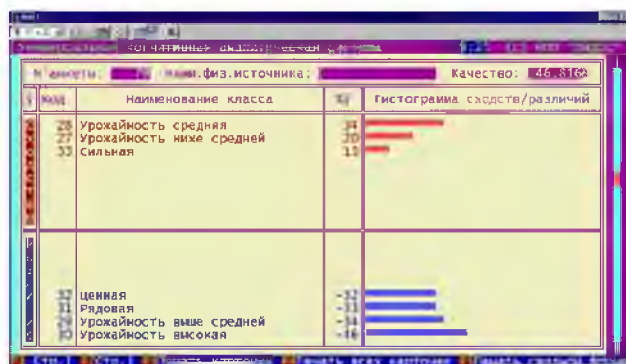


Рисунок 195. Пример карточки прогнозирования для конкретных условий выращивания

1. *Прогнозирование* того, какие хозяйственные результаты наиболее вероятны (а какие практически невозможны) на данном виде почв и с данными предшественниками, а также при условии применения имеющихся в распоряжении агротехнологий (рисунок 195). Указана мера сходства прогнозируемой ситуации с каждым будущим состоянием.

2. *Разработка рекомендаций по управляющим воздействиям*, т.е. консультирование по вопросам о том, какие виды почв, предшественники и *агротехнологии* должны быть, чтобы можно было рассчитывать с определенной уверенностью на заданный хозяйственный результат (рисунок 189).

Система позволяет оценивать степень достоверности своих прогнозов и рекомендаций по управлению, т.е. она не просто дает рекомендацию, но и количественно оценивает степень ее надежности. Кроме того, система дает характеристику влияния каждого технологического приема и рекомендации по замене желательных, но очень дорогих или не имеющих в наличии технологических приемов, другими, более дешевыми и доступными, и, при этом, имеющими сходное влияние на хозяйственные результаты.

Таким образом, данная методика позволяет "просматривать" различные варианты технологии, прогнозировать последствия применения различных технологических приемов, и на этой основе вырабатывать научно обоснованные рекомендации по выбору возделываемой культуры и оптимальной для поставленных целей агротехнологии.

В данном исследовании *в количественной форме* были обнаружены как уже известные закономерности по влиянию предшественников, почв, удобрений, способов вспашки и т.д. на результаты выращивания сельхозкультур, так и новые, ранее неизвестные.

### **Контрольные вопросы**

1. Что выбирается в качестве классов распознавания и что в качестве факторов в задаче прогнозирования урожая?
2. Что такое "шкала качества" и "шкала количества"? Какие градации у этих шкал?
3. Как классифицировать факторы, влияющие на урожай?
4. В чем состоит задача прогнозирования количественных и качественных результатов выращивания сельхозкультур?
5. В чем состоит задача поддержки принятия решений при выборе зон и подзон выращивания, культур для выращивания, агротехнологий?
6. Каким образом можно научно обоснованно заменять одни факторы другими?
7. С какими проблемами сталкивается разработчик и исследователь при решении задачи прогнозирования урожая?

### **Литература по лабораторной работе**

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

## ЛР-10: "Исследование случайной семантической информационной модели при различных объемах выборки"

### *Краткая теория*

Любая реальная информация о предметной области содержит как полезную информацию о закономерностях, так и шум. Соотношение между полезной информацией (сигналом) и шумом может быть различным:

- уровень сигнала может быть намного выше уровня шума, тогда шум можно считать несущественным;
- уровень сигнала может быть намного ниже уровня шума, тогда шум может существенно сказываться на свойствах модели.

Поэтому представляет интерес исследование семантической информационной модели, созданной на основе случайной обучающей выборки, в которой принадлежность анкет с описаниями объектов к классам и сам набор признаков в них – случайные.

Какая-то часть валидности обусловлена законами теории вероятностей, а какая-то – наличием закономерностей в предметной области и работой системы распознавания, причем в зависимости от параметров модели (размерности по классам и признакам и объема обучающей выборки).

Например, при увеличении объема выборки результат все ближе к предсказываемому теорией вероятностей. Но модель "борется" за повышение адекватности идентификации. И в результате получается валидность заметно выше, чем по теории вероятностей даже при довольно больших выборках.

Когда анализируешь величину интегральной валидности и оцениваешь ее в категориях "довольно хорошая", или "не достаточно высокая", то надо сравнивать ее с валидностью, получаемой по теории вероятностей. Например, если есть два класса, то валидность даже с неработающей системой распознавания должна быть 50% при *равновероятных* классах, а если классов 10, то валидность должна быть 10%. И только то, что свыше этого значения, предсказываемого теорией вероятности, можно отнести на счет закономерностей в предметной области и работы модели.

Если статистика мала и закон больших чисел не применим, то система "Эйдос" воспринимает шум как закономерности (причем даже иногда детерминистского характера, когда статистики вообще нет) и дает тем более высокую валидность модели, чем меньше статистика.

Получается, что о выявлении закономерностей в предметной области можно говорить только тогда, когда статистика достаточно велика, т.е. настолько велика, что модель может подавить или отсеять шум. Если бы в предметной области не было закономерностей (а был только шум), то валидность была бы

близка с предсказываемой теорией вероятностей, но фактически она значительно выше.

При увеличении объема обучающей выборки:

Во-первых, валидность должна стремиться не к нулю, а к величине, предсказываемой теорией вероятностей для равновероятных событий. Можно, конечно, ввести некую величину (каузальная валидность), как разность фактической валидности в системе "Эйдос" и теоретически предсказанной по теории вероятностей. Вот она уже будет стремиться к нулю.

Во-вторых, свойства шума таковы, что эта каузальная валидность должна стремиться к нулю и при внутренней, и при внешней валидности. Это должно происходить просто по определению шума (корреляция белого шума с белым шумом равна нулю), и потому, что интегральный критерий сходства в модели представляет собой *корреляцию*.

В-третьих, то, что как показывают численные эксперименты, каузальная валидность довольно медленно стремится к нулю, может означать, с одной стороны, невысокое качество генератора псевдослучайных чисел, а с другой стороны, – высокое качество модели распознавания, по-видимому, являющейся мощным средством выявления закономерностей в предметной области. Кстати, учитывая это, можно сравнивать различные генераторы "на степень их случайности".

Можно исследовать случайную модель с такими же параметрами, как какая-нибудь из реальных моделей (с таким же количеством классов, признаков, анкет) и сравнить их валидность. Получится некая величина. Можно считать, что разница между валидностью в реальном примере и случайной модели обусловлена наличием причинно-следственных связей в предметной области.

## **Задание**

1. Исследовать зависимость интегральной валидности семантической информационно-модели в зависимости от объема обучающей выборки при различном количестве классов и признаков.
2. Построить графики в Excel и дать их интерпретацию.

## **Пример решения**

В системе "Эйдос" в 1-й подсистеме в 5-м режиме встроена 4-я функция, обеспечивающая генерацию случайной модели (классификационных и описательных шкал и обучающей выборки) при заданных в диалоге параметрах модели (рисунок 170).



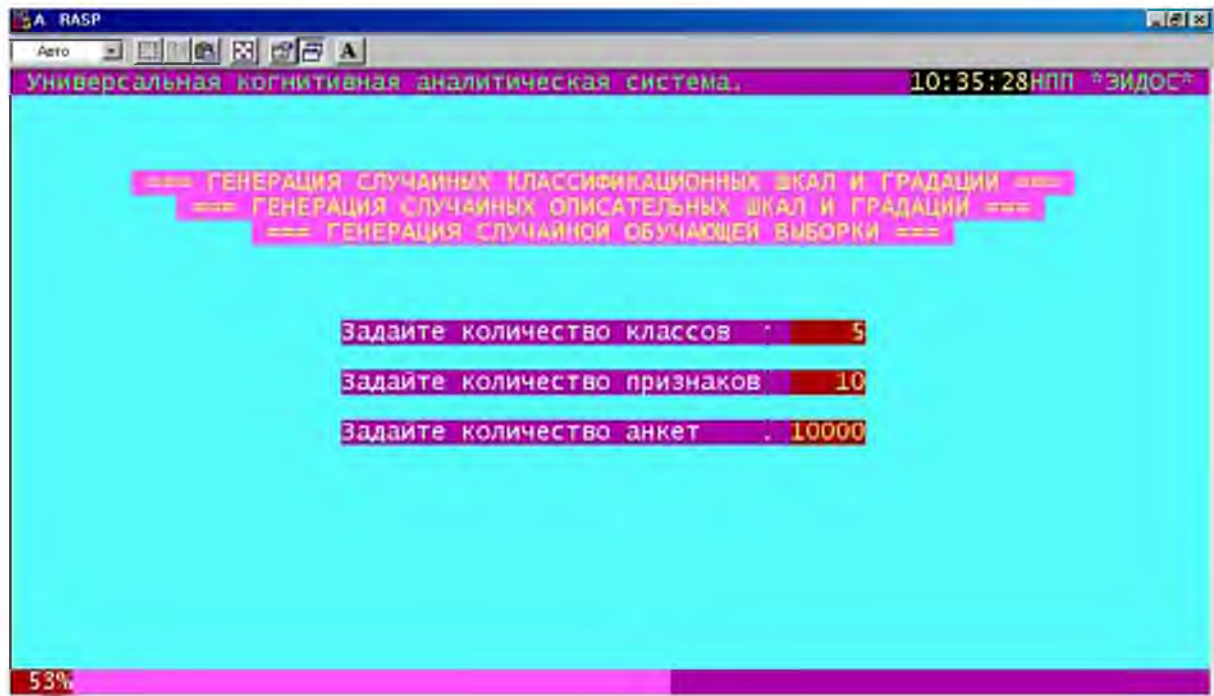


Рисунок 196. Интерфейс режима генерации случайной модели

Реальный исходный текст реализации этой функции приведен ниже.

```
***** ФОРМИРОВАНИЕ КЛАССИФИКАЦИОННЫХ И ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ, *****
***** А ТАКЖЕ ОБУЧАЮЩЕЙ ВЫБОРКИ СЛУЧАЙНЫМ ОБРАЗОМ *****
***** Луценко Е.В., 04/16/04 02:01pm *****

FUNCTION InpRandom()

***** БЛОК-1. ОТОБРАЖЕНИЕ ИНФОРМАЦИИ О ФУНКЦИЯХ ПРОГРАММНОГО ИНТЕРФЕЙСА *****

scr23 = SAVESCREEN(0,0,24,79)

SET CURSOR OFF
SET DATE ITALIAN
SET DECIMALS TO 15
SET ESCAPE On

Titul(.T.)

SHOWTIME(0,60,.F.,"rg+/n",.F.,.F.)

Mess1 = " === ГЕНЕРАЦИЯ СЛУЧАЙНЫХ КЛАССИФИКАЦИОННЫХ ШКАЛ И ГРАДАЦИЙ === "
Mess2 = " === ГЕНЕРАЦИЯ СЛУЧАЙНЫХ ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ === "
Mess3 = " === ГЕНЕРАЦИЯ СЛУЧАЙНОЙ ОБУЧАЮЩЕЙ ВЫБОРКИ === "
@4,40-LEN(Mess1)/2 SAY Mess1 COLOR "rg+/rb"
@5,40-LEN(Mess2)/2 SAY Mess2 COLOR "rg+/rb"
@6,40-LEN(Mess3)/2 SAY Mess3 COLOR "rg+/rb"

** БЛОК-3. ЗАДАНИЕ В ДИАЛОГЕ ДИАПАЗОНОВ СТОЛБЦОВ С КЛАССАМИ И ФАКТОРАМИ *****

N_Obj = 2
Mess = "Задайте количество классов : ####"
@10,40-LEN(Mess)/2 SAY Mess COLOR "w+/rb"
@10,52 GET N_Obj PICTURE "####" COLOR "rg+/r+"
N_Atr = 10
Mess = "Задайте количество признаков: ####"
@12,40-LEN(Mess)/2 SAY Mess COLOR "w+/rb"
@12,52 GET N_Atr PICTURE "####" COLOR "rg+/r+"
N_Ank = 100
Mess = "Задайте количество анкет : ####"
@14,40-LEN(Mess)/2 SAY Mess COLOR "w+/rb"
@14,52 GET N_Ank PICTURE "####" COLOR "rg+/r+"

SET CURSOR ON;READ;SET CURSOR OFF
```

```

IF LASTKEY()=27
  RESTSCREEN(0,0,24,79,scr23)
  RETURN
ENDIF

USE Setup EXCLUSIVE NEW
N_LogAnk = N_Obj          && Макс. кол-во логических анкет в одной физической
CLOSE ALL

** БЛОК-7. ГЕНЕРАЦИЯ КЛАССИФИКАЦИОННЫХ ШКАЛ И ГРАДАЦИЙ *****

Mess1 = " === ГЕНЕРАЦИЯ СЛУЧАЙНЫХ КЛАССИФИКАЦИОННЫХ ШКАЛ И ГРАДАЦИЙ === "
@4,40-LEN(Mess1)/2 SAY Mess1 COLOR "rg+*/rb"

@24,0 SAY REPLICATE("-",80) COLOR "rb/n"
x=0

CLOSE ALL
USE Object EXCLUSIVE NEW;ZAP
FOR j=1 TO N_Obj
  APPEND BLANK
  REPLACE Kod WITH j
  REPLACE Name WITH "Obj_"+ALLTRIM(STR(j,5))
  p=++x/(N_Obj+N_Atr+N_Ank)*100;p=IF(p<=100,p,100)
  @24,0 SAY STR(p,3)+"%" COLOR "w+/r+"
  @24,4 SAY REPLICATE("-",0.76*p) COLOR "rb+/n"
NEXT

** БЛОК-8. ГЕНЕРАЦИЯ ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ *****

Mess2 = " === ГЕНЕРАЦИЯ СЛУЧАЙНЫХ ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ === "
@5,40-LEN(Mess2)/2 SAY Mess2 COLOR "rg+*/rb"

USE Priz_ob EXCLUSIVE NEW;ZAP
USE Priz_per EXCLUSIVE NEW;ZAP

@24,0 SAY REPLICATE("-",80) COLOR "rb/n"

FOR i=1 TO N_Atr
  APPEND BLANK
  REPLACE Kod WITH i
  REPLACE Name WITH "Atr_"+ALLTRIM(STR(i,5))
  p=++x/(N_Obj+N_Atr+N_Ank)*100;p=IF(p<=100,p,100)
  @24,0 SAY STR(p,3)+"%" COLOR "w+/r+"
  @24,4 SAY REPLICATE("-",0.76*p) COLOR "rb+/n"
NEXT

** БЛОК-9. ГЕНЕРАЦИЯ ОБУЧАЮЩЕЙ ВЫБОРКИ *****

Mess3 = " === ГЕНЕРАЦИЯ СЛУЧАЙНОЙ ОБУЧАЮЩЕЙ ВЫБОРКИ === "
@6,40-LEN(Mess3)/2 SAY Mess3 COLOR "rg+*/rb"

USE ObInfZag EXCLUSIVE NEW;ZAP
USE ObInfKpr EXCLUSIVE NEW;ZAP

N_Rec = RECCOUNT()

FOR M_KodIst=1 TO N_Ank

  SELECT ObInfZag
  APPEND BLANK
  REPLACE Kod_ist WITH M_KodIst
  R = 1+INT(N_Obj*RANDOM()/65535)          && Код класса
  R = IF(R<1,1,R)
  R = IF(R>N_Obj,N_Obj,R)
  REPLACE Name_ist WITH "Ist-"+STRTRAN(STR(M_KodIst,5)," ", "0")+ "-"+STRTRAN(STR(R,5),"
", "0")
  FIELDPUT(3,R)

  ***** Генерация массива кодов признаков в для БД ObInfKpr
  A_Kpr := {}
  FOR j=1 TO N_Atr
    M_KodPr = 1+INT(N_Atr*RANDOM()/65535)    && Код признака
    M_KodPr = IF(M_KodPr<1,1,M_KodPr)
    M_KodPr = IF(M_KodPr>N_Atr,N_Atr,M_KodPr)
    IF ASCAN(A_Kpr,M_KodPr) = 0              && Если признак еще не встречался
      IF LEN(A_Kpr)+1 <= 4000

```

```

        AADD (A_Kpr,M_KodPr)
    ENDIF
ENDIF
NEXT
ASORT (A_Kpr)
***** Запись массива кодов признаков в БД ObInfKpr
SELECT ObInfKpr
APPEND BLANK
FIELDPUT (1,M_KodIst)
k=2
FOR j=1 TO LEN(A_Kpr)
    IF k <= 12
        FIELDPUT (k++,A_Kpr[j])
    ELSE
        APPEND BLANK
        FIELDPUT (1,M_KodIst)
        k=2
        FIELDPUT (k ,A_Kpr[j])
    ENDIF
NEXT
p=++x/(N_Obj+N_Atr+N_Ank)*100;p=IF (p<=100,p,100)
@24,0 SAY STR(p,3)+"%" COLOR "w+/r+"
@24,4 SAY REPLICATE("-",0.76*p) COLOR "rb+/n"
NEXT

@24,0 SAY REPLICATE("-",80) COLOR "rb/n"
GenNtxObj(.T.)
@24,0 SAY REPLICATE("-",80) COLOR "rb/n"
GenNtxPro(.T.)
@24,0 SAY REPLICATE("-",80) COLOR "rb/n"
GenNtxPrp(.T.)
@24,0 SAY REPLICATE("-",80) COLOR "rb/n"
GenNtxOin(.T.)
@24,0 SAY REPLICATE("-",80) COLOR "rb/n"

Mess = " ПРОЦЕСС ГЕНЕРАЦИИ ЗАВЕРШЕН УСПЕШНО !!! "
@24,40-LEN(Mess)/2 SAY Mess COLOR "rg+/rb"

INKEY(0)

RESTSCREEN(0,0,24,79,scr23)
CLOSE ALL

RETURN

```

### Алгоритм выполнения задания 1:

**Шаг 0.** Вход.

**Шаг 1.** Выполнить генерацию классификационных и описательных шкал и градаций, а также обучающей выборки при минимальном объеме обучающей выборки (например, при 2-х объектах).

**Шаг 2.** Произвести синтез модели.

**Шаг 3.** Скопировать обучающую выборку в распознаваемую.

**Шаг 4.** Произвести распознавание.

**Шаг 5.** Измерить внутреннюю интегральную валидность и занести информацию о параметрах случайной модели в Excel.

**Шаг 6.** Увеличить объем обучающей выборки на величину "Дельта".

**Шаг 7.** Если объем обучающей выборки меньше максимума, то перейти на шаг 2, иначе – на шаг 8.

**Шаг 8.** Выход.



Учащимся предлагается выполнить приведенный алгоритм самостоятельно.

### ***Контрольные вопросы***

1. Что такое "случайная модель"?
2. Почему необходимо исследовать свойства случайной модели?
3. Какие средства для генерации и исследования случайных моделей есть в системе "Эйдос"?

### ***Литература по лабораторной работе***

1. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. – 280с.
2. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

## ЧАСТЬ III. ЗАОЧНОЕ ОБУЧЕНИЕ, САМОСТОЯТЕЛЬНАЯ РАБОТА СТУДЕНТОВ И ПОДГОТОВКА К ЭКЗАМЕНУ

### 3.1. ИЗУЧЕНИЕ ДИСЦИПЛИНЫ ПРИ ЗАОЧНОЙ ФОРМЕ ОБУЧЕНИЯ

Изучение дисциплины "Интеллектуальные информационные системы при заочной форме обучения имеет свои *особенности*, которые в основном состоят в том, что из-за значительного сокращения и без того малого количества часов на лекции и лабораторные работы *существенно возрастает значение самостоятельной работы студентов*, т.е. работы с данным учебным пособием и рекомендованными в нем Internet-сайтами.

В *лекционном курсе* читается 4 обзорных лекции, названия которых совпадают с названиями разделов 1-й части:

1. Введение в интеллектуальные информационные системы.
2. Теоретические основы и эксплуатация универсальной когнитивной аналитической системы "Эйдос".
3. Принципы построения интеллектуальных информационных систем.
4. Применение и перспективы систем искусственного интеллекта.

*Практикум* базируется на универсальной когнитивной аналитической системе "Эйдос", разработанной автором пособия, и включает первые 2 лабораторные работы, рекомендованные в таблицах 40 и 41 на стр. 410:

- ЛР-1: "Прогнозирование учебных достижений студентов на основе их имеджевых фотороботов" (8 часов);
- ЛР-2: "Прогнозирование вероятных пунктов назначения железнодорожных составов" (4 часа).

### 3.2. ПРОГРАММА САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТА ПО ДИСЦИПЛИНЕ

#### 3.2.1. Общие положения

Ниже приводятся примерные наименования тем для самостоятельной работы студентов. Каждая тема изучается одним студентом индивидуально. На изучение каждой темы отводится 38 часов самостоятельной работы. По результатам работы оформляется реферат.

### 3.2.2. Наименования тем для самостоятельной работы студентов

Задания к самостоятельной работе (СР) студентов пронумерованы и снабжены индексом «Ю» или «Э», означающими, что эти задания рекомендуются студентам, обучающимся по «юридической» (Ю) или «экономической» (Э) специальностям. Все самостоятельные работы выполняются с применением системы «Эйдос», кроме тех случаев, когда другие системы указаны конкретно.

#### **Темы самостоятельной работы студентов:**

**СР-1-Ю.** Идентификация изображений местности по их вербальным описаниям.

**СР-2-Ю.** Оценка рисков правонарушений по признакам почерка (психографология).

**СР-3-ЭЮ.** Оценка рисков страхования и кредитования предприятий по их вербальным описаниям.

**СР-4-Ю.** Прогнозирование рисков совершения ДТП (дорожно-транспортных происшествий) по видам и времени на основе данных о владельце и автомобиле.

**СР-5-Э.** Прогнозирование успешности деятельности фирмы на основе оценки ее персонала.

**СР-6-Ю.** Прогнозирование продолжительности жизни пациентов, перенесших сердечный приступ, по данным эхокардиограммы на основе базы данных репозитория UCI.

**СР-7.** Классификация животных по внешним признакам на основе базы данных репозитория UCI.

**СР-8.** Диагностика фитопатологии по симптоматике и выработку рекомендаций по плану лечения на основе информации, содержащейся в учебниках.

**СР-9-Ю.** Идентификация изображений различных мест на территории КубГАУ по вербальным описаниям их фотографий (изображения взять с сайта КубГАУ: <http://kubagro.ru>, фотогалерея). СК-анализ семантической информационной модели.

**СР-10.** Прогнозирование успеваемости по ИИС на основе данных по социальному статусу студентов и их родителей.

**СР-11.** Прогнозирование направления деятельности фирмы на основе данных о расположении и внешнем виде ее офиса. СК-анализ семантической информационной модели.

**СР-12.** Выбор автомобиля для приобретения по его признакам (обучающую выборку взять на автомобильном рынке). СК-анализ семантической информационной модели.

**СР-13.** Выбор вариантов приобретения жилья по его признакам. СК-анализ семантической информационной модели.

**СР-14-Э.** Оценка важности различных видов городского транспорта и различных маршрутов в разрезе по остановкам. СК-анализ семантической информационной модели.

**CP-15.** Исследование систем: FineReader, Cunie Form и других систем ввода текстов со сканера. Исследовать зависимость качества распознавания текста от разрешения сканирования для разных систем. Оценку качества производить по количеству ошибок распознавания на одном и том же тексте. Составить рейтинг систем и версий, дать рекомендации. Оценить тоже самое, после использования After Scan.

**CP-16.** Исследование систем Stylus (Promt), Сократ, и других систем автоматизированного перевода. Сравнить качество автоматизированного перевода с русского языка на английский язык и обратно для текстов различной направленности (юридические, технические, художественные, стихи) и с различной длиной и сложностью предложений (статистика). Составить рейтинг систем и версий, дать рекомендации. Оценку качества перевода осуществлять путем обобщения экспертных оценок экспертов с разным уровнем компетентности (студенты).

**CP-17.** Исследовать реальную систему распознавание образов, идентификации и прогнозирования при решении задач лабораторных работ.

**CP-18.** Исследовать реальную систему поддержки принятия решений при решении задач лабораторных работ.

**CP-19.** Исследовать реальную экспертную систему при решении задач лабораторных работ.

**CP-20.** Исследовать реальную систему класса: "Нейронная сеть" на примере пакета NeuroOffice при решении задач лабораторных работ.

**CP-21.** Исследовать реальную систему, реализующую генетические алгоритмы при решении задач лабораторных работ.

**CP-22.** Исследовать реальную систему когнитивного моделирования при решении задач лабораторных работ.

**CP-23.** Исследовать реальную систему выявления знаний из опыта (эмпирических фактов) и интеллектуального анализа данных при решении задач лабораторных работ.

**CP-23+.** Решение задач идентификации и прогнозирования на основе данных репозитория UCI по следующим направлениям:

**Abalone**, Adult, Annealing, Anonymous Microsoft Web Data, Arrhythmia, Artificial Characters, **Audiologys**, Auto-Mpg, Automobile, Badges, Balance Scale, Balloons, Breast Cancer, **Wisconsin** Breast Cancers, **Pittsburgh** Bridges, Car Evaluation, Census Income, **Chesss**, Bach Chorales (time-series), Connect-4 Opening, Credit Screenings, Computer Hardware, Contraceptive Method Choice, **Coverttype** data, Cylinder Bands, Dermatology, Diabetes Data, The Second Data Generation Program - DGP/2, Document Understanding, **EBL** Domain Theories and Examples, Echocardiogram, **Ecoli**, Flags, Function Findings, Glass Identification, Haberman's Survival Data, Hayes-Roth, Heart Diseases, Hepatitis, Horse Colic, Housing (Boston), ICU Data, Image segmentation, Internet Advertisements, Ionosphere, Iris Plant, **Isolet** Spoken Letter Recognition, Kinship, Labor relations, LED Display Domains, Lenses, Letter Recognition, Liver-disorders, Logic-theorist, Lung Cancer, Lymphography, Mechanical Analysis Data, Meta-data, Mobile Robots, Molecular Biologys, MONK's Problems, Moral Reasoner, Multiple Features, Mushrooms, MUSKs, Nursery, **Othello** Domain Theory, Page Blocks Classification, **Pima** Indians Diabetes, Optical

Recognition of Handwritten Digits, Pen-Based Recognition of Handwritten Digits, Postoperative Patient, Primary Tumor, Qualitative Structure Activity Relationships (QSARs), **Quadrapped** Animals Data Generator, Servo, Shuttle Landing Control, Solar Flares, Soybeans, Challenger USA Space Shuttle O-Rings, Low Resolution Spectrometer, **Spambase**, **SPECT** and **SPECTF** hearts, Sponge, **Statlog** Projects, Student Loan Relational, Teaching Assistant Evaluation, **Tic-Tac-Toe** Endgame, Thyroid Disease, Trains, University, Congressional Voting Records, Water **Treatment** Plant, Waveform Data Generator, Wine Recognition, Yeast, **Zoo**, **Undocumented**s.

### 3.2.3. Примерная структура реферата, курсовой или дипломной работы и элементы научного исследования в самостоятельной работе

Результаты самостоятельной работы оформляются студентами в форме реферата, примерная структура которого приведена ниже. Реферат оформляется как научная работа.

**Во введении** кратко описывается основное содержание и структура работы.

**В первой главе** с условным наименованием: "Постановка задачи и пути ее решения":

- обосновывается актуальность проблемы или задачи, решению которой посвящена работа;
- формулируются и обосновываются требования к методу решения поставленной проблемы или задачи. Если требования имеют "разный вес" или "разную значимость", то для оценки относительной важности критериев можно использовать экспертные оценки в какой-либо количественной шкале;
- делается аналитический обзор (по литературе и материалам Internet) традиционных методов решения, причем каждый метод оценивается по обоснованным критериям;
- по результатам литературного обзора составляется рейтинг традиционных методов и делается вывод, что такой-то метод является наилучшим. Однако и он не полностью соответствует сформулированным требованиям. Поэтому в следующей главе Вы предлагаете и реализуете свою концепцию решения поставленной задачи, свободную от указанных ограничений.

**Во второй главе** с условным наименованием: "Предлагаемое решение задачи":

- предлагаемая концепция может заключаться в том, что Вы как-то развиваете или модифицируете наилучший из рассмотренных методов;
- затем Вы подробно описываете свое решение, конкретизирующее концепцию: описываете математическую модель; методику численных расчетов, т.е. структуры данных и алгоритмы, реализующие математическую модель; программную реализацию математического метода и методики численных расчетов;
- после этого описываете, как созданный Вами инструментарий позволил решить поставленную проблему или задачу, приводите образцы входных и вы-

ходных форм с численными расчетами, подтверждающими эффективность предложенного подхода.

**В третьей главе** с условным наименованием: "Эффективность и перспективы":

- анализируя причины эффективности предложенной технологии показываете и ее ограничения, а потом делаете вывод о ее более широкой применимости, чем только для решения сформулированной проблемы или задачи;

- описываете перспективы развития предложенной технологии, в частности, пути преодоления сформулированных ее ограничений.

**В заключении** делается вывод о том, что предложенный подход к решению поставленной задачи оказался успешным и перспективным.

В научной статье каждый из пунктов может быть представлен одним абзацем, кроме 6-го, который составляет основной объем статьи. Если описывать более подробно, может получиться соответственно, реферат, курсовая, дипломная, кандидатская (решена практически важная задача) или докторская (решена теоретически важная научная проблема и, на основе этого, ряд практически важных задач).

### **3.2.4. Методика использования репозитория UCI для оценки качества математических моделей систем искусственного интеллекта**

#### **3.2.4.1. Постановка задачи и пути ее решения**

Одним из важнейших направлений развития информационных систем является создание систем искусственного интеллекта (СИИ), в частности систем: с интеллектуальной обратной связью (биологическая обратная связь и семантический резонанс); распознавания образов; поддержки принятия решений; экспертных систем; нейронных сетей; генетических алгоритмов и машинной эволюции; когнитивного моделирования; выявления знаний из опыта (эмпирических фактов) и интеллектуального анализа данных (data mining), а также ряда других.

В связи с наличием многих альтернатив возникает необходимость оценки качества математических моделей и поддерживающих их систем искусственного интеллекта (ММ СИИ). *Сопоставимое* сравнение различных моделей актуально как для разработчиков моделей (адекватная *самооценка* своей работы), так и для потребителей моделей (адекватная *оценка* моделей и обоснованный выбор оптимальной из них по заданным критериям). *Сопоставимость* оценки может быть обеспечена, если модели будут протестированы на *одних и тех же исходных данных* по ряду типовых задач из области искусственного интеллекта.

Для этого необходимы:

1. *Свободный доступ* к тестовым исходным данным.
2. *Методика*, обеспечивающая преобразование исходных данных из формы, доступной через Internet, в форму, в которой они могут быть использованы в конкретной программной системе искусственного интеллекта, реализующей ту или иную математическую модель.

1-е условие выполнено сотрудниками Школы информации и компьютерных исследований Калифорнийского университета США (*School of Information & Computer Science University of California, Irvine, USA, <http://www.ics.uci.edu>*), которыми *создан* и размещен в Internet по адресу: *<http://www.ics.uci.edu/~mlearn/MLSummary.html>* банк исходных данных по задачам искусственного интеллекта.

На момент написания настоящей работы этот банк включает базы исходных данных по следующим задачам: *Abalone, Adult, Annealing, Anonymous Microsoft Web Data, Arrhythmia, Artificial Characters, Audiologys, Auto-Mpg, Automobile, Badges, Balance Scale, Balloons, Breast Cancer, Wisconsin Breast Cancers, Pittsburgh Bridges, Car Evaluation, Census Income, Chesss, Bach Chorales (time-series), Connect-4 Opening, Credit Screenings, Computer Hardware, Contraceptive Method Choice, Covertype data, Cylinder Bands, Dermatology, Diabetes Data, The Second Data Generation Program - DGP/2, Document Understanding, EBL Domain Theories and Examples, Echocardiogram, Ecoli, Flags, Function Findings, Glass Identification, Haberman's Survival Data, Hayes-Roth, Heart Diseases, Hepatitis, Horse Colic, Housing (Boston), ICU Data, Image segmentation, Internet Advertisements, Ionosphere, Iris Plant, Isolet Spoken Letter Recognition, Kinship, Labor relations, LED Display Domains, Lenses, Letter Recognition, Liver-disorders, Logic-theorist, Lung Cancer, Lymphography, Mechanical Analysis Data, Meta-data, Mobile Robots, Molecular Biologys, MONK's Problems, Moral Reasoner, Multiple Features, Mushrooms, MUSKs, Nursery, Othello Domain Theory, Page Blocks Classification, Pima Indians Diabetes, Optical Recognition of Handwritten Digits, Pen-Based Recognition of Handwritten Digits, Postoperative Patient, Primary Tumor, Qualitative Structure Activity Relationships (QSARs), Quadraped Animals Data Generator, Servo, Shuttle Landing Control, Solar Flares, Soybeans, Challenger USA Space Shuttle O-Rings, Low Resolution Spectrometer, Spambase, SPECT and SPECTF hearts, Sponge, Statlog Projects, Student Loan Relational, Teaching Assistant Evaluation, Tic-Tac-Toe Endgame, Thyroid Disease, Trains, University, Congressional Voting Records, Water Treatement Plant, Waveform Data Generator, Wine Recognition, Yeast, Zoo, Undocumented.*

Поэтому остается выполнить 2-е условие, а именно: *разработать типовую методику использования баз данных репозитория UCI для оценки качества математических моделей систем искусственного интеллекта, что и является основной задачей* данного раздела.

На первый взгляд решение этой задачи является относительно несложным, т.к. требует в основном знания и умения использования стандартных возможностей Internet-броузера, Word и Excel. Однако как показывает опыт, это вполне может представлять определенную сложность из-за большого числа операций преобразования формы информации и принципиальной неполной формализуемости этого процесса. Поэтому предмет данной данного раздела достаточно актуален.

Сформулируем *основные требования* к методикам, предназначенным для этих целей, а также критерии их оценки и сравнения:

1. *Высокая степень автоматизированности, т.е. минимизация затрат ручного труда.*

2. *Высокая скорость преобразования информации и, как следствие, – несущественность ее объема.*

3. *Высокая достоверность преобразования, т.е. отсутствие ошибок.*

Однако в литературе и в Internet не приводятся методики аналогичного назначения. Исходя из этого можно предположить, что в основном это преобразование осуществляется вручную, что не соответствует сформулированным требованиям ни по одному из приведенных критериев.

Поэтому предлагается методика, свободная от указанных ограничений. Основная идея этой методики состоит в том, что преобразование формы пред-

ставления исходных данных из HTML-формата непосредственно в базы данных системы осуществляется в два основных этапа:

– на 1-м этапе с использованием стандартных возможностей Word и Excel осуществляется преобразование исходных данных из HTML-формата в промежуточные DBF-таблицы;

– на 2-м этапе с помощью специально разработанного программного интерфейса осуществляется преобразование исходных данных из промежуточных DBF-таблиц в стандарт баз данных используемой системы искусственного интеллекта.

Рассмотрим данную методику подробнее на примере преобразования баз данных репозитория UCI по примеру **ZOO-database** в стандарт баз данных универсальной когнитивной аналитической системы "Эйдос" [81].

### 3.2.4.2. Преобразование исходных данных из HTML-формата в промежуточные DBF-таблицы

#### 3.2.4.2.1. Характеристика исходных данных

Из Internet по адресу: <http://ftp.ics.uci.edu/pub/machine-learning-databases/zoo/zoo.names> получаем исходную информацию: общее описание тестовой задачи (файл: `zoo_names.htm`) и обучающую выборку (файл: `zoo_data.htm`), которые приводятся далее без изменений.

**Общее описание задачи (файл: `zoo_names.htm`):**

#### *Zoo Database*

From Richard Forsyth

Artificial

7 classes of animals

17 attributes (besides name), 15 Boolean and 2 numeric-valued

No missing attribute values

[Ftp Access](#)

1. Title: Zoo database

2. Source Information

-- Creator: Richard Forsyth

-- Donor: Richard S. Forsyth

8 Grosvenor Avenue

Mapperley Park

Nottingham NG3 5DX

0602-621676

-- Date: 5/15/1990

3. Past Usage:

-- None known other than what is shown in Forsyth's PC/BEAGLE User's Guide.

4. Relevant Information:

-- A simple database containing 17 Boolean-valued attributes. The "type" attribute appears to be the class attribute. Here is a breakdown of which animals are in which type: (I find it unusual that there are 2 instances of "frog" and one of "girl"!)

Class# Set of animals:

- 
- |        |                                                                                                                                                                                                                                                                                                                                     |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 (41) | aardvark, antelope, bear, boar, buffalo, calf, cavy, cheetah, deer, dolphin, elephant, fruitbat, giraffe, girl, goat, gorilla, hamster, hare, leopard, lion, lynx, mink, mole, mongoose, opossum, oryx, platypus, polecat, pony, porpoise, puma, pussycat, raccoon, reindeer, seal, sealion, squirrel, vampire, vole, wallaby, wolf |
| 2 (20) | chicken, crow, dove, duck, flamingo, gull, hawk, kiwi, lark, ostrich, parakeet, penguin, pheasant, rhea, skimmer, skua, sparrow, swan, vulture, wren                                                                                                                                                                                |
| 3 (5)  | pitviper, seasnake, slowworm, tortoise, tuatara                                                                                                                                                                                                                                                                                     |



- 4 (13) bass, carp, catfish, chub, dogfish, haddock, herring, pike, piranha, seahorse, sole, stingray, tuna
- 5 (4) frog, frog, newt, toad
- 6 (8) flea, gnat, honeybee, housefly, ladybird, moth, termite, wasp
- 7 (10) clam, crab, crayfish, lobster, octopus, scorpion, seawasp, slug, starfish, worm
5. Number of Instances: 101
6. Number of Attributes: 18 (animal name, 15 Boolean attributes, 2 numerics)
7. Attribute Information: (name of attribute and type of value domain)
1. animal name: Unique for each instance
  2. hair: Boolean
  3. feathers: Boolean
  4. eggs: Boolean
  5. milk: Boolean
  6. airborne: Boolean
  7. aquatic: Boolean
  8. predator: Boolean
  9. toothed: Boolean
  10. backbone: Boolean
  11. breathes: Boolean
  12. venomous: Boolean
  13. fins: Boolean
  14. legs: Numeric (set of values: {0,2,4,5,6,8})
  15. tail: Boolean
  16. domestic: Boolean
  17. catsize: Boolean
  18. type: Numeric (integer values in range [1,7])
8. Missing Attribute Values: None
9. Class Distribution: Given above

### Обучающая выборка (файл: zoo_data.htm)

aardvark,1,0,0,1,0,0,1,1,1,1,0,0,4,0,0,1,1	antelope,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
bass,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,4	bear,1,0,0,1,0,0,1,1,1,1,0,0,4,0,0,1,1
boar,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1	buffalo,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
calf,1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,1,1	carp,0,0,1,0,0,1,0,1,1,0,0,1,0,1,1,0,4
catfish,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,4	cavy,1,0,0,1,0,0,0,1,1,1,0,0,4,0,1,0,1
cheetah,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1	chicken,0,1,1,0,1,0,0,0,1,1,0,0,2,1,1,0,2
chub,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,4	clam,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,7
crab,0,0,1,0,0,1,1,0,0,0,0,0,4,0,0,0,7	crayfish,0,0,1,0,0,1,1,0,0,0,0,0,6,0,0,0,7
crow,0,1,1,0,1,0,1,0,1,1,0,0,2,1,0,0,2	deer,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
dogfish,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,1,4	dolphin,0,0,0,1,0,1,1,1,1,1,0,1,0,1,0,1,1
dove,0,1,1,0,1,0,0,0,1,1,0,0,2,1,1,0,2	duck,0,1,1,0,1,1,0,0,1,1,0,0,2,1,0,0,2
elephant,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1	flamingo,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,1,2
flea,0,0,1,0,0,0,0,0,0,1,0,0,6,0,0,0,6	frog,0,0,1,0,0,1,1,1,1,1,0,0,4,0,0,0,5
frog,0,0,1,0,0,1,1,1,1,1,0,4,0,0,0,5	fruitbat,1,0,0,1,1,0,0,1,1,1,0,0,2,1,0,0,1
giraffe,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1	girl,1,0,0,1,0,0,1,1,1,1,0,0,2,0,1,1,1
gnat,0,0,1,0,1,0,0,0,0,1,0,0,6,0,0,0,6	goat,1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,1,1
gorilla,1,0,0,1,0,0,0,1,1,1,0,0,2,0,0,1,1	gull,0,1,1,0,1,1,1,0,1,1,0,0,2,1,0,0,2
haddock,0,0,1,0,0,1,0,1,1,0,0,1,0,1,0,0,4	hamster,1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,0,1
hare,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,0,1	hawk,0,1,1,0,1,0,1,0,1,1,0,0,2,1,0,0,2
herring,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,4	honeybee,1,0,1,0,1,0,0,0,0,1,1,0,6,0,1,0,6
housefly,1,0,1,0,1,0,0,0,0,1,0,0,6,0,0,0,6	kiwi,0,1,1,0,0,0,1,0,1,1,0,0,2,1,0,0,2
ladybird,0,0,1,0,1,0,1,0,0,1,0,0,6,0,0,0,6	lark,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,0,2
leopard,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1	lion,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
lobster,0,0,1,0,0,1,1,0,0,0,0,0,6,0,0,0,7	lynx,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
mink,1,0,0,1,0,1,1,1,1,1,0,0,4,1,0,1,1	mole,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,0,1
mongoose,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1	moth,1,0,1,0,1,0,0,0,0,1,0,0,6,0,0,0,6
newt,0,0,1,0,0,1,1,1,1,1,0,0,4,1,0,0,5	octopus,0,0,1,0,0,1,1,0,0,0,0,0,8,0,0,1,7
opossum,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,0,1	oryx,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,1,1
ostrich,0,1,1,0,0,0,0,0,1,1,0,0,2,1,0,1,2	parakeet,0,1,1,0,1,0,0,0,1,0,0,2,1,1,0,2
penguin,0,1,1,0,0,1,1,0,1,1,0,0,2,1,0,1,2	pheasant,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,0,2
pike,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,1,4	piranha,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,0,4
pitviper,0,0,1,0,0,0,1,1,1,1,1,0,0,1,0,0,3	platypus,1,0,1,1,0,1,1,0,1,1,0,0,4,1,0,1,1
polecat,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1	pony,1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,1,1
porpoise,0,0,0,1,0,1,1,1,1,0,0,1,0,1,0,1,1	puma,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
pussycat,1,0,0,1,0,0,1,1,1,1,0,0,4,1,1,1,1	raccoon,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1
reindeer,1,0,0,1,0,0,0,1,1,1,0,0,4,1,1,1,1	rhea,0,1,1,0,0,0,1,0,1,1,0,0,2,1,0,1,2
scorpion,0,0,0,0,0,0,1,0,0,1,1,0,8,1,0,0,7	seahorse,0,0,1,0,0,1,0,1,1,0,0,1,0,1,0,0,4

seal,1,0,0,1,0,1,1,1,1,0,1,0,0,0,1,1	sealion,1,0,0,1,0,1,1,1,1,0,1,2,1,0,1,1
seasnake,0,0,0,0,0,1,1,1,1,0,1,0,0,1,0,3	seawasp,0,0,1,0,0,1,1,0,0,0,1,0,0,0,0,7
skimmer,0,1,1,0,1,1,1,0,1,1,0,0,2,1,0,2	skua,0,1,1,0,1,1,1,0,1,1,0,0,2,1,0,2
slowworm,0,0,1,0,0,0,1,1,1,1,0,0,0,1,0,3	slug,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,7
sole,0,0,1,0,0,1,0,1,1,0,0,1,0,1,0,4	sparrow,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,2
squirrel,1,0,0,1,0,0,0,1,1,1,0,0,2,1,0,1	starfish,0,0,1,0,0,1,1,0,0,0,0,0,5,0,0,7
stingray,0,0,1,0,0,1,1,1,1,0,1,1,0,1,0,1,4	swan,0,1,1,0,1,1,0,0,1,1,0,0,2,1,0,1,2
termite,0,0,1,0,0,0,0,0,0,1,0,0,6,0,0,6	toad,0,0,1,0,0,1,0,1,1,1,0,0,4,0,0,5
tortoise,0,0,1,0,0,0,0,0,1,1,0,0,4,1,0,1,3	tuatara,0,0,1,0,0,0,1,1,1,1,0,0,4,1,0,0,3
tuna,0,0,1,0,0,1,1,1,1,0,0,1,0,1,0,1,4	vampire,1,0,0,1,1,0,0,1,1,1,0,0,2,1,0,0,1
vole,1,0,0,1,0,0,0,1,1,1,0,0,4,1,0,0,1	vulture,0,1,1,0,1,0,1,0,1,1,0,0,2,1,0,1,2
wallaby,1,0,0,1,0,0,0,1,1,1,0,0,2,1,0,1,1	wasp,1,0,1,0,1,0,0,0,0,1,1,0,6,0,0,6
wolf,1,0,0,1,0,0,1,1,1,1,0,0,4,1,0,1,1	worm,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,7
wren,0,1,1,0,1,0,0,0,1,1,0,0,2,1,0,0,2	

### Некоторые замечания к общему описанию задачи

Считаем необходимым согласиться с мнением модератора UCI, который отметил два несоответствия в общем описании задачи (см. п.4):

– во-первых, 18-й атрибут **"type"** на самом деле содержит информацию о принадлежности объекта к классу и, поэтому, относится не к описательным шкалам и градациям, а к классификационным;

– во-вторых, в 5-м классе 2 раза указан один и тот же объект **"frog"** (возможно, они разного пола?).

Кроме того, проанализировав общее описание задачи, представленное репозитарием UCI, мы также отмечаем ряд несоответствий. Так, 1-й атрибут **"animal name"** является прямым указанием на объект и не должен включаться в систему атрибутов, т.к. это делает задачу идентификации объектов тривиальной. Классы автором задачи никак не названы и мы дали им свои названия. Вызывает некоторое сомнение объединение в один класс, который мы условно назвали "Многоногие", представителей различных видов, например таких, как скорпион и осьминог.

#### 3.2.4.2.2. Этапы разработки семантической информационной модели

Разработка семантической информационной модели (СИМ) в системе "Эйдос" осуществляется в ряд этапов [81]:

- формализация предметной области (разработка классификационных и описательных шкал и градаций);
- подготовка и ввод обучающей выборки;
- синтез СИМ;
- оптимизация СИМ;
- проверка адекватности СИМ (измерение внутренней и внешней дифференциальной и интегральной валидности).

Если модель обладает достаточной адекватностью, то принимается решение о ее использовании в режиме идентификации и прогнозирования, а также об анализе предметной области путем исследования ее семантической информационной модели.

### 3.2.4.2.3. Формализация предметной области: разработка классификационных и описательных шкал и градаций

С учетом сформулированных замечаний к общему описанию задачи классификационные шкалы и градации будут иметь вид:

1. Млекопитающие.
2. Птицы.
3. Пресмыкающиеся.
4. Рыбы.
5. Земноводные.
6. Насекомые.
7. Многоногие.

**Описательные шкалы и градации**, приведенные в общем описании задачи, включают в основном булевы атрибуты, а также один количественный: код: 14, наименование: **legs** (количество ног). Этот атрибут мы преобразовали в шкалу с булевыми градациями. После этого, с учетом сделанных замечаний к общему описанию задачи, описательные шкалы и градации приняли вид, представленный в таблице 89.

**Таблица 89 – ОПИСАТЕЛЬНЫЕ ШКАЛЫ И ГРАДАЦИИ**

Код	Наименование	Код	Наименование	Код	Наименование
2	hair	10	backbone	18	legs-4
3	feathers	11	breathes	19	legs-5
4	eggs	12	venomous	20	legs-6
5	milk	13	fins	21	legs-7
6	airborne	14	legs-0	22	legs-8
7	aquatic	15	legs-1	23	tail
8	predator	16	legs-2	24	domestic
9	toothed	17	legs-3	25	catsize

Таблица 89 преобразуется из HTML-формата в Excel следующим образом:

1. Отмечаем блоком в Internet-браузере ту часть раздела 7 общего описания задачи (файл: **zoo_names.htm**), в которой перечислены атрибуты, копируем ее в буфер обмена, переходим в Word и вставляем из буфера обмена в документ.

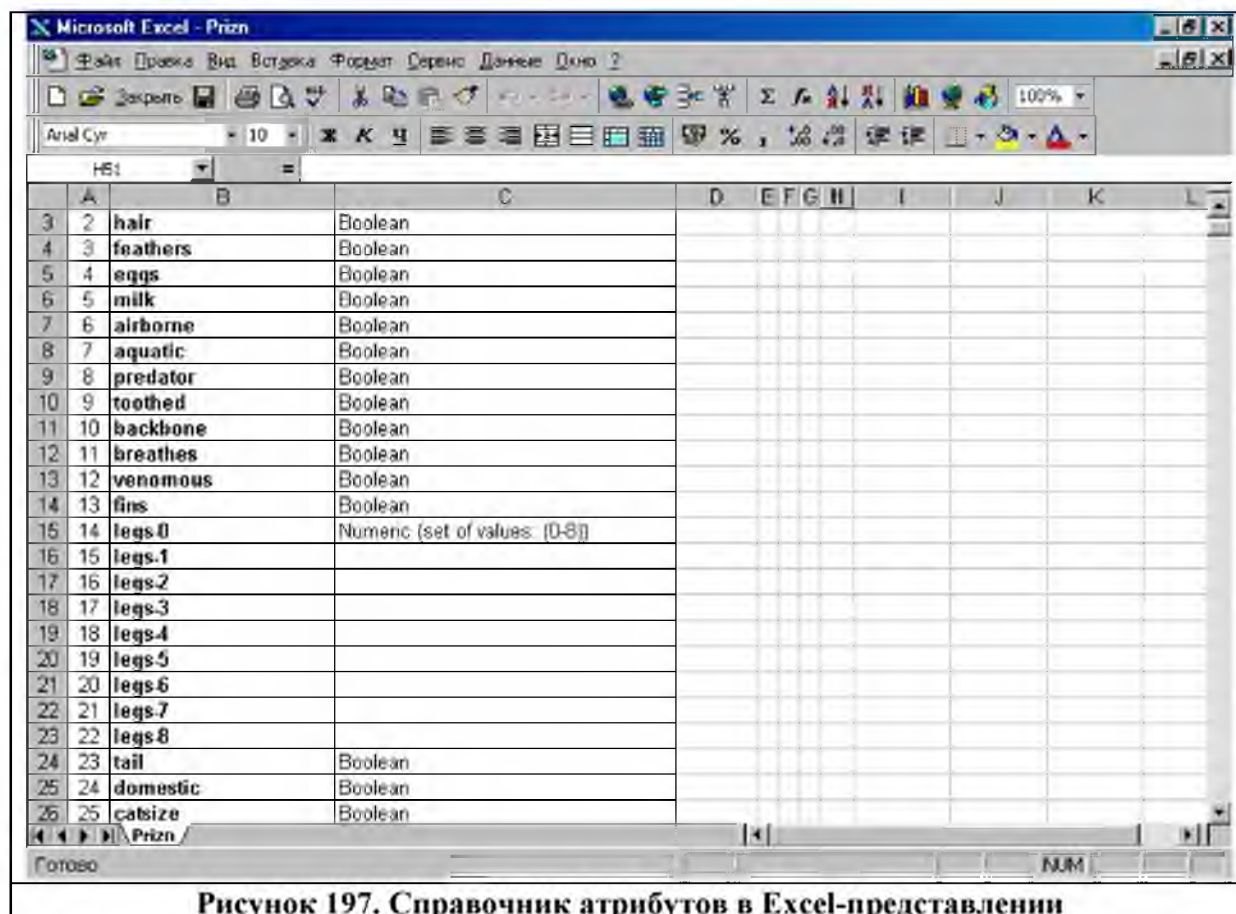
2. Записываем обучающую выборку в форме TXT-файла с именем **zoo_names.txt** в стандарте "Текст DOS с разбиением на строки".

3. Загружаем Excel и выполняем шаги:

- считываем файл **zoo_names.txt** в Excel, предварительно указав в окне "Тип файлов" вариант "Все файлы";
- задаем формат файла "DOS или OS/2 (PC-8) и нажимаем кнопку: "Далее";
- задаем символ-разделитель "Символ табуляции" и "считать последовательные разделители одним" и нажимаем кнопку "Готово". После некоторой

корректировки получаем вид описательных шкал и градаций, представленный на рисунке 197;

– записываем Excel-файл с описательными шкалами и градациями с именем **Prizn.xls**.



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - Prizn". The menu bar includes "Файл", "Правка", "Вид", "Вставка", "Формат", "Сервис", "Данные", and "Справка". The toolbar contains various icons for file operations, editing, and formatting. The active sheet is named "Prizn" and contains a table with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L
3	2	hair	Boolean									
4	3	feathers	Boolean									
5	4	eggs	Boolean									
6	5	milk	Boolean									
7	6	airborne	Boolean									
8	7	aquatic	Boolean									
9	8	predator	Boolean									
10	9	toothed	Boolean									
11	10	backbone	Boolean									
12	11	breathes	Boolean									
13	12	venomous	Boolean									
14	13	fins	Boolean									
15	14	legs0	Numeric (set of values: {0-8})									
16	15	legs1										
17	16	legs2										
18	17	legs3										
19	18	legs4										
20	19	legs5										
21	20	legs6										
22	21	legs7										
23	22	legs8										
24	23	tail	Boolean									
25	24	domestic	Boolean									
26	25	catsize	Boolean									

Рисунок 197. Справочник атрибутов в Excel-представлении

#### 3.2.4.2.4. Подготовка и ввод обучающей выборки

С обучающей выборкой осуществляем следующие преобразования формы представления:

1. Отмечаем блоком в Internet-броузере при просмотре файла **zoo_data.htm** весь файл, копируем его в буфер обмена, переходим в Word и вставляем из буфера обмена в документ.

2. Записываем обучающую выборку в форме TXT-файла с именем **zoo_data.txt** в стандарте "Текст DOS с разбиением на строки".

3. Загружаем Excel и выполняем шаги:

– считываем файл **zoo_data.txt** в Excel, предварительно указав в окне: "Тип файлов" вариант: "Все файлы", появляется 1-е окно (рисунок 198);

– в 1-м окне задаем формат файла: "DOS или OS/2 (PC-8)" и нажимаем кнопку: "Далее", появляется 2-е окно (рисунок 199);

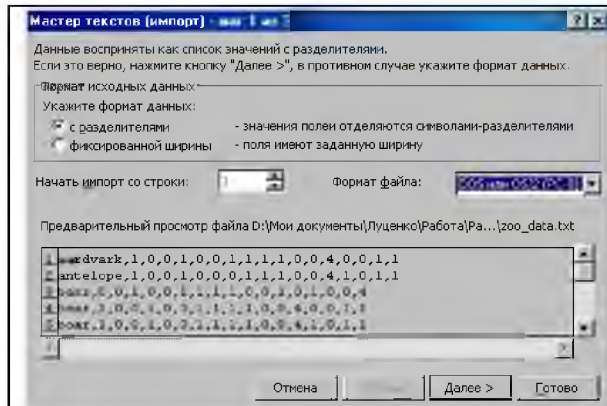


Рисунок 198. Excel-окно №1

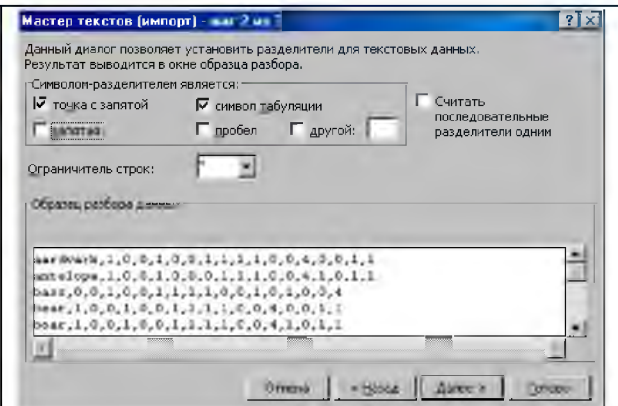


Рисунок 199. Excel-окно №2

- во 2-м окне задаем символ-разделитель "запятая" и нажимаем кнопку: "Далее", появляется 3-е окно (рисунок 200);
- нажимаем "Готово" и появляется 4-е окно (рисунок 201):

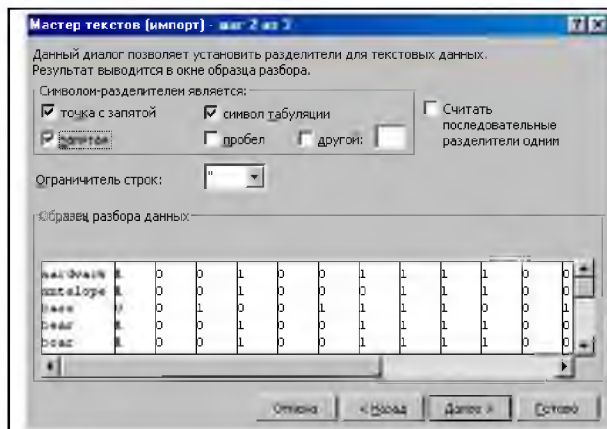


Рисунок 200. Excel-окно №3

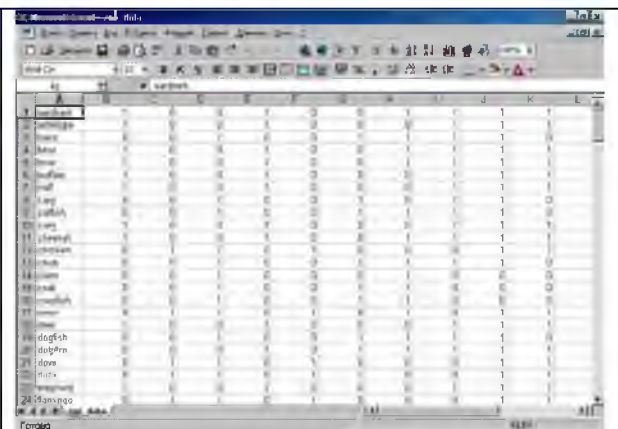


Рисунок 201. Excel-окно №4

4. Оформляем Excel-таблицу, представленную на рисунке 201, с использованием информации, содержащейся в общем описании задачи. В результате получаем таблицу (рисунок 202) и записываем ее в Excel-книге с именем: **zoo_data.xls**.

Microsoft Excel - zoo_data

Файл Правка Вид Вставка Формат Сервис Данные Окно ?

Закрывать

Arial Cyr 10

U39

		animal name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs {0-8}	tail	domestic	catsize	type {1-7}	kod animal name {1-101}
3																				
4	No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	1
5	1	aardvark	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1	1
6	2	antelope	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1	2
7	3	bass	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4	3
8	4	bear	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1	4
9	5	boar	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1	5
10	6	buffalo	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1	6
11	7	calf	1	0	0	1	0	0	0	1	1	1	0	0	4	1	1	1	1	7
12	8	carp	0	0	1	0	0	1	0	1	1	0	0	1	0	1	1	0	4	8
13	9	catfish	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4	9
14	10	cavy	1	0	0	1	0	0	0	1	1	1	0	0	4	0	1	0	1	10
15	11	cheetah	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1	11
16	12	chicken	0	1	1	0	1	0	0	0	1	1	0	0	2	1	1	0	2	12
17	13	chub	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4	13
18	14	clam	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	7	14
19	15	crab	0	0	1	0	0	1	1	0	0	0	0	0	4	0	0	0	7	15
20	16	crayfish	0	0	1	0	0	1	1	0	0	0	0	0	6	0	0	0	7	16

Готово

Рисунок 202. Вид фрагмента обучающей выборки в Excel

5. Создаем в Excel-книге `zoo_data.xls` еще один лист с именем `Kod`, в котором булевы значения, означающие "есть атрибут/нет атрибута", заменяются в описательных шкалах и градациях кодами атрибутов (рисунок 203):

Microsoft Excel - zoo_data

Файл Правка Вид Вставка Формат Сервис Данные Окно ?

Закрывать

Arial Cyr 10

O45 =zoo_data!O45+O\$4

		animal name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs {0-8}	tail	domestic	catsize	type {1-7}	kod animal name {1-101}	class kod type {1-7}	class kod name {1-101}
3	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	23	24	25	26	33	0	8
5	1	aardvark	2			5			8	9	10	11			18			25	26	33	1	8
6	2	antelope	2			5				9	10	11			18	23		25	26	34	1	9
7	3	bass			4		7	8	9	10				13	14	23		29	35	4	10	
8	4	bear	2			5			8	9	10	11			18			25	26	36	1	11
9	5	boar	2			5			8	9	10	11			18	23		25	26	37	1	12
10	6	buffalo	2			5				9	10	11			18	23		25	26	38	1	13
11	7	calf	2			5				9	10	11			18	23	24	25	26	39	1	14
12	8	carp			4		7			9	10			13	14	23	24	29	40	4	15	
13	9	catfish			4		7	8	9	10				13	14	23		29	41	4	16	
14	10	cavy	2			5				9	10	11			18		24	26	42	1	17	
15	11	cheetah	2			5			8	9	10	11			18	23	25	26	43	1	18	
16	12	chicken		3	4	6					10	11			16	23	24	27	44	2	19	
17	13	chub			4		7	8	9	10				13	14	23		29	45	4	20	
18	14	clam			4			8						14				32	46	7	21	
19	15	crab			4		7	8						18				32	47	7	22	
20	16	crayfish			4		7	8						20				32	48	7	23	

Готово

Рисунок 203. Лист кодов обучающей выборки в Excel



6. Создаем Excel-страницу для преобразования в DBF-стандарт: модифицируем формулу, заменяя пробелы (нет кода) нулями, с целью сохранения единого типа данных для столбцов таблицы, которые станут полями базы данных. Получаем таблицу, фрагмент которой приведен на рисунке 204.

		Attributes																	Class#		
	animal name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs (0-8)	tail	domestic	catsize	type (1-7)	kod animal name {1-101}	klass type {1-7}	klass kod name {1-101}
4	N 1	2	3	4	5	6	7	8	9	10	11	12	13	14	23	24	25	26	33	0	8
5	1 aardvark	2	0	0	5	0	0	8	9	10	11	0	0	18	0	0	25	26	33	1	8
6	2 antelope	2	0	0	5	0	0	0	9	10	11	0	0	18	23	0	25	26	34	1	9
7	3 bass	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0	0	29	35	4	10
8	4 bear	2	0	0	5	0	0	8	9	10	11	0	0	18	0	0	25	26	36	1	11
9	5 boar	2	0	0	5	0	0	8	9	10	11	0	0	18	23	0	25	26	37	1	12
10	6 buffalo	2	0	0	5	0	0	0	9	10	11	0	0	18	23	0	25	26	38	1	13
11	7 calf	2	0	0	5	0	0	0	9	10	11	0	0	18	23	24	25	26	39	1	14
12	8 carp	0	0	4	0	0	7	0	9	10	0	0	13	14	23	24	0	29	40	4	15
13	9 catfish	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0	0	29	41	4	16
14	10 cavy	2	0	0	5	0	0	0	9	10	11	0	0	18	0	24	0	26	42	1	17
15	11 cheetah	2	0	0	5	0	0	8	9	10	11	0	0	18	23	0	25	26	43	1	18
16	12 chicken	0	3	4	0	6	0	0	0	10	11	0	0	16	23	24	0	27	44	2	19
17	13 chub	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0	0	29	45	4	20
18	14 clam	0	0	4	0	0	0	8	0	0	0	0	0	14	0	0	0	32	46	7	21
19	15 crab	0	0	4	0	0	7	8	0	0	0	0	0	18	0	0	0	32	47	7	22
20	16 crayfish	0	0	4	0	0	7	8	0	0	0	0	0	20	0	0	0	32	48	7	23
21	17	0	2	4	0	6	0	8	0	10	11	0	0	16	22	0	0	27	49	2	24

Рисунок 204. Лист кодов обучающей выборки в Excel для преобразования в DBF-стандарт

Рисунок 204. Лист кодов обучающей выборки в Excel для преобразования в DBF-стандарт

7. Записываем лист "DBF" Excel-таблицы в DBF-стандарте:  
 – отмечаем блоком подтаблицу: A4:V105, в которой находятся коды атрибутов и классов;  
 – помещаем отмеченный блок в буфер обмена;  
 – выбираем режим "Файл – Сохранить как – Тип файла: DBF 4 (DBASE IV) – Сохранить".

8. Аналогично записываем в DBF-стандарте Excel-таблицу **Prizn.xls**.

В результате получаем промежуточные DBF-файлы с именами:

- **Zoo_data.dbf**: база данных с кодами признаков (атрибутов);
- **Prizn.dbf**: база данных с признаками (атрибутами).

Фрагменты этих баз данных, как они отображаются в просмотрщике, представлены на рисунках 205 и 206. Отметим, что эти промежуточные базы данных содержат всю необходимую и достаточную информацию в заданном стандарте для выполнения следующего – 2-го этапа работы.

nc - dbview  
dBASE View: D:\...1\zoo_data.dbf 1 / 101

N	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17
1	aardvark	2	0	0	5	0	0	8	9	10	11	0	0	18	0	0
2	antelope	2	0	0	5	0	0	0	9	10	11	0	0	18	23	0
3	bass	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0
4	bear	2	0	0	5	0	0	8	9	10	11	0	0	18	0	0
5	boar	2	0	0	5	0	0	8	9	10	11	0	0	18	23	0
6	buffalo	2	0	0	5	0	0	0	9	10	11	0	0	18	23	0
7	calf	2	0	0	5	0	0	0	9	10	11	0	0	18	23	24
8	carp	0	0	4	0	0	7	0	9	10	0	0	13	14	23	24
9	catfish	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0
10	cavy	2	0	0	5	0	0	0	9	10	11	0	0	18	0	24
11	cheetah	2	0	0	5	0	0	8	9	10	11	0	0	18	23	0
12	chicken	0	3	4	0	6	0	0	0	10	11	0	0	16	23	24
13	chub	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0
14	clam	0	0	4	0	0	0	8	0	0	0	0	0	14	0	0
15	crab	0	0	4	0	0	7	8	0	0	0	0	0	18	0	0
16	crayfish	0	0	4	0	0	7	8	0	0	0	0	0	20	0	0
17	crow	0	3	4	0	6	0	8	0	10	11	0	0	16	23	0
18	deer	2	0	0	5	0	0	0	9	10	11	0	0	18	23	0
19	dogfish	0	0	4	0	0	7	8	9	10	0	0	13	14	23	0
20	dolphin	0	0	0	5	0	7	8	9	10	11	0	13	14	23	0

Info 2 Fields 3 Record 5 Go to 6 Search 8 9 10 Quit

Рисунок 205. Фрагмент базы данных: Zoo_data.DBF

nc - dbview  
dBASE View: D:\...s\1\prizn.dbf 1 / 133

N1	N2	N3
1	animal name	Unique for each instance
2	hair	Boolean
3	feathers	Boolean
4	eggs	Boolean
5	milk	Boolean
6	airborne	Boolean
7	aquatic	Boolean
8	predator	Boolean
9	toothed	Boolean
10	backbone	Boolean
11	breathes	Boolean
12	venomous	Boolean
13	fins	Boolean
14	legs-0	Numeric (set of values: {0-8})
15	legs-1	
16	legs-2	
17	legs-3	
18	legs-4	
19	legs-5	
20	legs-6	

Info 2 Fields 3 Record 5 Go to 6 Search 8 9 10 Quit

Рисунок 206. Фрагмент базы данных: Prizn.DBF



### 3.2.4.3. Программный интерфейс для преобразования промежуточных DBF-файлов в базы данных системы "Эйдос"

Предлагается программный интерфейс, обеспечивающий автоматическое преобразование промежуточных DBF-файлов `Zoo_data.dbf` и `Prizn.dbf` в базы данных системы "Эйдос" (исходный текст на языке программирования – xBase приведен ниже):

```
*****
*** ФОРМИРОВАНИЕ КЛАССИФИКАЦИОННЫХ И ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ,
*** А ТАКЖЕ ОБУЧАЮЩЕЙ ВЫБОРКИ ИЗ DBF-Excel-файла РЕПОЗИТАРИЯ UCI ПО ЖИВОТНЫМ
*** http://ftp.ics.uci.edu/pub/machine-learning-databases/zoo/zoo.names
*** Луценко Е.В., 10/18/04 01:19pm *****
scr23 = SAVESCREEN(0,0,24,79)
SET CURSOR OFF
SET DATE ITALIAN
SET DECIMALS TO 15
SET ESCAPE On
FOR J=0 TO 24
    @J,0 SAY REPLICATE(" ",80) COLOR "rg+/N"
NEXT
SHOWTIME(0,60,.F.,"rg+/n",.F.,.F.)
Mess = " === ФОРМИРОВАНИЕ СПРАВОЧНИКОВ КЛАССИФИКАЦИОННЫХ ШКАЛ === "
@2,40-LEN(Mess)/2 SAY Mess COLOR "rg+/rb"
Vid = "Y"
@17,6 SAY "Включать в признаки коды наименования животного и его вида <Y/N>? #" COLOR "w+/rb"
*      012345678901234567890123456789012345678901234567890123456789012345678
*      0          10          20          30          40          50          60          70
@17,72 GET Vid PICTURE "X" COLOR "rg+/r"
SET CURSOR ON;READ;SET CURSOR OFF
IF Vid <> "Y" .AND. Vid <> "N"
    Vid = "N"
ENDIF
USE Object EXCLUSIVE NEW;ZAP
USE Zoo_data EXCLUSIVE NEW
ArObj := {}
AADD(ArObj,"МЛЕКОПИТАЮЩИЕ ")
AADD(ArObj,"ПТИЦЫ ")
AADD(ArObj,"ПРЕСМЫКАЮЩИЕСЯ?")
AADD(ArObj,"РЫБЫ ")
AADD(ArObj,"ЗЕМНОВОДНЫЕ ")
AADD(ArObj,"НАСЕКОМЫЕ ")
AADD(ArObj,"МНОГОНОГИЕ ")
SELECT Zoo_data
DBGOTOP()
DO WHILE .NOT. EOF()
    AADD(ArObj,FIELDGET(2))
    DBSKIP(1)
ENDDO
SELECT Object
DBGOTOP()
FOR j=1 TO LEN(ArObj)
    APPEND BLANK
    REPLACE Kod WITH j
    REPLACE Name WITH ArObj[j]
NEXT
CLOSE ALL
Mess = " ===== ФОРМИРОВАНИЕ СПРАВОЧНИКОВ ОПИСАТЕЛЬНЫХ ШКАЛ ===== "
USE Prizn EXCLUSIVE NEW
USE Priz_per EXCLUSIVE NEW;ZAP
ArPr := {}
SELECT Prizn
DBGOTOP()
DO WHILE .NOT. EOF()
    AADD(ArPr,FIELDGET(2))
    DBSKIP(1)
ENDDO
SELECT Priz_per
DBGOTOP()
```

```

FOR j=1 TO IF(Vid="Y",LEN(ArPr),25)
  APPEND BLANK
  REPLACE Kod WITH j
  REPLACE Name WITH ArPr[j]
NEXT
@24,0 SAY REPLICATE("-",80) COLOR "rb/n"
CLOSE ALL
Mess = " ===== ФОРМИРОВАНИЕ ОБУЧАЮЩЕЙ ВЫБОРКИ ===== "
CLOSE ALL
USE Zoo_data EXCLUSIVE NEW
USE ObInfZag EXCLUSIVE NEW;ZAP
USE ObInfKpr EXCLUSIVE NEW;ZAP
N_Rec = RECCOUNT()
DBGOTOP()
@24,0 SAY REPLICATE("-",80) COLOR "rb/n"
SELECT Zoo_data
DBGOTOP()
DO WHILE .NOT. EOF()
  ArObj := {}
  FOR j=1 TO 2
    AADD(ArObj,FIELDGET(j))
  NEXT
  FOR j=21 TO 22
    AADD(ArObj,FIELDGET(j))
  NEXT
  ArPr := {}
  FOR j=3 TO 20
    Mv = FIELDGET(j)
    IF Mv > 0
      IF Vid = "Y"
        AADD(ArPr,Mv)
      ELSE
        IF Mv <= 25
          AADD(ArPr,Mv)
        ENDIF
      ENDIF
    ENDIF
  NEXT
  ***** Запись массива кодов классов из БД Zoo_data в БД ObInfZag
  SELECT ObInfZag
  APPEND BLANK
  FOR j=1 TO LEN(ArObj)
    FIELDPUT(j,ArObj[j])
  NEXT
  ***** Запись массива кодов признаков из БД Zoo_data в БД ObInfKpr
  SELECT ObInfKpr
  APPEND BLANK
  FIELDPUT(1,ArObj[1])
  k=2
  FOR j=1 TO LEN(ArPr)
    IF k <= 12
      FIELDPUT(k++,ArPr[j])
    ELSE
      APPEND BLANK
      FIELDPUT(1,ArObj[1])
      k=2
      FIELDPUT(k,ArPr[j])
    ENDIF
  NEXT
  SELECT Zoo_data
  DBSKIP(1)
ENDDO
RESTSCREEN(0,0,24,79,scr23)
CLOSE ALL
QUIT

```

Программный интерфейс автоматически заполняет исходными данными следующие базы данных системы "Эйдос":

- **Object.dbf**: классы (классификационные шкалы и градации);
- **Priz_per.dbf**: атрибуты (описательные шкалы и градации);
- **ObInfZag.dbf**: обучающая выборка (главная база данных);
- **ObInfKpr.dbf**: обучающая выборка (связанная база данных).

В результате система "Эйдос" готова к синтезу семантической информационной модели и выполнению последующих этапов работ.

#### **3.2.4.4. Синтез семантической информационной модели**

Синтез СИМ состоит в расчете ряда баз данных, главной из которых является матрица информативностей. Этот синтез осуществляется на основе информации, содержащейся в файлах, перечисленных в предыдущем разделе. Для этих целей используется режим: "Обучение – Синтез семантической информационной модели – Автоматическое выполнение 1-2-3-4" (рисунок 207):

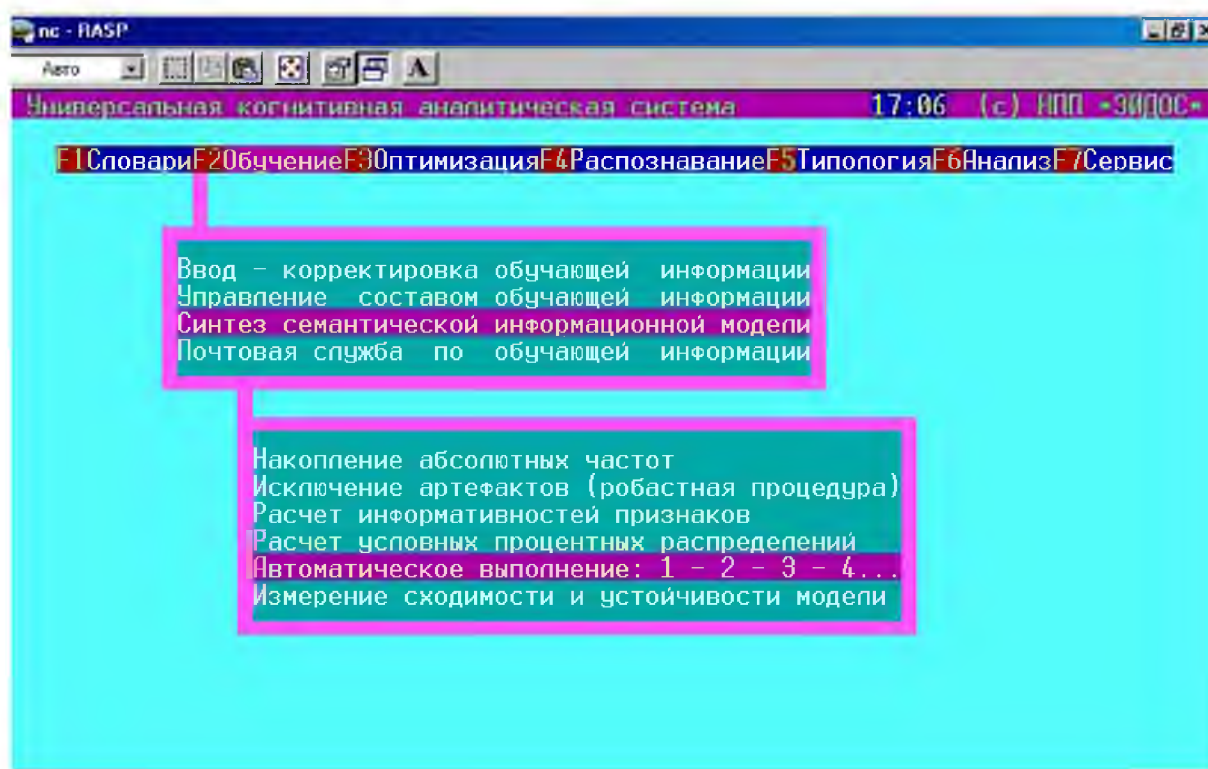


Рисунок 207. Режим: "Синтез СИМ"

#### **3.2.4.5. Оптимизация семантической информационной модели**

Оптимизация СИМ состоит в удалении из модели атрибутов с низкой селективной силой (по сути это Парето-оптимизация или ортонормирование) и осуществляется в режиме: "Исключение признаков с низкой селективной силой" подсистемы "Оптимизация" (рисунок 208).

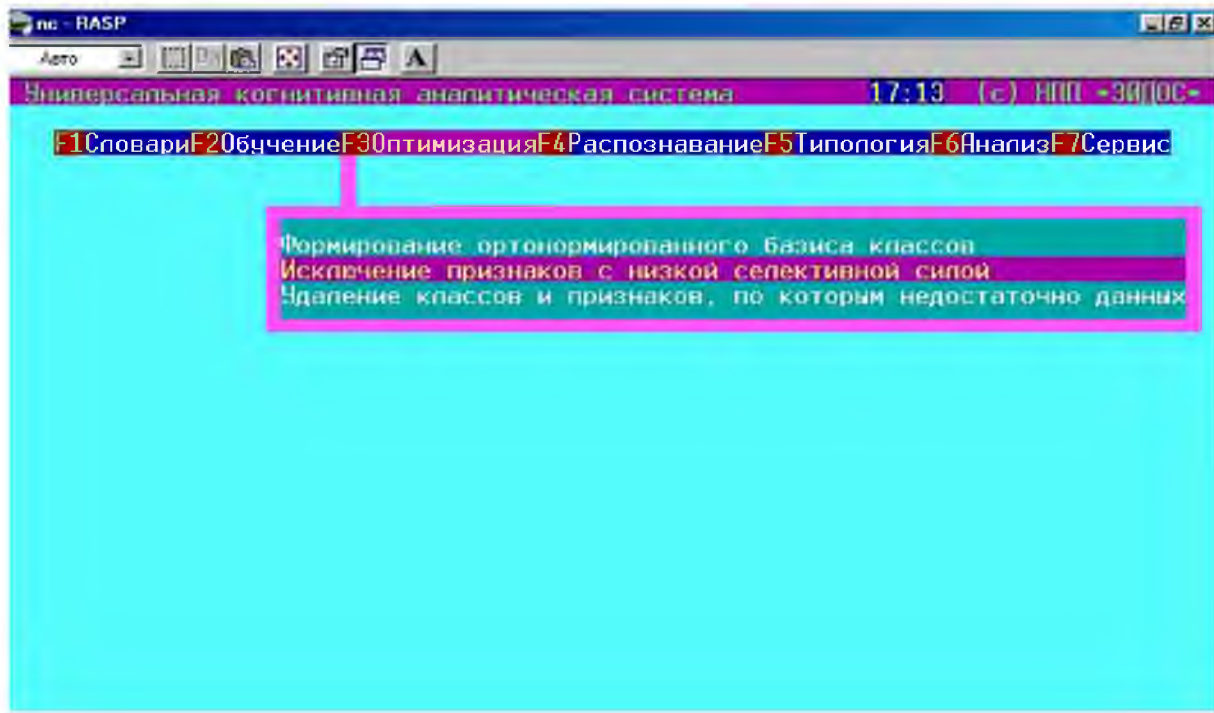


Рисунок 208. Подсистема: "Оптимизация семантической информационной модели"

В данном случае оптимизации СИМ не требуется, т.к. система признаков в модели содержит всего 25 булевых градаций, и все они, в основном, имеют достаточно высокую значимость, что видно из характерной формы Парето-диаграммы с отсутствием выраженной "полочки" (рисунок 209):

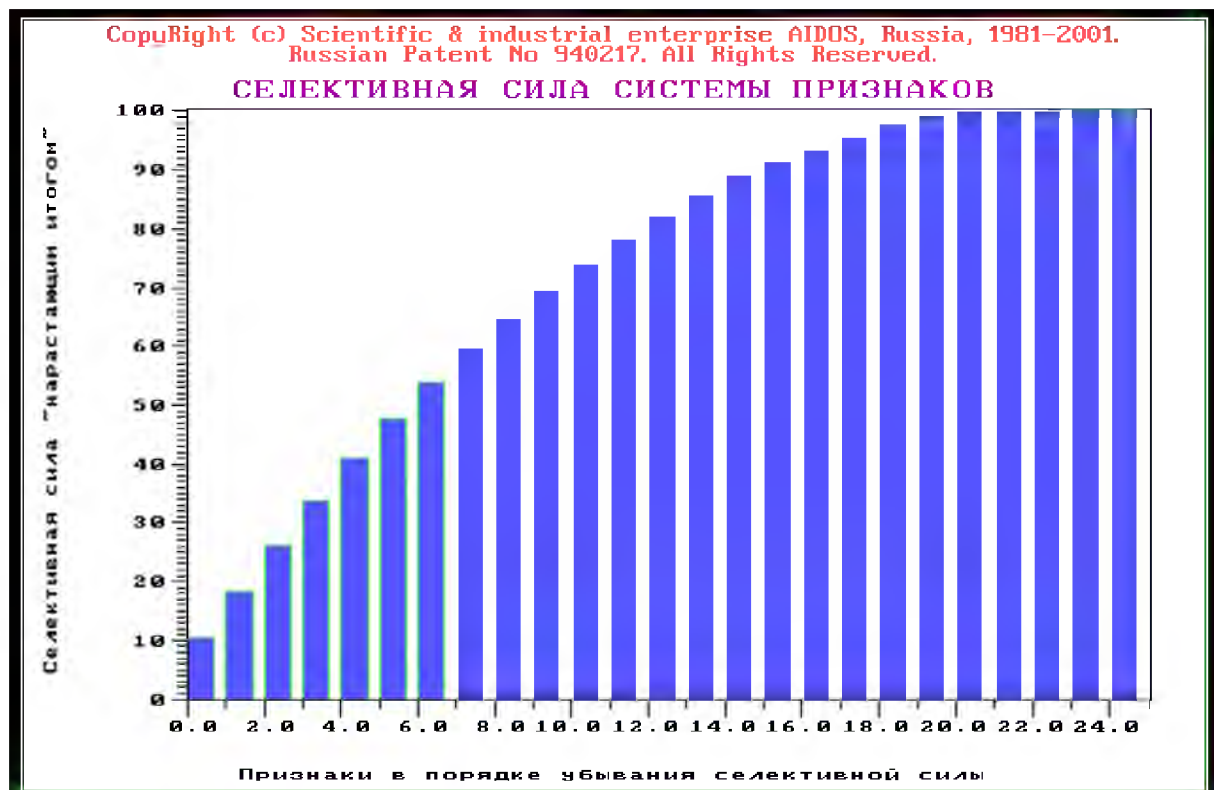


Рисунок 209. Парето-диаграмма атрибутов (накопительная, т.е. логистическая, кривая значимости атрибутов)

### 3.2.4.6. Проверка адекватности семантической информационной модели

Верификацию модели предлагается проверить путем расчета внутренней дифференциальной и интегральной валидности [81].

Необходимо отметить, что внутренняя валидность варианта семантической информационной модели, не учитывающей сделанные выше замечания к общему описанию задачи, составляет 100 %.

Для измерения валидности модели выполняются следующие действия:

1. Скопировать обучающую выборку в распознаваемую в подсистеме: "F2 Обучение – ввод корректировка обучающей информации – F5 Об.инф.->Расп.анк. – F2 Перезапись БД распознаваемых анкет – F1 Копировать всю БД".

2. Выполнить пакетное распознавание в подсистеме: "F4 Распознавание – Пакетное распознавание – Критерий сходства 1-й (корреляция)".

3. Измерить внутреннюю интегральную и дифференциальную валидность информационной модели в подсистеме: "F6 Анализ".

Результаты измерения внутренней валидности семантической информационной модели приведены в таблице 90:

**Таблица 90 – ИЗМЕРЕНИЕ ВАЛИДНОСТИ  
СЕМАНТИЧЕСКОЙ ИНФОРМАЦИОННОЙ МОДЕЛИ**

ВСЕГО анкет ( физических/логических ): 101/101 Распознано верно: 86  
Распознано ошибочно: 15 ВАЛИДНОСТЬ Системы распознавания: 85.1485%

21-10-04 17:25:14

г.Краснодар

N	Код	Наименования	Всего	Распозн	Распозн	Валидность
п/п	класса	к л а с с о в	анк.лог	верно	ошибочн	Системы (%)
1	2	ПТИЦЫ.....	20	20	0	100.0000
3	4	РЫБЫ.....	13	13	0	100.0000
5	5	ЗЕМНОВОДНЫЕ.....	4	4	0	100.0000
7	6	НАСЕКОМЫЕ.....	8	8	0	100.0000
9	1	МЛЕКОПИТАЮЩИЕ...	41	35	6	85.3659
11	3	ПРЕСМЫКАЮЩИЕСЯ?.	5	4	1	80.0000
13	7	МНОГОНОГИЕ.....	10	2	8	20.0000

Универсальная когнитивная аналитическая система

НПП *ЭЙДОС*

Обобщенные результаты распознавания представлены в таблице 91. Красным цветом и жирным шрифтом отмечены ошибочно идентифицированные объекты.

Таблица 91 – ИТОГОВЫЕ РЕЗУЛЬТАТЫ ИДЕНТИФИКАЦИИ

22-10-04

10:35:33

г. Краснодар

№ п/п	На-им. физ. источни-ка	Результаты идентификации														
		Идентифицирован как класс		Фактически является:		Уро-вень сход-ства %	Кач-во идент.	Ошибки по классам							Всего	
		Ко Д	Наименование	Ко Д	Наименование			1	2	3	4	5	6	7		
1	aardvark	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	45.153	58.327									0
2	antelope	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.860	72.096									0
3	bass	4	РЫБЫ	4	РЫБЫ	69.091	69.802									0
4	bear	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	45.153	58.327									0
5	boar	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0
6	buffalo	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.860	72.096									0
7	calf	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	50.381	79.729									0
8	carp	4	РЫБЫ	4	РЫБЫ	56.509	60.908									0
9	catfish	4	РЫБЫ	4	РЫБЫ	69.091	69.802									0
10	cavy	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	44.637	49.336									0
11	cheetah	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0
12	chicken	2	ПТИЦЫ	2	ПТИЦЫ	68.532	81.600									0
13	chub	4	РЫБЫ	4	РЫБЫ	69.091	69.802									0
14	clam	3	ПРЕСМЫКАЮЩИЕСЯ	7	МНОГОНОГИЕ	39.293	22.460								1	1
15	crab	5	ЗЕМНОВОДНЫЕ	7	МНОГОНОГИЕ	56.552	56.270								1	1
16	crayfish	5	ЗЕМНОВОДНЫЕ	7	МНОГОНОГИЕ	31.918	14.669								1	1
17	crow	2	ПТИЦЫ	2	ПТИЦЫ	61.940	73.682									0
18	deer	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.860	72.096									0
19	dogfish	4	РЫБЫ	4	РЫБЫ	57.215	71.018									0
20	dolphin	4	РЫБЫ	1	МЛЕКОПИТАЮЩИЕ	45.813	59.973	1								1
21	dove	2	ПТИЦЫ	2	ПТИЦЫ	68.532	81.600									0
22	duck	2	ПТИЦЫ	2	ПТИЦЫ	62.602	69.240									0
23	elephant	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.860	72.096									0
24	flamingo	2	ПТИЦЫ	2	ПТИЦЫ	58.945	78.981									0
25	flea	6	НАСЕКОМЫЕ.	6	НАСЕКОМЫЕ.	57.880	53.753									0
26	frog	5	ЗЕМНОВОДНЫЕ	5	ЗЕМНОВОДНЫЕ	56.991	58.285									0
27	frog	5	ЗЕМНОВОДНЫЕ	5	ЗЕМНОВОДНЫЕ	74.297	59.241									0
28	fruitbat	2	ПТИЦЫ	1	МЛЕКОПИТАЮЩИЕ	29.712	33.413	1								1
29	giraffe	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.860	72.096									0
30	girl	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	37.676	59.877									0
31	gnat	6	НАСЕКОМЫЕ.	6	НАСЕКОМЫЕ.	70.170	62.829									0
32	goat	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	50.381	79.729									0
33	gorilla	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	34.549	50.131									0
34	gull	2	ПТИЦЫ	2	ПТИЦЫ	53.136	62.059									0
35	haddock	4	РЫБЫ	4	РЫБЫ	72.286	74.190									0
36	hamster	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	45.203	61.201									0
37	hare	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	41.787	52.991									0
38	hawk	2	ПТИЦЫ	2	ПТИЦЫ	61.940	73.682									0
39	herring	4	РЫБЫ	4	РЫБЫ	69.091	69.802									0
40	honeybee	6	НАСЕКОМЫЕ.	6	НАСЕКОМЫЕ.	77.866	63.618									0
41	housefly	6	НАСЕКОМЫЕ.	6	НАСЕКОМЫЕ.	68.475	65.459									0
42	kiwi	2	ПТИЦЫ	2	ПТИЦЫ	45.650	56.201									0
43	ladybird	6	НАСЕКОМЫЕ.	6	НАСЕКОМЫЕ.	46.561	40.244									0
44	lark	2	ПТИЦЫ	2	ПТИЦЫ	72.585	82.063									0
45	leopard	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0
46	lion	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0

## Продолжение таблицы 91

№ п/п	На- им.физ. источни- ка	Результаты идентификации														
		Идентифицирован как класс		Фактически является:		Уро- вень сход- ства %	Кач- во идент .	Ошибки по классам							Всего	
		Ко д	Наименование	Ко д	Наименование			1	2	3	4	5	6	7		
47	lobster	5	ЗЕМНОВОДНЫЕ	7	МНОГОНОГИЕ	31.918	14.669								1	1
48	lynx	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0
49	mink	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	37.537	54.433									0
50	mole	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	40.754	51.203									0
51	mongoose	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0
52	moth	6	НАСЕКОМЫЕ.	6	НАСЕКОМЫЕ.	68.475	65.459									0
53	newt	5	ЗЕМНОВОДНЫЕ	5	ЗЕМНОВОДНЫЕ	31.652	34.274									0
54	octopus	7	МНОГОНОГИЕ	7	МНОГОНОГИЕ	32.232	37.177									0
55	opossum	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	40.754	51.203									0
56	oryx	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.860	72.096									0
57	ostrich	2	ПТИЦЫ	2	ПТИЦЫ	42.539	65.476									0
58	parakeet	2	ПТИЦЫ	2	ПТИЦЫ	68.532	81.600									0
59	penguin	2	ПТИЦЫ	2	ПТИЦЫ	25.030	42.429									0
60	pheasant	2	ПТИЦЫ	2	ПТИЦЫ	72.585	82.063									0
61	pike	4	РЫБЫ	4	РЫБЫ	57.215	71.018									0
62	piranha	4	РЫБЫ	4	РЫБЫ	69.091	69.802									0
63	pitviper	3	ПРЕСМЫКАЮЩИЕСЯ	3	ПРЕСМЫКАЮЩИЕСЯ	66.439	56.486									0
64	platypus	5	ЗЕМНОВОДНЫЕ	1	МЛЕКОПИТАЮЩИЕ	14.210	31.316	1								1
65	polecat	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0
66	pony	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	50.381	79.729									0
67	porpoise	4	РЫБЫ	1	МЛЕКОПИТАЮЩИЕ	45.813	59.973	1								1
68	puma	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0
69	pussycat	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	49.870	80.029									0
70	raccoon	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0
71	reindeer	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	50.381	79.729									0
72	rhea	2	ПТИЦЫ	2	ПТИЦЫ	33.019	56.339									0
73	scorpion	3	ПРЕСМЫКАЮЩИЕСЯ	7	МНОГОНОГИЕ	41.478	34.054								1	1
74	seahorse	4	РЫБЫ	4	РЫБЫ	72.286	74.190									0
75	seal	4	РЫБЫ	1	МЛЕКОПИТАЮЩИЕ	42.257	46.155	1								1
76	sealion	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	22.089	34.131									0
77	seasnake	3	ПРЕСМЫКАЮЩИЕСЯ	3	ПРЕСМЫКАЮЩИЕСЯ	52.601	33.745									0
78	seawasp	5	ЗЕМНОВОДНЫЕ	7	МНОГОНОГИЕ	51.925	13.781								1	1
79	skimmer	2	ПТИЦЫ	2	ПТИЦЫ	53.136	62.059									0
80	skua	2	ПТИЦЫ	2	ПТИЦЫ	53.136	62.059									0
81	slowworm	3	ПРЕСМЫКАЮЩИЕСЯ	3	ПРЕСМЫКАЮЩИЕСЯ	38.461	37.271									0
82	slug	4	РЫБЫ	7	МНОГОНОГИЕ	32.953	16.992								1	1
83	sole	4	РЫБЫ	4	РЫБЫ	72.286	74.190									0
84	sparrow	2	ПТИЦЫ	2	ПТИЦЫ	72.585	82.063									0
85	squirrel	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	29.979	36.725									0
86	starfish	7	МНОГОНОГИЕ	7	МНОГОНОГИЕ	48.217	40.232									0
87	stingray	4	РЫБЫ	4	РЫБЫ	51.509	40.028									0
88	swan	2	ПТИЦЫ	2	ПТИЦЫ	50.226	66.336									0
89	termite	6	НАСЕКОМЫЕ.	6	НАСЕКОМЫЕ.	57.880	53.753									0
90	toad	5	ЗЕМНОВОДНЫЕ	5	ЗЕМНОВОДНЫЕ	56.315	55.604									0
91	tortoise	5	ЗЕМНОВОДНЫЕ	3	ПРЕСМЫКАЮЩИЕСЯ	0.892	15.775			1						1
92	tuatara	3	ПРЕСМЫКАЮЩИЕСЯ	3	ПРЕСМЫКАЮЩИЕСЯ	20.568	22.593									0



Продолжение таблицы 91

№ п/п	На-им.физ. источни-ка	Результаты идентификации														
		Идентифицирован как класс		Фактически является:		Уро-вень сход-ства %	Кач-во идент	Ошибки по классам							Всего	
		Ко д	Наименование	Ко д	Наименование			1	2	3	4	5	6	7		
93	tuna	4	РЫБЫ	4	РЫБЫ	57.215	71.018									0
94	vampire	2	ПТИЦЫ	1	МЛЕКОПИТАЮЩИЕ	29.712	33.413	1								1
95	vole	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	41.787	52.991									0
96	vulture	2	ПТИЦЫ	2	ПТИЦЫ	49.581	73.319									0
97	wallaby	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	35.494	57.713									0
98	wasp	6	НАСЕКОМЫЕ.	6	НАСЕКОМЫЕ.	80.325	60.030									0
99	wolf	1	МЛЕКОПИТАЮЩИЕ	1	МЛЕКОПИТАЮЩИЕ	46.057	71.045									0
100	worm	4	РЫБЫ	7	МНОГОНОГИЕ	32.953	16.992								1	1
101	wren	2	ПТИЦЫ	2	ПТИЦЫ	72.585	82.063									0
ВСЕГО:								6	0	1	0	0	0	8	15	

Низкая достоверность идентификации класса многоногих, по-видимому, обусловлена тем, что в этот класс включены представители нескольких различных видов. Если их разделить на разные классы, то адекватность модели возрастет. Аналогичное предположение можно высказать по классу, который мы условно назвали "Пресмыкающиеся". Возможно, автор задачи разрабатывал ее с определенной долей иронии. Вместе с тем это никак не отражается на методике, предлагаемой в данном разделе.

Что касается класса "Млекопитающие", то, по-видимому, необходимо включить в модель дополнительные атрибуты, характерные именно для этого класса. Это следует из анализа результатов идентификации летучей мыши и дельфина (рисунки 210 и 211). Летучая мышь отнесена к птицам, а дельфин – к рыбам, т.к. по совокупности *использованных в модели* атрибутов они оказались наиболее похожими на обобщенные образы именно этих классов. Это подтверждают информационные портреты классов "Vampir" и "Dolphin", приведенные на рисунках 212 и 213. Вместе с тем необходимо обратить внимание на то, что в обоих случаях на втором месте по уровню сходства стоит правильный класс "Млекопитающие".



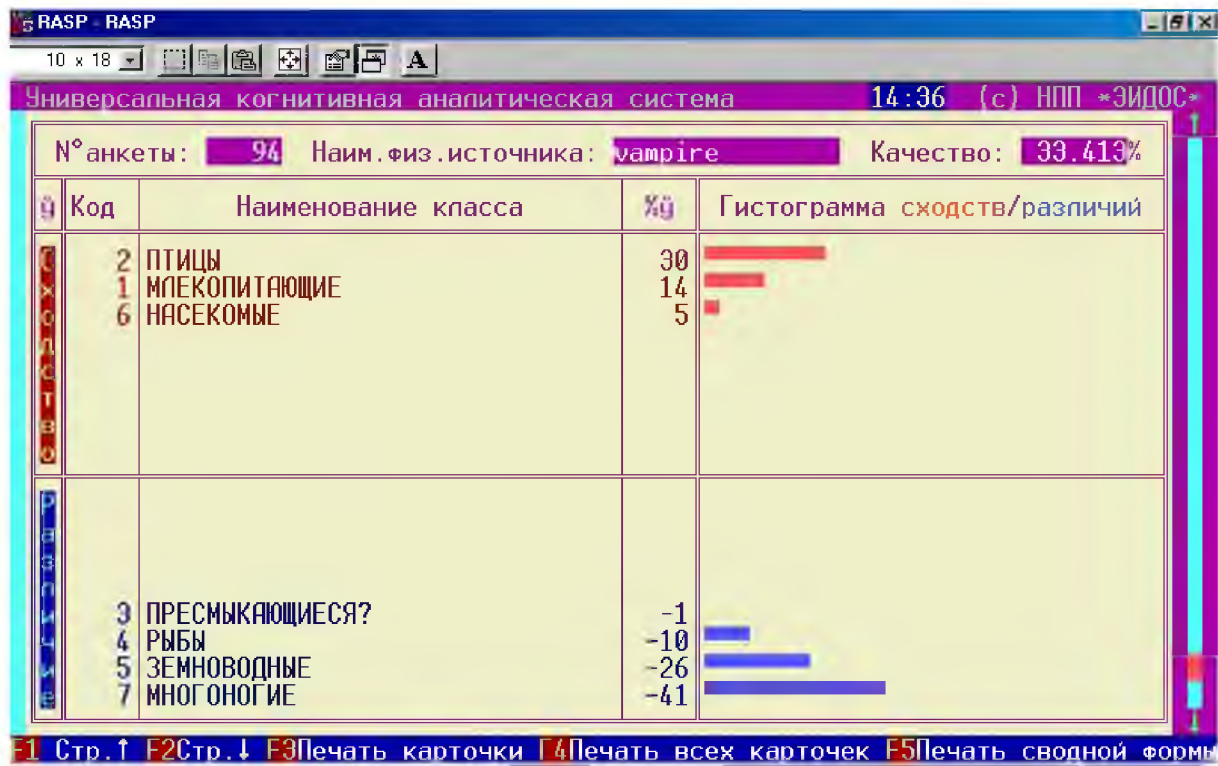


Рисунок 210. Результаты идентификации летучей мыши

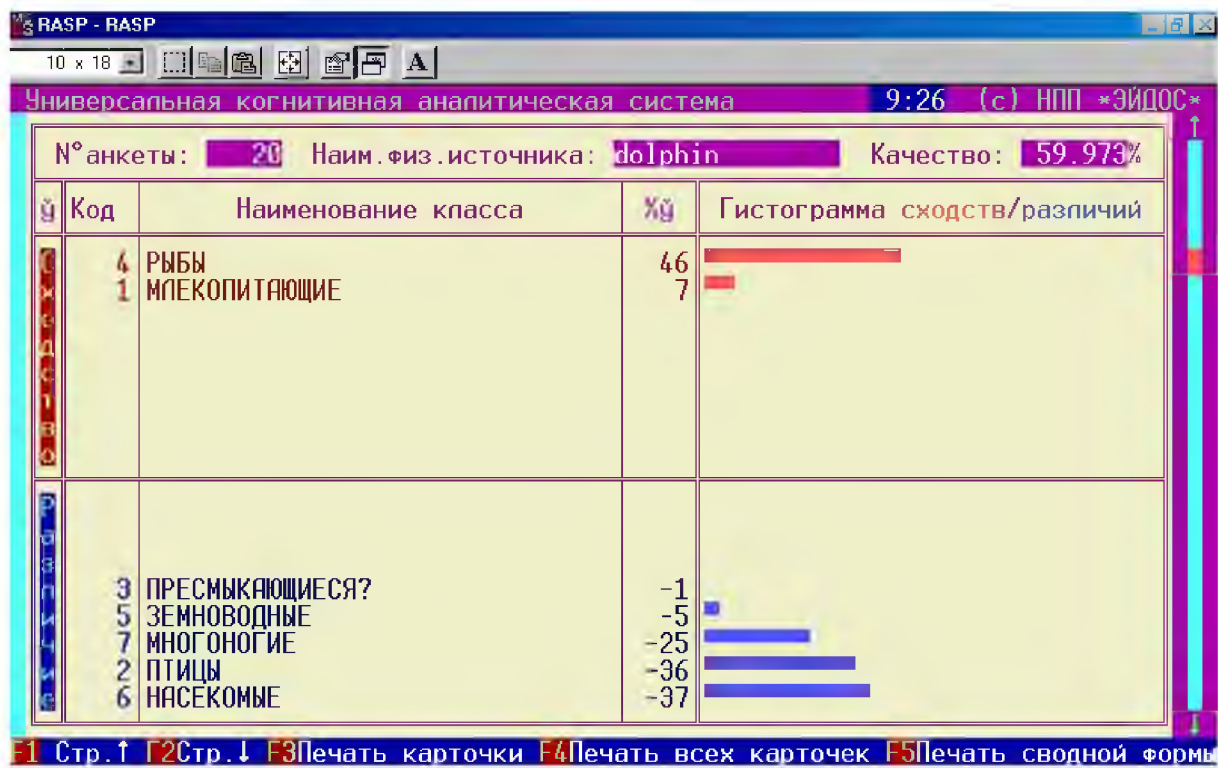
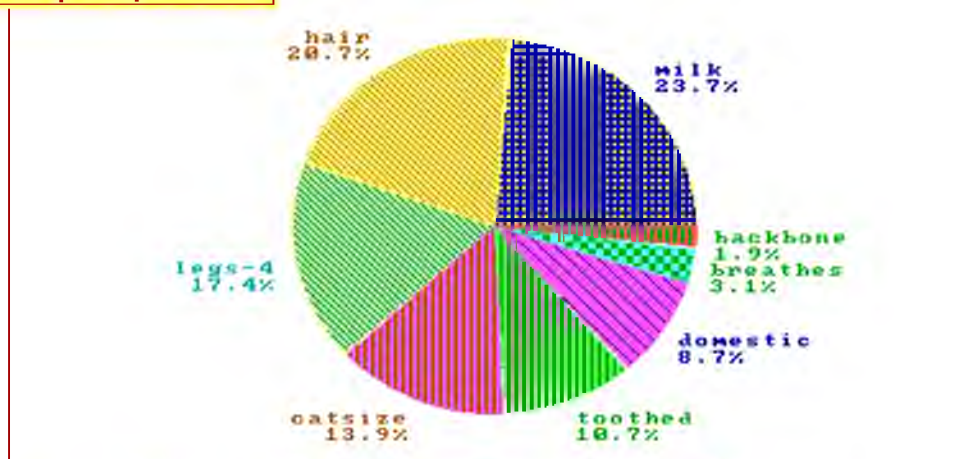


Рисунок 211. Результаты идентификации дельфина

Copyright (c) Scientific & industrial enterprise AIDOS, Russia, 1981-2001.  
Russian Patent No 940217. All Rights Reserved.

Информационный портрет класса распознавания:  
[1]-МЛЕКОПИТАЮЩИЕ ([0]-)

Фильтр: All, Positive



milk
  hair
  legs-4
  catsize
  toothed

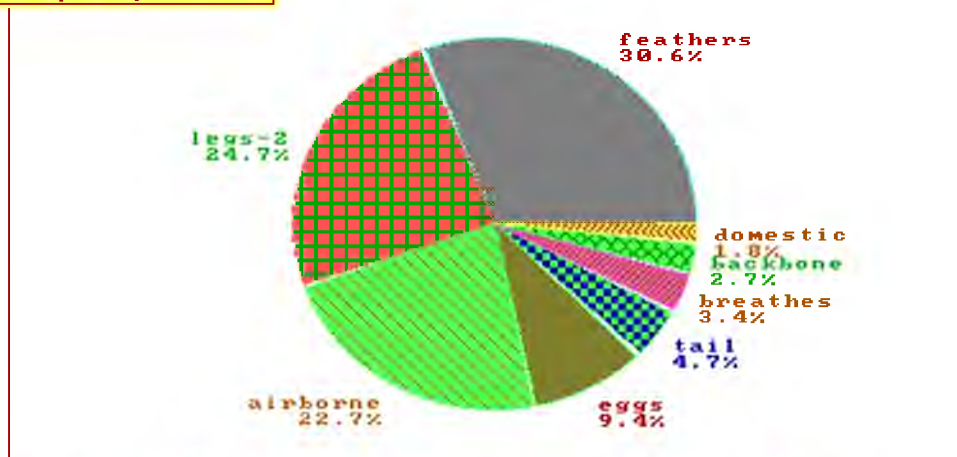
domestic
  breathes
  backbone

Рисунок 212. Информационный портрет класса: "Млекопитающие"

Copyright (c) Scientific & industrial enterprise AIDOS, Russia, 1981-2001.  
Russian Patent No 940217. All Rights Reserved.

Информационный портрет класса распознавания:  
[2]-ПТИЦЫ ([0]-)

Фильтр: All, Positive



feathers
  legs-2
  airborne
  eggs
  tail

breathes
  backbone
  domestic

Рисунок 213. Информационный портрет класса: "Птицы"

### 3.2.4.7. Анализ семантической информационной модели

Так как модель показала достаточно высокую степень адекватности, то исследование модели может корректным образом в определенных отношениях заменить изучение реального объекта (предметной области). Здесь мы не будем подробно освещать все возможности анализа модели, т.к. они подробно описаны в монографии [81]. Здесь мы ограничимся описанием лишь некоторых возможностей.

Результаты кластерно-конструктивного анализа классов представлены в графической форме семантической сети (рисунок 214).

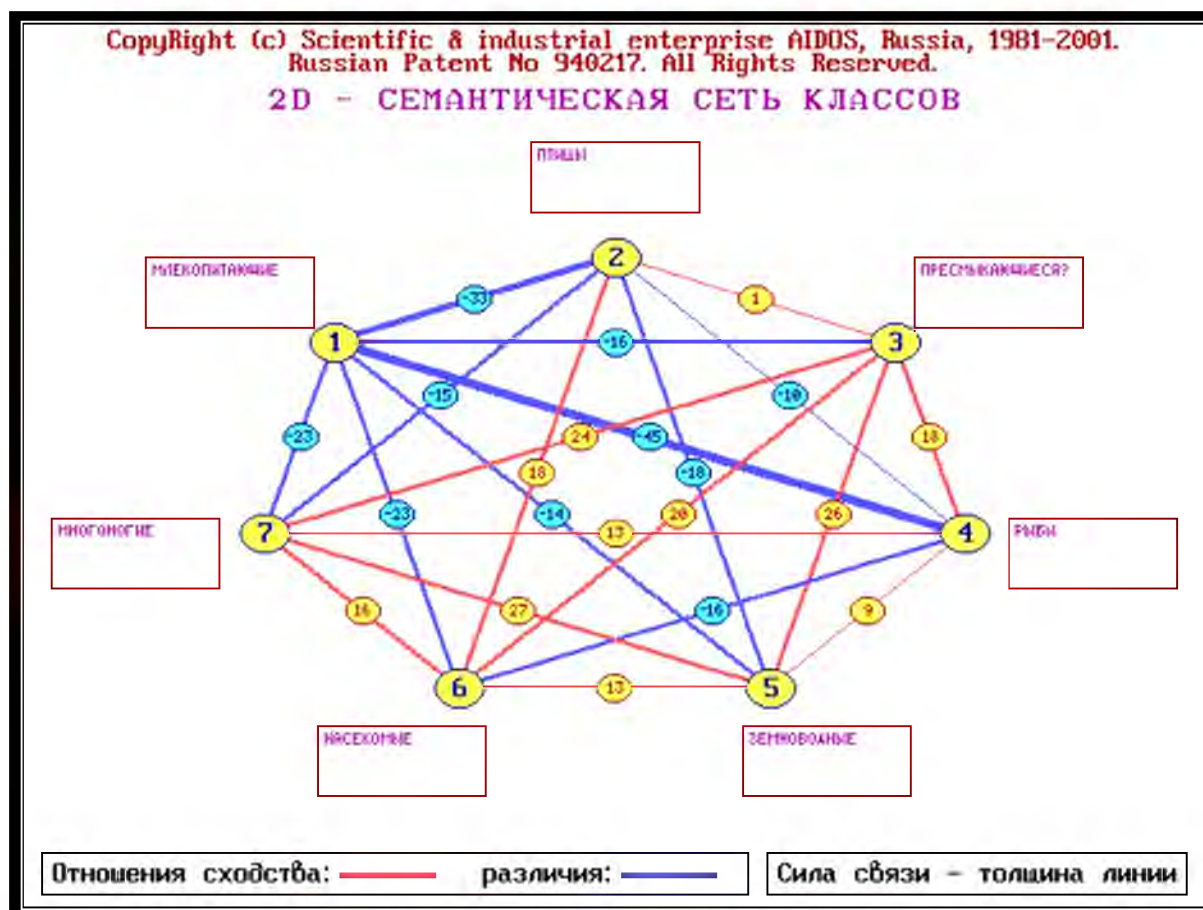


Рисунок 214. Семантическая сеть классов

Из этого рисунка видно, что млекопитающие в используемой системе признаков в наибольшей степени отличаются от рыб (конструкт: "Млекопитающие – рыбы", коды 1 и 4), а земноводные очень похожи на пресмыкающихся (кластер: коды 3 и 5).

Внутренняя структура любой линии на рисунке 214 может быть расшифрована и представлена в виде когнитивной диаграммы, одна из которых (в качестве примера) показана на рисунке 215.

Результаты кластерно-конструктивного анализа атрибутов приведены в графической форме семантической сети на рисунке 216.



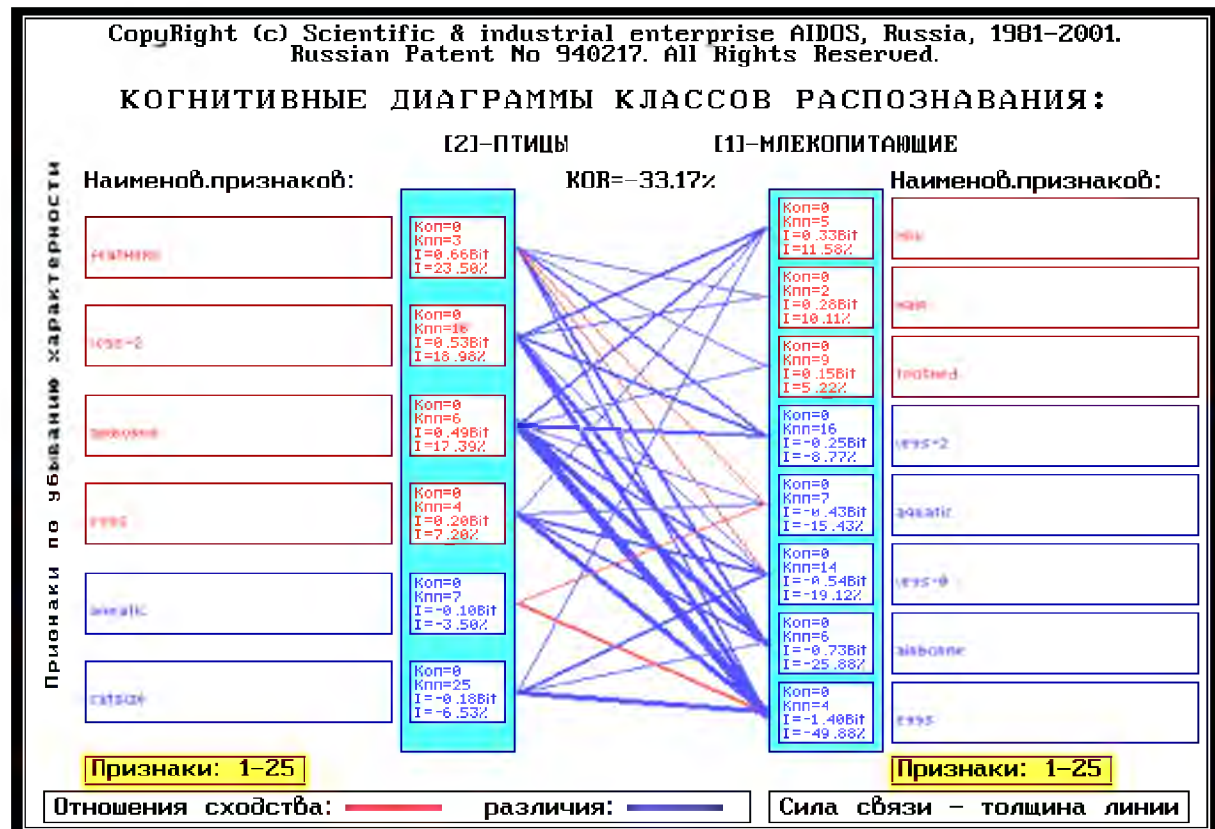


Рисунок 215. Расшифровка вклада атрибутов  
в сходство-различие классов: "Млекопитающие" и "Птицы"

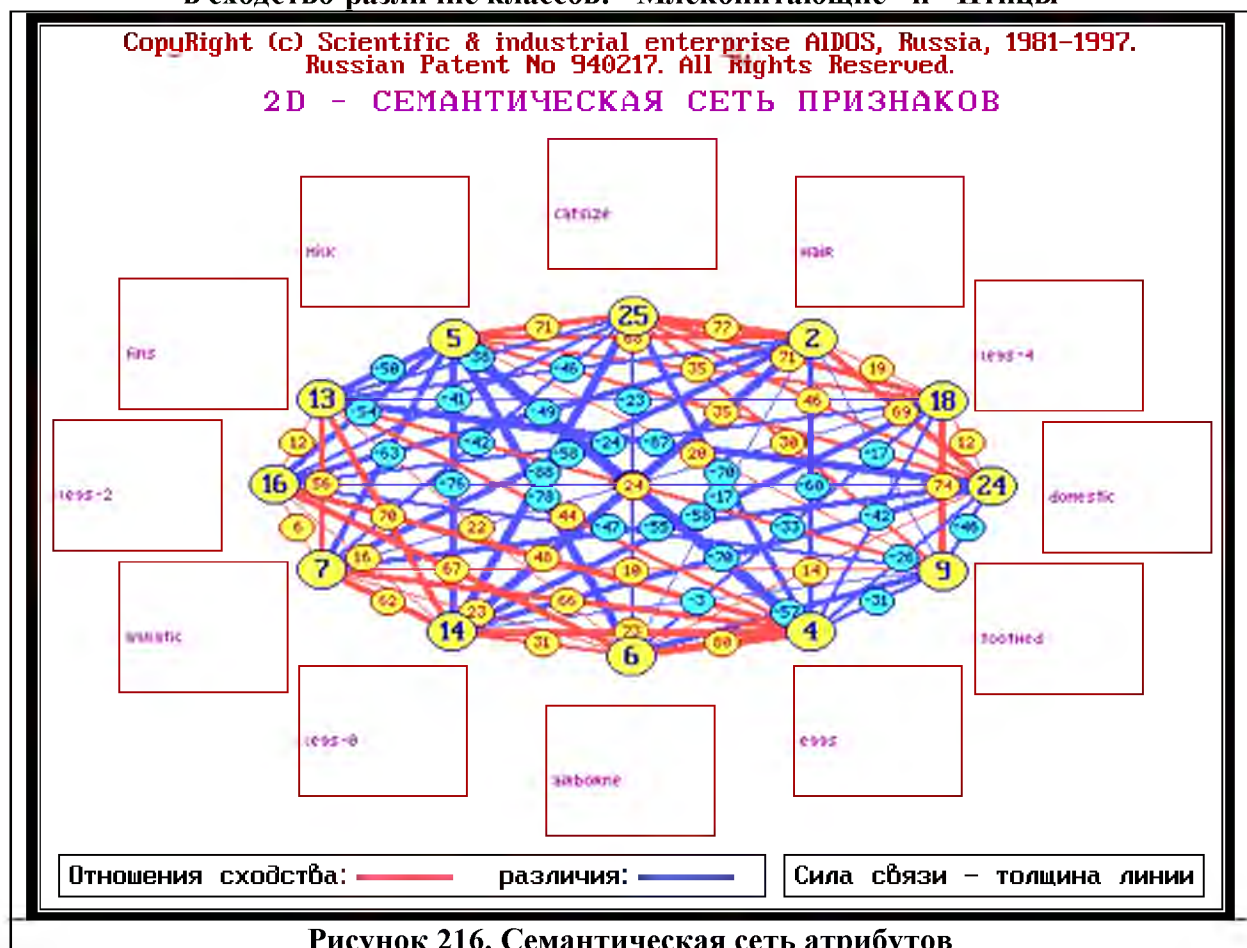


Рисунок 216. Семантическая сеть атрибутов

Из рисунка 216 видно, что атрибуты "Milk – eggs" (коды 5 и 4) образуют конструкт, как и, например, "наличие шерсти и отсутствие ног" (коды 2 и 14).

### **3.2.4.4. Резюме**

Таким образом, можно сделать вывод о том, что приведенная методика обеспечивает решение поставленной задачи, т.е. оценку качества математических моделей систем искусственного интеллекта путем использования баз данных репозитария UCI.

Аналогичный подход может быть эффективен и в случае использования других источников исходных данных (не UCI), а также других систем искусственного интеллекта (не системы "Эйдос"). Репозитарий UCI, семантическая информационная модель и реализующая ее универсальная когнитивная аналитическая система "Эйдос" рассмотрены здесь лишь в качестве примеров для демонстрации основных элементов предлагаемой типовой методики. В случае необходимости данная методика может быть развита или адаптирована для других случаев.

### **Список литературы**

1. Lutsenko E.V. Conceptual principles of the sistem (emergent) information theory & its application for the cognitive modelling of the active objects (entities) //2002 IEEE International Conference on Artificial Intelligence System (ICAIS 2002).—Computer society, IEEE, Los Alamos, California, Washington-Brussels-Tokyo, p.268–269.
2. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). — Краснодар: КубГАУ. 2002. —605 с.

### 3.3. ТЕОРЕТИЧЕСКИЕ ВОПРОСЫ

#### 3.3.1. Вопросы, выносящиеся на экзамен по дисциплине

На экзамен выносятся вопросы, приведенные как контрольные вопросы к лекциям и лабораторным работам.

##### ***Контрольные вопросы по лекции 1***

1. Основные положения информационно-функциональной теории развития техники.
2. Процессы труда и познания, как информационные процессы снятия неопределенности.
3. Организм человека и средства труда как информационные системы.
4. Законы развития техники.
5. Детерминация формы сознания человека функциональным уровнем средств труда.
6. Неизбежность возникновения компьютеров, информационных систем и систем искусственного интеллекта.
7. Информационная теория стоимости.
8. Связь количества и качества информации с меновой и потребительной стоимостью.
9. Информация, как сырье и как товар: абсолютная, относительная и аналитическая информация. Данные, информация, знания.
10. Стоимость и амортизация систем искусственного интеллекта и баз знаний.
11. Источники экономической эффективности систем искусственного интеллекта и интеллектуальной обработки данных с позиций информационной теории стоимости (повышение уровня системности и "охлаждение" объекта управления).
12. Интеллектуализация - генеральное направление и развития информационных технологий.
13. От электронных вычислительных машин к компьютерам. Функциональное определение компьютера.
14. Эволюция понятия: "Обработка информации" от информационного сырья к информационному продукту.
15. Эволюция технологий создания и поддержки информационных систем: автоматизация функций посредников.
16. Перспективы информационных технологий: интеллектуализация, создание самообучающихся, саморазвивающихся (эволюционирующих) и самовоспроизводящихся систем.

##### ***Контрольные вопросы по лекции 2***

1. Данные, информация, знания. Системно-когнитивный анализ как развитие концепции смысла Шенка-Абельсона.

2. Когнитивная концепция СК-анализа и синтез когнитивного конфигуратора.

3. Мышление как вычисление смысла и реализация операций со смыслом в инструментарии СК-анализа - системе "Эйдос".

4. Понятие: "Система искусственного интеллекта", место СИИ в классификации информационных систем.

5. Определение и классификация систем искусственного интеллекта, цели и пути их создания.

6. Тест Тьюринга и критерии "интеллектуальности" информационных систем. Может ли машина мыслить? Может ли искусственный интеллект превзойти своего создателя?

7. Классификация систем искусственного интеллекта.

8. Особенности технологии создания систем искусственного интеллекта (обучение, "социализация", как технологический этап).

9. Информационная модель деятельности специалиста и место систем искусственного интеллекта в этой деятельности.

10. Жизненный цикл системы искусственного интеллекта и критерии перехода между этапами этого цикла.

### ***Контрольные вопросы по лекции 3***

1. Системный анализ, как метод познания.

2. Принципы системного анализа.

3. Методы и этапы системного анализа.

4. Этапы когнитивного анализа.

5. Обобщенная схема системного анализа, ориентированного на интеграцию с когнитивными технологиями.

6. Когнитивная концепция и синтез когнитивного конфигуратора.

7. Понятие когнитивного конфигуратора и необходимость естественно-научной (формализуемой) когнитивной концепции.

8. Формализуемая когнитивная концепция

9. Когнитивный конфигуратор и базовые когнитивные операции системного анализа.

10. Задачи формализации базовых когнитивных операций системного анализа.

11. СК-анализ, как системный анализ, структурированный до уровня базовых когнитивных операций.

12. Место и роль СК-анализа в структуре управления.

13. Структура типовой АСУ.

14. Параметрическая модель адаптивной АСУ сложными системами.

15. Модель рефлексивной АСУ активными объектами и понятие мета-управления.

16. Двухконтурная модель РАСУ в АПК.

### ***Контрольные вопросы по лекции 4***

1. Предпосылки и теоретические основы системной теории информации.

2. Требования к математической модели и численной мере СТИ.

3. Выбор базовой численной меры СТИ.

4. Конструирование системной численной меры на основе базовой в СТИ.

5. Семантическая информационная модель СК-анализа.
6. Формализм динамики взаимодействующих семантических информационных пространств. Двухвекторное представление данных.
7. Применение классической теории информации К.Шеннона для расчета весовых коэффициентов и мер сходства.
8. Математическая модель метода распознавания образов и принятия решений, основанного на системной теории информации.
9. Некоторые свойства математической модели СК-анализа (сходимость, адекватность, устойчивость и др.).
10. Непараметричность модели. Робастные процедуры и фильтры для исключения артефактов в математической модели СК-анализа.
11. Зависимость информативностей факторов от объема обучающей выборки.
12. Зависимость адекватности семантической информационной модели от объема обучающей выборки (адекватность при малых и больших выборках).
13. Семантическая устойчивость модели СК-анализа.
14. Зависимость некоторых параметров модели СК-анализа от ее ортонормированности.
15. Взаимосвязь математической модели СК-анализа с другими моделями.
16. Взаимосвязь системной меры целесообразности информации со статистикой  $X^2$  и новая мера уровня системности предметной области.
17. Сравнение, идентификация и прогнозирование как разложение векторов объектов в ряд по векторам классов (объектный анализ).
18. Системно-когнитивный и факторный анализ. СК-анализ, как метод вариабельных контрольных групп.
19. Семантическая мера целесообразности информации и эластичность.
20. Связь семантической информационной модели с нейронными сетями.
21. Математический метод СК-анализа в свете идей интервальной бутстрепной робастной статистики объектов нечисловой природы.

### ***Контрольные вопросы по лекции 5***

1. Принципы формализации предметной области и подготовки эмпирических данных.
2. Иерархическая структура данных и последовательность численных расчетов в СК-анализе
3. Обобщенное описание алгоритмов СК-анализа
4. БКОСА-2.1. "Восприятие и запоминание исходной обучающей информации".
5. БКОСА-2.2. "Репрезентация. Сопоставление индивидуального опыта с коллективным (общественным)".
6. БКОСА-3.1.1. "Обобщение (синтез, индукция). Накопление первичных данных".
7. БКОСА-3.1.2. "Обобщение (синтез, индукция). Исключение артефактов".
8. БКОСА-3.1.3. "Обобщение (синтез, индукция). Расчет степени истинности содержательных смысловых связей между предпосылками и результатами (обобщенных таблиц решений)".



9. БКОСА-3.2. "Определение значимости шкал и градаций факторов, уровней Мерлина".

10. БКОСА-3.3. "Определение значимости шкал и градаций классов, уровней Мерлина".

11. БКОСА-4.1. "Абстрагирование факторов (снижение размерности семантического пространства факторов)".

12. БКОСА-4.2. "Абстрагирование классов (снижение размерности семантического пространства классов)".

13. БКОСА-5. "Оценка адекватности информационной модели предметной области".

14. БКОСА-7. "Сравнение, идентификация и прогнозирование. Распознавание состояний конкретных объектов (объектный анализ)".

15. БКОСА-9.1. "Дедукция и абдукция классов (семантический анализ обобщенных образов классов, решение обратной задачи прогнозирования)".

16. БКОСА-9.2. "Дедукция и абдукция факторов (семантический анализ факторов)".

17. БКОСА-10.1.1. "Классификация обобщенных образов классов".

18. БКОСА-10.1.2. "Формирование бинарных конструкторов классов".

19. БКОСА-10.1.3. "Визуализация семантических сетей классов".

20. БКОСА-10.2.1. "Классификация факторов".

21. БКОСА-10.2.2. "Формирование бинарных конструкторов факторов".

22. БКОСА-10.2.3. "Визуализация семантических сетей факторов".

23. БКОСА-10.3.1. "Содержательное сравнение классов".

25. БКОСА-10.3.2. "Расчет и отображение многозначных когнитивных диаграмм, в т.ч. диаграмм Вольфа Мерлина".

26. БКОСА-10.4.1. "Содержательное сравнение факторов".

27. БКОСА-10.4.2. "Расчет и отображение многозначных когнитивных диаграмм, в т.ч. инвертированных диаграмм Мерлина".

28. БКОСА-11. "Многовариантное планирование и принятие решения о применении системы управляющих факторов".

29. Детальные алгоритмы СК-анализа.

### ***Контрольные вопросы по лекции 6***

1. Назначение и состав системы "Эйдос".

2. Цели и основные функции системы "Эйдос".

3. Обобщенная структура системы "Эйдос".

4. Пользовательский интерфейс, технология разработки и эксплуатации приложений в системе "ЭЙДОС".

5. Начальный этап синтеза модели: когнитивная структуризация и формализация предметной области, подготовка исходных данных (подсистема "Словари") (БКОСА-1, БКОСА-2).

6. Синтез модели: пакетное обучение системы распознавания (подсистема "Обучение") (БКОСА-3).

7. Оптимизация модели (подсистема "Оптимизация") (БКОСА-4).

8. Верификация модели (оценка ее адекватности) (БКОСА-5).

9. Эксплуатация приложения в режиме адаптации и периодического синтеза модели (БКОСА-7, БКОСА-9, БКОСА-10).

10. Технические характеристики и обеспечение эксплуатации системы "ЭЙДОС" (версии 12.5).

11. Состав системы "Эйдос": Базовая система, системы окружения и программные интерфейсы импорта данных.

12. Отличия системы "Эйдос" от аналогов: экспертных и статистических систем.

13. Некоторые количественные характеристики системы "Эйдос".

14. Обеспечение эксплуатации системы "Эйдос".

15. АСК-анализ, как технология синтеза и эксплуатации рефлексивных АСУ активными объектами.

### ***Контрольные вопросы по лекции 7***

1. Интеллектуальные интерфейсы. Использование биометрической информации о пользователе в управлении системами.

2. Идентификация и аутентификация личности по почерку. Понятие клавиатурного почерка.

3. Соотношение психографологии и атрибуции текстов.

4. Идентификация и аутентификация личности пользователя компьютера по клавиатурному почерку.

5. Прогнозирование ошибок оператора по изменениям в его электроэнцефалограмме.

6. Системы с биологической обратной связью (БОС).

7. Мониторинг состояния сотрудников сборочного конвейера с целью обеспечения высокого качества продукции.

8. Компьютерные тренажеры, основанные на БОС, для обучения больных навыкам управления своим состоянием.

9. Компьютерные игры с БОС.

10. Системы с семантическим резонансом. Компьютерные (Ψ-технологии и интеллектуальный подсознательный интерфейс.

11. Системы виртуальной реальности и критерии реальности. Эффекты присутствия, деперсонализации и модификация сознания пользователя.

12. Классическое определение системы виртуальной реальности.

13. "Эффект присутствия" в виртуальной реальности.

14. Применения систем виртуальной реальности.

15. Модификация сознания и самосознания пользователя в виртуальной реальности.

16. Авторское определение системы виртуальной реальности.

17. Критерии реальности при различных формах сознания и их применение в виртуальной реальности.

18. Принципы эквивалентности (относительности) Галилея и Эйнштейна и критерии виртуальной реальности.

19. Системы с дистанционным телекинетическим интерфейсом.

### ***Контрольные вопросы по лекции 8***

1. Основные понятия и определения, связанные с системами распознавания образов.

2. Признаки и образы конкретных объектов, метафора фазового пространства.

3. Признаки и обобщенные образы классов.
4. Обучающая выборка и ее репрезентативность по отношению к генеральной совокупности. Ремонт (взвешивание) данных.
5. Основные операции: обобщение и распознавание.
6. Обучение с учителем (экспертом) и самообучение (кластерный анализ).
7. Верификация, адаптация и синтез модели.
8. Проблема распознавания образов.
9. Классификация методов распознавания образов.
10. Применение распознавания образов для идентификации и прогнозирования. Сходство и различие в содержании понятий "идентификация" и "прогнозирование".
11. Роль и место распознавания образов в автоматизации управления сложными системами.
12. Обобщенная структура системы управления.
13. Место системы идентификации в системе управления.
14. Управление как задача, обратная идентификации и прогнозированию.
15. Методы кластерного анализа.
16. Метод кластеризации: "Древовидная кластеризация".
17. Метод кластеризации: "Двухходовое объединение".
18. Метод кластеризации: "Метод К средних".

### ***Контрольные вопросы по лекции 9***

1. Многообразие задач принятия решений.
2. Принятие решений, как реализация цели.
3. Принятие решений, как снятие неопределенности (информационный подход).
4. Связь принятия решений и распознавания образов.
5. Классификация задач принятия решений.
6. Языки описания методов принятия решений.
7. Критериальный язык.
8. Язык последовательного бинарного выбора.
9. Обобщенный язык функций выбора.
10. Групповой выбор.
11. Выбор в условиях неопределенности.
12. Информационная (статистическая) неопределенность в исходных данных.
13. Неопределенность последствий.
14. Расплывчатая неопределенность.
15. Решение как компромисс и баланс различных интересов. О некоторых ограничениях оптимизационного подхода.
16. Экспертные методы выбора.
17. Юридическая ответственность за решения, принятые с применением систем поддержки принятия решений.
18. Условия корректности использования систем поддержки принятия решений.
19. Хранилища данных для принятия решений.

### ***Контрольные вопросы по лекции 10***

1. Экспертные системы, базовые понятия.
2. Экспертные системы, методика построения.
3. Этап-1 синтеза ЭС: "Идентификация".
4. Этап-2 синтеза ЭС: "Концептуализация".
5. Этап-3 синтеза ЭС: "Формализация".
6. Этап-4 синтеза ЭС: "Разработка прототипа".
7. Этап-5 синтеза ЭС: "Экспериментальная эксплуатация".
8. Этап-6 синтеза ЭС: "Разработка продукта".
9. Этап-7 синтеза ЭС: "Промышленная эксплуатация".

### ***Контрольные вопросы по лекции 11***

1. Биологический нейрон и формальная модель нейрона Маккалоки и Питтса.
2. Возможность решения простых задач классификации непосредственно одним нейроном.
3. Однослойная нейронная сеть и персептрон Розенблата.
4. Линейная разделимость и персептронная представляемость.
5. Многослойные нейронные сети.
6. Многослойный персептрон.
7. Модель Хопфилда.
8. Когнитрон и неоконитрон Фукушимы.
9. Проблемы и перспективы нейронных сетей.
10. Модель нелокального нейрона и нелокальные интерпретируемые нейронные сети прямого счета.
11. Метафора нейросетевого представления семантической информационной модели.
12. Соответствие основных терминов и понятий.
13. Гипотеза о нелокальности нейрона и информационная нейросетевая парадигма.
14. Решение проблемы интерпретируемости весовых коэффициентов (семантическая мера целесообразности информации и закон Фехнера).
15. Семантическая информационная модель, как нелокальная нейронная сеть.
16. Гипотеза о физической природе нелокального взаимодействия нейронов в нелокальной нейронной сети.
17. Решение проблемы интерпретируемости передаточной функции.
18. Решение проблемы размерности.
19. Решение проблемы линейной разделимости.
20. Моделирование причинно-следственных цепочек в нейронных сетях и семантической информационной модели.
21. Моделирование иерархических структур обработки информации.
22. Нейронные сети и СК-анализ.
23. Графическое отображение нейронов, Парето-подмножеств нелокальной нейронной сети, семантических сетей, когнитивных карт и диаграмм в системе "Эйдос".

### ***Контрольные вопросы по лекции 12***

1. Основные понятия, принципы и предпосылки генетических алгоритмов.
2. Пример работы простого генетического алгоритма.
3. Достоинства и недостатки генетических алгоритмов.
4. Примеры применения генетических алгоритмов.

### ***Контрольные вопросы по лекции 13***

1. Определение основных понятий: "Когнитивное моделирование" и "Классическая когнитивная карта", их связь с когнитивной психологией и гносеологией.
2. Когнитивная (познавательльно-целевая) структуризация знаний об исследуемом объекте и внешней для него среды на основе PEST-анализа и SWOT-анализа.
3. Разработка программы реализации стратегии развития исследуемого объекта на основе динамического имитационного моделирования (при поддержке программного пакета Ithink).

### ***Контрольные вопросы по лекции 14***

1. Интеллектуальный анализ данных (data mining).
2. Типы выявляемых закономерностей data mining: ассоциация, последовательность, классификация, кластеризация, прогнозирование.
3. Математический аппарат data mining: деревья решений, регрессионный анализ, нейронные сети, временные ряды.
4. Области применения технологий интеллектуального анализа данных: розничная торговля, банковская деятельность, страховой бизнес, производство, автоматизированные системы для интеллектуального анализа данных.

### ***Контрольные вопросы по лекции 15***

1. Обзор опыта применения АСК-анализа для управления и исследования социально-экономических систем.
2. Применение СИИ для синтеза и решения задач управления качеством подготовки специалистов.
3. Применение СИИ в экономических исследованиях.
4. Применение СИИ в социологических и политологических исследованиях.
5. Поддержка принятия решений по выбору агротехнологий, культур и пунктов выращивания.
6. Поддержка принятия решений по выбору агротехнологий.
7. Поддержка принятия решений по выбору культур и пунктов выращивания.
8. Применение СИИ для прогнозирования динамики сегмента рынка.
9. Когнитивная структуризация и формализация предметной области.
10. Формирование точечных и средневзвешенных прогнозов.
11. Оценка адекватности модели. Зависимость достоверности прогнозирования от разброса точечных прогнозов.
12. Детерминистские и бифуркационные участки траектории развития активных систем.

13. Прогнозирования времени перехода системы в бифуркационное состояние

20. Системно-когнитивный анализ семантической информационной модели.

21. Анализ динамики макроэкономических состояний городов и районов на уровне субъекта Федерации в ходе экономической реформы (на примере Краснодарского края) и прогнозирование уровня безработицы (на примере Ярославской области)

### ***Контрольные вопросы по лекции 16***

1. Ограничения АСК-анализа и обоснованное расширение области его применения на основе научной индукции.

2. Перспективы применения АСК-анализа в управлении.

3. Оценка рисков страхования и кредитования предприятий.

4. Мониторинг, прогнозирование и управление в финансово-экономической сфере.

5. Некоторые перспективные области применения АСК-анализа.

6. Развитие АСК-анализа.

7. Многоагентные распределенные системы обнаружения, накопления и использования знаний в Internet.

8. Развитие АСК-анализа с применением теории нечетких множеств и неклассической логики.

9. Динамика взаимодействующих семантических пространств и создание континуального АСК-анализа.

10. Перспективные области применения АСК-анализа и систем искусственного интеллекта.

### ***Контрольные вопросы по лабораторной работе 1***

1. Что такое классификационные и описательные шкалы и градации?

2. Какие существуют виды шкал?

3. Как шкалы связаны с конструктами и с познанием?

4. Чем обусловлена возможность текстового описания объектов обучающей и распознаваемой выборки на естественном языке?

### ***Контрольные вопросы по лабораторной работе 2***

1. Что понимается под формализацией задачи.

2. Как создать классификационные и описательные шкалы в системе "Эйдос"?

3. Как собрать исходную фактографическую информацию и ввести в систему обучающую выборку?

4. Как осуществить синтез и верификацию модели?

5. Как оценить ценность признаков для прогнозирования и выделить признаки, наиболее существенные для решения поставленной задачи?

6. Как провести анализ модели, чтобы ответить на следующие вопросы:

– как посещаемость занятий по системам искусственного интеллекта влияет на успеваемость по этой дисциплине?

– как сказывается пол на посещаемости?

– как выглядят конструкты "Пол", "Город-деревня", "Учебная группа", "Успеваемость", "Посещаемость"?

– какие студенты являются "типичными представителями" для своих учебных групп, а какие обладают своеобразием и выраженной индивидуальностью;

7. Как отобразить результаты анализа в графической форме нелокальных нейронов и семантических сетей признаков. На их основе построить классические когнитивные карты для хорошо и плохо успевающих студентов.

### ***Контрольные вопросы по лабораторной работе 3***

1. Какие работы выполняются на этапе формализации задачи?
2. Как в системе "Эйдос" ввести классификационные шкалы и градации, выбрав в качестве классов – различные уровни учебных достижений по различным дисциплинам, перечень которых взять из зачетной книжки?
3. Как в системе "Эйдос" ввести описательные шкалы и градации, используя характеристики подчеркика?
4. Каким образом подготовить и ввести в систему "Эйдос" обучающую выборку?
5. Как осуществить синтез и верификацию (измерение адекватности) семантической информационной модели в системе "Эйдос"?
6. Что включает системно-когнитивный анализ модели?
7. Как решаются задачи идентификации и прогнозирования в системе "Эйдос"?
8. В каких подсистемах и режимах системы "Эйдос" генерируются информационные портреты классов и факторов и отображаются в графической форме двумерных и трехмерных профилей классов и факторов?
9. Каким образом в системе "Эйдос" провести кластерно-конструктивный анализ классов и факторов и отобразить его в форме семантических сетей классов и факторов?
10. Как в системе "Эйдос" осуществить содержательное сравнение классов и факторов и отобразить результаты в форме когнитивных диаграмм классов и факторов?
11. Как в системе "Эйдос" построить нелокальные нейроны и интерпретируемые нейронные сети?
12. Как в системе "Эйдос" построить классические когнитивные модели и отобразить их в форме когнитивных карт?
13. Как в системе "Эйдос" построить интегральные когнитивные модели и отобразить в форме интегральных когнитивных карт?

### ***Контрольные вопросы по лабораторной работе 4***

1. Какие работы выполняются на этапе формализации задачи?
2. Как в системе "Эйдос" ввести классификационные шкалы и градации, выбрав в качестве классов – различные уровни учебных достижений по различным дисциплинам, перечень которых взять из зачетной книжки?
3. Как в системе "Эйдос" ввести описательные шкалы и градации, используя характеристики подчеркика?
4. Каким образом подготовить и ввести в систему "Эйдос" обучающую выборку?
5. Как осуществить синтез и верификацию (измерение адекватности) семантической информационной модели в системе "Эйдос"?



6. Что включает системно-когнитивный анализ модели?
7. Как решаются задачи идентификации и прогнозирования в системе "Эйдос"?
8. В каких подсистемах и режимах системы "Эйдос" генерируются информационные портреты классов и факторов и отображаются в графической форме двумерных и трехмерных профилей классов и факторов?
9. Каким образом в системе "Эйдос" провести кластерно-конструктивный анализ классов и факторов и отобразить его в форме семантических сетей классов и факторов?
10. Как в системе "Эйдос" осуществить содержательное сравнение классов и факторов и отобразить результаты в форме когнитивных диаграмм классов и факторов?
11. Как в системе "Эйдос" построить нелокальные нейроны и интерпретируемые нейронные сети?
12. Как в системе "Эйдос" построить классические когнитивные модели и отобразить их в форме когнитивных карт?
13. Как в системе "Эйдос" построить интегральные когнитивные модели и отобразить в форме интегральных когнитивных карт?
14. Кто такой "респондент"?
15. Что понимается под "социальным статусом" респондента?

### ***Контрольные вопросы по лабораторной работе 5***

1. Каким образом провести анализ устойчивости модели к неполноте информации и наличию шума?
2. Каким образом можно проверить способность модели правильно идентифицировать классы, один из которых является подмножеством другого?
3. Как оценить ценность букв для идентификации слов. Сравнить суммарную ценность для этой цели гласных и согласных букв?

### ***Контрольные вопросы по лабораторной работе 6***

1. Что такое атрибуция текстов?
2. Каким образом выполняется когнитивная структуризация предметной области?
3. В чем состоит формализацию предметной области и как ее осуществить в системе "Эйдос"?
4. Какие средства формирования обучающей выборки используются в системе "Эйдос" при решении задач атрибуции текстов?
5. В какой подсистеме и в каком режиме системы "Эйдос" осуществляется синтез семантической информационной модели и в чем он состоит?
6. В чем заключается оптимизация семантической информационной модели и как она осуществляется в системе "Эйдос"?
7. Как семантическая информационная модель проверяется на адекватность?
8. Как связана адекватность модели с внутренней и внешней, дифференциальной и интегральной валидностью?
9. Каким образом можно в системе "Эйдос" выполнить адаптацию модели и измерить, как изменилась ее адекватность?

10. Как в системе "Эйдос" осуществить пересинтез модели и измерить, как изменилась ее адекватность?

11. Чем отличается адаптация модели от пересинтеза в системе "Эйдос"?

12. В каких подсистемах и режимах системы "Эйдос" можно вывести информационные портреты?

13. В каких подсистемах и режимах системы "Эйдос" можно выполнить кластерно-конструктивный анализ модели?

### ***Контрольные вопросы по лабораторной работе 7***

1. Какие свойства натуральных чисел мы рассматривали в качестве их признаков?

2. Какие образом выполняется когнитивная структуризация предметной области?

3. В чем состоит формализацию предметной области и как ее осуществить в системе "Эйдос" при изучении свойств натуральных чисел?

4. Какие средства формирования обучающей выборки используются в системе "Эйдос" при изучении свойств натуральных чисел?

### ***Контрольные вопросы по лабораторной работе 8***

1. Что называется ортогональными проекциями тел?

2. В чем заключается сложность идентификации тел по их ортогональным проекциям?

3. За счет чего облегчается задача идентификации тел по их проекциям при наблюдении с нескольких точек или в движении?

### ***Контрольные вопросы по лабораторной работе 9***

1. Что выбирается в качестве классов распознавания и что в качестве факторов в задаче прогнозирования урожая?

2. Что такое "шкала качества" и "шкала количества"? Какие градации у этих шкал?

3. Как классифицировать факторы, влияющие на урожай?

4. В чем состоит задача прогнозирования количественных и качественных результатов выращивания сельхозкультур?

5. В чем состоит задача поддержки принятия решений при выборе зон и подзон выращивания, культур для выращивания, агротехнологий?

6. Каким образом можно научно обоснованно заменять одни факторы другими?

7. С какими проблемами сталкивается разработчик и исследователь при решении задачи прогнозирования урожая?

### ***Контрольные вопросы по лабораторной работе 10***

1. Что такое "случайная модель"?

2. Почему необходимо исследовать свойства случайной модели?

3. Какие средства для генерации и исследования случайных моделей есть в системе "Эйдос"?

### 3.3.2. Вопросы, относящиеся к дисциплине "ИИС", выносящиеся на государственный экзамен

1. Интеллектуальные информационные системы, как закономерный и неизбежный этап развития информационных систем.
2. Определение и критерии идентификации систем искусственного интеллекта. Тест Тьюринга.
3. Системы с интеллектуальной обратной связью.
4. Автоматизированные системы распознавания образов.
5. Системы поддержки принятия решений.
6. Экспертные системы.
7. Нейронные сети.
8. Генетические алгоритмы и моделирование эволюции.
9. Когнитивное моделирование.
10. Выявление знаний из опыта (эмпирических фактов) и интеллектуальный анализ данных (data mining).
11. Области применения систем искусственного интеллекта
12. Перспективы развития систем искусственного интеллекта, в т.ч. в Internet.
13. Абсолютная, относительная и аналитическая информация. Данные, информация, знания. Классификация СИИ.
14. Базы данных для поддержки принятия решений.
15. Автоматизированный системно-когнитивный анализ и универсальная когнитивная аналитическая система "Эйдос".

### 3.4. ПРАКТИЧЕСКИЕ ЗАДАНИЯ

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **прогнозирование продолжительности жизни пациентов, перенесших сердечный приступ, по данным эхокардиограммы на основе базы данных репозитория UCI и провести СК-анализ семантической информационной модели.**

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **принятие решения о выборе очередного хода в игре "крестики-нолики" в зависимости от расположения крестиков и ноликов и провести СК-анализ семантической информационной модели.**

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **классификацию животных по внешним признакам на основе базы данных репозитория UCI и провести СК-анализ семантической информационной модели**

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **диагностику фитопатологии по симптоматике и выработку рекомендаций по плану лечения на основе информации, содержащейся в учебнике, и провести СК-анализ семантической информационной модели**

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **идентификацию изображений различных мест на территории КубГАУ по вербальным описаниям их фотографий** (взять с сайта КубГАУ: <http://kubagro.ru>) и провести СК-анализ семантической информационной модели

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **прогнозирование успеваемости по ИИС на основе данных по социальному статусу их родителей** и провести СК-анализ семантической информационной модели

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **прогнозирование направления деятельности фирмы на основе данных о расположении и внешнем виде ее офиса** и провести СК-анализ семантической информационной модели

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **выбор автомобиля для приобретения по его признакам** (обучающую выборку взять на автомобильном рынке) и провести СК-анализ семантической информационной модели

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **выбор вариантов приобретения жилья по его признакам** и провести СК-анализ семантической информационной модели

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **идентификацию трехмерных тел (шар, куб, тетраэдр, конус, цилиндр, пирамида, призма и других) по их проекциям** и провести СК-анализ семантической информационной модели

Описать этапы разработки приложения в системе "Эйдос", обеспечивающее **оценку важности различных видов городского транспорта и различных маршрутов в разрезе по остановкам** и провести СК-анализ семантической информационной модели

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Идентификация трехмерных тел по полным и частичным наборам их ортогональных проекций (шар, куб, тетраэдр, конус, цилиндр, пирамида, призма)".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Идентификация и классификация натуральных чисел по их свойствам (делимости на натуральные делители)".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Идентификация и классификация студентов по их имиджу".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Прогнозирование пунктов назначения железнодорожных составов".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Прогнозирование количественных и качественных результатов выращивания зерновых колосовых".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Разработка обобщенных имеджевых фотороботов студентов по их успеваемости".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Идентификация изображений по их вербальным описаниям".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Атрибуция анонимных и псевдонимных текстов (определение вероятного авторства)".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Оценка рисков правонарушений по признакам подчерка (психографология)".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Оценка рисков страхования и кредитования предприятий по их описаниям".

Осуществить постановку задачи и формализацию предметной области, включая подготовку обучающей выборки, для решения задачи: "Оценка рисков совершения ДТП по данным о владельце и автомобиле".

Привести пример интегрального критерия.

Привести пример коллизии при двухэтапном групповом принятии решения.

Привести пример многокритеральной задачи принятия решений.

Привести пример, поясняющий различие между статистической и причинно-следственной связью.

Привести примеры критерия, параметра, фактора. Указать различия между ними.

Спланировать этапы исследования зависимости качества распознавания текста от разрешения сканирования в системах FineReader, Cunie Form и другие системах **разных версий**. Оценку качества производить по количеству ошибок распознавания. Составить рейтинг систем и версий, дать рекомендации. After Scan. Оценить тоже самое, после использования After Scan.

Спланировать этапы исследования качества переводов текстов в системах Stylus (Promt), Сократ, и других системах автоматизированного перевода **разных версий**. Сравнить качество автоматизированного перевода с русского языка на английский и обратно для текстов различной направленности (юридические, технические, художественные, стихи) и с различной длиной и сложностью предложений (статистика). Составить рейтинг систем и версий, дать рекомендации. Оценку качества перевода осуществлять путем обобщения экспертных оценок экспертов с разным уровнем компетентности (студенты).

Спланировать этапы исследования реальной системы выявления знаний из опыта (эмпирических фактов) и интеллектуального анализа данных при решении реальных задач.

Спланировать этапы исследования реальной системы класса: "Нейронная сеть" при решении задач.

Спланировать этапы исследования реальной системы когнитивного моделирования при решении реальных задач.

Спланировать этапы исследования реальной системы поддержки принятия решений при решении реальных задач.

Спланировать этапы исследования реальной системы распознавание образов, идентификации и прогнозирования при решении реальных задач.

Спланировать этапы исследования реальной системы, реализующей генетические алгоритмы при решении реальных задач.

Спланировать этапы исследования реальной экспертной системы при решении реальных задач.

## ЗАКЛЮЧЕНИЕ

В заключение хотелось бы отметить, что данное учебное пособие основано на материалах печати, Internet, а также на научных материалах кандидатской и докторской диссертаций автора, его научных работах, учебно-методических материалах, разработанных в ходе преподавания дисциплины "Интеллектуальные информационные системы" в течение 2002-2004 учебных годов на факультете прикладной информатики КубГАУ и опыте, полученном в ходе преподавания этой и других дисциплин.

Объем учебного пособия по ряду лекций оказался превышен и в будущем безусловно будет сокращен. Мы надеемся, что студенты сами обнаружат эти лекции и, прочитав их "по диагонали", смогут извлечь из них самую суть и в краткой форме донести ее до экзамена, способность к чему они уже не раз успешно демонстрировали. Кроме того в пособии некоторые учебные материалы дублируются в разных лекциях и лабораторных работах. Автор пошел на это сознательно, учитывая, что студенты посещают не все занятия и "повторение - мать учения".

Мы надеемся, что в дальнейшем данное учебное пособие будет совершенствоваться, как и универсальная когнитивная аналитическая система "Эйдос" и методология и технология системно-когнитивного анализа, на которых оно основано. Все критические замечания и пожелания, направленные на совершенствование данного учебного пособия, будут с благодарностью приняты.

*Автор: проф. Луценко Е.В.*  
Краснодар. 22.08.2005.



## ЛИТЕРАТУРА

### ОСНОВНАЯ ЛИТЕРАТУРА

1. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.
2. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности: 351400 "Прикладная информатика (по отраслям)". – Краснодар: КубГАУ. 2004. – 633 с.
3. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
4. Луценко Е.В. Универсальная автоматизированная система распознавания образов "Эйдос" (версия 4.1).-Краснодар: КЮИ МВД РФ, 1995.- 76с.
5. Симанков В.С., Луценко Е.В. Адаптивное управление сложными системами на основе теории распознавания образов. Монография (научное издание). – Краснодар: ТУ КубГТУ, 1999. - 318с.
6. Симанков В.С., Луценко Е.В., Лаптев В.Н. Системный анализ в адаптивном управлении: Монография (научное издание). /Под науч. ред. В.С.Симанкова. – Краснодар: ИСТЭК КубГТУ, 2001. – 258с.

### ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

1. Lutsenko E. V. Conceptual principles of the system (emergent) information theory & its application for the cognitive modelling of the active objects (entities), 2002 IEEE International Conference on Artificial Intelligence System (ICAIS 2002). –Computer society, IEEE, Los Alamos, California, Washington-Brussels-Tokyo, p. 268-269.
2. Адаменко А., Кучуков А.. Логическое программирование и Visual-Prolog. – СПб: «БХВ-Петербург», 2003.
3. Акофф Р., Эмери Ф.. О целеустремленных системах. – М.: Советское радио, 1974.
4. Алиев Р.А., Абдикеев Н.М., Шахназаров М.М.. Производственные системы с искусственным интеллектом. - М.: Радио и связь, 1990.
5. Анищик Т.А., Галлиев К.С., Кочкин Н.Н., Лойко В.И., Лаптев В.Н., Луценко Е.В. и др. Информатика: компьютерный практикум для студентов агрономических, инженерных, юридических и экономических специальностей (Компьютерный практикум) Компьютерный практикум для студентов агрономических, инженерных, юридических и экономических специальностей (Под. ред. В.И.Лойко). – Краснодар: ФГОУ ВПО "КубГАУ", 2004. – 128 с.
6. Арбиб М.. Метафорический мозг. - М.: Мир, 1976.
7. Бакурадзе Л.А., Луценко Е.В., Самсонов Г.А. Руководящие материалы по эксплуатации первой очереди автоматизированной информационно-управляющей си-



- стемы агропромышленного комплекса (АИУС-АПК) с применением ПЭКВМ "Искра-226", Курск: Тип. Курского облисполкома, 1986. –109 с. (ДСП).
8. Барановская Т.П., Василенко И.И., Кондратьев В.Ю., Курносов С.А., Лаптев В.Н., Лойко В.И., Луценко Е.В. Методическое пособие по итоговой государственной аттестации выпускников (Методическое пособие) Методическое пособие по итоговой государственной аттестации выпускников специальности: 351400 – «Прикладная информатика в юриспруденции». /Под. ред. В.И. Лойко. – Краснодар: ФГОУ ВПО "КубГАУ", 2004. – 49 с.
  9. Берштейн Л.С. Нечеткие модели для экспертных систем в САПР. М.: Энергоатомиздат, 1991.-136 с.
  10. Бранский В.П. Философские основания проблемы синтеза релятивистских и квантовых принципов. –Л: ЛГУ, 1973. –175с.
  11. Братко И. Программирование на языке ПРОЛОГ для искусственного интеллекта. - М.: Мир, 1990.
  12. Васильев В.И. Распознающие системы. Справочник. - Киев, Наукова думка, 1983.
  13. Васильев Л.Г. Три парадигмы понимания: анализ литературы вопроса. <http://newasp.omskreg.ru/intellect/f54.htm>.
  14. Гаазе-Рапопорт М.Г., Поспелов Д.А.. От амебы до робота: модели поведения. – М.: Наука, 1987.
  15. Гаврилов А.В.. Гибридные интеллектуальные системы. – Новосибирск: НГТУ, 2003.
  16. Гаврилов А.В.. Лабораторный практикум по нейронным сетям. Ч.1. - Новосибирск, НГТУ, 2000.
  17. Гаврилов А.В.. Системы искусственного интеллекта. Уч. пособие, ч. 1. - Новосибирск, НГТУ, 2000, 2001.
  18. Гаврилова Т.А., Хорошевский В.Ф.. Базы знаний интеллектуальных систем. – СПб: Питер, 2000.
  19. Гаврилова Т.А., Червинская К.Р.. Извлечение и структурирование знаний для экспертных систем. - М.: Радио и связь, 1992.
  20. Гаек П., Гавранек Т.. Автоматическое образование гипотез. - М.: Наука, 1984.
  21. Галушкин А.. Нейрокомпьютеры. М.: ИПРЖР, 2000.
  22. Галушкин А.. Теория нейронных сетей. М.:ИПРЖР, 2000.
  23. Гладун В.П.. Планирование решений. - Киев, Наукова думка, 1987.
  24. Горбань А.Н., Россиев Д.А.. Нейронные сети на персональном компьютере. - Новосибирск, Наука, 1996.
  25. Горбань А.Н.. Обучение нейронных сетей. - М.: СП Параграф, 1990.
  26. Гренандер У.. Лекции по теории образов. В 3-х кн. – М.: Мир, 1983.
  27. Джан Роберт Г., Данн Бренда Дж. Границы реальности. (Роль сознания в физическом мире). /Пер. с англ. - М.: Объединенный институт высоких температур РАН, 1995. - 287с.
  28. Джексон П. Введение в экспертные системы. - М., СПб., Киев: "Вильямс", 2001.
  29. Драгавцев В.А., Драгавцева И.А., Лопатина Л.М. Управление продуктивностью сельскохозяйственных культур на основе закономерностей их генетических и фенотипических изменений при смене лимитов внешней среды. – Краснодар. СКЗ-НИИСиВ, 2004. – 208 с.
  30. Драгавцева И.А. и др. Персик на Юге России и Украины. –Краснодар: СКЗ-НИИСиВ, 2001. –120с.

31. Драгавцева И.А. Экологические основы оптимального размещения абрикоса на Северном Кавказе. Дисс...д.с./х.н. (06.01.07 – Плодоводство) – Краснодар: КубГАУ, 1981. – 328 с.
32. Драгавцева И.А. Экологические ресурсы продуктивности абрикоса на юге России. – Краснодар: 1999. – 94с.
33. Драгавцева И.А., Лопатина Л.М., Луценко Е.В. Автоматизация системного анализа продуктивности плодовых культур Юга России, Научные труды Российской академии сельскохозяйственных наук. – 2005. – №4. – С. 11-14.
34. Драгавцева И.А., Луценко Е.В., Запорожец Н.М., Луценко Н.Е. Новые подходы к районированию плодовых культур на Юге России с применением компьютерного моделирования, Организационно-экономический механизм инновационного процесса и приоритетные проблемы научного обеспечения развития отрасли (Материалы Всероссийской научно-практической конференции 3-4 февраля 2003 г. СКЗНИИСиВ, Краснодар). – Краснодар. – 2003. – С. 74-76.
35. Драгавцева И.А., Луценко Е.В., Лопатина Л.М., Луценко Н.Е. Применение системного анализа для прогнозирования успешности выращивания сельскохозяйственных культур (на примере плодовых), В сб.: "Формы и методы повышения эффективности координации исследований для ускорения процесса передачи реальному сектору экономики завершенных разработок". – Краснодар: СКЗНИИСиВ, 2002. – С. 62-67.
36. Драгавцева И.А., Луценко Е.В., Луценко Н.Е., Лопатина Л.М. Применение автоматизированного системного анализа для прогноза продуктивности плодовых культур на Юге страны, Материалы Всероссийской научно-практической конференции. – Краснодар: СКЗНИИСВ, 2002. – С. 8-11.
37. Дюбуа Д., Прад А.. Теория возможностей. Приложения к представлению знаний в информатике. - М.: Радио и связь, 1990.
38. Дюк В.А. Компьютерная психодиагностика. - СПб: Братство, 1994. - 365с.
39. Дюк В.А. Обработка данных на ПК в примерах. – СПб: Питер, 1997. – 240с.
40. Егоров Е.А., Драгавцева И.А., Лопатина Л.М., Луценко Е.В. и др. Интенсивные технологии возделывания плодовых культур, Монография (научное издание): СКЗНИИСиВ. – Краснодар. – 2004. – 394 с.
41. Емельянов-Барковский Л.Б.. Интеллектуальная квазибиологическая система. – М.: Наука, 1990. – 112с.
42. Ерофеев А.А., Поляков А.О.. Интеллектуальные системы управления. СПб: Издательство СПбГТУ, 1999.
43. Завгородний В.В. и Мельников Ю.Н. Идентификация по клавиатурному почерку, "Банковские Технологии" №9, 1998.
44. Загоруйко Н.Г.. Прикладные методы анализа данных и знаний. – Новосибирск, 1999.
45. Заде Л.. Понятие о лингвистической переменной и его применение к принятию решений. - М.: Мир, 1976.
46. Интеллектуализация ЭВМ - // В уч. пос. Перспективы развития вычислительной техники в 11 кн. Кн. 2. - М.: Высшая школа, 1989.
47. Интеллектуальные системы и их моделирование. - М.:Наука, 1986.
48. Искусственный интеллект. Применение в интегрированных производственных системах. Под ред. Э.Кьюсиака. - М.: Машиностроение, 1991.
49. Искусственный интеллект. Справочник в 3-х томах. - М.: Радио и связь, 1990.
50. Каллан Р. Основные концепции нейронных сетей. - М.: "Вильямс", 2001.
51. Калустов А.А., Луценко Е.В. Применение автоматизированного системно-когнитивного анализа для совершенствования методов компьютерной селекции

- подсолнечника, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №02(10). – Режим доступа: <http://ej.kubagro.ru/2005/02/10/p10.asp>
52. Кандрашина Е.Ю., Литвинцева А.В., Поспелов Д.А.. Представление знаний о времени и пространстве в интеллектуальных системах. - М.: Наука, 1989.
  53. Кива В. Данные, информация, знания. <http://vlak.webzone.ru/rus/it/knowledge.html>.
  54. Клыков Ю.И. Луценко Е.В., и др. Разработка и управление реализацией городских целевых комплексных программ. ГСГ 00.004-85 (Стандарт) Городской стандарт. – Краснодар: КФ ПНИИПОУ, 1985. – 107 с.
  55. Ковальски Р.. Логика в решении проблем. - М.: Наука, 1990.
  56. Козлов Ю.М.. Адаптация и обучение в робототехнике. – М.: Наука, 1990.
  57. Комарцова Л.Г., Максимов А.В. Нейрокомпьютеры. - М.: Изд-во МГТУ им. Н.Э.Баумана, 2002.
  58. Комашинский В.И., Смирнов В.И. Нейронные сети и их применение в системах управления и связи. – М.: Горячая линия – Телеком, 2002.
  59. Корнеев В.В., Гарев А.Ф., Васютин С.В., Райх В.В.. Базы данных. Интеллектуальная обработка информации. - М.: "Нолидж", 2000.- 2-е изд., 2001.
  60. Корноушенко Е.К., Максимов В.И. Управление процессами в слабоформализованных средах при стабилизации графовых моделей среды. Труды ИПУ, вып.2, 1998.
  61. Кохонен Т.. Ассоциативная память. – М.: Мир, 1980.
  62. Круглов В.В., Борисов В.В.. Искусственные нейронные сети. Теория и практика. - М.: Горячая линия-Телеком, 2001.
  63. Кузнецов В.Е.. Представление в ЭВМ неформальных процедур. - М.: Наука, 1989.
  64. Кузнецов И.П.. Кибернетические диалоговые системы. – М.: Наука, 1976.
  65. Кузнецов И.П.. Механизмы обработки семантической информации. – М.: Наука, 1978.
  66. КуССуль Э.М. Ассоциативные нейроподобные структуры. - Киев, Наукова думка, 1990.
  67. Лаптев В.Н., Луценко Е.В. Информационные технологии в юриспруденции (Методические указания) Методические указания по подготовке курсовых работ (для студентов специальностей 3514003 – Прикладная информатика в юриспруденции). – Краснодар: ФГОУ ВПО "КубГАУ", 2003. – 62 с.
  68. Левин Р., Дранг Д., Эдельсон Б.. Практическое введение в технологию искусственного интеллекта и экспертных систем с иллюстрациями на Бейсике. - М.: Финансы и статистика, 1990.
  69. Лийв Э. Х. Инфодинамика. Обобщённая энтропия и негэнтропия. - Таллинн, 1998. - 200 с.
  70. Линдсей П., Норман Д.. Переработка информации у человека. – М.: Мир, 1974.
  71. Логический подход к искусственному интеллекту. - М.: Мир, 1990.
  72. Лойко В.И., Курносое С.А., Лаптев В.Н., Луценко Е.В., Мироненко В.Л. Базы данных (Методические указания) Методические указания по подготовке курсовых работ (для студентов специальностей 3514001 – Прикладная информатика в экономике и 3514003 – Прикладная информатика в юриспруденции). – Краснодар: ФГОУ ВПО "КубГАУ", 2004. – 62 с.
  73. Лойко В.И., Лаптев В.Н., Луценко Е.В., Постный В.А. Вычислительные сети, системы и телекоммуникации (Методические указания) Методические указания по подготовке курсовых работ для студентов специальностей 3514001 – Прикладная информатика в экономике и 3514003 – Прикладная информатика в юриспруденции. Краснодар: КубГАУ, 2003. – 46 с.

74. Лопатина Л.М., Драгавцева И.А., Луценко Е.В. Концептуальная постановка задачи: "Прогнозирование количественных и качественных результатов выращивания заданной культуры в заданной точке", Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №05(7). – Режим доступа: <http://ej.kubagro.ru/2004/05/08/p08.asp>
75. Лорьер Ж.-Л. Системы искусственного интеллекта. - М: Мир, 1991.
76. Луценко Е.В. Автоматизация когнитивных операций системного анализа, В сб.: "Проблемы совершенствования систем защиты информации, энергоснабжения военных объектов и образовательных технологий подготовки специалистов". Материалы II межвузовской научно-технической конференции. – Краснодар: КВИ, 2001. – С. 131-133.
77. Луценко Е.В. Автоматизированная система "ЭЙДОС", как инструмент для разработки и эксплуатации новых психологических тестов (Тезисы) В сб.: "Теоретические и прикладные проблемы социально-психологической и медико-педагогической служб". Материалы IX региональной научно-практической конференции. – Краснодар: КО РПО, 1995. – С. 47.
78. Луценко Е.В. Автоматизированная система распознавания образов, математическая модель и опыт применения, В сб.: "В.И.Вернадский и современность (к 130-летию со дня рождения)". Тезисы научно-практической конференции. – Краснодар: КНА, 1993. – С. 37-42.
79. Луценко Е.В. Автоматизированный когнитивный системный анализ влияния ДПО на качество подготовки специалистов, В сб.: "Опыт и проблемы совершенствования региональной системы дополнительного профессионального образования в области управления качеством". Тезисы докладов научно-практической конференции. – Краснодар: Академия СМС, 2001. – С. 38-39.
80. Луценко Е.В. Автоматизированный когнитивный системный анализ фондового рынка, В сб.: "Проблемы экономического и социального развития России". Материалы Всероссийской научно-практической конференции. – Пенза: ПГУ, 2001. – С. 206-209.
81. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем):, Монография (научное издание). – Краснодар: КубГАУ, 2002. – 605 с.
82. Луценко Е.В. Автоматизированный системно-когнитивный анализ в экономике, Изв. вузов. Северо-Кавказский регион. Технические науки. – 2003. – №1. – С. 189-194.
83. Луценко Е.В. Адаптивная система обработки данных «ВЕГА» на базе мини-ЭВМ «Искра-226», Деп. в ВИНТИ 25.10.84. №7180-84деп. – 18с.
84. Луценко Е.В. Адаптивная система обработки данных «ВЕГА», ИЛ №6-85. – Краснодар: КЦНТИ, 1985. – 4с.
85. Луценко Е.В. Адаптивная система обработки данных, Автоматизация управления транспортно-заготовительными работами в РАПО. Экспресс-информация. Выпуск 14. – М.: ЦНИИ ТЭИ приборостроения, 1985. – 4с.
86. Луценко Е.В. АСК-анализ информационной безопасности организационно-технических систем, Межвузовский сборник научных трудов, том 1. – Краснодар: КВИ, 2003. – С. 87-89.
87. Луценко Е.В. АСК-анализ как метод выявления когнитивных функциональных зависимостей в многомерных зашумленных фрагментированных данных, Науч-

- ный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №03(11). – Режим доступа: <http://ej.kubagro.ru/2005/03/19/p19.asp>
88. Луценко Е.В. Атрибуция анонимных и псевдонимных текстов в системно-когнитивном анализе, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(5). – Режим доступа: <http://ej.kubagro.ru/2004/03/03/p03.asp>
  89. Луценко Е.В. Атрибуция текстов, как обобщенная задача идентификации и прогнозирования, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №02(2). – Режим доступа: <http://ej.kubagro.ru/2003/02/13/p13.asp>
  90. Луценко Е.В. Беседы об искусстве превращения жемчуга в алмаз (Мастеру, звезда которого светит из будущего), Краснодар: НПП "ЭЙДОС", 1995. – 138 с.
  91. Луценко Е.В. Взаимосвязь эластичности и системной меры целесообразности информации, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(1). – Режим доступа: <http://ej.kubagro.ru/2003/01/05/p05.asp>
  92. Луценко Е.В. Виртуализация общества как основной информационный аспект глобализации (основы информационно-функциональной теории развития техники и информационной теории стоимости), Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №01(9). – Режим доступа: <http://ej.kubagro.ru/2005/01/02/p02.asp>
  93. Луценко Е.В. Возможности прогнозирования учебных достижений студентов на основе АСК-анализа их имеджевых фотороботов, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №02(4). – Режим доступа: <http://ej.kubagro.ru/2004/02/13/p13.asp>
  94. Луценко Е.В. Гипотеза о мировом компьютере, ментальном программировании и НЛО, В сб.: "Новые идеи и гипотезы (тезисы докладов конференции)". – Краснодар: КДНТ, 1990. – С. 168-174.
  95. Луценко Е.В. Гипотеза о физической природе НЛО и тунгусского феномена, В сб.: "Новые идеи и гипотезы (тезисы докладов конференции)". – Краснодар: КДНТ, 1990. – С. 178-183.
  96. Луценко Е.В. Идентификация слов по входящим в них буквам с применением системно-когнитивного анализа, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №02(4). – Режим доступа: <http://ej.kubagro.ru/2004/02/12/p12.asp>
  97. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности: 351400 "Прикладная информатика (по отраслям)". – Краснодар: КубГАУ. 2004. – 633 с.
  98. Луценко Е.В. Интерференция последствия выбора в результате одновременного осуществления альтернатив и необходимость разработки системной (эмерджентной) теории информации, Проблемы совершенствования систем защиты информации, образовательного процесса и электроснабжения военных объектов: Межвузовский сборник научных работ. – 2002. – №3. – С.72-74.
  99. Луценко Е.В. Информационно-функциональная теория развития техники, закон повышения качества базиса и детерминация формы сознания человека функциональным уровнем технологической среды. Труды Кубанского государственного аграрного университета. – Выпуск № 420 (448), – КубГАУ, Краснодар, 2005. – С.218-236.
  100. Луценко Е.В. Исследование адекватности, сходимости и семантической устойчивости системно-когнитивной модели активных объектов, Проблемы совершенствования систем защиты информации, образовательного процесса и электро-

- снабжения военных объектов: Межвузовский сборник научных работ. – 2002. – №3. – С. 64-70.
101. Луценко Е.В. Когнитивная аналитическая система "ЭЙДОС-6.0" и система "ЭЙДОС-Ψ" – адекватный инструментарий для психологических служб МВД, В сб.: "Актуальные проблемы социально-правовой подготовки специалистов и перспективы совершенствования системы комплектования ОВД". Материалы межвузовской научно-практической конференции (16-18.05.1997). Часть 1. –Краснодар: КЮИ МВД РФ, 1997. – С. 65-69.
  102. Луценко Е.В. Когнитивная аналитическая система "ЭЙДОС-6.2", как инструмент исследования интегральной индивидуальности по Вольфу Мерлину, В сб.: "Актуальные проблемы социально-правовой подготовки специалистов и перспективы совершенствования системы комплектования ОВД". Материалы межвузовской научно-практической конференции (16-18.05.1997). Часть 2. –Краснодар: КЮИ МВД РФ, 1997. – С. 136-141.
  103. Луценко Е.В. Концептуальные основы системной (эмерджентной) теории информации и ее применение для когнитивного моделирования активных объектов, Ж-л "Перспективные информационные технологии и интеллектуальные системы", №1/2003. – Таганрог: ТГРТУ, 2003. – С. 23-27.
  104. Луценко Е.В. Критерии реальности и принцип эквивалентности виртуальной и "истинной" реальности, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №06(8). – Режим доступа: <http://ej.kubagro.ru/2004/06/10/p10.asp>.
  105. Луценко Е.В. Магнит грядущего (о небиологических формах жизни, природе НЛО и перспективах сознания), Краснодар: Агропромполиграфист, 1990. – 70с.
  106. Луценко Е.В. Математическая модель автоматизированной системы распознавания образов (Тезисы) В сб.: "Тезисы докладов VIII Всесоюзного съезда психологов". – М.: Наука, 1989. – С.35.
  107. Луценко Е.В. Математический метод СК-анализа в свете идей интервальной бутстрепной робастной статистики объектов нечисловой природы, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(3). – Режим доступа: <http://ej.kubagro.ru/2004/01/13/p13.asp>
  108. Луценко Е.В. Методика использования репозитария UCI для оценки качества математических моделей систем искусственного интеллекта, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №02(2). – Режим доступа: <http://ej.kubagro.ru/2003/02/12/p12.asp>
  109. Луценко Е.В. Нелокальные интерпретируемые нейронные сети прямого счета, как инструмент системно-когнитивного анализа, Изв. вузов. Северо-Кавказский регион. Технические науки. Приложение №3. – 2003. – С. 3-12.
  110. Луценко Е.В. О пути перехода технократической цивилизации к постапокалиптическому обществу, В сб.: "Глобальные проблемы человечества" Тезисы докладов конференции. – Краснодар: КНА, 1994. – С. 31-36.
  111. Луценко Е.В. Персональная система обработки информации "ВЕГА-М" для мини-ЭВМ "Искра-226", ИЛ о научно-техническом достижении №87-12. – Краснодар: КЦНТИ, 1987. Гос. рег. № 01.06.01.Н15.06.03.05.02. – 4с.
  112. Луценко Е.В. Персональная технологическая система "ВЕГА-3" для ЭВМ "Искра-226", ИЛ №174-89. –Краснодар: КЦНТИ, 1989. – 4с.
  113. Луценко Е.В. Разработка методологии синтеза адаптивных АСУ сложными объектами на основе применения моделей распознавания образов и принятия решений (Автореферат кандидатской диссертации) Автореферат диссертации на соискание ученой степени к.т.н. (05.13.06). – Краснодар: КубГТУ, 1999. – 24 с.



114. Луценко Е.В. Разработка методологии синтеза адаптивных АСУ сложными объектами на основе применения моделей распознавания образов и принятия решений (Кандидатская диссертация) Диссертация на соискание ученой степени к.т.н. (05.13.06). – Краснодар: КубГТУ, 1999. – 187 с.
115. Луценко Е.В. Расчет эластичности объектов информационной безопасности на основе системной теории информации, Ж-л "Безопасность информационных технологий". – М.: МИФИ, 2003. – №2. – С. 82-90.
116. Луценко Е.В. Рефлексивная модель управления качеством подготовки специалиста, В сб.: "Проблемы совершенствования систем защиты информации, энергоснабжения военных объектов и образовательных технологий подготовки специалистов". Материалы II межвузовской научно-технической конференции. – Краснодар: КВИ, 2001. – С. 129-131.
117. Луценко Е.В. Селиверстов В.В. Разработка профессиональных оптимальных адаптивных тестов на основе интеллектуальной технологии "ЭЙДОС", В сб.: "Современные компьютерные технологии обучения". Материалы 2-й межвузовской научно-методической конференции. – Краснодар: КВВАУ, 1998. – С. 32-34.
118. Луценко Е.В. Синтез адаптивных систем управления индивидуальным обучением на базе интеллектуальной системы "ЭЙДОС", В сб.: "Современные компьютерные технологии обучения". Материалы 2-й межвузовской научно-методической конференции. – Краснодар: КВВАУ, 1998. – С. 27-30.
119. Луценко Е.В. Синтез экстремальных систем "человек–машина" на основе принципа многоуровневой адаптивности, Труды Краснодарского ВАИ: Межвузовский научно-методический сборник. Вып. 4. – Краснодар: КВАИ, 1999. – С. 186-190.
120. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(1). – Режим доступа: <http://ej.kubagro.ru/2003/01/11/p11.asp>
121. Луценко Е.В. Системно-когнитивный анализ в управлении АПК (Автореферат докторской диссертации) Автореферат диссертации на соискание ученой степени д. э. н. (08.00.13). – Краснодар: КубГАУ, 2003. – 47 с.
122. Луценко Е.В. Системно-когнитивный анализ в управлении АПК (Докторская диссертация) Диссертация на соискание ученой степени д. э. н. (08.00.13). – Краснодар: КубГАУ, 2003. – 467 с.
123. Луценко Е.В. Системно-когнитивный анализ детерминистско-бифуркационной динамики активных систем, Проблемы совершенствования систем защиты информации, образовательного процесса и электроснабжения военных объектов: Межвузовский сборник научных работ. – 2002. – №3. – С. 50-53.
124. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка-Абельсона, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(5). – Режим доступа: <http://ej.kubagro.ru/2004/03/04/p04.asp>
125. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (На примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"), Краснодар: КЮИ МВД РФ, 1996. – 278 с.
126. Луценко Е.В. Теоретические основы системной (эмерджентной) теории информации, Проблемы совершенствования систем защиты информации, образовательного процесса и электроснабжения военных объектов: Межвузовский сборник научных работ. – 2002. – №3. – С. 84-93.

127. Луценко Е.В. Технологический модуль поддержки диалога персональной системы обработки информации ВЕГА-М, ИЛ №274-87. – Краснодар: Краснодарский ЦНТИ, 1987.
128. Луценко Е.В. Технология разработки автогенерируемых программ в системе Бэйсик 02 персональной ЭВМ "Искра-226", ИЛ №273-87. –Краснодар: КЦНТИ, 1987. – 4с.
129. Луценко Е.В. Типовая методика и инструментарий когнитивной структуризации и формализации задач в СК-анализе, // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(3). – Режим доступа: <http://ej.kubagro.ru/2004/01/16/p16.asp>
130. Луценко Е.В. Универсальная автоматизированная система распознавания образов "ЭЙДОС-4", ИЛ №438-93. –Краснодар: КЦНТИ, 1993. – 4с.
131. Луценко Е.В. Универсальная автоматизированная система распознавания образов "ЭЙДОС": опыт и перспективы применения, В сб.: "Состояние и связи криминалистики и теории оперативно-розыскной деятельности ОВД". Материалы всероссийской научно-практической конференции (14-16.06.1995). – Краснодар: КЮИ МВД РФ, 1995. – С. 168-177 (ДСП).
132. Луценко Е.В. Универсальная автоматизированная система распознавания образов "ЭЙДОС" (версия 4.1), Краснодар: КЮИ МВД РФ, 1995. – 76 с.
133. Луценко Е.В. Универсальный модуль ввода-вывода файлов данных в системе Бэйсик 02 персональной ЭВМ "Искра-226", ИЛ №6-87. – Краснодар: КЦНТИ, 1987. – 4с.
134. Луценко Е.В. Утилита проектирования оверлейных структур в среде Clipper, ИЛ №101-91. –Краснодар: КЦНТИ, 1991. – 4с.
135. Луценко Е.В. Численный расчет эластичности объектов информационной безопасности на основе системной теории информации, Краснодар: КЮИ МВД РФ, 2003. – С. 43-54.
136. Луценко Е.В., Бакурадзе Л.А. Математическая модель и алгоритм решения задачи оперативного планирования и управления в условиях РАПО, Деп. в ВИНТИ 25.09.84. №2650-85деп. – 14с.
137. Луценко Е.В., Бакурадзе Л.А., Гнездюков Н.Д. Программное обеспечение оперативного планирования транспортно-заготовительного процесса на уборке сахарной свеклы с использованием микро-ЭВМ, ИЛ №79-84. –Краснодар: КЦНТИ, 1984. – 2с.
138. Луценко Е.В., Босенко А.А. Использование редактора программ и стандартных инструкций ВМО Бэйсик 02 ПЭКВМ "Искра-226" для работы с текстами, ИЛ №512-84. –Краснодар: КЦНТИ, 1984. – 4с.
139. Луценко Е.В., Босенко А.А. Автоматизированная система обработки данных на базе ПЭКВМ "Искра-226", ИЛ №557-84. –Краснодар: КЦНТИ, 1984. – 2с.
140. Луценко Е.В., Босенко А.А. Контекстное редактирование программ в системе Бэйсик 02 ПЭКВМ «Искра-226», ИЛ №347-84. –Краснодар: КЦНТИ, 1984. – 4с.
141. Луценко Е.В., Босенко А.А. Обработка ошибок на ПЭКВМ «Искра-226», ИЛ №155-84. –Краснодар: КЦНТИ, 1984. – 4с.
142. Луценко Е.В., Босенко А.А. Организация вычислимого перехода в системе Бэйсик 02 (версия 15.02.84) ПЭКВМ "Искра-226", ИЛ №513-84. –Краснодар: КЦНТИ, 1984. – 4с.
143. Луценко Е.В., Власенко А.В. Методика изучения влияния дополнительного профессионального образования на качество подготовки специалистов, В сб.: "Опыт и проблемы совершенствования региональной системы дополнительного профессионального образования в области управления качеством". Тезисы докладов



- научно-практической конференции. – Краснодар: Академия СМС, 2001. – С. 23-24.
144. Луценко Е.В., Гнездюков Н.Д. Оперативный контроль работы предприятий РАПО в период уборки урожая с применением ПЭКВМ "Искра-226", ИЛ №73-84. – Краснодар: КЦНТИ, 1984. – 4с.
  145. Луценко Е.В., Гнездюков Н.Д. Параллельное использование ресурсов ПЭКВМ «Искра-226» в адаптивной системе обработки данных «ВЕГА», Деп. в ВИНТИ 31.01.85. №3393-85деп. – 8с.
  146. Луценко Е.В., Гнездюков Н.Д. Программное обеспечение станции подготовки данных на базе ПЭКВМ «Искра-226», Автоматизация управления транспортно-заготовительными работами в РАПО. Экспресс-информация. Выпуск 15. – М.: ЦНИИ ТЭИ приборостроения, 1985. – 3с.
  147. Луценко Е.В., Гнездюков Н.Д. Реализация обобщенного оператора Convert на языке ассемблер ПЭВМ "Искра-226", ИЛ № 175-89. – Краснодар: КЦНТИ, 1989. – 4с.
  148. Луценко Е.В., Лаптев В.Н. Некоторые проблемы и перспективы управления качеством подготовки специалистов, В сб.: "Проблемы совершенствования систем защиты информации, энергоснабжения военных объектов и образовательных технологий подготовки специалистов". Материалы II межвузовской научно-технической конференции. – Краснодар: КВИ, 2001. – С. 133-135.
  149. Луценко Е.В., Лаптев В.Н. Изучение некоторых понятий гносеологии и психологии на примере когнитивной аналитической системы «ЭЙДОС»:, В сб.: "Современные компьютерные технологии обучения". Материалы 2-й межвузовской научно-методической конференции. – Краснодар: КВВАУ, 1998. – С. 34-38.
  150. Луценко Е.В., Лаптев В.Н. Применение когнитивной аналитической системы "ЭЙДОС" для прогнозирования успешности будущей профессиональной деятельности абитуриентов и учащихся вузов, В сб.: "Современные компьютерные технологии обучения". Материалы 2-й межвузовской научно-методической конференции. – Краснодар: КВВАУ, 1998. – С. 30-32.
  151. Луценко Е.В., Лаптев В.Н., Третьяк В.Г. Прогнозирование качества специальной деятельности методом подсознательного (подпорогового) тестирования на основе семантического резонанса, В сб.: "Проблемы совершенствования систем защиты информации, энергоснабжения военных объектов и образовательных технологий подготовки специалистов". Материалы II межвузовской научно-технической конференции. – Краснодар: КВИ, 2001. – С. 127-128.
  152. Луценко Е.В., Лебедев А.Н. Диагностика и прогнозирование профессиональных и творческих способностей методом АСК-анализа электроэнцефалограмм в системе "Эйдос", Межвузовский сборник научных трудов, том 1. – Краснодар: КВИ, 2003. – С. 227-229.
  153. Луценко Е.В., Лопатина Л.М. Создание автоматизированной системы мониторинга, анализа, прогноза и управления продуктивностью сельскохозяйственных культур, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №02(2). – Режим доступа: <http://ej.kubagro.ru/2003/02/07/p07.asp>
  154. Луценко Е.В., Рябикина З.М., Некрасов С.Д. Личность и профессия: проблема самоактуализации, В сб.: "Психологические проблемы самореализации личности" / Под ред. О.Г.Кукосяна. – Краснодар: КубГУ, 1997. – С. 127-140.
  155. Луценко Е.В., Третьяк В.Г. Анализ профессиональных траекторий специалистов с применением системы "Эйдос", Личность и ее бытие (социально-психологические аспекты бытия личности в местном сообществе): сборник научных работ / Под ред. З.И.Рябикиной. – Краснодар: КубГУ, 2002. – С. 43-49.

156. Любарский Ю.Я. Интеллектуальные информационные системы. - М.: Наука, 1990.
157. Люгер Дж.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем. – М.: «Вильямс», 2003.
158. Максимов В.И., Качаев С.В., Корноушенко Е.К. Концептуальное моделирование и мониторинг проблемных и конфликтных ситуаций при целенаправленном развитии региона. В сб. "Современные технологии управления для администраций городов и регионов". Фонд "Проблемы управления", М. 1998.
159. Максимов В.И., Корноушенко Е.К. Аналитические основы применения когнитивного подхода при решении слабоструктурированных задач. Труды ИПУ, вып.2, 1998.
160. Максимов В.И., Корноушенко Е.К. Знание – основа анализа. Банковские технологии, № 4, 1997.
161. Максимов В.И., Корноушенко Е.К., Качаев С.В. Анализ ситуации и компенсация теневых аспектов в свободной торговле. В сб. "Современные технологии управления для администраций городов и регионов". Фонд "Проблемы управления", М. 1998.
162. Максимов В.И., Корноушенко Е.К., Качаев С.В., Григорян А.К. Когнитивный подход к анализу проблемы демонополизации в транспортном комплексе. Труды ИПУ, вып.2, 1998.
163. Малышев Н.Г., Берштейн Л.С., Боженюк А.В.. Нечеткие модели для экспертных систем в САПР. - М.: Энергоатомиздат, 1991.
164. Марселлус Д. Программирование экспертных систем на Турбо-Прологе. - М.: Финансы и статистика, 1994.
165. Медведев В.С., Потемкин В.Г.. Нейронные сети MATLAB 6. – М.: Диалог-МИФИ, 2002.
166. Мелихов А.Н., Бернштейн Л.С., Коровин С.Я.. Ситуационные советующие системы с нечеткой логикой. - М.: Наука, 1990.
167. Минский М.. Фреймы для представления знаний. - М.: Энергия, 1979.
168. Моделирование языковой деятельности в интеллектуальных системах. Под ред. А.Е.Кибрика и А.С.Нариньяни. - М.: Наука, 1987.
169. Моргенштерн И. Психографология. –СПб.: Питер, 1994. –350 с.
170. Назаров А.В., Лоскутов А.И.. Нейросетевые алгоритмы прогнозирования и оптимизации систем. – СПб.: Наука и техника, 2003.
171. Накано Э.. Введение в робототехнику. – М.: Мир, 1988.
172. Нейроинформатика. / А.Н. Горбань, В.Л. Дунин-Барковский, А.Н. Кирдин и др. – Новосибирск: Наука. Сибирское предприятие РАН, 1998. – 296с.
173. Нейрокомпьютеры и интеллектуальные роботы. Под ред. Н.М. Амосова. – Киев, Наукова думка, 1991.
174. Нейронные сети. STATISTICA Neural Networks. - М.: Горячая линия - Телеком, 2000.
175. Нейропрограммы. Уч. пособие. – Красноярск, 1994.
176. Нечеткие множества в моделях управления и искусственного интеллекта/А.Н. Аверкин, И.З. Батыршин, А.Ф. Блишун, В.Б. Силов, В.Б. Тарасов. Под ред. Д.А. Поспелова.- М.:Наука. Гл. ред. физ.-мат. лит., 1986.- 312 с.
177. Нечеткие множества и теория возможностей. Под ред. Р.Ягера. - М.: Радио и связь, 1986
178. Нильсон Н.. Принципы искусственного интеллекта. - М.: Радио и связь, 1985.
179. Обработка знаний. - М: Мир, 1990.

180. Обработка нечеткой информации в системах принятия решений./ А.Н.Борисов, А.В.Алексеев, и др..- М.: Радио и связь, 1989. – 304с.
181. Осипов Г.С. Приобретение знаний интеллектуальными системами. М., «Наука», 1997.
182. Осовский С.. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2002.
183. Пат. № 2000610164. РФ. Адаптивная система анализа и прогнозирования состояний сложных систем "ДЕЛЬТА". /В.С.Симанков (Россия), Е.В.Луценко (Россия); Заяв. № 2000610164. Оpubл. 03.03.2000. – 50с.
184. Пат. № 2003610433 РФ. Автоматизированная система мониторинга, анализа и прогнозирования развития сельхозкультур "ПРОГНОЗ-АГРО". / И. А. Драгавцева (Россия), Е.В.Луценко (Россия), Л.М.Лопатина (Россия); Заяв. № 2002611927 РФ. Оpubл. от 18.02.2003. – 50с.
185. Пат. № 2003610986 РФ. Универсальная когнитивная аналитическая система "ЭЙДОС" /Е.В.Луценко (Россия); Заяв. № 2003610510 РФ. Оpubл. от 22.04.2003. – 50с.
186. Пат. № 2003610987 РФ. Автоматизированная система комплексной обработки данных психологического тестирования "ЭЙДОС-Ψ" / Е.В.Луценко (Россия), С.Д.Некрасов (Россия); Заяв. № 2003610511 РФ. Оpubл. от 22.04.2003. – 50с.
187. Пат. № 2003620035 РФ. База данных автоматизированной системы мониторинга, анализа и прогнозирования развития сельхозкультур "ПРОГНОЗ-АГРО". / И. А. Драгавцева (Россия), Е.В.Луценко (Россия), Л.М.Лопатина (Россия); Заяв. № 2002620178 РФ. Оpubл. от 20.02.2003. – 50с.
188. Пат. № 940217. РФ. Универсальная автоматизированная система распознавания образов "ЭЙДОС". /Е.В.Луценко (Россия); Заяв. № 940103. Оpubл. 11.05.94. – 50с.
189. Пат. № 940328 РФ. Универсальная автоматизированная система анализа, мониторинга и прогнозирования состояний многопараметрических динамических систем "ЭЙДОС-Т". /Е.В.Луценко (Россия); Заяв. № 940324. Оpubл. 18.08.94. – 50с.
190. Пат. № 940334. РФ. Универсальная автоматизированная система анализа и прогнозирования ситуаций на фондовом рынке "ЭЙДОС-фонд" /Е.В.Луценко (Россия), Б.Х.Шульман (США); Заяв. № 940336. Оpubл. 23.08.94. – 50с.
191. Пенроуз Р.. Новый ум короля: о компьютерах, мышлении и законах физики. – М.: Едиториал УРСС, 2003. – 384с.
192. Перегудов Ф.И., Тарасенко Ф.П. Введение в системный анализ: Учебное пособие. – М.: Высшая школа, 1997. – 389с.
193. Петрик А.А., Лобанов В.Г., Симанков В.С., Луценко Е.В. Адаптивная система управления университетом (Тезисы) В сб.: "Информационные технологии и телекоммуникации в образовании". Тезисы докладов 2-й Международной выставки-конференции. – М.: ВВЦ, 2000. – С. 30-31.
194. Попов Э.В., И.Б. Фоминых, Е.Б. Кисель, М.Д.Шапот. Статические и динамические экспертные системы. - М.: Финансы и статистика, 1996.
195. Попов Э.В.. Общение с ЭВМ на естественном языке. - М.: Наука, 1986.
196. Попов Э.В.. Экспертные системы. - М.: Наука, 1987.
197. Поспелов Д.А.. Моделирование рассуждений. - М.: Радио и связь, 1989.
198. Построение экспертных систем. Под ред. Ф. Хейес-Рота, Д. Уотермена, Д. Лена-та. - М.: Мир, 1987.
199. Потапова Р.К.. Речевое управление роботом. – М.: Радио и связь, 1989.
200. Представление и использование знаний. - М: Мир, 1989.
201. Прибрам К.. Языки мозга. – М.: Прогресс, 1975.
202. Приобретение знаний. - М: Мир, 1990.

203. Психология машинного зрения. – Под ред. П. Уинстона, М.: Мир, 1978.
204. Райков А.Н. Аналитическим службам - информационные технологии. /Ваш выбор. 1994. № 4. - С.28-29.
205. Райков А.Н. Гносеологическая декомпозиция процессов рефлексивного управления. /"Рефлексивное управление". Тезисы международного симпозиума (17-19.10.2000). – М.: Ин-т психол. РАН, 2000. – С.89-90.
206. Райков А.Н. Интеллектуальные информационные технологии и системы. В 2-х частях. – М.: МИРЭА, 1998. – 213с.
207. Райков А.Н. Интеллектуальные информационные технологии: Учебное пособие. – М.: МГИРЭА(ТУ), 2000. - 96с.
208. Реальность и прогнозы искусственного интеллекта. -М.: Мир, 1987.
209. Розенблатт Ф. Принципы нейродинамики (перцептрон и теория механизмов мозга). // М.: "Мир", 1965.—480с.
210. Рубашкин В.Ш.. Представление и анализ смысла в интеллектуальных информационных системах. - М.: Наука, 1989.
211. Сайт: ИПУ РАН, Сектор-51 "Когнитивный анализ и моделирование ситуаций": <http://www.ipu.ru/labs/lab51/projects.htm>.
212. Сафронова Т.И., Луценко Е.В. Исследование семантической информационной модели управления качеством грунтовых вод на рисовых оросительных системах, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №05(7). – Режим доступа: <http://ej.kubagro.ru/2004/05/14/p14.asp>
213. Сафронова Т.И., Луценко Е.В. Когнитивная структуризация и формализация задачи управления качеством грунтовых вод на рисовых оросительных системах, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №05(7). – Режим доступа: <http://ej.kubagro.ru/2004/05/13/p13.asp>
214. Сафронова Т.И., Луценко Е.В. Проблема управления качеством грунтовых вод на рисовых оросительных системах и концепция ее решения, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №05(7). – Режим доступа: <http://ej.kubagro.ru/2004/05/11/p11.asp>
215. Сафронова Т.И., Луценко Е.В. Синтез, оптимизация и верификация семантической информационной модели управления качеством грунтовых вод на рисовых оросительных системах, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №05(7). – Режим доступа: <http://ej.kubagro.ru/2004/05/12/p12.asp>
216. Сигеру Омату, Марзуки Халид, Рубия Юсоф. Нейроуправление и его приложения. М.: ИПРЖР, 2000.
217. Симанков В.С., Луценко Е.В. Адаптивная автоматизированная система управления качеством обучения, В сб.: "Информационные технологии и системы". Материалы III Всероссийской. научно-технической конференции. – Воронеж: ГГТА, 1999. – С. 155-157.
218. Симанков В.С., Луценко Е.В. Адаптивная система анализа и прогнозирования состояний сложных систем "Дельта", В сб.: "Теория конфликта и ее приложения". Материалы I Всероссийской научно-технической конференции. – Воронеж: ВГТА, 2000. – С. 4-5.
219. Симанков В.С., Луценко Е.В. Адаптивное управление сложными системами на основе теории распознавания образов, Монография (научное издание). – Краснодар: ТУ КубГТУ, 1999. –318 с.
220. Симанков В.С., Луценко Е.В. Исследование эффективности управления обучением с применением адаптивной информационной модели, В сб.: "Компьютерные

- технологии в науке и образовании XXI века". Материалы III международной открытой сессий "Modus Academicus". – Ульяновск: УлГУ, 1999. – С. 242-246.
221. Симанков В.С., Луценко Е.В. Моделирование принятия решений в адаптивных АСУ сложными системами на основе теории информации, Информационные технологии. – 1999. – №2. – С. 8-14.
  222. Симанков В.С., Луценко Е.В. О требованиях, предъявляемых к тестовым заданиям и тестам по предметной обученности (Тезисы) В сб.: "Интеллектуальные информационные системы". Тезисы докладов 2-го Всероссийского совещания-семинара. – Воронеж: ВГТУ, 2000. – С. 35-36.
  223. Симанков В.С., Луценко Е.В. Применение кластерно-конструктивного анализа и теории информации для выявления конфликтующих коалиций в сложных организационных системах, В сб.: "Теория конфликта и ее приложения". Материалы I Всероссийской научно-технической конференции. – Воронеж: ВГТА, 2000. – С. 11-12.
  224. Симанков В.С., Луценко Е.В. Применение семантической теории информации для обобщения и сравнительно-типологического анализа исторического опыта стран, "Вековой поиск модели хозяйственного развития России". Тезисы докладов III-й международной научно-практической конференции. – Волжский: ВГУ, 2000. – С. 21-23.
  225. Симанков В.С., Луценко Е.В. Решение задач управления качеством обучения с применением интеллектуальной системы "Дельта", В сб.: "Информационные технологии и телекоммуникации в образовании". Тезисы докладов 2-й Международной выставки-конференции. – М.: ВВЦ, 2000. – С. 52-53.
  226. Симанков В.С., Луценко Е.В. Синтез адаптивных АСУ сложными системами с применением моделей распознавания образов, Автоматизация и современные технологии. – 1999. – №1. – С. 32-37.
  227. Симанков В.С., Луценко Е.В., Зайцев И.В. Синтез автоматизированных систем управления автономными комбинированными фотоветроэлектротехническими установками, Деп. в ВИНТИ 27.11.98, №3500-В98. – 22 с.
  228. Симанков В.С., Луценко Е.В., Зайцев И.В. Синтез адаптивных САУ ФВЭУ с применением моделей распознавания образов и принятия решений, основанных на теории информации (Тезисы) В сб.: "Компьютерные технологии в науке, проектировании и производстве". Тезисы докладов I-й Всероссийской научно-технической конференции. Часть 17. – Н.Новгород: НГТУ, 1999. – С. 12.
  229. Симанков В.С., Луценко Е.В., Лаптев В.Н. Системный анализ в адаптивном управлении, Монография (научное издание) / Под науч. ред. В.С.Симанкова. – Краснодар: ИСТЭК, 2001. – 258 с.
  230. Системы управления базами данных и знаний. Справочное издание под ред. А.Н. Наумова. - М.: Финансы и статистика, 1991.
  231. Словарь по кибернетике. /Под ред. В.М.Глушкова. – К.: Наукова думка, 1979. - 502с.
  232. Смирнов И., Безносюк Е., Журавлёв А. Психотехнологии: Компьютерный психо-семантический анализ и психокоррекция на неосознаваемом уровне. - М.: Изд. группа Прогресс-Культура, 1995. - 416с.
  233. Соьер Б., Фостер Д.Л. Программирование экспертных систем на Паскале. - М.: Финансы и статистика, 1990.
  234. Соломатин Н.М.. Информационные семантические системы. - // В уч. пос. Перспективы развития вычислительной техники в 11 кн. Кн. 1. - М.: Высшая школа, 1989.

235. Сотник С.Л. Конспект лекций по курсу "Основы проектирования систем искусственного интеллекта": (1997-1998), <http://neuroschool.narod.ru/books/sotnik.html>.
236. Статические и динамические экспертные системы. Э.В. Попов, И.Б. Фоминых, Е.Б. Кисель, М.Д. Шапот. – М.: Финансы и статистика, 1996.
237. Таунсенд К., Фохт Д. Проектирование и программная реализация экспертных систем на персональных ЭВМ. – М.: Финансы и статистика, 1990.
238. Тельнов Ю.Ф. Интеллектуальные информационные системы в экономике. (Учебное пособие) - М., 2002.
239. Тельнов Ю.Ф. Интеллектуальные информационные системы. (Учебное пособие) - М., 2000.
240. Темников Ф.Е., Афонин В.А., Дмитриев В.И. Теоретические основы информационной техники. – М.: Энергия, 1979. – 511с.
241. Терехов С.А. Лекции по теории и приложениям искусственных нейронных сетей. Лаборатория Искусственных Нейронных Сетей НТО-2, ВНИИТФ, Снежинск ([http://alife.narod.ru/lectures/neural/Neu_index.htm](http://alife.narod.ru/lectures/neural/Neu_index.htm)).
242. Техническая имитация интеллекта. - // В уч. пос. Робототехника и гибкие автоматизированные производства в 9 кн. Кн. 6. - М.: Высшая школа, 1986.
243. Техническое зрение роботов. – Под ред. А.Пью. М.: Машиностроение, 1987.
244. Тимофеев А.В.. Адаптивные робототехнические комплексы. – Л.: Машиностроение, 1988.
245. Ткачев А.Н., Луценко Е.В. Гуманистическая экономика, качество жизни и цели региональной администрации, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(6). – Режим доступа: <http://ej.kubagro.ru/2004/04/18/p18.asp>
246. Ткачев А.Н., Луценко Е.В. Исследование многоуровневой семантической информационной модели влияния инвестиций на уровень качества жизни населения региона, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(6). – Режим доступа: <http://ej.kubagro.ru/2004/04/19/p19.asp>
247. Ткачев А.Н., Луценко Е.В. Качество жизни населения, как интегральный критерий оценки эффективности деятельности региональной администрации, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №02(4). – Режим доступа: <http://ej.kubagro.ru/2004/02/14/p14.asp>
248. Ткачев А.Н., Луценко Е.В. Постановка задачи и синтез многоуровневой модели влияния инвестиций на качество жизни, Труды Кубанского государственного аграрного университета. 2002. Вып. 401 (429), юбилейный. – С. 314-326.
249. Ткачев А.Н., Луценко Е.В. Формальная постановка задачи и синтез многоуровневой семантической информационной модели влияния инвестиций на уровень качества жизни населения региона, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(6). – Режим доступа: <http://ej.kubagro.ru/2004/04/17/p17.asp>
250. Третьяк В.Г., Александров С.Г., Луценко Е.В. Модель профессионально-значимых качеств личности сотрудников ОВД, Вестник учебного отдела КЮИ МВД РФ. –2001. – №1. – С. 37-40.
251. Тыгу Э.Х.. Концептуальное программирование. - М.: Наука, 1984.
252. Убейко В.Н.. Экспертные системы. – М.: МАИ, 1992.
253. Уинстон П.. Искусственный интеллект. М.: Мир, 1980.
254. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. М.: Мир. 1992.
255. Уотерман Д.. Руководство по экспертным системам. - М.: Мир, 1989.
256. Фу К., Гонсалес Р., Ли К.. Робототехника. – М.: Мир, 1989.



257. Хагуров А.А., Луценко Е.В., Бакурадзе Л.А. Системный подход к социальному планированию развития города, Проблемы управления развитием городов: сборник докладов научно-практической конференции. – Краснодар: Краснодарский горисполком, 1985. – С. 53-61 (ДСП).
258. Хоггер К.. Введение в логическое программирование. М.: Мир, 1988.
259. Хорн Б.К.П.. Зрение роботов. – М.: Мир, 1989.
260. Хофман И. Активная память. – М.: Прогресс, 1986.
261. Хювенен Э., Сеппянен Й.. Мир Лиспа. В 2-х томах. - М.: Мир, 1990.
262. Цехмистро И.З. Поиски квантовой концепции физических оснований сознания. – Харьков: ХГУ, 1981. - 275с.
263. Частиков А.П., Гаврилова Т.А., Белов Д.Л.. Разработка экспертных систем. Среда CLIPS. – СПб: БХВ-Петербург, 2003.
264. Шагас Ч. Вызванные потенциалы мозга в норме и патологии. –М.: Мир, 1975. – 314с.
265. Шенк Р.. Обработка концептуальной информации. - М.: Энергия, 1980.
266. Шитиков В.К., Розенберг Г.С., Зинченко Т.Д. Количественная гидроэкология: методы системной идентификации. – Тольятти: ИЭВБ РАН, 2003. – 463 с.
267. Щукин Т.Н., Дорохов В. Б., Лебедев А.Н., Луценко Е.В. ЭЭГ прогноз успешности выполнения психомоторного теста при снижении уровня бодрствования: постановка задачи, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(6). – Режим доступа: <http://ej.kubagro.ru/2004/04/20/p20.asp>
268. Щукин Т.Н., Дорохов В. Б., Лебедев А.Н., Луценко Е.В. ЭЭГ прогноз успешности выполнения психомоторного теста при снижении уровня бодрствования: описание эксперимента, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(6). – Режим доступа: <http://ej.kubagro.ru/2004/04/21/p21.asp>
269. Щукин Т.Н., Дорохов В. Б., Лебедев А.Н., Луценко Е.В. ЭЭГ прогноз успешности выполнения психомоторного теста при снижении уровня бодрствования: анализ результатов исследования, Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №04(6). – Режим доступа: <http://ej.kubagro.ru/2004/04/22/p22.asp>
270. Экспертные системы. Принципы работы и примеры. Под ред. Р. Форсайта.- М.: Радио и связь, 1987.
271. Элти Дж., М. Кумбс. Экспертные системы. Концепции и примеры. – М.: Финансы и статистика, 1987.
272. Янг Дж.Ф.. Робототехника. – Л.: Машиностроение, 1979.
273. Ярушкина Н.Г.. Основы теории нечетких и гибридных систем. – М.: Финансы и статистика, 2004.

## INTERNET-САЙТЫ

Сайт автора учебного пособия: <http://Lc.Kubagro.ru>, выход на страничку о системе "Эйдос": <http://lc.kubagro.ru/aidos/eidos.htm>.

<http://ej.kubagro.ru>, статьи в электронном Научном журнале КубГАУ о применении системы "Эйдос" для решения задач СИИ в различных предметных областях: <http://ej.kubagro.ru/a/viewaut.asp?id=11>

**Базы данных репозитория UCI:**

<http://www.ics.uci.edu/~mlearn/MLRepository.html>

<http://datadiver.nw.ru/>

<http://www.dialog-21.ru>  
<http://alephegg.narod.ru/Refs/Diagonal.htm>  
<http://www.orc.ru/~stasson/neurox.html>  
<http://www.statsoft.ru/home/textbook/modules/stneunet.html>  
<http://www.finbridge.ru/net.shtml>  
<http://alife.narod.ru/>  
[http://www.businessstest.ru/default.asp?topic_id=3](http://www.businessstest.ru/default.asp?topic_id=3)

**А также сайты, обнаруживаемые в поисковых системах Yandex.ru, Aport.ru и Rambler.ru по запросам:**

- "Интеллектуальная обработка данных (data mining)";
- "Распознавание образов";
- "Поддержка принятия решений";
- "Экспертные системы".
- "Когнитивное моделирование";
- "Нейронные сети";
- "Генетические алгоритмы";
- "Моделирование эволюции (машинная эволюция)";
- "Клавиатурный почерк";
- "Биометрическая идентификация пользователя";
- "Биологическая обратная связь";
- "Семантический резонанс".



## КРАТКИЙ СЛОВАРЬ ТЕРМИНОВ ПО СК-АНАЛИЗУ И СИСТЕМАМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В данном небольшом толковом словаре по терминам СИИ мы ни в коей мере не претендуем на его полноту и исчерпывающий характер (да это и вряд ли возможно) и приводим лишь определения тех терминов, которые введены автором данного учебного пособия, а также тех, у которых автором изменены или модифицированы формулировки.

**Принятие решения** есть действие над множеством альтернатив, в результате которого исходное множество альтернатив сужается. Это действие называется "**выбор**".

**Экспертная система (ЭС)** – это программа, которая в определенных отношениях заменяет эксперта или группу экспертов в той или иной предметной области.

**Клавиатурный почерк** – система индивидуальных особенностей начертаний и динамики воспроизведения букв, слов и предложений на клавиатуре.

**Система, оснащенная интеллектуальным интерфейсом** – это система, способная вести себя по-разному в зависимости от результатов идентификации пользователя, его профессионального уровня и текущего психофизиологического состояния.

**Аутентификация** – это проверка, действительно ли пользователь является тем, за кого себя выдает. При этом пользователь должен предварительно сообщить о себе идентификационную информацию: свое имя и пароль, соответствующий названному имени.

**Идентификация** – это установление его личности.

**Почерк** – это система индивидуальных особенностей начертания и динамики воспроизведения букв, слов и предложений вручную различными людьми или на различных устройствах печати.

**Система с биологической обратной связью (БОС)** – это система, поведение которой зависит от психофизиологического (биологического) состояния пользователя.

**Система с семантическим резонансом** – это система, поведение которой зависит от состояния сознания пользователя и его психологической реакции на смысловые стимулы, в т.ч. неосознаваемые.

**Виртуальная реальность (ВР)** – модельная трехмерная (3D) окружающая среда, создаваемая компьютерными средствами и реалистично реагирующая на взаимодействие с пользователями.

**Эффект присутствия** – это создаваемая для пользователя иллюзия его присутствия в смоделированной компьютером среде, при этом создается полное впечатление "присутствия" в виртуальной среде, очень сходное с ощущением присутствия в обычном "реальном" мире.

**Система виртуальной реальности (ВР)** – это система, обеспечивающая:

- *генерацию полиперцептивной модели реальности* в соответствии с математической моделью этой реальности, реализованной в программной системе;

- *погружение пользователя в модель реальности* путем подачи на все или основные его перцептивные каналы – органы восприятия, программно-управляемых по величине и содержанию воздействий: зрительного, слухового, тактильного, термического, вкусового и обонятельного и других;

- *управление системой* путем использования виртуального "образа Я" пользователя и виртуальных органов управления системой (интерфейса), на которые он воздействует, представляющие собой зависящую от пользователя часть модели реальности;

- *реалистичную реакцию* моделируемой реальности на виртуальное воздействие и управление со стороны пользователя;

- *разрыв отождествления* пользователя со своим "Образом Я" из обычной реальности (*деперсонализация*), и отождествление себя с "виртуальным образом Я", генерируемым системой виртуальной реальности (*модификация сознания и самосознания пользователя*);

- *эффект присутствия* пользователя в моделируемой реальности в своем "виртуальном образе Я", т.е. эффект личного участия пользователя в наблюдаемых виртуальных событиях;

- *положительные результаты применения критериев реальности*, т.е. функциональную замкнутость и самодостаточность виртуальной реальности, вследствие чего никакими действиями внутри виртуальной реальности, осуществляемыми над ее объектами, в т.ч. объектами виртуального интерфейса, с помощью своего виртуального тела, невозможно установить, "истинная" эта реальность или виртуальная.

**Критерий реальности** – это самосогласованность реальности, т.е. получение одной и той же информации качественно различными способами и по различным каналам связи (принцип наблюдаемости):

- согласованность реальности самой с собой во времени;

- согласованность и взаимное подтверждение информации от различных органов восприятия, которые обычно реагируют на различные формы материи и часто являются парными (зрение, слух, обоняние) и расположенными в различных точках пространства.

**Принцип эквивалентности виртуальной и истинной реальности:** виртуальная система отсчета, локализованная в полнофункциональной виртуальной реальности полностью физически эквивалентна физической системе отсчета, локализованной в "истинной реальности", т.е. никакими действиями внутри виртуальной реальности, осуществляемыми над ее объектами, в т.ч. объектами виртуального интерфейса, с помощью своего виртуального тела, невозможно установить, "истинная" эта реальность или виртуальная (Луценко Е.В., 2004).

**Взвешивание данных или ремонт обучающей выборки** – это операция, в результате которой частное распределение объектов по классам в обучающей выборке максимально, на сколько это возможно, приближается либо к частотному распределению генеральной совокупности (если оно известно из независимых источников), либо к равномерному.

**Обобщение** – это операция формирования обобщенных образов классов на основе описаний конкретных объектов, входящих в обучающую выборку.

**Распознавание** – это операция сравнения и определения степени сходства образа данного конкретного объекта с образами других конкретных объектов или с обобщенными образами классов, в результате которой формируется рейтинг объектов или классов по убыванию сходства с распознаваемым объектом.

**Обучение с учителем** – это процесс формирования обобщенных образов классов, на основе обучающей выборки, содержащей характеристики конкретных объектов как в описательных, так и в классификационных шкалах и градациях.

**Шкала** – это способ классификации объектов по наименованиям или степени выраженности некоторого свойства. Понятие шкалы тесно связано с ключевым понятием когнитивной психологии: понятием конструкта, более того, практически является синонимом или формальным аналогом этого понятия.

**Градация** – это положение на шкале (или интервал, диапазон), соответствующее наименованию или определенной степени выраженности свойства.

**Обучение без учителя или самообучение** – это процесс формирования обобщенных образов классов, на основе обучающей выборки, содержащей характеристики конкретных объектов, причем только в описательных шкалах и градациях.

**Верификация модели** – это операция установления степени ее адекватности (валидности) путем сравнения результатов идентификации конкретных объектов с их фактической принадлежностью к обобщенным образам классов.

**Адаптация модели** – это количественное уточнение модели, не требующее изменения классификационных и описательных шкал и градаций, а лишь объема обучающей выборки.

**Синтез (или повторный синтез – пересинтез) модели** – качественное уточнение модели, путем учета в модели объектов и факторов, ранее не входивших ни в обучающую выборку, ни в генеральную совокупность, по отношению к которой данная обучающая выборка репрезентативна.

**Кластеризация** – это операция автоматической классификации, в ходе которой объекты объединяются в группы (кластеры) таким образом, что внутри групп различия между объектами минимальны, а между группами – максимальны. При этом в ходе кластеризации не только определяется состав кластеров, но и сам их набор и границы.

**Системный анализ** – современный метод теоретического познания и программно-целевого управления, в котором объект управления рассматривается как *система*.

**Система** – представляет собой совокупность *элементов* различных уровней иерархии (подсистемы), *связанных* между собой каналами взаимодействия, объединенных в единое структурно-функциональное целое, обеспечивающее им преимущества в достижении общей *цели* системы и целей подсистем за счет *системного эффекта*.

**Системный (эмерджентный) эффект** – наличие у системы качественно новых, *эмерджентных свойств*, которые не сводятся к сумме свойств ее частей.

**Эмерджентность свойств** – так как все свойства есть свойства тех или иных систем, то можно предположить, что все свойства без исключения имеют эмерджентную природу, т.е. любое свойство основано на уровне Реальности этим свойством не обладающим. Например, свойство "быть соленым" основано на свойствах Na и Cl, по отдельности этим свойством ни в коей мере не обладающими.

**Уровень системности** – степень отличия свойств системы от суммы свойств ее частей.

**Диалектика: структура-свойство-отношение** – подсистемы различных уровней иерархии могут рассматриваться с внешней точки зрения как неделимое целое, не имеющее частей, т.е. элемент, обладающий *свойством* вступать во взаимодействие с другими элементами, а с внутренней точки зрения – как имеющие *структуру*, состоящую из элементов более низкого уровня иерархии, объединенных определенными видами *взаимосвязей*. Сам канал взаимодействия может рассматриваться как *отношение* элементов, которые с помощью него взаимодействуют, или как система, обладающая определенной *структурой*, включающая среду передачи и объекты, перемещающиеся в этой среде и переносящие субстанцию взаимодействия, например: вещество, стоимость, энергию или информацию.

**Системно-когнитивный анализ (СК-анализ)** – системный анализ, структурированный по базовым когнитивным операциям.

**Автоматизированный системно-когнитивный анализ (АСК-анализ)** – автоматизированный СК-анализ, т.е. системный анализ, автоматизированный путем структурирования по базовым когнитивным операциям системного анализа (*БКОСА*) и включающий: формализуемую когнитивную концепцию, математическую модель, методику численных расчетов и реализующий их программный инструментарий, в качестве которого в настоящее время выступает универсальная когнитивная аналитическая система "Эйдос". АСК-анализ предложен в 2002 году Е.В.Луценко.

#### **Компоненты АСК-анализа:**

- формализуемая когнитивная концепция и следующий из нее когнитивный конфигуратор;
- теоретические основы, методология, технология и методика СК-анализа;
- математическая модель СК-анализа, основанная на системном обобщении семантической меры целесообразности информации А. Харкевича;
- методика численных расчетов, в универсальной форме реализующая математическую модель СК-анализа, включающая иерархическую структуру данных и 24 детальных алгоритма 10 БКОСА;
- специальное инструментальное программное обеспечение, реализующее математическую модель и численный метод СК-анализа – Универсальная когнитивная аналитическая система "Эйдос";
- методика, технология и результаты синтеза рефлексивных АСУ активными объектами на основе АСК-анализа.

#### **Этапы АСК-анализа –**

- 1) когнитивная структуризация предметной области;

2) формализация предметной области (конструирование классификационных и описательных шкал и градаций);

3) подготовка обучающей выборки (ввод данных мониторинга в базу прецедентов);

4) синтез семантической информационной модели (СИМ);

5) оптимизация СИМ;

6) проверка адекватности СИМ (измерение внутренней и внешней, дифференциальной и интегральной валидности);

7) системно-когнитивный анализ СИМ, исследование моделируемого объекта путем исследования его модели:

– решение задач идентификации и прогнозирования;

– генерация информационных портретов классов и факторов, т.е. решение обратной задачи прогнозирования, поддержка принятия решений по управлению (результаты отображаются в графической форме двухмерных и трехмерных профилей классов и факторов);

– кластерно-конструктивный анализ классов и факторов (результаты отображаются в форме семантических сетей классов и факторов);

– содержательное сравнение классов и факторов (результаты отображаются в форме когнитивных диаграмм классов и факторов);

– изучение системы детерминации состояний моделируемого объекта, не-локальные нейроны и интерпретируемые нейронные сети;

– построение классических когнитивных моделей (когнитивных карт).

– построение интегральных когнитивных моделей (интегральных когнитивных карт).

**Математическая модель АСК-анализа** основана на системной теории информации (СТИ).

**Системная теория информации (СТИ)** – Отличия СТИ от классической теории информации Больцмана-Найквиста-Хартли-Шеннона обусловлены отличиями понятия "*система*" от понятия "*множество*". СТИ рассматривает в качестве элементов не только первичные элементы множества, но и элементы, представляющие собой подсистемы различных уровней иерархии, образующиеся за счет *взаимодействия* первичных элементов, а также учитывает понятие *цели*. В рамках СТИ предложено системное обобщение семантической меры информации Харкевича, которое удовлетворяет принципу соответствия с мерой Хартли в детерминистском случае, как и мера Шеннона в случае равновероятных событий, чем преодолена несогласованность семантической теории информации и классической теории информации Шеннона. Так как данная мера учитывает понятие цели, то она является количественной мерой знаний. В рамках СТИ предложены гипотезы "О возрастании эмерджентности", следующие из нее: "О природе сложности системы", и "О видах системной информации".

**Гипотеза "О возрастания эмерджентности"**: "Чем больше элементов в системе, тем большую долю содержащейся в ней информации составляет информация, содержащаяся во взаимосвязях ее элементов".

**Гипотеза "О природе сложности системы"**: сложность системы определяется количеством содержащейся в ней информации.

**Гипотеза "О видах системной информации":** системная информация включает две составляющие:

- зависящую от количества элементов системы;
- зависящую как от количества элементов системы, так и от сложности взаимосвязей между ними.

**Методика численных расчетов АСК-анализа** включает: структуры входных данных (формализация предметной области), промежуточных и выходных данных, а также алгоритмы БКОСА.

**Программный инструментарий АСК-анализа** – универсальная когнитивная аналитическая система "Эйдос".

**Базовые когнитивные операции системного анализа (БКОСА)** – когнитивные операции, образующие *когнитивный конфигуратор*.

**Когнитивный конфигуратор** – минимальный полный набор познавательных (когнитивных от: "cognition" – "познание", англ.) операций, к которым сводятся различные процессы познания, в т.ч. системный анализ, как метод познания. В *формализуемой когнитивной концепции* выявлено 10 таких операций, каждая из которых достаточно элементарна для формализации и программной реализации: 1) присвоение имен; 2) восприятие; 3) обобщение (синтез, индукция); 4) абстрагирование; 5) оценка адекватности модели; 6) сравнение, идентификация и прогнозирование; 7) дедукция и абдукция; 8) классификация и генерация конструкторов; 9) содержательное сравнение; 10) планирование и принятие решений об управлении.

**Формализуемая когнитивная концепция** – когнитивная концепция, предложенная с целью разработки СК-анализа. Из данной концепции выводятся структура когнитивного конфигулятора, система базовых когнитивных операций и обобщенная схема системного анализа, структурированного до уровня базовых когнитивных операций (СК-анализ).

Рассматривает процесс познания, как многоуровневую иерархическую систему обработки информации в которой когнитивные структуры каждого уровня являются результатом интеграции структур предыдущего уровня. На 1-м уровне этой системы находятся дискретные элементы потока чувственного восприятия, которые на 2-м уровне интегрируются в чувственный образ конкретного объекта. Те, в свою очередь, на 3-м уровне интегрируются в обобщенные образы классов и факторов, образующие на 4-м уровне кластеры, а на 5-м конструкторы. Система конструкторов на 6-м уровне образуют текущую парадигму реальности (т.е. человек познает мир путем синтеза и применения конструкторов). На 7-м же уровне обнаруживается, что текущая парадигма не единственно-возможная.

Ключевым для когнитивной концепции является понятие *факта*, под которым понимается соответствие дискретного и интегрального элементов познания (т.е. элементов разных уровней интеграции-иерархии), обнаруженное на опыте. *Факт* рассматривается как квант *смысла*, что является основой для его формализации. Таким образом, происхождение смысла связывается со своего рода "разностью потенциалов", существующей между смежными уровнями интеграции-иерархии обработки информации в процессах познания. Между когнитивными структурами разных уровней иерархии существует отношение "дискретное – интегральное". Именно это служит основой формализации смысла.

**АСК-анализ в экономике** – новый математический и инструментальный метод экономики, характеризующийся универсальной непараметрической математической моделью, основанной на системной теории информации, наличием методики численных расчетов и программного инструментария (система "Эйдос").

**Адекватность модели** – это ее способность правильно идентифицировать объекты. Понятие адекватности имеет свою структуру, включающую понятия внутренней и внешней, дифференциальной и интегральной *валидности* ("валидность" и переводится как "правильность").

**Внутренняя валидность** – способность модели правильно идентифицировать объекты, входящие в обучающую выборку.

**Внешняя валидность** – способность модели верно идентифицировать объекты, не входящие в обучающую выборку, но относящиеся к генеральной совокупности, по отношению к которой она репрезентативна.

**Интегральная валидность** – это валидность, средневзвешенная по всей обучающей выборке.

**Дифференциальная валидность** – это способность правильно идентифицировать отдельные классы.

**Инструментарий СК-анализа** – универсальная когнитивная аналитическая система "Эйдос".

**Показатели валидности:**

– *идентифицировано верно* – это количество объектов обучающей выборки, которые идентифицированы как классы, к которым они действительно относятся;

– *идентифицировано ошибочно* – это количество объектов обучающей выборки, которые идентифицированы как классы, к которым они в действительности не относятся (ошибка идентификации);

– *неидентифицировано верно* – это количество объектов обучающей выборки, которые неидентифицированы как классы, к которым они действительно не относятся;

– *неидентифицировано ошибочно* – это количество объектов обучающей выборки, которые неидентифицированы как классы, к которым они в действительности относятся (ошибка неидентификации).

**Идентификация** – количественная оценка степени сходства конкретного объекта или его состояния с классом по признакам, которые относятся к тому же моменту времени, что и состояние.

**Прогнозирование** – количественная оценка степени сходства конкретного объекта или его состояния с классом по признакам, причем признаки относятся к более раннему времени, чем состояние.

**Информационный портрет класса** – это список факторов, ранжированных в порядке убывания силы их влияния на переход объекта управления в состояние, соответствующее данному классу. Информационный портрет класса отражает систему его детерминации. Генерация информационного портрета класса представляет собой решение обратной задачи прогнозирования, т.к. при прогнозировании по системе факторов определяется спектр наиболее вероятных будущих состояний объекта управления, в которые он может перейти под влия-

нием данной системы факторов, а в информационном портрете мы наоборот, по заданному будущему состоянию объекта управления определяем систему факторов, детерминирующих это состояние, т.е. вызывающих переход объекта управления в это состояние. В начале информационного портрета класса идут факторы, оказывающие положительное влияние на переход объекта управления в заданное состояние, затем факторы, не оказывающие на это существенного влияния, и далее – факторы, препятствующие переходу объекта управления в это состояние (в порядке возрастания силы препятствования). Информационные портреты классов могут быть *отфильтрованы* по диапазону факторов, т.е. мы можем отобразить влияние на переход объекта управления в данное состояние не всех отраженных в модели факторов, а только тех, коды которых попадают в определенный диапазон, например, относящиеся к определенным описательным шкалам.

**Информационный (семантический) портрет фактора** – это список классов, ранжированный в порядке убывания силы влияния данного фактора на переход объекта управления в состояния, соответствующие данным классам. Информационный портрет фактора называется также его *семантическим портретом*, т.к. в соответствии с концепцией смысла системно-когнитивного анализа, являющейся обобщением концепции смысла Шенка-Абельсона, *смысл фактора состоит в том, какие будущие состояния объекта управления он детерминирует*. Сначала в этом списке идут состояния объекта управления, на переход в которые данный фактор оказывает наибольшее влияние, затем состояния, на которые данный фактор не оказывает существенного влияния, и далее состояния – переходу в которые данный фактор препятствует. Информационные портреты факторов могут быть *отфильтрованы* по диапазону классов, т.е. мы можем отобразить влияние данного фактора на переход объекта управления не во все возможные будущие состояния, а только в состояния, коды которых попадают в определенный диапазон, например, относящиеся к определенным классификационным шкалам.

**Функция влияния** представляет собой график зависимости вероятностей перехода объекта управления в будущие состояния под влиянием различных значений некоторого фактора. Если взять несколько информационных портретов факторов, соответствующих градациям одной описательной шкалы, отфильтровать их по диапазону градаций некоторой классификационной шкалы и взять из каждого информационного портрета по одному состоянию, на переход в которое объекта управления данная градация фактора оказывает наибольшее влияние, то мы и получим зависимость, отражающую вероятность перехода объекта управления в будущие состояния под влиянием различных значений некоторого фактора, т.е. функцию влияния. Функции влияния являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой "Эйдос". Необходимо отметить, что на вид функций влияния математической моделью СК-анализа не накладывается никаких ограничений, в частности, они могут быть и *нелинейные*.

**Двухмерные и трехмерные профили классов и факторов** – это графические диаграммы, отображающие силу и направление влияния различных факторов на переход объекта управления в различные состояния. *Профилем класса*



называется графическое отображение столбца матрицы информативностей, соответствующего данному классу. *Профилем признака (фактора)* называется графическое отображение строки матрицы информативностей, соответствующего данному признаку. *Информативности факторов* означают силу и направление влияния данного фактора на переход системы в состояние, соответствующее данному классу.

**Кластерно-конструктивный анализ** – это математический метод анализа данных, обеспечивающий:

- выявление классов, наиболее сходных по системе их детерминации и объединение их в кластеры;
- выявление кластеров классов, наиболее сильно отличающиеся по системе их детерминации и построение из них полюсов конструкторов классов, при этом остальные кластеры включаются в конструкторы в качестве промежуточных между полюсами;
- выявление факторов, наиболее сходных по детерминируемым ими классам и объединение их в кластеры;
- выявление кластеров факторов, наиболее сильно отличающиеся по детерминируемым ими классам и построение из них полюсов конструкторов факторов, при этом остальные кластеры включаются в конструкторы в качестве промежуточных между полюсами.

Состояния объекта управления, соответствующие классам, включенным в один кластер, могут быть достигнуты одновременно, т.е. являются *совместимыми (коалиционными)* по детерминирующим их факторам. Состояния объекта управления, соответствующие классам, образующим полюса конструктора, не могут быть достигнуты одновременно, т.е. являются противоположными по детерминирующим их факторам (*антагонистическими*).

Факторы, включенные в один кластер, оказывают сходное влияние на поведение объекта управления и могут, при необходимости, быть использованы для замены друг друга. Факторы, образующие полюса конструктора, оказывают противоположное влияние на поведение объекта управления.

*Кластерно-конструктивный анализ классов* позволяет сравнить их по сходству системы детерминации и отобразить эту информацию в наглядной графической форме семантической сети классов.

*Кластерно-конструктивный анализ факторов* позволяет сравнить факторы по сходству их влияния на переход объекта в будущие состояния и отобразить эту информацию в наглядной графической форме семантической сети факторов.

**Когнитивные диаграммы классов (факторов)** – это графические диаграммы, позволяющие отобразить в чем конкретно состоит сходство и различие любых двух классов (или любых двух факторов), т.е. детально увидеть структуру каждой линии связи в семантической сети. Когнитивные диаграммы представляет собой графическое изображение обобщенного коэффициента корреляции профилей классов (или факторов), при этом каждая линия, вносящая вклад в сходство или различие соответствует одному слагаемому, ее цвет – знаку, а толщина – модулю этого слагаемого.

**Нелокальные нейроны и интерпретируемые нейронные сети** позволяют в наглядной форме отобразить систему детерминации будущих состояний. *Нелокальный нейрон* представляет собой будущее состояние объекта управления с изображением наиболее сильно влияющих на него факторов с указанием силы и направления (способствует-препятствует) их влияния. Нейронная сеть представляет собой совокупность взаимосвязанных нейронов. В классических нейронных сетях связь между нейронами осуществляется по входным и выходным сигналам, а в нелокальных нейронных сетях – на основе общего информационного поля, реализуемого семантической информационной моделью. Система "Эйдос" обеспечивает построение любого подмножества многослойной нейронной сети с заданными или выбираемыми по заданным критериям рецепторами и нейронами, связанными друг с другом связями любого уровня опосредованности.

**Классические когнитивные карты** являются графической формой представления фрагментов СИМ, объединяющей достоинства таких форм, как нейроны и семантические сети факторов. Классическая когнитивная карта представляет собой нейрон, соответствующий некоторому состоянию объекта управления с рецепторами, каждый из которых соответствует фактору в определенной степени способствующему или препятствующему переходу объекта в это состояние. Рецепторы соединены связями как с нейроном, так и друг с другом. Связи рецепторов с нейроном отражают силу и направление влияния факторов, а связи рецепторов друг с другом, отображаемые в форме семантической сети факторов, – сходство и различие между рецепторами по характеру их влияния на объект управления. Таким образом, классическая когнитивная карта представляет собой нейрон с семантической сетью факторов, изображенные на одной диаграмме.

**Обобщенные когнитивные карты** позволяют объединить в одной графической форме семантические сети классов и факторов, объединенных нейронной сетью. Если объединить несколько классических когнитивных карт на одной диаграмме и изобразить на ней также связи между нейронами в форме семантической сети классов, то получим обобщенную (интегральную) когнитивную карту. Система "Эйдос" обеспечивает построение любого подмножества многоуровневой семантической информационной модели с заданными или выбираемыми по заданным критериям рецепторами и нейронами, связанными друг с другом связями любого уровня опосредованности в форме классических и обобщенных когнитивных карт. В частности, в системе полуавтоматически формируется задание на генерацию подмножеств обобщенной когнитивной карты.

**Луценко Евгений Вениаминович**

профессор, доктор экономических наук, кандидат технических наук,  
профессор кафедры компьютерных технологий и систем  
Кубанского государственного аграрного университета

# **ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ**

Допущено

Министерством сельского хозяйства Российской Федерации  
в качестве учебного пособия  
для студентов высших учебных заведений,  
обучающихся по специальности  
080801 "Прикладная информатика (по областям)"

Литературный редактор: Авторская правка  
Оригинал-макет: Е. В. Луценко

Лицензия ИД № 02334 от 14 июля 2000 г.

Подписано в печать 03.05.2004. Формат 60×84. Бумага типографская.

Печ. л.: – 52,3 Заказ № _____ Тираж _____ экз.

Отпечатано в типографии Кубанского государственного аграрного университета  
350044, г. Краснодар, ул. Калинина, 13