

УДК 303.732.4

UDC 303.732.4

**МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ
ВЫЯВЛЕНИЯ, ПРЕДСТАВЛЕНИЯ И
ИСПОЛЬЗОВАНИЯ ЗНАНИЙ В АСК-
АНАЛИЗЕ И ИНТЕЛЛЕКТУАЛЬНОЙ
СИСТЕМЕ «ЭЙДОС»****METHODOLOGICAL ASPECTS OF
DETECTION, REPRESENTATION AND USAGE
OF KNOWLEDGE IN COMPUTERIZED
SYSTEM-COGNITIVE ANALYSIS AND
INTELLECTUAL "EIDOS" SYSTEM**

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор
Кубанский государственный аграрный универси-
тет, Россия, 350044, Краснодар, Калинина, 13,
prof.lutsenko@gmail.com

Lutsenko Evgeny Veniaminovich
Dr. Sci.Econ., Cand.Tech.Sci., professor
Kuban State Agrarian University, Krasnodar, Russia

В статье на небольшом и наглядном численном примере подробно рассматриваются методологические аспекты технологии выявления знаний из эмпирических данных, представления знаний и их использования для решения задач прогнозирования, принятия решений и исследования предметной области в системно-когнитивном анализе (СК-анализ) и его программном инструментарии – интеллектуальной системе «Эйдос»

In this article, on a small and evident numerical example, methodological aspects of a process engineering of detection of knowledge from the trial-and-error data explicitly are considered, representation of knowledge and its usage for problem solving of forecasting, decision making and data domain examination in system-cognitive analysis (SC-analysis) and its programmatic toolkit - intellectual "Eidos" system are shown

Ключевые слова: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, БАЗА ЗНАНИЙ, КОГНИТИВНАЯ ФУНКЦИЯ, СЕМАНТИЧЕСКАЯ ИНФОРМАЦИОННАЯ МОДЕЛЬ

Keywords: COMPUTERIZED SYSTEM-COGNITIVE ANALYSIS, KNOWLEDGE BASE, COGNITIVE FUNCTION, SEMANTIC INFORMATION MODEL

**“Истинное знание – это знание причин”
Френсис Бэкон (1561–1626 гг.)**

Интеллектуальные системы – это автоматизированные системы, обеспечивающие *выявление* знаний из эмпирических данных, хранение и накопление их в различных формах *представления*, а также их *использование* для решения различных задач. Современный уровень развития теории и практики искусственного интеллекта и динамика развития этого научного и технологического направления таковы, что, по-видимому, можно обоснованно говорить о его затяжном кризисе, более того, о том, что его развитие возможно пошло по тупиковому пути. У автора есть развитые конкретные глубоко аргументированные представления о путях выхода из этого кризиса, связанные с решением ключевых вопросов о том, может ли мыслить объект и какими структурами поддерживается функция мышления у людей и какими структурами она в принципе может поддерживаться в технических системах. Парадоксальность ситуации заключается в том, что, казалось бы, само собой разумеющиеся и очевидные ответы на эти «простые» вопросы, скорее всего, являются *неверными*. Однако обсуждение этих путей и вопросов далеко выходит за рамки данной работы, т.к. требуют углубленного анализа закономерностей развития человека, технологии и общества, а также некоторых изменений в современных мировоззренческих концепциях и научных парадигмах¹.

¹ <http://ru.wikipedia.org/wiki/Парадигма>

Здесь же отметим лишь, что выявление, представление и использование знаний безусловно является *проблемой*, и в различных интеллектуальных системах эта проблема решаются (или не решаются) по-разному².

Далее рассмотрим вариант решения этой проблемы в автоматизированном системно-когнитивном анализе (АСК-анализ) и его программном инструментарии – интеллектуальной системе «Эйдос».

Прежде всего, кратко рассмотрим соотношение содержания понятий: «данные», «информация» и «знания».

Данные – это информация, рассматриваемая безотносительно к ее смысловому содержанию, находящаяся на носителях или в каналах связи и представленная в определенной системе кодирования или на определенном языке (т.е. в формализованном виде).

Информация – это *осмысленные* данные. Смысл, семантика, содержание (согласно концепции смысла Шенка-Абельсона [6]) – это знание причинно-следственных зависимостей.

Знания – это информация, *полезная* для достижения целей (рисунок 1).

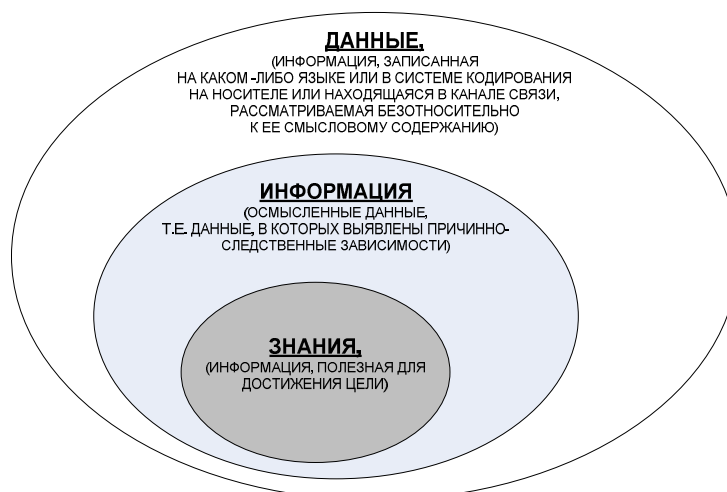


Рисунок 1. Соотношение содержания понятий: «данные», «информация», «знания»

Знания могут быть представлены в различных формах, характеризующихся различной *степенью формализации*:

- вообще неформализованные знания, т.е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);
- знания, формализованные в естественном вербальном языке;
- знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними;
- знания в форме технологий, организационных производственных, социально-экономических и политических структур;
- знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных си-

² См., например: <http://www.aiportal.ru/>

стемах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно **последовательно повышать степень формализации** исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

- преобразовать исходные данные в информацию;
- преобразовать информацию в знания;
- использовать знания для решения задач прогнозирования, принятия решений и исследования предметной области.

Для этого в АСК-анализе предусмотрены следующие этапы [2]:

1. Когнитивная структуризация предметной области, при которой определяется, что мы хотим прогнозировать и на основе чего (конструирование классификационных и описательных шкал).

2. Формализация предметной области (8):

- разработка градаций классификационных и описательных шкал (номинального, порядкового и числового типа);
- использование разработанных на предыдущих этапах классификационных и описательных шкал и градаций для формального описания (кодирования) исследуемой выборки.

3. Синтез и верификация (оценка степени адекватности) модели.

4. Если модель адекватна, то ее использование для решения задач идентификации, прогнозирования и принятия решений, а также для исследования моделируемой предметной области.

Рассмотрим, как реализуются эти этапы на простом наглядном примере, который положен в основу лабораторной работы №1 по дисциплине: «Интеллектуальные информационные системы», преподаваемой автором в Кубанском государственном аграрном университете [14]. Этот пример интересен тем, что рассматриваемая в нем задача легко решается также и с помощью «естественного интеллекта», что позволяет сравнить результаты его работы с работой автоматизированной интеллектуальной системы и увидеть как она работает.

Данная задача взята из книги Д.Мичи и Р.Джонстона "Компьютер – творец" [1], (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 (рисунок 2).

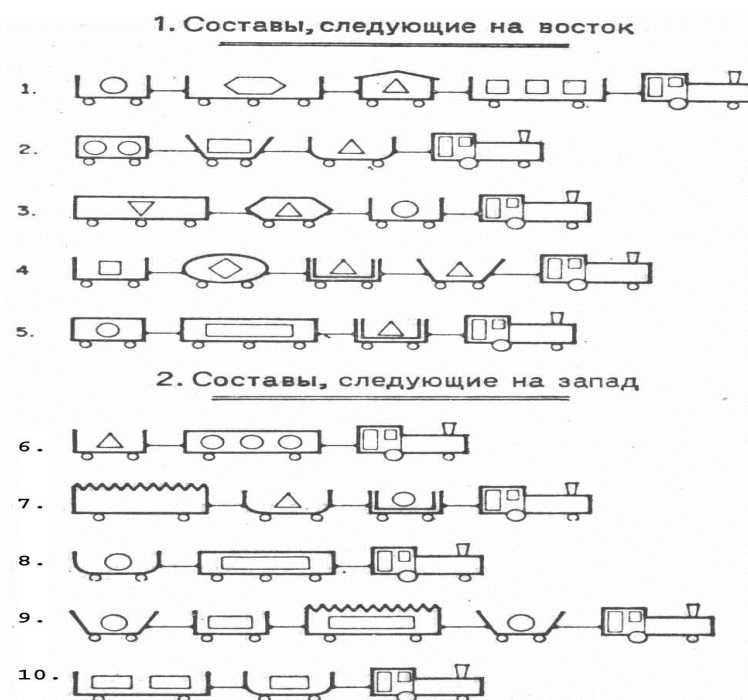


Рисунок 2. Исходные данные по примеру в графическом виде

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 (фактов), представляющих в совокупности исследуемую выборку.
2. Каждый из объектов исследуемой выборки представляет собой систему, имеющую сложную многоуровневую структуру признаков (экстенционально описание).
3. Для каждого из объектов исследуемой выборки известно, к каким обобщенным категориям (классам) он относится (интенционально описание).
4. Необходимо сформировать модель, обеспечивающую идентификацию объектов по их признакам, т.е. 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.

Совокупность экстенционального и интенционального описания каждого объекта, по сути, представляет собой его *определение* через подведе-

ние под более общее понятие и выделение специфических признаков. Например, так определяется понятие «млекопитающее»: это животное (более общее понятие), выкармливающее своих детей молоком (специфический признак). На основе ряда определений конкретных объектов путем их *обобщения* можно получить определения классов. Если привести в качестве примеров исследуемой выборки множество различных животных, как млекопитающих, так и других, каждый из таких примеров определить множеством признаков и построить модель, то окажется, что наиболее характерным признаком млекопитающих является не наличие шерсти или когтей, а именно вскармливание детенышей молоком.

Первым делом *вручную*³ преобразуем исходные данные из графической формы, представленной на рисунке 2, в форму Excel-таблицы исходных данных (таблица 1):

Таблица 1 – EXCEL-ТАБЛИЦА ИСХОДНЫХ ДАННЫХ

Источник информации	Классификационные шкалы		Описательные шкалы						
	Состав следует на	Наименование состава	Форма вагона	Длина вагона	Количество осей вагона	Грузоподъемность вагона	Вид стенок вагона	Вид крыши вагона	Вид груза (кол-во и вид)
Сост-01, var-1	ВОСТОК	Состав-01	Прямоугольная	Короткий	2	40,0	Одинарные	Отсутствует	1 овал
Сост-02, var-1	ВОСТОК	Состав-02	Прямоугольная	Короткий	2	40,0	Одинарные	Прямая	2 овала
Сост-03, var-1	ВОСТОК	Состав-03	Прямоугольная	Длинный	3	80,0	Одинарные	Прямая	1 перевернутый треугольник
Сост-04, var-1	ВОСТОК	Состав-04	Прямоугольная	Короткий	2	40,0	Одинарные	Отсутствует	1 квадрат
Сост-05, var-1	ВОСТОК	Состав-05	Прямоугольная	Короткий	2	40,0	Одинарные	Прямая	1 овал
Сост-06, var-1	ЗАПАД	Состав-06	Прямоугольная	Короткий	2	40,0	Одинарные	Отсутствует	1 треугольник
Сост-07, var-1	ЗАПАД	Состав-07	Прямоугольная	Длинный	2	60,0	Одинарные	Гофрированная	Отсутствует
Сост-08, var-1	ЗАПАД	Состав-08	U-образная	Короткий	2	30,0	Одинарные	Отсутствует	1 овал
Сост-09, var-1	ЗАПАД	Состав-09	V-образная	Короткий	2	30,0	Одинарные	Отсутствует	1 овал
Сост-10, var-1	ЗАПАД	Состав-10	Прямоугольная	Длинный	2	60,0	Одинарные	Отсутствует	2 прямоугольника
Сост-01, var-2	ВОСТОК	Состав-01	Прямоугольная	Длинный	3	80,0	Одинарные	Отсутствует	1 ромб
Сост-02, var-2	ВОСТОК	Состав-02	V-образная	Короткий	2	30,0	Одинарные	Отсутствует	1 прямоугольник
Сост-03, var-2	ВОСТОК	Состав-03	Ромбовидная	Короткий	2	40,0	Одинарные	Прямая	1 треугольник
Сост-04, var-2	ВОСТОК	Состав-04	Овальная	Короткий	2	40,0	Одинарные	Овальная	1 ромб
Сост-05, var-2	ВОСТОК	Состав-05	Прямоугольная	Длинный	3	80,0	Одинарные	Прямая	1 длинный прямоугольник
Сост-06, var-2	ЗАПАД	Состав-06	Прямоугольная	Длинный	2	60,0	Одинарные	Прямая	3 овала
Сост-07, var-2	ЗАПАД	Состав-07	U-образная	Короткий	2	30,0	Одинарные	Отсутствует	1 треугольник
Сост-08, var-2	ЗАПАД	Состав-08	Прямоугольная	Длинный	3	80,0	Одинарные	Прямая	1 длинный прямоугольник
Сост-09, var-2	ЗАПАД	Состав-09	Прямоугольная	Короткий	2	40,0	Одинарные	Отсутствует	1 прямоугольник
Сост-10, var-2	ЗАПАД	Состав-10	U-образная	Короткий	2	30,0	Одинарные	Отсутствует	1 прямоугольник
Сост-01, var-3	ВОСТОК	Состав-01	Прямоугольная	Короткий	2	40,0	Одинарные	Треугольная	1 треугольник
Сост-02, var-3	ВОСТОК	Состав-02	U-образная	Короткий	2	30,0	Одинарные	Отсутствует	1 треугольник
Сост-03, var-3	ВОСТОК	Состав-03	Прямоугольная	Короткий	2	40,0	Одинарные	Отсутствует	1 овал
Сост-04, var-3	ВОСТОК	Состав-04	Прямоугольная	Короткий	2	40,0	Двойные	Отсутствует	1 треугольник
Сост-05, var-3	ВОСТОК	Состав-05	Прямоугольная	Короткий	2	40,0	Двойные	Отсутствует	1 треугольник
Сост-07, var-3	ЗАПАД	Состав-07	Прямоугольная	Короткий	2	40,0	Двойные	Отсутствует	1 овал
Сост-09, var-3	ЗАПАД	Состав-09	Прямоугольная	Длинный	2	60,0	Одинарные	Гофрированная	1 длинный прямоугольник
Сост-01, var-4	ВОСТОК	Состав-01	Прямоугольная	Длинный	2	60,0	Одинарные	Отсутствует	3 квадрата
Сост-04, var-4	ВОСТОК	Состав-04	U-образная	Короткий	2	30,0	Одинарные	Отсутствует	1 треугольник
Сост-09, var-4	ЗАПАД	Состав-09	V-образная	Короткий	2	30,0	Одинарные	Отсутствует	1 овал

³ То, что в данном случае преобразование исходных данных из графической формы в табличную осуществляется *вручную* не является каким-либо ограничением, т.к. нет никаких принципиальных проблем автоматизировать подобное преобразование.

Итак, исходные данные по задаче – это Excel-таблица №1.

Процедура преобразования исходных данных в информацию – это анализ данных, состоящий из двух шагов:

- выявление в исходных данных *фактов* или *событий*;
- выявление причинно-следственных связей (зависимостей) между этими событиями.

Фактически для преобразования исходных данных в информацию необходимо:

1. Разработать классификационные и описательные шкалы и градации.

2. С использованием классификационных и описательных шкал и градаций *закодировать* исходные данные, в результате чего получится обучающая выборка, состоящая из *фактов*, представляющих собой примеры в единстве экстенционального и интенционального описания.

3. Произвести расчет матриц абсолютных частот, условных и безусловных процентных распределений и матрицы информативностей, отражающей причинно-следственные связи между значениями факторов и принадлежностью объектов к классам.

Таким образом, информация по задаче – это исходные данные плюс классификационные и описательные шкалы и градации, обучающая выборка, а также матрицы частот, процентных распределений и информативностей.

Процедура преобразования информации в знания – это оценка полезности информации для достижения цели.

Значит знания по задаче – это информация плюс цель и оценка степени полезности информации для достижения этой цели.

Знания получаются из информации, когда мы классифицируем будущие состояния объекта управления как желательные (целевые) и нежелательные.

Банк данных – это базы данных плюс система *управления базами данных* (СУБД) (стандартные термины). СУБД – это, по сути, *система управления данными*.

Информационный банк – это информационные базы плюс информационные системы (предлагается стандартизировать эти термины). Информационная система – это, по сути, *система управления информацией*.

Банк знаний – это базы знаний плюс интеллектуальные системы (стандартные термины). Интеллектуальная система – это, по сути, *система управления знаниями*.

Существует очевидная параллель между терминами и понятиями, связанными с данными, информацией и знаниями, наглядно представленная в таблице 2.

Таблица 2 – ПАРАЛЛЕЛЬ МЕЖДУ ПОНЯТИЯМИ И ТЕРМИНАМИ, КАСАЮЩИМИСЯ ДАННЫХ, ИНФОРМАЦИИ И ЗНАНИЙ

Объект	Субъект	Система
База данных (БД)	Система управления базами данных (СУБД)	Банк данных=БД+СУБД
Информационная база (ИБ)	Информационная система (система управления информационными базами – СУИБ)	Информационный банк=ИБ+СУИБ
База знаний (БЗ)	Интеллектуальная система (система управления базами знаний – СУБЗ)	Банк знаний=БЗ+СУБЗ

Автор предлагает «узаконить», т.е. стандартизировать термины, отмеченные в таблице 2 красным цветом. Это позволит упорядочить все эти термины в единой стройной системе, построенной на основе соотношения содержания понятий «данные», «информация» и «знания».

Это актуально, т.к. в настоящее время существуют явная путаница в использовании этих понятий, встречающаяся даже в названиях соответствующих дисциплин: «Управление знаниями», «Интеллектуальные информационные системы», «Представление знаний в информационных системах». Например, дисциплина «Управление знаниями» является *гуманитарной* и в ней изучаются слабо формализованные, не основанные на применении автоматизированных интеллектуальных систем, этапы, формы и методы управления знаниями⁴. Вместе с тем название этой дисциплины явно соотносится с названием дисциплины «Управление данными». Интеллектуальные системы часто некорректно называются интеллектуальными информационными системами, с тем же успехом их можно было бы называть: «Интеллектуальные СУБД», но лучше и правильнее было бы называть их как предложено: «Системы управления базами знаний». Дисциплина «Алгоритмы и структуры данных» соотносится с дисциплиной «Представление знаний в информационных системах», хотя ясно, что они представляются не в информационных, а в интеллектуальных системах. В настоящее время дисциплина «Интеллектуальные информационные системы» по своему содержанию включает «Представление знаний в информационных системах», тогда как из вышеизложенного ясно, что они должны соотноситься по своему содержанию также, как СУБД и «Модели баз данных» (в которых обычно преподается лишь одна реляционная модель). Отметим также, что если применить определение знаний к моделям, описываемым в дис-

⁴ Типичные вопросы, изучаемые в этой дисциплине: стратегия управления знаниями предприятия; организационная культура в контексте управления знаниями; измерение интеллектуального капитала; корпоративные знания: как ими управлять; интеграция знаний предприятия; бизнес держится на знаниях, сам того не зная; новые программы корпоративного обучения в среде управления знаниями: опыт зарубежных компаний; менеджмент знаний: подход к внедрению; общепринятых заблуждений об управлении знаниями (knowledge management)

циплине «Представление знаний в информационных системах», то обнаруживается, что иногда в ней описываются не модели баз знаний, а модели баз данных или информационные модели. В частности это видно на примере семантических сетей, которые, по сути, представляют собой инфологическую модель реляционной базы данных.

По мнению автора дисциплины «Управление знаниями» и «Представление знаний в интеллектуальных системах» по сути, представляют собой две части одной дисциплины и должны отражать не способы управления знаниями различной степени формализации (как в настоящее время), а описание автоматизированных интеллектуальных систем и баз знаний.

Существует дисциплина: «Алгоритмы и структуры данных». Предлагается ввести аналогичные дисциплины: «Алгоритмы и информационные структуры» (в АСК-анализе – это формализация предметной области и синтез модели) и «Алгоритмы структурирования знаний» (по содержанию близко к когнитологии, инженерии знаний, представлению знаний)».

Факт наличия причинно-следственных зависимостей может быть установлен методом хи-квадрат, а ее вид – многофакторным анализом. Однако факторный анализ позволяет обрабатывать данные лишь очень небольших размерностей (по числу факторов) и предъявляет чрезвычайно жесткие требования к наличию полных повторностей всех вариантов сочетаний факторов в исходных данных (т.е. данные не должны быть фрагментарными), что на практике выполнить удается крайне редко.

Поэтому большой интерес представляют другие подходы к решению **задачи** выявления в эмпирических данных причинно-следственных зависимостей и их вида, отражения выявленных зависимостей в наглядной графической и аналитической форме.

Рассмотрим вариант решения этой задачи, развиваемый в СК-анализе и реализованный в системе Эйдос».

Для этого сформулируем **требования** к форме представления данных, информации и знаний, позволяющие оценить *степень их пригодности* для решения задач прогнозирования и принятия решений, а также исследования предметной области (например, кластерного анализа).

Прежде всего, результаты решения вышеперечисленных задач должны быть **инвариантны** относительно:

- *единиц измерения* градаций факторов (признаков);
- *типов шкал*, используемых для формализации классов и факторов (номинальные, порядковые и числовые);
- *различных статистических характеристик исходной выборки*: частотных распределений объектов по классам (обобщенным категориям), частотных распределений градаций факторов, различий в количестве при-

знаков в описаниях объектов исследуемой выборки, различий в суммарном количестве признаков по классам.

Кроме того, форма представления должна обеспечивать решение вышеперечисленных задач с минимальными дополнительными затратами ручного труда, а это значит, что *вся предварительная обработка должна быть максимально автоматизирована.*

Эти требования можно рассматривать и как *критерии* выбора наиболее подходящей для решения вышеперечисленных задач формы представления данных, информации и знаний.

Рассмотрим **влияние единиц измерения в исходной выборке на результаты решения задач** прогнозирования и принятия решений, а также исследования предметной области (например, кластерного анализа).

Если в исходных данных какие-то значения выражены в больших единицах измерения, то их числовые значения будут малыми, и наоборот, если единицы измерения мелкие, то числовые значения – большие. Большие значения оказывают большее влияние на результаты математической обработки, чем малые, и *это приводит к возникновению зависимости результатов решения задач идентификации, прогнозирования и принятия решений, а также кластерного анализа, от выбранных размерностей исходных данных, что, на взгляд автора, совершенно неприемлемо и указывает на то, что такое решение нельзя признать корректным и даже вообще решением.* По этой же причине некорректно совместно обрабатывать сами исходные данные, представленные в **различных** единицах измерения (натуральных или ценовых), например, складывать расстояния, представленные в километрах и в метрах, а затем прибавлять к ним тонны и килограммы, а затем еще и безразмерные величины. Вроде это очевидно, но, как это ни удивительно, но как показывает опыт на практике это довольно часто делается, а потом еще на основе подобного «анализа» делаются и выводы. Очень странно, что обычно на это *не обращают никакого внимания* при использовании исходных данных, представленных в различных единицах измерения. Например, даже в таких популярных (причем, совершенно заслуженно) системах, как SPSS, в подсистеме кластерного анализа приводятся примеры кластерного анализа над исходными данными, представленными в различных единицах измерения.

Для решения поставленной задачи в АСК-анализе проводится *последовательное повышение степени формализации исходных данных до уровня, обеспечивающего их обработку на компьютере в программной системе.* После выполнения когнитивной структуризации и формализации предметной области осуществляется синтез модели.

Рассмотрим на нашем простом примере, как осуществляется формализация предметной области и преобразование исходных данных в информацию и знания.

В системе «Эйдос» есть подсистема _15, содержащая большое количество различных программных интерфейсов для импорта в систему «Эйдос» исходных данных из внешних баз данных различных стандартов (рисунок 3):

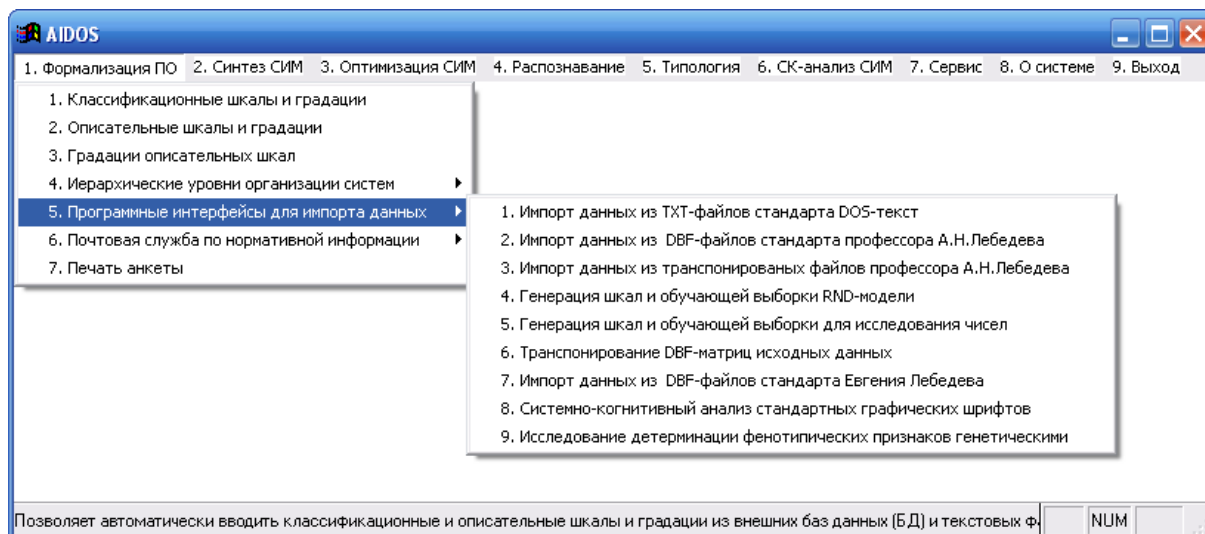


Рисунок 3. Меню выхода на подсистему _15 системы «Эйдос»

Для импорта исходных из таблиц, стандарта таблицы 1 и автоматизированной *формализации предметной области* служит программный интерфейс _152. На рисунке 4 приведен Help этого режима, в котором разъясняются требования к файлу исходных данных (поэтому в тексте мы повторять их не будем), а на рисунке 5 – меню задания параметров импорта данных из внешних баз данных в систему «Эйдос». В первой экранной форме на рисунке 5 задаются параметры преобразования, а на второй приведена таблица, характеризующая модель, которая будет создана в результате применения этих параметров. Если пользователя что-либо не устраивает в этих результатах, то он имеет возможность скорректировать параметры преобразования.

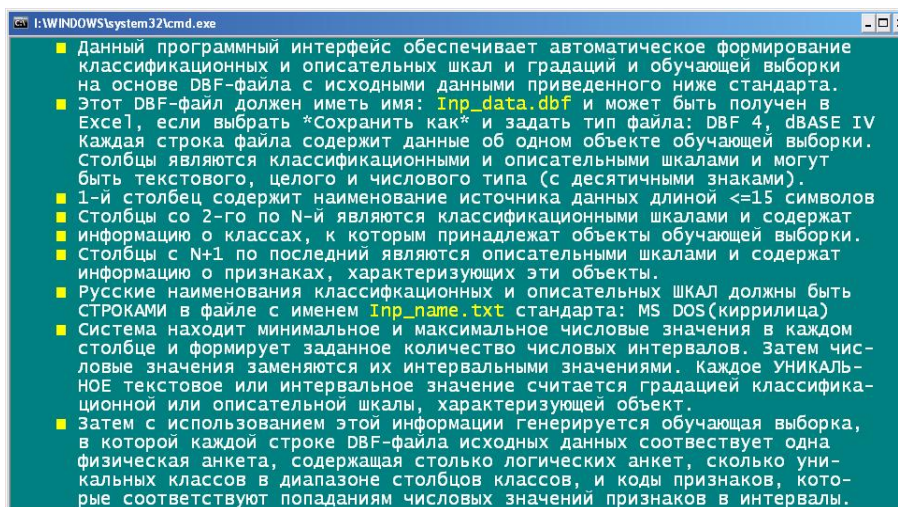


Рисунок 4. Help режима _152 системы «Эйдос»

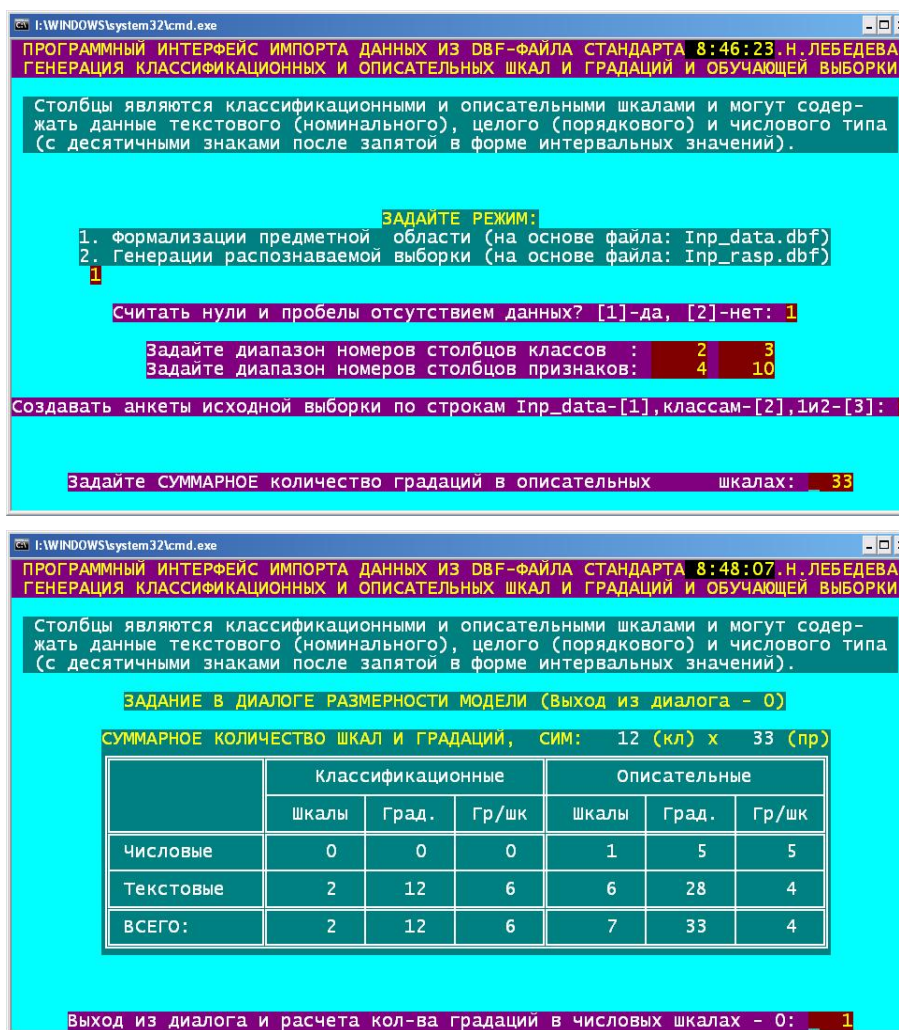


Рисунок 5. Меню режима _152 задания параметров импорта данных из внешних баз данных в систему «Эйдос»

В результате работы режима _152 системы «Эйдос» на основе заданных параметров модели автоматически формируются справочники классификационных и описательных шкал и градаций номинального (текстового), порядкового (целочисленного) и числового типа (последние – в форме интервальных значений), а также исходная (обучающая) выборка (таблицы 3-7). На шкалах номинально типа определены отношения только тождества и различия. На порядковых шкалах, кроме того, определены отношения больше и меньше между градациями. На числовых шкалах, кроме того, есть начало отсчета и единица измерения и над градациями определены все арифметические операции. Необходимо отметить, что формализация числовых значений в виде интервальных значений является вполне естественной и хорошо обоснованной, т.к. фактически результатом измерения является не просто число, а число, заданное с определенной точностью или погрешностью, т.е. относящееся к некоторому интервалу. Количество интервалов на числовой шкале должно определяться таким образом, чтобы все они были представлены как минимум 5 примерами, что

считается минимальной статистикой. Это значит, что если исследуемая выборка мала, то интервалы должны быть велики, и погрешность модели будет вынужденно велика, и наоборот, если по мере увеличения объема исходной выборки интервалы могут быть уменьшены и точность модели возрастает. Иначе говоря, невозможно точно отразить предметную область, если у нас недостаточно данных. Это можно считать каким-то вариантом теоремы Котельникова об отсчетах⁵.

Таблица 3 – СПРАВОЧНИК КЛАССИФИКАЦИОННЫХ ШКАЛ

KOD	NAME
1	СОСТАВ СЛЕДУЕТ НА
2	НАИМЕНОВАНИЕ СОСТАВА

Таблица 4 – СПРАВОЧНИК КЛАССИФИКАЦИОННЫХ ШКАЛ И ГРАДАЦИЙ

KOD	NAME
1	СОСТАВ СЛЕДУЕТ НА-ВОСТОК
2	СОСТАВ СЛЕДУЕТ НА-ЗАПАД
3	НАИМЕНОВАНИЕ СОСТАВА-Состав-01
4	НАИМЕНОВАНИЕ СОСТАВА-Состав-02
5	НАИМЕНОВАНИЕ СОСТАВА-Состав-03
6	НАИМЕНОВАНИЕ СОСТАВА-Состав-04
7	НАИМЕНОВАНИЕ СОСТАВА-Состав-05
8	НАИМЕНОВАНИЕ СОСТАВА-Состав-06
9	НАИМЕНОВАНИЕ СОСТАВА-Состав-07
10	НАИМЕНОВАНИЕ СОСТАВА-Состав-08
11	НАИМЕНОВАНИЕ СОСТАВА-Состав-09
12	НАИМЕНОВАНИЕ СОСТАВА-Состав-10

Градации второй классификационной шкалы, т.е. все градации с 3-й по 12-ю удалены вручную, т.к. интересует не определение номера состава, а его идентификация с обобщенными образами классов составов идущих на восток и на запад.

Таблица 5 – СПРАВОЧНИК ОПИСАТЕЛЬНЫХ ШКАЛ

KOD	NAME
1	ФОРМА ВАГОНА
2	ДЛИНА ВАГОНА
3	КОЛИЧЕСТВО ОСЕЙ ВАГОНА
4	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА
5	ВИД СТЕНОК ВАГОНА
6	ВИД КРЫШИ ВАГОНА
7	ВИД ГРУЗА (КОЛ-ВО И ВИД)
8	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ

Восьмая шкала введена *вручную*, т.к. соответствующие признаки являются признаками второго уровня иерархии, если рассматривать состав, как систему, т.е. это не признаки вагонов, а признаки состава в целом.

⁵ <http://ru.wikipedia.org/wiki/Теорема%20Котельникова>

Таблица 6 – СПРАВОЧНИК ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ

KOD	NAME	Примечание: тип шкалы
1	ФОРМА ВАГОНА-U-образная	Номинальный (текстовый)
2	ФОРМА ВАГОНА-V-образная	
3	ФОРМА ВАГОНА-Овальная	
4	ФОРМА ВАГОНА-Прямоугольная	
5	ФОРМА ВАГОНА-Ромбовидная	
6	ДЛИНА ВАГОНА-Длинный	Номинальный (текстовый)
7	ДЛИНА ВАГОНА-Короткий	
8	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-002	Порядковый (целочисленный)
9	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-003	
10	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00}	Числовой (интервальные значения)
11	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00}	
12	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00}	
13	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00}	
14	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00}	
15	ВИД СТЕНОК ВАГОНА-Двойные	Номинальный (текстовый)
16	ВИД СТЕНОК ВАГОНА-Одинарные	
17	ВИД КРЫШИ ВАГОНА-Гофрированная	
18	ВИД КРЫШИ ВАГОНА-Овальная	
19	ВИД КРЫШИ ВАГОНА-Отсутствует	
20	ВИД КРЫШИ ВАГОНА-Прямая	
21	ВИД КРЫШИ ВАГОНА-Треугольная	
22	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 длинный прямоугольник	Номинальный (текстовый)
23	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 квадрат	
24	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 овал	
25	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 перевернутый треугольник	
26	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 прямоугольник	
27	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 ромб	
28	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 треугольник	
29	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 овала	
30	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 прямоугольника	
31	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 квадрата	
32	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 овала	
33	ВИД ГРУЗА (КОЛ-ВО И ВИД)-Отсутствует	

Таблица 7 – ИСХОДНАЯ (ОБУЧАЮЩАЯ) ВЫБОРКА

Код объекта	Наименование объекта	Коды классов	Коды признаков										
			4	7	8	10	11	16	19	24	4	6	9
33	Состав-01	1	4	7	8	10	11	16	19	24	4	6	9
			16	19	27	4	7	8	10	11	16	21	28
			6	8	12	13	16	19	31	36			
34	Состав-02	1	4	7	8	10	11	16	20	29	2	7	8
			16	19	26	1	7	8	10	16	19	28	35
			35										
35	Состав-03	1	4	6	9	14	16	20	25	5	7	8	10
			16	20	28	4	7	8	10	11	16	19	24
			35										
36	Состав-04	1	4	7	8	10	11	16	19	23	3	7	8
			11	16	18	27	4	7	8	10	11	15	19
			1	7	8	10	16	19	28	36			
37	Состав-05	1	4	7	8	10	11	16	20	24	4	6	9
			16	20	22	4	7	8	10	11	15	19	28
			35										
38	Состав-06	2	4	7	8	10	11	16	19	28	4	6	8
			13	16	20	32	34						
			35										
39	Состав-07	2	4	6	8	12	13	16	17	33	1	7	8
			16	19	28	4	7	8	10	11	15	19	24
			35										
40	Состав-08	2	1	7	8	10	16	19	24	4	6	9	14
			20	22	34								
			35										
41	Состав-09	2	2	7	8	10	16	19	24	4	7	8	10
			16	19	26	4	6	8	12	13	16	17	22
			7	8	10	16	19	24	36				
42	Состав-10	2	4	6	8	12	13	16	19	30	1	7	8
			16	19	26	34							
			35										

Обучающая выборка состоит из трех таблиц баз данных:

- первая включает коды объектов выборки и наименование источника данных;
- вторая содержит коды классов, к которым принадлежит объект;
- третья содержит коды признаков объекта.

Первая таблица связана со второй и третьей отношением «один ко многим».

Объекты исходной выборки формируются путем кодирования строк таблицы исходных данных (таблица 1) с применением справочников классификационных и описательных шкал и градаций (таблицы 4 и 6), но при этом могут формироваться и объединенные объекты из строк по классам. Суммарное количество классификационных и описательных шкал, с которым работает режим _152, не ограничено, но на практике составляет не более 256, что связано с ограничением MS Excel 2003. В более поздних версиях MS Excel это ограничение снято, но из них исключен XLS-DBF-конвертер. Нет никаких принципиальных проблем снять все эти ограничения и путем разработки небольших специализированных программ, объединяющих листы MS Excel (что и делалось при необходимости автором) или использования имеющихся конвертеров. Суммарное количество градаций классификационных шкал, как и градаций описательных шкал в текущей версии системы «Эйдос» ограничено 4000, но в будущих версиях это ограничение *планируется* снять.

База данных, представленная в таблице 1, рассматриваемая *совместно* с таблицами 3, 4, 5, 6 и 7, является *результатом формализации* предметной области.

После формализации предметной области осуществляется синтез и верификация (оценка достоверности) модели, а также повышение ее эффективности [2]. Синтез модели включает расчет на основе эмпирических данных, представленных в исследуемой выборке, следующих матриц (таблицы 8, 9, 10):

- матрицы абсолютных частот (большинство статических систем этим и ограничиваются);
- матрицы условных и безусловных процентных распределений (в некоторых системах это также делается);
- матрицы информативностей или матрицы знаний (что осуществляется только в АСК-анализе).

Таблица 8 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ

Код	Наименование	Восток	Запад	Сумма
1	ФОРМА ВАГОНА-У-образная	2	3	5
2	ФОРМА ВАГОНА-V-образная	1	1	2
3	ФОРМА ВАГОНА-Овальная	1		1
4	ФОРМА ВАГОНА-Прямоугольная	11	8	19
5	ФОРМА ВАГОНА-Ромбовидная	1		1
6	ДЛИНА ВАГОНА-Длинный	4	5	9
7	ДЛИНА ВАГОНА-Короткий	13	8	21
8	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-002	14	12	26
9	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-003	3	1	4
10	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00}	11	6	17
11	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00}	9	2	11
12	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00}	1	3	4
13	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00}	1	4	5
14	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00}	1	1	2
15	ВИД СТЕНОК ВАГОНА-Двойные	2	1	3
16	ВИД СТЕНОК ВАГОНА-Одинарные	15	11	26
17	ВИД КРЫШИ ВАГОНА-Гофрированная		2	2
18	ВИД КРЫШИ ВАГОНА-Овальная	1		1
19	ВИД КРЫШИ ВАГОНА-Отсутствует	10	9	19
20	ВИД КРЫШИ ВАГОНА-Прямая	5	2	7
21	ВИД КРЫШИ ВАГОНА-Треугольная	1		1
22	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 длинный прямоугольник	1	2	3
23	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 квадрат	1		1
24	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 овал	3	4	7
25	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 перевернутый треугольник	1		1
26	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 прямоугольник	1	2	3
27	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 ромб	2		2
28	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 треугольник	5	2	7
29	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 овала	1		1
30	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 прямоугольника		1	1
31	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 квадрата	1		1
32	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 овала		1	1
33	ВИД ГРУЗА (КОЛ-ВО И ВИД)-Отсутствует		1	1
34	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-2		3	3
35	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-3	3	1	4
36	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-4	2	1	3
	Кол-во объектов обуч.выборки	5	5	10

Матрица абсолютных частот (таблица 1) по сути, является таблицей сопряженности⁶. Таблица 1, рассматриваемая *совместно* с таблицами 3, 4, 5, 6, 7, а также 8, 9 и 10, в соответствии с терминологией, предлагаемой в таблице 2, является уже не базой данных, а *информационной базой*.

Если же среди классов выделить целевые и нежелательные, то таблица 10 может рассматриваться уже как *база знаний*, т.к. содержит количественные оценки степени полезности (и вредности) информации для достижения целей.

Рассмотрим, используя вышеперечисленные критерии, в какой степени эти матрицы *пригодны* для решения задач прогнозирования и приня-

⁶ См.: <http://www.machinelearning.ru/wiki/index.php?title=Таблица%20сопряженности>

тия решений, а также исследования предметной области (например, кластерного анализа) и какую работу необходимо выполнять вручную и автоматизировать, чтобы повысить их пригодность для этого.

Матрица абсолютных частот отражает, сколько раз каждая градация факторов встречается у объектов каждого класса.

Проблема размерностей при расчете матрицы абсолютных частот решается тем, что *сами размерные исходные данные с использованием шкал различных типов (номинальных, порядковых и числовых) заменяются на факты их встречи*, т.е. на частоты встреч тех или иных их *интервальных значений* [31] в различных группах, соответствующих классам. Фактом является наблюдение определенного экстенционального значения (признака, градации фактора) у объекта исходной выборки, относящегося к некоторой интенциональной категории (классу).

Однако вышеперечисленные задачи решать на основе абсолютных частот можно только в том случае, если по каждому классу в исходных данных было приведено *одинаковое* количество примеров, что на практике встречается крайне редко и является трудно достижимым при сборе исходных данных, за исключением случая жестко спланированного управляемого эксперимента (обычно очень небольшой размерности). Можно, конечно, вручную учитывать это различие, однако реально это возможно сделать только на моделях очень небольшой размерности и требует специальных усилий (работы).

Чтобы результаты решения вышеперечисленных задач *не зависели* от количества примеров по разным классам (т.е. были *инвариантны* относительно формы частотных распределений примеров по классам, частотного распределения признаков и др.) можно с помощью формул (1) перейти от матрицы абсолютных частот к матрице условных и безусловных процентных распределений (матрице *относительных частот* или частостей⁷) (таблица 9).

$$P_{ij} = \frac{N_{ij}}{N_i}; N_i = \sum_{j=1}^W N_{ij}; \quad (1)$$

где:

N_{ij} – суммарное количество наблюдений факта: "действовал i -й фактор и объект перешел в j -е состояние";

N_i – суммарное количество встреч i -го фактора у всех объектов;

W – количество классов (мощность множества будущих состояний объекта управления).

⁷ **Частота** (абсолютная частота) – количество элементов совокупности, которые имеют данное значение признака. **Частость** (относительная частота) – отношение частоты к общему количеству исследуемых элементов, т.е. объему совокупности.

Таблица 9 – МАТРИЦА УСЛОВНЫХ И БЕЗУСЛОВНЫХ ПРОЦЕНТНЫХ РАСПРЕДЕЛЕНИЙ (СИМ-2)

Код	Наименование	В группе: «Восток»	В группе: «Запад»	По всей выборке
1	ФОРМА ВАГОНА-У-образная	40,0	60,0	50,0
2	ФОРМА ВАГОНА-У-образная	20,0	20,0	20,0
3	ФОРМА ВАГОНА-Овальная	20,0		10,0
4	ФОРМА ВАГОНА-Прямоугольная	220,0	160,0	190,0
5	ФОРМА ВАГОНА-Ромбовидная	20,0		10,0
6	ДЛИНА ВАГОНА-Длинный	80,0	100,0	90,0
7	ДЛИНА ВАГОНА-Короткий	260,0	160,0	210,0
8	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-002	280,0	240,0	260,0
9	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-003	60,0	20,0	40,0
10	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00}	220,0	120,0	170,0
11	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00}	180,0	40,0	110,0
12	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00}	20,0	60,0	40,0
13	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00}	20,0	80,0	50,0
14	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00}	20,0	20,0	20,0
15	ВИД СТЕНОК ВАГОНА-Двойные	40,0	20,0	30,0
16	ВИД СТЕНОК ВАГОНА-Одинарные	300,0	220,0	260,0
17	ВИД КРЫШИ ВАГОНА-Гофрированная		40,0	20,0
18	ВИД КРЫШИ ВАГОНА-Овальная	20,0		10,0
19	ВИД КРЫШИ ВАГОНА-Отсутствует	200,0	180,0	190,0
20	ВИД КРЫШИ ВАГОНА-Прямая	100,0	40,0	70,0
21	ВИД КРЫШИ ВАГОНА-Треугольная	20,0		10,0
22	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 длинный прямоугольник	20,0	40,0	30,0
23	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 квадрат	20,0		10,0
24	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 овал	60,0	80,0	70,0
25	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 перевернутый треугольник	20,0		10,0
26	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 прямоугольник	20,0	40,0	30,0
27	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 ромб	40,0		20,0
28	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 треугольник	100,0	40,0	70,0
29	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 овала	20,0		10,0
30	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 прямоугольника		20,0	10,0
31	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 квадрата	20,0		10,0
32	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 овала		20,0	10,0
33	ВИД ГРУЗА (КОЛ-ВО И ВИД)-Отсутствует		20,0	10,0
34	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-2		60,0	30,0
35	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-3	60,0	20,0	40,0
36	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-4	40,0	20,0	30,0
	Кол-во объектов обуч.выборки	5	5	10

При неограниченном увеличении объема выборки частоты стремятся (сходятся) к теоретическим вероятностям, как своим пределам, поэтому частоты можно считать эмпирическими вероятностями. Способ, которым частоты приближаются к вероятностям, называется *сходимостью* модели. В системе «Эйдос» реализован специальный режим, позволяющий исследовать *сходимость модели*, в том числе *скорость сходимости* и погрешность различия частоты и вероятности при различных объемах исследуемой выборки. Учитывая все это при достаточно больших выборках, по мнению авторов, допустимо вместо термина «частота» использовать термин «условная вероятность», тем более что в аналитических выражениях обычно оперируют именно вероятностями⁸.

⁸ Процентные распределения отличаются от вероятностных тем, что вероятности умножены на 100.

Однако и при решении вышеперечисленных задач на основе матрицы условных и безусловных процентных распределений приходится **вручную осуществлять сравнение** условных относительных частот, является определенной **работой** и реально возможно только на моделях очень малой размерности и требует довольно больших специальных усилий. Поэтому есть смысл **автоматизировать** и это сравнение, так, чтобы в нашем распоряжении была матрица, содержащая уже сами **результаты** сравнения условных относительных частот в количественной форме.

Для того чтобы реализовать эту автоматизацию необходимо **выбрать базу сравнения и способ сравнения**, т.е. ответить на два вопроса:

– с чем **сравнивать** условные относительные частоты: друг с другом или с безусловными частотами;

– **каким способом сравнивать** условные относительные частоты: с помощью вычитания или с помощью деления.

Если в модели есть всего два класса, то можно сравнивать условные относительные частоты как друг с другом, так и с безусловными частотами, т.к. это одинаково как по трудоемкости (затрачиваемым вычислительным ресурсам), так и по результатам сравнения. Если же в модели хотя бы три класса, то уже возникают определенные затруднения в том, как сравнить условные процентные распределения по ним, а если их сотни или тысячи, то это становится даже в теоретическом плане непонятным. Поэтому в [2] предлагается **использовать в качестве базы для сравнения (нормы) условных относительных частот их взвешенное среднее по всей исследуемой выборке или безусловные частоты (2)**:

$$I_{ij} = \text{Log}_2 \frac{P_{ij}}{P_i} \quad (2)$$

где:

– P_{ij} – вероятность перехода объекта в j -е состояние при условии действия на него i -го значения фактора;

– P_i – вероятность наблюдения i -го значения фактора по всей выборке;

Подставим в выражение (2), представляющее собой вариант формулы А.Харкевича, значения вероятностей, выраженные через фактически наблюдаемые абсолютные частоты из (3)

$$P_{ij} = \frac{N_{ij}}{N_j}; P_i = \frac{N_i}{N}; N_i = \sum_{j=1}^M N_{ij}; N_j = \sum_{i=1}^W N_{ij}; N = \sum_{i=1}^W \sum_{j=1}^M N_{ij} \quad (3)$$

где:

W – количество классов (мощность множества будущих состояний объекта управления)

M – максимальный уровень сложности смешанных состояний объекта управления;

N_{ij} – суммарное количество наблюдений факта: "действовал i -й фактор и объект перешел в j -е состояние";

N_j – в СИМ-1 (семантической информационной модели) суммарное количество встреч различных факторов у объектов, перешедших в j -е состояние;

N_i – в суммарное количество встреч i -го фактора у всех объектов;

N – суммарное количество встреч различных факторов у всех объектов.

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}N}{N_iN_j} \right) \quad (4)$$

Выражение (4) дает количество информации о принадлежности объекта к j -му классу, если он обладает i -м признаком, выраженное через абсолютные частоты из таблицы 8.

Это решение, приводящее к выражению (2), соответствует принятому в статистике методу средних и отклонений от средних и представляет собой косвенное или опосредованное *сравнение* условных процентных распределений друг с другом, т.к. база сравнения рассчитывается с их использованием.

Кроме того, важно отметить, что получающееся в результате этого выражение (4) для количества информации **тождественно** выражению (6), получающемуся путем сравнения фактически наблюдаемой абсолютной частоты встреч признака в определенной группе (классе) с теоретически ожидаемой частотой его наблюдения по методу хи-квадрат [2, раздел 3.4]⁹, т.е. если принять, что:

N_{ij} – *фактическое* количество встреч i -го признака у объектов j -го класса;

T_{ij} – *теоретически* ожидаемое количество встреч i -го признака у объектов j -го класса:

$$T_{ij} = \frac{N_iN_j}{N} \quad (5)$$

Подставив выражение (4) в (3) получаем:

$$I_{ij} = \text{Log}_2 \left(\frac{N_{ij}}{T_{ij}} \right) \quad (6)$$

или:

$$I_{ij} = \text{Log}_2(N_{ij}) - \text{Log}_2(T_{ij}) \quad (7)$$

Что касается вопроса о том, вычитание или деление для этого сравнения использовать, то этот вопрос не является принципиальным, т.к. раз-

⁹ <http://lc.kubagro.ru/aidos/aidos02/3.4.htm>

личие между вычитанием и делением сводится к выбору *единиц измерения* результатов сравнения: если взять логарифм от *отношения*, то получится *разность* логарифмов делимого и делителя.

Из выражения (6) для количества знаний следует:

$$\begin{cases} \text{если } N_{ij} < T_{ij} \text{ то } c_{ij}^2 > 0, I_{ij} < 0 \\ \text{если } N_{ij} = T_{ij} \text{ то } c_{ij}^2 = 0, I_{ij} = 0 \\ \text{если } N_{ij} > T_{ij} \text{ то } c_{ij}^2 > 0, I_{ij} > 0 \end{cases}$$

Если фактическая вероятность наблюдения i-го признака при предъявлении объекта j-го класса равна теоретически 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 c^2 по сравнению с теоретической величиной.**

В работе [2, раздел 3.4]¹⁰ предлагается основанная на выражении (6) и вышеприведенной интерпретации мера количества знаний в базе знаний, представляющая собой *количественную меру степени выраженности закономерностей в предметной области:*

$$H = \sqrt[2]{\frac{1}{(M \cdot W - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2} \quad (8)$$

где:

$$\bar{I} = \frac{1}{M \cdot W} \sum_{j=1}^W \sum_{i=1}^M I_{ij} \quad \text{– средняя информативность признаков по матрице информативностей (при увеличении объема выборки стремится к нулю).}$$

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** когнитивного пространства классов и пространства атрибутов.

Интересно отметить, что в американском Internet-ресурсе¹¹

¹⁰ <http://lc.kubagro.ru/aidos/aidos02/3.4.htm>

¹¹ <http://www.heritagehealthprize.com/c/hhp/Details/Rules>



предлагается следующая метрика для оценки достоверности модели (качества алгоритма):

$$\varepsilon = \sqrt{\frac{1}{n} \sum_i^n [\log(p_i + 1) - \log(a_i + 1)]^2} \tag{9}$$

Where:

1. *i* is a member;
2. *n* is the total number of members;
3. *p* is the predicted number of days spent in hospital for member *i* in the test period;
4. *a* is the actual number of days spent in hospital for member *i* in the test period.
5. log is the *natural* logarithm function.

Эта метрика (количественная мера) весьма сходна с выражением (8) из работы [2] и, по существу, содержит под квадратным корнем сумму квадратов выражения (7), если считать, что теоретически ожидаемое значение величины – это и есть ее прогнозируемое значение:

$$e = \sqrt[2]{\frac{1}{n} \sum_{ij}^n [Log_2(N_{ij}) - Log_2(T_{ij})]^2} \tag{10}$$

Если переписать выражение (9) с учетом вышесказанного и выражения (7), то получим:

$$e = \sqrt[2]{\frac{1}{n} \sum_{ij}^n I_{ij}^2} \tag{11}$$

Выражение (11) практически совпадает с выражением (8) из работы автора [2] (изданной в 2002 году), если учесть, что при увеличении объема выборки среднее по *I_{ij}* *стремится к нулю*. Это значит, что информационная мера сходства, используемая в АСК-анализе и системе «Эйдос», тесно связана с энтропийной мерой сходства¹²

Переход от матрицы абсолютных частот к матрице условных и безусловных процентных распределений обеспечивает инвариантность результатов решения вышеперечисленных задач от формы частотного распределения примеров по классам, однако при этом никак не решается вопрос о зависимости этих результатов от *размерностей* различных града-

¹² См.: <http://yandex.ru/yandsearch?text=энтропийная%20мера%20сходства> , <http://www.mce.su/eng/archive/mce14/sect283/authors/person2421/doc12304/> и др.

ций факторов (признаков) и типов шкал, используемых для формализации факторов.

Проблему размерностей можно было бы решить, перейдя к *стандартизированным величинам*¹³ или отношениям условных и безусловных вероятностей. Например, *формулу Байеса*¹⁴ можно рассматривать как дающую количественную оценку степени влияния фактора на наступление некоторого события. Отношение условной вероятности наблюдения некоторого значения фактора в группе (классе) к безусловной вероятности его наблюдения по всей исследуемой выборке также можно рассматривать как количественную меру силы и направления его влияния на переход объекта в состояние, соответствующее классу, т.е. как количественную оценку силы и направления причинно-следственной связи между ними.

Возникает вопрос о том, каким образом формально описать влияние на объект не отдельных значений факторов, а всей их системы. Для того чтобы это сделать введем понятие частных критериев и интегрального критерия.

Частным критерием будем называть выраженное в количественной форме влияние отдельного значения фактора на переход объекта в различные состояния.

Это значит, что отношение условной вероятности наблюдения некоторого значения фактора в группе (классе) к безусловной вероятности его наблюдения по всей исследуемой выборке можно, рассматривать как частный критерий.

Тогда, если значение фактора *способствует* переходу объекта в некоторое состояние, то отношение условной вероятности наблюдения этого значения фактора в группе (классе), соответствующей данному состоянию, будет *больше* безусловной вероятности его наблюдения по всей исследуемой выборке и этот критерий будет иметь значение больше 1.

Если значение фактора *препятствует* переходу объекта в некоторое состояние, то отношение условной вероятности наблюдения этого значения фактора в группе (классе), соответствующей данному состоянию, будет *меньше* безусловной вероятности его наблюдения по всей исследуемой выборке и этот критерий будет иметь значение меньше 1.

Если же значение фактора *никак не влияет на* переход объекта в некоторое состояние, то отношение условной вероятности наблюдения этого значения фактора в группе (классе), соответствующей данному состоянию, будет *равно* безусловной вероятности его наблюдения по всей исследуемой выборке и этот критерий будет иметь значение равное 1.

¹³ При стандартизации вектора каждая его координата заменяется на отношение ее разности со средним по всем координатам к среднеквадратичному отклонению координат от среднего.

¹⁴ <http://ru.wikipedia.org/wiki/Теорема%20Байеса>

Интегральным критерием будем называть некоторое аналитическое выражение от частных критериев, которое количественно отражает силу влияния *системы факторов* на переход объекта в различные состояния.

Моделируемый объект является *линейным*, если результат совместного действия на него совокупности факторов является *суммой* результатов влияния на него каждого из этих факторов в отдельности, т.е. выполняется *принцип суперпозиции*¹⁵ факторов. Чем меньше интенсивность взаимодействия между факторами в объекте, *тем ближе система факторов к множеству* [29] и тем ближе объект к линейному. Таким образом, для линейных объектов можно обоснованно считать, что взаимодействие между факторами в этих объектах отсутствует, т.е. по сути можно считать, что на них действует не система факторов, а множество факторов.

Для *линейных* объектов интегральный критерий, отражающий совместное влияние факторов на объект, можно представить в форме *суммы* влияния каждого из этих факторов в отдельности, т.е. в форме суммы частных критериев, т.е. для линейных объектов оправданно и обоснованно использовать *аддитивный интегральный критерий*.

Приведенные выше количественные меры силы и направления причинно-следственных связей очень неудобны для использования подобных в качестве частных критериев, в основном потому, что в случае *отсутствия* влияния фактора они равны 1. В результате в аддитивном интегральном критерии будет присутствовать некое слагаемое, равное количеству недействующих факторов, и для каждого класса это слагаемое будет свое. В результате подобный интегральный критерий окажется просто непригодным для оценки влияния совокупности факторов на поведение объекта.

Поэтому эти частные критерии необходимо *нормировать* так, чтобы в случае отсутствия влияния он принимали значение равное нулю, а не единице. Есть много вариантов осуществить подобную нормировку, из которых наиболее очевидными являются:

- вычесть 1 из отношения условной вероятности к безусловной;
- взять логарифм от отношения условной вероятности к безусловной.

Первый вариант нормировки приводит к показателям типа ROI¹⁶ (количественная оценка степени полезности инвестиций) и различным его обобщениям.

Второй вариант сразу приводит к семантической мере целесообразности информации А.Харкевича. Из этих вариантов для количественной оценки степени полезности информации для достижения целей по мнению автора предпочтительным является применение меры А.Харкевича [2]. Это

¹⁵ <http://slovari.yandex.ru/~книги/БСЭ/Суперпозиции%20принцип/>

¹⁶ <http://ru.wikipedia.org/wiki/ROI>

связано с тем, что использование логарифма в этой мере позволяет привлечь огромный пласт научных понятий, связанных с данными, информацией и знаниями, что является для нас очень ценным.

Очень важно, что этот подход позволяет автоматически решить проблему сопоставимой обработки многих факторов, измеряемых в различных единицах измерения, т.к. в этом подходе рассматриваются не сами факторы, какой бы природы они не были и какими бы шкалами не формализовались, а количество информации, которое в них содержится о поведении моделируемого объекта (таблица 10):

Таблица 10 – МАТРИЦА ИНФОРМАТИВНОСТЕЙ (БАЗА ЗНАНИЙ) В МИЛЛИБИТАХ¹⁷ (СИМ-2)

Код	Наименование	В группе: «Восток»	В группе: «Запад»	Дискриминантная сила признака
1	ФОРМА ВАГОНА-U-образная	-322	263	414
2	ФОРМА ВАГОНА-V-образная			
3	ФОРМА ВАГОНА-Овальная	1000		707
4	ФОРМА ВАГОНА-Прямоугольная	212	-248	325
5	ФОРМА ВАГОНА-Ромбовидная	1000		707
6	ДЛИНА ВАГОНА-Длинный	-170	152	228
7	ДЛИНА ВАГОНА-Короткий	308	-392	495
8	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-002	107	-115	157
9	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-003	585	-1000	1121
10	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00}	372	-503	618
11	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00}	710	-1459	1534
12	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00}	-1000	585	1121
13	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00}	-1322	678	1414
14	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00}			
15	ВИД СТЕНОК ВАГОНА-Двойные	415	-585	707
16	ВИД СТЕНОК ВАГОНА-Одинарные	206	-241	316
17	ВИД КРЫШИ ВАГОНА-Гофрированная		1000	707
18	ВИД КРЫШИ ВАГОНА-Овальная	1000		707
19	ВИД КРЫШИ ВАГОНА-Отсутствует	74	-78	107
20	ВИД КРЫШИ ВАГОНА-Прямая	515	-807	935
21	ВИД КРЫШИ ВАГОНА-Треугольная	1000		707
22	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 длинный прямоугольник	-585	415	707
23	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 квадрат	1000		707
24	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 овал	-222	193	293
25	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 перевернутый треугольник	1000		707
26	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 прямоугольник	-585	415	707
27	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 ромб	1000		707
28	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 треугольник	515	-807	935
29	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 овала	1000		707
30	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 прямоугольника		1000	707
31	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 квадрата	1000		707
32	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 овала		1000	707
33	ВИД ГРУЗА (КОЛ-ВО И ВИД)-Отсутствует		1000	707
34	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-2		1000	707
35	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-3	585	-1000	1121
36	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-4	415	-585	707
	Средне-квадратичное отклонение	594	608	

Для расчета таблицы 10 используется СИМ-2, в которой N_j представляет собой суммарное количество встреч объектов, относящихся к j -му классу. Численные эксперименты показали незначительное отличие СИМ-

¹⁷ Использование миллибитов оправдано тем, что это позволяет лучше использовать разрядную сетку за счет отсутствия ведущего нуля и десятичной точки.

1 от СИМ-2, но в данной статье в численном примере используется СИМ-2, как более наглядная.

После синтеза модели обязательно осуществляется ее верификация, т.е. измеряется ее достоверность (валидность, адекватность). Обычно это делают, решая различные задачи с помощью созданной модели и оценивая качество их решения. Это могут быть задачи распознавания (идентификации и прогнозирования), поддержки принятия решений и исследования предметной области.

Идентификация – это 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. Различие в математической модели алгоритмах решения задач идентификации и прогнозирования минимальны.

Обычно достоверность модели оценивается путем оценки качества решения задачи идентификации, как наиболее простых. При этом могут использоваться различные подходы к выбору объектов для синтеза модели и для идентификации:

1. Простейшим вариантом является использование всех объектов исследуемой выборки как для синтеза модели, так и для ее верификации. Этот подход дает несколько завышенную оценку достоверности модели.

2. Бутстрепный метод предполагает использование одной части объектов исследуемой выборки для синтеза модели, а другой для ее верификации. Причем способов разделения исследуемой выборки на эти две части также существует много. Объекты, на которых проверяется достоверность модели, исключаются из выборки, на основе которой она создается, чтобы исключить их влияние на создаваемую модель. Успешность идентификации таких объектов означает, что закономерности взаимосвязи между признаками объектов и их принадлежностью к классам, выявленные на исследуемой выборке, действуют и для этих объектов. Необходимо отметить, что *относительное влияние новых объектов на модель уменьшается с увеличением объема выборки, поэтому при больших объемах выборки вполне оправдано использовать первый подход, а бутстрепный метод актуален только на малых выборках.*

3. Наиболее серьезная и убедительная проверка достоверности модели осуществляется когда для синтеза модели используется обучающая выборка, а для оценки достоверности модели *новые* данные, которых вообще не существовало на момент синтеза модели.

В системе «Эйдос» реализованы возможности для всех этих методов верификации модели. Измерение достоверности созданной модели путем оценки качества решения задачи идентификации, т.е. численная оценка эмпирической вероятности ошибок не идентификации и ошибки ложной

идентификации как в целом по всей выборке, так и в разрезе по каждому классу распознавания, показало, что модель имеет 100% достоверность (рисунок 6):

- ИЗМЕРЕНИЕ АДЕКВАТНОСТИ <ДИФФЕРЕНЦИАЛЬНОЙ И ИНТЕГРАЛЬНОЙ ВАЛИДНОСТИ> СЕМАНТИЧЕСКОЙ ИНФОРМАЦИОННОЙ МОДЕЛИ
- Всего физических анкет: 10 <100% для п.15>
 Всего логических анкет: 10
4. Средняя достоверность идентификации логических анкет с учетом сходства : 44.459%
 5. Среднее сходство логических анкет, правильно отнесенных к классу : 21.773%
 6. Среднее сходство логических анкет, ошибочно не отнесенных к классу : 0.000%
 7. Среднее сходство логических анкет, ошибочно отнесенных к классу : 0.000%
 8. Среднее сходство логических анкет, правильно не отнесенных к классу : 22.686%
 9. Средняя достоверность идентификации логических анкет с учетом кол-ва : 100.000%
 10. Среднее количество физич-х анкет, действительно относящихся к классу: 5.000 <100% для п.11 и п.12>
 Среднее количество физич-х анкет, действительно не относящихся к классу: 5.000 <100% для п.13 и п.14>
 Всего физических анкет: 10.000 <100% для п.15>
 11. Среднее количество и % лог-их анкет, правильно отнесенных к классу: 5.000, т.е. 100.000%
 12. Среднее количество и % лог-их анкет, ошибочно не отнесенных к классу: 0.000, т.е. 0.000% <Ошибка 1-го рода>
 13. Среднее количество и % лог-их анкет, ошибочно отнесенных к классу: 0.000, т.е. 0.000% <Ошибка 2-го рода>
 14. Среднее количество и % лог-их анкет, правильно не отнесенных к классу: 5.000, т.е. 100.000%
 15. Средневзвешенная вероятность случайного угадывания принадлежности объекта к классу (< % >): 50.000
 16. Средневзвешенная эффективность применения модели по сравнению со случ. угадыванием (раз): 2.000
 17. Обобщенная достоверность модели <D1+D2>/2: 100.000%. Обобщенная ошибка <E1+E2>/2: 0.000%

23-05-11 08:20:35

г. Краснодар

N п/п	Код класса	Наименование класса	Кол-во лог. анк. дейст-но относящихся к классу	Количество логических анкет правильно или ошибочно отнесенных или не отнесенных к классу				Вероятн. случайного угадывания (%) =NL/NFA	Эффективн. модели по срав. со случ. угадыв. (<раз>)
				Правильн. отнесен.	Ошибочно не отнес.	Ошибочно отнесен.	Правильн. не отнес.		
1	2	3	10	11	12	13	14	15	16
1	1	СОСТАВ СЛЕДУЕТ НА-ВОСТОК	5	5	0	0	5	50.000	2.000
2	2	СОСТАВ СЛЕДУЕТ НА-ЗАПАД	5	5	0	0	5	50.000	2.000
		Ср.взв.значения	5.0	5.0	0.0	0.0	5.0	50.000	2.000

Универсальная когнитивная аналитическая система

НПП «ЭЙДОС»

Рисунок 6. Выходная форма системы «Эйдос» с оценкой достоверности модели

Если модель адекватна, т.е. верно отражает моделируемую предметную область, то исследование этой модели корректно считать исследованием самой моделируемой предметной области. Исследование модели включает кластерно-конструктивный анализ классов и факторов, семантические сети, когнитивные диаграммы, классические и обобщенные (интегральные) когнитивные карты, нелокальные нейроны, когнитивные функции [10, 15] и т.д. Всего система «Эйдос» генерирует более 50 текстовых форм и более 55 видов различных графических форм, лишь нескорые из которых имеют аналоги в MS Excel. Вопросы исследования моделей в АСК-анализе подробно освещены в работах [2, 14], монографиях и статьях автора¹⁸.

Решение задачи идентификации осуществляется в АСК-анализе и системе «Эйдос» на основе леммы Неймана-Пирсона¹⁹ следующим образом [2]. Количество информации в признаке о принадлежности обладающего им объекта к классу (таблица 10), рассматривается как частный критерий. Если известно, что у объекта не один признак, а система признаков, то считается что интегральным критерием, дающим количественную оценку степени принадлежности (или непринадлежности) данного объекта к

¹⁸ Ссылки на эти работы есть на сайте автора: <http://lc.kubagro.ru/>

¹⁹ Из ряда гипотез предпочтительной является та, в пользу которой больше информации.

классам, является *суммарное количество информации* об этом, содержащееся в его системе признаков (1):

$$I_j = \sum_{i=1}^M I_{ij} L_i, \quad (7)$$

где:

$\mathbf{I}_{ij} = \{I_{ij}\}$ – вектор j -го состояния объекта управления (j -го класса), координаты которого, т.е. частные критерии, рассчитываются согласно выражения (4);

$\mathbf{L}_i = \{L_i\}$ – вектор состояния предметной области, включающий все виды факторов, характеризующих объект управления, возможные управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\mathbf{L}_i = \begin{cases} 1, & \text{если } i - \text{е значение фактора встречается } 1 \text{ раз;} \\ n, & \text{если } i - \text{е значение фактора встречается } n \text{ раз;} \\ 0, & \text{если } i - \text{е значение фактора не встречается.} \end{cases}$$

В реализованной модели значения координат вектора состояния предметной области (объекта обучающей или распознаваемой выборки) принимались равными либо равным 1 (фактор действует), либо равным 0 (фактор не действует).

Таким образом, интегральный критерий (1) представляет собой суммарное количество информации, содержащееся в факторах различной природы (т.е. факторах характеризующих объект управления, управляющее воздействие и окружающую среду) о переходе активного объекта управления в целевое состояние [2]. Важно отметить, что предложенный интегральный критерий имеет *неметрическую* природу (он представляет собой скалярное произведение векторов класса и объекта²⁰) и поэтому его применение корректно²¹ в неортонормированном пространстве с неевклидовой метрикой, каким, как правил и является когнитивное пространство, построенное на классификационных шкалах, как осях²². Кроме того, данный интегральный критерий по своей математической форме *совпадает* с аргументом активационной функцией нейрона²³ в нейронных сетях, если интерпретировать весовые коэффициенты на рецепторах как количество информации в соответствии с выражением (4) [16].

²⁰ <http://ru.wikipedia.org/wiki/Скалярное%20произведение>

²¹ В отличие, например от Евклидова расстояния, которое чаще всего применяется на практике даже без проверки пространства на ортонормированность, не говоря уже о проверке метрики на евклидовость, т.е. некорректно. Поэтому если эта мера работает успешно, то только потому, что случайно повезло.

²² Когнитивное пространство может быть построено и на описательных шкалах, как осях, а также на сочетании одной классификационной и нескольких описательных шкал [см. 2, раздел 7.14]:

<http://lc.kubagro.ru/aidos/aidos02/7.14.htm>

²³ <http://ru.wikipedia.org/wiki/Искусственный%20нейрон>

На рисунке 7 представлены некоторые экранные формы результатов идентификации:

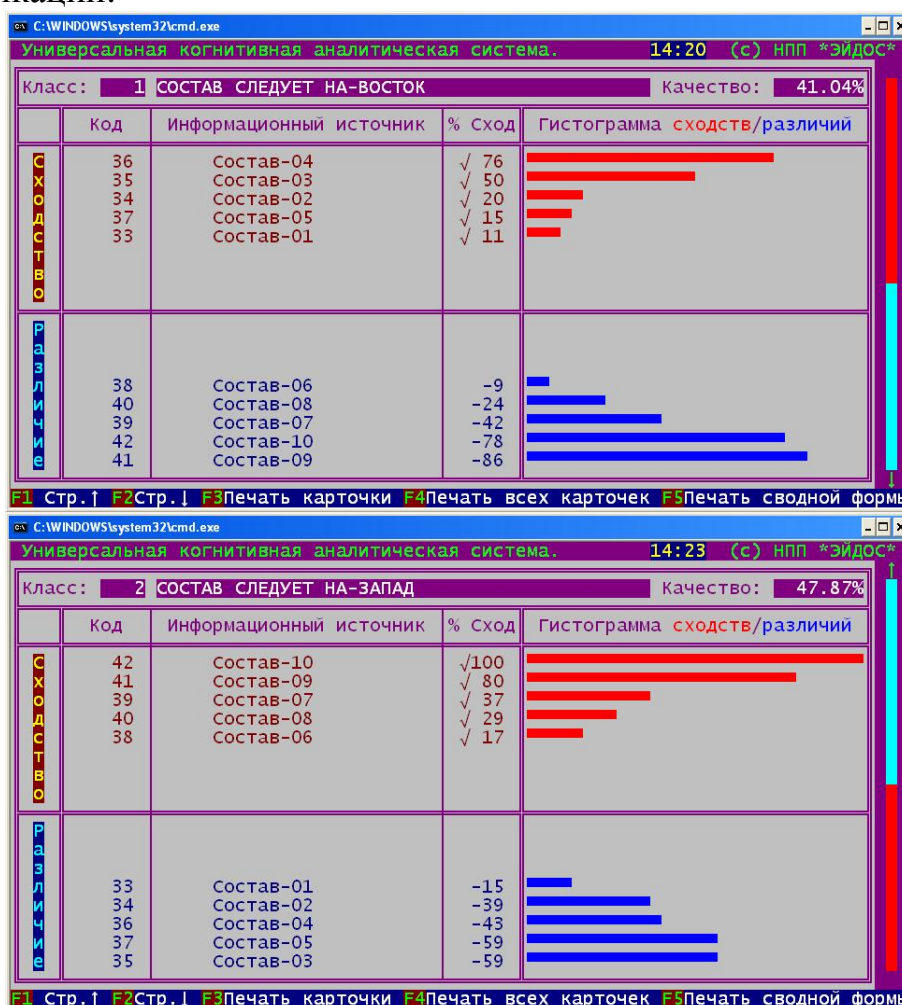


Рисунок 7. Примеры экранных форм результатов идентификации в системе «Эйдос»

Интегральный критерий (7) имеет и другую интерпретацию: как весовой коэффициент при разложении вектора идентифицируемого объекта в ряд по векторам классов [2, раздел 3.5]²⁴.

Задача принятия решений является *обратной* задачей по отношению к задаче идентификации и прогнозирования, т.е. если при прогнозировании по значениям факторов определяется степень принадлежности объектов к классам, то при принятии решений – наоборот, по заданному целевому состоянию (классу) вырабатываются рекомендации по системе факторов, которые обуславливают переход системы в состояние, соответствующее этому классу.

В интеллектуальной системе «Айдос» есть много различных выходных форм, содержащих результаты решения задачи принятия решений: не-локальные нейроны, информационные портреты и другие.

²⁴ <http://lc.kubagro.ru/aidos/aidos02/3.5.htm>

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 влиянием данной системы факторов, а в информационном портрете наоборот, по заданному будущему состоянию объекта управления определяется система 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 от *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.

Информационные портреты классов: «Восток» и «Запад» приведены в таблицах 11 и 12.

Таблица 11 – ИНФОРМАЦИОННЫЙ ПОРТРЕТ КЛАССА: «ВОСТОК»

NUM	KOD	NAME	INFBIT
1	3	ФОРМА ВАГОНА-Овальная	0,10415
2	5	ФОРМА ВАГОНА-Ромбовидная	0,10415
3	18	ВИД КРЫШИ ВАГОНА-Овальная	0,10415
4	21	ВИД КРЫШИ ВАГОНА-Треугольная	0,10415
5	23	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 квадрат	0,10415
6	25	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 перевернутый треугольник	0,10415
7	27	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 ромб	0,10415
8	29	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 овала	0,10415
9	31	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 квадрата	0,10415
10	11	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00}	0,06710
11	9	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-003	0,05103
12	35	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-3	0,05103
13	20	ВИД КРЫШИ ВАГОНА-Прямая	0,04202
14	28	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 треугольник	0,04202
15	15	ВИД СТЕНОК ВАГОНА-Двойные	0,02928
16	36	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-4	0,02928
17	10	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00}	0,02377
18	7	ДЛИНА ВАГОНА-Короткий	0,01560
19	4	ФОРМА ВАГОНА-Прямоугольная	0,00324
20	16	ВИД СТЕНОК ВАГОНА-Одинарные	0,00259
21	8	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-002	-0,01015
22	19	ВИД КРЫШИ ВАГОНА-Отсутствует	-0,01436
23	2	ФОРМА ВАГОНА-V-образная	-0,02383
24	14	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00}	-0,02383
25	6	ДЛИНА ВАГОНА-Длинный	-0,04558
26	24	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 овал	-0,05229
27	1	ФОРМА ВАГОНА-U-образная	-0,06503
28	22	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 длинный прямоугольник	-0,09870
29	26	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 прямоугольник	-0,09870

30	12	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00}	-0,15181
31	13	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00}	-0,19301

Таблица 12 – ИНФОРМАЦИОННЫЙ ПОРТРЕТ КЛАССА: «ЗАПАД»

NUM	KOD	NAME	INFBIT
1	17	ВИД КРЫШИ ВАГОНА-Гофрированная	0,15535
2	30	ВИД ГРУЗА (КОЛ-ВО И ВИД)-2 прямоугольника	0,15535
3	32	ВИД ГРУЗА (КОЛ-ВО И ВИД)-3 овала	0,15535
4	33	ВИД ГРУЗА (КОЛ-ВО И ВИД)-Отсутствует	0,15535
5	34	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-2	0,15535
6	13	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00}	0,11415
7	12	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00}	0,10223
8	22	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 длинный прямоугольник	0,08049
9	26	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 прямоугольник	0,08049
10	1	ФОРМА ВАГОНА-У-образная	0,06103
11	24	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 овал	0,05203
12	6	ДЛИНА ВАГОНА-Длинный	0,04682
13	2	ФОРМА ВАГОНА-V-образная	0,02737
14	14	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00}	0,02737
15	19	ВИД КРЫШИ ВАГОНА-Отсутствует	0,01739
16	8	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-002	0,01259
17	16	ВИД СТЕНОК ВАГОНА-Одинарные	-0,00347
18	4	ФОРМА ВАГОНА-Прямоугольная	-0,00436
19	7	ДЛИНА ВАГОНА-Короткий	-0,02284
20	10	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00}	-0,03694
21	15	ВИД СТЕНОК ВАГОНА-Двойные	-0,04749
22	36	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-4	-0,04749
23	20	ВИД КРЫШИ ВАГОНА-Прямая	-0,07595
24	28	ВИД ГРУЗА (КОЛ-ВО И ВИД)-1 треугольник	-0,07595
25	9	КОЛИЧЕСТВО ОСЕЙ ВАГОНА-003	-0,10061
26	35	КОЛИЧЕСТВО ВАГОНОВ В СОСТАВЕ-3	-0,10061
27	11	ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00}	-0,15941

Из приведенных в таблицах 11 и 12 информационных портретов обобщенных образов классов «Восток» и «Запад» видно, что система искусственного интеллекта выявила на основе приведенных примеров (рисунок 2) наиболее характерные и наиболее нехарактерные признаки этих категорий, которые выявил бы и естественный интеллект. Очень важно отметить, что в данном случае (т.е. в случае применения АСК-анализа и системы «Эйдос») *результаты работы искусственного интеллекта понятны и естественны для естественного интеллекта эксперта.*

С другой стороны искусственный интеллект способен выявлять знания из эмпирических данных *очень большой размерности*, на столько большой, что человеку не хватит и нескольких жизней, чтобы хотя бы только их прочитать, если тратить на чтение весь рабочий день, не говоря уже о выявлении в них каких-то закономерностей или знаний [15].

Кроме того, искусственный интеллект выявляет знания в *количественной* форме, тогда как естественный, только на качественном уровне и в слабо-формализованном виде.

Если мы ставим перед собой цель найти поезд, идущий на запад по его признакам, то эта степень характерности признаков (т.е. количество информации, которое в них содержится о том, что поезд идет на запад или на восток) превращается в знания, позволяющие нам достичь цели, которую мы ставим.

Если проанализировать признаки, содержащиеся в таблицах 11 и 12, сравнив содержащееся в них количество информации о принадлежности или непринадлежности обладающих ими объектов к тем или иным классам, то можно условно разделить их на три основных группы:

- *детерминистские*, содержащие очень большое количество информации;
- *статистические*, содержащие среднее количество информации;
- *практически бесполезные*, содержащие очень малое количество информации.

Числовые шкалы формализуются в системе «Эйдос» в виде интервальных значений, которые нумеруются от минимального значения к максимальному. Для удобства пользователей в наименования градаций числовых шкал, как классификационных, так и описательных, включены условные обозначения номера интервального значения типа: 1/5, что означает: «Первое интервальное значение из пяти».

Необходимо также отметить, что представление о полностью линейных объектах (системах) является абстракцией и реально все объекты являются принципиально нелинейными. Вместе с тем для большинства систем нелинейные эффекты можно считать эффектами второго и более высоких порядков и такие системы *в первом приближении* можно считать линейными. Возможны различные модели *взаимодействия факторов*, в частности, развиваемые в форме системного обобщения теории множеств [29]. Этот подход в перспективе может стать одним из вариантов развития теории нелинейных систем.

Отметим, что математическая модель АСК-анализа (системная теория информации) *органично* учитывает принципиальную нелинейность всех объектов. Это проявляется в нелокальности нейронной сети системы «Эйдос» [30], приводящей к зависимости *всех* информативностей от *любого* изменения в исходных данных, а не как в методе обратного распростра-

нения ошибки²⁵. В результате *значения матрицы информативностей количественно отражают факторы не как множество, а как систему.*

Объект может перейти в некоторое будущее состояние под действием различного количества факторов, но какая бы система факторов не обуславливала (детерминировала) этот переход, в ней не может содержаться информации больше, чем можно получить, точно узнав, что объект переходит в данное состояние. Это количество информации в АСК-анализе называется «Теоретически максимальное количество информации» и определяется только количеством классов (будущих состояний объекта), которые в детерминистском случае равновероятны, т.к. между классами и факторами выполняется взаимнооднозначное соответствие, когда каждое будущее состояние однозначно определяется единственным фактором. Формула А.Харкевича видоизменена в работе [2] таким образом, чтобы удовлетворять принципу соответствия с формулой Р.Хартли в детерминистском случае. Поэтому, чем меньше факторов, тем жестче ими детерминировано поведение объекта, и наоборот, чем больше этих факторов, тем меньше влияние каждого из них на поведение объекта. Например, если переход объекта в некоторое состояние однозначно определяется единственным фактором, то добавление в модель еще одного *точно такого же* фактора приводит к тому, что в сумме эти два фактора будут оказывать тоже самое влияние, которое делится между ними поровну.

Так в математической модели АСК-анализа учитывается *взаимодействие* факторов и отличие *системы* факторов от *множества* факторов [29], являющееся источником нелинейности моделируемого объекта.

Итак, *в матрице информативностей количественно отражены сила и направление влияния каждого значения фактора на переход объекта в каждое из состояний, а также учтено, что совокупность факторов является системой, а не множеством, т.е. учтены взаимодействие факторов и нелинейность моделируемого объекта. Результаты решения задач идентификации, прогнозирования, принятия решений и научного исследования моделируемой предметной области (в частности кластерно-конструктивного анализа), на основе матрицы информативностей **инвариантны** относительно формы частотного распределения объектов исследуемой выборки по классам, единиц измерения значений факторов и типа шкал, используемых для формализации факторов.*

Это позволяет корректно использовать в АСК-анализе аддитивный интегральный критерий в форме *суммы* частных критериев не только для линейных, но и для нелинейных объектов.

²⁵ <http://ru.wikipedia.org/wiki/Метод%20обратного%20распространения%20ошибки>

Различие между матрицей информативностей и матрицей знаний. Если в модели отражены лишь причинно-следственные связи между факторами и будущими состояниями объекта, но не отражена степень желательности ли нежелательности этих будущих состояний, то мы имеем дело с матрицей информативностей. Если же некоторые из будущих событий классифицируются как желательные, т.е. целевые, а другие как нежелательные, то появляется возможность количественной оценки степени полезности информации о действии факторов для перевода объекта в эти состояния, т.е. для преобразования информации в знания.

Процесс преобразования информации в знания – это процесс оценки степени полезности информации для достижения желаемых будущих состояний, т.е. целей. Таким образом, база знаний количественно отражает степень полезности (а также бесполезности и вредности) факторов для достижения целей: она содержит знания в количественной форме о величине и направлении влияния каждого значения фактора на перевод объекта в каждое из будущих состояний, как желаемое, так и нежелательное.

Соотношение различных моделей представления знаний, в т.ч. процедурной и декларативной, таково, что можно обоснованно говорить о том, что одни и те же знания могут быть более или менее полно и адекватно представлены с помощью большинства моделей и различие между ними не столь велико и принципиально, как обычно принято думать.

Это различие можно сравнить с различием между различными языками, на которых фразы, имеющие один и тот же смысл, звучат или выглядят (в текстовой форме) совершенно по-разному.

Не представляет исключения и модель представления знаний, принятая в системно-когнитивном анализе и системе «Эйдос», которая имеет много общего со многими моделями представления знаний. Например, база знаний (матрица информативностей) системы «Эйдос» очень напоминает матрицу весовых коэффициентов нейронных сетей, но в отличие от нее имеет четкую научно обоснованную интерпретацию коэффициентов и рассчитывается прямым счетом, а не путем итерационного подбора методом обратного распространения ошибки. Коэффициенты матрицы информативностей представляют собой количество информации в признаке о принадлежности обладающего им объекта к классу. О смысле же весовых коэффициентов нейронной сети идут научные дискуссии. Кроме того база знаний системы «Эйдос» очень напоминает таблицу принятия решений и на основе нее действительно принимаются решения о принадлежности объектов к классам (задача идентификации и прогнозирования) или о том,

какие значения факторов необходимы для перевода объекта в заданное целевое состояние (задача принятия решений) (таблица 11)

Таблица 13 – СОПОСТАВЛЕНИЕ МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ АСК-АНАЛИЗА И СИСТЕМЫ «ЭЙДОС» С ДРУГИМИ МОДЕЛЯМИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ

Модель системно-когнитивного анализа и интеллектуальной системы «Эйдос»	Классификационные шкалы и градации	Описательные шкалы и градации	Конкретный образ объекта исследуемой выборки	База знаний (декларативное и процедурное представление знаний), прямые и обратные правдоподобные рассуждения	Обобщенный образ класса	Кластеры, могут отображаться в форме дерева и семантической сети	Конструкт как система наиболее непохожих классов с спектром промежуточных по уровню сходства классов
Логическая модель (детерминистская Аристотелевская логика)	Бинарные (дихотомические) справочники классов	Бинарные (дихотомические) справочники признаков	Бинарный вектор объекта	Правила логического вывода	---	---	---
Логическая модель (нечеткая логика Л.Заде)	Номинальные, порядковые и числовые справочники классов	Номинальные, порядковые и числовые справочники признаков	Вектор объекта с указанием степени выраженности у него признаков	Нечеткие правила логического вывода	---	---	---
Фреймовая модель	Имена фреймов	Слоты и шпации	Фреймы-экземпляры	Процедуры формирования фреймов прототипов на основе фреймов-экземпляров	Фреймы-образцы, или прототипы	---	---
Процедурная модель (экспертные системы)	---	---	---	Продукционное представление правил вывода	---	---	---
Семантические сети	---	Свойства и их значения	Элемент класса	Отношения между классами	Класс	Граф результатов кластерного анализа	---
Нейронные сети	Множество нейронов	Множество рецепторов	Объект обучающей выборки	Матрица весовых коэффициентов	Нейрон с весовыми коэффициентами	Нейрон 2-го слоя сети	---

Факт – это единство экстенционального и интенционального описания *события*, обнаруженного эмпирическим путем, т.е. по сути, *факт это определение события*. Пример факта: «Кошка кормит котят молоком». Пример определения в науке: «Млекопитающее – это животное (более общее, интенциональное понятие), вскармливающее своих детей молоком (экстенциональный специфический признак)».

Закономерности – это причинно-следственные зависимости, выявленные на исследуемой выборке и распространяемые лишь на саму эту выборку.

Эмпирический закон – это закономерности, выявленные на исследуемой выборке и распространяемые на некоторую более широкую предметную область, в которой действуют *те же причины их существования*, что и в исследуемой выборке. Эта более широкая предметная область называется *генеральной совокупностью*, по отношению к которой исследуемая выборка *репрезентативна*.

Очень важно понимать, что генеральная совокупность является более широкой, чем исследуемая выборка, не только в пространстве, но и во времени. *Периоды* времени, в течение которых закономерности в предметной области существенно не меняются, называются *периодами эргодичности*. Границы между периодами эргодичности называются точками бифуркации. Используя эту терминологию можно сказать, что *генеральная совокупность эргодична по отношению к исследуемой выборке, а граница генеральной совокупности состоит из точек бифуркации*.

Большой интерес представляет *анализ* глубоких аналогий и параллелей между рассмотренными выше понятиями автоматизированного системно-когнитивного анализа и некоторыми базовыми понятиями физики²⁶. Ниже предпринимается попытка сформулировать эти понятия физики в терминах АСК-анализа. Это сразу позволяет увидеть неожиданные аспекты понимания и варианты развития этих понятий.

Принцип относительности Галилея и Эйнштейна формулируется следующим образом: все физические процессы в *инерциальных системах отсчёта*²⁷ протекают одинаково, независимо от того, неподвижна ли система или она находится в состоянии равномерного и прямолинейного движения²⁸.

В терминах АСК-анализа принцип относительности можно сформулировать следующим образом: закономерности, выявленные в исследуемой выборке в одной области *пространства и времени*, выполняются и в других областях, в которых действуют *те же причины их существования*,

²⁶ Прежде всего, имеются в виду такие понятия, как: пространство и время, законы сохранения, инерциальная система отсчета, законы природы.

²⁷ <http://ru.wikipedia.org/wiki/Инерциальная%20система%20отсчёта>

²⁸ <http://ru.wikipedia.org/wiki/Принцип%20относительности>

что и в исследуемой выборке. В частности, эти причины могут состоять в том, что виртуальная реальность сконструирована таким образом, чтобы в ней выполнялся принцип относительности [17]²⁹.

В фундаментальной теореме Нетер³⁰ доказывається, что **причинами существования** законов сохранения являются однородность и изотропность пространства и времени (их симметрии) (таблица 12):

Таблица 14 – СИММЕТРИИ ПРОСТРАНСТВА И ВРЕМЕНИ И ОБУСЛОВЛЕННЫЕ ИМИ ЗАКОНЫ СОХРАНЕНИЯ³¹

Преобразования	Инвариантность	Закон сохранения
Трансляции времени	Консервативность	...энергии
Изотропия времени	Изотропия времени	...энтропии
Трансляции пространства	Однородность	...импульса
Вращения	Изотропность пространства	...момента импульса
Группа Лоренца	Лоренц-инвариантность	...интервала

Так как «Инерциальной называется система отсчёта, по отношению к которой пространство является однородным и изотропным, а время – однородным»³², то можно сделать вывод: *законы сохранения и инерциальные системы отсчета имеют общие причины существования* – это симметрии пространства и времени.

Учитывая этот вывод, переформулируем принцип относительности в АСК-анализе так: закономерности, выявленные в исследуемой выборке в одной области *пространства и времени*, тем точнее выполняются в других областях, чем меньше отличаются свойства *пространства и времени* (метрические и топологические) в этих других областях от исследуемой.

Эргодичными по отношению к исследуемой выборке будем называть те области пространства и времени, в которых действуют те же закономерности, что в исследуемой выборке.

Неэргодичные области пространства и времени существенно отличаются по своим свойствам друг от друга и разделены границами, состоящими из точек, которые мы будем называть *точками бифуркации*.

Используя эти термины, принцип относительности в АСК-анализе можно сформулировать следующим образом: *эмпирические закономерности, выявленные на исследуемой выборке в определенной области пространства и времени будут выполняться и в других областях, эргодичных по отношению к данной области и будут нарушаться при переходе в неэргодичные по отношению к ней области.*

²⁹ С другой стороны в любой реализации виртуальной реальности практически неизбежны некоторые погрешности, вследствие которых принцип относительности может иногда нарушаться, может быть очень незначительно. Это означает, что нарушение принципа относительности может указывать на виртуальную природу нашей реальности. Есть и другие соображения в пользу этой гипотезы: например квантовые явления, в частности квантовое движение, парадокс Эйнштейна-Подольского-Розена и др.

³⁰ <http://ru.wikipedia.org/wiki/Теорема%20Нетер>

³¹ Источник информации: <http://ru.wikipedia.org/wiki/Принцип%20относительности>

³² <http://ru.wikipedia.org/wiki/Инерциальная%20система%20отсчёта>

Таким образом, мы можем сформулировать научную *гипотезу* «О существовании границ применимости принципа относительности и законов сохранения»: принцип относительности и законы сохранения *выполняются* для эргодичных по отношению к исследуемой выборке областей пространства и времени и *нарушаются* при переходе через границу бифуркации в неэргодичную по отношению к ней область.

Естественно, возникает вопрос о возможности экспериментальной проверки этой гипотезы. Для этого необходимо экспериментально выйти к границам бифуркации исследуемой области пространства и времени и за их границы в неэргодичную область. Из одной области пространства и времени можно перейти в другую область путем изменения исследуемых периодов времени, смещения и поворота в пространстве. Сочетание смещений и поворотов позволяет изменить пространственно-временной масштаб исследуемой предметной области.

Если эмпирические закономерности в исследуемой выборке выявлены за определенный период *времени*, то они считаются действующими и в другие периоды, в течение которых действуют *те же причины существования* этих закономерностей, что и в исследуемой выборке, т.е. в периоды времени, эргодичные по отношению к исследуемой выборке. Это значит, что вообще говоря на протяжении очень малых или очень больших периодов времени по сравнению с исследуемым периодом эти закономерности могут и нарушаться.

Ясно, что неограниченное перемещение в определенном направлении или изменение пространственного масштаба исследуемой предметной области, в конце концов, приведет к выходу за границы области эргодичности и это потребует пересмотра законов, открытых в исследуемой предметной области. Например, при уменьшении масштаба до размеров атомов и элементарных частиц мы переходим в область квантовых явлений, а при увеличении масштаба до размеров галактики и более мы сталкиваемся с явлениями, для объяснения которых необходимо либо пересмотреть законы гравитации, либо ввести не наблюдаемые в настоящее время сущности, типа темной материи или энергии.

Но является ли *фактически* наше физическое пространство времени однородным и изотропным, и если да или нет, то на сколько именно, т.е. в какой степени?

Не смотря на то, что на принципе относительности во многом основана современная наука, он далеко не так очевиден.

Существует *легенда*, согласно которой Галилео Галилей³³ исследовал закономерности свободного падения, сбрасывая шары из разных материалов одинакового размера, но разной массы с Пизанской башни и измеряя время падения по собственному пульсу. Однако найденные Галилеем за-

³³ <http://ru.wikipedia.org/wiki/Галилей,%20Галилео>

кономерности свободного падения действовали за миллиарды лет до него и будут действовать через миллиарды лет после него и не только в Италии, но и в других странах и даже не только на Земле, и не только в нашей галактике.

Само время изменяет абсолютно все, и даже, наверное, само себя, поэтому, скорее всего, существуют различные неэргодичные по отношению друг к другу периоды времени. Нет ничего постоянного в этом мире, кроме разве что самого этого непостоянства. Возможно, с течением времени меняется и гравитационная постоянная³⁴, и другие фундаментальные константы. Конечно, это происходит очень и очень медленно. Тем ни менее есть надежда, что современные сверхточные информационно-измерительные технологии позволят уловить это изменение. Скорее всего, есть области Вселенной, в которых значения этих фундаментальных констант иные, чем в нашей области, да и в любой из этих областей они могут незначительно изменяться в разных областях.

Если, в соответствии с моделью, принятой в общей теории относительности (ОТО) А.Эйнштейна, рассматривать гравитацию как искривление пространства-времени, то ясно, что структура (метрика, а возможно и топология) пространства-времени зависит от распределения масс, а оно далеко не равномерно, значит и пространство-время не однородно и не изотропно на столько же, как и распределение масс. Следовательно, базовое предположение, на котором основана теорема Нётер об однородности и изотропности пространства-времени, является абстракцией и реально выполняется лишь с некоторой точностью или погрешностью, как, по-видимому, и вытекающие из этих свойств пространства-времени законы сохранения импульса, энергии и момента количества движения.

Профессор Майкл Лонго³⁵ с коллегами обобщив, знания о распределении спиральных галактик в метagalактике, совершили неожиданное открытие: в северной небесной полусфере Земли более распространены левозакрученные спирали, а в южной – правозакрученные (хотя для южной части это и не столь выражено). Результаты этого исследования говорит о том, структура пространства метagalактики, при детальном ее рассмотрении, является асимметричной, а не строго однородной и изотропной. Ось вращения Земли сориентирована в пространстве таким образом, что это различие наибольшее для ее северного и южного полушарий. Из этого факта следуют далеко идущие выводы. Например, известно, что сам феномен существования белковой жизни на Земле связан с ярко выраженной асимметрией (киральностью) белковых молекул.

³⁴ См., например: Миллюков В.К. Изменяется ли гравитационная постоянная? - Природа, 1986, №6, с. 96-104.

³⁵ <http://www.technologyreview.com/blog/arxiv/23410/>

Эмпирический закон является феноменологическим, т.е. внешним описанием зависимости последствий от причин, который не раскрывает механизма или способа, которым реализуется эта зависимость.

Научный закон – это содержательная интерпретация механизма действия эмпирического закона, т.е. *способа* преобразования причин в следствия. Научный закон является содержательным *объяснением* и интерпретацией эмпирического закона. Это объяснение, когда оно разрабатывается, не сразу становится научным законом, а сначала имеет статус научной гипотезы и приобретает статус научного закона лишь после того, как *на практике*, т.е. эмпирически, подтверждаются предсказания существования новых, ранее неизвестных явлений, сделанные на основе научной гипотезы. Таким образом, научный закон – это научная гипотеза, адекватность и прогностическая сила которой подтверждены (верифицированы) эмпирически. Процесс преобразования научной гипотезы в научный закон – это процесс подтверждения на практике адекватности этой научной гипотезы.

Необходимо подчеркнуть, что существует принципиальная возможность создания *неограниченного*³⁶ количества различных моделей, одинаково хорошо (адекватно, верно, достоверно) отражающих одну и ту же предметную область. Это является еще одной причиной некорректности гипостазирования лишь одной из них, т.к. реальность *одновременно* не может быть такой, какой она отражается во всех этих различных моделях (или может, о чем писал Эверетт³⁷). Иногда такие модели и действительно созданы, тогда это проявляется особенно наглядно. Естественно возникает вопрос о *критериях выбора* одной модели, в определенном смысле или по определенным обоснованным критериям «наилучшей» из многих. Среди этих критериев следует отметить адекватность, удовлетворение принципу соответствия и широту адекватно отражаемой предметной области, а также ее простоту и красоту. *Из многих моделей предпочтительная та, которая более адекватна (точнее), и та, которая адекватно отражает более широкую предметную область, т.е. включает в себя на основе принципа соответствия другие известные модели*, а также более простая и красивая модель. Однако часто бывает, что разработка многих моделей (научных теорий) весьма затруднительна (и может занимать не одну сотню лет) и есть или известна всего лишь одна-единственная модель. В этом случае эта единственная модель автоматически начинает считаться наилучшей (из всех известных).

Поэтому ученые, разрабатывающие новые перспективные модели, стремятся к созданию более общих моделей, т.е. моделей, имеющих более широкую область адекватности, что обеспечивает этим моделям конку-

³⁶ По крайней мере, в настоящее время неизвестно каких-либо ограничений на количество различных моделей одной и той предметной области. Так, например, число 4 можно получить в результате теоретически *неограниченного* количества различных арифметических операций: $2*2$, $2+2$, $8/2$, и т.д. и т.д.

³⁷ <http://ru.wikipedia.org/wiki/Многомировая%20интерпретация>

рентные преимущества перед другими моделями. Обычно более общие модели являются и более точными, чем частные.

Расширение области адекватности модели осуществляется путем включения в исходные данные новых (обычно ранее неизвестных, но иногда и давно всем известных) фактов, не принадлежащих генеральной совокупности, по отношению к которой исследуемая выборка репрезентативна, с последующим пересинтезом модели, формулированием новых эмпирических и научных законов.

Это значит, что ученые, действительно стремящиеся к поиску истины и развитию науки, должны буквально охотиться за такими новыми фактами, не вписывающимися в традиционные общепринятые научные теории, концепции и парадигмы реальности. Но в жизни мы часто видим обратное, когда отрицается само существование подобных фактов на том основании, что «этого не может быть, т.к. не может быть никогда» или потому, что они «не вписываются» в частные неоправданно гипостазированные модели. Этот феномен объясняется с позиций теории научных революций Томаса Куна [12]³⁸.

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 8 обозначены цифрами:

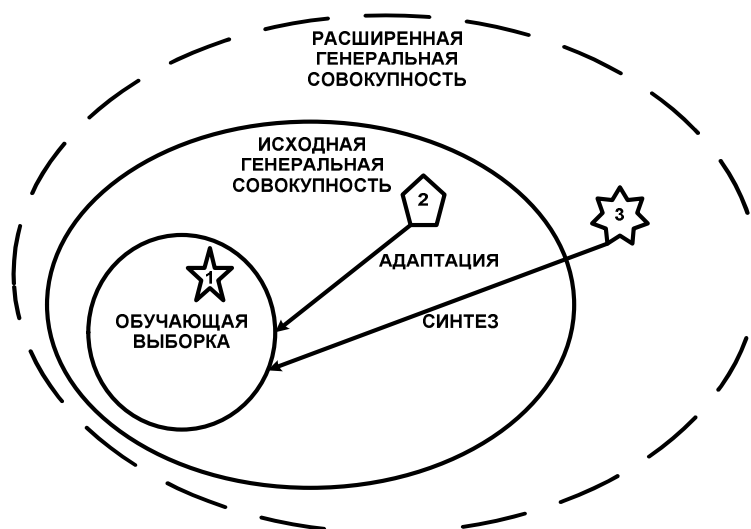


Рисунок 8. К пояснению смысла понятий:

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

³⁸ т.е. они просто пожинают плоды социализации прошлых этапов развития науки, см. <http://www.humanities.edu.ru/db/msg/81548>

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).

Высшая форма использования информации – это знания, т.е. ее использование для достижения целей (управления). На рисунке 9 приведена схема автоматизированной системы управления (АСУ), в которой процесс выявления знаний из предметной области (познание) входит непосредственно в цикл управления:

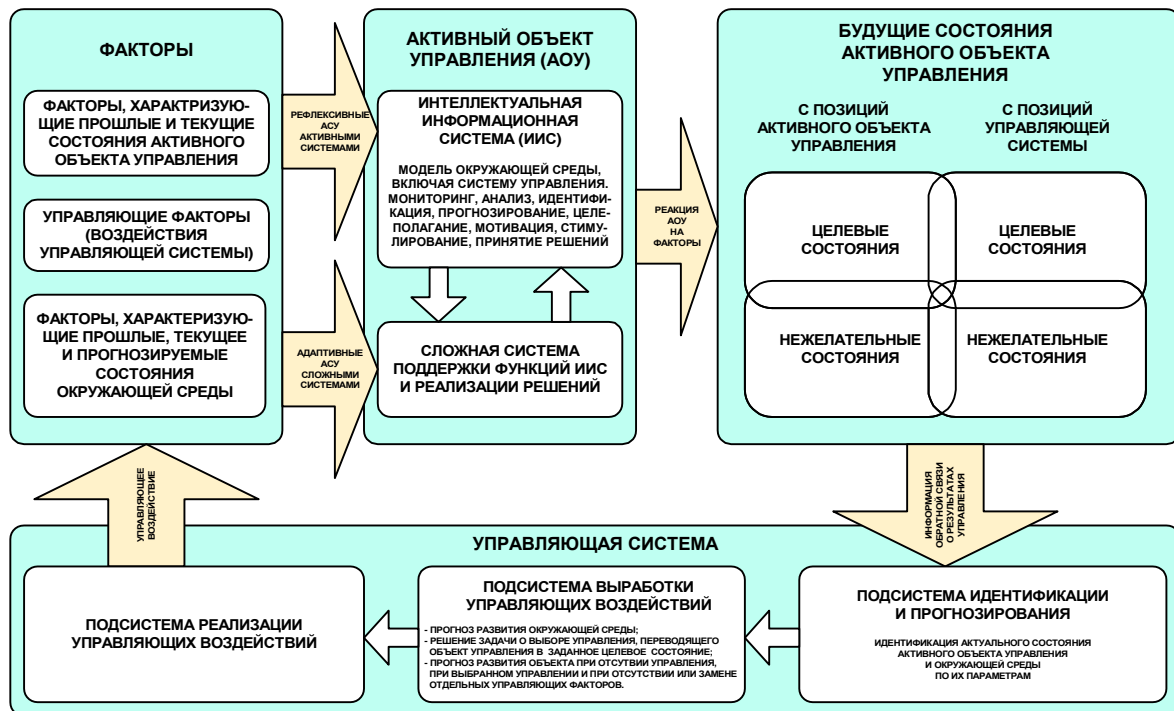


Рисунок 9. Структура предметной области и рефлексивной АСУ активными объектами в АСК-анализе, в которой процесс выявления знаний из предметной области (познание) входит непосредственно в цикл управления [2]

Подробнее цикл выявления знаний из эмпирических данных и их применения для прогнозирования и управления (принятия решений) в АСК-анализе показан на рисунке 10.

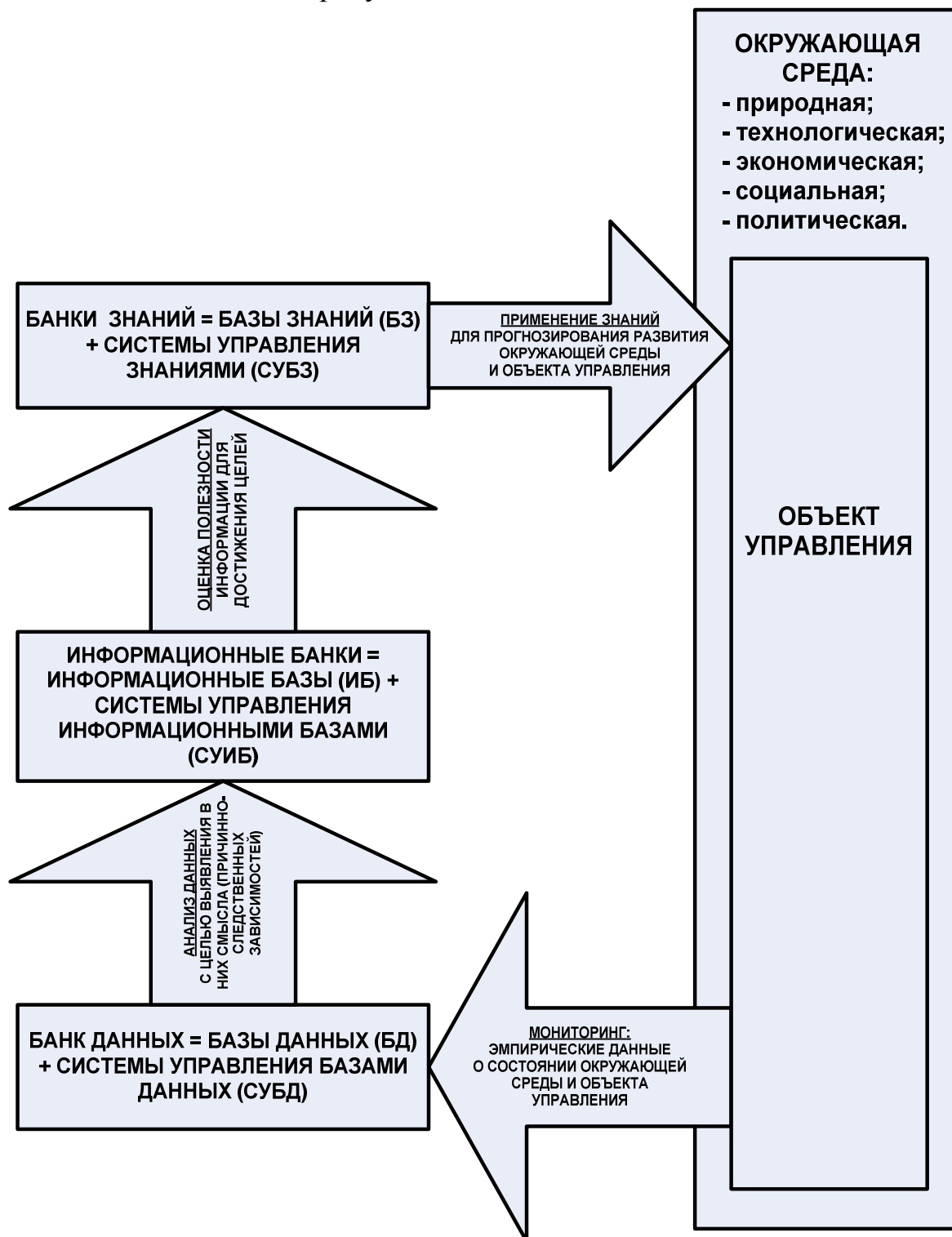


Рисунок 10. Цикл выявления знаний из эмпирических данных и их применения для прогнозирования и управления (принятия решений) в АСК-анализе

Если модель имеет высокую степень адекватности и, особенно, если нет альтернативных достоверных моделей, то возникает соблазн неоправданно и необоснованно считать, *что реальность устроена именно таким образом, какой она отражается в этой наилучшей по сформулированным выше критериям модели* или научной теории, т.е. *возникает соблазн необоснованно придать онтологический статус абстрактной модели*. В этом состоит широко распространенная малозаметная (когда ее совершаешь) ошибка познания, называемая «Гипостазирование³⁹». Однако эта ошибка влечет за собой целый шлейф весьма заметных последствий, важнейшим из которых является отрицание существования фактов, закономерностей и эмпирических законов, не вписывающихся в те или иные научные теории, даже если эти факты в буквальном смысле слова *очевидны*.

Например, апологеты воздухоплавания отрицали возможность создания летательных аппаратов тяжелее воздуха, не смотря на птиц, которые садились и взлетали перед ними (или даже смотря на них, но не осознавая, *что* они видят перед собой)⁴⁰. При этом они исходили из того, что принцип действия летательных аппаратов может быть основан только на законе Архимеда, как это следовало из единственной известной им научной теории полета. Если бы эти апологеты воздухоплавания отрицали возможность летательных аппаратов тяжелее воздуха, принцип действия которых основан на законе Архимеда, то к ним бы не было никаких претензий. Иначе говоря, эти апологеты упустили из внимания, что могут существовать (а мы уже знаем, что и действительно существуют) и другие принципы полета: в частности, баллистический, аэродинамический, ракетный, электромагнитный, на которых может быть основан принцип действия летательных аппаратов тяжелее воздуха, причем *эти аппараты ни в коей мере не нарушают закон Архимеда (как почему-то думали апологеты и поэтому отрицали возможность их создания) и полностью ему подчиняются*.

Признание существования факта не зависит от обнаружения закономерности. Признание существования закономерности не зависит от обнаружения соответствующего эмпирического закона. Признание существования эмпирических законов не зависит от наличия верифицированной содержательной интерпретации или научного закона, а если она есть, то от того, является ли она «правильной» или «неправильной» по тем или иным критериям или по чьему-то мнению. Таким образом, признание существования факта не зависит от наличия их интерпретации или теории, которая

³⁹ <http://yandex.ru/yandsearch?text=гипостазирование>

⁴⁰ Совершенно аналогично сейчас спорят о возможности создания искусственного интеллекта, совершенно не замечая того, что люди обладают интеллектом и, безусловно, созданы по определенной технологии. А между тем хорошо известно, что если что-то создано по одной технологии, то оно может быть создано и по другим технологиям.

его объясняет, и *отсутствие такой теории не является основанием для отрицания существования или непризнания существования факта.*

Конечно, часто проще и выгоднее не разрабатывать новые теоретические объяснения «неудобных» фактов, а просто отрицать само их существование по принципу: «Есть факт – есть проблема, нет факта – нет проблемы». Подобное отрицание чаще всего наблюдается на этапе социализации науки⁴¹.

От обнаружения фактов до создания адекватной теории их объясняющей, могут пройти сотни и даже тысячи лет, более того, со временем меняются и сами критерии достоверности теории или того, что вообще можно признать «объяснением». Поэтому в разное время различные и даже противоречащие друг другу теории считаются правильными объяснениями одних и тех же фактов.

Выводы. Целесообразность использования именно матрицы информативностей (базы знаний) для визуализации когнитивных функций и решения других задач (идентификации, прогнозирования, поддержки принятия решений и исследования предметной области) состоит в следующем. На основе сравнения абсолютных частот можно делать какие-либо выводы только в том случае, когда по разным классам приведено одинаково количество примеров. В общем случае это количество примеров по разным классам *всегда разное*, поэтому матрица абсолютных частот сама по себе непригодна для решения вышеперечисленных задач. Но на практике иногда встречается ситуация, когда это количество мало отличается по разным классам и тогда использование этой матрицы не дает большой ошибки и позволяет составить более-менее достоверное представление о предметной области. Получить модель предметной области, *инвариантную* относительно различий в количестве примеров по классам, можно перейдя от матрицы абсолютных частот к матрице условных и безусловных процентных распределений⁴². Использование матрицы условных и безусловных процентных распределений позволяет получить такой же уровень достоверности выводов о предметной области, какой матрица абсолютных частот *в случае* равного количества примеров по разным классам. Однако для того, чтобы получить эти выводы необходимо *сравнивать* условные процентные распределения друг с другом. При небольших размерностях моделей это сравнение может быть проведено вручную, правда лишь на качественном уровне, однако при реальных встречающихся на практике размерностях вручную это сделать не представляется возможным. *Поэтому в АСК-анализе принято решение автоматизировать количественное срав-*

⁴¹ <http://www.humanities.edu.ru/db/msg/81548>

⁴² При увеличении объема выборки относительные частоты (выраженные в процентах) стремятся к вероятностям, как пределам. При достаточно больших объемах выборки различие относительных частот от вероятностей невелико и на практике приемлемо вместо терминов «относительные и абсолютные частоты» использовать термины «условные и безусловные вероятности». Под процентными распределениями понимаются именно частоты, а не вероятности.

нение условных процентных распределений. При этом возник вопрос о том, как именно их сравнивать: друг с другом или с какой-то базой сравнения. Если в модели всего два класса, то приемлем вариант сравнения условных процентных распределений по ним друг с другом. Но когда классов всего три, то уже не очень понятно как это делать, если же их сотни, тысячи или десятки тысяч, то это становится вообще непонятным. В АСК-анализе этот вопрос решен путем сравнения условных процентных распределений по классам с безусловным процентным распределением по всей выборке, которое и выступает базой (нормой) сравнения. Кроме того, важно отметить, что получающееся в результате этого выражение для количества информации **тождественно** выражению, получающемуся путем сравнения фактически наблюдаемой абсолютной частоты встреч признака в определенной группе к теоретически ожидаемой частоте его наблюдения в методе хи-квадрат. Таким образом, *матрица информативностей представляет собой **результат** выполнения процедуры⁴³ автоматизированного сравнения условных процентных распределений признаков по классам с их безусловным процентным распределением по всей выборке.* По сути это результат **нормировки** условных процентных распределений с использованием в качестве нормы безусловного процентного распределения. Поэтому использование матрицы информативностей освобождает исследователя-аналитика от необходимости выполнения огромной рутинной работы (которую он как правило и не может выполнить вручную, а когда может, то лишь на качественном, а не количественном уровне) по сравнению условных процентных распределений друг с другом или с какой-либо базой сравнения и обеспечивает более высокий уровень достоверности выводов, чем использование матрицы условных и безусловных процентных распределений.

Таким образом, *когнитивные функции являются адекватным математическим инструментом для формального представления причинно-следственных зависимостей.* Когнитивные функции представляют собой многозначные интервальные функции многих аргументов, в которых различные значения функции в различной степени соответствуют различным значениям аргументов, причем количественной мерой этого соответствия выступает знания, т.е. информация о причинно-следственных зависимостях в эмпирических данных, полезная для достижения целей. Многочисленные исследования [9, 10] подтверждают, что *метод и программный инструментарий визуализации когнитивных функций* позволяют наглядно увидеть такие причинно-следственные закономерности предметной области, которые другими методами выявить и описать весьма проблематично.

В качестве **перспективных** направлений исследований и разработок в области АСК-анализа отметим следующие. Матрица информативностей

⁴³ Поэтому *матрица информативностей относится к декларативной форме представления знаний, но сама она получается в результате применения процедурной модели*

представляет собой *исчерпывающий* перечень (т.е. в *декларативной* форме представления знаний) результатов идентификации объектов, описанных *одним* признаком (градацией описательной шкалы) с классами (градациями классификационных шкал). С этой точки зрения она состоит из *двумерных* подматриц, образуемых всеми сочетаниями описательных и классификационных шкал по две (одна описательная и одна классификационная шкала). Но матрица информативностей содержит всю необходимую и достаточную информацию *для расчета* (т.е. в *процедурной* форме представления знаний) *всех возможных результатов идентификации объектов с классами* (градациями классификационных шкал), описанных не одним, всеми сочетаниями по два, три, ... N признаков из M , где M – количество описательных шкал (если градации в шкалах *альтернативные*). Если реально выполнить эти процедуры распознавания для всех возможных объектов⁴⁴, то *результаты идентификации можно наглядно представить в декларативной форме в виде значений элементов многомерных матриц (когнитивных тензоров или тензоров знаний⁴⁵, описывающих когнитивное пространство⁴⁶), образованных одной классификационной шкалой и несколькими описательными*. Координатные плоскости этих многомерных матриц являются подматрицами матрицы информативностей с общей классификационной шкалой и разными описательными шкалами. Эти подматрицы визуализируются в базовой системе «Эйдос» (режим _54) и в системе «Эйдос-астра» в форме когнитивных функций [9, 10]. В АСК-анализе в качестве частных критериев принято рассматривать признаки (градации описательных шкал), точнее не сами признаки, а количество информации, содержащейся в признаках о принадлежности обладающих ими объектов к классам, а в качестве интегрального критерия – *сумму информативностей частных критериев*, т.е. свертку или скалярное произведение в координатной форме [2]. Многомерные когнитивные структуры в этих подматрицах отражают результаты идентификации объектов со всеми возможными сочетаниями признаков⁴⁷ из n по m , которые, следуя работе [11], будем называть Эйдосами, а их визуализацию – *многомерными когнитивными функциями*. Это название оправдано тем, что сечения Эйдосов координатными плоскостями являются уже известными когнитивными функциями [9, 10].

⁴⁴ В текущей версии базовой системы «Эйдос» в целях экономии времени расчетов это делается не для всех возможных объектов, а только для тех, которые представлены в распознаваемой выборке.

⁴⁵ См.: <http://ru.wikipedia.org/wiki/Тензор>. Компонентами когнитивного тензора является количество знаний в градации описательной шкалы о градации классификационной шкалы, определяемое согласно выражения {4} на стр.19 данной статьи. Это открывает путь описания когнитивного пространства с использованием математического аппарата и понятий общей теории относительности (ОТО) А.Эйнштейна.

⁴⁶ С этой точки зрения физическое пространство-время может рассматриваться как *подпространство* (частный случай или проекция) когнитивного пространства [13].

⁴⁷ В этой модели будем считать, что градации каждой описательной шкалы являются *альтернативными* признаками, т.е. не могут встречаться одновременно.

В многомерном когнитивном пространстве можно ввести понятия, аналогичные понятиям кинематики и динамики. Если в качестве описательных шкал в семантической информационной модели [2] использовать пространственно-временные координаты, то АСК-анализ можно будет рассматривать как инструмент для *информационного описания физических процессов и явлений*, в котором роль пространства-времени будет играть когнитивное пространство-время. Между физическим и когнитивным пространством-временем существует тесная взаимосвязь и вообще между ними больше общего, чем различий [13]. Предлагается также добавить к пространственно-временным координатам еще две координаты, количественно отражающие «уровень системности» и «степень детерминированности» системы. В результате получится наука, которую можно было бы назвать «*Информационная динамика систем*».

Материалы данной статьи могут быть использованы при разработке интеллектуальных систем, а также при проведении лабораторных работ по дисциплинам: «Интеллектуальные информационные системы» для специальности: 080801.65 – Прикладная информатика (по областям) и «Представление знаний в информационных системах» для специальности: 230201.65 – Информационные системы и технологии.

Библиографический список

1. Мичи Д., Джонстон Р. Компьютер – творец. – М.: Мир, 1987. – 251 с.
2. Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп. – Краснодар: КубГАУ, 2006. – 318с.
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. – 605с⁴⁸.
4. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее время / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(54). – Шифр Информрегистра: 0420900012\0110. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>
5. Луценко Е.В. Универсальная когнитивная аналитическая система "ЭЙДОС". Пат. № 2003610986 РФ. Заяв. № 2003610510 РФ. Оpubл. от 22.04.2003.
6. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(5). – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/04.pdf>
7. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(1). – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>
8. Луценко Е.В. Типовая методика и инструментарий когнитивной структуризации и формализации задач в СК-анализе / Е.В. Луценко // Политематический сетевой электронный

⁴⁸ Для удобства читателей ряд работ из списка приведен на сайте автора: <http://lc.kubagro.ru/>

- научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(3). – Режим доступа: <http://ej.kubagro.ru/2004/01/pdf/16.pdf>
9. Луценко Е.В. Когнитивные функции как адекватный инструмент для формального представления причинно-следственных зависимостей / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №09(63). С. 1 – 23. – Шифр Информрегистра: 0421000012\0233. – Режим доступа: <http://ej.kubagro.ru/2010/09/pdf/01.pdf>, 1,438 у.п.л.
 10. Луценко Е.В. Метод визуализации когнитивных функций – новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №03(67). С. 240 – 282. – Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>, 2,688 у.п.л.
 11. Луценко Е.В. Неформальная постановка и обсуждение задач, возникающих при системном обобщении теории множеств на основе системной теории информации (Часть 2-я: задачи 4–9) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №04(38). С. 26 – 65. – Шифр Информрегистра: 0420800012\0049. – Режим доступа: <http://ej.kubagro.ru/2008/04/pdf/03.pdf>, 2,5 у.п.л.
 12. Кун Т., Структура научных революций, М., Прогресс, 1977.
 13. Луценко Е.В. Неформальная постановка и обсуждение задач, возникающих при системном обобщении теории множеств на основе системной теории информации (Часть 1-я: задачи 1–3) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №03(37). С. 154 – 185. – Шифр Информрегистра: 0420800012\0031. – Режим доступа: <http://ej.kubagro.ru/2008/03/pdf/12.pdf>, 2 у.п.л.
 14. Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп. – Краснодар: КубГАУ, 2006. – 318с. – Режим доступа: http://lc.kubagro.ru/aidos/aidos06_lab/index.htm
 15. Луценко Е.В. Развитие интеллектуальной системы «Эйдос-астра», снимающее ограничения на размерность баз знаний и разрешение когнитивных функций / Е.В. Луценко, А.П. Трунев, Е.А. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №05(69). С. 353 – 377. – Режим доступа: <http://ej.kubagro.ru/2011/05/pdf/31.pdf>, 1,562 у.п.л.
 16. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(1). С. 79 – 91. – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>, 0,812 у.п.л.
 17. Луценко Е.В. Критерии реальности и принцип эквивалентности виртуальной и "истинной" реальности / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №06(8). С. 70 – 88. – Режим доступа: <http://ej.kubagro.ru/2004/06/pdf/10.pdf>, 1,188 у.п.л.