

УДК 004.8

5.2.2. Математические, статистические и инструментальные методы в экономике (физико-математические науки, экономические науки)

**МАТЕМАТИЧЕСКАЯ МОДЕЛЬ  
АВТОМАТИЗИРОВАННОГО СИСТЕМНО-  
КОГНИТИВНОГО АНАЛИЗА И СИСТЕМЫ  
«ЭЙДОС» В КВАНТОРНОМ И МАТРИЧНОМ  
ПРЕДСТАВЛЕНИИ**

Луценко Евгений Вениаминович

д.э.н., к.т.н., профессор

[Web of Science ResearcherID S-8667-2018](https://www.researcherid.org/rid/S-8667-2018)

Scopus Author ID: 57188763047

РИНЦ SPIN-код: 9523-7101

[prof.lutsenko@gmail.com](mailto:prof.lutsenko@gmail.com) <http://lc.kubagro.ru>

[https://www.researchgate.net/profile/Eugene\\_Lutsenko](https://www.researchgate.net/profile/Eugene_Lutsenko)

*Кубанский Государственный Аграрный университет имени И.Т.Трубилина, Краснодар, Россия*

Головин Никита Сергеевич

Студент

РИНЦ SPIN-код: 4735-4214

[nikitagolovin416@gmail.com](mailto:nikitagolovin416@gmail.com)

<http://nickup.byethost22.com>

*Элитная частная экономическая школа, г.Нови-Сад, Сербия*

В работе описаны кванторные и матричные модели Автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – системы «Эйдос», которые позволяют корректно интегрировать числовые и текстовые данные, выявлять причинно-следственные связи и поддерживать принятие решений в условиях неопределенности. Модели обеспечивают унификацию обработки данных, включая зашумленные и фрагментированные, с различными типами измерительных шкал и единиц измерения. Описаны алгоритмы синтеза и верификации моделей, а также методы прогнозирования и поддержки принятия решений. Предложенные подходы находят применение в междисциплинарных исследованиях, включая экономику, технические науки, сельское хозяйство, биологию, психологию и педагогику, медицину и многие другие направления науки. Работа демонстрирует эффективность метода в обработке больших данных, разработке моделей и создании интеллектуальных приложений, подтверждая универсальность и практическую ценность системы «Эйдос»

Ключевые слова: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, АСК-АНАЛИЗ, ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА «ЭЙДОС», КОГНИТИВНОЕ МОДЕЛИРОВАНИЕ, КВАНТОРНО-МАТРИЧНЫЕ МОДЕЛИ, АНАЛИЗ ДАННЫХ, ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ  
<http://dx.doi.org/10.21515/1990-4665-206-012>

UDC 004.8

5.2.2. Mathematical, statistical and instrumental methods in economics (physical and mathematical sciences, economic sciences)

**MATHEMATICAL MODEL OF AUTOMATED  
SYSTEM-COGNITIVE ANALYSIS AND THE  
EIDOS SYSTEM IN A QUANTIFIER-MATRIX  
REPRESENTATION**

Lutsenko Evgeniy Veniaminovich

Doctor of Economics, Cand.Tech.Sci., Professor

[Web of Science ResearcherID S-8667-2018](https://www.researcherid.org/rid/S-8667-2018)

Scopus Author ID: 57188763047

RSCI SPIN-code: 9523-7101

[prof.lutsenko@gmail.com](mailto:prof.lutsenko@gmail.com) <http://lc.kubagro.ru>

[https://www.researchgate.net/profile/Eugene\\_Lutsenko](https://www.researchgate.net/profile/Eugene_Lutsenko)

*Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia*

Golovin Nikita Sergeevich

Student

RSCI SPIN code: 4735-4214

[nikitagolovin416@gmail.com](mailto:nikitagolovin416@gmail.com)

[http://nickup.byethost22.com/](http://nickup.byethost22.com)

*Elite private economic school, Novi Sad, Serbia*

The article describes quantifier and matrix models of Automated System Cognitive Analysis (ASC-analysis) and its software toolkit, the Eidos system, which allow for the correct integration of numerical and textual data, identify cause-effect relationships and support decision-making in conditions of uncertainty. The models provide a unification of data processing, including noisy and fragmented data, with different types of measuring scales and units of measurement. Algorithms for the synthesis and verification of models, as well as methods for forecasting and decision support are described. The proposed approaches are used in interdisciplinary research, including economics, technical sciences, agriculture, biology, psychology and pedagogy, medicine and many other fields of science. The work demonstrates the effectiveness of the method in processing big data, developing models and creating intelligent applications, confirming the versatility and practical value of the Eidos system

Keywords: AUTOMATED SYSTEM-COGNITIVE ANALYSIS, ASC-ANALYSIS, INTELLIGENT EIDOS SYSTEM, COGNITIVE MODELING, QUANTIFIER-MATRIX MODELS, DATA ANALYSIS, DECISION SUPPORT

## СОДЕРЖАНИЕ

<b>1. ВВЕДЕНИЕ .....</b>	<b>3</b>
1.1. ОПИСАНИЕ ИССЛЕДУЕМОЙ ПРЕДМЕТНОЙ ОБЛАСТИ .....	3
1.2. ОБЪЕКТ И ПРЕДМЕТ ИССЛЕДОВАНИЯ .....	3
1.3. ПРОБЛЕМА, РЕШАЕМАЯ В РАБОТЕ И ЕЕ АКТУАЛЬНОСТЬ.....	3
1.4. ЦЕЛЬ РАБОТЫ.....	3
<b>2. МЕТОД.....</b>	<b>4</b>
2.1. АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ (АСК-АНАЛИЗ) КАК МЕТОД РЕШЕНИЯ ПРОБЛЕМЫ.....	4
2.2. СИСТЕМА «ЭЙДОС» – ИНСТРУМЕНТАРИЙ АСК-АНАЛИЗА.....	4
2.3. ЭТАПЫ АСК-АНАЛИЗА – ЗАДАЧИ РАБОТЫ .....	4
<b>3. РЕЗУЛЬТАТЫ (РЕШЕНИЕ ЗАДАЧ) .....</b>	<b>4</b>
3.1. <i>Задача-1.</i> КОГНИТИВНАЯ СТРУКТУРИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ. ДВЕ ИНТЕРПРЕТАЦИИ КЛАССИФИКАЦИОННЫХ И ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ .....	4
3.2. <i>Задача-2.</i> ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ .....	4
3.3. <i>Задача-3.</i> СИНТЕЗ СТАТИСТИЧЕСКИХ И СИСТЕМНО-КОГНИТИВНЫХ МОДЕЛЕЙ. МНОГОПАРАМЕТРИЧЕСКАЯ ТИПИЗАЦИЯ И ЧАСТНЫЕ КРИТЕРИИ ЗНАНИЙ.....	13
3.4. <i>Задача-4.</i> ВЕРИФИКАЦИЯ МОДЕЛЕЙ .....	36
3.5. <i>Задача-5.</i> ВЫБОР НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ .....	46
3.6. <i>Задача-6.</i> СИСТЕМНАЯ ИДЕНТИФИКАЦИЯ И ПРОГНОЗИРОВАНИЕ .....	46
3.7. <i>Задача-7.</i> ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ.....	50
3.8. <i>Задача-8.</i> ИССЛЕДОВАНИЕ ОБЪЕКТА МОДЕЛИРОВАНИЯ ПУТЕМ ИССЛЕДОВАНИЯ ЕГО МОДЕЛИ ....	55
<b>4. ОБСУЖДЕНИЕ .....</b>	<b>69</b>
4.1. СРАВНЕНИЕ С ДРУГИМИ СИСТЕМАМИ.....	69
4.2. ДОСТУПНОСТЬ И УДОБСТВО .....	70
4.3. ТЕХНОЛОГИЧЕСКИЕ ОСОБЕННОСТИ .....	70
4.4. ИНТЕГРАЦИЯ ЛОКАЛЬНЫХ И ОБЛАЧНЫХ РЕШЕНИЙ.....	71
4.5. ПРАКТИЧЕСКОЕ ЗНАЧЕНИЕ .....	71
Выводы.....	71
<b>5. ВЫВОДЫ И ПЕРСПЕКТИВЫ.....</b>	<b>71</b>
<b>ЛИТЕРАТУРА .....</b>	<b>72</b>

## **1. ВВЕДЕНИЕ**

Данная статья является кратким вариантом работы [1]. Все ссылки на литературные источники даны в соответствии со списком литературы работы [1].

### **1.1. Описание исследуемой предметной области**

Предметная область работы охватывает концептуальные и практические аспекты исследования многопараметрических систем, в которых необходимо выявление причинно-следственных связей, формализация знаний и поддержка принятия решений.

Сложные системы, рассматриваемые в исследовании, включают биологические, социально-экономические, технические и природные объекты. Особое внимание уделено междисциплинарному подходу, обеспечивающему интеграцию методов когнитивного моделирования, искусственного интеллекта, системного анализа и обработки информации.

Данная предметная область охватывает широкий спектр задач, от автоматизации аналитических процессов до создания уникальных методов для работы с разнородными данными. Исследуемая методология обладает высокой гибкостью и применимостью в различных научных и практических сферах, что подтверждается многолетним опытом использования системы «Эйдос» в решении задач экономического анализа, биологии, медицины и других областей.

### **1.2. Объект и предмет исследования**

Объект исследования – автоматизированные методы и технологии анализа данных

Предмет исследования – кванторно-матричные модели АСК-анализа и системы «Эйдос».

### **1.3. Проблема, решаемая в работе и ее актуальность**

Дополнительные пояснения даны в работе [1].

### **1.4. Цель работы**

Целью работы является кванторно-матричное описание математической модели Автоматизированного системно-когнитивного анализа (АСК-анализа) и интеллектуальной системы «Эйдос».

## 2. МЕТОД

### 2.1. Автоматизированный системно-когнитивный анализ (АСК-анализ) как метод решения проблемы

Автоматизированный системно-когнитивный анализ (АСК-анализ) предложен *проф.Е.В.Луценко* в 2002 году в ряде статей 1997-2001 годов<sup>1</sup> и фундаментальной монографии [2].

Дополнительные пояснения даны в работе [1].

### 2.2. Система «Эйдос» – инструментарий АСК-анализа

Дополнительные пояснения даны в работе [1].

### 2.3. Этапы АСК-анализа – задачи работы

Дополнительные пояснения даны в работе [1].

## 3. РЕЗУЛЬТАТЫ (РЕШЕНИЕ ЗАДАЧ)

### 3.1. Задача-1. Когнитивная структуризация предметной области. Две интерпретации классификационных и описательных шкал и градаций

Дополнительные пояснения даны в работе [1].

### 3.2. Задача-2. Формализация предметной области

Дополнительные пояснения даны в работе [1].

#### 3.2.1. Представление исходных данных

В системе «Эйдос» (режим 2.3.2) есть 6 автоматизированных программных интерфейсов (API), обеспечивающих ввод в систему числовых, текстовых и графических внешних данных (см. рисунок ниже).

2.3.2. Программные интерфейсы с внешними базами данных	2.3.2.1. Импорт данных из текстовых файлов
	2.3.2.2. Универсальный программный интерфейс импорта данных в систему
	2.3.2.3. Импорт данных из транспонированных внешних баз данных
	2.3.2.4. Оцифровка изображений по внешним контурам
	2.3.2.5. Оцифровка изображений по всем пикселям и спектру
	2.3.2.6. Сценарный АСК-анализ символьных и числовых рядов
	2.3.2.7. Транспонирование файлов исходных данных
	2.3.2.8. Объединение нескольких файлов исходных данных в один
	2.3.2.9. Разбиение TXT-файла на файлы-абзацы
	2.3.2.10. CSV => DBF конвертер системы "Эйдос"
	2.3.2.11. Прогноз событий по астропараметрам по Н.А.Чердниченко
	2.3.2.12. Прогнозирование землетрясений методом Н.А.Чердниченко
	2.3.2.13. Чемпионат RAIF-Challenge 2017-API-bank
	2.3.2.14. Чемпионат RAIF-Challenge 2017-API-retail
	2.3.2.15. Вставка промежуточных строк в файл исходных данных Inp_data

Автоматизированные программные интерфейсы системы «Эйдос»

<sup>1</sup> <http://lc.kubagro.ru/aidos/Sprab0802.pdf> (см. с публикации № 48).

Каждый из этих интерфейсов имеет много различных настроек, придающих им гибкость и возможность тщательной тонкой настройки параметров процессов конвертации.

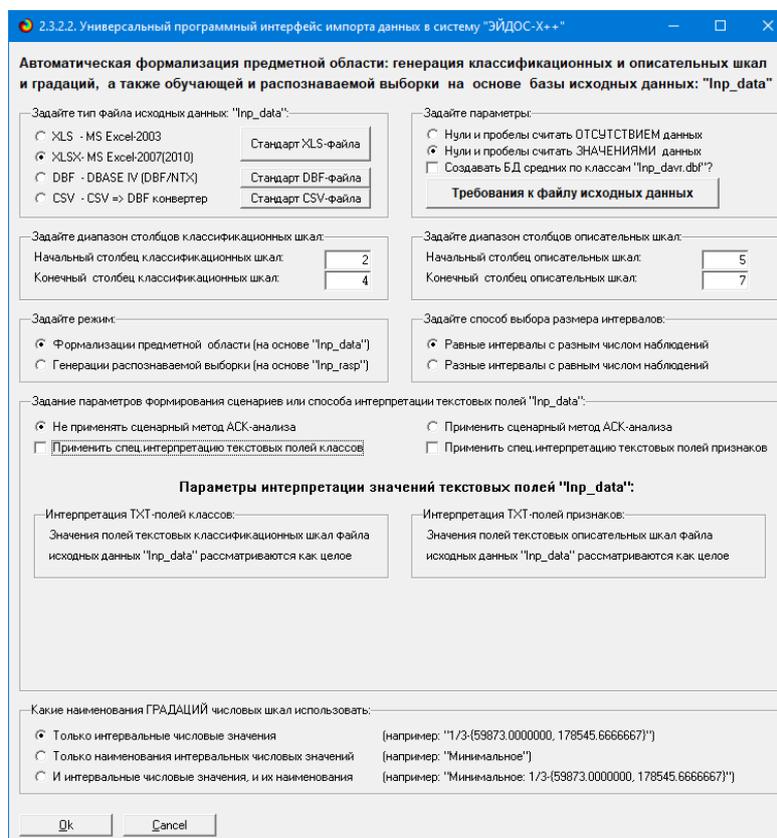
Но в данной работе в качестве примера мы рассмотрим один из наиболее развитых интерфейсов, обеспечивающий ввод в систему внешних табличных данных: API-2.3.2.2. С работой других программных интерфейсов можно ознакомиться в тематических подборках публикаций по адресу: <http://lc.kubagro.ru/aidos/Aidos-X.htm>.

Таблица исходных данных для API-2.3.2.2 должна иметь вид (см. таблицу):

Стандарт таблицы исходных данных для API-2.3.2.2

Наименование объектов обучающей выборки	Наименование 1-й классификационной шкалы	Наименование 2-й классификационной шкалы	...	Наименование 1-й описательной шкалы	Наименование 2-й описательной шкалы	...
1-й объект обучающей выборки	Значение градации классификационной шкалы	Значение градации классификационной шкалы	...	Значение градации описательной шкалы	Значение градации описательной шкалы	...
2-й объект обучающей выборки	Значение градации классификационной шкалы	Значение градации классификационной шкалы	...	Значение градации описательной шкалы	Значение градации описательной шкалы	...
...	...	...	...	...	...	...

Главная экранная форма управления API-2.3.2.2 приведены на рисунке 3:



Главная экранная форма управления API-2.3.2.2

На рисунках ниже приведены экранные формы подробного хелпа режима 2.3.2.2:

Помощь по режиму 2.3.2.2 для случая Excel-файлов исходных данных
— □ ×

Режим 2.3.2.2: Универсальный программный интерфейс импорта данных из внешней базы данных "Inp\_data.xls" в систему "Эйдос-Х++" и формализации предметной области.

- Данный программный интерфейс обеспечивает формализацию предметной области, т.е. анализ файла исходных данных Inp\_data.xls(x), формирование классификационных и описательных шкал и градаций, а затем кодирование файла исходных с их использованием.
- Файл исходных данных должен иметь имя: Inp\_data.xls(x), а файл распознаваемой выборки имя: Inp\_rasp.xls(x). Файлы Inp\_data.xls(x) и Inp\_rasp.xls(x) должны находиться в папке ../AIDOS-X/AID\_DATA/Inp\_data/. Эти файлы имеют совершенно одинаковую структуру.
- 1-я строка этого файла должна содержать наименования колонок на любом языке, в т.ч. и русском. Эти наименования должны быть во всех колонках, при этом переносы по словам разрешены, а объединение ячеек, разрыв строки знак абзаца не допускаются. Эти наименования должны быть короткими, но понятными, т.к. они будут в выходных формах, а к ним еще будут добавляться наименования градаций. В числовых шкалах надо ОБЯЗАТЕЛЬНО указывать единицы измерения и число знаков после запятой в колонке должно быть ОДИНАКОВОЕ.
- 1-я колонка содержит наименование объекта обучающей выборки или наименование наблюдения. Оно может быть длинным: до 255 символов.
- Каждая строка этого файла, начиная со 2-й, содержит данные об одном объекте обучающей выборки или одном наблюдении. В MS Excel-2003 в листе может быть до 65536 строк и до 256 колонок. В листе MS Excel-2010 и более поздних возможно до 1048576 строк и 16384 колонок.
- Столбцы, начиная со 2-го, являются классификационными и описательными шкалами и могут быть текстового (номинального / порядкового) или числового типа (с десятичными знаками после запятой).
- Столбцу присваивается числовой тип, если все значения его ячеек числового типа. Если хотя бы одно значение является текстовым (не числом, в т.ч. пробелом), то столбцу присваивается текстовый тип. Это означает, что нули должны быть указаны нулями, а не пробелами.
- Столбцы со 2-го по N-й являются классификационными шкалами (выходными параметрами) и содержат данные о классах (будущих состояниях объекта управления), к которым принадлежат объекты обучающей выборки.
- Столбцы с N+1 по последний являются описательными шкалами (свойствами или факторами) и содержат данные о признаках (т.е. значениях свойств или значениях факторов), характеризующих объекты обучающей выборки.
- В результате работы режима формируется файл INP\_NAME.TXT стандарта MS DOS (кириллица), в котором наименования классификационных и описательных шкал являются СТРОКАМИ. Система формирует классификационные и описательные шкалы и градации. Для этого в каждом числовом столбце система находит минимальное и максимальное числовые значения и формирует заданное количество числовых интервалов, после чего числовые значения заменяются их интервальными значениями. В текстовых столбцах система находит уникальные текстовые значения. Каждое УНИКАЛЬНОЕ интервальное числовое или текстовое значение считается градацией классификационной или описательной шкалы, характеризующей объект. В каждой шкале ее градации сортируются по алфавиту. С использованием шкал и градаций кодируются исходные данные в результате чего генерируется обучающая выборка, каждый объект которой соответствует одной строке файла исходных данных INP\_DATA и содержит коды классов, соответствующие фактам совпадения числовых или уникальных текстовых значений классов с градациями классификационных шкал и коды признаков, соответствующие фактам совпадения числовых или уникальных текстовых значений признаков с градациями описательных шкал
- Распознаваемая выборка формируется на основе файла INP\_RASP аналогично, за исключением того, что классификационные и описательные шкалы и градации не создаются, а используются ранее созданные в модели, и базы распознаваемой выборки могут не включать коды классов, если столбцы классов в файле INP\_RASP были пустыми. Структура файла INP\_RASP должна быть такая же, как INP\_DATA, т.е. они должны ПОЛНОСТЬЮ совпадать по наименованиям столбцов, но могут иметь разное количество строк с разными значениями в них.

Принцип организации таблицы исходных данных:

Наименование объекта обучающей выборки	Наименование 1-й классификационной шкалы	Наименование 2-й классификационной шкалы	...	Наименование 1-й описательной шкалы	Наименование 2-й описательной шкалы	...
1-й объект обучающей выборки (1-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
2-й объект обучающей выборки (2-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
...	...	...	...	...	...	...

**Определения основных терминов и профилактика типичных ошибок при подготовке Excel-файла исходных данных**

Первая экранная форма хелпа API-2.3.2.2

Помощь по режиму 2.3.2.2 для случая Excel-файлов исходных данных

Режим 2.3.2.2: Универсальный программный интерфейс импорта данных из внешней базы данных "Inp\_data.xls(x)" в систему "Эйдос-X++"  
**ТЕРМИНЫ АСК-АНАЛИЗА И СИСТЕМЫ "ЭЙДОС":**

Шкала представляет собой способ формализации предметной области. Используются числовые и текстовые шкалы, при этом текстовые могут быть номинальными и порядковыми. На номинальных шкалах есть только отношения эквивалентности и неэквивалентности, на порядковых, кроме того еще отношения "больше", "меньше", а на числовых - кроме того могут выполняться все арифметические операции. Каждый объект выборки (наблюдение) описан с одной стороны своими признаками, а с другой - принадлежностью к некоторым обобщающим категориям (классам). Такая структура описания называется онтологией или фреймом экземпляром и является базовой для всех моделей представления знаний. В АСК-анализе и системе "Эйдос" используются три интерпретации шкал и градаций: универсальная, статическая и динамическая:

- в универсальной интерпретации: признаки - это градации описательных шкал;
- в статической интерпретации: описательная шкала - это свойство, а градация (признак) - это степень выраженности этого свойства;
- в динамической интерпретации: описательная шкала - это фактор, а градация (признак) - это значение фактора;
- в универсальной интерпретации: классы - это градации классификационных шкал;
- в статической интерпретации: классификационная шкала - способ классификации обобщающих категорий (классов), к которым в настоящем времени по отношению к признакам относятся состояния объекта моделирования;
- в динамической интерпретации: классификационная шкала - способ классификации обобщающих категорий (классов), к которым в будущем времени по отношению к признакам относятся состояния объекта прогнозирования или управления;

**ПРОФИЛАКТИКА ОШИБОК В ФАЙЛЕ ИСХОДНЫХ ДАННЫХ:**

- 1-я строка файла "Inp\_data.xls(x)" должна содержать наименования колонок. Эти наименования должны быть во всех колонках, при этом переносы по словам разрешены, а объединение ячеек, разрыв строки знак абзаца и неалфавитные символы не допускаются. Эти наименования должны быть короткими, но понятными, т.к. они будут в выходных формах, а к ним еще будут добавляться наименования градаций. В числовых шкалах надо обязательно указывать единицы измерения. Число знаков после запятой в числовой колонке должно быть одинаковым.
- 1-я колонка содержит наименование объекта обучающей выборки или наименование наблюдения. Оно может быть длинным: до 255 символов.
- Столбцы, начиная со 2-го, являются классификационными и описательными шкалами и могут быть текстового (номинального / порядкового) или числового типа (со знаками после запятой). Чтобы текстовая шкала была порядковой, нужно чтобы при сортировке по алфавиту градации этой шкалы образовывали осмысленную последовательность от минимального значения до максимального. Например, текстовая шкала "Размер" с градациями: "очень малое", "малое", "среднее", "большое", "очень большое", будет номинальной шкалой, т.к. при сортировке по алфавиту они расположатся в порядке: "большое", "малое", "очень большое", "очень малое", "среднее". Чтобы шкала "Размер" стала порядковой нужно в этих градациях присвоить следующие значения: "1/5-очень малое", "2/5-малое", "3/5-среднее", "4/5-большое", "5/5-очень большое".
- Столбцу присваивается числовой тип, если все значения его ячеек числового типа. Если хотя бы одно значение является текстовым (не числом, в т.ч. пробелом), то столбцу присваивается текстовый тип. Это означает, что нули должны быть указаны нулями, а не пробелами. Если в системе "Эйдос" в режимах 2.1, 2.2 посмотреть на градации классификационных и описательных шкал, которые должны быть числовыми, то сразу будет видно, в какой форме представлены числа: числовыми диапазонами или прямо числами. Если числовыми диапазонами, значит в файле исходных данных в этом отношении все правильно, если же числами, то возможно в Excel-файле нужно заменить десятичные точки на запятые, а также найти и исправить нечисловые данные в числовых по смыслу колонках. Быстро найти их можно перейдя на последнюю строку файла исходных данных и задав расчет суммы колонки. В формуле будет видно с какой строки идет расчет суммы. Если со 2-й, то значит все верно, иначе будет указана строка, в которой находится нечисловое значение.
- Система "Эйдос" работает с областью данных файла исходных данных, которую можно выделить блоком, поставив курсор в ячейку A1, нажав Ctrl+Home, а затем зажав клавиши Shift+Ctrl нажать End. Если этот блок выходит за пределы области таблицы, фактически занятой данными надо скопировать эту фактическую область данных в буфер обмена, создать новый лист и скопировать в него, а исходный лист удалить.
- Иногда бывает полезно сбросить все форматирование Excel-таблицы исходных данных. Это можно сделать в MS Excel. А можно скопировать таблицу в MS Word, а потом обратно в MS Excel.

Принцип организации таблицы исходных данных:

Наименование объекта обучающей выборки	Наименование 1-й классификационной шкалы	Наименование 2-й классификационной шкалы	...	Наименование 1-й описательной шкалы	Наименование 2-й описательной шкалы	...
1-й объект обучающей выборки (1-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
2-й объект обучающей выборки (2-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
...	...	...	...	...	...	...

Вторая экранная форма хелпа API-.2.3.2.2

Обучающая выборка, заданная в виде таблицы 1, может быть формально описана с использованием понятий векторов, матриц и кванторов следующим образом.

1-я колонка таблицы 1 – это информация, идентифицирующая наблюдение.

Потом идут колонки на желтом фоне – классификационные шкалы, градации которых являются способами группировки наблюдений по обобщающим категориям – классам, это результаты влияния факторов. Количество классификационных шкал –  $C$ .

Затем идут колонки на белом фоне – описательные шкалы, факторы, влияющие на объект моделирования. Количество факторов –  $F$ .

И классификационные, и описательные шкалы могут быть всех типов.

Шкалы могут быть текстового и числового типа.

Шкалы текстового типа могут быть номинальные и порядковые, а также бинарные (дихотомические).

Шкалы числового типа с единицей измерения. Они могут быть интервальными и шкалами отношений.

В интервальных числовых шкалах нет начала отсчета.

В числовых шкалах отношений есть начало отсчета.

Каждое наблюдение описывается двумя векторами: вектором значений факторов и вектором результатов влияния этих значений факторов.

Обучающая выборка, представленная в виде таблицы (таблица 1), описывается в терминах векторов, матриц и кванторов следующим образом.

#### 1. Структура данных

1. Первая колонка не является шкалой – это информация, идентифицирующая наблюдение.

2. Желтые колонки – классификационные шкалы, определяющие обобщающие категории (классы), результат влияния факторов. Количество классификационных шкал обозначим как  $C$ .

3. Белые колонки – описательные шкалы (факторы), определяющие характеристики объекта моделирования. Количество факторов обозначим как  $F$ .

Каждое наблюдение состоит из двух частей:

- Вектор факторов ( $x_i$ ) описывает значения факторов для одного наблюдения.

- Вектор результатов ( $y_i$ ) отражает результаты влияния факторов.

#### 2. Типы шкал

Все шкалы, как классификационные, так и описательные, делятся на текстовые и числовые:

- Текстовые шкалы:

- Номинальные ("категория" $_j$ ).
- Порядковые ( $r_1 < r_2 < \dots < r_k$ ).
- Бинарные ( $\{0,1\}$ ).
- Числовые шкалы:
- Интервальные ( $\mathbb{R}$ , без нулевой точки).
- Шкалы отношений ( $\mathbb{R}^+$ , с нулевой точкой).

Формальное описание обучающей выборки

### 1. Общая структура данных

Пусть:

- $N$  количество наблюдений (строк в таблице).
- $C$  количество классификационных шкал (обобщающие категории, классы).
- $F$  количество факторов (описывающих шкал).
- $\mathbf{X}$  матрица размерности  $N \times F$ , где каждая строка вектор значений факторов для одного наблюдения.
- $\mathbf{Y}$  матрица размерности  $N \times C$ , где каждая строка вектор значений классификационных шкал для одного наблюдения.

Наблюдение  $i$  ( $i \in \{1, \dots, N\}$ ) описывается двумя векторами:

- $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iF})^T \in \mathbb{X}^F$  вектор значений факторов.
- $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iC})^T \in \mathbb{Y}^C$  вектор результатов их влияния (классов).

Здесь:

- $\mathbb{X}^F$  пространство значений факторов (включая числовые и текстовые шкалы).
- $\mathbb{Y}^C$  пространство классов, т.е. результатов влияния факторов.

### 2. Представление данных в матричной форме

Обучающая выборка представляется в виде пары матриц:

$$\mathcal{D} = (\mathbf{X}, \mathbf{Y})$$

где:

- $\mathbf{X} = [\mathbf{x}_1^T; \mathbf{x}_2^T; \dots; \mathbf{x}_N^T] \in \mathbb{R}^{N \times F}$  матрица факторов;
- $\mathbf{Y} = [\mathbf{y}_1^T; \mathbf{y}_2^T; \dots; \mathbf{y}_N^T] \in \mathbb{R}^{N \times C}$  матрица классов.

Типы данных и шкал

Факторы ( $\mathbf{X}$ )

Каждый фактор  $x_{ij}$  ( $i = 1, \dots, N; j = 1, \dots, F$ ) принимает значение из множества, зависящего от типа шкалы:

- Для номинальных шкал:  $x_{ij} \in \{\text{категория}_1, \text{категория}_2, \dots\}$ .
- Для порядковых шкал:  $x_{ij} \in \{r_1, r_2, \dots, r_k\}$ , где  $r_1 < r_2 < \dots < r_k$ .
- Для бинарных шкал:  $x_{ij} \in \{0,1\}$ .
- Для числовых шкал:
- Интервальные:  $x_{ij} \in \mathbb{R}$ , но без нулевой точки.

- Шкалы отношений:  $x_{ij} \in \mathbb{R}^+$ , с нулевой точкой.

Классы ( $\mathbf{Y}$ )

Каждый элемент  $y_{ij}$  ( $i = 1, \dots, N; j = 1, \dots, C$ ) соответствует результату влияния факторов и принимает значение из:

- Для номинальных шкал:  $y_{ij} \in \{\text{класс}_1, \text{класс}_2, \dots\}$ .
- Для порядковых шкал:  $y_{ij} \in \{r_1, r_2, \dots, r_k\}$ , где  $r_1 < r_2 < \dots < r_k$ .
- Для бинарных шкал:  $y_{ij} \in \{0, 1\}$ .
- Для числовых шкал:
- Интервальные:  $y_{ij} \in \mathbb{R}$ , но без нулевой точки.
- Шкалы отношений:  $y_{ij} \in \mathbb{R}^+$ , с нулевой точкой.

Формализация с кванторами

Для каждого наблюдения  $i$ :

$$\exists \mathbf{x}_i \in \mathbb{X}^F, \exists \mathbf{y}_i \in \mathbb{Y}^C, \text{ такие что } (\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{D}.$$

Множество всех наблюдений:

$$\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i) \mid i = 1, \dots, N\}.$$

Совместное описание факторов и классов

$$\forall i \in \{1, \dots, N\}, \exists (\mathbf{x}_i, \mathbf{y}_i) \in \mathbb{X}^F \times \mathbb{Y}^C: (\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{D}.$$

Пример визуализации

Для обучающей выборки из 3 наблюдений ( $N = 3$ ), 2 классификационных шкал ( $C = 2$ ) и 4 факторов ( $F = 4$ ), структура данных может быть представлена следующим образом:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \end{bmatrix}.$$

Здесь строки матрицы  $\mathbf{X}$  соответствуют значениям факторов, а строки  $\mathbf{Y}$  результатам влияния.

Заключение

Формальное представление обучающей выборки с помощью матриц и кванторов позволяет унифицировать описание данных различных типов.

### 3.2.2. Формирование классификационных и описательных шкал и градаций

Для формализации процесса разработки справочников, кодирования данных и формирования базы событий с использованием кванторов и векторно-матричного представления:

Сначала разрабатываются справочники классификационных и описательных шкал и градаций.

Затем исходные данные кодируются с их помощью и получается база событий.

1. Формализация справочников классификационных и описательных шкал

Пусть:

-  $\mathcal{C} = \{c_1, c_2, \dots, c_C\}$  множество классификационных шкал (индексы шкал), где каждая шкала  $c_j$  имеет конечное множество градаций  $G_{c_j} = \{g_{c_j,1}, g_{c_j,2}, \dots, g_{c_j,k_{c_j}}\}$ .

-  $\mathcal{F} = \{f_1, f_2, \dots, f_F\}$  множество описательных шкал (индексы факторов), где каждая шкала  $f_i$  имеет множество градаций  $G_{f_i} = \{g_{f_i,1}, g_{f_i,2}, \dots, g_{f_i,k_{f_i}}\}$ .

- Для каждой шкалы  $s \in \mathcal{C} \cup \mathcal{F}$  существует отображение:

$$\forall g_{s,m} \in G_s, \exists \text{ код } k_{s,m} \in \mathbb{Z}^+,$$

где  $k_{s,m}$  уникальный идентификатор градации  $g_{s,m}$  из справочника.

**3.2.3. Кодирование исходных данных с использованием классификационных и описательных шкал и градаций и формирование обучающей выборки**

2. Кодирование исходных данных

Пусть исходные данные представлены в виде наблюдений:

-  $N$  число наблюдений;

-  $\mathbf{X} \in \mathbb{R}^{N \times (C+F)}$  матрица наблюдений, где:

- первые  $C$  столбцов значения классификационных шкал;

- последние  $F$  столбцов значения описательных шкал.

Процесс кодирования:

Для наблюдения  $n$ , шкалы  $s$ , и значения  $x_{n,s}$ :

- Если шкала  $s$  номинальная или порядковая (текстовая):

$$\forall x_{n,s}, \exists k_{s,m} \in \mathbb{Z}^+, \text{ такой, что } x_{n,s} = g_{s,m}.$$

Тогда  $x_{n,s}$  заменяется кодом  $k_{s,m}$ .

Процесс кодирования числовых шкал

Если шкала  $s$  числовая (интервальная или шкала отношений):

- Каждое значение  $x_{n,s}$  преобразуется в интервал:

$$x_{n,s} \in \mathbb{R}, \text{ либо } x_{n,s} \in \mathbb{R}^+.$$

Значения заменяются на их интервальные коды, заданные разбиением диапазона значений на конечное число интервалов  $I_s = \{I_{s,1}, I_{s,2}, \dots, I_{s,K_s}\}$ , где  $K_s$  число интервалов, которое задается в диалоге.

Алгоритм кодирования:

1. Для каждой числовой шкалы  $s$ :

- Определяется диапазон значений  $[\min_s, \max_s]$ , где  $\min_s = \min(x_{n,s})$ ,  $\max_s = \max(x_{n,s})$ .

- Диапазон разбивается на  $K_s$  равных интервалов с разным числом наблюдений или разных (адаптивных) интервалов с одинаковым в целых числах числом наблюдений (по решению пользователя):

$$I_{s,k} = [a_k, a_{k+1}), \text{ где } a_k = \min_s + \frac{(k-1) \cdot (\max_s - \min_s)}{K_s}, k \in \{1, \dots, K_s\}.$$

2. Для каждого значения  $x_{n,s}$ :

- Определяется интервал  $I_{s,k}$ , которому оно принадлежит:

$$x_{n,s} \in I_{s,k} \Rightarrow x_{n,s} = k, \text{ где } k \in \{1, \dots, K_s\}.$$

Таким образом,  $x_{n,s}$  заменяется на код  $k$ , соответствующий номеру интервала.

Пример:

Пусть  $s$  шкала с диапазоном значений  $[0,100]$ , и задано  $K_s = 5$ . Интервалы будут:  $I_{s,1} = [0,20)$ ,  $I_{s,2} = [20,40)$ ,  $I_{s,3} = [40,60)$ ,  $I_{s,4} = [60,80)$ ,  $I_{s,5} = [80,100]$ . Если  $x_{n,s} = 45$ , то  $x_{n,s} \in I_{s,3}$ , и значение заменяется на  $k = 3$ .

Результат кодирования для числовых шкал

После кодирования числовые значения в матрице  $\mathbf{X}$  заменяются на интервальные коды. Закодированная матрица данных:

$$\mathbf{X}_{\text{код}} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1(C+F)} \\ x_{21} & x_{22} & \dots & x_{2(C+F)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{N(C+F)} \end{bmatrix},$$

где  $x_{n,s}$  для числовых шкал это номер соответствующего интервала  $k$ .

Формирование базы событий

База событий  $\mathbf{E}$ , как и ранее, включает идентификаторы наблюдений  $ID_n$  и закодированные данные:

$$\mathbf{E} = \begin{bmatrix} ID_1 & x_{11} & x_{12} & \dots & x_{1(C+F)} \\ ID_2 & x_{21} & x_{22} & \dots & x_{2(C+F)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ID_N & x_{N1} & x_{N2} & \dots & x_{N(C+F)} \end{bmatrix}.$$

Где числовые шкалы представлены интервалами, что унифицирует данные и делает их готовыми для дальнейшего анализа.

Формально:

$$\forall n \in \{1, \dots, N\}, \exists (ID_n, \mathbf{x}_n), \mathbf{x}_n \in \mathbb{R}^{C+F}, \mathbf{x}_n = \{x_{n1}, x_{n2}, \dots, x_{n(C+F)}\}.$$

Это гарантирует, что каждая строка базы событий связана с уникальным наблюдением  $n$ .

Итог:

Полученная база событий  $\mathbf{E}$  содержит закодированные данные, где классификационные и описательные шкалы унифицированы, что позволяет использовать их для последующего анализа, обработки и моделирования.

### 3.3. Задача-3. Синтез статистических и системно-когнитивных моделей. Многопараметрическая типизация и частные критерии знаний

Дополнительные пояснения даны в работе [1].

#### 3.3.1. Суть математической модели АСК-анализа

1. Определение матрицы абсолютных частот

Матрица абсолютных частот  $N$  рассчитывается для всех комбинаций значений классов и факторов, где:

$$N \in \mathbb{R}^{C \times F};$$

$N_{ij}$  – количество наблюдений, принадлежащих классу  $j$ , при значении фактора  $i$ .

Каждый элемент матрицы  $N$  задаётся следующим образом:

$$N_{ij} = \sum_{k=1}^N \delta(y_{kj}, x_{ki}),$$

где:

-  $\delta(y_{kj}, x_{ki})$  индикаторная функция, равная 1, если наблюдение  $k$  принадлежит классу  $j$  и имеет значение фактора  $i$ , и 0 в противном случае:

$$\delta(y_{kj}, x_{ki}) = \begin{cases} 1, & \text{если } y_{kj} = 1 \text{ и } x_{ki} = 1, \\ 0, & \text{в противном случае.} \end{cases}$$

Это классическая **булева индикаторная функция (или характеристическая функция)**, которая проверяет выполнение условия и возвращает 1, если условие выполнено, и 0 в противном случае.

Источник и смысл переменных  $y_{kj}$  и  $x_{ki}$  в индикаторной функции

Переменные  $y_{kj}$  и  $x_{ki}$  вводятся как способ формального описания принадлежности наблюдений к классам и значений факторов в задаче анализа данных. Они используются для построения матрицы  $N_{ij}$ , которая отражает взаимосвязь между классами и значениями факторов.

1.  $y_{kj}$ :

- Переменная  $y_{kj}$  описывает, принадлежит ли наблюдение  $k$  классу  $j$ .

- Если  $y_{kj} = 1$ , то наблюдение  $k$  принадлежит классу  $j$ .

- Если  $y_{kj} = 0$ , то наблюдение  $k$  не принадлежит классу  $j$ .

Переменная  $y_{kj}$  обычно берётся из классификации данных. Например, если мы анализируем данные о животных, где классы  $j$  это "млекопитающие", "птицы", "рептилии" и т.д., то  $y_{kj}$  указывает, к какому классу  $j$  относится каждое наблюдение  $k$ .

2.  $x_{ki}$ :

- Переменная  $x_{ki}$  описывает, имеет ли наблюдение  $k$  значение фактора  $i$ .

- Если  $x_{ki} = 1$ , то у наблюдения  $k$  присутствует значение фактора  $i$ .

- Если  $x_{ki} = 0$ , то значение фактора  $i$  у наблюдения  $k$  отсутствует.

Факторы  $i$  могут быть любыми характеристиками наблюдений. Например, если мы изучаем животных, факторы  $i$  могут быть "наличие шерсти", "наличие крыльев", "наличие чешуи" и т.д.

Пример происхождения: Предположим, у нас есть набор данных о 5 животных:

- Наблюдения  $k$ : каждое животное (например, собака, воробей, змея и т.д.).
- Классы  $j$ : типы животных (например, млекопитающие, птицы, рептилии).
- Факторы  $i$ : характеристики животных (например, наличие шерсти, крыльев, чешуи).

Наблюдение (k)	Млекопитающие (y_k1)	Птицы (y_k2)	Наличие шерсти (x_k1)	Наличие крыльев (x_k2)
1	1	0	1	0
2	0	1	0	1
3	0	0	0	0
4	1	0	1	1
5	0	1	0	1

Итог:

- Переменные  $y_{kj}$  и  $x_{ki}$  происходят из структуры данных, где:
- $y_{kj}$  это информация о принадлежности наблюдений  $k$  к классам  $j$  (классификация).
- $x_{ki}$  это информация о наличии значений факторов  $i$  у наблюдений  $k$  (характеристики).
- Они задаются на этапе предобработки данных, чтобы математически формализовать задачу анализа.

Таким образом, матрица  $N$  (матрица абсолютных частот) описывает полное распределение наблюдений по классам и значениям факторов на основе данных обучающей выборки.

На основе таблицы матрицы абсолютных частот рассчитываются матрицы условных и безусловных процентных распределений (таблицы приведены ниже).

Для того чтобы рассчитать матрицы условных и безусловных процентных распределений на основе матрицы абсолютных частот, используем кванторный и векторно-матричный формализм.

**Матрица абсолютных частот (статистическая модель ABS)**

		Классы					Сумма
		1	...	$j$	...	$W$	
Значения факторов	1	$N_{11}$		$N_{1j}$		$N_{1W}$	
	...						
	$i$	$N_{i1}$		$N_{ij}$		$N_{iW}$	$N_{i\Sigma} = \sum_{j=1}^W N_{ij}$
	...						
	$M$	$N_{M1}$		$N_{Mj}$		$N_{MW}$	
Суммарное количество признаков по классу				$N_{\Sigma j} = \sum_{i=1}^M N_{ij}$			$N_{\Sigma\Sigma} = \sum_{i=1}^W \sum_{j=1}^M N_{ij}$
Суммарное количество объектов обучающей выборки по классу				$N_{\Sigma j}$			$N_{\Sigma\Sigma} = \sum_{j=1}^W N_{\Sigma j}$

**Примечания**

Предположим, что в матрице абсолютных частот  $N$ :

-  $N_{ij}$  количество наблюдений, принадлежащих классу  $j$  при значении фактора  $i$ .

-  $M$  количество значений факторов.

-  $C$  количество классов.

Матричная форма:

$$N = \begin{pmatrix} N_{11} & N_{12} & \dots & N_{1C} \\ N_{21} & N_{22} & \dots & N_{2C} \\ \vdots & \vdots & \ddots & \vdots \\ N_{M1} & N_{M2} & \dots & N_{MC} \end{pmatrix}$$

где  $N \in \mathbb{R}^{M \times C}$ .

**1. Безусловные процентные распределения**

Безусловные распределения характеризуют вероятности появления признаков и классов без учета других факторов.

**1.1. Безусловная вероятность класса**

Безусловная вероятность класса  $y_j$  это вероятность того, что наблюдение принадлежит классу  $y_j$ , независимо от значений факторов  $x_i$ .

Для того чтобы найти безусловную вероятность класса, нужно сложить все значения в столбце  $j$  матрицы  $N$  и нормализовать по общему количеству наблюдений:

$$P(y_j) = \frac{\sum_{i=1}^M N_{ij}}{\sum_{i=1}^M \sum_{j=1}^C N_{ij}}$$

где:

-  $\sum_{i=1}^M N_{ij}$  это общее количество наблюдений в классе  $y_j$ ,

-  $\sum_{i=1}^M \sum_{j=1}^C N_{ij}$  это общее количество наблюдений.

Векторно-матричная запись:

Общая безусловная вероятность класса  $P(\mathbf{y})$  для всех классов может быть записана как:

$$P(\mathbf{y}) = \frac{\mathbf{N}\mathbf{1}_M}{\sum_{i=1}^M \sum_{j=1}^C N_{ij}},$$

где:

-  $\mathbf{1}_M$  это вектор размерности  $M$  (вектор из единиц, т.е.  $[1,1, \dots, 1]^T$ ),

-  $\mathbf{N}\mathbf{1}_M$  вектор, содержащий суммы по столбцам матрицы  $N$ , т.е.

$\sum_{i=1}^M N_{ij}$  для каждого класса  $j$ .

## 1.2. Безусловная вероятность признака

Безусловная вероятность признака  $x_i$  это вероятность того, что для произвольного наблюдения наблюдается значение признака  $x_i$ , независимо от того, к какому классу оно принадлежит.

Для этого нужно сложить все значения в строке  $i$  и нормализовать по общему числу наблюдений:

$$P(x_i) = \frac{\sum_{j=1}^C N_{ij}}{\sum_{i=1}^M \sum_{j=1}^C N_{ij}},$$

где:

-  $\sum_{j=1}^C N_{ij}$  это общее количество наблюдений для признака  $x_i$ .

Векторно-матричная запись:

Общая безусловная вероятность признака  $P(\mathbf{x})$  для всех признаков может быть записана как:

$$P(\mathbf{x}) = \frac{\mathbf{N}^T \mathbf{1}_C}{\sum_{i=1}^M \sum_{j=1}^C N_{ij}},$$

где:

-  $\mathbf{1}_C$  вектор размерности  $C$  (вектор из единиц, т.е.  $[1,1, \dots, 1]^T$ ),

-  $\mathbf{N}^T \mathbf{1}_C$  вектор, содержащий суммы по строкам матрицы  $N$ , т.е.

$\sum_{j=1}^C N_{ij}$  для каждого признака  $i$ .

## 2. Условные процентные распределения

Условные распределения показывают вероятности появления значений признаков при условии, что наблюдение принадлежит определенному классу.

### 2.1. Условная вероятность признака $x_i$ при классе $y_j$

Условная вероятность признака  $x_i$ , если мы знаем, что наблюдение принадлежит классу  $y_j$ , вычисляется по формуле:

$$P(x_i | y_j) = \frac{N_{ij}}{\sum_{k=1}^M N_{kj}},$$

где:

- $N_{ij}$  количество наблюдений с признаком  $x_i$  в классе  $y_j$ ,
- $\sum_{k=1}^M N_{kj}$  это общее количество наблюдений в классе  $y_j$ .

Векторно-матричная запись:

Для всех признаков  $x_i$  при условии класса  $y_j$ , матрица условных вероятностей  $P(x | y)$  может быть записана как:

$$P(x | y) = \frac{N}{N\mathbf{1}_M},$$

где:

- $N$  матрица абсолютных частот,
- $N\mathbf{1}_M$  вектор, содержащий суммы по строкам матрицы  $N$ , то есть  $\sum_{k=1}^M N_{kj}$ .

Каждая строка матрицы  $P(x | y)$  нормализуется так, чтобы сумма элементов в строке была равна 1.

Итоговая структура

1. Безусловные распределения: - Безусловные вероятности классов и признаков можно рассчитать с использованием нормировки по соответствующим суммам.

2. Условные распределения: - Условные вероятности признаков при классе нормализуются по количеству наблюдений в классе.

Таким образом, расчет матриц условных и безусловных процентных распределений сводится к суммированию и нормировке элементов матрицы абсолютных частот  $N$ .

**Матрица условных и безусловных процентных распределений (статистические модели PRC1 и PRC2)**

		Классы					Безусловная вероятность признака
		1	...	j	...	w	
Значения факторов	1	$P_{11}$		$P_{1j}$		$P_{1w}$	
	...						
	i	$P_{i1}$		$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$		$P_{iw}$	$P_{i\Sigma} = \frac{N_{i\Sigma}}{N_{\Sigma\Sigma}}$
	...						
	M	$P_{M1}$		$P_{Mj}$		$P_{Mw}$	
Безусловная вероятность класса				$P_{\Sigma j}$			

Затем на основе матрицы абсолютных частот и матриц условных и безусловных процентных распределений с использованием частных критериев, знаний приведенных в таблице ниже, рассчитываются матрицы 7 системно-когнитивных моделей (таблицы приведены ниже).

В таблице частных критериев приведены формулы:

– для сравнения фактических и теоретических абсолютных частот;  
– для сравнения условных и безусловных относительных частот («вероятностей»).

И это **сравнение** в таблицах 8 и 9 осуществляется двумя возможными способами: путем **вычитания** и путем **деления**.

Обозначения к таблице:

*i* – значение прошлого параметра;

*j* – значение будущего параметра;

*N<sub>ij</sub>* – количество встреч *j*-го значения будущего параметра при *i*-м значении прошлого параметра;

*M* – суммарное число значений всех прошлых параметров;

*W* – суммарное число значений всех будущих параметров.

*N<sub>i</sub>* – количество встреч *i*-м значения прошлого параметра по всей выборке;

*N<sub>j</sub>* – количество встреч *j*-го значения будущего параметра по всей выборке;

*N* – количество встреч *j*-го значения будущего параметра при *i*-м значении прошлого параметра по всей выборке.

*I<sub>ij</sub>* – частный критерий знаний: количество знаний в факте наблюдения *i*-го значения прошлого параметра о том, что объект перейдет в состояние, соответствующее *j*-му значению будущего параметра;

$\Psi = \text{Log}(W)/\text{Log}(N)$  – нормировочный коэффициент (Е.В.Луценко, 1979), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли;

*P<sub>i</sub>* – безусловная относительная частота встречи *i*-го значения прошлого параметра в обучающей выборке;

*P<sub>ij</sub>* – условная относительная частота встречи *i*-го значения прошлого параметра при *j*-м значении будущего параметра.

**Различные аналитические формы частных критериев знаний, применяемые в АСК-анализе и системе «Эйдос»**

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	Через относительные частоты	Через абсолютные частоты
<b>ABS</b> , матрица абсолютных частот, $N_{ij}$ - фактическое число встреч $i$ -го признака у объектов $j$ -го класса; $\bar{N}_{ij}$ - теоретическое число встреч $i$ -го признака у объектов $j$ -го класса; $N_i$ – суммарное количество признаков в $i$ -й строке; $N_j$ – суммарное количество признаков или объектов обучающей выборки в $j$ -м классе; $N$ – суммарное количество признаков по всей выборке (таблица 7)	$N_{ij}$ – фактическая частота; $N_i = \sum_{j=1}^W N_{ij}; N_j = \sum_{i=1}^M N_{ij}; N = \sum_{i=1}^W \sum_{j=1}^M N_{ij};$ $\bar{N}_{ij} = \frac{N_i N_j}{N}$ – теоретическая (ожидаемая) частота.	
<b>PRC1</b> , матрица условных $P_{ij}$ и безусловных $P_i$ процентных распределений, в качестве $N_j$ используется суммарное количество признаков по классу	---	$P_{ij} = \frac{N_{ij}}{N_j}; P_i = \frac{N_i}{N}$
<b>PRC2</b> , матрица условных $P_{ij}$ и безусловных $P_i$ процентных распределений, в качестве $N_j$ используется суммарное количество объектов обучающей выборки по классу	---	
<b>INF1</b> , частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета вероятностей: $N_j$ – суммарное количество признаков по $j$ -му классу. Вероятность того, что если у объекта $j$ -го класса обнаружен признак, то это $i$ -й признак	$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}}{\bar{N}_{ij}} = \Psi \times \text{Log}_2 \frac{N_{ij} N}{N_i N_j}$
<b>INF2</b> , частный критерий: количество знаний по А.Харкевичу, 2-й вариант расчета вероятностей: $N_j$ – суммарное количество объектов по $j$ -му классу. Вероятность того, что если предъявлен объект $j$ -го класса, то у него будет обнаружен $i$ -й признак.	$I_{ij} = \Psi \times (\text{Log}_2 P_{ij} - \text{Log}_2 P_i)$	$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}}{\bar{N}_{ij}} = \Psi \times (\text{Log}_2 N_{ij} - \text{Log}_2 \bar{N}_{ij})$
<b>INF3</b> , частный критерий: Хи-квадрат: разности между фактическими и теоретически ожидаемыми абсолютными частотами	---	$I_{ij} = N_{ij} - \bar{N}_{ij} = N_{ij} - \frac{N_i N_j}{N}$
<b>INF4</b> , частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: $N_j$ – суммарное количество признаков по $j$ -му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}}{\bar{N}_{ij}} - 1 = \frac{N_{ij} N}{N_i N_j} - 1$
<b>INF5</b> , частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: $N_j$ – суммарное количество объектов по $j$ -му классу		
<b>INF6</b> , частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: $N_j$ – суммарное количество признаков по $j$ -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N} = \frac{N_{ij} N - N_i N_j}{N_j N}$
<b>INF7</b> , частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: $N_j$ – суммарное количество объектов по $j$ -му классу		

**Матрица системно-когнитивной модели в общем виде**

		Классы					Значимость фактора
		1	...	j	...	W	
Значения факторов	1	$I_{11}$		$I_{1j}$		$I_{1W}$	$\sigma_{1\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	i	$I_{i1}$		$I_{ij}$		$I_{iW}$	$\sigma_{i\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	M	$I_{M1}$		$I_{Mj}$		$I_{MW}$	$\sigma_{M\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

Ниже в таблицах ниже приведены матрицы конкретных системно-когнитивных моделей (СК-моделей).

**Матрица системно-когнитивных моделей INF1 и INF2**

		Классы					Значимость фактора
		1	...	j	...	W	
Значения факторов	1	$I_{11}$		$I_{1j}$		$I_{1W}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	i	$I_{i1}$		$I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$		$I_{iW}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	M	$I_{M1}$		$I_{Mj}$		$I_{MW}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$\sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

**Матрица системно-когнитивной модели INF3**

		Классы					Значимость фактора
		1	...	<i>j</i>	...	<i>W</i>	
Значения факторов	1	$I_{11}$		$I_{1j}$		$I_{1W}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	<i>i</i>	$I_{i1}$		$I_{ij} = N_{ij} - \frac{N_i N_j}{N}$		$I_{iW}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	<i>M</i>	$I_{M1}$		$I_{Mj}$		$I_{MW}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$\sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

**Матрица системно-когнитивных моделей INF4 и INF5**

		Классы					Значимость фактора
		1	...	<i>j</i>	...	<i>W</i>	
Значения факторов	1	$I_{11}$		$I_{1j}$		$I_{1W}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	<i>i</i>	$I_{i1}$		$I_{ij} = \frac{P_{ij}}{P_i} - 1$		$I_{iW}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	<i>M</i>	$I_{M1}$		$I_{Mj}$		$I_{MW}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$\sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

**Матрица системно-когнитивных моделей INF6 и INF7**

		Классы					Значимость фактора
		1	...	<i>j</i>	...	<i>W</i>	
Значения факторов	1	$I_{11}$		$I_{1j}$		$I_{1W}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{1j} - \bar{I}_1)^2}$
	...						
	<i>i</i>	$I_{i1}$		$I_{ij} = P_{ij} - P_i$		$I_{iW}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
	...						
	<i>M</i>	$I_{M1}$		$I_{Mj}$		$I_{MW}$	$\sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{Mj} - \bar{I}_M)^2}$
Степень редукции класса		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$\sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

**3.3.2. Кванторно-матричная математическая модель взаимосвязи меры  $\chi$ -квадрат Карла Пирсона с семантической мерой целесообразности информации Александра Харкевича, с коэффициентом возврата инвестиций (ROI) и коэффициентом взаимосвязи**

**3.3.2.1. Постановка задачи и обозначения**

Сформулируем прямые и обратные математические взаимосвязи между всеми системно-когнитивными моделями (СК-модели):

1. Мера  $\chi$ -квадрат  $I_{\chi^2}$  (СК-модель INF3);
2. Семантическая мера Харкевича  $I_H$  (СК-модели INF1 и INF2);
3. Коэффициент возврата инвестиций  $I_{ROI}$  (СК-модели INF4 и INF5);
4. Коэффициент взаимосвязи:  $I_{REL}$  (СК-модели INF6 и INF7).

Найдем конкретный вид прямых и обратных функций:

Прямые функции	Обратные функции
$I_{\chi^2} = f_{12}(I_H)$	$I_H = g_{21}(I_{\chi^2})$
$I_{\chi^2} = f_{13}(I_{ROI})$	$I_{ROI} = g_{31}(I_{\chi^2})$
$I_{\chi^2} = f_{14}(I_{REL})$	$I_{REI} = g_{41}(I_{\chi^2})$
$I_H = f_{23}(I_{ROI})$	$I_{ROI} = g_{32}(I_H)$
$I_H = f_{24}(I_{REL})$	$I_{REI} = g_{42}(I_H)$
$I_{ROI} = f_{34}(I_{REL})$	$I_{REI} = g_{43}(I_{ROI})$

Используем обозначения:

$N_{ij}$  – фактическая частота встреч  $i$  – го признака в  $j$  – м классе,

$N_i = \sum_{j=1}^W N_{ij}$  – суммарное количество встреч  $i$  – го признака,

$N_j = \sum_{i=1}^M N_{ij}$  – суммарное количество встреч признаков в  $j$  – м классе,

$N = \sum_{i=1}^W \sum_{j=1}^M N_{ij}$  – суммарное количество встреч признаков по всей выборке,

$\bar{N}_{ij} = \frac{N_i N_j}{N}$  – теоретическая (ожидаемая) частота.

$P_{ij} = \frac{N_{ij}}{N_j}$  – условная вероятность встречи  $i$  – го признака в  $j$  – м классе,

$P_i = \frac{N_i}{N}$  – безусловная вероятность встречи  $i$  – го признака по всей выборке,

$\frac{N_{ij}}{\bar{N}_{ij}} = \frac{P_{ij}}{P_i}$ .

### 3.3.2.2. Формулы для частных критериев

1. Мера  $\chi$ -квадрат  $I_{\chi^2}$  статистические отклонения:

Мера  $\chi$ -квадрат оценивает отклонение фактической частоты  $N_{ij}$  от теоретической (ожидаемой) частоты  $\bar{N}_{ij}$ :

$$I_{\chi^2} = N_{ij} - \bar{N}_{ij}.$$

2. Семантическая мера Харкевича  $I_H$  значимость информации:

Семантическая мера  $I_H$  зависит от отношения фактической частоты  $N_{ij}$  к теоретической частоте  $\bar{N}_{ij}$  через логарифм:

$$I_H = \Psi \cdot \log_2 \frac{N_{ij}}{\bar{N}_{ij}}.$$

3. Коэффициент возврата инвестиций  $I_{ROI}$  экономическая эффективность:

Коэффициент ROI оценивает экономическую эффективность через вероятность признаков  $P_{ij}$  и  $P_i$ , где:

$$P_{ij} = \frac{N_{ij}}{N_j}, \quad P_i = \frac{N_i}{N}.$$

Тогда коэффициент ROI выражается как:

$$I_{ROI} = \frac{P_{ij} - P_i}{P_i} = \frac{\frac{N_{ij}}{N_j} - \frac{N_i}{N}}{\frac{N_i}{N}} = \frac{N_{ij}}{\bar{N}_{ij}} - 1.$$

4. Коэффициент взаимосвязи  $I_{REI}$

$$I_{REI} = P_{ij} - P_i = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$$

### 3.3.2.3. Очевидные взаимосвязи, обнаруживающиеся при переходе от линейных шкал к логарифмическим

Из таблицы 10 мы видим, что:

1. Количество информации по Александру Харкевичу отличается от хи-квадрат только тем, что для измерения фактических и теоретических частот используется логарифмическая шкала:

$$\text{Inf3:} \quad I_{ij} = N_{ij} - \bar{N}_{ij} = N_{ij} - \frac{N_i N_j}{N}$$

$$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}}{\bar{N}_{ij}} = \Psi \times \text{Log}_2 \frac{N_{ij} N}{N_i N_j}$$

Inf1, Inf2:

$$I_{ij} = \Psi \times \text{Log}_2 \frac{N_{ij}}{\bar{N}_{ij}} = \Psi \times (\text{Log}_2 N_{ij} - \text{Log}_2 \bar{N}_{ij})$$

2. Количество информации по Александру Харкевичу отличается от коэффициента взаимосвязи только тем, что для измерения условных и безусловных относительных частот используется логарифмическая шкала:

$$\text{Inf1, Inf2:} \quad I_{ij} = \Psi \times \text{Log}_2 \frac{P_{ij}}{P_i}$$

$$I_{ij} = \Psi \times (\text{Log}_2 P_{ij} - \text{Log}_2 P_i)$$

Inf6, Inf7:

$$I_{ij} = P_{ij} - P_i$$

### 3.3.2.4. Прямая и обратная связь величин $I_{\chi^2}$ и $I_H$

Для того чтобы найти вид функции  $f_{12}$ , которая связывает величины  $I_{\chi^2}$  и  $I_H$ :

$$I_{\chi^2} = f_{12}(I_H)$$

нужно понять, как они соотносятся друг с другом на основе их математических определений. Рассмотрим выражения для каждой из величин:

1. Величина  $I_{\chi^2}$  (элемент мера хи-квадрат):

$$I_{\chi^2} = N_{ij} - \bar{N}_{ij},$$

где:

-  $N_{ij}$  это некоторое наблюдаемое значение в таблице сопряженности (или других данных),

-  $\bar{N}_{ij}$  это ожидаемое значение, рассчитанное с учетом модели или гипотезы.

2. Величина  $I_H$  (энтропия):

$$I_H = \Psi \cdot \log_2 \left( \frac{N_{ij}}{\bar{N}_{ij}} \right),$$

где:

-  $\Psi$  некоторый коэффициент масштабирования (константа),

-  $\log_2$  логарифм по основанию 2.

Теперь, чтобы выразить  $I_{\chi^2}$  через  $I_H$ , можно попытаться провести аналогию или преобразование через  $N_{ij}$  и  $\bar{N}_{ij}$ . Для этого начнем с того, что выразим отношение  $\frac{N_{ij}}{\bar{N}_{ij}}$  через  $I_H$ .

Из уравнения для  $I_H$  получаем:

$$I_H = \Psi \cdot \log_2 \left( \frac{N_{ij}}{\bar{N}_{ij}} \right),$$

что можно решить относительно отношения:

$$\frac{N_{ij}}{\bar{N}_{ij}} = 2^{I_H/\Psi}.$$

Теперь, чтобы найти зависимость  $I_{\chi^2}$  от  $I_H$ , можно переписать  $N_{ij}$  через  $\bar{N}_{ij}$ :

$$N_{ij} = \bar{N}_{ij} \cdot 2^{I_H/\Psi}.$$

Подставим это в выражение для  $I_{\chi^2}$ :

$$I_{\chi^2} = \bar{N}_{ij} \cdot 2^{I_H/\Psi} - \bar{N}_{ij}.$$

Вынесем  $\bar{N}_{ij}$  за скобки:

$$I_{\chi^2} = \bar{N}_{ij} (2^{I_H/\Psi} - 1).$$

Таким образом, зависимость  $I_{\chi^2}$  от  $I_H$  имеет вид:

$$\boxed{I_{\chi^2} = \bar{N}_{ij} (2^{I_H/\Psi} - 1).}$$

Где  $\bar{N}_{ij}$  и  $\Psi$  это параметры, которые могут быть известны или заданы для конкретной задачи.

Чтобы найти обратную функцию  $I_H = g_{21}(I_{\chi^2})$ , нужно решить уравнение для  $I_H$  в терминах  $I_{\chi^2}$ , исходя из полученной зависимости:

$$I_{\chi^2} = \bar{N}_{ij} (2^{I_H/\Psi} - 1).$$

Шаги:

1. Выразим  $2^{I_H/\Psi}$ :

Из исходного уравнения:

$$I_{\chi^2} = \bar{N}_{ij} (2^{I_H/\Psi} - 1),$$

прибавим 1 к обеим частям:

$$I_{\chi^2} + 1 = \bar{N}_{ij} \cdot 2^{I_H/\Psi}.$$

Теперь поделим обе части на  $\bar{N}_{ij}$ :

$$\frac{I_{\chi^2} + 1}{\bar{N}_{ij}} = 2^{I_H/\Psi}.$$

2. Применим логарифм:

Теперь применим логарифм по основанию 2 к обеим частям:

$$\log_2 \left( \frac{I_{\chi^2} + 1}{\bar{N}_{ij}} \right) = \frac{I_H}{\Psi}.$$

3. Решим относительно  $I_H$ :

Теперь умножим обе части на  $\Psi$ , чтобы выразить  $I_H$ :

$$I_H = \Psi \cdot \log_2 \left( \frac{I_{\chi^2} + 1}{\bar{N}_{ij}} \right).$$

Таким образом, обратная функция для  $I_H$  от  $I_{\chi^2}$  имеет вид:

$$I_H = g_{21}(I_{\chi^2}) = \Psi \cdot \log_2 \left( \frac{I_{\chi^2} + 1}{\bar{N}_{ij}} \right).$$

$$I_H = \Psi \cdot \log_2 \left( \frac{I_{\chi^2} + 1}{\bar{N}_{ij}} \right).$$

Это и есть искомая обратная функция  $g_{21}(I_{\chi^2})$ .

### 3.3.2.5. Прямая и обратная связь величин $I_{\chi^2}$ и $I_{ROI}$

Найдём конкретный вид функций  $I_{\chi^2} = f_{13}(I_{ROI})$  и  $I_{ROI} = g_{31}(I_{\chi^2})$ , используя данные формулы и выражения.

1. Определения:

$$- I_{\chi^2} = N_{ij} - \bar{N}_{ij},$$

$$- I_{ROI} = \frac{P_{ij} - P_i}{P_i},$$

$$- P_{ij} = \frac{N_{ij}}{N_j},$$

$$- P_i = \frac{N_i}{N}.$$

2. Выразим  $I_{\chi^2}$  через  $I_{ROI}$ :

Подставим  $P_{ij} = \frac{N_{ij}}{N_j}$  и  $P_i = \frac{N_i}{N}$  в формулу  $I_{ROI}$ :

$$I_{ROI} = \frac{\frac{N_{ij}}{N_j} - \frac{N_i}{N}}{\frac{N_i}{N}}.$$

Приведём числитель к общему знаменателю:

$$I_{ROI} = \frac{\frac{N_{ij}}{N_j} - \frac{N_i}{N}}{\frac{N_i}{N}} = \frac{\frac{N_{ij}N - N_iN_j}{N_jN}}{\frac{N_i}{N}}.$$

Упростим выражение:

$$I_{ROI} = \frac{N}{N_i} \cdot \frac{N_{ij}N - N_iN_j}{N_jN}.$$

Сократим N в числителе и знаменателе:

$$I_{ROI} = \frac{N_{ij}N - N_iN_j}{N_iN_j}.$$

Теперь выразим  $N_{ij}$  через  $I_{ROI}$ :

$$N_{ij} = \frac{N_iN_j}{N} \cdot (1 + I_{ROI}).$$

Подставим это значение  $N_{ij}$  в формулу для  $I_{\chi^2}$ :

$$I_{\chi^2} = N_{ij} - \bar{N}_{ij}.$$

Обычно  $\bar{N}_{ij} = \frac{N_iN_j}{N}$ , так как это ожидаемое значение в таблице сопряженности. Таким образом:

$$I_{\chi^2} = \bar{N}_{ij} \cdot (1 + I_{ROI}) - \bar{N}_{ij}$$

Вынесем  $\bar{N}_{ij}$  за скобки:

$$I_{\chi^2} = \bar{N}_{ij} \cdot ((1 + I_{ROI}) - 1).$$

Упростим выражение:

$$I_{\chi^2} = \bar{N}_{ij} \cdot I_{ROI}.$$

Вид функции  $f_{13}(I_{ROI})$ :

$$I_{\chi^2} = f_{13}(I_{ROI}) = \bar{N}_{ij} \cdot I_{ROI},$$

$$\boxed{I_{\chi^2} = \bar{N}_{ij} \cdot I_{ROI}}$$

3. Обратная функция  $I_{ROI} = g_{31}(I_{\chi^2})$ :

Теперь выразим  $I_{ROI}$  через  $I_{\chi^2}$ . Из формулы выше:

$$I_{\chi^2} = \bar{N}_{ij} \cdot I_{ROI}.$$

Разделим обе стороны на  $\bar{N}_{ij}$ :

$$I_{ROI} = \frac{I_{\chi^2}}{\bar{N}_{ij}}.$$

Вид функции  $g_{31}(I_{\chi^2})$ :

$$I_{ROI} = g_{31}(I_{\chi^2}) = \frac{I_{\chi^2}}{\bar{N}_{ij}},$$

$$\boxed{I_{ROI} = \frac{I_{\chi^2}}{\bar{N}_{ij}}}$$

Итоговые формулы:

1. Прямая функция:

$$I_{\chi^2} = \bar{N}_{ij} \cdot I_{ROI}.$$

2. Обратная функция:

$$I_{ROI} = \frac{I_{\chi^2}}{\bar{N}_{ij}}.$$

### 3.3.2.6. Прямая и обратная связь величин $I_{\chi^2}$ и $I_{REL}$

Выразим  $I_{\chi^2}$  через  $I_{REL}$  и найдем обратную зависимость. Начнем с заданных формул:

Заданные величины:

1. Квадратичное отклонение:

$$I_{\chi^2} = N_{ij} - \bar{N}_{ij},$$

где  $N_{ij}$  наблюдаемое значение, а  $\bar{N}_{ij}$  ожидаемое значение.

2. Разностное энтропийное отклонение, т.е. коэффициент взаимосвязи (REL):

$$I_{REL} = P_{ij} - P_i,$$

где:

$$P_{ij} = \frac{N_{ij}}{N_j}, \quad P_i = \frac{N_i}{N},$$

и:

-  $N_j = \sum_i N_{ij}$  сумма по строкам для j-й колонки,

-  $N_i = \sum_j N_{ij}$  сумма по столбцам для i-й строки,

-  $N = \sum_{ij} N_{ij}$  общий объем.

Найдем  $I_{\chi^2}$  через  $I_{REL}$ :

1. Подставим  $P_{ij} = \frac{N_{ij}}{N_j}$  и  $P_i = \frac{N_i}{N}$  в формулу  $I_{REL}$ :

$$I_{REL} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}.$$

2. Выразим  $N_{ij}$  через  $I_{REL}$ :

Умножим обе части на  $N_j$ , чтобы убрать знаменатель:

$$N_{ij} = N_j \cdot I_{REL} + N_j \cdot \frac{N_i}{N}.$$

3. Подставим это выражение  $N_{ij}$  в формулу для  $I_{\chi^2}$ :

$$I_{\chi^2} = N_{ij} - \bar{N}_{ij}.$$

Ожидаемое значение  $\bar{N}_{ij}$  обычно определяется как  $\bar{N}_{ij} = \frac{N_i N_j}{N}$ .

Подставим  $N_{ij}$  из предыдущего шага:

$$I_{\chi^2} = \left( N_j \cdot I_{REL} + N_j \cdot \frac{N_i}{N} \right) - \frac{N_i N_j}{N}.$$

Упростим выражение:

$$I_{\chi^2} = N_j \cdot I_{REL}.$$

Вид функции  $I_{\chi^2} = f_{14}(I_{REL})$ :

$$I_{\chi^2} = f_{14}(I_{REL}) = N_j \cdot I_{REL},$$

$$\boxed{I_{\chi^2} = N_j \cdot I_{REL}.}$$

Найдем обратную функцию  $I_{REL} = g_{41}(I_{\chi^2})$ :

Рассмотрим зависимость:

$$I_{\chi^2} = N_j \cdot I_{REL}.$$

Выразим  $I_{REL}$  через  $I_{\chi^2}$ :

$$\boxed{I_{REL} = \frac{I_{\chi^2}}{N_j},}$$

Итоговые функции:

$$1. I_{\chi^2} = f_{14}(I_{REL}) = N_j \cdot I_{REL},$$

$$2. I_{REL} = g_{41}(I_{\chi^2}) = \frac{I_{\chi^2}}{N_j}.$$

### 3.3.2.7. Прямая и обратная связь величин $I_H$ и $I_{ROI}$

Для нахождения функций  $I_H = f_{23}(I_{ROI})$  и  $I_{ROI} = g_{32}(I_H)$ , начнем с того, что выразим взаимосвязь между величинами  $I_H$  и  $I_{ROI}$  через их определения.

Шаг 1: Зависимость  $I_H$  от  $I_{ROI}$

Имеем:

$$I_H = \Psi \cdot \log_2 \frac{P_{ij}}{P_i},$$

и

$$I_{ROI} = \frac{P_{ij} - P_i}{P_i}.$$

Из второго уравнения выразим  $P_{ij}$ :

$$P_{ij} = P_i \cdot (1 + I_{ROI}).$$

Подставим это в выражение для  $I_H$ :

$$I_H = \Psi \cdot \log_2 \frac{P_i \cdot (1 + I_{ROI})}{P_i}.$$

Сокращая  $P_i$ , получаем:

$$\boxed{I_H = \Psi \cdot \log_2(1 + I_{ROI}).}$$

Шаг 2: Зависимость  $I_{ROI}$  от  $I_H$

Теперь выразим  $I_{ROI}$  из уравнения:

$$I_H = \Psi \cdot \log_2(1 + I_{ROI}).$$

Разделим обе части на  $\Psi$ :

$$\frac{I_H}{\Psi} = \log_2(1 + I_{ROI}).$$

Применим экспоненциальную функцию с основанием 2:

$$2^{I_H/\Psi} = 1 + I_{ROI}.$$

Вычтем 1 из обеих частей:

$$I_{ROI} = 2^{I_H/\Psi} - 1.$$

Итоговые функции

1. Функция  $I_H = f_{23}(I_{ROI})$ :

$$I_H = \Psi \cdot \log_2(1 + I_{ROI}).$$

2. Функция  $I_{ROI} = g_{32}(I_H)$ :

$$I_{ROI} = 2^{I_H/\Psi} - 1.$$

### 3.3.2.8. Прямая и обратная связь величин $I_H$ и $I_{REL}$

Для нахождения функций  $I_H = f_{24}(I_{REL})$  и  $I_{REL} = g_{42}(I_H)$ , выразим зависимости, исходя из заданных формул.

Определения:

1. Величина  $I_H$ :

$$I_H = \Psi \cdot \log_2\left(\frac{P_{ij}}{P_i}\right),$$

где:

$$P_{ij} = \frac{N_{ij}}{N_j}, \quad P_i = \frac{N_i}{N}.$$

2. Величина  $I_{REL}$ :

$$I_{REL} = P_{ij} - P_i.$$

Выражение  $I_H = f_{24}(I_{REL})$ :

1. Подставим  $P_{ij} = P_i + I_{REL}$  (из выражения для  $I_{REL}$ ) в формулу для  $I_H$ :

$$I_H = \Psi \cdot \log_2\left(\frac{P_i + I_{REL}}{P_i}\right).$$

2. Упростим дробь:

$$I_H = \Psi \cdot \log_2\left(1 + \frac{I_{REL}}{P_i}\right).$$

Таким образом:

$$I_H = f_{24}(I_{REL}) = \Psi \cdot \log_2\left(1 + \frac{I_{REL}}{P_i}\right),$$

$$I_H = \Psi \cdot \log_2\left(1 + \frac{I_{REL}}{P_i}\right).$$

Выражение  $I_{REL} = g_{42}(I_H)$ :

1. Начнем с формулы для  $I_H$ :

$$I_H = \Psi \cdot \log_2\left(1 + \frac{I_{REL}}{P_i}\right).$$

2. Выразим логарифм:

$$\frac{I_H}{\Psi} = \log_2\left(1 + \frac{I_{REL}}{P_i}\right).$$

3. Преобразуем логарифм в показательную форму:

$$2^{I_H/\Psi} = 1 + \frac{I_{REL}}{P_i}.$$

4. Выразим  $I_{REL}$ :

$$I_{REL} = P_i \cdot (2^{I_H/\Psi} - 1).$$

Таким образом:

$$I_{REL} = g_{42}(I_H) = P_i \cdot (2^{\frac{I_H}{\Psi}} - 1),$$

$$I_{REL} = P_i \cdot (2^{\frac{I_H}{\Psi}} - 1).$$

Итоговые формулы:

1.  $I_H = f_{24}(I_{REL})$ :

$$I_H = \Psi \cdot \log_2 \left( 1 + \frac{I_{REL}}{P_i} \right).$$

2.  $I_{REL} = g_{42}(I_H)$ :

$$I_{REL} = P_i \cdot (2^{I_H/\Psi} - 1).$$

### 3.3.2.9. Прямая и обратная связь величин $I_{ROI}$ и $I_{REL}$

Для нахождения функций  $f_{34}(I_{REL})$  и  $g_{43}(I_{ROI})$ , нам нужно выразить  $I_{ROI}$  через  $I_{REL}$  и наоборот.

Шаг 1: Определения

$$1. I_{ROI} = \frac{P_{ij} - P_i}{P_i} = \frac{\frac{N_{ij} - N_i}{N_j} - \frac{N_i}{N}}{\frac{N_i}{N}}. \quad 2. I_{REL} = P_{ij} - P_i = \frac{N_{ij}}{N_j} - \frac{N_i}{N}.$$

Шаг 2: Выразим  $I_{ROI}$  через  $I_{REL}$

Имеем связь между  $I_{ROI}$  и  $I_{REL}$ :

$$I_{ROI} = \frac{I_{REL}}{P_i}.$$

Так как  $P_i = \frac{N_i}{N}$ , окончательный вид функции:

$$I_{ROI} = f_{34}(I_{REL}) = \frac{I_{REL}}{P_i} = \frac{I_{REL}}{\frac{N_i}{N}} = \frac{N}{N_i} \cdot I_{REL},$$

$$I_{ROI} = \frac{N}{N_i} \cdot I_{REL}.$$

Шаг 3: Выразим  $I_{REL}$  через  $I_{ROI}$

Из определения  $I_{ROI}$ :

$$I_{ROI} = \frac{I_{REL}}{P_i}.$$

Умножим обе части на  $P_i$ :

$$I_{REL} = I_{ROI} \cdot P_i.$$

Подставим  $P_i = \frac{N_i}{N}$ :

$$I_{REL} = g_{43}(I_{ROI}) = I_{ROI} \cdot \frac{N_i}{N},$$

$$I_{REL} = I_{ROI} \cdot \frac{N_i}{N}.$$

Итоговые функции

$$1. I_{ROI} = f_{34}(I_{REL}) = \frac{N}{N_i} \cdot I_{REL}.$$

$$2. I_{REL} = g_{43}(I_{ROI}) = I_{ROI} \cdot \frac{N_i}{N}.$$

### 3.3.2.10. Сводная таблица прямых и обратных функций, связывающих частные критерии

Итак, прямые и обратные функции, связывающие частные критерии, имеют следующий конкретный вид:

Прямые функции	Обратные функции
$I_{\chi^2} = \bar{N}_{ij} (2^{I_H/\Psi} - 1)$	$I_H = \Psi \cdot \log_2 \left( \frac{I_{\chi^2} + 1}{\bar{N}_{ij}} \right)$
$I_{\chi^2} = \bar{N}_{ij} \cdot I_{ROI}$	$I_{ROI} = \frac{I_{\chi^2}}{\bar{N}_{ij}}$
$I_{\chi^2} = N_j \cdot I_{REL}$	$I_{REL} = \frac{I_{\chi^2}}{N_j}$
$I_H = \Psi \cdot \log_2 (1 + I_{ROI})$	$I_{ROI} = 2^{I_H/\Psi} - 1$
$I_H = \Psi \cdot \log_2 \left( 1 + \frac{I_{REL}}{P_i} \right)$	$I_{REL} = P_i \cdot (2^{I_H/\Psi} - 1)$
$I_{ROI} = \frac{I_{REL}}{P_i}$	$I_{REL} = I_{ROI} \cdot P_i$

Из данной таблицы видно, что все частные критерии, кроме  $I_H$  и  $I_{ROI}$  (количество информации по Харкевичу и коэффициент возврата инвестиций) выражаются друг через друга с использованием значений из матриц абсолютных частот и условных и безусловных процентных распределений и только кроме  $I_H$  и  $I_{ROI}$  выражаются непосредственно друг через друга.

### 3.3.3. Средневзвешенный частный критерий (коллектив решающих правил) на основе $\chi$ -квадрат, семантической мерой целесообразности информации, коэффициента возврата инвестиций и коэффициента взаимосвязи

Кванторно-матричная математическая модель

1. Мера  $\chi$ -квадрат Карла Пирсона (INF3):

$$I_{\chi^2} = N_{ij} - \bar{N}_{ij},$$

где  $\bar{N}_{ij} = \frac{N_i N_j}{N}$  теоретически ожидаемая частота, а  $N_i, N_j, N$  суммарные частоты для признака  $i$ , класса  $j$  и всей выборки соответственно.

2. Семантическая мера Харкевича (INF1):

$$I_H = \Psi \cdot \log_2 \frac{N_{ij}}{\bar{N}_{ij}},$$

где  $\Psi$  коэффициент значимости информации.

3. Коэффициент возврата инвестиций (ROI):

$$I_{ROI} = \frac{P_{ij} - P_i}{P_i} = \frac{\frac{N_{ij}}{N_j} - \frac{N_i}{N}}{\frac{N_i}{N}},$$

где:

-  $P_{ij} = \frac{N_{ij}}{N_j}$  условная вероятность,

-  $P_i = \frac{N_i}{N}$  безусловная вероятность.

Нормировка весовых коэффициентов

Коэффициенты  $w_1, w_2, w_3$  являются нормировочными, т.е. их сумма равна единице:

$$w_1 + w_2 + w_3 = 1,$$

и их значения выбираются в зависимости от достоверности моделей, например:

- Если модель  $\chi$ -квадрат показывает высокую точность,  $w_1$  будет доминировать.

- Если значимость информации является критичной, увеличивается  $w_2$ .

- Если важна экономическая интерпретация, приоритет отдаётся  $w_3$ .

Задача выбора весов может быть решена через оптимизацию, например:

$$w_1, w_2, w_3 = \operatorname{argmax} \text{Достоверность модели на обучающей выборке.}$$

Обобщённое выражение модели Объединённый критерий записывается как:

$$I_{\text{общ}} = w_1 \cdot I_{\chi^2} + w_2 \cdot I_H + w_3 \cdot I_{ROI},$$

где:

- $w_1, w_2, w_3 \geq 0$ ,
- $w_1 + w_2 + w_3 = 1$ .

#### Интерпретация

- $w_1$ : отражает вклад статистического анализа ( $\chi$ -квадрат).
- $w_2$ : показывает значимость семантической меры.
- $w_3$ : оценивает эффективность системы с точки зрения инвестиций (ROI).

Эта модель позволяет гибко настраивать вес каждого критерия в зависимости от достоверности данных и целей анализа.

Для оценки достоверности моделей используется F-метрика Ван Ризбергера, т.е. весовые коэффициенты  $w_1, w_2, w_3$  могут быть определены на основе этой метрики, учитывающей гармоническое среднее между точностью (precision) и полнотой (recall).

Включение F-метрики в модель

Формула F-метрики:

$$F_{\beta} = \frac{(1 + \beta^2) \cdot \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}},$$

где:

- Precision (точность): доля релевантных данных среди всех предсказанных как релевантные,
- Recall (полнота): доля релевантных данных среди всех релевантных,
- $\beta$ : параметр, определяющий приоритет между точностью ( $\beta < 1$ ) и полнотой ( $\beta > 1$ ).

Применение F-метрики для весовых коэффициентов: Каждой модели ( $I_{\chi^2}, I_H, I_{ROI}$ ) может быть назначен вес на основе её F-метрики:

$$w_k = \frac{F_{\beta,k}}{\sum_{k=1}^3 F_{\beta,k}},$$

где  $k \in \{1,2,3\}$  соответствует  $\chi$ -квадрату, мере Харкевича и ROI соответственно. Это гарантирует нормировку весов:

$$w_1 + w_2 + w_3 = 1.$$

Итоговая модель с нормировкой через F-метрику Объединённый критерий записывается следующим образом:

$$I_{\text{общ}} = w_1 \cdot I_{\chi^2} + w_2 \cdot I_H + w_3 \cdot I_{ROI},$$

где:

- $w_1 = \frac{F_{\beta,\chi^2}}{F_{\beta,\chi^2} + F_{\beta,H} + F_{\beta,ROI}}$ ,
- $w_2 = \frac{F_{\beta,H}}{F_{\beta,\chi^2} + F_{\beta,H} + F_{\beta,ROI}}$ ,

$$- W_3 = \frac{F_{\beta,ROI}}{F_{\beta,\chi^2} + F_{\beta,H} + F_{\beta,ROI}}$$

Пример расчёта весов 1. Для каждой модели ( $I_{\chi^2}, I_H, I_{ROI}$ ) рассчитываются:

- Precision ( $P_k$ ),
- Recall ( $R_k$ ),
- F-метрика ( $F_{\beta,k}$ ).

2. Вес  $w_k$  для каждой модели определяется как отношение её F-метрики к сумме всех F-метрик.

3. Подставленные веса используются в общей формуле  $I_{общ}$ .

Интерпретация:

- Достоверность модели  $\chi$ -квадрат ( $F_{\beta,\chi^2}$ ) показывает, насколько статистическое отклонение данных точно оценивается.

- Мера Харкевича ( $F_{\beta,H}$ ) отражает семантическую релевантность информации.

- Коэффициент ROI ( $F_{\beta,ROI}$ ) учитывает экономическую эффективность.

Таким образом, использование F-метрики Ван Ризбергера позволяет объективно нормировать веса критериев на основе их качества в задачах анализа данных.

### 3.3.4. Конфигуратор системно-когнитивных моделей АСК-анализа и интеллектуальной системы «Эйдос»

#### Конфигуратор системно-когнитивных моделей АСК-анализа и интеллектуальной системы «Эйдос»

	Способ сравнения	Нормировка не требуется	Нормировка к 0 путем взятия логарифма	Нормировка к 0 путем вычитания 1
Сравнение фактических и теоретических абсолютных частот	Путем деления	---	INF1, INF2, Александра Харкевича	INF4, INF5, Коэффициент возврата инвестиций ROI
	Путем вычитания	INF3, $\chi$ -квадрат Карла Пирсона	---	---
Сравнение условных и безусловных относительных частот	Путем деления	---	INF1, INF2, Александра Харкевича	INF4, INF5, Коэффициент возврата инвестиций ROI
	Путем вычитания	INF6, INF7	---	---

Отметим, что как значимость значения фактора, степень детерминированности класса и ценность или качество модели в АСК-анализе рассматривается вариабельность значений частных критериев этого значения фактора, класса или модели в целом (см. таблицу ниже).

**Уточнение терминологии АСК-анализа**

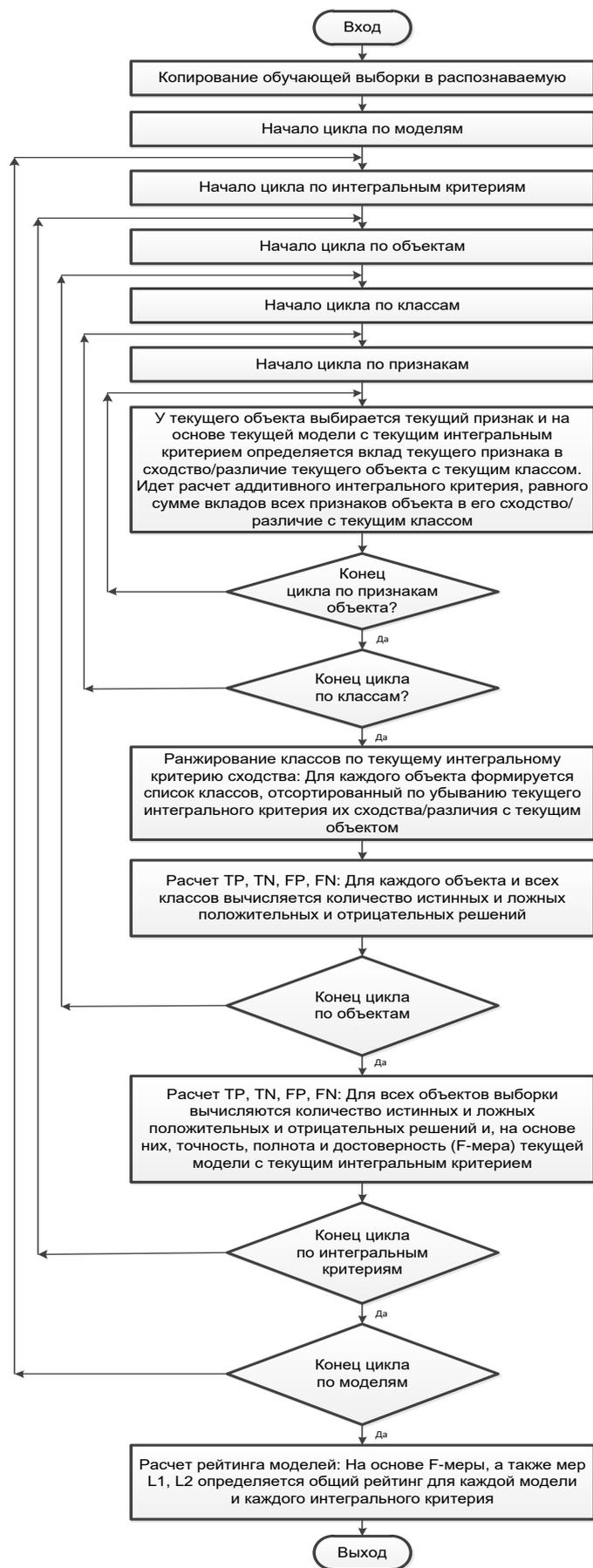
№	Традиционные термины (синонимы)	Новый термин	Формула
1	1. Значимость значения фактора (признака). 2. Дифференцирующая мощность значения фактора (признака). 3. Ценность значения фактора (признака) для решения задачи идентификации и других задач	Корень из информационной мощности значения фактора	$\sigma_{i\Sigma} = \sqrt[2]{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
2	1. Степень детерминированности класса. 2. Степень обусловленности класса.	Корень из информационной мощности класса	$\sigma_{\Sigma j} = \sqrt[2]{\frac{1}{M-1} \sum_{i=1}^M (I_{ij} - \bar{I}_j)^2}$
3	1. Качество модели. 2. Ценность модели. 3. Степень сформированности модели. 4. Количественная мера степени выраженности закономерностей в моделируемой предметной области	Корень из информационной мощности модели	$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

**3.4. Задача-4. Верификация моделей**

**3.4.1. Алгоритм оценки достоверностей моделей путем решения задачи распознавания объектов обучающей выборки с разными интегральными критериями**

Алгоритм оценки достоверностей моделей путем идентификации объектов обучающей выборки приведен на рисунке ниже:

1. Начало алгоритма.
2. Копирование обучающей (тренировочной) выборки в распознаваемую (тестовую).
3. Начало цикла по моделям.
4. Выбирается текущая модель.
5. Начало цикла по интегральным критериям.
6. Выбирается текущий интегральный критерий.
7. Начало цикла по объектам.
8. Выбирается текущий объект обучающей выборки.
9. Начало цикла по классам.
10. Выбирается текущий класс.
11. Начало цикла по признакам объекта.
12. У текущего объекта выбирается текущий признак и на основе текущей модели с текущим интегральным критерием определяется вклад текущего признака в сходство/различие текущего объекта с текущим классом. Идет расчет аддитивного интегрального критерия, равного сумме вкладов всех признаков объекта в его сходство/различие с текущим классом.
13. Конец цикла по признакам объекта. Все признаки? Если нет, то переход на шаг 12, иначе на шаг 14.
14. Конец цикла по классам. Все классы? Если нет, то переход на шаг 10, иначе на шаг 15.



15. Ранжирование классов по текущему интегральному критерию сходства: Для каждого объекта формируется список классов, отсортированный по убыванию текущего интегрального критерия их сходства/различия с текущим объектом.

16. Расчет TP, TN, FP, FN: Для каждого объекта и всех классов вычисляется количество истинных и ложных положительных и отрицательных решений.

17. Конец цикла по объектам. Все объекты? Если нет, то переход на шаг 8, иначе на шаг 18.

18. Расчет TP, TN, FP, FN: Для всех объектов выборки вычисляются количество истинных и ложных положительных и отрицательных решений и, на основе них, точность, полнота и достоверность (F-мера) текущей модели с текущим интегральным критерием.

19. Конец цикла по интегральным критериям. Все интегральные критерии? Если нет, то переход на шаг 6, иначе на шаг 20.

20. Конец цикла по моделям. Все модели? Если нет, то переход на шаг 4, иначе на шаг 21.

21. Расчет рейтинга моделей: На основе F-меры, а также мер L1, L2 определяется общий рейтинг для каждой модели и каждого интегрального критерия.

22. Конец алгоритма.

### **3.4.2. Кванторно-матричное математическое описание алгоритма оценки достоверности моделей для классической F-меры Ван Ризбергена**

Для описания предложенного алгоритма оценки достоверностей моделей в кванторно-матричном формализме, структурируем все циклы и операции с использованием кванторов общности ( $\forall$ ) и существования ( $\exists$ ), а также ввести обозначения для матриц данных и выполняемых операций.

#### **3.4.2.1. Условные обозначения**

1. X матрица признаков объектов, где элемент  $x_{ij}$  значение j-го признака i-го объекта.

2. Y вектор истинных меток классов для объектов обучающей выборки.

3. M множество моделей, где каждая модель  $m \in M$ .

4. K множество интегральных критериев, где каждый критерий  $k \in K$ .

5. C множество классов, где  $c \in C$ .

6. TP, TN, FP, FN метрики качества классификации (истинно положительные, истинно отрицательные, ложно положительные, ложно отрицательные результаты).

7. F-мера мера оценки качества модели, вычисляемая на основе TP, TN, FP, FN.

- 8.  $R(m, k)$  рейтинг модели  $m$  с интегральным критерием  $k$ .
- 9.  $S(i, c, m, k)$  интегральный критерий сходства объекта  $i$  с классом  $c$ , вычисляемый моделью  $m$  с критерием  $k$ .
- 10.  $P(i, j, c, m, k)$  вклад  $j$ -го признака объекта  $i$  в сходство с классом  $c$ , определяемый моделью  $m$  с критерием  $k$ .

### 3.4.2.2. Математическая формализация

1. Копирование выборки:

$$X_{\text{test}} \leftarrow X_{\text{train}}, \quad Y_{\text{test}} \leftarrow Y_{\text{train}}$$

2. Циклы по моделям, критериям, объектам, классам и признакам:

$$\forall m \in M, \forall k \in K, \forall i \in \{1, \dots, N\}, \forall c \in C, \forall j \in \{1, \dots, D\}, \exists P(i, j, c, m, k):$$

- Вклад текущего признака:

$$P(i, j, c, m, k) = f(x_{ij}, c, m, k)$$

- Суммирование вкладов признаков (аддитивный интегральный критерий):

$$S(i, c, m, k) = \sum_{j=1}^D P(i, j, c, m, k)$$

3. Ранжирование классов для объекта  $i$ :

$$\forall i \in \{1, \dots, N\}, \exists C_{\text{ranked}}(i, m, k):$$

$$C_{\text{ranked}}(i, m, k) = \text{SortDescending}(S(i, c, m, k) \mid c \in C)$$

4. Вычисление TP, TN, FP, FN:

Для каждого объекта и класса:

Пусть:

- $Y_i$  истинная метка класса объекта  $i$ .
- $c_{\text{predicted}}(i, m, k)$  предсказанный класс объекта  $i$  моделью  $m$  с критерием  $k$ .
- $\delta(\text{condition})$  индикаторное выражение, принимающее значение 1, если условие выполняется, и 0, если нет.
- $N$  общее количество объектов в выборке.
- $C$  множество всех классов.

Формула для TP (истинно положительные):

$$\forall m \in M, \forall k \in K, \exists TP(m, k):$$

$$TP(m, k) = \sum_{i=1}^N \delta(Y_i = c_{\text{predicted}}(i, m, k) \wedge Y_i = c_{\text{true}})$$

Формула для TN (истинно отрицательные):

$$\forall m \in M, \forall k \in K, \exists TN(m, k):$$

$$TN(m, k) = \sum_{i=1}^N \sum_{\substack{c \in C \\ c \neq Y_i}} \delta(c_{\text{predicted}}(i, m, k) \neq c \wedge Y_i \neq c)$$

Формула для FP (ложно положительные):

$$\forall m \in M, \forall k \in K, \exists FP(m, k):$$

$$FP(m, k) = \sum_{i=1}^N \delta(c_{\text{predicted}}(i, m, k) = c \wedge Y_i \neq c)$$

Формула для FN (ложно отрицательные):

$$\forall m \in M, \forall k \in K, \exists FN(m, k):$$

$$FN(m, k) = \sum_{i=1}^N \delta(c_{\text{predicted}}(i, m, k) \neq c \wedge Y_i = c)$$

Таким образом, данные формулы позволяют детализировать все аспекты работы алгоритма классификации, включая корректные и ошибочные предсказания.

### 3.4.3. Завершение алгоритма: рейтинг моделей

Расчет F-меры, а также мер L1 и L2, которые будут описаны ниже, производится по одним и тем же формулам:

$$\text{Precision}(m, k) = \frac{TP(m, k)}{TP(m, k) + FP(m, k)},$$

$$\text{Recall}(m, k) = \frac{TP(m, k)}{TP(m, k) + FN(m, k)}$$

$$F(m, k), L1(m, k), L2(m, k) = \frac{2 \cdot \text{Precision}(m, k) \cdot \text{Recall}(m, k)}{\text{Precision}(m, k) + \text{Recall}(m, k)}$$

Расчет средневзвешенного рейтинга моделей:

$$R(m, k) = g(F(m, k), L1(m, k), L2(m, k))$$

где L1, L2 предложенные автором дополнительные метрики, являющиеся обобщениями классической F-меры, оценивающие достоверность модели (будут подробнее описаны ниже).

Результаты формализуются как:

$$\text{Output: } \{R(m, k) \mid m \in M, k \in K\}$$

Эти результаты могут быть представлены в виде таблицы, в которой строки соответствуют различным моделям и интегральным критериям, а колонки отражают F-меру, а также ее компоненты: TP, TN, FP, FN, Precision и Recall (режим 3.4 системы «Эйдос»).

В завершение алгоритма оценки достоверности моделей, полученные значения мер для каждой модели и каждого интегрального критерия сравниваются, и модели ранжируются в порядке убывания их достоверности. Этот рейтинг предоставляет пользователю возможность выбрать наиболее достоверную модель для дальнейшего применения в задачах классификации, прогнозирования и других аналитических задачах.

### 3.4.4. Кванторное представление

Обобщенное представление алгоритма:

$$\forall m \in M, \forall k \in K, \forall i \in \{1, \dots, N\}, \forall c \in C, \forall j \in \{1, \dots, D\},$$

1. Вклад признака:

$$\exists P(i, j, c, m, k): P(i, j, c, m, k) = f(x_{ij}, c, m, k)$$

2. Сумма вкладов:

$$\exists S(i, c, m, k): S(i, c, m, k) = \sum_{j=1}^D P(i, j, c, m, k)$$

3. Ранжирование классов:

$$\exists C_{\text{ranked}}(i, m, k): C_{\text{ranked}}(i, m, k) = \text{SortDescending}(S(i, c, m, k) \mid c \in C)$$

4. Качество модели:

$$\exists R(m, k): R(m, k) = g(F(m, k), L1(m, k), L2(m, k))$$

Таким образом, весь алгоритм представлен через вложенные кванторы и матричные операции.

### 3.4.5. Кванторно-матричное математическое описание алгоритма оценки достоверности моделей для меры L1 – нечеткого мультиклассового обобщения классической F-меры Ван Ризбергена

#### 3.4.3.1. Необходимость и принцип нечеткого мультиклассового обобщения классической F-меры Ван Ризбергена

В АСК-анализе и системе "Эйдос" было разработано нечеткое мультиклассовое обобщение классической F-меры Ван Ризбергена [50]. Это связано с двумя основными причинами:

1. Многоклассовая принадлежность: В системе "Эйдос" каждый объект может принадлежать сразу нескольким классам одновременно.

2. Нечеткость оценки принадлежности: Используемая функция принадлежности (значение интегрального критерия)  $S(i, c, m, k)$ , определяющая степень сходства/различия объекта  $i$  с классом  $c$ , вычисляемая в модели  $m$  с интегральным критерием  $k$ , является нечеткой.

Значение  $S(i, c, m, k)$  интерпретируется следующим образом:

-  $S(i, c, m, k) > 0$ : мера уверенности системы в принадлежности объекта классу.

-  $S(i, c, m, k) < 0$ : мера уверенности системы в непринадлежности объекта классу.

-  $S(i, c, m, k) = 0$ : полная неопределённость.

Интегральный критерий  $S(i, c, m, k)$  является адекватной мерой степени истинности решения, т.к. чем он выше, тем больше доля истинных решений.

Для обеспечения адекватности описания меры достоверности моделей в системе "Эйдос" в классических формулах Ван Ризбергена вместо индукторной функции, принимающей значения 1 или 0,

используется  $S(i, c, m, k)$ . Это позволяет учитывать нечеткость при вычислении метрик TP, TN, FP, FN:

- TP (True Positive): если  $S(i, c, m, k) > 0$ , и объект действительно принадлежит классу, то к TP добавляется  $S(i, c, m, k)$ .

- TN (True Negative): если  $S(i, c, m, k) < 0$ , и объект действительно не принадлежит классу, то к TN добавляется  $S(i, c, m, k)$ .

- FP (False Positive): если  $S(i, c, m, k) > 0$ , но объект не принадлежит классу, то к FP добавляется  $S(i, c, m, k)$ .

- FN (False Negative): если  $S(i, c, m, k) < 0$ , но объект принадлежит классу, то к FN добавляется  $S(i, c, m, k)$ .

### 3.4.3.2. Математическая формализация

Для полного математического описания алгоритма оценки достоверности моделей можно использовать кванторное представление. Это позволяет формализовать условия принадлежности и не принадлежности объектов к классам в терминах логических кванторов.

Рассмотрим следующие выражения для каждой модели  $m$  и каждого интегрального критерия  $k$ :

- True Positive (TP):

$$TP(m, k) = \sum_{i \in P_c} S(i, c, m, k)$$

где  $P_c$  множество объектов, действительно принадлежащих классу  $c$ , и  $S(i, c, m, k) > 0$ .

- True Negative (TN):

$$TN(m, k) = \sum_{i \in N_c} S(i, c, m, k)$$

где  $N_c$  множество объектов, действительно не принадлежащих классу  $c$ , и  $S(i, c, m, k) < 0$ .

- False Positive (FP):

$$FP(m, k) = \sum_{i \in N_c} S(i, c, m, k)$$

где  $S(i, c, m, k) > 0$ , но объект  $i$  не принадлежит классу  $c$ .

- False Negative (FN):

$$FN(m, k) = \sum_{i \in P_c} S(i, c, m, k)$$

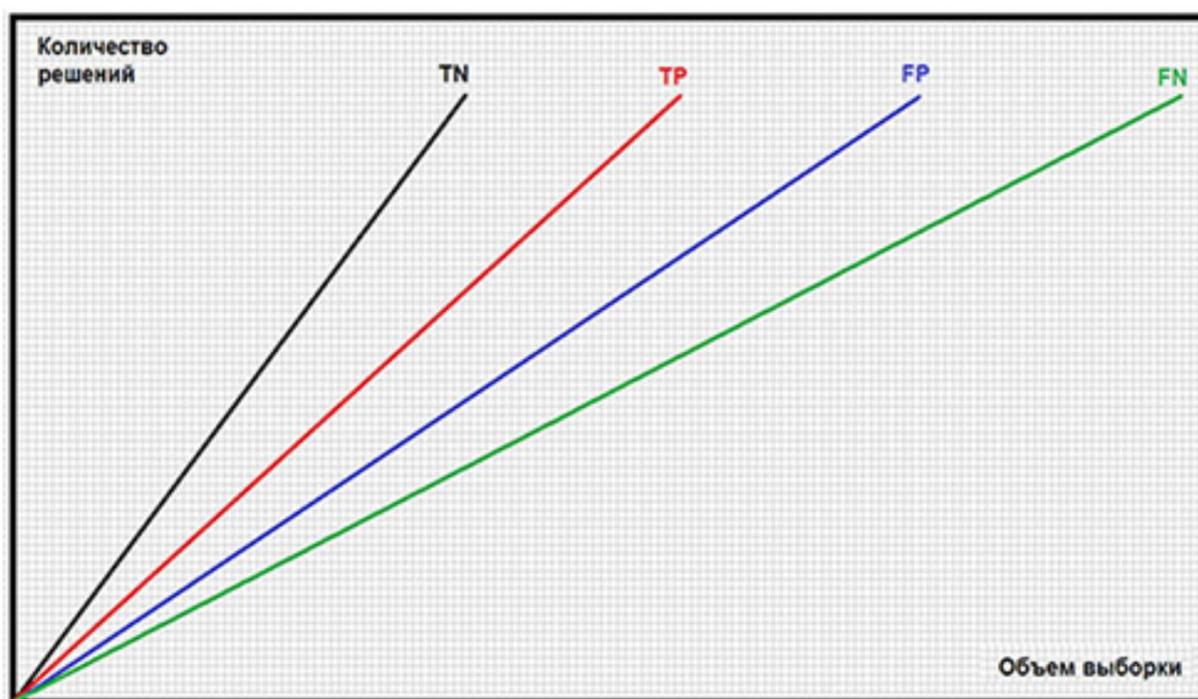
где  $S(i, c, m, k) < 0$ , но объект  $i$  принадлежит классу  $c$ .

Эти выражения позволяют строго математически описать процесс оценки достоверности модели, учитывая нечеткость и мультиклассовость.

### 3.4.6. Кванторно-матричное математическое описание алгоритма оценки достоверности моделей для меры L2 – обобщения меры L1, инвариантного относительно объемов данных

#### 3.4.4.1. Необходимость и принцип обобщения меры L1, инвариантного относительно объемов данных

В работе [9] автором предложено инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергера в АСК-анализе и системе "Эйдос". Необходимость разработки данного обобщения обусловлена тем, что и классическая F-мера Ван Ризбергера, и ее нечеткое мультиклассовое обобщение L1, предложенное проф.Е.В.Луценко [50], основаны на компонентах: TP, TN, FP, FN, которые являются сумматорами, линейно растущими при увеличении объема выборки:



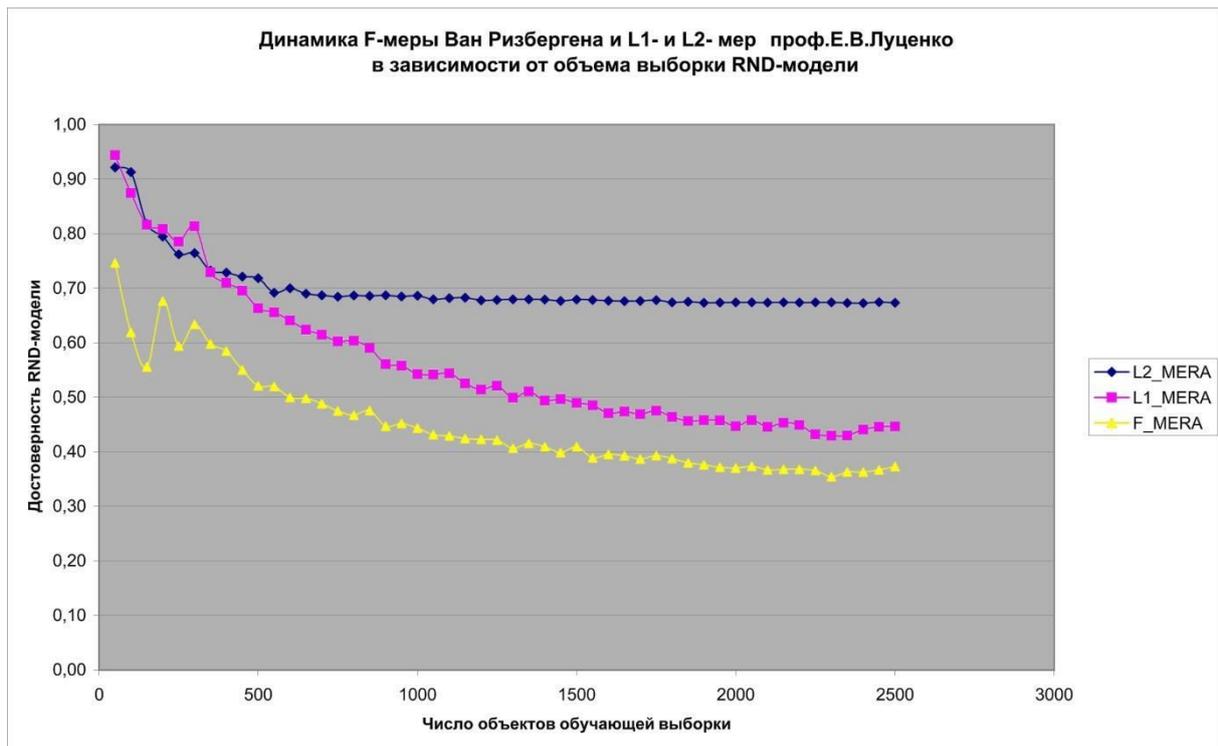
Естественно, возникает вопрос о том, как зависят Precision и Recall:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$$

а также сами мер F и L1 от объема выборки.

$$F(m, k), L1(m, k) = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Этот вопрос был исследован на численных моделях в работе [9] (рисунок 10 из работы [9]):



Оказалось, что классическая F-мера ведет себя крайне неустойчиво при малых выборках менее 500 наблюдений: ее поведение при таком числе наблюдений имеет вид затухающего колебания с начальной амплитудой в 20% достоверности. Фактически это означает, при малых выборках классической F-мерой Ван Ризбергена пользоваться вообще нельзя.

Предложенная мера L1, являющаяся нечетким мультиклассовым обобщением классической F-меры, в этом смысле ведет себя значительно более разумно и логично, изменяясь плавно и предсказуемо.

Но и классическая F-мера, и мера L1, являющаяся нечетким мультиклассовым обобщением классической F-меры, при увеличении объема выборки ведут себя очень сходно: плавно все медленнее и медленнее уменьшаясь и асимптотически стремясь к некому пределу при неограниченном увеличении объема выборки. Стабилизация их значений происходит при 3000 наблюдений и более. При этом значение достоверности модели уменьшается весьма значительно: для F-меры уменьшается от 75% до 35%, т.е. на 40%, а L1-меры от 95% до 45%, т.е. на 50%. Такие погрешности в определении достоверности моделей не могут быть признаны приемлемыми и допустимыми.

По этим причинам в работе [9] автором было предложено нечеткое мультиклассовое обобщение классической меры Ван Ризбергена, инвариантное относительно объемов данных.

**Основная идея** этого обобщения очень проста и основана на большой предельной теореме: раз компоненты: TP, TN, FP, FN, являются сумматорами, линейно зависящими от объема выборки, то надо перейти от абсолютных величин сумматоров к относительным величинам: частостям,

которые асимптотически стремятся к вероятностям наблюдения различных исходов идентификации (TP, TN, FP, FN) при неограниченном увеличении объема выборки.

### 3.4.4.2. Математическая формализация

Для построения инвариантной относительно объемов данных L2-меры достоверности моделей используется модификация формул оценки точности и полноты, в которых учитываются средние значения модулей уровней сходства для истинно и ложно классифицированных объектов. Обозначим:

-  $S(i, c, m, k)$  уровень сходства объекта  $i$  с классом  $c$ , рассчитанный в модели  $m$  по интегральному критерию  $k$ .

-  $P_c$  множество объектов, действительно принадлежащих классу  $c$ .

-  $N_c$  множество объектов, не принадлежащих классу  $c$ .

Вычисление компонент

1. Средний модуль сходства для истинно положительных решений (ATP):

$$ATP(m, k) = \frac{\sum_{i \in P_c, S(i, c, m, k) > 0} S(i, c, m, k)}{|P_c|}$$

2. Средний модуль сходства для истинно отрицательных решений (ATN):

$$ATN(m, k) = \frac{\sum_{i \in N_c, S(i, c, m, k) < 0} S(i, c, m, k)}{|N_c|}$$

3. Средний модуль сходства для ложно положительных решений (AFP):

$$AFP(m, k) = \frac{\sum_{i \in N_c, S(i, c, m, k) > 0} S(i, c, m, k)}{|N_c|}$$

4. Средний модуль сходства для ложно отрицательных решений (AFN):

$$AFN(m, k) = \frac{\sum_{i \in P_c, S(i, c, m, k) < 0} S(i, c, m, k)}{|P_c|}$$

Определение точности и полноты На основе введенных компонент определяются:

- A-точность модели (PA):

$$PA(m, k) = \frac{ATP(m, k)}{ATP(m, k) + AFP(m, k)}$$

- A-полнота модели (RA):

$$RA(m, k) = \frac{ATP(m, k)}{ATP(m, k) + AFN(m, k)}$$

Формула L2-меры L2-мера определяется аналогично F-мере и L1-мере, но с использованием A-точности и A-полноты:

$$L2(m, k) = \frac{2 \cdot PA(m, k) \cdot RA(m, k)}{PA(m, k) + RA(m, k)}$$

Эти формулы обеспечивают корректную оценку достоверности модели независимо от объема данных, так как при расчёте используются средние значения модулей уровней сходства, нормализованные относительно числа объектов в каждой категории.

### **3.5. Задача-5. Выбор наиболее достоверной модели**

Дополнительные пояснения даны в работе [1].

### **3.6. Задача-6. Системная идентификация и прогнозирование**

Дополнительные пояснения даны в работе [1].

#### **3.6.1. Интегральный критерий «Сумма знаний»**

**Интегральный критерий «Сумма знаний»** представляет собой суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний, представленных в help режима 5.5:

$$I_j = (\vec{I}_{ij}, \vec{L}_i).$$

В выражении круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид свертки:

$$I_j = \sum_{i=1}^M I_{ij} L_i,$$

где:  $M$  – количество градаций описательных шкал (признаков);

$\vec{I}_{ij} = \{I_{ij}\}$  – вектор состояния  $j$ -го класса;

$\vec{L}_i = \{L_i\}$  – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где } n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или  $n$ , если он присутствует у объекта с интенсивностью  $n$ ,

т.е. представлен  $n$  раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

### 3.6.2. Интегральный критерий «Семантический резонанс знаний»

Интегральный критерий «Семантический резонанс знаний» представляет собой *нормированное* суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний, представленных в help режима 3.3 и имеет вид:

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^M (I_{ij} - \bar{I}_j) (L_i - \bar{L}),$$

где:

$M$  – количество градаций описательных шкал (признаков);  $\bar{I}_j$  – средняя информативность по вектору класса;  $\bar{L}$  – среднее по вектору объекта;

$\sigma_j$  – среднеквадратичное отклонение частных критериев знаний вектора класса;  $\sigma_l$  – среднеквадратичное отклонение по вектору распознаваемого объекта.

$\bar{I}_{ij} = \{I_{ij}\}$  – вектор состояния  $j$ -го класса;  $\bar{L}_i = \{L_i\}$  – вектор состояния распознаваемого объекта (состояния или явления), включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив-локатор), т.е.:

$$\bar{L}_i = \begin{cases} 1, & \text{если } i - \text{й фактор действует;} \\ n, & \text{где } n > 0, \text{ если } i - \text{й фактор действует с истинностью } n; \\ 0, & \text{если } i - \text{й фактор не действует.} \end{cases}$$

В текущей версии системы «Эйдос-Х++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или  $n$ , если он присутствует у объекта с интенсивностью  $n$ , т.е. представлен  $n$  раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

Приведенное выражение для интегрального критерия «Семантический резонанс знаний» получается непосредственно из выражения для критерия «Сумма знаний» после замены координат перемножаемых векторов их стандартизированными значениями:

$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}$ ,  $L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_l}$ . Поэтому по своей сути он также является скалярным произведением (сверткой) двух стандартизированных (единичных) векторов класса и объекта.

Существуют и много других способов нормирования, например, путем применения вейвлетов и сплайнов, в частности линейной интерполяции:  $I_{ij} \rightarrow \frac{I_{ij} - I_j^{\min}}{I_j^{\max} - I_j^{\min}}$ ,  $L_i \rightarrow \frac{L_i - L^{\min}}{L^{\max} - L^{\min}}$ . Это позволяет предложить неограниченное количество других видов интегральных критериев. Но результаты их применения едва ли они будут существенно отличаться от уже существующих, поэтому они не реализованы в системе «Эйдос».

Для формализации данного подхода можно записать нормирование в следующем виде:

1. Линейная интерполяция:

Пусть  $X = \{x_1, x_2, \dots, x_n\}$  набор исходных данных, тогда нормирование через линейную интерполяцию может быть выражено следующим образом:

$$\forall i \exists a_i, b_i (\hat{x}_i = a_i x_i + b_i), \text{ где } \hat{x}_i \in [0, 1].$$

2. Нормирование через вейвлет-преобразование:

Нормированные значения  $\hat{X} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n\}$  вычисляются на основе разложения через матрицу вейвлетов  $W$ :

$$\exists W \forall X (\hat{X} = WX).$$

3. Сплайн-нормирование:

При использовании сплайнов нормирование можно представить как задачу нахождения параметров функции  $S(x)$ , удовлетворяющей граничным условиям:

$$\forall i \exists S(x) (S(x_i) = \hat{x}_i \wedge S'(x) \text{ гладкая}).$$

Несмотря на возможность применения вышеуказанных методов, их реализация в системе «Эйдос» не представлена, так как они не дают существенного выигрыша по сравнению с уже существующими методами.

### 3.6.3. Важные математические свойства интегральных критериев

Данные интегральные критерии обладают очень интересными *математическими свойствами*, которые обеспечивают им важные достоинства:

#### 3.6.3.1. Неметрическая природа

*Во-первых*, интегральный критерий имеет **неметрическую** природу, т.е. он является мерой сходства векторов класса и объекта, но не расстоянием между ними, а косинусом угла между ними, т.е. это межвекторное или информационное расстояние. Поэтому его применение

является корректным в **неортономмированных** пространствах, которые, как правило, и встречаются на практике и в которых применение Евклидова расстояния (теоремы Пифагора) является некорректным.

Интегральный критерий обладает неметрической природой, т.е. он является мерой сходства векторов класса и объекта, а не расстоянием между ними. Математически это выражается следующим образом:

$$\forall \vec{x}, \vec{y} \exists \text{Sim}(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|},$$

где  $\text{Sim}(\vec{x}, \vec{y})$  – косинус угла между векторами  $\vec{x}$  и  $\vec{y}$ , характеризующий их сходство, а не расстояние между ними. Это делает применение критерия корректным в неортономмированных пространствах, где Евклидово расстояние ( $d(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|$ ) использовать некорректно.

### 3.6.3.2. Фильтрация белого шума

*Во-вторых*, данный интегральный критерий является **фильтром**, подавляющим белый шум, который всегда присутствует в эмпирических исходных данных и в моделях, созданных на их основе. Это свойство подавлять белый шум проявляется у данного критерия тем ярче, чем больше в модели градаций описательных шкал.

Интегральный критерий является фильтром, подавляющим белый шум  $\epsilon$ , присутствующий в эмпирических данных  $\vec{x} = \vec{x}_{\text{ист}} + \vec{\epsilon}$ . Это свойство усиливается с увеличением числа классов, градаций описательных шкал и объема выборки.

Формально:

$$\exists \vec{w} \forall \vec{x}, \text{Sim}(\vec{x}, \vec{w}) = \max(\text{Sim}(\vec{x}_{\text{ист}}, \vec{w})), \text{ где } \vec{x}_{\text{ист}} = \vec{x} - \vec{\epsilon}.$$

### 3.6.3.3. Связь с нечеткой логикой

*В-третьих*, интегральный критерий сходства представляет собой количественную меру сходства/различия конкретного объекта с обобщенным образом класса и имеет тот же смысл, что и **функция принадлежности** элемента множеству в нечеткой логике Лотфи Заде. **Однако** в нечеткой логике эта функция задается исследователем априорно путем выбора из нескольких возможных вариантов, а в АСК-анализе и его программном инструментарии – интеллектуальной системе «Эйдос» она рассчитывается в соответствии с хорошо обоснованной математической моделью непосредственно на основе эмпирических данных.

Интегральный критерий сходства можно интерпретировать как функцию принадлежности  $\mu(x)$  элемента  $x$  множеству  $C$ :

$$\forall x, \exists \mu(x) = \text{Sim}(\vec{x}, \vec{w}_C),$$

где  $\vec{w}_C$  – обобщенный образ класса  $C$ . В отличие от нечеткой логики, где  $\mu(x)$  задается априорно, в системе «Эйдос» она вычисляется на основе эмпирических данных.

### 3.6.3.4. Самооценка уверенности

В-четвертых, кроме того значение интегрального критерия сходства представляет собой адекватную самооценку **степени уверенности** системы в положительном или отрицательном решении о принадлежности/непринадлежности объекта к классу или **риска ошибки** при таком решении.

Значение интегрального критерия сходства  $\text{Sim}(\vec{x}, \vec{w})$  интерпретируется как мера уверенности системы в принятом решении. Для классификации объекта  $\vec{x}$ :

$$\text{Уверенность} = \text{Sim}(\vec{x}, \vec{w}_{C_{\max}}), \text{ где } C_{\max} = \text{argmax}_C \text{Sim}(\vec{x}, \vec{w}_C).$$

### 3.6.3.5. Расчет коэффициентов разложения

В-пятых, по сути, при распознавании происходит расчет коэффициентов  $I_j$  разложения функции объекта  $L_i$  в ряд по функциям классов  $I_{ij}$ , т.е. определяется **вес** каждого обобщенного образа класса в образе объекта, что подробнее описано в работах [2, 3, 4].

В процессе распознавания объект  $L_i$  представляется в виде разложения по функциям классов  $I_{ij}$ :

$$L_i = \sum_j I_j I_{ij},$$

где  $I_j$  – коэффициенты весов, характеризующие вклад обобщенного образа класса  $j$  в образ объекта  $L_i$ .

Режим идентификации и прогнозирования системы «Эйдос» реализуется в рамках режима 4.1.2. Более подробное описание математической модели представлено в работах [2, 3, 4].

## 3.7. Задача-7. Поддержка принятия решений

### 3.7.1. Упрощенный вариант принятия решений как обратная задача прогнозирования, позитивный и негативный информационные портреты классов, SWOT-анализ в кванторно-матричном представлении

Дополнительные пояснения даны в работе [1].

Задачи прогнозирования и принятия решений относятся друг к другу как прямая и обратная задачи:

– при прогнозировании по значениям факторов, действующих на объект моделирования, определяется, в какое будущее состояние он перейдет под их действием:

$$Y = f(X),$$

где

$X = [x_1, x_2, \dots, x_n]$  вектор значений факторов;

$Y = [y_1, y_2, \dots, y_m]$  вектор будущего состояния объекта;

$f(\cdot)$  функция прогнозирования, задаваемая моделью.

– при принятии решений, наоборот, по будущему целевому состоянию объекта моделирования определяются значения факторов, которые обуславливают его переход в это будущее целевое состояние:

$$X = f^{-1}(Y),$$

где  $f^{-1}(\cdot)$  обратная функция, определяющая необходимые значения факторов.

Таким образом, задача принятия решений является обратной по отношению к задаче прогнозирования. Однако это справедливо лишь в упрощенном случае: при использовании, например, SWOT-анализа (режим 4.4.8 системы «Эйдос») [10].

Рассмотрим SWOT-анализ в кванторно-матричном представлении.

Пусть объект моделирования описывается матрицей факторов  $X$  и их влиянием на переход в целевое состояние, представленным матрицей  $W$ , где:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, \quad W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mn} \end{bmatrix},$$

где  $x_{ij}$  значение  $j$ -го фактора для  $i$ -го объекта, а  $w_{ij}$  коэффициент влияния  $j$ -го фактора на переход объекта в состояние  $i$ .

Целевое состояние объекта моделирования определяется следующим образом:

$$Y = X \cdot W^T,$$

Где

$Y$  вектор результирующего состояния.

При проведении SWOT-анализа значения факторов делятся на: -  
Позитивные влияния (сильные стороны):

$$S = \{x_{ij} \mid w_{ij} > 0\}, \quad w_{ij} \in W,$$

где

-  $S$  множество позитивных факторов, которые способствуют переходу объекта моделирования и управления в целевое состояние.

- Негативные влияния (слабости):

$$W = \{x_{ij} \mid w_{ij} < 0\}, \quad w_{ij} \in W,$$

где  $W$  множество факторов, которые препятствуют переходу в целевое состояние.

На SWOT-диаграмме:

-  $S$  визуализируются красным цветом с толщиной линии, пропорциональной значению  $|w_{ij}|$ ;

-  $W$  синим цветом, с аналогичной пропорциональностью.

Важным математическим свойством является автоматическое формирование SWOT-диаграмм на основе статистических и системно-

когнитивных моделей. Алгоритм отображения формально записывается как:

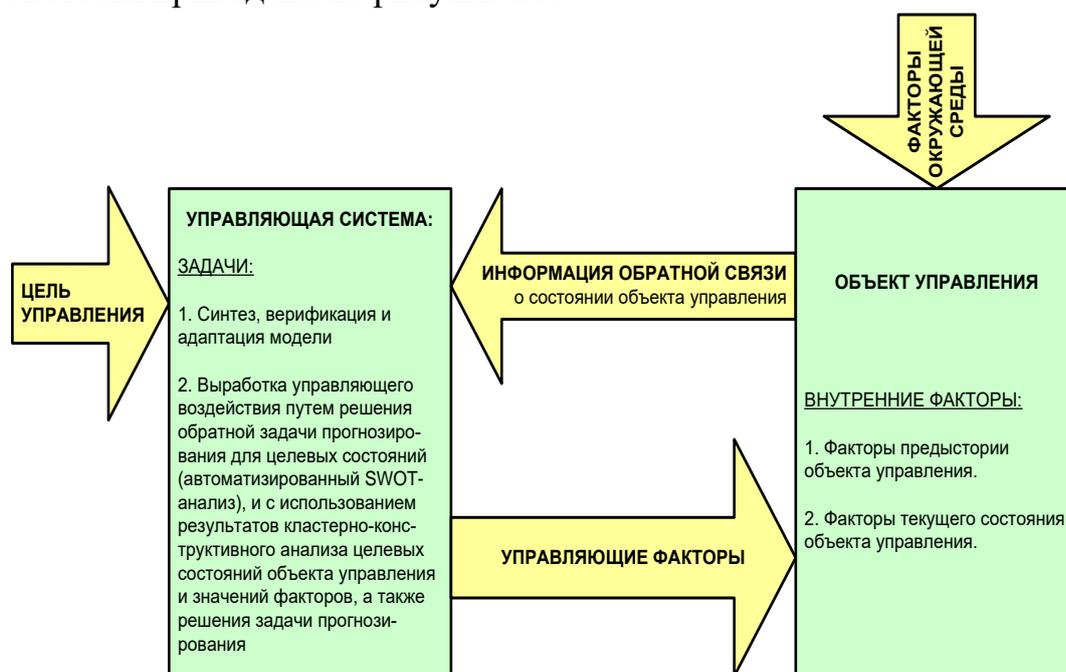
$$SWOT(X, W) = \begin{cases} \text{Red line} & \text{if } w_{ij} > 0, \\ \text{Blue line} & \text{if } w_{ij} < 0. \end{cases}$$

Таким образом, система «Эйдос» исключает субъективизм, используя строго математическое представление для формирования позитивных и негативных информационных портретов классов.

### 3.7.2. Развитый алгоритм принятия решений в адаптивных интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»

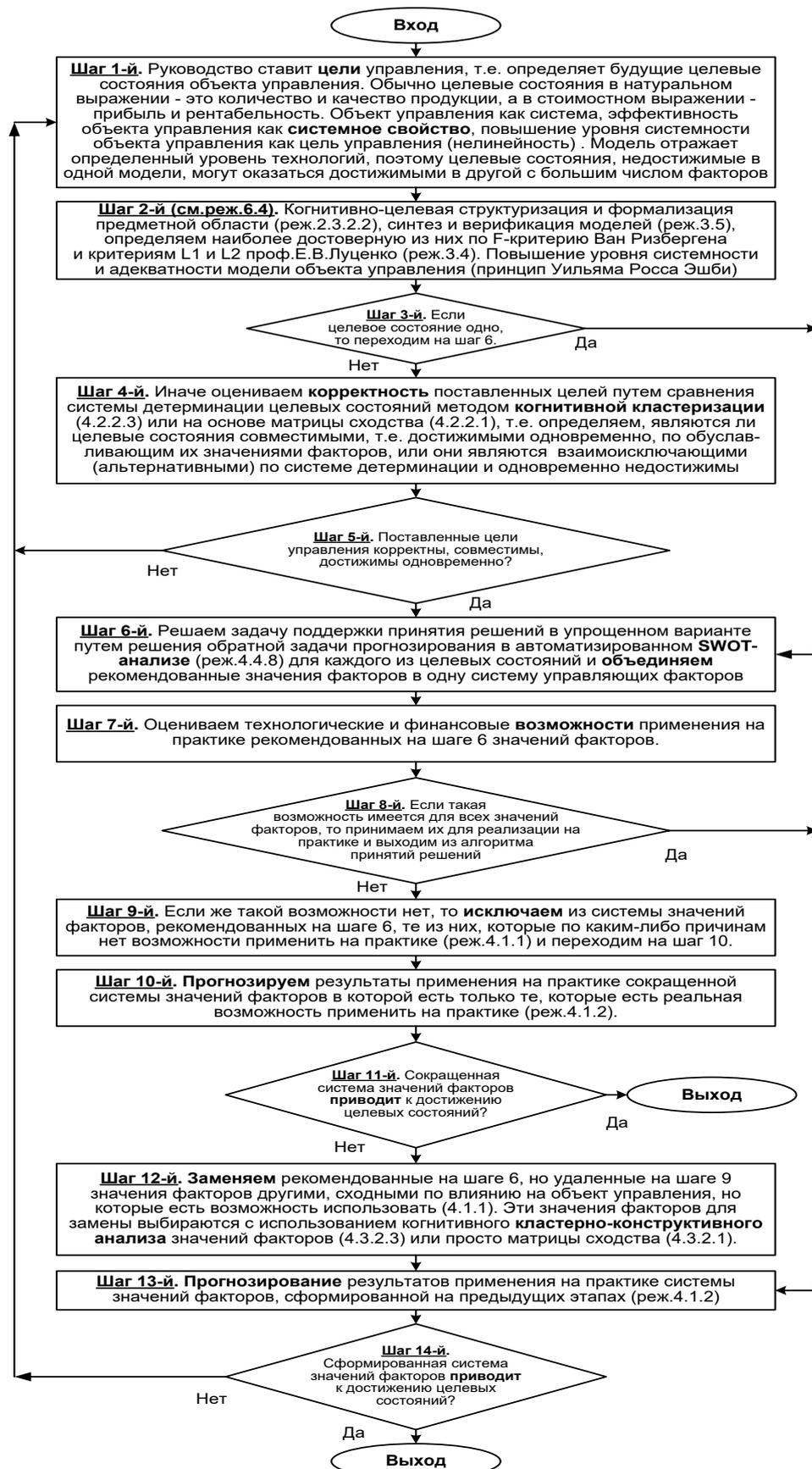
Дополнительные пояснения даны в работе [1].

Приведенный выше развитый алгоритм принятия решений может быть применен в адаптивных интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос». Принципиальная схема подобной системы приведена на рисунке 21:



1. Решения всегда принимаются на основе модели. Модели могут быть различной степени формализации: интуитивные (субъективные осознанные и неосознанные) неформализованные модели, вербализованные модели, лингвистические модели (различные структуры текста), алгоритмические модели и модели данных, статистические и информационные модели, математические (аналитические) модели. Формализация нужна чтобы передавать модели людям и техническим системам.  
 2. Виды управления: оперативное, тактическое, стратегическое. Что это значит в экономических и технических системах управления.  
 3. Различие между АСУ и САУ: участие человека в реальном времени в принятии решений. Кто несет ответственность за ошибочные решения. Адаптивность: принцип дуальности управления Александра Фельдбаума.  
 4. Критерий различия управляющих факторов от факторов окружающей среды с точки зрения управляющей системы и объекта управления. Иерархическая структура окружающей среды. Мы прогнозируем курс рубля на завтра, а ЦБ принимает решение об этом, для нас это фактор окружающей среды, а для ЦБ - это управляющий фактор.  
 5. Решение задачи принятия решений путем многократного многовариантного решения задачи прогнозирования быстро приводит к комбинаторному взрыву при увеличении количества факторов. Обычно в реальных задачах очень большое количество факторов. Поэтому при реальном количестве факторов задача принятия решений может быть решена только путем решения обратной задачи прогнозирования, т.е. SWOT-анализа. Однако в SWOT-анализе задается только одно целевое состояние и некоторые рекомендуемые значения факторов не могут быть применены по технологическим и финансовым причинам. Поэтому необходимо их исключить или заменить на основе результатов кластерно-конструктивного анализа значений факторов и оценить адекватность такого варианта решения путем прогнозирования результатов применения такой измененной системы значений факторов.

### Принципиальная схема адаптивной интеллектуальной системы управления на основе АСК-анализа и системы «Эйдос»



Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»

### **Математическое описание этапов алгоритма**

Приведенный алгоритм позволяет адаптивно управлять объектом на основе АСК-анализа и системы «Эйдос», учитывая технологические и финансовые ограничения, а также обеспечивая достижение целевых состояний.

**Шаг 1.** Постановка целей управления. Целевые состояния объекта управления задаются как множество  $T = \{t_1, t_2, \dots, t_m\}$ , где  $t_i$   $i$ -е целевое состояние. Цели могут быть выражены в натуральных или стоимостных показателях.

**Шаг 2.** Когнитивно-целевая структуризация и формализация предметной области. Формализация предметной области включает синтез моделей  $M = \{M_1, M_2, \dots, M_k\}$ , где каждая модель  $M_j$  описывает объект управления. Верификация моделей проводится на основе  $F$ -критерия Ван Ризбергена и критериев  $L1$  и  $L2$ :

$$F(M_j) = \frac{\text{Точность модели}}{\text{Сложность модели}}$$

Выбирается модель с максимальным значением  $F(M_j)$ .

**Шаг 3.** Проверка количества целевых состояний. Если  $|T| = 1$ , переход к шагу 6. Иначе к шагу 4.

**Шаг 4.** Оценка корректности поставленных целей. Корректность целей оценивается через матрицу сходства  $S$ , где  $S_{ij}$  косинусное расстояние между векторами факторов, обуславливающих целевые состояния  $t_i$  и  $t_j$ :

$$S_{ij} = \frac{f_i \cdot f_j}{\|f_i\| \cdot \|f_j\|}$$

Цели считаются совместимыми, если  $S_{ij} \geq \theta$ , где  $\theta$  пороговое значение.

**Шаг 5.** Проверка совместимости целей. Если все цели совместимы, переход к шагу 6. Иначе к шагу 1.

**Шаг 6.** Решение обратной задачи прогнозирования. Для каждого целевого состояния  $t_i$  решается обратная задача прогнозирования, результатом которой является набор факторов  $X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$ . Объединенный набор факторов:

$$X = \bigcup_{i=1}^m X_i$$

**Шаг 7.** Оценка технологических и финансовых возможностей. Для каждого фактора  $x_{ij} \in X$  оценивается возможность его применения на практике. Если все факторы применимы, переход к шагу 13. Иначе к шагу 9.

**Шаг 8.** Применение факторов на практике. Если все факторы применимы, они принимаются для реализации, и осуществляется переход к шагу 13.

**Шаг 9.** Исключение неприменимых факторов. Из набора  $X$  исключаются факторы, которые невозможно применить на практике, формируя сокращенный набор  $X'$ .

**Шаг 10.** Прогнозирование результатов применения сокращенного набора факторов. Для сокращенного набора  $X'$  прогнозируются результаты  $Y' = f(X')$ .

**Шаг 11.** Проверка достижимости целевых состояний. Если  $Y'$  соответствует целевым состояниям  $T$ , алгоритм завершается. Иначе переход к шагу 12.

**Шаг 12.** Замена неприменимых факторов. Неприменимые факторы заменяются сходными по влиянию на объект управления. Новый набор факторов  $X''$  формируется на основе матрицы сходства:

$$X'' = \{x''_{ij} \mid \operatorname{argmax}_{x''_{ij}} S(x_{ij}, x''_{ij})\}$$

**Шаг 13.** Прогнозирование результатов применения системы факторов. Для системы факторов  $X''$  прогнозируются результаты  $Y'' = f(X'')$ .

**Шаг 14.** Проверка достижимости целевых состояний. Если  $Y''$  соответствует целевым состояниям  $T$ , алгоритм завершается. Иначе переход к шагу 1.

### **3.8. Задача-8. Исследование объекта моделирования путем исследования его модели**

#### **3.8.1. Инвертированные SWOT-диаграммы значений описательных шкал (семантические потенциалы)**

Дополнительные пояснения даны в работе [1].

Для формализации семантических потенциалов можно использовать следующие математические представления:

##### **3.8.1.1. Матричное представление**

Для более удобной интерпретации и визуализации влияния значений факторов на состояния объекта моделирования используется матричное представление.

Пусть:

- $A = [a_{ij}]$  матрица семантических потенциалов,
- $i$  индекс значений фактора,
- $j$  индекс классов будущих состояний,
- $a_{ij}$  сила влияния  $i$ -го значения фактора на  $j$ -й класс.

Формула:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

где

$m$  количество градаций описательной шкалы,

$n$  количество классов классификационной шкалы.

Элементы матрицы  $a_{ij}$  вычисляются по формуле:

$$a_{ij} = g(x_i, y_j),$$

где  $g(x_i, y_j)$  мера семантического потенциала, вычисляемая в зависимости от модели (например, в системе «Эйдос» это статистические и системно-когнитивные модели: Abs, Prc1, Prc2, Inf1, Inf3, Inf4, Inf5, Inf6, Inf7, описанные в разделе 3.3 данной работы).

Таким образом, кванторное и матричное представления позволяют формализовать и детализировать интерпретацию семантических потенциалов, что обеспечивает их применение для аналитики и принятия решений.

Для удобства анализа могут вводиться нормированные значения матрицы. В приведенных обозначениях нормированные значения матрицы можно записать следующим образом:

$$a_{ij}^{\text{norm}} = \frac{a_{ij}}{\sum_{k=1}^m \sum_{l=1}^n a_{kl}},$$

где:

-  $a_{ij}^{\text{norm}}$  нормированное значение элемента матрицы семантических потенциалов,

-  $a_{ij}$  исходное значение элемента матрицы,

-  $m$  количество градаций описательной шкалы,

-  $n$  количество классов классификационной шкалы,

-  $\sum_{k=1}^m \sum_{l=1}^n a_{kl}$  сумма всех элементов матрицы  $A$ .

Нормированное значение  $a_{ij}^{\text{norm}}$  интерпретируется как доля или вес влияния конкретной пары  $(x_i, y_j)$  в общей семантической модели. Это позволяет легко сравнивать вклад отдельных значений факторов в классификацию различных будущих состояний.

Эти формулы делают анализ инвертированных SWOT-диаграмм более строгим, обеспечивая возможности для количественного и визуального анализа семантических потенциалов в контексте многомерных классификационных задач. ““

### 3.8.1.2. Кванторное представление

Для определения силы и направления влияния значения фактора можно использовать логико-математическое выражение с кванторами всеобщности и существования:

Формула:

$$\forall x \in X, \exists y \in Y: f(x, y) \geq \theta,$$

где:

- X множество значений фактора (градации описательных шкал),
- Y множество классов, соответствующих будущим состояниям объекта моделирования (градаций классификационных шкал),
- $f(x, y)$  функция силы и направления влияния значения фактора  $x$  на переход объекта моделирования в состояние  $y$ ,
- $\theta$  пороговое значение, модуль значения частого критерия, задающее минимальную силу влияния (по умолчанию это значение равно нулю).

### 3.8.2. Кластерно-конструктивный анализ классов

Дополнительные пояснения даны в работе [1].

#### 3.8.2.1. Матричное представление

В системе «Эйдос» (в режиме 4.2.2.1) рассчитывается матрица сходства классов по системе их детерминации. Пусть множества классов обозначим через  $C = C_1, C_2, \dots, C_n$ , а функция сходства между классами определяется как  $S(C_i, C_j)$ , где  $S$ : нормированная определенным образом матрица сходства:

$$S = \begin{bmatrix} S(C_1, C_1) & S(C_1, C_2) & \dots & S(C_1, C_n) \\ S(C_2, C_1) & S(C_2, C_2) & \dots & S(C_2, C_n) \\ \vdots & \vdots & \ddots & \vdots \\ S(C_n, C_1) & S(C_n, C_2) & \dots & S(C_n, C_n) \end{bmatrix}$$

На основе этой матрицы рассчитываются и выводятся три основные формы:

1. Круговая 2d-когнитивная диаграмма классов (режим 4.2.2.2): Классы размещаются на окружности, а взаимосвязи между ними отображаются в виде графа, где каждая связь весом  $S(C_i, C_j)$  указывает на степень сходства между классами  $C_i$  и  $C_j$ . Формально:

$$\forall i, j (1 \leq i, j \leq n) \exists \text{ ребро } (C_i, C_j), \text{ если } S(C_i, C_j) > 0.$$

2. Агломеративные дендрограммы (режим 4.2.2.3): Используется иерархическая кластеризация, где расстояние между классами определяется как:

$$d(C_i, C_j) = 1 - S(C_i, C_j), \quad d(C_i, C_i) = 0.$$

Алгоритм агломеративной кластеризации последовательно объединяет кластеры с минимальным расстоянием, формируя древовидную структуру.

3. График изменения межкластерных расстояний (режим 4.2.2.3): Для анализа когнитивных объединений классов строится график, где по оси абсцисс откладывается количество кластеров, а по оси ординат среднее или максимальное расстояние между объединяемыми кластерами на каждом этапе.

Матрица сходства также используется при расчете других выходных форм. Например, для построения когнитивной сети (графа), где вводятся

веса на ребрах, определяется суммарная степень связанности классов  $R(C_i)$  через формулу:

$$R(C_i) = \sum_{j=1}^n S(C_i, C_j).$$

### 3.8.2.2. Кванторное представление:

1. Формирование матрицы сходства (показан пример нормировки значений матрицы сходства:  $[0,1]$ ):

$$\forall C_i, C_j \in C, S(C_i, C_j) = f(C_i, C_j), f: C \times C \rightarrow [0,1],$$

где  $f$  функция сходства, основанная на когнитивной модели.

2. Построение когнитивного графа:

$$\forall C_i, C_j (C_i \neq C_j) \exists (C_i \rightarrow C_j), \text{ если } S(C_i, C_j) > \varepsilon,$$

где  $\varepsilon$ , минимальный порог для отображения связи (по умолчанию он равен нулю).

3. Иерархическая кластеризация:

$$\exists K \subseteq C, \forall C_i, C_j \in K, d(C_i, C_j) \leq d_{\max},$$

где  $K$  кластер, а  $d_{\max}$  максимально допустимое расстояние внутри кластера.

Эти математические выражения позволяют формализовать алгоритмы кластерно-конструктивного анализа классов в системе «Эйдос».

### 3.8.3. Кластерно-конструктивный анализ значений описательных шкал

Дополнительные пояснения даны в работе [1].

#### 3.8.3.1. Матричное представление

Матрица сходства  $S$  представляется в виде:

$$S = \begin{bmatrix} 1 & S_{12} & S_{13} & \dots & S_{1n} \\ S_{21} & 1 & S_{23} & \dots & S_{2n} \\ S_{31} & S_{32} & 1 & \dots & S_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & S_{n3} & \dots & 1 \end{bmatrix},$$

где элементы  $S_{ij}$  задают степень сходства между признаками  $f_i$  и  $f_j$ .

Для когнитивной кластеризации строится дендрограмма на основе преобразования матрицы сходства в расстояния:

$$D_{ij} = 1 - S_{ij}.$$

Полученная матрица расстояний  $D$  используется для алгоритма агломеративной кластеризации, визуализируемой в виде деревьев (дендрограмм).

Таким образом, кванторное и матричное представления формализуют процесс кластерно-конструктивного анализа, обеспечивая

строгое математическое описание расчета и визуализации когнитивных моделей признаков.

### 3.8.3.2. Кванторное представление

Процесс построения матрицы сходства можно описать следующим образом. Пусть множество признаков задано как  $F = \{f_1, f_2, \dots, f_n\}$ , а шкалы значений признаков как  $S = \{s_1, s_2, \dots, s_m\}$ . Тогда сходство между двумя признаками  $f_i$  и  $f_j$  определяется следующим выражением:

$$\forall i, j \in \{1, 2, \dots, n\}, \quad S_{ij} = \phi(f_i, f_j),$$

где  $\phi(f_i, f_j)$  функция семантического сходства.

Далее для каждой пары признаков проводится кластерный анализ, который можно выразить через кванторное условие:

$$\exists C_k \subseteq F: \forall f_i, f_j \in C_k, \quad S_{ij} \geq \theta,$$

где  $\theta$  пороговое значение сходства (по умолчанию оно равно нулю),  $C_k$  кластер признаков.

### 3.8.4. Модель знаний системы «Эйдос» и нелокальные нейроны

Дополнительные пояснения даны в работе [1].

#### 3.8.4.1. Матричное представление

В матричном виде нейросеть представляется весовой матрицей  $W$ :

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nm} \end{bmatrix}.$$

Активации классов вычисляются через матричное умножение:

$$A = R \cdot W^T,$$

где

$R = [r_1, r_2, \dots, r_m]$  вектор входных признаков,

$A = [A(c_1), A(c_2), \dots, A(c_n)]$  вектор активаций классов.

#### 3.8.4.2. Кванторное представление

Пусть задано множество классов  $C = \{c_1, c_2, \dots, c_n\}$  и множество рецепторов (признаков)  $R = \{r_1, r_2, \dots, r_m\}$ . Тогда активация нейрона  $c_i$  определяется выражением:

$$\forall i \in \{1, 2, \dots, n\}, \quad A(c_i) = f\left(\sum_{j=1}^m w_{ij} \cdot r_j\right),$$

где  $w_{ij}$  весовой коэффициент влияния рецептора  $r_j$  на нейрон  $c_i$ , а  $f$  функция активации.

Для полносвязной нейросети:

$$\forall i, j \in \{1, 2, \dots, n\}, \quad w_{ij} \neq 0.$$

Для заданного класса  $c_i$  содержательная интерпретация веса  $w_{ij}$  описывается через теорию информации:

$$w_{ij} = I(r_j; c_i),$$

где  $I(r_j; c_i)$  частный критерий модели, отражающий силу и направление связи между рецептором  $r_j$  и классом  $c_i$ . В системе «Эйдос» это статистические и системно-когнитивные модели: Abs, Prc1, Prc2, Inf1, Inf3, Inf4, Inf5, Inf6, Inf7, описанные в разделе 3.3 данной работы.

Таким образом, гибридная модель системы «Эйдос» формализуется через матричные операции и кванторные выражения, что обеспечивает строгую математическую основу для нечеткого декларативного представления знаний.

### **3.8.5. Гибридная модель системы «Эйдос» как полносвязная нелокальная нейронная сеть прямого счета**

#### **3.8.5.1. Формулы в матричном представлении**

В матричной форме модель системы «Эйдос» записывается как:

$$Z = \sigma(W \cdot R + B),$$

где:

- $Z$  вектор выходных значений нейронов слоя,
- $W$  матрица весов размерности  $m \times n$ , где  $m$  количество нейронов,  $n$  количество рецепторов,
- $R$  вектор входных значений рецепторов, -  $B$  вектор смещений размерности  $m$ ,
- $\sigma(x)$  векторная функция активации, применяемая покомпонентно.

Критерий детерминации нейронов.

Порядок расположения нейронов на визуализации слоя определяется модулем силы их детерминации:

$$|D_i| = \sum_{j=1}^n |w_{ij} \cdot r_j|,$$

где  $D_i$  детерминация нейрона  $i$ .

#### **3.8.5.2. Формулы в кванторном представлении**

Общая модель влияния рецепторов на классы описывается следующим образом:

$$\forall c_i \in C, \forall r_j \in R, \exists w_{ij} \in W: I(r_j; c_i) = w_{ij},$$

где:

- $C$  множество классов,
- $R$  множество рецепторов,
- $W$  множество весов, отражающих силу связи между рецепторами и классами.

Суммарная активация нейрона  $z_i$  слоя вычисляется как:

$$z_i = \sigma \left( \sum_{j=1}^n w_{ij} \cdot r_j + b_i \right),$$

где:

- $\sigma(x)$  функция активации,
- $b_i$  смещение (bias),
- $n$  количество входов (рецепторов).

Данный подход позволяет гибко визуализировать и анализировать влияние различных факторов на выходные значения нейронов, а также моделировать взаимодействия в многослойных полносвязных нейронных сетях.

### 3.8.6. 3D-интегральные когнитивные карты

3d-интегральная когнитивная карта является отображением на одном рисунке: Когнитивной диаграммы классов в верхней части рисунка. Линии связи в конативной диаграмме классов отображают цветом знак связи (красным - сходные, синим - отличающиеся), а толщиной – модуль силы связи между классами. Когнитивной диаграммы значений факторов (признаков) в нижней части рисунка. Линии связи в конативной диаграмме значений факторов отображают цветом знак связи (красным - сходные, синим - отличающиеся), а толщиной – модуль силы связи между значениями факторов. Соединяющего их одного слоя нейронной сети (режим 4.4.12 системы «Эйдос»), в котором отображаются связи между рецепторами нейронов в нижней части и нейронами (классами) в верхней части. Линии связи отображают цветом знак связи (красным - активирующая, синим - тормозящая), а толщиной – модуль силы связи.

3D-интегральная когнитивная карта является отображением на одном рисунке:

#### 3.8.6.1. Когнитивная диаграмма классов

Отображается в верхней части рисунка в виде круговой диаграммы, в вершинах которой находятся классы, соединенные линиями связи. Линии связи в когнитивной диаграмме классов отображают цветом знак связи:

- Красным сходные.
- Синим отличающиеся.
- Толщина линии отображает модуль силы связи между классами.

**Матричное представление связей:**

$$S = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{bmatrix},$$

где  $S_{ij}$  элемент матрицы, отображающий силу связи между  $C_i$  и  $C_j$ .

**Формальное кванторное описание:**

Пусть множество классов задано как  $C = \{c_1, c_2, \dots, c_n\}$ , тогда:

$$\forall i, j \in \{1, \dots, n\} (c_i \rightarrow c_j) \Rightarrow \begin{cases} \text{цвет линии: red,} & \text{если } s_{ij} > 0, \\ \text{цвет линии: blue,} & \text{если } s_{ij} < 0, \end{cases}$$

где  $s_{ij}$  сила связи между классами  $c_i$  и  $c_j$ .

**3.8.6.2. Когнитивная диаграмма значений факторов (признаков)**

Отображается в нижней части рисунка в виде круговой диаграммы, в вершинах которой находятся значения факторов, соединенные линиями связи. Линии связи в когнитивной диаграмме значений факторов отображают цветом знак связи:

- Красным сходные.
- Синим отличающиеся.
- Толщина линии отображает модуль силы связи между значениями факторов.

**Матричное представление связей:**

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mm} \end{bmatrix}.$$

**Формальное кванторное описание:**

Пусть множество факторов задано как  $F = \{f_1, f_2, \dots, f_m\}$ , тогда:

$$\forall i, j \in \{1, \dots, m\} (f_i \rightarrow f_j) \Rightarrow \begin{cases} \text{цвет линии: red,} & \text{если } w_{ij} > 0, \\ \text{цвет линии: blue,} & \text{если } w_{ij} < 0, \end{cases}$$

где  $w_{ij}$  сила связи между факторами  $f_i$  и  $f_j$ .

**3.8.6.3. Соединяющий вершины когнитивных диаграмм один слой нейронной сети**

Связи между рецепторами нейронов в нижней части и нейронами (классами) в верхней части. Линии связи отображают цветом знак связи:

- Красным активирующая.
- Синим тормозящая.
- Толщина линии отображает модуль силы связи.

**Матричное представление:**

$$V = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \dots & v_{pn} \end{bmatrix}.$$

**Формальное кванторное описание:**

Пусть  $R = \{r_1, r_2, \dots, r_p\}$  рецепторы, а  $C = \{c_1, c_2, \dots, c_n\}$  нейроны. Связь между ними задаётся матрицей весов  $V$ :

$$\forall i \in \{1, \dots, p\}, \forall j \in \{1, \dots, n\} (r_i \rightarrow c_j) \Rightarrow \begin{cases} \text{цвет линии: red,} & \text{если } v_{ij} > 0, \\ \text{цвет линии: blue,} & \text{если } v_{ij} < 0, \end{cases}$$

где  $v_{ij}$  вес связи между рецептором  $r_i$  и нейроном  $c_j$ .

Таким образом, 3D-интегральная когнитивная карта объединяет когнитивные диаграммы классов, значений факторов и связи между ними через слой нейронной сети, что позволяет визуализировать и анализировать сложные взаимосвязи.

### 3.8.7. 2D-интегральные когнитивные карты содержательного сравнения классов (опосредованные нечеткие правдоподобные рассуждения)

Дополнительные пояснения даны в работе [1].

#### 3.8.7.1. Матричное представление

Пусть:

-  $A_1 = [aInfAtr1[1], \dots, aInfAtr1[i]]$  вектор информативностей признаков для 1-го класса.

-  $A_2 = [aInfAtr2[1], \dots, aInfAtr2[j]]$  вектор информативностей признаков для 2-го класса.

-  $S = [s_{ij}]$  матрица сходства признаков между классами.

1. Теоретически максимальная информативность:

$$mTeorMaxInf = \frac{\log(N_{Cls})}{\log(2)}$$

2. Информативность в процентах (векторы):

$$MInfPerTM1 = \frac{A_1}{mTeorMaxInf} \cdot 100,$$

$$MInfPerTM2 = \frac{A_2}{mTeorMaxInf} \cdot 100$$

3. Информативность в стандартизированных величинах:

$$MInfStand1 = \frac{A_1 - mSr1}{mDi1}, \quad MInfStand2 = \frac{A_2 - mSr2}{mDi2}$$

4. Сила связи (матрица бит):

$$MRelBit = S \circ (A_1 \cdot A_2^T),$$

где  $\circ$  поэлементное умножение.

5. Максимально возможная сила связи (скаляр):

$$mMaxRelBit = mTeorMaxInf^2$$

6. Сила связи в процентах (матрица):

$$MRelPercTM = \frac{MRelBit}{mMaxRelBit} \cdot 100$$

7. Сила связи в стандартизированных величинах:

$$MRelStand = \frac{S \circ (MInfStand1 \cdot MInfStand2^T)}{2}$$

### 3.8.7.2. Кванторное представление

Кванторное представление формул основывается на использовании кванторов существования ("∃") и всеобщности ("∀"). Формулы интерпретируются в терминах множеств, элементов и логических операций.

1. Теоретически максимальная информативность:

$$\forall N_{Cls} \in \mathbb{N}, \exists m_{TeorMaxInf} \in \mathbb{R}: m_{TeorMaxInf} = \frac{\log(N_{Cls})}{\log(2)}$$

2. Информативность признаков в процентах от теоретического максимума:

$$\forall i, \exists m_{InfPerTM1} \in \mathbb{R}: m_{InfPerTM1} = \frac{a_{InfAtr1}[i]}{m_{TeorMaxInf}} \cdot 100$$

$$\forall j, \exists m_{InfPerTM2} \in \mathbb{R}: m_{InfPerTM2} = \frac{a_{InfAtr2}[j]}{m_{TeorMaxInf}} \cdot 100$$

3. Информативность признаков в стандартизированных величинах:

$$\forall i, \exists m_{InfStand1} \in \mathbb{R}: m_{InfStand1} = \frac{a_{InfAtr1}[i] - m_{Sr1}}{m_{Di1}}$$

$$\forall j, \exists m_{InfStand2} \in \mathbb{R}: m_{InfStand2} = \frac{a_{InfAtr2}[j] - m_{Sr2}}{m_{Di2}}$$

4. Сила связи (в битах):

$$\forall i, j, \exists m_{RelBit} \in \mathbb{R}: m_{RelBit} = m_{K_{12}} \cdot a_{InfAtr1}[i] \cdot a_{InfAtr2}[j]$$

5. Максимально возможная сила связи (в битах):

$$\exists m_{MaxRelBit} \in \mathbb{R}: m_{MaxRelBit} = m_{TeorMaxInf}^2$$

6. Сила связи в процентах от теоретического максимума:

$$\forall i, j, \exists m_{RelPercTM} \in \mathbb{R}: m_{RelPercTM} = \frac{m_{RelBit}}{m_{MaxRelBit}} \cdot 100$$

7. Сила связи в стандартизированных величинах:

$$\forall i, j, \exists m_{RelStand} \in \mathbb{R}: m_{RelStand} = \frac{m_{K_{12}} \cdot m_{InfStand1} \cdot m_{InfStand2}}{2}$$

Таким образом, кванторное представление акцентирует логическую интерпретацию, а матричное вычислительную реализацию на основе линейной алгебры.

### 3.8.8. 2D-интегральные когнитивные карты содержательного сравнения значений факторов (опосредованные нечеткие правдоподобные рассуждения)

Дополнительные пояснения даны в работе [1].

#### 3.8.8.1. Матричное представление

Обозначим:

- A1 вектор информативности классов первого признака:

$$A1 = [a_{InfCls1}[i]].$$

- A2 вектор информативности классов второго признака:

$$A2 = [aInfCls2[j]].$$

- М матрица сходства классов:  $M = [mK12]$ .

- mTeorMaxInf скалярное значение:  $mTeorMaxInf = \frac{\log(N_{cls})}{\log(2)}$ .

1. Информативность в процентах (векторы):

$$InfPerTM1 = \frac{A1}{mTeorMaxInf} \cdot 100, \quad InfPerTM2 = \frac{A2}{mTeorMaxInf} \cdot 100$$

2. Стандартизированная информативность (векторы):

$$InfStand1 = \frac{A1 - mSr1}{mDi1}, \quad InfStand2 = \frac{A2 - mSr2}{mDi2}$$

3. Сила связи в битах (матрица):

$$RelBit = M \cdot (A1 \otimes A2),$$

где  $\otimes$  поэлементное умножение векторов.

4. Максимальная теоретическая сила связи (скаляр):

$$mMaxRelBit = mTeorMaxInf^2$$

5. Сила связи в процентах (матрица):

$$RelPercTM = \frac{RelBit}{mMaxRelBit} \cdot 100$$

6. Сила связи в стандартизированных величинах (матрица):

$$RelStand = \frac{M \cdot (InfStand1 \otimes InfStand2)}{2}$$

### 3.8.8.2. Кванторное представление

1. Теоретически максимальная информативность:

$$\forall i \forall j \ mTeorMaxInf = \frac{\log(N_{cls})}{\log(2)}$$

2. Информативность классов в процентах:

$$\forall i \ mInfPerTM1 = \frac{aInfCls1[i]}{mTeorMaxInf} \cdot 100$$

$$\forall j \ mInfPerTM2 = \frac{aInfCls2[j]}{mTeorMaxInf} \cdot 100$$

3. Стандартизированная информативность:

$$\forall i \ mInfStand1 = \frac{aInfCls1[i] - mSr1}{mDi1}$$

$$\forall j \ mInfStand2 = \frac{aInfCls2[j] - mSr2}{mDi2}$$

4. Сила связи в битах:

$$\forall i \forall j \ mRelBit = mK12 \cdot aInfCls1[i] \cdot aInfCls2[j]$$

5. Максимальная теоретическая сила связи:

$$mMaxRelBit = 1 \cdot mTeorMaxInf^2$$

6. Сила связи в процентах:

$$\forall i \forall j \ mRelPercTM = \frac{mRelBit}{mMaxRelBit} \cdot 100$$

7. Сила связи в стандартизированных величинах:

$$\forall i \forall j mRelStand = \frac{mK12 \cdot mInfStand1 \cdot mInfStand2}{2}$$

Данные формулы в кванторной форме показывают взаимосвязь через универсальные кванторы, а в матричной форме демонстрируют вычисления через линейную алгебру, что может быть полезно для реализации в вычислительных системах.

### 3.8.9. Когнитивные функции

Дополнительные пояснения даны в работе [1].

#### 3.8.9.1. Матричное представление

Когнитивная функция для объекта моделирования можно выразить в виде матрицы влияния:

$$K = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{m1} & k_{m2} & \dots & k_{mn} \end{bmatrix},$$

где  $k_{ij}$  степень влияния  $i$ -го значения фактора на  $j$ -е состояние объекта, вычисляемая в соответствии с моделью (в системе «Эйдос» это статистические и системно-когнитивные модели: Abs, Prc1, Prc2, Inf1, Inf3, Inf4, Inf5, Inf6, Inf7, описанные в разделе 3.3 данной работы).

#### 3.8.9.2. Кванторное представление

Когнитивную функцию можно представить как:

$$\forall x \in X, \exists y \in Y: K(x, y) = \text{вес}(x, y),$$

где

$X$  множество градаций описательной шкалы;

$Y$  множество градаций классификационной шкалы;

$K(x, y)$  функция, задающая степень влияния  $x$  на  $y$ .

Когнитивные функции являются одним из наиболее мощных и наглядных средств когнитивной графики, имеющих в системе «Эйдос». Они позволяют отобразить силу и направление влияния каждого значения каждого фактора на переход объекта моделирования в каждое из будущих состояний, соответствующих классам.

Пример обобщенного представления когнитивной функции: Если рассматривать когнитивную функцию как обобщение вероятностной функции распределения, то её можно записать:

$$\forall x \in X, \forall y \in Y: P(x \rightarrow y) = K(x, y),$$

Где

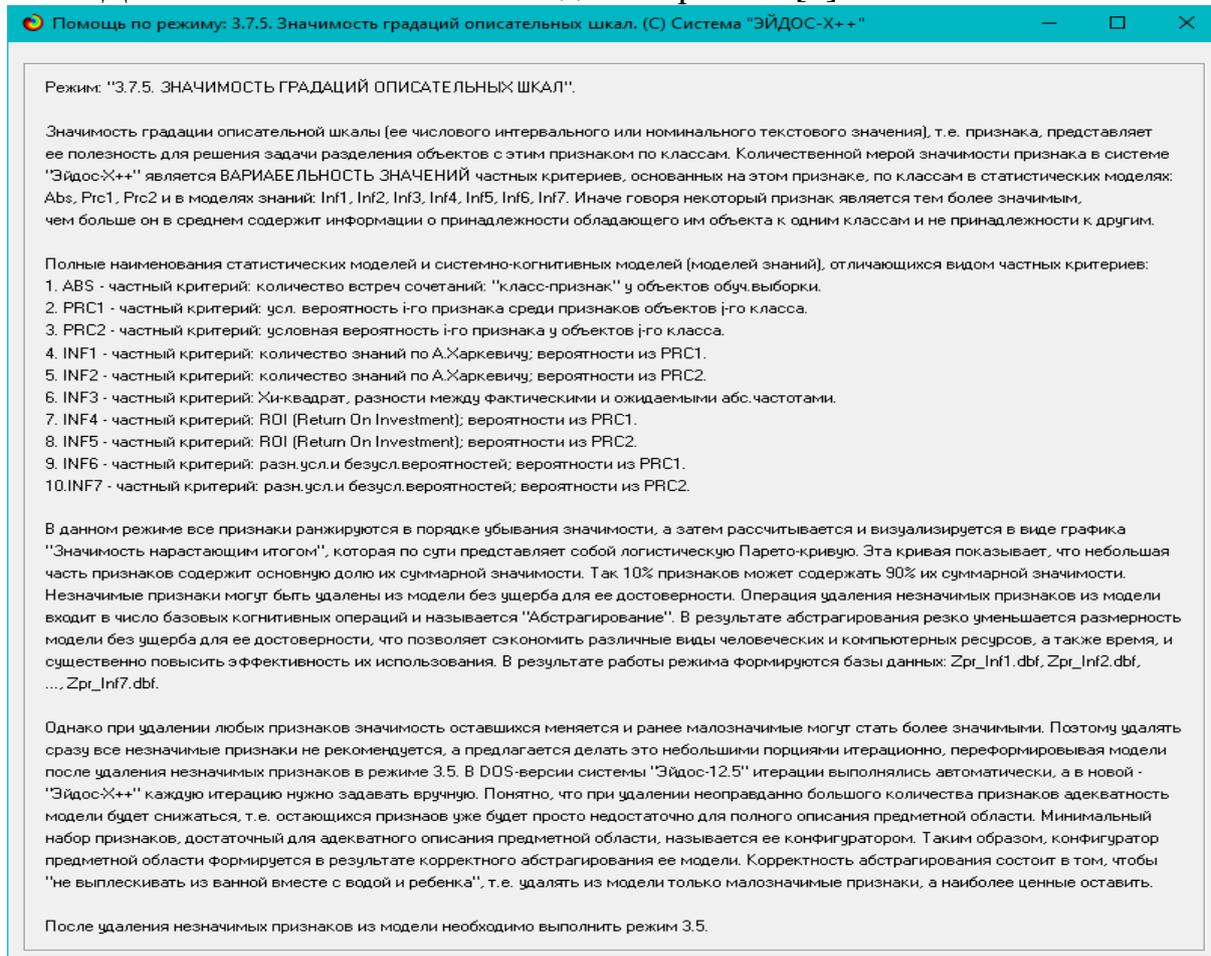
$P(x \rightarrow y)$  вероятность перехода объекта из состояния  $x$  в состояние  $y$ ;

$K(x, y)$  количественная мера этой вероятности.

Дополнительные пояснения даны в работе [1].

### 3.8.10. Значимость описательных шкал и их градаций

Дополнительные пояснения даны в работе [1].



Дополнительные пояснения даны в работе [1].

#### 3.8.10.1. Матричное представление значимости градаций

Значимость каждой градации  $g_j$  описательной шкалы  $S_i$  можно представить как сумму абсолютных значений её влияния на все классы:

$$G_i = \begin{bmatrix} g_{i1} \\ g_{i2} \\ \vdots \\ g_{im} \end{bmatrix}, \quad g_{ij} = \sum_{k=1}^n |k_{ij}|,$$

где

$g_{ij}$  значимость j-й градации i-й описательной шкалы,

$k_{ij}$  степень влияния градации на идентификацию классов.

Значимость всей шкалы в матричном виде является средним от значимости всех градаций этой описательной шкалы  $S_i$ :

$$\text{Значимость}(S_i) = \frac{1}{m} \sum_{j=1}^m g_{ij},$$

где  $m$  количество градаций в шкале  $S_i$ .

### 3.8.10.2. Кванторное представление значимости

Для всех градаций  $g_j$  описательной шкалы  $S$ :

$$\forall g_j \in S, \exists c_k \in C: \text{Значимость}(g_j) = \sum_{k=1}^n |I(g_j, c_k)|,$$

где

$I(g_j, c_k)$  количество информации, передаваемое градацией  $g_j$  о классе  $c_k$ ,

$C$  множество всех классов.

Парето-кривая селективной силы: Сумма селективной силы системы значений факторов нарастающим итогом:

$$P(i) = \sum_{j=1}^i \text{Значимость}(g_j), \quad i = 1, 2, \dots, m,$$

где  $P(i)$  накопленная селективная сила для первых  $i$  градаций, упорядоченных по убыванию значимости.

Описание режимов системы «Эйдос»:

- Режим 3.7.5 отображает значимость градаций шкалы в виде графиков, где каждый столбец соответствует  $g_{ij}$ .

- Режим 3.7.4 вычисляет среднюю значимость всей описательной шкалы как агрегированную метрику.

### 3.8.11. Степень детерминированности классов и классификационных шкал

Дополнительные пояснения даны в работе [1].

#### 3.8.11.1. Матричное представление

Степень детерминированности для конкретного класса  $C_j$  может быть определена на основе элементов столбца  $j$ -й матрицы модели:

$$D(C_j) = 1 - \frac{\sigma_j}{\sigma_{\max}},$$

где: -  $D(C_j)$  степень детерминированности класса  $C_j$ ; -  $\sigma_j$  дисперсия значений факторов в столбце, соответствующем классу  $C_j$ ; -  $\sigma_{\max}$  максимальная дисперсия по всем столбцам матрицы.

Для всей классификационной шкалы  $S$  степень детерминированности вычисляется как среднее арифметическое:

$$D(S) = \frac{1}{n} \sum_{j=1}^n D(C_j),$$

где  $n$  – количество классов в шкале.

### 3.8.11.2. Кванторное представление:

Степень детерминированности класса  $C_j$  может быть выражена следующим образом:

$$\exists C_j \in S: D(C_j) = 1 - \frac{\sigma_j}{\sigma_{\max}},$$

где  $S$  – множество всех классов шкалы.

Для всей классификационной шкалы степень детерминированности записывается:

$$\forall S \subseteq \mathcal{C}, \exists D(S): D(S) = \frac{1}{n} \sum_{j=1}^n D(C_j),$$

где  $\mathcal{C}$  – множество всех возможных классов.

Интерпретация:

- Если степень детерминированности  $D(C_j) \rightarrow 1$ , это означает, что класс  $C_j$  практически однозначно определяется значениями факторов.

- Если  $D(C_j) \rightarrow 0$ , это указывает на высокую вариабельность значений факторов, что затрудняет прогнозирование принадлежности идентифицируемого объекта к данному классу.

- Среднее значение  $D(S)$  для классификационной шкалы характеризует общую предсказуемость всех классов шкалы.

Таким образом, степень детерминированности класса позволяет количественно оценить, насколько хорошо модель в системе «Эйдос» способна прогнозировать состояния объекта, соответствующие каждому классу и в целом для классификационной шкалы.

## 4. ОБСУЖДЕНИЕ

Развитие автоматизированного системно-когнитивного анализа (АСК-анализа) и внедрение универсальной когнитивной аналитической системы «Эйдос-Х++» представляют собой значительный вклад в развитие технологий искусственного интеллекта (ИИ) в России и мире. Важность системы заключается в ее уникальных характеристиках, универсальности и практической применимости в различных областях.

### 4.1. Сравнение с другими системами

На фоне множества современных решений в области ИИ, таких как ML.NET, OpenAI, Neural University и другие, «Эйдос-Х++» выделяется своей универсальностью и ориентацией на широкий круг пользователей. В

отличие от многих специализированных платформ, эта система разработана в общем подходе, независимом от предметной области, что позволяет применять ее для решения задач в экономике, биологии, инженерии и других дисциплинах.

#### 4.2. Доступность и удобство

Одним из ключевых преимуществ «Эйдос-X++» является ее доступность. Открытая лицензия (CC BY-SA 4.0) обеспечивает свободный доступ к системе и ее исходным текстам, что способствует популяризации науки, а также адаптации и доработке системы пользователями под их конкретные нужды. Кроме того, система предлагает «нулевой порог входа», что делает ее подходящей для пользователей без опыта программирования и знаний в области ИИ.

У желающих есть все возможности для изучения АСК-анализа и системы «Эйдос» на своем компьютере и их применения для исследований и разработок интеллектуальных облачных Эйдос-приложений в различных предметных областях.

Для этого надо скачать систему с сайта разработчика по ссылке на странице: [http://lc.kubagro.ru/Installation\\_Eidos.php](http://lc.kubagro.ru/Installation_Eidos.php).

Для создания собственных интеллектуальных облачных Эйдос-приложений можно воспользоваться инструкцией:

[http://lc.kubagro.ru/aidos/How\\_to\\_make\\_your\\_own\\_cloud\\_Eidos-application.pdf](http://lc.kubagro.ru/aidos/How_to_make_your_own_cloud_Eidos-application.pdf).

По различным аспектам применения данной технологии есть более **320** видео-занятий, с которыми можно ознакомиться по ссылкам, приведенным на странице:

[http://lc.kubagro.ru/Video\\_lessons\\_by\\_Prof.E.V.Lutsenko/Catalog.php](http://lc.kubagro.ru/Video_lessons_by_Prof.E.V.Lutsenko/Catalog.php).

Каталог интеллектуальных облачных Эйдос-приложений, содержащий более **415** датасетов и описаний решения задач в системе Эйдос находится по ссылке:

[http://lc.kubagro.ru/Source\\_data\\_applications/WebAppls.html](http://lc.kubagro.ru/Source_data_applications/WebAppls.html).

Любое из этих приложений можно установить для изучения на своем компьютере (при условии, что на нем есть Internet и не заблокирован FTP доступ) в режиме 1.3 системы «Эйдос».

Кроме того в режиме 1.3 доступно 30 локальных учебных приложений, доступных без Internet, скачиваемых по ссылке [http://lc.kubagro.ru/Installation\\_Eidos.php](http://lc.kubagro.ru/Installation_Eidos.php) вместе с инсталляцией системы в полных архивах.

#### 4.3. Технологические особенности

«Эйдос-X++» обладает мощным инструментарием для работы с данными, включая 6 автоматизированных интерфейсов ввода информации из различных источников: таблиц, текстов и графики. Способность обрабатывать неполные, зашумленные и взаимозависимые данные,

измеряемые в различных шкалах и единицах, позволяет применять систему в реальных условиях, где данные редко соответствуют строгим стандартам. Это делает «Эйдос-Х++» универсальным инструментом для моделирования, прогнозирования и принятия решений.

#### **4.4. Интеграция локальных и облачных решений**

Особое внимание уделено интеграции локальных и облачных решений, что предоставляет пользователям доступ к широкому спектру учебных и научных приложений. В настоящее время доступны 30 локальное и более 415 облачных приложений, что подтверждает востребованность и актуальность системы среди исследователей и практиков.

#### **4.5. Практическое значение**

Доказательством эффективности системы служат многолетние акты внедрения, начиная с 1987 года. Возможность создания моделей в реальном времени и их последующее использование для анализа, идентификации и прогнозирования обеспечивает широкий спектр применения системы «Эйдос-Х++» как в научной, так и в прикладной среде.

#### **Выводы**

Несмотря на наличие множества достойных решений в области ИИ, система «Эйдос-Х++» занимает уникальное место благодаря своей универсальности, доступности и практической направленности. Ее развитие и дальнейшее внедрение могут стать важным шагом на пути к созданию открытых и доступных технологий, способных решать сложные междисциплинарные задачи.

#### **5. ВЫВОДЫ И ПЕРСПЕКТИВЫ**

Результаты проведенного исследования подтвердили универсальность и высокую эффективность методологии автоматизированного системно-когнитивного анализа (АСК-анализ) и интеллектуальной системы «Эйдос». Применение разработанной методологии позволяет интегрировать данные различных типов, формализовать причинно-следственные связи и решать сложные задачи моделирования в условиях фрагментированности и неопределенности информации.

Кванторно-матричная модель, лежащая в основе подхода, обеспечивает унификацию описания данных, что делает возможным междисциплинарное применение в экономике, технических науках, биологии, сельском хозяйстве, психологии и педагогике, медицине и в других областях.

Одним из ключевых преимуществ системы «Эйдос» является ее доступность и адаптируемость. Открытая лицензия и широкий спектр учебных и научных приложений делают систему востребованной как среди профессионалов, так и среди пользователей, не обладающих специализированной подготовкой. Кроме того, использование графических процессоров (GPU) для ускорения вычислений делает систему подходящей для работы с большими данными.

Перспективы дальнейшего развития системы связаны с углублением исследований в области когнитивного моделирования и искусственного интеллекта. Планируется развитие алгоритмов, обеспечивающих большую точность и скорость анализа, а также расширение функционала для интеграции с современными облачными платформами. Дополнительным направлением является совершенствование мультязычной поддержки и разработка новых интеллектуальных приложений, ориентированных на решение прикладных задач.

Таким образом, методология АСК-анализа и система «Эйдос» представляют собой мощный инструмент для познания и решения задач в различных предметных областях. Их дальнейшее развитие способствует укреплению научного и практического потенциала междисциплинарных исследований, а также популяризации передовых технологий искусственного интеллекта.

Работа может быть основой для лабораторных работ и научных исследований по применению систем искусственного интеллекта, в частности лингвистического АСК-анализа для решения задач в различных предметных областях [1-67].

## **ЛИТЕРАТУРА**

1. Луценко Е.В., Головин Н.С. Математическая модель Автоматизированного системно-когнитивного анализа и системы «Эйдос» в кванторном и матричном представлении // January 2025, DOI: [10.13140/RG.2.2.10299.94246](https://doi.org/10.13140/RG.2.2.10299.94246), License [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/), <https://www.researchgate.net/publication/387670175>

## **References**

1. Lucenko E.V., Golovin N.S. Matematicheskaya model` Avtomatizirovannogo sistemno-kognitivnogo analiza i sistemy` «E`jdos» v kvantornom i matrichnom predstavlenii // January 2025, DOI: [10.13140/RG.2.2.10299.94246](https://doi.org/10.13140/RG.2.2.10299.94246), License [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/), <https://www.researchgate.net/publication/387670175>