

**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ
ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ И.Т. ТРУБИЛИНА»**

Факультет прикладной информатики

Кафедра Компьютерных систем и технологий

КУРСОВАЯ РАБОТА

по дисциплине: Интеллектуальные системы и технологии

на тему: «АСК-анализ демографических и поведенческих факторов оттока
клиентов»

Выполнил студент группы: ИТ2341 Фотеску Николай Константинович

Допущен к защите _____

Руководитель проекта д.э.н., к.т.н., профессор Луценко Е.В. (

(подпись, расшифровка подписи)



Защищен _____

(дата)

Оценка отлично

Краснодар
2025

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ
УНИВЕРСИТЕТ ИМЕНИ И. Т. ТРУБИЛИНА»

Факультет прикладной информатики

РЕЦЕНЗИЯ
на курсовую работу

Студента Фотеску Николая Константиновича курса

2 очной формы обучения группы ИТ2341

Направление подготовки 09.04.02 «Информационные системы и технологии»

Наименование темы «АСК-анализ демографических и поведенческих факторов оттока клиентов»

Рецензент: Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор

(Ф.И.О., ученое звание и степень, должность)

Оценка качества выполнения курсовой работы

| № п/п | Показатель | Оценка соответствия («неудовлетворительно», «удовлетворительно», «хорошо», «отлично») |
|-------|---|--|
| 1 | Степень полноты обзора состояния проблемы и корректность постановки цели и задач исследования | 5 |
| 2 | Уровень и корректность использования в работе различных методов исследований | 5 |
| 3 | Степень комплексности работы, применения в ней знаний общепрофессиональных и специальных дисциплин | 5 |
| 4 | Ясность, четкость, последовательность и обоснованность изложения | 5 |
| 5 | Применение современных технологий обработки информации | 5 |
| 6 | Качество оформления работы (общий уровень грамотности, стиль изложения, качество иллюстраций, соответствие требованиям по оформлению) | 5 |
| 7 | Уровень освоения компетенций в результате выполнения курсовой работы (проекта) | 5 |
| 8 | Ответы на вопросы при защите | 5 |

Достоинства работы _____

Недостатки работы _____

Итоговая оценка при защите отлично

Рецензент  (Е. В. Луценко)

«18» февраля 2023 г.

РЕФЕРАТ

Курсовая работа содержит: 80 страниц, 34 рисунка, 14 таблиц, 15 литературных источников.

Ключевые слова: АСК-АНАЛИЗ, ОТТОК КЛИЕНТОВ, ДЕМОГРАФИЧЕСКИЕ ФАКТОРЫ, ПОВЕДЕНЧЕСКИЕ ФАКТОРЫ, УДЕРЖАНИЕ КЛИЕНТОВ, ПРОГНОЗИРОВАНИЕ ОТТОКА, СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА «ЭЙДОС», МЕТРИЗАЦИЯ ДАННЫХ, КЛАССИФИКАЦИОННЫЕ ШКАЛЫ.

Основной задачей данной работы является проведение автоматизированного системно-когнитивного анализа (АСК-анализа) для выявления демографических и поведенческих факторов, влияющих на отток клиентов.

Для достижения этой цели необходимо решить ряд задач, связанных с формированием обобщенных образов классов, идентификацией конкретных клиентов с соответствующими классами принятия решений, а также исследованием моделируемой предметной области через анализ построенной модели.

АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ ДЕМОГРАФИЧЕСКИХ И ПОВЕДЕНЧЕСКИХ ФАКТОРОВ ОТТОКА КЛИЕНТОВ

Фотеску Николай Константинович
студент факультета ПИ, группы
ИТ2341

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

В современных условиях высокой конкуренции на рынке услуг и товаров проблема оттока клиентов становится одной из ключевых для компаний различных сфер деятельности. Отток клиентов может быть вызван множеством причин, связанных как с внутренними процессами компании, так и с индивидуальными характеристиками самих клиентов. Среди основных факторов выделяются демографические особенности (возраст, пол, доход) и поведенческие паттерны. Для комплексного анализа влияния этих факторов на уровень оттока предложен метод Автоматизированного системно-когнитивного анализа (АСК-анализ). Данная методология позволяет не только выявить корреляционные связи между различными параметрами, но и построить когнитивные модели, которые помогают понять механизмы принятия решений клиентами о прекращении сотрудничества с компанией. Для реализации анализа используется программный инструментальный интеллектуальной системы «Эйдос», который обеспечивает автоматизацию процессов обработки данных и визуализации результатов. В статье представлен конкретный пример применения АСК-анализа для исследования демографических и поведенческих факторов, влияющих на отток клиентов. В рамках исследования использовались как табличные формы представления данных, так и графические модели, позволяющие наглядно продемонстрировать сложные зависимости между параметрами. Пример может быть применен как учебный материал для освоения методологии АСК-анализа, а также для разработки практических рекомендаций по снижению уровня оттока клиентов и повышению их лояльности.

Ключевые слова: АСК-АНАЛИЗ, СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА «ЭЙДОС», ДЕМОГРАФИЧЕСКИЕ ФАКТОРЫ, ПОВЕДЕНЧЕСКИЕ ХАРАКТЕРИСТИКИ, ОТТОК КЛИЕНТОВ, УПРАВЛЕНИЕ ЛОЯЛЬНОСТЬЮ.

AUTOMATED SYSTEM-COGNITIVE ANALYSIS OF DEMOGRAPHIC AND BEHAVIORAL FACTORS OF CUSTOMER OUTFLOW

Fotescu Nikolai Konstantinovich
student of the faculty of PI, groups
IT2341

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

In modern conditions of high competition in the market of services and goods, the problem of customer churn is becoming one of the key issues for companies in various fields of activity. Customer churn can be caused by a variety of reasons related to both the internal processes of the company and the individual characteristics of the customers themselves. Demographic features (age, gender, income) and behavioral patterns are among the main factors. For a comprehensive analysis of the influence of these factors on the level of outflow, the method of Automated system-cognitive analysis (ASK-analysis) is proposed. This methodology allows not only to identify correlations between various parameters, but also to build cognitive models that help to understand the mechanisms of customer decision-making about termination of cooperation with the company. To implement the analysis, the software toolkit of the intelligent Eidos system is used, which provides automation of data processing and visualization of results. The article provides a specific example of the use of ASK analysis to study demographic and behavioral factors affecting customer churn. The study used both tabular forms of data representation and graphical models to visually demonstrate complex dependencies between parameters. The example can be used as a training material for mastering the methodology of ASK analysis, as well as for developing practical recommendations to reduce the level of customer churn and increase their loyalty.

Keywords: ASK-ANALYSIS, SYSTEM-COGNITIVE ANALYSIS, INTELLIGENT SYSTEM "EIDOS", DEMOGRAPHIC FACTORS, BEHAVIORAL CHARACTERISTICS, CUSTOMER OUTFLOW, LOYALTY MANAGEMENT.

СОДЕРЖАНИЕ

| | |
|--|----|
| 1. INTRODUCTION (ВВЕДЕНИЕ)..... | 6 |
| 1.1. ОПИСАНИЕ ИССЛЕДУЕМОЙ ПРЕДМЕТНОЙ ОБЛАСТИ..... | 6 |
| 1.2. ОБЪЕКТ И ПРЕДМЕТ ИССЛЕДОВАНИЯ | 6 |
| 1.3. ПРОБЛЕМА, РЕШАЕМАЯ В РАБОТЕ И ЕЕ АКТУАЛЬНОСТЬ..... | 7 |
| 1.4. ЦЕЛЬ РАБОТЫ | 7 |
| 2. METHODS (МЕТОДЫ) | 7 |
| 2.1. ОБОСНОВАНИЕ ТРЕБОВАНИЙ К МЕТОДУ РЕШЕНИЯ ПРОБЛЕМЫ | 7 |
| 2.2. ЛИТЕРАТУРНЫЙ ОБЗОР МЕТОДОВ РЕШЕНИЯ ПРОБЛЕМЫ, ИХ ХАРАКТЕРИСТИКА И ОЦЕНКА СТЕПЕНИ СООТВЕТСТВИЯ ОБОСНОВАННЫМ ТРЕБОВАНИЯМ | 8 |
| 2.3. АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ (АСК-АНАЛИЗ) КАК МЕТОД РЕШЕНИЯ ПРОБЛЕМЫ..... | 8 |
| 2.4. СИСТЕМА «ЭЙДОС» – ИНСТРУМЕНТАРИЙ АСК-АНАЛИЗА | 10 |
| 2.5. ЦЕЛЬ И ЗАДАЧИ РАБОТЫ..... | 17 |
| 3. RESULTS (РЕЗУЛЬТАТЫ) | 19 |
| 3.1. <i>Задача-1. Когнитивная структуризация предметной области. Две интерпретации классификационных и описательных шкал и градаций.....</i> | 19 |
| 3.2. <i>Задача-2. Формализация предметной области</i> | 20 |
| 3.3. <i>Задача-3. Синтез статистических и системно-когнитивных моделей. Многопараметрическая типизация и частные критерии знаний</i> | 26 |
| 3.4. <i>Задача-4. Верификация моделей.....</i> | 37 |
| 3.4.1. <i>Универсальная постановка задачи, не зависящая от предметной области.....</i> | 37 |
| 3.5. <i>Задача-5. Выбор наиболее достоверной модели</i> | 41 |
| 3.6. <i>Задача-6. Системная идентификация и прогнозирование</i> | 42 |
| 3.6.1. <i>Интегральный критерий «Сумма знаний».....</i> | 43 |
| 3.6.2. <i>Интегральный критерий «Семантический резонанс знаний».....</i> | 43 |
| 3.6.3. <i>Важные математические свойства интегральных критериев</i> | 45 |
| 3.6.4. <i>Конкретное решение задачи идентификации и прогнозирования в данной работе в системе «Эйдос»</i> | 45 |
| 3.7. <i>Задача-7. Поддержка принятия решений</i> | 48 |
| 3.7.1. <i>Упрощенный вариант принятия решений как обратная задача прогнозирования, позитивный и негативный информационные портреты классов, SWOT-анализ.....</i> | 48 |
| 3.7.2. <i>Развитый алгоритм принятия решений в адаптивных интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»</i> | 50 |
| 3.8. <i>Задача-8. Исследование объекта моделирования путем исследования его модели.....</i> | 54 |
| 3.8.1. <i>Инвертированные SWOT-диаграммы значений описательных шкал (семантические потенциалы).....</i> | 54 |
| 3.8.2. <i>Кластерно-конструктивный анализ классов</i> | 56 |
| 3.8.4. <i>Модель знаний системы «Эйдос» и нелокальные нейроны</i> | 59 |
| 3.8.5. <i>Нелокальная нейронная сеть.....</i> | 61 |
| 3.8.6. <i>3D-интегральные когнитивные карты</i> | 63 |
| 3.8.7. <i>2D-интегральные когнитивные карты содержательного сравнения классов (опосредованные нечеткие правдоподобные рассуждения)</i> | 64 |
| 3.8.8. <i>Когнитивные функции</i> | 65 |
| 3.8.9. <i>Значимость описательных шкал и их градаций.....</i> | 71 |
| 3.8.11. <i>Степень детерминированности классов и классификационных шкал</i> | 73 |
| 3.8.12. <i>Агломеративная кластеризация значений факторов.....</i> | 75 |
| 4. DISCUSSION (ОБСУЖДЕНИЕ) | 76 |
| 5. CONCLUSIONS (ВЫВОДЫ) | 76 |

1. INTRODUCTION (ВВЕДЕНИЕ)

1.1. Описание исследуемой предметной области

Отток клиентов представляет собой одну из наиболее острых проблем для современных компаний, оказывающих товары или услуги. Утрата клиентской базы напрямую влияет на финансовые показатели бизнеса, снижает его конкурентоспособность и требует дополнительных затрат на привлечение новых клиентов. Исследование причин оттока становится важным этапом в разработке стратегий удержания и повышения лояльности клиентов.

Предметная область исследования охватывает два ключевых аспекта: демографические и поведенческие характеристики клиентов. К демографическим факторам относятся такие параметры, как возраст, пол, уровень дохода, семейное положение и место проживания. Эти данные помогают выделить группы риска, которые с большей вероятностью могут прекратить сотрудничество с компанией. Поведенческие факторы включают частоту взаимодействия с продуктом или услугой, историю покупок, уровень удовлетворенности качеством сервиса и степень лояльности бренду.

Особое внимание уделяется взаимосвязи между этими характеристиками и решениями клиентов об отказе от услуг компании. Например, молодые клиенты могут быть более склонны к экспериментам с новыми брендами, тогда как пожилые люди чаще остаются лояльными проверенным компаниям. Аналогично, низкая частота использования сервиса или недостаточный уровень персонализации предложений может стать причиной оттока.

Целью анализа является выявление ключевых закономерностей, определяющих отток клиентов, и разработка рекомендаций для их минимизации. Для этого применяется Автоматизированный системно-когнитивный анализ (АСК-анализ), который позволяет комплексно исследовать взаимосвязи между различными факторами и построить модели прогнозирования оттока.

1.2. Объект и предмет исследования

Объектом исследования выступает процесс оттока клиентов в современных компаниях, предоставляющих товары или услуги. Этот процесс рассматривается как комплексное явление, обусловленное влиянием различных внутренних и внешних факторов, таких как демографические особенности клиентов, их поведенческие паттерны и уровень удовлетворенности взаимодействием с компанией.

Предметом исследования являются демографические и поведенческие факторы, оказывающие влияние на решение клиентов

прекратить сотрудничество с компанией. В рамках предметной области изучаются взаимосвязи между характеристиками клиентов (возраст, пол, доход, частота покупок, лояльность и др.) и вероятностью их оттока. Особое внимание уделяется анализу причинно-следственных связей, которые позволяют выявить ключевые драйверы оттока и определить группы риска среди клиентской базы.

1.3. Проблема, решаемая в работе и ее актуальность

Проблема оттока клиентов является одной из наиболее острых для современных компаний, особенно в условиях высокой конкуренции и динамично меняющегося рынка. Потеря клиентов не только снижает доходы бизнеса, но и увеличивает затраты на привлечение новых, что негативно сказывается на общей рентабельности компании. При этом многие организации сталкиваются с трудностями в выявлении истинных причин оттока, так как они часто обусловлены сложным сочетанием демографических и поведенческих факторов, которые сложно анализировать традиционными методами.

Актуальность данной работы заключается в необходимости разработки эффективного подхода к анализу и прогнозированию оттока клиентов на основе их демографических и поведенческих характеристик. Современные методы анализа данных, такие как Автоматизированный системно-когнитивный анализ (АСК-анализ), позволяют выявить скрытые закономерности и причинно-следственные связи, которые невозможно обнаружить с помощью стандартных инструментов. Это дает возможность компаниям своевременно выявлять группы риска среди клиентов и принимать превентивные меры для их удержания.

Решение данной проблемы имеет важное практическое значение, так как позволяет оптимизировать маркетинговые и управленческие стратегии, повысить уровень лояльности клиентов и снизить затраты на поддержание клиентской базы. Таким образом, внедрение АСК-анализа в практику исследований оттока клиентов способствует повышению конкурентоспособности компаний и устойчивости их бизнеса в долгосрочной перспективе.

1.4. Цель работы

Цель работы – выявить демографические и поведенческие факторы оттока клиентов с помощью АСК-анализа и разработать рекомендации для его снижения.

2. METHODS (МЕТОДЫ)

2.1. Обоснование требований к методу решения проблемы

Из специфики поставленной проблемы сопоставимости обработки в одной модели исходных, представленных в разных типах шкал числовых и текстовых (лингвистических) и в разных единицах измерения, вытекают следующие **требования** к методу решения проблемы:

1. Метод должен обеспечивать устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных (неточных) взаимозависимых (нелинейных) данных большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения.

2. Метод решения проблемы не должен предъявлять чрезмерно жестких требований к исходным данным, которые практически невозможно выполнить, а должен обеспечивать обработку тех данных, которые реально есть.

3. Метод должен не теоретически, а реально на практике решать поставленную проблему, а значит, он должен иметь поддерживающий его программный инструментарий, находящийся в полном открытом бесплатном доступе.

2.2. Литературный обзор методов решения проблемы, их характеристика и оценка степени соответствия обоснованным требованиям

Поиск в Internet математических методов и реализующих их программных систем, **одновременно** удовлетворяющих всем требованиям, обоснованным в п.2.1 данной работы показал, что альтернатив Автоматизированному системно-когнитивному анализу и его программному инструментарию – системе «Эйдос» в настоящее время практически нет.

2.3. 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 годов¹ и фундаментальной монографии .

Сам термин: «Автоматизированный системно-когнитивный анализ (АСК-анализ)» был предложен проф.Е.В.Луценко в 2001 году. На тот момент этот термин вообще не встречался в Internet. Сегодня по соответствующему запросу в Гугле находится около 23000 сайтов с этим сочетанием слов².

Примечание: Ниже приведено очень краткое описание АСК-анализа и системе «Эйдос». Это описание может выглядеть как нескромность и самовосхваление. Но автор просит

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

² [https://yandex.ru/search/?text=Автоматизированный%2Всистемно-когнитивный%2Ванализ%2В\(АСК-анализ\)&lr=35&clid=2327117-18&win=360](https://yandex.ru/search/?text=Автоматизированный%2Всистемно-когнитивный%2Ванализ%2В(АСК-анализ)&lr=35&clid=2327117-18&win=360)

читателей понять его правильно. Это сделано исключительно для тех довольно многочисленных читателей, которые впервые слышат об этом методе и системе.

АСК-анализ включает:

- теоретические основы, в частности базовую формализуемую когнитивную концепцию;
- математическую модель, основанную на системном обобщении теории информации (СТИ);
- методику численных расчетов (структуры баз данных и алгоритмы их обработки);
- программный инструментарий, в качестве которого в настоящее время выступает универсальная когнитивная аналитическая система «Эйдос» (интеллектуальная система «Эйдос»).

Более подробно математический метод АСК-анализа описан в работе и ряде других. Около половины из 706 опубликованных автором научных работ посвящены теоретическим основам АСК-анализа и его практическим применениям в ряде предметных областей. На момент написания данной работы автором опубликовано 50 монографий, 27 учебных пособий, в т.ч. 3 учебных пособия с грифами УМО и Министерства, получено 34 патента РФ на системы искусственного интеллекта, 360 публикации в изданиях, входящих в перечень ВАК РФ и приравненных им (по данным РИНЦ), 17 публикаций в изданиях, входящих в ядро РИНЦ, 6 статей в журналах, входящих в WoS, 7 публикаций в журналах, входящих в Скопус³.

Три монографии включены в фонды библиотеки конгресса США⁴.

АСК-анализ и система "Эйдос" были успешно применены в 8 докторских диссертациях по экономическим, техническим, биологическим наукам и в 8 кандидатских диссертациях по экономическим, техническим, психологическими и медицинским наукам, еще несколько докторских и кандидатских диссертаций по техническим, экономическим, филологическим и медицинским наукам с применением АСК-анализа находятся в стадии выхода на защиту.

Автор является основателем междисциплинарной научной школы: «Автоматизированный системно-когнитивный анализ»⁵. Научная школа: "Автоматизированный системно-когнитивный анализ" является междисциплинарным научным направлением на пересечении, по крайней мере, трех научных специальностей (согласно недавно утвержденной новой номенклатуры научных специальностей ВАК РФ⁶). Основные научные специальности, которым соответствует научная школа:

- 5.12.4. Когнитивное моделирование;

³ <http://lc.kubagro.ru/aidos/Sprab0802.pdf>

⁴ <https://catalog.loc.gov/vwebv/search?searchArg=Lutsenko+E.V.> (и кликнуть: "Search")

⁵ <https://www.famous-scientists.ru/school/1608>

⁶ <https://www.garant.ru/products/ipo/prime/doc/400450248/>

- 1.2.1. Искусственный интеллект и машинное обучение;
- 2.3.1. Системный анализ, управление и обработка информации.

Научная школа: "Автоматизированный системно-когнитивный анализ" включает следующие междисциплинарные научные направления:

- Автоматизированный системно-когнитивный анализ числовых и текстовых табличных данных ;
- Автоматизированный системно-когнитивный анализ текстовых данных [32];
- Спектральный и контурный автоматизированный системно-когнитивный анализ изображений;
- Сценарный автоматизированный системно-когнитивный анализ временных и динамических рядов.

Приводить здесь ссылки на все эти работы здесь вряд ли целесообразно. Отметим лишь, что у автора есть личный сайт и страничка в РесечГейт и РИНЦ, на которых можно получить более полную информацию о методе АСК-анализа и системе «Эйдос». Краткая информация об АСК-анализе и системе «Эйдос» есть в материале: http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf.

На сайте автора размещены тематические подборки публикаций по применению АСК-анализа и системы «Эйдос» в различных предметных областях.

Решение поставленной в работе проблемы сопоставимости в АСК-анализе и системе «Эйдос» обеспечивается *путем* метризации номинальных шкал, т.е. повышения их степени формализации до уровня числовых шкал. Сама метризация номинальных шкал достигается *путем* вычисления количества информации, содержащегося в градациях номинальных шкал о получении той или иной урожайности. Для работы с лингвистическими переменными применяются стандартные возможности АСК-анализа.

2.4. Система «Эйдос» – инструментарий АСК-анализа

Конечно, на системе «Эйдос» как говорят «Свет клином не сошелся». Существует много очень достойных систем искусственного интеллекта. Чтобы лично убедиться в этом достаточно самостоятельно осуществить поиск в Internet, просто посмотреть файлы: [NCKR-1](#), [NCKR-2](#), [NCKR-3](#), [NCKR-4](#) или пройти по ссылкам: <https://learn.microsoft.com/ru-ru/dotnet/machine-learning/how-does-ml-dotnet-work>, <http://chat.openai.com/>, <https://poe.com/>, <https://neural-university.ru/>, <https://dzen.ru/a/ZCKZRKvrlEMBWOк8>, <https://ora.ai/>, <https://ora.ai/explore?path=trending>, <https://ora.ai/eugene-lutsenko/aidos>, <https://poe.com/Aidos-X>, <https://rudalle.ru/>, <https://bard.google.com/>, <https://chatbot.theb.ai/>, <https://problembo.com/ru/services> (может пригодиться

- почта на 10 минут: <https://10minutemail.net/>), <https://poe.com/GPT-3.5-Turbo-Instruct>, <https://www.seaart.ai/home>, <https://ui.chatai.com/>.

И все же Универсальная когнитивная аналитическая система «Эйдос-Х++» отличается от большинства из этих систем, по крайней мере, некоторыми из следующих своих параметров:

- является универсальной и может быть применена во многих предметных областях, т.к. разработана в универсальной постановке, не зависящей от предметной области (<http://lc.kubagro.ru/aidos/index.htm>) и имеет 6 автоматизированных программных интерфейсов (API) ввода данных из внешних источников данных различных типов: таблиц, текстов и графики. Система «Эйдос» является автоматизированной системой, т.е. предполагает непосредственное участие человека в реальном времени в процессе создания моделей и их использования для решения задач идентификации, прогнозирования, принятия решений и исследования предметной области путем исследования ее модели (автоматические системы работают без такого участия человека);

- является одной из первых и наиболее популярных отечественных систем искусственного интеллекта персонального уровня, т.е. не требует от пользователя специальной подготовки в области технологий искусственного интеллекта и программирования: есть акт внедрения системы «Эйдос» 1987 года (<http://lc.kubagro.ru/aidos/aidos02/PR-4.htm>);

- реально работает, обеспечивает устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных взаимозависимых (нелинейных) данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения (т.е. не предъявляет жестких требований к данным, которые невозможно выполнить, а обрабатывает те данные, которые есть);

- имеет «нулевой порог входа», содержит большое количество интеллектуальных локальных (т.е. поставляемых с инсталляцией) и облачных учебных и научных Эйдос-приложений (в настоящее время их 31 и более 426, соответственно: http://lc.kubagro.ru/Source_data_applications/WebAppls.htm) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf, http://lc.kubagro.ru/Presentation_LutsenkoEV.pdf);

- поддерживает on-line среду накопления знаний и обмена ими, широко используется во всем мире (<http://lc.kubagro.ru/map5.php>);



- обеспечивает мультиязычную поддержку интерфейса на 51 языке. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;

- наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решения этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA, т.е. поддерживать язык OpenGL);

- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач идентификации, прогнозирования, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе: http://lc.kubagro.ru/aidos/aidos18_LLS/aidos18_LLS.pdf);

- хорошо имитирует человеческий стиль мышления и является инструментом познания: дает результаты анализа, понятные экспертам на основе их опыта, интуиции и профессиональной компетенции, если эти эксперты уже есть, а если их еще нет, то она все равно дает верные результаты познания, что будет признано будущими экспертами, когда они появятся;

- вместо того, чтобы предъявлять к исходным данным практически неосуществимые требования (вроде нормальности распределения, абсолютной точности и полных повторностей всех сочетаний значений факторов и их полной независимости и аддитивности) автоматизированный системно-когнитивный анализ (АСК-анализ) предлагает без какой-либо предварительной обработки осмыслить те данные, которые есть, и, тем самым, преобразовать их в информацию, а

затем преобразовать эту информацию в знания путем ее применения для достижения целей (т.е. для принятия решений и управления) и решения задач классификации, поддержки принятия решений и содержательного эмпирического исследования моделируемой предметной области.

В чем сила подхода, реализованного в системе Эйдос? В том, что она реализует подход, эффективность которого не зависит от того, что мы думаем о предметной области и думаем ли вообще. Она формирует модели непосредственно на основе эмпирических данных, а не на основе наших представлений о механизмах реализации закономерностей в этих данных. Именно поэтому Эйдос-модели эффективны даже если наши представления о предметной области ошибочны или вообще отсутствуют.

В этом и слабость этого подхода, реализованного в системе Эйдос. Модели системы Эйдос - это феноменологические модели, отражающие эмпирические закономерности в фактах обучающей выборки, т.е. они не отражают причинно-следственного механизма детерминации, а только сам факт и характер детерминации. Содержательное объяснение этих эмпирических закономерностей формулируется уже экспертами на теоретическом уровне познания в содержательных научных законах.

We are briefly describing a new innovative method of artificial intelligence: Automated system-cognitive analysis (ASC-analysis), which has its own software tools – intelligent system called "Eidos" (open source software).

В разработке системы «Эйдос» были следующие этапы:

1-й этап, «подготовительный»: 1979-1992 годы. Математическая модель системы "Эйдос" разработана в 1979 и впервые прошла экспериментальную апробацию в 1981 году (первый расчет на компьютере на основе модели). С 1981 по 1992 система "Эйдос" неоднократно реализовалась на платформе Wang (на компьютерах Wang-2200C). В 1987 году впервые получен [акт внедрения](#) на одну из ранних версий системы «Эйдос», реализованную в среде персональной технологической системы «Вега-М» разработки автора (см.2-й акт).

2-й этап, «эра IBM PC и MS DOS»: 1992-2012 годы. Для IBM-совместимых персональных компьютеров система "Эйдос" впервые реализована на языках CLIPPER-87 и CLIPPER-5.01 (5.02) в 1992 году, а в 1994 году уже были получены [свидетельства РосПатента](#), первые в Краснодарском крае и, возможно, в России на системы искусственного интеллекта (слева приведена титульная видеограмма финальной DOS-версии системы «Эйдос-12.5», июнь 2012 года). С тех пор и до настоящего времени система непрерывно совершенствуется на IBM PC.

3-й этап, «эра MS Windows xp, 8, 7»: 2012-2020 годы. С июня 2012 по 14.12.2020 система «Эйдос» развивалась на языке [Аляска-1.9](#) + [Экспресс++](#) + библиотека для работы с Internet xb2net. Система «Эйдос-X1.9» хорошо работала на всех версиях MS Windows кроме Windows-10, которая требовала специальной настройки. Наиболее трудоемкие в

вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решение этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA).

4-й этап, «эра MS Windows-10»: 2020-2021 годы. С 13.12.2020 года по настоящее время система «Эйдос» развивается на языке [Аляска-2.0](#) + [Экспресс++](#). Библиотека xh2net в ней больше не используется, т.к. все возможности работы с Internet входят в [базовые возможности языка программирования](#).

5-й этап, «эра Больших данных, информации и знаний»: 2022 год. С 2022 года автор и разработчик системы «Эйдос» проф.Е.В.Луценко вплотную занялся разработкой профессиональной версии системы «Эйдос» на языке xBase++eXpress++Advantage Database Server (ADS), обеспечивающей обработку больших данных, информации и знаний (Big Data, Big Information, Big Knowledge).

6-й этап, «эра Больших данных, информации и знаний»: с 2023 года по настоящее время. С 2023 развитие системы «Эйдос» будет осуществляться на языках Питон (Python), а также [Аляска-2.0](#) + [Экспресс++](#).

[Скачать и запустить систему «Эйдос-X++» \(самую новую на текущий момент версию\) или обновление системы до текущей версии.](#) Это наиболее полная на данный момент незащищенная от несанкционированного копирования портативная (portable) версия системы (не требующая инсталляции) с полными [исходными текстами](#) текущей версии (за исключением ключей доступа к ftp-серверу системы «Эйдос» и ключей лицензионного программного обеспечения), находящаяся в полном открытом бесплатном доступе (около 180 Мб). Обновление имеет объем около 10 Мб. [Кредо](#). Лаборатория в [ResearchGate](#) по АСК-анализу и системе «Эйдос».

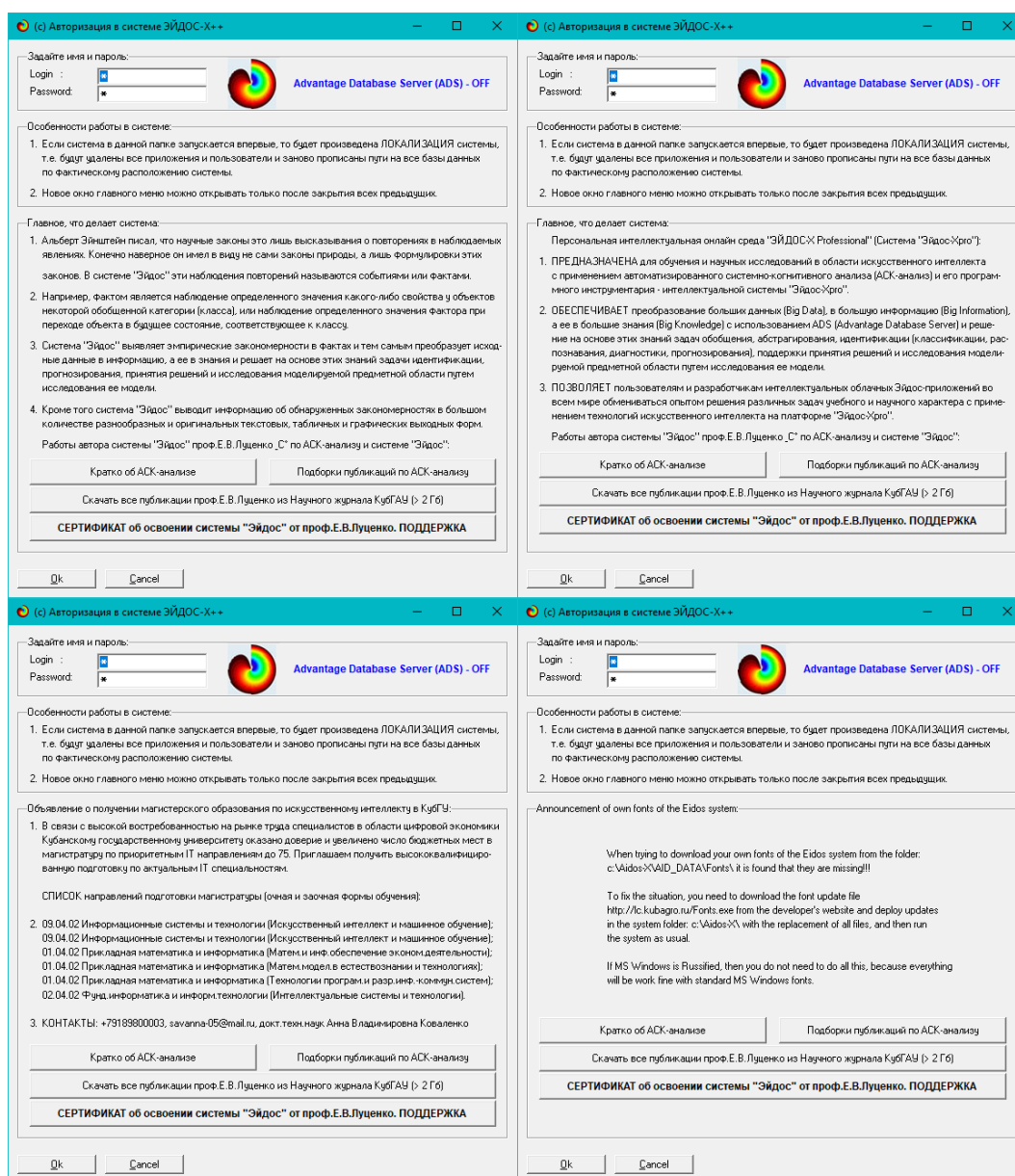
[Задание-инструкция для учащихся по разработке собственного интеллектуального облачного Эйдос-приложения](#)⁷

На рисунке 1 приведена титульная видеogramма DOS-версии системы «Эйдос», а на рисунке 2 – титульные видеogramмы текущей версии системы «Эйдос» (их в настоящее время 7 и они меняются по очереди при каждом запуске системы):

⁷ http://lc.kubagro.ru/aidos/How_to_make_your_own_cloud_Eidos-application.pdf



Рисунок 1. Титульная видеोगрамма DOS-версии системы «Эйдос» (до 2012 года)⁸



⁸ http://lc.kubagro.ru/pic/aidos_titul.jpg

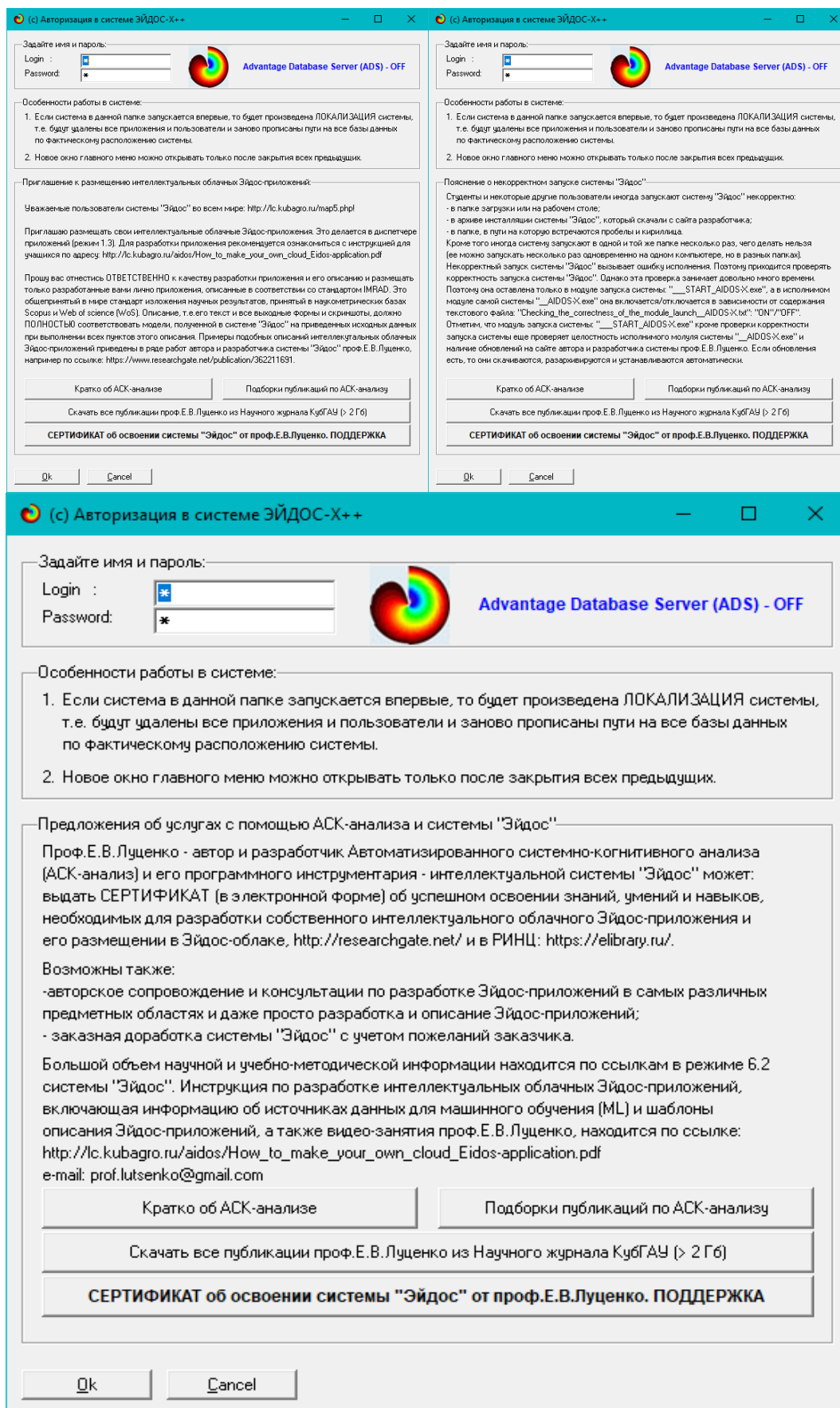


Рисунок 2. Титульные видеогаммы текущей версии системы «Эйдос»

2.5. Цель и задачи работы

Целью работы является решение поставленной проблемы.

Достижение поставленной цели в АСК-анализе обеспечивается решением следующих *задач* и подзадач, которые получаются в результате декомпозиции цели и являются *этапами* ее достижения:

Задача-1. Когнитивная структуризация предметной области.

Задача-2. Формализация предметной области.

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

Задача-4. Верификация моделей.

Задача-5. Выбор наиболее достоверной модели.

Задача-6. Системная идентификация и прогнозирование.

Задача-7. Поддержка принятия решений (упрощенный вариант принятия решений как обратная задача прогнозирования, позитивный и негативный информационные портреты классов, SWOT-анализ; развитый алгоритм принятия решений в АСК-анализе).

Задача-8 исследование объекта моделирования путем исследования его модели, включает ряд подзадач:

8.1) инвертированные SWOT-диаграммы значений описательных шкал (семантические потенциалы);

8.2) кластерно-конструктивный анализ классов;

8.3) кластерно-конструктивный анализ значений описательных шкал;

8.4) модель знаний системы «Эйдос» и нелокальные нейроны;

8.5) нелокальная нейронная сеть;

8.6) 3d-интегральные когнитивные карты;

8.7) 2d-интегральные когнитивные карты содержательного сравнения классов (опосредованные нечеткие правдоподобные рассуждения);

8.8) 2d-интегральные когнитивные карты содержательного сравнения значений факторов (опосредованные нечеткие правдоподобные рассуждения);

8.9) когнитивные функции;

8.10) значимость описательных шкал и их градаций;

8.11) степень детерминированности классов и классификационных шкал).

Для данной работы особое значение имеет решение подзадачи 8.1, связанной с анализом демографических и поведенческих факторов, влияющих на отток клиентов, так как это позволяет детально исследовать влияние различных характеристик клиентов на их решение прекратить сотрудничество с компанией, включая роль возраста, пола, частоты взаимодействия с сервисом и уровня удовлетворенности.

На рисунке 3 приведена последовательность преобразования исходных данных в информацию, а ее в знания и применения этих знаний для решения различных задач в системе «Эйдос»:

Последовательность обработки данных, информации и знаний в системе «Эйдос»,
повышение уровня системности данных, информации и знаний,
повышение уровня системности моделей

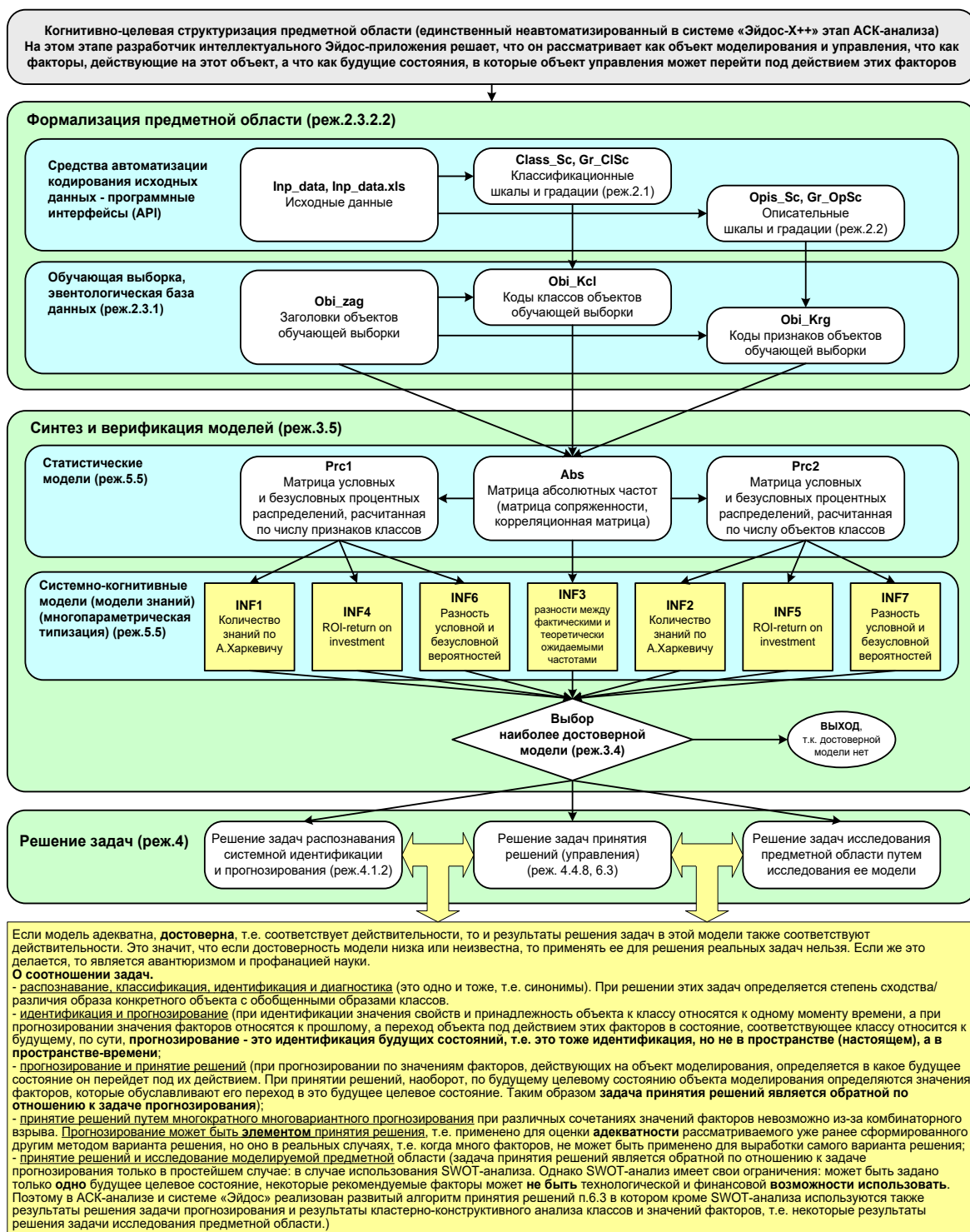


Рисунок 3. Последовательность преобразования исходных данных в информацию, а ее в знания и применения этих знаний для решения различных задач в системе «Эйдос» (этапы АСК-анализа)

3. RESULTS (РЕЗУЛЬТАТЫ)

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

На этапе когнитивно-целевой структуризации предметной области мы неформализуемым путем решаем на качественном уровне, что будем рассматривать в качестве объекта моделирования, что в качестве факторов, действующих на моделируемый объект (причин), а что в качестве результатов действия этих факторов (последствий). По сути это постановка решаемой проблемы.

Описательные шкалы служат для формального описания факторов, а классификационные – результатов их действия на объект моделирования. Шкалы могут быть числовые и текстовые. Текстовые шкалы могут быть номинальные и порядковые.

Когнитивная структуризация предметной области является первым и единственным неавтоматизированным в системе «Эйдос» этапом АСК-анализа, т.е. все последующие этапы АСК анализа в ней полностью автоматизированы.

В АСК-анализе и системе «Эйдос» применяется две интерпретации классификационных и описательных шкал и градаций: **статичная и динамичная** и соответствующая терминология (обобщающая, статичная и динамичная). Есть также обобщающая интерпретация и соответствующая ей терминология.

Статичная интерпретация и терминология:

- градации классификационных шкал – это обобщающие категории видов объектов (классы);
- описательные шкалы – свойства объектов, градации описательных шкал – значения свойств (признаки) объектов.

Динамичная интерпретация и терминология:

- градации классификационных шкал – это обобщающие категории будущих состояний объекта моделирования (классы), описывающие результаты действия факторов на объект моделирования в натуральном и стоимостном выражении: например, количество и качество продукции, прибыль и рентабельность;
- описательные шкалы – факторы, действующие на объект моделирования, градации описательных шкал – значения факторов, действующие на объект моделирования.

Обобщающая терминология:

- классификационные шкалы и градации;
- описательные шкалы и градации.

В данной работе в качестве объекта моделирования выступает отток клиентов, а в качестве факторов рассматриваются следующие характеристики: пол, возрастная категория, наличие партнера, иждивенцы, использование телефонной связи, подключение нескольких линий, доступ к интернет-услугам, наличие онлайн-защиты, защита устройства, техническая поддержка, стриминговые услуги, тип контракта, безбумажный расчет, метод оплаты, ежемесячные платежи и общая сумма платежей (таблица 1). В качестве результата действия этих факторов рассматривается вероятность оттока клиентов (таблица 2).

Таблица 1 – Описательные шкалы (факторы)

| KOD_OPSC | NAME_OPSC |
|----------|----------------|
| 1 | Gender |
| 2 | SENIORCITIZEN |
| 3 | PARTNER |
| 4 | DEPENDENTS |
| 5 | PHONESERVICE |
| 6 | PAYMENTMETHOD |
| 7 | MONTHLYCHARGES |
| 8 | TOTALCHARGES |

Таблица 2 – Классификационные шкалы (результаты действия факторов)

| KOD_CLSC | NAME_CLSC |
|----------|-----------|
| 1 | Churn |

3.2. Задача-2. Формализация предметной области

На этапе формализации предметной области разрабатываются классификационные и описательные шкалы и градации, а затем исходные данные кодируются с их использованием, в результате чего получается обучающая выборка. Обучающая выборка, по сути, представляет собой исходные данные, *нормализованные* с помощью классификационных и описательных шкал и градаций. Этим самым подготавливается все необходимое для выполнения следующего этапа АСК-анализа: синтеза и верификации статистических и системно-когнитивных моделей.

В системе «Эйдос» имеется большое количество разнообразных автоматизированных программных интерфейсов (API), обеспечивающих ввод в систему внешних данных различных типов: текстовых, табличных и графических, а также других, которые могут быть представлены в этом виде, например, аудио или данные электроэнцефалограммы (ЭЦГ) или кардиограммы (ЭКГ).

Этим обеспечивается возможность комфортного для пользователя применения системы «Эйдос» для проведения научных исследований и решения практических задач в самых различных предметных областях и направлениях науки, практически везде, где человек применяет естественный интеллект.

В качестве источника исходных данных в данной работе используем Excel-таблицы из работы (см. таблицы 3 и 4):

Таблица 3 – Исходные данные по влиянию различных факторов на вероятность оттока клиентов

| ID | gender | SeniorCitizen | Partner | Dependents | MonthlyCharges | TotalCharges | Churn |
|------------|--------|---------------|---------|------------|---------------------------|--------------|-------|
| 7590-VHVEG | Female | No | Yes | No | Electronic check | 30 | No |
| 5575-GNVDE | Male | No | No | No | Mailed check | 57 | No |
| 3668-QPYBK | Male | No | No | No | Mailed check | 54 | Yes |
| 7795-CFOCW | Male | No | No | No | Bank transfer (automatic) | 42 | No |
| 9237-HQITU | Female | No | No | No | Electronic check | 71 | Yes |
| 9305-CDSKC | Female | No | No | No | Electronic check | 100 | Yes |
| 1452-KIOVK | Male | No | No | Yes | Credit card (automatic) | 89 | No |
| 6713-OKOMC | Female | No | No | No | Mailed check | 30 | No |
| 7892-POOKP | Female | No | Yes | No | Electronic check | 105 | Yes |
| 6388-TABGU | Male | No | No | Yes | Bank transfer (automatic) | 56 | No |
| 9763-GRSKD | Male | No | Yes | Yes | Mailed check | 50 | No |
| 7469-LKBCI | Male | No | No | No | Credit card (automatic) | 19 | No |
| 8091-TTVAX | Male | No | Yes | No | Credit card (automatic) | 100 | No |
| 0280-XJGEX | Male | No | No | No | Bank transfer (automatic) | 104 | Yes |
| 5129-JLPIS | Male | No | No | No | Electronic check | 106 | No |
| 3655-SNQYZ | Female | No | Yes | Yes | Credit card (automatic) | 113 | No |
| 8191-XWSZG | Female | No | No | No | Mailed check | 21 | No |
| 9959-WOFKT | Male | No | No | Yes | Bank transfer (automatic) | 107 | No |
| 4190-MFLUW | Female | No | Yes | Yes | Credit card (automatic) | 55 | Yes |
| 4183-MYFRB | Female | No | No | No | Electronic check | 90 | No |
| 8779-QRDMV | Male | Yes | No | No | Electronic check | 40 | No |

Используя стандартные возможности MS Excel, *исходные данные из таблицы 3 представим в виде, стандартном для системы «Эйдос»* (таблица 4):

Таблица 3 – Excel-таблица исходных данных в стандарте системы «Эйдос»

| ID | gender | SeniorCitizen | Partner | Dependents | MonthlyCharges | TotalCharges | Churn |
|------------|--------|---------------|---------|------------|---------------------------|--------------|-------|
| 7590-VHVEG | Female | No | Yes | No | Electronic check | 30 | No |
| 5575-GNVDE | Male | No | No | No | Mailed check | 57 | No |
| 3668-QPYBK | Male | No | No | No | Mailed check | 54 | Yes |
| 7795-CFOCW | Male | No | No | No | Bank transfer (automatic) | 42 | No |
| 9237-HQITU | Female | No | No | No | Electronic check | 71 | Yes |
| 9305-CDSKC | Female | No | No | No | Electronic check | 100 | Yes |
| 1452-KIOVK | Male | No | No | Yes | Credit card (automatic) | 89 | No |
| 6713-OKOMC | Female | No | No | No | Mailed check | 30 | No |
| 7892-POOKP | Female | No | Yes | No | Electronic check | 105 | Yes |
| 6388-TABGU | Male | No | No | Yes | Bank transfer (automatic) | 56 | No |
| 9763-GRSKD | Male | No | Yes | Yes | Mailed check | 50 | No |
| 7469-LKBCI | Male | No | No | No | Credit card (automatic) | 19 | No |
| 8091-TTVAX | Male | No | Yes | No | Credit card (automatic) | 100 | No |
| 0280-XJGEX | Male | No | No | No | Bank transfer (automatic) | 104 | Yes |
| 5129-JLPIS | Male | No | No | No | Electronic check | 106 | No |
| 3655-SNQYZ | Female | No | Yes | Yes | Credit card (automatic) | 113 | No |
| 8191-XWSZG | Female | No | No | No | Mailed check | 21 | No |
| 9959-WOFKT | Male | No | No | Yes | Bank transfer (automatic) | 107 | No |
| 4190-MFLUW | Female | No | Yes | Yes | Credit card (automatic) | 55 | Yes |
| 4183-MYFRB | Female | No | No | No | Electronic check | 90 | No |
| 8779-QRDMV | Male | Yes | No | No | Electronic check | 40 | No |

Таблица 4 имеет следующую структуру:

- каждая строка описывает одно наблюдение с определенным сочетанием значений факторов и вероятностью оттока клиента.

- каждое наблюдение описывается двумя способами: с одной стороны, значениями факторов, влияющих на объект моделирования (лингвистические и числовые переменные, градации описательных шкал, представленные на бесцветном фоне в таблице 4), а с другой стороны, результатами воздействия этих факторов, то есть вероятностью оттока клиентов, выраженными в текстовых и числовых шкалах в различных единицах измерения (выделено желтым фоном). Такая структура описания наблюдений в технологиях искусственного интеллекта называется «онтологией», а в рамках модели представления знаний Марвина Мински (1975) именуется «фрейм-экземпляр».

- 1-я колонка – не является шкалой и содержит номер наблюдения или другую идентифицирующую информацию.

- 18-ая колонка классификационная шкала – это шкала *текстового* типа описывающие *результаты* действия факторов в различных единицах измерения (таблица 4), в данном случае отток клиента. В общем случае в исходных данных может быть значительно больше классификационных шкал, описывающих результаты воздействия факторов на объект моделирования как в натуральном, так и в стоимостном выражении. Например, это могут быть показатели, связанные с оттоком клиентов: количество ушедших клиентов, их доля в общей базе, упущенная прибыль или снижение рентабельности бизнеса. В системе «Эйдос» существует не очень жесткое ограничение на суммарное количество градаций классификационных шкал: их должно быть не более 2032;

- колонки с 2-й по 7-ю – это описательные шкалы, формализующие факторы, действующие на объект моделирования (таблица 4). Эти шкалы имеют числовой и текстовый тип и их градациями являются лингвистические и числовые переменные;

- при вводе данных в систему «Эйдос» нули и пробелы в исходных данных могут рассматриваться как значащие или как отсутствие данных. 1-й вариант и будет использован в данной работе.

Математическая модель системы «Эйдос» обеспечивает устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных (фрагментированных) зашумленных взаимозависимых (нелинейных) данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения.

Таким образом, система «Эйдос» не предъявляет чрезмерно жестких практически невыполнимых требований к исходным данным, а

обрабатывает те данные, которые реально есть, например, подобные представленным в таблице 4.

В системе «Эйдос» есть 6 основных автоматизированных программных интерфейсов (API), обеспечивающих ввод в систему и интеллектуальную обработку числовых, текстовых и графических данных, представленных в виде таблиц и файлов. Возможна обработка и других видов данных (например, данных землетрясений, ЭЭГ, ЭКГ, аудио и видео), которые возможно представить в этих форматах (рисунок 4).

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 транспортированных 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. Транспортирование 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

Рисунок 4. Автоматизированные программные интерфейсы (API) системы «Эйдос»

Для ввода исходных данных, представленных в таблице 4, в систему «Эйдос», используется один из ее автоматизированных программных интерфейсов (API), а именно универсальный автоматизированный программный интерфейс ввода данных из файлов MS Excel (API-2.3.2.2).

Требования API-2.3.2.2 к исходным данным подробно описаны в хелпах этого режима (рисунки 5):

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.MBI" в систему "Эйдос" и в форматизации предметной области.

Данный программный интерфейс обеспечивает форматизацию предметной области, т.е. анализ файла исходных данных (Inp_data.MBI), формирование классификационных и описательных шкал и градаций, а также кодирование файла исходных данных с их использованием.

Файл исходных данных должен иметь имя Inp_data.MBI, а файл результатов выбора имени Inp_data.MBI.Result. Файлы Inp_data.MBI и Inp_data.MBI.Result должны находиться в папке \AIMDOS\IABD_DATA\Inp_data.MBI. 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 переводы по словам разрешены, а объединение строк или абзаца не допускается. Эти наименования должны быть короткими, но понятными, т.е. они будут в выводе формы, а в ней еще будут добавлены наименования градаций. В числовой шкале 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 длиннее до 255 символов.

Каждая строка этого файла, начиная со 2-й, содержит данные об объекте обучающей выборки или о 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 колонки.

Столбцы, начиная со 2-го, являются классификационными и описательными шкалами и могут быть текстового (именованного / порядкового) или числового типа (до десятичных знаков после запятой).

Столбцы присваиваются числовой тип, если все значения его имеют числового типа. Если хотя бы одно значение является текстовым (не числом, в т.ч. пробелом), то столбцы присваивается текстовый тип. Это означает, что нули должны быть указаны нулями, а не пробелами.

Столбцы со 2-го по N-й являются классификационными шкалами (выделены параметрами) и содержат данные о классе (в текущем состоянии 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 значений факторов), характеризующих объекты обучающей выборки.

В результате работы режима формируется файл Inp_NAME.TXT стандарта MS DOS (биномиаль), в котором наименования классификационных и описательных шкал выделены СИМВОЛИЧЕСКИ. Система формирует классификационные и описательные шкалы и градаций. Для этого в каждой числовой столбце система находит минимальное и максимальное числовые значения и формирует заданное количество числовых интервалов, после чего числовые значения заменяются интервальной 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 градаций 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 классы соответствующих фактов соединения числовых или уникальных текстовых значений признаков с градациями описательных шкал и код признаков, соответствующие фактам соединения числовых или уникальных текстовых значений признаков с градациями описательных шкал.

Разомножения выбора формируются на основе файла Inp_RASP. Аналогично, за исключением того, что классификационные и описательные шкалы и градаций не создаются, а используются ранее созданные в модели. В файлы результатов выбора могут не входить классы классов, если столбцы классов в файле Inp_RASP были пустыми. Структура файла Inp_RASP должна быть такой 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

Помощь по режиму 2.3.2.2 для случая Excel-файлов исходных данных

Шкала классификации (классы) формируется градациями объекта и используется числовыми и текстовыми шкалами, факт и тип (переводы) для указывать наименование и кодирования. На наименованиях шкалы есть только отношение значимости и независимости, на переводы (кроме того есть отношение "Большой", "меньше", а на числовые - кроме того могут выполняться все арифметические операции). Каждый объект выбора (наблюдение) описан с одной стороны своими признаками, а с другой - принадлежностью к некоторым обучающим категориям (классам). Такая структура описания называется оптимальной или факторной записью и является базовой для всех типовых представлений знаний.

В АСК-анализе системы "Эйдос" используются три интерпретационных шкалы и градаций: универсальная, статическая и динамическая.

- в универсальной интерпретации: признаки - это градации описательных шкал;
- в статической интерпретации: описательная шкала - это свойства, а градация (признак) - это степень выраженности этого свойства;
- в динамической интерпретации: описательная шкала - это факторы, а градация (признак) - это значение фактора;
- в универсальной интерпретации: классы - это градации классификационных шкал;
- в статической интерпретации: классификационная шкала - способ классификации объектов обучающих категорий (классов), к которым в настоящее время по отношению к признакам относятся состояния объекта прогнозирования или управления;
- в динамической интерпретации: классификационная шкала - способ классификации обучающих категорий (классов), к которым в будущем прогнозируются события.

ПРОГРАММНЫЙ СВОБОДЫ В ФАЙЛЕ ИСХОДНЫХ ДАННЫХ:

- 1-я строка файла "Inp_data.MBI" должна содержать наименования колонок. Эти наименования должны быть во всех колонках; при этом переводы по словам разрешены, а объединение строк или абзаца не допускается. Эти наименования должны быть короткими, но понятными, т.е. они будут в выводе формы, а в ней еще будут добавлены наименования градаций. В числовой шкале 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 длиннее до 255 символов.
- Столбцы, начиная со 2-го, являются классификационными и описательными шкалами и могут быть текстового (именованного / порядкового) или числового типа (до десятичных знаков после запятой). Чтобы текстовая шкала была порядковой, надо чтобы при сортировке по алфавиту градаций этой шкалы образовывались последовательно от минимального значения до максимального. Например, текстовая шкала "Размер" с градациями "меньше", "максимум", "среднее", "большое", "очень большое", "бывает номинальной шкалой, т.е. при сортировке по алфавиту она расположится в порядке "большое", "максимум", "очень большое", "меньше", "среднее". Чтобы шкала "Размер" стала порядковой надо в эти градации присвоить следующие значения "1,5 очень малое", "2,5 малое", "3,5 среднее", "4,5 большое", "5,5 очень большое".
- Столбцы присваиваются числовой тип, если все значения его имеют числового типа. Если хотя бы одно значение является текстовым (не числом, в т.ч. пробелом), то столбцы присваивается текстовый тип. Это означает, что нули должны быть указаны нулями, а не пробелами.
- Если в системе "Эйдос" в режиме 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 интервалами или просто числами. Если числовыми интервалами, значит в файле исходных данных в этой столбце все правильно, если же числами, то колонки в Excel-файле надо заменить десятичными точкой и запятой, а также найти и исправить нечисловые данные в числовых по ссылке колонок. Быстро найти и можно перейти на последнюю строку файла исходных данных и сразу расчитать сумму колонок. В форме будет видно в какой строке 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 Shift+Home, а затем нажав клавишу Shift+Ctrl+начало End. Если этот блок выведет за пределы области таблицы, функция запятой данных надо отключить эту функциональную область данных в буфер обмена, создав новый лист и скопировав в него, а классификация лист "дальше".
- 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 вставить 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 |

Рисунок 5. Хелпы API-2.3.2.2 системы «Эйдос»

Экранные формы управления API-2.3.2.2 системы «Эйдос» с *реальными параметрами*, использованными в данной работе, приведены на рисунках 6:

2.3.2.2. Универсальный программный интерфейс импорта данных в систему "ЭЙДОС-X++"

Автоматическая формализация предметной области: генерация классификационных и описательных шкал и градаций, а также обучающей и распознаваемой выборки на основе базы исходных данных: "Inp_data"

Задайте тип файла исходных данных: "Inp_data":

XLS - MS Excel-2003 Стандарт XLS-файла

XLSX- MS Excel-2007(2010) Стандарт XLSX-файла

DBF - DBASE IV (DBF/NTX) Стандарт DBF-файла

CSV - CSV => DBF конвертер Стандарт CSV-файла

Задайте параметры:

Нули и пробелы считать ОТСУТСТВИЕМ данных

Нули и пробелы считать ЗНАЧЕНИЯМИ данных

Создавать БД средних по классам "Inp_davr.dbf"?

Требования к файлу исходных данных

Задайте диапазон столбцов классификационных шкал:

Начальный столбец классификационных шкал:

Конечный столбец классификационных шкал:

Задайте диапазон столбцов описательных шкал:

Начальный столбец описательных шкал:

Конечный столбец описательных шкал:

Задайте режим:

Формализации предметной области (на основе "Inp_data")

Генерации распознаваемой выборки (на основе "Inp_rasp")

Задайте способ выбора размера интервалов:

Равные интервалы с разным числом наблюдений

Разные интервалы с равным числом наблюдений

Задание параметров формирования сценариев или способа интерпретации текстовых полей "Inp_data":

Не применять сценарный метод АСК-анализа Применить сценарный метод АСК-анализа

Применить спец. интерпретацию текстовых полей классов Применить спец. интерпретацию текстовых полей признаков

Параметры интерпретации значений текстовых полей "Inp_data":

Интерпретация TXT-полей классов:

Значения полей текстовых классификационных шкал файла исходных данных "Inp_data" рассматриваются как целое

Интерпретация TXT-полей признаков:

Значения полей текстовых описательных шкал файла исходных данных "Inp_data" рассматриваются как целое

Какие наименования ГРАДАЦИЙ числовых шкал использовать:

Только интервальные числовые значения (например: "1/3-{59873.0000000, 178545.6666667}")

Только наименования интервальных числовых значений (например: "Минимальное")

И интервальные числовые значения, и их наименования (например: "Минимальное: 1/3-{59873.0000000, 178545.6666667}")

2.3.2.2. Задание размерности модели системы "ЭЙДОС-Х++"

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ: (равные интервалы)
 Количество градаций классификационных и описательных шкал в модели, т.е.: [2 классов x 18 признаков]

| Тип шкалы | Количество классификационных шкал | Количество градаций классификационных | Среднее количество градаций на класс.шкалу | Количество описательных шкал | Количество градаций описательных шкал | Среднее количество градаций на опис.шкалу |
|-----------|-----------------------------------|---------------------------------------|--|------------------------------|---------------------------------------|---|
| Числовые | 0 | 0 | 0,00 | 2 | 10 | 5,00 |
| Текстовые | 1 | 2 | 2,00 | 4 | 8 | 2,00 |
| ВСЕГО: | 1 | 2 | 2,00 | 6 | 18 | 3,00 |

Задайте количество числовых диапазонов (интервалов, градаций) в шкале: _____
 В описательных шкалах:

2.3.2.2. Процесс импорта данных из внешней БД "Inp_data" в систему "ЭЙДОС-Х++"

Стадии исполнения процесса

1/3: Формирование классификационных и описательных шкал и градаций на основе БД "Inp_data"- Готово
 2/3: Генерация обучающей выборки и базы событий "EventsKO" на основе БД "Inp_data"- Готово
 3/3: Переиндексация всех баз данных нового приложения- Готово

ПРОЦЕСС ФОРМАЛИЗАЦИИ ПРЕДМЕТНОЙ ОБЛАСТИ ЗАВЕРШЕН УСПЕШНО !!!

Прогноз времени исполнения

Начало: 16:54:52 **Окончание: 16:55:14**

100.00000%

Прошло: 0:00:22 **Осталось: 0:00:00**

Рисунок 6. Экранные формы управления API-2.3.2.2 системы «Эйдос»

Отметим, что в системе «Эйдос» обычно используются базы данных с расширением «dbf». Они открываются в MS Excel или могут быть конвертированы в файлы xlsx с помощью онлайн-сервисов или в режиме 5.12 (этот режим системы «Эйдос» написан на Питоне).

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

Математические модели, на основе которых рассчитываются статистические и системно-когнитивные модели автоматизированного системно-когнитивного анализа (АСК-анализ) и интеллектуальной системы «Эйдос», подробно описаны в ряде монографий и статей автора. Поэтому в данной работе мы рассмотрим эти вопросы очень кратко, акцентируя внимание лишь на математической взаимосвязи коэффициента возврата инвестиций (ROI) с мерой χ -квадрат Карла Пирсона и с семантической мерой целесообразности информации Александра Харкевича.

Отметим, что модели системы «Эйдос» основаны на матрице абсолютных частот, отражающей число встреч градаций описательных шкал по градациям классификационных шкал (фактов). Но для решения всех задач используется не непосредственно сама эта матрица, а матрицы условных и безусловных процентных распределений и системно-когнитивные модели, которые рассчитываются на ее основе и отражают какое количество информации содержится в факте наблюдения определенной градации описательной шкалы о том, что объект моделирования перейдет в состояние, соответствующее определенной градации классификационной шкалы (классу).

Математическая модель АСК-анализа и системы «Эйдос» основана на системной нечеткой интервальной математике и обеспечивает сопоставимую обработку больших объемов фрагментированных и зашумленных взаимосвязанных (нелинейных) данных, представленных в различных типах шкал (дихотомических, номинальных, порядковых и числовых) и различных единицах измерения.

Суть математической модели АСК-анализа состоит в следующем.

Непосредственно на основе эмпирических данных рассчитывается матрица абсолютных частот (таблица 5):

Таблица 5– 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 | | | | $N_{\Sigma j}$ | | | $N_{\Sigma\Sigma} = \sum_{j=1}^W N_{\Sigma j}$ |

На основе таблицы 5 рассчитываются матрицы условных и безусловных процентных распределений (таблица 6).

Таблица 6 – 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 | | | | $P_{\Sigma j}$ | | | |

Отметим, что в АСК-анализе и его программном инструментарии интеллектуальной системе «Эйдос» используется два способа расчета матриц условных и безусловных процентных распределений:

1-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество признаков по классу;

2-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество объектов обучающей выборки по классу.

На практике часто встречается существенная **несбалансированность** данных, под которой понимается сильно отличающееся количество наблюдений объектов обучающейся выборки, относящихся к различным градациям одной классификационной или описательной шкалы. Поэтому решать задачу на основе непосредственно матрицы абсолютных частот (таблица 8) было бы очень неразумно (за исключением меры взаимосвязи хи-квадрат) и переход от абсолютных частот к условным и безусловным относительным частотам (частостям) (таблица 6) является весьма обоснованным и логичным.

Этот переход полностью снимает проблему **несбалансированности** данных, т.к. в последующем анализе используется не матрица абсолютных частот (таблица 5), а матрицы условных и безусловных процентных распределений (таблица 6), а также матрицы системно-когнитивных моделей, рассчитываемые на основе матрицы абсолютных частот и матрицы условных и безусловных процентных распределений. Этот подход снимает также проблему обеспечения сопоставимости обработки в

одной модели исходных данных, представленных в различных видах шкал (дихотомических, номинальных, порядковых и числовых) и в разных единицах измерения. В системе «Эйдос» этот подход применяется всегда при решении любых задач.

Затем на основе таблиц 5 и 6 с использованием частных критериев, знаний приведенных таблице 7, рассчитываются матрицы 7 системно-когнитивных моделей (таблица 8).

Таблица 7– 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 |
| ABS , 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 частота. | |
| PRC1 , матрица условных P_{ij} и безусловных P_i процентных распределений, в качестве N_j используется суммарное количество признаков по классу | ... | $P_{ij} = \frac{N_{ij}}{N_j}; P_i = \frac{N_i}{N}$ |
| PRC2 , матрица условных P_{ij} и безусловных P_i процентных распределений, в качестве N_j используется суммарное количество объектов обучающей выборки по классу | ... | $P_{ij} = \frac{N_{ij}}{N_j}; P_i = \frac{N_i}{N}$ |
| INF1 , 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 | $I_{ij} = \Psi \times \log_2 \frac{P_{ij}}{P_i}$ | $I_{ij} = \Psi \times \log_2 \frac{N_{ij}}{\bar{N}_{ij}} = \Psi \times \log_2 \frac{N_{ij} N}{N_i N_j}$ |
| INF2 , 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. | $I_{ij} = \Psi \times \log_2 \frac{P_{ij}}{P_i}$ | $I_{ij} = \Psi \times \log_2 \frac{N_{ij}}{\bar{N}_{ij}} = \Psi \times \log_2 \frac{N_{ij} N}{N_i N_j}$ |
| INF3 , частный критерий: Хи-квадрат: разности между фактическими и теоретически ожидаемыми абсолютными частотами | ... | $I_{ij} = N_{ij} - \bar{N}_{ij} = N_{ij} - \frac{N_i N_j}{N}$ |
| INF4 , частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу | $I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$ | $I_{ij} = \frac{N_{ij}}{\bar{N}_{ij}} - 1 = \frac{N_{ij} N}{N_i N_j} - 1$ |
| INF5 , частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j -му классу | $I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$ | $I_{ij} = \frac{N_{ij}}{\bar{N}_{ij}} - 1 = \frac{N_{ij} N}{N_i N_j} - 1$ |
| INF6 , частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу | $I_{ij} = P_{ij} - P_i$ | $I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N} = \frac{N_{ij} N - N_i N_j}{N_j N}$ |
| INF7 , частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j -му классу | $I_{ij} = P_{ij} - P_i$ | $I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N} = \frac{N_{ij} N - N_i N_j}{N_j N}$ |

Обозначения к таблице:

i – значение прошлого параметра;

j – значение будущего параметра;

N_{ij} – 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_i – количество встреч i -м значения прошлого параметра по всей выборке;

N_j – 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_{ij} – 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;

Ψ – нормировочный коэффициент (Е.В.Луценко, 2002), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли;

P_i – безусловная относительная частота встречи i -го значения прошлого параметра в обучающей выборке;

P_{ij} – условная относительная частота встречи i -го значения прошлого параметра при j -м значении будущего параметра.

В таблице 7 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**.

Таблица 8 – Матрица системно-когнитивной модели

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

Количество частных критериев знаний и основанных на них системно-когнитивных моделей (таблица 7), применяемых в настоящее время в системе «Эйдос» равно 7 определяется тем, что они получаются путем всех возможных вариантов сравнения фактических и теоретических абсолютных частот, условных и безусловных относительных частот путем вычитания и путем деления, и при этом N_j рассматривается как суммарное количество или признаков, или объектов обучающей выборки в j -м классе, а **нормировка к нулю** (для аддитивных интегральных критериев), если нет связи между наличием признака и принадлежностью объекта к классу,

осуществляется либо логарифмированием, либо вычитанием единицы (таблица 9).

Таблица 9– 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, χ^2 -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 | --- | --- |

Обратим особое внимание на то, что сравнение фактических и теоретических абсолютных частот путем деления приводит при нормировках к нулю (что нужно для применения аддитивных интегральных критериев) путем взятия логарифма и путем вычитания 1 к *тем же самым* моделям, что и сравнение условных и безусловных относительных частот путем деления с теми же самыми способами нормировки. Таким образом, если на основе матрицы абсолютных частот рассчитать матрицы условных и безусловных процентных распределений, а затем сравнить фактические абсолютные частоты с теоретическими путем вычитания и деления, а также сравнить условные и безусловные относительные частоты также путем вычитания и деления и провести нормировку к 0 путем взятия логарифма и путем вычитания 1, то получается 3 статистических модели: матрица абсолютных частот и две матрицы относительных частот, т.е. условных и безусловных процентных распределений, а также всего 7 системно-когнитивных моделей. Других же системно-когнитивных моделей, рассчитываемых на основе приведенных статистических моделей, просто нет. Это и есть конфигуратор статистических и когнитивных моделей в смысле В.А.Лефевра. *Под конфигуратором В.А.Левфер понимал минимальный полный набор понятийных шкал или конструктов, т.е. понятий, достаточный для адекватного описания предметной области*⁹. Необходимо отметить, что все эти модели рассчитываются в интеллектуальной системе «Эйдос».

Когда мы сравниваем фактические и теоретические абсолютные частоты путем вычитания у нас получается частный критерий знаний: «хи-квадрат» (СК-модель INF3), когда же мы сравниваем их путем деления, то у нас получается частный критерий: «количество информации по

9 См. 1.2.1.2.1.1. Определение понятия конфигуратора, http://lc.kubagro.ru/aidos/aidos06_lec/index.htm

А.Харкевичу» (СК-модели INF1, INF2) или «коэффициент возврата инвестиций ROI» - Return On Investment (СК-модели INF4, INF5) в зависимости от способа нормировки.

Когда же мы сравниваем условные и безусловные относительные частоты путем вычитания у нас получается частный критерий знаний: «коэффициент взаимосвязи» (СК-модели INF6, INF7), когда же мы сравниваем их путем деления, то у нас получается частный критерий: «количество информации по А.Харкевичу» (СК-модели INF1, INF2).

Таким образом, мы видим, что все частные критерии знаний тесно взаимосвязаны друг с другом. Особенно интересна связь знаменитого критерия хи-квадрат К.Пирсона с замечательной мерой количества информации А.Харкевича и с известным в экономике коэффициентом ROI.

Вероятность рассматривается как предел, к которому стремится относительная частота (отношение количества благоприятных исходов к числу испытаний) при неограниченном увеличении количества испытаний. Ясно, что вероятность – это математическая абстракция, которая никогда не встречается на практике (также, как и другие математические и физические абстракции, типа математической точки, материальной точки, бесконечно малой и т.п.). На практике встречается только относительная частота. Но она может быть весьма близкой к вероятности. Например, при 480 наблюдений различие между относительной частотой и вероятностью (погрешность) составляет около 5%, при 1250 наблюдениях – около 2.5%, при 10000 наблюдениях – 1%.

Суть этих методов в том, что вычисляется количество информации в значении фактора о том, что объект моделирования перейдет под его действием в определенное состояние, соответствующее классу. Это позволяет сопоставимо и корректно обрабатывать разнородную информацию о наблюдениях объекта моделирования, представленную в различных типах измерительных шкал и различных единицах измерения.

На основе системно-когнитивных моделей, представленных в таблице 8 (отличаются частыми критериями, приведенными в таблице 10), решаются задачи идентификации (классификации, распознавания, диагностики, прогнозирования), поддержки принятия решений (обратная задача прогнозирования), а также задача исследования моделируемой предметной области путем исследования ее системно-когнитивной модели.

Отметим, что как значимость значения фактора, степень детерминированности класса и ценность или качество модели в АСК-анализе рассматривается варибельность значений частных критериев этого значения фактора, класса или модели в целом (таблица 10).

Таблица 10 – 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 | $H = \sqrt[2]{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$ |

Численно эта вариабельность может измеряться разными способами, например, средним отклонением модулей частных критериев от среднего, дисперсией или среднеквадратичным отклонением или его квадратом. В системе «Эйдос» принят последний вариант, т.к. эта величина совпадает с мощностью сигнала, в частности мощностью информации, а в АСК-анализе все модели рассматриваются как источник информации об объекте моделирования. Поэтому есть все основания уточнить традиционную терминологию АСК-анализа (таблица 10):

Итак, в разделе раскрывается простая Математическая взаимосвязь меры χ -квадрат Карла Пирсона с коэффициентом возврата инвестиций (ROI) и с семантической мерой целесообразности информации Александра Харкевича. Эта взаимосвязь обнаруживается, если на основе матрицы абсолютных частот рассчитать матрицы условных и безусловных процентных распределений, а затем сравнить фактические абсолютные частоты с теоретическими путем вычитания и деления, а также сравнить условные и безусловные относительные частоты также путем вычитания и деления и выполнить нормировку к нулю путем взятия логарифма или вычитания 1. При этом получается 3 статистических модели: матрица абсолютных частот и две матрицы относительных частот, т.е. условных и безусловных процентных распределений, а также всего 7 системно-когнитивных моделей. Именно 7, а не большее количество системно-когнитивных моделей в итоге получается потому, что модели, получающиеся в результате сравнения фактических и теоретических абсолютных частот путем деления и нормировки к нулю путем взятия логарифма или вычитания 1 **тождественно совпадают** с моделями,

получающимися путем сравнения условных и безусловных относительных частот путем деления и нормировки к нулю путем взятия логарифма или вычитания 1. Это и есть конфигуратор статистических и когнитивных моделей в смысле В.А.Лефевра, содержащий минимальное количество моделей, позволяющих полно описать моделируемую предметную область.

Показательно, примечательно и весьма замечательно, что *модель меры χ -квадрат Карла Пирсона из статистики оказалась математически тесно связанной с коэффициентом возврата инвестиций (ROI), применяемой в экономике в теории управления портфелем инвестиций и с мерой информации Александра Харкевича из семантической теории информации и теории управления знаниями. Все эти модели рассчитываются в интеллектуальной системе «Эйдос».*

В системе «Эйдос» синтез моделей производится в режиме 3.5 (рисунок 7):

3.5. Синтез и верификация моделей

Задайте модели для синтеза и верификации

Статистические базы:

1. ABS - частный критерий: количество встреч сочетаний: "класс-признак" у объектов обуч.выборки

Задайте источник данных для расчета модели ABS:

Обучающая выборка Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7

Задайте значение фона в матрице абсолютных частот:

2. PRC1 - частный критерий: усл. вероятность i-го признака среди признаков объектов j-го класса

3. PRC2 - частный критерий: условная вероятность i-го признака у объектов j-го класса

Системно-когнитивные модели (базы знаний):

4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1

5. INF2 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC2

6. INF3 - частный критерий: χ -квадрат, разности между фактическими и ожидаемыми абс. частотами

7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1

8. INF5 - частный критерий: ROI (Return On Investment); вероятности из PRC2

9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вероятности из PRC1

10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; вероятности из PRC2

Текущая модель

1. ABS

2. PRC1

3. PRC2

4. INF1

5. INF2

6. INF3

7. INF4

8. INF5

9. INF6

10. INF7

Параметры копирования обучающей выборки в распознаваемую (бутстрепный подход):

Какие объекты обуч. выборки копировать:

Копировать всю обучающую выборку

Копировать только текущий объект

Копировать каждый N-й объект

Копировать N случайных объектов

Копировать объекты от N1 до N2 (fastest)

Вообще не менять распознаваемую выборку

Пояснение по алгоритму верификации

Измеряется внутренняя достоверн. модели

Выполнить:

Синтез и верификацию

Только верификацию

Только синтез (xBase++)

Только синтез (Python)

Задайте процессор

CPU GPU

Задайте алгоритм:

Классика - дольше

Упрощенно-быстрее

Использование только наиболее достоверных результатов распознавания: Rasp.dbf и целесообразность применения бутстрепного подхода

Расчетный размер БД результатов распознавания Rasp.dbf равен 110094 байт, т.е.: 0.0051267 % от MAX-возможного, (от 2Гб)

УЧИТЫВАТЬ только наиболее достоверные результаты распознавания с МОДУЛЕМ инт. крит. "Резонанс знаний" выше %

В примении бутстрепного подхода нет необходимости. Синтез и верификация моделей будут выполнены на основе всей выборки.

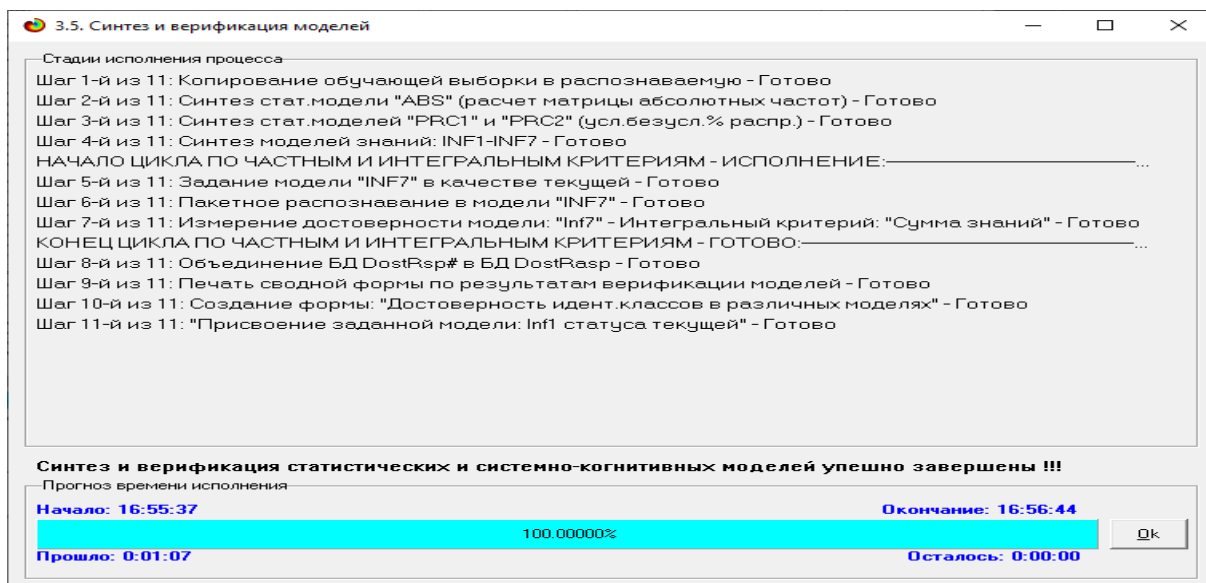


Рисунок 7. Экранные формы режима синтеза и верификации моделей

В результате работы режима 3.5 создано 3 статистических и 7 системно-когнитивных моделей, некоторые из которых приведены на рисунках 8-11:

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"

| Код признака | Наименование описательной шкалы и градации | 1. CHURN NO | 2. CHURN YES | 3. CHURN YES | Сумма | Среднее | Средн. квадрат. |
|--------------|--|-------------|--------------|--------------|---------|---------|-----------------|
| 1086.0 | TOTALCHARGES-7373.8000000 | | | 1.0 | 1.0 | 0.33 | 0.58 |
| 1087.0 | TOTALCHARGES-7382.2500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1088.0 | TOTALCHARGES-7405.5000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1089.0 | TOTALCHARGES-7422.1000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1090.0 | TOTALCHARGES-7459.0500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1091.0 | TOTALCHARGES-7511.6500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1092.0 | TOTALCHARGES-7532.1500000 | | | 1.0 | 1.0 | 0.33 | 0.58 |
| 1093.0 | TOTALCHARGES-7589.8000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1094.0 | TOTALCHARGES-7611.5500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1095.0 | TOTALCHARGES-7611.8500000 | | | 1.0 | 1.0 | 0.33 | 0.58 |
| 1096.0 | TOTALCHARGES-7616.0000000 | | | 1.0 | 1.0 | 0.33 | 0.58 |
| 1097.0 | TOTALCHARGES-7848.5000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1098.0 | TOTALCHARGES-7849.8500000 | | | 1.0 | 1.0 | 0.33 | 0.58 |
| 1099.0 | TOTALCHARGES-7853.7000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1100.0 | TOTALCHARGES-7878.3000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1101.0 | TOTALCHARGES-7882.2500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1102.0 | TOTALCHARGES-7895.1500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1103.0 | TOTALCHARGES-7930.5500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1104.0 | TOTALCHARGES-7943.4500000 | | | 1.0 | 1.0 | 0.33 | 0.58 |
| 1105.0 | TOTALCHARGES-7966.9000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1106.0 | TOTALCHARGES-7984.1500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1107.0 | TOTALCHARGES-8003.8000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1108.0 | TOTALCHARGES-8041.6500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1109.0 | TOTALCHARGES-8046.8500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1110.0 | TOTALCHARGES-8065.6500000 | | | 1.0 | 1.0 | 0.33 | 0.58 |
| 1111.0 | TOTALCHARGES-8071.0500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1112.0 | TOTALCHARGES-8126.6500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1113.0 | TOTALCHARGES-8129.3000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1114.0 | TOTALCHARGES-8196.4000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1115.0 | TOTALCHARGES-8220.4000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1116.0 | TOTALCHARGES-8248.5000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1117.0 | TOTALCHARGES-8306.0500000 | | | 1.0 | 1.0 | 0.33 | 0.58 |
| 1118.0 | TOTALCHARGES-8310.5500000 | | | 1.0 | 1.0 | 0.33 | 0.58 |
| 1119.0 | TOTALCHARGES-8399.1500000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1120.0 | TOTALCHARGES-8404.9000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1121.0 | TOTALCHARGES-8468.2000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| 1122.0 | TOTALCHARGES-8476.5000000 | | 1.0 | | 1.0 | 0.33 | 0.58 |
| | Сумма числа признаков | 64.0 | 12896.0 | 4624.0 | 17584.0 | | |
| | Среднее | 0.1 | 11.5 | 4.1 | | 5.22 | |
| | Среднеквадратичное отклонение | 0.4 | 61.3 | 22.8 | | | 38.05 |
| | Сумма числа объектов обуч.выборки | 4.0 | 806.0 | 289.0 | 1099.0 | | |

Рисунок 8. Статистическая модель «ABS», матрица абсолютных частот (фрагмент)

5.5. Модель: "3. PRC2 - частный критерий: условная вероятность i-го признака у объектов j-го класса"

| Код признака | Наименование описательной шкалы и градации | 1. CHURN NO | 2. CHURN NO | 3. CHURN YES | Безусл. вероятн | Среднее | Средн. квадрат |
|--------------|--|-------------|-------------|--------------|-----------------|---------|----------------|
| 1086.0 | TOTALCHARGES-7379.8000000 | | | 0.346 | 0.091 | 0.115 | 0.215 |
| 1087.0 | TOTALCHARGES-7382.2500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1088.0 | TOTALCHARGES-7405.5000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1089.0 | TOTALCHARGES-7422.1000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1090.0 | TOTALCHARGES-7459.0500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1091.0 | TOTALCHARGES-7511.6500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1092.0 | TOTALCHARGES-7532.1500000 | | | 0.346 | 0.091 | 0.115 | 0.215 |
| 1093.0 | TOTALCHARGES-7589.8000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1094.0 | TOTALCHARGES-7611.5500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1095.0 | TOTALCHARGES-7611.8500000 | | | 0.346 | 0.091 | 0.115 | 0.215 |
| 1096.0 | TOTALCHARGES-7616.0000000 | | | 0.346 | 0.091 | 0.115 | 0.215 |
| 1097.0 | TOTALCHARGES-7848.5000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1098.0 | TOTALCHARGES-7849.8500000 | | | 0.346 | 0.091 | 0.115 | 0.215 |
| 1099.0 | TOTALCHARGES-7853.7000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1100.0 | TOTALCHARGES-7878.3000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1101.0 | TOTALCHARGES-7882.2500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1102.0 | TOTALCHARGES-7895.1500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1103.0 | TOTALCHARGES-7930.5500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1104.0 | TOTALCHARGES-7943.4500000 | | | 0.346 | 0.091 | 0.115 | 0.215 |
| 1105.0 | TOTALCHARGES-7966.9000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1106.0 | TOTALCHARGES-7984.1500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1107.0 | TOTALCHARGES-8003.8000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1108.0 | TOTALCHARGES-8041.6500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1109.0 | TOTALCHARGES-8046.8500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1110.0 | TOTALCHARGES-8065.6500000 | | | 0.346 | 0.091 | 0.115 | 0.215 |
| 1111.0 | TOTALCHARGES-8071.0500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1112.0 | TOTALCHARGES-8126.6500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1113.0 | TOTALCHARGES-8129.3000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1114.0 | TOTALCHARGES-8196.4000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1115.0 | TOTALCHARGES-8220.4000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1116.0 | TOTALCHARGES-8248.5000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1117.0 | TOTALCHARGES-8306.0500000 | | | 0.346 | 0.091 | 0.115 | 0.215 |
| 1118.0 | TOTALCHARGES-8310.5500000 | | | 0.346 | 0.091 | 0.115 | 0.215 |
| 1119.0 | TOTALCHARGES-8399.1500000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1120.0 | TOTALCHARGES-8404.9000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1121.0 | TOTALCHARGES-8468.2000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| 1122.0 | TOTALCHARGES-8476.5000000 | 0.124 | | | 0.091 | 0.041 | 0.086 |
| | Сумма | 1600.000 | 1600.000 | 1600.000 | 4800.000 | | |
| | Среднее | 1.426 | 1.426 | 1.426 | | 1.426 | |
| | Среднеквадратичное отклонение | 9.337 | 7.605 | 7.905 | | | 8.314 |
| | Сумма числа объектов общ. выборки | 4.000 | 806.000 | 289.000 | 1099.000 | | |

Рисунок 9. Статистическая модель «PRC2», матрица условных и безусловных процентных распределений (фрагмент)

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"

| Код признака | Наименование описательной шкалы и градации | 1. CHURN NO | 2. CHURN NO | 3. CHURN YES | Сумма | Среднее | Средн. квадрат |
|--------------|--|-------------|-------------|--------------|----------|---------|----------------|
| 1086.0 | TOTALCHARGES-7379.8000000 | | | 0.217 | 0.217 | 0.072 | 0.125 |
| 1087.0 | TOTALCHARGES-7382.2500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1088.0 | TOTALCHARGES-7405.5000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1089.0 | TOTALCHARGES-7422.1000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1090.0 | TOTALCHARGES-7459.0500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1091.0 | TOTALCHARGES-7511.6500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1092.0 | TOTALCHARGES-7532.1500000 | | | 0.217 | 0.217 | 0.072 | 0.125 |
| 1093.0 | TOTALCHARGES-7589.8000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1094.0 | TOTALCHARGES-7611.5500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1095.0 | TOTALCHARGES-7611.8500000 | | | 0.217 | 0.217 | 0.072 | 0.125 |
| 1096.0 | TOTALCHARGES-7616.0000000 | | | 0.217 | 0.217 | 0.072 | 0.125 |
| 1097.0 | TOTALCHARGES-7848.5000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1098.0 | TOTALCHARGES-7849.8500000 | | | 0.217 | 0.217 | 0.072 | 0.125 |
| 1099.0 | TOTALCHARGES-7853.7000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1100.0 | TOTALCHARGES-7878.3000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1101.0 | TOTALCHARGES-7882.2500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1102.0 | TOTALCHARGES-7895.1500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1103.0 | TOTALCHARGES-7930.5500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1104.0 | TOTALCHARGES-7943.4500000 | | | 0.217 | 0.217 | 0.072 | 0.125 |
| 1105.0 | TOTALCHARGES-7966.9000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1106.0 | TOTALCHARGES-7984.1500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1107.0 | TOTALCHARGES-8003.8000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1108.0 | TOTALCHARGES-8041.6500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1109.0 | TOTALCHARGES-8046.8500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1110.0 | TOTALCHARGES-8065.6500000 | | | 0.217 | 0.217 | 0.072 | 0.125 |
| 1111.0 | TOTALCHARGES-8071.0500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1112.0 | TOTALCHARGES-8126.6500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1113.0 | TOTALCHARGES-8129.3000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1114.0 | TOTALCHARGES-8196.4000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1115.0 | TOTALCHARGES-8220.4000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1116.0 | TOTALCHARGES-8248.5000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1117.0 | TOTALCHARGES-8306.0500000 | | | 0.217 | 0.217 | 0.072 | 0.125 |
| 1118.0 | TOTALCHARGES-8310.5500000 | | | 0.217 | 0.217 | 0.072 | 0.125 |
| 1119.0 | TOTALCHARGES-8399.1500000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1120.0 | TOTALCHARGES-8404.9000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1121.0 | TOTALCHARGES-8468.2000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| 1122.0 | TOTALCHARGES-8476.5000000 | 0.050 | | | 0.050 | 0.017 | 0.029 |
| | Сумма | 4.279 | 39.429 | 60.328 | 104.036 | | |
| | Среднее | 0.004 | 0.035 | 0.054 | | 0.031 | |
| | Среднеквадратичное отклонение | 0.054 | 0.025 | 0.094 | | | 0.067 |
| | Сумма числа объектов общ. выборки | 4.000 | 806.000 | 289.000 | 1099.000 | | |

Рисунок 10. Системно-когнитивная модель «INF1», матрица информативностей (по А.Харкевичу) (фрагмент)

5.5. Модель: "6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс. частотами"

| Код признака | Наименование описательной шкалы и градации | 1. CHURN NO | 2. CHURN NO | 3. CHURN YES | Сумма | Среднее | Средн. квадрат |
|--------------|--|-------------|-------------|--------------|----------|---------|----------------|
| 1086.0 | TOTALCHARGES-7379.8000000 | -0.004 | -0.733 | 0.737 | | | 0.735 |
| 1087.0 | TOTALCHARGES-7382.2500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1088.0 | TOTALCHARGES-7405.5000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1089.0 | TOTALCHARGES-7422.1000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1090.0 | TOTALCHARGES-7459.0500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1091.0 | TOTALCHARGES-7511.6500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1092.0 | TOTALCHARGES-7532.1500000 | -0.004 | -0.733 | 0.737 | | | 0.735 |
| 1093.0 | TOTALCHARGES-7589.8000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1094.0 | TOTALCHARGES-7611.9500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1095.0 | TOTALCHARGES-7611.9500000 | -0.004 | -0.733 | 0.737 | | | 0.735 |
| 1096.0 | TOTALCHARGES-7616.0000000 | -0.004 | -0.733 | 0.737 | | | 0.735 |
| 1097.0 | TOTALCHARGES-7848.9000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1098.0 | TOTALCHARGES-7849.9500000 | -0.004 | -0.733 | 0.737 | | | 0.735 |
| 1099.0 | TOTALCHARGES-7853.7000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1100.0 | TOTALCHARGES-7878.3000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1101.0 | TOTALCHARGES-7882.2500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1102.0 | TOTALCHARGES-7895.1500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1103.0 | TOTALCHARGES-7930.9500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1104.0 | TOTALCHARGES-7943.4500000 | -0.004 | -0.733 | 0.737 | | | 0.735 |
| 1105.0 | TOTALCHARGES-7966.9000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1106.0 | TOTALCHARGES-7984.1500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1107.0 | TOTALCHARGES-8003.8000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1108.0 | TOTALCHARGES-8041.6500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1109.0 | TOTALCHARGES-8046.9500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1110.0 | TOTALCHARGES-8065.6500000 | -0.004 | -0.733 | 0.737 | | | 0.735 |
| 1111.0 | TOTALCHARGES-8071.0500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1112.0 | TOTALCHARGES-8126.6500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1113.0 | TOTALCHARGES-8129.3000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1114.0 | TOTALCHARGES-8196.4000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1115.0 | TOTALCHARGES-8220.4000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1116.0 | TOTALCHARGES-8248.5000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1117.0 | TOTALCHARGES-8306.0500000 | -0.004 | -0.733 | 0.737 | | | 0.735 |
| 1118.0 | TOTALCHARGES-8310.9500000 | -0.004 | -0.733 | 0.737 | | | 0.735 |
| 1119.0 | TOTALCHARGES-8399.1500000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1120.0 | TOTALCHARGES-8404.9000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1121.0 | TOTALCHARGES-8468.2000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| 1122.0 | TOTALCHARGES-8476.5000000 | -0.004 | 0.267 | -0.263 | | | 0.265 |
| | Сумма | | | | | | |
| | Среднее | | | | | | |
| | Среднеквадратичное отклонение | 0.184 | 3.204 | 3.192 | | | 2.613 |
| | Сумма числа объектов обуч.выборки | 4.000 | 806.000 | 289.000 | 1099.000 | | |

Рисунок 11 . Системно-когнитивная модель «INF3», матрица Хи-квадрат (по К.Пирсону) (фрагмент)

Полученные модели корректно использовать для решения задач только в том случае, если они достаточно достоверны (адекватны), т.е. верно отражают моделируемую предметную область. Поэтому в следующем разделе оценим достоверность созданных статистических и системно-когнитивных моделей.

3.4. Задача-4. Верификация моделей

3.4.1. Универсальная постановка задачи, не зависящая от предметной области

Оценка достоверности моделей в системе «Эйдос» осуществляется путем решения задачи классификации объектов обучающей выборки по обобщенным образам классов и подсчета количества истинных и ложных положительных и отрицательных решений по F-мере Ван Ризбергера, а также по критериям L1- L2-мерам проф. Е.В.Луценко, которые предложены для того, чтобы смягчить или полностью преодолеть некоторые недостатки F-меры.

Достоверность моделей можно оценивать и путем решения других задач, например, задач прогнозирования, выработки управляющих решений, исследования объекта моделирования путем исследования его

модели. Но это более трудоемко и даже всегда возможно, особенно на экономических и политических моделях.

В режиме 3.4 системы «Эйдос» и ряде других изучается достоверность каждой частной модели в соответствии с этими мерами достоверности.

В соответствии с критерием достоверности L1-мерой проф.Е.В.Луценко наиболее достоверной является СК-модель ABS с интегральным критерием: «Сумма знаний»: **L1=0.760** при максимуме 1 (рисунок 12). Эту модель и будем использовать для решения поставленных в работе задач.

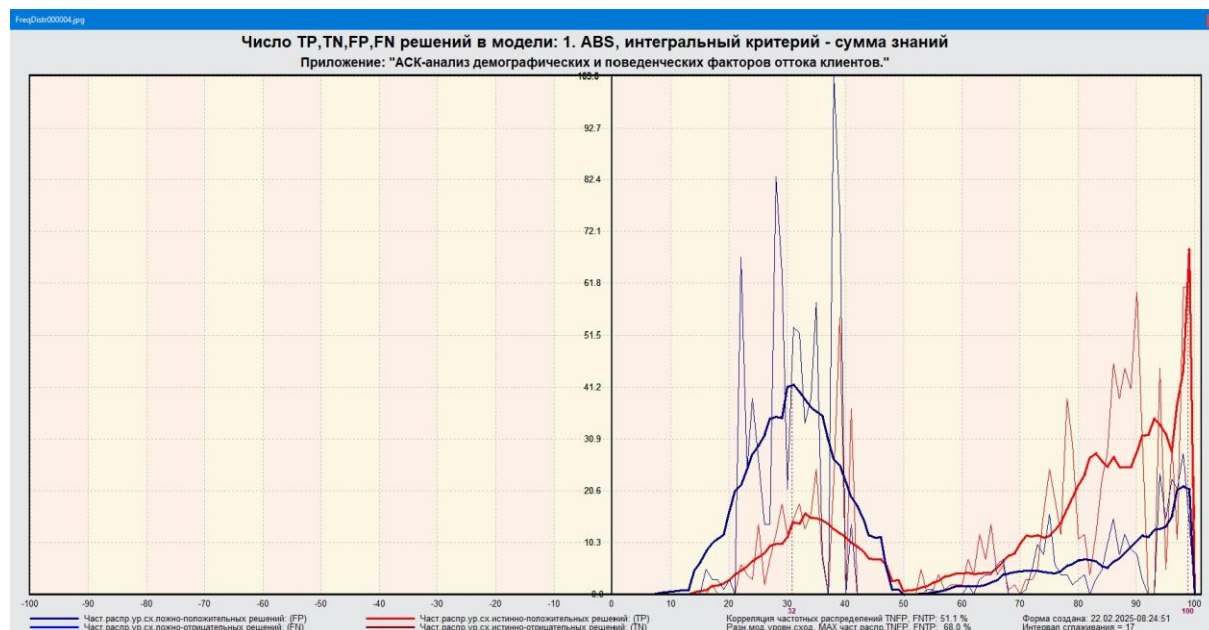
3.4. Обобщенная форма по доств.моделям при разн.крит. Текущая модель: "INF1"

| Наименование модели и частного критерия | Интегральный критерий | Точность модели | Полнота модели | Фигера Ван Рызбергена | Сумма модуль-уровней сход. истинно-полож. решений (STP) | Сумма модуль-уровней сход. истинно-отриц. решений (STN) | Сумма модуль-уровней сход. ложно-полож. решений (SFP) | Сумма модуль-уровней сход. ложно-отриц. решений (SFN) | S-Точность модели | S-Полнота модели | L1-мера проф. Е.В.Луценко | Средний уровень истинно-п. решений |
|---|-----------------------------------|-----------------|----------------|-----------------------|---|---|---|---|-------------------|------------------|---------------------------|------------------------------------|
| 1.ABS - частный критерий: количество встреч соематий "Клас... | Корреляция абс частот с обр... | 0.500 | 0.983 | 0.663 | 629.271 | 3.417 | 557.049 | 2.172 | 0.530 | 0.997 | 0.692 | 0 |
| 1.ABS - частный критерий: количество встреч соематий "Клас... | Значка абс частот по призна... | 0.500 | 1.000 | 0.667 | 789.107 | | 499.322 | | 0.612 | 1.000 | 0.760 | 0 |
| 2.PRC1 - частный критерий: усл. вероятность того признака сред... | Корреляция усл отн частот с о... | 0.500 | 0.983 | 0.663 | 629.271 | 3.417 | 557.049 | 2.172 | 0.530 | 0.997 | 0.692 | 0 |
| 2.PRC1 - частный критерий: усл. вероятность того признака сред... | Сумма усл отн частот по призна... | 0.500 | 1.000 | 0.667 | 846.562 | | 830.351 | | 0.505 | 1.000 | 0.671 | 0 |
| 3.PRC2 - частный критерий: условная вероятность того признака... | Корреляция усл отн частот с о... | 0.500 | 0.983 | 0.663 | 629.257 | 3.417 | 557.044 | 2.172 | 0.530 | 0.997 | 0.692 | 0 |
| 3.PRC2 - частный критерий: условная вероятность того признака... | Сумма усл отн частот по призна... | 0.500 | 1.000 | 0.667 | 846.562 | | 830.351 | | 0.505 | 1.000 | 0.671 | 0 |
| 4.INF1 - частный критерий: количество знаний по А.Харкевичу: в... | Семантический резонанс зна... | 0.502 | 0.548 | 0.524 | 209.792 | 172.130 | 191.375 | 160.388 | 0.523 | 0.567 | 0.544 | 0 |
| 4.INF1 - частный критерий: количество знаний по А.Харкевичу: в... | Сумма знания | 0.671 | 0.532 | 0.593 | 112.796 | 277.831 | 67.475 | 66.856 | 0.626 | 0.628 | 0.627 | 0 |
| 5.INF2 - частный критерий: количество знаний по А.Харкевичу: в... | Семантический резонанс зна... | 0.502 | 0.548 | 0.524 | 209.790 | 172.129 | 191.374 | 160.390 | 0.523 | 0.567 | 0.544 | 0 |
| 5.INF2 - частный критерий: количество знаний по А.Харкевичу: в... | Сумма знания | 0.671 | 0.532 | 0.593 | 112.796 | 277.831 | 67.475 | 66.856 | 0.626 | 0.628 | 0.627 | 0 |
| 6.INF3 - частный критерий: "Зинвадат, разности между факти... | Семантический резонанс зна... | 0.559 | 0.559 | 0.559 | 257.751 | 257.751 | 180.861 | 180.861 | 0.588 | 0.588 | 0.588 | 0 |
| 6.INF3 - частный критерий: "Зинвадат, разности между факти... | Сумма знания | 0.559 | 0.559 | 0.559 | 257.751 | 257.751 | 180.861 | 180.861 | 0.588 | 0.588 | 0.588 | 0 |
| 7.INF4 - частный критерий: ROI (Return On Investment); вероимо... | Семантический резонанс зна... | 0.510 | 0.510 | 0.510 | 200.813 | 200.813 | 174.385 | 174.385 | 0.535 | 0.535 | 0.535 | 0 |
| 7.INF4 - частный критерий: ROI (Return On Investment); вероимо... | Сумма знания | 0.608 | 0.608 | 0.608 | 164.586 | 217.234 | 110.656 | 58.007 | 0.598 | 0.739 | 0.661 | 0 |
| 8.INF5 - частный критерий: ROI (Return On Investment); вероимо... | Семантический резонанс зна... | 0.510 | 0.510 | 0.510 | 200.813 | 200.813 | 174.385 | 174.385 | 0.535 | 0.535 | 0.535 | 0 |
| 8.INF5 - частный критерий: ROI (Return On Investment); вероимо... | Сумма знания | 0.608 | 0.608 | 0.608 | 164.586 | 217.234 | 110.656 | 58.007 | 0.598 | 0.739 | 0.661 | 0 |
| 9.INF6 - частный критерий: равн.усли безул.вероятностей; вер... | Семантический резонанс зна... | 0.559 | 0.559 | 0.559 | 257.751 | 257.751 | 180.861 | 180.861 | 0.588 | 0.588 | 0.588 | 0 |
| 9.INF6 - частный критерий: равн.усли безул.вероятностей; вер... | Сумма знания | 0.559 | 0.559 | 0.559 | 257.751 | 257.751 | 180.861 | 180.861 | 0.588 | 0.588 | 0.588 | 0 |
| 10.INF7 - частный критерий: равн.усли безул.вероятностей; ве... | Семантический резонанс зна... | 0.559 | 0.559 | 0.559 | 257.751 | 257.751 | 180.861 | 180.861 | 0.588 | 0.588 | 0.588 | 0 |
| 10.INF7 - частный критерий: равн.усли безул.вероятностей; ве... | Сумма знания | 0.559 | 0.559 | 0.559 | 167.824 | 182.347 | 168.600 | 77.111 | 0.499 | 0.685 | 0.577 | 0 |
| 10.INF7 - частный критерий: равн.усли безул.вероятностей; ве... | Семантический резонанс зна... | 0.559 | 0.559 | 0.559 | 257.751 | 257.751 | 180.861 | 180.861 | 0.588 | 0.588 | 0.588 | 0 |
| 10.INF7 - частный критерий: равн.усли безул.вероятностей; ве... | Сумма знания | 0.559 | 0.559 | 0.559 | 167.824 | 182.347 | 168.600 | 77.111 | 0.499 | 0.685 | 0.577 | 0 |

Помощь по мерам достоверности | Помощь по частотным распределениям | TR|TN|FP|FN | (TP|FP)/(TN|FN) | (T-F)/(T+F)*100 | Задать интервал сглаживания

Рисунок 12. Экранные формы режима измерения достоверности моделей 3.4

На рисунках 13 приведены частотные распределения количества истинных и ложных, положительных и отрицательных решений в наиболее достоверной по L1-мере проф.Е.В.Луценко СК-модели ABS.



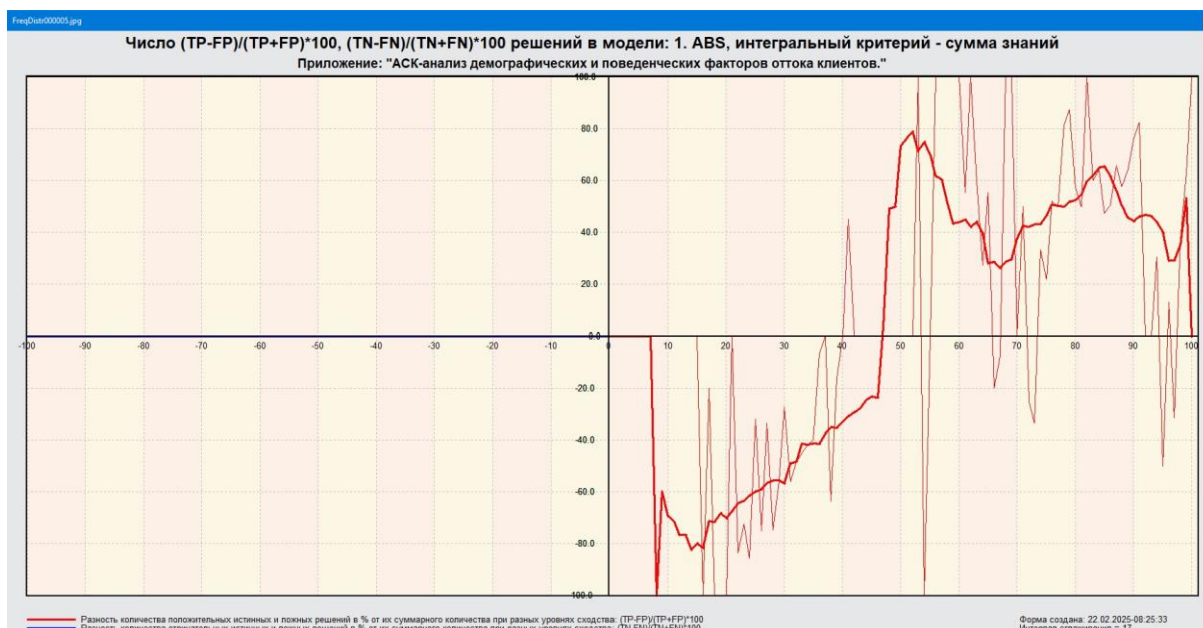


Рисунок 13. Частотные распределения количества истинных и ложных, положительных и отрицательных решений в наиболее достоверной по L1-мере проф.Е.В.Луценко СК-модели ABS

Из этих частотных распределений видно, что в наиболее достоверной по критерию достоверности L1-мерой проф.Е.В.Луценко СК-модели ABS:

- График показывает, что количество отрицательных ложных решений (FN) практически равно нулю на всех уровнях сходства. Это подтверждает, что в данной задаче отрицательные ложные решения не встречаются или встречаются крайне редко;

- с увеличением уровня сходства, доля истинных положительных решений (TP) также возрастает. Это подтверждает, что уровень сходства является адекватной внутренней мерой системы «Эйдос» для оценки достоверности решений;

- высокий уровень сходства коррелирует с высокой долей истинных положительных решений, что делает его надежным индикатором для самооценки и аудита степени достоверности решений и уровня риска ошибочного решения;

- ***чем выше уровень сходства, тем больше доля истинных решений. Поэтому уровень сходства является адекватной внутренней мерой системы «Эйдос», так сказать адекватной мерой самооценки или аудита степени достоверности решений и уровня риска ошибочного решения.***

На рисунках 14 приведены экранные формы хелпов режима 3.4, в которых подробно объясняется смысл этого режима. Эти формы приводятся в данной работе вместо более детального описания данного режима.

Помощь по режимам: 3.4, 4.1.3.#: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++"

ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Предположим, модель дает такой прогноз, что выпадет все: и 1, и 2, и 3, и 4, и 5, и 6. Понятно, что из всего этого выпадет лишь что-то одно. В этом случае модель не предскажет, что не выпадет, но зато она обязательно предскажет, что выпадет. Однако при этом очень много объектов будет отнесено к классам, к которым они не относятся. Тогда вероятность истинно-положительных решений у модели будет 1/6, а вероятность ложно-положительных решений - 5/6. Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.

ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Представим себе, что мы выбрасываем кубик с 6 гранями, и модель предсказывает, что ничего не выпадет, т.е. не выпадет ни 1, ни 2, ни 3, ни 4, ни 5, ни 6, но что-то из этого, естественно, обязательно выпадет. Конечно, модель не предсказала, что выпадет, зато она очень хорошо предсказала, что не выпадет. Вероятность истинно-отрицательных решений у модели будет 5/6, а вероятность ложно-отрицательных решений - 1/6. Такой прогноз гораздо достовернее, чем положительный псевдопрогноз, но тоже бесполезен.

ИДЕАЛЬНЫЙ ПРОГНОЗ.
Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.

РЕАЛЬНЫЙ ПРОГНОЗ.
На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществятся один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представьте себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В. Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {-1, +1} и {0, 1}:

$$L_a = \frac{TP + TN - FP - FN}{TP + TN + FP + FN} \quad (\text{нормировка: } \{-1, +1\})$$

$$L_b = \frac{1 + (TP + TN - FP - FN)}{(TP + TN + FP + FN)} / 2 \quad (\text{нормировка: } \{0, 1\})$$

где количество: TP - истинно-положительных решений; TN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;

Классическая F-мера достоверности моделей Ван Ризбергера (колонка выделена ярко-голубым фоном):
 $F\text{-мера} = 2(Precision \cdot Recall) / (Precision + Recall)$ - достоверность модели
 $Precision = TP / (TP + FP)$ - точность модели;
 $Recall = TP / (TP + FN)$ - полнота модели;

L1-мера проф. Е.В. Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СУММ уровней сходства (колонка выделена ярко-зеленым фоном):
 $L1\text{-мера} = 2(SPrecision \cdot SRecall) / (SPrecision + SRecall)$
 $SPrecision = STP / (STP + SFP)$ - точность с учетом сумм уровней сходства;
 $SRecall = STP / (STP + SFN)$ - полнота с учетом сумм уровней сходства;
 STP - Сумма модулей сходства истинно-положительных решений; SFN - Сумма модулей сходства истинно-отрицательных решений;
 SFP - Сумма модулей сходства ложно-положительных решений; SFN - Сумма модулей сходства ложно-отрицательных решений.

L2-мера проф. Е.В. Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желтым фоном):
 $L2\text{-мера} = 2(APrecision \cdot ARecall) / (APrecision + ARecall)$
 $APrecision = ATP / (ATP + AFP)$ - точность с учетом средних уровней сходства;
 $ARecall = ATP / (ATP + AFN)$ - полнота с учетом средних уровней сходства;
 ATP-SFP/TP - Среднее модулей сходства истинно-положительных решений; AFN-SFN/FN - Среднее модулей сходства истинно-отрицательных решений;
 AFP-SFP/FP - Среднее модулей сходства ложно-положительных решений; AFN-SFN/FN - Среднее модулей сходства ложно-отрицательных решений.

Строки с максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке.

Из графиков частотных распределений истинно-положительных, истинно-отрицательных, ложно-положительных и ложно-отрицательных решений видно, что чем выше модуль уровня сходства, тем больше доля истинных решений. Это значит, что модуль уровня сходства является адекватной мерой степени истинности решения и степени уверенности системы в этом решении. Поэтому система "Эйдос" имеет адекватный критерий достоверности собственных решений, с помощью которого она может отфильтровать заведомо ложные решения.

Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергера в АСК-анализе и системе "Эйдос" / Е.В. Луценко // Политематический сетевой с электронной научный журнал Кубанского государственного аграрного университета [Научный журнал КубГАУ] [Электронный ресурс] - Краснодар: КубГАУ, 2017. - №02(126). С. 1 - 32. -IDA [article ID]: 1261702001. -Режим доступа: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 у.п.л.

Помощь по режиму 3.4. (С) Система "ЭЙДОС-Х++"

Режим: 4.1.3.11. РАСЧЕТ И ГРАФИЧЕСКАЯ ВИЗУАЛИЗАЦИЯ ЧАСТОТНЫХ РАСПРЕДЕЛЕНИЙ УРОВНЕЙ СХОДСТВА:

По нажатию кнопок: [TP,TN,FP,FN], [(TP-FP)/(TN-FN)], [(T-F)/(T+F)*100] отображаются графики частотных распределений для модели и интегрального критерия той строки, на которой в экранной форме 3.4 стоит курсор. По клику на кнопку: [(T-F)/(T+F)*100] выводятся графики частотных распределений: (TP-FP)/(TP+FP)*100 и (TN-FN)/(TN+FN)*100, где:
 TP-True-Positive; TN-True-Negative; FP-False Positive; FN-False-Negative, количество истинных и ложных положительных и отрицательных решений.

Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергера в АСК-анализе и системе "Эйдос" / Е.В. Луценко // Политематический сетевой с электронной научный журнал Кубанского государственного аграрного университета [Научный журнал КубГАУ] [Электронный ресурс] - Краснодар: КубГАУ, 2017. - №02(126). С. 1 - 32. -IDA [article ID]: 1261702001. -Режим доступа: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 у.п.л.

Примерные графики TP,TN,FP,FN, а также F-меры и критериев L1, L2 при увеличении объема выборки:

Количество решений

Объем выборки

Динамика F-меры Ван Ризбергера и L1, L2 мер проф.Е.В.Луценко в зависимости от объема выборки RND-модели

Достоверность, %

Число объектов обучающей выборки

— L2_MERA
— L1_MERA
— F_MERA

Рисунок 14. Экранные формы хелпов режима измерения достоверности моделей

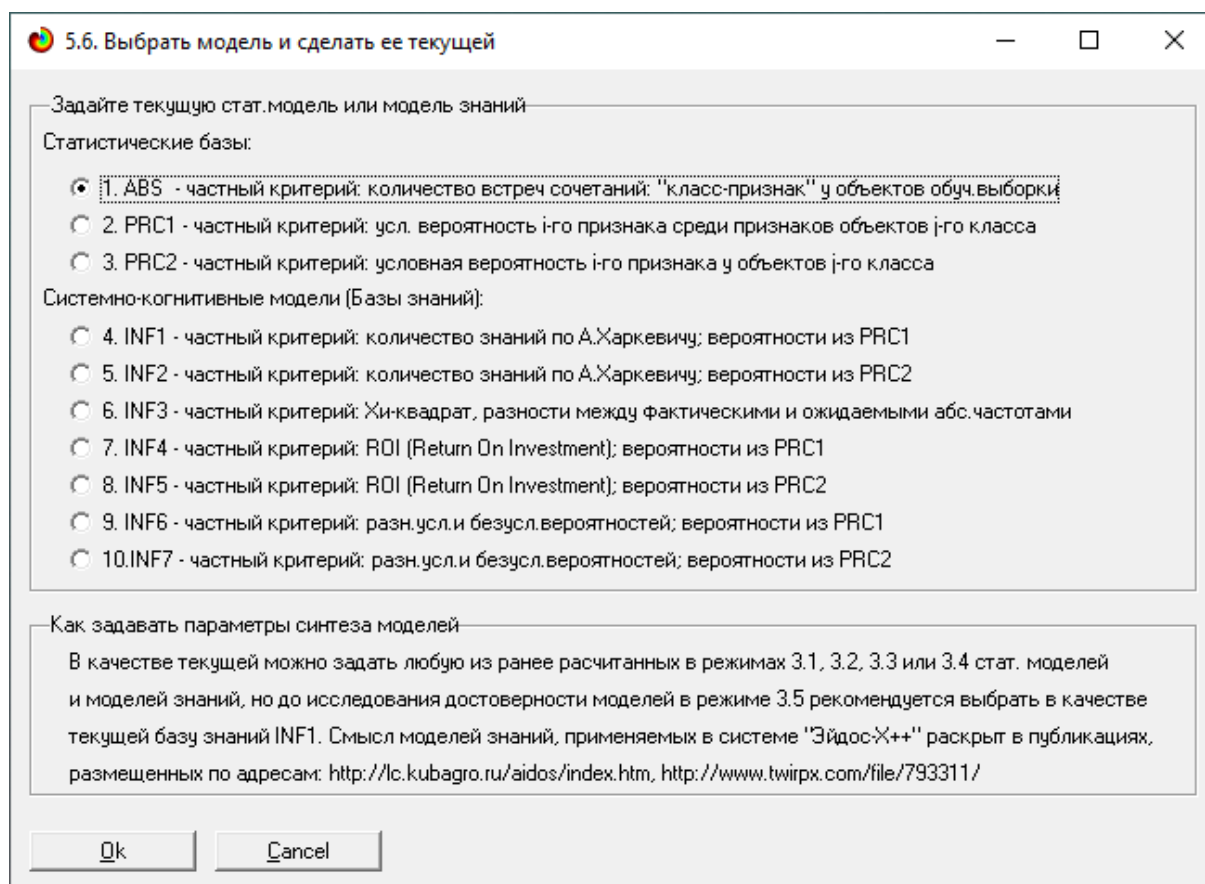
3.5. Задача-5. Выбор наиболее достоверной модели

Все последующие задачи решаются в наиболее достоверной модели.

Причины этого просты. Если модель достоверна, то:

- идентификация объекта с классом достоверна, т.е. модель относит объекты к классам, к которым они действительно принадлежат;
- прогнозирование достоверно, т.е. действительно наступают те события, которые прогнозируются;
- принятие решений адекватно (достоверно), т.е. после реализации принятых управляющих решений объект управления действительно переходит в целевые будущие состояния;
- исследование достоверно, т.е. полученные в результате исследования модели объекта моделирования выводы могут быть с полным основанием отнесены к объекту моделирования.

Технически сам выбор наиболее достоверной модели и задание ее текущей осуществляется в режиме 5.6 системы «Эйдос» и проходит быстро (рисунки 15). Это необходимо делать лишь для решения задачи идентификации и прогнозирования (в режиме 4.1.2), которая требует и потребляет наибольшие вычислительные ресурсы и поэтому решается только для модели, заданной в качестве текущей. Все остальные расчеты проводятся в системе «Эйдос» сразу во всех моделях.



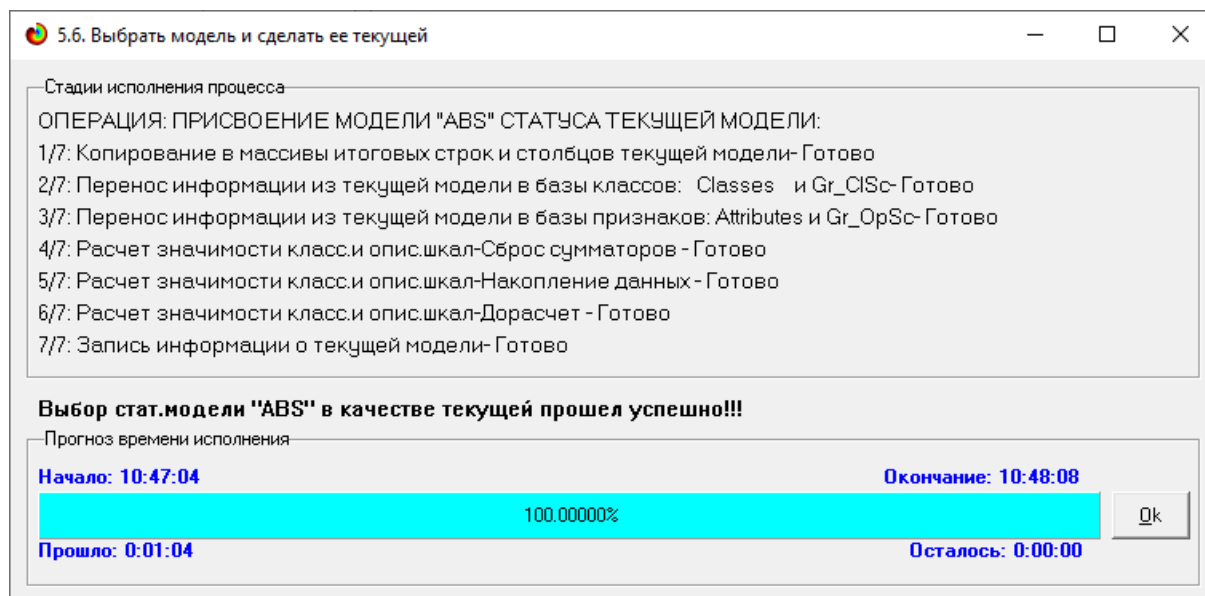


Рисунок 15. Задание СК-модели ABS в качестве текущей

3.6. Задача-6. Системная идентификация и прогнозирование

При решении задачи идентификации каждый объект распознаваемой выборки сравнивается по всем своим признакам с каждым из обобщенных образов классов. Смысл решения задачи идентификации заключается в том, что при определении принадлежности конкретного объекта к обобщенному образу классу об этом конкретном объекте *по аналогии становится известно все, что известно об объектах этого класса, по крайней мере, самое существенное о них, т.е. чем они отличаются от объектов других классов.*

Задачи идентификации и прогнозирования взаимосвязаны и мало чем отличаются друг от друга. Главное различие между ними в том, что при идентификации значения свойств и принадлежность объекта к классу относятся к одному моменту времени, а при прогнозировании значения факторов относятся к прошлому, а переход объекта под действием этих факторов в состояние, соответствующее классу относится к будущему (рисунок 3).

Задача решается в модели, заданной в качестве текущей, т.к. является весьма трудоемкой в вычислительном отношении. Правда с разработкой и реализацией в системе «Эйдос» высокоэффективных алгоритмов распознавания и использованием графического процессора (GPU) для расчетов эта проблема практически снялась.

Сравнение осуществляется путем применения *неметрических интегральных критериев*, которых в настоящее время в системе «Эйдос» используется два. Эти интегральные критерии интересны тем, что корректны¹⁰ в неортонормированных пространствах, которые всегда и встречаются на практике, и являются фильтрами подавления шума.

¹⁰ В отличие от Евклидова расстояния, которое используется для подобных целей наиболее часто

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:

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

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;

σ_j – среднеквадратичное отклонение частных критериев знаний вектора класса; σ_l – среднеквадратичное отклонение по вектору распознаваемого объекта.

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

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

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).

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

$$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}, \quad L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_l}.$$

Поэтому по своей сути он также является скалярным

произведением двух стандартизированных (единичных) векторов класса и объекта. Существуют и много других способов нормирования, например, путем применения вейвлетов и сплайнов, в частности линейной интерполяции: $I_{ij} \rightarrow \frac{I_{ij} - I_j^{\min}}{I_j^{\max} - I_j^{\min}}, \quad L_i \rightarrow \frac{L_i - L^{\min}}{L^{\max} - L^{\min}}$. 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».

3.6.3. Важные математические свойства интегральных критериев

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

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

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.

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.

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.

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.

3.6.4. Конкретное решение задачи идентификации и прогнозирования в данной работе в системе «Эйдос»

В АСК-анализе разработаны, а в системе «Эйдос» реализованы развитые методы прогнозирования, основанные на сценарном методе АСК-анализа или сценарном АСК-анализе. Но в задачи данной работы не входит их подробное рассмотрение, тем более, что они подробно

освящены и на теоретическом уровне, и с детальными численными примерами в работах и в ряде других.

Поэтому в данной работе рассмотрим стандартный несколько упрощенный вариант решения задачи идентификации и прогнозирования. Запустим режим 4.1.2 системы «Эйдос» (рисунок 16).

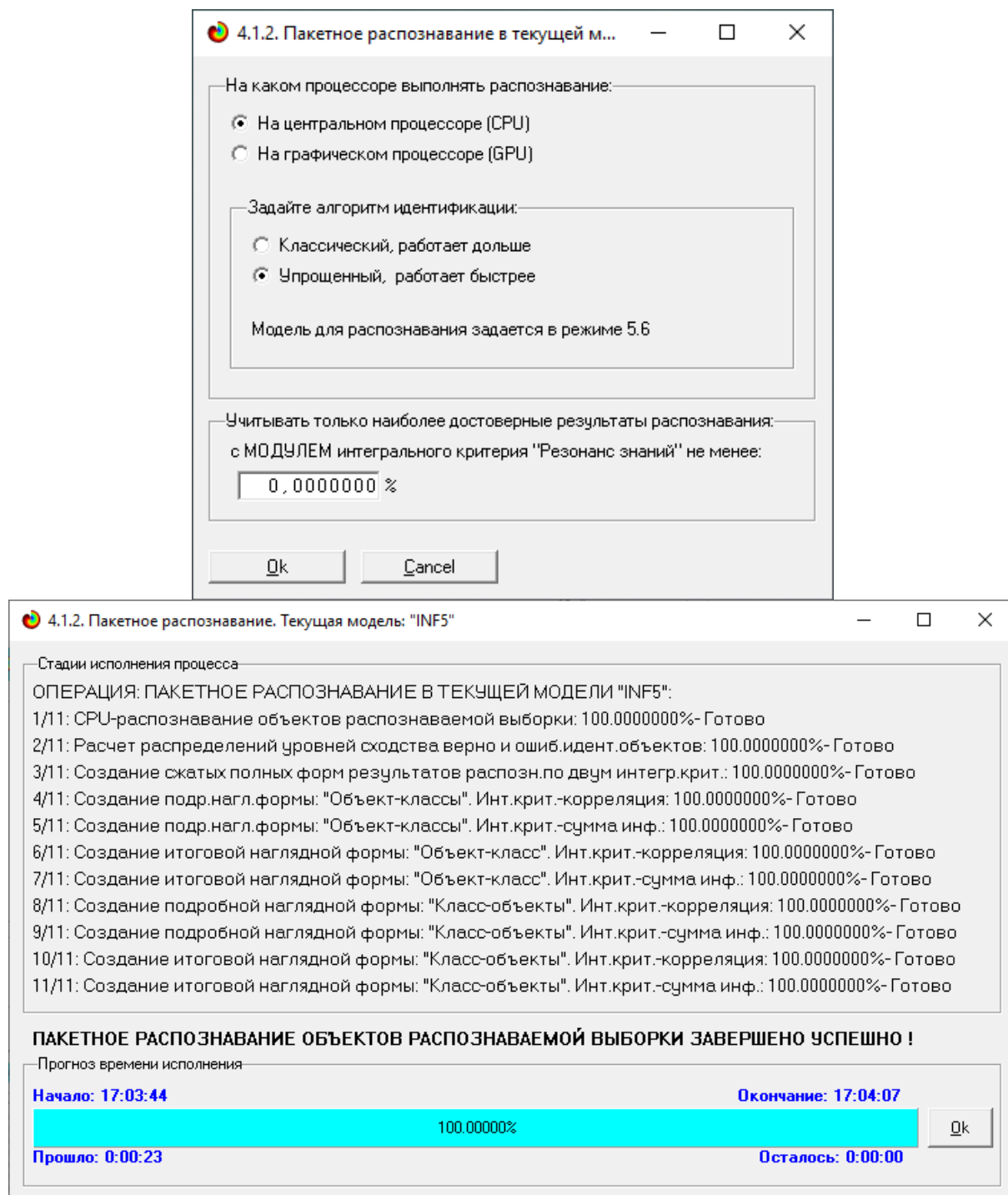


Рисунок 16. Экранные формы режима 4.1.2 идентификации и прогнозирования

По результатам решения задачи идентификации и прогнозирования в системе «Эйдос» выдается большое количество разнообразных выходных форм, которых в настоящее время 14 (рисунок 17).

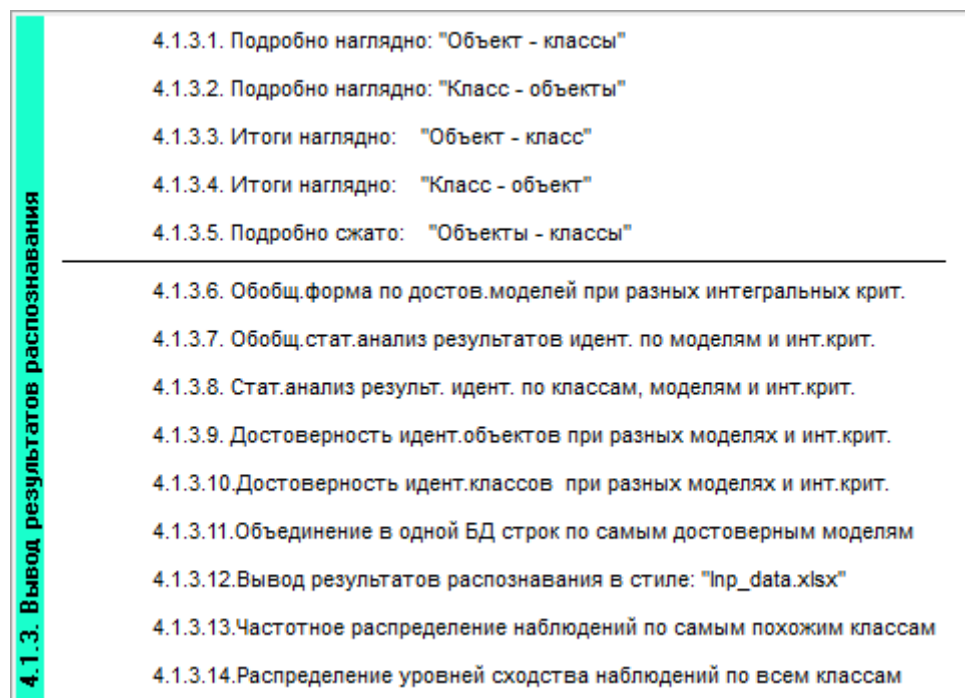


Рисунок 17. Меню подсистемы 4.1.3 системы «Эйдос», обеспечивающей вывод форм по результатам решения задачи идентификации и прогнозирования

Из этих выходных форм рассмотрим только две: 4.1.3.1 и 4.1.3.2 (рисунок 18).

4.1.3.1. Визуализация результатов распознавания в отношении: "Объект-классы", Текущая модель: "ABS"

| Расознаваемые объекты | |
|-----------------------|--------------|
| Код | Наим объекта |
| 1 | 7590-VHVEG |
| 2 | 5575-GNVDE |
| 3 | 3668-QPYBK |
| 4 | 7795-CFDCW |
| 5 | 9237-HQITU |
| 6 | 9305-CDSKC |
| 7 | 1452-KIOVK |
| 8 | 6713-OKDMC |
| 9 | 7892-POOKP |
| 10 | 6388-TABGU |
| 11 | 9763-GRSKD |
| 12 | 7469-LKBCI |
| 13 | 8091-TTVAX |
| 14 | 0280-XJGEX |
| 15 | 5129-JLPIS |
| 16 | 3655-SNQYZ |
| 17 | 8191-XWSZG |
| 18 | 9959-WDFKT |
| 19 | 4190-MFLUW |
| 20 | 4183-MYFRB |
| 21 | 8779-QRDMV |
| 22 | 1680-VDcww |
| 23 | 1066-JKSGK |

| Интегральный критерий сходства: "Семантический резонанс знаний" | | | | |
|---|---------------------|----------|------|----------------------|
| Код | Наименование класса | Сходство | Ф... | Сходство |
| 2 | CHURN-Yes | 79,22... | | ████████████████████ |
| 1 | CHURN-No | 74,58... | v | ████████████████████ |

| Интегральный критерий сходства: "Сумма знаний" | | | | |
|--|---------------------|----------|------|----------------------|
| Код | Наименование класса | Сходство | Ф... | Сходство |
| 1 | CHURN-No | 82,17... | v | ████████████████████ |
| 2 | CHURN-Yes | 32,41... | | ████████████████████ |

Помощь | 9 классов | Классы с MaxMin УрСх | 9 классов с MaxMin УрСх | ВСЕ классы | ВКЛ. фильтр по класс. шкале | ВЫКЛ. фильтр по класс. шкале | Граф. диаграммы

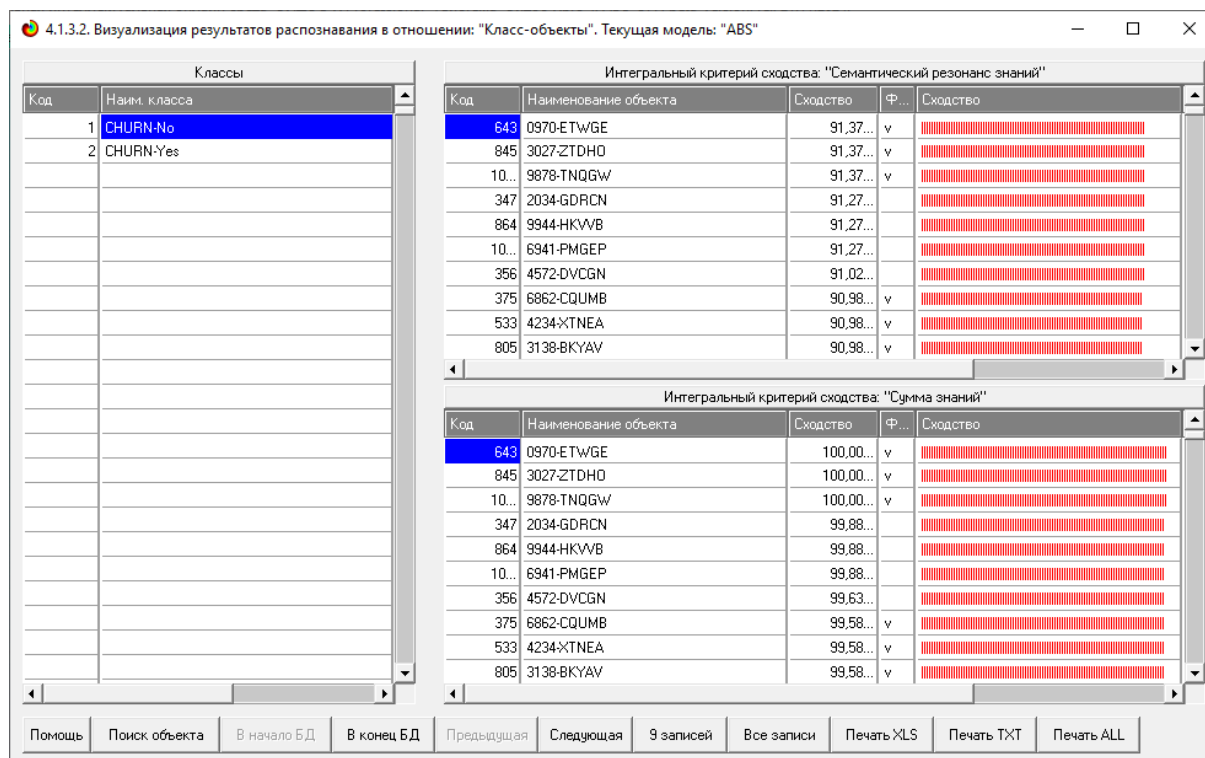


Рисунок 18. Некоторые экранные формы результатов идентификации и прогнозирования 4.1.3 системы «Эйдос»

Эти выходные формы, учитывая сказанное выше об интегральных критериях системы «Эйдос», как говорят «интуитивно понятны» и не требуют особых комментариев.

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

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

Таким образом, задача принятия решений является *обратной* по отношению к задаче прогнозирования. Но это так только в простейшем случае: в случае использования SWOT-анализа (режим 4.4.8 системы «Эйдос») (рисунки 19).

На первом рисунке 19 приведена экранная форма задания в диалоге параметров отображения SWOT-диаграммы. На этой экранной форме в верхнем окне пользователь курсором выбирает исследуемый класс, внизу

слева задает модель для исследования, а справа внизу задает отображать ли SWOT-диаграмму. Кроме того, пользователь может включить или выключить фильтры по факторам и посмотреть помощь по режиму. При включении фильтра по фактору, на котором стоит курсор, на экранных формах отображается влияние только значения этого фактора.

Графические выходные формы, приведенные на рисунках 19, интуитивно понятны и не требуют особых комментариев. Отметим лишь, что на SWOT-диаграммах наглядно показаны знак и сила влияния каждого значения фактора на переход объекта моделирования в состояние, соответствующее классу, выбранному в верхнем окне. Знак показан цветом, а сила влияния – толщиной линии.

Слева на SWOT-диаграмме приведены значения факторов, способствующих переходу объекта моделирования в состояние, соответствующее классу, выбранному в верхнем окне (показаны красным цветом), а справа – препятствующих этому переходу (показаны синим цветом). Сила влияния каждого значения фактора на поведение объекта моделирования показана толщиной линии.

4.4.8. Количественный автоматизированный SWOT-анализ классов средствами АСК-анализа в системе "Эйдос"

Выбор класса, соответствующего будущему состоянию объекта управления

| Код | Наименование класса | Редукция клас... | N объектов [абс.] | N объектов [%] |
|-----|---------------------|------------------|-------------------|----------------|
| 1 | CHURN-No | 61,3747767 | 12896 | 73,6073059 |
| 2 | CHURN-Yes | 22,8760842 | 4624 | 26,3926941 |

SWOT-анализ класса:1 "CHURN-No" в модели:1 "ABS"

Способствующие факторы и сила их влияния

| Код | Наименование фактора и его интервального значения | Сила влияния |
|-----|---|--------------|
| 10 | PHONESERVICE-Yes | 732.0... |
| 3 | SENIORCITIZEN-No | 683.0... |
| 7 | DEPENDENTS-No | 541.0... |
| 33 | PAPERLESSBILLING-Yes | 468.0... |
| 5 | PARTNER-No | 409.0... |
| 2 | GENDER-Male | 406.0... |
| 1 | GENDER-Female | 400.0... |
| 6 | PARTNER-Yes | 397.0... |
| 29 | CONTRACT-Month-to-month | 394.0... |
| 11 | MULTIPLELINES-No | 383.0... |
| 17 | ONLINESECURITY-No | 377.0... |
| 23 | TECHSUPPORT-No | 364.0... |
| 15 | INTERNETSERVICE-Fiber optic | 359.0... |
| 20 | DEVICERENTITION-No | 351.0... |

Препятствующие факторы и сила их влияния

| Код | Наименование фактора и его интервального значения | Сила влияния |
|-----|---|--------------|
| 0 | | |

ВКЛЮЧИТЬ фильтр по фактору ВЫКЛЮЧИТЬ фильтр по фактору ВКЛЮЧИТЬ фильтр по фактору ВЫКЛЮЧИТЬ фильтр по фактору

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 SWOT-диаграмма

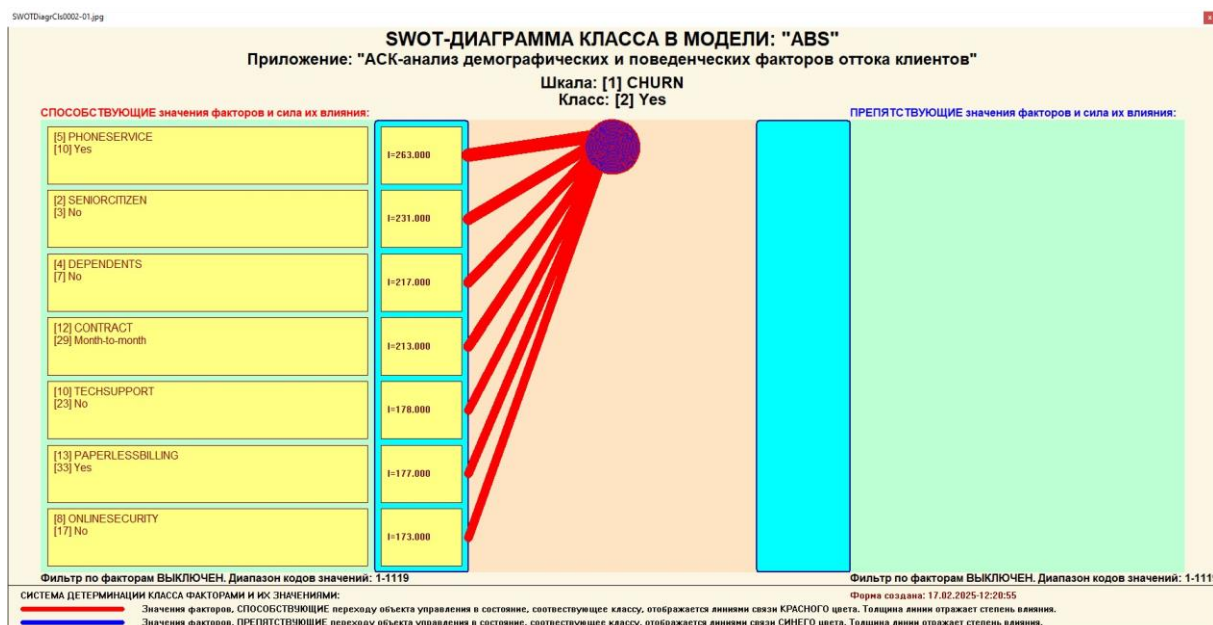
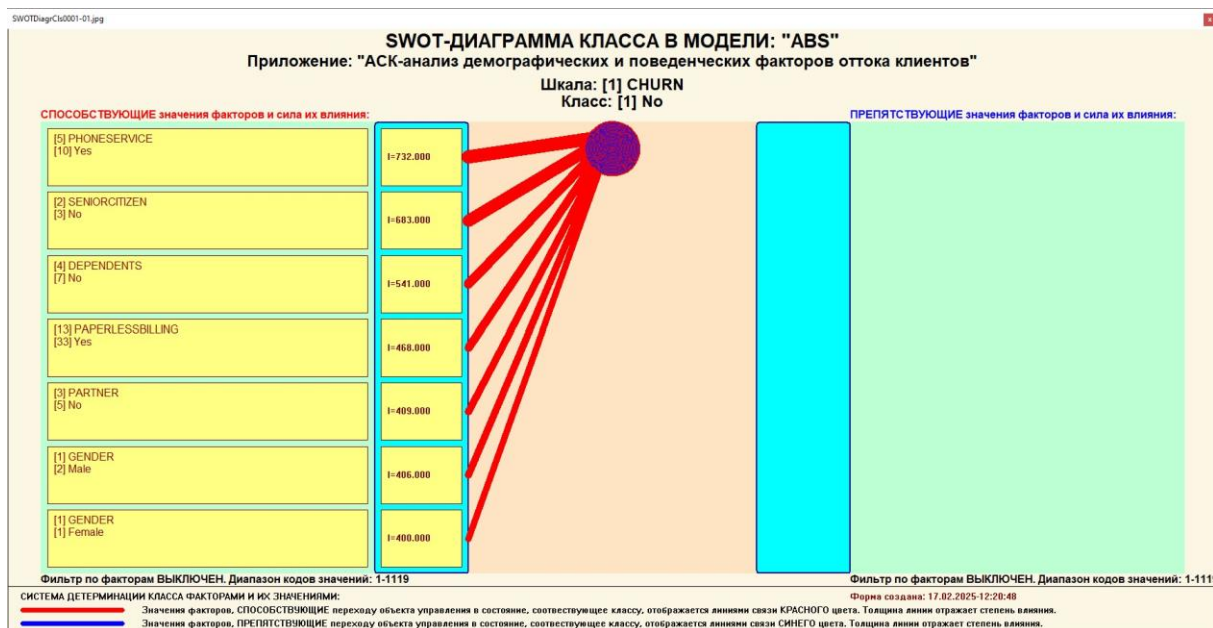


Рисунок 19. Примеры экранной формы режима автоматизированного SWOT-анализа (режим 4.4.8 системы «Эйдос»)

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

Однако SWOT-анализ (режим 4.4.8 системы «Эйдос») имеет свои ограничения: может быть задано только одно будущее целевое состояние, целевые состояния могут быть недостижимыми одновременно (альтернативными) или совместимыми по системе обуславливающих их значений факторов, некоторые рекомендуемые факторы может не быть технологической и финансовой возможности использовать и возможно

надо искать им замену, примерно так же влияющую на объект моделирования.

Поэтому в АСК-анализе и системе «Эйдос» реализован развитый алгоритм принятия решений (режим 6.3) в котором кроме SWOT-анализа используются также результаты решения задачи прогнозирования и результаты кластерно-конструктивного анализа классов и значений факторов, т.е. некоторые результаты решения задачи исследования предметной области. Этот алгоритм описан в работах и в ряде других работ.

Приведем этот алгоритм и в данной работе (рисунок 20).

Шаг 1-й. Руководство ставит **цели** управления, т.е. определяет будущие целевые состояния объекта управления. Обычно целевые состояния в натуральном выражении – это количество и качество продукции, а в стоимостном выражении – прибыль и рентабельность. Объект управления как система, эффективность объекта управления как **системное свойство**, повышение уровня системности объекта управления как цель управления (нелинейность). Модель отражает определенный уровень технологий, поэтому целевые состояния, недостижимые в одной модели, могут оказаться достижимыми в другой с большим числом факторов.

Шаг 2-й (см. реж.6.4). Когнитивно-целевая структуризация и формализация предметной области (реж.2.3.2.2), синтез и верификация моделей (реж.3.5), определяем наиболее достоверную из них по F-критерию Ван Ризбергена и критериям L1 и L2 проф.Е.В.Луценко (реж.3.4) [9]. Повышение уровня системности и адекватности модели объекта управления (принцип Уильяма Росса Эшби).

Шаг 3-й. Если целевое состояние одно, то переходим на шаг 6, а иначе на шаг 4.

Шаг 4-й. Иначе оцениваем **корректность** поставленных целей путем сравнения системы детерминации целевых состояний методом **когнитивной кластеризации** (4.2.2.3) или на основе матрицы сходства (4.2.2.1), т.е. определяем, являются ли целевые состояния совместимыми, т.е. достижимыми одновременно, по обуславливающим их значениями факторов, или они являются взаимоисключающими (альтернативными) по системе детерминации и одновременно недостижимы.

Шаг 5-й. Поставленные цели управления корректны, совместимы, достижимы одновременно? Если да – переход на шаг 6, иначе на шаг 1.

Шаг 6-й. Решаем задачу поддержки принятия решений в упрощенном варианте путем решения обратной задачи прогнозирования в автоматизированном **SWOT-анализе** (реж.4.4.8) для каждого из целевых состояний и **объединяем** рекомендованные значения факторов в одну систему управляющих факторов [10].

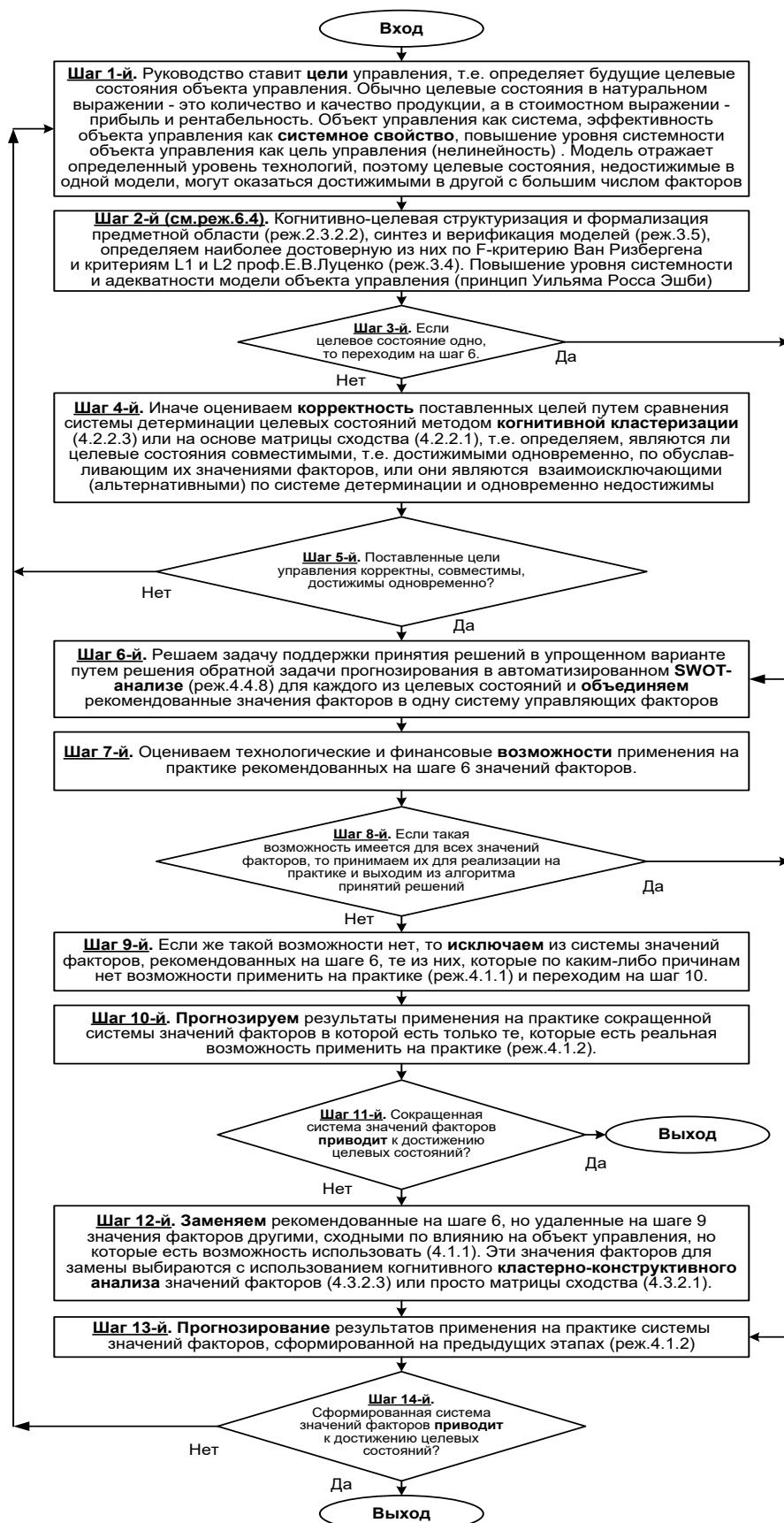


Рисунок 20. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»

Шаг 7-й. Оцениваем технологические и финансовые **возможности** применения на практике рекомендованных на шаге 6 значений факторов.

Шаг 8-й. Если такая возможность имеется для всех значений факторов, то принимаем их для реализации на практике и переходим на шаг 13 для проверки эффективности принятых решений, а иначе переходим на шаг 9.

Шаг 9-й. Если же такой возможности нет, то **исключаем** из системы значений факторов, рекомендованных на шаге 6, те из них, которые по каким-либо причинам нет возможности применить на практике (реж.4.1.1) и переходим на шаг 10.

Шаг 10-й. **Прогнозируем** результаты применения на практике сокращенной системы значений факторов в которой есть только те, которые есть реальная возможность применить на практике (реж.4.1.2).

Шаг 11-й. Сокращенная система значений факторов **приводит** к достижению целевых состояний? Если да, то выход из алгоритма принятия решений, а иначе переход на шаг 12.

Шаг 12-й. **Заменяем** рекомендованные на шаге 6, но удаленные на шаге 9 значения факторов другими, сходными по влиянию на объект управления, но которые есть возможность использовать (4.1.1). Эти значения факторов для замены выбираются с использованием когнитивного **кластерно-конструктивного анализа** значений факторов (4.3.2.3) или просто матрицы сходства (4.3.2.1).

Шаг 13-й. **Прогнозирование** результатов применения на практике системы значений факторов, сформированной на предыдущих этапах (реж.4.1.2)

Шаг 14-й. Сформированная система значений факторов **приводит** к достижению целевых состояний? Если да – то выход из алгоритма принятия решений, а иначе переход на шаг 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.

Как мы видим в развитом алгоритме принятия решений широко используются результаты конкретного решения различных задач: и задачи прогнозирования, и некоторых задач исследования объекта моделирования путем исследования его модели. Необходимо особо отметить, что система «Эйдос» поддерживает решение всех этих задач, которые необходимо решать в развитом алгоритме принятия решений. Потому ниже кратко рассмотрим решение этих и некоторых других задач.

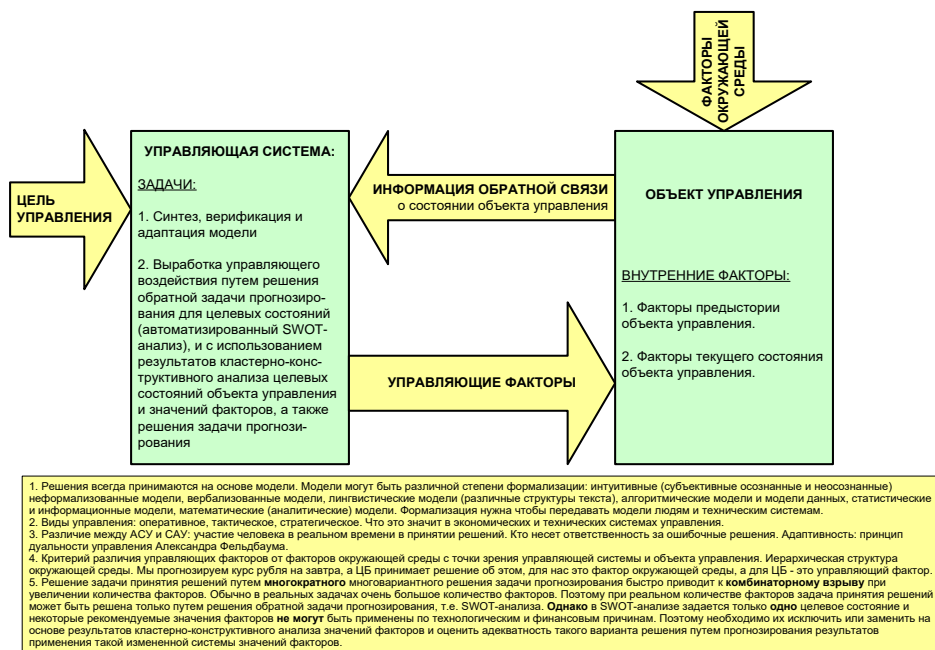


Рисунок 21. Принципиальная схема адаптивной интеллектуальной системы управления на основе АСК-анализа и системы «Эйдос»

Необходимо отметить, что модели системы Эйдос – это феноменологические модели, отражающие эмпирические закономерности в фактах обучающей выборки, т.е. они не отражают причинно-следственного *механизма* детерминации, а только сам факт и характер детерминации (силу и направление влияния факторов). *Содержательное* объяснение конкретных механизмов этих эмпирических закономерностей формулируется уже экспертами на теоретическом уровне познания в содержательных научных законах.

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

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

Инвертированные SWOT-диаграмм (предложены автором в работе), отражают силу и направление влияния конкретной градации описательной шкалы на переход объекта моделирования в состояния, соответствующие градациям классификационных шкал (классы). Это и есть *смысл* (семантический потенциал) этой градации описательной шкалы. Инвертированные SWOT-диаграммы выводятся в режиме 4.4.9 системы «Эйдос».

Примеры инвертированных SWOT-диаграмм приведены на рисунках 22 для некоторых значений факторов:

4.4.9 Количественный автоматизированный SWOT-анализ значений факторов средствами АСК-анализа в системе "Эйдос"

Выбор значения фактора, оказывающего влияние на переход объекта управления в будущие состояния

| Код | Наименование значения фактора |
|-----|-------------------------------|
| 1 | GENDER-Female |
| 2 | GENDER-Male |
| 3 | SENIORCITIZEN-No |
| 4 | SENIORCITIZEN-Yes |
| 5 | PARTNER-No |
| 6 | PARTNER-Yes |

SWOT-анализ значения фактора: 1 "GENDER-Female" в модели: 1 "ABS"

СПОСОБСТВУЕТ:

| Код | Состояния объекта управления, переходу в которые данное значение фактора СПОСОБСТВУЕТ | Сила влияния |
|-----|---|--------------|
| 1 | CHURN-No | 400... |
| 2 | CHURN-Yes | 139... |

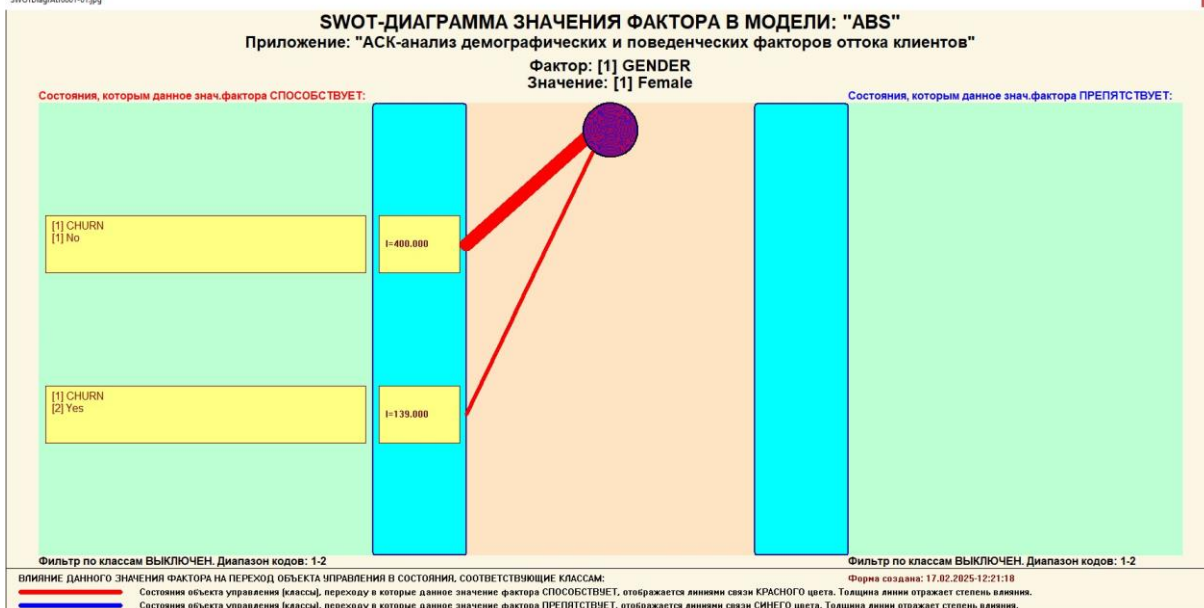
ПРЕПЯТСТВУЕТ:

| Код | Состояния объекта управления, переходу в которые данное значение фактора ПРЕПЯТСТВУЕТ | Сила влияния |
|-----|---|--------------|
| 0 | | |

ВКЛЮЧИТЬ фильтр по кл.шкале ВЫКЛЮЧИТЬ фильтр по кл.шкале

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 SWOT-диаграмма

SWOTdiagAt0001-01.jpg



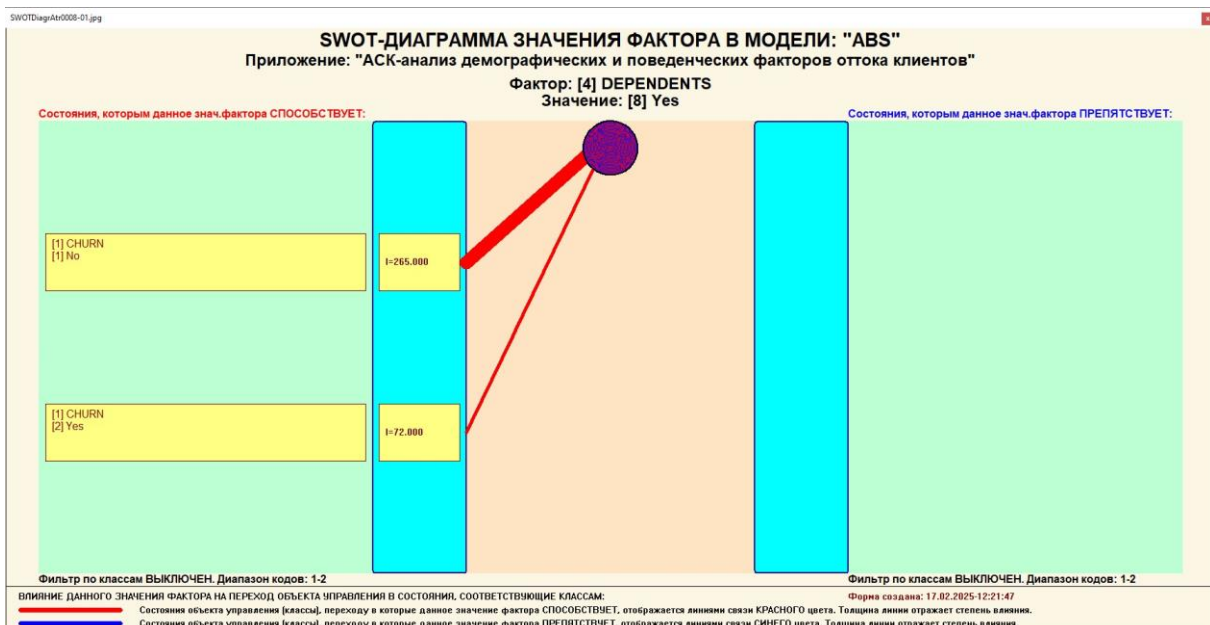
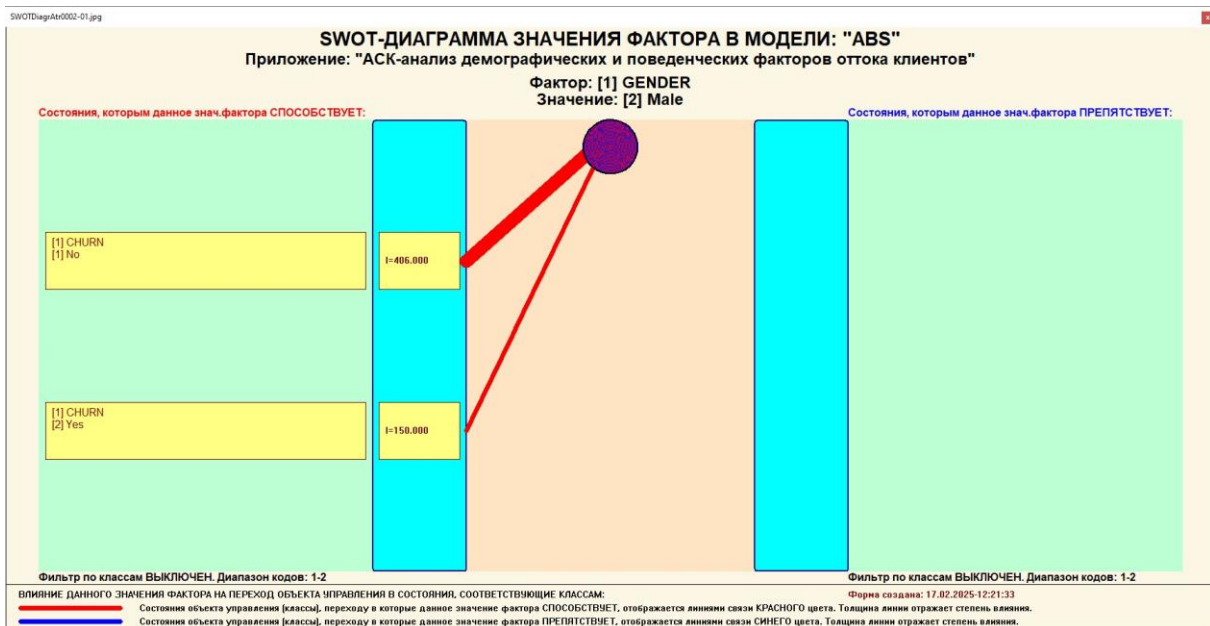


Рисунок 22. Примеры инвертированных SWOT-диаграмм влияния значений факторов на переход объекта моделирования в будущие состояния, соответствующие классам

Приведенные на рисунке 22 инвертированные SWOT-диаграммы исчерпывающим образом отражают силу и направление влияния каждого значения каждого фактора на переход объекта моделирования в состояния, соответствующие различным классам.

3.8.2. Кластерно-конструктивный анализ классов

В системе «Эйдос» (в режиме 4.2.2.1, рисунок 23) рассчитывается матрица сходства классов (таблица 11) по системе их детерминации и на основе этой матрицы рассчитывается и выводится три основных формы:

– круговая 2d-когнитивная диаграмма классов (режим 4.2.2.2) (рисунок 24);

Эта матрица сходства (таблица 11) используются и при расчете некоторых других выходных форм.

На рисунке 23 представлены экранные формы режима 4.2.2.1, обеспечивающего расчет матрицы сходства классов по системе их детерминации, т.е. по обуславливающим их значениям факторов:

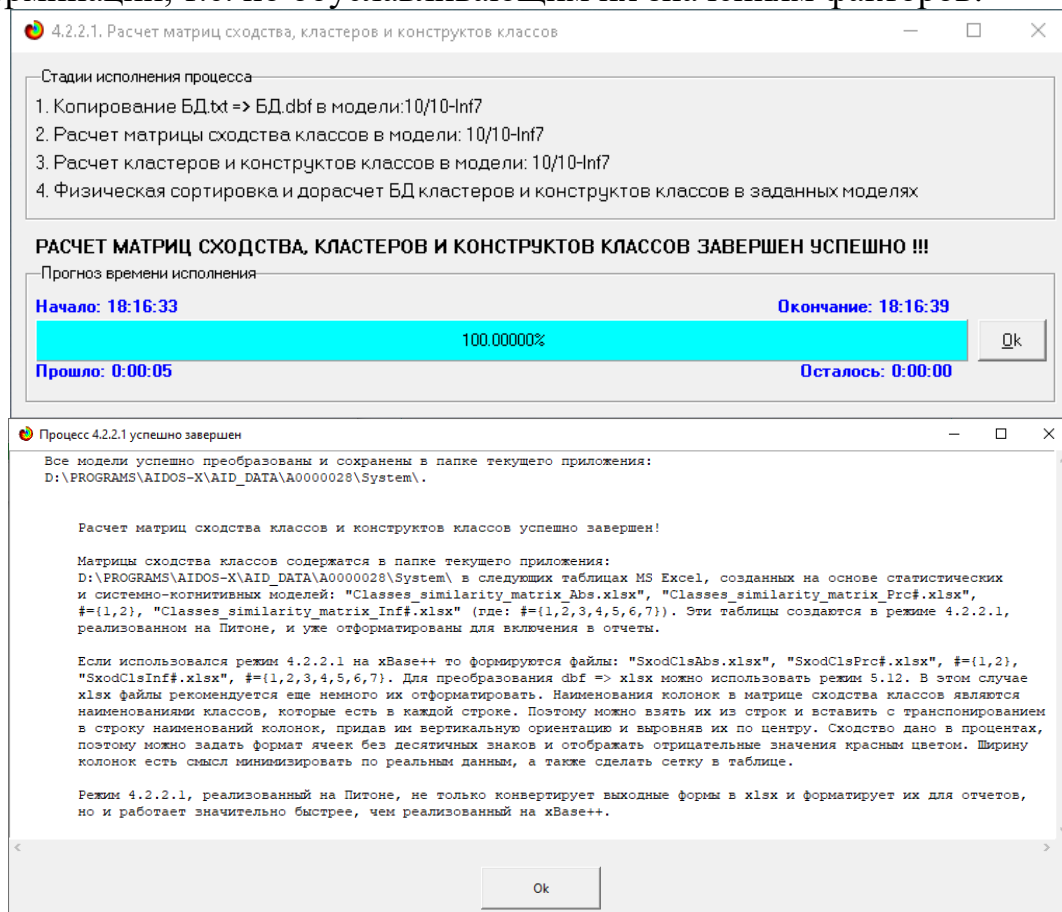


Рисунок 23. Экранные формы режима 4.2.2.1, обеспечивающего расчет матрицы сходства классов

Таблица 11 – Матрица сходства классов в СК-модели ABS (полностью)

| | 1-CHURN-No | 2-CHURN-Yes |
|-------------|------------|-------------|
| 1-CHURN-No | 100 | 98,20003 |
| 2-CHURN-Yes | 98,20003 | 100 |

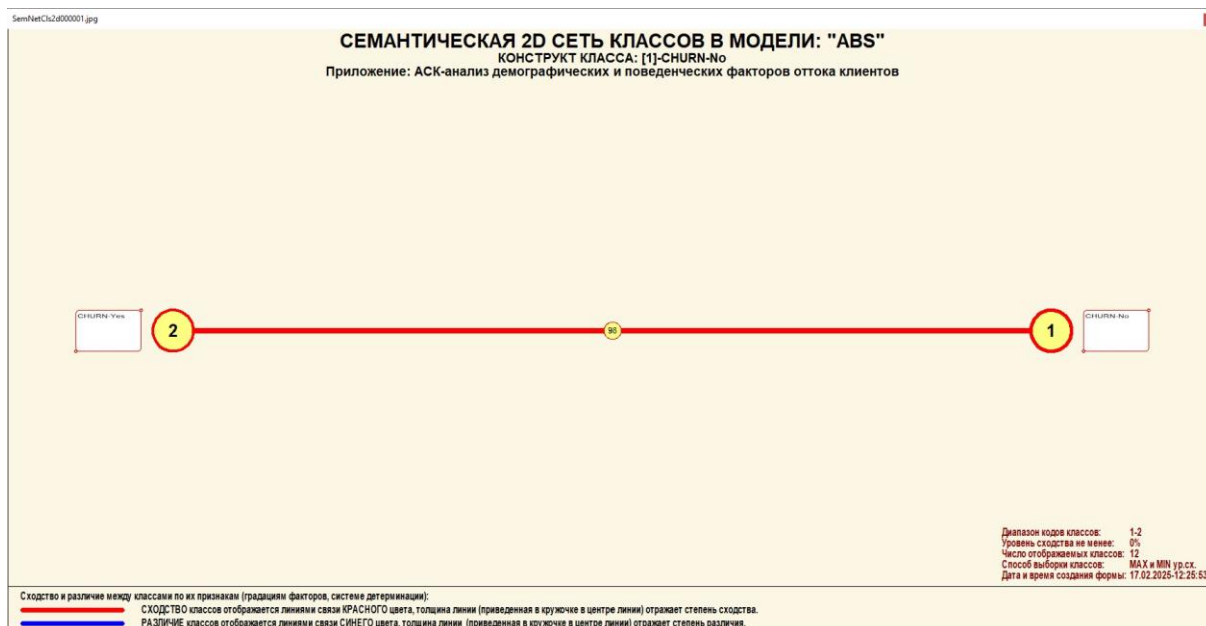


Рисунок 24. Круговая 2d-когнитивная диаграмма классов (режим 4.2.2.2)

В системе «Эйдос» (в режиме 4.3.2.1, рисунок 25) рассчитывается матрица сходства признаков (таблица 12) по их смыслу и на основе этой матрицы рассчитывается и выводится четыре основных формы:

– круговая 2d-когнитивная диаграмма признаков (режим 4.3.2.2) рисунок 26);

Эта матрица сходства (таблица 12) используются и при расчете некоторых других выходных форм.

На рисунке 25 представлены экранные формы режима 4.3.2.1, обеспечивающего расчет матрицы сходства значений факторов по силе и направлению их влияния на переход объекта моделирования в различные будущие состояния, соответствующие классам:

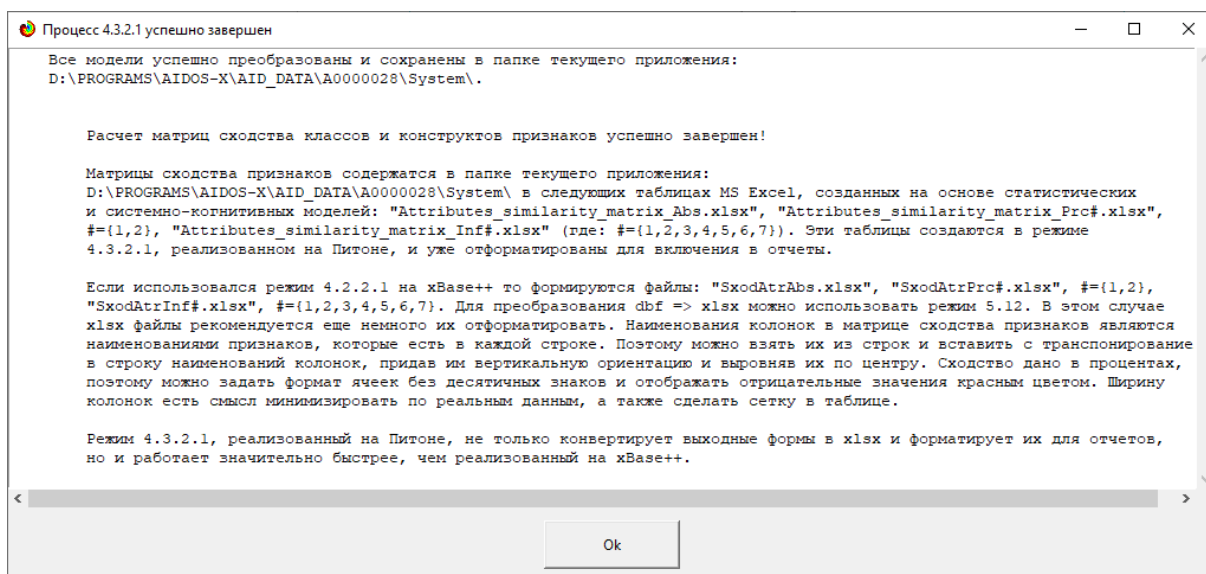


Рисунок 25. Экранные формы режима 4.3.2.1, обеспечивающего расчет матриц сходства значений факторов

Таблица 12 – Матрица сходства значений факторов в СК-модели ABS (фрагмент)

| | 1 - GENDER-Female | 2 - GENDER-Male | 3 - SENIORCITIZEN-No | 4 - SENIORCITIZEN-Yes | 5 - PARTNER-No | 6 - PARTNER-Yes | 7 - DEPENDENTS-No |
|-----------------------|-------------------|-----------------|----------------------|-----------------------|----------------|-----------------|-------------------|
| 1 - GENDER-Female | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 2 - GENDER-Male | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3 - SENIORCITIZEN-No | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 4 - SENIORCITIZEN-Yes | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 5 - PARTNER-No | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

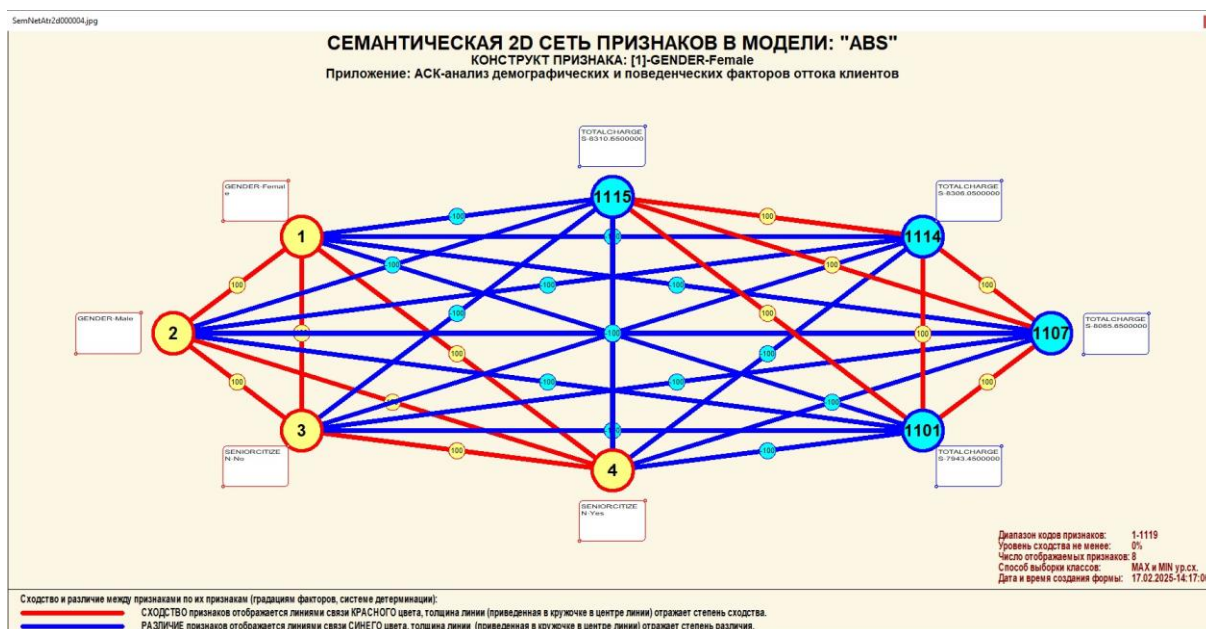


Рисунок 26. Круговая 2d-когнитивная диаграмма значений факторов в СК-модели ABS (режим 4.3.2.2)

3.8.4. Модель знаний системы «Эйдос» и нелокальные нейроны

Модель знаний системы «Эйдос» относится к **нечетким декларативным гибридным** моделям и объединяет в себе некоторые положительные особенности нейросетевой и фреймовой моделей представления знаний.

Классы в этой модели соответствуют нейронам и фреймам, а признаки рецепторам и шпациям (описательные шкалы – слотам).

От фреймовой модели представления знаний модель системы «Эйдос» отличается своей эффективной и простой программной реализацией, полученной за счет того, что разные фреймы отличаются друг от друга не набором слотов и шпаций, а лишь информацией в них.

Поэтому в системе «Эйдос» при увеличении числа фреймов само количество баз данных не увеличивается, а увеличивается лишь их размерность. Это является очень важным свойством моделей системы «Эйдос», существенно облегчающим и упрощающим программную реализацию и быстрое действие системы.

От нейросетевой модели представления знаний модель системы «Эйдос» отличается тем, что:

1) весовые коэффициенты на рецепторах не подбираются итерационным методом обратного распространения ошибки, а рассчитываются методом прямого счета на основе хорошо теоретически обоснованной модели, основанной на **теории информации** (это напоминает байесовские сети);

2) весовые коэффициенты имеют хорошо теоретически обоснованную **содержательную интерпретацию**, основанную на теории информации;

3) нейросеть является **нелокальной**, как сейчас говорят «полносвязной».

В системе «Эйдос» нелокальные нейроны визуализируются (режим 4.4.10 системы «Эйдос») в виде специальных графических форм, на которых сила и направление влияния рецепторов нейрона на степень его активации/торможения, которые отображаются соответственно в форме цвета и толщины дендрита (рисунок 27). В форме управления визуализацией есть возможность задавать фильтры по факторам, которые надо визуализировать.

4.4.10.Графическое отображение нелокального нейрона в системе "Эйдос"

Выбор нелокального нейрона (класса) для визуализации

| Код | Наименование нелокального нейрона (класса) |
|-----|--|
| 1 | CHURN-No |
| 2 | CHURN-Yes |

Подготовка визуализации нейрона: 1 "CHURN-No" в модели: 1 "ABS"

АКТИВИРУЮЩИЕ рецепторы и сила их влияния

| Код | Наименование фактора и его интервального значения | Сила влияния |
|-----|---|--------------|
| 10 | PHONESERVICE-Yes | 732.0... |
| 3 | SENIORCITIZEN-No | 683.0... |
| 7 | DEPENDENTS-No | 541.0... |
| 33 | PAPERLESSBILLING-Yes | 468.0... |
| 5 | PARTNER-No | 409.0... |
| 2 | GENDER-Male | 406.0... |
| 1 | GENDER-Female | 400.0... |
| 6 | PARTNER-Yes | 397.0... |
| 29 | CONTRACT-Month-to-month | 394.0... |
| 11 | MULTIPURPOSE... | 300.0... |

ТОРМОЗЯЩИЕ рецепторы и сила их влияния

| Код | Наименование фактора и его интервального значения | Сила влияния |
|-----|---|--------------|
| 0 | | |

ВКЛЮЧИТЬ фильтр по фактору ВЫКЛЮЧИТЬ фильтр по фактору

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7

НЕЙРОН Максимальное количество отображаемых рецепторов: 999
Минимальный вес.коэф. отображаемых рецепторов: 0,000

Сортировать рецепторы:
 по информативности
 по модулю информативности

Отображать рецепторы:
 с наименованиями
 только с кодами

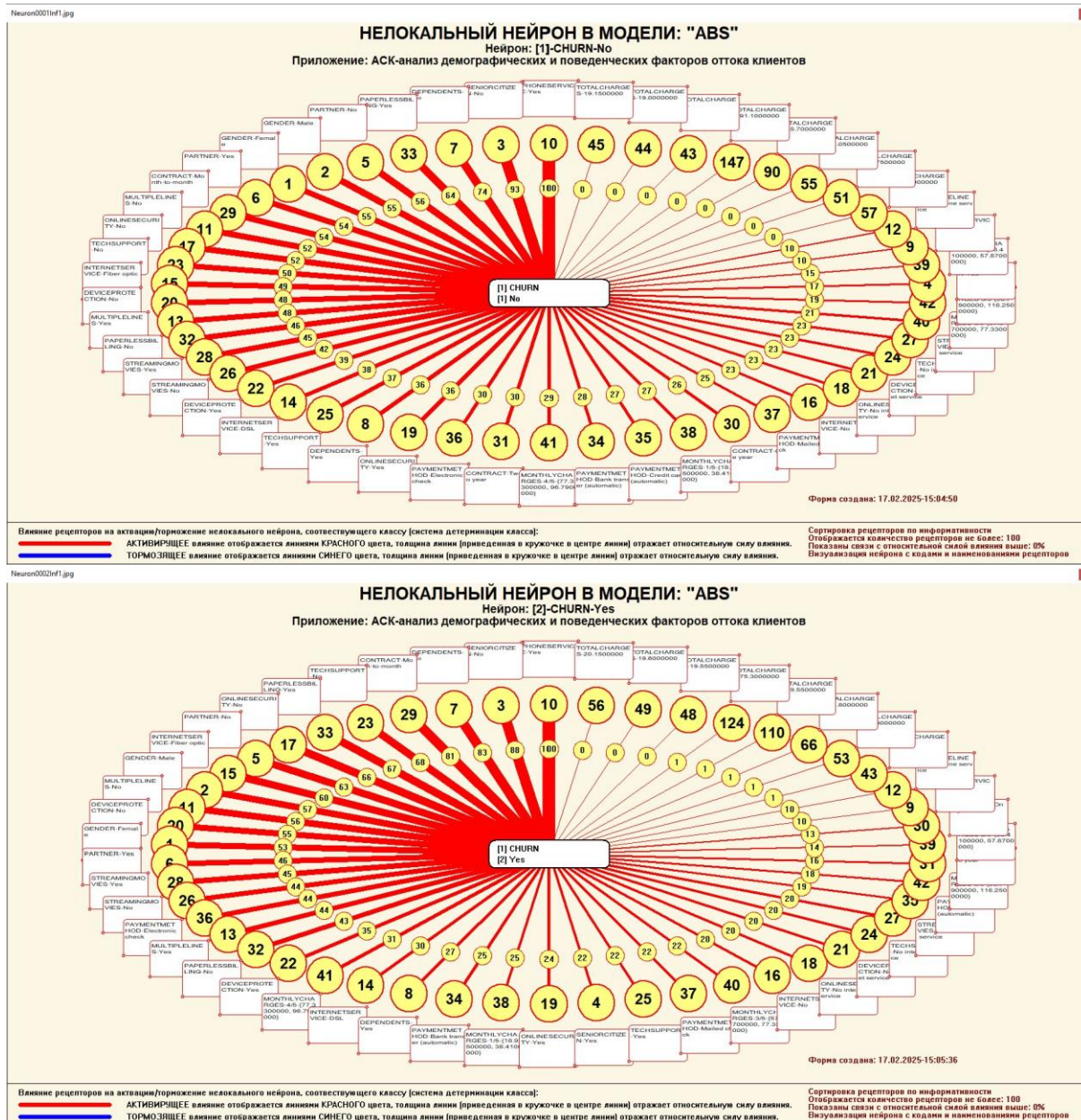


Рисунок 27. Примеры нелокальных нейронов, соответствующие классам

3.8.5. Нелокальная нейронная сеть

В системе «Эйдос» есть возможность построения моделей, соответствующих многослойным нейронным сетям.

Есть также возможность визуализации любого одного слоя нелокальной нейронной сети (режим 4.4.11 системы «Эйдос»).

Такой слой в наглядной форме отражает силу и направление влияния рецепторов ряда нейрона на степень их активации/торможения в форме цвета и толщины дендритов.

Нейроны на изображении слоя нейронной сети расположены слева направо в порядке убывания модуля суммарной силы их детерминации рецепторами, т.е. слева находятся результаты, наиболее жестко

обусловленные действующими на них значениями факторов, а справа – менее жестко обусловленные (рисунок 28). В форме управления визуализацией есть возможность задавать фильтры по факторам, которые надо визуализировать.

4.4.11. Отображение Парето-подмножеств одного слоя нелокальной нейронной сети в системе "Эйдос"

Выбор нелокальных нейронов (классов) для визуализации в нейросети

| Sel | Код | Наименование нелокального нейрона (класса) |
|-------------------------------------|-----|--|
| <input checked="" type="checkbox"/> | 1 | CHURN-No |
| <input checked="" type="checkbox"/> | 2 | CHURN-Yes |

Помощь Максимальное количество отображаемых нейронов: ClearSet Диапазон кодов отображаемых нейронов: -
 Максимальное количество отображаемых связей: Диапазон кодов отображаемых рецепторов: -

Подготовка визуализации нейрона: 1 "CHURN-No" в модели: 1 "ABS"

АКТИВИРУЮЩИЕ рецепторы и сила их влияния **ТОРМОЗЯЩИЕ рецепторы и сила их влияния**

| Код | Наименование фактора и его интервального значения | Сила влияния |
|-----|---|--------------|
| 10 | PHONESERVICE-Yes | 732.0... |
| 3 | SENIORCITIZEN-No | 683.0... |
| 7 | DEPENDENTS-No | 541.0... |
| 33 | PAPERLESSBILLING-Yes | 468.0... |
| 5 | PARTNER-No | 409.0... |
| 2 | GENDER-Male | 406.0... |
| 1 | GENDER-Female | 400.0... |
| 6 | PARTNER-Yes | 397.0... |
| 29 | CONTRACT-Month-to-month | 394.0... |
| 11 | MULTIPLELINES-M... | 390.0... |

| Код | Наименование фактора и его интервального значения | Сила влияния |
|-----|---|--------------|
| 0 | | |

ВКЛЮЧИТЬ фильтр по фактору ВЫКЛЮЧИТЬ фильтр по фактору ВКЛЮЧИТЬ фильтр по фактору ВЫКЛЮЧИТЬ фильтр по фактору

НейроСеть Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 Максимальное количество отображаемых рецепторов: Отображать связи с интенсивностью >= % от макс.:

Сортировать связи: по модулю информативности по информативности и знаку Отображать наименования: нейронов рецепторов

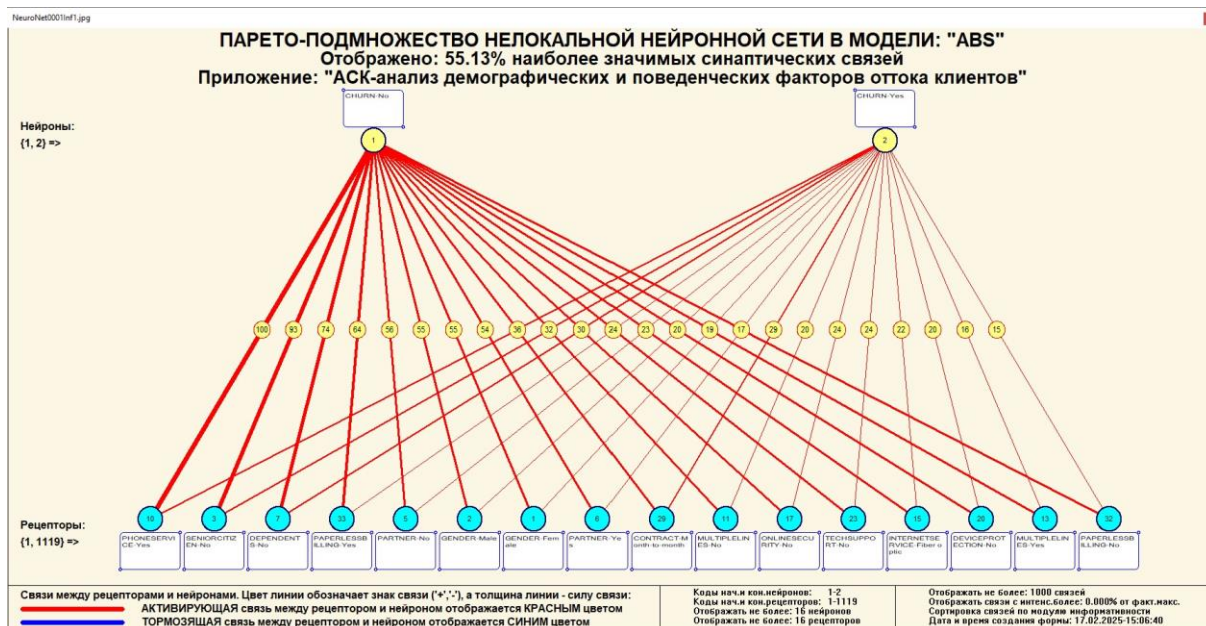


Рисунок 28. Нейронная сеть в СК-модели ABS

3.8.6. 3D-интегральные когнитивные карты

3d-интегральная когнитивная карта является отображением на одном рисунке когнитивной диаграммы классов вверху и когнитивной диаграммы значений факторов внизу и соединяющего их одного слоя нейронной сети (режим 4.4.12 системы «Эйдос») (рисунок 29).

4.4.12. Отображение Парето-подмножеств одного слоя интегральной когнитивной карты в системе "Эйдос"

Выбор нелокальных нейронов (классов) для визуализации в когнитивной карте

| Sel | Код | Наименование нелокального нейрона (класса) |
|-------------------------------------|-----|--|
| <input checked="" type="checkbox"/> | 1 | CHURN-No |
| <input type="checkbox"/> | 2 | CHURN-Yes |

Помощь Максимальное количество отображаемых нейронов: ClearSet Диапазон кодов отображаемых нейронов: -
 Максимальное количество отображаемых связей: Диапазон кодов отображаемых рецепторов: -

Подготовка визуализации нейрона:1 "CHURN-No" в модели:1 "ABS"

АКТИВИРУЮЩИЕ рецепторы и сила их влияния

| Код | Наименование фактора и его интервального значения | Сила влияния |
|-----|---|--------------|
| 10 | PHONESERVICE-Yes | 732.0... |
| 3 | SENIORCITIZEN-No | 683.0... |
| 7 | DEPENDENTS-No | 541.0... |
| 33 | PAPERLESSBILLING-Yes | 468.0... |
| 5 | PARTNER-No | 409.0... |
| 2 | GENDER-Male | 406.0... |
| 1 | GENDER-Female | 400.0... |
| 6 | PARTNER-Yes | 397.0... |
| 29 | CONTRACT-Month-to-month | 394.0... |
| 11 | MULTIBUSINESS- | 300.0... |

ТОРМОЗЯЩИЕ рецепторы и сила их влияния

| Код | Наименование фактора и его интервального значения | Сила влияния |
|-----|---|--------------|
| 0 | | |

ВКЛЮЧИТЬ фильтр по фактору ВЫКЛЮЧИТЬ фильтр по фактору ВКЛЮЧИТЬ фильтр по фактору ВЫКЛЮЧИТЬ фильтр по фактору

Когн. карта Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 Максимальное количество отображаемых рецепторов:
 Отображать связи с интенсивностью >= % от макс.:

Сортировать связи: Отображать наименования:
 по модулю информативности нейронов
 по информативности и знаку рецепторов

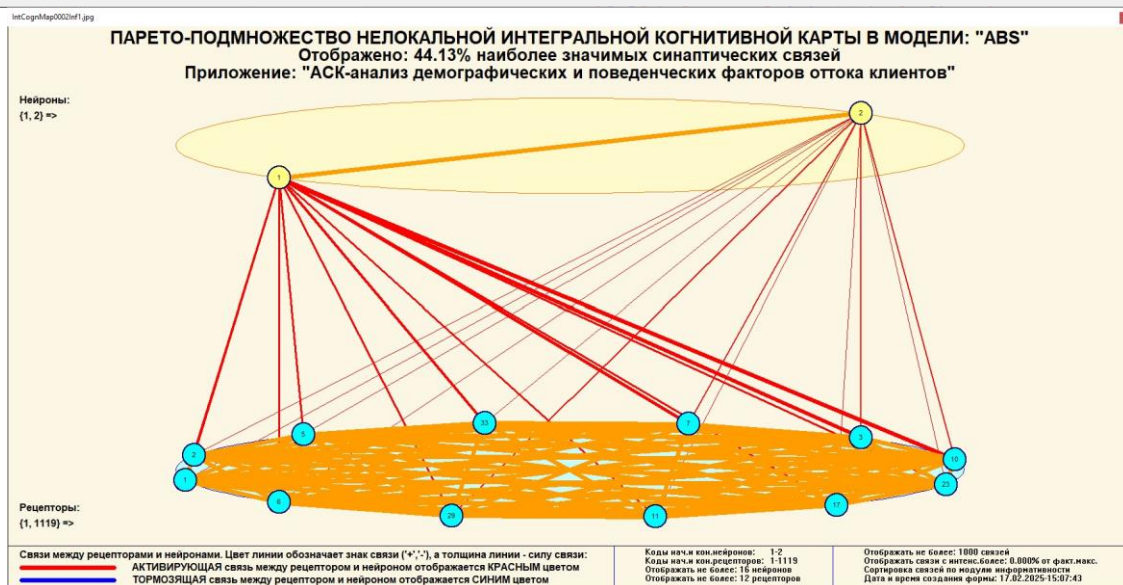


Рисунок 29 3d-когнитивная диаграмма классов и признаков (режим 4.4.12)

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

В 2d-когнитивных диаграммах сравнения классов по системе их детерминации видно, насколько сходны или насколько отличаются друг от друга классы по значениям обуславливающих их факторов.

Однако мы не видим из этой диаграммы, чем именно конкретно сходны и чем именно отличаются эти классы по значениям обуславливающих их факторов.

Это мы можем увидеть из когнитивной диаграммы содержательного сравнения классов, которая отображается в режиме 4.2.3 системы «Эйдос».

2D-интегральные когнитивные карты содержательного сравнения классов являются примерами опосредованных нечетких правдоподобных логических заключений, о которых может быть одним из первых писал Дьердь Пойа. Впервые об автоматизированной реализации рассуждений подобного типа в интеллектуальной системе «Эйдос» написано в 2002 году в работе на странице 521¹¹. Позже об этом писалось в работе¹² и ряде других работ автора, поэтому здесь подробнее рассматривать этот вопрос нецелесообразно.

Пример опосредованных правдоподобных рассуждений.

Допустим нам известно, что один человек имеет голубые глаза, а другой черные волосы. Спрашивается, эти признаки вносят вклад в сходство или в различие этих двух людей? В АСК-анализе и системе «Эйдос» этот вопрос решается так. В модели на основе кластерно-конструктивного анализа классов и значений факторов (признаков) известно, насколько те или иные признаки сходны или отличаются по их влиянию на объект моделирования. Поэтому понятно, что человек с голубыми глазами вероятнее всего блондин, а брюнет, скорее всего, имеет темные глаза. Так что понятно, что эти признаки вносят вклад в различие этих двух людей.

Всего системой в данной модели генерируется 4 формы содержательного сравнения классов. Так как каждый из 2 классов сравнивается со всеми остальными, в т.ч. с собой, то всего получается $2^2=4$ подобных диаграмм.

¹¹ https://www.elibrary.ru/download/elibrary_18632909_64818704.pdf, Таблица 7. 17, стр. 521

¹² <http://ej.kubagro.ru/2013/07/pdf/15.pdf>, стр.44.

3.8.8. Когнитивные функции

Когнитивные функции являются обобщением классического математического понятия функции на основе системной теории информации и предложены Е.В.Луценко в 2005 году.

Когнитивные функции отображают, какое количество информации содержится в градациях описательной шкалы о переходе объекта моделирования в состояния, соответствующие градациям классификационной шкалы. При этом в статистических и системно-когнитивных моделях в каждой градации описательной шкалы содержится информация обо всех градациях классификационной шкалы, т.е. *каждому значению аргумента соответствуют все значения функции, но соответствуют в разной степени, причем как положительной, так и отрицательной, которая отображается цветом.*

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

В системе «Эйдос» когнитивные функции отображаются в режиме 4.5 (рисунки 30). Первая экранная форма данного режима представляет собой краткий хелп, поясняющий смысл понятия «Когнитивная функция», а также позволяющий выйти на экранную форму системы «Эйдос» с действующими гиперссылками на работы по когнитивным функциям, а также страницы сайта автора со списком этих работ и работ автора по выявлению, представлению и использованию знаний, логике и методологии научного познания.

Необходимо отметить, что модели системы «Эйдос» – это *феноменологические* модели, отражающие *эмпирические* закономерности в фактах обучающей выборки, т.е. они отражают причинно-следственные связи, но не отражают *механизма детерминации*, а только сам факт и характер детерминации. Содержательное объяснение этих эмпирических закономерностей формулируется уже экспертами на теоретическом уровне познания в содержательных научных законах.

4.5. Генерация, визуализация и запись когнитивных функций системы "Эйдос"

Задайте статистические и/или системно-когнитивные модели для генерации когнитивных функций:

Статистические базы:

- 1. ABS - частный критерий: количество встреч сочетаний: "класс-признак" у объектов обуч.выборки
- 2. PRC1 - частный критерий: усл. вероятность i-го признака среди признаков объектов j-го класса
- 3. PRC2 - частный критерий: условная вероятность i-го признака у объектов j-го класса

Системно-когнитивные модели (Базы знаний):

- 4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1
- 5. INF2 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC2
- 6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс. частотами
- 7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1
- 8. INF5 - частный критерий: ROI (Return On Investment); вероятности из PRC2
- 9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вероятности из PRC1
- 10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; вероятности из PRC2

Задайте виды когнитивных функций для генерации, визуализации и записи:

- 1. Сетка триангуляции Делоне без цветовой заливки.
- 2. Сглаженные изолинии триангуляции Делоне без цветовой заливки.
- 3. Сетка триангуляции Делоне с цветовой заливкой.
- 4. Сглаженные изолинии триангуляции Делоне с цветовой заливкой.
- 5. Сглаженная цветочная заливка изолиний с заданным количеством градаций цвета.

Задайте дополнительные параметры визуализации когнитивных функций:

- Соединять ли точки с максимальным количеством информации линией КРАСНОГО цвета?
- Соединять ли точки с минимальным количеством информации линией СИНЕГО цвета?

Задайте количество градаций уровня (цвета и изолиний) когнитивных функций:

Задайте количество пикселей на дюйм в изображениях когнитивных функций:

Задайте паузу в секундах между визуализациями когнитивных функций:

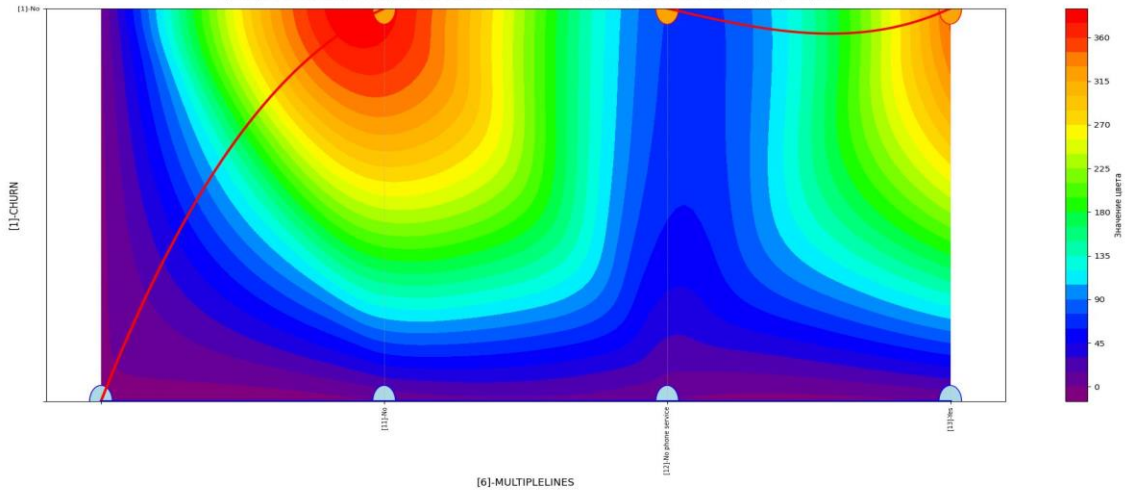
Задайте размер шрифта для наименований градаций шкал X и Y:

Визуализация когнитивных функций new Визуализация когнитивных функций old

Работы по когнитивным функциям-1 Работы по когнитивным функциям-2

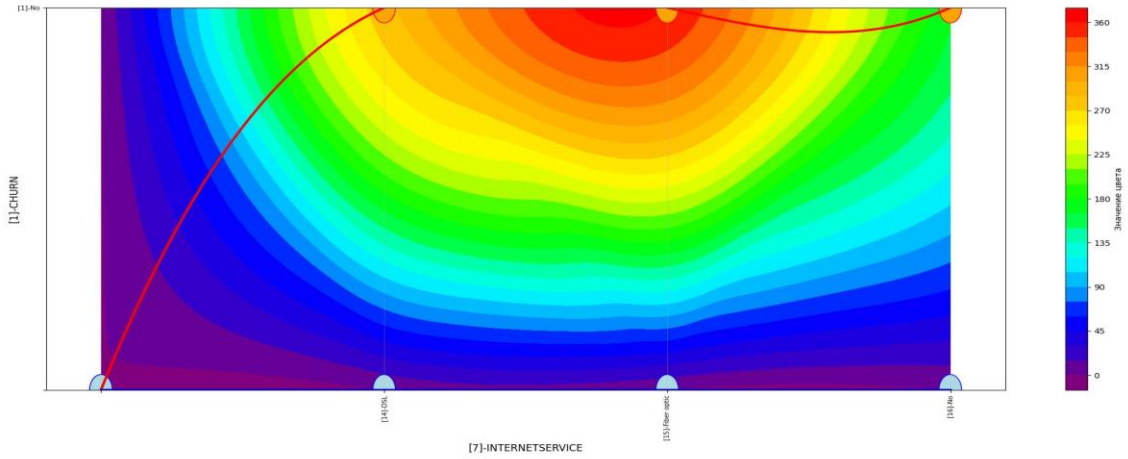
КОГНИТИВНАЯ ФУНКЦИЯ

Приложение: 'АСК-анализ демографических и поведенческих факторов оттока клиентов'
 Модель: '1. ABS - частный критерий: количество встреч сочетаний: "класс-признак" у объектов обуч.выборки'
 Функция: '5. Сглаженная цветовая заливка изолиний с заданным количеством градаций цвета'



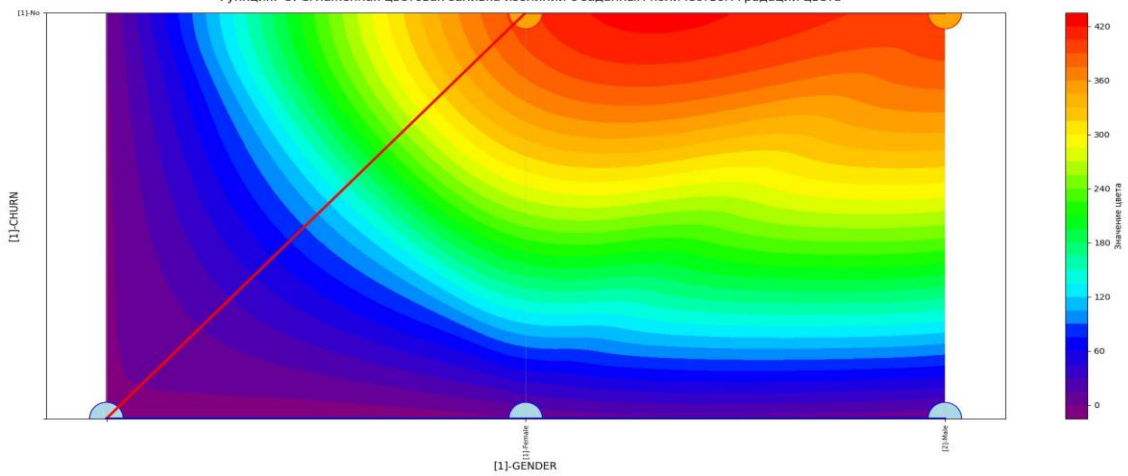
КОГНИТИВНАЯ ФУНКЦИЯ

Приложение: 'АСК-анализ демографических и поведенческих факторов оттока клиентов'
 Модель: '1. ABS - частный критерий: количество встреч сочетаний: "класс-признак" у объектов обуч.выборки'
 Функция: '5. Сглаженная цветовая заливка изолиний с заданным количеством градаций цвета'



КОГНИТИВНАЯ ФУНКЦИЯ

Приложение: 'АСК-анализ демографических и поведенческих факторов оттока клиентов'
 Модель: '1. ABS - частный критерий: количество встреч сочетаний: "класс-признак" у объектов обуч.выборки'
 Функция: '5. Сглаженная цветовая заливка изолиний с заданным количеством градаций цвета'



4.5. Визуализация когнитивных функций

Что такое когнитивная функция:

Визуализация прямых, обратных, позитивных, негативных, полностью и частично редуцированных когнитивных функций. Когнитивная функция представляет собой графическое отображение силы и направления влияния различных значений некоторого фактора на переходы объекта управления в будущие состояния, соответствующие классам. Когнитивные функции представляют собой новый перспективный инструмент отражения и наглядной визуализации закономерностей и эмпирических законов. Разработка содержательной научной интерпретации когнитивных функций представляет собой способ познания природы, общества и человека. Когнитивные функции могут быть: прямые, отражающие зависимость классов от признаков, обобщающие информационные портреты признаков; обратные, отражающие зависимость признаков от классов, обобщающие информационные портреты классов; позитивные, показывающие чему способствуют система детерминации; негативные, отражающие чему препятствуют система детерминации; средневзвешенные, отражающие совокупное влияние всех значений факторов на поведение объекта (причем в качестве весов наблюдений используется количество информации в значении аргумента о значениях функции) различной степенью редукции или степенью детерминации, которая отражает в графической форме (в форме полосы) количество знаний в аргументе о значении функции и является аналогом и обобщением доверительного интервала. Если отобразить подматрицу матрицы знания, отображая цветом силу и направление влияния каждой градации некоторой описательной шкалы на переход объекта в состояния, соответствующие классам некоторой классификационной шкалы, то получим нередуцированную когнитивную функцию. Когнитивные функции являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой "Эйдос". Необходимо отметить, что на вид функций влияния математической моделью АСК-анализа не накладывается никаких ограничений, в частности, они могут быть и не дифференцируемые.

Луценко Е.В. Метод визуализации когнитивных функций - новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2011. - №03(67). С. 240 - 282. - Шифр Информрегистра: 0421100012\0077. . 2,688 у.п.л. - Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>

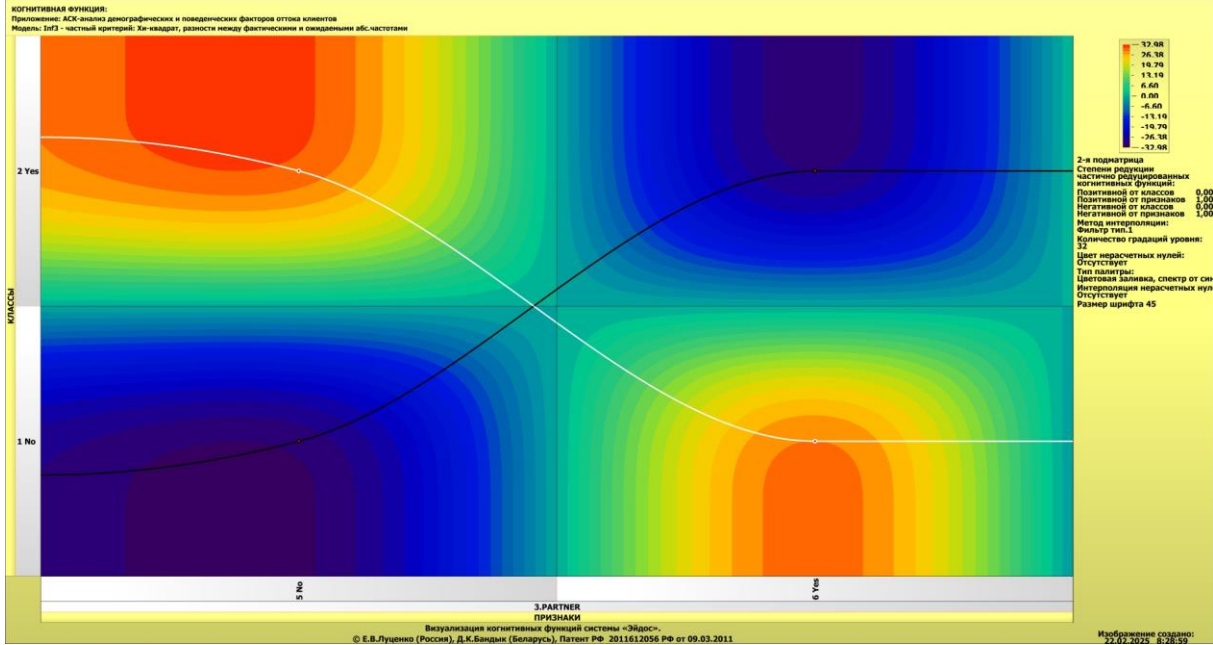
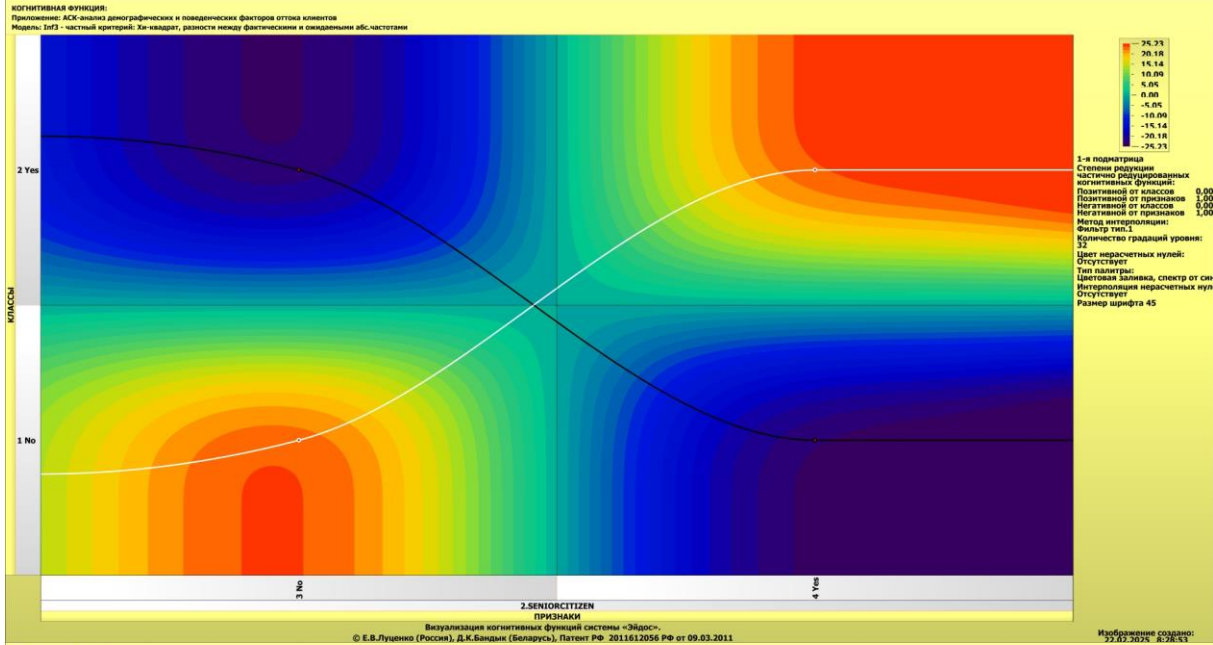
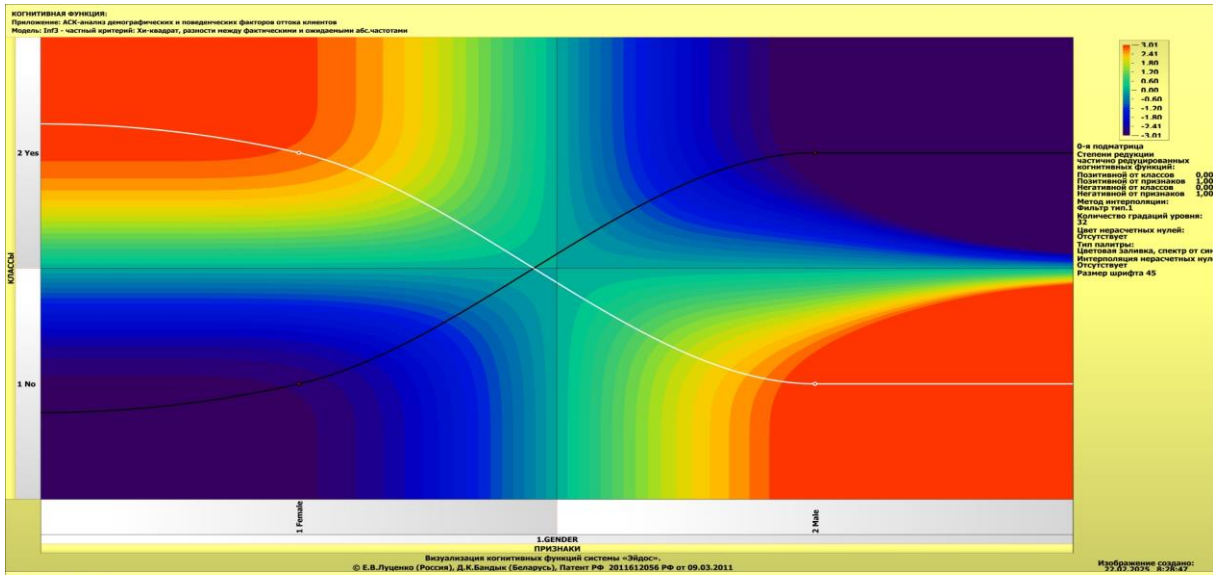
Задайте нужный режим:

Визуализации когнитивных функций

Литератур.ссылки на работы по когнитивным функциям

Литератур.ссылки на работы по когнитивным функциям

Литератур.ссылки на работы по управлению знаниями



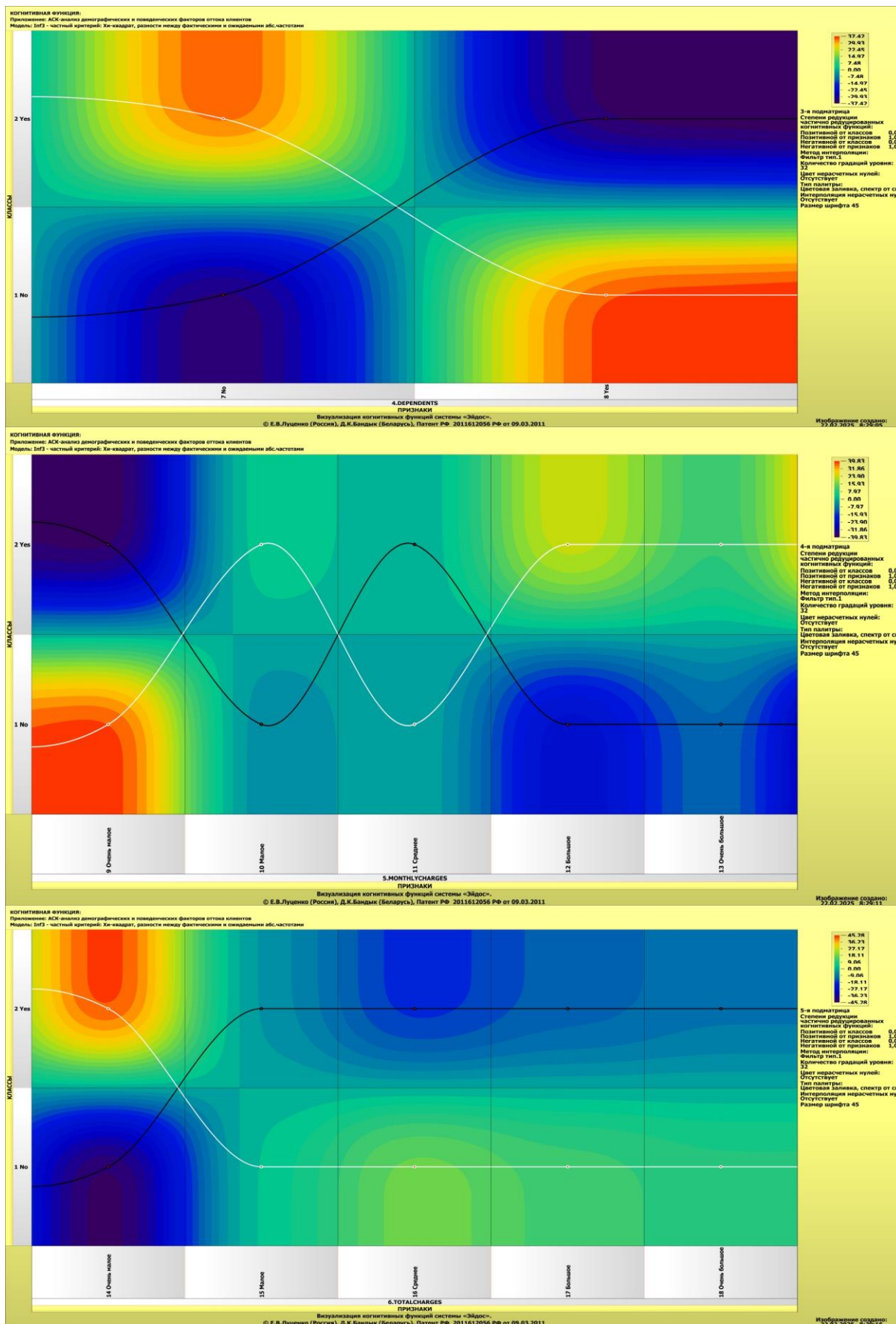


Рисунок 30. Примеры когнитивных функций в СК-модели ABS

Как уже отмечалось содержательное объяснение когнитивных функций на теоретическом уровне познания – это дело специалистов в той предметной области, к которой относится предмет моделирования.

3.8.9. Значимость описательных шкал и их градаций

В АСК-анализе все факторы рассматриваются с одной единственной точки зрения: сколько информации содержится в их значениях о переходе объекта моделирования и управления, на который они действуют, в определенное будущее состояние, описываемое классом (градация классификационной шкалы), и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: единицах количества информации.

Значимость (селективная сила) градаций описательных шкал в АСК-анализе – это вариабельность частных критериев в статистических и системно-когнитивных моделях, например, в модели Inf1, это вариабельность информативностей (режим 3.7.5 системы «Эйдос»):

Значимость всей описательной шкалы является средним от степени значимости ее градаций (режим 3.7.4 системы «Эйдос»).

Если рассортировать все градации факторов (признаки) в порядке убывания селективной силы и получить сумму селективной силы системы значений факторов нарастающим итогом, то получим Парето-кривую.

На рисунке 31 приведена Парето-кривая силы влияния значений факторов на поведение объекта моделирования в СК-модели ABS:

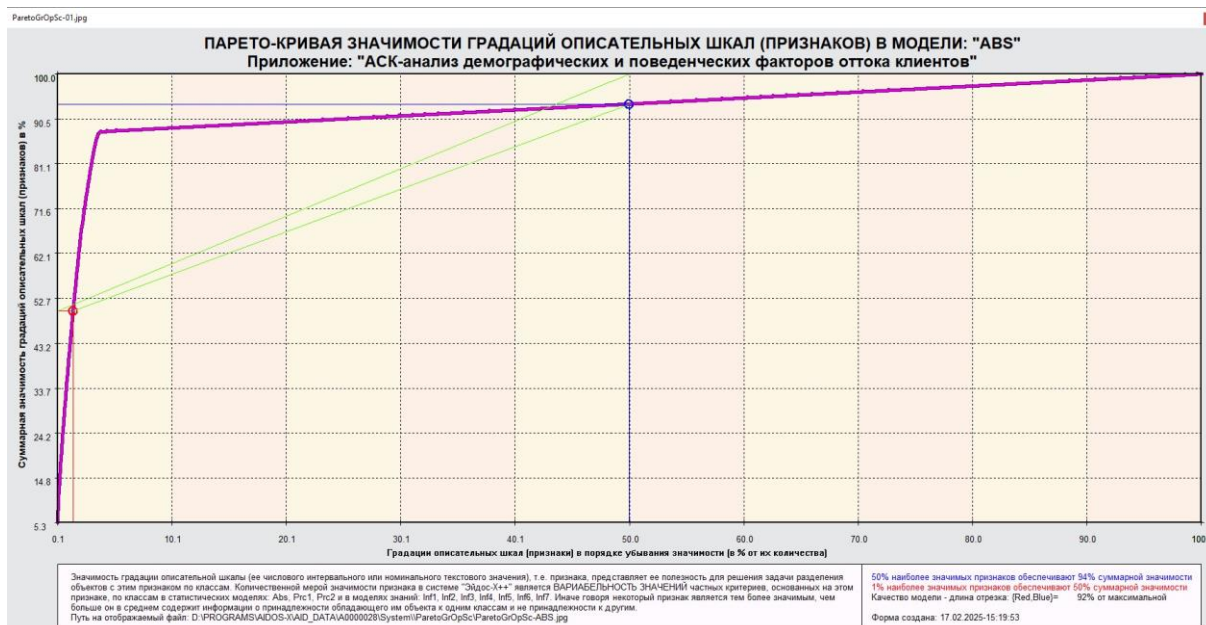


Рисунок 31. Парето-кривая силы влияния значений факторов на поведение объекта моделирования в СК-модели ABS

Из рисунка 31 видно, что 1% наиболее ценных значений факторов обеспечивает половину суммарного влияния всех значений факторов, а

половина наиболее ценных значений факторов обеспечивает 50% суммарного влияния. На рисунке 32 система «Эйдос» привела рейтинг качества статистических и системно-когнитивных моделей. Кроме того, на этом рисунке приведены имена Excel-файлов с информацией о силе и направлении влияния значений факторов в разных моделях.

В таблице 13 представлены исходные данные для построения кумулятивной кривой на рисунке 31. Из таблицы 13 видно, какую долю от суммарного влияния на переход объекта моделирования в будущее состояния, соответствующие классам, имеет каждое значение каждого фактора.

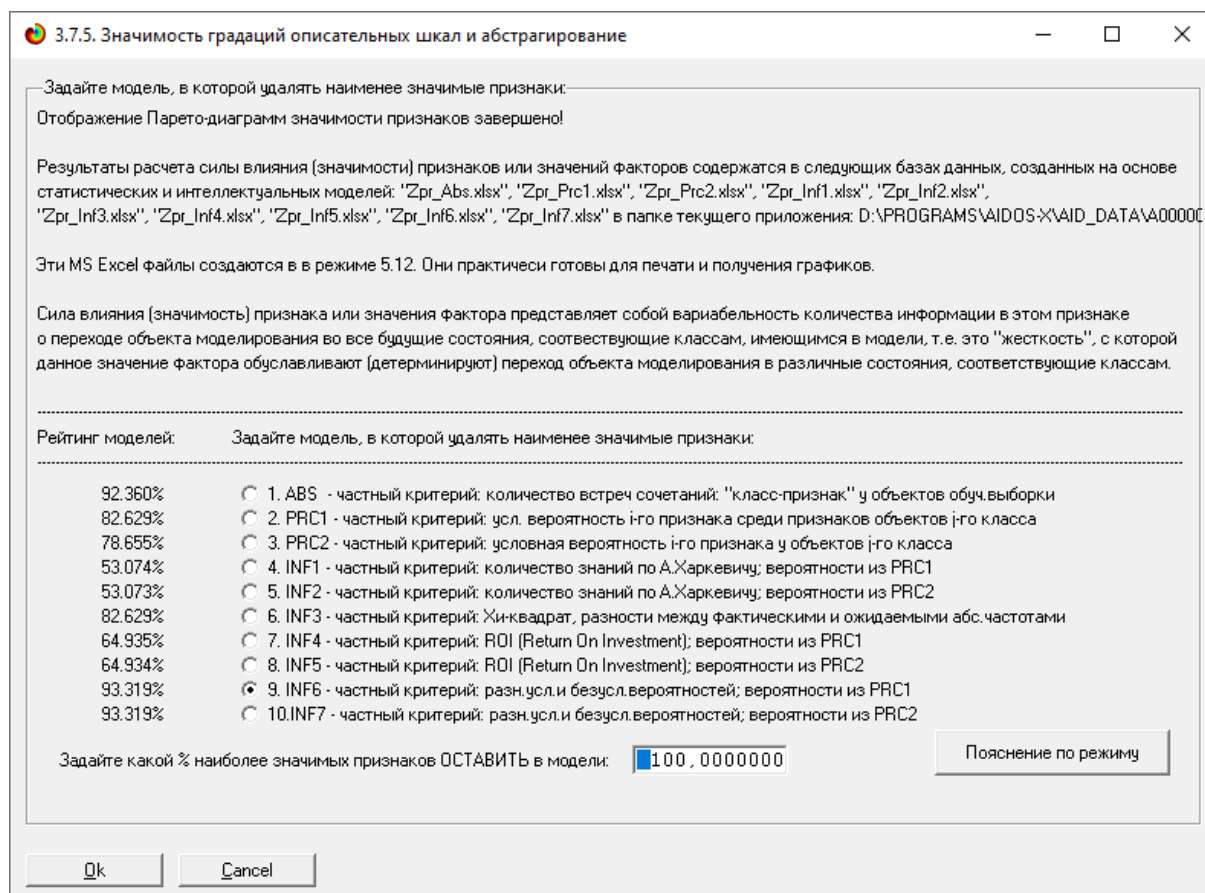


Рисунок 32. Рейтинг качества статистических и системно-когнитивных моделей и имена Excel-файлов с информацией о силе и направлении влияния значений факторов в этих моделях

Таблица 13 – Сила влияния значений факторов на поведение объекта моделирования в СК-модели ABS

| № | №% | Код значения фактора | Наименование фактора и его значения | Кода фактора | Значимость, % | Значимость кумулятивно, % |
|-----|------------|----------------------|-------------------------------------|--------------|---------------|---------------------------|
| 1 | 0,0893655 | 10 | PHONESERVICE-Yes | 5 | 331,6330804 | 331,6330804 |
| 2 | 0,1787310 | 3 | SENIORCITIZEN-No | 2 | 319,6122651 | 651,2453455 |
| 178 | 15,9070599 | 180 | TOTALCHARGES-145.1500000 | 16 | 0,7071068 | 5586,8506804 |
| 179 | 15,9964254 | 181 | TOTALCHARGES-146.6000000 | 16 | 0,7071068 | 5587,5577872 |

| | | | | | | |
|------|-------------|-----|---------------------------|----|-----------|--------------|
| 737 | 65,8623771 | 741 | TOTALCHARGES-2681.1500000 | 16 | 0,7071068 | 5982,1233816 |
| 738 | 65,9517426 | 742 | TOTALCHARGES-2684.8500000 | 16 | 0,7071068 | 5982,8304884 |
| 1118 | 99,9106345 | 224 | TOTALCHARGES-220.4500000 | 16 | 0,0000000 | 6249,4097520 |
| 1119 | 100,0000000 | 546 | TOTALCHARGES-1305.9500000 | 16 | 0,0000000 | 6249,4097520 |

На экранной форме рисунка 33 приведены имена Excel-файлов с информацией о силе влияния факторов в разных моделях.

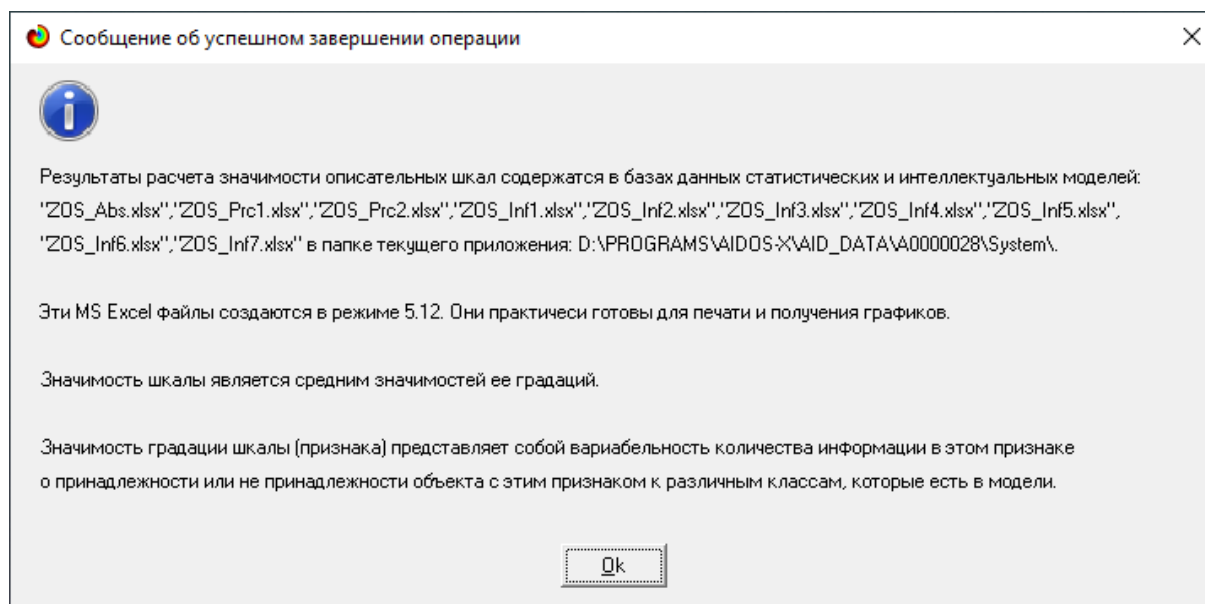


Рисунок 33. Имена Excel-файлов с информацией о силе влияния факторов в статистических и системно-когнитивных моделях

3.8.11. Степень детерминированности классов и классификационных шкал

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается **степенью варибельности значений факторов** (градаций описательных шкал) в колонке матрицы модели, соответствующей данному классу (режим 3.7.3 системы «Эйдос»).

Чем выше степень детерминированности класса, тем более достоверно он прогнозируется по значениям факторов.

Степень детерминированности (обусловленности) всей классификационной шкалы является средним от степени детерминированности ее градаций, т.е. классов (режим 3.7.2 системы «Эйдос»).

На рисунках 34 приведены экранные формы режимов 3.7.2 и 3.7.3 системы «Эйдос», содержащие информацию о степени детерминированности (обусловленности) состояний объекта моделирования действующими на него факторами:



Рисунок 34. Экранные формы режимов 3.7.2 и 3.7.3 системы «Эйдос»

В таблице 14 представлены исходные данные для построения кумулятивной кривой на рисунке 34.

Таблица 14 – Степень детерминированности классов в СК-модели ABS

| № | Код | Наименование класса | Значимость, % | Значимость кумулятивно, % |
|---|-----|---------------------|---------------|---------------------------|
| 1 | 1 | CHURN-No | 61,4000000 | 61,4000000 |
| 2 | 2 | CHURN-Yes | 22,9000000 | 84,3000000 |

В таблице 14 приведена информация о степени детерминированности классов значениями факторов в системно-когнитивной модели ABS. Степень детерминированности классификационных шкал является средним от степени детерминированности их градаций.

3.8.12. Агломеративная кластеризация значений факторов

Сначала выполняется режим расчета матриц сходства 4.3.2.1, а затем режим генерации агломеративных дендрограмм значений факторов 4.3.2.3 (рисунки 35).

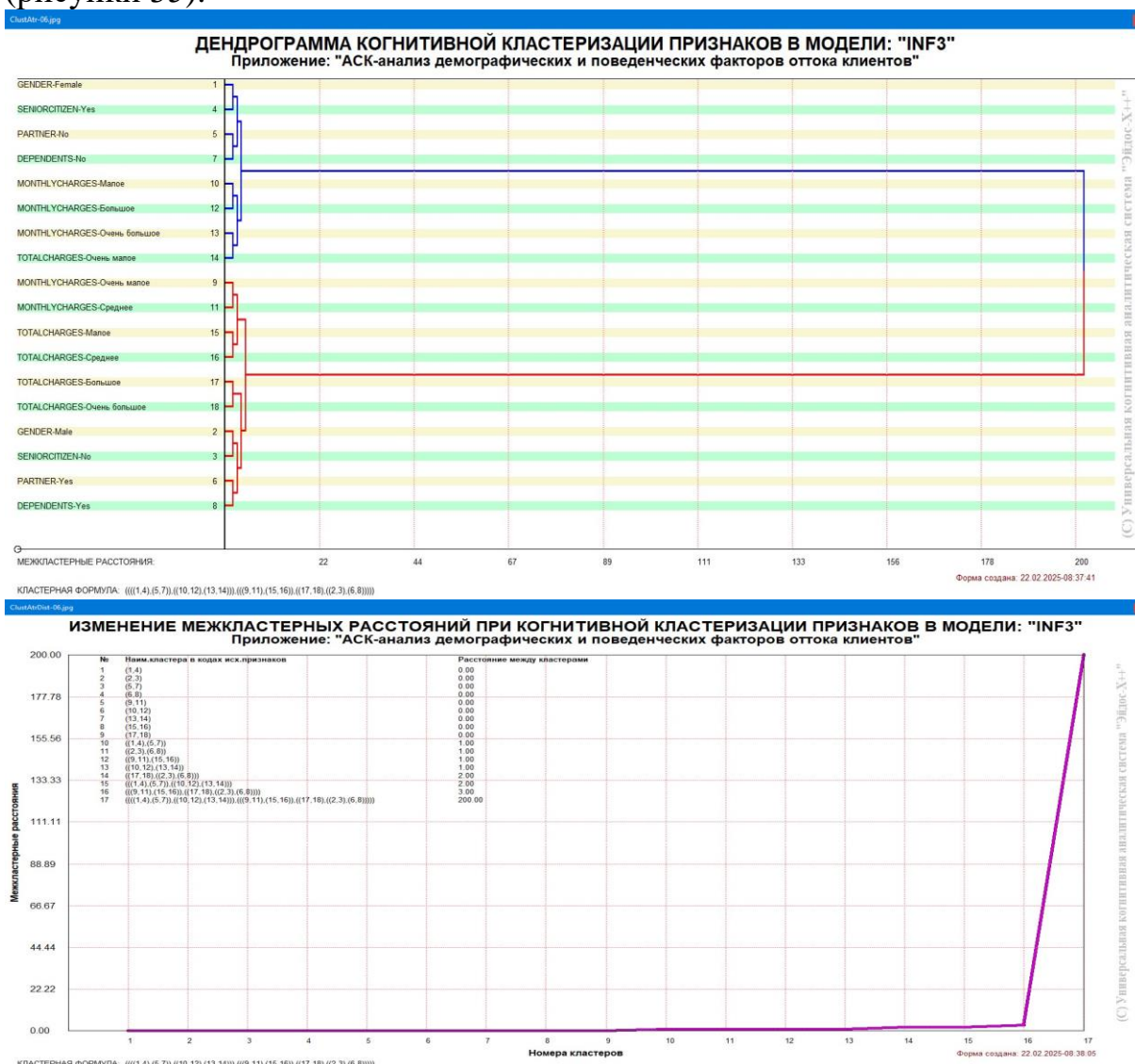


Рисунок 35. Экранные формы режима 4.3.2.3 системы «Эйдос»

4. DISCUSSION (ОБСУЖДЕНИЕ)

В ходе выполнения работы были успешно решены поставленные задачи, что позволило достичь заявленных целей. Для этого был применён метод Автоматизированного системно-когнитивного анализа (АСК-анализ) с использованием программного инструмента – системы «Эйдос». Проведённый анализ полностью согласуется с исходными данными, на которых строилось исследование. При этом использование АСК-анализа и системы «Эйдос» существенно расширяет возможности для прогнозирования, принятия решений и углублённого изучения моделируемой предметной области по сравнению с традиционными подходами. Это делает данные методы эффективным инструментом для дальнейших научных и практических исследований.

Основные достижения работы:

- Разработана методология корректного построения сопоставимых системно-когнитивных моделей, которые могут работать с разнородными данными, включая лингвистические переменные и числовые показатели, выраженные в различных единицах измерения;
- подтверждена возможность применения таких моделей для решения широкого спектра задач, включая прогнозирование, поддержку принятия решений и детальное исследование особенностей моделируемой предметной области.

Для углубления и расширения результатов рекомендуется:

- Увеличить объём исходных данных, чтобы повысить точность и репрезентативность моделей;
- ввести дополнительные классификационные шкалы;
- практическая и теоретическая ценность предложенных решений очевидна. Их актуальность подтверждается успешным опытом применения подобных методик в смежных областях. Развитие данного направления позволит значительно улучшить качество прогнозов, повысить обоснованность принимаемых решений и углубить понимание исследуемых процессов.

5. CONCLUSIONS (ВЫВОДЫ)

Понимание причин, по которым клиенты прекращают сотрудничество с компанией, позволяет разработать эффективные стратегии удержания и повысить лояльность потребителей. Однако анализ причин оттока часто сталкивается с рядом сложностей, таких как разнородность данных (демографические, поведенческие и другие типы переменных), их различная природа (категориальные и числовые данные) и необходимость выявления сложных взаимосвязей между факторами.

Отток клиентов может быть обусловлен множеством факторов, включая демографические характеристики (возраст, пол, уровень дохода), поведенческие паттерны (частота покупок, средний чек, активность на платформе) и внешние условия (экономическая ситуация, конкурентное давление). Традиционные методы анализа данных, такие как регрессионный анализ или кластеризация, часто оказываются недостаточно гибкими для работы с разнообразными типами данных и не позволяют выявить скрытые закономерности, влияющие на решения клиентов.

Использование АСК-анализа и системы «Эйдос» открывает новые возможности для понимания причин оттока клиентов и разработки эффективных стратегий удержания. Этот подход может быть особенно полезен для компаний, работающих в цифровой среде, где поведенческие данные играют ключевую роль. Кроме того, предложенный метод может быть использован для научных исследований в области маркетинга, управления и анализа данных, способствуя развитию теории и практики в этих областях.

Таким образом, применение АСК-анализа для исследования демографических и поведенческих факторов оттока клиентов представляет собой перспективное направление, которое может значительно повысить эффективность бизнес-процессов и способствовать устойчивому развитию компаний в условиях современной конкурентной среды.

REFERENCES (ЛИТЕРАТУРА)

1. Луценко, Е. В. Сценарный и спектральный автоматизированный системно-когнитивный анализ / Е. В. Луценко. – Краснодар : Кубанский государственный аграрный университет им. И.Т. Трубилина, 2021. – 288 с. – DOI 10.13140/RG.2.2.22981.37608. – EDN ZQLITW.
2. Орлов, А. И. Анализ данных, информации и знаний в системной нечеткой интервальной математике / А. И. Орлов, Е. В. Луценко. – Краснодар : Кубанский государственный аграрный университет им. И.Т. Трубилина, 2022. – 405 с. – ISBN 978-5-907550-62-9. – DOI 10.13140/RG.2.2.15688.44802. – EDN OQULUW.
3. Луценко, Е. В. Автоматизация функционально-стоимостного анализа и метода "Директ-костинг" на основе АСК-анализа и системы "Эйдос" (автоматизация управления натуральной и финансовой эффективностью затрат без содержательных технологических и финансово-экономических расчетов на основе информационных и когнитивных технологий и теории управления)1 / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2017. – № 131. – С. 1-18. – DOI 10.21515/1990-4665-131-001. – EDN ZRXVFN.
4. Луценко, Е. В. Системное обобщение принципа Эшби и повышение уровня системности модели объекта познания как необходимое условие адекватности процесса его познания / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2020. – № 163. – С. 100-134. – DOI 10.21515/1990-4665-163-009. – EDN SWKGWY.

5. Луценко, Е. В. Эффективность объекта управления как его эмерджентное свойство и повышение уровня системности как цель управления / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2021. – № 165. – С. 77-98. – DOI 10.13140/RG.2.2.11887.25761. – EDN UMTAMT.

6. Луценко, Е. В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2003. – № 1. – С. 76-88. – EDN JWXLKT.

7. Работы проф.Е.В.Луценко & С^о по выявлению, представлению и использованию знаний, логике и методологии научного познания: http://lc.kubagro.ru/aidos/Works_on_identification_presentation_and_use_of_knowledge.htm

8. Пойа Дьердь. Математика и правдоподобные рассуждения. // под редакцией С.А.Яновской. Пер. с английского И.А.Вайнштейна., М., Наука, 1975 — 464 с., <http://ilib.mccme.ru/djvu/polya/rassuzhdenija.htm>

9. Луценко, Е. В. Системно-когнитивный анализ как развитие концепции смысла Шенка - Абельсона / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2004. – № 5. – С. 14-35. – EDN JWXMKX.

10. Работы проф.Е.В.Луценко & С^о по когнитивным функциям: http://lc.kubagro.ru/aidos/Works_on_cognitive_functions.htm

11. Луценко, Е. В. Системы представления и приобретения знаний / Е. В. Луценко, В. И. Лойко, В. Н. Лаптев. – Краснодар : Экоинвест, 2018. – 513 с. – ISBN 978-5-94215-415-8. – EDN UZZBLC.

12. Луценко, Е. В. Проблемы и перспективы теории и методологии научного познания и автоматизированный системно-когнитивный анализ как автоматизированный метод научного познания, обеспечивающий содержательное феноменологическое моделирование / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2017. – № 127. – С. 1-60. – DOI 10.21515/1990-4665-127-001. – EDN YLZTMX.

13. Влияние сроков применения борной кислоты на генеративную деятельность косточковых культур / Т. Н. Дорошенко, Л. Г. Рязанова, Н. В. Захарчук, Д. В. Максимцов // Плодоводство и виноградарство Юга России. – 2016. – № 41(5). – С. 121-130. – EDN WKBFHT, <https://journalkubansad.ru/pdf/16/05/12.pdf>

14. Монографии по АСК-анализу: http://lc.kubagro.ru/aidos/Aidos-X.htm#_Toc128746370

15. Некоторые учебники и учебные пособия проф.Е.В.Луценко: http://lc.kubagro.ru/aidos/Aidos-X.htm#_Toc128746372.