

УДК 004.8

06.01.01 – Общее земледелие, растениеводство
(сельскохозяйственные науки)

АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ ЗАВИСИМОСТИ УРОВНЯ РАЗВИТИЯ РАКОВЫХ ОПУХОЛЕЙ ЛЕГКИХ ОТ РАЗЛИЧНЫХ ФАКТОРОВ

Гулаков Сергей Юрьевич

Шипулин Семен Владимирович

Студенты КУБГАУ группы ИТ2201

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

Данная работа является продолжением серии работ автора по применению Автоматизированного системно-когнитивного анализа (ACK-анализ) для решения широкого спектра задач в области агрономии, т.е. по когнитивной медицине. В работе решается задача выявления зависимости уровня развития раковых опухолей легких от различных факторов. На основе знания этих зависимостей решаются задачи прогнозирования, принятия решений и исследования моделируемой предметной области путем исследования ее системно-когнитивной модели. Спецификой данной задачи является то, что все независимые переменные (факторы) являются лингвистическими (категориальными) переменными. Поэтому для решения данной задачи применяется лингвистический ACK-анализ, т.е. когнитивная математическая лингвистика. При этом зависимые переменные, т.е. результаты влияния факторов, измеряются в числовых шкалах. Таким образом, в работе строится гибридная модель, включающая как номинальные (текстовые), так и числовые шкалы. Сопоставимость обработки данных разных типов, представленных в разных типах шкал и разных единицах измерения обеспечивается путем метризации номинальных шкал, т.е. повышения их степени формализации до уровня числовых шкал. Приводится краткое описание ACK-анализа и его программного инструментария – интеллектуальной системы «Эйдос». Работа может быть основой для лабораторных работ по применению систем искусственного интеллекта, в частности лингвистического ACK-анализа для решения задач в области когнитивной агрономии.

Ключевые слова: ЛИНГВИСТИЧЕСКИЙ ACK-АНАЛИЗ, ЛИНГВИСТИЧЕСКИЙ АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, КОГНИТИВНАЯ АГРОНОМИЯ, ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА «ЭЙДОС»,

УДК 004.8

01/06/01 - General farming, crop production (agricultural sciences)

AUTOMATED SYSTEM-COGNITIVE ANALYSIS OF THE DEPENDENCE OF THE LEVEL OF DEVELOPMENT OF LUNG CANCER TUMORS ON VARIOUS FACTORS

Gulakov Sergey Yurievich

Shipulin Semyon Vladimirovich

Students of KUBGAU group IT2201

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

This work is a continuation of the author's series of works on the use of Automated system-cognitive analysis (ASC-analysis) to solve a wide range of problems in the field of agronomy, i.e. cognitive agronomy. The paper solves the problem of identifying the dependence of the level of development of lung cancer tumors on various factors. Based on the knowledge of these dependencies, the tasks of forecasting, decision-making and research of the simulated subject area are solved by studying its system-cognitive model. The specificity of this task is that all independent variables (factors) are linguistic (categorical) variables. Therefore, linguistic ASC-analysis, i.e. cognitive mathematical linguistics, is used to solve this problem. At the same time, dependent variables, i.e. the results of the influence of factors, are measured in numerical scales. Thus, a hybrid model is being built in the work, including both nominal (text) and numerical scales. Comparability of data processing of different types presented in different types of scales and different units of measurement is ensured by metrization of nominal scales, i.e. increasing their degree of formalization to the level of numerical scales. A brief description of the ASC-analysis and its software tools - the intelligent system "Eidos" is given. The work can be the basis for laboratory work on the use of artificial intelligence systems, in particular linguistic ASC-analysis for solving problems in the field of cognitive agronomy.

Keywords: LINGUISTIC ASK-ANALYSIS, LINGUISTIC AUTOMATED SYSTEMIC COGNITIVE ANALYSIS, COGNITIVE AGRONOMY, INTELLIGENT SYSTEM "EIDOS",

СОДЕРЖАНИЕ

1. INTRODUCTION (ВВЕДЕНИЕ).....	3
1.1. ОПИСАНИЕ ИССЛЕДУЕМОЙ ПРЕДМЕТНОЙ ОБЛАСТИ	3
1.2. ОБЪЕКТ И ПРЕДМЕТ ИССЛЕДОВАНИЯ	3
1.3. ПРОБЛЕМА, РЕШАЕМАЯ В РАБОТЕ И ЕЕ АКТУАЛЬНОСТЬ.....	3
1.4. ЦЕЛЬ РАБОТЫ	3
2. METHODS (МЕТОДЫ).....	3
2.1. ОБОСНОВАНИЕ ТРЕБОВАНИЙ К МЕТОДУ РЕШЕНИЯ ПРОБЛЕМЫ	3
2.2. ЛИТЕРАТУРНЫЙ ОБЗОР МЕТОДОВ РЕШЕНИЯ ПРОБЛЕМЫ, ИХ ХАРАКТЕРИСТИКА И ОЦЕНКА СТЕПЕНИ СООТВЕТСТВИЯ ОБОСНОВАННЫМ ТРЕБОВАНИЯМ	4
2.3. АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ (АСК-АНАЛИЗ) КАК МЕТОД РЕШЕНИЯ ПРОБЛЕМЫ.....	4
2.4. СИСТЕМА «ЭЙДОС» - ИНСТРУМЕНТАРИЙ АСК-АНАЛИЗА	6
2.5. ЦЕЛЬ И ЗАДАЧИ РАБОТЫ	10
3. RESULTS (РЕЗУЛЬТАТЫ).....	13
3.1. ЗАДАЧА-1. КОГНИТИВНАЯ СТРУКТУРИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ. ДВЕ ИНТЕРПРЕТАЦИИ КЛАССИФИКАЦИОННЫХ И ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ	13
3.2. ЗАДАЧА-2. ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ	14
3.3. ЗАДАЧА-3. СИНТЕЗ СТАТИСТИЧЕСКИХ И СИСТЕМНО-КОГНИТИВНЫХ МОДЕЛЕЙ. МНОГОПАРАМЕТРИЧЕСКАЯ ТИПИЗАЦИЯ И ЧАСТНЫЕ КРИТЕРИИ ЗНАНИЙ.....	21
3.4. ЗАДАЧА-4. ВЕРИФИКАЦИЯ МОДЕЛЕЙ	32
3.5. ЗАДАЧА-5. ВЫБОР НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	35
3.6. ЗАДАЧА-6. СИСТЕМНАЯ ИДЕНТИФИКАЦИЯ И ПРОГНОЗИРОВАНИЕ	36
3.6.1. Интегральный критерий «Сумма знаний»	37
3.6.2. Интегральный критерий «Семантический резонанс знаний».....	37
3.6.3. Важные математические свойства интегральных критериев.....	38
3.6.4. Решение задачи идентификации и прогнозирования в системе «Эйдос».....	39
3.7. ЗАДАЧА-7. ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ.....	42
3.7.1. Упрощенный вариант принятия решений как обратная задача прогнозирования, позитивный и негативный информационные портреты классов, SWOT-анализ.....	42
3.7.2. Развитый алгоритм принятия решений в адаптивных интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»	44
3.8. ЗАДАЧА-8. ИССЛЕДОВАНИЕ ОБЪЕКТА МОДЕЛИРОВАНИЯ ПУТЕМ ИССЛЕДОВАНИЯ ЕГО МОДЕЛИ ..	48
3.8.1. Инвертированные SWOT-диаграммы значений описательных шкал (семантические потенциалы)	48
3.8.2. Кластерно-конструктивный анализ классов	51
3.8.3. Кластерно-конструктивный анализ значений описательных шкал	53
3.8.4. Модель знаний системы «Эйдос» и нелокальные нейроны.....	56
3.8.5. Нелокальная нейронная сеть.....	59
3.8.6. 3D-интегральные когнитивные карты	61
3.8.7. 2D-интегральные когнитивные карты содержательного сравнения классов (опосредованные нечеткие правдоподобные рассуждения).....	62
3.8.8. 2D-интегральные когнитивные карты содержательного сравнения значений факторов (опосредованные нечеткие правдоподобные рассуждения).....	65
3.8.9. Когнитивные функции	68
3.8.10. Значимость описательных шкал и их градаций.....	74
3.8.11. Степень детерминированности классов и классификационных шкал	76
4. DISCUSSION (ОБСУЖДЕНИЕ)	77
5. CONCLUSIONS (ВЫВОДЫ)	78
REFERENCES (ЛИТЕРАТУРА)	79

1. INTRODUCTION (ВВЕДЕНИЕ)

1.1. Описание исследуемой предметной области

Данная работа является продолжением серии работ автора по применению Автоматизированного системно-когнитивного анализа (АСК-анализ) для решения широкого спектра задач в области медицины. В работе решается задача выявления зависимости показателей развития рака лёгких от различных факторов: анатомических факторов, внешней среды, привычек пациента и т.д. На основе знания этих зависимостей решаются разнообразные задачи прогнозирования, принятия решений и исследования моделируемой предметной области путем исследования ее системно-когнитивной модели (СК-модель).

1.2. Объект и предмет исследования

Объект исследования – выявление зависимостей развития раковых опухолей лёгких от различных факторов.

Предмет исследования – выявление зависимостей развития раковых опухолей лёгких от анатомических факторов, внешней среды, привычек пациента и т.д.

1.3. Проблема, решаемая в работе и ее актуальность

Спецификой данной задачи является то, что все независимые переменные имеют очень низкую степень формализации, т.к. формализуются в виде **лингвистических** (категориальных) переменных, тогда как результаты влияния этих факторов, т.е. различные медицинские показатели пациентов измеряются в числовых шкалах.

Таким образом, в работе решается **проблема** построения гибридной модели, включающей как номинальные (текстовые), так и числовые шкалы и обеспечивающей **сопоставимость** обработки данных разных типов, представленных в разных типах шкал и разных единицах измерения.

1.4. Цель работы

Целью работы является решение поставленной проблемы.

Достижение поставленной цели обеспечивается решением ряда **задач** и подзадач, которые являются **этапами** достижения цели. Конкретная формулировка этих задач зависит от метода решения проблемы, поэтому обоснованно мы сформулируем их в конце раздела, т.е. после обоснованного выбора и описания метода решения проблемы.

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 очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения.

2. Иначе говоря, метод решения проблемы не должен предъявлять жестких требований к исходным данным, которые невозможно выполнить, а должен обеспечивать обработку тех данных, которые реально есть.

3. Метод должен 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 программных систем, *одновременно*:

- находящихся в полном открытом бесплатном доступе;
 - обеспечивающих сопоставимую обработку числовой и текстовой информации в одной модели, дает следующие результаты;

дал, что альтернатив Автоматизированному системно-когнитивному языку и его программному инструментарию – системе «Эйдос» в настоящее время здесь нет [4].

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 [2].

Сам термин: «Автоматизированный системно-когнитивный анализ (ACK-анализ)» был предложен проф. Е. В. Луценко. На тот момент он вообще не встречался в Internet. Сегодня по соответствующему запросу в Яндексе находится 9 миллионов сайтов с этим сочетанием слов².

АСК-анализ включает:

- теоретические основы, в частности базовую формализуемую когнитивную концепцию;

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

- 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 [3] и ряде других [5]. Около половины из 665 опубликованных автором научных работ посвящены теоретическим основам АСК-анализа и его практическим применением в ряде предметных областей. На момент написания данной работы автором опубликовано более 40 монографий, 27 учебных пособий, в т.ч. 3 учебных пособия с грифами УМО и Министерства, получен 32 патент РФ на системы искусственного интеллекта, 346 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 [Scopus](#)³ [6, 7, 8].

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, психологическим и медицинским наукам, еще несколько докторских и кандидатских диссертаций 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 номенклатуре научных специальностей ВАК РФ⁶). Основные научные специальности, которым соответствует научная школа:

- 5.12.4. Когнитивное моделирование;
- 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;

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/>

- Автоматизированный системно-когнитивный анализ текстовых данных;
- 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 [7] и страничка в РесечГейт [8], 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.

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 [6]. Для работы с лингвистическими переменными применяется лингвистический АСК-анализ [4].

2.4. Система «Эйдос» - инструментарий АСК-анализа

Существует много систем искусственного интеллекта. Универсальная когнитивная аналитическая система «Эйдос» отличается от них следующими параметрами:

- 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>). Система «Эйдос» является автоматизированной системой, т.е. предполагает непосредственное участие человека в реальном времени при решении задач идентификации, прогнозирования, принятия решений и исследования предметной области (автоматические системы работают без такого участия человека);

- находится в полном открытом бесплатном доступе (http://lc.kubagro.ru/aidos/_Aidos-X.htm), причем с актуальными исходными текстами (http://lc.kubagro.ru/_AidosALL.txt): открытая лицензия: [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) (<https://creativecommons.org/licenses/by-sa/4.0/>), и это означает, что ей могут пользоваться все, кто пожелает, без какого-либо дополнительного разрешения со стороны первичного правообладателя – автора системы «Эйдос» проф. Е.В.Луценко (отметим, что система «Эйдос» создана полностью с использованием только лицензионного инструментального программного обеспечения и на нее имеется 31 свидетельство РосПатента РФ);

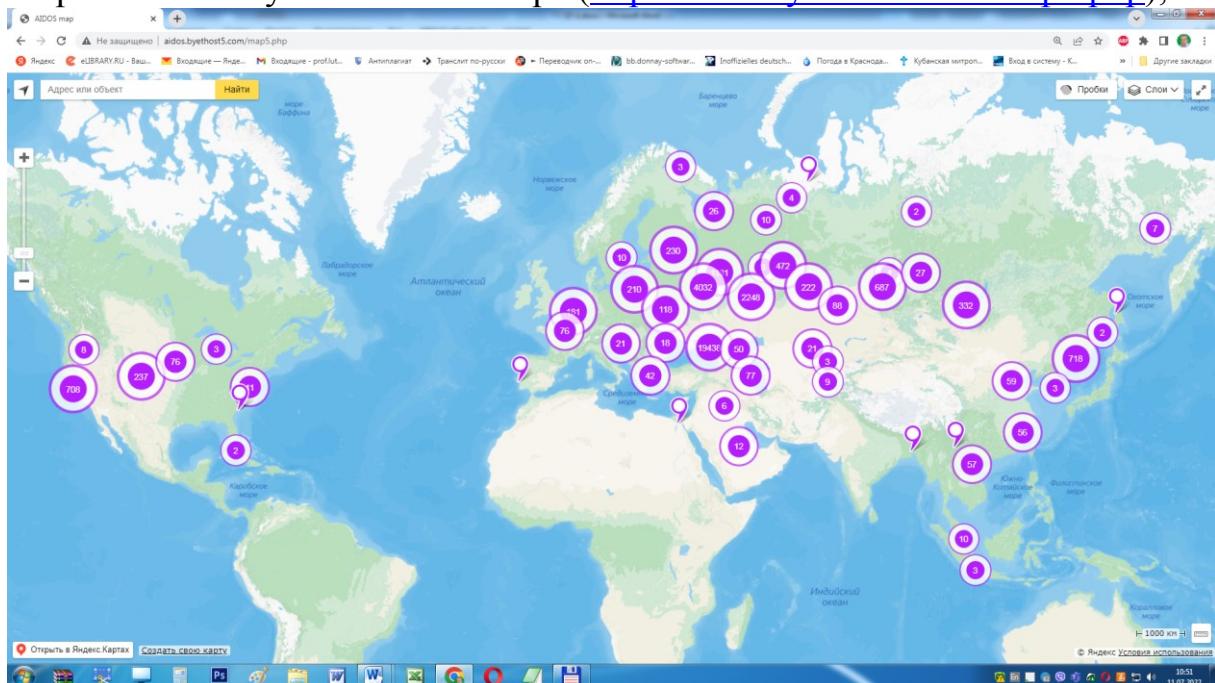
- является одной из первых отечественных систем искусственного интеллекта персонального уровня, т.е. не требует от пользователя специальной подготовки в области технологий искусственного интеллекта:

«имеет нулевой порог входа» (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 локальных (поставляемых с инсталляцией) и облачных учебных и научных Эйдос-приложений (в настоящее время их 31 и более 336, соответственно: http://aidos.byethost5.com/Source_data_applications/WebAppls.htm) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);

- поддерживает on-line среду накопления знаний и обмена ими, широко используется во всем мире (<http://aidos.byethost5.com/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);

– обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач классификации, 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: дает результаты анализа, понятные экспертам на основе их опыта, интуиции и профессиональной компетенции;

– 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 эти данные и тем самым преобразовать их в информацию, а затем преобразовать эту информацию в знания путем ее применения для достижения целей (т.е. для 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 законах⁷.

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

⁷ Ссылка на это краткое описание системы «Эйдос» на английском языке:
http://lc.kubagro.ru/aidos/The_Eidos_en.htm

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 по ссылке: <http://lc.kubagro.ru/aidos/aidos02/PR-4.htm>).

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 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).

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 xb2net 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»: с 2022 года по настоящее время. С 2022 года автор и разработчик системы «Эйдос» проф. Е.В.Луценко вплотную занялся разработкой профессиональной версии системы «Эйдос» на языке Аляска+Экспресс, обеспечивающей обработку больших данных, информации и знаний (Big Data, Big Information, Big Knowledge) с использованием ADS (Advantage Database Server), а также на языке C# (Visual Studio | C#).

На рисунке 1 приведена титульная видеограмма DOS-версии системы «Эйдос», а на рисунке 2 – текущей версии системы «Эйдос»:



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

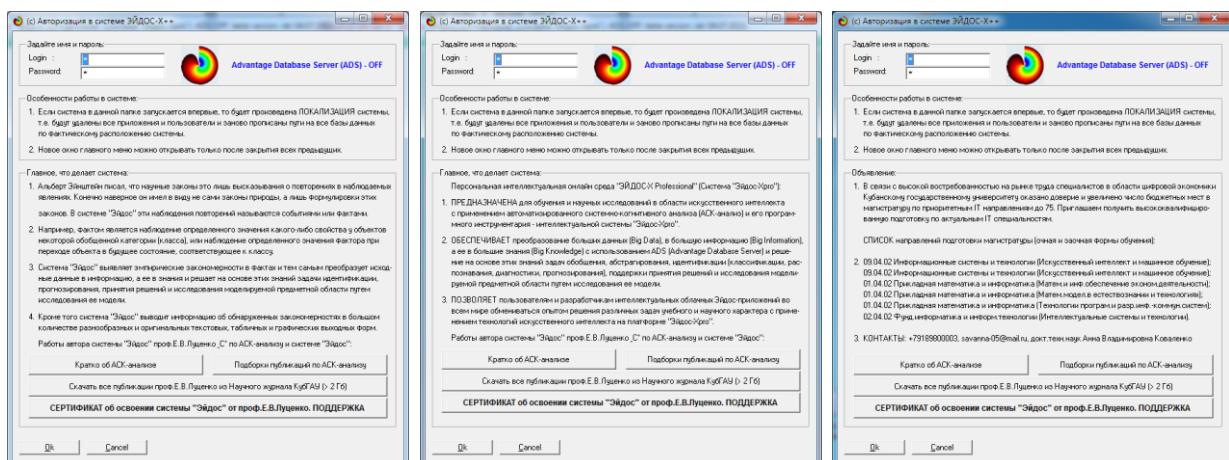


Рисунок 2. Титульные видеограммы текущей версии системы «Эйдос»

2.5. Цель и задачи работы

Целью работы является решение поставленной проблемы.

Как уже показано выше, для работы с лингвистическими переменными целесообразно применить лингвистический АСК-анализ [4].

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

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

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*:

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

2) кластерно-конструктивный анализ классов;

3) кластерно-конструктивный анализ значений описательных шкал;

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

5) нелокальная нейронная сеть;

6) 3d-интегральные когнитивные карты;

7) 2d-интегральные когнитивные карты содержательного сравнения классов (опосредованные нечеткие правдоподобные рассуждения);

8) 2d-интегральные когнитивные карты содержательного сравнения значений факторов (опосредованные нечеткие правдоподобные рассуждения);

9) когнитивные функции;

10) значимость описательных шкал и их градаций;

11) степень детерминированности классов и классификационных шкал).

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**

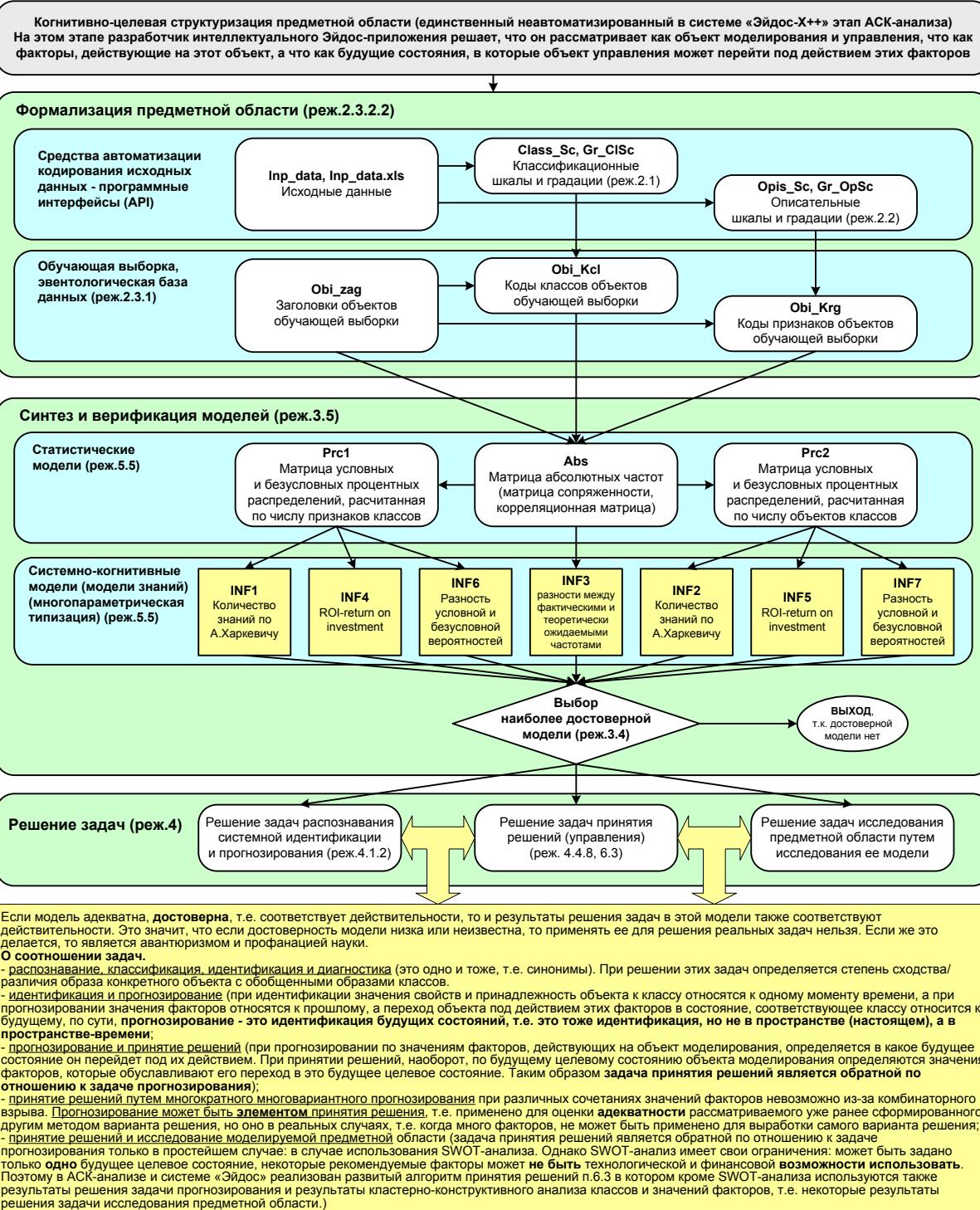


Рисунок 3. Последовательность преобразования исходных данных в информацию, а ее в знания и применения этих знаний для решения различных задач в системе «Эйдос»

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 неавтоматизированном 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** выступают пациенты, в качестве **факторов** анатомических, факторов внешней среды, привычек пациента и т.д. (таблица 1), а в качестве **результатов** действия этих факторов уровень развития раковых опухолей лёгких (таблица 2):

Таблица 1 – Описательные шкалы (факторы)

KOD_OPSC	NAME_OPSC
1	AGE
2	GENDER
3	AIR POLLUTION
4	ALCOHOL USE
5	DUST ALLERGY
6	OCCUPATIONAL HAZARDS
7	GENETIC RISK
8	CHRONIC LUNG DISEASE
9	BALANCED DIET
10	OBESITY
11	SMOKING
12	PASSIVE SMOKER
13	CHEST PAIN
14	COUGHING OF BLOOD
15	FATIGUE
16	WEIGHT LOSS
17	SHORTNESS OF BREATH
18	WHEEZING
19	SWALLOWING DIFFICULTY
20	CLUBBING OF FINGER NAILS
21	FREQUENT COLD
22	DRY COUGH
23	SNORING

Источник: c:\Aidos-X\AID_DATA\A0000006\System\Opis_Sc.dbf

Таблица 2 – Классификационные шкалы (результаты действия факторов)

KOD_CLSC	NAME_CLSC
1	LEVEL

Источник: c:\Aidos-X\AID_DATA\A0000006\System\Class_Sc.dbf

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 в самых различных направлениях науки и решения практических задач в самых различных предметных областях, практически почти везде, где человек применяет естественный интеллект.

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 6).

2.3.2. Программные интерфейсы с внешними базами данных
2.3.2.1. Импорт данных из текстовых файлов
2.3.2.2. Универсальный программный интерфейс импорта данных в систему
2.3.2.3. Импорт данных из транспонированных внешних баз данных
2.3.2.4. Оцифровка изображений по внешним контурам
2.3.2.5. Оцифровка изображений по всем пикселям и спектру
2.3.2.6. Сценарный АСК-анализ символьных и числовых рядов
2.3.2.7. Транспонирование файлов исходных данных
2.3.2.8. Объединение нескольких файлов исходных данных в один
2.3.2.9. Разбиение TXT-файла на файлы-абзацы
2.3.2.10. CSV => DBF конвертер системы "Эйдос"
2.3.2.11. Прогноз событий по астропараметрам по Н.А.Чередниченко
2.3.2.12. Прогнозирование землетрясений методом Н.А.Чередниченко
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 5, 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 7):

Экранные формы управления API-2.3.2.2 системы «Эйдос» с *реальными параметрами*, использованными в данной работе, приведены на рисунках 8.

В таблицах 6, 7, 8 приведены классификационные и описательные шкалы и градации, а также обучающая выборка, сформированные API-2.3.2.2 при параметрах, показанных на рисунках 8.

Для классификационных шкал на 3-м рисунке 8 приведено также количество наблюдений для каждого интервального значения (градации) и его размер. За счет того, что интервальные значения имеют разные размеры удается преодолеть **несбалансированность данных**, т.к. число наблюдений в каждом интервальном значении некоторой шкалы получается равным с точностью 1 (т.к. число наблюдений – всегда целое число).

*Под **несбалансированностью данных** понимается неравномерность распределения значений свойств объекта моделирования или действующих на него факторов по диапазону изменения значений числовых шкал.*

Помощь по режиму 2.3.2.2 для случая Excel-файлов исходных данных

Режим 2.3.2.2: Универсальный программный интерфейс импорта данных из внешней базы данных "Inp_data.xls" в систему "Эйдос-X++" и формализации предметной области.

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

Принцип организации таблицы исходных данных:

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

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

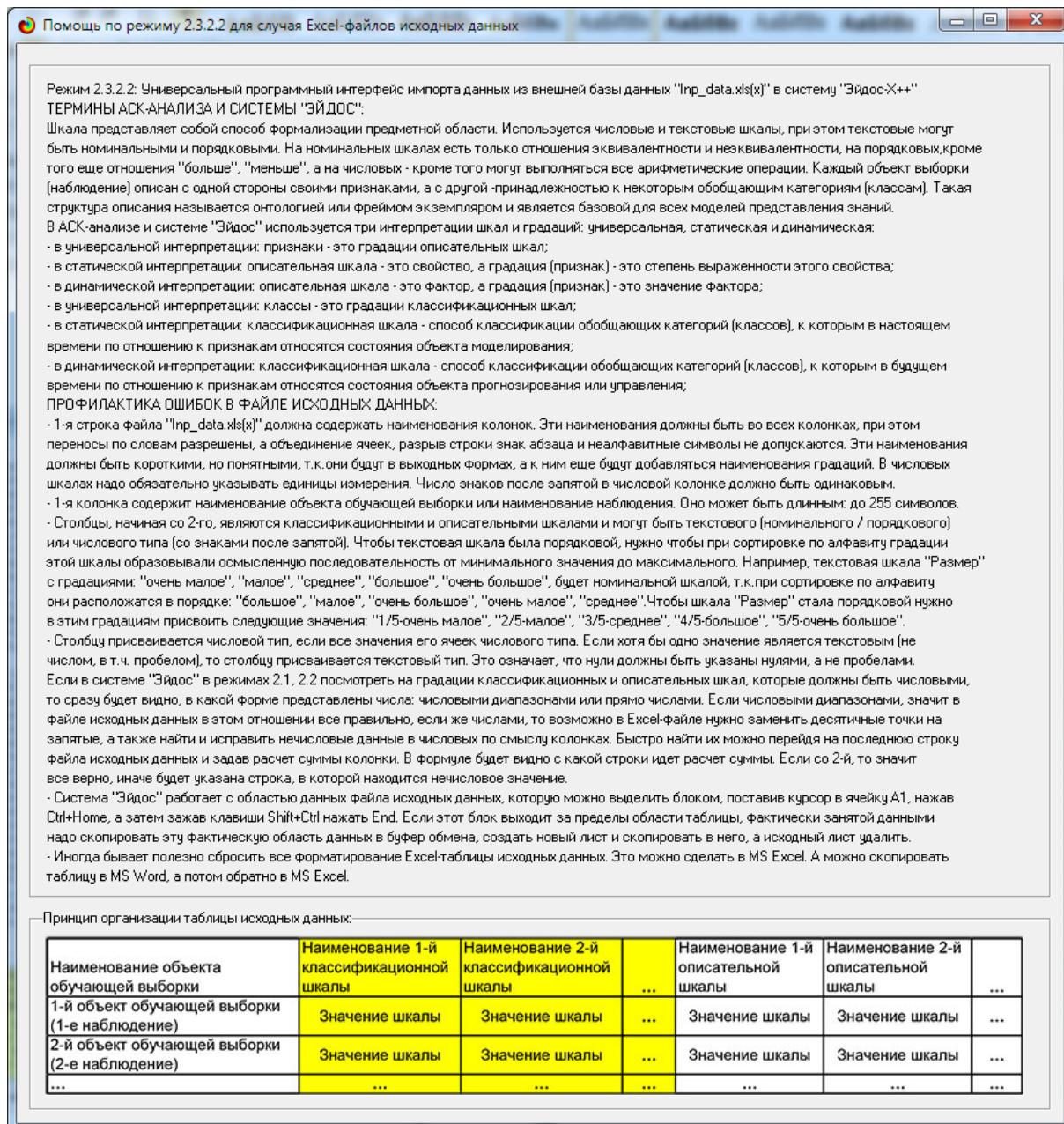


Рисунок 5. Хелпы API-2.3.2.2 системы «Эйдос»

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

— □ ×

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 XLSX- MS Excel-2007(2010) DBF - DBASE IV (DBF/NTX) CSV - CSV => DBF конвертер

Стандарт XLS-файла
Стандарт DBF-файла
Стандарт CSV-файла

Задайте параметры:

Нули и пробелы считать ОТСУТСТВИЕМ данных
 Нули и пробелы считать ЗНАЧЕНИЯМИ данных
 Создавать БД средник 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 спец.интерпретацию текстовых полей классов

Применить сценарный метод АСК-анализа
 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)"]

Ok **Cancel**

 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.: [3 классов x 171 признаков]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс.шкалу	Количество описательных шкал	Количество градаций описательных	Среднее количество градаций на опис.шкалу
Числовые	0	0	0,00	1	5	5,00
Текстовые	1	3	3,00	22	166	7,55
ВСЕГО:	1	3	3,00	23	171	7,43

Задайте количество числовых диапазонов (интервалов, градаций) в шкале:

В описательных шкалах:

Пересчитать шкалы и градации

Выйти на создание модели

2.3.2.2. Параметры классификационных и описательных шкал и градаций

ПАРАМЕТРЫ ШКАЛ И ГРАДАЦИЙ С АДАПТИВНЫМИ ГРАНИЦАМИ И ПРИМЕРНО РАВНЫМ КОЛИЧЕСТВОМ НАБЛЮДЕНИЙ ПО ГРАДАЦИЯМ с коррекцией ошибки округления числа наблюдений по интервалу градации при переходе к следующей градации

КЛАССИФИКАЦИОННАЯ ШКАЛА: код: [1], наим.:" ПЛОТНОСТЬ ПОЧВЫ В СЛОЕ 0-10 CM (Г/СМ3)", набл.на шкалу (всего): 28, тип/число градаций в шкале: "Равное число событий в интервалах"

1 Нам.градации: 1/3-(1.1500000, 1.2200000), размер интервала=0.0700000, расч./факт.число наблюдений на градациях: 9/9

2 Нам.градации: 2/3-(1.2200000, 1.3100000), размер интервала=0.0900000, расч./факт.число наблюдений на градациях: 9/9

3 Нам.градации: 3/3-(1.3100000, 1.4800000), размер интервала=0.1700000, расч./факт.число наблюдений на градациях: 10/10

КЛАССИФИКАЦИОННАЯ ШКАЛА: код: [2], наим.:" ПЛОТНОСТЬ ПОЧВЫ В СЛОЕ 10-20 CM (Г/СМ3)", набл.на шкалу (всего): 28, тип/число градаций в шкале: "Равное число событий в интервалах"

4 Нам.градации: 1/3-(1.1700000, 1.3000000), размер интервала=0.1300000, расч./факт.число наблюдений на градациях: 9/9

5 Нам.градации: 2/3-(1.3000000, 1.3600000), размер интервала=0.0600000, расч./факт.число наблюдений на градациях: 9/9

6 Нам.градации: 3/3-(1.3600000, 1.5400000), размер интервала=0.1800000, расч./факт.число наблюдений на градациях: 10/10

КЛАССИФИКАЦИОННАЯ ШКАЛА: код: [3], наим.:" ПЛОТНОСТЬ ПОЧВЫ В СЛОЕ 20-30 CM (Г/СМ3)", набл.на шкалу (всего): 28, тип/число градаций в шкале: "Равное число событий в интервалах"

7 Нам.градации: 1/3-(1.2500000, 1.3200000), размер интервала=0.0700000, расч./факт.число наблюдений на градациях: 9/9

8 Нам.градации: 2/3-(1.3200000, 1.3800000), размер интервала=0.0600000, расч./факт.число наблюдений на градациях: 9/9

9 Нам.градации: 3/3-(1.3800000, 1.5200000), размер интервала=0.1400000, расч./факт.число наблюдений на градациях: 10/10

КЛАССИФИКАЦИОННАЯ ШКАЛА: код: [4], наим.:" РАЗМЕР АГРЕГАТОВ >25<10 (%)", набл.на шкалу (всего): 36, тип/число градаций в шкале: "Равное число событий в интервалах"/3

10 Нам.градации: 1/3-(50.3000000, 52.9000000), размер интервала=2.6000000, расч./факт.число наблюдений на градациях: 12/12

11 Нам.градации: 2/3-(52.9000000, 65.6000000), размер интервала=12.7000000, расч./факт.число наблюдений на градациях: 12/12

12 Нам.градации: 3/3-(65.6000000, 70.0000000), размер интервала=4.4000000, расч./факт.число наблюдений на градациях: 12/12

КЛАССИФИКАЦИОННАЯ ШКАЛА: код: [5], наим.:" РАЗМЕР АГРЕГАТОВ <0,25>>10 (%)", набл.на шкалу (всего): 36, тип/число градаций в шкале: "Равное число событий в интервалах"/3

13 Нам.градации: 1/3-(30.0000000, 33.7000000), размер интервала=3.7000000, расч./факт.число наблюдений на градациях: 12/12

14 Нам.градации: 2/3-(33.7000000, 41.8000000), размер интервала=8.1000000, расч./факт.число наблюдений на градациях: 12/12

15 Нам.градации: 3/3-(41.8000000, 49.7000000), размер интервала=7.9000000, расч./факт.число наблюдений на градациях: 12/12

КЛАССИФИКАЦИОННАЯ ШКАЛА: код: [6], наим.:" КОЭФФИЦИЕНТ СТРУКТУРНОСТИ", набл.на шкалу (всего): 36, тип/число градаций в шкале: "Равное число событий в интервалах"/3

16 Нам.градации: 1/3-(1.0000000, 1.1000000), размер интервала=0.1000000, расч./факт.число наблюдений на градациях: 12/12

17 Нам.градации: 2/3-(1.1000000, 1.9000000), размер интервала=0.8000000, расч./факт.число наблюдений на градациях: 12/12

18 Нам.градации: 3/3-(1.9000000, 2.3000000), размер интервала=0.4000000, расч./факт.число наблюдений на градациях: 12/12

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [1], наим.:" ОБРАБОТКА ПОЧВЫ", тип шкалы/число градаций в шкале: "Равное число событий в интервалах"/4

1 Нам.градации: 1/4-1.Вспашка (контроль)

2 Нам.градации: 2/4-2.Измелевание

3 Нам.градации: 3/4-3.Дисковое лущение

4 Нам.градации: 4/4-4.Прямой посев

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [2], наим.:" УДОБРЕНИЯ", тип шкалы/число градаций в шкале: "Равное число событий в интервалах"/3

5 Нам.градации: 1/3-N2OP80 + N40

6 Нам.градации: 2/3-N4OP160 + N40

7 Нам.градации: 3/3-Без удобрений (контроль)

ОПИСАТЕЛЬНАЯ ШКАЛА: код: [3], наим.:" ФАЗА ВЕГЕТАЦИИ", тип шкалы/число градаций в шкале: "Равное число событий в интервалах"/4

8 Нам.градации: 1/4-1.Всходы

9 Нам.градации: 2/4-2.Весенне кущение

10 Нам.градации: 3/4-3.Колючение

11 Нам.градации: 4/4-4.Полная спелость

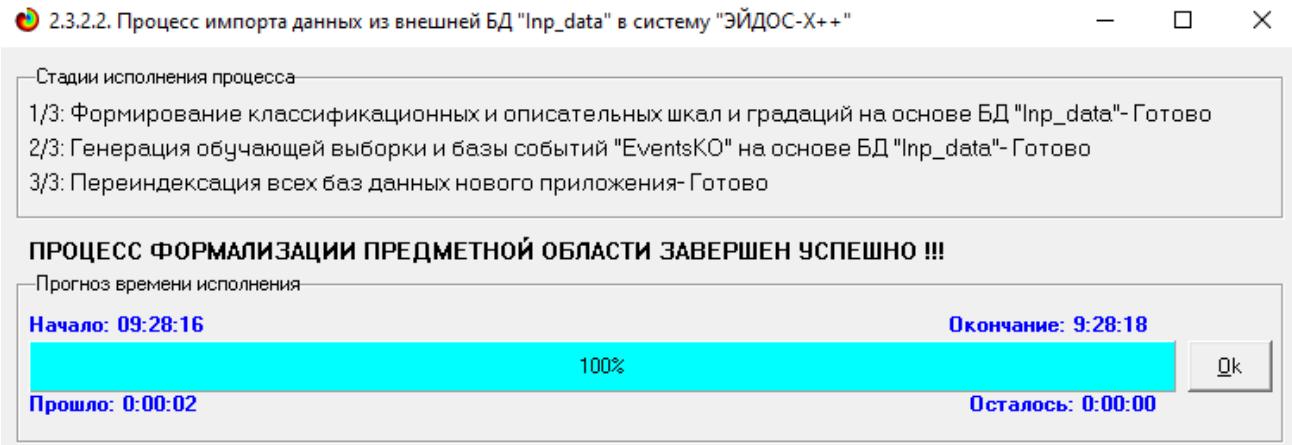


Рисунок 6. Экранные формы управления API-2.3.2.2 системы «Эйдос»

Таблица 3 – Классификационные шкалы и градации (числовые шкалы)

KOD_CLS	NAME_CLS	KOD_CLSC	N_CHRCLSC
1	LEVEL-1-High	1	5
2	LEVEL-2-Medium	1	5
3	LEVEL-3-Low	1	5

Источник: c:\Aidos-X\AID_DATA\A0000002\System\Classes.dbf

Таблица 4 – Описательные шкалы и градации (лингвистические переменные)

KOD_ATR	NAME_ATR	KOD_OPSC	N_CHROPSC	MIN_GRINT	MAX_GRINT	AVR_GRINT	INT_INF	SUM_II	SII_PERC	RANG	ABS	PERC_FIZ	UNIVERSAL	DATE	TIME
1	AGE-1/5-[14.000000, 25.800000]	1	3			0,0802330				43	17,2690763				
2	AGE-2/5-[25.800000, 37.600000]	1	3			0,0486298				105	42,1686747				
3	AGE-3/5-[37.600000, 49.400000]	1	3			0,0429275				66	26,5060241				
4	AGE-4/5-[49.400000, 61.200000]	1	3			0,0638228				17	6,8273092				
5	AGE-5/5-[61.200000, 73.000000]	1	3			0,1294869				18	7,2289157				
6	GENDER-Female	2	6			0,0527098				110	44,1767068				
7	GENDER-Male	2	6			0,0469600				139	55,8232932				
8	AIR POLLUTION-High1	3	13			0,1019015				7	2,8112450				
9	AIR POLLUTION-High2	3	13			0,1019015				4	1,6064257				
10	AIR POLLUTION-Low1	3	13			0,0809843				31	12,4497992				
11	AIR POLLUTION-Low2	3	13			0,1659758				51	20,4819277				
12	AIR POLLUTION-Low3	3	13			0,0829726				43	17,2690763				
13	AIR POLLUTION-Medium1	3	13			0,0895579				25	10,0401606				
14	AIR POLLUTION-Medium2	3	13			0,0550040				5	2,0080321				
15	AIR POLLUTION-Medium3	3	13			0,2553959				83	33,3333333				
16	ALCOHOL USE-High1	4	11			0,2160822				44	17,6706827				
17	ALCOHOL USE-High2	4	11			0,0540390				46	18,4738956				
18	ALCOHOL USE-Low1	4	11			0,0692435				40	16,0642570				
19	ALCOHOL USE-Low2	4	11			0,0584876				42	16,8674699				
20	ALCOHOL USE-Low3	4	11			0,1605567				18	7,2289157				
21	ALCOHOL USE-Medium1	4	11			0,1038660				14	5,624900				
22	ALCOHOL USE-Medium2	4	11			0,0771829				25	10,0401606				
23	ALCOHOL USE-Medium3	4	11			0,1723428				20	8,0321285				
24	DUST ALLERGY-High1	5	12			0,0707202				101	40,5622490				
25	DUST ALLERGY-High2	5	12			0,1312175				3	1,2048193				
26	DUST ALLERGY-Low1	5	12			0,1312174				12	4,8192771				
27	DUST ALLERGY-Low2	5	12			0,1536154				15	6,0240964				
28	DUST ALLERGY-Low3	5	12			0,0692435				20	8,0321285				
29	DUST ALLERGY-Medium1	5	12			0,0692435				35	14,0562249				
30	DUST ALLERGY-Medium2	5	12			0,1909250				32	12,8514056				
31	DUST ALLERGY-Medium3	5	12			0,1426949				31	12,4497992				
32	OCCUPATIONAL HAZARDS-High1	6	20			0,2025896				93	37,3493976				
33	OCCUPATIONAL HAZARDS-High2	6	20			0,1174639				7	2,8112450				
34	OCCUPATIONAL HAZARDS-Low1	6	20			0,1312174				10	4,0160643				
35	OCCUPATIONAL HAZARDS-Low2	6	20			0,0562463				30	12,0481928				
36	OCCUPATIONAL HAZARDS-Low3	6	20			0,2137290				39	15,6626506				
37	OCCUPATIONAL HAZARDS-Medium1	6	20			0,0919874				25	10,0401606				

Источник: c:\Aidos-X\AID_DATA\A0000002\System\Attributes.dbf

Таблица 5 – Обучающая выборка (полностью)

NAME_OBJ	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17	N18	N19	N20	N21	N22	N23	N24	N25	MEMOCLS	MEMOATR
P1	2	7	11	21	30	37	43	49	56	65	72	79	88	99	106	115	121	129	138	145	153	161	169	3		
P10	1	7	12	18	30	36	44	49	56	63	71	81	88	98	104	114	118	127	141	146	152	158	167	2		
P100	2	7	13	22	31	38	45	51	60	61	71	80	90	94	102	110	119	129	136	148	157	158	167	1		
P1000	2	7	8	16	24	32	46	47	54	61	68	76	84	94	107	113	122	128	139	149	157	158	170	1		
P101	3	7	15	17	24	32	40	53	54	61	69	76	84	95	106	113	123	128	139	146	155	160	168	1		
P102	2	7	13	22	31	38	45	51	60	61	71	80	90	94	102	110	119	129	136	148	157	158	167	1		
P103	4	6	11	21	30	37	43	49	56	65	72	79	88	99	106	115	121	129	138	145	153	161	169	3		
P104	2	6	12	18	29	36	42	50	58	64	70	81	89	96	106	113	121	131	137	146	154	162	168	3		
P105	2	6	13	22	31	38	46	52	59	66	75	83	92	100	104	115	122	129	139	150	153	162	166	2		
P106	3	7	11	20	29	35	44	50	57	64	71	80	90	99	104	113	123	133	140	148	153	159	170	2		
P107	3	7	15	16	24	32	40	53	54	61	68	77	84	93	108	114	121	126	135	146	155	163	168	1		
P108	5	6	15	17	24	32	40	53	54	61	68	77	84	93	103	117	124	126	137	148	154	159	169	1		
P109	3	6	13	22	31	39	45	51	60	67	75	83	92	101	108	114	121	131	138	145	151	163	171	2		
P11	2	7	15	16	24	32	46	47	54	61	68	76	84	94	107	113	122	128	139	149	157	158	170	1		
P110	2	6	12	18	29	35	43	49	57	64	71	79	90	97	105	113	122	131	136	149	153	164	167	3		
P111	5	7	14	23	31	38	46	52	60	66	69	82	91	100	107	114	125	129	136	146	152	164	167	2		
P112	1	7	12	18	30	36	44	49	56	63	71	81	88	98	104	114	118	127	141	146	152	158	167	2		
P113	2	7	15	16	24	32	46	47	54	61	68	76	84	94	107	113	122	128	139	149	157	158	170	1		
P114	2	7	15	16	24	32	40	47	60	61	68	76	84	93	102	116	118	133	134	143	151	164	167	1		
P115	1	7	11	21	30	39	45	52	58	67	74	81	92	100	108	114	121	128	139	142	153	159	171	2		
P116	1	7	15	17	24	32	46	47	54	64	69	76	86	101	108	113	124	129	138	146	152	158	171	1		
P117	4	6	13	22	31	38	45	51	60	61	71	80	90	94	102	110	119	129	136	148	157	158	167	1		
P118	5	7	15	17	24	32	40	53	54	61	69	76	84	95	106	113	123	128	139	146	155	160	168	1		
P119	2	6	15	16	24	32	40	53	54	61	68	76	84	93	105	110	125	126	141	142	153	161	166	1		
P12	2	7	15	16	24	32	40	47	60	61	68	76	84	93	102	116	118	133	134	143	151	164	167	1		
P120	5	7	15	17	24	32	40	53	56	65	70	79	90	98	105	110	125	132	136	144	154	162	167	2		
P121	3	6	11	18	30	36	42	50	56	65	70	81	88	99	109	110	121	132	135	145	154	160	168	2		
P122	1	7	12	19	29	35	43	49	57	64	71	79	89	98	107	116	125	132	140	148	157	163	169	2		
P123	2	7	15	16	24	32	40	53	54	61	73	77	84	93	107	115	124	133	140	149	155	164	170	1		
P124	2	6	10	23	24	38	43	49	60	63	72	80	88	97	106	114	118	126	139	143	151	158	170	2		
P125	2	6	11	23	27	36	46	53	60	65	75	77	84	101	108	116	123	133	140	148	157	163	165	1		
P126	3	7	11	21	30	39	45	52	58	67	68	76	88	98	102	110	118	130	135	144	152	164	167	1		
P127	2	6	10	23	24	33	40	53	54	61	72	81	85	93	106	113	125	131	137	147	152	160	166	2		
P128	2	7	11	21	30	37	43	49	56	65	72	79	88	99	106	115	121	129	138	145	153	161	169	3		
P129	1	6	12	18	29	36	42	50	58	64	70	81	89	96	106	113	121	131	137	146	154	162	168	3		
P13	1	7	11	21	30	39	45	52	58	67	74	81	92	100	108	114	121	128	139	142	153	159	171	2		
P130	2	6	12	18	29	35	43	49	57	64	71	79	90	97	105	113	122	131	136	149	153	164	167	3		
P131	2	7	15	16	25	32	46	47	54	63	73	80	84	94	105	114	125	131	137	147	152	160	166	3		

Источник: c:\Aidos-X\AID_DATA\A0000002\System\EventsKO.dbf

Отметим, что в системе «Эйдос» обычно используются базы данных с расширением «dbf». Они открываются в MS Excel или могут быть конвертированы в файлы xls,xlsx с помощью онлайн-сервисов.

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

Синтез и верификация статистических и системно-когнитивных моделей (СК-моделей) моделей осуществляется в режиме 3.5 системы «Эйдос». Математические модели, на основе которых рассчитываются статистические и СК-модели, подробно описаны в ряде монографий и статей автора. Поэтому в данной работе мы рассмотрим эти вопросы очень кратко. Отметим лишь, 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) (рисунок 3).

Математическая модель АСК-анализа и системы «Эйдос» основана на системной нечеткой интервальной математике и обеспечивает [2, 3] сопоставимую обработку больших объемов фрагментированных и зашумленных взаимозависимых данных, представленных в различных типах шкал (номинальных, порядковых и числовых) и различных единицах измерения.

Структура математической модели АСК-анализа состоит в следующем.

Непосредственно на основе эмпирических данных (см. Help режима 2.3.2.2) рассчитывается матрица абсолютных частот (таблица 9).

На ее основе рассчитываются матрицы условных и безусловных процентных распределений (таблица 10).

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_{Σ_j} используется суммарное количество признаков по классу;

2-й способ: в качестве N_{Σ_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 9) было бы очень неразумно и переход от абсолютных частот к условным и безусловным относительным частотам (частостям) (таблица 10) является весьма обоснованным и логичным.

Таблица 6 – 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			$N_{\Sigma j}$				$N_{\Sigma\Sigma} = \sum_{j=1}^w N_{\Sigma j}$

Таблица 7 – 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		$P_{\Sigma j}$				

Этот переход полностью снимает проблему несбалансированности данных, т.к. в последующем анализе используется не матрица абсолютных частот (таблица 9), а матрицы условных и безусловных процентных распределений (таблица 10) и матрицы системно-когнитивных моделей (СК-модели, таблица 12), в частности матрица информативностей.

Этот подход снимает также проблему обеспечения сопоставимости обработки в одной модели исходных данных, представленных в различных видах шкал (номинальных, порядковых и числовых) и в разных единицах измерения [6].

В системе «Эйдос» этот подход применяется *всегда* при решении любых задач.

Затем на основе таблиц 9 и 10 с использованием *частных критериеv, знаний* приведенных таблице 11, рассчитываются матрицы семи системно-когнитивных моделей (таблица 12).

В таблице 11 приведены формулы:

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

И это сравнение в таблицах 9 и 10 осуществляется двумя возможными способами: путем **вычитания** и путем **деления**.

Количество частных критериев знаний и основанных на них системно-когнитивных моделей (таблица 11), применяемых в настоящее время в системе «Эйдос» равное 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**.

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 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).

Таблица 8 – 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)		$N_i = \sum_{j=1}^W N_{ij}; N_j = \sum_{i=1}^M N_{ij}; N = \sum_{i=1}^W \sum_{j=1}^M N_{ij};$ $N_{ij} - \text{фактическая частота},$ $\bar{N}_{ij} = \frac{N_i N_j}{N} - \text{теоретическая частота}.$
PRC1 , матрица условных P_{ij} и безусловных P_i процентных	--	$P_{ij} = \frac{N_{ij}}{N_j}; P_i = \frac{N_i}{N}$

распределений, в качестве N_j используется суммарное количество признаков по классу		
PRC2 , матрица условных P_{ij} и безусловных P_i процентных распределений, в качестве N_j используется суммарное количество объектов обучающей выборки по классу		
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.		
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		
INF6 , частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: N_j – суммарное количество признаков по j -му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$
INF7 , частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: N_j – суммарное количество объектов по j -му классу		

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.

Таким образом, мы видим, что **все частные критерии знаний тесно взаимосвязаны друг с другом**. Особенно интересна связь знаменитого

критерия хи-квадрат 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%.

Таблица 9 – Матрица системно-когнитивной модели

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

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 [6].

На основе системно-когнитивных моделей, представленных в таблице 12 (отличаются частыми критериями, приведенными в таблице 11), 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 вариабельность значений частных критерииев этого значения фактора, класса или модели в целом (таблица 12).

Численно эта вариабельность может измеряться разными способами, например средним отклонением модулей частных критерииев 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 13):

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	$\sigma_{\Sigma} = \sqrt{\frac{1}{W-1} \sum_{j=1}^W (I_{ij} - \bar{I}_i)^2}$
2	1. Степень детерминированности класса. 2. Степень обусловленности класса.	Корень из информационной мощности класса	$\sigma_{\Sigma_j} = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (I_{ij} - \bar{I}_j)^2}$
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	$H = \sqrt{\frac{1}{(W \cdot M - 1)} \sum_{j=1}^W \sum_{i=1}^M (I_{ij} - \bar{I})^2}$

В системе «Эйдос» синтез моделей производится в режиме 3.5 (рисунок 9):

 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 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: Хи-квадрат, разности между фактическими и ожидаемыми abs.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 безусл.вероятностей; вероятности из PRC1

10. INF7 - частный критерий: разн. усл. и безусл.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

[Пояснение по алгоритму верификации](#)

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:

[Подробнее](#)

Удалять из обуч.выборки скопированные объекты:

Не удалять
 Удалять

Измеряется внутренняя достоверн. модели

Для каждой заданной модели выполнить:

Синтез и верификацию
 Только верификацию
 Только синтез

Задайте процессор:

CPU GPU

Задайте алгоритм:

Классика - дальше
 Упрощенно-быстрее

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

Расчетный размер БД результатов распознавания Rasp.dbf равен 37544 байт, т.е.: 0.0017483 % от MAX-возможного, (от 2Гб)

Задайте, сколько % от исходной БД Rasp.dbf ОСТАВИТЬ, удаляя наименее достоверные результаты распознавания:

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

[Ok](#) [Cancel](#)

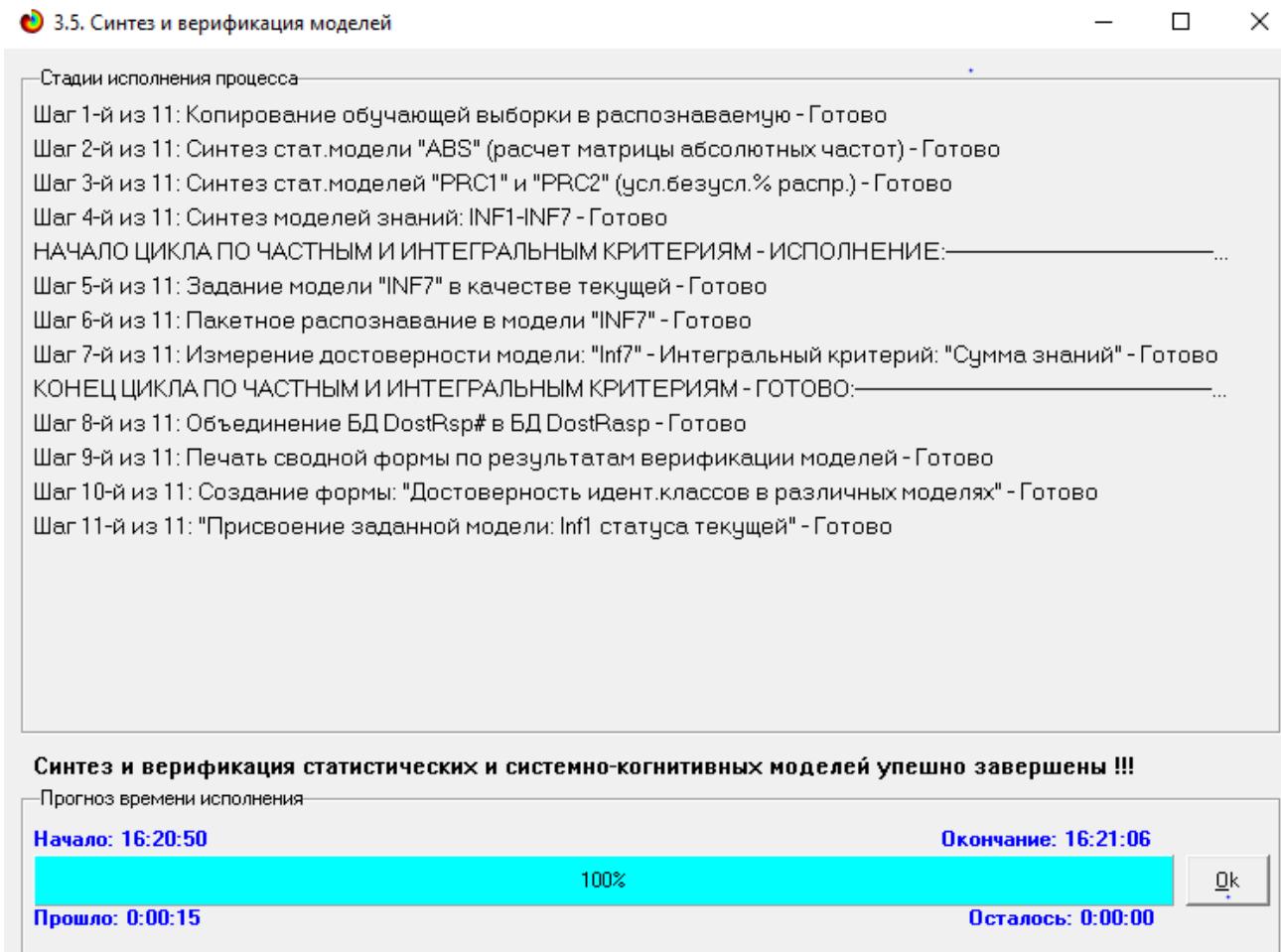


Рисунок 7. Экранные формы режима синтеза и верификации моделей

В результате работы режима 3.5 создано 3 статистических и 7 системно-когнитивных моделей, некоторые из которых приведены на рисунках 10-13:

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"							
Код признака	Наименование описательной шкалы и градации	1. LEVEL 1 HIGH	2. LEVEL 2 MEDIUM	3. LEVEL 3 LOW	Сумма	Среднее	Средн. квад.
1.0	AGE-1/5(14.000000, 25.800000)	10.0	15.0	18.0	43.0	14.33	4.04
2.0	AGE-2/5(25.800000, 37.600000)	51.0	26.0	28.0	105.0	35.00	13.89
3.0	AGE-3/5(37.600000, 49.400000)	21.0	28.0	17.0	66.0	22.00	5.57
4.0	AGE-4/5(49.400000, 61.200000)	6.0	4.0	7.0	17.0	5.67	1.53
5.0	AGE-5/5(61.200000, 73.000000)	7.0	9.0	2.0	18.0	6.00	3.61
6.0	GENDER-Female	32.0	35.0	43.0	110.0	36.67	5.69
7.0	GENDER-Male	63.0	47.0	29.0	139.0	46.33	17.01
8.0	AIR POLLUTION-High1	7.0			7.0	2.33	4.04
9.0	AIR POLLUTION-High2	4.0			4.0	1.33	2.31
10.0	AIR POLLUTION-Low1		22.0	9.0	31.0	10.33	11.06
11.0	AIR POLLUTION-Low2	6.0	19.0	26.0	51.0	17.00	10.15
12.0	AIR POLLUTION-Low3		15.0	28.0	43.0	14.33	14.01
13.0	AIR POLLUTION-Medium1	14.0	7.0	4.0	25.0	8.33	5.13
14.0	AIR POLLUTION-Medium2			3.0	5.0	1.67	1.53
15.0	AIR POLLUTION-Medium3	64.0	16.0	3.0	83.0	27.67	32.13
16.0	ALCOHOL USE-High1	41.0			44.0	14.67	22.85
17.0	ALCOHOL USE-High2	30.0	16.0		46.0	15.33	15.01
18.0	ALCOHOL USE-Low1		16.0	24.0	40.0	13.33	12.22
19.0	ALCOHOL USE-Low2		19.0	23.0	42.0	14.00	12.29
20.0	ALCOHOL USE-Low3		3.0	15.0	18.0	6.00	7.94
21.0	ALCOHOL USE-Medium1	3.0	4.0	7.0	14.0	4.67	2.08
22.0	ALCOHOL USE-Medium2	18.0	7.0		25.0	8.33	9.07
23.0	ALCOHOL USE-Medium3	3.0	17.0		20.0	6.67	9.07
24.0	DUST ALLERGY-High1	71.0	30.0		101.0	33.67	35.64
25.0	DUST ALLERGY-High2				3.0	1.00	1.73
26.0	DUST ALLERGY-Low1				12.0	4.00	6.93
27.0	DUST ALLERGY-Low2		3.0		12.0	5.00	6.24

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

Код признака	Наименование описательной шкалы и градации	1. LEVEL 1 HIGH	2. LEVEL 2 MEDIUM	3. LEVEL 3 LOW	Безусл. вероятн.	Среднее	Средн. квадр.
1.0	AGE-1/5-(14.0000000, 25.8000000)	10.526	18.293	25.000	17.269	17.940	7.254
2.0	AGE-2/5-(25.8000000, 37.6000000)	53.684	31.707	38.889	42.169	41.427	11.217
3.0	AGE-3/5-(37.6000000, 49.4000000)	22.105	34.146	23.611	26.506	26.621	6.571
4.0	AGE-4/5-(49.4000000, 61.2000000)	6.316	4.878	9.722	6.827	6.972	2.499
5.0	AGE-5/5-(61.2000000, 73.0000000)	7.368	10.976	2.778	7.229	7.041	4.120
6.0	GENDER-Female	33.684	42.683	59.722	44.177	45.363	13.235
7.0	GENDER-Male	66.316	57.317	40.278	55.823	54.637	13.235
8.0	AIR POLLUTION-High1	7.368			2.811	2.456	4.265
9.0	AIR POLLUTION-High2	4.211			1.606	1.404	2.442
10.0	AIR POLLUTION-Low1		26.829	12.500	12.450	13.110	13.436
11.0	AIR POLLUTION-Low2	6.316	23.171	36.111	20.482	21.866	14.951
12.0	AIR POLLUTION-Low3		18.293	38.889	17.269	19.061	19.467
13.0	AIR POLLUTION-Medium1	14.737	8.537	5.556	10.040	9.610	4.695
14.0	AIR POLLUTION-Medium2		3.659	2.778	2.008	2.145	1.920
15.0	AIR POLLUTION-Medium3	67.368	19.512	4.167	33.333	30.349	32.976
16.0	ALCOHOL USE-High1	43.158		4.167	17.671	15.775	23.817
17.0	ALCOHOL USE-High2	31.579	19.512		18.474	17.030	15.946
18.0	ALCOHOL USE-Low1		19.512	33.333	16.064	17.615	16.758
19.0	ALCOHOL USE-Low2		23.171	31.944	16.867	18.372	16.515
20.0	ALCOHOL USE-Low3		3.659	20.833	7.229	8.164	11.134
21.0	ALCOHOL USE-Medium1	3.158	4.878	9.722	5.622	5.919	3.415
22.0	ALCOHOL USE-Medium2	18.947	8.537		10.040	9.161	9.500
23.0	ALCOHOL USE-Medium3	3.158	20.732		8.032	7.963	11.181
24.0	DUST ALLERGY-High1	74.737	36.585		40.562	37.107	37.382
25.0	DUST ALLERGY-High2			4.167	1.205	1.389	2.416
26.0	DUST ALLERGY-Low1			16.667	4.819	5.556	9.633
27.0	DUST ALLERGY-Low2	3.158		16.667	6.024	6.608	8.864

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

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

Код признака	Наименование описательной шкалы и градации	1. LEVEL 1 HIGH	2. LEVEL 2 MEDIUM	3. LEVEL 3 LOW	Сумма	Среднее	Средн. квадр.
1.0	AGE-1/5-(14.000000, 25.800000)	-0.091	0.011	0.068	-0.012	-0.004	0.080
2.0	AGE-2/5-(25.800000, 37.600000)	0.044	-0.052	-0.015	-0.023	-0.008	0.049
3.0	AGE-3/5-(37.600000, 49.400000)	-0.033	0.046	-0.021	-0.008	-0.003	0.043
4.0	AGE-4/5-(49.400000, 61.200000)	-0.014	-0.062	0.065	-0.011	-0.004	0.064
5.0	AGE-5/5-(61.200000, 73.000000)	0.004	0.076	-0.175	-0.095	-0.032	0.129
6.0	GENDER-Female	-0.050	-0.006	0.055	-0.001	0.000	0.053
7.0	GENDER-Male	0.032	0.005	-0.060	-0.023	-0.008	0.047
8.0	AIR POLLUTION-High1	0.176			0.176	0.059	0.102
9.0	AIR POLLUTION-High2	0.176			0.176	0.059	0.102
10.0	AIR POLLUTION-Low1		0.141	0.001	0.141	0.047	0.081
11.0	AIR POLLUTION-Low2	-0.215	0.023	0.104	-0.089	-0.030	0.166
12.0	AIR POLLUTION-Low3		0.011	0.149	0.159	0.053	0.083
13.0	AIR POLLUTION-Medium1	0.070	-0.030	-0.108	-0.068	-0.023	0.090
14.0	AIR POLLUTION-Medium2		0.110	0.059	0.169	0.056	0.055
15.0	AIR POLLUTION-Medium3	0.129	-0.098	-0.381	-0.350	-0.117	0.255
16.0	ALCOHOL USE-High1	0.164			-0.265	-0.101	0.216
17.0	ALCOHOL USE-High2	0.098	0.010		0.108	0.036	0.054
18.0	ALCOHOL USE-Low1		0.036	0.134	0.169	0.056	0.069
19.0	ALCOHOL USE-Low2		0.058	0.117	0.175	0.058	0.058
20.0	ALCOHOL USE-Low3		-0.125	0.194	0.069	0.023	0.161
21.0	ALCOHOL USE-Medium1	-0.106	-0.026	0.100	-0.031	-0.010	0.104
22.0	ALCOHOL USE-Medium2	0.116	-0.030		0.087	0.029	0.077
23.0	ALCOHOL USE-Medium3	-0.171	0.174		0.003	0.001	0.172
24.0	DUST ALLERGY-High1	0.112	-0.019		0.093	0.031	0.071
25.0	DUST ALLERGY-High2			0.227	0.227	0.076	0.131
26.0	DUST ALLERGY-Low1			0.227	0.227	0.076	0.131
27.0	DUST ALLERGY-Low2	-0.118		0.186	0.068	0.023	0.154

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

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

Код признака	Наименование описательной шкалы и градации	1. LEVEL 1 HIGH	2. LEVEL 2 MEDIUM	3. LEVEL 3 LOW	Сумма	Среднее	Средн. квадр.
1.0	AGE-1/5-(14.000000, 25.800000)	-6.406	0.839	5.566			6.030
2.0	AGE-2/5-(25.800000, 37.600000)	10.940	-8.578	-2.361			9.971
3.0	AGE-3/5-(37.600000, 49.400000)	-4.181	6.265	-2.084			5.526
4.0	AGE-4/5-(49.400000, 61.200000)	-0.486	-1.598	2.084			1.889
5.0	AGE-5/5-(61.200000, 73.000000)	0.133	3.072	-3.205			3.141
6.0	GENDER-Female	-9.968	-1.225	11.193			10.633
7.0	GENDER-Male	9.968	1.225	-11.193			10.633
8.0	AIR POLLUTION-High1	4.329	-2.305	-2.024			3.752
9.0	AIR POLLUTION-High2	2.474	-1.317	-1.157			2.144
10.0	AIR POLLUTION-Low1	-11.827	11.791	0.036			11.809
11.0	AIR POLLUTION-Low2	-13.458	2.205	11.253			12.502
12.0	AIR POLLUTION-Low3	-16.406	0.839	15.566			16.002
13.0	AIR POLLUTION-Medium1	4.462	-1.233	-3.229			3.991
14.0	AIR POLLUTION-Medium2	-1.908	1.353	0.554			1.700
15.0	AIR POLLUTION-Medium3	32.333	-11.333	-21.000			28.416
16.0	ALCOHOL USE-High1	24.213	-14.490	-9.723			21.104
17.0	ALCOHOL USE-High2	12.450	0.851	-13.301			12.897
18.0	ALCOHOL USE-Low1	-15.261	2.827	12.434			14.062
19.0	ALCOHOL USE-Low2	-16.024	5.169	10.855			14.166
20.0	ALCOHOL USE-Low3	-6.867	-2.928	9.795			8.709
21.0	ALCOHOL USE-Medium1	-2.341	-0.610	2.952			2.699
22.0	ALCOHOL USE-Medium2	8.462	-1.233	-7.229			7.918
23.0	ALCOHOL USE-Medium3	-4.631	10.414	-5.783			9.037
24.0	DUST ALLERGY-High1	32.466	-3.261	-29.205			30.964
25.0	DUST ALLERGY-High2	-1.145	-0.988	2.133			1.848
26.0	DUST ALLERGY-Low1	-4.578	-3.952	8.530			7.394
27.0	DUST ALLERGY-Low2	-2.723	-4.940	7.663			6.728

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

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

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 Ризбергена, а также по критериям L1- L2-мерам проф. Е.В.Луценко, которые предложены для того, чтобы смягчить или полностью преодолеть некоторые недостатки F-меры [12].

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 F-мерой Ван Ризбергена наиболее достоверной является СК-модель INF4 с интегральным критерием: «Семантический резонанс знаний»: F=0.867 (1-й рисунок 14).

В соответствии с критерием достоверности L1-мерой проф. Е.В.Луценко наиболее достоверной является СК-модель INF3 (хи-квадрат К.Пирсона) с интегральным критерием: «Сумма знаний»: L1=0.961 (2-й рисунок 14). **Эту модель и будем использовать для решения поставленных в работе задач.**

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	Гипотеза Ван Ризбергена	Сумма модел. уловлен. сущ., истинно-пол.	Сумма модел. уловлен. сущ., истинно-отриц.	Сумма модел. уловлен. сущ., ложноПол.	Сумма модел. уловлен. сущ., истинно-отриц. решений (SFP)	Сумма модел. уловлен. сущ., истинно-пол.	S-Точность модели	S-Полнота модели	L1-мера проф. Е.В.Луценко
1. ABS - частный критерий количества встреч сомнений: "числ. встреч сомнений" / общее количество встреч сомнений	249	247	232	266	2	0.481	0.992	0.648	156.203	25.405	55.283	0.014	0.739	1.000	0.850		
1. ABS - частный критерий количества встреч сомнений: "числ. встреч сомнений" / общее количество встреч сомнений	249	249	498	0.333	1.000	0.500	150.797	143.756		0.512	1.000	0.677					
2. PRCT - частный критерий ул. вероятность что признака сущ. в сущности	249	247	232	266	2	0.481	0.992	0.648	156.203	25.405	55.284	0.014	0.739	1.000	0.850		
2. PRCT - частный критерий ул. вероятность что признака сущ. в сущности	249	249	498	0.333	1.000	0.500	167.227	166.232		0.501	1.000	0.668					
3. PRC2 - частный критерий условная вероятность что признака сущ. в сущности	249	247	232	266	2	0.481	0.992	0.648	156.203	25.405	55.283	0.014	0.739	1.000	0.850		
3. PRC2 - частный критерий условная вероятность что признака сущ. в сущности	249	249	498	0.333	1.000	0.500	167.227	166.232		0.501	1.000	0.668					
4. INF - частный критерий условная вероятность что признака сущ. в сущности	249	225	452	46	24	0.830	0.904	0.865	113.207	208.999	3.987	4.439	0.966	0.962	0.984		
4. INF - частный критерий количество знаний по А.Хоркевич, в. Семантический резонанс энз.	249	241	381	117	8	0.673	0.968	0.794	132.157	148.273	14.420	2.185	0.902	0.984	0.941		
5. INF - частный критерий количество знаний по А.Хоркевич, в. Семантический резонанс энз.	249	225	452	46	24	0.830	0.904	0.865	113.207	208.999	3.987	4.439	0.966	0.962	0.984		
5. INF - частный критерий количество знаний по А.Хоркевич, в. Семантический резонанс энз.	249	241	381	117	8	0.673	0.968	0.794	132.157	148.273	14.420	2.185	0.902	0.984	0.941		
6. INF - частный критерий Хиккерт, различия между фактами	249	237	361	137	12	0.634	0.952	0.761	136.256	168.411	16.931	0.639	0.889	0.995	0.539		
6. INF - частный критерий Хиккерт, различия между фактами	249	237	361	137	12	0.634	0.952	0.761	113.946	125.749	12.232	0.430	0.903	0.996	0.347		
7. INF - частный критерий ROI (Return On Investment), вероятн.	249	240	495	7	9	0.972	0.984	0.968	114.143	205.275	0.655	2.209	0.994	0.981	0.988		
7. INF - частный критерий ROI (Return On Investment), вероятн.	249	249	311	187	0.571	1.000	0.727	123.104	38.240	24.609	0.833	1.000	0.909				
8. INF - частный критерий ROI (Return On Investment), вероятн.	249	240	491	2	9	0.972	0.984	0.968	114.143	205.275	0.655	2.209	0.994	0.981	0.988		
8. INF - частный критерий ROI (Return On Investment), вероятн.	249	249	311	187	0.571	1.000	0.727	123.104	38.240	24.609	0.833	1.000	0.909				
9. INF - частный критерий разн. безул.вероятност. вер.	249	226	373	125	23	0.644	0.908	0.753	123.548	141.523	14.116	3.018	0.897	0.976	0.935		
9. INF - частный критерий разн. безул.вероятност. вер.	249	237	316	182	12	0.566	0.952	0.710	124.440	60.171	28.655	0.551	0.813	0.996	0.895		
10. INF - частный критерий разн. безул.вероятност. вер.	249	226	373	125	23	0.644	0.908	0.753	123.548	141.523	14.116	3.018	0.897	0.976	0.935		
10. INF - частный критерий разн. безул.вероятност. вер.	249	237	316	182	12	0.566	0.952	0.710	124.440	60.171	28.655	0.551	0.813	0.996	0.895		

Рисунок 12. Экранные формы режима измерения достоверности моделей 3.4

На рисунках 15 приведены частотные распределения количества истинных и ложных, положительных и отрицательных решений в наиболее достоверной по L1-мере проф. Е.В.Луценко СК-модели INF3.

Из этих частотных распределений видно, что в наиболее достоверной по критериюм достоверности L1-мерой проф.Е.В.Луценко СК-модели INF3:

- отрицательные ложные решения вообще практически не встречаются не встречаются, за исключением 1-го случая при уровне различия –10%;
- при уровнях сходства меньше 40% преобладают ложные положительные решения, а при более высоких уровнях сходства – истинные положительные решения. При уровнях сходства выше 50% ложных положительных решений вообще нет;
- *чем выше уровень сходства, тем большее 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.*

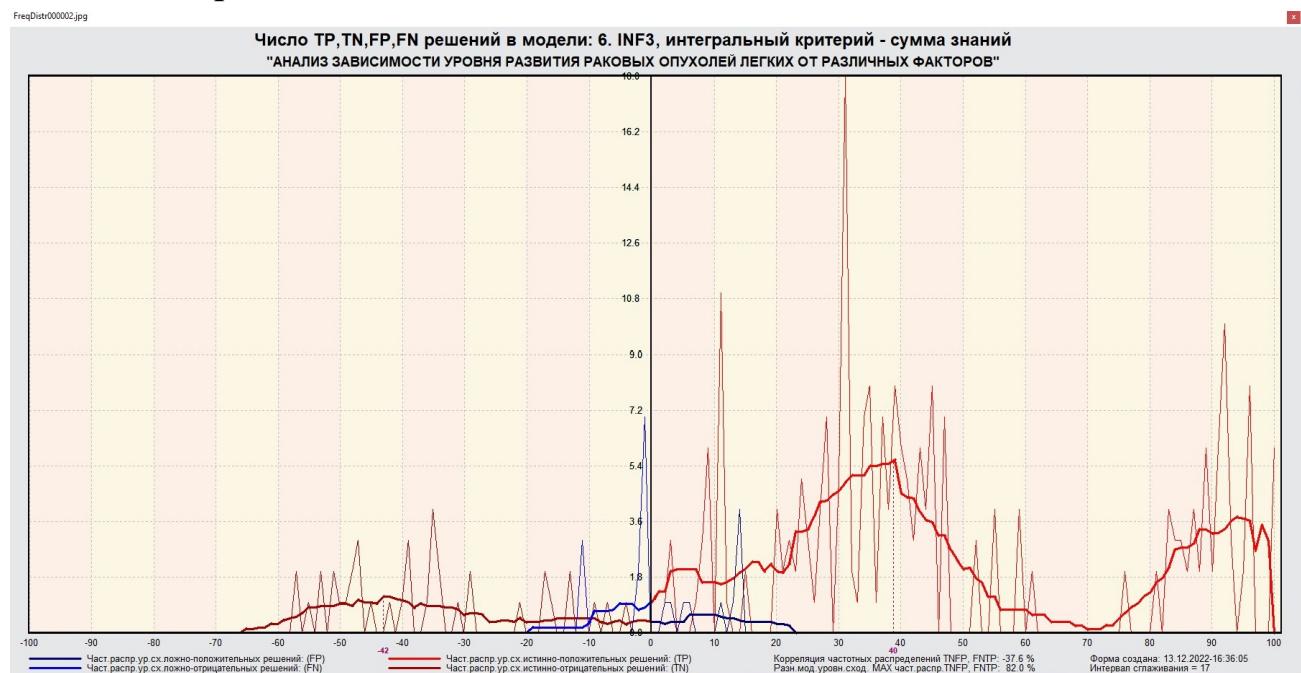


Рисунок 13. Частотные распределения количества истинных и ложных, положительных и отрицательных решений в наиболее достоверной по L1-мере проф.Е.В.Луценко СК-модели INF3

На рисунках 16 приведены экранные формы хелпов режима 3.4, в которых подробно объясняется смысл этого режима. Эти формы приводятся в работе вместо более детального описания данного режима.

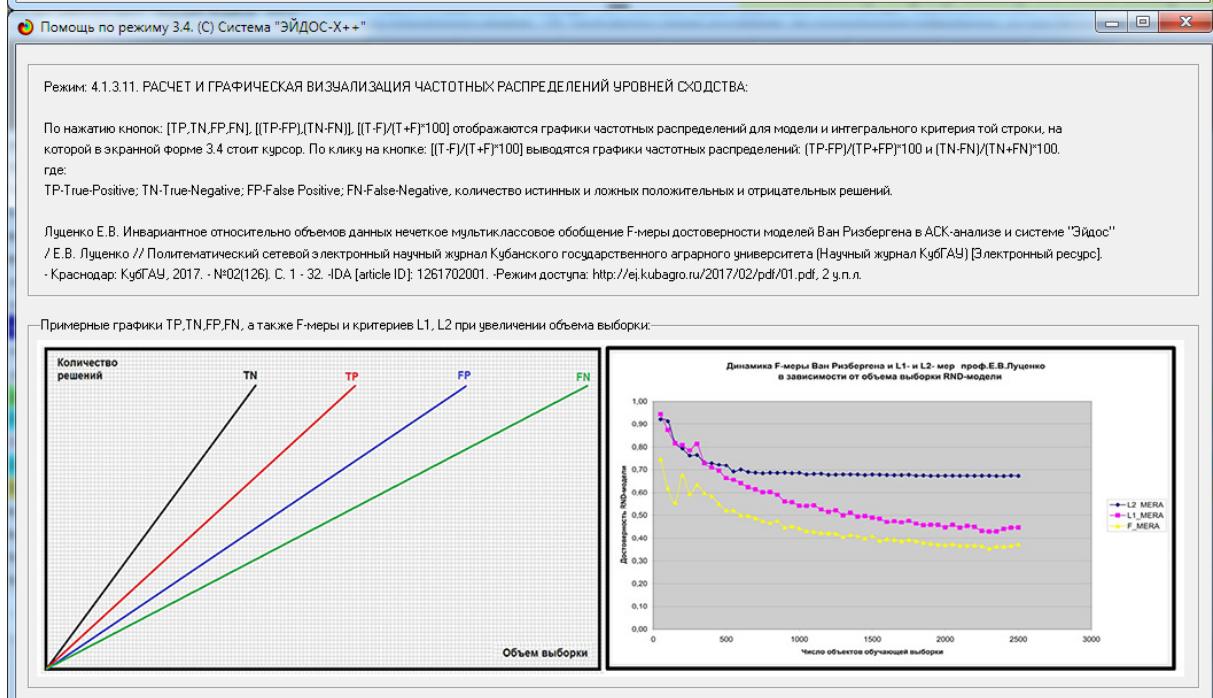
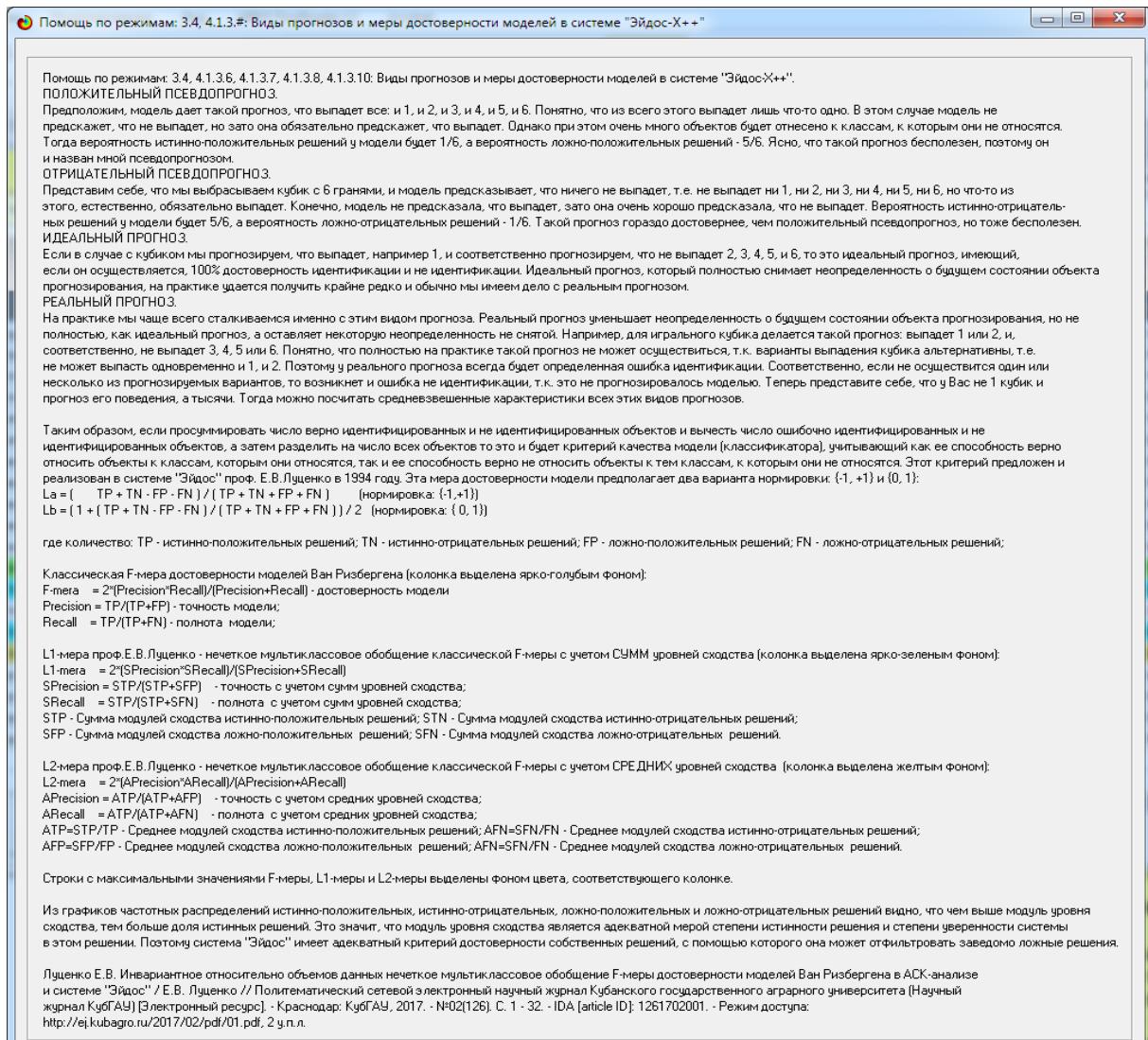


Рисунок 14. Экранные формы хелпов режима измерения достоверности моделей

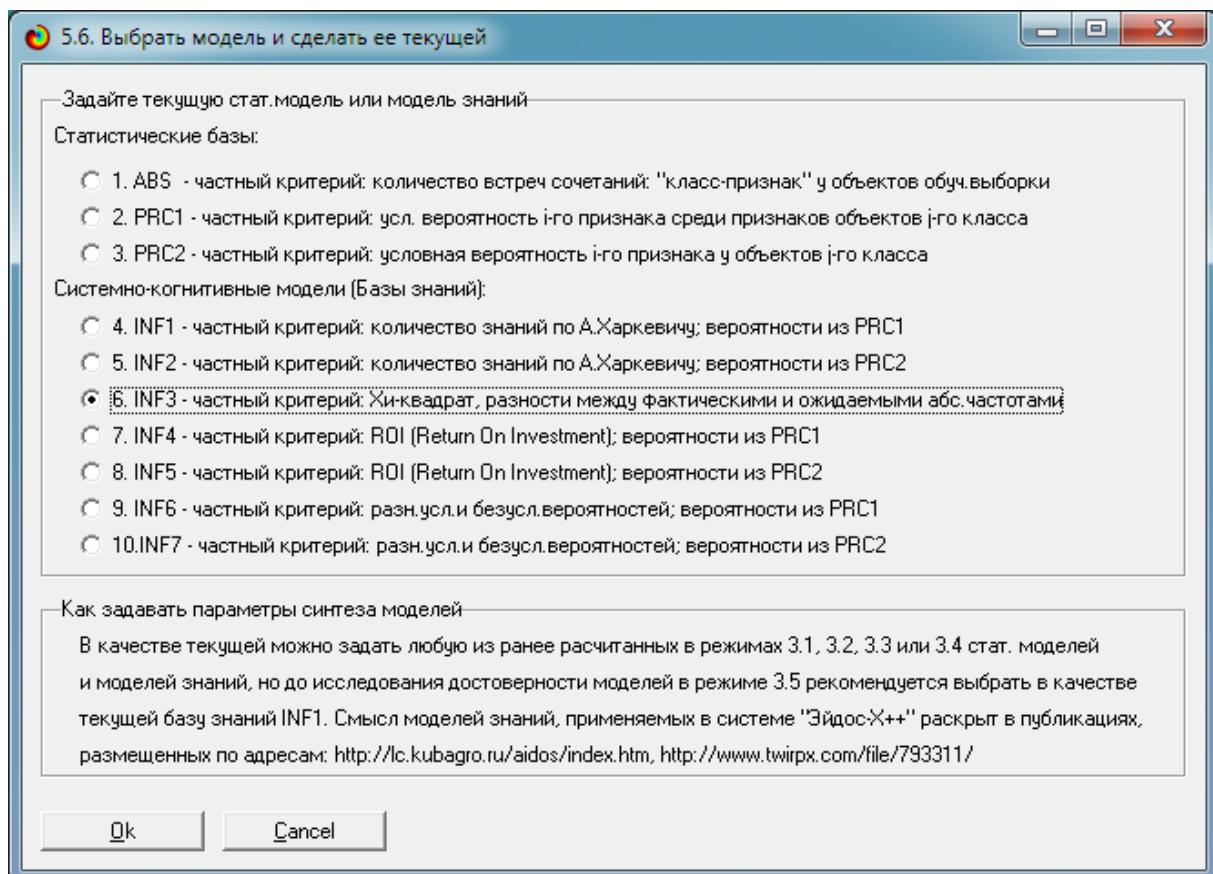
3.5. Задача-5. Выбор наиболее достоверной модели

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 17). 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.



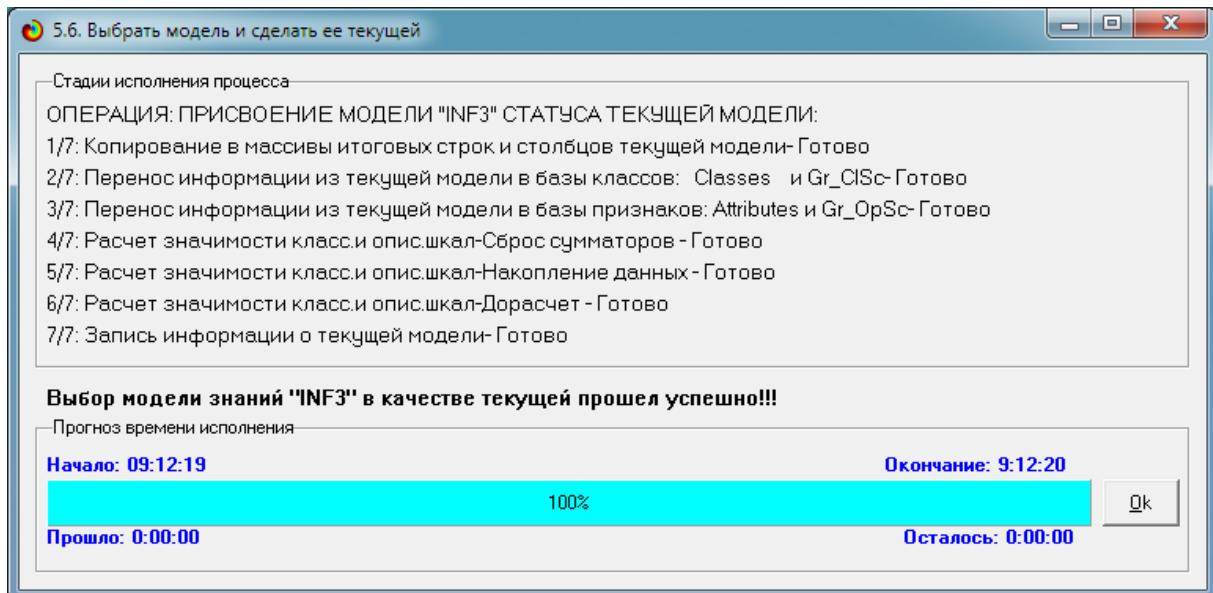


Рисунок 15. Задание СК-модели INF3 в качестве текущей

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» высокоеффективных 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

корректны⁹ в неортонормированных пространствах, которые всегда и встречаются на практике, и являются фильтрами подавления шума.

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.:

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

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:

⁹ В отличие от Евклидова расстояния, которое используется для подобных целей наиболее часто

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

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.

$\vec{I}_j = \{I_{ij}\}$ – 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 стандартизованными значениями:

$I_{ij} \rightarrow \frac{I_{ij} - \bar{I}_j}{\sigma_j}$, $L_i \rightarrow \frac{L_i - \bar{L}}{\sigma_l}$. Поэтому по своей сути он также является скалярным произведением двух стандартизованных (единичных) векторов класса и объекта. Существуют и много других способов нормирования, например, путем применения сплайнов, в частности линейной интерполяции:

$I_{ij} \rightarrow \frac{I_{ij} - I_j^{\min}}{I_j^{\max} - I_j^{\min}}$, $L_i \rightarrow \frac{L_i - L^{\min}}{L^{\max} - L^{\min}}$. Это позволяет предложить неограниченное количество других видов интегральных критериев. Но результаты их применения едва ли они будут существенно отличаться от уже существующих, поэтому они не реализованы в системе «Эйдос».

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 являются 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 являются **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 монографии [11, 12].

На рисунках 17 приведены экранные формы режима идентификации и прогнозирования 4.1.2 системы «Эйдос»:

3.6.4. Решение задачи идентификации и прогнозирования 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 освещены и на теоретическом уровне, и с детальными численными примерами в работах [13, 14] и в ряде других¹⁰.

¹⁰ См., например: http://lc.kubagro.ru/aidos/Works_on_Scenario_ASC-analysis.htm

Поэтому в данной работе рассмотрим стандартный несколько упрощенный вариант решения задачи идентификации и прогнозирования. Запустим режим 4.1.2 системы «Эйдос» (рисунок 18):

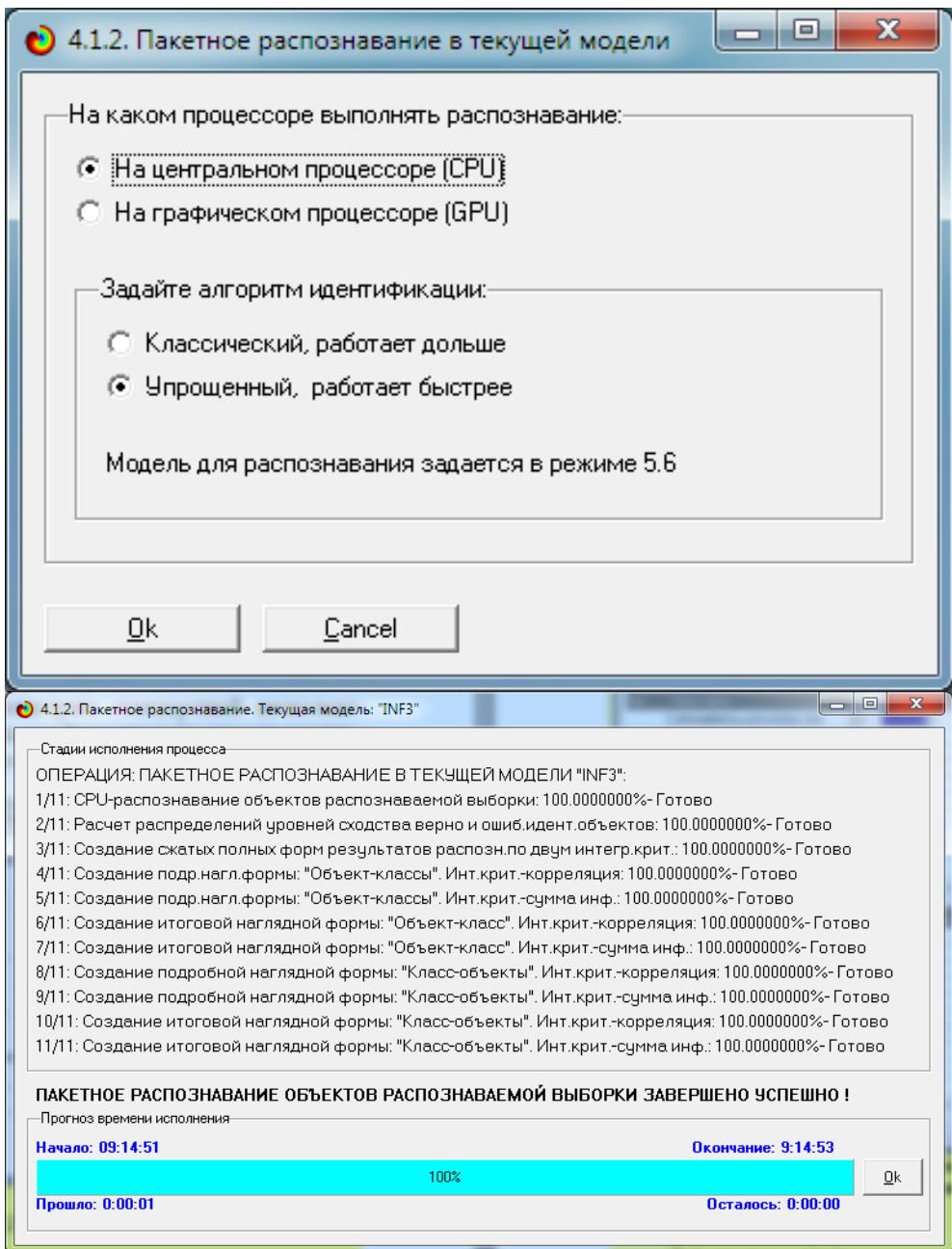


Рисунок 16. Экранные формы режима 4.1.2 идентификации и прогнозирования

По результатам решения задачи идентификации и прогнозирования в системе «Эйдос» выдается большое количество разнообразных выходных форм, которых в настоящее время 12 (рисунок 19):

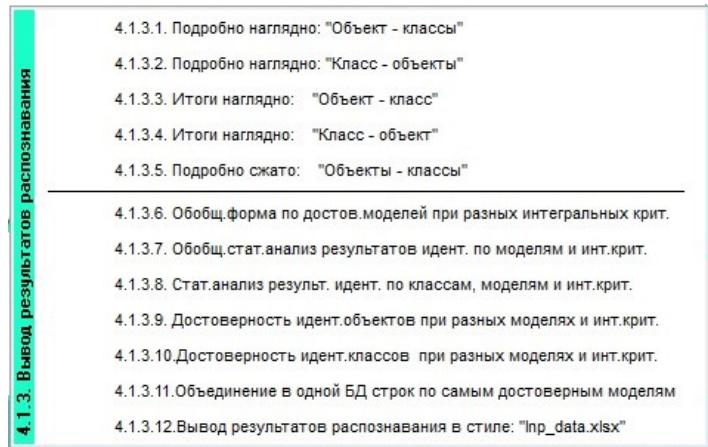


Рисунок 17. Меню подсистемы 4.1.3 системы «Эйдос», обеспечивающей вывод форм по результатам решения задачи идентификации и прогнозирования

Из этих выходных форм рассмотрим только две: 4.1.3.1 и 4.1.3.2 (рисунок 20):

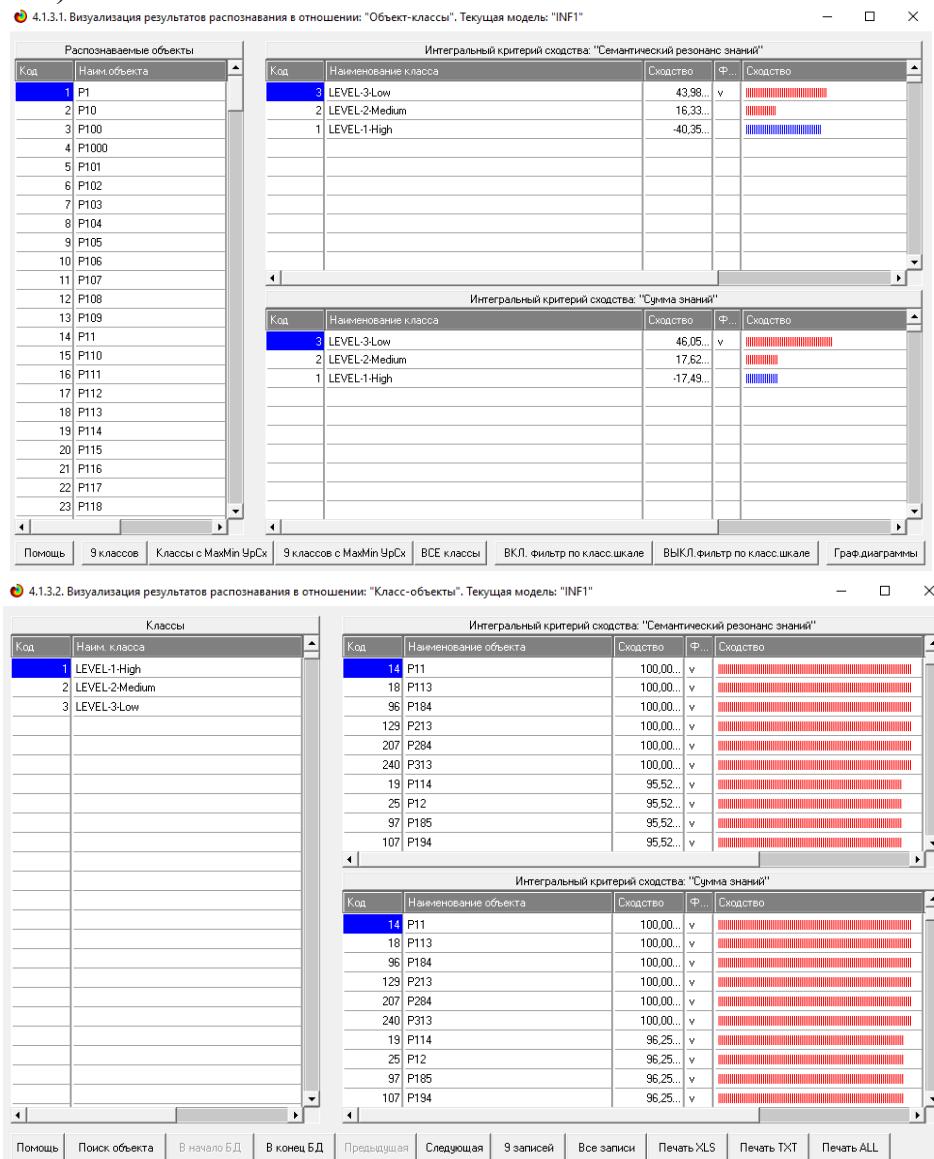


Рисунок 18. Некоторые экранные формы результатов идентификации и прогнозирования 4.1.3 системы «Эйдос»

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») [15] (рисунки 21).

Выходные формы, приведенные на рисунках 21, 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.

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	LEVEL-1-High	0,000000	0	0,000000
2	LEVEL-2-Medium	0,000000	0	0,000000
3	LEVEL-3-Low	0,000000	0	0,000000

SWOT-анализ класса:1 "LEVEL-1-High" в модели:6 "INF3"

Способствующие факторы и сила их влияния		
Код	Наименование фактора и его интервального значения	Сила влияния
61	OBESITY-High1	48.663
84	CHEST PAIN-High1	38.622
32	OCCUPATIONAL HAZARDS-High1	35.518
54	BALANCED DIET-High1	34.004
24	DUST ALLERGY-High1	32.466
15	AIR POLLUTION-Medium3	32.333
76	PASSIVE SMOKER-High1	31.542
68	SMOKING-High1	28.161
16	ALCOHOL USE-High1	24.213
40	GENETIC RISK-High1	23.478
93	COUGHING OF BLOOD-High1	21.450
53	CHRONIC LUNG DISEASE-Medium3	20.767
102	FATIGUE-High2	17.936
158	DRY COUGH-High1	17.450

Препятствующие факторы и сила их влияния		
Код	Наименование фактора и его интервального значения	Сила влияния
79	PASSIVE SMOKER-Low2	-24.418
56	BALANCED DIET-Low2	-20.984
42	GENETIC RISK-Low2	-18.313
81	PASSIVE SMOKER-Medium1	-16.787
70	SMOKING-Low1	-16.787
49	CHRONIC LUNG DISEASE-Low2	-16.406
12	AIR POLLUTION-Low3	-16.406
19	ALCOHOL USE-Low2	-16.024
131	WHEEZING-Medium1	-15.643
99	COUGHING OF BLOOD-Medium1	-15.643
43	GENETIC RISK-Low3	-15.643
18	ALCOHOL USE-Low1	-15.261
65	OBESITY-Medium1	-14.932
132	WHEEZING-Medium2	-14.880

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

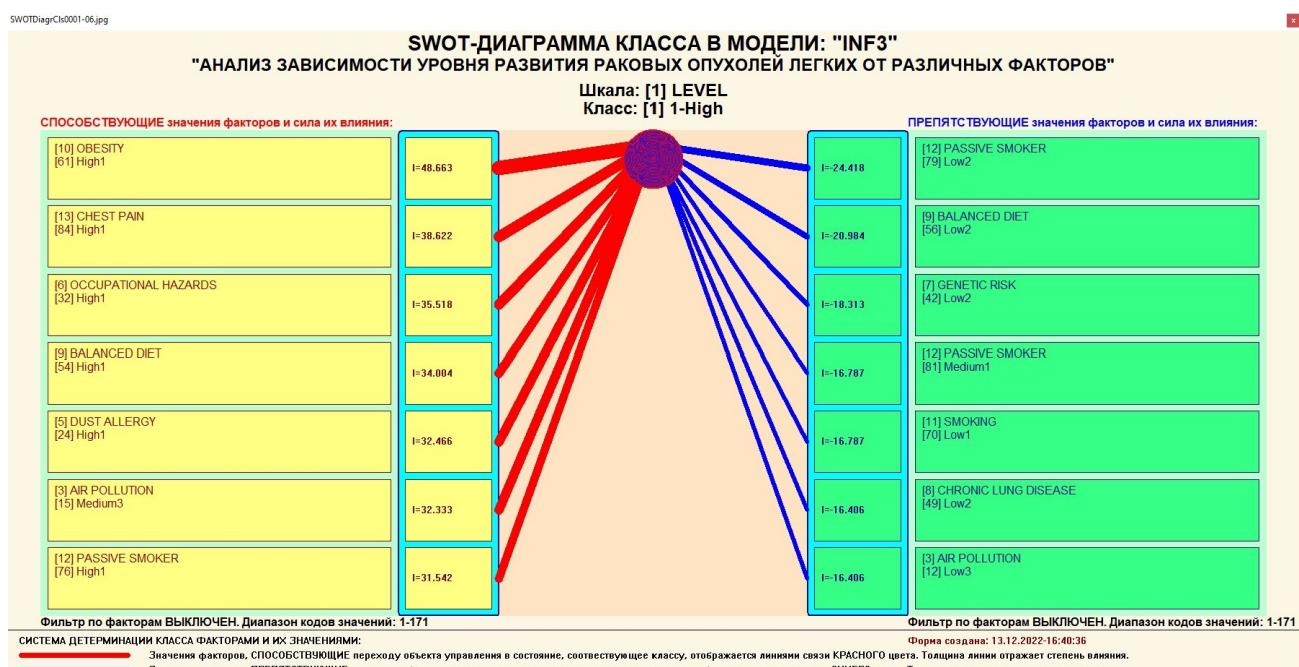


Рисунок 19. Примеры экранной формы режима автоматизированного SWOT-анализа (режим 4.4.8 системы «Эйдос»)

На первом рисунке 21 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.

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.

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 [13, 14, 16] и в ряде других работ.

Приведем этот алгоритм и в данной работе (рисунок 22).

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 [16, 17, 18].

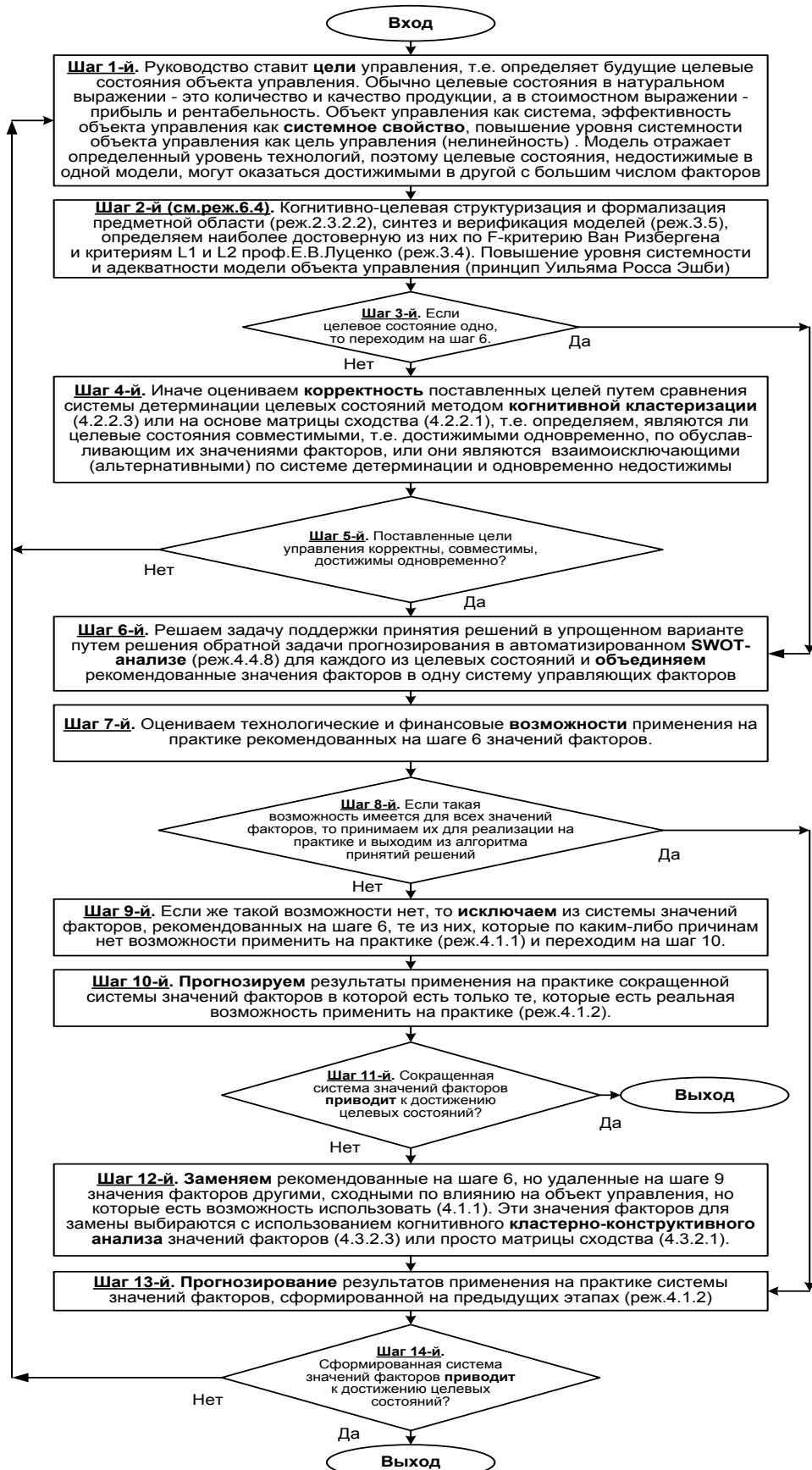


Рисунок 20. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»

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) [5]. Повышение уровня системности и адекватности модели объекта управления (принцип Уильяма Росса Эшби) [17].

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 совместимы, т.е. достижимы 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 [15].

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) [19].

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 23:

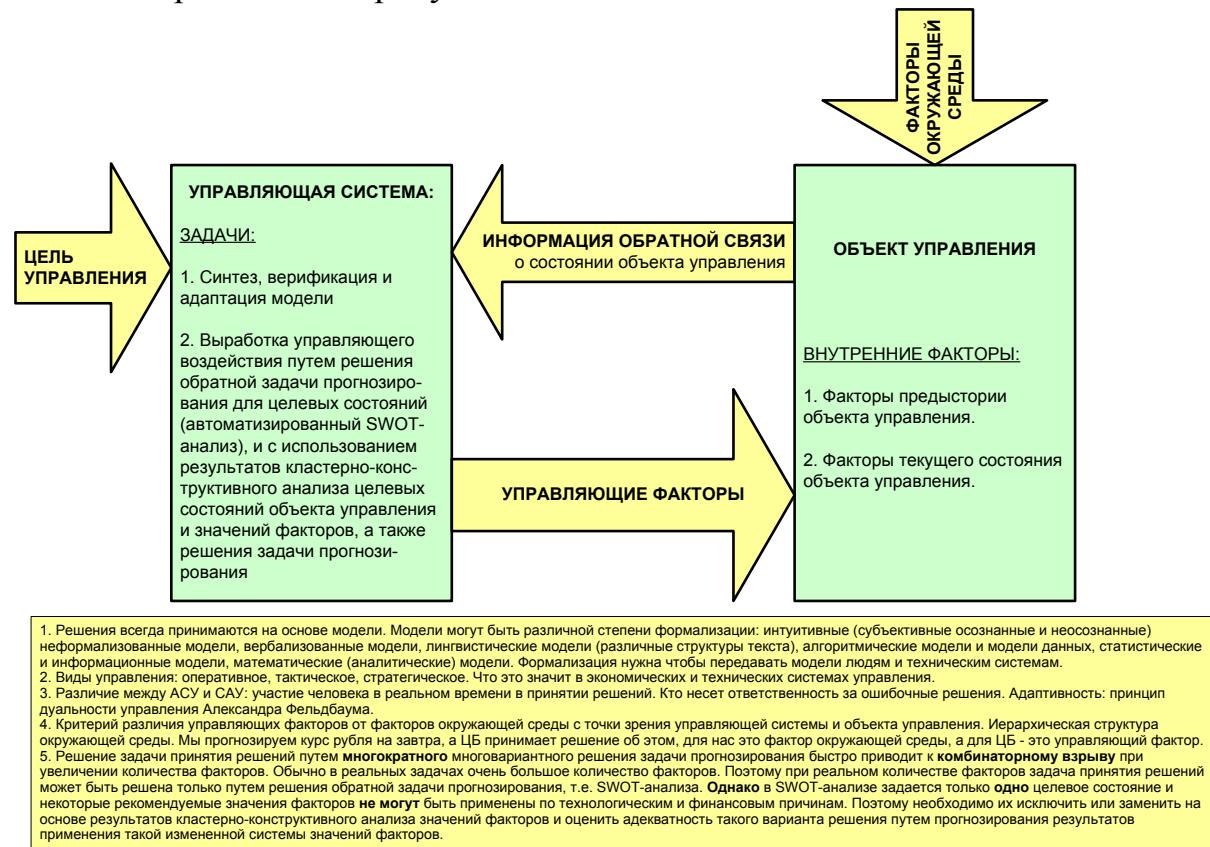


Рисунок 21. Принципиальная схема адаптивной интеллектуальной системы управления на основе АСК-анализа и системы «Эйдос»

Как мы видим в развитом алгоритме принятия решений широко используются результаты 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 задач, которые необходимо решать в развитом алгоритме принятия решений. Потому ниже кратко рассмотрим решение этих и некоторых других задач.

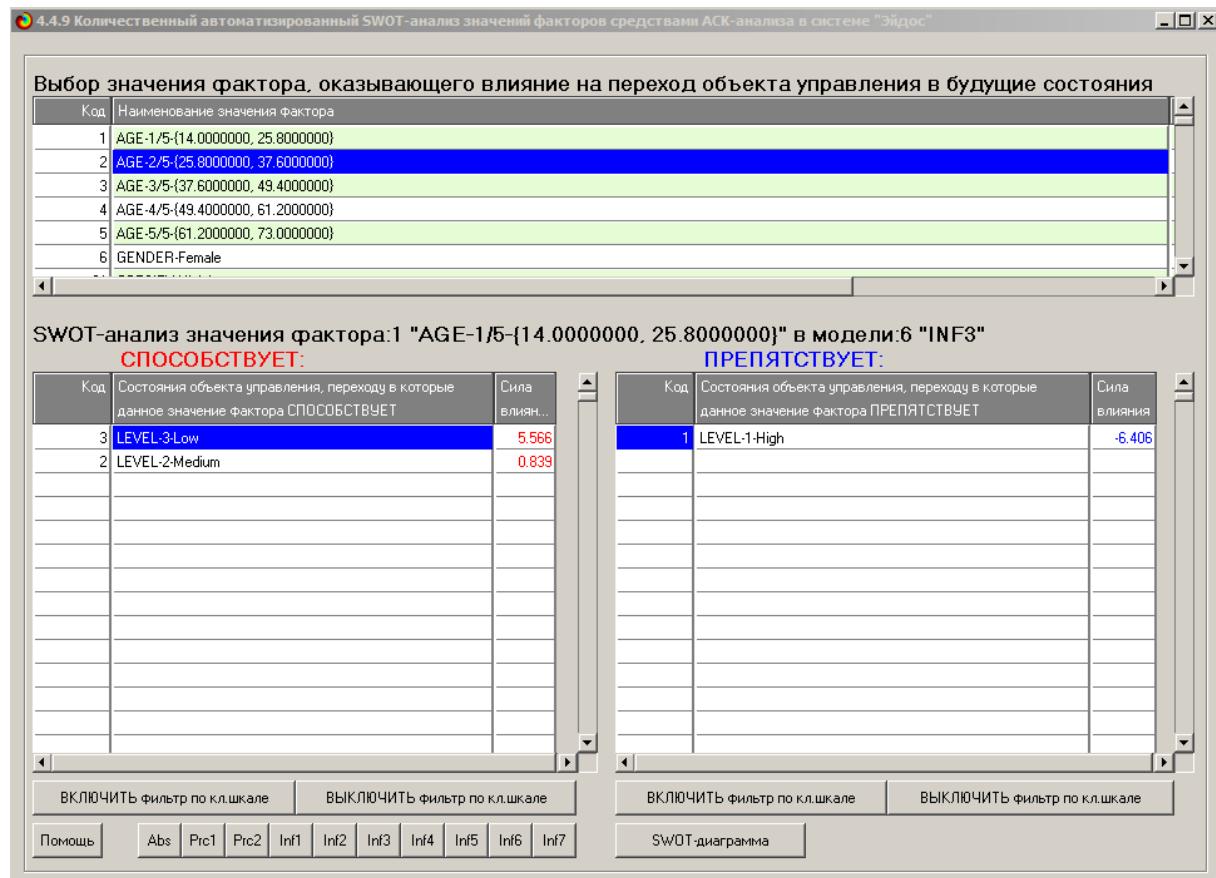
Необходимо отметить, что модели системы Эйдос – это феноменологические модели, отражающие эмпирические закономерности в фактах обучающей выборки, т.е. они не отражают причинно-следственного механизма детерминации, а только сам факт и характер детерминации. Содержательное объяснение этих эмпирических закономерностей формулируется уже экспертами на теоретическом уровне познания в содержательных научных законах [21, 22].

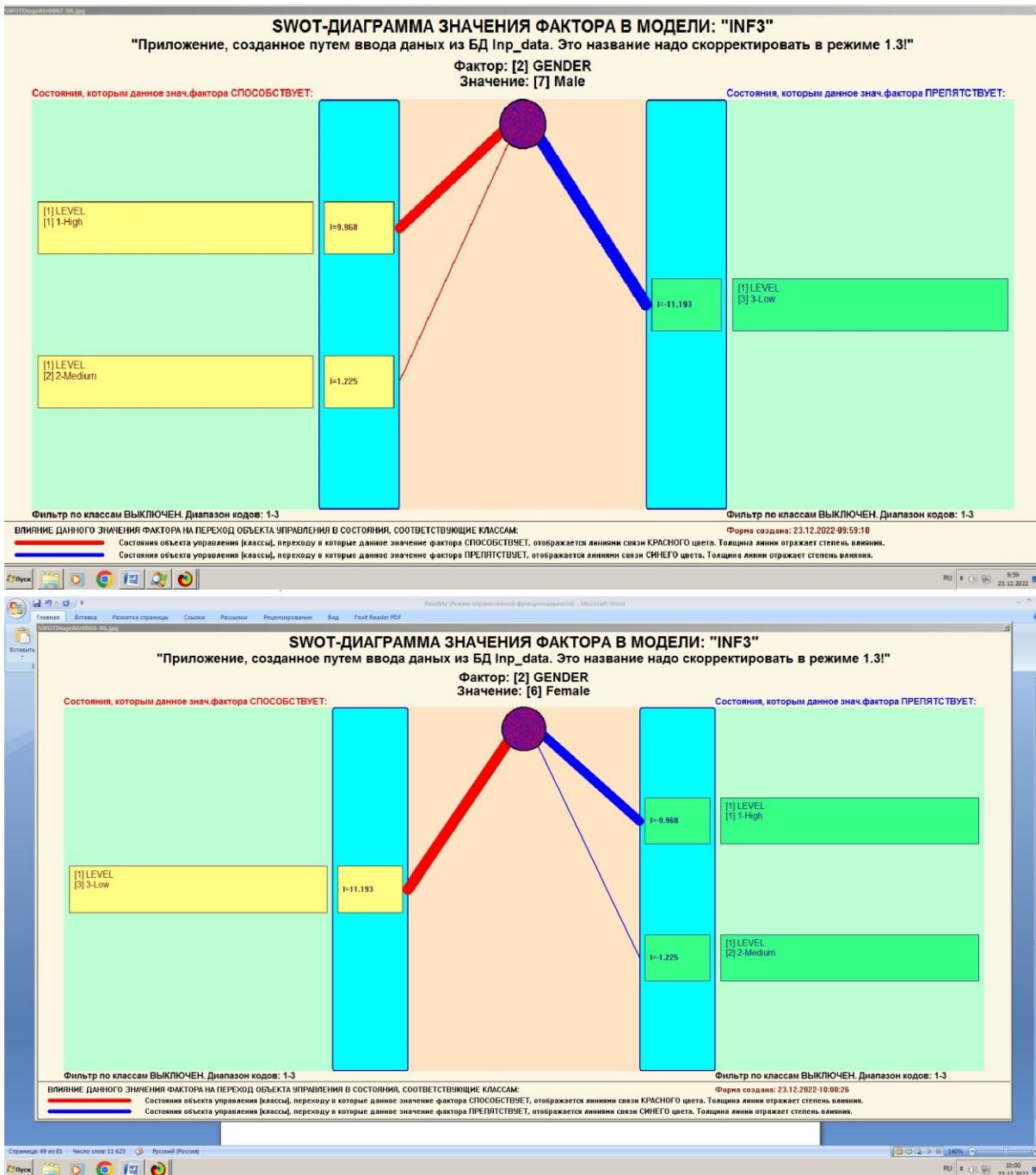
3.8. Задача-8. Исследование объекта моделирования путем исследования его модели

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

Инвертированные SWOT-диаграммы (предложены автором в работе [15]), 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».

Инвертированные SWOT-диаграммы для каждого значения фактора, которые представляют собой лингвистические переменные, приведены ниже на рисунке 24:





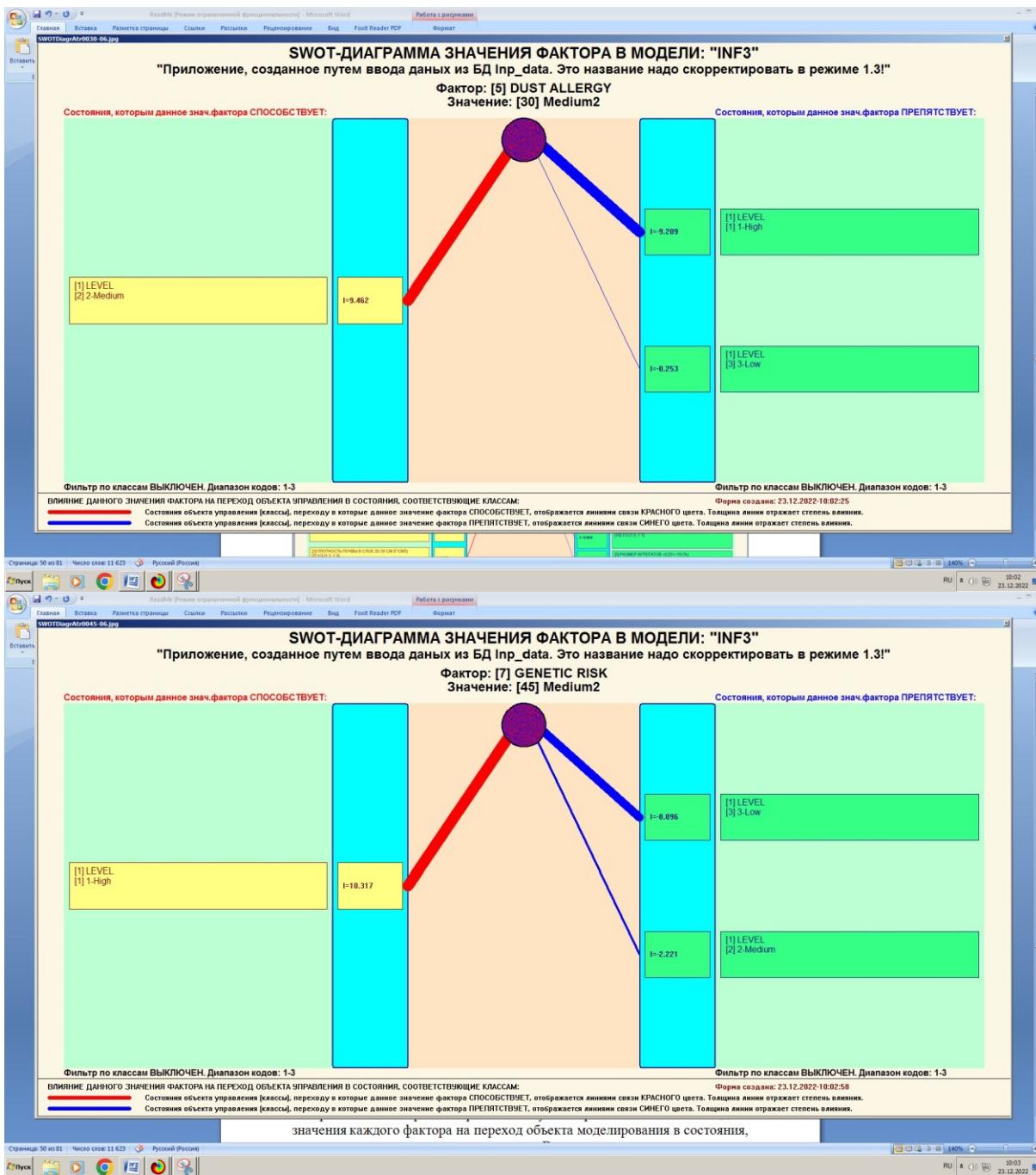


Рисунок 22. Примеры инвертированных SWOT-диаграмм влияния значений факторов на переход объекта моделирования в будущие состояния, соответствующие различным классам

Приведенные на рисунке 24 инвертированные SWOT-диаграммы исчерпывающим образом отражают силу и направление влияния каждого значения каждого фактора на переход объекта моделирования в состояния, соответствующие различным классам. Во многом это и есть решение проблемы, поставленной в работе.

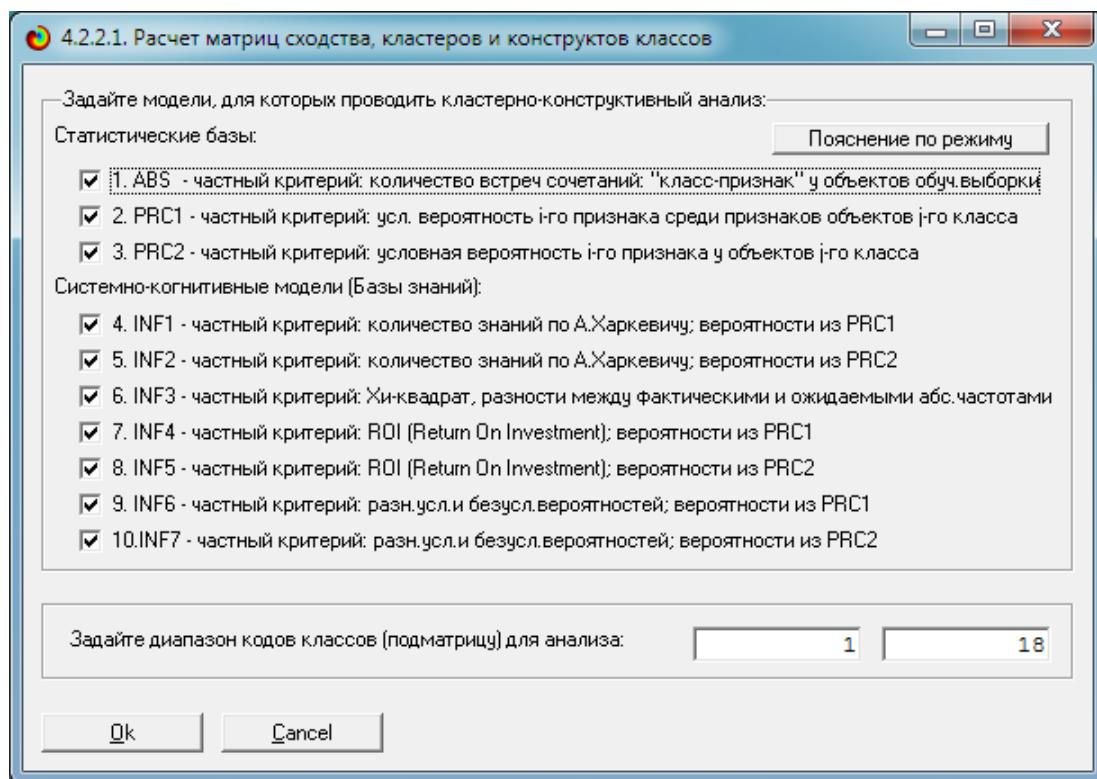
3.8.2. Кластерно-конструктивный анализ классов

В системе «Эйдос» (в режиме 4.2.2.1, рисунок 25) рассчитывается матрица сходства классов (таблица 14) 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 26);
- агломеративные дендрограммы, полученные в результате **когнитивной (истинной) кластеризации классов** (предложена автором в 2011 году в работе [19]) (режим 4.2.2.3) (рисунок 27);
- график изменения межкластерных расстояний (режим 4.2.2.3) (рисунок 28).

Эта матрица сходства (таблица 14) 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:



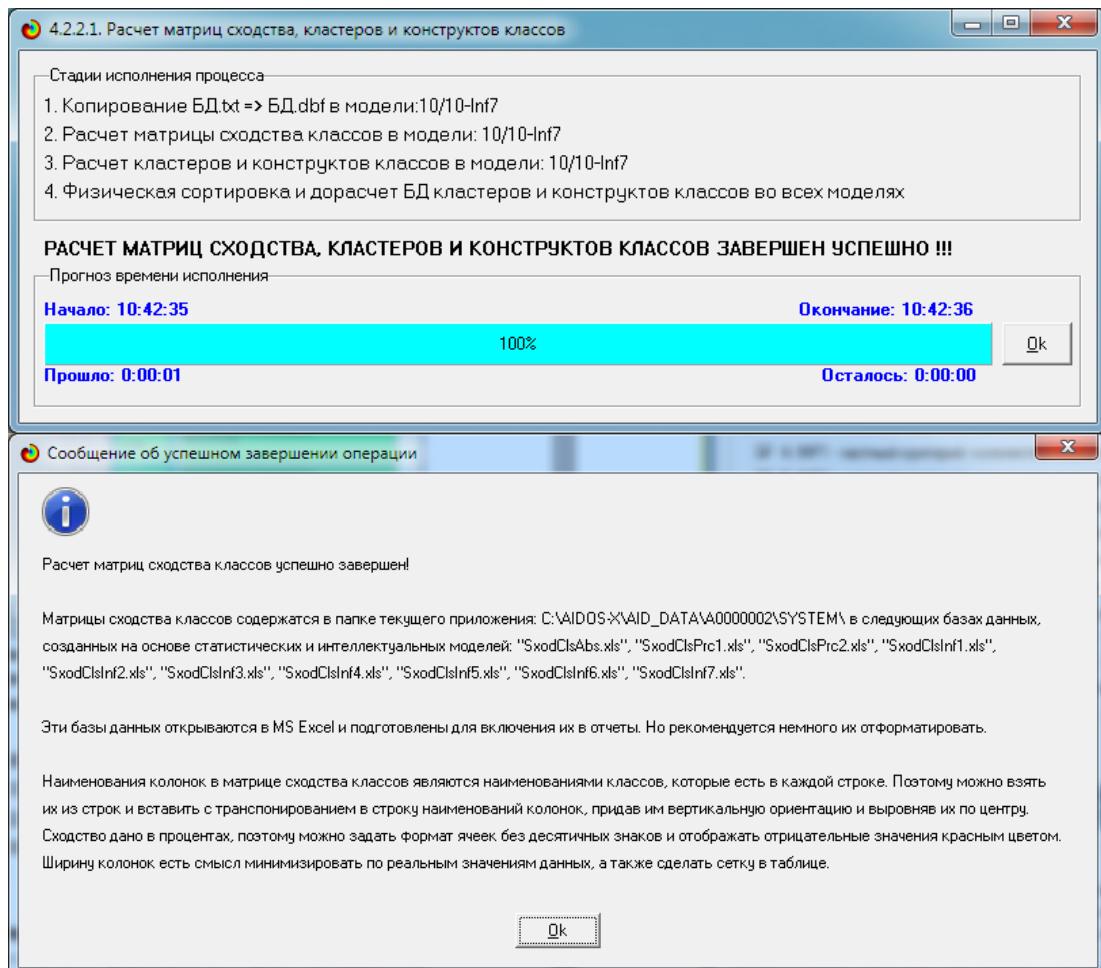


Рисунок 23. Экранные формы режима 4.2.2.1, обеспечивающего расчет матриц сходства классов

Таблица 11 – Матрица сходства классов в СК-модели INF3 (полностью)

KOD_CLS	KOD_CLSC	NAME_CLS	N1	N2	N3
1	1	LEVEL-1-High	0,0000000	0,0000000	0,0000000
2	1	LEVEL-2-Medium	0,0000000	0,0000000	0,0000000
3	1	LEVEL-3-Low	0,0000000	0,0000000	100,0000000

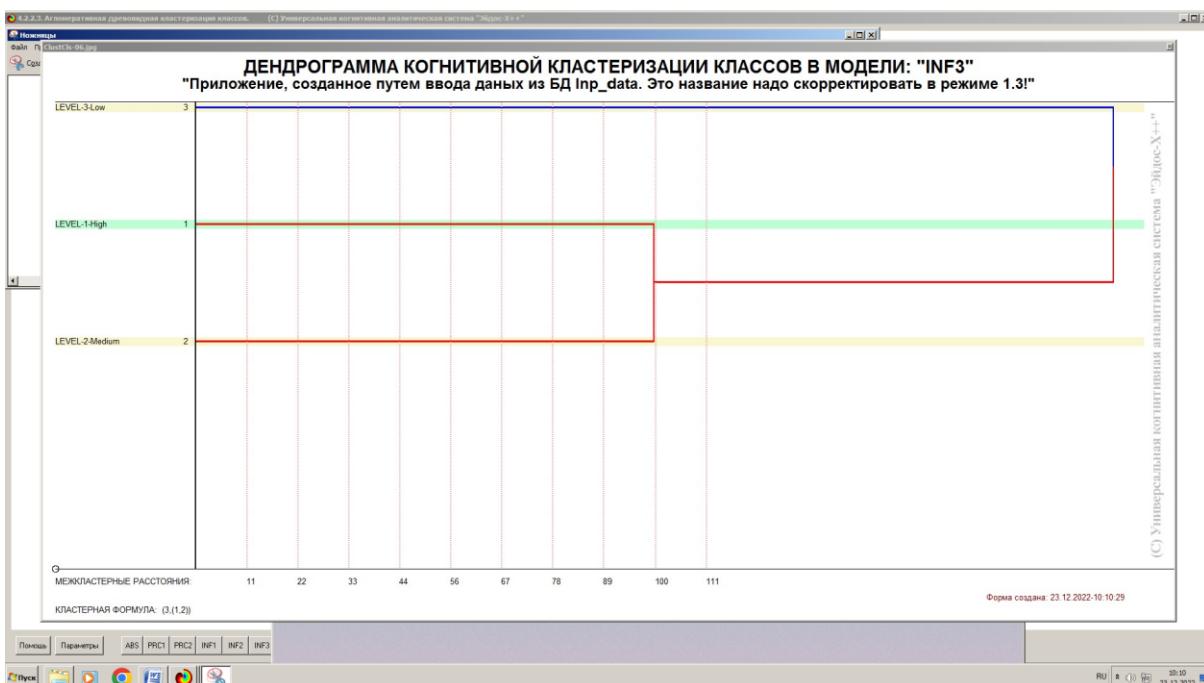


Рисунок 24. Агломеративная дендрограмма, полученная в результате когнитивной (истинной) кластеризации классов (режим 4.2.2.3)

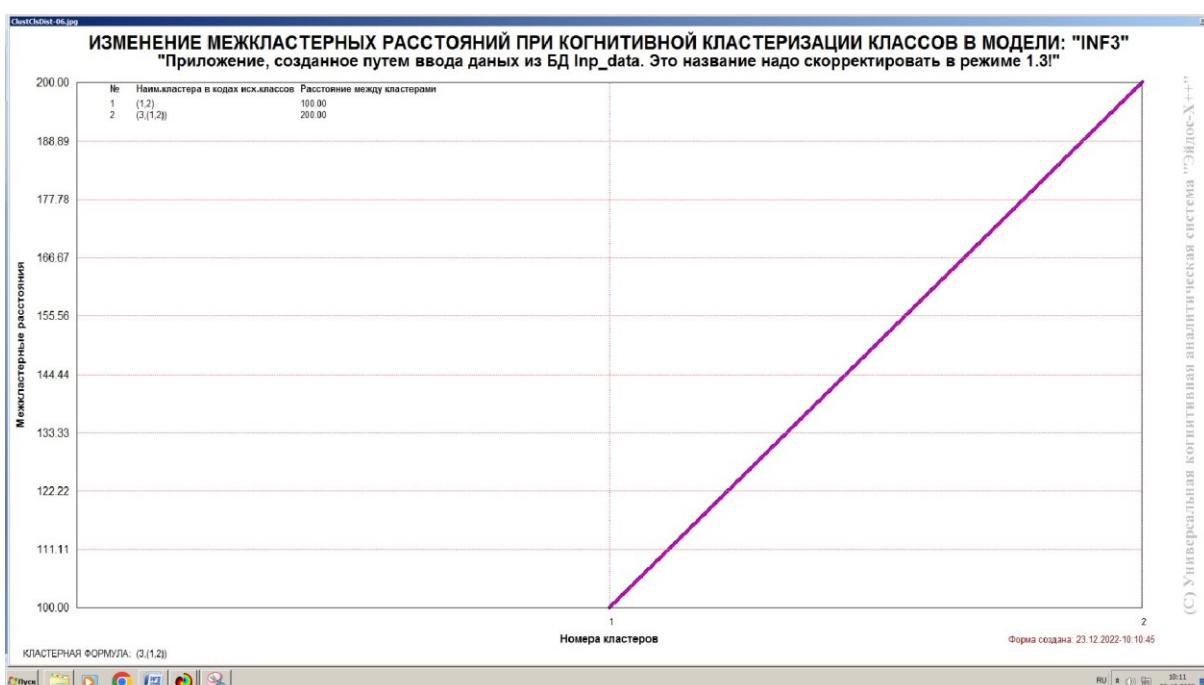


Рисунок 25. График изменения межкластерных расстояний (режим 4.2.2.3)

3.8.3. Кластерно-конструктивный анализ значений описательных шкал

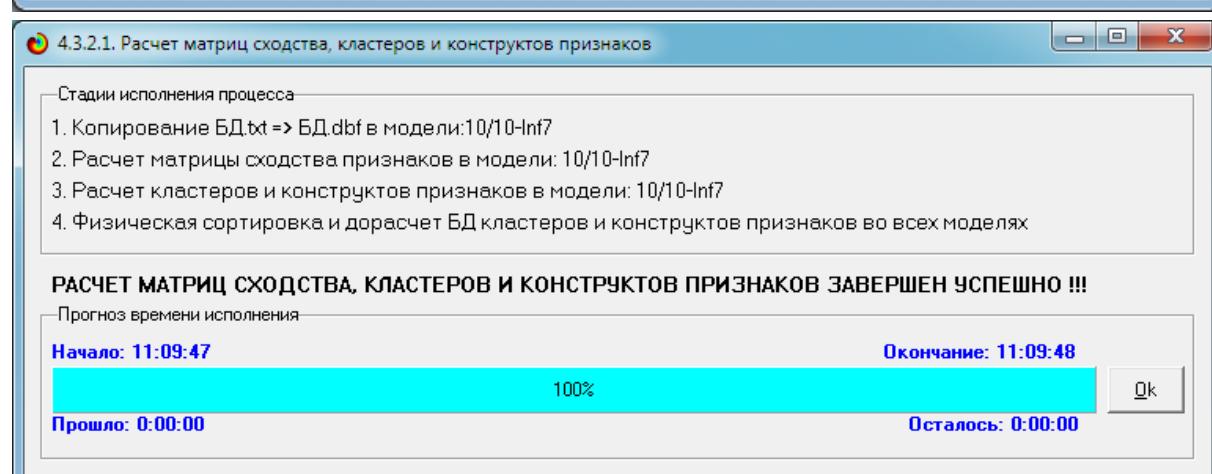
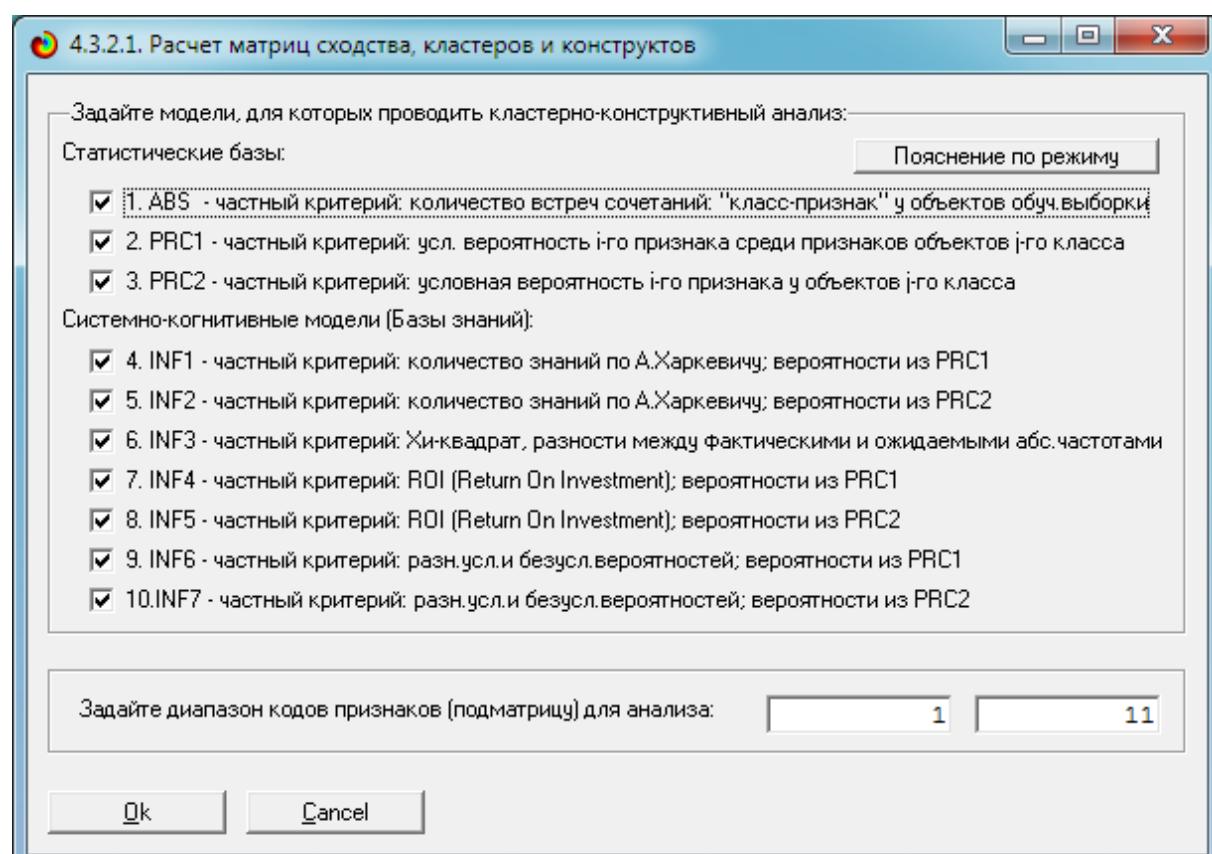
В системе «Эйдос» (в режиме 4.3.2.1, рисунок 29) рассчитывается матрица сходства признаков (таблица 15) по их смыслу и на основе этой матрицы рассчитывается и выводится четыре основных формы:

- круговая 2d-когнитивная диаграмма признаков (режим 4.3.2.2) (рисунок 29);

- агломеративные дендрограммы, полученные в результате **когнитивной (истинной) кластеризации признаков** (предложена автором в 2011 году в работе [19]) (режим 4.3.2.3) рисунок 30);
- график изменения межклластерных расстояний (режим 4.3.2.3) рисунок 31).

Эта матрица сходства (таблица 15) используются и при расчете некоторых других выходных форм.

На рисунке 29 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:



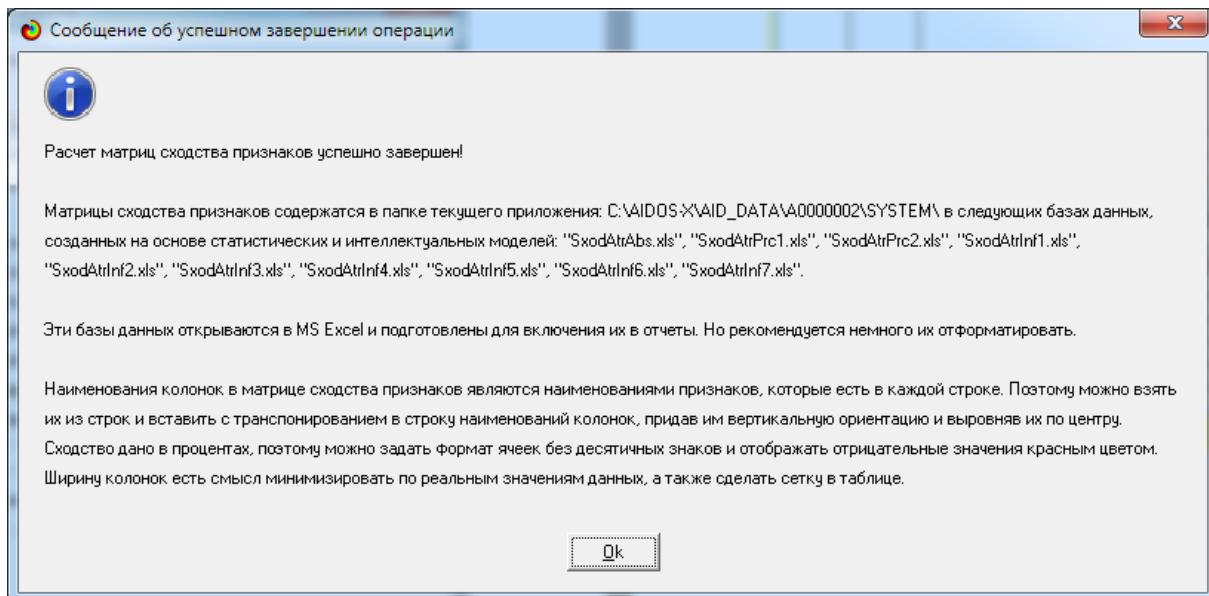


Рисунок 26. Экранные формы режима 4.3.2.1, обеспечивающего расчет матриц сходства значений факторов

Таблица 12 – Матрица сходства значений факторов в СК-модели INF1 (полностью)

KOD_ATR	KOD_OPSC	NAME_ATR	N1	N2	N3	N4	N5	N6	N7	N8	N9
1		1 AGE-1/5-[14.0000000, 25.8000000]	100,0000000	-72,5343515	29,5585983	48,6838794	-56,6694589	96,6745011	-92,3300182	-93,4267520	-93,4267520
2		1 AGE-2/5-[25.8000000, 37.6000000]	-72,5343515	100,0000000	-87,2028673	28,6175140	-15,6133108	-52,517581	40,5314690	92,3125038	92,3125038
3		1 AGE-3/5-[37.6000000, 49.4000000]	29,5585983	-87,2028673	100,0000000	-69,6585125	61,9604677	4,1441603	9,4002393	-61,6796188	-61,6796188
4		1 AGE-4/5-[49.4000000, 61.2000000]	48,6838794	28,6175140	-69,6585125	100,0000000	-99,5583236	69,4037718	-78,4988327	-14,3374194	-14,3374194
5		1 AGE-5/5-[61.2000000, 73.0000000]	-56,6694589	-15,6133108	61,9604677	-99,5583236	100,0000000	-75,8562327	83,9682737	23,5654011	23,5654011
6		2 GENDER-Female	96,6745011	-52,5172581	4,1441603	69,4037718	-75,8562327	100,0000000	-99,0821093	-81,2007729	-81,2007729
7		2 GENDER-Male	-92,3300182	40,5314690	9,4002393	-78,4988327	83,9682737	-99,0821093	100,0000000	72,5657181	72,5657181
8		3 AIR POLLUTION-High1	-93,4267520	92,3125038	-61,6796188	-14,3374194	-81,2007729	72,5657181	100,0000000	100,0000000	100,0000000
9		3 AIR POLLUTION-High2	-93,4267520	92,3125038	-61,6796188	-14,3374194	23,5654011	-81,2007729	72,5657181	100,0000000	100,0000000
10		3 AIR POLLUTION-Low1	16,2824442	-79,7304256	99,0969123	-78,2565882	72,0660933	-9,4919369	22,8617772	-50,3936000	-50,3936000
11		3 AIR POLLUTION-Low2	99,3134822	-80,0888201	40,5305140	38,1319576	-46,6424900	93,0192607	-87,2033786	-96,9563830	-96,9563830
12		3 AIR POLLUTION-Low3	81,4460882	-19,1349016	-31,3548847	90,3327250	-93,9608780	93,5762354	-97,4841682	-55,4034191	-55,4034191
13		3 AIR POLLUTION-Medium1	-99,5934137	66,0381447	-20,8325113	-56,3547341	63,8613520	-98,55852736	95,4145698	89,8347273	89,8347273
14		3 AIR POLLUTION-Medium2	66,6722854	-99,6661943	90,9076526	32,6431923	23,6249980	45,3944849	-32,9228960	-88,8653301	-88,8653301
15		3 AIR POLLUTION-Medium3	55,5851617	-77,7843820	66,7326483	73,42299872	-99,9335073	98,5233468	83,2748201	83,2748201	83,2748201
16		4 ALCOHOL USE-High1	-95,6961192	49,4344707	-0,5616368	-71,9387065	78,1421729	-99,9357921	99,5028317	79,0574642	79,0574642
17		4 ALCOHOL USE-High2	-96,3291650	88,3519731	88,51196878	-23,4472689	32,4702889	-86,2601590	78,6298601	99,5696604	99,5696604
18		4 ALCOHOL USE-Low1	91,2067690	-37,5296372	-12,2123087	80,2196237	-85,4706961	98,6601344	-99,9599941	-70,5906203	-70,5906203
19		4 ALCOHOL USE-Low2	-60,4666721	98,6864502	-60,4666721	13,7371837	62,1556672	-69,2356465	99,5361531	-97,3220219	-86,4391047
20		4 ALCOHOL USE-Low3	46,9939891	26,6770343	-70,4348401	99,9814833	-99,3592279	68,0055304	-77,2921622	-12,4303280	-12,4303280
21		4 ALCOHOL USE-Medium1	95,8501343	-49,8891826	1,0709382	71,5657690	-77,8069503	99,9535433	-99,4480819	-79,3842025	-79,3842025
22		4 ALCOHOL USE-Medium2	-84,8151278	97,9876915	-75,6784274	4,9822939	4,4163536	-68,4465525	57,9630738	98,1296124	98,1296124
23		4 ALCOHOL USE-Medium3	62,7314048	-99,1110903	92,9393338	-37,4845267	28,6151905	40,7782880	-28,0091263	-86,3766094	-86,3766094
24		5 DUST ALLERGY-High1	-87,8237780	96,6227456	-71,6450091	-0,9835272	10,3670306	-72,6729868	62,7201164	99,1030807	99,1030807
25		5 DUST ALLERGY-High2	77,5934912	-12,8575847	-37,3270335	92,8765203	-95,9462550	91,1456717	-95,8702941	-50,0000000	-50,0000000

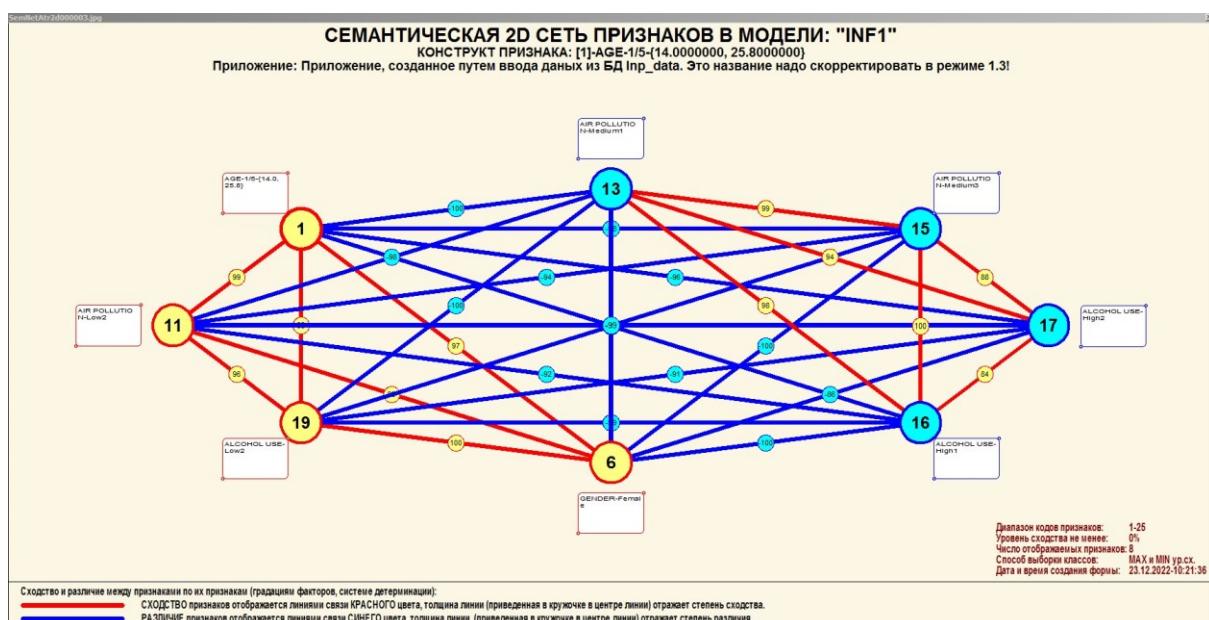


Рисунок 27. Круговая 2d-когнитивная диаграмма значений факторов в СК-модели INF1 (режим 4.3.2.2)

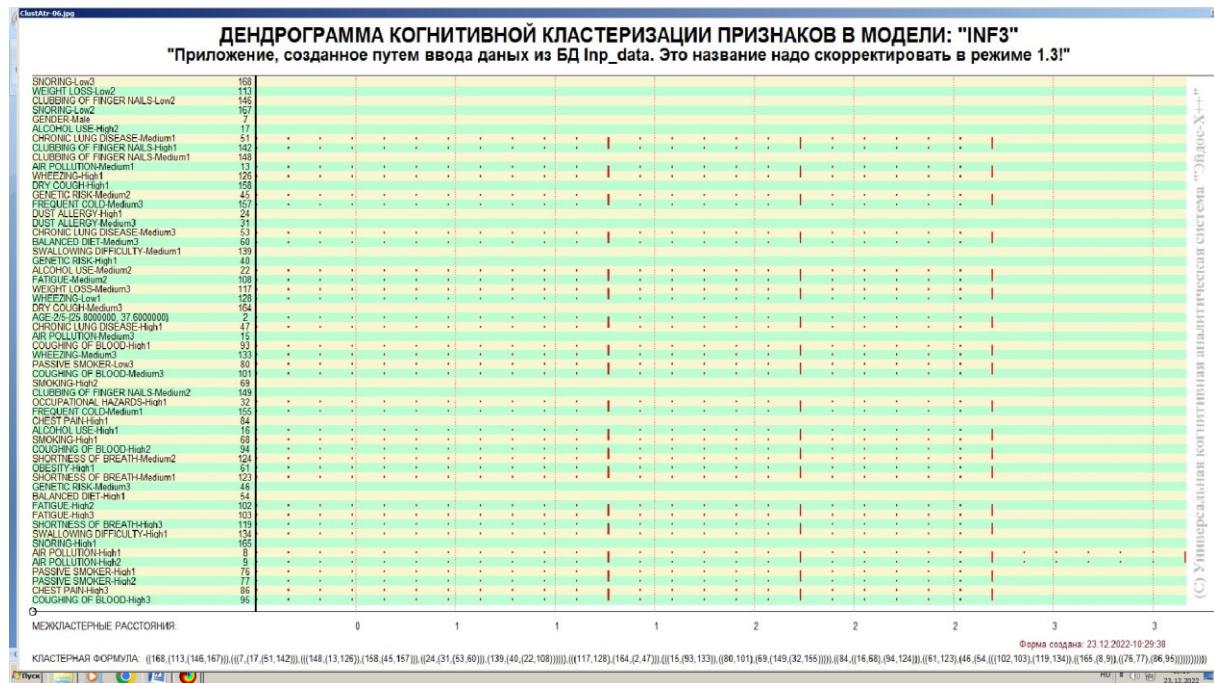


Рисунок 28. Агломеративная дендрограмма, полученная в результате когнитивной (истинной) кластеризации признаков (режим 4.3.2.3)

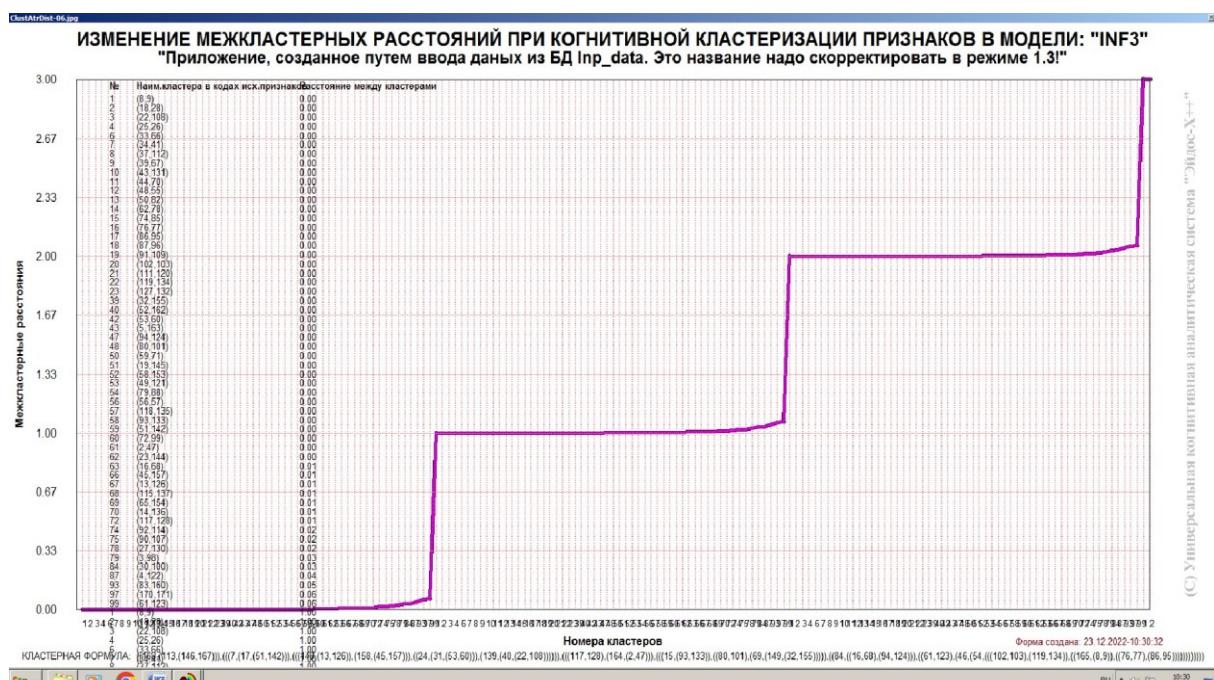


Рисунок 29. График изменения межкластерных расстояний (режим 4.3.2.3)

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.

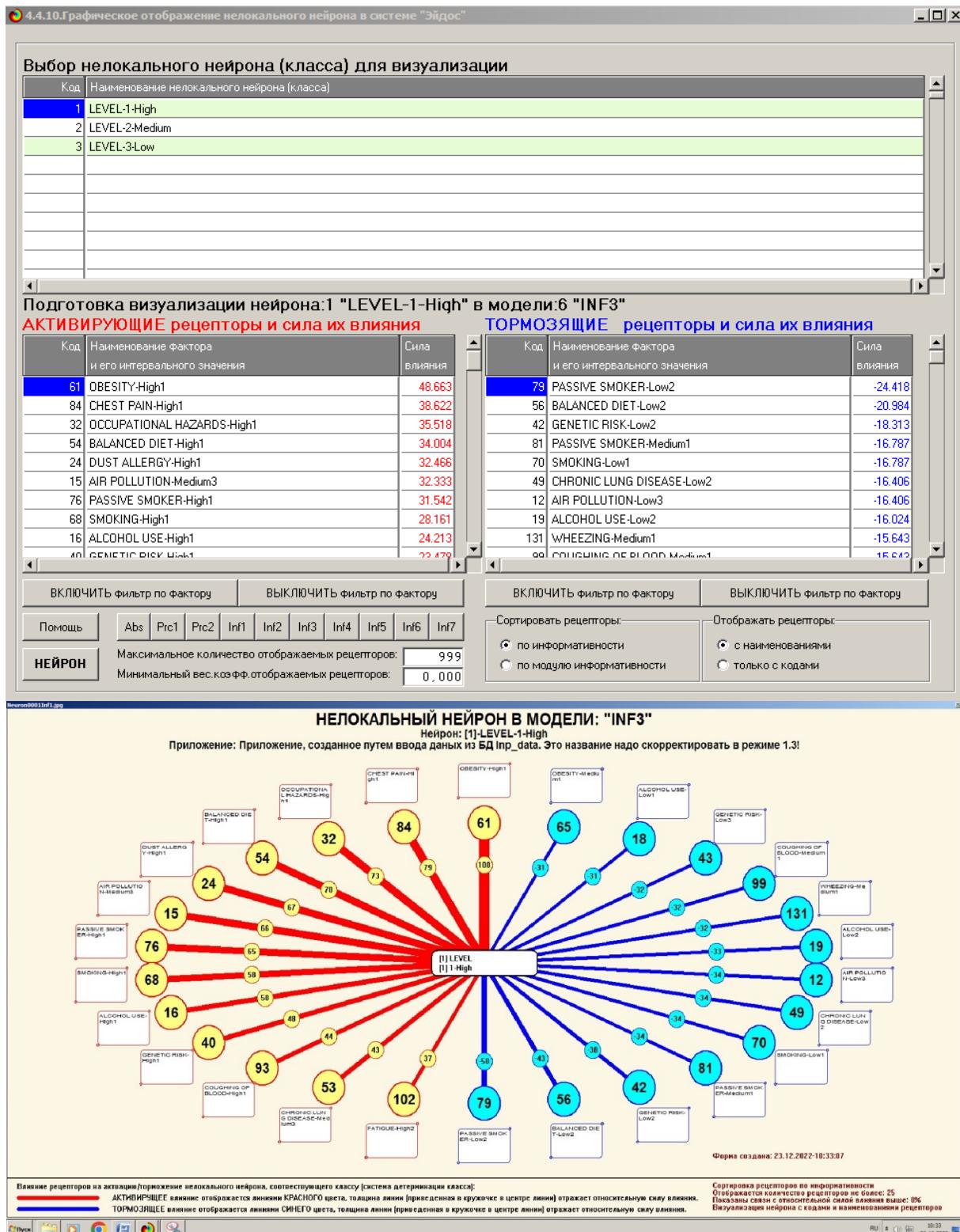
Классы в этой модели соответствуют нейронам и фреймам, а признаки рецепторам и шпациям (описательные шкалы – слотам).

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 реализации.

От нейросетевой модели представления знаний модель системы «Эйдос» отличается тем, что [20]:

- 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 33). В форме управления визуализацией есть возможность задавать фильтры по факторам, которые надо визуализировать.



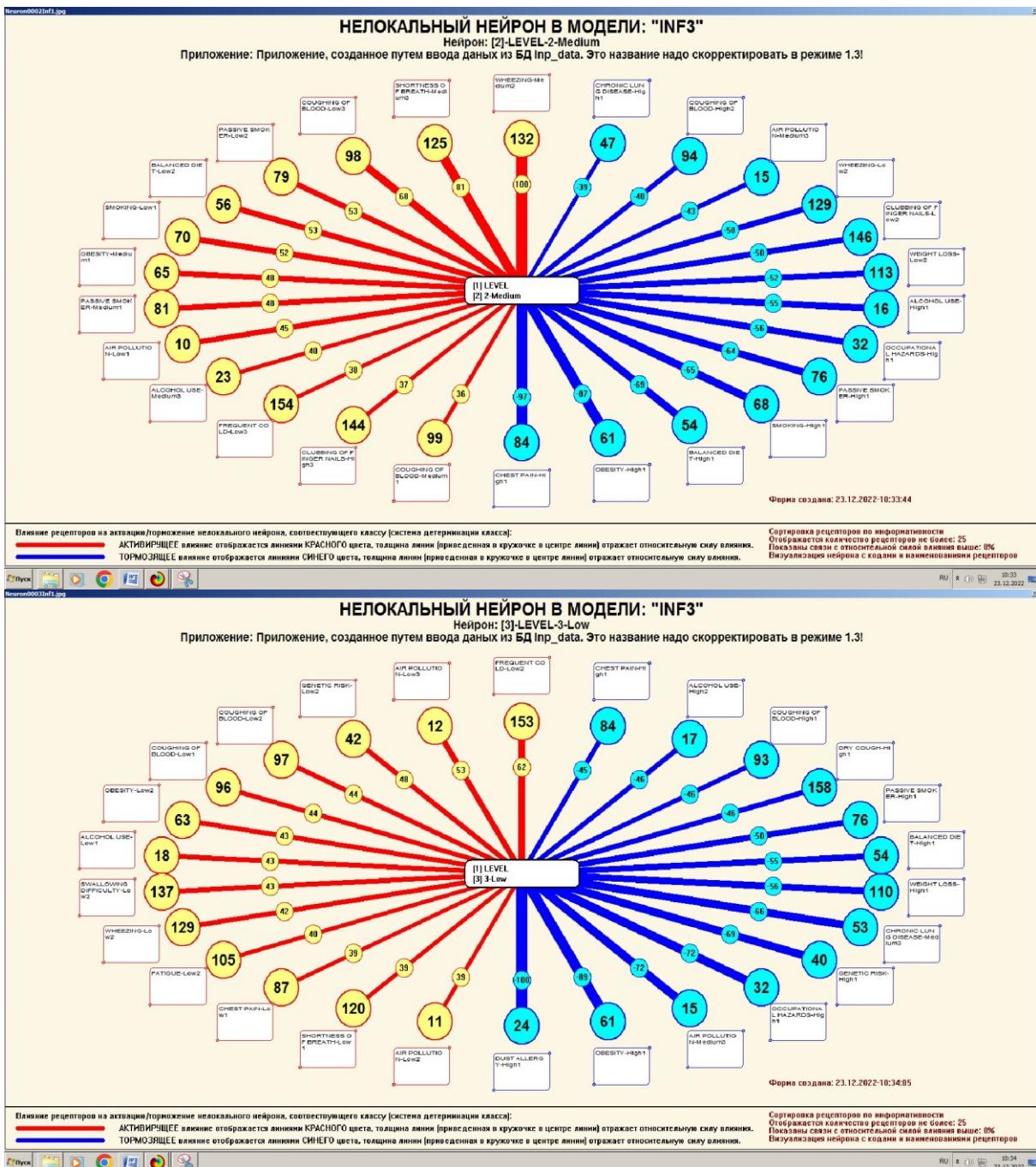


Рисунок 30. Примеры нелокальных нейронов, соответствующие классам

3.8.5. Нелокальная нейронная сеть

В системе «Эйдос» есть возможность построения моделей, соответствующих многослойным нейронным сетям [20].

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 34). В форме управления визуализацией есть возможность задавать фильтры по факторам, которые надо визуализировать.

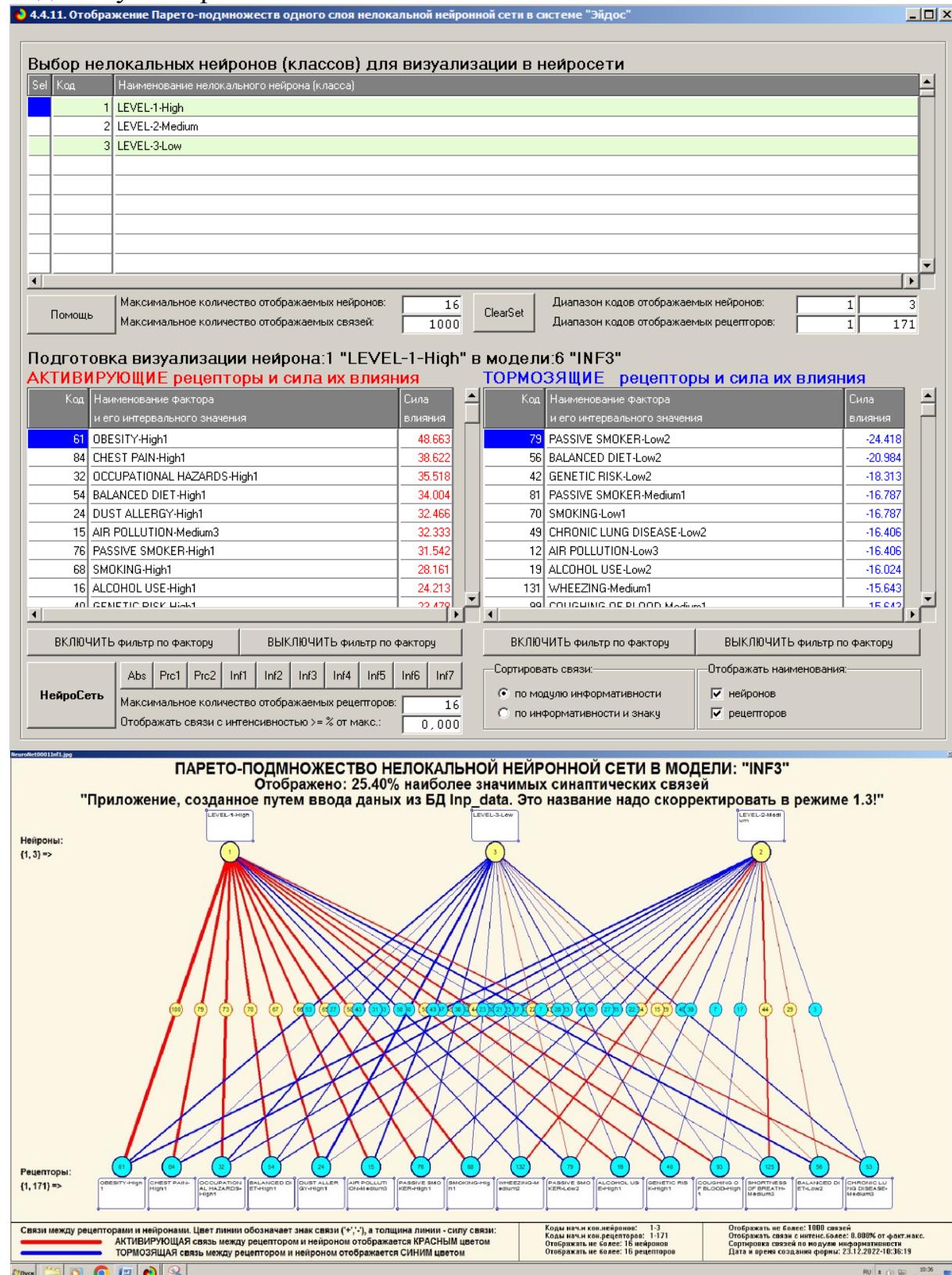
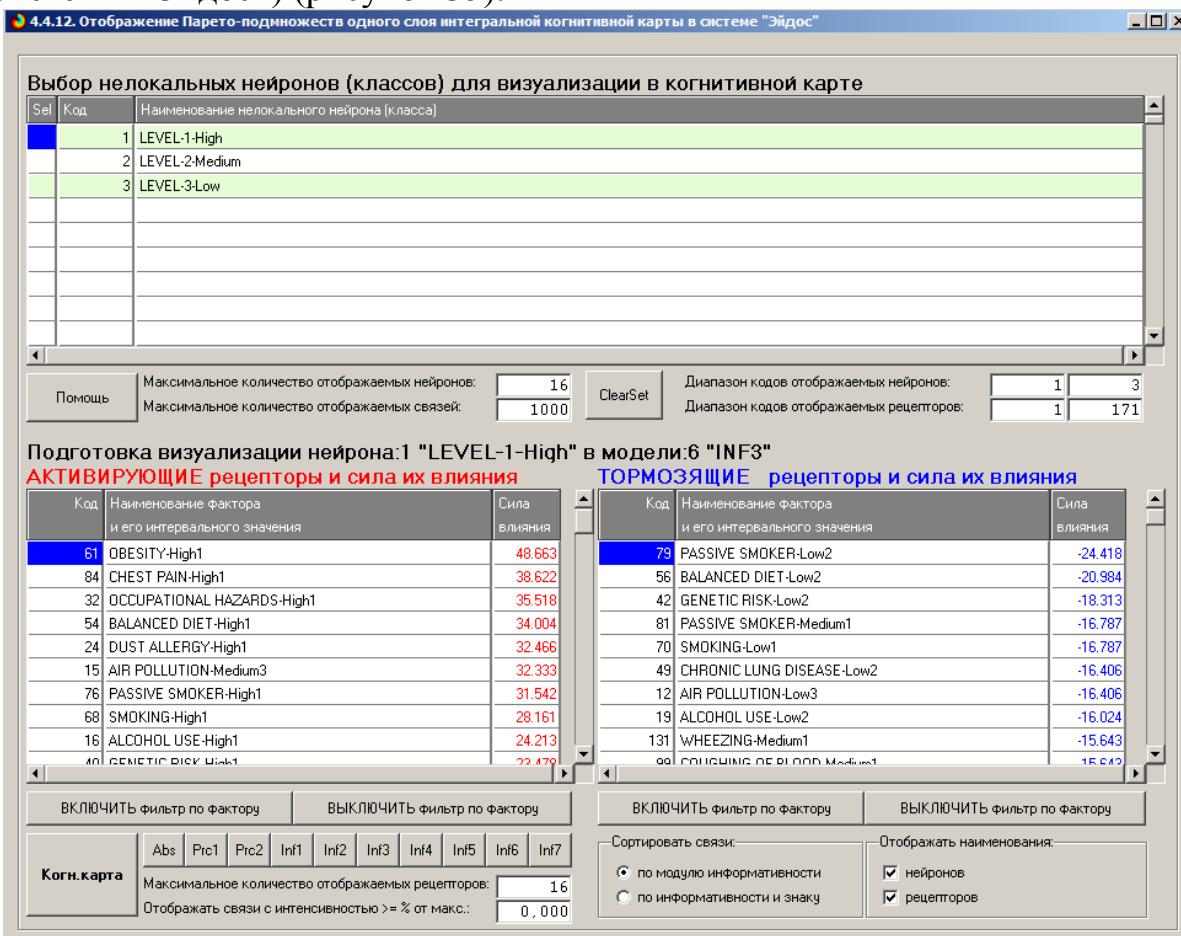


Рисунок 31. Нейронная сеть в СК-модели INF3

3.8.6. 3D-интегральные когнитивные карты

3d-интегральная когнитивная карта является отображением на одном рисунке когнитивной диаграммы классов (рисунок 26) вверху и когнитивной диаграммы значений факторов (рисунок 30) внизу и соединяющего их одного слоя нейронной сети (рисунок 34) (режим 4.4.12 системы «Эйдос») (рисунок 35):



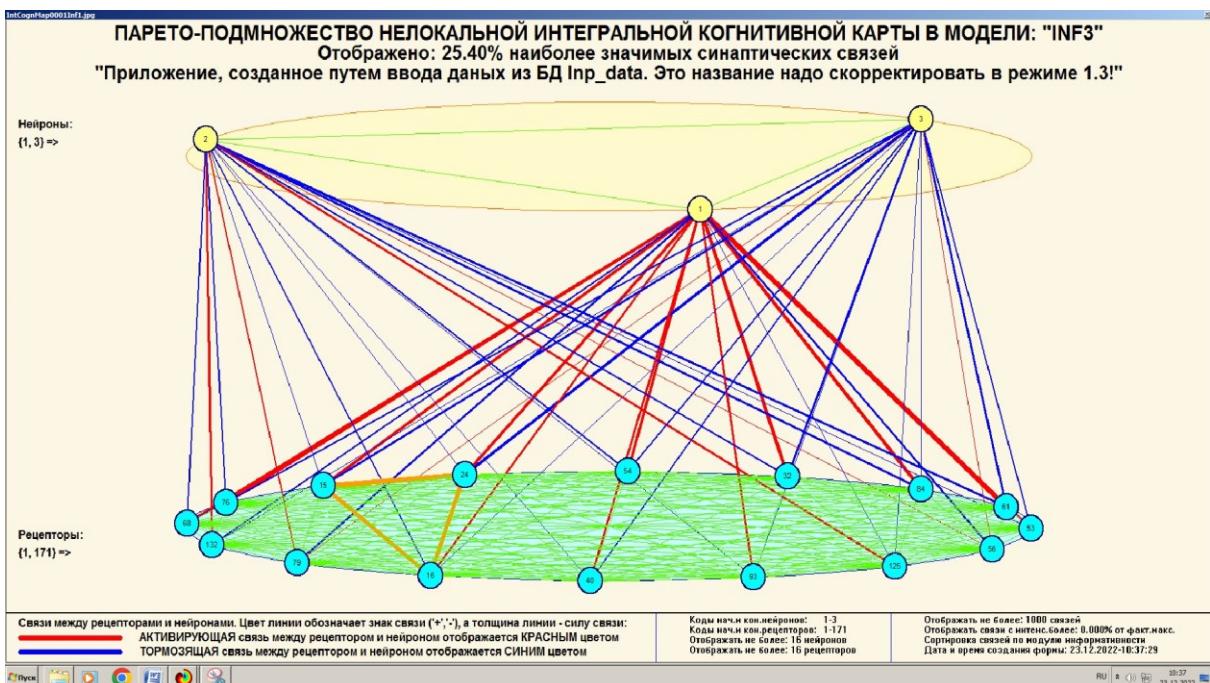


Рисунок 32. 3d-когнитивная диаграмма классов и признаков (режим 4.4.12)

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 [23]. Впервые об автоматизированной реализации рассуждений подобного типа в интеллектуальной системе «Эйдос» написано в 2002 году в работе [2] на странице 521¹¹. Позже об этом писалось в работе [3]¹² 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

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.

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.

Примеры экранной формы управления и нескольких 2d-интегральных когнитивных карт содержательного сравнения классов по их системе детерминации приведены ниже на рисунках 36. Всего системой в данной модели генерируется 324 подобных диаграмм, поэтому, естественно, все они не приводятся. Но пользователь при желании всегда может скачать и установить систему «Эйдос» с сайта разработчика: http://lc.kubagro.ru/aidos/_Aidos-X.htm, а затем в диспетчере приложений (режим 1.3) скачать и установить интеллектуальное облачное Эйдос-приложение №335 и получить в нем все выходные формы, как описано в данной статье.

4.2.3. Когнитивные диаграммы классов. Задание параметров генерации выходных форм

Выбор классов для когнитивной диаграммы

Задайте коды двух классов, для левого и правого информационных портретов когнитивной диаграммы по очереди выбирая их курсором в таблице и кликая на соответствующей кнопке ниже нее

Код	Наименование класса
0	ВСЕ КЛАССЫ
1	LEVEL-1-High
2	LEVEL-2-Medium
3	LEVEL-3-Low

Выбор кода класса левого инф.портрета Выбор кода класса правого инф.портрета

Выбор способа фильтрации признаков в информационных портретах когнитивной диаграммы

Задайте коды двух описательных шкал, для левого и правого информационных портретов когнитивной диаграммы по очереди выбирая их курсором в таблице и кликая на соответствующей кнопке ниже нее

Код	Наименование описательной шкалы	Минимальный код градации	Максимальный код градации
0	ВСЕ ОПИСАТЕЛЬНЫЕ ШКАЛЫ	1	171
1	AGE	1	5
2	GENDER	6	7
3	AIR POLLUTION	8	15
4	ALCOHOL USE	16	23
5	DUST ALLERGY	24	31

Выбор кода описательной шкалы левого инф.портрета Выбор кода описательной шкалы правого инф.портрета

Задайте модели, в которых проводить расчеты когнитивных диаграмм

Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7

Задайте max количество отображаемых связей: Помощь

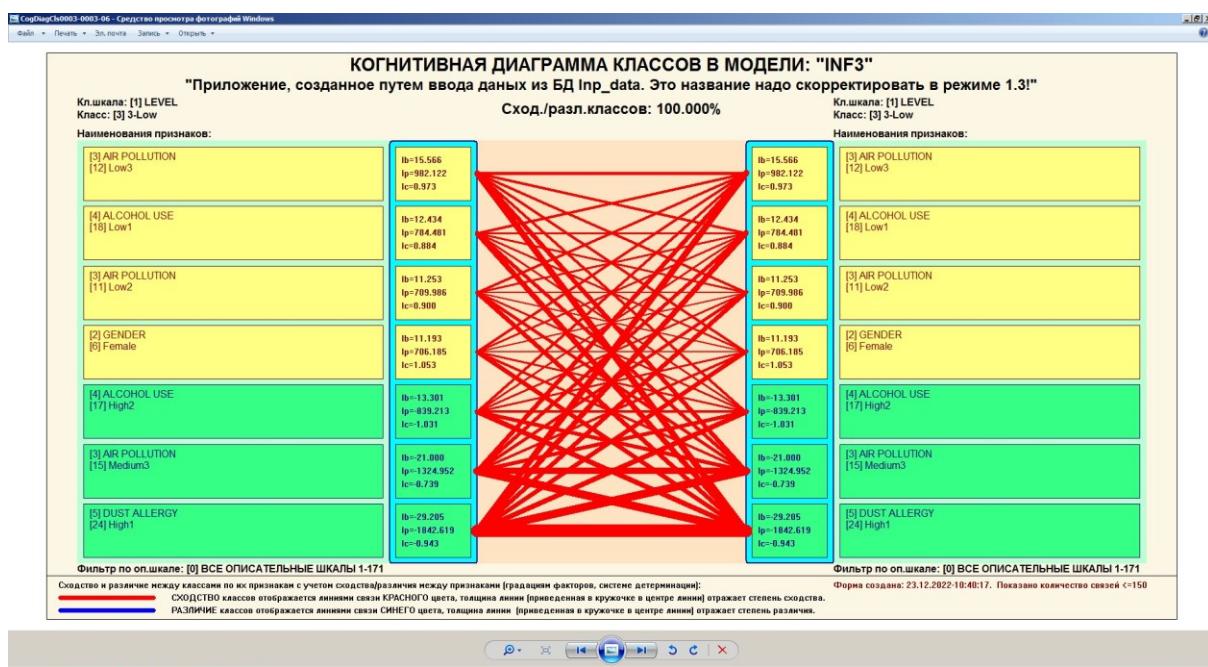
В диалоге заданы следующие параметры расчета когнитивных диаграмм:

Класс для левого инф.портрета: [0] ВСЕ КЛАССЫ
Класс для правого инф.портрета: [0] ВСЕ КЛАССЫ
Описат.шкала для левого инф.портрета: [0] ВСЕ ОПИСАТЕЛЬНЫЕ ШКАЛЫ
Описат.шкала для правого инф.портрета: [0] ВСЕ ОПИСАТЕЛЬНЫЕ ШКАЛЫ
Модели, заданные для расчета: Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7

Задайте режим вывода когнитивных диаграмм:

Показать все диаграммы с остановкой
 Записать все диаграммы без показа

Ok Cancel



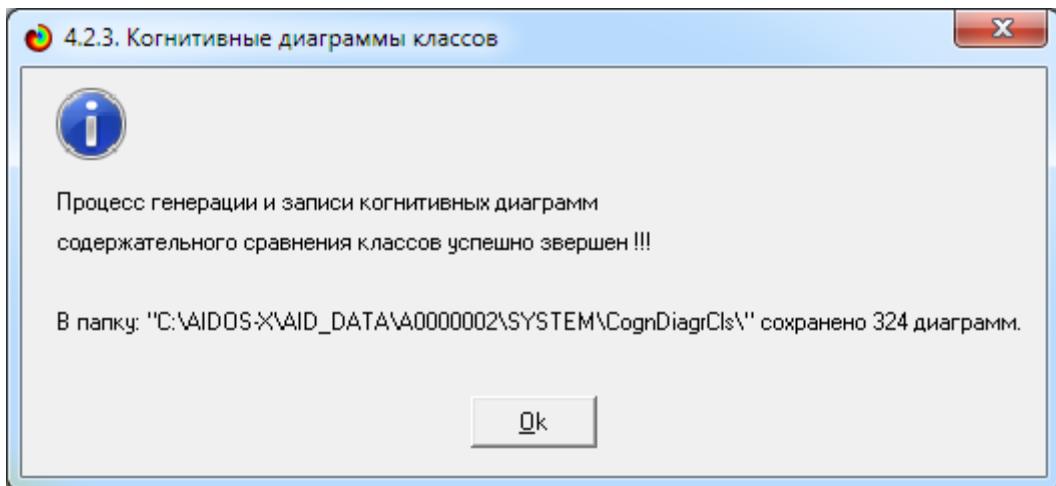


Рисунок 33. Примеры 2d-интегральных когнитивных карт содержательного сравнения классов по их системе детерминации в СК-модели INF3

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

Из 2d-когнитивных диаграммах сравнения значений факторов по их влиянию на объект моделирования, т.е. на его переходы в состояния, соответствующие классам вполне понятно, насколько сходны или отличаются любые два значения факторов по их смыслу.

Напомним, что смысл, согласно концепции смысла Шенка-Абельсона, используемой в АСК-анализе, состоит в знании причин и последствий [24].

Однако из этой диаграммы не видно, чем именно конкретно сходны или отличаются значения факторов по их смыслу.

Это видно из когнитивных диаграмм, которые можно получить в режиме 4.3.3 системы «Эйдос».

Примеры экранной формы управления и нескольких 2d-интегральных когнитивных карт содержательного сравнения значений факторов по их силе и направлению их влияния на переход объекта моделирования в будущие состояния, соответствующие классам, приведены ниже на рисунках 37.

4.3.3. Когнитивные диаграммы признаков. Задание параметров генерации выходных форм

Выбор признаков для когнитивной диаграммы

Задайте коды двух признаков, для левого и правого информационных портретов когнитивной диаграммы по очереди выбирая их курсором в таблице и кликая на соответствующей кнопке ниже нее

Код	Наименование признака
0	ВСЕ ПРИЗНАКИ
1	AGE-1/5-{14.000000, 25.800000}
2	AGE-2/5-{25.800000, 37.600000}
3	AGE-3/5-{37.600000, 49.400000}
4	AGE-4/5-{49.400000, 61.200000}
5	AGE-5/5-{61.200000, 73.000000}
6	GENDER-Female
7	GENDER-Male

Выбор кода признака левого инф.портрета Выбор кода признака правого инф.портрета

Выбор способа фильтрации классов в информационных портретах когнитивной диаграммы

Задайте коды двух классификационных шкал, для левого и правого информационных портретов когнитивной диаграммы по очереди выбирая их курсором в таблице и кликая на соответствующей кнопке ниже нее

Код	Наименование классификационной шкалы	Минимальный код градации	Максимальный код градации
0	ВСЕ КЛАССИФИКАЦИОННЫЕ ШКАЛЫ	1	3
1	LEVEL	1	3

Выбор кода классификационной шкалы левого инф.портрета Выбор кода классификационной шкалы правого инф.портрета

Задайте модели, в которых проводить расчеты когнитивных диаграмм:

Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7

Задайте max количество отображаемых связей:

В диалоге заданы следующие параметры расчета когнитивных диаграмм:

Признак для левого инф.портрета: [0] ВСЕ ПРИЗНАКИ
Признак для правого инф.портрета: [0] ВСЕ ПРИЗНАКИ
Классиф.шкала для левого инф.портрета: [0] ВСЕ КЛАССИФИКАЦИОННЫЕ ШКАЛЫ
Классиф.шкала для правого инф.портрета: [0] ВСЕ КЛАССИФИКАЦИОННЫЕ ШКАЛЫ
Модели, заданные для расчета: Abs, Prc1, Prc2, Inf1, Inf2, Inf3, Inf4, Inf5, Inf6, Inf7

Задайте режим вывода когнитивных диаграмм:

Показать все диаграммы с остановкой
 Записать все диаграммы без показа

Ok **Cancel**

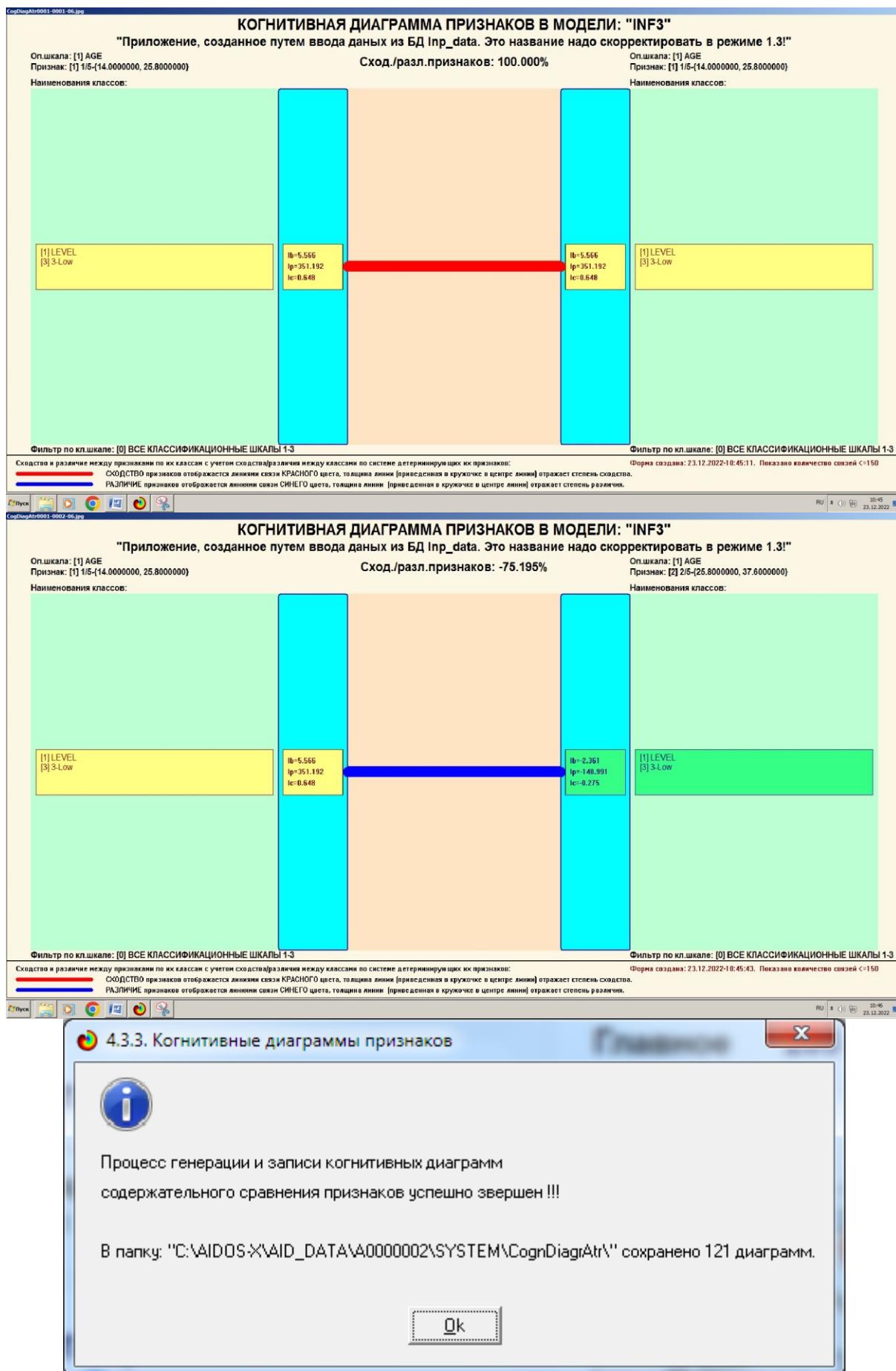


Рисунок 34. Примеры 2d-интегральных когнитивных карт содержательного сравнения значений факторов по их влиянию на переход объекта моделирования в состояния, соответствующие классам в СК-модели INF3

Всего системой в данной модели генерируется 121 подобных диаграмм, поэтому, естественно, все они не приводятся. Но пользователь при желании всегда может скачать и установить систему «Эйдос» с сайта разработчика: http://lc.kubagro.ru/aidos/_Aidos-X.htm, а затем в диспетчере приложений (режим 1.3) скачать и установить интеллектуальное облачное Эйдос-приложение №**335** и получить в нем все выходные формы, как описано в данной статье.

3.8.9. Когнитивные функции

Когнитивные функции являются обобщением классического математического понятия функции на основе системной теории информации и предложены Е.В.Луценко в 2005 году [25, 26].

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 фактора на переход объекта моделирования в каждое из будущих состояний.

В системе «Эйдос» когнитивные функции отображаются в режиме 4.5 (рисунки 38). 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 [21]. Содержательное объяснение этих

эмпирических закономерностей формулируется уже экспертами на теоретическом уровне познания в содержательных научных законах [22].

4.5. Визуализация когнитивных функций

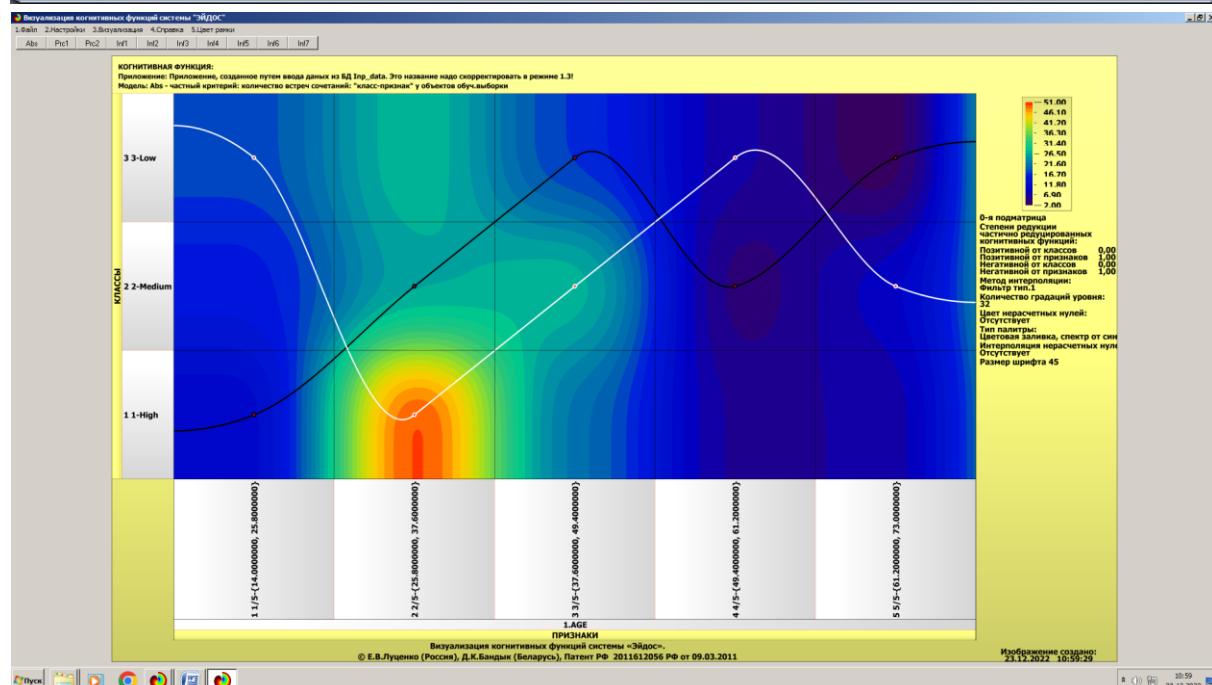
Что такое когнитивная функция:

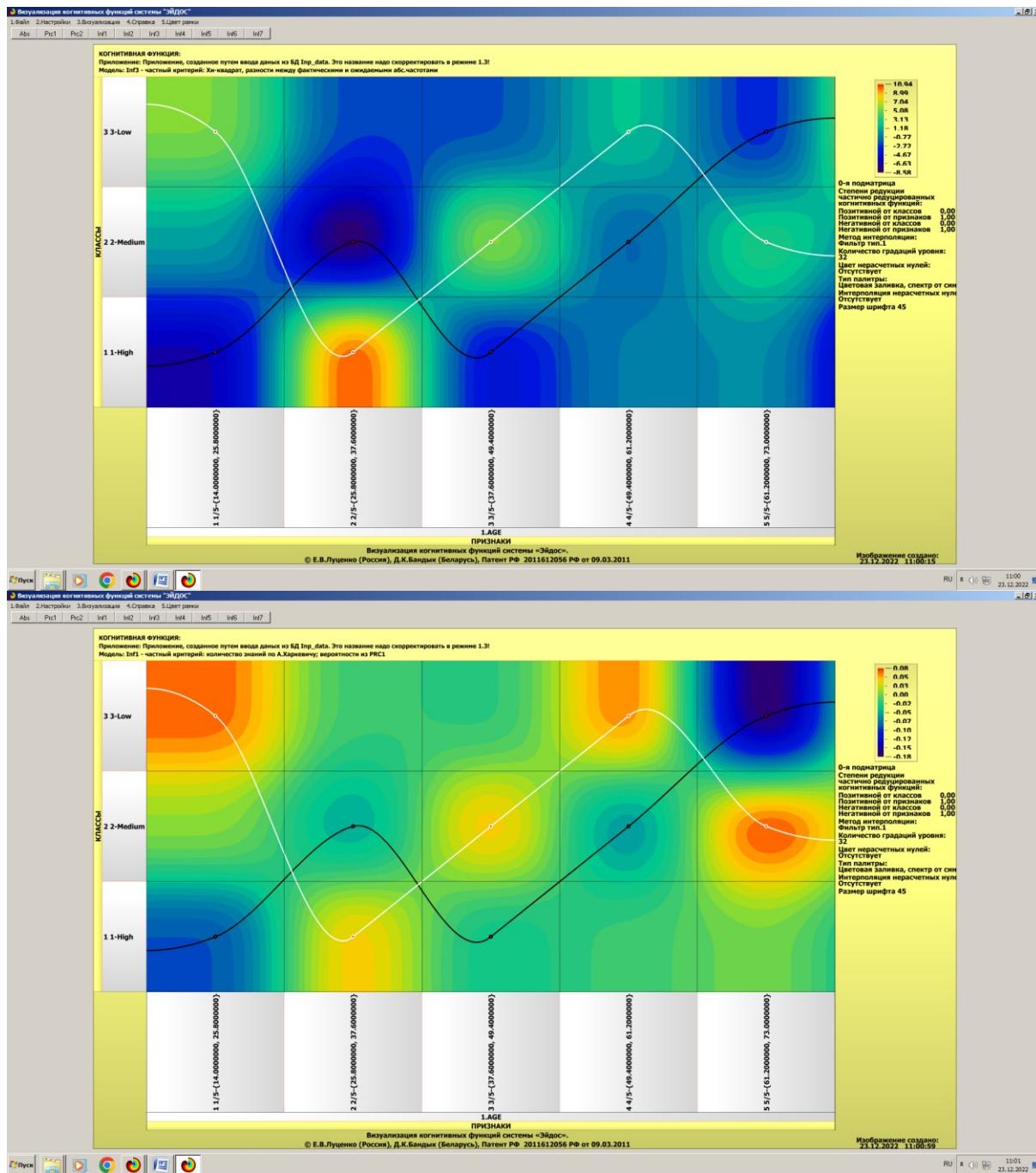
Визуализация прямых, обратных, позитивных, негативных, полностью и частично редуцированных когнитивных функций Когнитивная функция представляет собой графическое отображение силы и направления влияния различных значений некоторого фактора на переходы объекта управления в будущие состояния, соответствующие классам. Когнитивные функции представляют собой новый перспективный инструмент отражения и наглядной визуализации закономерностей из 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.

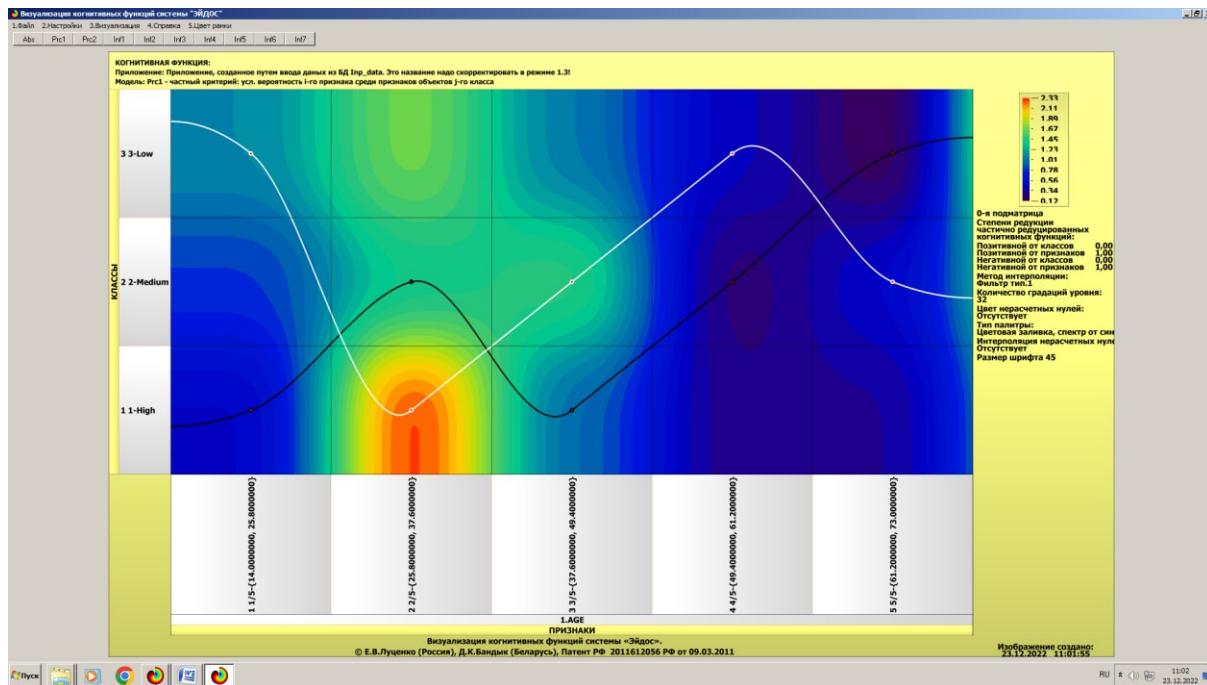
Луценко Е.В. Метод визуализации когнитивных функций - новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета [Научный журнал КубГАУ] [Электронный 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>

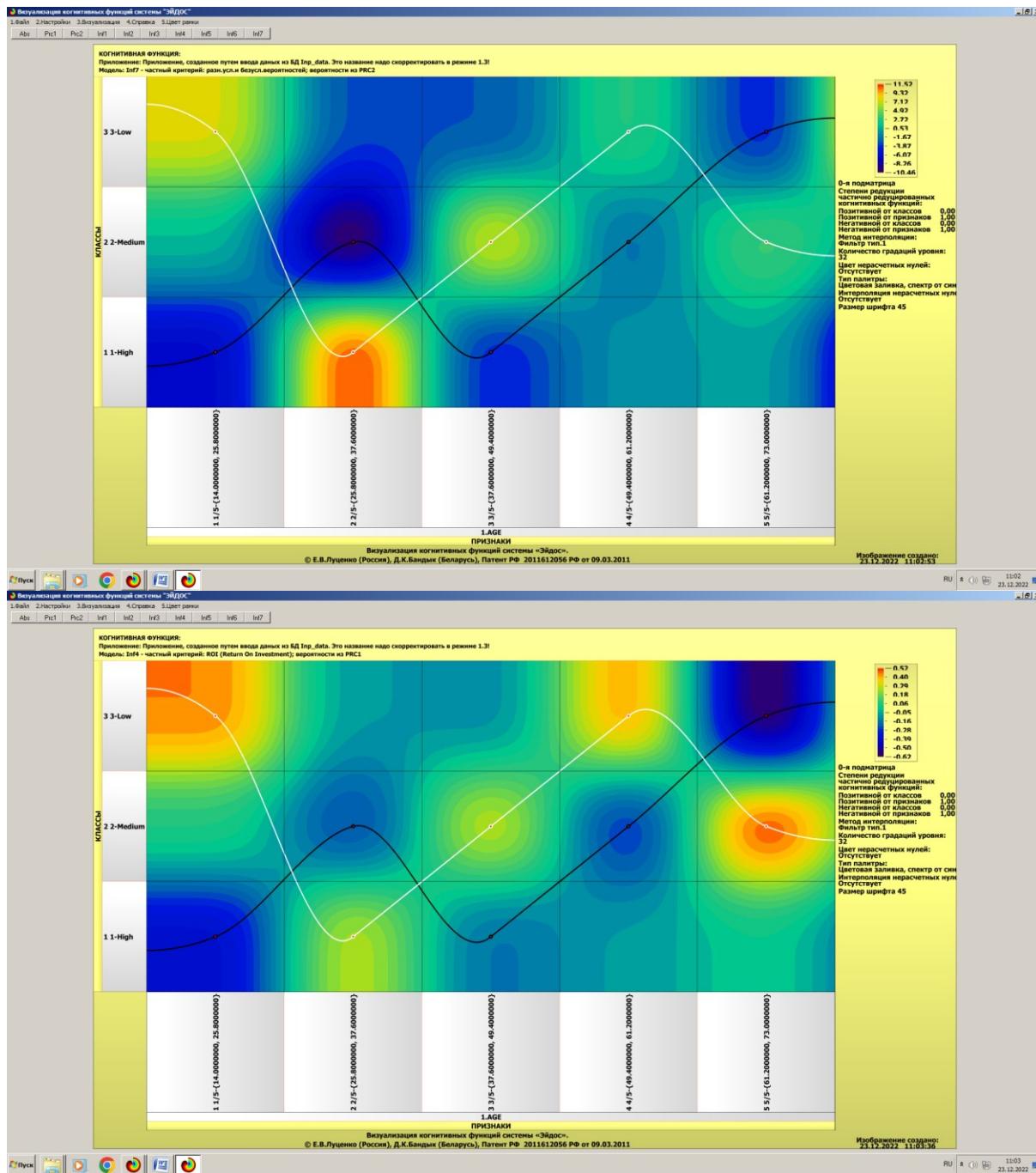
Задайте нужный режим:

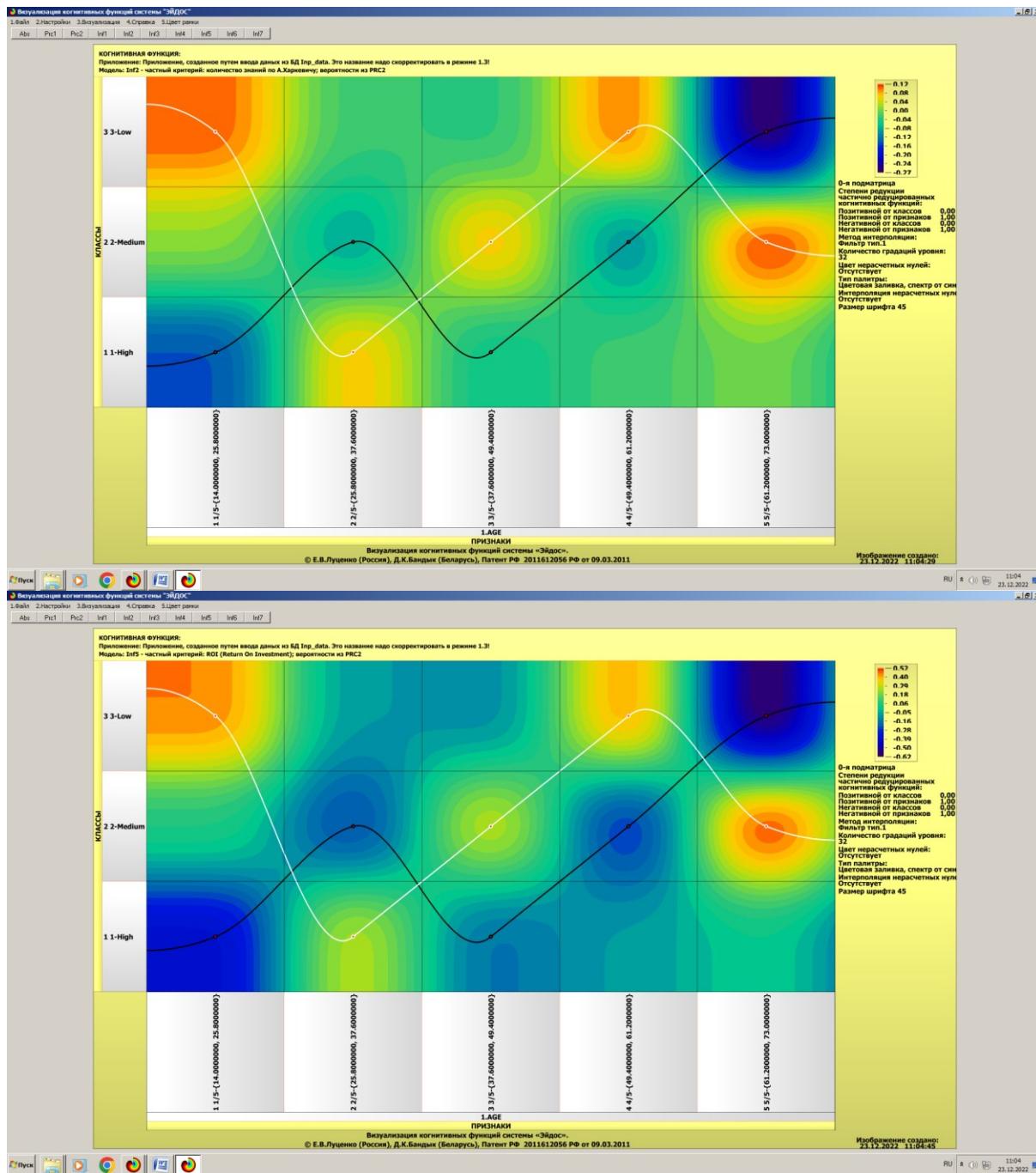
Визуализации когнитивных функций	Литератур.ссылки на работы по когнитивным функциям
Литератур.ссылки на работы по когнитивным функциям	Литератур.ссылки на работы по управлению знаниями











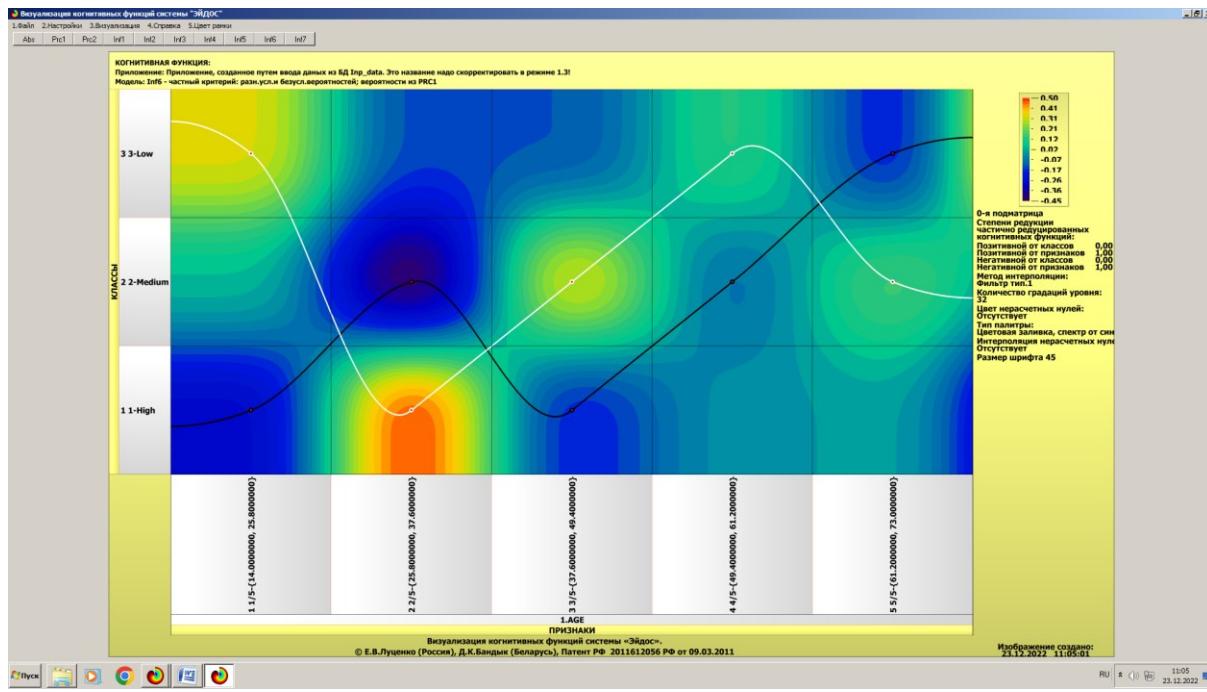


Рисунок 35. Примеры когнитивных функций в СК-модели INF1

Как уже отмечалось содержательное объяснение когнитивных функций на теоретическом уровне познания – это дело специалистов в той предметной области, к которой относится предмет моделирования [22].

3.8.10. 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 [6].

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.

Таблица 13 – Сила влияния значений факторов на поведение объекта моделирования в СК-модели INF3

NUM	NUM_PRC	KOD_OPSC	NAME_OPSC	N_GROPS	KODGR_MIN	KODGR_MAX	ZNACH_OS	ZN_OSNIT	ZNACH_PRC	ZN_PRCNIT
1	4,3478261	10 OBESITY		7	61	67	13,6183948	13,6183948	6,1894168	6,1894168
2	8,6956522	9 BALANCED DIET		7	54	60	12,5657029	26,1840977	5,7109794	11,9003961
3	13,0434783	12 PASSIVE SMOKER		8	76	83	11,8582209	38,0423186	5,3894363	17,2898324
4	17,3913043	14 COUGHING OF BLOOD		9	93	101	11,4505411	49,4928597	5,2041501	22,4939825
5	21,7391304	4 ALCOHOL USE		8	16	23	11,3238007	60,8166604	5,1465479	27,6405305
6	26,0869565	7 GENETIC RISK		7	40	46	11,0243854	71,8410458	5,0104668	32,6509973
7	30,4347826	18 WHEEZING		8	126	133	10,9777959	82,8188417	4,9892924	37,6402898
8	34,7826087	8 CHRONIC LUNG DISEASE		7	47	53	10,7096528	93,5284945	4,8674242	42,5077140
9	39,1304348	2 GENDER		2	6	7	10,6333665	104,1618610	4,8327529	47,3404668
10	43,4782609	5 DUST ALLERGY		8	24	31	10,2099018	114,3717628	4,6402926	51,9807595
11	47,8260870	17 SHORTNESS OF BREATH		8	118	125	10,2013264	124,5730892	4,6363952	56,6171547
12	52,1739130	6 OCCUPATIONAL HAZARDS		8	32	39	10,1437238	134,7168130	4,6102154	61,2273701
13	56,5217391	3 AIR POLLUTION		8	8	15	10,0394830	144,7562960	4,5628391	65,7902092
14	60,8696562	13 CHEST PAIN		9	84	92	9,5993269	154,3556229	4,3627928	70,1530020
15	65,2173913	11 SMOKING		8	68	75	9,5630127	163,9186356	4,3462884	74,4992904
16	69,5652174	21 FREQUENT COLD		7	151	157	9,1602347	173,0788703	4,1632300	78,6625204
17	73,9130435	15 FATIGUE		8	102	109	9,1211958	182,2000661	4,1454873	82,8080077
18	78,2606896	22 DRY COUGH		7	158	164	8,3187557	190,5188218	3,7807867	86,5887944
19	82,6086957	16 WEIGHT LOSS		8	110	117	7,1435128	197,6623346	3,2466512	89,8354456
20	86,9565217	20 CLUBBING OF FINGER NAILS		9	142	150	6,9446053	204,6069399	3,1562498	92,9916954
21	91,3043478	19 SWALLOWING DIFFICULTY		8	134	141	5,5551573	210,1620972	2,5247604	95,5164558
22	95,6521739	1 AGE		5	1	5	5,3112904	215,4733876	2,4139254	97,9303813
23	100,0000000	23 SNORING		7	165	171	4,5537223	220,0271099	2,0696187	100,0000000

На экранной форме рисунка 40 приведены имена Excel-файлов с информацией о силе и направлении влияния значений факторов в разных моделях:

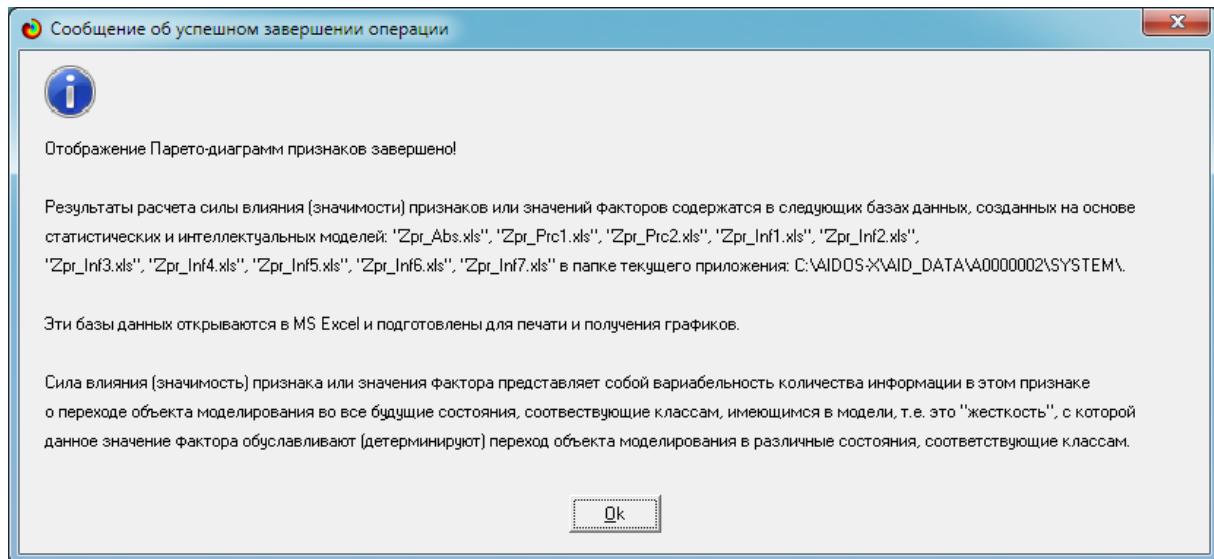


Рисунок 36. Имена Excel-файлов с информацией о силе влияния значений факторов в разных моделях

На экранной форме рисунка 41 приведены имена Excel-файлов с информацией о силе влияния факторов в разных моделях.

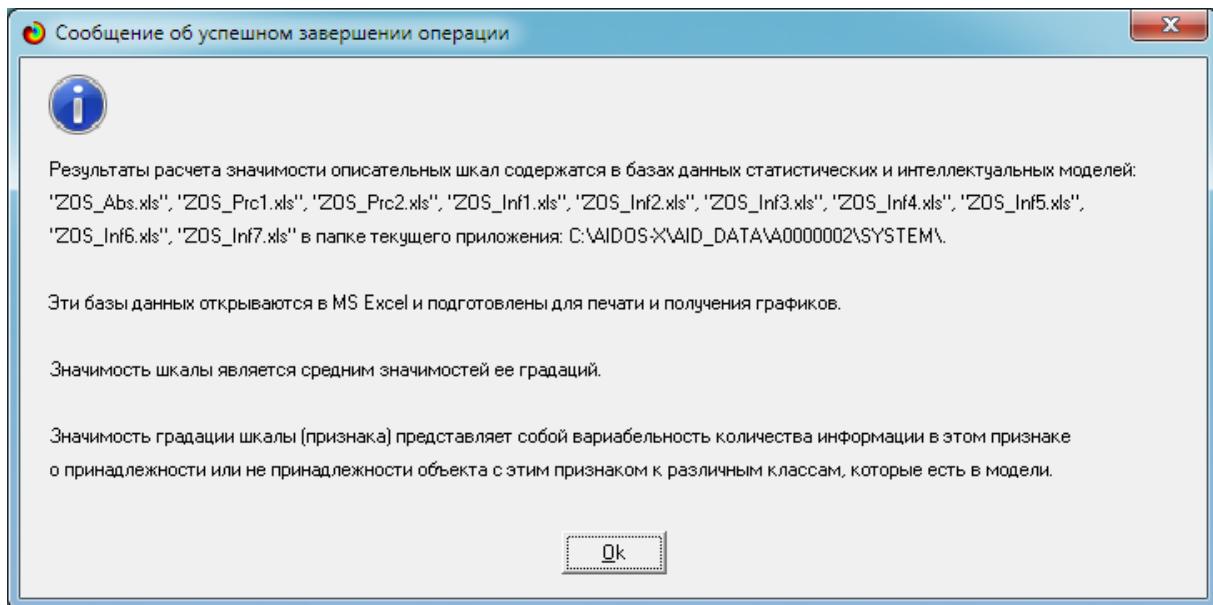


Рисунок 37. имена Excel-файлов с информацией о силе влияния факторов в разных моделях

3.8.11. Степень детерминированности классов и классификационных шкал

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается *степенью вариабельности 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 38 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:

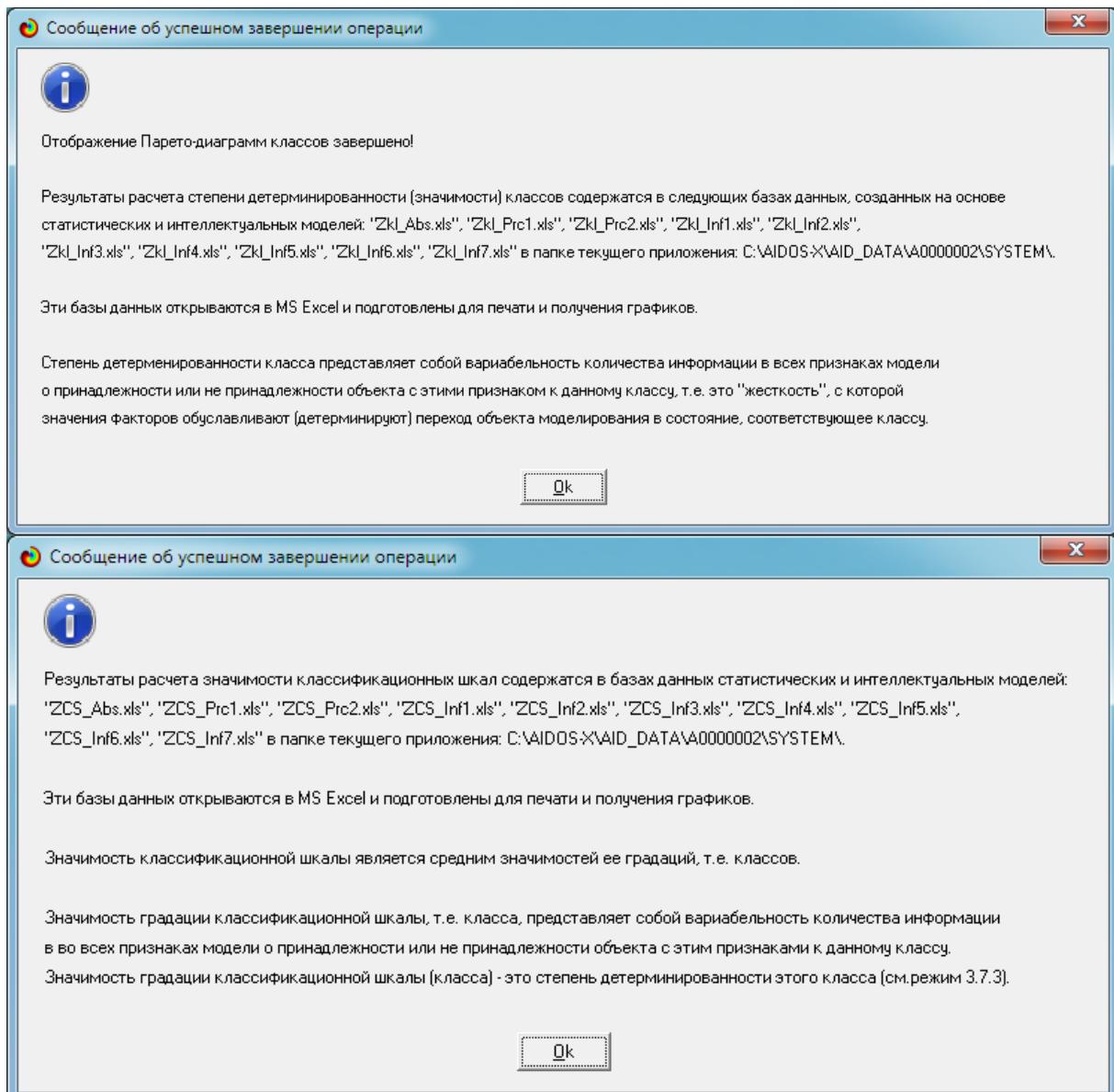


Рисунок 38. Экранные формы режимов 3.7.2 и 3.7.3 системы «Эйдос»

4. DISCUSSION (ОБСУЖДЕНИЕ)

Полученные результаты можно оценить как успешно решающие сформулированную в работе проблему и обеспечивающие достижение поставленной в работе цели. Эти результаты получены путем применения лингвистического Автоматизированного системно-когнитивного анализа (лингвистический АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос».

Анализ полученных результатов, проведенный в данной работе, полностью согласуется с результатами работы [10], на исходных данных которой они основаны. С другой стороны применение АСК-анализа и системы «Эйдос» весьма существенно расширяет возможности решения задач прогнозирования, принятия решений и исследования моделируемой предметной области, по сравнению с методами, применяемыми в

работе [10]. Поэтому есть все основания рекомендовать применение АСК-анализа и системы «Эйдос» для проведения дальнейших углубленных исследований.

Достижением данной работы является:

1. Возможность построения системно-когнитивных моделей предметной области на основе исходных данных, содержащих лингвистические переменные.

2. Возможность применения системно-когнитивных моделей для решения задач прогнозирования, принятия решений и исследования моделируемой предметной области.

В качестве перспективы продолжения исследований можно было бы рекомендовать существенно увеличить объем исходных данных, количество исследуемых факторов, а также количество классификационных шкал и их градаций (классов) для описания будущих состояний объекта моделирования.

Например, можно было бы исследовать в создаваемых системно-когнитивных моделях, не только технологические, но и природно-климатические факторы.

Рекомендуется ввести классификационные шкалы, отражающие влияние исследуемых факторов на объект моделирования не только в натуральном выражении (количество и качество различных видов продукции), но и в стоимостном выражении (прибыль и рентабельность, как общая по предприятию, так и в разрезе по видам продукции).

Перспективность и ценность результатов подобных исследований и разработок для теории и практики не вызывает особых сомнений, что подтверждается работами автора в этой области [1-26].

У желающих есть все возможности для изучения данной работы и для дальнейших исследований с применением АСК-анализа и системы «Эйдос» на своем компьютере.

Для этого надо скачать систему с сайта разработчика по ссылке на странице: http://lc.kubagro.ru/aidos/_Aidos-X.htm, а затем в диспетчере приложений (режим 1.3) установить интеллектуальное облачное Эйдос-приложение №335. По различным аспектам применения данной технологии есть большое количество видео-занятий (около 300), с которыми можно ознакомиться по ссылкам, приведенным на странице: http://lc.kubagro.ru/aidos/How_to_make_your_own_cloud_Eidos-application.pdf.

5. CONCLUSIONS (ВЫВОДЫ)

В работе решена задача выявление зависимостей развития раковых опухолей лёгких от различных факторов. На основе знания этих зависимостей решены задачи прогнозирования, принятия решений и

исследования моделируемой предметной области путем исследования ее системно-когнитивной модели.

Спецификой данной задачи является то, что все независимые переменные являются лингвистическими (категориальными) переменными. Поэтому для решения данной задачи применяется лингвистический АСК-анализ, т.е. когнитивная математическая лингвистика.

Таким образом, в работе строится гибридная модель, включающая как номинальные (текстовые), так и числовые шкалы. Сопоставимость обработки данных разных типов, представленных в разных типах шкал и разных единицах измерения обеспечивается путем метризации номинальных шкал, т.е. повышения их степени формализации до уровня числовых шкал [6].

В работе приводится краткое описание АСК-анализа и его программного инструментария – интеллектуальной системы «Эйдос». Работа может быть основой для лабораторных работ и научных исследований по применению систем искусственного интеллекта, в частности лингвистического АСК-анализа для решения задач в области *когнитивной агрономии*.

REFERENCES (ЛИТЕРАТУРА)

1. Работы проф.Е.В.Луценко & С° по тематике, связанной с АПК, частности с когнитивной агрономией: http://lc.kubagro.ru/aidos/Work_with_agricultural.htm
2. Луценко, Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами : (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем) / Е. В. Луценко. – Краснодар : Кубанский государственный аграрный университет имени И.Т. Трубилина, 2002. – 605 с. – ISBN 5-94672-020-1. – EDN OCZFHC.
3. Орлов, А. И. Системная нечеткая интервальная математика / А. И. Орлов, Е. В. Луценко. – Краснодар : Кубанский государственный аграрный университет, 2014. – 600 с. – ISBN 978-5-94672-757-0. – EDN RZXZ.
4. Работы проф.Е.В.Луценко по АСК-анализу текстов, т.е. по когнитивной математической лингвистике: http://lc.kubagro.ru/aidos/Works_on_ASK-analysis_of_texts.htm
5. Работы проф.Е.В.Луценко по информационным мерам уровня системности (коэффициентам эмержентности) и системному обобщению математики http://lc.kubagro.ru/aidos/Work_on_emergence.htm
6. Луценко, Е. В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе "Эйдос" / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2013. – № 92. – С. 61-71. – EDN RNEGHR.
7. Сайт Е.В.Луценко: <http://lc.kubagro.ru/>.
8. Страница Е.В.Луценко в РесечГейт <https://www.researchgate.net/profile/Eugene-Lutsenko>

9. Страница Е.В.Луценко в РИНЦ: https://elibrary.ru/author_profile.asp?id=123162.
10. Горпинченко, К. Н. Прогнозирование и принятие решений по выбору агротехнологий в зерновом производстве с применением методов искусственного интеллекта (на примере СК-анализа) / К. Н. Горпинченко, Е. В. Луценко. – Краснодар : Кубанский государственный аграрный университет имени И.Т. Трубилина, 2013. – 168 с. – ISBN 978-5-94672-644-3. – EDN RAIMQL.
11. Луценко, Е. В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе "Эйдос" / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2017. – № 126. – С. 1-32. – DOI 10.21515/1990-4665-126-001. – EDN XXXBDV.
12. 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.
13. 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.
14. Луценко, Е. В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы "Эйдос-Х++" / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2014. – № 101. – С. 1367-1409. – EDN SZVWRV.
15. 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.
16. 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.
17. 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.
18. Луценко, Е. В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе "Эйдос") / Е. В. Луценко, В. Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2011. – № 71. – С. 27-74. – EDN OIGYBB.
19. 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.

20. 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.

21. Работы проф.Е.В.Луценко & С° 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

22. 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>

23. 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.

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

25. 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.

26. База данных, на основе которой был проведен автоматизированный системно-когнитивный анализ зависимости уровня развития раковых опухолей легких от различных факторов: <https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link>