МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ ИМЕНИ И. Т. ТРУБИЛИНА»

Факультет прикладной информатики

Кафедра компьютерных технологий и систем

КУРСОВАЯ РАБОТА

по дисциплине: Интеллектуальные системы и технологии
на тему: АСК-анализ рейтинга фильмов в онлайн кинотеатре «Кинопоиск» на основе данных с портала Kaggle
Выполнил студент группы: ИТ2041 Омельченко Дмитрий Александрович
Допущен к защите: _____

Руководитель проекта: д.э.н., к. т. н., профессор Луценко Е. В. (_____)
(подпись, расшифровка подписи)

Защищен_____20.02.2022_____
(дата)
Оценка_____отлично_____

Краснодар 2022

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

Факультет прикладной информатики

РЕЦЕНЗИЯ
на курсовую работу

Студента Омельченко Дмитрий Александрович
курса 2 очной формы обучения группы ИТ2041
Направление подготовки 09.04.02 «Информационные системы и технологии»
Наименование темы «АСК-анализ рейтинга фильмов в онлайн кинотеатре «Кинопоиск» на основе данный портала Kaggle»
Рецензент: Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор
(Ф.И.О., ученое звание и степень, должность)

Оценка качества выполнения курсовой работы

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

Достоинства работы _____
_____
_____
Недостатки работы _____
_____
_____
Итоговая оценка при защите _____отлично_____
Рецензент _____ (Е. В. Луценко)
«20» февраля 2022 г.

РЕФЕРАТ

Курсовая работа содержит: 64 страницы, 37 рисунков, 15 литературных источников.

Ключевые слова: СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, AIDOS-X, КОГНИТИВНЫЙ АНАЛИЗ, МОДЕЛИ, ШКАЛЫ, КЛАССЫ.

Целью работы является провести автоматизированный системно-когнитивный анализ рейтинга фильмов в онлайн кинотеатре «Кинопоиск» на основе данный портала Kaggle. Добиться этого можно анализом методов формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами, принятия решений и исследования моделируемой предметной области путем исследования модели.

СОДЕРЖАНИЕ

СОДЕРЖАНИЕ **4**

1. ВВЕДЕНИЕ **5**

1.1. Описание исследуемой предметной области 5
1.2. Объект и предмет исследования 5
1.3. Проблема, решаемая в работе и ее актуальность 5
1.4. Цель и задачи работы 8

2. МЕТОДЫ **10**

2.1. Обоснование требований к методу решения проблемы 10
2.3. Автоматизированный системно-когнитивный анализ (АСК-анализ) 12
2.4. Система «Эйдос» - инструментарий АСК-анализа 14

3. РЕЗУЛЬТАТЫ **20**

3.1. Задача-1. Когнитивная структуризация предметной области. Две интерпретации классификационных и описательных шкал и градаций 20
3.2. Задача-2. Формализация предметной области 21
3.3. Задача-3. Синтез статистических и системно-когнитивных моделей. Многопараметрическая типизация и частные критерии знаний 26
3.4. Задача-4. Верификация моделей 32
3.5. Задача-5. Выбор наиболее достоверной модели 37
3.6. Задача-6. Системная идентификация и прогнозирование 40
3.6.1. Интегральный критерий «Сумма знаний» *44*
3.6.2. Интегральный критерий «Семантический резонанс знаний» *45*
3.6.3. Важные математические свойства интегральных критериев *47*
3.7. Задача-7. Поддержка принятия решений 48
3.7.1. Упрощенный вариант принятия решений как обратная задача прогнозирования, позитивный и негативный информационные портреты классов, SWOT-анализ *48*
3.7.2. Развитый алгоритм принятия решений в АСК-анализе *49*
3.8. Задача-8. Исследование объекта моделирования путем исследования его модели 49
3.8.1. Инвертированные SWOT-диаграммы значений описательных шкал (семантические потенциалы) *49*
3.8.2. Кластерно-конструктивный анализ классов *50*
3.8.3. Кластерно-конструктивный анализ значений описательных шкал *52*
3.8.4. Модель знаний системы «Эйдос» и нелокальные нейроны *54*
3.8.5. Нелокальная нейронная сеть *55*
3.8.6. 3D-интегральные когнитивные карты *56*
3.8.7. 2D-интегральные когнитивные карты содержательного сравнения классов (опосредованные нечеткие правдоподобные рассуждения) *56*
3.8.8. 2D-интегральные когнитивные карты содержательного сравнения значений факторов (опосредованные нечеткие правдоподобные рассуждения) *58*
3.8.9. Значимость описательных шкал и их градаций *59*
3.8.10. Степень детерминированности классов и классификационных шкал *60*

4. ОБСУЖДЕНИЕ **61**

ЗАКЛЮЧЕНИЕ **62**

СПИСОК ЛИТЕРАТУРЫ **63**

1. ВВЕДЕНИЕ

1.1. Описание исследуемой предметной области

Современная информационные системы получают все большее развитие благодаря технологиям искусственного интеллекта. Оценка качества математических моделей некоторых из них не выносит критики. В данной курсовой работе рассмотрено решение задачи АСК-анализ рейтинга фильмов в онлайн кинотеатре «Кинопоиск» на основе данный портала Kaggle.

1.2. Объект и предмет исследования

Целью данной курсовой работы является.

Задачами, поставленными в данной курсовой работе, являются:

– подготовка исходных данных и формализация предметной области;

– синтез и верификация статистических и системно-когнитивных моделей и выбор наиболее достоверной модели;

– решение различных задач в наиболее достоверной модели: прогнозирование, поддержка принятия решений, исследование полученных моделей.

Объектом исследования данной работы является выборка данных рейтингов фильмов в онлайн кинотеатре «Кинопоиск».

1.3. Проблема, решаемая в работе и ее актуальность

Проблема – это несоответствие фактического положения дел и желаемого (целевого) в предмете исследования, противоречие между фактическим и желаемым, и это несоответствие и противоречие неприемлемо.

Если в работе не ставится и не решается какая-либо проблема, то эта работа *не актуальна,* т.е. вообще не нужна и непонятно зачем она написана. Поэтому проблема, решаемая в работе, обязательно должна быть сформулирована в самом ее начале. Предыдущие разделы являются подготовительными и создают для этого необходимые предпосылки, чтобы постановка проблемы была обоснованной и убедительной.

О соотношении содержания понятий «Проблема» и «Задача».

Задача – это *простая* проблема, а проблема – это *сложная* задача.

Сложность относительна, т.е. зависит от степени *компетентности* исследователя и степени его *информированности* и возможностей информационного *поиска*. Если исследователю известен метод разрешения проблемной ситуации, то для него это задача, а если неизвестен или он вообще не существует, то проблема.

Предмет кандидатской диссертации – это решение задачи, путем применения *известного* метода, но *адаптированного*, доработанного для конкретного предмета исследования. В этой адаптации, проведенной лично автором исследования, и состоит *научная новизна*. Практическая значимость решения задачи состоит в том, что ранее она не была решена в предмете исследования и это решение может быть применено в объекте исследования. Кандидатская диссертация – это прикладное научное исследование.

Предмет докторской диссертации – это решение проблемы, путем разработки нового метода. Иначе говоря, поиск метода решения проблемы, соответствующего обоснованным требованиям, не дал положительного результата и пришлось разработать собственный метод. Научная новизна работы состоит в разработке нового ранее неизвестного метода, обеспечивающего решение не только проблемы, решаемой в работе, но и сходных проблем в других предметных областях. Практическая значимость решения проблемы состоит в том, что это решение может быть применено не только в предмете и объекте исследования, но и в других предметных областях, а также в том, что для последующих поколений исследователей эта

проблемная ситуация уже будет квалифицироваться не как проблема, а как задача. Таким образом, докторская диссертация имеет межотраслевое звучание и переводит решаемую проблему в класс задачи. Докторская диссертация – это фундаментальное научное исследование.

Интересно, что исторически первое логически безупречное широко известное нам описание проблемной ситуации, по-видимому, дал Будда в своих четырех благородных истинах: существует страдание; существует причина страдания – желание; существует прекращение страдания – нирвана; существует путь, ведущий к прекращению страдания – восьмеричный путь.

Констатация проблемной ситуации дана Буддой в следующих словах: «Существует страдание» (фактическая ситуация), «Существует прекращение страдания» (целевое состояние). Способ перехода от нежелательного фактического состояния к желательному целевому был ранее неизвестен, и Будда сообщает, что «Существует путь, ведущий к прекращению страдания» (известный как восьмеричный путь). Этот путь состоит в последовательном устранении причин страданий (восьмеричный путь), в качестве которых Будда указывает на желания. Страдания могут быть устранены двумя способами:1) созданием внешней ситуации, не вызывающей страданий; 2) таким изменением самого себя, в результате которого любая внешняя ситуация не вызывает страданий. Будда описывает 2-й способ.

Результатом данный работы можно считать получение теоретических и практических знаний в области анализа работы систем искусственного интеллекта и анализа результата их работы.

Курсовая работа состоит из введения, двух глав, заключения, списка литературы и приложения. Общий объем работы составляет 67 страницы.

1.4. Цель и задачи работы

Не смотря на описанные выше различия между задачей и проблемой обычно в научных исследованиях говорят о решении проблемы.

Это связано с тем, что обычно после формулировки решаемой проблемы ставится цель ее решить, а затем осуществляется декомпозиция цели в последовательность задач, решение которых является этапами достижения цели.

А иначе бы получалось, что есть ряд подзадач, являющихся этапами решения задачи, что чисто стилистически неудобно. Поставленные задачи обычно становятся подразделами научной работы, например, подразделами последующих глав диссертации и разделов статьи.

Поскольку для решения поставленной проблемы используется автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий, в качестве которого в настоящее время выступает интеллектуальная система «Эйдос», то достижение поставленной цели обеспечивается решением следующих задач и подзадач, которые являются этапами достижения цели:

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

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 (Инвертированные SWOT-диаграммы значений описательных шкал (семантические потенциалы); Кластерно-конструктивный анализ классов; Кластерно-конструктивный анализ значений описательных шкал; Модель знаний системы «Эйдос» и нелокальные нейроны; Нелокальная нейронная сеть; 3d-интегральные когнитивные карты; 2d-интегральные когнитивные карты содержательного сравнения классов (опосредованные нечеткие правдоподобные рассуждения); 2d-интегральные когнитивные карты содержательного сравнения значений факторов (опосредованные нечеткие правдоподобные рассуждения); Когнитивные функции; Значимость описательных шкал и их градаций; Степень детерминированности классов и классификационных шкал).

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

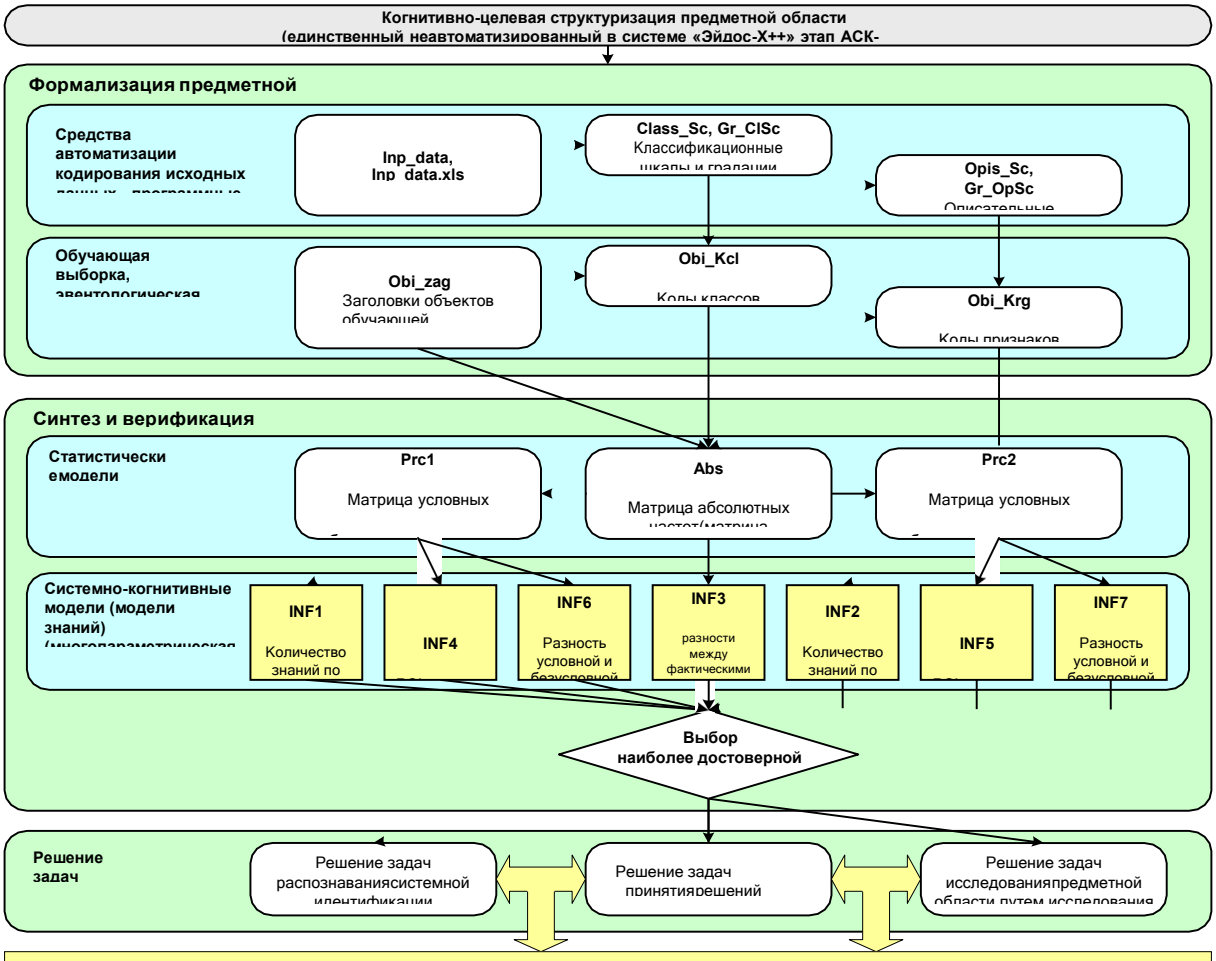


Рисунок 1. Последовательность решения задачв АСК-анализе и системе «Эйдос»

Рассмотрим решение поставленных задач в подробном численном примере.

2. МЕТОДЫ

2.1. Обоснование требований к методу решения проблемы

В качестве метода исследования, решения проблемы и достижения цели было решено использовать новый метод искусственного интеллекта:

Автоматизированный системно-когнитивный анализ или АСК-анализ. Основной причиной выбора АСК-анализа является то, что он включает теорию и метод количественного выявления в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал и единицах измерения.

Очень важным является также то, что АСК-анализ имеет свой развитый доступный программный инструментарий, в качестве которого в настоящее время выступает Универсальная когнитивная аналитическая система Aidos-X. Система «Эйдос» выгодно отличается от других интеллектуальных систем следующими параметрами:

- разработана в универсальной постановке, не зависящей от предметной области. Поэтому она является универсальной и может быть применена во многих предметных областях, в которых не требуется автоматического, т. е. без непосредственного участия человека в реальном времени решения задач идентификации, прогнозирования, приятия решений и исследования предметной области;
- находится в полном открытом бесплатном доступе причем с актуальными исходными текстами;
- 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»;

– 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);

– содержит большое количество локальных (поставляемых с инсталляцией) и облачных учебных и научных Эйдос-приложений;

– поддерживает on-line среду накопления знаний и широко используется во всем мире;

– 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 решение этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA);

– 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;

– 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 повторности 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.

2.3. Автоматизированный системно-когнитивный анализ (АСК-анализ)

Автоматизированный системно-когнитивный анализ (АСК-анализ) предложен проф.Е.В.Луценко в 2002 году в ряде статей и фундаментальной монографии [1]. Сам термин: «Автоматизированный системно-когнитивный анализ (АСК-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 слов[1].

АСК-анализ включает:

– теоретические основы, в частности базовую формализуемую когнитивную концепцию;

– математическую модель, основанную на системном обобщении теории информации (СТИ);

– методику численных расчетов (структуры баз данных и алгоритмы их обработки);

– программный инструментарий, в качестве которого в настоящее время выступает универсальная когнитивная аналитическая система «Эйдос» (интеллектуальная система «Эйдос»).

Более подробно АСК-анализ описан в работах [2, 3] и ряде других. Около половины из 650 опубликованных автором научных работ посвящены теоретическим основам АСК-анализа и его практическим применениям в ряде предметных областей. На момент написания данной работы автором опубликовано более 38 монографий, 27 учебных пособий, в т.ч. 3 учебных пособия с грифами УМО и Министерства, получен 31 патент РФ на системы искусственного интеллекта, 334 публикации в изданиях, входящих в перечень ВАК РФ и приравненных им (по данным РИНЦ), 6 статей в журналах, входящих в WoS, 5 публикаций в журналах, входящих в Скопус[2].

Три монографии включены в фонды библиотеки конгресса США[3].

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

Автор является основателем междисциплинарной научной школы: «Автоматизированный системно-когнитивный анализ»[4]. Научная школа: "Автоматизированный системно-когнитивный анализ" является междисциплинарным научным направлением на пересечении по крайней мере трех научных специальностей (согласно недавно утвержденной новой номенклатуры научных специальностей ВАК РФ[5]). 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;

– 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 [4] и страничка в РесечГейт [5], на которых можно получить более полную информацию о методе АСК-анализа. Краткая информация об АСК-анализе и системе «Эйдос» есть в материале: http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf.

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, приятия 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/), 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 300, соответственно: http://aidos.byethost5.com/Source_data_applications/WebAppls.htm) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);

– поддерживает online 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 решение этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA);

– 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 законах[6].

В разработке системы «Эйдос» были следующие этапы:

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

2-й этап, «эра IBM PC и MS DOS»: 1992-2012 годы. Для IBM-совместимых персональных компьютеров система "Эйдос" впервые реализована на языках CLIPPER-87 и CLIPPER-5.01 (5.02) в 1992 году, а в 1994 году уже были получены свидетельства РосПатента, первые в Краснодарском крае и, возможно, в России на системы искусственного

интеллекта (слева приведена титульная видеограмма финальной DOS-версии системы «Эйдос-12.5», июнь 2012 года). С тех пор и до настоящего времени система непрерывно совершенствуется на IBM PC.

3-й этап, «эра MS Windows xp, 8, 7»: 2012-2020 годы. С июня 2012 по 14.12.2020 система «Эйдос» развивалась на языке Аляска-1.9 + Экспресс++ + библиотека для работы с Internet xb2net. Система «Эйдос-X1.9» хорошо работала на всех версиях MS Windows кроме Windows-10, которая требовала специальной настройки. Наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решение этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA).

4-й этап, «эра MS Windows-10»: 2020-2021 годы. С 13.12.2020 года по настоящее время система «Эйдос» развивается на языке Аляска-2.0 + Экспресс++. Библиотека 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»:

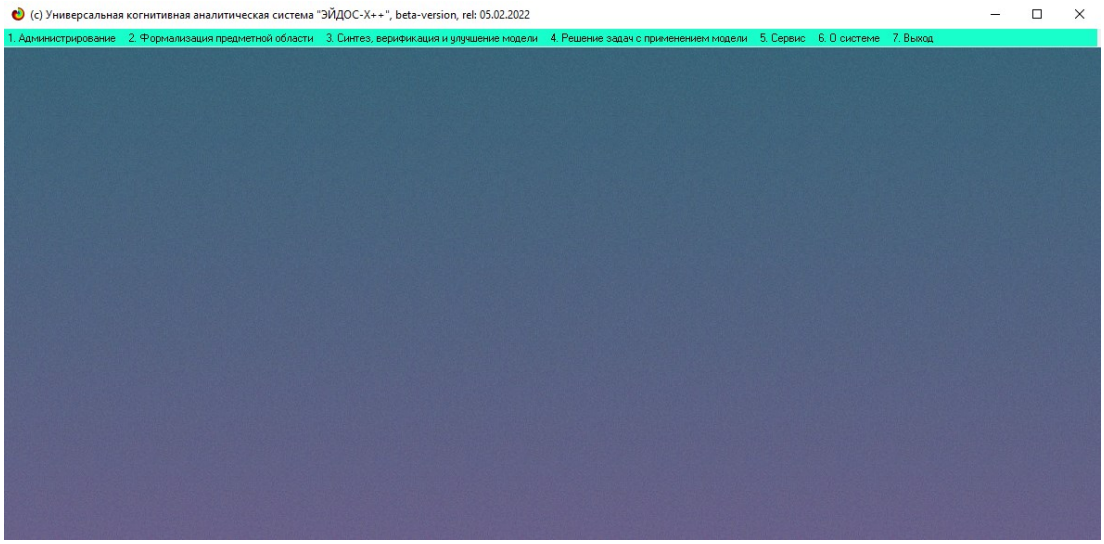


Рисунок 1. Титульная видеограмма DOS-версии системы «Эйдос» [7]

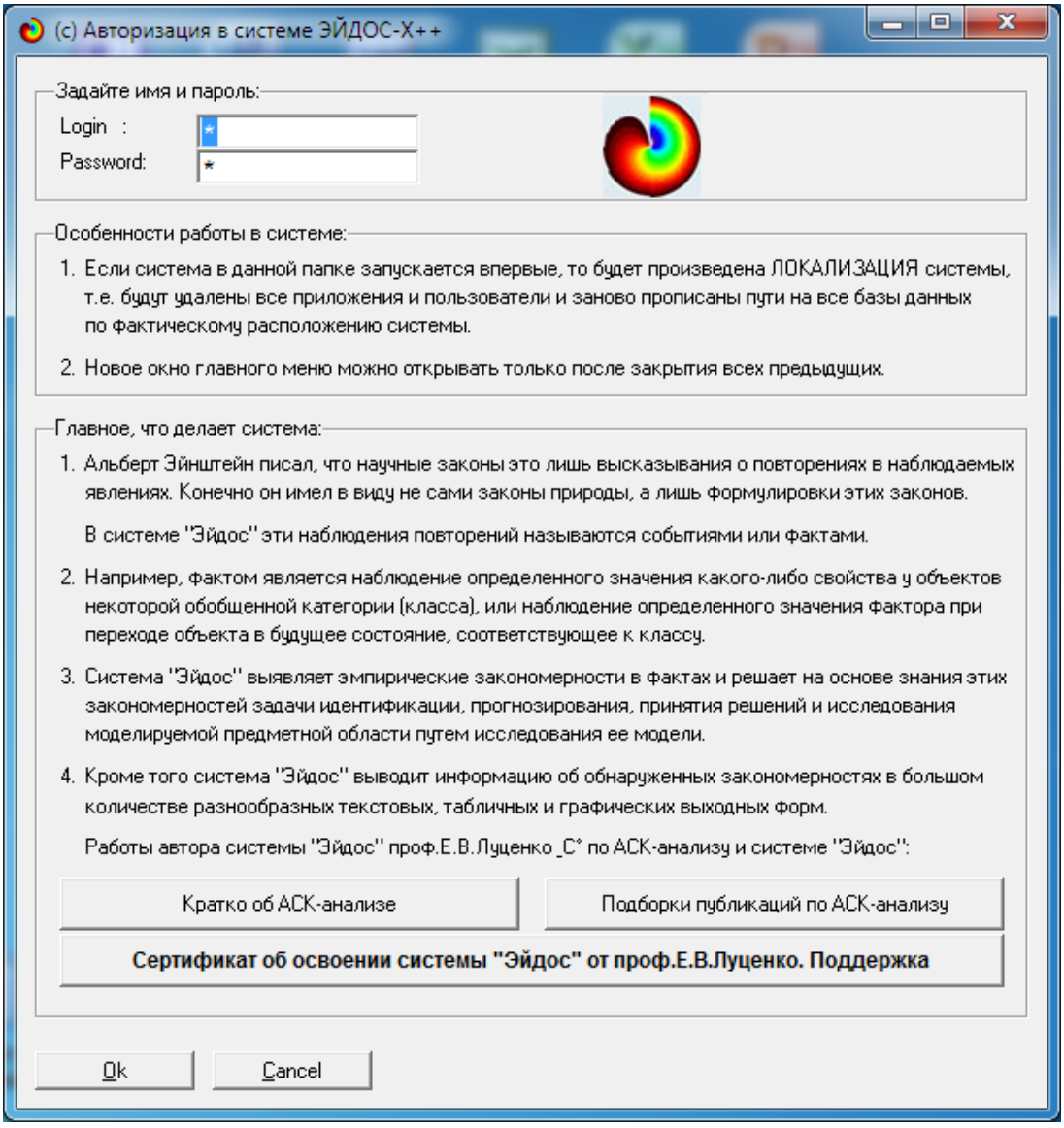


Рисунок 2. Титульная видеограмма текущей версии системы «Эйдос»

3. 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 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).

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);

– 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;

Для импорта обучающей выборки в систему AIDOS-X необходимо скопировать ее в папку Inp_data и переименовать в Inp_data.xlsx, после этого можно запустить саму программу и универсальный программный интерфейс импорта данных в систему (режим 2.3.2.2), результат заполнения которого представлен на рисунке 3.

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.

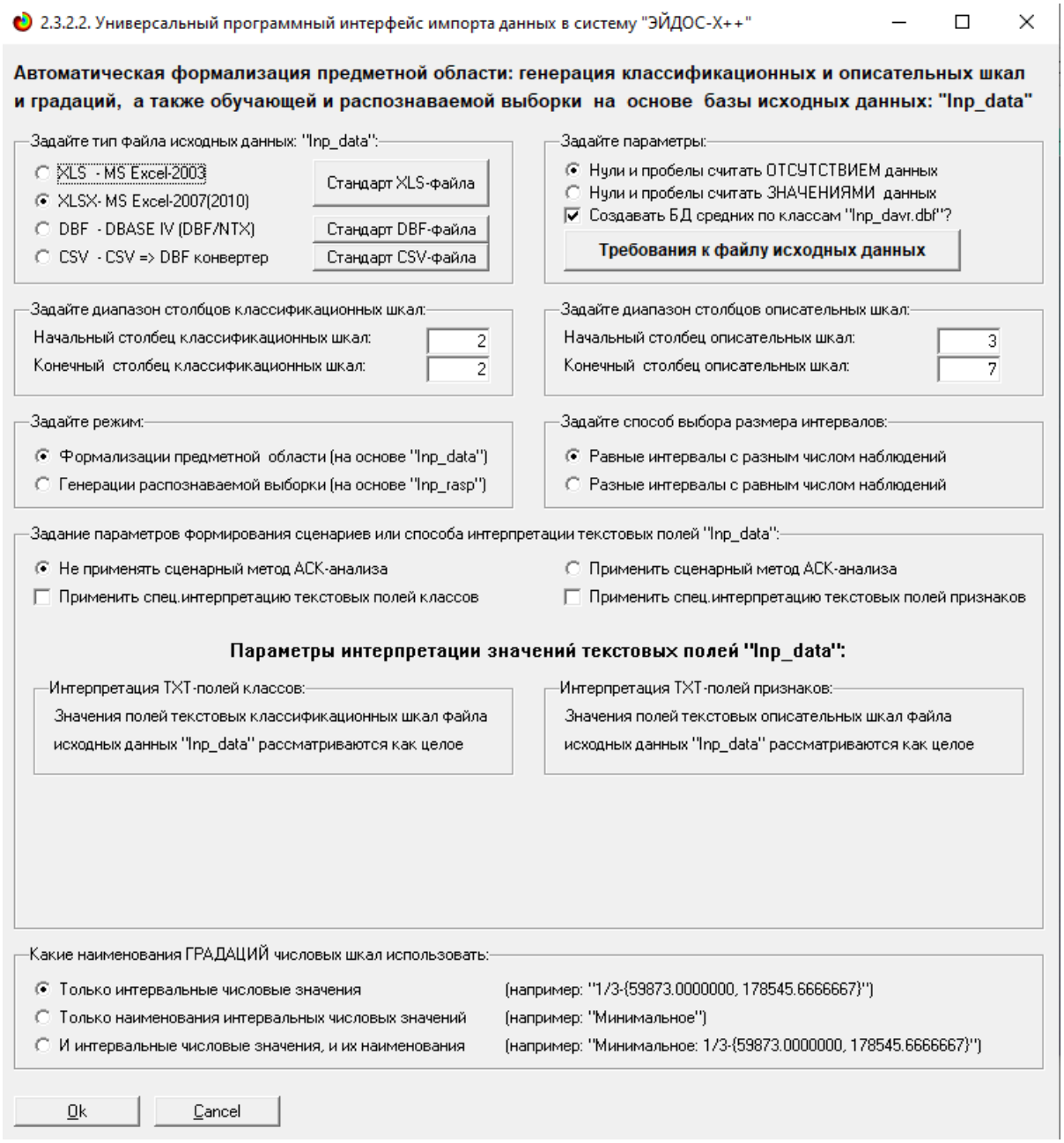


Рисунок 3 – Интерфейс импорта

Следует выделить следующие настройки:

– Тип файла – XLSX;

– Классификационная шкала – 2;

– Описательные шкалы – 3-10.

После этого приложение просит задать размерности модели системы, оставляем всё как есть (рисунок 4).

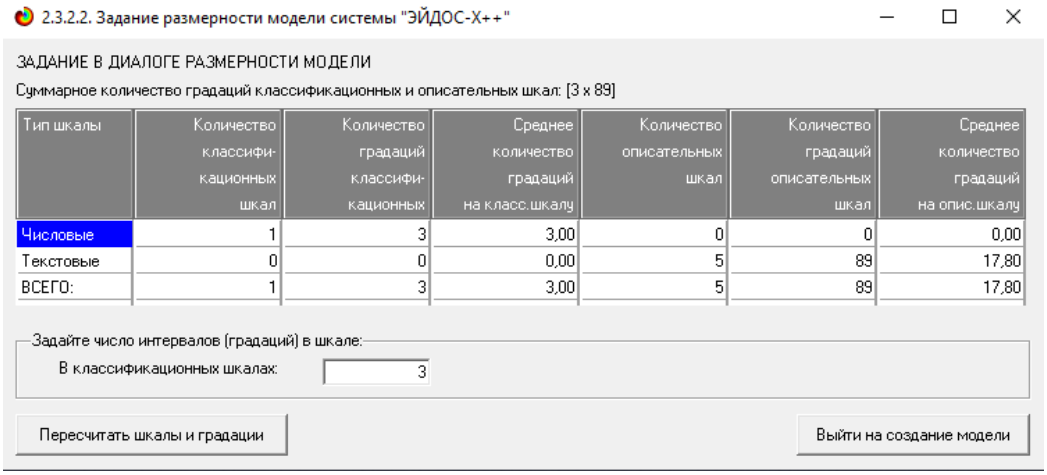


Рисунок 4 – Задание размерностей системы

Процесс импорта данных из внешнего файла в систему представлен на рисунке 5.

Рисунок 5 – Импорт данных

После загрузки данных система автоматически нашла классификационные шкалы, которые можно посмотреть в режиме 2.1 (рисунок 6) и описательные шкалы, которые можно посмотреть в режиме 2.2 (рисунок 7).

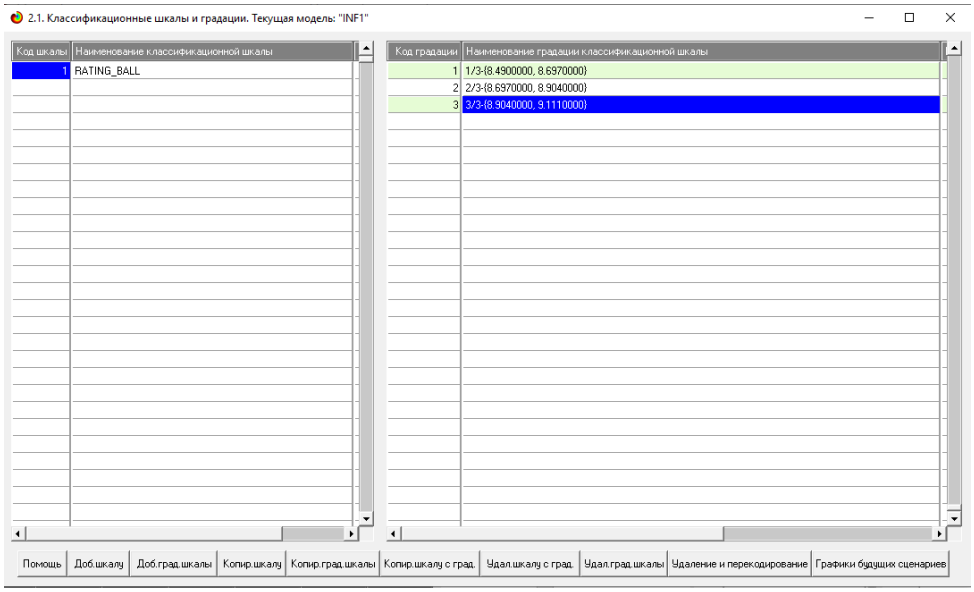


Рисунок 6 – Классификационные шкалы

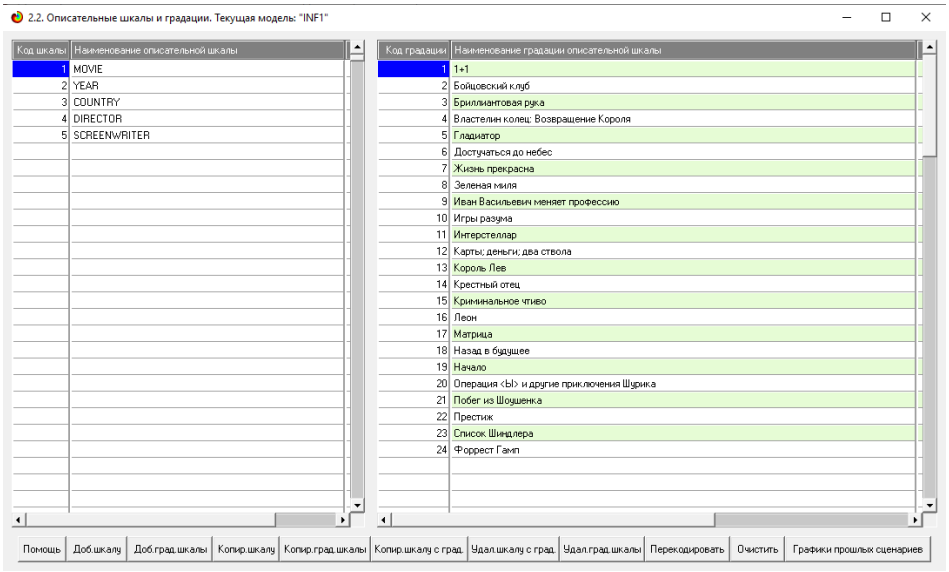


Рисунок 7 – Описательные шкалы

Так же существует возможность ручной корректировки выгруженных данных и добавление новых объектов, которая открывается с помощью режима 2.3.1 (рисунок 8).

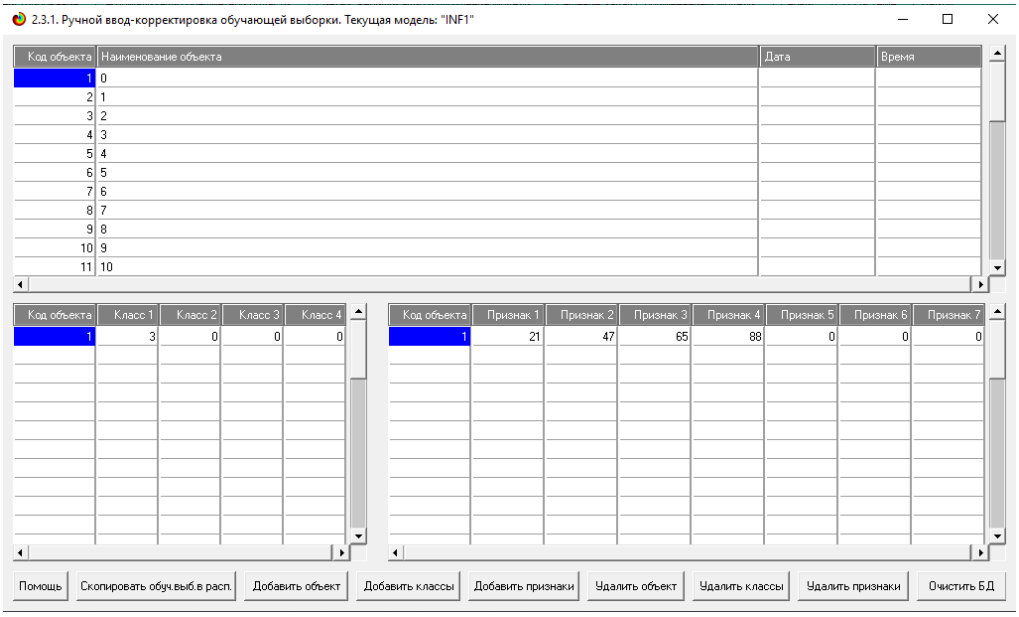


Рисунок 8 – Ручная корректировка

Выборка происходит по наименованию турнира «rating_ball», итоговую таблицу можно увидеть на рисунке 1.

	A	B	C	D	E	F	G	H
1	rating	rating_ball	movie	year	country	director	screenwriter	actors
2	0	9,111	Побег из Шоушенка	1994 Year	США	Фрэнк Дарабонт	Фрэнк Дарабонт; Стивен Кинг	Тим Роббинс; Морган Фриман; Боб Гантон; Уильям Сэдлер; Клэнси Браун
3	1	9,062	Зеленая миля	1999 Year	США	Фрэнк Дарабонт	Фрэнк Дарабонт; Стивен Кинг	Том Хэнкс; Дэвид Морс; Бонни Хант; Майкл Кларк Дункан; Джеймс Кромуэлл
4	2	8,913	Форрест Гамп	1994 Year	США	Роберт Земекис	Эрик Рот; Уинстон Грум	Том Хэнкс; Робин Райт; Салли Филд; Гэри Синиз; Майкелти Уильямсон
5	3	8,817	Список Шиндлера	1993 Year	США	Стивен Спилберг	Стивен Зеллиан; Томас Кенилли	Лиам Нисон; Бен Кингсли; Рэйф Файнс; Кэролайн Гудолл; Эмбет Дэвидц
6	4	8,807	1+1	2011 Year	Франция	Оливье Накаш; Эрик Толедано	Оливье Накаш; Филипп Поццо ди Борго; Эрик Толедано	Франсуа Клюзе; Омар Си; Анн Ле Ни; Одри Флеро; Жозефин де Мо
7	5	8,662	Начало	2010 Year	США	Кристофер Нолан	Кристофер Нолан	Леонардо ДиКаприо; Джозеф Гордон-Левитт; Эллен Пейдж; Том Харди; Кэн Ватанабэ
8	6	8,681	Леон	1994 Year	Франция	Люк Бессон	Люк Бессон	Жан Рено; Гари Олдман; Натали Портман; Дэнни Айелло; Питер Эппел
9	7	8,772	Король Лев	1994 Year	США	Роджер Аллерс; Роб Минкофф	Ирен Меччи; Джонатан Робертс; Линда Вулвертон; ...	Мэттью Бродерик; Джереми Айронс; Нэйтан Лейн; Эрни Сабелла; Джеймс Эрл Джонс
10	8	8,645	Бойцовский клуб	1999 Year	США	Дэвид Финчер	Джим Улс; Чак Паланик	Эдвард Нортон; Брэд Питт; Хелена Бонем Картер; Мит Лоаф; Зэк Гренье
11	9	8,782	Иван Васильевич меняет профессию	1973 Year	СССР	Леонид Гайдай	Владлен Бахнов; Леонид Гайдай; Михаил Булгаков	Александр Демьяненко; Юрий Яковлев; Леонид Куравлёв; Наталья Крачковская; Савелий Крамаров
12	10	8,624	Жизнь прекрасна	1997 Year	Италия	Роберто Бениньи	Винченцо Черами; Роберто Бениньи	Роберто Бениньи; Николетта Браски; Джорджио Кантарини; Джустино Дурано; Серджио Бини Бустрик
13	11	8,629	Достучаться до небес	1997 Year	Германия	Томас Ян	Томас Ян; Тиль Швайгер	Тиль Швайгер; Ян Йозеф Лиферс; Тьерри Ван Вервеке; Мориц Бляйбтрой; Хуб Стапель
						Фрэнсис Форд	Марио Пьюзо; Фрэнсис Форд	Марлон Брандо; Аль Пачино; Джеймс Каан;

Рисунок 1 – Фрагмент обучающей выборки

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).

В данной курсовой работе используются данные с электронного ресурса Kaggle на тему «Kinopoisk-TOP250 Russian»:

https://www.kaggle.com/alexandertesemnikov/kinopoisktop250russiandataset/code

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.

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).

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

Суть математической модели АСК-анализа состоит в следующем.

Непосредственно на основе эмпирических данных (см. Help режима 2.3.2.2) рассчитывается матрица абсолютных частот (таблица 1).

Таблица 1 – Матрица абсолютных частот (статистическая модель ABS)

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

На ее основе рассчитываются матрицы условных и безусловных процентных распределений (таблица 2).

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

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

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

На практике часто встречается существенная несбалансированность данных, под которой понимается сильно отличающееся количество объектов обучающейся выборки, относящихся к различным классам. Поэтому решать задачу на основе непосредственно матрицы абсолютных частот (таблица 1) было бы очень неразумно и переход от абсолютных частот к условным и безусловным относительным частотам (частостям) является весьма обоснованным и логичным.

Таблица 2 – 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, а матрицы условных и безусловных процентных распределений, и матрицы системно-когнитивных моделей (СК-модели, таблица 4), 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 [1].

В системе «Эйдос» это осуществляется всегда при решении любых задач.

Затем на основе таблицы 2 с использованием частных критериев, знаний, приведенных таблице 3, рассчитываются матрицы системно-когнитивных моделей (таблица 4).

Таблица 3– 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; $\overline{N}_{ij}$ - 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 1)	$N_i=\sum_{j=1}^{W}N_{ij};\ N_j=\sum_{i=1}^{M}$ $N_{ij}-$ *фактическая* $\overline{N}_{ij}=\frac{N_iN_j}{N}-$ *теор*	
PRC1, матрица условных P_{ij} и безусловных P_i процентных распределений, в качестве N_j используется суммарное количество признаков по классу	---	$P_{ij}=\frac{N_{ij}}{N_j};\ P_i$
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: *Nj* – суммарное количество признаков по *j*-му классу. Вероятность того, что если у объекта j-го класса обнаружен признак, то это i-й признак	$I_{ij}=\Psi\ I_{ij}=\Psi\times Log$	
INF2, частный критерий: количество знаний по А.Харкевичу, 2-й вариант расчета вероятностей: *Nj* – 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	---	$I_{ij}=N_{ij}-\overline{N}_{ij}$
INF4, частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: *Nj* – суммарное количество признаков по *j*-му классу	$I_{ij}=\frac{P_{ij}}{P_i}\ I_{ij}=\frac{N_{ij}}{\overline{N}_{ij}}-$	
INF5, частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: *Nj* – суммарное количество объектов по *j*-му классу		
INF6, частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: *Nj* – суммарное количество признаков по *j*-му классу	$I_{ij}=1\ I_{ij}=\frac{N_{ij}}{N_j}-\frac{N}{N}$	
INF7, частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: *Nj* – суммарное количество объектов по *j*-му классу		

Обозначения к таблице 3:

i – значение прошлого параметра;
j - значение будущего параметра;
N_{ij} – количество встреч j-го значения будущего параметра при i-м значении прошлого параметра;
M – суммарное число значений всех прошлых параметров;
W - суммарное число значений всех будущих параметров.
N_i – количество встреч i-м значения прошлого параметра по всей выборке;
N_j – количество встреч j-го значения будущего параметра по всей выборке;
N – количество встреч j-го значения будущего параметра при i-м значении прошлого параметра по всей выборке.
I_{ij} – частный критерий знаний: количество знаний в факте наблюдения i-го значения прошлого параметра о том, что объект перейдет в состояние, соответствующее j-му значению будущего параметра;
Ψ – нормировочный коэффициент (Е.В.Луценко, 2002), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли;
P_i – безусловная относительная частота встречи i-го значения прошлого параметра в обучающей выборке;

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

В таблице 3 приведены формулы:

– для сравнения фактических и теоретических абсолютных частот;

– для сравнения условных и безусловных относительных частот («вероятностей»).

И это сравнение в таблице 3 осуществляется двумя возможными способами: путем вычитания и путем деления.

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

Таким образом, мы видим, что все частные критерии знаний тесно взаимосвязаны друг с другом. Особенно интересна связь знаменитого критерия хи-квадрат 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 частой и вероятностью (погрешность) составляет около 5%, при 1250 наблюдениях – около 2.5%, при 10000 наблюдениях – 1%.

Таблица 4 – Матрица системно-когнитивной модели

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

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 [1].

На основе системно-когнитивных моделей, представленных в таблице 9 (отличаются частыми критериями, приведенными в таблице 8), 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 [10-64].

Отметим, что как значимость значения фактора, степень детерминированности класса и ценность или качество модели в АСК-анализе

рассматривается вариабельность значений частных критериев этого значения фактора, класса или модели в целом (таблица 4).

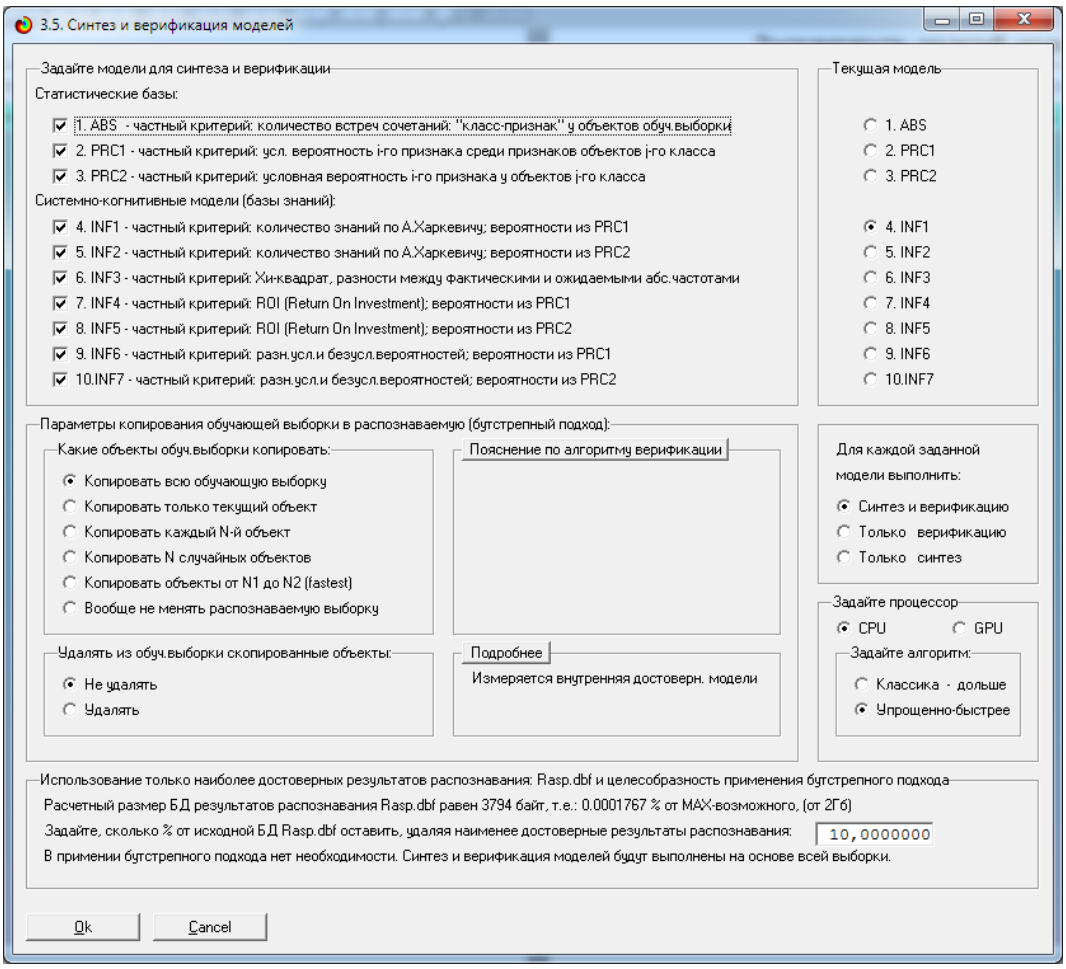
Численно эта вариабельность может измеряться разными способами, например, средним отклонением модулей частных критериев от среднего, дисперсией или среднеквадратичным отклонением, или его квадратом. В системе «Эйдос» принят последний вариант, т.к. эта величина совпадает с мощностью сигнала, в частности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

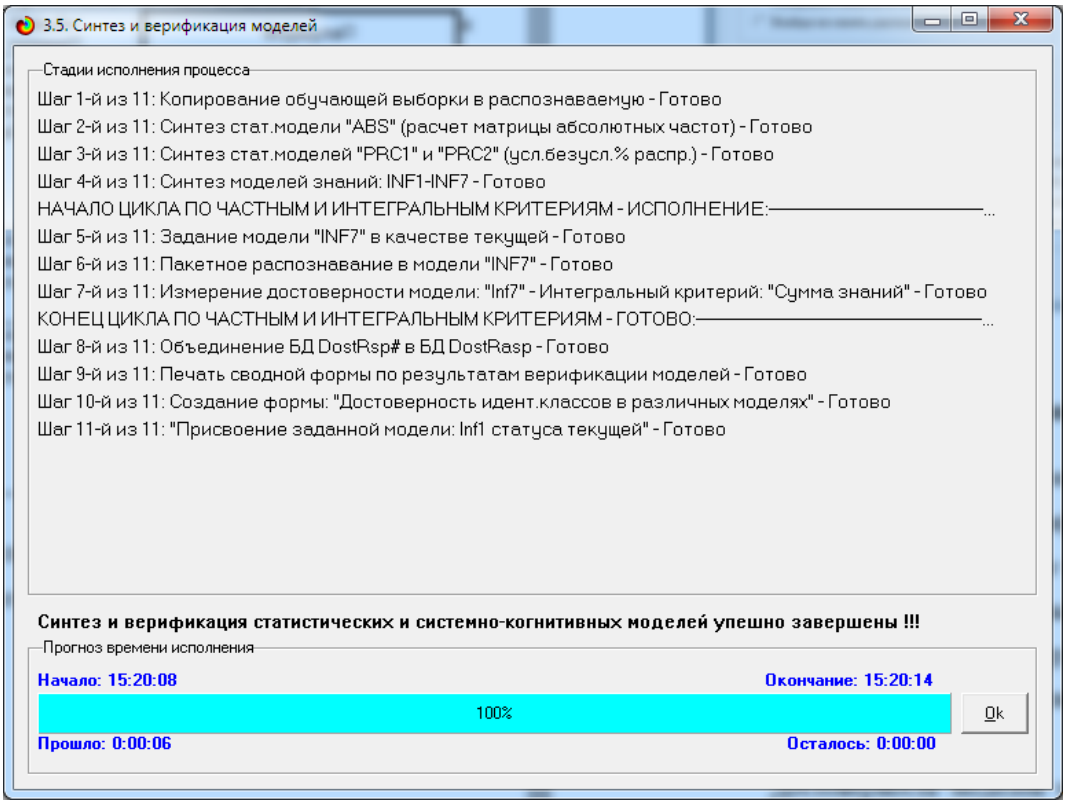
Поэтому есть все основания уточнить традиционную терминологию АСК-анализа (таблица 5):

Таблица 5 – Уточнение терминологии АСК-анализа

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

Синтез моделей осуществляется врежиме5 3.5 системы «Эйдос» при параметрах, приведенных на скриншоте экранной формы управления данным режимом, приведенной ниже:


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
2. PRC1 - частный критерий: усл. вероятность i-го признака среди признаков объектов j-го класса
3. PRC2 - частный критерий: условная вероятность i-го признака у объектов j-го класса
Системно-когнитивные модели (базы знаний):
4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1
5. INF2 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC2
6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами
7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1
8. INF5 - частный критерий: ROI (Return On Investment); вероятности из PRC2
9. INF6 - частный критерий: разн.усл.и 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 (бутстрепный подход):
Какие объекты обуч.выборки копировать:
Копировать всю обучающую выборку
Копировать только текущий объект
Копировать каждый N-й объект
Копировать N случайных объектов
Копировать объекты от N1 до N2 (fastest)
Вообще не менять распознаваемую выборку
Пояснение по алгоритму верификации
Для каждой заданной модели выполнить:
Синтез и верификацию
Только верификацию
Только синтез
Задайте процессор
CPU
GPU
Задайте алгоритм:
Классика - дольше
Упрощенно-быстрее
Удалять из обуч.выборки скопированные объекты:
Не удалять
Удалять
Подробнее
Измеряется внутренняя достоверн. модели
Использование только наиболее достоверных результатов распознавания: Rasp.dbf и целесообразность применения бутстрепного подхода
Расчетный размер БД результатов распознавания Rasp.dbf равен 3794 байт, т.е.: 0.0001767 % от MAX-возможного, (от 2Гб)
Задайте, сколько % от исходной БД Rasp.dbf оставить, удаляя наименее достоверные результаты распознавания:
10,0000000
В примении бутстрепного подхода нет необходимости. Синтез и верификация моделей будут выполнены на основе всей выборки.
Ok
Cancel


3.5. Синтез и верификация моделей
Стадии исполнения процесса
Шаг 1-й из 11: Копирование обучающей выборки в распознаваемую - Готово
Шаг 2-й из 11: Синтез стат.модели "ABS" (расчет матрицы абсолютных частот) - Готово
Шаг 3-й из 11: Синтез стат.моделей "PRC1" и "PRC2" (усл.безусл.% распр.) - Готово
Шаг 4-й из 11: Синтез моделей знаний: INF1-INF7 - Готово
НАЧАЛО ЦИКЛА ПО ЧАСТНЫМ И ИНТЕГРАЛЬНЫМ КРИТЕРИЯМ - ИСПОЛНЕНИЕ:
Шаг 5-й из 11: Задание модели "INF7" в качестве текущей - Готово
Шаг 6-й из 11: Пакетное распознавание в модели "INF7" - Готово
Шаг 7-й из 11: Измерение достоверности модели: "Inf7" - Интегральный критерий: "Сумма знаний" - Готово
КОНЕЦ ЦИКЛА ПО ЧАСТНЫМ И ИНТЕГРАЛЬНЫМ КРИТЕРИЯМ - ГОТОВО:
Шаг 8-й из 11: Объединение БД DostRsp# в БД DostRasp - Готово
Шаг 9-й из 11: Печать сводной формы по результатам верификации моделей - Готово
Шаг 10-й из 11: Создание формы: "Достоверность идент.классов в различных моделях" - Готово
Шаг 11-й из 11: "Присвоение заданной модели: Inf1 статуса текущей" - Готово
Синтез и верификация статистических и системно-когнитивных моделей успешно завершены !!!
Прогноз времени исполнения
Начало: 15:20:08
Окончание: 15:20:14
100%
Ok
Прошло: 0:00:06
Осталось: 0:00:00

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 [8].

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.

С результатами верификации моделей, отличающихся частными критериями, можно ознакомиться в режиме 3.4 системы «Эйдос», они представлены на рисунке 9.

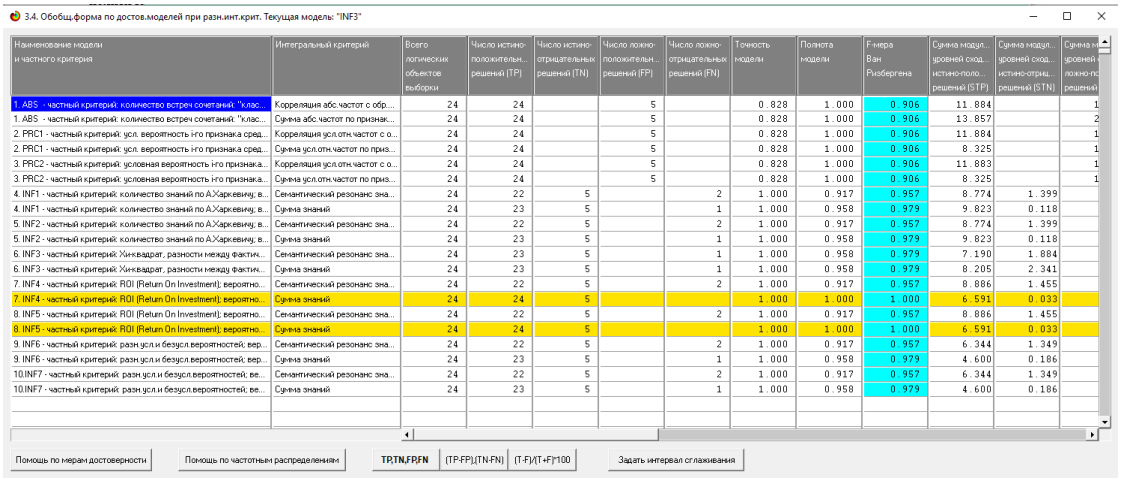


Рисунок 9 – Оценки достоверности моделей

Для оценки достоверности моделей в АСК-анализе и систему Aidos-X используется F-мера Ван Ризбергена и L-мера, представляющая собой ее

нечеткое мультиклассовое обобщение, предложенное профессором Луценко (рисунок 10). Наиболее достоверными оказались модели INF3, INF4 и INF5, однако рассматривать далее будем только модель PRC2. Точно данной модели по F-мере Ван Ризбергена составляет 0.293, а по L1-мере профессора Луценко – 0.802. L1-мера, предложенная профессором Луценко является более достоверной, в сравнении с F-мерой Ван Ризбергена. Таким образом, уровень достоверности прогнозирования с применением модели выше экспертных оценок, достоверность которых считается равной примерно 100%.

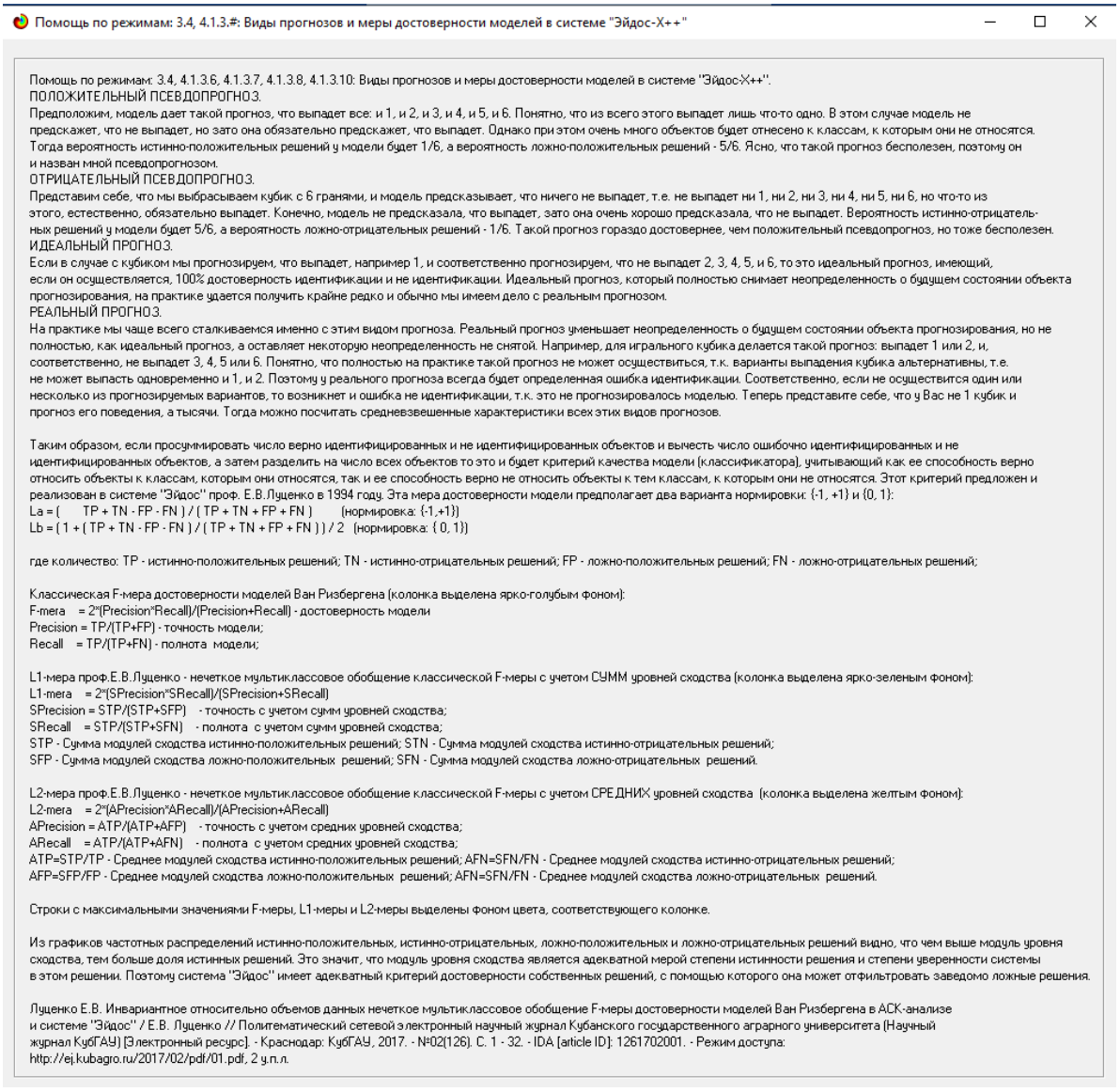


Рисунок 10 – Виды прогнозов и меры достоверностей в системе «Эйдос»

Также необходимо обратить внимание на то, что статистические модели, как правило, дают более низкую средневзвешенную достоверность

идентификации и не идентификации, чем модели знаний, и практически никогда – более высокую. Этим и оправдано применение моделей знаний из 23 интеллектуальных технологий. На рисунке 11 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и не идентифицированных ситуаций в наиболее достоверной модели INF3.

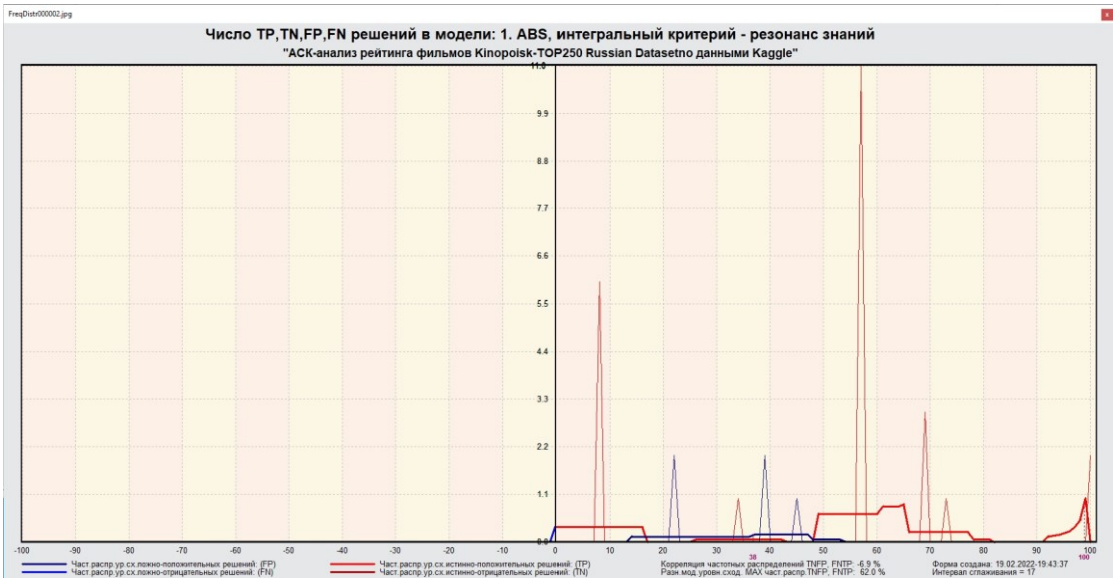


Рисунок 11 – Частотные распределения верно и ошибочно идентифицированных и не идентифицированных состояний объекта моделирования в зависимости от уровня сходства в модели INF3

Левое распределение, большее по амплитуде включает истинно-отрицательные и ложноположительные решения, а правое, меньшее по амплитуде, включает ложные отрицательные и истинно-положительные решения. Сдвиг этих распределений относительно друг друга и другие различия между ними и позволяют решать задачу идентификации видов матчей и другие задачи.

Видно, что для отрицательных решений количество истинных решений всегда превосходит количество ложных, причем при уровнях различия 60% и больше 70% ложные отрицательные решения практически отсутствуют.

Для положительных решений картина примерно схожа, за исключением того, что от 15% до примерно 55% количество ложных решений больше числа истинных, но далее идет на спад и ложные решения перестают встречаться после 70%.

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 осуществляется в режиме 5.6 системы «Эйдос» и проходит быстро. Это необходимо лишь для решения задачи идентификации и прогнозирования (в режиме 4.1.2), которая требует наибольшие вычислительные ресурсы и поэтому решается только для модели, заданной текущей. Все остальные расчеты проводятся в системе «Эйдос» сразу во всех моделях.

В соответствии с технологией АСК-анализа необходимо задать текущую модель в режиме 5.6, в ее качестве выберем наиболее достоверную модель INF3, что показано на рисунке 12.

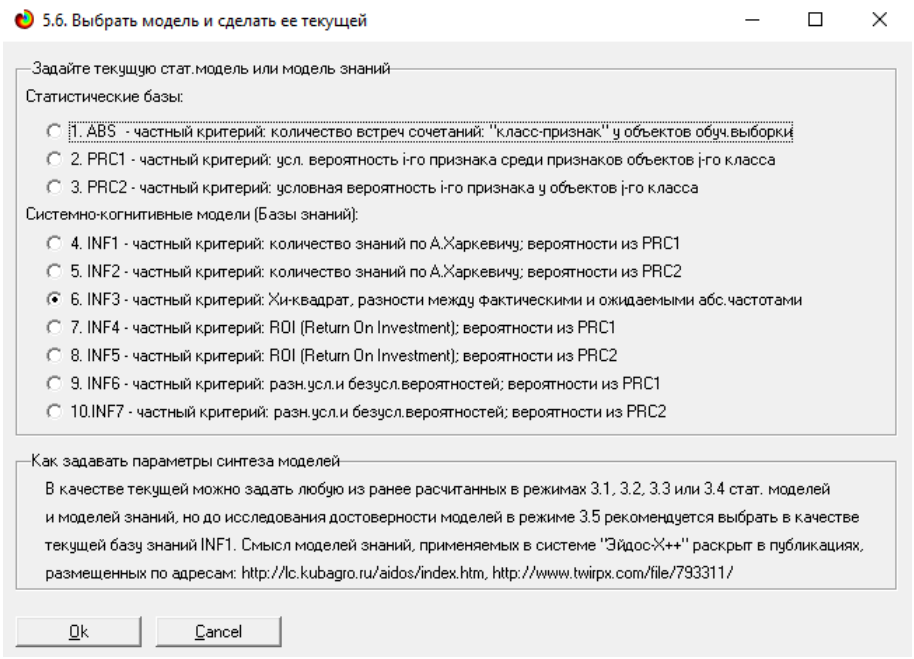


Рисунок 12 – Выбор текущей модели

После этого необходимо провести пакетное распознавание текущей модели в режиме 4.1.2 (рисунок 13)

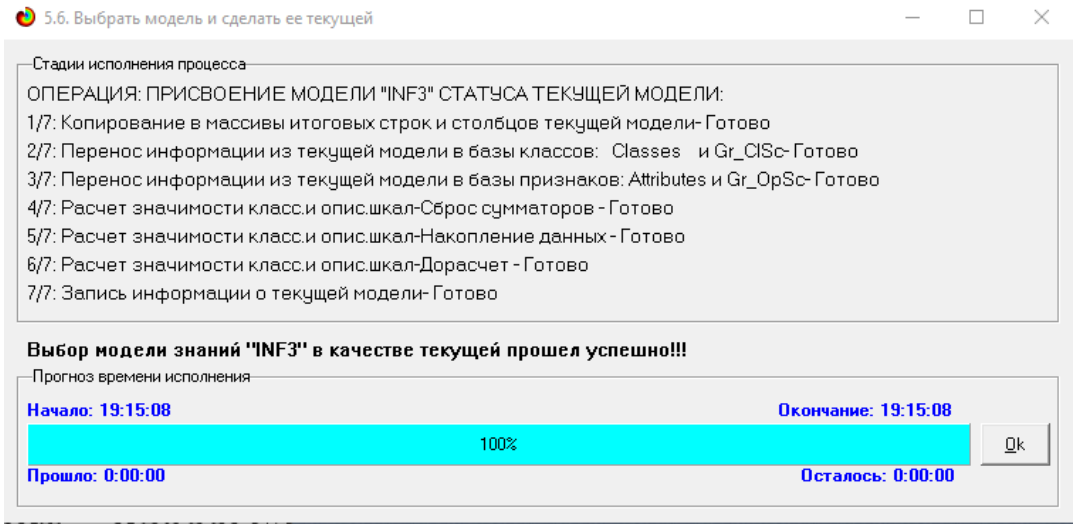
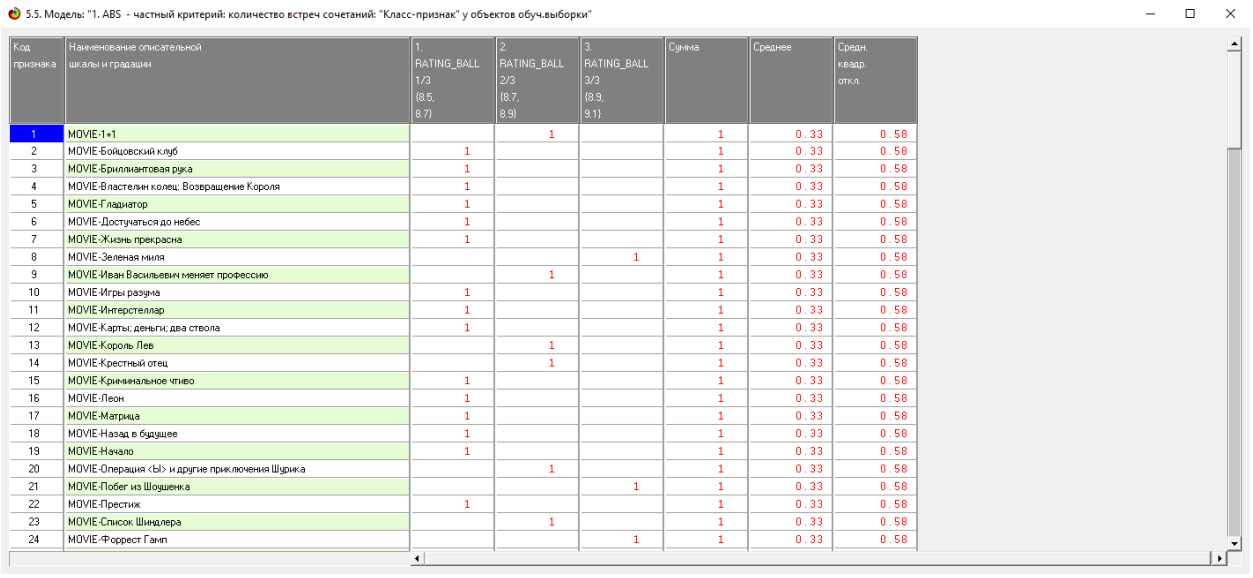


Рисунок 13 – Пакетное распознавание текущей модели

Решение задачи идентификации может быть рассмотрено на нескольких моделях, в которых рассчитано количество информации по А. Харкевичу, которое было сгенерировано по принадлежности идентифицируемого объекта к каждому из классов, учитывая признаки объекта.

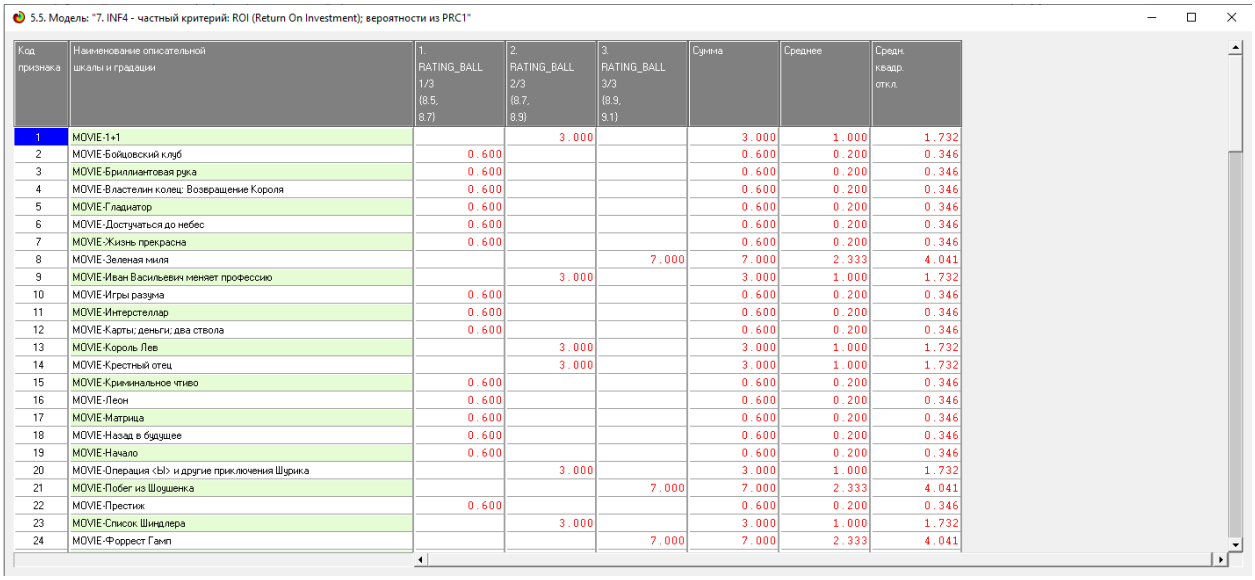
Таким образом, частные критерии представляют собой формулы для преобразования матрицы абсолютных частот, которая представлена на рисунке 14 и открытой с помощью режима 5.5.1, в матрицы условных и безусловных процентных распределений, и матрицы знаний.



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

Код признака	Наименование описательной шкалы и градации	1. RATING_BALL 1/3 (8.5, 8.7)	2. RATING_BALL 2/3 (8.7, 8.9)	3. RATING_BALL 3/3 (8.9, 9.1)	Сумма	Среднее	Средн. квадр. откл.
1	MOVIE-1+1		1		1	0.33	0.58
2	MOVIE-Бойцовский клуб	1			1	0.33	0.58
3	MOVIE-Бриллиантовая рука	1			1	0.33	0.58
4	MOVIE-Властелин колец: Возвращение Короля	1			1	0.33	0.58
5	MOVIE-Гладиатор	1			1	0.33	0.58
6	MOVIE-Достучаться до небес	1			1	0.33	0.58
7	MOVIE-Жизнь прекрасна	1			1	0.33	0.58
8	MOVIE-Зеленая миля			1	1	0.33	0.58
9	MOVIE-Иван Васильевич меняет профессию		1		1	0.33	0.58
10	MOVIE-Игры разума	1			1	0.33	0.58
11	MOVIE-Интерстеллар	1			1	0.33	0.58
12	MOVIE-Карты, деньги, два ствола	1			1	0.33	0.58
13	MOVIE-Король Лев		1		1	0.33	0.58
14	MOVIE-Крестный отец		1		1	0.33	0.58
15	MOVIE-Криминальное чтиво	1			1	0.33	0.58
16	MOVIE-Леон	1			1	0.33	0.58
17	MOVIE-Матрица	1			1	0.33	0.58
18	MOVIE-Назад в будущее	1			1	0.33	0.58
19	MOVIE-Начало	1			1	0.33	0.58
20	MOVIE-Операция «Ы» и другие приключения Шурика		1		1	0.33	0.58
21	MOVIE-Побег из Шоушенка			1	1	0.33	0.58
22	MOVIE-Престиж	1			1	0.33	0.58
23	MOVIE-Список Шиндлера		1		1	0.33	0.58
24	MOVIE-Форрест Гамп			1	1	0.33	0.58

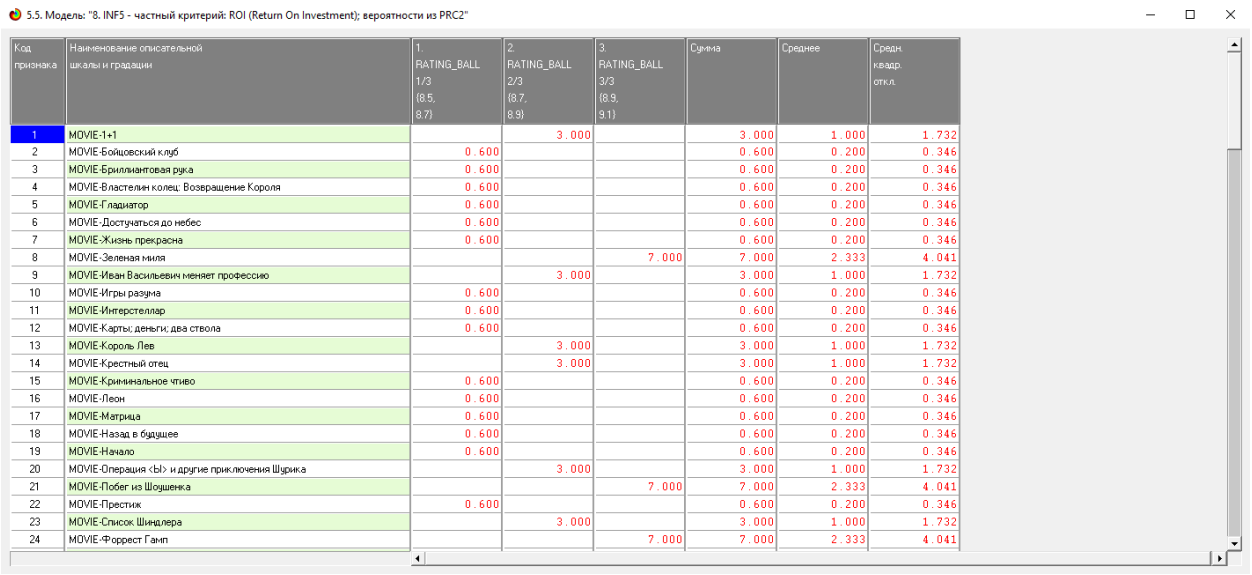
Рисунок 14 – Матрица абсолютных частот



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

Код признака	Наименование описательной шкалы и градации	1. RATING_BALL 1/3 (8.5, 8.7)	2. RATING_BALL 2/3 (8.7, 8.9)	3. RATING_BALL 3/3 (8.9, 9.1)	Сумма	Среднее	Средн. квадр. откл.
1	MOVIE-1+1		3.000		3.000	1.000	1.732
2	MOVIE-Бойцовский клуб	0.600			0.600	0.200	0.346
3	MOVIE-Бриллиантовая рука	0.600			0.600	0.200	0.346
4	MOVIE-Властелин колец: Возвращение Короля	0.600			0.600	0.200	0.346
5	MOVIE-Гладиатор	0.600			0.600	0.200	0.346
6	MOVIE-Достучаться до небес	0.600			0.600	0.200	0.346
7	MOVIE-Жизнь прекрасна	0.600			0.600	0.200	0.346
8	MOVIE-Зеленая миля			7.000	7.000	2.333	4.041
9	MOVIE-Иван Васильевич меняет профессию		3.000		3.000	1.000	1.732
10	MOVIE-Игры разума	0.600			0.600	0.200	0.346
11	MOVIE-Интерстеллар	0.600			0.600	0.200	0.346
12	MOVIE-Карты, деньги, два ствола	0.600			0.600	0.200	0.346
13	MOVIE-Король Лев		3.000		3.000	1.000	1.732
14	MOVIE-Крестный отец		3.000		3.000	1.000	1.732
15	MOVIE-Криминальное чтиво	0.600			0.600	0.200	0.346
16	MOVIE-Леон	0.600			0.600	0.200	0.346
17	MOVIE-Матрица	0.600			0.600	0.200	0.346
18	MOVIE-Назад в будущее	0.600			0.600	0.200	0.346
19	MOVIE-Начало	0.600			0.600	0.200	0.346
20	MOVIE-Операция «Ы» и другие приключения Шурика		3.000		3.000	1.000	1.732
21	MOVIE-Побег из Шоушенка			7.000	7.000	2.333	4.041
22	MOVIE-Престиж	0.600			0.600	0.200	0.346
23	MOVIE-Список Шиндлера		3.000		3.000	1.000	1.732
24	MOVIE-Форрест Гамп			7.000	7.000	2.333	4.041

Рисунок 15 – Фрагмент модели PRC1



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

Код признака	Наименование описательной шкалы и градации	1. RATING_BALL 1/3 (8.5, 8.7)	2. RATING_BALL 2/3 (8.7, 8.9)	3. RATING_BALL 3/3 (8.9, 9.1)	Сумма	Среднее	Средн. квадр. откл.
1	MOVIE-1+1		3.000		3.000	1.000	1.732
2	MOVIE-Бойцовский клуб	0.600			0.600	0.200	0.346
3	MOVIE-Бриллиантовая рука	0.600			0.600	0.200	0.346
4	MOVIE-Властелин колец: Возвращение Короля	0.600			0.600	0.200	0.346
5	MOVIE-Гладиатор	0.600			0.600	0.200	0.346
6	MOVIE-Достучаться до небес	0.600			0.600	0.200	0.346
7	MOVIE-Жизнь прекрасна	0.600			0.600	0.200	0.346
8	MOVIE-Зеленая миля			7.000	7.000	2.333	4.041
9	MOVIE-Иван Васильевич меняет профессию		3.000		3.000	1.000	1.732
10	MOVIE-Игры разума	0.600			0.600	0.200	0.346
11	MOVIE-Интерстеллар	0.600			0.600	0.200	0.346
12	MOVIE-Карты, деньги, два ствола	0.600			0.600	0.200	0.346
13	MOVIE-Король Лев		3.000		3.000	1.000	1.732
14	MOVIE-Крестный отец		3.000		3.000	1.000	1.732
15	MOVIE-Криминальное чтиво	0.600			0.600	0.200	0.346
16	MOVIE-Леон	0.600			0.600	0.200	0.346
17	MOVIE-Матрица	0.600			0.600	0.200	0.346
18	MOVIE-Назад в будущее	0.600			0.600	0.200	0.346
19	MOVIE-Начало	0.600			0.600	0.200	0.346
20	MOVIE-Операция «Ы» и другие приключения Шурика		3.000		3.000	1.000	1.732
21	MOVIE-Побег из Шоушенка			7.000	7.000	2.333	4.041
22	MOVIE-Престиж	0.600			0.600	0.200	0.346
23	MOVIE-Список Шиндлера		3.000		3.000	1.000	1.732
24	MOVIE-Форрест Гамп			7.000	7.000	2.333	4.041

Рисунок 16 – Фрагмент модели INF5

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.

Задача решается в модели, заданной в качестве текущей, т.к. является весьма трудоемкой в вычислительном отношении. Правда с 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 корректны[8] в неортонормированных пространствах, которые всегда и встречаются на практике, и являются фильтрами подавления шума.

Режим 4.1.3 системы «Эйдос» обеспечивает отображение результатов идентификации и прогнозирования в различных формах:

– Подробно наглядно: "Объект – классы";

– Подробно наглядно: "Класс – объекты";

– Итоги наглядно: "Объект – классы";

– Итоги наглядно: "Класс – объекты";

– Подробно сжато: "Объект – классы";

– Обобщенная форма по достоверности моделей при разных интегральных критериях;

– Обобщенный статистический анализ результатов идентификации по моделям и интегральным критериям;

– Статистический анализ результатов идентификации по классам, моделям и интегральным критериям;

– Распознавание уровня сходства при разных моделях и интегральных критериях;

–Достоверность идентификации классов при разных моделях и интегральных критериях.

Рассмотрим подробнее режимы работы 4.1.3.1 и 4.1.3.2, результаты которых представлены на рисунке 17 и 18.

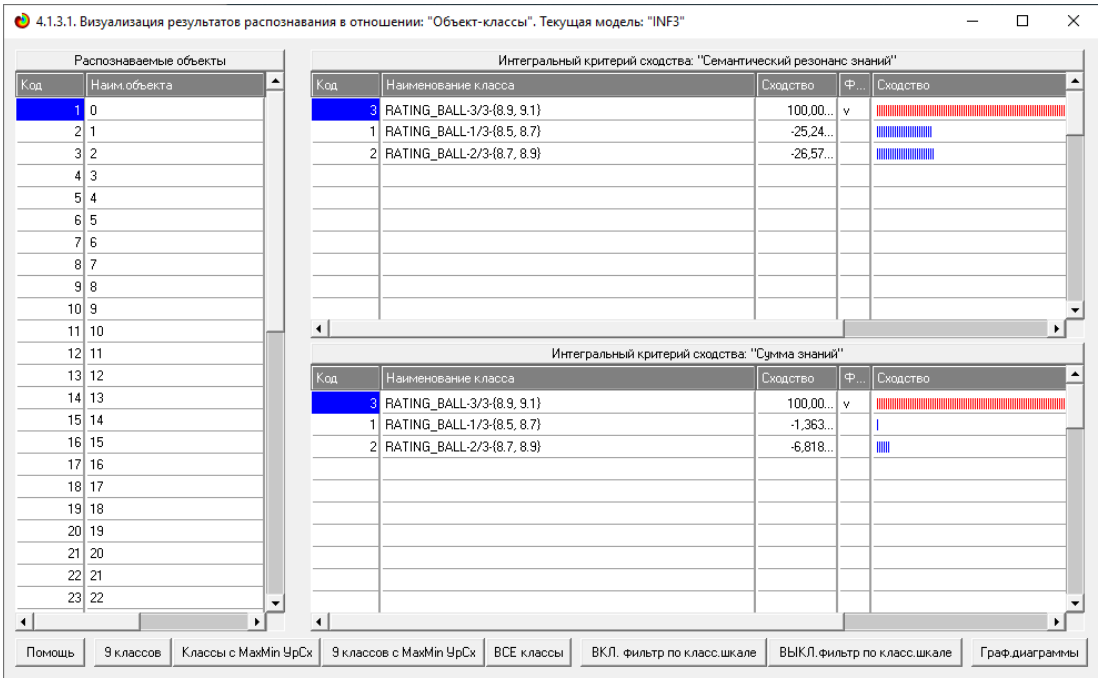


Рисунок 17 – Результат режима работы 4.1.3.1, идентификация объектов в модели INF3

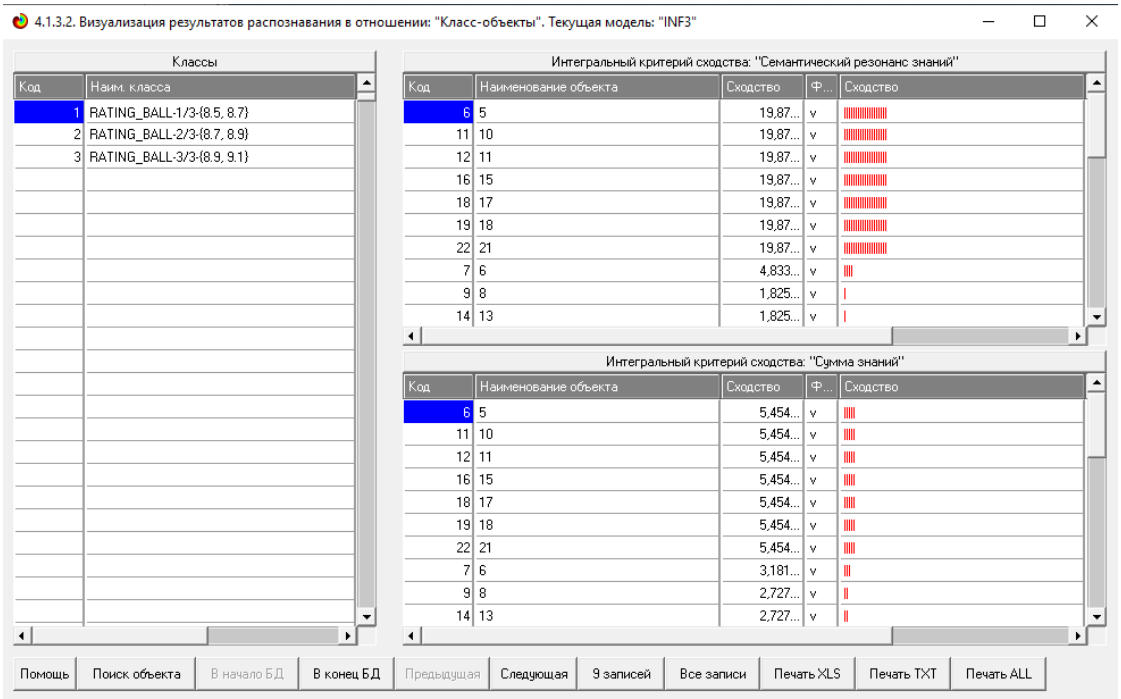


Рисунок 18 – Результат режима работы 4.1.3.2, идентификация классов в модели INF3

Из рисунков выше видно, что результаты идентификации являются отличными.

Для синтеза и верификации моделей создан режим 3.5, после его запуска задается модель, которая помечается текущей (рисунок 19).

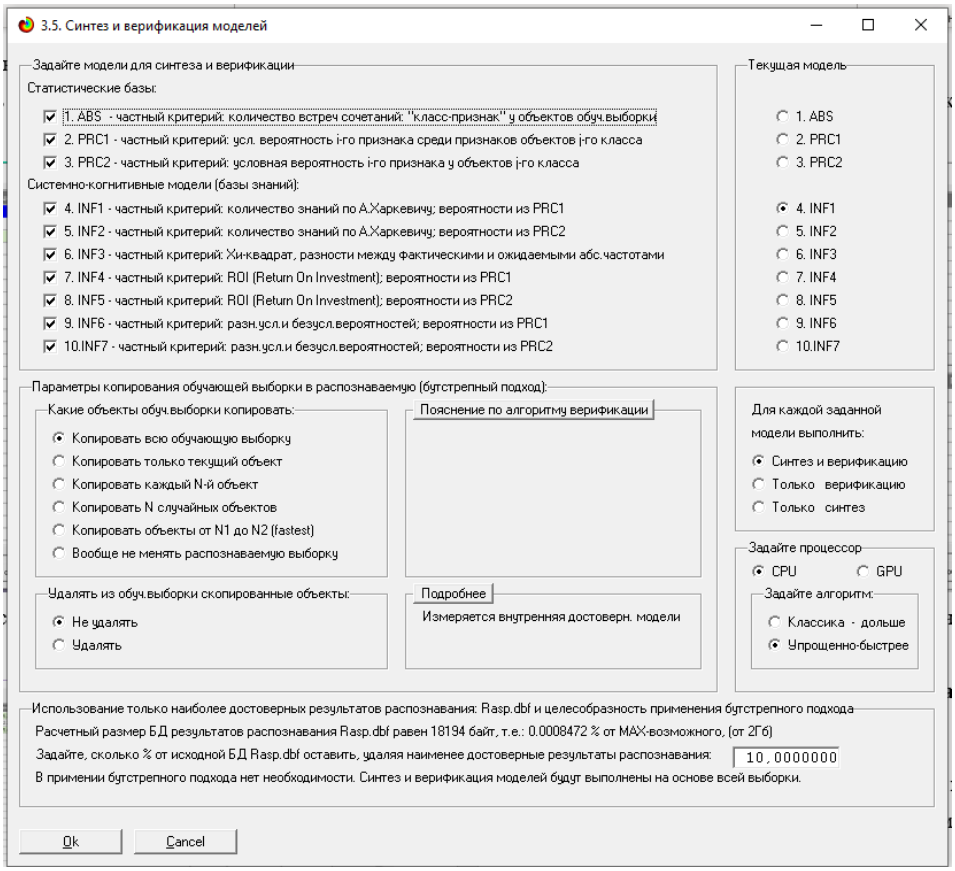


Рисунок 19 – Режим синтеза моделей

Данный режим содержит множество различных методов верификации, но мы используем параметры, которые система предлагает по умолчанию. Стадия процесса исполнения режима 3.5 и прогноз времени исполнения показаны на рисунке 20.

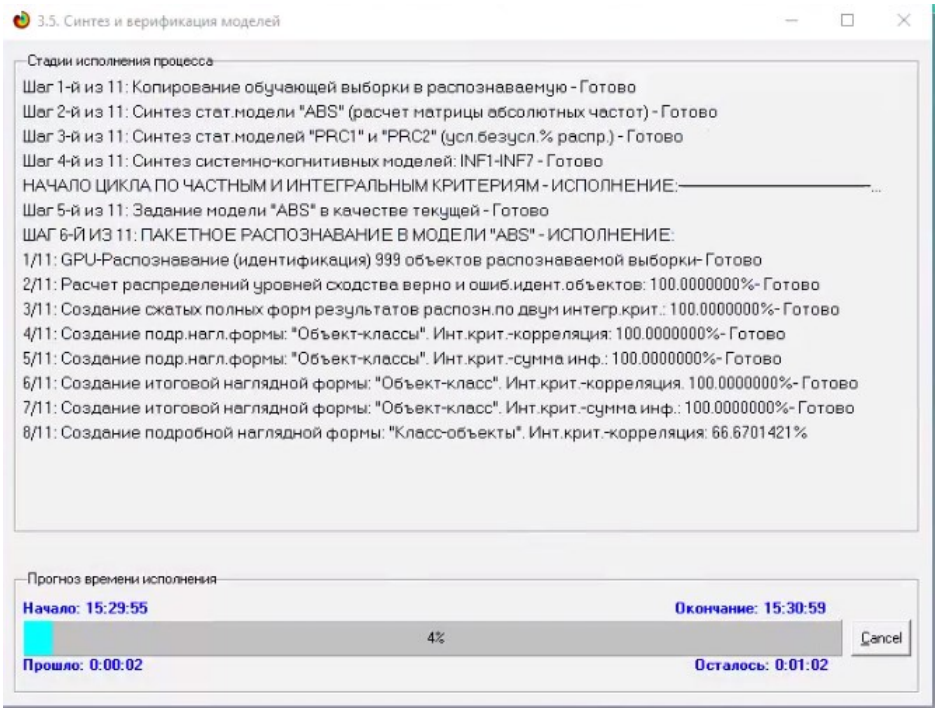


Рисунок 20 – Исполнение режима синтеза моделей

Следует заметить, что синтез и верификация всех моделей занял 10 секунд. После данного этапа можно приступить к выбору наиболее достоверной модели.

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.:

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

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

3.6.2. Интегральный критерий «Семантический резонанс знаний»

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

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

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

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.

$\vec{I}_{\bar{y}} = \{I_{\bar{y}}\}$ – 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.:

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

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

Приведенное выражение для интегрального критерия «Семантический резонанс знаний» получается непосредственно из выражения для критерия «Сумма знаний» после замены координат перемножаемых векторов их стандартизированными значениями: $I_{\bar{y}} \to \frac{I_{\bar{y}} - \bar{I}_j}{\sigma_j}$, $L_i \to \frac{L_i - \bar{L}}{\sigma_l}$. Поэтому по своей сути он также является скалярным произведением двух стандартизированных (единичных) векторов класса и объекта. Существуют и много других способов нормирования, например, путем применяя сплайнов, в частности линейной интерполяции: $I_{\bar{y}} \to \frac{I_{\bar{y}} - I_j^{\min}}{I_j^{\max} - I_j^{\min}}$, $L_i \to \frac{L_i - L^{\min}}{L^{\max} - L^{\min}}$. Это позволяет

предложить другие виды интегральных критериев. Но они в настоящее время 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: Луценко Е. В. Сценарный и спектральный автоматизированный системно-когнитивный анализ: научная монография / Е. В. Луценко. – Краснодар: КубГАУ, 2021. – 288 с., ISBN 978-5-907474-67-3,
DOI: 10.13140/RG.2.2.22981.37608, https://www.researchgate.net/publication/353555996

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») [9].

3.7.2. Развитый алгоритм принятия решений в АСК-анализе

Однако SWOT-анализ (режим 4.4.8 системы «Эйдос») имеет свои ограничения: может быть задано только одно будущее целевое состояние, некоторые рекомендуемые факторы может не быть технологической и финансовой возможности использовать.

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.

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

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

Инвертированные SWOT-диаграмм 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».

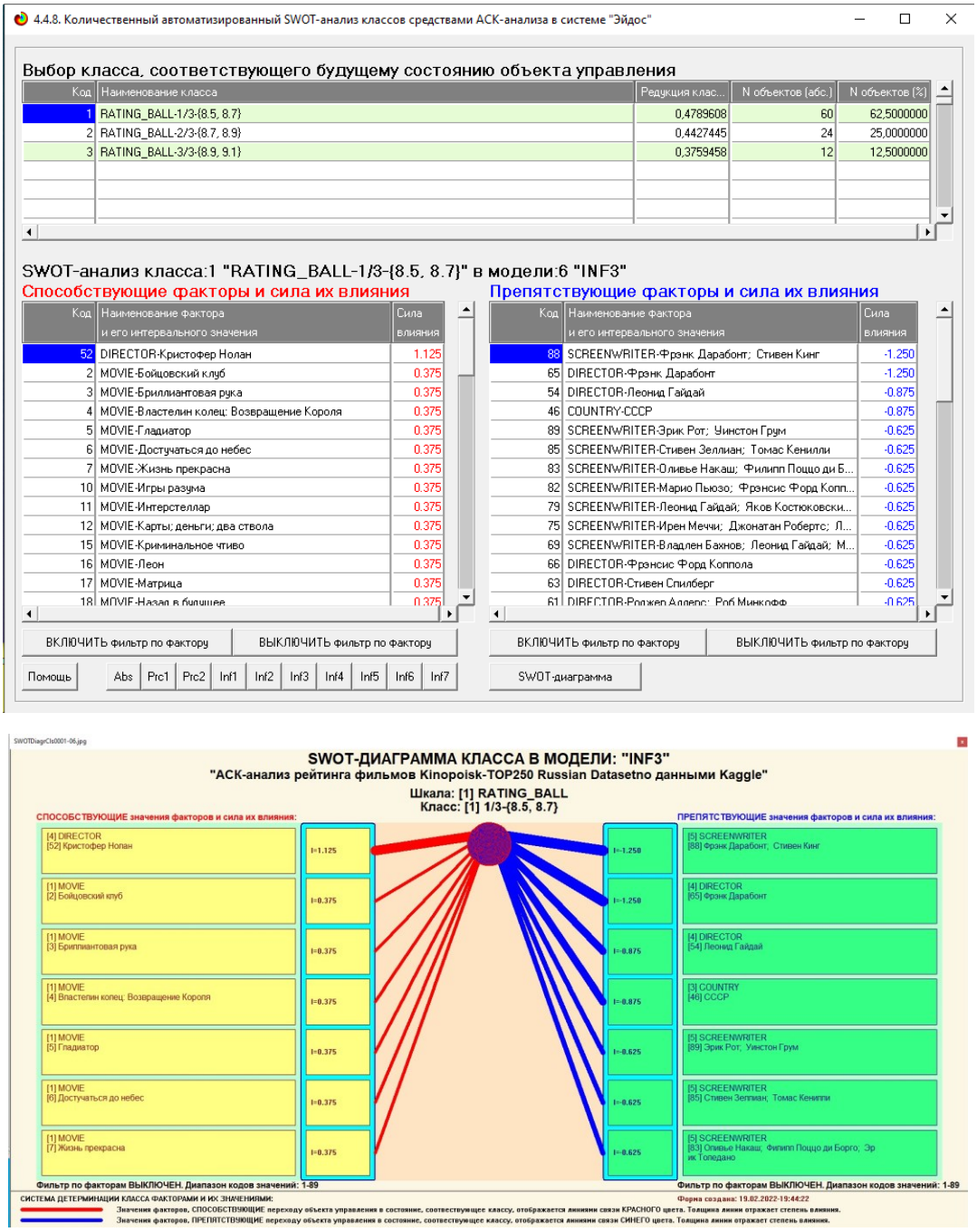


Рисунок 21 – Количественный SWOT-анализ значений в модели INF3

3.8.2. Кластерно-конструктивный анализ классов

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

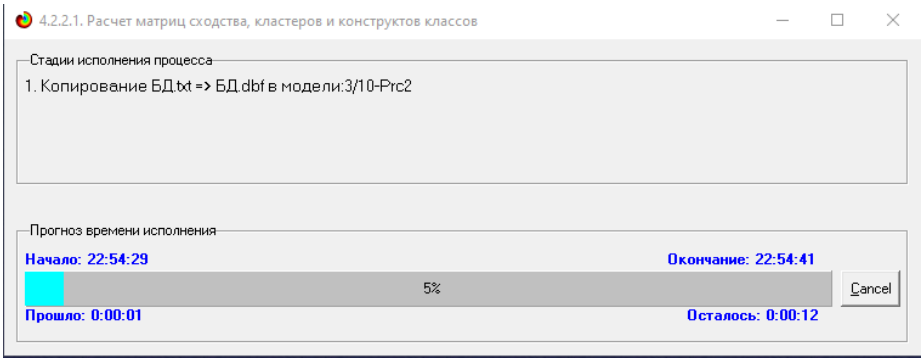


Рисунок 22 – Расчет матриц сходства, кластеров и конструкторов классов

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

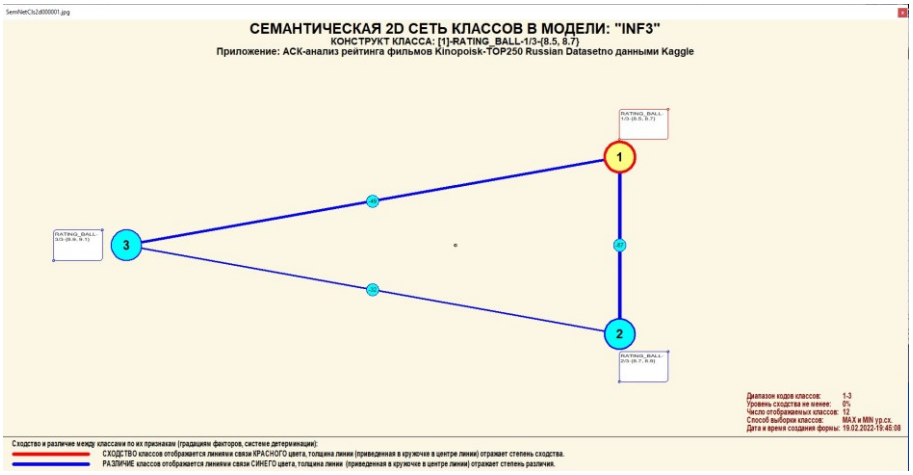


Рисунок 23 – Круговая 2d-когнитивная диаграмма классов

– агломеративных дендрограмм, полученных в результате когнитивной (истинной) кластеризации классов (режим 4.2.2.3);

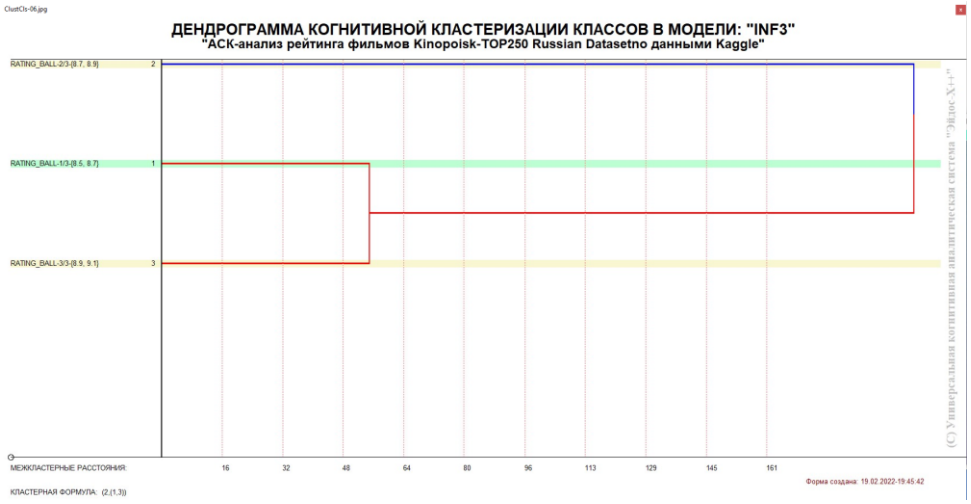


Рисунок 24 – Дендограмма когнитивной кластеризации классов

– график изменения межкластерных расстояний (режим 4.2.2.3);

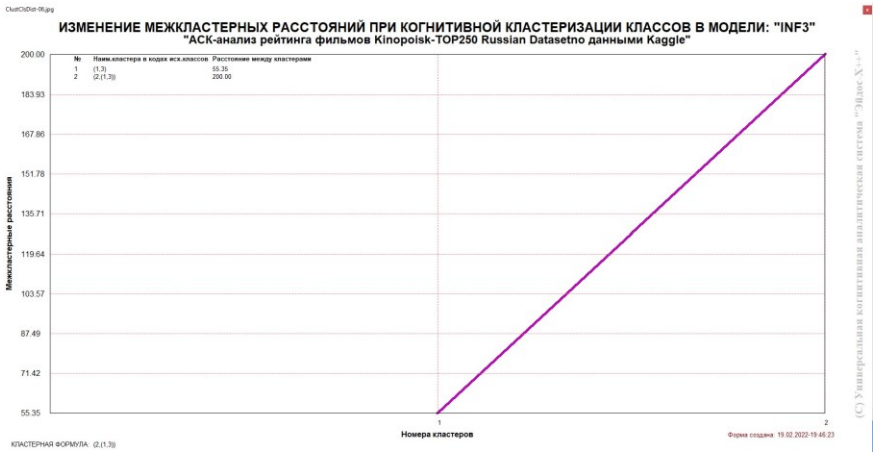


Рисунок 25 – График изменения межкластерных расстояний

3.8.3. Кластерно-конструктивный анализ значений описательных шкал

В системе «Эйдос» (в режиме 4.3.2.1) рассчитывается матрица сходства признаков по системе их смыслу и на основе этой матрицы рассчитывается и выводится три основных формы:

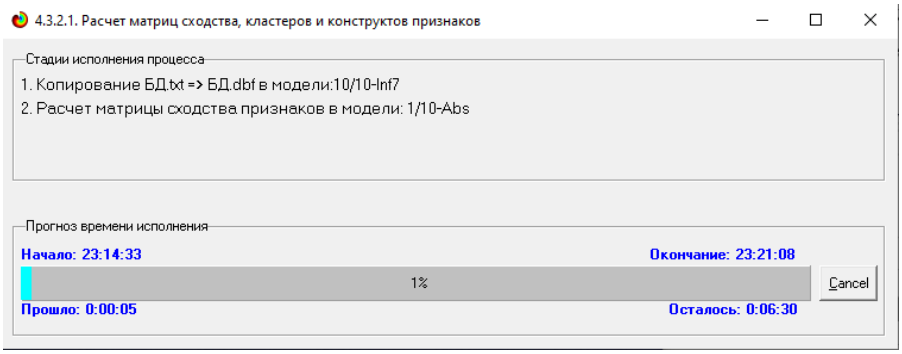


Рисунок 26 – Расчет матриц сходства, кластеров и конструкторов признаков

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

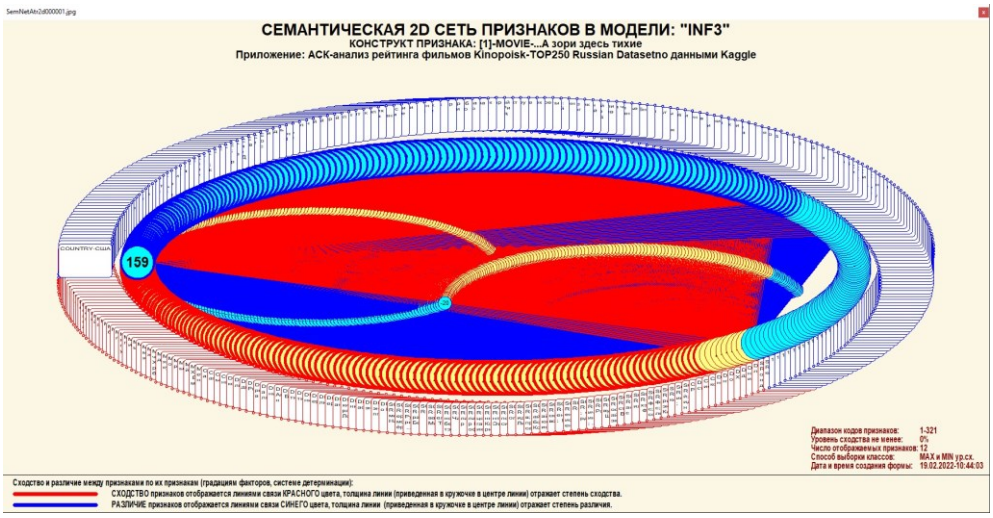


Рисунок 27 – круговая 2d-когнитивная диаграмма признаков

– агломеративных дендрограмм, полученных в результате когнитивной (истинной) кластеризации признаков (режим 4.3.2.3);

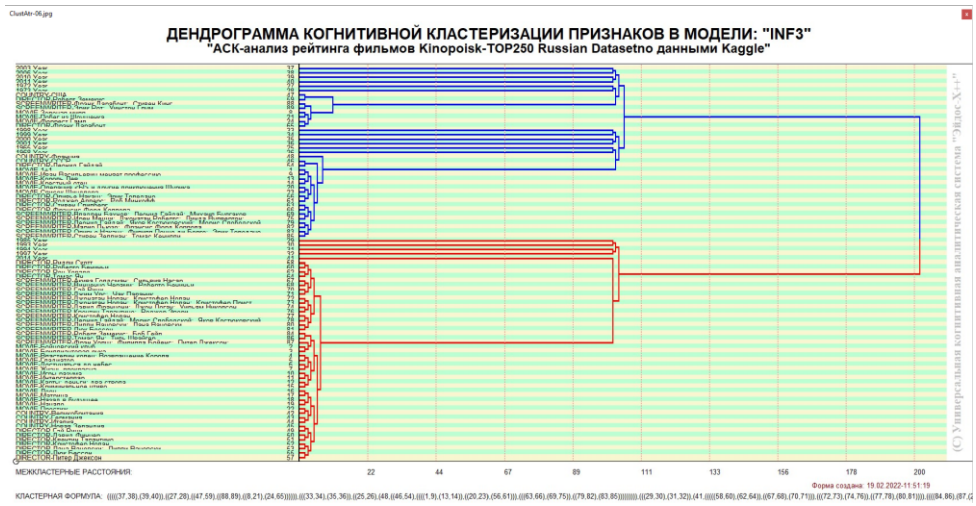


Рисунок 28 – дендограмма когнитивной кластеризации признаков

– график изменения межкластерных расстояний (режим 4.3.2.3);

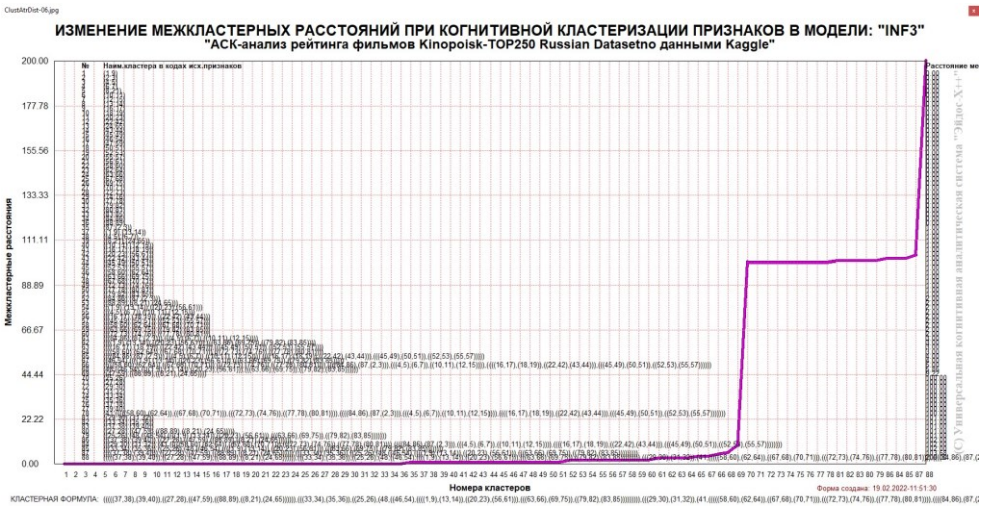


Рисунок 29 – график изменения межкластерных расстояний

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 [12]:

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 отображается в форме цвета и толщины дендрита.

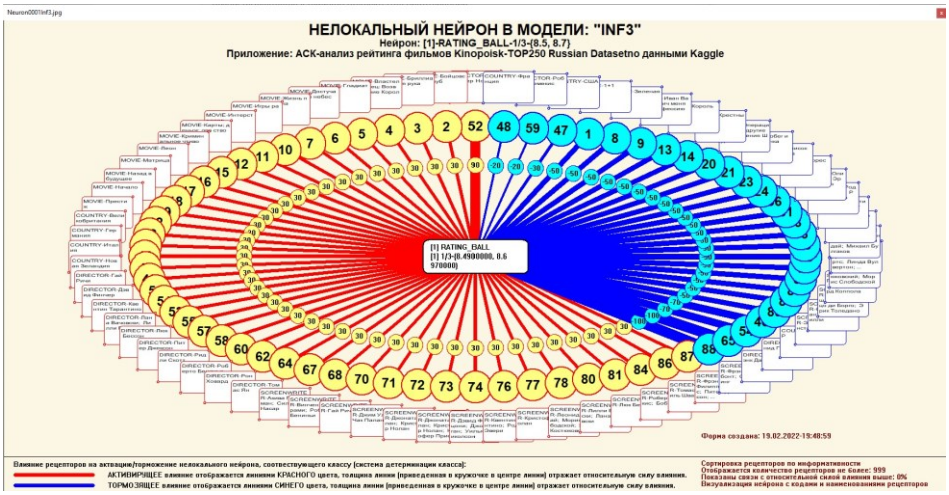


Рисунок 30 – Нелокальный нейрон в модели «INF3»

3.8.5. Нелокальная нейронная сеть

В системе «Эйдос» есть возможность построения моделей, соответствующих многослойным нейронным сетям [12].

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.

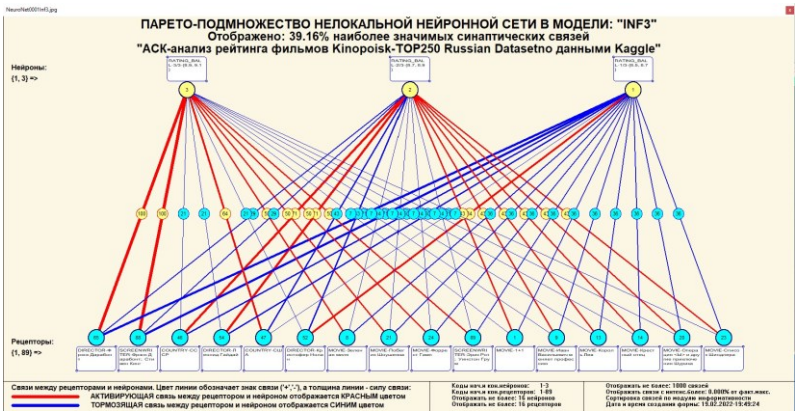


Рисунок 31 – Паретто – подмножество нелокальной нейронной сети

3.8.6. 3D-интегральные когнитивные карты

3d-интегральная когнитивная карта является отображением на одном рисунке когнитивных диаграмм классов и значений факторов вверху и внизу соответственно и одного слоя нейронной сети (режим 4.4.12 системы «Эйдос»).

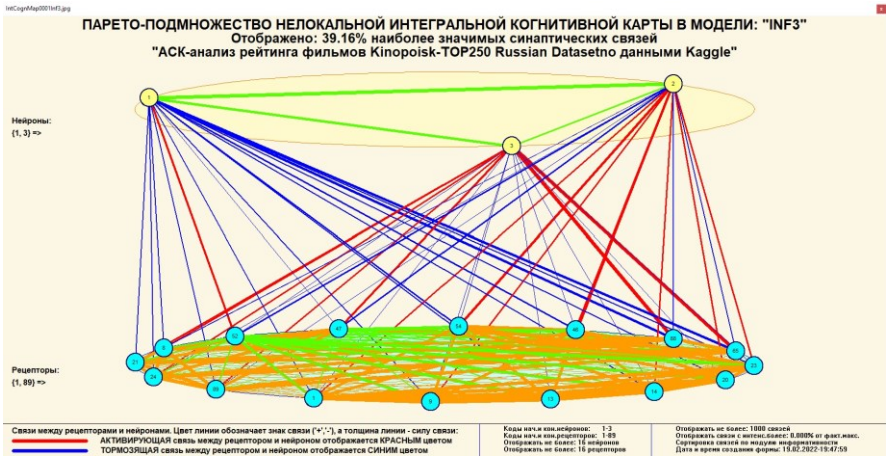


Рисунок 32 – 3d-когнитивная диаграмма классов и признаков

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 первым писал Дьердь Пойа [13]. Впервые об автоматизированной реализации рассуждений подобного типа в интеллектуальной системе «Эйдос» написано в 2002 году в работе [1] на странице 521[9]. Позже об этом писалось в работе [7][10] и ряде других работ автора, поэтому здесь подробнее рассматривать этот вопрос нецелесообразно.

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

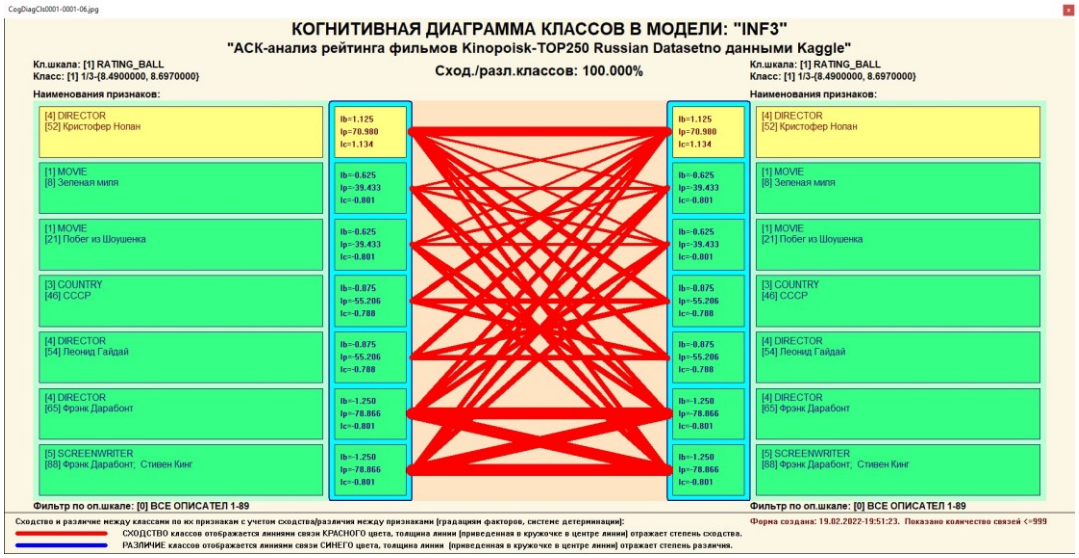


Рисунок 33 – Когнитивная диаграмма классов в модели

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

Из 2d-когнитивных диаграмм 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 [14].

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».

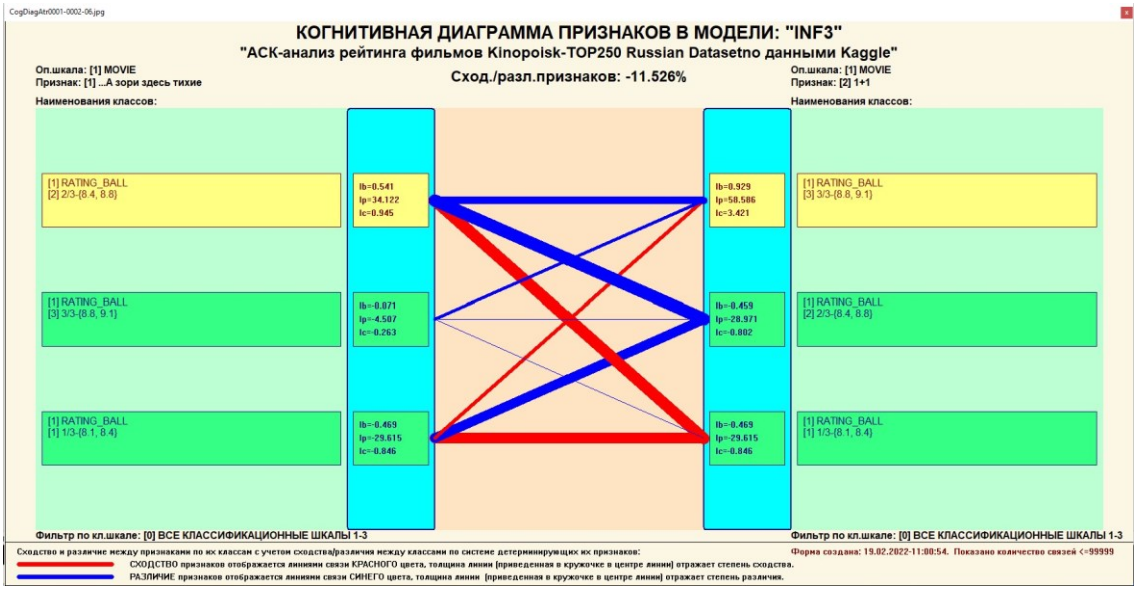


Рисунок 34 – Когнитивная диаграмма признаков в модели

3.8.9. Значимость описательных шкал и их градаций

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

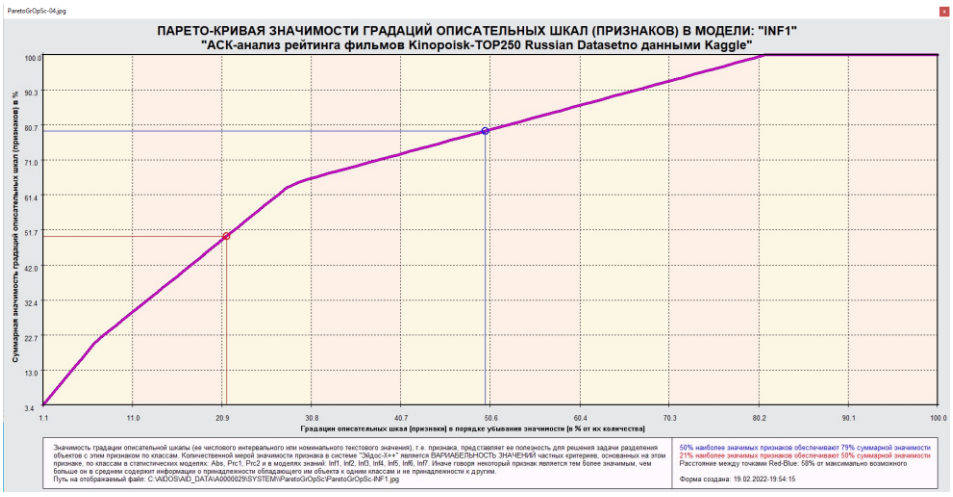


Рисунок 35 – Парето-кривая значимости градаций описательных шкал

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

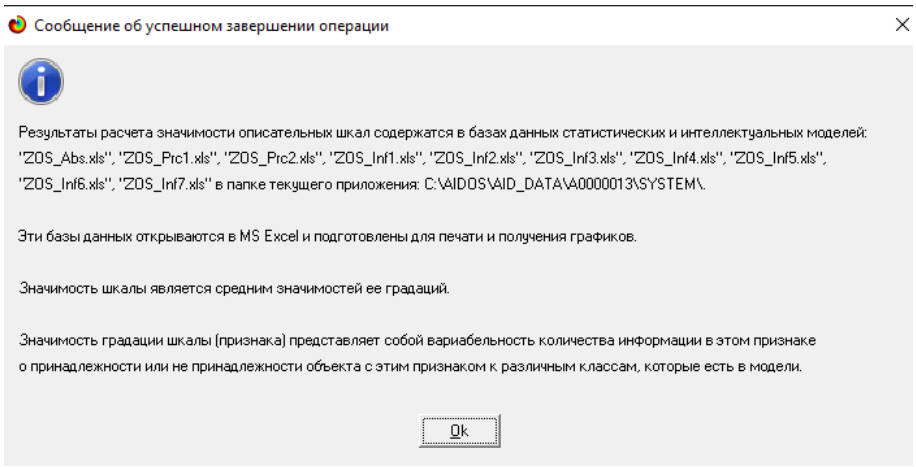


Рисунок 36 – Среднее от степени значимости

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.

3.8.10. 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.

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»).

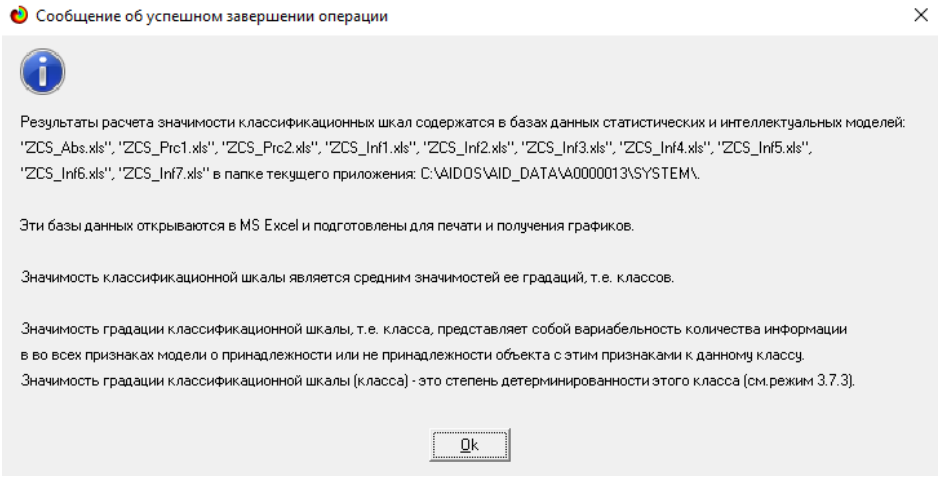


Рисунок 37 – Среднее от степени детерминированности

4. ОБСУЖДЕНИЕ

Цель данной курсовой работы была успешно достигнута в рамках исследования. Однако исходные данные, использованные для ее решения, имеют ряд ограничений: они имеют ограниченный объем, а также неполный перечень данных о фильмах, их длительности, оценки с разных сайтов и форумов, а так же просмотров в онлайн кинотеатре.

Методы АСК-анализа и его программный инструментарий интеллектуальная система «Эйдос» могут быть применены и для решения задач прогнозирования, принятия решений и исследования моделируемой предметной области путем исследования ее модели и в других предметных областях.

ЗАКЛЮЧЕНИЕ

Целью работы было провести автоматизированный системно-когнитивный рейтинга фильмов в онлайн кинотеатре «Кинопоиск» на основе данный портала Kaggle.

Для этого были изучены методы формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами. Построение моделей было осуществлено с помощью системы искусственного интеллекта Aidos-X, наиболее достоверной моделью оказалась модель INF3, точность модели составила 0,995.

АСК-анализ, использованный в данной работе, позволяет:

– сформировать обобщенные лингвистические образы классов (семантические ядра) на основе фрагментов или примеров, относящихся к ним текстов на любом языке;

– количественно сравнить лингвистический образ конкретного героя, или описание объекта, процесса с обобщенными лингвистическими образами групп (классов);

– сравнить обобщенные лингвистические образы классов друг с другом и создавать их кластеры и конструкты;

– исследовать моделируемую предметную область путем исследования лингвистической системно-когнитивной модели.

Все это можно делать для любого естественного или искусственного языка, или системы кодирования.

СПИСОК ЛИТЕРАТУРЫ

1. Луценко Е.В. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос» и системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №02(116). С. 1 – 60. – IDA [article ID]: 1161602001. – Режим доступа: http://ej.kubagro.ru/2016/02/pdf/01.pdf, 3,75 у.п.л.3.

2. Луценко Е.В. АСК-анализ, моделирование и идентификация живых существ на основе их фенотипических признаков / Е.В. Луценко, Ю.Н. Пенкина // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2014. – №06(100). С. 1346 – 1395. – IDA [article ID]: 1001406090. – Режим доступа: http://ej.kubagro.ru/2014/06/pdf/90.pdf, 3,125 у.п.л.

3. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

4. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: http://lc.kubagro.ru/, свободный. - Загл. с экрана. Яз.рус.

5. Луценко Е.В. Количественная оценка степени манипулирования индексом Хирша и его модификация, устойчивая к манипулированию / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №07(121). С. 202 – 234. – IDA [article ID]: 1211607005. – Режим доступа: http://ej.kubagro.ru/2016/07/pdf/05.pdf, 2,062 у.п.л.. http://dx.doi.org/10.21515/1990-4665-121-0057.

6. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос-X++» Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 328 – 356. – IDA [article ID]: 0831209025. – Режим доступа: http://ej.kubagro.ru/2012/09/pdf/25.pdf, 1,812 у.п.л.

7. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: http://ej.kubagro.ru/2003/01/pdf/11.pdf, 0,812 у.п.л.

8. Луценко Е.В. Синтез семантических ядер научных специальностей ВАК РФ и автоматическая классификации статей по научным специальностям с применением АСК-анализа и интеллектуальной системы «Эйдос» (на примере Научного журнала КубГАУ и его научных специальностей: механизации, агрономии и ветеринарии) / Е.В. Луценко, Н.В. Андрафанова, Н.В. Потапова // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2019. – №01(145). С. 31 – 102. – IDA [article ID]: 1451901033. – Режим доступа: http://ej.kubagro.ru/2019/01/pdf/33.pdf, 4,5 у.п.л.

9. Луценко Е.В. Эффективность объекта управления как его эмерджентное свойство и повышение уровня системности как цель управления / Луценко Е.В. //

Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2021. – №01(165). – Режим доступа: http://ej.kubagro.ru/2021/01/pdf/09.pdf, 1,313 у.п.л. – IDA [article ID]: 1652101009. http://dx.doi.org/10.21515/1990-4665-165-009

10. 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 (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: http://ej.kubagro.ru/2013/08/pdf/58.pdf, 1,562у.п.л.

11. Луценко Е.В. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос» / Е.В. Луценко, Е.К. Печурина, А.Э. Сергеев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №06(160). С. 95 – 114. – IDA [article ID]: 1602006009. – Режим доступа: http://ej.kubagro.ru/2020/06/pdf/09.pdf, 1,25 у.п.л.

12. Луценко Е.В., Подсистема агломеративной когнитивной кластеризации классов системы «Эйдос» ("Эйдос-кластер"). Пат. № 2012610135 РФ. Заяв. № 2011617962 РФ 26.10.2011. Опубл. От 10.01.2012. – Режим доступа: http://lc.kubagro.ru/aidos/2012610135.jpg, 3,125 у.п.л.

13. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: http://ej.kubagro.ru/2003/01/pdf/11.pdf, 0,812у.п.л.

14. Луценко Е.В. Метод визуализации когнитивных функций – новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №03(067). С. 240 – 282. – Шифр Информрегистра: 0421100012\0077, IDA [article ID]: 0671103018. – Режим доступа: http://ej.kubagro.ru/2011/03/pdf/18.pdf, 2,688 у.п.л.

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