

УДК 4.08

**Автоматизированный системно-когнитивный анализ характеристик жанров фильмов**

Семенова Линда Владиславовна,  
студент факультета ПИ, группы ИТ2002  
[linaseeds@gmail.com](mailto:linaseeds@gmail.com)

*Кубанский государственный аграрный университет имени И.Т.Трубилина, Краснодар, Россия*

Целью данной работы является изучение характеристик жанров фильмов. Достижение данной цели представляет большой личный интерес. Для нас это позволит получить знания в работе с универсальной когнитивной аналитической системой «Эйдос-Х++», а также получить зачет. Для достижения поставленной цели применяется Автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальная система «Эйдос».

Ключевые слова: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, АСК-АНАЛИЗ, СИСТЕМА «ЭЙДОС».

**Automated system-cognitive analysis of the characteristics of genres of films**

Semenova Linda Vladislavovna,  
student of the faculty of PI, group IT2002  
[linaseeds@gmail.com](mailto:linaseeds@gmail.com)

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

The purpose of this work is to study the characteristics of genres of films. The achievement of this goal is of great personal interest. For us, this will allow us to gain knowledge in working with the universal analytical system "Eidos-X ++", as well as get a credit. To achieve this goal, the Automated System-Cognitive Analysis (ASK-analysis) and its software tools are used - the intelligent system "Eidos".

Key words: AUTOMATED SYSTEM-COGNITIVE ANALYSIS, ASC-ANALYSIS, "EIDOS" SYSTEM.

## **Введение**

**Целью** данной работы является изучение характеристик жанров фильмов.

Достижение данной цели представляет большой личный интерес. Для нас это позволяет получить знания в работе с универсальной когнитивной аналитической системой «Эйдос-Х++», а также получить зачет.

АСК-анализ предполагает, что для достижения поставленной цели необходимо решить следующие задачи, которые получаются путем декомпозиции цели и являются этапами ее достижения:

Задача 1: когнитивная структуризация предметной области.

Задача 2: подготовка исходных данных и формализация предметной области.

Задача 3: синтез и верификация статистических и системно-когнитивных моделей и выбор наиболее достоверной модели.

Задача 4: решение различных задач в наиболее достоверной модели:

- подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация);

- подзадача 4.2. Поддержка принятия решений;

- подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели (когнитивные диаграммы классов и значений факторов, агломеративная когнитивная кластеризация классов и значений факторов, нелокальные нейроны и нейронные сети, 3d-интегральные когнитивные карты, когнитивные функции).

Эти задачи, по сути, представляют собой **этапы** автоматизированного системно-когнитивный анализа (ACK-анализ), который поэтому и предлагается применить для их решения.

ACK-анализ представляет собой метод искусственного интеллекта, разработанный проф. Е.В. Луценко в 2002 году для решения широкого класса задач идентификации, прогнозирования, классификации, диагностики, поддержки принятия решений и исследования моделируемой предметной области путем исследования ее модели. ACK-анализ доведен до **инновационного** уровня благодаря тому, что имеет свой программный инструментарий – универсальную когнитивную аналитическую систему «Эйдос-Х++» (система «Эйдос»).

**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 (<http://lc.kubagro.ru/aidos/index.htm>), в которых не требуется автоматического, т.е. без непосредственного участия человека в реальном времени решения задач идентификации, прогнозирования, принятия решений и исследования предметной области;

- находится в полном открытом бесплатном доступе ([http://lc.kubagro.ru/aidos/\\_Aidos-X.htm](http://lc.kubagro.ru/aidos/_Aidos-X.htm)), причем с актуальными исходными текстами ([http://lc.kubagro.ru/\\_AIDOS-X.txt](http://lc.kubagro.ru/_AIDOS-X.txt));

- 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>);

- 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 (в настоящее время их 31 и 240, соответственно) ([http://lc.kubagro.ru/aidos/Presentation\\_Aidos-online.pdf](http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf));

- поддерживает on-line среду накопления знаний и широко используется во всем мире (<http://aidos.byethost5.com/map5.php>);

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

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

- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе: [http://lc.kubagro.ru/aidos/aidos18\\_LLS/aidos18\\_LLS.pdf](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 на теоретическом уровне познания в теоретических научных законах.

Всем этим и обусловлен выбор АСК-анализа и его программного инструментария – интеллектуальной системы «Эйдос» в качестве метода и инструмента решения поставленной проблемы и достижения цели работы (рисунок 1).

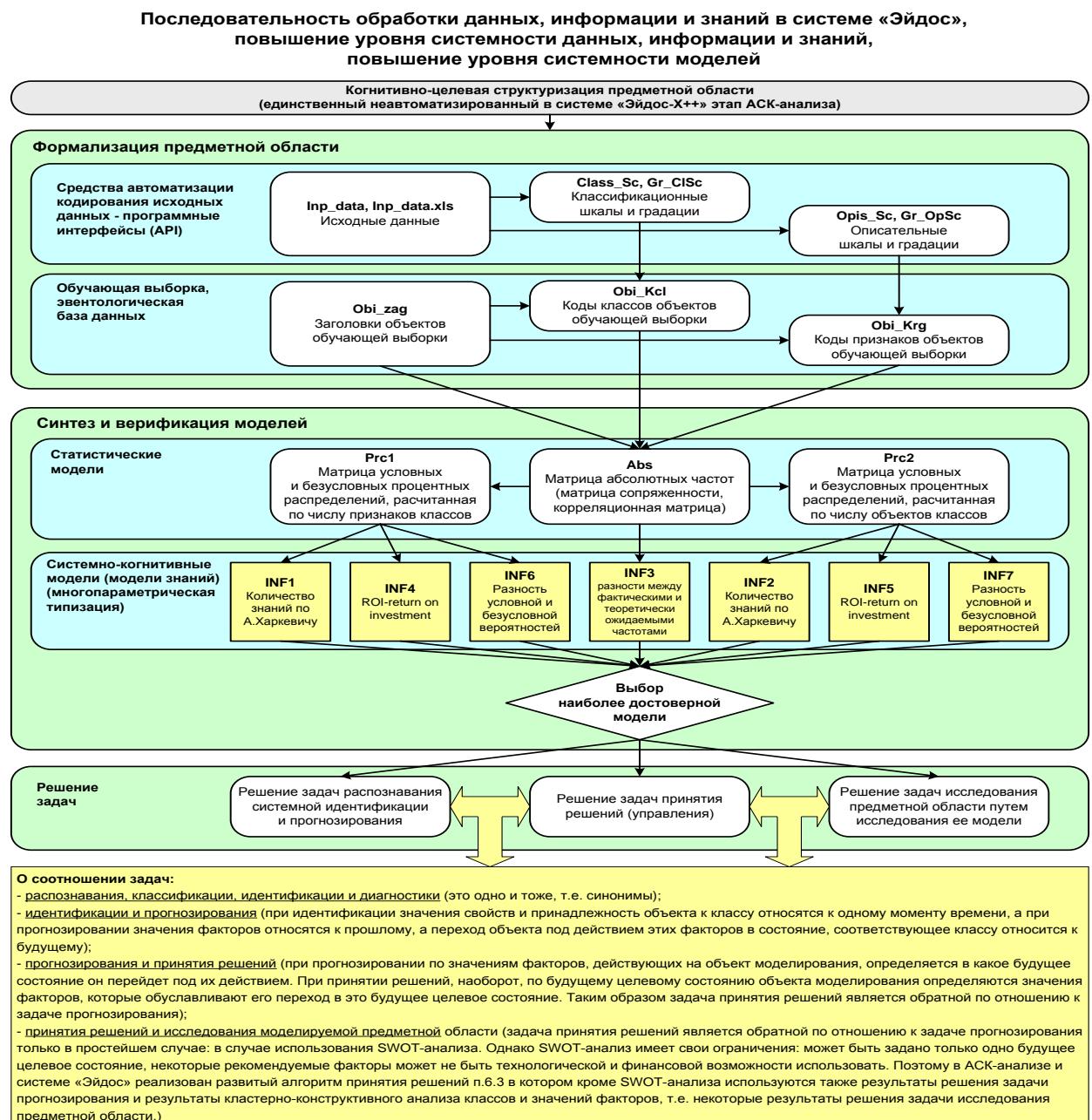


Рисунок 1. Последовательность решения задач в АСК-анализе и системе «Эйдос»

Рассмотрим решение поставленных задач в подробном численном примере.

## **Задача 1: когнитивная структуризация предметной области**

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).

При этом необходимо отметить, что системно-когнитивные модели (СК-модели) отражают лишь сам факт наличия зависимостей между значениями факторов и результатами их действия. Но они не отражают причин и механизмов такого влияния.

Это значит:

– во-первых, что содержательная интерпретация СК-моделей – это компетенция специалистов-экспертов хорошо разбирающихся в данной предметной области. Иногда встречается ситуация, когда и то, что на первый взгляд являются причинами, и то, что казалось бы является их последствиями, на самом деле являются последствиями неких глубинных причин, которых мы не видим и никоим образом непосредственно не отражаем в модели;

– во-вторых, даже если содержательной интерпретации не разработано, то в принципе это не исключает возможности пользоваться ими на практике для достижения заданных результатов и поставленных целей, т.е. для управления.

В данной работе в качестве классификационных шкал выберем категорию жанры, как она обозначается в самом списке (таблица 1), а в качестве факторов, влияющих на эти результаты – список фильмов, дата выпуска фильма, возрастное ограничение фильма, рейтинг фильма, есть ли этот фильм в Netflix(таблица 2):

Таблица 1 – Классификационная шкала

Код	Наименование
1	Genre

Таблица 2 – Описательные шкалы

Код	Наименование
1	Year
2	Rating
3	IMDB_Rating
4	Netflix

## **Задача 2: подготовка исходных данных и формализация предметной области**

Исходные данные для данной статьи (рисунок 2) получены из списков популярных фильмов, а также из других подробных списков фильмам, найденных в интернете.

	A	B	C	D	E	F
1	Titles	Genre	Year	Rating	IMDB_Rating	Netflix
2	Breaking Bad	thriller	2008	18+	9,5	yes
3	Game of Thrones	fantasy	2011	18+	9,3	no
4	Rick and Morty	comedy	2013	18+	9,2	no
5	Dark	thriller	2017	16+	8,8	yes
6	Stranger Things	horror	2016	16+	8,8	yes
7	Avatar: The Last Airbender	anime	2005	7+	9,2	yes
8	Sherlock	detective	2010	16+	9,1	yes
9	Chernobyl	drama	2019	18+	9,4	no
10	Friends	comedy	1994	16+	8,9	no
11	The Office	comedy	2005	16+	8,9	yes
12	Better Call Saul	drama	2015	18+	8,7	yes
13	The Wire	thriller	2002	18+	9,3	no
14	Supernatural	fantasy	2005	16+	8,4	no
15	Fargo	thriller	2014	18+	8,9	no
16	Black Mirror	fantasy	2011	18+	8,8	yes
17	Band of Brothers	drama	2001	18+	9,4	no
18	Attack on Titan	drama	2013	16+	8,8	no
19	House	drama	2004	16+	8,7	no
20	Money Heist	thriller	2017	18+	8,4	yes
21	Peaky Blinders	drama	2013	18+	8,8	yes
22	Mr. Robot	thriller	2015	18+	8,5	no
23	Vikings	drama	2013	18+	8,5	no
24	The Sopranos	drama	1999	18+	9,2	no
25	The Simpsons	comedy	1989	7+	8,7	no
26	The Mandalorian	fantasy	2019	7+	8,7	no
27	Lost	fantasy	2004	16+	8,3	no
28	Westworld	fantasy	2016	18+	8,7	no
29	The Umbrella Academy	fantasy	2019	16+	8,1	yes
30	Narcos	drama	2015	18+	8,8	yes
31	True Detective	detective	2014	18+	9	no
32	Marvel's Daredevil	fantasy	2015	18+	8,6	yes
33	Brooklyn Nine-Nine	comedy	2013	16+	8,4	no
34	The Walking Dead	horror	2010	18+	8,2	yes

Рисунок 2 – Исходные данные для ввода в систему «Эйдос»

Затем с параметрами, показанными на рисунке 3, запустим режим 2.3.2.2 системы «Эйдос», представляющий собой автоматизированный программный интерфейс (API) с внешними данными табличного типа.

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

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

Задайте тип файла исходных данных: "Inp\_data":

- XLS - MS Excel-2003
- XLSX - MS Excel-2007(2010)
- DBF - DBASE IV (DBF/NTX)
- CSV - CSV => DBF конвертер

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

Задайте параметры:

- Нули и пробелы считать ОТСУТСТВИЕМ данных
- Нули и пробелы считать ЗНАЧЕНИЯМИ данных
- Создавать БД средних по классам "Inp\_davg.dbf"?
- Требования к файлу исходных данных

Задайте диапазон столбцов классификационных шкал:

Начальный столбец классификационных шкал: 2  
Конечный столбец классификационных шкал: 2

Задайте диапазон столбцов описательных шкал:

Начальный столбец описательных шкал: 3  
Конечный столбец описательных шкал: 6

Задайте режим:

- Формализации предметной области (на основе "Inp\_data")
- Генерации распознаваемой выборки (на основе "Inp\_asp")

Задание параметров формирования сценариев или способа интерпретации текстовых полей "Inp\_data":

- Не применять сценарий метод АСК-анализа
- Применить сценарий метод АСК-анализа

**Параметры интерпретации значений текстовых полей "Inp\_data":**

Интерпретация TXT-полей классов:  
Значения полей текстовых классификационных шкал файла исходных данных "Inp\_data" рассматриваются как целое

Интерпретация TXT-полей признаков:  
Значения полей текстовых описательных шкал файла исходных данных "Inp\_data" рассматриваются как целое

Какие наименования ГРАДАЦИЙ числовых шкал использовать:

- Только интервальные числовые значения [например: "1/3-(59873.000000, 178545.6666667)"]
- Только наименования интервальных числовых значений [например: "Минимальное"]
- И интервальные числовые значения, и их наименования [например: "Минимальное: 1/3-(59873.000000, 178545.6666667)"]

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

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [7 x 11]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс.шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис.шкалу
Числовые	0	0	0,00	2	6	3,00
Текстовые	1	7	7,00	2	5	2,50
ВСЕГО:	1	7	7,00	4	11	2,75

Задайте число интервалов (градаций) в шкале:

В описательных шкалах: 3

Пересчитать шкалы и градации Параметры числ.шкал и градаций Выйти на создание модели

**2.3.2.2. Процесс импорта данных из внешней БД "Inp\_data" в систему "ЭЙДОС-Х++"**

Стадии исполнения процесса

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

ПРОЦЕСС ФОРМАЛИЗАЦИИ ПРЕДМЕТНОЙ ОБЛАСТИ ЗАВЕРШЕН УСПЕШНО !!!

Прогноз времени исполнения

Начало: 13:50:26	Окончание: 13:50:29
100%	
Прошло: 0:00:03	Осталось: 0:00:00

Ok

Рисунок 3. Экранные формы программного интерфейса (API) 2.3.2.2 системы «Эйдос» с внешними данными табличного типа

Обратим внимание на то, что заданы аддативные интервалы, учитывающее неравномерность распределения данных по диапазону значений, что важно при относительно небольшом числе наблюдений. Если бы интервалы были заданы равными по величине, то в них бы учитывалось сильно отличающееся число наблюдений, а в некоторых интервалах их бы могло не оказаться вовсе. В описательных шкалах задано 3 числовых интервальных значения.

На рисунке 4 приведен Help данного режима, в котором объясняется принцип организации таблицы исходных данных для данного режима. Здесь же обратим внимание на то, что в таблице 3 значения параметров могут быть представлены как числовыми, так и текстовыми значениями.

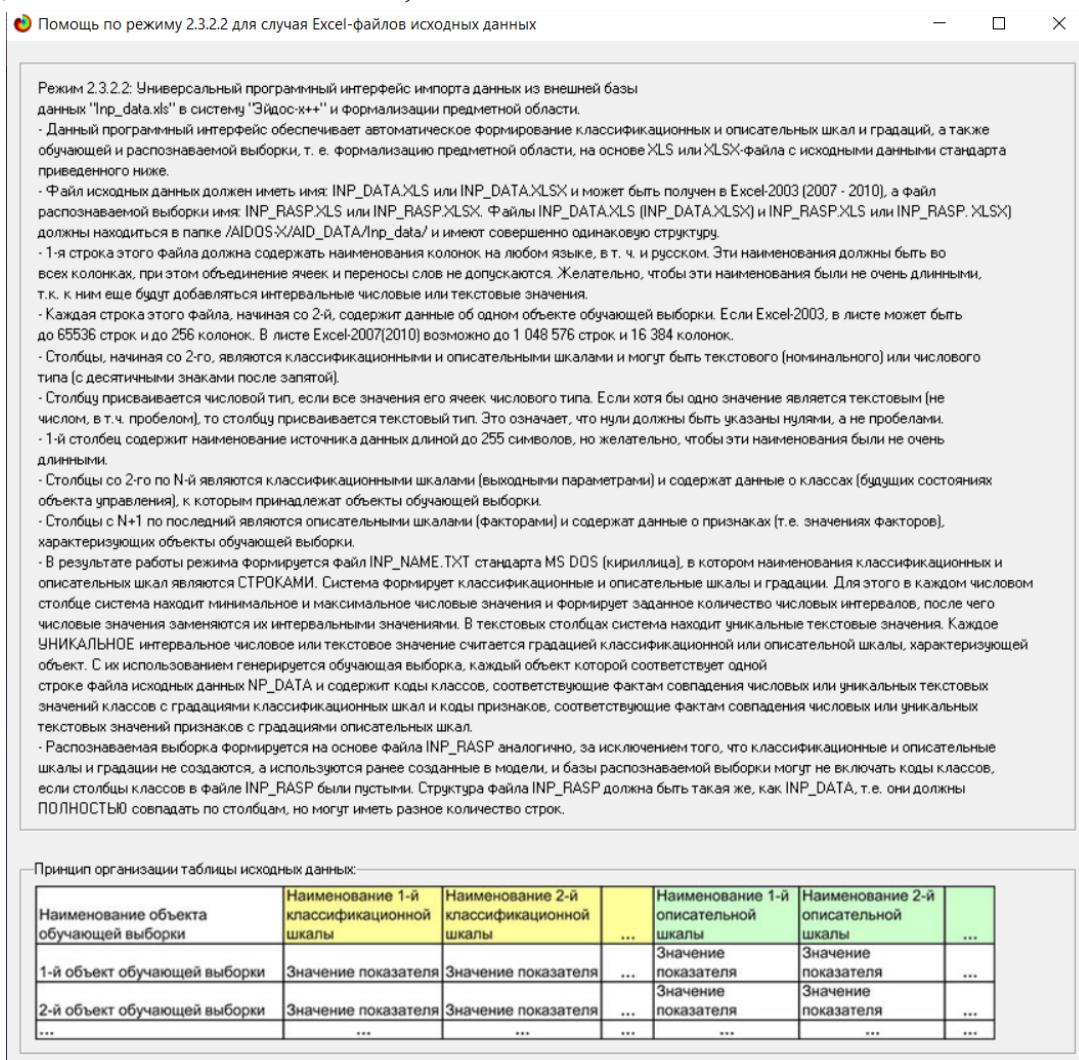


Рисунок 4. Экранная форма HELP программного интерфейса (API) 2.3.2.2

В результате работы режима сформирована 1 классификационная шкала с суммарным количеством градаций (классов) 7 (таблица 3) и 4 описательных шкал с суммарным числом градаций 11 (таблица 4). С использованием классификационных и описательных шкал и градаций исходные данные (рисунок 2) были закодированы и в результате получена обучающая выборка (таблица 5):

Таблица 3 – Классификационные шкалы и градации жанров фильмов

Код	Название
1	GENRE-1/7-anime
2	GENRE-2/7-comedy
3	GENRE-3/7-detective
4	GENRE-4/7-drama
5	GENRE-5/7-fantasy
6	GENRE-6/7-horror
7	GENRE-7/7-thriller

Таблица 4 – Описательные шкалы и градации жанров фильмов

Код	Название
1	1/3-{1989.0000000, 2007.0000000}
2	2/3-{2007.0000000, 2013.0000000}
3	3/3-{2013.0000000, 2019.0000000}
4	1/3-16+
5	2/3-18+
6	3/3-7+
7	1/3-{8.1000000, 8.5000000}
8	2/3-{8.5000000, 8.8000000}
9	3/3-{8.8000000, 9.5000000}
10	1/2-no
11	2/2-yes

Таблица 5 – Обучающая выборка

NAME_OBJ	N2	N3	N4	N5	N6
Breaking Bad	7	2	5	9	11
Game of Thrones	5	2	5	9	10
Rick and Morty	2	2	5	9	10
Dark	7	3	4	8	11
Stranger Things	6	3	4	8	11
Avatar: The Last Airbender	1	1	6	9	11
Sherlock	3	2	4	9	11
Chernobyl	4	3	5	9	10
Friends	2	1	4	9	10
The Office	2	1	4	9	11
Better Call Saul	4	3	5	8	11
The Wire	7	1	5	9	10
Supernatural	5	1	4	7	10
Fargo	7	3	5	9	10
Black Mirror	5	2	5	8	11
Band of Brothers	4	1	5	9	10
Attack on Titan	4	2	4	8	10
House	4	1	4	8	10
Money Heist	7	3	5	7	11
Peaky Blinders	4	2	5	8	11
Mr. Robot	7	3	5	7	10
Vikings	4	2	5	7	10
The Sopranos	4	1	5	9	10
The Simpsons	2	1	6	8	10
The Mandalorian	5	3	6	8	10

Lost	5	1	4	7	10
Westworld	5	3	5	8	10
The Umbrella Academy	5	3	4	7	11
Narcos	4	3	5	8	11
True Detective	3	3	5	9	10
Marvel's Daredevil	5	3	5	8	11
Brooklyn Nine-Nine	2	2	4	7	10
The Walking Dead	6	2	5	7	11
Parks and Recreation	2	2	4	8	10
Mindhunter	7	3	5	8	11
South Park	2	1	5	8	10
Suits	4	2	4	7	10
Hannibal	3	2	5	7	11
Dexter	7	1	5	8	10
Community	2	2	6	7	10
Fringe	7	2	4	7	10
Modern Family	4	2	4	7	10
The Big Bang Theory	2	1	4	7	10
Firefly	5	1	4	9	10
Mad Men	4	1	4	8	10
Fullmetal Alchemist: Brotherhood	1	2	5	9	10
Ozark	7	3	5	7	11
The Witcher	5	3	5	7	11
Lucifer	5	3	4	7	11
It's Always Sunny in Philadelphia	2	1	5	8	10

Обучающая выборка (таблица 5), по сути, представляет собой нормализованные исходные данные, т.е. таблицу исходных данных (рисунок 2), закодированную с помощью классификационных и описательных шкал и градаций (таблицы 3 и 4). Таким образом, созданы все необходимые и достаточные условия для выполнения следующего этапа АСК-анализа: т.е. для синтеза и верификации моделей.

### **Задача 3: синтез и верификация статистических и системно-когнитивных моделей и выбор наиболее достоверной из них для решения задач**

Синтез и верификация статистических и системно-когнитивных моделей (СК-моделей) моделей осуществляется в режиме 3.5 системы «Эйдос» (рисунок 5).

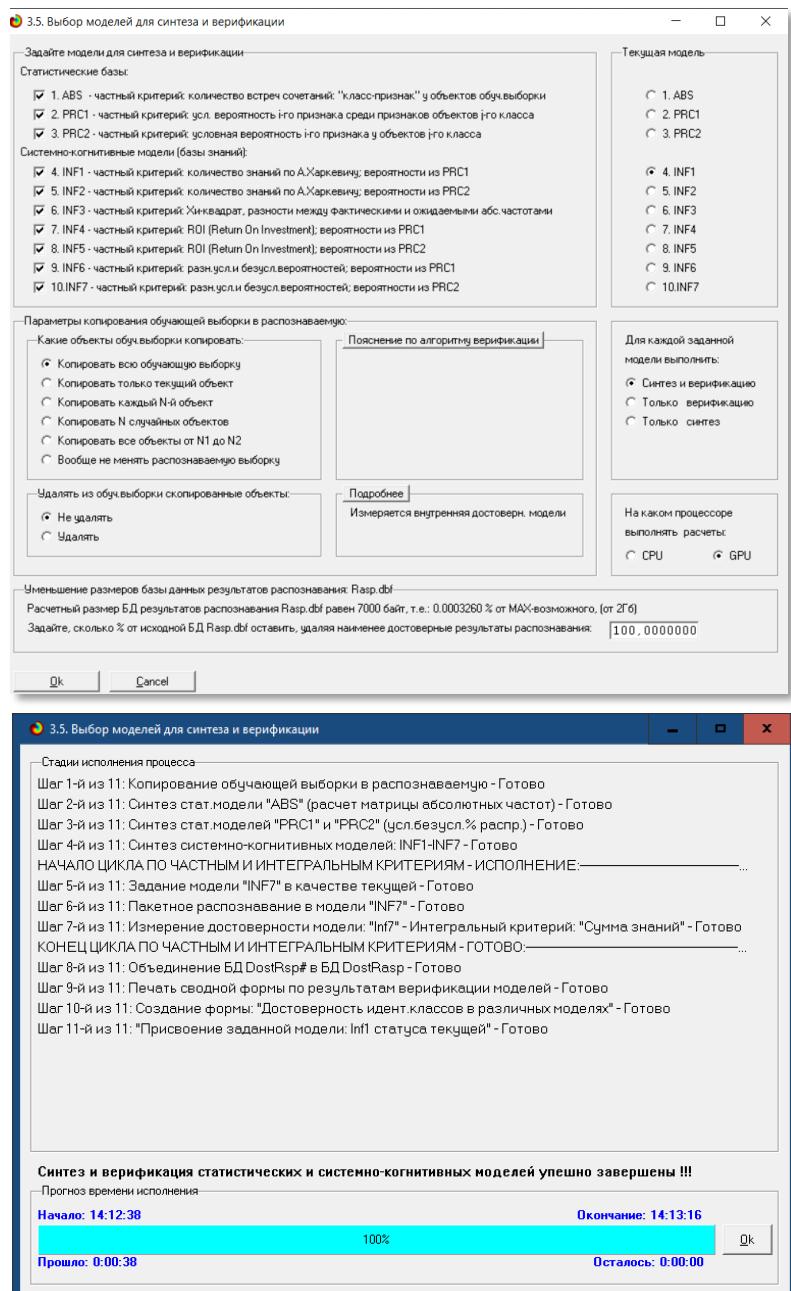


Рисунок 5. Экранные формы режима синтеза и верификации статистических и системно-когнитивных моделей системы «Эйдос»

Обратим внимание на то, что на рисунке 5 в правом нижнем углу окна задана опция: «Расчеты проводить на графическом процессоре (GPU)».

Из рисунка 5 видно, что весь процесс синтеза и верификации моделей занял 30 секунд. Отметим, что при синтезе и верификации моделей использовался графический процессор (GPU) видеокарты. На центральном процессоре (CPU) выполнение этих операций занимает значительно большее время (на некоторых задачах это происходит в десятки, сотни и даже тысячи раз дольше). Таким образом, неграфические вычисления на графических процессорах видеокарты делает возможной обработку больших объемов

исходных данных за разумное время. В процесс синтеза и верификации моделей осуществляется также расчет 10 выходных форм, на что уходит более 99% времени исполнения.

Фрагменты самих созданных статистических и системно-когнитивных моделей (СК-модели) приведены на рисунках 6, 7:

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"											
Код признака	Наклонение описательной шкалы и градации	1. GENRE 1/7 ANIME	2. GENRE 2/7 COMEDY	3. GENRE 3/7 DETECTIVE	4. GENRE 4/7 DRAMA	5. GENRE 5/7 FANTASY	6. GENRE 6/7 HORROR	7. GENRE 7/7 THRILLER	Сумма	Среднее	Средн. квадр. откл.
1	YEAR-1/3(1989.0000000, 2007.0000000)	1	6		4	3		2	16	2.29	2.21
2	YEAR-2/3(2007.0000000, 2013.0000000)	1	4	2	5	2	1	2	17	2.43	1.51
3	YEAR-3/3(2013.0000000, 2019.0000000)			1	3	6	1	6	17	2.43	2.64
4	RATING-1/3!+			5	1	5	1	2	19	2.71	2.21
5	RATING-2/3!+	1	3	2	7	5	1	8	27	3.86	2.85
6	RATING-3/3!+	1	2			1		4	4	0.57	0.79
7	IMDB_RATING-1/3-(8.1000000, 8.5000000)		3	1	3	5	1	4	17	2.43	1.81
8	IMDB_RATING-2/3-(8.5000000, 8.8000000)		4		6	4	1	3	18	2.57	2.30
9	IMDB_RATING-3/3-(8.8000000, 9.5000000)	2	3	2	3	2		3	15	2.14	1.07
10	NETFLIX1/2/no	1	9	1	9	6		5	31	4.43	3.82
11	NETFLIX2/2/yes	1	1	2	3	5	2	5	19	2.71	1.70
Сумма числа признаков											
Среднее											
Среднеквадратичное отклонение											
Сумма числа объектов обуч.выборки											

Рисунок 6. Матрица абсолютных частот (фрагмент)

5.5. Модель: "б. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми abs.частотами"											
Код признака	Наклонение описательной шкалы и градации	1. GENRE 1/7 ANIME	2. GENRE 2/7 COMEDY	3. GENRE 3/7 DETECTIVE	4. GENRE 4/7 DRAMA	5. GENRE 5/7 FANTASY	6. GENRE 6/7 HORROR	7. GENRE 7/7 THRILLER	Сумма	Среднее	Средн. квадр. откл.
1	YEAR-1/3(1989.0000000, 2007.0000000)	0.360	2.800	-0.960	0.160	-0.520	-0.640	-1.200		1.356	
2	YEAR-2/3(2007.0000000, 2013.0000000)	0.320	0.600	0.980	0.920	-1.740	0.320	-1.400	0.000	1.107	
3	YEAR-3/3(2013.0000000, 2019.0000000)	-0.680	-3.400	-0.020	-1.080	2.260	0.320	2.600	0.000	2.048	
4	RATING-1/3!+	-0.760	1.200	-0.140	0.440	0.820	0.240	-1.800	0.000	1.017	
5	RATING-2/3!+	-0.080	-2.400	0.380	0.520	-0.940	-0.080	2.600	0.000	1.518	
6	RATING-3/3!+	0.840	1.200	-0.240	-0.960	0.120	-0.160	-0.800		0.796	
7	IMDB_RATING-1/3-(8.1000000, 8.5000000)	-0.680	-0.400	-0.020	-1.080	1.260	0.320	0.600		0.800	
8	IMDB_RATING-2/3-(8.5000000, 8.8000000)	-0.720	0.400	-1.080	1.680	0.040	0.280	-0.600		0.923	
9	IMDB_RATING-3/3-(8.8000000, 9.5000000)	1.400		1.100	-0.600	-1.300	-0.600			0.964	
10	NETFLIX1/2/no	-0.240	2.800	-0.860	1.560	-0.820	-1.240	-1.200	0.000	1.566	
11	NETFLIX2/2/yes	0.240	-2.800	0.860	-1.560	0.820	1.240	1.200	0.000	1.566	
Сумма											
Среднее											
Среднеквадратичное откл.значение											

Рисунок 7. Модель INF3 (фрагмент)

Отметим, что в АСК-анализе и СК-моделях степень выраженности различных свойств объектов наблюдения рассматривается с единственной точки зрения: с точки зрения того, какое **количество информации** содержится в них о том, к каким обобщающим категориям будут принадлежать или не принадлежать эти объекты. Поэтому не играет никакой роли, в каких единицах измерения измеряются те или иные свойства объектов наблюдения, а также в каких единицах измеряются результаты влияния этих свойств, натуральных, в процентах или стоимостных. Это и есть решение проблемы сопоставимости в АСК-анализе и системе «Эйдос», отличающее их от других интеллектуальных технологий.

### Верификация статистических и системно-когнитивных моделей

Оценка достоверности моделей в системе «Эйдос» осуществляется путем решения задачи классификации объектов обучающей выборки по

обобщенным образом классов и подсчета количества истинных положительных и отрицательных, а также 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. В режиме 3.4 системы «Эйдос» изучается достоверность каждой частной модели в соответствии с этими мерами достоверности (рисунок 8).

3.4. Обобщ.форма по достов.моделей при разн.кнкт.крит. Текущая модель: "INF1"													
Наименование модели и частного критерия	Интегральный критерий	F-мера Ван Ризбергена	Сумма модел... уровней сход... истинно-поло... решений (STP)	Сумма модел... уровней сход... истинно-отриц... решений (STN)	Сумма модел... уровней сход... ложно-поло... решений (SFP)	Сумма модел... уровней сход... ложно-отриц... решений (SFN)	S-точность модели	S-Полнота модели	L1-мера проф. Е.В.Луценко	Средний модель... уровень сход... истинно-поло... решений	Средний модель... уровень сход... истинно-отриц... решений	Средний модель... уровень ложно-... решений	Средний модель... уровень ложно-... решений
1. ABS - частный критерий: количество встреч совпадений "класс... Сумма abs.частот по признакам..."	Корреляция abs.частот с обр...	0.343	25.399	41.967	70.067	0.289	0.266	0.989	0.419	0.540	0.341	0	0
1. ABS - частный критерий: количество встреч совпадений "класс... Сумма abs.частот по признакам..."	Сумма abs.частот по признакам...	0.250	35.231	122.462	122.462	0.223	0.223	1.000	0.365	0.705	0	0	0
2. PRC1 - частный критерий: усл. вероятность i-го признака сред... Сумма усл.отн.частот по признакам...	Корреляция усл.отн.частот с о...	0.343	25.399	41.967	70.067	0.289	0.266	0.989	0.419	0.540	0.341	0	0
2. PRC1 - частный критерий: усл. вероятность i-го признака сред... Сумма усл.отн.частот по признакам...	Сумма усл.отн.частот по признакам...	0.250	36.723	173.157	0.175	1.000	0.298	0.734	0.341	0	0	0	0
3. PRC2 - частный критерий: условная вероятность i-го признака... Сумма усл.отн.частот по признакам...	Корреляция усл.отн.частот с о...	0.343	25.399	41.967	70.067	0.289	0.266	0.989	0.419	0.540	0.341	0	0
3. PRC2 - частный критерий: условная вероятность i-го признака... Сумма усл.отн.частот по признакам...	Сумма усл.отн.частот по признакам...	0.250	36.723	173.157	0.175	1.000	0.298	0.734	0.341	0	0	0	0
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в... Сумма усл.отн.частот по признакам...	Корреляция abs.частот с обр...	0.376	18.660	76.941	38.665	3.410	0.326	0.845	0.470	0.504	0.405	0	0
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в... Сумма усл.отн.частот по признакам...	Сумма знаний	0.293	9.063	22.629	34.094	1.593	0.210	0.851	0.337	0.232	0.184	0	0
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в... Сумма усл.отн.частот по признакам...	Корреляция abs.частот с обр...	0.376	18.660	76.941	38.665	3.410	0.326	0.845	0.470	0.504	0.405	0	0
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в... Сумма усл.отн.частот по признакам...	Сумма знаний	0.293	9.063	22.629	34.094	1.593	0.210	0.851	0.337	0.232	0.184	0	0
6. INF3 - частный критерий: Хи-квадрат: различия между фактич... Семантический резонанс зна...	Хи-квадрат: различия между фактич...	0.408	21.666	71.909	47.884	2.333	0.312	0.903	0.463	0.504	0.395	0	0
6. INF3 - частный критерий: Хи-квадрат: различия между фактич... Сумма знаний	Хи-квадрат: различия между фактич...	0.408	14.662	41.287	26.127	1.502	0.343	0.907	0.497	0.341	0.227	0	0
7. INF4 - частный критерий: ROI (Return On Investment): вероятно... Сумма знаний	ROI (Return On Investment): вероятно...	0.400	17.595	72.730	36.292	3.276	0.327	0.843	0.471	0.463	0.367	0	0
7. INF4 - частный критерий: ROI (Return On Investment): вероятно... Семантический резонанс зна...	Сумма знаний	0.289	6.178	6.412	24.304	0.435	0.203	0.934	0.333	0.151	0.060	0	0
8. INF5 - частный критерий: ROI (Return On Investment): вероятно... Сумма знаний	ROI (Return On Investment): вероятно...	0.400	17.595	72.730	36.292	3.276	0.327	0.843	0.471	0.463	0.367	0	0
8. INF5 - частный критерий: ROI (Return On Investment): вероятно... Семантический резонанс зна...	Сумма знаний	0.289	6.178	6.412	24.304	0.435	0.203	0.934	0.333	0.151	0.060	0	0
9. INF6 - частный критерий: различия безузл.вероятностей: вер... Сумма знаний	Семантический резонанс зна...	0.398	19.975	73.657	43.698	2.791	0.314	0.877	0.462	0.512	0.382	0	0
9. INF6 - частный критерий: различия безузл.вероятностей: вер... Сумма знаний	Сумма знаний	0.303	12.104	16.175	46.112	0.888	0.208	0.932	0.340	0.281	0.148	0	0
10. INF7 - частный критерий: различия безузл.вероятностей: вер... Сумма знаний	Семантический резонанс зна...	0.398	19.975	73.657	43.698	2.791	0.314	0.877	0.462	0.512	0.382	0	0
10. INF7 - частный критерий: различия безузл.вероятностей: вер... Сумма знаний	Сумма знаний	0.303	12.104	16.175	46.112	0.888	0.208	0.932	0.340	0.281	0.148	0	0

Рисунок 8. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена и L1-критерию проф. Е.В.Луценко

Из рисунка 8 мы видим, что в данном интеллектуальном приложении по F-критерию Ван Ризбергена наиболее достоверной является СК-модель INF3 с интегральным критерием «Сумма знаний» ( $F=0,408$  при максимуме 0,907), что является довольно хорошим результатом, по критерию L1 проф. Е.В.Луценко та же модель является наиболее достоверной ( $L1=0,497$  при максимуме 0,907), что является отличным результатом.

**Это подтверждает наличие и адекватное отражение в СК-модели INF3 сильной причинно-следственной зависимости между видами жанров фильмов и их характеристиками.**

На рисунке 9 приведено частотное распределение числа истинных и ложных положительных и отрицательных решений по результатам идентификации характеристик жанров фильмов в СК-модели INF3 по данным обучающей выборки:

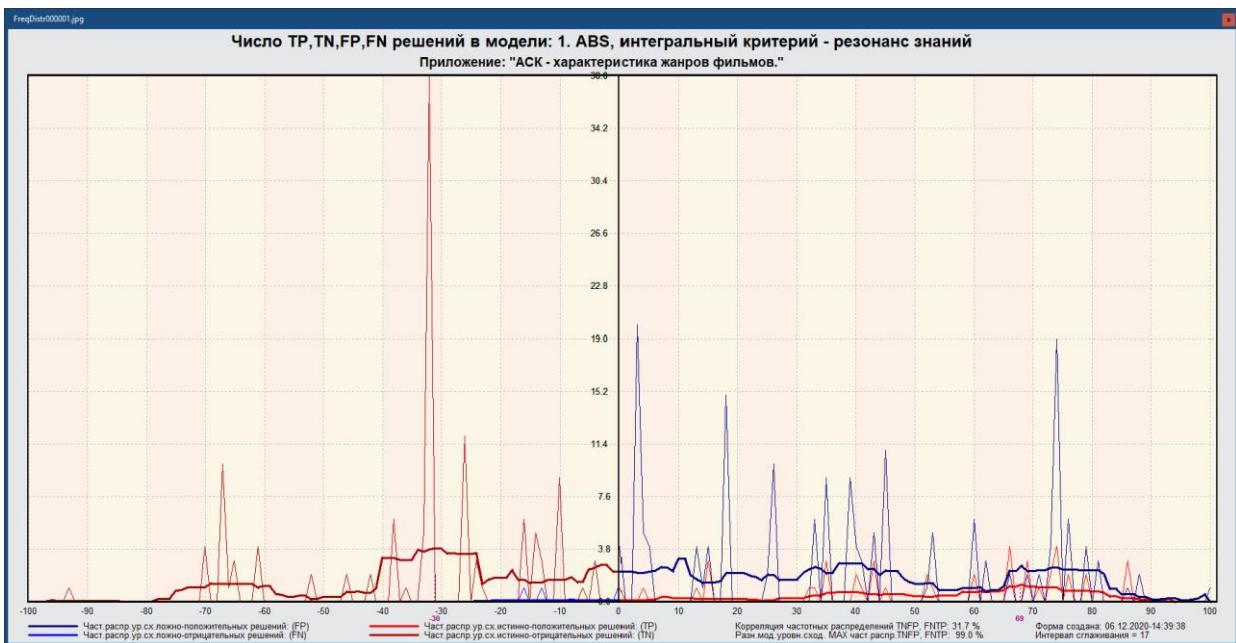


Рисунок 9. Частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности в СК-модели Inf3

Рисунок 9 содержит изображения двух частотных распределений, похожих на нормальные, сдвинутых относительно друг друга по фазе.

Левое распределение, большее по амплитуде включает истинно-отрицательные и ложно-положительные 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 жанров фильмов по его характеристикам и другие задачи.

Видно, что для отрицательных решений количество истинных решений всегда значительно превосходит количество ложных решений, причем при уровнях различия больше примерно 5% ложные отрицательные решения вообще практически отсутствуют.

Видно также, что для положительных решений картина более сложная и включает 3 диапазона уровней сходства

1) при уровнях сходства от 0% до 65% количество ложных решений больше числа истинных;

2) при уровнях сходства от 65% до примерно 95% есть и истинные и ложные положительные решения, но число истинных решений больше числа ложных и доля истинных решений возрастает при увеличении уровня сходства;

3) при уровнях сходства выше 95% ложные положительные решения не встречаются.

На рисунке 10 приведен Help по режиму 3.4, в котором описаны меры достоверности моделей, применяемые в системе «Эйдос»:

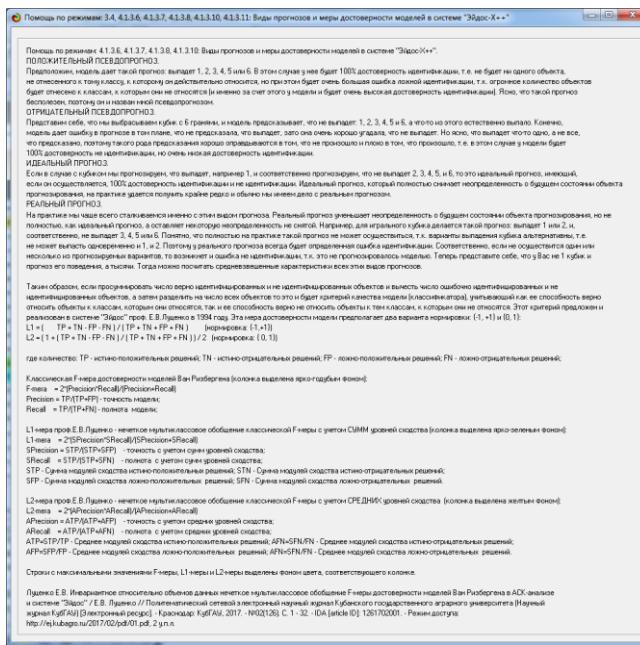


Рисунок 10. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена и L1- и L2-критериям проф. Е.В.Луценко

### Выбор наиболее достоверной модели и присвоение ей статуса текущей

В соответствии со схемой обработки данных, информации и знаний в системе «Эйдос» (рисунок 1), присвоим СК-модели INF3 статус текущей модели. Для этого запустим режим 5.6 с параметрами, приведенными на экранной форме (рисунок 11):

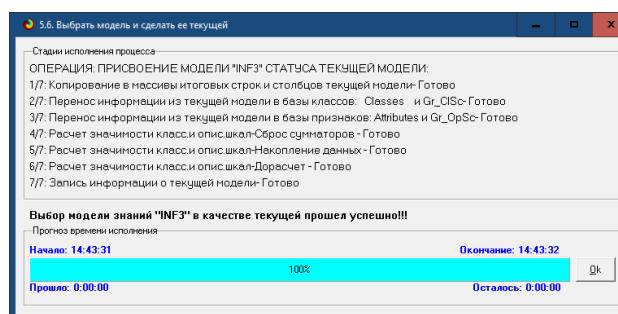
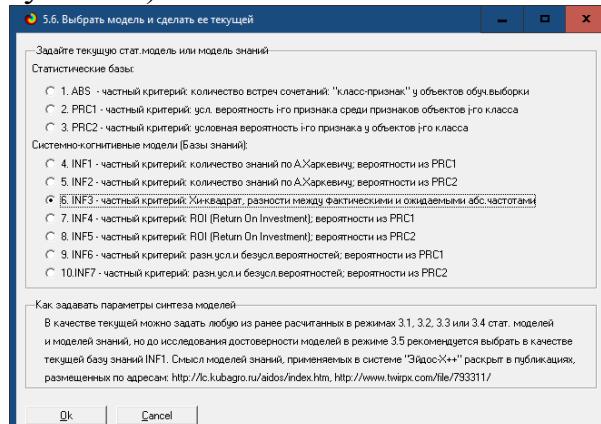


Рисунок 11. Экранные формы придания наиболее достоверной СК-модели Inf3 статуса текущей модели

## **Задача 4: решение различных задач в наиболее достоверной модели**

### **Подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация)**

Решим задачу системной идентификации, т.е. определение жанров фильмов на основе обучающей выборки в наиболее достоверной СК-модели INF3 на GPU (рисунок 12).

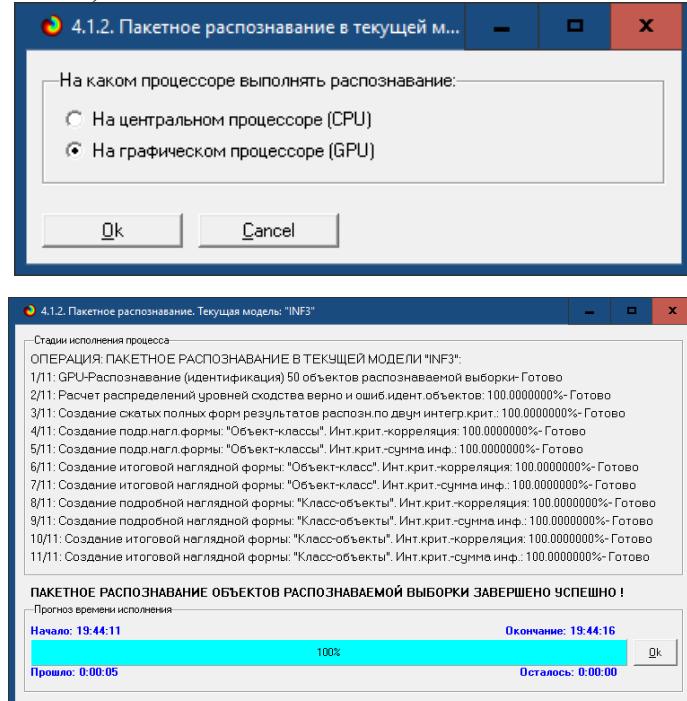


Рисунок 12. Экранные формы, которые отображают процесс решения задачи системной идентификации в текущей модели

Из рисунка 12 видно, что процесс идентификации занял 5 секунды.

Отметим, что 95% этого времени заняла не сама идентификация на GPU, а создание 11 выходных форм на основе результатов этого прогнозирования. Эти формы отражают результаты прогнозирования в различных разрезах и обобщениях:

Приведем две из этих 10 форм: 4.1.3.1 и 4.1.3.2 (рисунок 13).

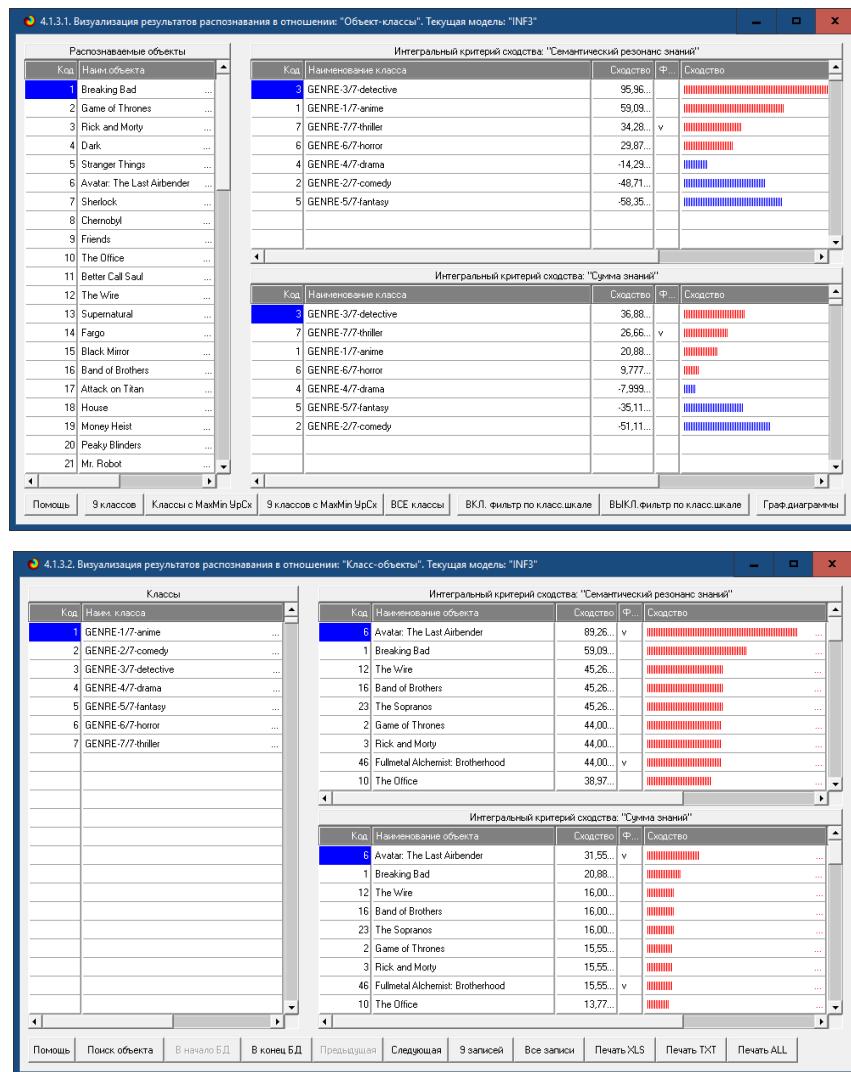


Рисунок 13. Выходные формы по результатам идентификации видов жанров по их характеристикам

Символ « $\vee$ » стоит против тех результатов идентификации, которые подтвердились на опыте, т.е. соответствуют факту. Из рисунка 13 видно, что результаты идентификации являются отличными, естественно при учете информации из рисунка 9 о том, что достоверные прогнозы в данной модели имеют уровень сходства выше 35%, т.е., по сути, результаты с более низким уровнем сходства надо просто игнорировать.

## Подзадача 4.2. Поддержка принятия решений (SWOT-анализ)

При принятии решений определяется сила и направление влияния значений факторов на принадлежность состояний объекта моделирования к тем или иным классам, соответствующим различным будущим состояниям. По сути, это решение задачи SWOT-анализа.

Применительно к задаче, решаемой в данной работе, SWOT-анализ показывает степень влияния различных значений характеристик жанров на их вид.

В системе «Эйдос» в режиме 4.4.8 поддерживается решение этой задачи. При этом **выявляется система детерминации заданного класса**, т.е. система значений факторов, обуславливающих переход объекта моделирования и управления в состояние, соответствующее данному классу, а также препятствующих этому переходу.

На рисунках 14 приведены SWOT-диаграммы наглядно отражающие силу и направление влияния различных значений характеристик жанров на их вид.

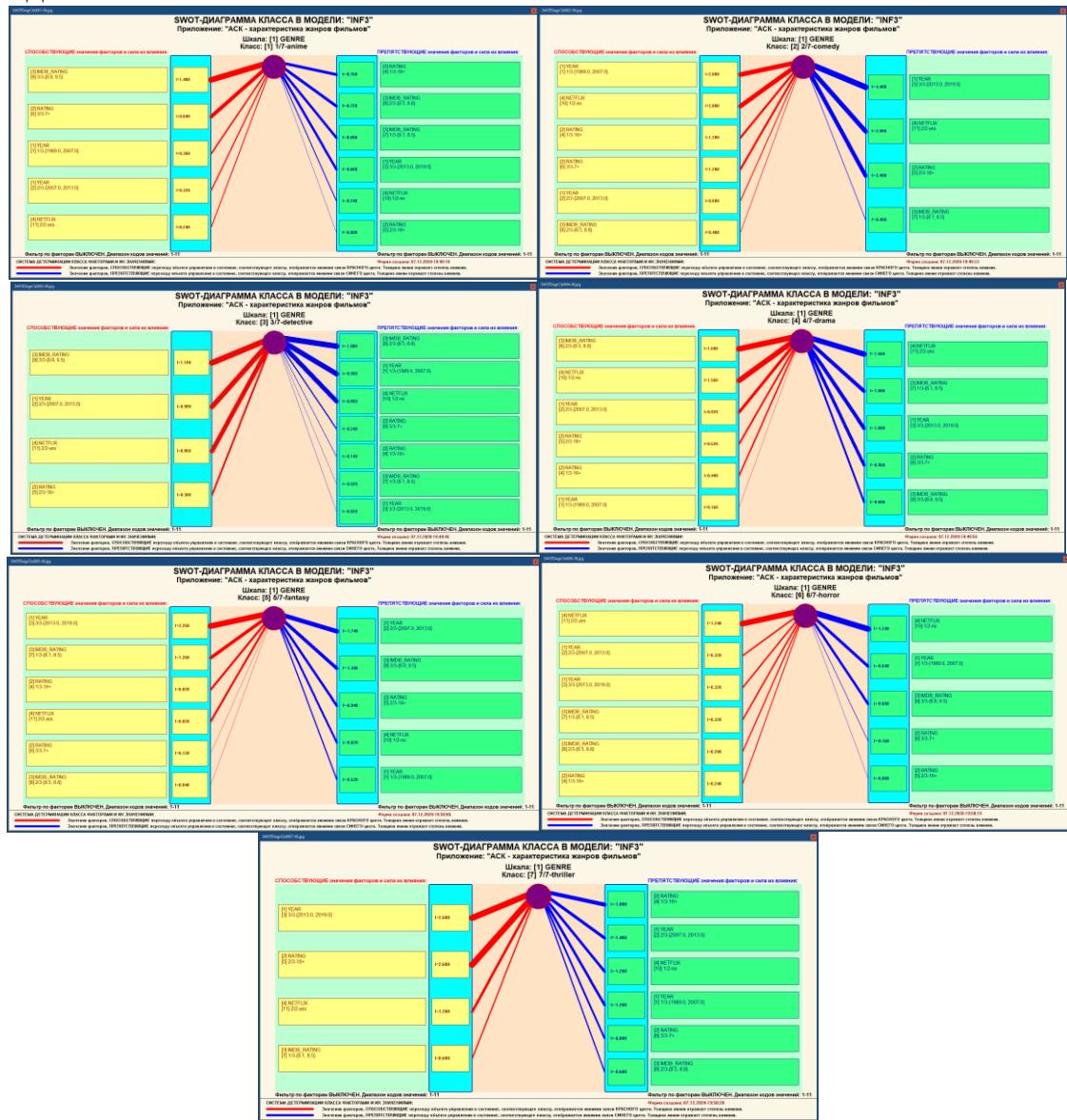


Рисунок 14. SWOT-диаграммы детерминации характеристик жанров фильмов

Эти SWOT-диаграммы наглядно отражают силу и направление влияния различных значений характеристик жанров фильмов на их вид.

Отметим также, что система «Эйдос» обеспечивала решение этой задачи *всегда*, т.е. даже в самых ранних DOS-версиях и в реализациях системы «Эйдос» на других языках и типах компьютеров. Например, первый акт внедрения системы «Эйдос», где об этом упоминается в явном виде, датируется 1987 годом, а первый подобный расчет относится к 1981 году.

Но тогда SWOT-диаграммы назывались позитивным и негативным информационными портретами классов.

Информация о системе значений факторов, обуславливающих переход объекта моделирования в различные будущие состояния, соответствующие классам, может быть приведена не только в диаграммах, показанных на рисунках 14, но и во многих других табличных и графических формах, которые в данной работе не приводятся только из-за ограниченности ее объема. В частности в этих формах может быть выведена значительно более полная информация (в т. ч. вообще вся имеющаяся в модели). Подобная подробная информация содержится в базах данных, расположенных по пути: \Aidos-X\AID\_DATA\A0000001\System\SWOTCls####Inf3.DBF, где: «####» – код класса с ведущими нулями. Эти базы открываются в MS Excel.

На рисунке 15 приведены примеры нескольких инвертированных SWOT-диаграмм, отражающих силу и направление влияния различных жанров фильмов на их характеристики.

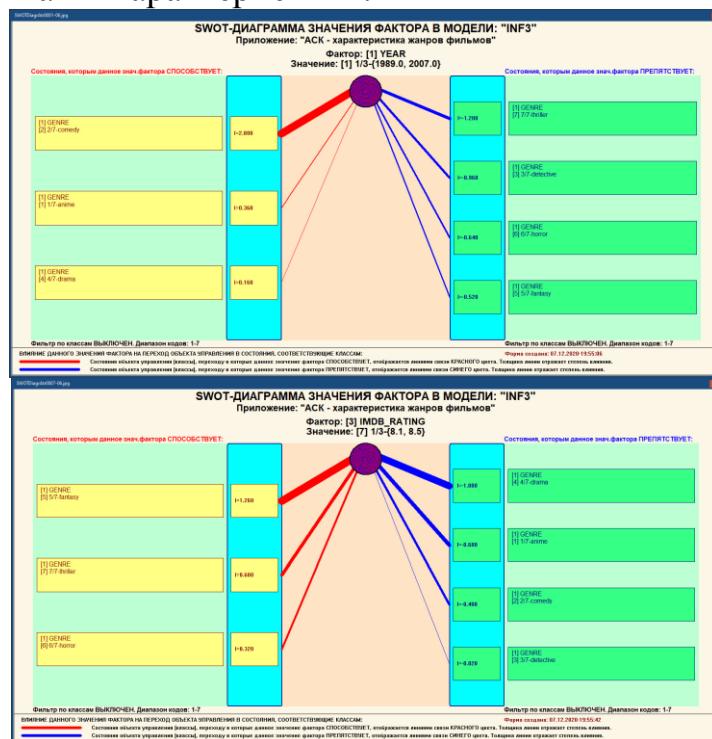


Рисунок 15. Примеры SWOT-диаграмм, отражающих силу и направление влияния различных жанров фильмов на их характеристики

Из первого примера видно, что комедии более характерен триллер и детектив.

В заключение отметим, что SWOT-анализ является широко известным и общепризнанным метод стратегического планирования. Однако это не

мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых чаще всего является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего опыта и профессиональной компетенции. Но возможности экспертов имеют свои ограничения и часто по различным причинам они не могут и не хотят это сделать. Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов. Эта проблема может решаться путем автоматизации функций экспертов, т.е. путем измерения силы и направления влияния факторов непосредственно на основе эмпирических данных. Подобная технология разработана давно, ей уже более 30 лет, но, к сожалению, она сравнительно малоизвестна – это интеллектуальная система «Эйдос».

### **Подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели**

Если модель предметной области достоверна, то исследование модели можно считать исследованием самого моделируемого объекта, т.е. результаты исследования модели корректно относить к самому объекту моделирования, «переносить на него».

В системе «Эйдос» есть довольно много возможностей для такого исследования, но в данной работе из-за ограничений на ее объем мы рассмотрим лишь результаты кластерно-конструктивного анализа классов и признаков (когнитивные диаграммы и дендрограммы), а также нелокальные нейроны, нелокальные нейронные сети, 3d-интегральные когнитивные карты и когнитивные функции.

#### **4.3.1. Когнитивные диаграммы классов**

Эти диаграммы отражают сходство/различие классов. Мы получаем их в режимах 4.2.2.1 и 4.2.2.2 (рисунок 16).

Отметим также, что на когнитивной диаграмме, приведенной на рисунке 16, показаны **количественные** оценки сходства/различия жанров фильмов по связанным с ними значениям их характеристик. Важно, что эти результаты сравнения получены с применением системно-когнитивной модели, созданной *непосредственно на основе эмпирических данных*, а не как традиционно делается на основе экспертных оценок неформализуемым путем на основе опыта, интуиции и профессиональной компетенции.

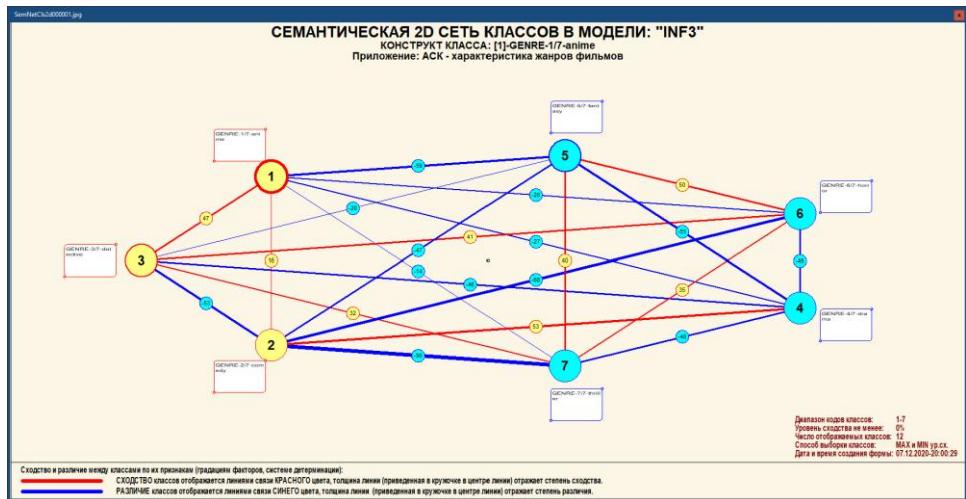


Рисунок 16. Когнитивная диаграмма классов, отражающая сходства/различия жанров фильмов по связанным с ними значениям их характеристик.

В системе «Эйдос» есть возможность управлять параметрами формирования и вывода изображения, приведенного на рисунке 16. Для этого используется диалоговое окно, приведенное на рисунке 17.

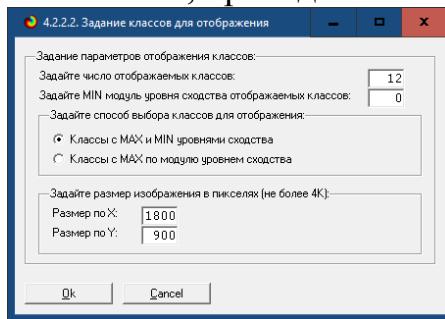


Рисунок 17. Диалоговое окно управления параметрами формирования и вывода изображения когнитивной диаграммы классов

### 4.3.2. Агломеративная когнитивная кластеризация классов

Информация о сходстве/различии классов, содержащаяся в матрице сходства, может быть визуализирована не только в форме, когнитивных диаграмм, пример которой приведен на рисунке 16, но и в форме агломеративных дендрограмм, полученных в результате **когнитивной кластеризации** (рисунок 18):

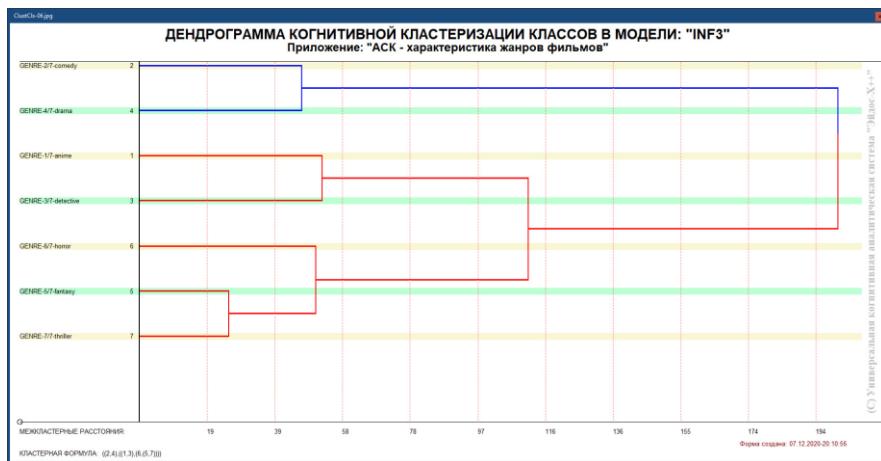


Рисунок 18. Дендрограмма когнитивной агломеративной кластеризации, отражающая сходство/различие жанров фильмов

Из рисунков 16 и 18 мы видим, что некоторые жанры фильмов сходны по детерминирующей их системе значений характеристик, и, следовательно, корректно ставить задачу их одновременного достижения, а другие по системе значений этих параметров сильно отличаются, и, следовательно, являются взаимоисключающими, т.е. альтернативными и цель их одновременного достижения является некорректной и недостижимой, т.к. для достижения одного из альтернативных результатов необходимы одни значения характеристик, а для достижения другого – совершенно другие, которые не могут наблюдаться одновременно с первыми.

Из дендрограммы когнитивной агломеративной кластеризации классов, приведенной на рисунке 18, мы видим также, что все классы образуют два противоположных кластера, являющихся полюсами конструкта, по системе значений обуславливающих значениям параметров их характеристик.

На рисунке 19 мы видим график изменения межклассовых расстояний:

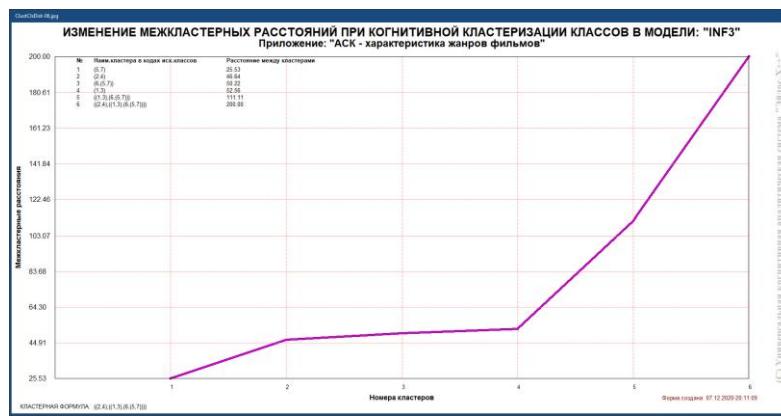


Рисунок 19. График изменения межклассовых расстояний

### 4.3.3. Когнитивные диаграммы значений факторов

Эти диаграммы отражают сходство/различие значений параметров характеристик по их смыслу, т.е. по содержащейся в них информации о

жанров фильмов. Эти диаграммы мы получаем в режимах 4.3.2.1 и 4.3.2.2 (рисунок 20).

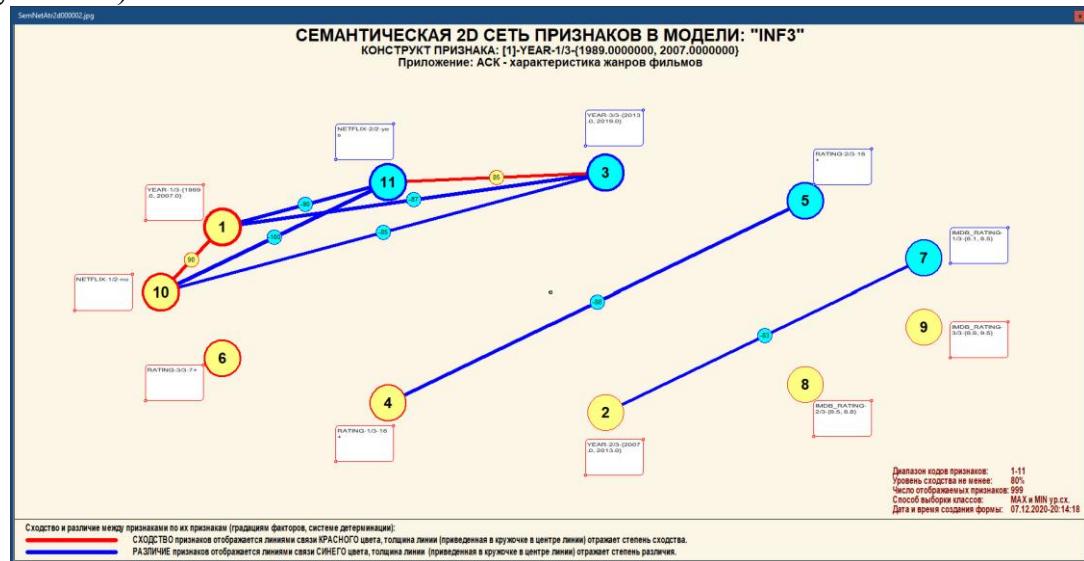


Рисунок 20. сходство/различие жанров фильмов в соответствии с их характеристиками

Из рисунка 20 видно, что все значения факторов образуют два крупных кластера, противоположных по их смыслу. Эти кластеры образуют полюса конструкта.

Отметим, что на когнитивной диаграмме, приведенной на рисунке 20, показаны **количественные** оценки сходства/различия значений факторов, полученные с применением системно-когнитивной модели, созданной *непосредственно на основе эмпирических данных*, а не как традиционно делается на основе экспертных оценок неформализуемым путем на основе опыта, интуиции и профессиональной компетенции.

Диаграмма, приведенная на рисунке 20, получена при параметрах, приведенных на рисунке 21.

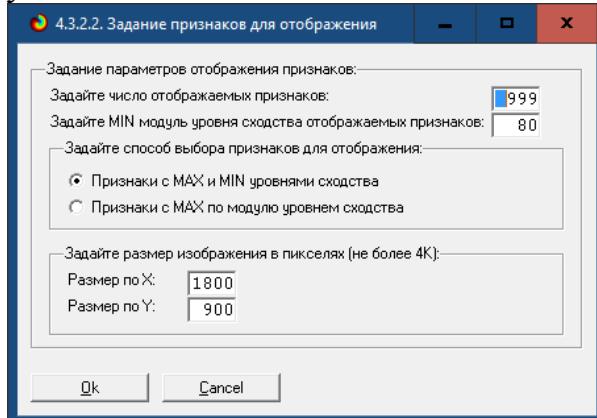


Рисунок 10. Параметры отображения когнитивной диаграммы, приведенной на рисунке 20

#### 4.3.4. Агломеративная когнитивная кластеризация значений факторов

На рисунке 22 приведена агломеративная дендрограмма когнитивной кластеризации значений факторов и график изменения межклusterных расстояний, полученные на основе той же матрицы сходства признаков по их смыслу, что и в когнитивных диаграммах, пример которой приведен на рисунке 20.

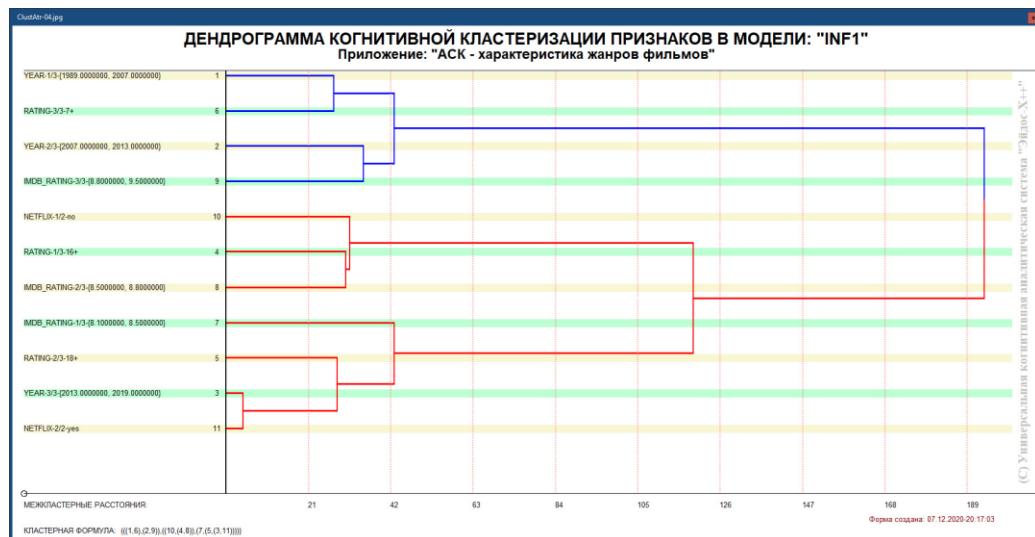


Рисунок 22. Дендрограмма агломеративной когнитивной кластеризации признаков

Из дендрограммы на рисунке 22 мы видим, что все значения факторов образуют 2 четко выраженных кластера, объединенных в полюса конструкта (показаны синими и красным цветами).

Хорошо видна группировка признаков по их смыслу, т.е. по содержащейся в них информации о жанре фильмов. **Значения факторов на полюсах конструкта факторов (рисунок 22) обуславливают переход объекта моделирования в состояния, соответствующие классам, представленным на полюсах конструкта классов (рисунки 18 и 20).**

На рисунке 23 приведен график межклusterных расстояний значений признаков.

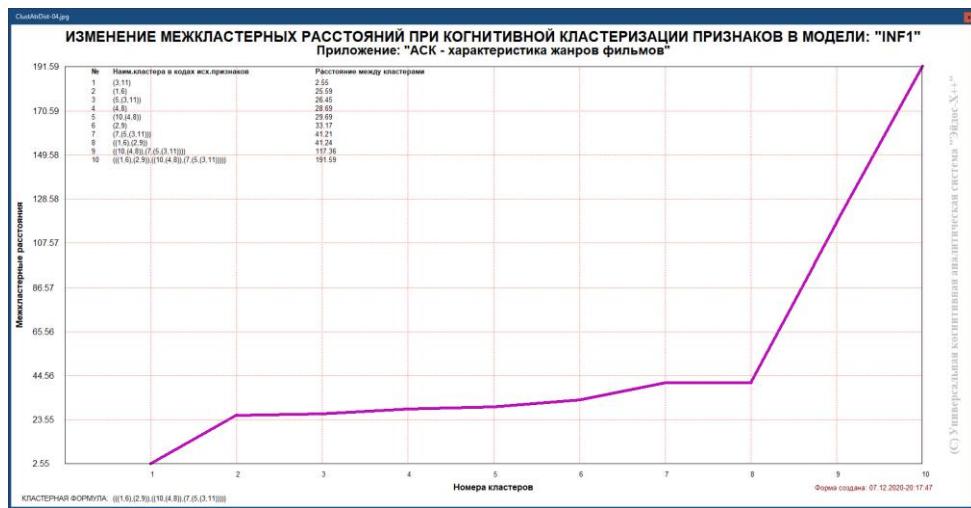


Рисунок 23. График изменения межклusterных расстояний при когнитивной кластеризации значений факторов

#### 4.3.5. Нелокальные нейроны и нелокальные нейронные сети

На рисунке 24 приведён пример нелокального нейрона, а на рисунке 25 - фрагмент одного слоя нелокальной нейронной сети:

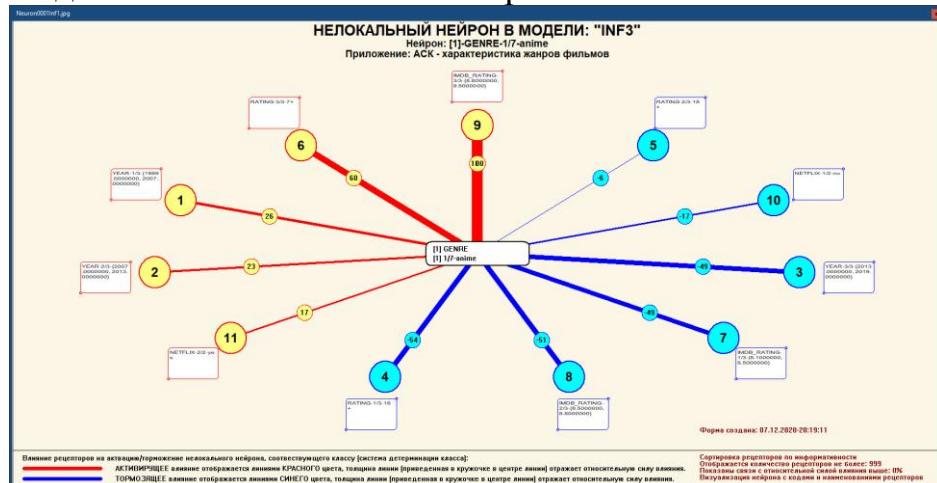


Рисунок 24. Пример нелокального нейрона, отражающего силу и направление влияния значений жанров фильмов на их вид

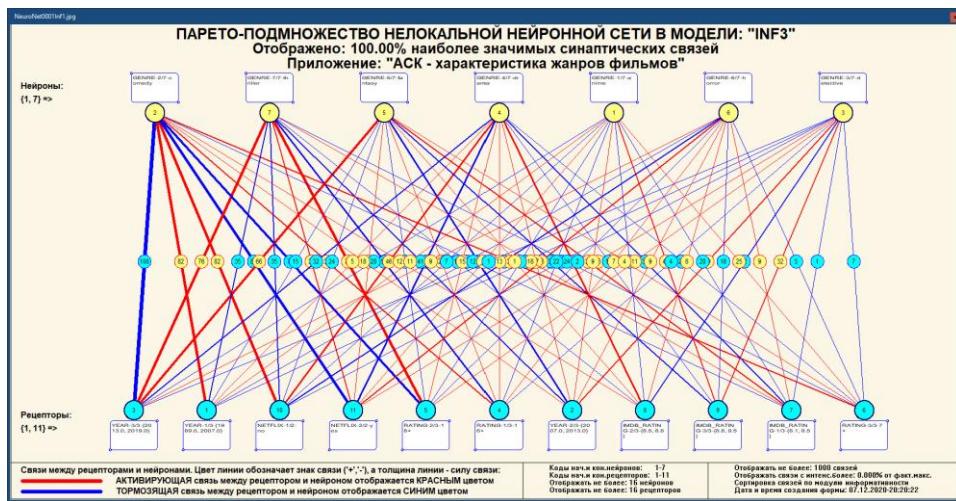


Рисунок 25. Один слой нелокальной нейронной сети, отражающий силу и направление влияния значений жанров фильмов на их вид (фрагмент 100% сети)

В приведенном фрагменте слоя нейронной сети нейроны соответствуют жанров фильмов, а рецепторы – их характеристикам. Нейроны расположены слева направо в порядке убывания силы детерминации, т.е. слева находятся результаты, наиболее жестко обусловленные обуславливающими их значениями факторами, а справа – менее жестко обусловленные.

Модель знаний системы «Эйдос» относится к ***нечетким декларативным*** гибридным моделям и объединяет в себе некоторые 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.

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

- 1) весовые коэффициенты на рецепторах не подбираются итерационным методом обратного распространения ошибки, а считаются прямым счетом 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».

#### 4.3.6. 3d-интегральные когнитивные карты

На рисунке 26 приведен фрагмент 3d-интегральной когнитивной карты, отражающий фрагмент около 100% СК-модели Inf3.

3d-интегральная когнитивная карта является отображением на одном рисунке когнитивных диаграмм классов и значений факторов, отраженных соответственно на рисунках 16 и 20, и одного слоя нейронной сети, приведенного на рисунке 25.

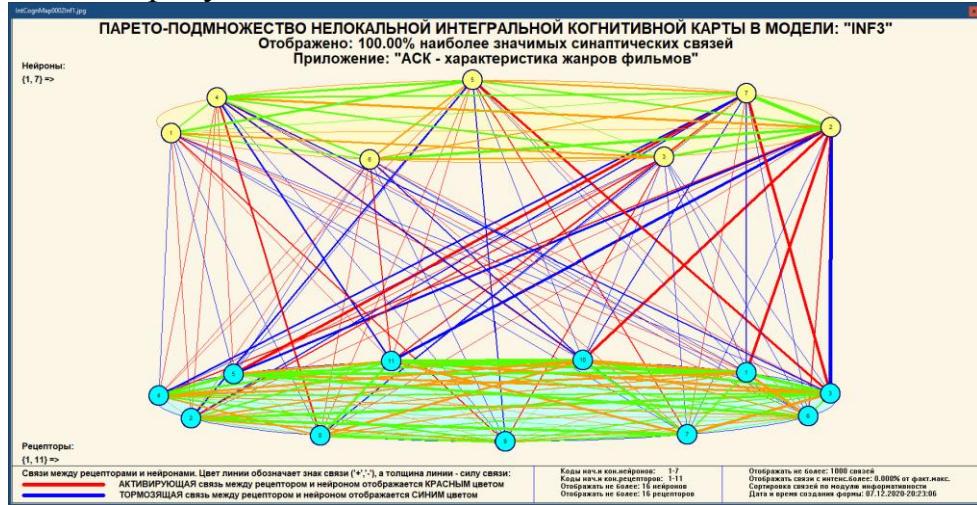


Рисунок 26. 3d-интегральная когнитивная карта в СК-модели Inf3

#### 4.3.7. Когнитивные функции

Вместо описания того, что представляют собой когнитивные функции, приведем help соответствующего режима системы «Эйдос» (рисунок 27).

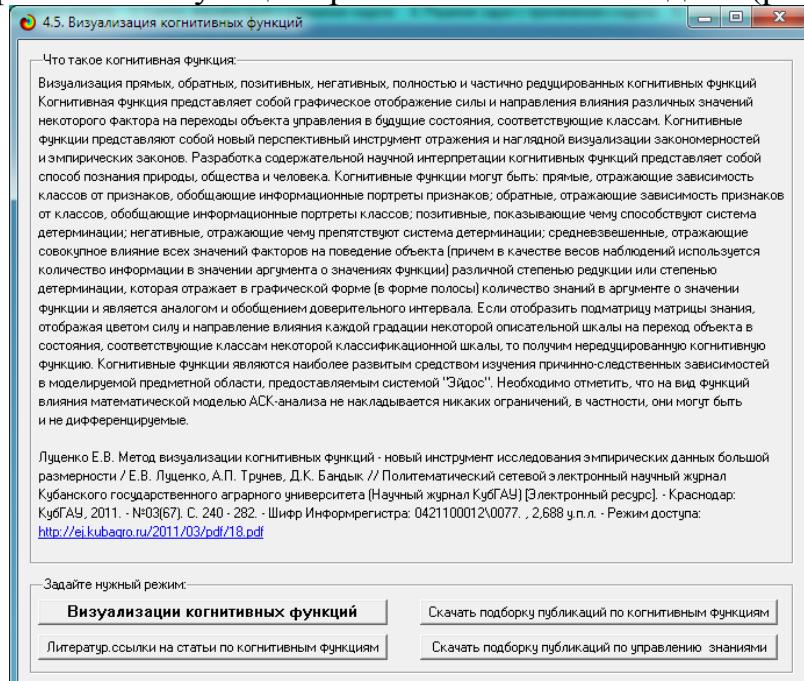


Рисунок 27. Help режима визуализации когнитивных функций

Когнитивная функция представляет собой графическое отображение силы и направления влияния различных значений некоторого фактора (признаков) на переходы объекта управления в будущие состояния, соответствующие классам. Классы являются градациями классификационных шкал.

Когнитивные функции представляют собой новый перспективный инструмент отражения и наглядной визуализации эмпирических закономерностей и эмпирических законов. Разработка содержательной научной интерпретации когнитивных функций представляет собой способ познания природы, общества и человека.

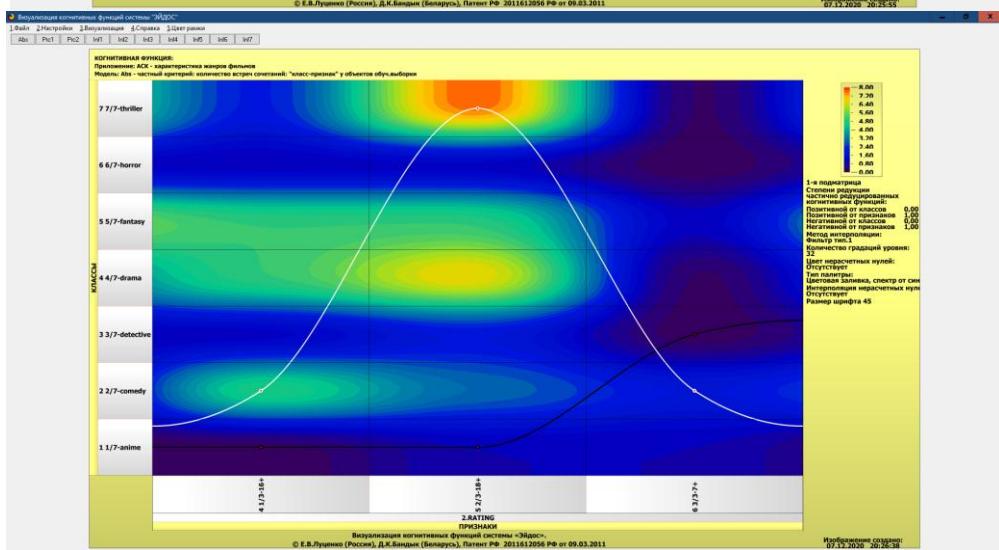
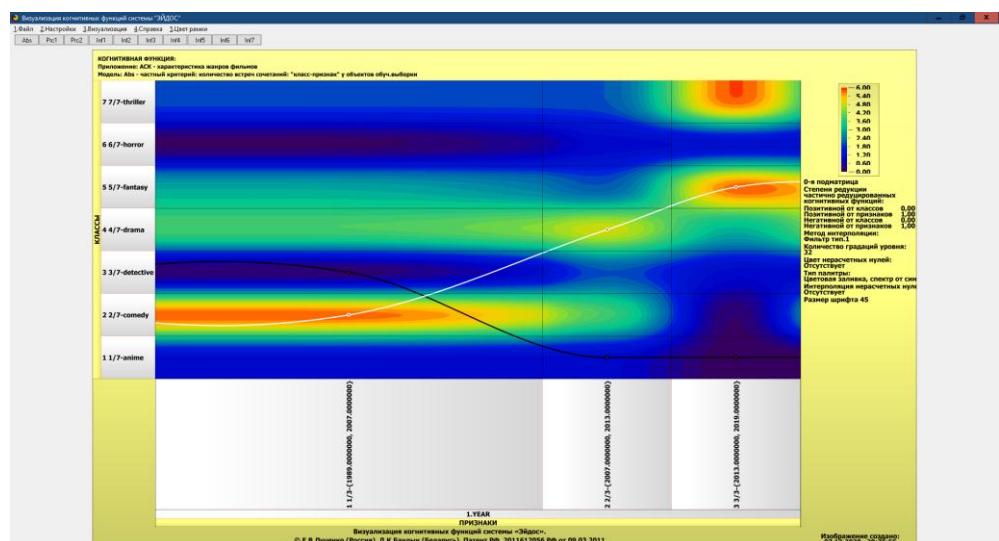
Когнитивные функции могут быть: прямые, отражающие зависимость классов от признаков, обобщающие информационные портреты признаков; обратные, отражающие зависимость признаков от классов, обобщающие информационные портреты классов; позитивные, показывающие чему способствуют система детерминации (обозначены белой линией); негативные, отражающие чему препятствуют система детерминации (обозначены черной линией); средневзвешенные, отражающие совокупное влияние всех значений факторов на поведение объекта (причем в качестве весов наблюдений используется количество информации в значении аргумента о значениях функции) различной степенью редукции или степенью детерминации, которая отражает в графической форме (в форме полосы разной толщины) количество знаний в аргументе о значении функции и является аналогом и обобщением доверительного интервала.

Если отобразить подматрицу матрицы знания, отображая цветом силу и направление влияния каждой градации некоторой описательной шкалы на переход объекта в состояния, соответствующие классам некоторой классификационной шкалы, то получим нередуцированную когнитивную функцию.

Когнитивные функции являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой "Эйдос".

Необходимо отметить, что **на вид функций влияния математической моделью АСК-анализа не накладывается никаких ограничений**, в частности, они могут быть и не дифференцируемые.

На рисунках 28 приведены когнитивные функции, наглядно отражающие силу и направление влияния значений (т.е. степени выраженности) различных жанров фильмов на их вид (класс).



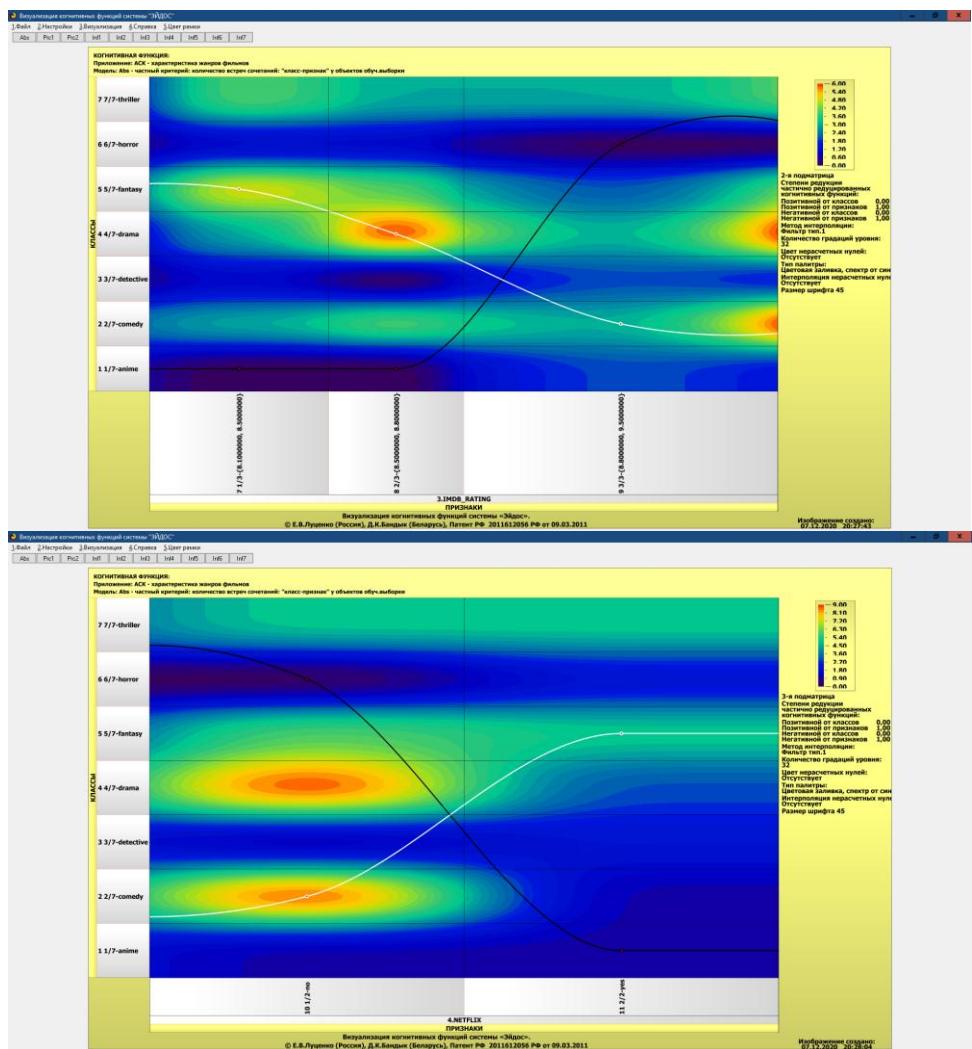


Рисунок 28. Примеры когнитивных функций в СК-модели INF3, отражающих силу и направление влияния значений жанров фильмов на их вид

Из когнитивных функций, приведенных на рисунке 28, хорошо видно, что **зависимости между значениями жанров фильмов и их классом в основном имеют довольно предсказуемый характер.**

#### 4.3.8. Сила и направление влияния значений жанров фильмов и сила влияния этих характеристик на жанры фильмов

На рисунках 6, 7 приведены фрагменты некоторых статистических и системно-когнитивных моделей, отражающих моделируемую предметную область.

Строки матриц моделей соответствуют значениям факторов, т.е. значениям жанров фильмов (градации описательных шкал).

Колонки матриц моделей соответствуют различным классам, отражающим жанры фильмов (градации классификационных шкал).

Числовые значения в ячейках матриц моделей, находящихся на пересечении строк и колонок, отражают направление (знак) и силу влияния

конкретного значения жанра фильма, соответствующего сроке, на конкретное значение его класса (вида).

Если какое-то значение характеристики слабо влияет на жанр фильма, то в соответствующей строке матрицы модели будут малые по модулю значения разных знаков, если же влияние сильное – то и значения будут большие по модулю разных знаков.

Если значение какой-либо характеристики способствует получению некоторого определенного жанра фильма, то в соответствующей этому результату ячейке матрицы модели будут положительные значения, если же понижает – то и значения будут отрицательные.

Из этого следует, что суммарную силу влияния того или иного значения жанров фильмов на его класс (т.е. ценность данного значения характеристики для решения задачи квалиметрии и других задач) можно количественно оценивать *степенью вариабельности значений* в строке матрицы модели, соответствующей этому значению жанров фильмов.

Существует много мер вариабельности значений: это и среднее модулей отклонения от среднего, и дисперсия, и среднеквадратичное отклонение и другие. В АСК-анализе и системе «Эйдос» для этой цели принято использовать среднеквадратичное отклонение. Численно оно равно стандартному отклонению и вычисляется по той же формуле, но мы предпочитаем не использовать термин «стандартное отклонение», т.к. он предполагает нормальность распределения исследуемых последовательностей чисел, а значит и проверку соответствующих статистических гипотез.

Самая правая колонка в матрицах моделей на рисунках 6, 7 содержит количественную оценку вариабельности значений строки модели (среднеквадратичное отклонение), которая и представляет собой ценность значения жанров фильмов, соответствующего строке, для решения задачи квалиметрии и других задач, рассмотренных в работе.

Если рассортировать матрицу модели по этой самой правой колонке в порядке убывания, а потом просуммировать значения в ней нарастающим итогом, то получим логистическую Парето-кривую, отражающую зависимость ценности модели от числа наиболее ценных признаков в ней (рисунок 29, таблица 6).

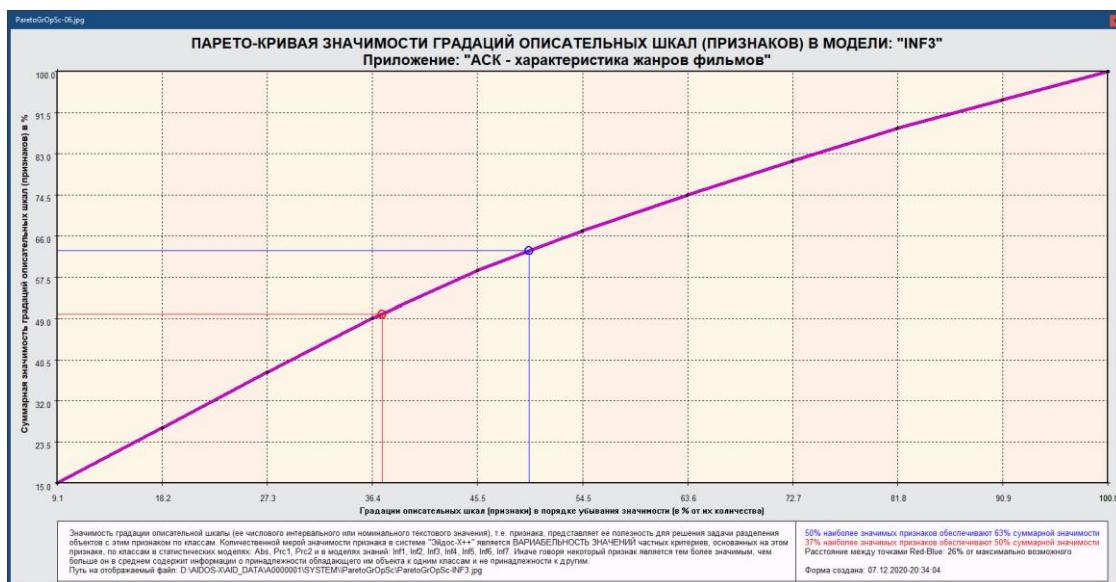


Рисунок 29. Парето-кривая значимости градаций описательных шкал

Таблица 6 – Парето-таблица значимости градаций описательных шкал, т.е. сила влияния значений характеристик жанров фильмов на их класс в СК-модели INF3

№	Код	Название признака	Код описательных шкал	Значимость, %	Значимость нарастающим итогом
1	3	YEAR-3/3-{2013.0000000, 2019.0000000}	1	14,9881887	14,9881887
2	10	NETFLIX-1/2-no	4	11,4647989	26,4529876
3	11	NETFLIX-2/2-yes	4	11,4647989	37,9177865
4	5	RATING-2/3-18+	2	11,1129832	49,0307697
5	1	YEAR-1/3-{1989.0000000, 2007.0000000}	1	9,9270982	58,9578679
6	2	YEAR-2/3-{2007.0000000, 2013.0000000}	1	8,1061764	67,0640443
7	4	RATING-1/3-16+	2	7,4408848	74,5049291
8	9	IMDB_RATING-3/3-{8.8000000, 9.5000000}	3	7,0585881	81,5635172
9	8	IMDB_RATING-2/3-{8.5000000, 8.8000000}	3	6,7529286	88,3164458
10	7	IMDB_RATING-1/3-{8.1000000, 8.5000000}	3	5,8549221	94,1713678
11	6	RATING-3/3-7+	2	5,8286322	100,0000000

Из рисунка 29 и таблицы 6 видно, что 50% наиболее ценных для решения задачи идентификации характеристик жанров фильмов обуславливают 63% суммарной ценности, а 50% суммарной ценности обеспечиваются 37% наиболее ценных характеристик жанров фильмов.

Обращаем внимание, что наиболее ценным является год выхода фильма, а наименее ценным – ограничение возраста в фильме, причем наименее ценный фактор является бессмысленным.

Из таблицы 6 видно, что наиболее сильное влияние на жанр фильма оказывают значения характеристик:

1. YEAR-3/3-{2013.0000000, 2019.0000000}
2. NETFLIX-1/2-no
3. NETFLIX-2/2-yes
4. RATING-2/3-18+
5. YEAR-1/3-{1989.0000000, 2007.0000000}

а наиболее низкое:

1. RATING-1/3-16+
2. IMDB\_RATING-3/3-{8.8000000, 9.5000000}
3. IMDB\_RATING-2/3-{8.5000000, 8.8000000}
4. IMDB\_RATING-1/3-{8.1000000, 8.5000000}
5. RATING-3/3-7+

Ценность же характеристик жанров фильмов (всей описательной шкалы или фактора), для решения этих задач можно количественно оценивать как среднее от ценности значений этого параметра (таблица 7).

Таблица 7 – Парето-таблица значимости описательных шкал, т.е. сила влияния характеристик жанров фильмов на их класс в СК-модели INF3

№	Код	Название описательной шкалы	Значимость, %	Значимость нарастающим итогом
1	4	NETFLIX	30,8567342	30,8567342
2	1	YEAR	29,6250156	60,4817498
3	2	RATING	21,8746191	82,3563688
4	3	IMDB_RATING	17,6436312	100,0000000

Из таблицы 7 видно, что наиболее сильное влияние на жанр фильма оказывают Netflix и ограничение возраста в фильме, а наиболее низкое – рейтинг фильма и когда вышел фильм.

#### 4.3.9. Степень детерминированности жанров фильмов

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается **степенью вариабельности значений факторов** (градаций описательных шкал) в колонке матрицы модели, соответствующей данному классу (таблица 8).

В данной работе у нас классами являются жанры фильмов, а значениями градаций описательных шкал – их характеристики.

На рисунке 30 мы видим Парето-кривую степени детерминированности классов значениями характеристик нарастающим итогом.

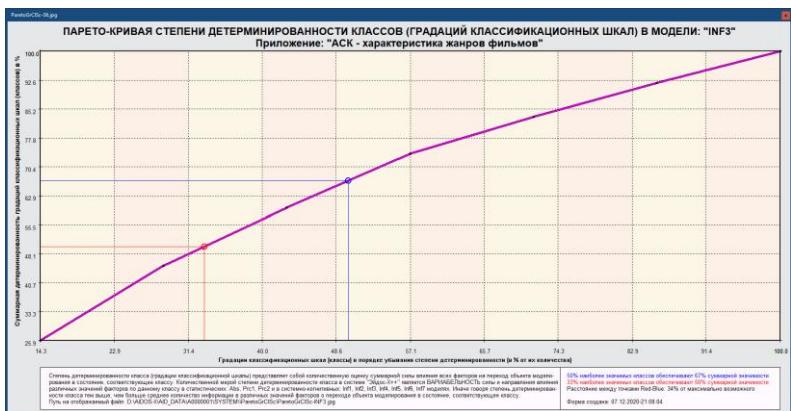


Рисунок 30. Парето-кривая степени детерминированности классов

Таблица 8 – Парето-таблица степеней детерминированности (обусловленности) жанров фильмов в СК-модели INF3

№	Код класса	Название класса	Значимость, %	Значимость нарастающим итогом
1	2	GENRE-2/7-comedy	25,8830437	25,8830437
2	7	GENRE-7/7-thriller	19,1875160	45,0705597
3	5	GENRE-5/7-fantasy	14,8422684	59,9128281
4	4	GENRE-4/7-drama	13,8101174	73,7229455
5	3	GENRE-3/7-detective	9,4830273	83,2059728
6	1	GENRE-1/7-anime	8,7204548	91,9264276
7	6	GENRE-6/7-horror	8,0735724	100,0000000

Из таблицы 8 мы видим, что значения характеристик наиболее сильно детерминируют (обуславливают) такие жанры фильмов, как комедии (comedy) и триллеры (thriller), а наиболее слабо – аниме (anime) и хоррор (horror). При этом степень детерминированности наиболее и наименее детерминированных классов отличается в три раза, что довольно существенно.

Чем выше степень детерминированности жанров фильмов значениями их характеристик, тем легче определить этот класс по параметрам жанра.

## 7. Выводы

Как показывает анализ результатов численного эксперимента предложенное и реализованное в системе «Эйдос» решение поставленных задач является вполне эффективным, что позволяет обоснованно утверждать, что цель работы достигнута, поставленная проблема решена.

В результате проделанной работы, с помощью системы «Эйдос» были созданы 3 статистические и 7 системно-когнитивных моделей, в которых непосредственно на основе эмпирических данных сформированы обобщенные образы классов по различным жанрам фильмов, изучено влияние характеристик жанров фильмов на эти классы, и, на основе этого, решены задачи идентификации, классификации и исследования моделируемой предметной области путем исследования ее модели.

## **Список литературы**

1. 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. <http://elibrary.ru/item.asp?id=18632909>
2. Луценко Е.В. Универсальный информационный вариационный принцип развития систем / Е.В. Луценко // Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 117 – 193. – Шифр Информрегистра: 0420800012\0091, IDA [article ID]: 0410807010. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/10.pdf>, 4,812 у.п.л.
3. Lutsenko E. V. System analysis and decision-making (Automated system-cognitive analysis and solving problems of identification, decision-making and research of the simulated subject area): textbook / E. V. Lutsenko. - Krasnodar: ECSC "Eidos", 2020. - 1031 p. // August 2020, DOI: [10.13140/RG.2.2.27247.05289](https://doi.org/10.13140/RG.2.2.27247.05289), License [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/), [https://www.researchgate.net/publication/343998862 SYSTEM ANALYSIS AND DECISION MAKING Automated system-cognitive analysis and solving problems of identification decision-making and research of the simulated subject area](https://www.researchgate.net/publication/343998862_SYSTEM_ANALYSIS_AND_DECISION_MAKING_Automated_system-cognitive_analysis_and_solving_problems_of_identification_decision-making_and_research_of_the_simulated_subject_area), см. учебный вопрос-2.8.5. Повышение уровня системности объекта управления как цель управления.
4. Lutsenko E.V. On higher forms of consciousness, the prospects of man, technology and society (selected works) // August 2019, DOI: [10.13140/RG.2.2.21336.24320](https://doi.org/10.13140/RG.2.2.21336.24320), License [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/), [https://www.researchgate.net/publication/335057548 On HIGHER FORMS of CONSCIOUSNESS the PROSPECTS of MAN TECHNOLOGY AND SOCIETY selected works](https://www.researchgate.net/publication/335057548_On_HIGHER_FORMS_of_CONSCIOUSNESS_the_PROSPECTS_of_MAN_TECHNOLOGY_AND_SOCIETY_selected_works)
5. Lutsenko E.V. ABOUT THE INTERFACE: "SOUL-COMPUTER» (artificial intelligence: problems and solutions within the system information and functional paradigm of society development) // April 2019, DOI: [10.13140/RG.2.2.23132.85129](https://doi.org/10.13140/RG.2.2.23132.85129), [https://www.researchgate.net/publication/332464278 ABOUT THE INTERFACE SOUL-COMPUTER artificial intelligence problems and solutions within the system information and functional paradigm of society development](https://www.researchgate.net/publication/332464278_ABOUT_THE_INTERFACE_SOUL-COMPUTER_artificial_intelligence_problems_and_solutions_within_the_system_information_and_functional_paradigm_of_society_development)
6. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. <http://elibrary.ru/item.asp?id=21358220>
7. Луценко Е.В. Количественные меры возрастания эмерджентности в процессе эволюции систем (в рамках системной теории информации) / Е.В. Луценко // Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2006. – №05(021). С. 355 – 374. – Шифр Информрегистра: 0420600012\0089, IDA [article ID]: 0210605031. – Режим доступа: <http://ej.kubagro.ru/2006/05/pdf/31.pdf>, 1,25 у.п.л.
8. Луценко Е.В. Существование, несуществование и изменение как эмерджентные свойства систем // Квантовая Магия, том 5, вып. 1, стр. 1215-1239, 2008. <http://quantmagic.narod.ru/volumes/VOL512008/p1215.html>
9. Луценко Е.В. Проблемы и перспективы теории и методологии научного познания и автоматизированный системно-когнитивный анализ как автоматизированный метод научного познания, обеспечивающий содержательное феноменологическое моделирование / Е.В. Луценко // Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №03(127). С. 1 – 60. – IDA [article ID]: 1271703001. – Режим доступа: <http://ej.kubagro.ru/2017/03/pdf/01.pdf>, 3,75 у.п.л.
10. Сайт проф.Е.В.Луценко: <http://lc.kubagro.ru/>

11. Страница Е.В.Луценко: [https://www.researchgate.net/profile/Eugene\\_Lutsenko](https://www.researchgate.net/profile/Eugene_Lutsenko)
12. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-Х++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 у.п.л.
13. ПАМЯТИ ПРОФЕССОРА О.Г. КУКОСЯНА // – Режим доступа: <https://www.elibrary.ru/item.asp?id=20266263>
14. Луценко Е.В. Формирование субъективных (виртуальных) моделей физической и социальной реальности сознанием человека и неоправданное приятие им онтологического статуса (гипостазирование) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №09(113). С. 1 – 32. – IDA [article ID]: 1131509001. – Режим доступа: <http://ej.kubagro.ru/2015/09/pdf/01.pdf>, 2 у.п.л.
15. Луценко Е.В. Принципы и перспективы корректной содержательной интерпретации субъективных (виртуальных) моделей физической и социальной реальности, формируемых сознанием человека / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №01(115). С. 22 – 75. – IDA [article ID]: 1151601003. – Режим доступа: <http://ej.kubagro.ru/2016/01/pdf/03.pdf>, 3,375 у.п.л.
16. Луценко Е.В. Сценарный АСК-анализ как метод разработки на основе эмпирических данных базисных функций и весовых коэффициентов для разложения в ряд функции состояния объекта или ситуации по теореме А.Н.Колмогорова (1957) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №07(161). С. 76 – 120. – IDA [article ID]: 1612007009. – Режим доступа: <http://ej.kubagro.ru/2020/07/pdf/09.pdf>, 2,812 у.п.л.
17. 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>, 1,562 у.п.л.
18. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.
19. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.