

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 курсовой проект (работу)

Студента (ки) Маляревского Александра Владимировича
курса 3 очной (заочной) формы обучения
направления подготовки Информационные системы и технологии
направленность(профиль) Интеллектуальные информационные системы и технологии.

Наименование темы «Автоматизированный системно-когнитивный анализ рейтинга фильмов сайта IMDB на основе данных репозитория».

Рецензент: Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор
(Ф.И.О., ученое звание и степень, должность)

Оценка качества выполнения курсового проекта (работы)

/п	Показатель	Оценка соответствия заданию (по 5-и балльной шкале)
1.	Актуальность тематики работы	5
2.	Степень полноты обзора состояния проблемы и корректность постановки цели и задач исследования	5
3.	Уровень и корректность использования в работе различных методов исследований	4
4.	Степень комплексности работы, применения в ней знаний общепрофессиональных и специальных дисциплин	5
5.	Ясность, четкость, последовательность и обоснованность изложения	4
6.	Применение современных технологий обработки информации	5
7.	Качество оформления работы (общий уровень грамотности, стиль изложения, качество иллюстраций, соответствие требованиям по оформлению)	4
8.	Ответы на вопросы при защите	5

Достоинства работы

Актуальность темы применения современных математических моделей и программного инструментария для проведения АСК-анализа на основе базы данных

Недостатки работы

Отдельные погрешности в оформлении

Итоговая оценка при защите 5 (отлично)

Рецензент 

Е.В. Луценко

« » 201

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ
ФЕДЕРАЦИИ

ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ
УНИВЕРСИТЕТ
ИМЕНИ И.Т. ТРУБИЛИНА»

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

к курсовой работе

по дисциплине: ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ
на тему:

**Автоматизированный системно-когнитивный анализ рейтинга
фильмов сайта IMDB на основе данных репозитория**

выполнил студент группы ИТ1421 **Маляревский Александр
Владимирович**

Допущен к защите _____

Руководитель проекта _____ Луценко Евгений Вениаминович,
д.э.н., к.т.н., профессор
(подпись, расшифровка подписи)

Нормоконтролер _____ Николаева Ирина Валентиновна,
к.т.н., доцент
(подпись, расшифровка подписи)

Защищена _____ Оценка _____

(дата)

Члены комиссии _____ В.И. Лойко
_____ Е.В. Луценко
_____ И. В. Николаева
(подпись, дата, расшифровка подписи)

Краснодар

2017 г.

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ И.Т. ТРУБИЛИНА»

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

УТВЕРЖДАЮ:

Зав. кафедрой КТС _____ В. И. Лойко

**ЗАДАНИЕ на
курсовую работу**

Студенту: ИТ1401 группы 3 курса
Факультета прикладной информатики
Специальности: 09.03.02 Информационные системы и технологии
(шифр)

Маляревскому Александру Владимировичу
(Ф.И.О.)

Тема проекта: **Автоматизированный системно-когнитивный
анализ рейтинга фильмов сайта IMDb на основе данных
репозитория**

Содержание задания: Проанализировать методы формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами, принятия решений и исследования моделируемой предметной области путем исследования ее модели

Объем работы:

а) пояснительная записка к работе _____ листа формата А4 б)
графическая часть _____ лист формата А4

Рекомендуемая литература: Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для бакалавриата. 7-е изд., перераб. и доп.- Краснодар: КубГАУ – 2016, – 615 с., в электронном виде на сайте автора: <http://lc.kubagro.ru/aidos/p14.htm>

Срок выполнения проекта: с “ ” _____ по “ ” _____ 2017 г.

Срок защиты: _____ “ ” _____ 2017 г.

Дата выдачи задания: _____ “ ” _____ 2017 г.

Дата сдачи проекта на кафедру: _____ “ ” _____ 2017 г.

Руководитель проекта: _____ Луценко Евгений Вениаминович,

д.э.н., к.т.н., профессор
(подпись, Ф.И.О., звание, степень)

Задание принял студент _____ “ ” _____ 2017 г.

(подпись, дата)

Краснодар
2017 г.

СОДЕРЖАНИЕ

Введение.....	5
1. Синтез и верификация моделей.....	6
1.1. Описание решения.....	6
1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel.....	6
1.3. Ввод выборки в систему Aidos-X.....	8
1.4. Синтез и верификация статистических и интеллектуальных моделей.....	12
1.5. Виды моделей системы «Aidos».....	13
1.5. Результаты верификации моделей.....	15
2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	18
2.1. Выделение основных задач.....	18
2.2. Выбор модели и пакетное распознавание.....	18
2.3 Визуализация когнитивных функций.....	22
2.4 SWOT и PEST матрицы и диаграммы.....	24
2.5 Нелокальные нейронные сети и нейтроны.....	27
Заключение.....	29
Список используемых источников.....	30

ВВЕДЕНИЕ

Создание систем искусственного интеллекта является одним из важных и перспективных направлений развития современных информационных технологий. Так как существует множество альтернатив систем искусственного интеллекта, то возникает необходимость оценки качества математических моделей этих систем. В данной работе рассмотрен автоматизированный системно-когнитивный анализ рейтинга фильмов сайта IMDb на основе данных репозитория UCI.

Для достижения поставленной цели необходимы свободный доступ к тестовым исходным данным и методика, которая поможет преобразовать эти данные в форму, которая необходима для работы в системе искусственного интеллекта. Сборник баз данных Kaggle - это то что нам нужно.

В курсовой работе использована база данных «movie dataset» из банка исходных данных по задачам искусственного интеллекта – репозитория Kaggle.

Для решения задачи используем стандартные возможности Microsoft Office Word и Excel, а также систему искусственного интеллекта "Aidos-X++".

1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

1.1. Описание решения

В соответствии с методологией АСК-анализа решение поставленной задачи проведем в четыре этапа:

1. Преобразование исходных данных из csv-формата в промежуточные файлы MS Excel.
2. Преобразование исходных данных из промежуточных файлов MS Excel в базы данных системы "Aidos".
3. Синтез и верификация моделей предметной области.
4. Применение моделей для решения задач идентификации, прогнозирования и исследования предметной области.

1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel

С электронного ресурса kaggle.com возьмем базу данных movie dataset.

В этом файле содержится данные о 5000 лучших фильмов по версии IMDB. Он содержит:

- Название фильма.
- Проголосовавшие пользователи.
- Контент-рейтинг.
- Цвет фильма.
- Страна.
- Кинорежиссер.
- Главный актер.
- Жанр
- Язык

- Бюджет
- Дата выхода

В качестве разделителей используются запятые. Из всех вышеперечисленных характеристик нас интересуют:

- Название фильма.
- Проголосовавшие пользователи.
- Контент-рейтинг.
- Страна.
- Жанр.

Это нам позволит понять, фильмы какого жанра предпочитают смотреть пользователи, учитывая контент-рейтинг, а также в каких странах больше всего пользователей смотрят фильмы.

На таблицу 1 представлен фрагмент обучающей выборки:

Фильм	Фильм	Проголосовавшие пользователи	Контент-рейтинг	Цвет	Страна	Кинорежиссер	actor_1_name	genres
AvatarB	AvatarB	886204	PG-13	Color	USA	James Cameron	CCH Pounder	Action Adventure Fantasy Sci-Fi
Pirates of the Caribbean: At World's EndB	Pirates of the Caribbean : At World's EndB	471220	PG-13	Color	USA	Gore Verbinski	Johnny Depp	Action Adventure Fantasy
SpectreB	SpectreB	275868	PG-13	Color	UK	Sam Mendes	Christoph Waltz	Action Adventure Thriller
The Dark Knight RisesB	The Dark Knight RisesB	1144337	PG-13	Color	USA	Christopher Nolan	Tom Hardy	Action Thriller
John CarterB	John CarterB	212204	PG-13	Color	USA	Andrew Stanton	Daryl Sabara	Action Adventure Sci-Fi
Spider-Man 3B	Spider-Man 3B	383056	PG-13	Color	USA	Sam Raimi	J.K. Simmons	Action Adventure Romance
TangledB	TangledB	294810	PG	Color	USA	Nathan Greno	Brad Garrett	Adventure Animation Comedy Family Fantasy Musical Romance
Avengers: Age of UltronB	Avengers: Age of UltronB	462669	PG-13	Color	USA	Joss Whedon	Chris Hemsworth	Action Adventure Sci-Fi
Harry Potter and the Half-Blood PrinceB	Harry Potter and the Half-Blood PrinceB	321795	PG	Color	UK	David Yates	Alan Rickman	Adventure Family Fantasy Mystery
Batman v Superman: Dawn of JusticeB	Batman v Superman : Dawn of JusticeB	371639	PG-13	Color	USA	Zack Snyder	Henry Cavill	Action Adventure Sci-Fi
Superman ReturnsB	Superman ReturnsB	240396	PG-13	Color	USA	Bryan Singer	Kevin Spacey	Action Adventure Sci-Fi
Quantum of SolaceB	Quantum of SolaceB	330784	PG-13	Color	UK	Marc Forster	Giancarlo Giannini	Action Adventure
Pirates of the Caribbean: Dead Man's ChestB	Pirates of the Caribbean : Dead Man's ChestB	522040	PG-13	Color	USA	Gore Verbinski	Johnny Depp	Action Adventure Fantasy
The Lone RangerB	The Lone RangerB	181792	PG-13	Color	USA	Gore Verbinski	Johnny Depp	Action Adventure Western
Man of SteelB	Man of SteelB	548573	PG-13	Color	USA	Zack Snyder	Henry Cavill	Action Adventure Fantasy Sci-Fi

The Chronicles of Narnia: Prince CaspianB	The Chronicle s of Narnia: Prince CaspianB	149922	PG	Color	USA	Andrew Adamson	Peter Dinklage	Action Adventure Family Fantasy
The AvengersB	The Avengers B	995415	PG-13	Color	USA	Joss Whedon	Chris Hemsworth	Action Adventure Sci-Fi
Pirates of the Caribbean: On Stranger TidesB	Pirates of the Caribbean : On Stranger TidesB	370704	PG-13	Color	USA	Rob Marshall	Johnny Depp	Action Adventure Fantasy
Men in Black 3B	Men in Black 3B	268154	PG-13	Color	USA	Barry Sonnenfeld	Will Smith	Action Adventure Comedy Family Fantasy Sci-Fi
The Hobbit: The Battle of the Five ArmiesB	The Hobbit: The Battle of the Five ArmiesB	354228	PG-13	Color	New Zealand	Peter Jackson	Aidan Turner	Adventure Fantasy
The Amazing Spider-ManB	The Amazing Spider-ManB	451803	PG-13	Color	USA	Marc Webb	Emma Stone	Action Adventure Fantasy
Robin HoodB	Robin HoodB	211765	PG-13	Color	USA	Ridley Scott	Mark Addy	Action Adventure Drama History
The Hobbit: The Desolation of SmaugB	The Hobbit: The Desolation of SmaugB	483540	PG-13	Color	USA	Peter Jackson	Aidan Turner	Adventure Fantasy
The Golden CompassB	The Golden Compass B	149019	PG-13	Color	USA	Chris Weitz	Christopher Lee	Adventure Family Fantasy

Таблица 1 – Фрагмент обучающей выборки

Классификационные шкалы следующие: фильм, проголосовавшие пользователи. Остальные – описательные.

1.3. Ввод выборки в систему Aidos-X

Теперь, когда мы имеем обучающую выборку в формате *.xlsx, можно импортировать ее в систему Aidos-X.

Скопируем выборку в папку C:\Work\Aidos-X\AID_DATA\Inp_data\ и переименуем ее в Inp_data. Затем, запустив систему, воспользуемся универсальным программным интерфейсом для ввода данных из внешних баз табличного вида (режим 2.3.2.2). Настройка ввода будет такая, как на рисунке 1 – ввод обучающей выборки.

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 Стандарт XLS-файла
☒ XLSX - MS Excel-2007(2010)
☐ DBF - DBASE IV (DBF/NTX) Стандарт DBF-файла
☐ CSV - Comma-Separated Values Стандарт CSV-файла

Задайте параметры:

☒ Нули и пробелы считать ОТСУТСТВИЕМ данных
☐ Нули и пробелы считать ЗНАЧЕНИЯМИ данных
☒ Создавать БД средних по классам "Inp_davr.dbf"?
 Требования к файлу исходных данных

Задайте диапазон столбцов классификационных шкал:

Начальный столбец классификационных шкал:
 Конечный столбец классификационных шкал:

Задайте диапазон столбцов описательных шкал:

Начальный столбец описательных шкал:
 Конечный столбец описательных шкал:

Задайте режим:

☒ Формализации предметной области (на основе "Inp_data")
☐ Генерации распознаваемой выборки (на основе "Inp_rasp")

Задайте способ выбора размера интервалов:

☒ Равные интервалы с разным числом наблюдений
☐ Разные интервалы с равным числом наблюдений

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

☒ Не применять сценарный метод АСК-анализа ☐ Применить сценарный метод АСК-анализа
☐ Применить спец. интерпретацию текстовых полей классов ☐ Применить спец. интерпретацию текстовых полей признаков

Параметры интерпретации значений текстовых полей "Inp_data":

Интерпретация TXT-полей классов:

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

Интерпретация TXT-полей признаков:

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

Какие наименования ГРАДАЦИЙ числовых шкал использовать:

☒ Только интервальные числовые значения (например: "1/3-{59873.0000000, 178545.6666667}")
☐ Только наименования интервальных числовых значений (например: "Минимальное")
☐ И интервальные числовые значения, и их наименования (например: "Минимальное: 1/3-{59873.0000000, 178545.6666667}")

Ok Cancel

Рисунок 1 – Ввод обучающей выборки

Выделим следующие настройки:

- Тип файла – xlsx.
- Классификационные шкалы – 2-3.
- Описательные шкалы – 4-12.
- В качестве классов рассматриваются значения полей целиком.

Так как предварительно числовые шкалы были разбиты на интервалы так, как это необходимо для исследования, перерасчет шкал после ввода выборки производить не надо.

После импорта автоматически формируются классификационные и описательные шкалы, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Этим самым полностью автоматизировано выполняется этап «Формализация предметной области».

Классификационные шкалы можно просмотреть в режиме 2.1 (рисунок 2).

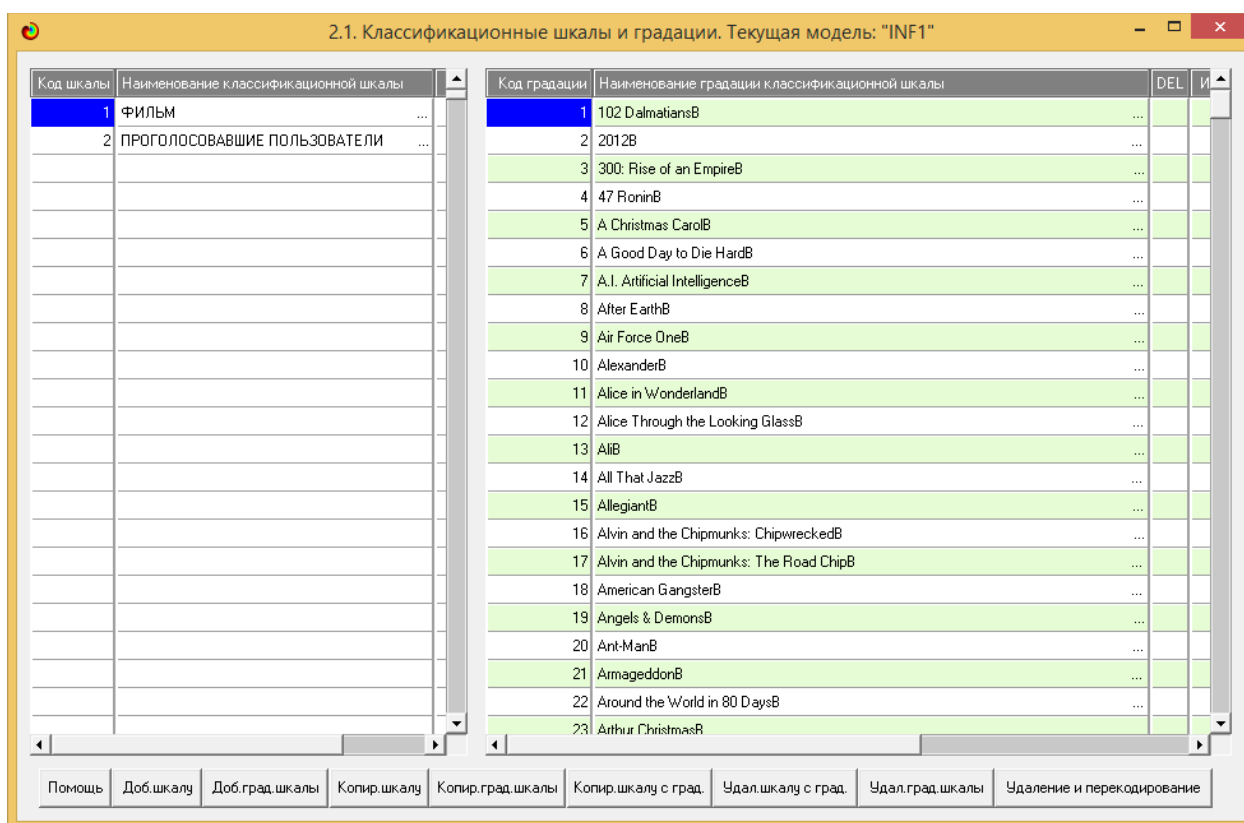


Рисунок 2 – Классификационные шкалы

Описательные шкалы – в режиме 2.2 (рисунок 3).

Для ручного ввода-корректировки выборки существует режим 2.3.1, его можно увидеть на рисунке 4. Установка значений описательных и классификационных шкал объектов осуществляется по их номерам.

2.2. Описательные шкалы и градации. Текущая модель: "INF1"

Код шкалы	Наименование описательной шкалы	Код градации	Наименование градации описательной шкалы
1	КОНТЕНТ-РЕЙТИНГ	1	G
2	ЦВЕТ	2	PG-13
3	СТРАНА	3	PG
4	КИНОРЕЖИССЕР	4	R
5	ACTOR_1_NAME	5	TV-14
6	GENRES	6	TV-G
7	LANGUAGE	7	TV-MA
8	BUDGET	8	TV-PG
9	TITLE_YEAR		

Помощь Доб. шкалу Доб. град. шкалы Копир. шкалу Копир. град. шкалы Копир. шкалу с град. Удал. шкалу с град. Удал. град. шкалы Перекодировать Очистить

Рисунок 3 – Описательные шкалы

2.3.1. Ручной ввод-корректировка обучающей выборки. Текущая модель: "INF1"

Код объекта	Наименование объекта	Дата	Время
1	AvatarB	...	...
2	Pirates of the Caribbean: At World's EndB	...	...
3	SpectreB	...	...
4	The Dark Knight RisesB	...	...
5	John CarterB	...	...
6	Spider-Man 3B	...	...
7	TangledB	...	...
8	Avengers: Age of UltronB	...	...
9	Harry Potter and the Half-Blood PrinceB	...	...
10	Batman v Superman: Dawn of JusticeB	...	...

Код объекта	Класс 1	Класс 2	Класс 3	Класс 4
1	27	496	0	0

Код объекта	Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
1	2	10	21	116	317	569	747
1	753	0	0	0	0	0	0

Помощь Скопировать обуч. выб. в расп. Добавить объект Добавить классы Добавить признаки Удалить объект Удалить классы Удалить признаки Очистить БД

Рисунок 4 – Ручной ввод-корректировка выборки

Тем самым создаются все необходимые и достаточные предпосылки для выявления связей между значениями факторов и результатами их совместного системного воздействия.

1.4. Синтез и верификация статистических и интеллектуальных моделей

Далее запускаем режим 3.5 – Синтез и верификация моделей. Настраиваем так, как показано на рисунке 5.

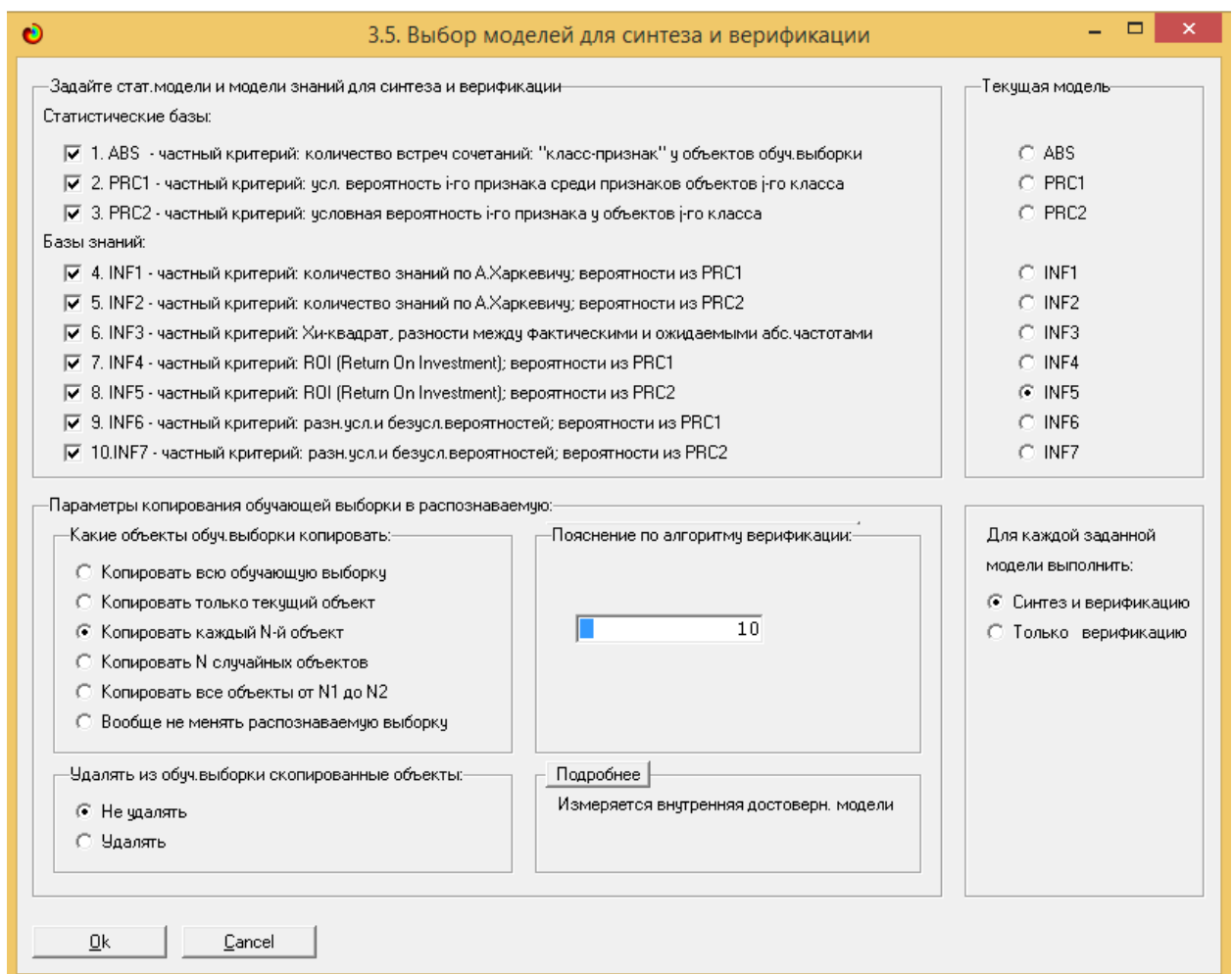


Рисунок 5 – Настройка синтеза и верификации моделей

Так как в выборке около 5000 значений, для распознавания будем использовать лишь 500. Это будет более наглядным представлением, т.к. информация по некоторым фильмам частично либо полностью отсутствует.

Также это обусловлено устаревшими фильмами. После построения моделей нажимаем «ОК» на форме, которая показана на рисунке 6.

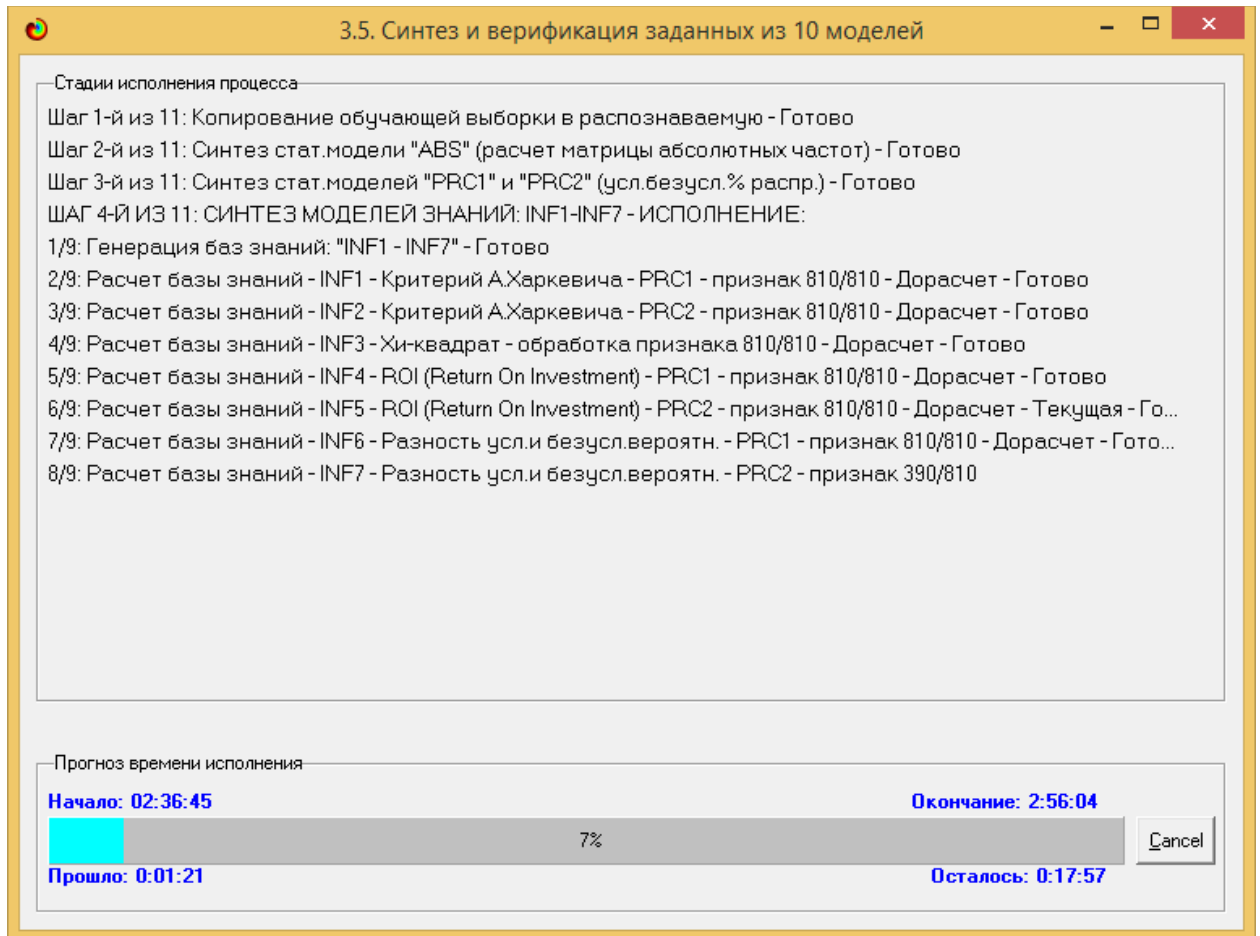


Рисунок 6 – Завершение процесса синтеза и верификации моделей

После этого можно перейти непосредственно к выбору наиболее достоверной модели.

1.5. Виды моделей системы «Aidos»

Рассмотрим решение задачи идентификации на примере модели INF1, в которой рассчитано количество информации по А.Харкевичу, которое мы получаем о принадлежности идентифицируемого объекта к каждому из классов, если знаем, что у этого объекта есть некоторый признак.

По сути, частные критерии представляют собой просто формулы для преобразования матрицы абсолютных частот (таблица 2) в матрицы

условных и безусловных процентных распределений, и матрицы знаний (таблицы 3 и 4) (проф. В.И.Лойко).

Таблица 2 – Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений (фрагменты)

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

Код признака	Наименование описательной шкалы и градации	1. ФИЛЬМ 102 DALMATIA...	2. ФИЛЬМ 2012B	3. ФИЛЬМ 300: RISE OF AN EMPIREB	4. ФИЛЬМ 47 RONINB	5. ФИЛЬМ A CHRISTMAS CAROLB	6. ФИЛЬМ A GOOD DAY TO DIE HARDB	7. ФИЛЬМ A.I. ARTIFICIAL INTELLIGE...	8. ФИЛЬМ AFTER EARTHb	9. ФИЛЬМ AIR FORCE ONEB	10. ФИЛЬМ ALEXANDE...	11. ФИЛЬМ ALICE IN WONDERL...	12. ФИЛЬМ ALICE THROUGH THE LOOKING GLASSB	13. ФИЛЬМ ALIB
1	КОНТЕНТ-РЕЙТИНГ-G	1												
2	КОНТЕНТ-РЕЙТИНГ-PG-13		1		1			1	1					
3	КОНТЕНТ-РЕЙТИНГ-PG					1						1	1	
4	КОНТЕНТ-РЕЙТИНГ-R			1			1				1			1
5	КОНТЕНТ-РЕЙТИНГ-TV-14													
6	КОНТЕНТ-РЕЙТИНГ-TV-G													
7	КОНТЕНТ-РЕЙТИНГ-TV-MA													
8	КОНТЕНТ-РЕЙТИНГ-TV-PG													
9	ЦВЕТ-Black and White													
10	ЦВЕТ-Color	1	1	1	1	1	1	1	1	1	1	1	1	
11	СТРАНА-Australia													
12	СТРАНА-Belgium													
13	СТРАНА-Canada													
14	СТРАНА-China													
15	СТРАНА-France										1			
16	СТРАНА-Germany													

Таблица 3 – Матрица информативностей (модель INF1) в битах

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

Код признака	Наименование описательной шкалы и градации	1. ФИЛЬМ 102 DALMATIA...	2. ФИЛЬМ 2012B	3. ФИЛЬМ 300: RISE OF AN EMPIREB	4. ФИЛЬМ 47 RONINB	5. ФИЛЬМ A CHRISTMAS CAROLB	6. ФИЛЬМ A GOOD DAY TO DIE HARDB	7. ФИЛЬМ A.I. ARTIFICIAL INTELLIGE...	8. ФИЛЬМ AFTER EARTHb	9. ФИЛЬМ AIR FORCE ONEB	10. ФИЛЬМ ALEXANDE...	11. ФИЛЬМ ALICE IN WONDERL...	12. ФИЛЬМ ALICE THROUGH THE LOOKING GLASSB	13. ФИЛЬМ ALIB	14. ФИЛЬМ AL JAC...
1	КОНТЕНТ-РЕЙТИНГ-G	3.240													
2	КОНТЕНТ-РЕЙТИНГ-PG-13		0.597		0.597			0.597	0.597						
3	КОНТЕНТ-РЕЙТИНГ-PG					1.432						1.432	1.432		
4	КОНТЕНТ-РЕЙТИНГ-R			1.862			1.862			1.862	1.862			1.997	
5	КОНТЕНТ-РЕЙТИНГ-TV-14														
6	КОНТЕНТ-РЕЙТИНГ-TV-G														
7	КОНТЕНТ-РЕЙТИНГ-TV-MA														
8	КОНТЕНТ-РЕЙТИНГ-TV-PG														
9	ЦВЕТ-Black and White														
10	ЦВЕТ-Color	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012		
11	СТРАНА-Australia														
12	СТРАНА-Belgium														
13	СТРАНА-Canada														
14	СТРАНА-China														
15	СТРАНА-France														
16	СТРАНА-Germany										4.058				

Таблица 4 – Матрица знаний (модель INF3)

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

Код признака	Наименование описательной шкалы и градации	1. ФИЛЬМ 102 DALMATIA...	2. ФИЛЬМ 2012B	3. ФИЛЬМ 300. RISE OF AN EMPIREB	4. ФИЛЬМ 47 RONINB	5. ФИЛЬМ A CHRISTMAS CAROLB	6. ФИЛЬМ A GOOD DAY TO DIE HARDB	7. ФИЛЬМ AI. ARTIFICIAL INTELLIGE...	8. ФИЛЬМ AFTER EARTHB	9. ФИЛЬМ AIR FORCE ONEB	10. ФИЛЬМ ALEXANDER...	11. ФИЛЬМ ALICE IN WONDERL...	12. ФИЛЬМ ALICE THROUGH THE LOOKING GLASSB	13. ФИЛЬМ ALIB	14. ФИЛЬМ AL JAL
1	КОНТЕНТ-РЕЙТИНГ-G	0.960	-0.040	-0.040	-0.040	-0.040	-0.040	-0.040	-0.040	-0.040	-0.040	-0.040	-0.040	-0.040	-0.035
2	КОНТЕНТ-РЕЙТИНГ-PG-13	-0.553	0.447	-0.553	0.447	-0.553	0.447	-0.553	0.447	-0.553	-0.553	-0.553	-0.553	-0.553	-0.484
3	КОНТЕНТ-РЕЙТИНГ-PG	-0.242	-0.242	-0.242	-0.242	0.758	-0.242	-0.242	-0.242	-0.242	-0.242	0.758	0.758	-0.212	
4	КОНТЕНТ-РЕЙТИНГ-R	-0.158	-0.158	0.842	-0.158	-0.158	0.842	-0.158	-0.158	0.842	0.842	-0.158	-0.158	0.862	
5	КОНТЕНТ-РЕЙТИНГ-TV-14	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
6	КОНТЕНТ-РЕЙТИНГ-TV-G	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
7	КОНТЕНТ-РЕЙТИНГ-TV-MA	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
8	КОНТЕНТ-РЕЙТИНГ-TV-PG	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
9	ЦВЕТ-Black and White														
10	ЦВЕТ-Color	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	-0.864
11	СТРАНА-Australia	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.010
12	СТРАНА-Belgium	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
13	СТРАНА-Canada	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
14	СТРАНА-China	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
15	СТРАНА-France	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.006
16	СТРАНА-Germany	-0.018	-0.018	-0.018	-0.018	-0.018	-0.018	-0.018	-0.018	-0.018	0.982	-0.018	-0.018	-0.018	-0.016

1.5. Результаты верификации моделей

Результаты верификации моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями, приведены на рисунке 7.

4.1.3.6. Обобщенная форма по достов.моделям при разн.инт.крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Всего признаков объектной выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	Ф-мера Ван Ризенбергена	Сумма модулей усреднений по истинно-положительным решениям (ST)	Сумма модулей усреднений по истинно-отрицательным решениям (ST)	Сумма модулей усреднений по ложно-положительным решениям (SEP)	Сумма модулей усреднений по ложно-отрицательным решениям (SF)
1. ABS - частный критерий: количество встреч сочетаний "хлас...	Корреляция абс частот с абс частот по признакам...	92	92		24212		0.004	1.000	0.408	78.006		10542.731	
2. PRC1 - частный критерий: количество встреч сочетаний "хлас...	Сумма абс частот по признакам...	92	92		24212		0.004	1.000	0.408	31.597		92.434	
3. PRC1 - частный критерий: усл. вероятность его признака сред...	Корреляция усл.отн частот с усл.отн частот по признакам...	92	92		24212		0.004	1.000	0.408	78.006		10542.731	
4. PRC1 - частный критерий: усл. вероятность его признака сред...	Сумма усл.отн частот по признакам...	92	92		24212		0.004	1.000	0.408	68.995		10686.181	
5. PRC2 - частный критерий: условная вероятность его признака...	Корреляция усл.отн частот с усл.отн частот по признакам...	92	92		24212		0.004	1.000	0.408	78.006		10542.730	
6. PRC2 - частный критерий: усл. вероятность его признака...	Сумма усл.отн частот по признакам...	92	92		24212		0.004	1.000	0.408	68.795		10608.950	
7. INF1 - частный критерий: количество знаний по А\Харкевичу; в...	Семантический резонанс зна...	92	67	8783	15429	25	0.004	0.728	0.409	41.483	39.507	679.307	0.831
8. INF1 - частный критерий: количество знаний по А\Харкевичу; в...	Сумма знаний	92	87	48	24164	5	0.004	0.946	0.407	35.748	1.747	662.631	0.120
9. INF2 - частный критерий: количество знаний по А\Харкевичу; в...	Семантический резонанс зна...	92	67	8906	15306	25	0.004	0.728	0.409	41.510	36.293	680.685	0.821
10. INF2 - частный критерий: количество знаний по А\Харкевичу; в...	Сумма знаний	92	87	47	24165	5	0.004	0.946	0.407	35.775	1.698	662.497	0.111
11. INF3 - частный критерий: Хи-квадрат, разности между факти...	Семантический резонанс зна...	92	83	13337	10875	9	0.008	0.902	0.415	55.468	1450.143	1527.159	0.714
12. INF3 - частный критерий: Хи-квадрат, разности между факти...	Сумма знаний	92	83	13337	10875	9	0.008	0.902	0.415	20.874	217.135	197.448	1.181
13. INF4 - частный критерий: ROI (Return On Investment); веро...	Семантический резонанс зна...	92	63	22932	1280	29	0.047	0.685	0.488	36.817	147.412	62.067	1.251
14. INF4 - частный критерий: ROI (Return On Investment); веро...	Сумма знаний	92	89	43	24169	3	0.004	0.967	0.407	13.292	0.009	22.686	0.001
15. INF5 - частный критерий: ROI (Return On Investment); веро...	Семантический резонанс зна...	92	63	22922	1290	29	0.047	0.685	0.487	36.822	146.981	62.114	1.241
16. INF5 - частный критерий: ROI (Return On Investment); веро...	Сумма знаний	92	89	40	24172	3	0.004	0.967	0.407	14.885	0.009	25.461	0.001
17. INF6 - частный критерий: разн.усли и безусл.вероятностей; вер...	Семантический резонанс зна...	92	83	2153	22059	9	0.004	0.902	0.407	52.156	18.778	2315.676	0.881
18. INF6 - частный критерий: разн.усли и безусл.вероятностей; вер...	Сумма знаний	92	83	27	24185	9	0.003	0.902	0.407	38.463	0.494	2062.700	0.021
19. INF7 - частный критерий: разн.усли и безусл.вероятностей; вер...	Семантический резонанс зна...	92	84	2533	21679	8	0.004	0.913	0.408	53.315	13.921	2317.855	0.471
20. INF7 - частный критерий: разн.усли и безусл.вероятностей; вер...	Сумма знаний	92	86	25	24187	6	0.004	0.935	0.407	38.462	0.427	2037.096	0.011

Рисунок 7 – Результаты верификации моделей

Наиболее достоверной в данном приложении оказалась модель INF5 при интегральном критерии «Семантический резонанс знаний», если учитывать F-меру Ван Ризенбергена. Точность модели составляет 0,488, то есть уровень достоверности прогнозирования с применением модели выше,

чем экспертных оценок, достоверность которых считается равной примерно 48,4%.

Также статистические модели обычно дают более низкую средневзвешенную достоверность идентификации и неидентификации, чем модели знаний, и практически никогда – более высокую. Соответственно, это оправдывает применение интеллектуальных технологий и моделей знаний. Посмотрим, каковы частные распределения уровней сходства и различия для верно и ошибочно неидентифицированных и идентифицированных ситуаций в наиболее достоверной модели INF5 (рисунок 8). Для этого зайдём в режим 4.1.3.11, выберем модель INF5 и запросим график по сумме знаний.

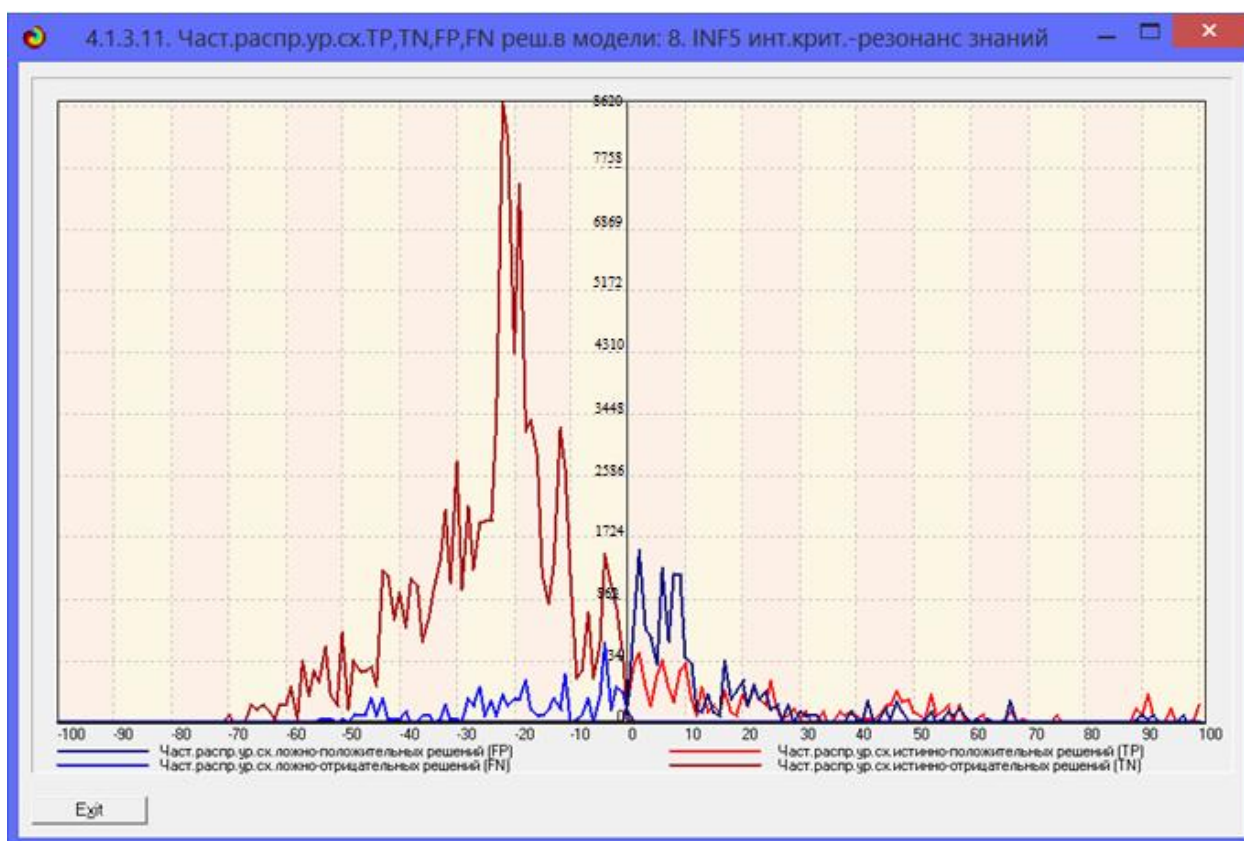


Рисунок 8 – Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF4

Как можно видеть, при модели INF4 непринадлежность объекта к классу лучше определяется, чем принадлежность. При этом существует вероятность ложного определения непринадлежности или принадлежности.

Выберем модель INF4 в качестве основной и перейдем к решению задач идентификации.

2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Выделение основных задач

Если смотреть с позиции людей, занимающиеся разработкой фильма, то какие проблемы можно выявить?

Во-первых, необходимо обратить внимание на востребованность жанров для людей определенной страны. Например, в некоторых странах люди предпочитают смотреть фильмы жанра Sci-fi, но кинематограф страны предлагает, например, только Action-фильмы. Поэтому необходимо проанализировать рынок и выявить предпочтения людей.

Во-вторых, необходимо подобрать правильный контент, чтобы он вписывался в рамки контент-рейтинга.

В-третьих, следует обратить внимание на рынок со стороны потребителя. Будет ли создаваемый фильм уместным, интересно ли его будет смотреть и т.д.

Таким образом, в первую очередь, нас интересуют следующие связи:

- Фильм – проголосовавшие пользователи.
- Фильм – страна.
- Фильм – контент-рейтинг.
- Жанр – язык.
- Фильм – бюджет.

2.2. Выбор модели и пакетное распознавание

Зададим модель INF5 в качестве текущей в режиме 5.6 (рисунок 9), после чего проведем пакетное распознавание в режиме 4.1.2 (рисунок 10).

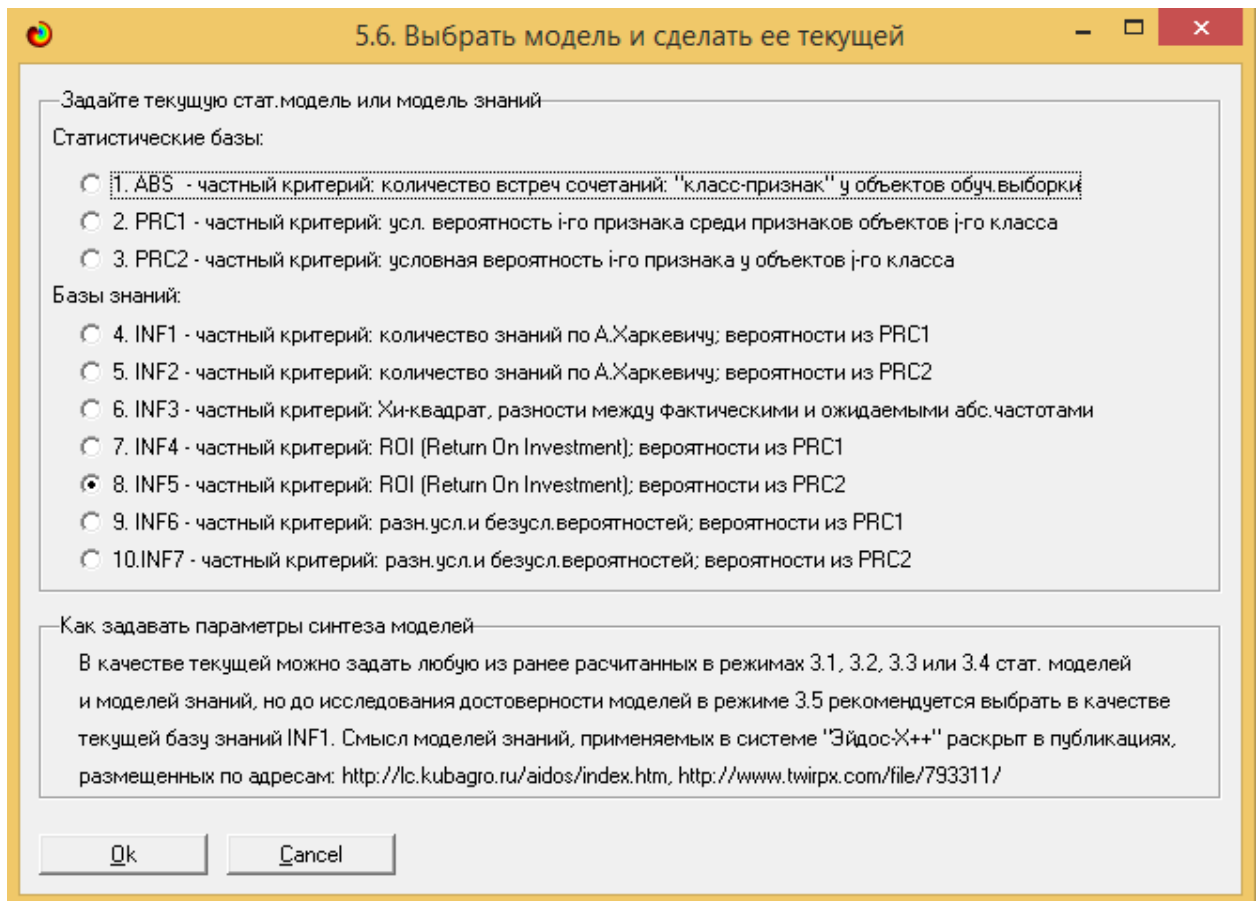


Рисунок 9 – Выбор текущей модели

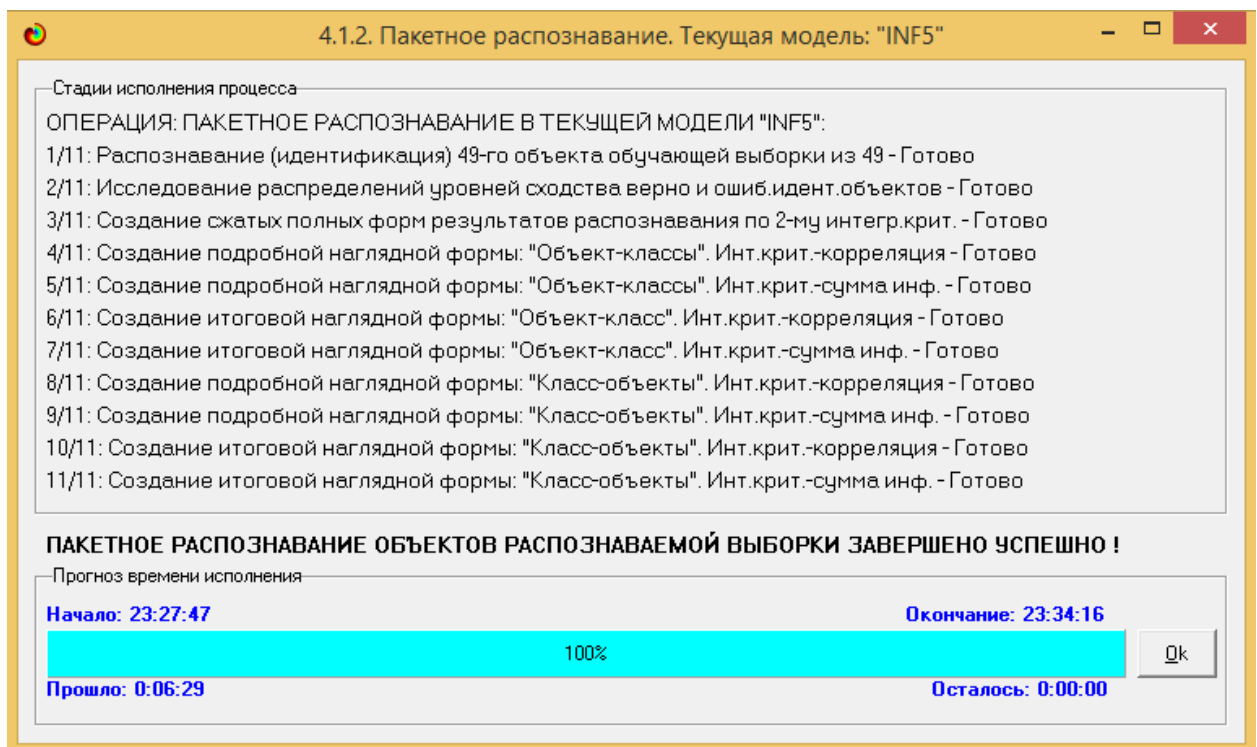


Рисунок 10 – Пакетное распознавание в новой модели

В результате пакетного распознавания в текущей модели создается ряд баз данных, которые визуализируются в выходных экранных формах, отражающих результаты решения задачи идентификации и прогнозирования.

Режим 4.1.3 системы «Aidos-X» обеспечивает отображение результатов идентификации и прогнозирования в различных формах:

1. Подробно наглядно: "Объект – классы".
2. Подробно наглядно: "Класс – объекты".
3. Итоги наглядно: "Объект – классы".
4. Итоги наглядно: "Класс – объекты".
5. Подробно сжато: "Объект – классы".
6. Обобщенная форма по достоверности моделей при разных интегральных критериях.
7. Обобщенный статистический анализ результатов идентификации по моделям и интегральным критериям.
8. Статистический анализ результатов идентификации по классам, моделям и интегральным критериям.
9. Распознавание уровня сходства при разных моделях и интегральных критериях.
10. Достоверность идентификации классов при разных моделях и интегральных критериях.

Ниже кратко рассмотрим некоторые из них.

На рисунках 11 и 12 приведены примеры прогнозов высокой и низкой достоверности классов, а также итоги распознавания классов объектов в модели INF5:

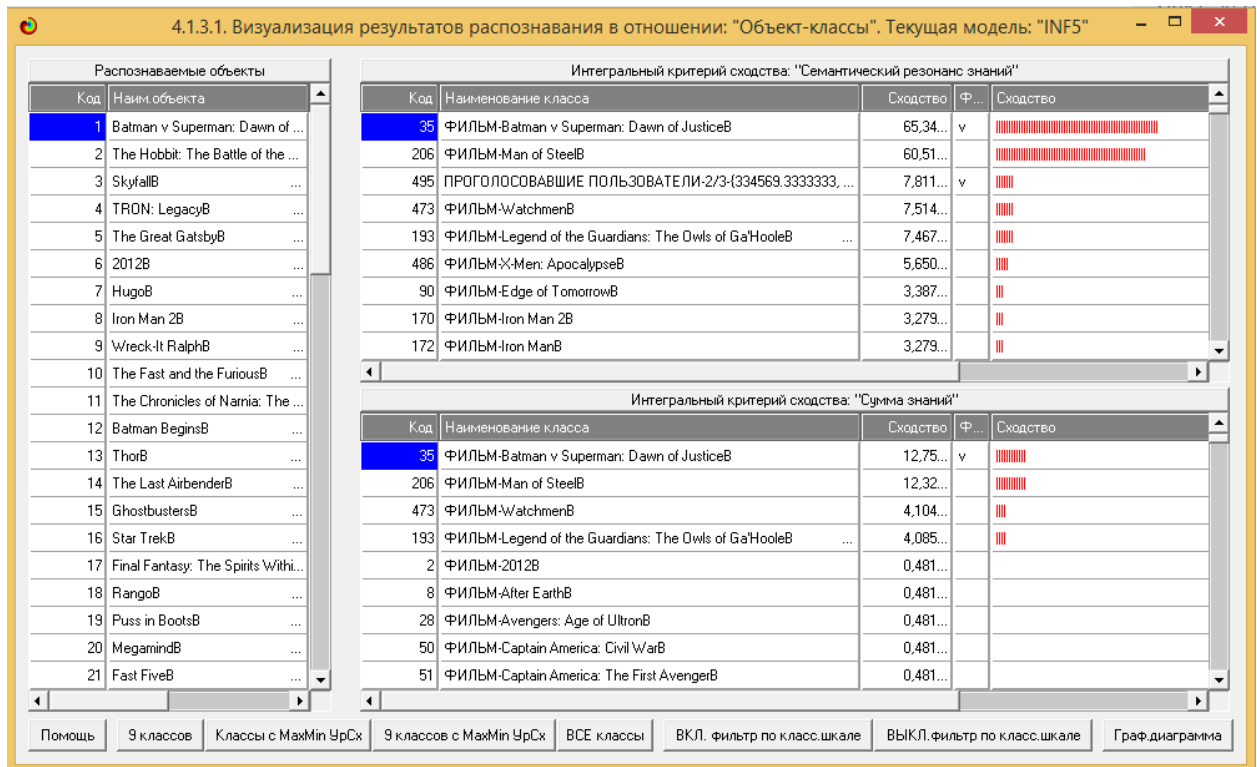


Рисунок 11 – Прогнозы достоверности классов объектов

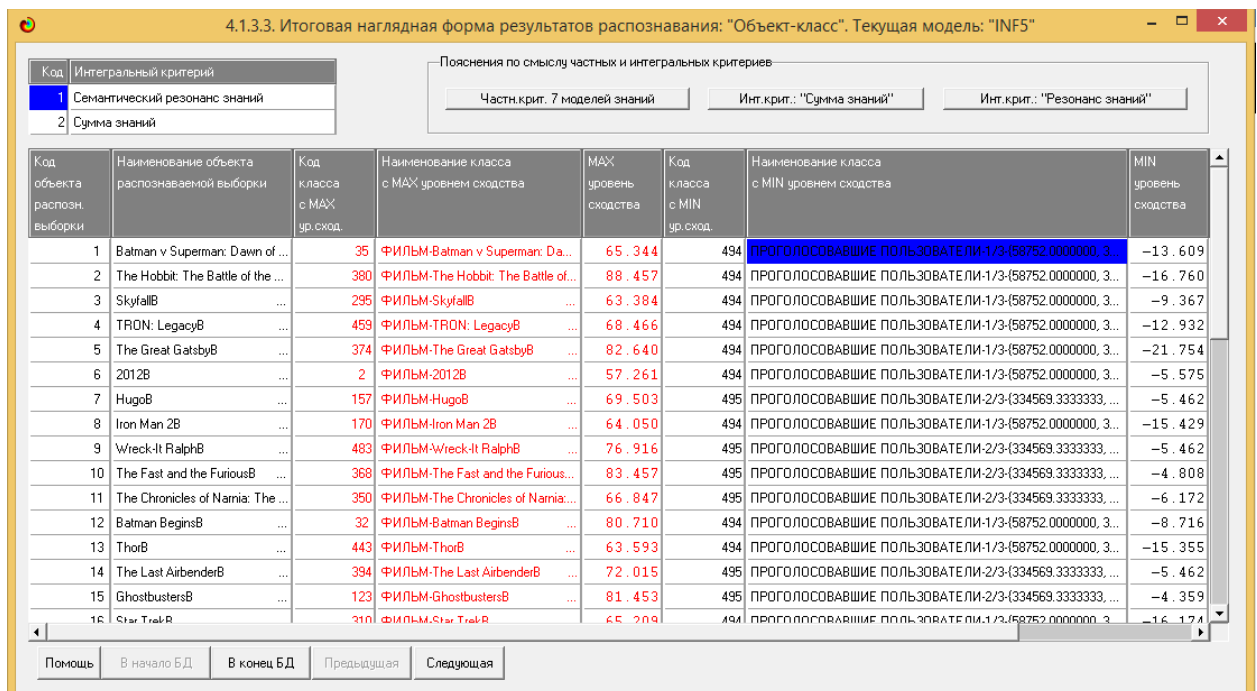


Рисунок 12 – Итоги распознавания классов объектов

2.3 Визуализация когнитивных функций

Применительно к задачам, рассматриваемым в данной работе, когнитивная функция показывает, какое количество информации содержится в различных значениях факторов о том, что объект моделирования перейдет в те или иные будущие состояния.

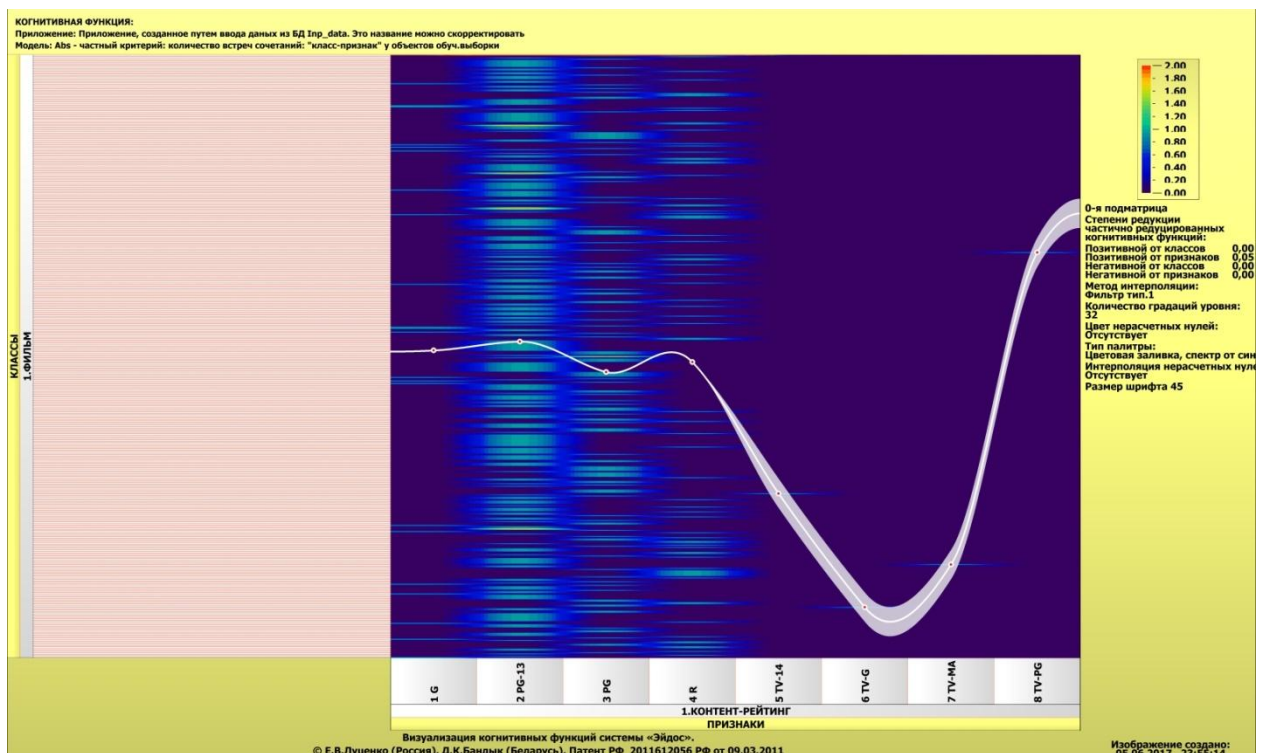


Рисунок 13 – Визуализация зависимости фильмов от контент-рейтинга

Рассмотрим визуализацию когнитивных функций в рамках определенных задач.

Зависимость фильмов от контент-рейтинга представлена на рисунке 13.

Из него видно, что большая часть фильмов выпускается с рейтингами PG-13 и TV-PG. PG-13-настоятельное предостережение родителям, некоторый материал может быть неподходящим для детей до 13 лет. Это связано с определенными жанрами фильмов. Так, например, большинство

action-фильмов имеет рейтинг PG-13, так любимый взрослыми людьми. То же можно сказать о фильмах фантастики и триллерах.

Следует обратить внимание, что на втором месте находится рейтинг G-допускаются зрители всех возрастов. К это рейтингу можно отнести анимационные фильмы, приключенческие фильмы, которые предпочитают смотреть как дети, так и взрослые.

Теперь перейдем к зависимости «проголосовавшие пользователи-страна». Она отражена на рисунке 14.

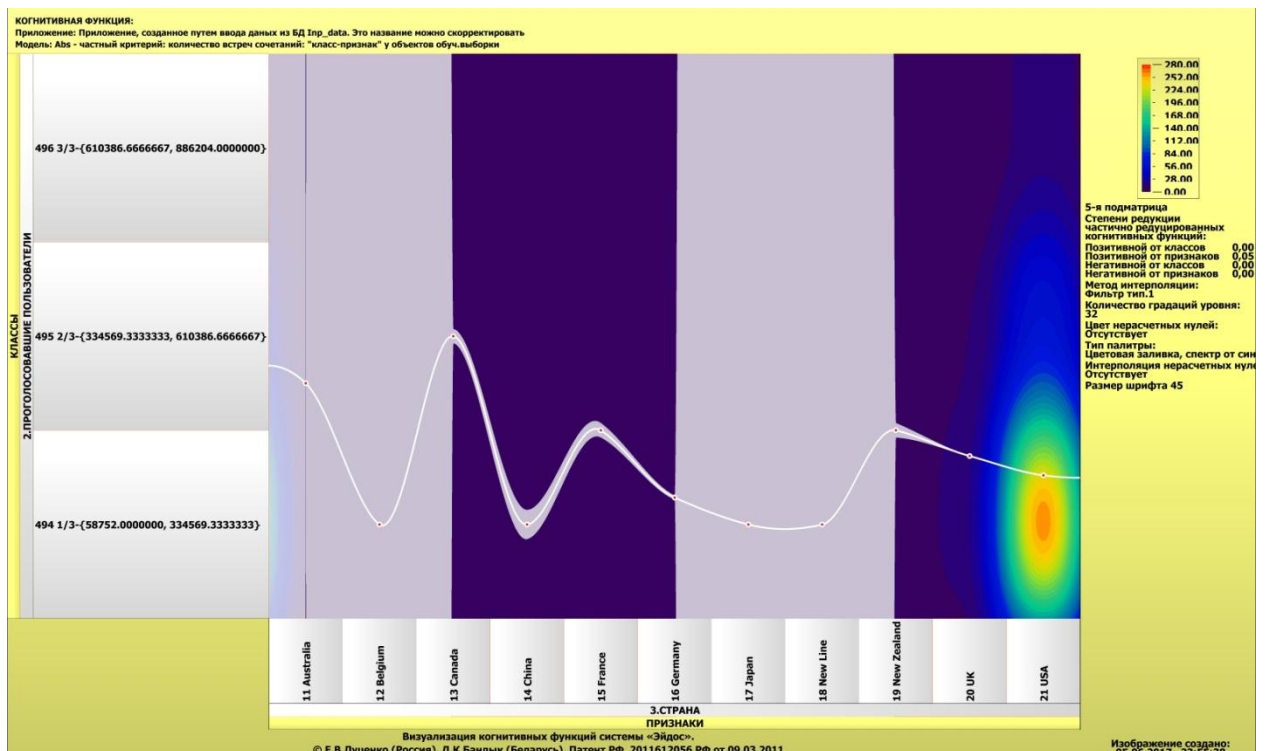


Рисунок 14 – Визуализация зависимости проголосовавших пользователей от страны

Обратимся к графику визуализации зависимости. На нем видно, что основная часть проголосовавших пользователей из США. Это связано со многими факторами, я приведу лишь некоторые из них. Во-первых, в США ежегодно выпускается более 700 фильмов такими студиями, как Сони, Уорнер, Дисней, Юниверсал, Фокс, Парамаунт и т.д. Во-вторых, следует обратить внимание на официальные языки США - английский, испанский, гавайский, французский, что означает, что фильмы,

произведенные только в странах с официальным английским языком (а их около 20), будут понятны людям и в США. Стоит отметить, что именно живая озвучка актеров гораздо приятнее, чем дубляж, который может не передать смысла некоторых выражений, а так же лишит фильма изначального искрометного юмора и даже смысла. В-третьих, кинематограф в США в общем сильнее развит, чем во многих странах. Например, на голосование за определенный фильм сильно влияет нативная реклама, реклама за счет другой продукции, и даже реклама в заведениях общепита. Стоит отметить, что за голосование за определенный фильм люди могут получать какие-нибудь подарки, например внутриигровые, или, например, дополнительное количество очков в заведениях общепита, за которые в будущем можно приобрести еду и напитки.

2.4 SWOT и PEST матрицы и диаграммы

SWOT-анализ является широко известным и общепризнанным методом 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 является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего профессионального опыта и 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.

В версии системы под MS Windows: «Aidos-X++» предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (рисунок 17).

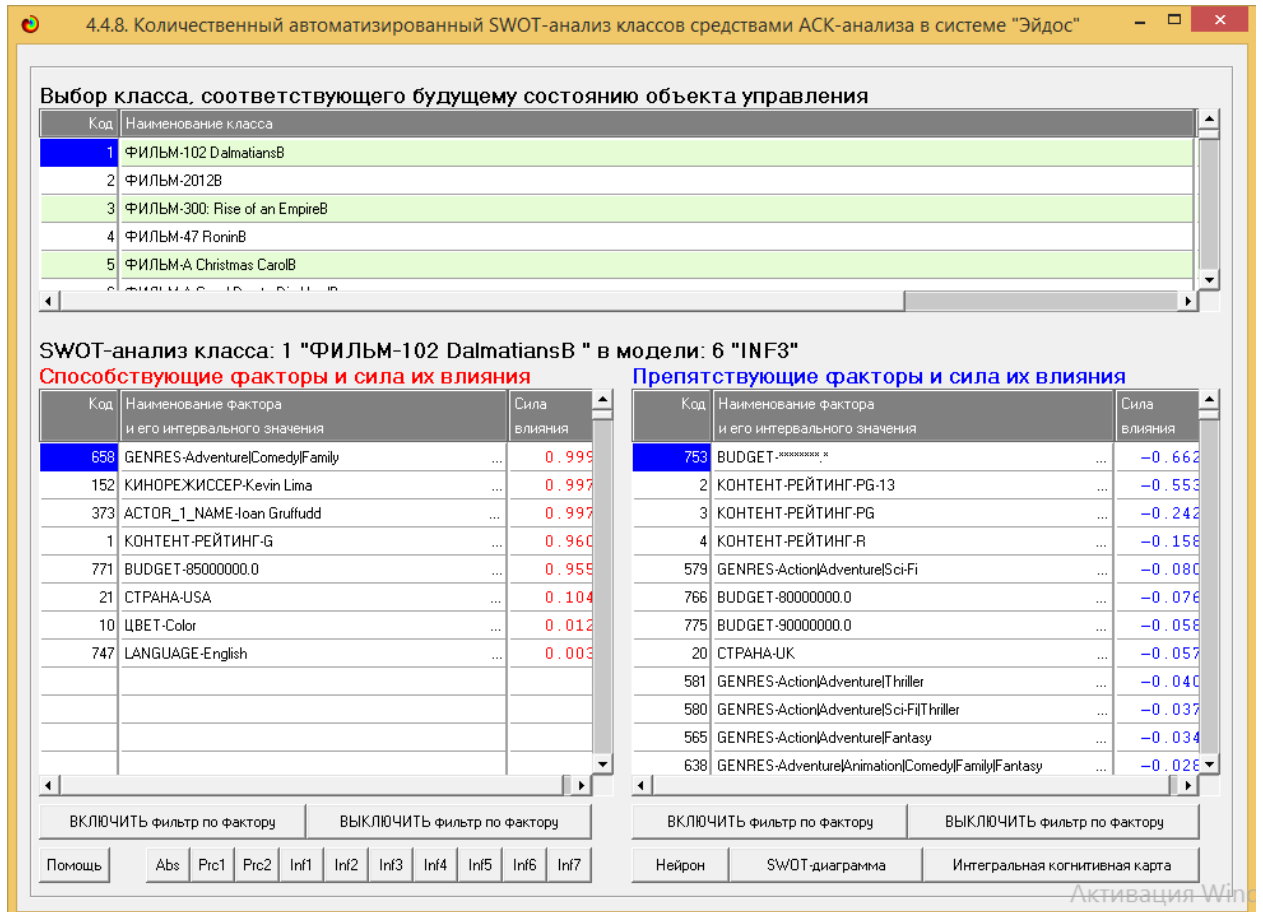


Рисунок 15 – Пример SWOT-Матрицы в модели INF5

На рисунке 18 приведен пример инвертированной SWOT-матрицы.

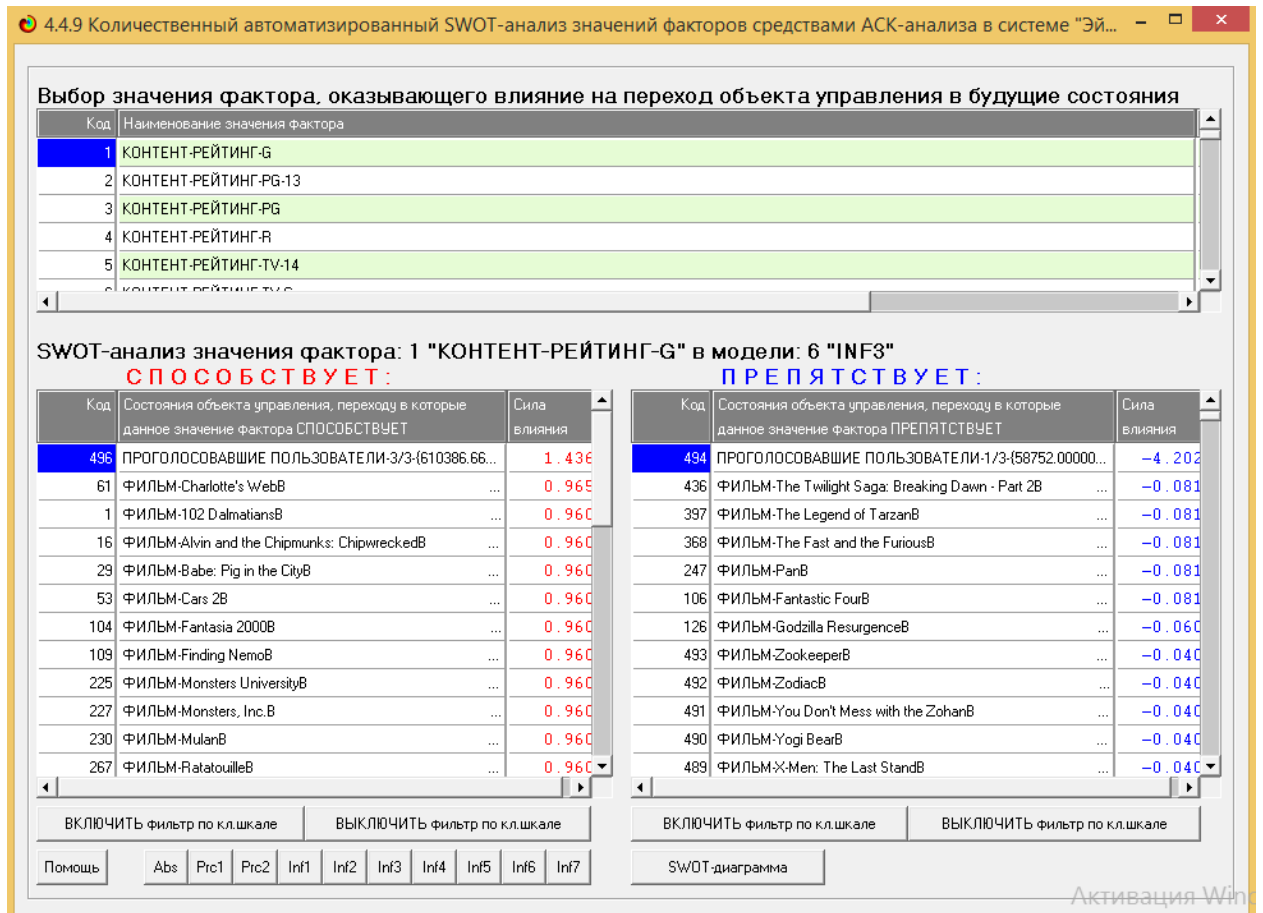


Рисунок 16 – Пример инвертированной SWOT-Матрицы в модели INF5

Возьмем, например, фильм "Опасные пассажиры поезда 123" и построим SWOT-матрицу по ней (рисунок 19):

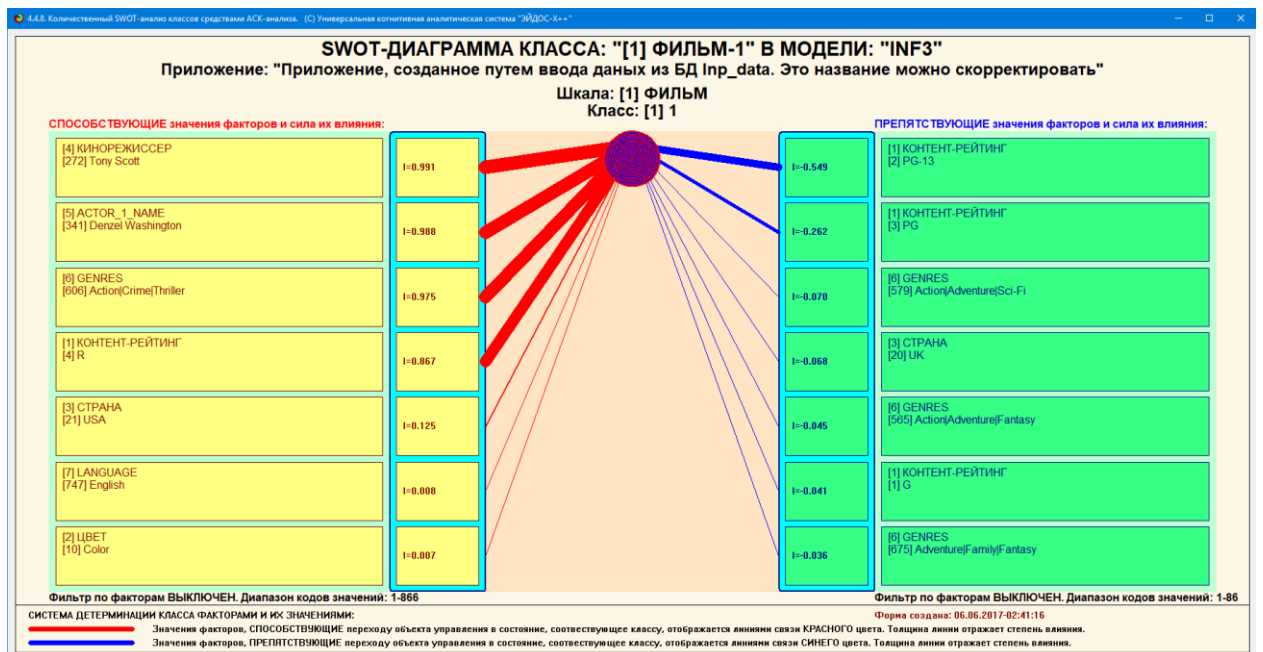


Рисунок 19 – SWOT-матрица, построенная для фильма "Опасные пассажиры поезда 123"

Как можно видеть, несмотря на непопулярность фильма, проголосовавших пользователей-149998. Это связано, возможно, с сильным актерским составом - Дензел Вашингтон, Джон Траволта, Джон Туртурро. Стоит обратить внимание также на контент-рейтинг фильма-PG-13, то есть необходимо присутствие родителей. Причина - жанры: action, crime, thriller.

2.5 Нелокальные нейронные сети и нейтроны

Каждому классу системно-когнитивной модели соответствует нелокальный нейрон, совокупность которых образует не локальную нейронную сеть.

Рассмотрим пару примеров, возвращаясь к нашим задачам.

На рисунке 19 изображен нелокальный нейрон для фильма Wall-E.

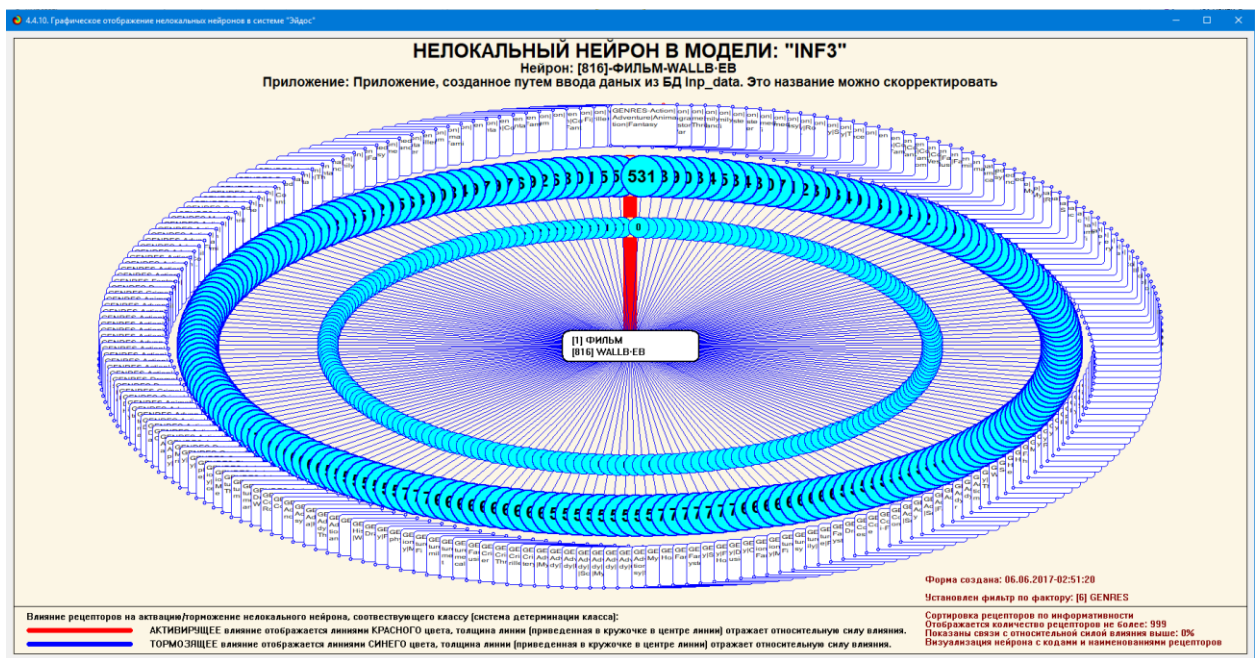


Рисунок 19 – Нейрон для фильма Wall-E

На основании этого нейрона проанализируем фильм Wall-E. Это мультипликационный фильма киностудии Pixar Animation Studios. Фильм имеет рейтинг G-зрители всех возрастов. За данный анимационный фильм проголосовало огромное количество пользователей - 718317, что почти в 8 раз больше чем за фильм из предыдущего примера. На основе этого нейрона

можно определить, что люди, просмотревшие Wall-E предпочитают такие жанры, как Animation, Adventure, Fantasy. Эти жанры нравятся абсолютно всем возрастным категориям.

На основании нелокальных нейронов строится нейронная сеть, пример которой представлен на рисунке 20.

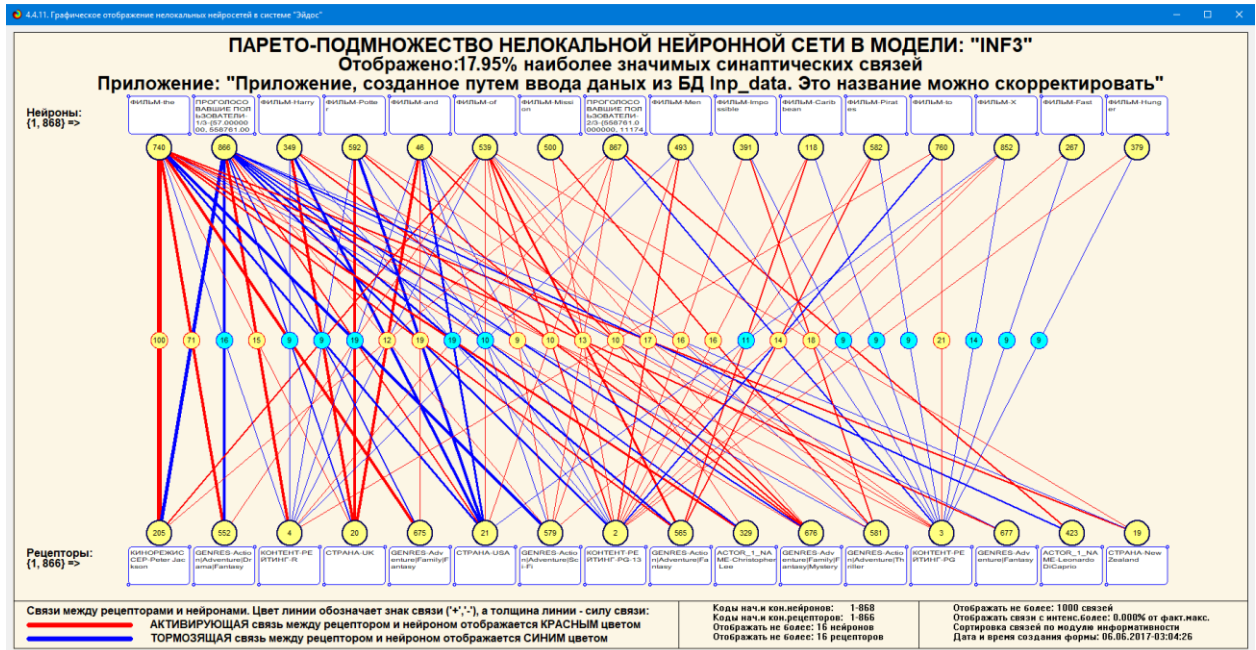


Рисунок 20 – Пример нелокальной нейронной сети в модели INF5

По ней можно, определить какой кинорежиссер больше всего нравится пользователям. Как видим, это Питер Джексон, что неудивительно, т.к. он является режиссером киотрилогии "Властелин колец", созданной по роману Дж. Р. Р. Толкина. Киотрилогия под его контролем выиграла 17 премий Оскара из 30 номинаций, что является потрясающим результатом. Также он был режиссером трилогии "Хоббита", фильма "Кинг-конг". Все вышеперечисленные фильмы совмещают в себе самые популярные жанры, которые так нравятся людям.

ЗАКЛЮЧЕНИЕ

Так как существует множество систем искусственного интеллекта, то возникает необходимость сопоставимой оценки качества их математических моделей. Одним из вариантов решения этой задачи является тестирование различных системы на общей базе исходных данных, для чего очень удобно использовать общедоступную базу Kaggle.

В данной курсовой работе был показан пример использования общедоступной базы фильмов IMDB для анализа рейтинга фильмов. При этом наиболее достоверной в данном приложении оказалась модель INF5, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Семантический резонанс». Точность модели составляет 0,488, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 48,4%.

При рассмотрении модели в перспективе ретроанализа можно убедиться в достоверности моделей: история изменений предпочтений людей в жанрах фильмов подтверждает выводы, получаемые в процессе выполнения задач курсовой работы.

На основе базы данных Kaggle, рассмотренной в данной курсовой работе, можно построить модели прогнозирования не с помощью АСК-анализа и реализующей его системы «Aidos-X», а с применением других математических методов и реализующих их программных систем, а затем сравнить их качество.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Луценко Е.В. Методика использования репозитория UCI для оценки качества математических моделей систем искусственного интеллекта / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: http://lc.kubagro.ru/My_training_schedule.doc КубГАУ, 2003. – №02(002). С. 120 – 145. – IDA [article ID]: 0020302012. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 у.п.л. 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.

2. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 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.

5. [Электронный ресурс]. Статья "baza dannix sotrudnikov": <http://allexcel.ru/gotovye-tablitsy-excel-besplatno>, свободный. - Загл. с экрана. Яз. англ.

6. Официальный сайт World of Warcraft [Электронный ресурс]. Режим доступа: <http://eu.battle.net/wow/ru>, свободный. - Загл. с экрана. Яз. рус.

7. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее время / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 48 – 77. – Шифр Информрегистра: 0420900012\0110, IDA [article ID]: 0540910004. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>, 1,875 у.п.л.

8. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос- Х++» / 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.

9. Луценко Е.В., Боровко А.Ю. ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА И КЛАССОВ СОЛНЕЧНЫХ ВСПЫШЕК НА ОСНОВЕ ИХ ПРЕДЫСТОРИИ ПО ДАННЫМ РЕПОЗИТОРИЯ UC1 С ПРИМЕНЕНИЕМ АСК- АНАЛИЗА И ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ «ЭЙДОС» // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – IDA [article ID]: 0831209025. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/99.pdf>