

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) Кравченко Александра Сергеевича
курса 3 очной (заочной) формы обучения
направления подготовки Информационные системы и технологии
направленность(профиль)Интеллектуальные информационные системы и технологии.
Наименование темы «Исследование рейтинга лучших музыкальных альбомов с использованием АСК-анализа и системы Эйдос».

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/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 И ТЕХНОЛОГИИ

на тему:

**Исследование рейтинга лучших музыкальных альбомов с
использованием АСК-анализа и системы Эйдос**

выполнил студент группы ИТ1421 **Кравченко Александр Сергеевич**

Допущена к защите

Руководитель проекта Луценко Евгений Вениаминович, д.э.н., к.т.н.,
профессор

(подпись, расшифровка подписи)

Нормоконтролер Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор

(подпись, расшифровка подписи)

Защищена _____

Оценка _____

(дата)

Члены комиссии	_____	В.И. Лойко
	_____	Е.В. Луценко
	_____	В.Н. Лаптев

(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: ИТ1421 группы 3 курса
Факультета Прикладной информатики
Специальности: 09.03.02 Информационные системы и технологии
(шифр)

Кравченко Александру Сергеевичу
(Ф.И.О.)

Тема проекта: **Исследование рейтинга лучших музыкальных альбомов с использованием АСК системы Эйдос**

Содержание задания: Проанализировать современные методы и средства формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами

Объем работы:

а) пояснительная записка к работе _____ 31 _____ листа формата А4
б) графическая часть _____ 21 _____ лист формата А4

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

Срок выполнения проекта: с “ _____ ” _____ по “ _____ ” _____ 20__ г.

Срок защиты: “ _____ ” _____ 20__ г.

Дата выдачи задания: “ _____ ” _____ 20__ г.

Дата сдачи проекта на кафедру: “ _____ ” _____ 20__ г.

Руководитель проекта: Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор
(подпись, Ф.И.О., звание, степень)

Задание принял студент _____
(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	17
2.1. Выделение основных задач.....	17
2.2. Выбор модели и пакетное распознавание	17
2.3 Визуализация когнитивных функций	21
2.4 SWOT и PEST матрицы и диаграммы	23
2.5 Нелокальные нейронные сети и нейтроны.....	26
2.6 Кластерный и конструктивный анализ.....	27
ЗАКЛЮЧЕНИЕ	29
СПИСОК ИСПОЛЬЗОВАННЫХ 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 рассмотрено решение задачи по определению музыкальных предпочтений людей по версии журнала «Rolling Stone».

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. Удачным выбором является сборник баз данных Kaggle.

В данной курсовой работе использована база данных «wow avatar history» 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: greatest_albums.csv по ссылке: <https://www.kaggle.com/notgibs/500-greatest-albums-of-all-time-rolling-stone>.

Это база данных объединения лучших альбомов всего времени, по мнению журнала «Rolling Stone». Она содержит:

- Артиста.
- Год.
- Альбом.
- Жанр.
- Поджанр.

Это нам позволит понять, какие жанры и поджанры музыки предпочитают люди, какие группы самые популярные и в какой период времени были выпущены самые популярные альбомы.

Преобразуем файлы в формат *.xlsx, в итоге получается следующая таблица для работы:

Таблица 1 – Фрагмент обучающей выборки

Artist	Album	Year	Album	Genre	Subgenre
A Tribe Called Quest	The Low End Theory	1991	The Low End Theory	Hip Hop	Conscious
ABBA	The Definitive Collection	2001	The Definitive Collection	Electronic, Pop	Europop, Synth-pop, Disco
AC/DC	Back in Black	1980	Back in Black	Rock	Hard Rock
AC/DC	Highway to Hell	1979	Highway to Hell	Rock	Hard Rock
Aerosmith	Rocks	1976	Rocks	Rock	Hard Rock, Classic Rock
Aerosmith	Toys in the Attic	1975	Toys in the Attic	Rock	Hard Rock, Blues Rock
Al Green	Greatest Hits	1975	Greatest Hits	Funk / Soul	Soul
Al Green	I'm Still in Love With You	1972	I'm Still in Love With You	Funk / Soul	Rhythm & Blues, Soul
Al Green	Call Me	1973	Call Me	Funk / Soul	Rhythm & Blues, Soul
Albert King	Born Under a Bad Sign	1967	Born Under a Bad Sign	Funk / Soul, Blues	Electric Blues, Rhythm & Blues, Soul
Alice Cooper	Love It to Death	1971	Love It to Death	Rock	Classic Rock
Amy Winehouse	Back to Black	2006	Back to Black	Funk / Soul, Pop	Soul
Arcade Fire	Funeral	2004	Funeral	Rock	Indie Rock
Arctic Monkeys	Whatever People Say I Am, That's What I'm Not	2006	Whatever People Say I Am, That's What I'm Not	Rock	Indie Rock
Aretha Franklin	I Never Loved a Man the Way I Love You	1967	I Never Loved a Man the Way I Love You	Funk / Soul	Soul

Aretha Franklin	Lady Soul	1968	Lady Soul	Funk / Soul	Soul
B.B. King	Live at the Regal	1965	Live at the Regal	Blues	Chicago Blues
B.B. King	Live in Cook County Jail	1971	Live in Cook County Jail	Blues	Electric Blues
Barry White	Can't Get Enough	1974	Can't Get Enough	Funk / Soul	Soul, Disco
Beastie Boys	Paul's Boutique	1989	Paul's Boutique	Hip Hop, Rock, Funk / Soul	Alternative Rock, Pop Rap, Psychedelic
Beastie Boys	Licensed to Ill	1986	Licensed to Ill	Hip Hop	None
Beck	Odelay	1996	Odelay	Electronic, Hip Hop, Funk / Soul, Pop	Electro, Downtempo, Hip Hop, Disco, Afrobeat, Abstract
Beck	Sea Change	2002	Sea Change	Rock	Alternative Rock, Post Rock
Big Brother & the Holding Company	Cheap Thrills	1968	Cheap Thrills	Rock	Folk Rock, Blues Rock, Psychedelic Rock

Классификационная шкала следующая: артист, альбом. 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 d:\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.

Стоит отметить следующие настройки:

- Тип файла – **xlsx**.
- Классификационные шкалы – 2-3.
- Описательные шкалы – 4-6.

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 - Comma-Separated Values

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_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":

В качестве классов рассматриваются:

☒ Значения полей целиком
☐ Элементы значений полей - слова
☐ Элементы значений полей - символы

☒ Выделять уникальные значения и сортировать
☐ Не выделять уникальных значений и не сортировать

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

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).

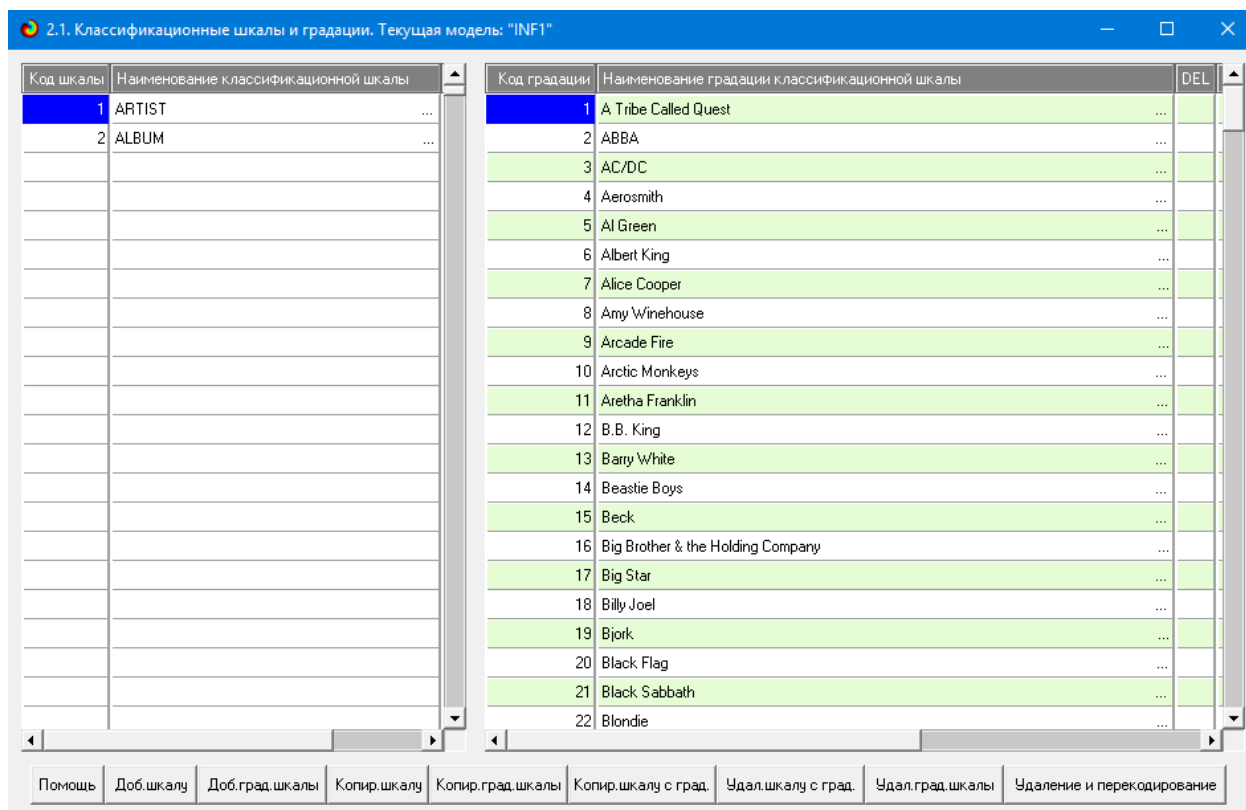


Рисунок 2 – Классификационные шкалы

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.

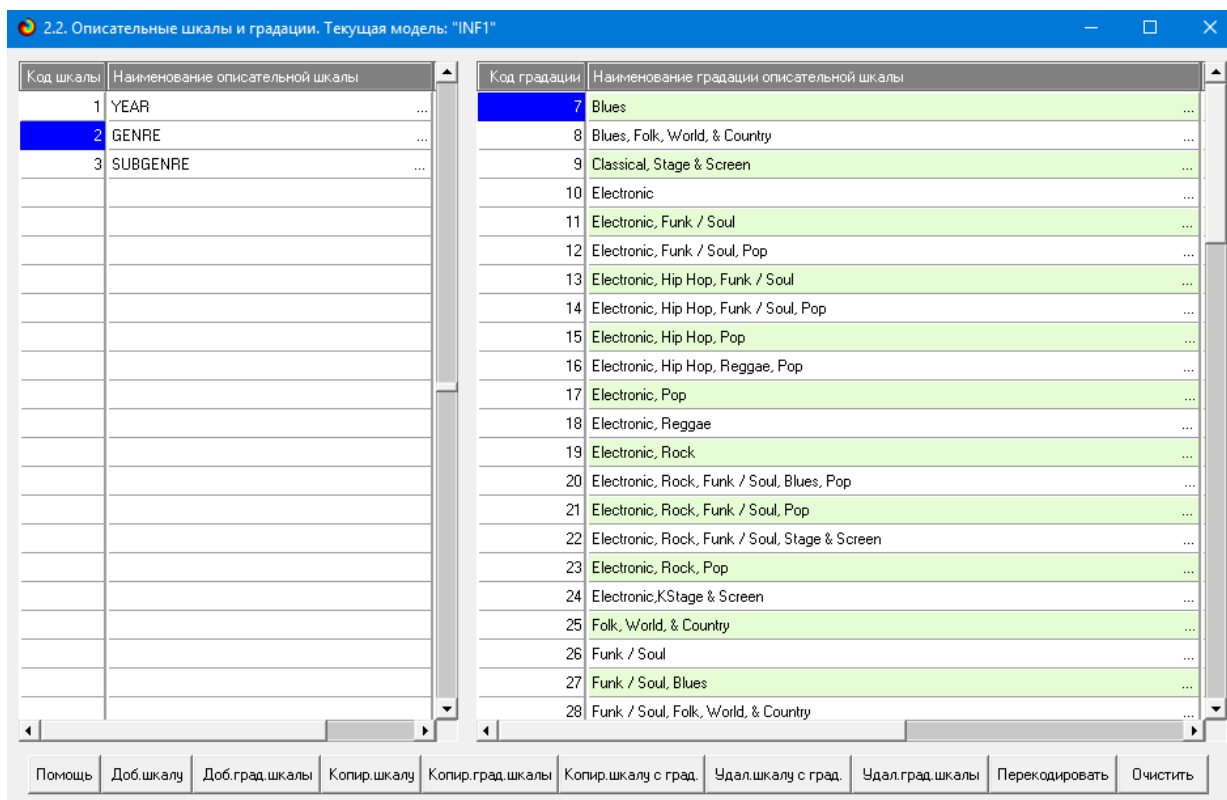


Рисунок 3 – Описательные шкалы

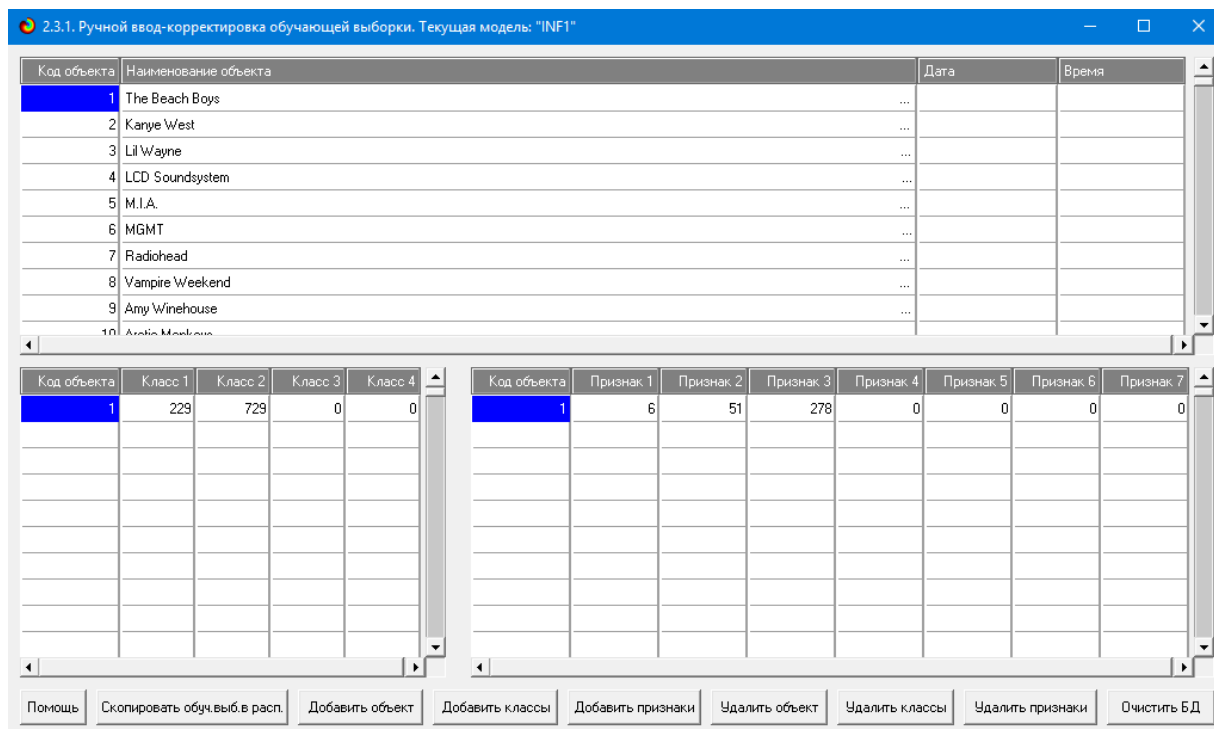


Рисунок 4 – Ручной ввод-корректировка выборки

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.

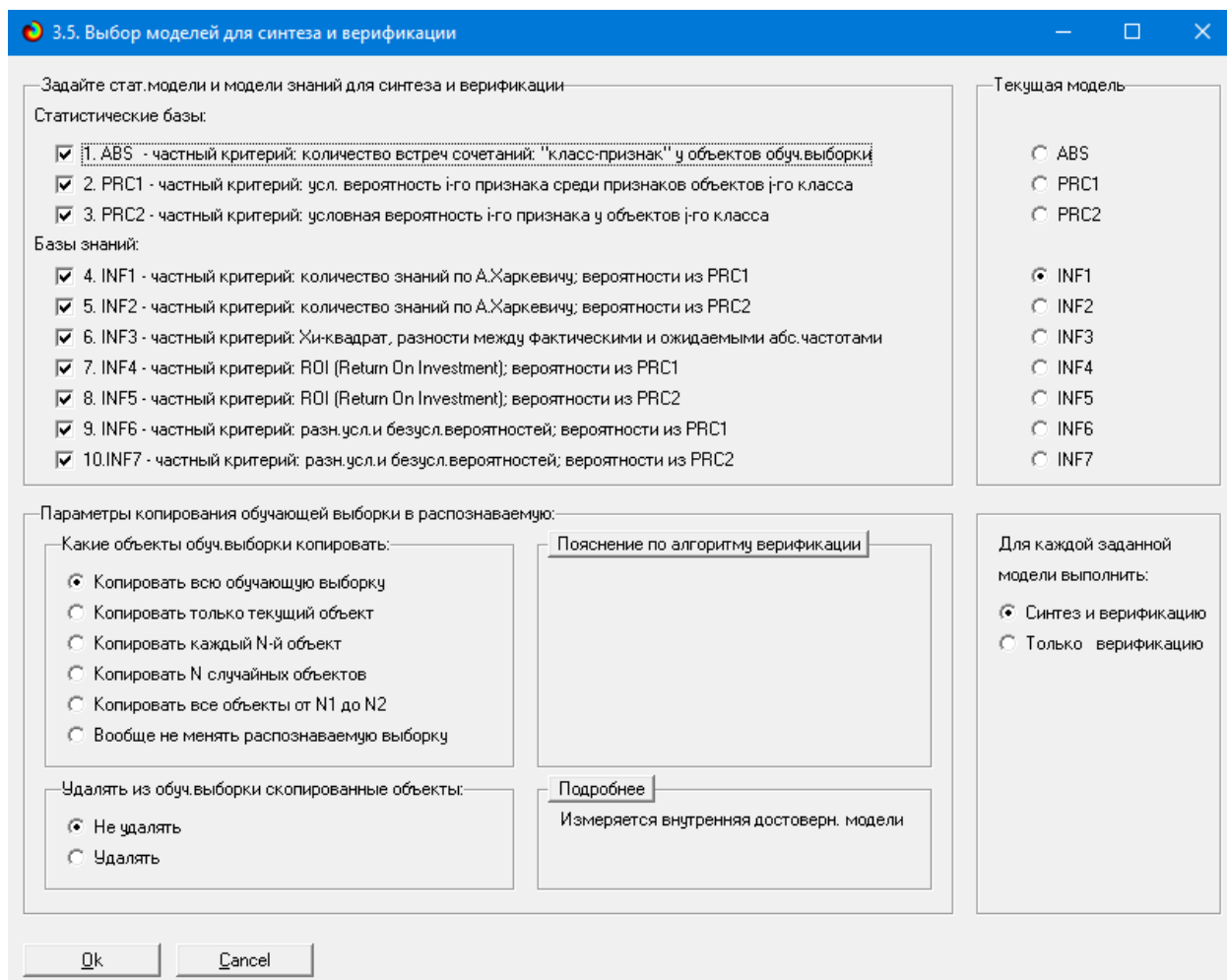


Рисунок 5 – Настройка синтеза и верификации моделей

После построения моделей нажимаем «ОК» на форме, которая показана на рисунке 6.

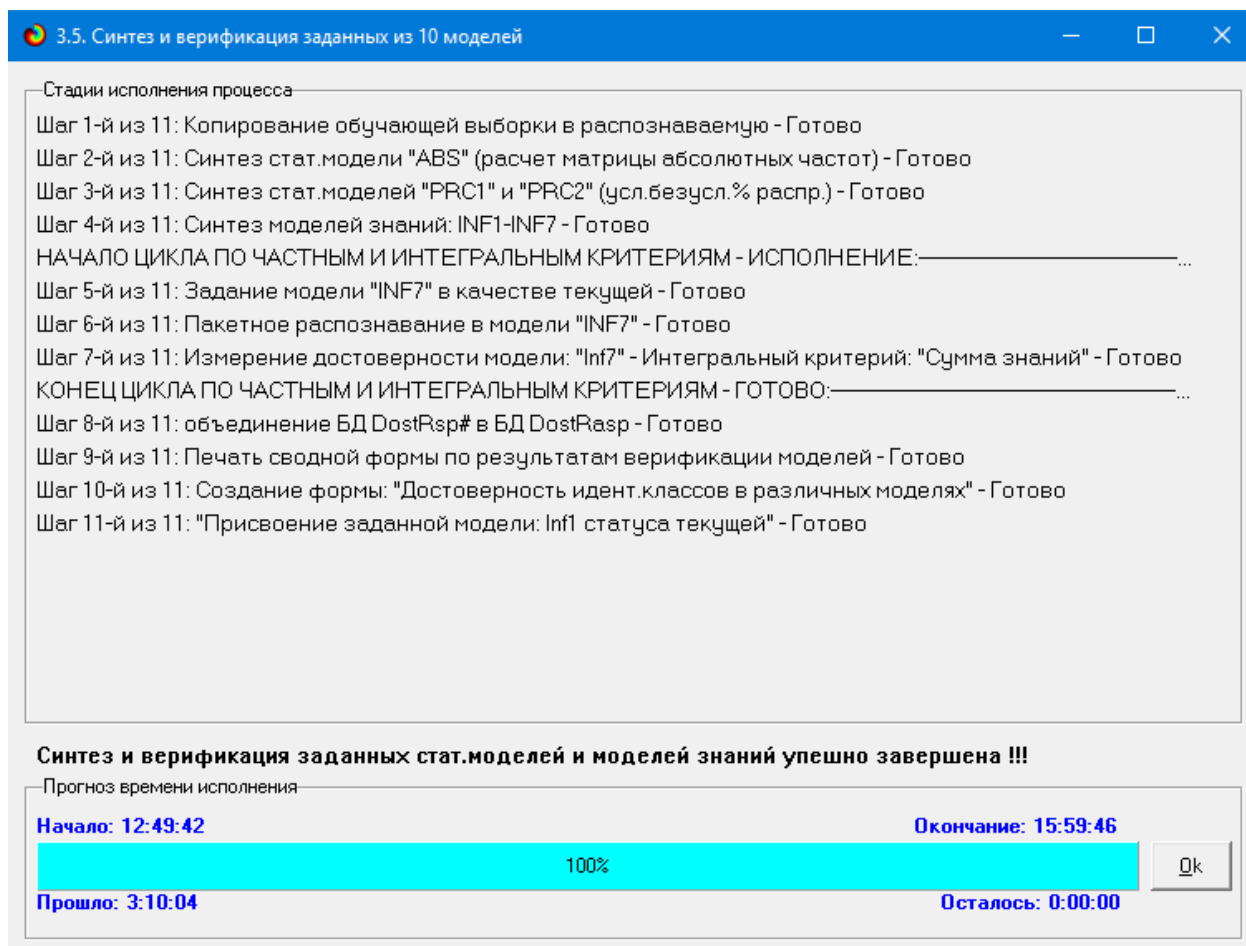


Рисунок 6 – Завершение процесса синтеза и верификации моделей

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).

[illegible][illegible]

Таблица 4 – Матрица знаний (модель INF3)

5.5. Модель: "6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами"															
Код признака	Наименование описательной шкалы и градации	1. ARTIST A TRIBE CALLED QUEST	2. ARTIST ABBA	3. ARTIST AC/DC	4. ARTIST AEROSMITH	5. ARTIST AL GREEN	6. ARTIST ALBERT KING	7. ARTIST ALICE COOPER	8. ARTIST AMY WINEHOUSE	9. ARTIST ARCADE FIRE	10. ARTIST ARCTIC MONKEYS	11. ARTIST ARETHA FRANKLIN	12. ARTIST B.B. KING	13. ARTIST BARRY WHITE	14. ARTIST BEATLES
1	YEAR-1/6(1955.000000, 1964.333333)	...	-0.051	-0.051	-0.101	-0.101	-0.152	-0.051	-0.051	-0.051	-0.051	-0.051	-0.101	-0.101	-0.051
2	YEAR-2/6(1964.333333, 1973.666667)	...	-0.368	-0.368	-0.736	-0.736	0.896	0.632	0.632	-0.368	-0.368	-0.368	1.264	1.264	-0.368
3	YEAR-3/6(1973.666667, 1983.000000)	...	-0.240	-0.240	1.521	1.521	0.281	-0.240	-0.240	-0.240	-0.240	-0.240	-0.479	-0.479	0.760
4	YEAR-4/6(1983.000000, 1992.333333)	...	0.837	-0.163	-0.326	-0.326	-0.488	-0.163	-0.163	-0.163	-0.163	-0.163	-0.326	-0.326	-0.163
5	YEAR-5/6(1992.333333, 2001.666667)	...	-0.127	0.873	-0.255	-0.255	-0.382	-0.127	-0.127	-0.127	-0.127	-0.127	-0.255	-0.255	-0.127
6	YEAR-6/6(2001.666667, 2011.000000)	...	-0.052	-0.052	-0.103	-0.103	-0.155	-0.052	0.948	0.948	0.948	-0.103	-0.103	-0.103	-0.052
7	GENRE-Blues	...	-0.016	-0.016	-0.032	-0.032	-0.049	-0.016	-0.016	-0.016	-0.016	-0.016	-0.032	1.968	-0.016
8	GENRE-Blues, Folk, World, & Country	...	-0.002	-0.002	-0.004	-0.004	-0.006	-0.002	-0.002	-0.002	-0.002	-0.002	-0.004	-0.004	-0.002
9	GENRE-Classical, Stage & Screen	...	-0.001	-0.001	-0.002	-0.002	-0.003	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002	-0.002	-0.001
10	GENRE-Electronic	...	-0.014	-0.014	-0.028	-0.028	-0.042	-0.014	-0.014	-0.014	-0.014	-0.014	-0.028	-0.028	-0.014
11	GENRE-Electronic, Funk / Soul	...	-0.004	-0.004	-0.008	-0.008	-0.012	-0.004	-0.004	-0.004	-0.004	-0.004	-0.008	-0.008	-0.004
12	GENRE-Electronic, Funk / Soul, Pop	...	-0.002	-0.002	-0.004	-0.004	-0.006	-0.002	-0.002	-0.002	-0.002	-0.002	-0.004	-0.004	-0.002
13	GENRE-Electronic, Hip Hop, Funk / Soul	...	-0.002	-0.002	-0.004	-0.004	-0.006	-0.002	-0.002	-0.002	-0.002	-0.002	-0.004	-0.004	-0.002
14	GENRE-Electronic, Hip Hop, Funk / Soul, Pop	...	-0.002	-0.002	-0.004	-0.004	-0.006	-0.002	-0.002	-0.002	-0.002	-0.002	-0.004	-0.004	-0.002
15	GENRE-Electronic, Hip Hop, Pop	...	-0.002	-0.002	-0.004	-0.004	-0.006	-0.002	-0.002	-0.002	-0.002	-0.002	-0.004	-0.004	-0.002
16	GENRE-Electronic, Hip Hop, Reggae, Pop	...	-0.002	-0.002	-0.004	-0.004	-0.006	-0.002	-0.002	-0.002	-0.002	-0.002	-0.004	-0.004	-0.002
17	GENRE-Electronic, Pop	...	-0.010	0.990	-0.020	-0.020	-0.030	-0.010	-0.010	-0.010	-0.010	-0.010	-0.020	-0.020	-0.010

1.5. Результаты верификации моделей

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

4.13.6. Обобщ.форма по достов.моделям при разн.инт.крит. Текущая модель: "INF1"														
Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризенбергена	Сумма мод. уровней схож. истинно-полож. решений (ST ₊)	Сумма мод. уровней схож. истинно-отриц. решений (ST ₋)	Сумма мод. уровней схож. ложно-полож. решений (SFP)	Сумма мод. уровней схож. ложно-отриц. решений (SF)	
1. ABS - частный критерий: количество встреч сомнений; "х/вас."	Корреляция абс.частот с обр...	493	493	109856	84401		0.006	1.000	0.120	457.866	947.081	32837.123		
1. ABS - частный критерий: количество встреч сомнений; "х/вас."	Сумма абс.частот по признакам...	493	493	109856	84401		0.006	1.000	0.120	114.556		6984.111		
2. PRCI - частный критерий: усл. вероятность i-го признака сред.	Корреляция усл.отн.частот с о...	493	493	109856	84401		0.006	1.000	0.120	457.866	947.081	32837.126		
2. PRCI - частный критерий: усл. вероятность i-го признака сред.	Сумма усл.отн.частот по признакам...	493	493	109856	84401		0.006	1.000	0.120	434.054		32228.849		
3. PRCI2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн.частот с о...	493	493	109856	84401		0.006	1.000	0.120	457.864	947.080	32837.055		
3. PRCI2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн.частот по признакам...	493	493	109856	84401		0.006	1.000	0.120	434.054		32228.849		
4. INF1 - частный критерий: количество знаний по АХаркевичу; в.	Семантический резонанс зна...	493	493	111849	82408		0.006	1.000	0.120	368.071	883.135	10477.609		
4. INF1 - частный критерий: количество знаний по АХаркевичу; в.	Сумма знаний	493	493	110776	83481		0.006	1.000	0.120	265.461	24.793	6917.689		
5. INF2 - частный критерий: количество знаний по АХаркевичу; в.	Семантический резонанс зна...	493	493	111849	82408		0.006	1.000	0.120	368.071	883.135	10477.610		
5. INF2 - частный критерий: количество знаний по АХаркевичу; в.	Сумма знаний	493	493	110776	83481		0.006	1.000	0.120	265.461	24.793	6917.688		
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Семантический резонанс зна...	493	493	112980	81277		0.006	1.000	0.120	445.476	18109.654	19283.890		
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Сумма знаний	493	493	112980	81277		0.006	1.000	0.120	168.603	6133.169	5964.567		
7. INF4 - частный критерий: ROI (Return On Investment); вероимо...	Семантический резонанс зна...	493	493	154616	39641		0.012	1.000	0.402	285.441	809.187	2086.372		
7. INF4 - частный критерий: ROI (Return On Investment); вероимо...	Сумма знаний	493	493	110730	83527		0.006	1.000	0.201	67.414	0.141	175.065		
8. INF5 - частный критерий: ROI (Return On Investment); вероимо...	Семантический резонанс зна...	493	493	154616	39641		0.012	1.000	0.402	285.441	809.187	2086.372		
8. INF5 - частный критерий: ROI (Return On Investment); вероимо...	Сумма знаний	493	493	110730	83527		0.006	1.000	0.201	67.414	0.141	175.064		
9. INF6 - частный критерий: разн.усл.и безуслов.вероятностей; вер...	Семантический резонанс зна...	493	493	111611	82646		0.006	1.000	0.120	443.735	1022.651	24785.933		
9. INF6 - частный критерий: разн.усл.и безуслов.вероятностей; вер...	Сумма знаний	493	493	110737	83520		0.006	1.000	0.120	355.414	44.148	19934.425		
10.INF7 - частный критерий: разн.усл.и безуслов.вероятностей; ве...	Семантический резонанс зна...	493	493	111611	82646		0.006	1.000	0.120	443.735	1022.651	24785.933		
10.INF7 - частный критерий: разн.усл.и безуслов.вероятностей; ве...	Сумма знаний	493	493	110737	83520		0.006	1.000	0.120	355.414	44.148	19934.425		

Рисунок 7 – Результаты верификации моделей

Наиболее достоверной в данном приложении оказалась модель INF4 при интегральном критерии «Семантический резонанс знаний», если учитывать F-меру Ван Ризенбергена. Точность модели составляет 0,402, то есть уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 40,2%.

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 INF4 (рисунок 8). Для этого зайдём в режим 4.1.3.11, выберем модель INF4 и запросим график по сумме знаний.

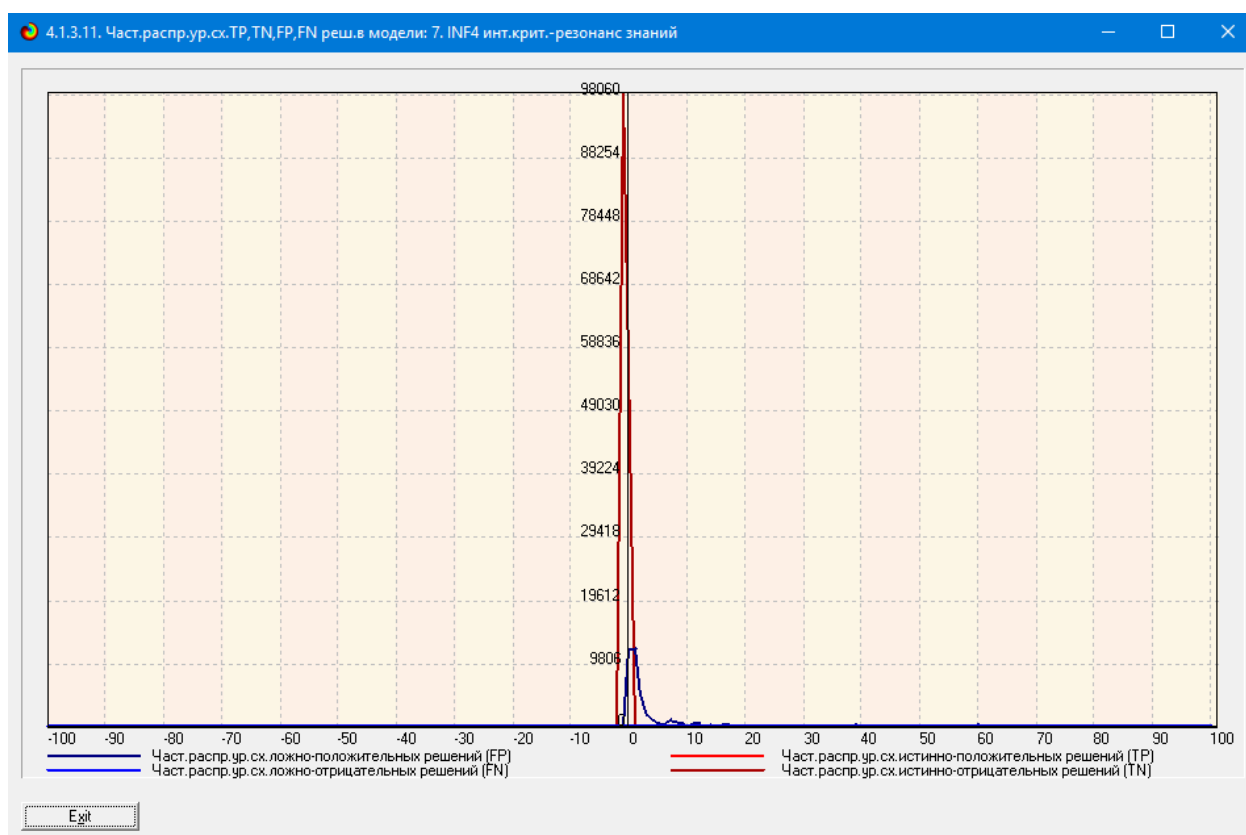


Рисунок 8 – Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF4

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 музыканта, то какие задачи при анализе баз данных можно выделить?

Во-первых, зависимость выбора жанра музыки. Зачастую популярность песни зависит от выбора музыкантом того или иного жанра музыки, если выбрать не популярный жанр, то можно остаться без большого количества поклонников, а для музыканта это один из главных критериев.

Во-вторых, жанр музыки зависит напрямую от годов жизни музыканта, песня должна соответствовать трендам и проблемам текущего времени.

В-третьих, название альбома должно быть громким, легким и брутальным, чтобы при прочтении названия можно было испытать волнение и страсть.

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

- Артист – альбом
- Альбом – жанр
- Артист – год
- Жанр – поджанр

2.2. Выбор модели и пакетное распознавание

Зададим модель INF4 в качестве текущей в режиме 5.6 (рисунок 9), после чего проведем пакетное распознавание в режиме 4.1.2 (рисунок 10).

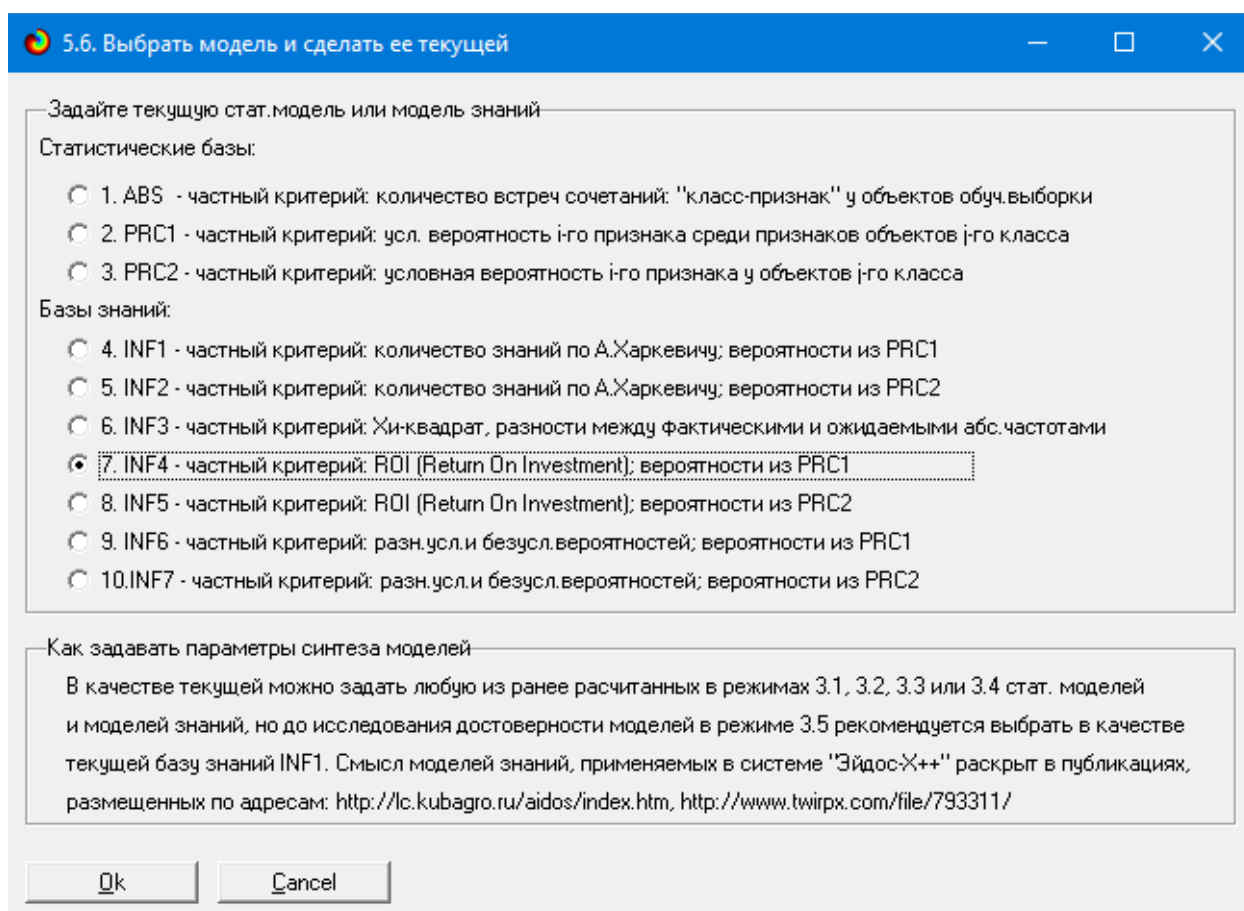


Рисунок 9 – Выбор текущей модели

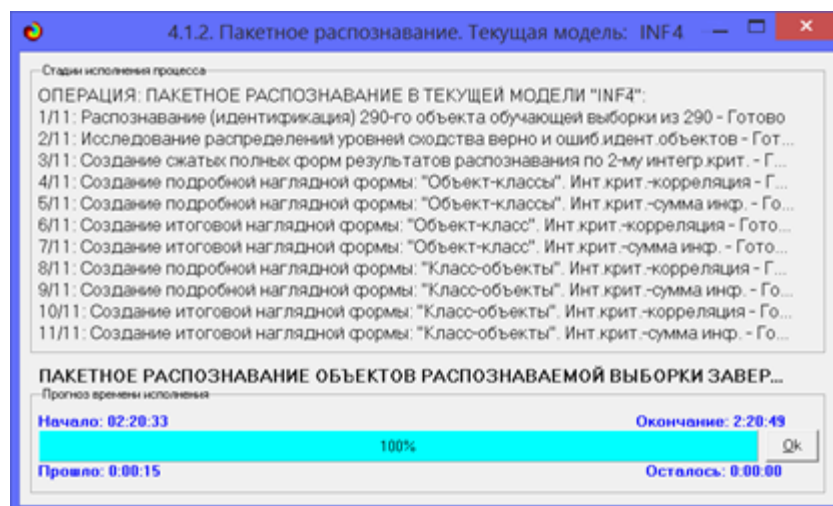


Рисунок 10 – Пакетное распознавание в новой модели

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 INF4:

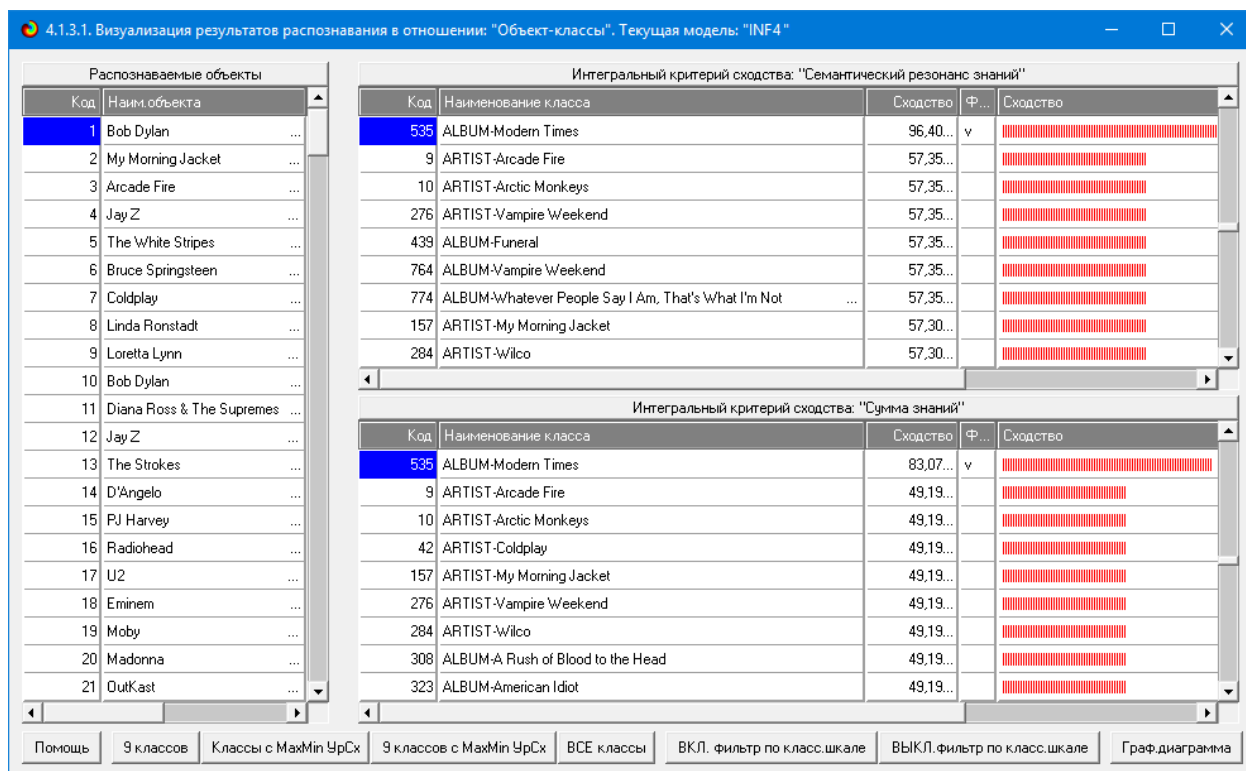


Рисунок 11 – Прогнозы достоверности классов объектов

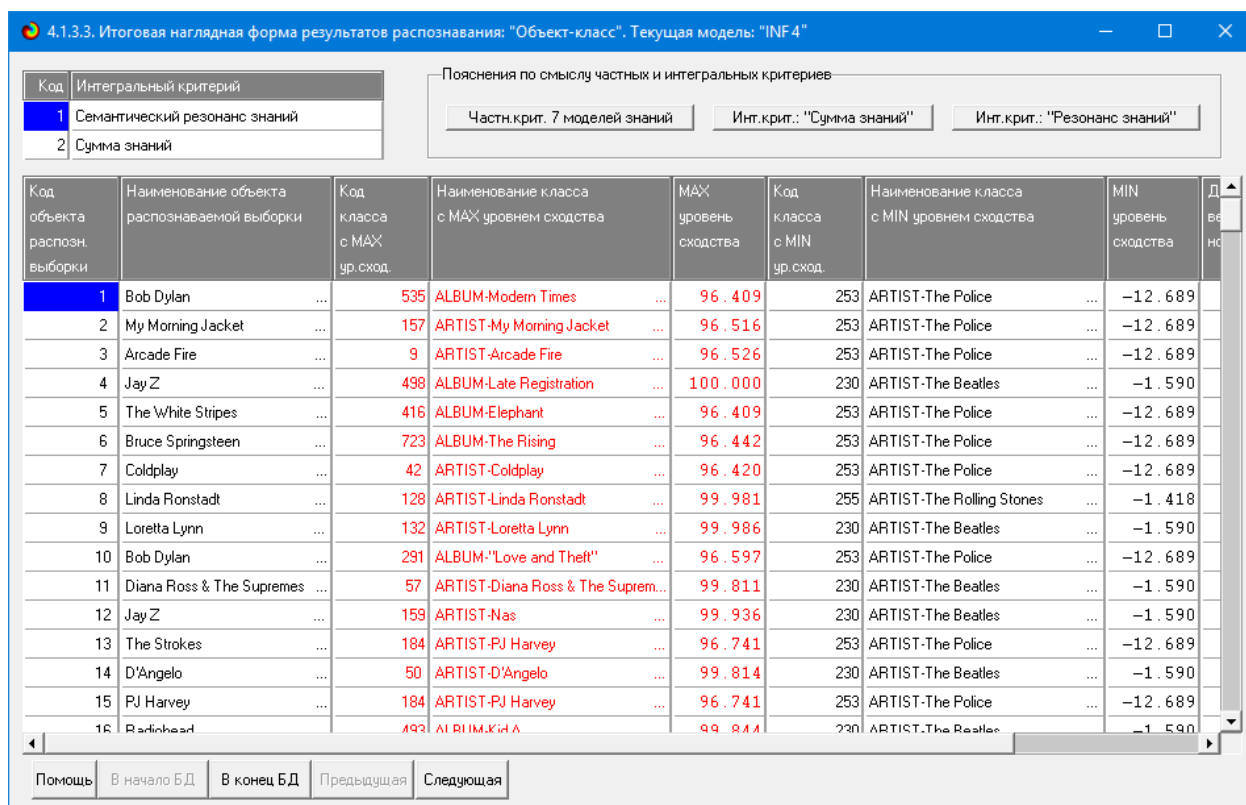


Рисунок 12 – Итоги распознавания классов объектов

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.

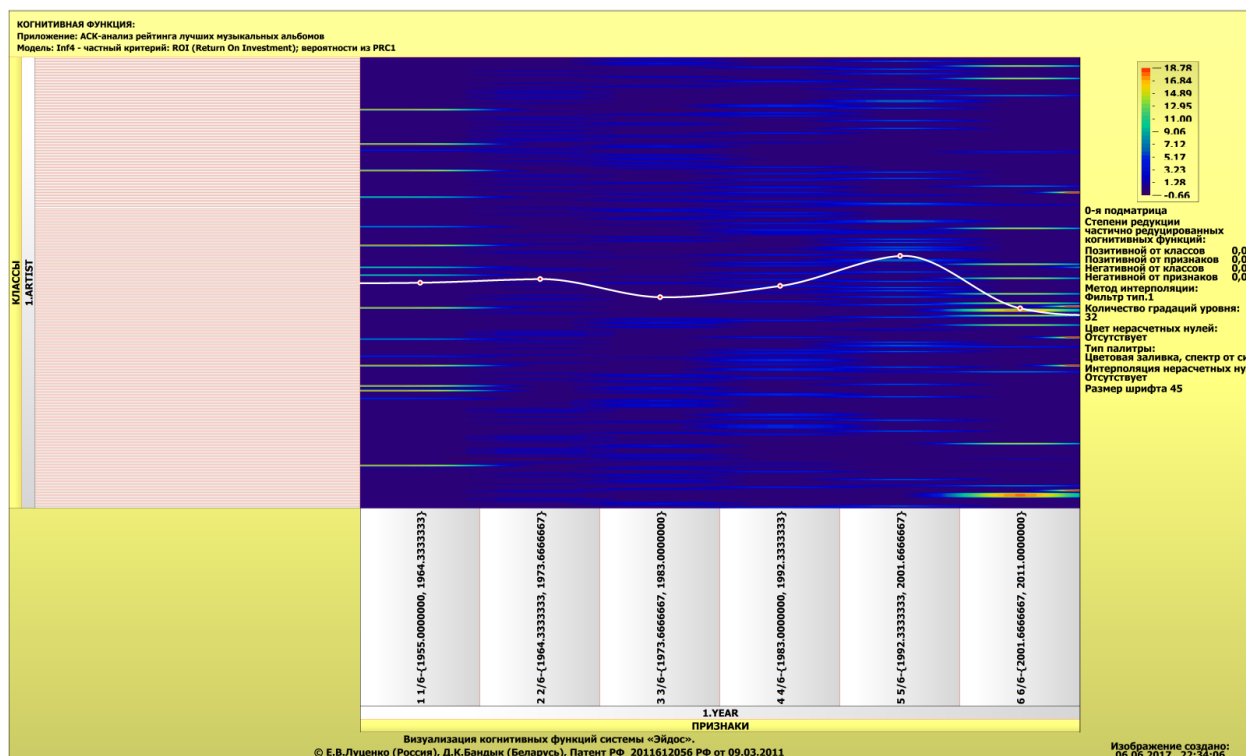


Рисунок 13 – Визуализация зависимости артистов от года

Рассмотрим визуализацию когнитивных функций в рамках определенных задач.

Зависимость популярности артистов от года представлена на рисунке 13. Из него видно, что в период с 1955 г. по 1964 г. были очень популярные артисты. В период с 1964 г. по 2001 г. не было очень популярных музыкальных групп, они находились примерно на одном уровне популярности, а начиная с 2001 г. по 2011 г. четко выделяется тенденция «выстреливания» групп, эти группы за короткий промежуток времени становились знаменитыми.

Теперь перейдем к зависимости «артист-жанр». Она отражена на рисунке 14.

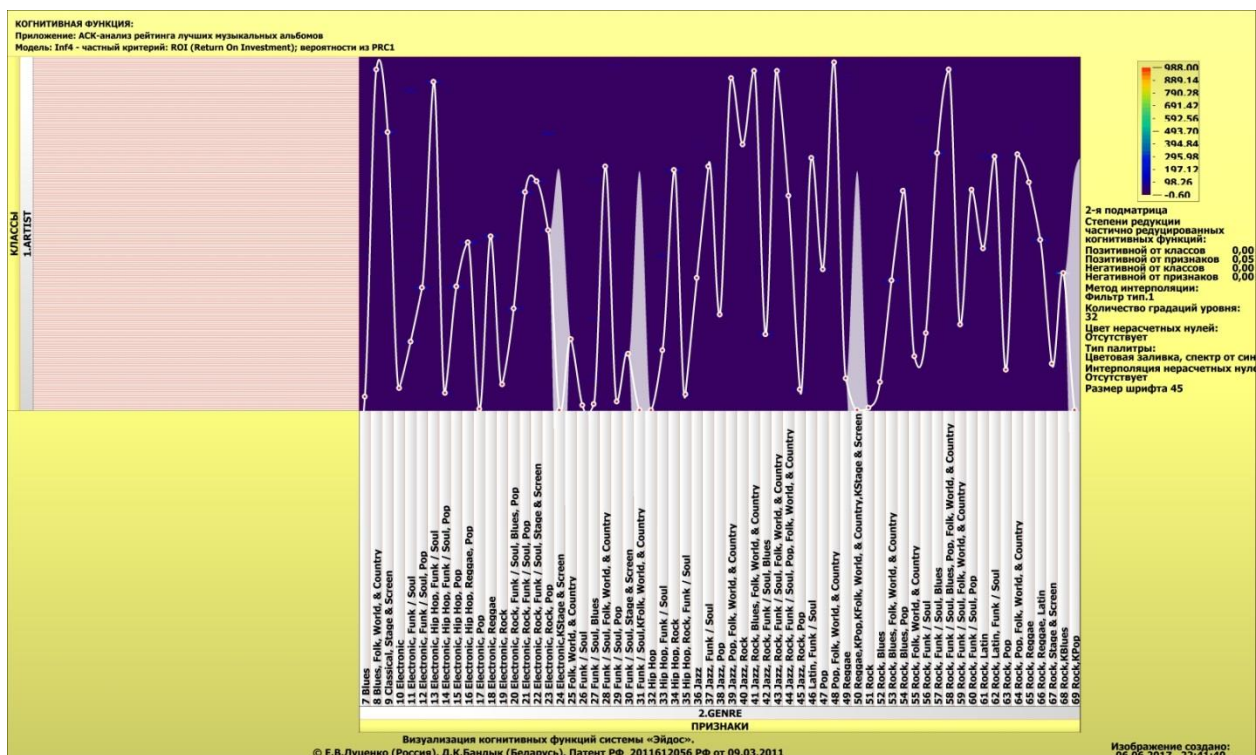


Рисунок 14 – Визуализация зависимости артиста от жанра

Blues и Country были довольно популярны на протяжении достаточно большого промежутка времени. Дело в том, что в афроамериканской среде к началу XX века сформировался стиль, оказавший впоследствии глобальное влияние на всю популярную музыку. Это Blues. В среде «белых» американцев как ответ «черным» сформировался музыкальный стиль, известный сейчас как «Country». В его основе европейские (преимущественно англо-кельтские) народные песни. Никакого негритянского воздействия на этот стиль как не было раньше, так нет и в наши дни, этот «заповедник белых» существует по сей день в почти неизменном виде. Нововведения произошли лишь с появлением электрогитары, содержание музыки при этом не изменилось.

С появлением электрогитары стали появляться новые жанры музыки такие как: electronic, funk, hip-hop, rock. Это был совершенно новый и необычный для европейского слуха манера игры музыкантов. Исполнение было очень резвым и даже несколько «диким» для того времени. Сейчас нас этим не удивить, мы слышали всякое, но на заре XX века это было очень круто.

Исходя из данных журнала «Rolling Stone» преимущественно лидируют жанры Rock, Hip-Hop и Electronic.

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 15).

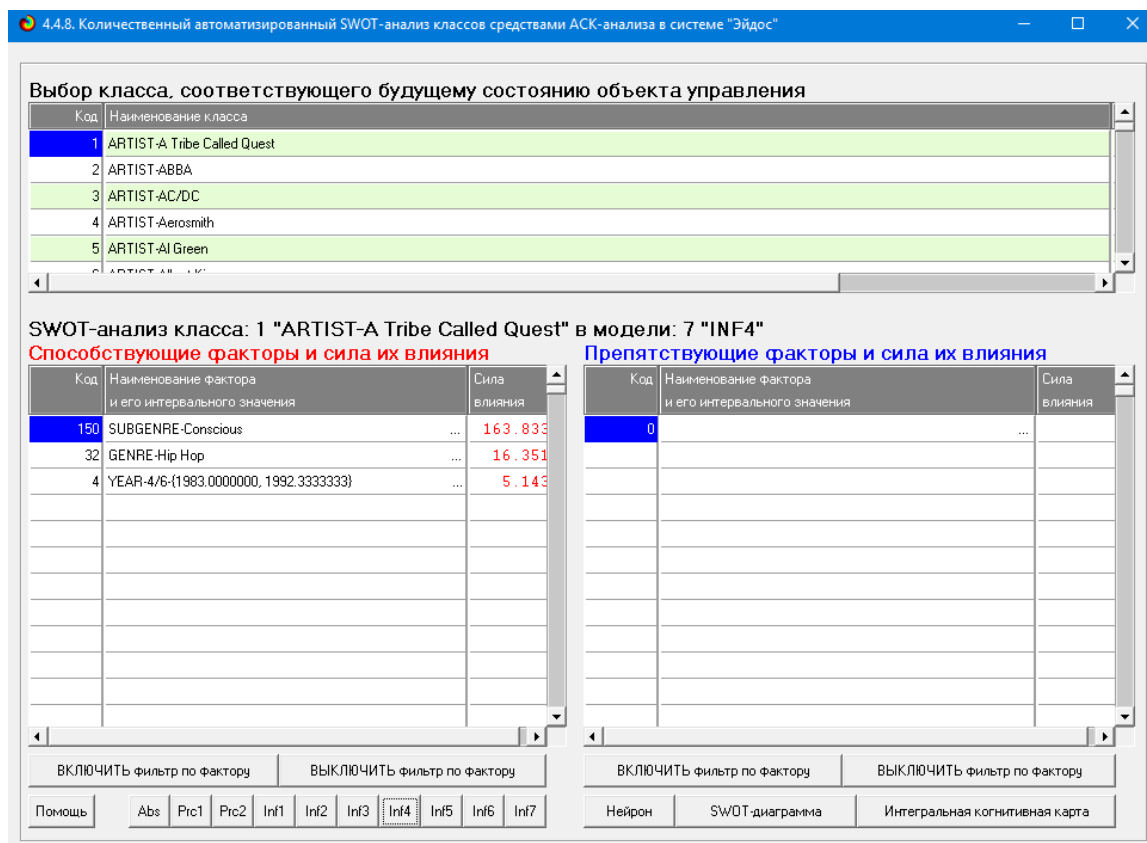


Рисунок 15 – Пример SWOT-Матрицы в модели INF4

На рисунке 16 приведен пример инвертированной SWOT-матрицы.

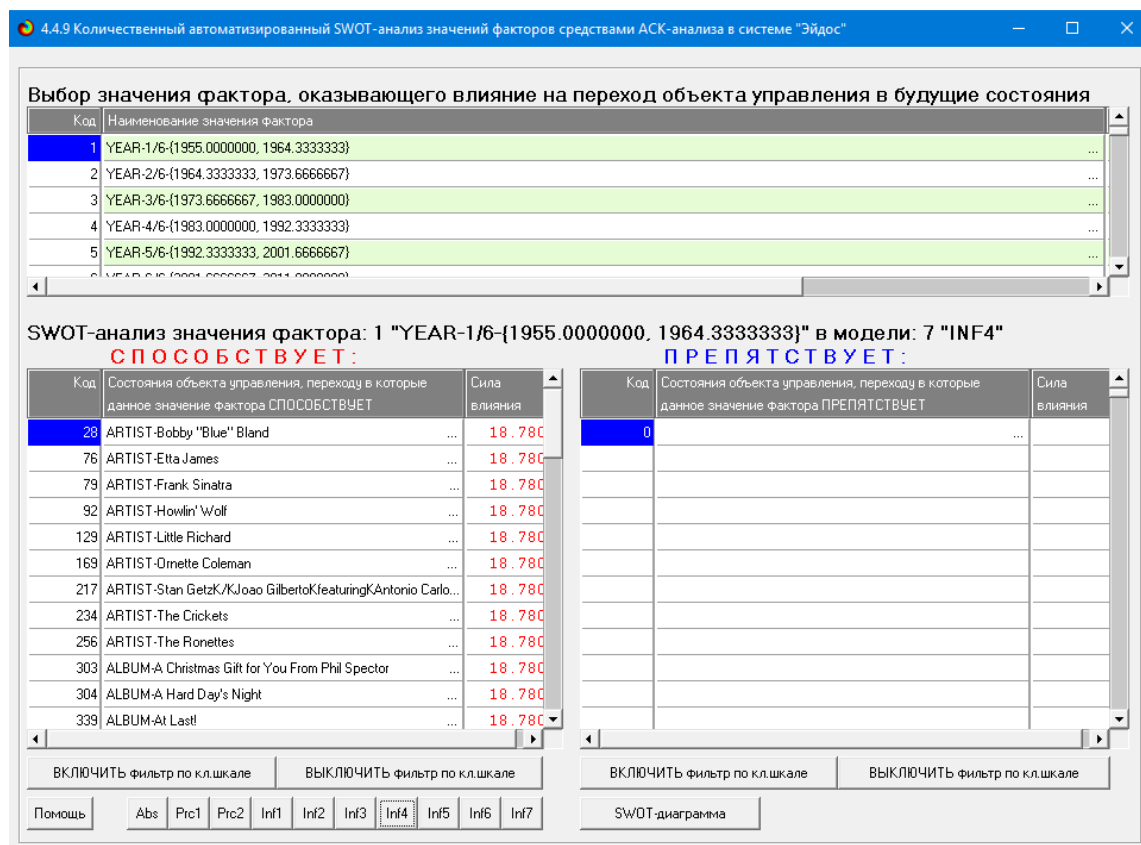


Рисунок 16 – Пример инвертированной SWOT-Матрицы в модели INF4

На примере группы «AC/DC» посмотрим в каких годах группа была популярна, какой жанр и поджанр предпочитала.

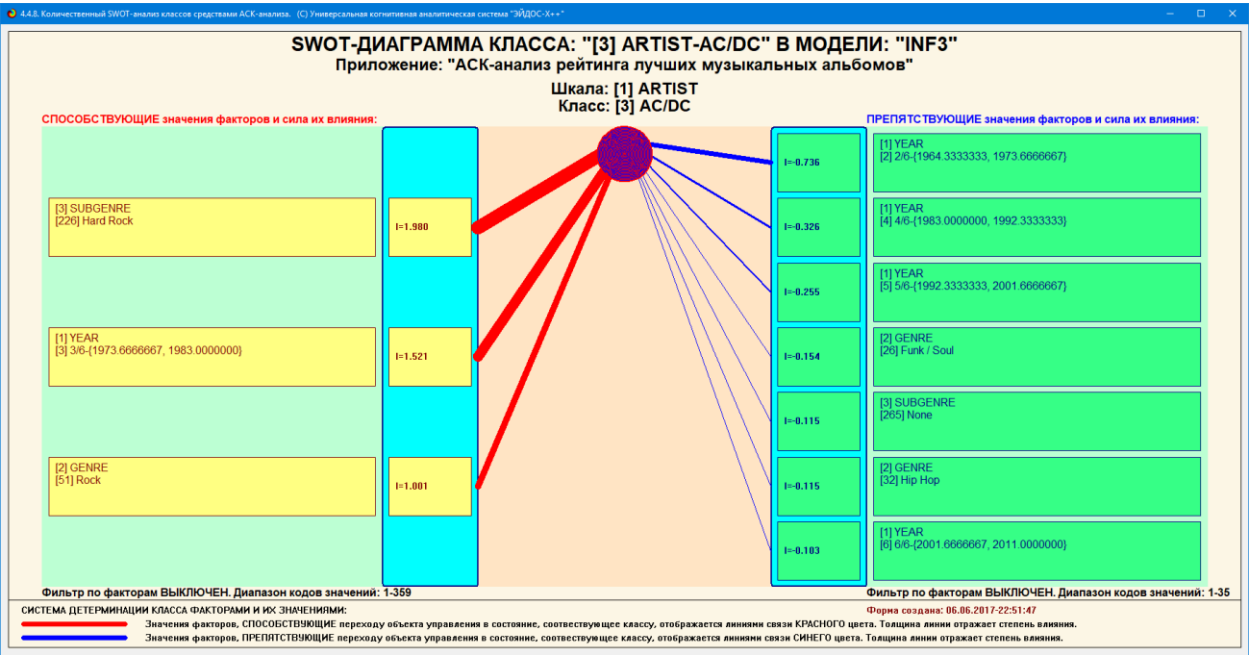


Рисунок 17 – SWOT-матрица, построенная для «AC/DC»

Как можно видеть, группа была популярна в 1973-1983 г., их любимый жанр Rock, а поджанр Hard Rock.

А теперь узнаем, кого вероятнее всего может встретиться в списке топ исполнителей, построив SWOT-диаграмму для (рисунок 18).

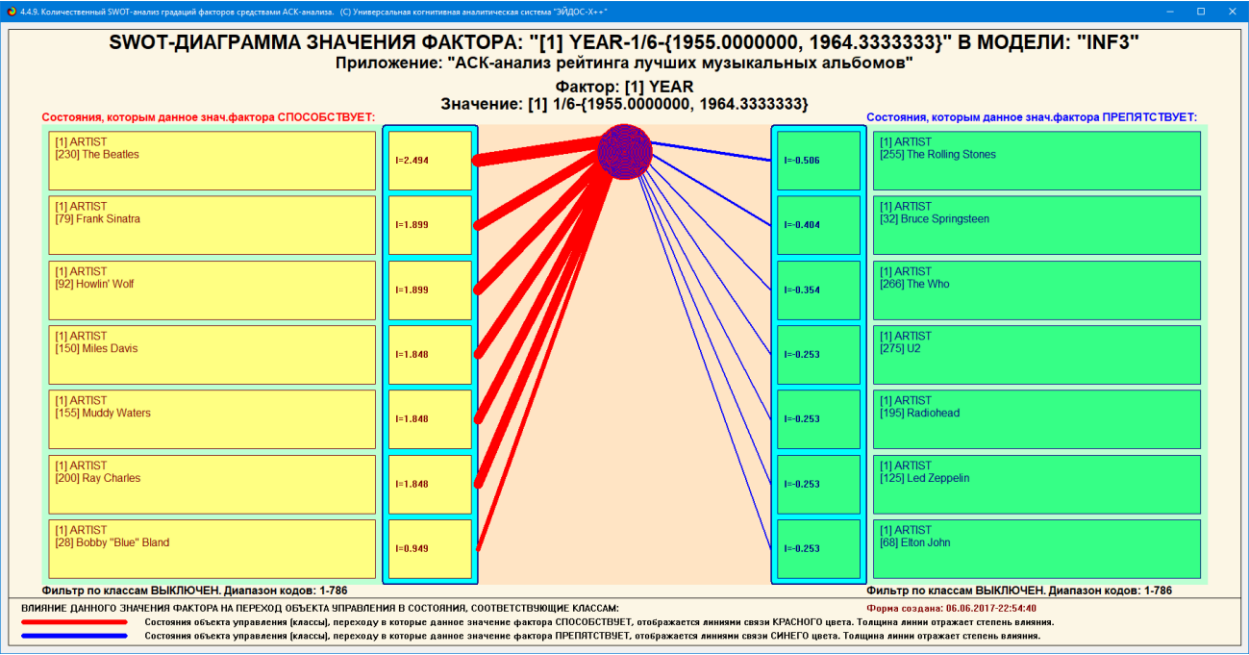


Рисунок 18 – SWOT-диаграмма признаков топ исполнителей

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 группы AC/DC.



Рисунок 19 – Нейрон для артиста AC/DC

На его основании можно определить, что AC/DC предпочитает жанры Rock, Hard Rock, именно эти жанры на сегодняшний день пользуется большей популярностью, чем другие.

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

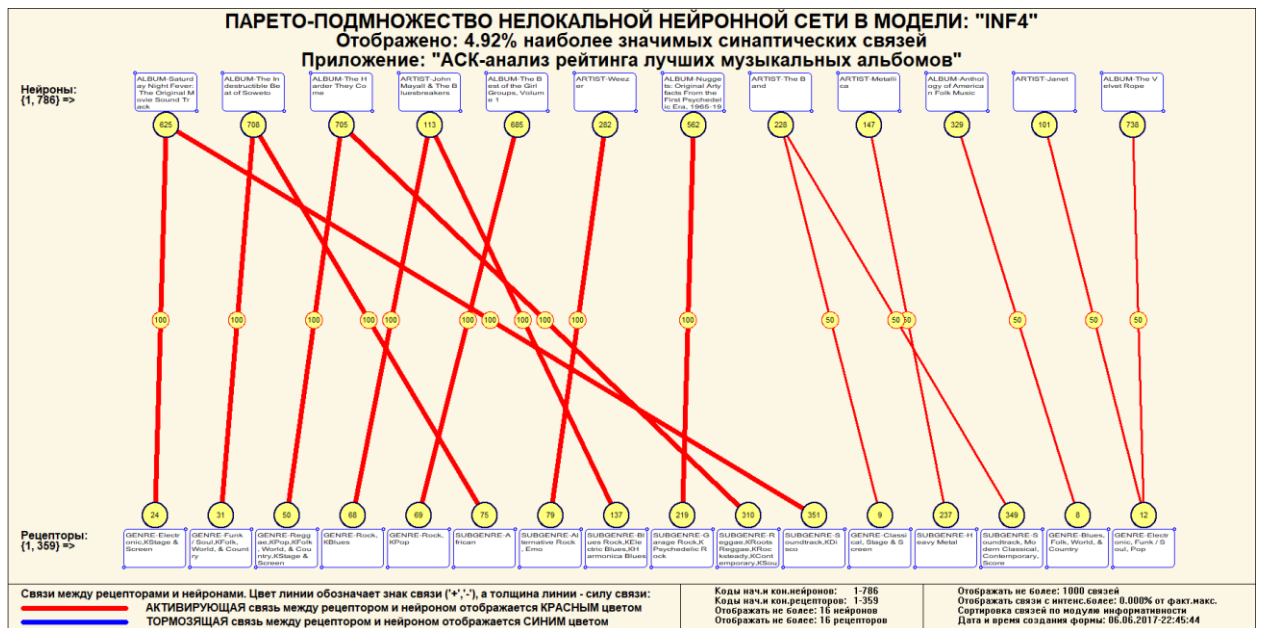


Рисунок 20 – Пример нелокальной нейронной сети в модели INF4

По ней можно, в основном, определить зависимость выбора самых популярных артистов жанра и поджанра.

2.6 Кластерный и конструктивный анализ

В режиме 4.2.2.2 после расчета матриц сходства, кластеров и конструкторов (4.2.2.1) строим 2D сеть классов в выбранной модели знаний (рисунок 21).

По этой модели мы определяем популярных артистов. Jackie Wilson был определен как самый популярный музыкант в процессе проведения анализа из-за его наиболее частом появлении в журнале и из-за его влияния на других артистов. Также можно видеть, что многие артисты тоже достаточно сильно влияют друг на друга, но не так сильно как Jackie Wilson. Связано это с тем, что он работал на стыке жанров ритм-энд-блюза и рок-н-ролла, которые повлияли на развитие музыки в будущем.

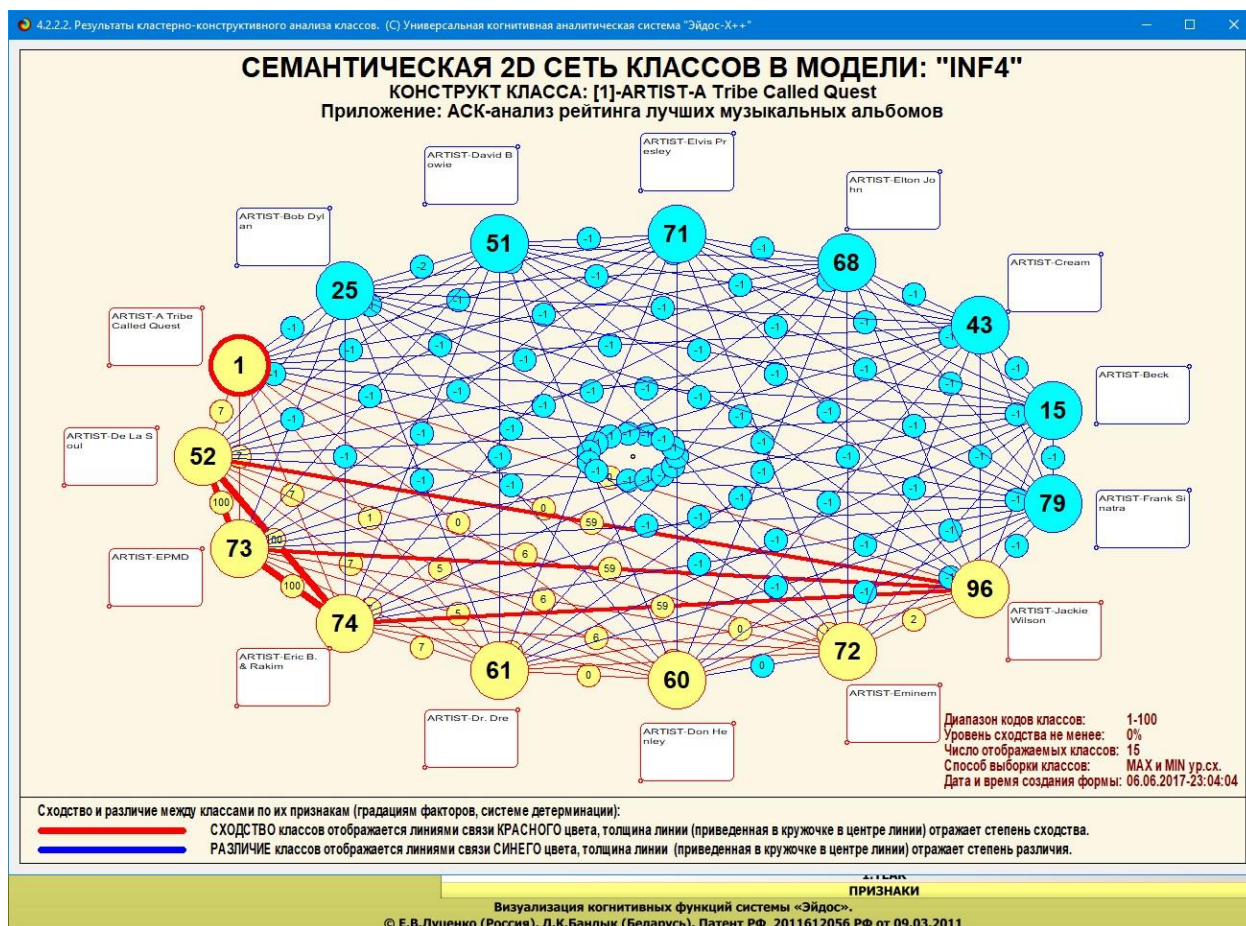


Рисунок 21 – Семантическая 2D сеть классов

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 музыкантов и их альбомов для построения модели поведения и предпочтения игроков. При этом наиболее достоверной в данном приложении оказалась модель INF4, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Семантический резонанс». Точность модели составляет 0,402, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 40,2%.

При рассмотрении модели в перспективе ретроанализа можно убедиться в достоверности моделей: история изменений контента журнала 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://ic.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 Rolling Stone [Электронный ресурс]. Режим доступа: <http://rollingstone.com>, 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 UCI 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>