

Федеральное государственное бюджетное образовательное
учреждение высшего образования

**КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ ИМ. И.Т.
ТРУБИЛИНА**

Факультет прикладной информатики

КУРСОВАЯ РАБОТА

по дисциплине: Системы представления и приобретения знаний.

на тему: «Исследование количества преступников по штату Техас с использованием АСК-анализа и системы "Эйдос"»

выполнил студент группы ИТз 1841 Кравченко Александр Сергеевич

Допущен к защите

Руководитель проекта: Луценко Евгений Вениаминович

Защищен _____

(дата)

Оценка _____

Краснодар

2019 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	4
1.1. Описание решения	4
1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel	4
1.3. Ввод выборки в систему Aidos-X.....	6
1.4. Синтез и верификация статистических и интеллектуальных моделей.....	10
1.5. Виды моделей системы «Aidos».....	11
1.5. Результаты верификации моделей	13
2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	15
2.1. Выделение основных задач.....	15
2.2. Выбор модели и пакетное распознавание	15
2.3 Визуализация когнитивных функций	19
2.4 SWOT и PEST матрицы и диаграммы	20
2.5 Нелокальные нейронные сети и нейтроны.....	23
2.6 Кластерный и конструктивный анализ	25
ЗАКЛЮЧЕНИЕ	26
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	27

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.

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 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](https://www.kaggle.com) возьмем базу данных: `greatest_albums.csv`.

Это база данных содержит данные преступников которые совершили преступления. Она содержит:

- Возраст.
- Год.
- Расу.
- Штат.
- Уровень образования
- Число преступлений.

Это нам позволит понять, в какое время было совершено больше преступлений, и проанализировать данные преступников такие как возраст, уровень образования и др.

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

Таблица 1 – Фрагмент обучающей выборки

Age	Year	Race	County	Age (when received)	Education level (Highest Grade Completed)	Native County (0=In state, 1 = From out of state)	Previous Crime (0=No, 1=Yes)
47	2017	Hispanic	Hidalgo	28	11	1	0
38	2017	White	Bee	22	8	0	1
46	2017	Black	Bexar	34	10	0	0
61	2017	White	Tarrant	36	9	0	1
44	2017	Hispanic	Bexar	22	10	0	1
43	2017	Black	Dallas	30	12	0	1
48	2017	White	Tarrant	37	12	0	0
58	2016	White	Houston	45	9	1	0
38	2016	Hispanic	Hidalgo	21	8	0	0
33	2016	White	Hunt	24	NA	0	0
58	2016	White	Harris	40	8	0	0
43	2016	Hispanic	Collin	19	NA	0	NA
35	2016	White	Wharton	27	12	0	0
43	2016	White	Harris	30	6	0	1
36	2015	Black	Madison	22	9	0	NA
33	2015	Hispanic	Dallas	20	8	0	NA

35	2015	Hispanic	Harris	20	8	0	0
27	2015	Hispanic	Nueces	22	10	0	NA
46	2015	Black	Smith	33	11	0	1
67	2015	White	Grayson	36	16	1	0

Классификационная шкала следующая: возраст, год. 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 – 4-5.
- Описательные шкалы – 6-19.

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}")

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).

2.1. Классификационные шкалы и градации. Текущая модель: "INF1"

Код шкалы	Наименование классификационной шкалы	Код градации	Наименование градации классификационной шкалы	DEL
1	AGE	1	1/3-{24.0000000, 38.3333333}	
2	YEAR	2	2/3-{38.3333333, 52.6666667}	
		3	3/3-{52.6666667, 67.0000000}	

Помощь Доб. шкалу Доб. град. шкалы Копир. шкалу Копир. град. шкалы Копир. шкалу с град. Удал. шкалу с град. Удал. град. шкалы Удаление и перекодирование

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.

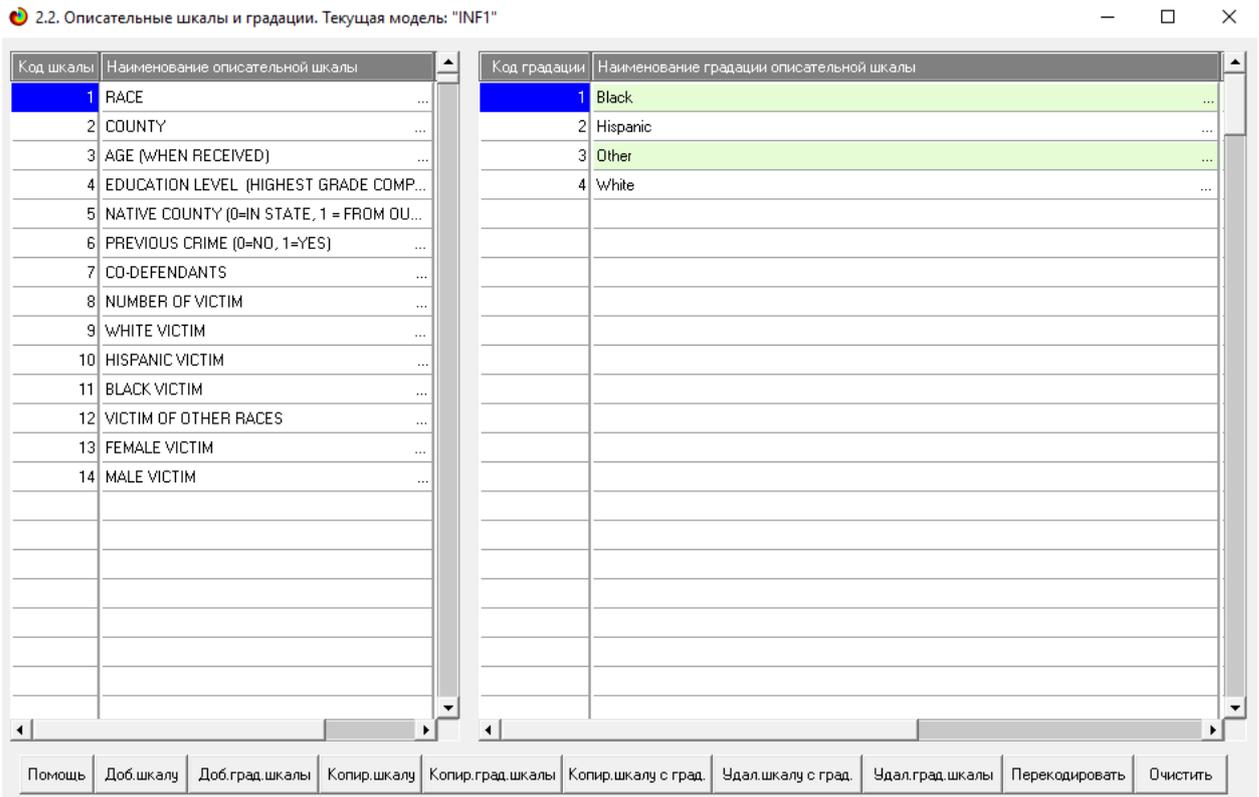


Рисунок 3 – Описательные шкалы

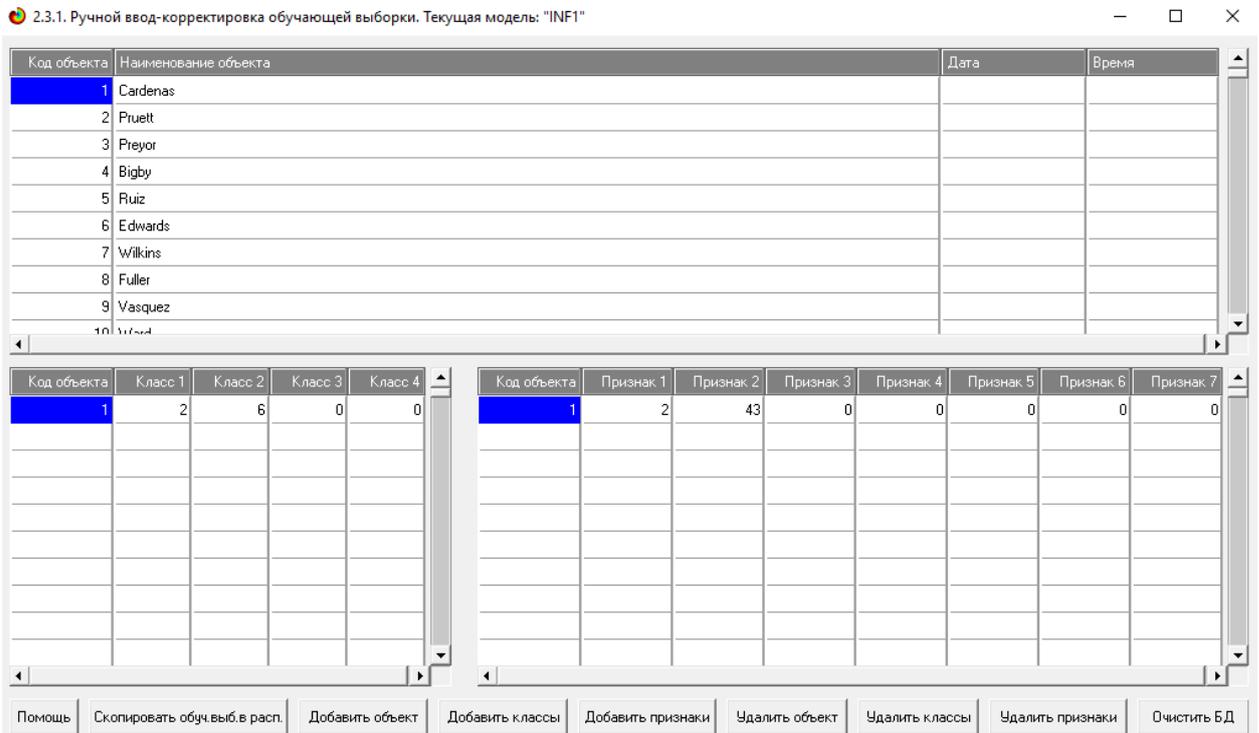


Рисунок 4 – Ручной ввод-корректировка выборки

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.

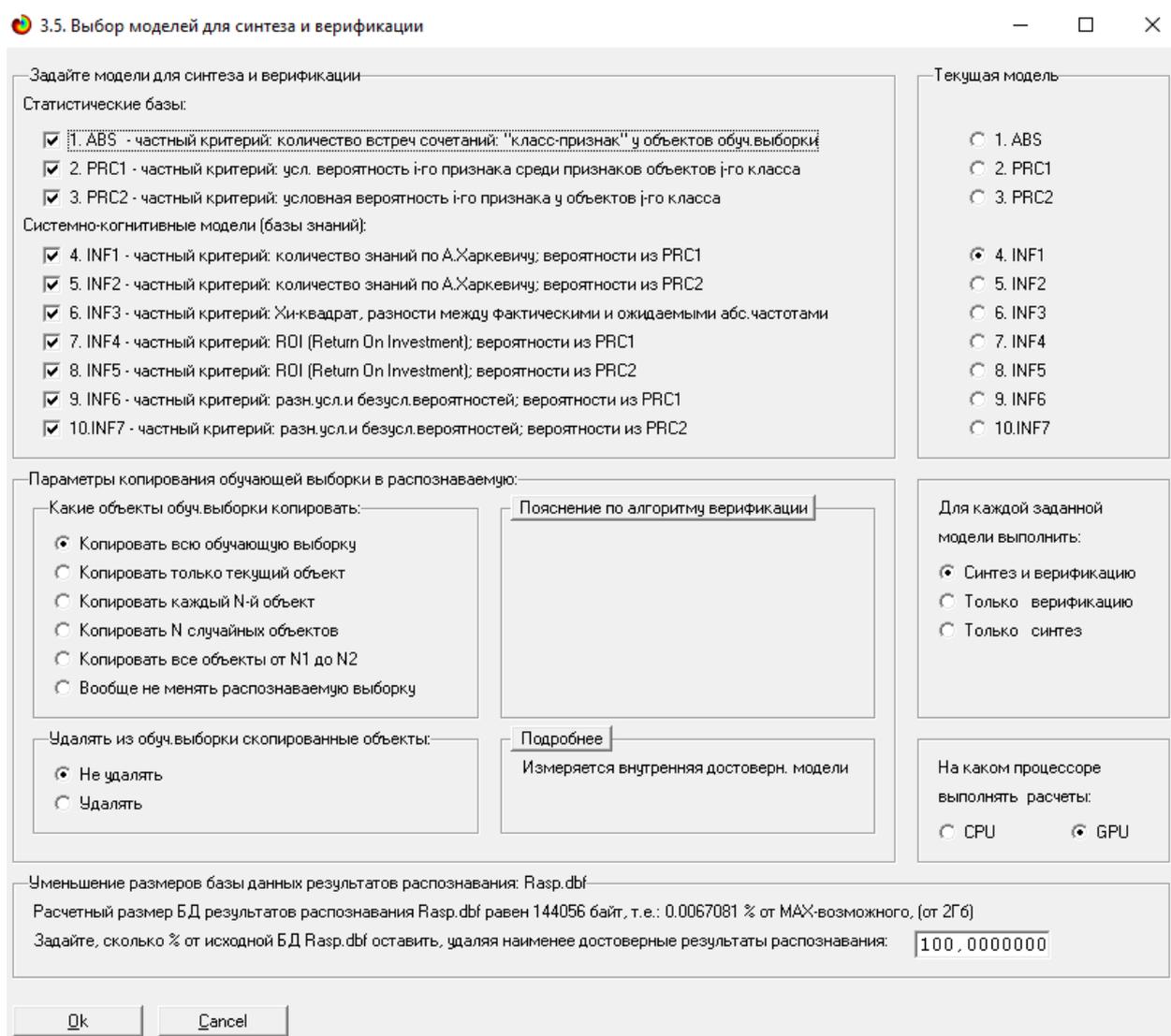


Рисунок 5 – Настройка синтеза и верификации моделей

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

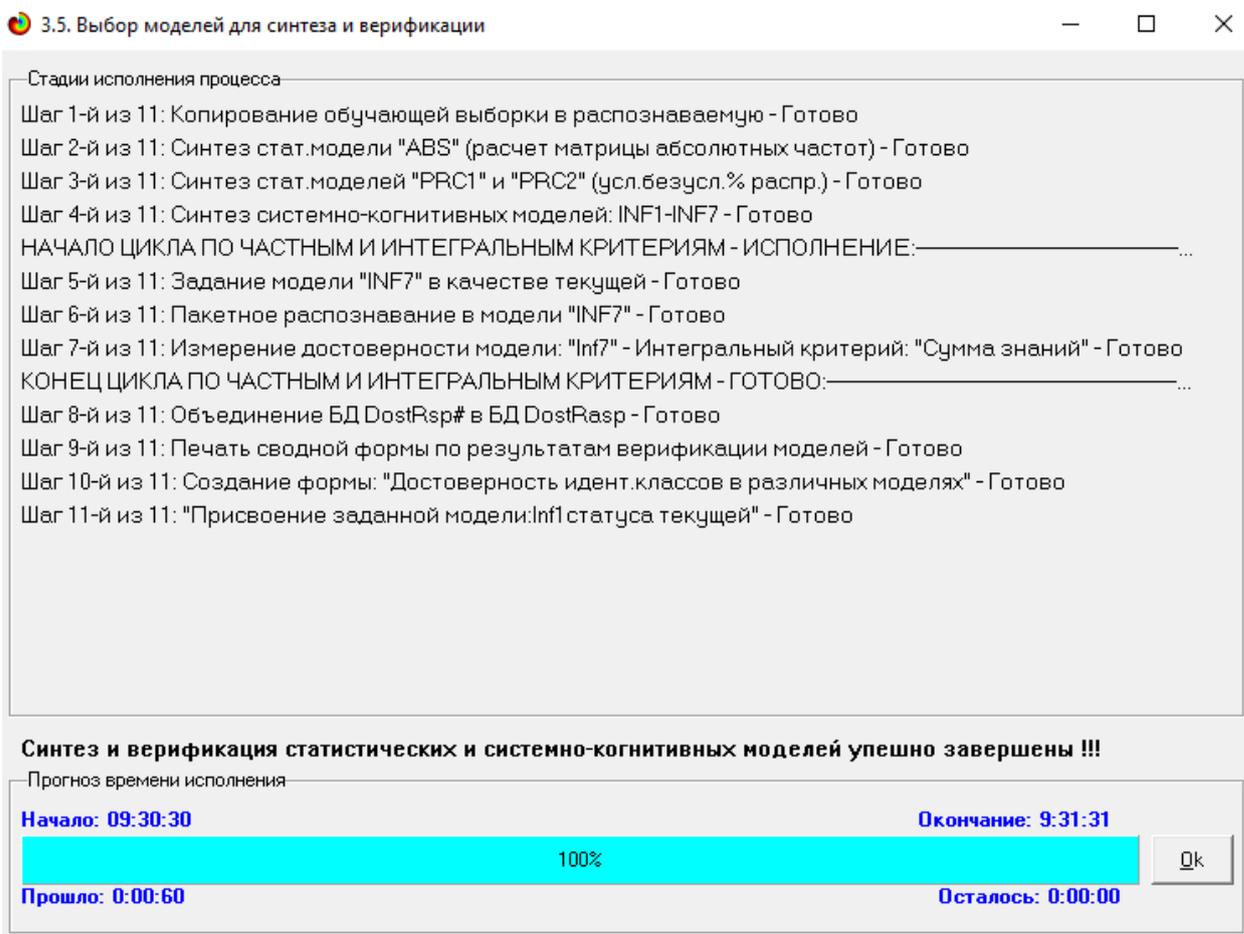


Рисунок 6 – Завершение процесса синтеза и верификации моделей

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).

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 обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. AGE 1/3 (24.0, 38.3)	2. AGE 2/3 (38.3, 52.7)	3. AGE 3/3 (52.7, 67.0)	4. YEAR 1/3 (1982.0, 1993.7)	5. YEAR 2/3 (1993.7, 2005.3)	6. YEAR 3/3 (2005.3, 2017.0)	Сумма	Среднее	Средн. квадрат. откл.
1	RACE-Black	120	70	8	25	96	77	396	66.00	42.42
2	RACE-Hispanic	60	40	3	14	38	51	206	34.33	21.81
3	RACE-Other		2			2		4	0.67	1.03
4	RACE-White	99	108	35	32	148	62	484	80.67	45.65
5	COUNTY-Anderson	2	2			3	1	8	1.33	1.21
6	COUNTY-Aransas		1			1		2	0.33	0.52
7	COUNTY-Atascosa		1			1		2	0.33	0.52
8	COUNTY-Bailey	1					1	2	0.33	0.52
9	COUNTY-Bastrop	1				1		2	0.33	0.52
10	COUNTY-Bee	2					2	4	0.67	1.03
11	COUNTY-Bell		3		1	1	1	6	1.00	1.10
12	COUNTY-Bexar	28	15	1	3	17	24	88	14.67	10.89
13	COUNTY-Bowie	3	2			3	2	10	1.67	1.37
14	COUNTY-Brazoria	2	2		2	1	1	8	1.33	0.82
15	COUNTY-Brazos	7	5		4	6	2	24	4.00	2.61
16	COUNTY-Brown		1			1		2	0.33	0.52
17	COUNTY-Caldwell		1			1		2	0.33	0.52
18	COUNTY-Cameron	3	2	1	3	2	1	12	2.00	0.89
19	COUNTY-Chambers		1		1			2	0.33	0.52
20	COUNTY-Cherokee	3				1	2	6	1.00	1.26
21	COUNTY-Clay	1				1		2	0.33	0.52
22	COUNTY-Collin	4	1	2		4	3	14	2.33	1.63
23	COUNTY-Comal	1		1		2		4	0.67	0.82

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. AGE 1/3 (24.0, 38.3)	2. AGE 2/3 (38.3, 52.7)	3. AGE 3/3 (52.7, 67.0)	4. YEAR 1/3 (1982.0, 1993.7)	5. YEAR 2/3 (1993.7, 2005.3)	6. YEAR 3/3 (2005.3, 2017.0)	Сумма	Среднее	Средн. квадрат. откл.
1	RACE-Black	0.044	-0.034	-0.211	-0.212	0.011	0.099	-0.304	-0.051	0.132
2	RACE-Hispanic	0.031	-0.004	-0.315	-0.108	-0.076	0.175	-0.377	-0.063	0.172
3	RACE-Other		0.294			0.239		0.534	0.089	0.139
4	RACE-White	-0.080	0.040	0.192	-0.197	0.084	-0.033	0.005	0.001	0.135
5	COUNTY-Anderson	-0.017	0.075			0.149	-0.041	0.166	0.028	0.071
6	COUNTY-Aransas		0.294			0.239		0.534	0.089	0.139
7	COUNTY-Atascosa		0.294			0.239		0.534	0.089	0.139
8	COUNTY-Bailey	0.202					0.397	0.599	0.100	0.166
9	COUNTY-Bastrop	0.202				0.239		0.442	0.074	0.115
10	COUNTY-Bee	0.202					0.397	0.599	0.100	0.166
11	COUNTY-Bell		0.294		0.095	-0.108	0.050	0.332	0.055	0.135
12	COUNTY-Bexar	0.059	-0.046	-0.393	-0.406	-0.061	0.205	-0.641	-0.107	0.246
13	COUNTY-Bowie	0.041	0.005			0.078	0.107	0.231	0.039	0.046
14	COUNTY-Brazoria	-0.017	0.075		0.223	-0.199	-0.041	0.042	0.007	0.139
15	COUNTY-Brazos	0.032	0.018		0.095	0.020	-0.169	-0.004	-0.001	0.089
16	COUNTY-Brown		0.294			0.239		0.534	0.089	0.139
17	COUNTY-Caldwell		0.294			0.239		0.534	0.089	0.139
18	COUNTY-Cameron	-0.017	-0.053	0.236	0.223	-0.108	-0.169	0.113	0.019	0.171
19	COUNTY-Chambers		0.294		0.442			0.737	0.123	0.196
20	COUNTY-Cherokee	0.202				-0.108	0.269	0.363	0.061	0.143
21	COUNTY-Clay	0.202				0.239		0.442	0.074	0.115
22	COUNTY-Collin	0.025	-0.320	0.407		0.063	0.129	0.304	0.051	0.234
23	COUNTY-Comal	-0.017		0.584		0.239		0.806	0.134	0.241

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. AGE 1/3 (24.0, 38.3)	2. AGE 2/3 (38.3, 52.7)	3. AGE 3/3 (52.7, 67.0)	4. YEAR 1/3 (1982.0, 1993.7)	5. YEAR 2/3 (1993.7, 2005.3)	6. YEAR 3/3 (2005.3, 2017.0)	Сумма	Среднее	Средн. квадрат. откл.
1	RACE-Black	15.574	-7.960	-7.614	-23.836	3.201	20.634	0.000	16.541	
2	RACE-Hispanic	5.677	-0.555	-5.122	-11.404	-10.274	21.679	0.000	12.364	
3	RACE-Other	-1.055	1.213	-0.158	-0.493	1.063	-0.569	0.000	0.928	
4	RACE-White	-28.632	12.716	15.916	-27.688	34.579	-6.891	0.000	25.474	
5	COUNTY-Anderson	-0.110	0.425	-0.315	-0.987	1.125	-0.139		0.714	
6	COUNTY-Aransas	-0.527	0.606	-0.079	-0.247	0.531	-0.285		0.464	
7	COUNTY-Atascosa	-0.527	0.606	-0.079	-0.247	0.531	-0.285		0.464	
8	COUNTY-Bailey	0.473	-0.394	-0.079	-0.247	-0.469	0.715		0.485	
9	COUNTY-Bastrop	0.473	-0.394	-0.079	-0.247	0.531	-0.285		0.402	
10	COUNTY-Bee	0.945	-0.787	-0.158	-0.493	-0.937	1.431	0.000	0.970	
11	COUNTY-Bell	-1.582	1.819	-0.237	0.260	-0.406	0.146	0.000	1.106	
12	COUNTY-Bexar	4.794	-2.324	-2.470	-7.852	-3.622	11.474		6.942	
13	COUNTY-Bowie	0.363	0.031	-0.394	-1.233	0.657	0.577		0.717	
14	COUNTY-Brazoria	-0.110	0.425	-0.315	1.013	-0.875	-0.139		0.649	
15	COUNTY-Brazos	0.671	0.275	-0.946	1.040	0.376	-1.416		0.964	
16	COUNTY-Brown	-0.527	0.606	-0.079	-0.247	0.531	-0.285		0.464	
17	COUNTY-Caldwell	-0.527	0.606	-0.079	-0.247	0.531	-0.285		0.464	
18	COUNTY-Cameron	-0.164	-0.362	0.527	1.520	-0.812	-0.708		0.884	
19	COUNTY-Chambers	-0.527	0.606	-0.079	0.753	-0.469	-0.285		0.551	
20	COUNTY-Cherokee	1.418	-1.181	-0.237	-0.740	-0.406	1.146	0.000	1.048	
21	COUNTY-Clay	0.473	-0.394	-0.079	-0.247	0.531	-0.285		0.402	
22	COUNTY-Collin	0.308	-1.756	1.448	-1.727	0.719	1.007		1.399	
23	COUNTY-Comal	-0.055	-0.787	0.842	-0.493	1.063	-0.569		0.778	

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. Обобщ.форма по достов.моделей 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 ложноположительных решений (FP)	Число ложноотрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризенбергена	Сумма между. уровней сход. истинно-полож. решений (STR)	Сумма между. уровней сход. истинно-отриц. решений (STN)	Сумма уровней ложно-полож. решений
1. ABS - частный критерий: количество встреч сочетаний "клас..."	Корреляция абс частот с обр...	1090	1090	2	2178		0.334	1.000	0.500	543.066	0.034	920
1. ABS - частный критерий: количество встреч сочетаний "клас..."	Сумма абс частот по признак...	1090	1090		2180		0.333	1.000	0.500	298.523		397
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн частот с о...	1090	1090	2	2178		0.334	1.000	0.500	543.066	0.034	920
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн частот по при...	1090	1090		2180		0.333	1.000	0.500	295.788		520
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн частот с о...	1090	1090	2	2178		0.334	1.000	0.500	543.066	0.034	920
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн частот по при...	1090	1090		2180		0.333	1.000	0.500	163.511		297
4. INF1 - частный критерий: количество знаний по АХаркевичу в...	Семантический резонанс зна...	1090	546	1697	483	544	0.531	0.501	0.515	82.452	360.276	41
4. INF1 - частный критерий: количество знаний по АХаркевичу в...	Сумма знаний	1090	794	1385	795	296	0.500	0.728	0.593	61.663	142.258	37
5. INF2 - частный критерий: количество знаний по АХаркевичу в...	Семантический резонанс зна...	1090	530	1605	575	560	0.480	0.486	0.483	64.993	223.496	54
5. INF2 - частный критерий: количество знаний по АХаркевичу в...	Сумма знаний	1090	705	1291	889	385	0.442	0.647	0.525	39.946	72.165	37
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Семантический резонанс зна...	1090	749	1310	870	341	0.463	0.687	0.553	371.323	535.642	411
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Сумма знаний	1090	749	1310	870	341	0.463	0.687	0.553	135.819	205.798	115
7. INF4 - частный критерий: ROI (Return On Investment), вероятн...	Семантический резонанс зна...	1090	468	1874	306	622	0.605	0.429	0.502	68.190	213.091	19
7. INF4 - частный критерий: ROI (Return On Investment), вероятн...	Сумма знаний	1090	831	1263	917	259	0.475	0.762	0.586	57.468	51.313	34
8. INF5 - частный критерий: ROI (Return On Investment), вероятн...	Семантический резонанс зна...	1090	423	1827	353	667	0.545	0.388	0.453	60.252	172.621	31
8. INF5 - частный критерий: ROI (Return On Investment), вероятн...	Сумма знаний	1090	731	1250	930	359	0.440	0.671	0.531	32.991	20.473	27
9. INF6 - частный критерий: разн.усл.и безуслов.вероятностей; вер...	Семантический резонанс зна...	1090	731	1332	848	359	0.463	0.671	0.548	371.727	540.550	413
9. INF6 - частный критерий: разн.усл.и безуслов.вероятностей; вер...	Сумма знаний	1090	749	1309	871	341	0.462	0.687	0.553	102.714	212.738	143
10. INF7 - частный критерий: разн.усл.и безуслов.вероятностей; ве...	Семантический резонанс зна...	1090	645	1323	857	445	0.429	0.592	0.498	238.908	419.809	306
10. INF7 - частный критерий: разн.усл.и безуслов.вероятностей; ве...	Сумма знаний	1090	653	1199	981	437	0.400	0.599	0.479	44.174	54.338	71

Рисунок 7 – Результаты верификации моделей

Наиболее достоверной в данном приложении оказалась модель INF1 при интегральном критерии «Семантический резонанс знаний», если учитывать F-меру Ван Ризенбергена. Точность модели составляет 0,593, что есть уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 59,3%.

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 INF1 (рисунок 8). Для этого зайдём в режим 4.1.3.11, выберем модель INF1 и запросим график по сумме знаний.

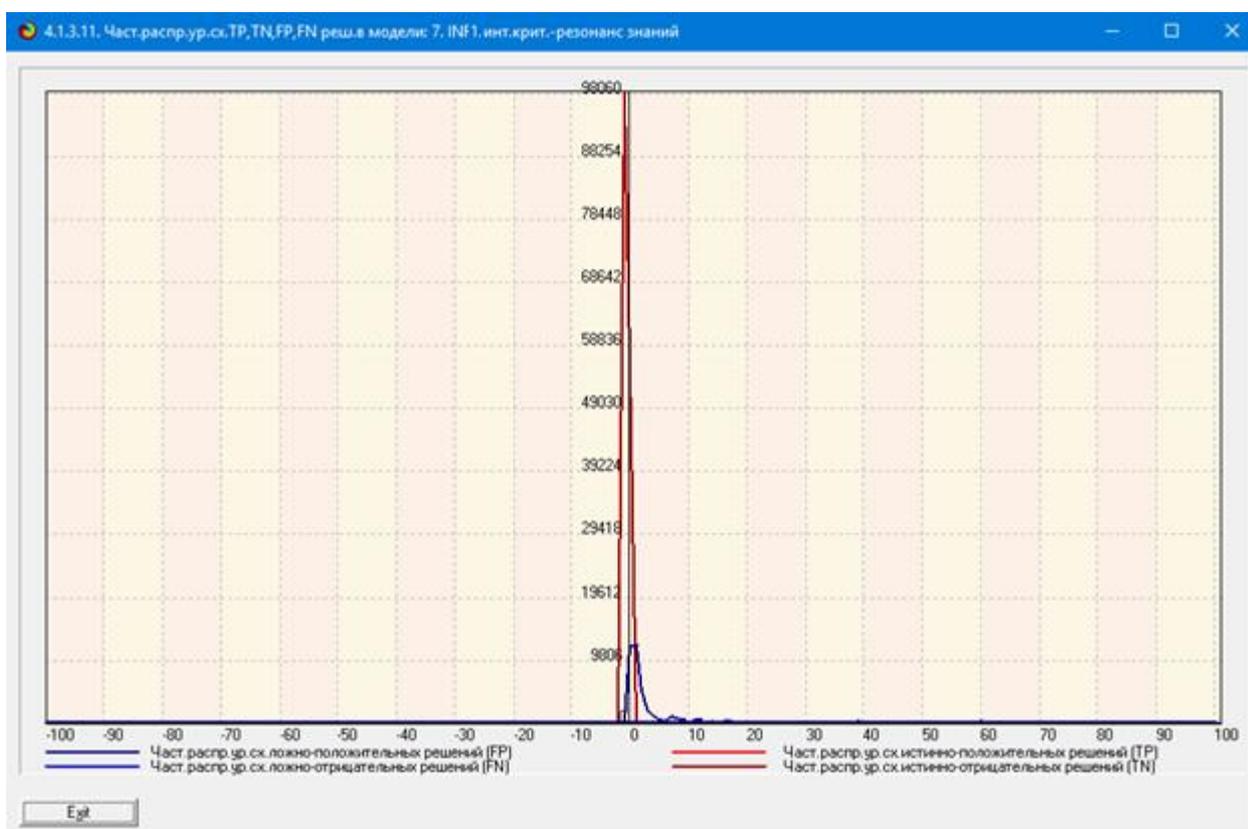


Рисунок 8 – Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF1

Как можно видеть, при модели INF1 непринадлежность объекта к классу лучше определяется, чем принадлежность. При этом существует вероятность ложного определения непринадлежности или принадлежности.

Выберем модель INF1 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 представителя закона, то какие задачи при анализе баз данных можно выделить?

Во-первых, зависимость преступлений от штата. Зачастую количество полицейских зависит от количества преступлений в области, если выбрать не напряженный штат или область, то можно остаться без поддержки в момент беспорядков.

Во-вторых, можно вычислить какие люди чаще совершают преступления и в каком возрасте.

В-третьих, статистика по годам может показать как работают правоохранительные органы в том или ином штате. Таким образом, в первую очередь, нас интересуют следующие связи:

- Год – штат
- Возраст – раса
- Раса – уровень образования

2.2. Выбор модели и пакетное распознавание

Зададим модель INF1 в качестве текущей в режиме 5.6 (рисунок 9), после чего проведем пакетное распознавание в режиме 4.1.2 (рисунок 10).

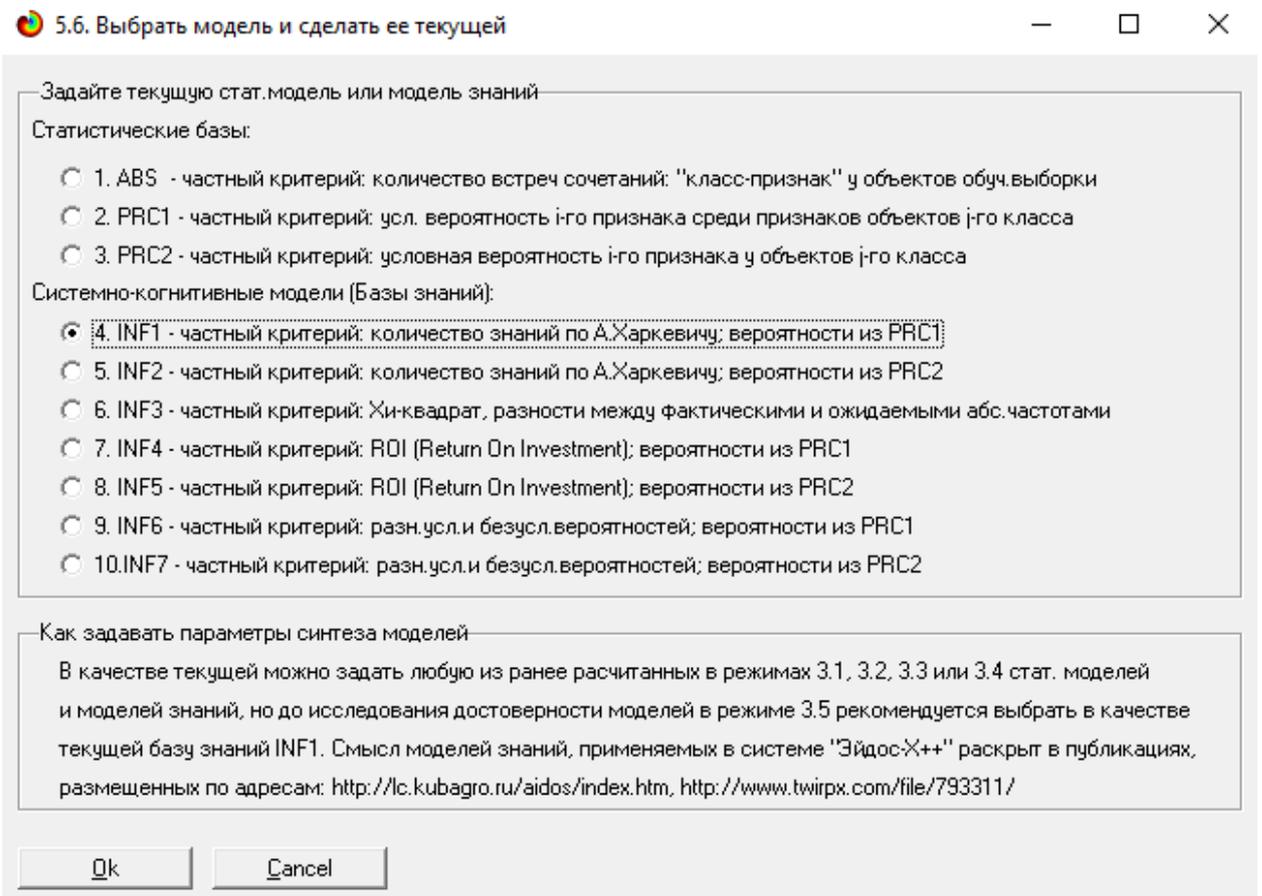


Рисунок 9 – Выбор текущей модели

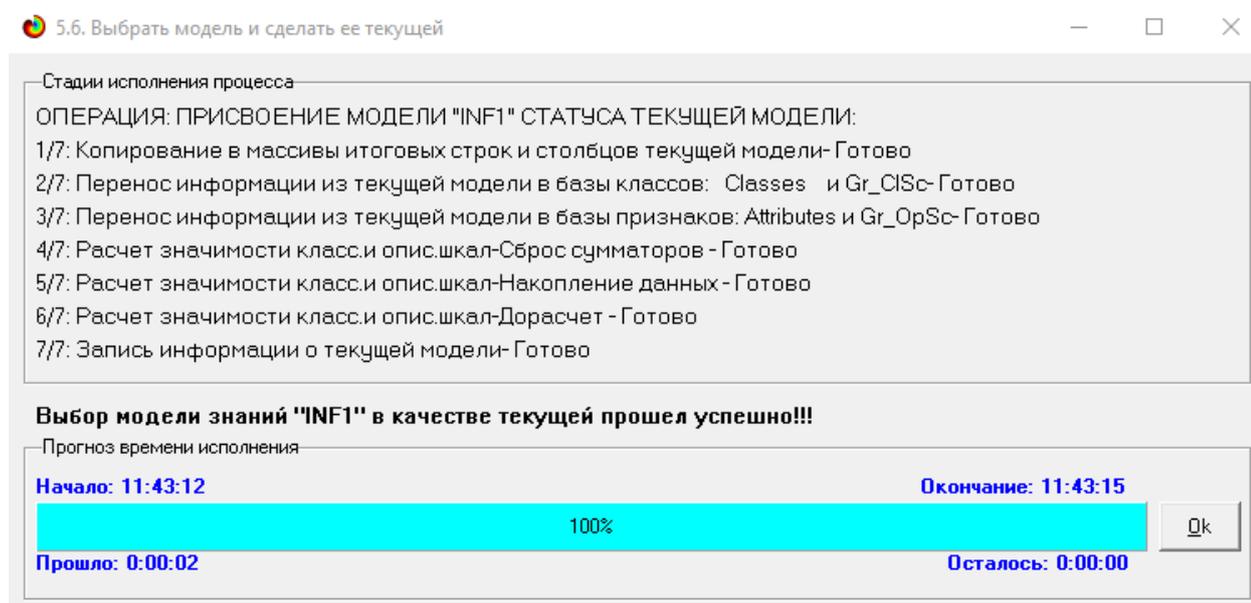


Рисунок 10 – Пакетное распознавание в новой модели

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 INF1:

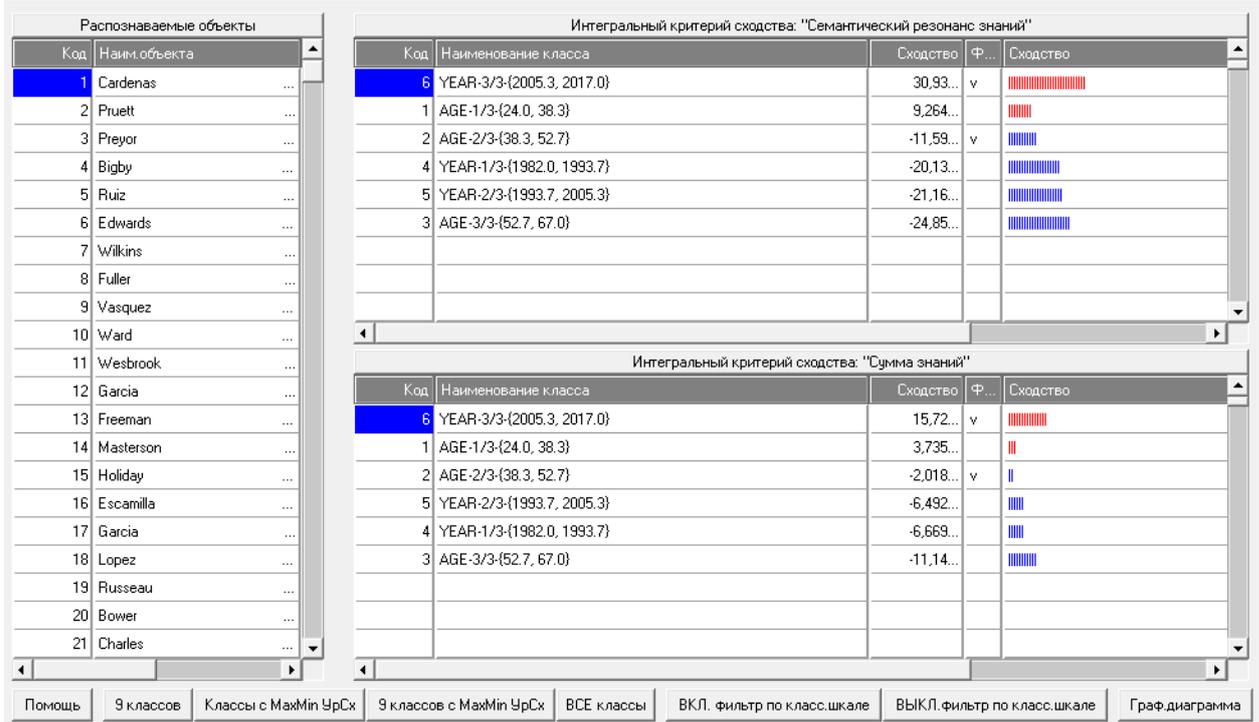


Рисунок 11 – Прогнозы достоверности классов объектов

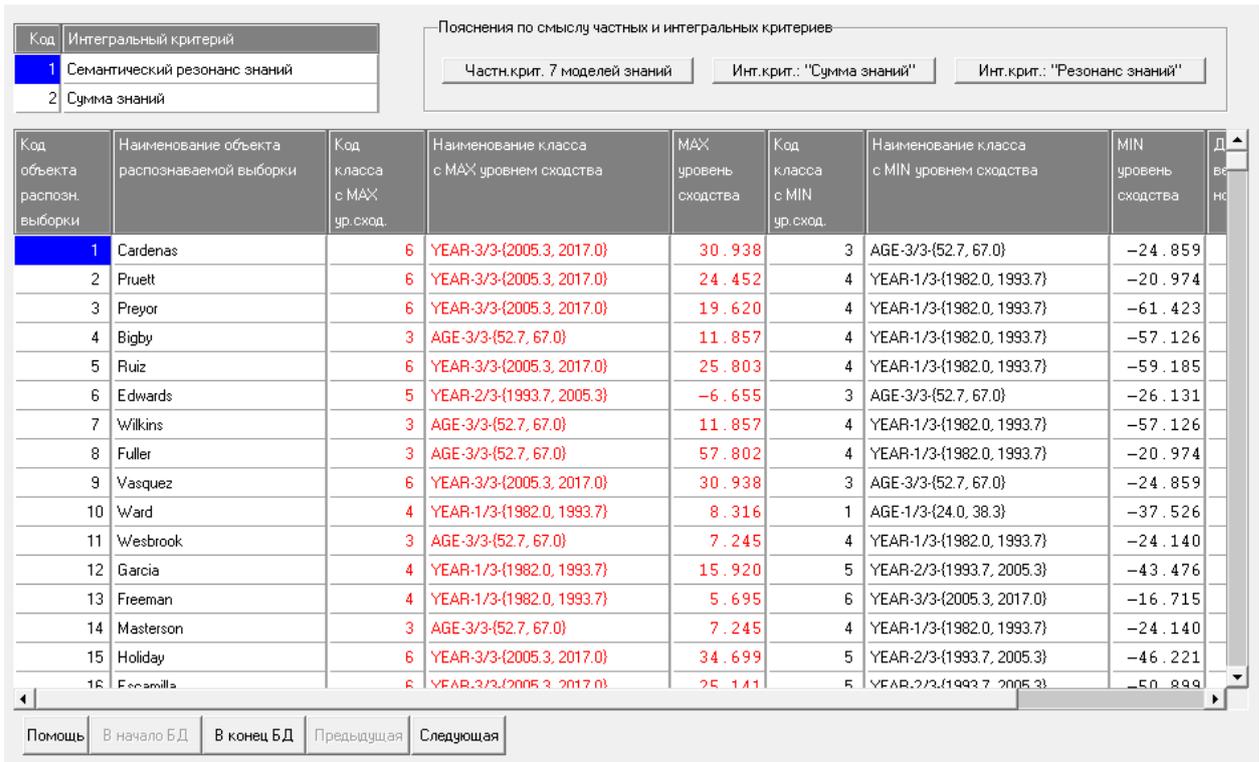


Рисунок 12 – Итоги распознавания классов объектов

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.

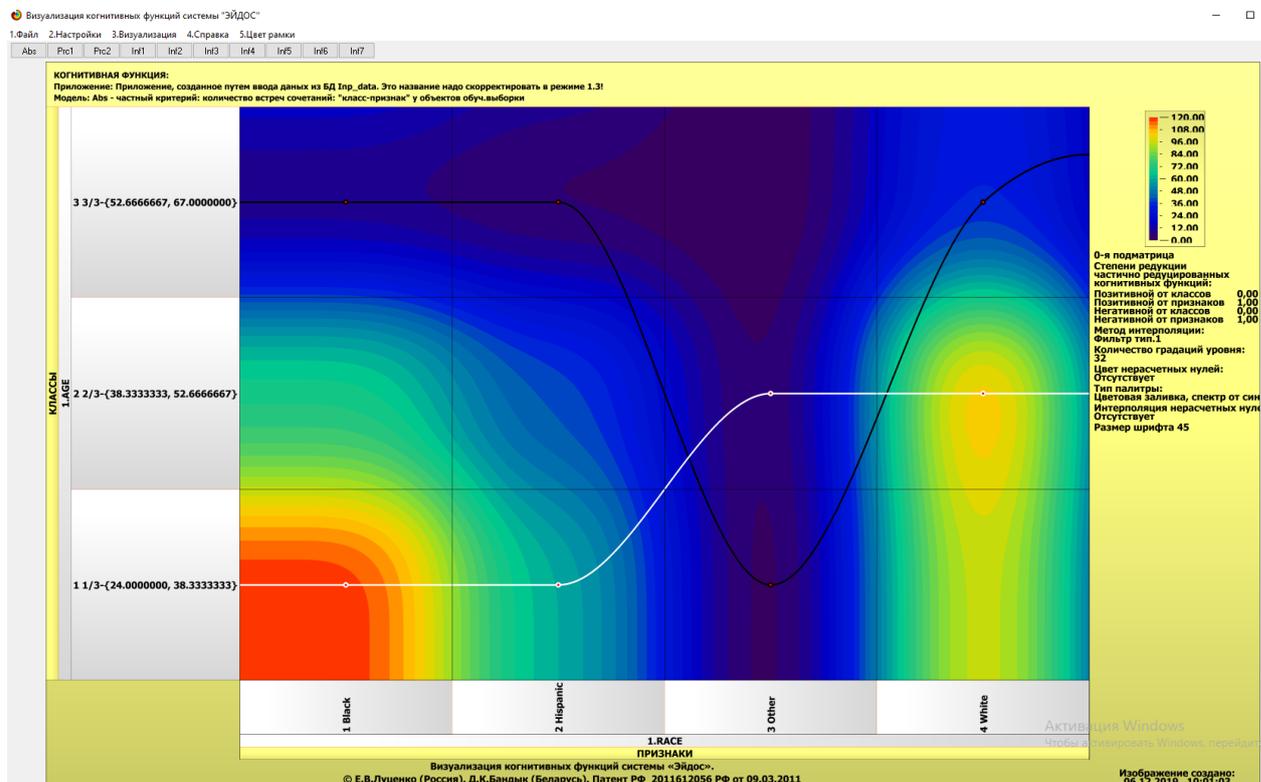


Рисунок 13 – Визуализация зависимости расы от возраста

Рассмотрим визуализацию когнитивных функций в рамках определенных задач.

Зависимость расы артистов от возраста представлена на рисунке 13. Из него видно, что люди с черным цветом кожи больше совершают преступления с 20 лет до 36, а с белым цветом с 36 до 52. Теперь перейдем к зависимости «возраст-штат». Она отражена на рисунке 14.

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).

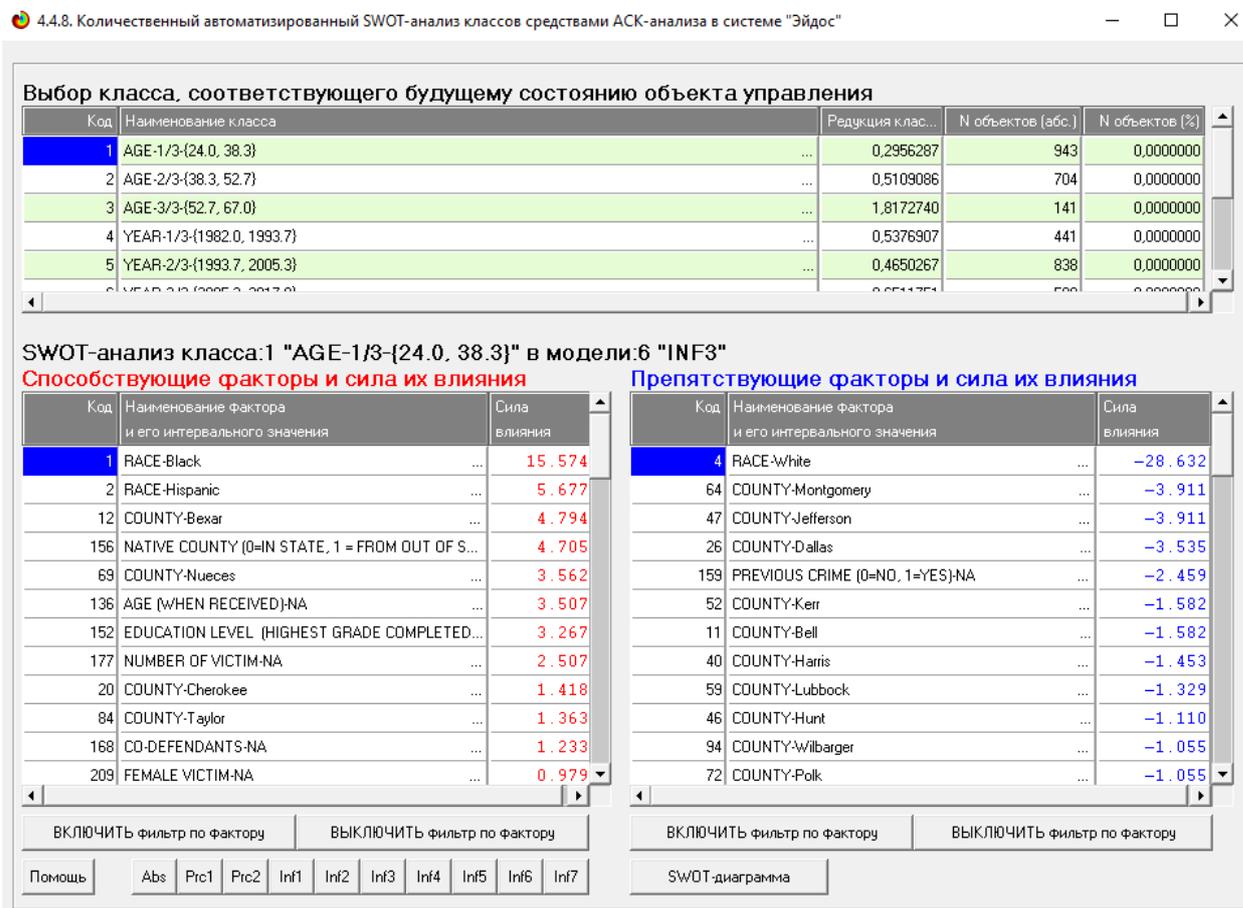


Рисунок 15 – Пример SWOT-Матрицы в модели INF1

На рисунке 16 приведен пример инвертированной SWOT-матрицы.

Выбор значения фактора, оказывающего влияние на переход объекта управления в будущие состояния

Код	Наименование значения фактора
1	RACE-Black
2	RACE-Hispanic
3	RACE-Other
4	RACE-White
5	COUNTY-Anderson
6	COUNTY...

SWOT-анализ значения фактора: 1 "RACE-Black" в модели: 6 "INF3"

СПОСОБСТВУЕТ:

Код	Состояния объекта управления, переходу в которые данное значение фактора СПОСОБСТВУЕТ	Сила влияния
6	YEAR-3/3-(2005.3, 2017.0)	20.634
1	AGE-1/3-(24.0, 38.3)	15.574
5	YEAR-2/3-(1993.7, 2005.3)	3.201

ПРЕПЯТСТВУЕТ:

Код	Состояния объекта управления, переходу в которые данное значение фактора ПРЕПЯТСТВУЕТ	Сила влияния
4	YEAR-1/3-(1982.0, 1993.7)	-23.836
2	AGE-2/3-(38.3, 52.7)	-7.960
3	AGE-3/3-(52.7, 67.0)	-7.614

ВКЛЮЧИТЬ фильтр по кл.шкале ВЫКЛЮЧИТЬ фильтр по кл.шкале

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 SWOT-диаграмма

Рисунок 16 – Пример инвертированной SWOT-Матрицы в модели INF1

На примере группы «Tribe Called Quest» посмотрим в каких годах группа была популярна, какой жанр и поджанр предпочитала.

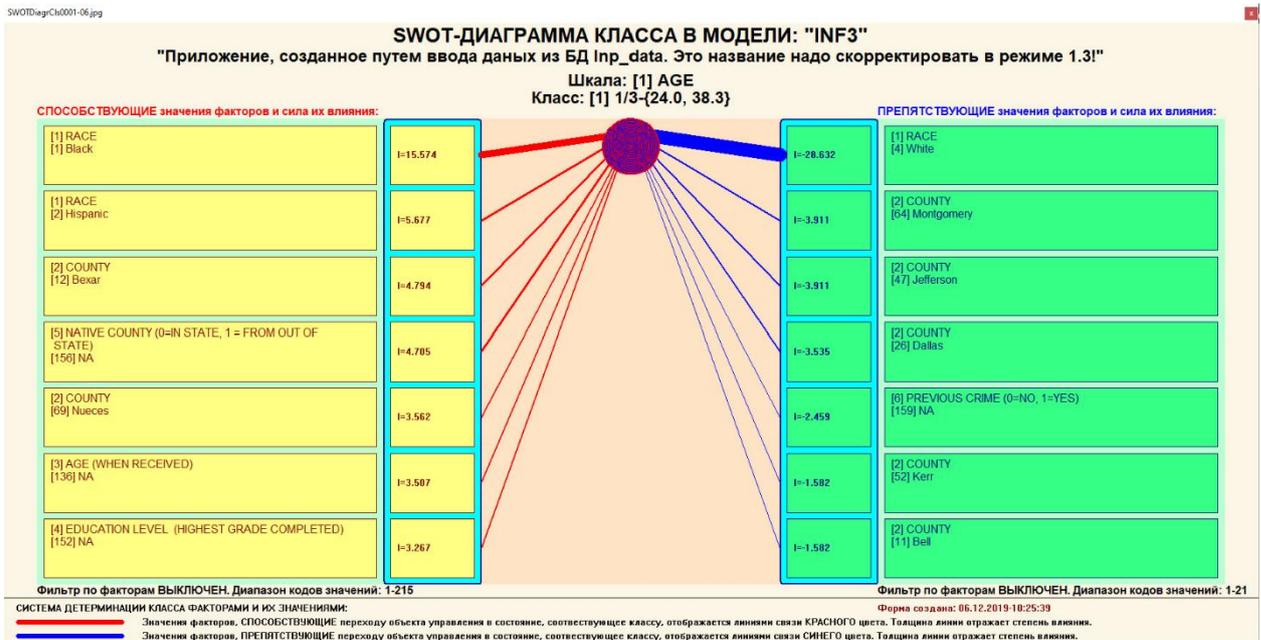


Рисунок 17 – SWOT-матрица, построенная для «Tribe Called Quest»

Как можно видеть, группа была популярна в 1983-1992 г., их любимым жанр hip-hop, а поджанр Conscious.

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

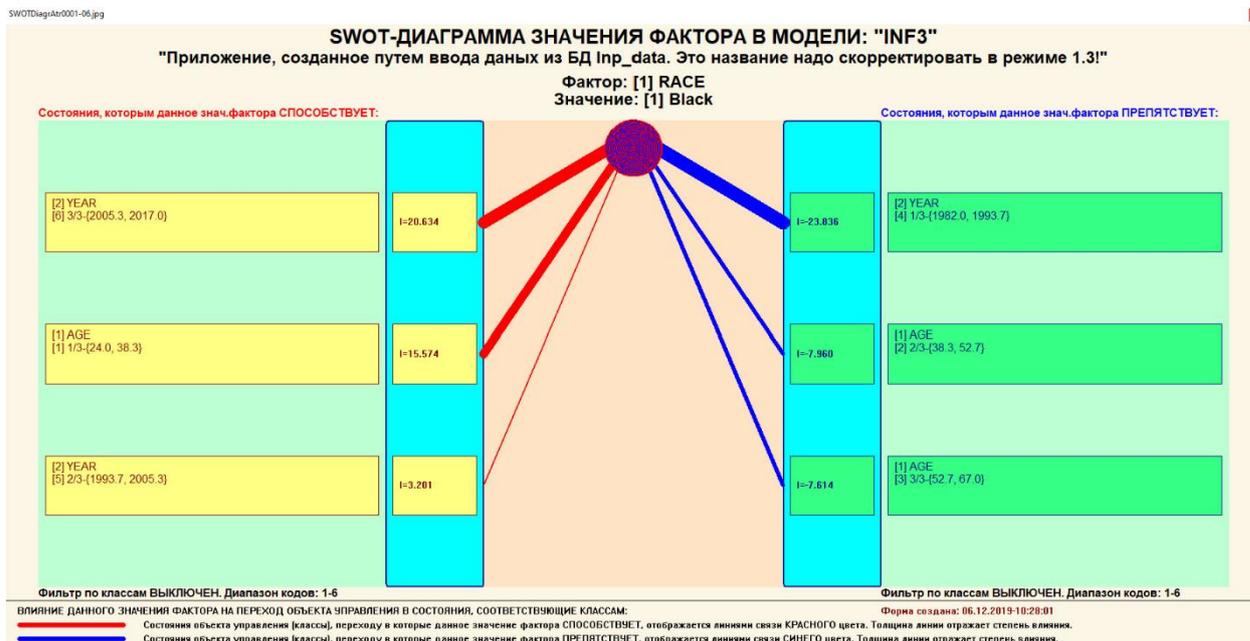


Рисунок 18 – SWOT-диаграмма признаков топ исполнителей

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 возраста.

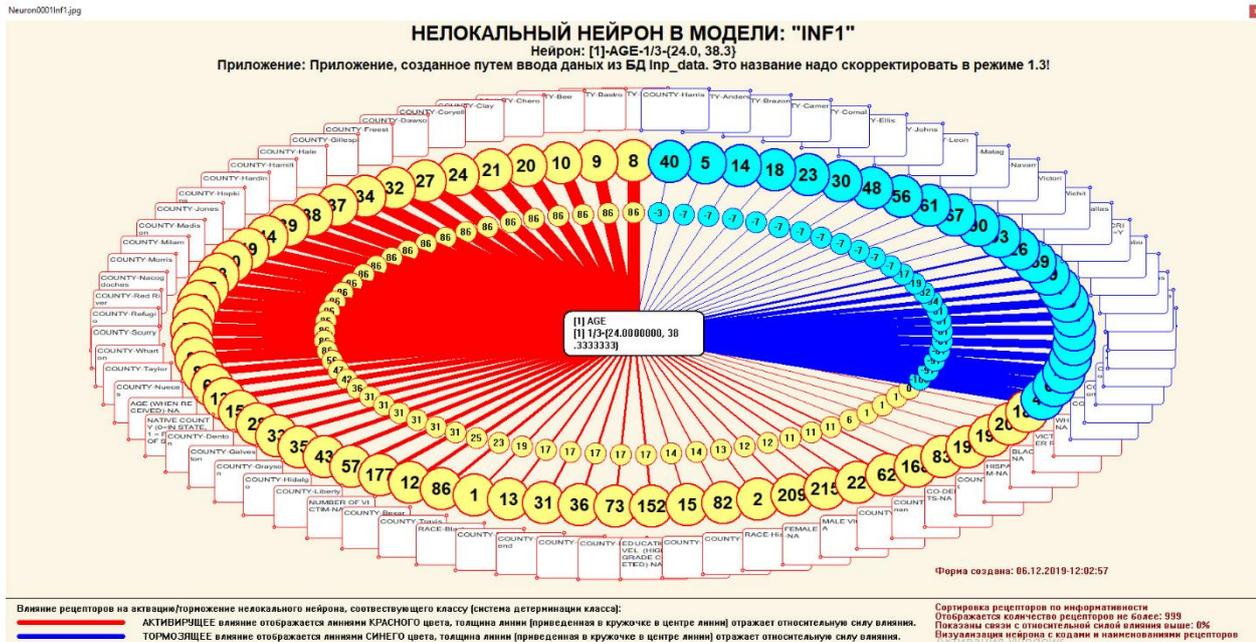


Рисунок 19 – Нейрон для возраста

На его основании можно определить в каком штате в каком возрасте происходят преступления.

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

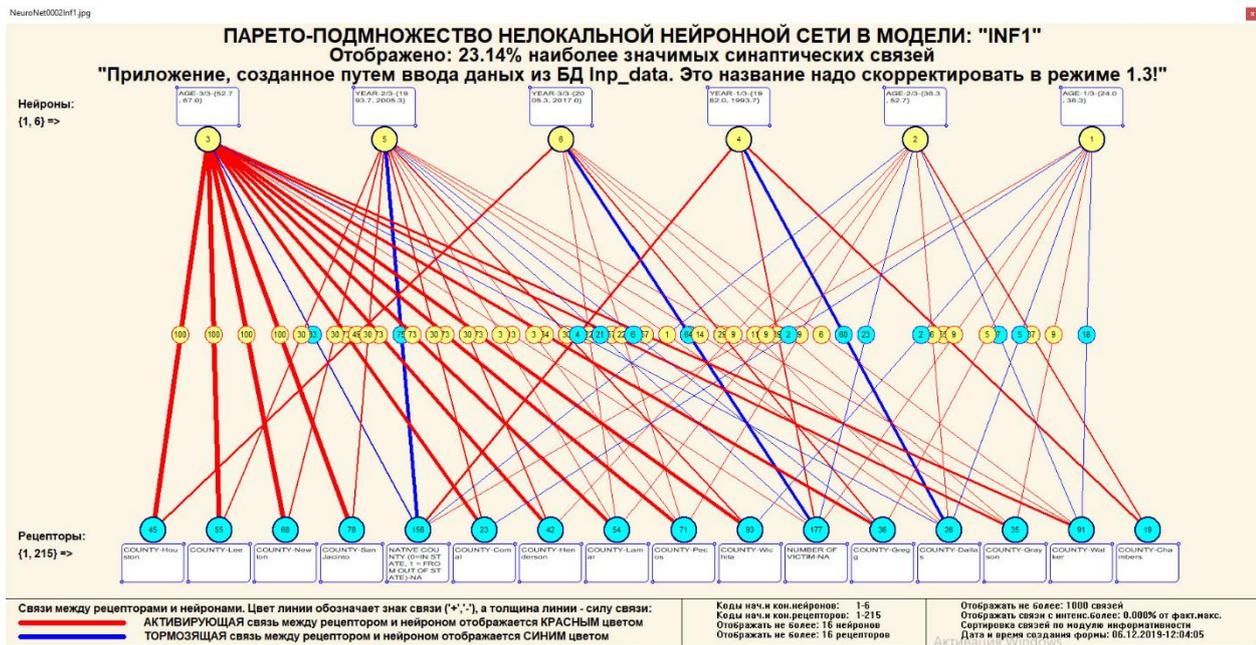


Рисунок 20 – Пример нелокальной нейронной сети в модели INF1

По ней можно, в основном, определить зависимость возраста и штата преступлений.

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 что самыми криминальными временами были 2005 по 2017 год, а возраст преступников варьируется от 52 до 67 лет.

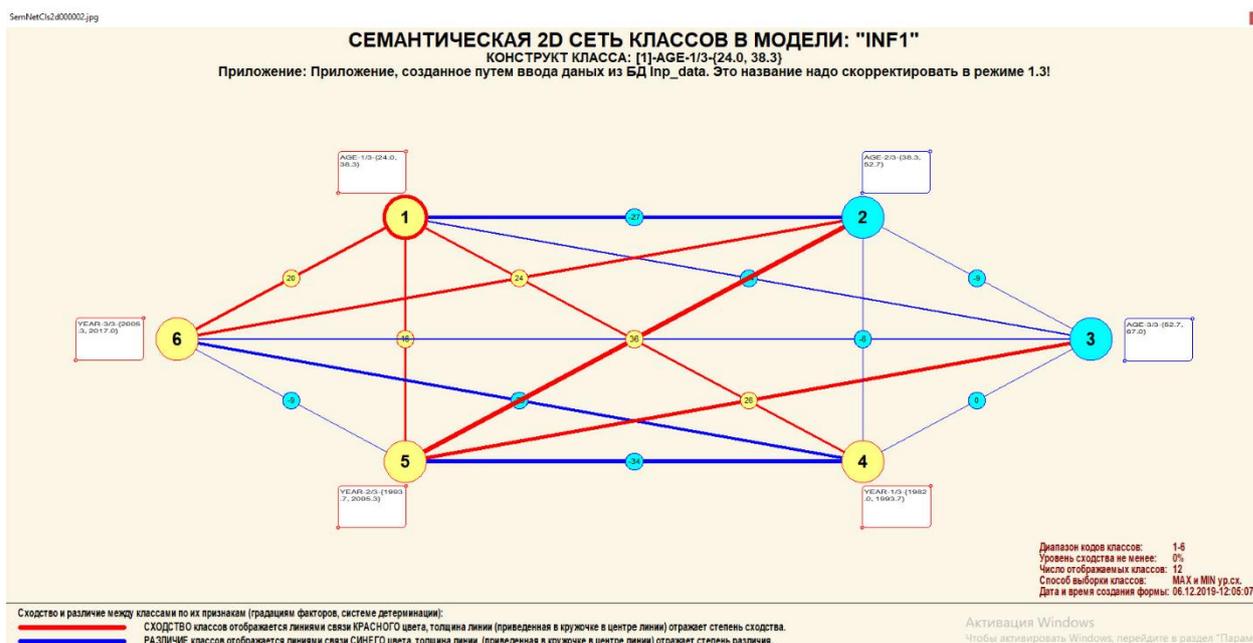


Рисунок 21 – Семантическая 2D сеть классов

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 преступников для построения модели поведения и предпочтения игроков. При этом наиболее достоверной в данном приложении оказалась модель INF1, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Семантический резонанс». Точность модели составляет 0,593, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 59,3%.

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

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>

