

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 Ворожейкина Дмитрия Сергеевича курса 3 очной формы обучения
направления подготовки Информационные системы и технологии.

направленность (профиль) академический бакалавриат.

Наименование темы «Разработка и исследование системно когнитивной модели
зависимости совершенных преступлений от различных факторов в городе Солт
Лейк Сити, США».

Рецензент:

(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	5
4.	Степень комплексности работы, применения в ней знаний общепрофессиональных и специальных дисциплин	4
5.	Ясность, четкость, последовательность и обоснованность изложения	5
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

Применение современных математических

Недостатки работы

Недостаточно качественное оформление

Итоговая оценка при защите _____

Рецензент _____

(Е.В. Луценко)

« ____ » _____ 201 ____ г.

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:

**Разработка и исследование системно когнитивной модели зависимости
совершенных преступлений от различных факторов в городе Солт Лейк
Сити, США**

выполнил студент группы ИТ1401

Ворожейкин Дмитрий Сергеевич

Допущена 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)

Ворожейкину Дмитрию Сергеевичу
(Ф.И.О.)

Тема проекта: **Разработка и исследование системно когнитивной модели зависимости совершенных преступлений от различных факторов в городе Солт Лейк Сити, США**

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.....	6
1.СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	7
1.1. Описание решения.....	7
1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel	7
1.3. Синтез и верификация статистических и интеллектуальных моделей	14
1.4. Виды моделей системы «Эйдос»	16
1.5. Результаты верификации моделей	17
2.РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	24
2.1. Решение задачи идентификации	24
2.2. Когнитивные функции	27
2.3. SWOT и PEST матрицы и диаграммы	29
2.4. Нелокальные Нейронные Сети И Нейроны.....	31
2.5 Кластерный и Конструктивный Анализ.....	33
ЗАКЛЮЧЕНИЕ	34
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	35

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 поиску зависимости состава продуктов быстрого питания от выбранной категории.

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 «menu» 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 "Эйдос-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 "Эйдос".
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

Из электронного ресурса баз данных Kaggle по ссылке <https://www.kaggle.com/foenix/slc-crime> возьмем базу данных преступлений – «database.csv», которую оставим без изменений.

Общее описание задачи:

1. Характер преступления
2. Дата и время инцидента
3. Место преступления
4. День недели

Столбцы 1-3 описательные шкалы.

Столбец 4 является классификационной шкалой. Этот столбец показывает день недели преступления.

Обучающая выборка:

Таблица 1 – Inp_data.xls

n	description	occurred	Место	День недели
1	STOLEN VEHICLE	01/19/2016 05:49:39 PM	700 N 1200 W	вторник
2	LARCENY	01/21/2016 06:41:12 PM	1900 W 1700 S	четверг
3	LARCENY	01/22/2016 06:40:18 PM	1100 E BRICKYARD RD	пятница
4	LARCENY	01/24/2016 12:38:55 PM	1800 S REDWOOD RD	суббота
5	LARCENY	01/24/2016 12:40:20 PM	1800 S REDWOOD RD	суббота
6	LARCENY	01/24/2016 12:40:41 PM	1800 S REDWOOD RD	суббота
7	LARCENY	01/24/2016 12:39:15 PM	1800 S REDWOOD RD	суббота
8	LARCENY	01/24/2016 12:41:06 PM	1800 S REDWOOD RD	суббота
9	LARCENY	01/24/2016 12:41:33 PM	1800 S REDWOOD RD	суббота
10	LARCENY	01/24/2016 12:41:59 PM	1800 S REDWOOD RD	суббота
11	ASSAULT	01/23/2016 01:31:37 PM	500 N SIR PHILIP DR	суббота
12	FRAUD	01/25/2016 11:06:21 AM	1100 E 500 S	суббота
13	PUBLIC PEACE	01/22/2016 02:11:21 PM	1400 S FOOTHILL DR	суббота
14	ESCAPE	01/23/2016 01:05:36 PM	200 S 500 W	суббота
15	LARCENY	01/23/2016 10:27:21 AM	700 S 600 W	суббота
16	ESCAPE	01/25/2016 01:46:39 PM	100 E 100 S	суббота
17	PUBLIC ORDER	01/25/2016 06:53:38 AM		суббота
18	STOLEN VEHICLE	01/25/2016 12:10:33 PM	1300 E 500 S	суббота
19	LARCENY	01/25/2016 03:18:20 PM	900 S 300 W	суббота
20	ASSAULT	01/23/2016 04:38:56 PM	1200 E TOMAHAWK DR	суббота
21	PUBLIC ORDER	01/23/2016 11:30:50 AM	300 N K ST	суббота
22	NONREPTABL TA	01/23/2016 04:00:18 PM		суббота
23	LARCENY	01/24/2016 06:04:21 PM	500 S 1000 W	суббота
24	LARCENY	01/25/2016 10:16:34 AM	1800 S REDWOOD RD	суббота
25	STOLEN VEHICLE	01/26/2016 09:16:58 AM	400 S 400 W	суббота
26	BURGLARY	01/26/2016 07:23:21 AM	000 E 700 S	суббота
27	ESCAPE	01/26/2016 01:59:26 PM	2400 S 700 E	суббота
28	DRUGS	01/23/2016 10:58:10 PM	200 S RIO GRANDE ST	суббота

Всего в базе 5 столбцов, 1 классификационный и 4 описательных.

Поскольку ввод исходных данных в систему «Эйдос» планируется осуществить с помощью ее универсального программного интерфейса импорта данных из внешних баз данных, который работает с файлами MS Excel, то преобразуем данные из csv-файла в xls-файл, для удобства дальнейшего редактирования и выполним следующие операции.

Конвертируем исходный файл из csv формата в xls и запишем его с именем: Inp_data.xls в папку: c:\Aidos-X\AID_DATA\Inp_data\. В результате получим таблицу исходных данных, полностью подготовленную для обработки в системе «Эйдос» и записанную в нужную папку в виде файла нужного типа с нужным именем.

Автоматизированная формализация предметной области путем импорта исходных данных из внешних баз данных в систему "Эйдос".

Для загрузки базы исходных данных в систему «Эйдос» необходимо воспользоваться универсальным программным интерфейсом для ввода данных из внешних баз данных табличного вида, т.е. режимом 2.3.2.2.

Рисунок 1. Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

В экранной форме, приведенной на рисунке 5, задать настройки, показанные на рисунке:

- "Задайте тип файла исходных данных Inp_data": "XLS – MS Excel-2003";
- "Задайте диапазон столбцов классификационных шкал": "Начальный столбец классификационных шкал" – 2, "Конечный столбец классификационных шкал" – 2(последний столбец в таблице);
- "Задайте диапазон столбцов описательных шкал": "Начальный столбец описательных шкал" – 3, "Конечный столбец описательных шкал" – 24;
- "Задание параметров формирования сценариев или способа интерпретации текстовых полей": "Не применять сценарный метод АСК-анализа".

После нажать кнопку "ОК". Далее открывается окно, где размещена информация о размерности модели (рисунок 2). В этом окне необходимо нажать кнопку "Выйти на создание модели".

2.3.2.2. Задание размерности модели системы "Эйдос-Х++"

ИНФОРМАЦИЯ О РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [7 x 1697]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс.шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис.шкалу
Числовые	0	0	0,00	0	0	0,00
Текстовые	1	7	7,00	3	1697	565,67
ВСЕГО:	1	7	7,00	3	1697	565,67

Задайте число интервалов (градаций) в шкале:

Пересчитать шкалы и градации

Выйти на создание модели

Рисунок 2. Задание размерности модели системы "Эйдос"

Далее открывается окно, отображающее стадию процесса импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос" (рисунок 3), а также прогноз времени завершения этого процесса. В том окне необходимо

дождаться завершения формализации предметной области и нажать кнопку "ОК".

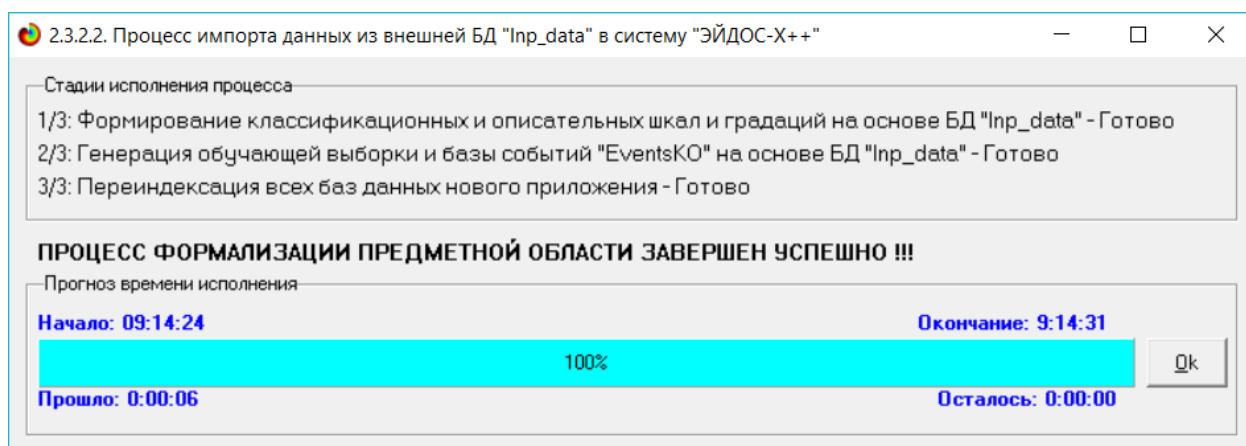


Рисунок 3. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

В результате формируются классификационные и описательные шкалы и градации, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Этим самым полностью автоматизировано выполняется 2-й этап АСК- анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1 (рисунок 7).

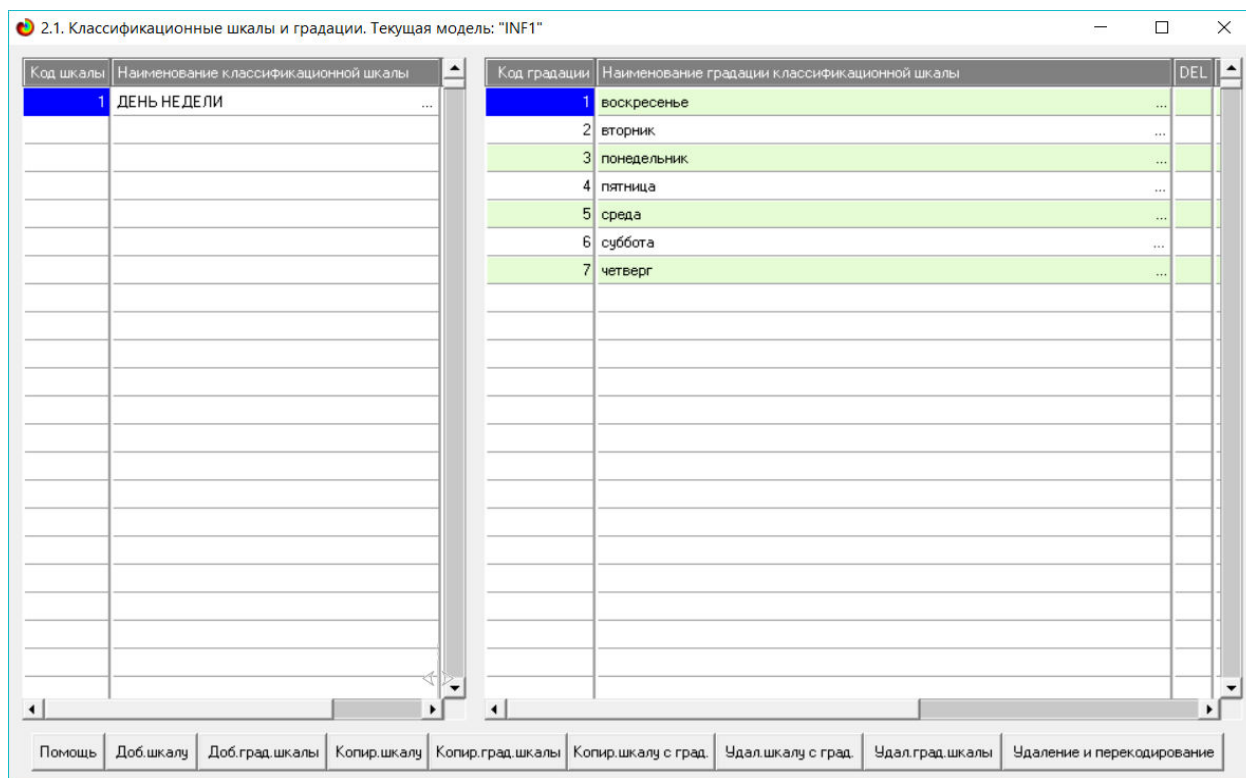


Рисунок 4. Классификационные шкалы и градации (фрагмент)

Для просмотра описательных шкал и градаций необходимо запустить режим 2.2 (рисунок 5):

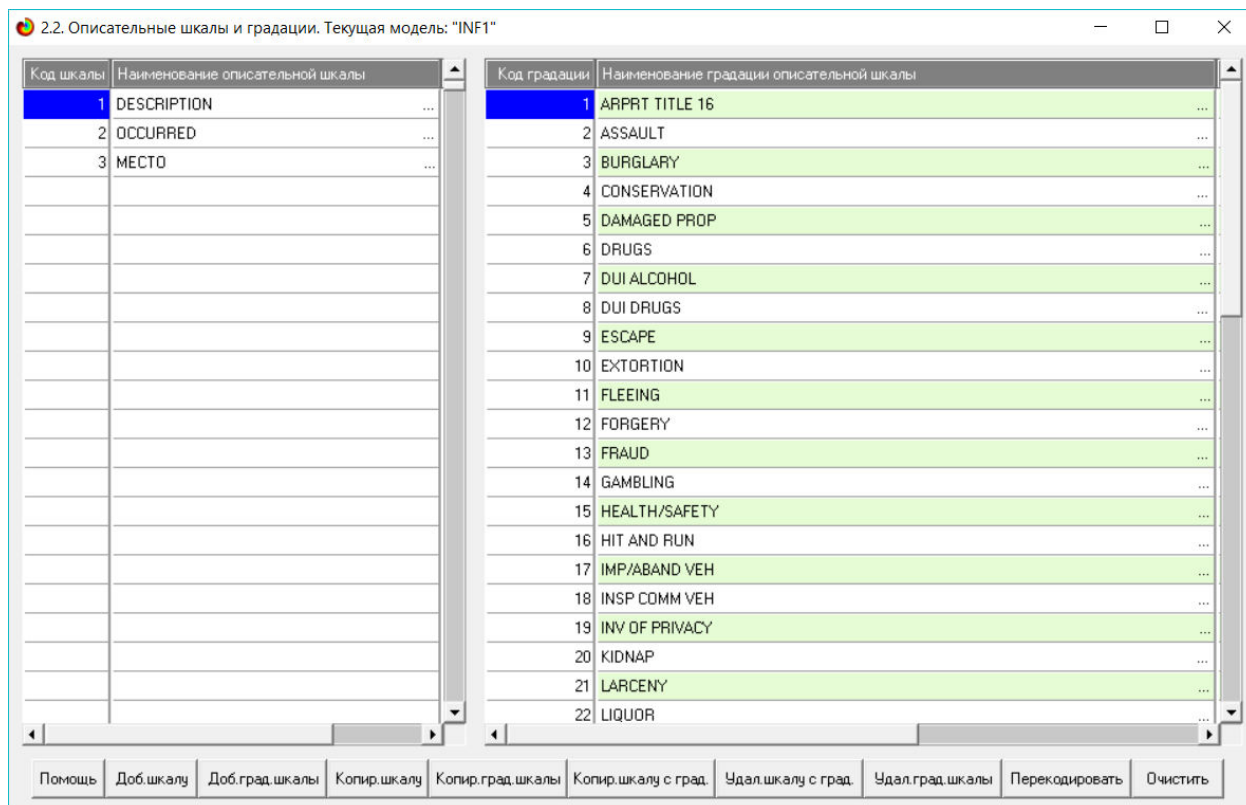


Рисунок 5. Описательные шкалы и градации (фрагмент)

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

Код объекта	Наименование объекта	Дата	Время
1	1	...	
2	2	...	
3	3	...	
4	4	...	
5	5	...	
6	6	...	
7	7	...	
8	8	...	
9	9	...	
10	10	...	

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

Код объекта	Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
1	34	37	1628	0	0	0	0

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

Рисунок 6. Обучающая выборка (фрагмент)

Тем самым создаются все необходимые и достаточные предпосылки для выявления силы и направления причинно-следственных связей между значениями факторов и результатами их совместного системного воздействия (с учетом нелинейности системы).

1.3. Синтез и верификация статистических и интеллектуальных моделей

Далее запускаем режим 3.5, в котором задаются модели для синтеза и верификации, а также задается модель, которой по окончании режима присваивается статус текущей (рисунок 7).

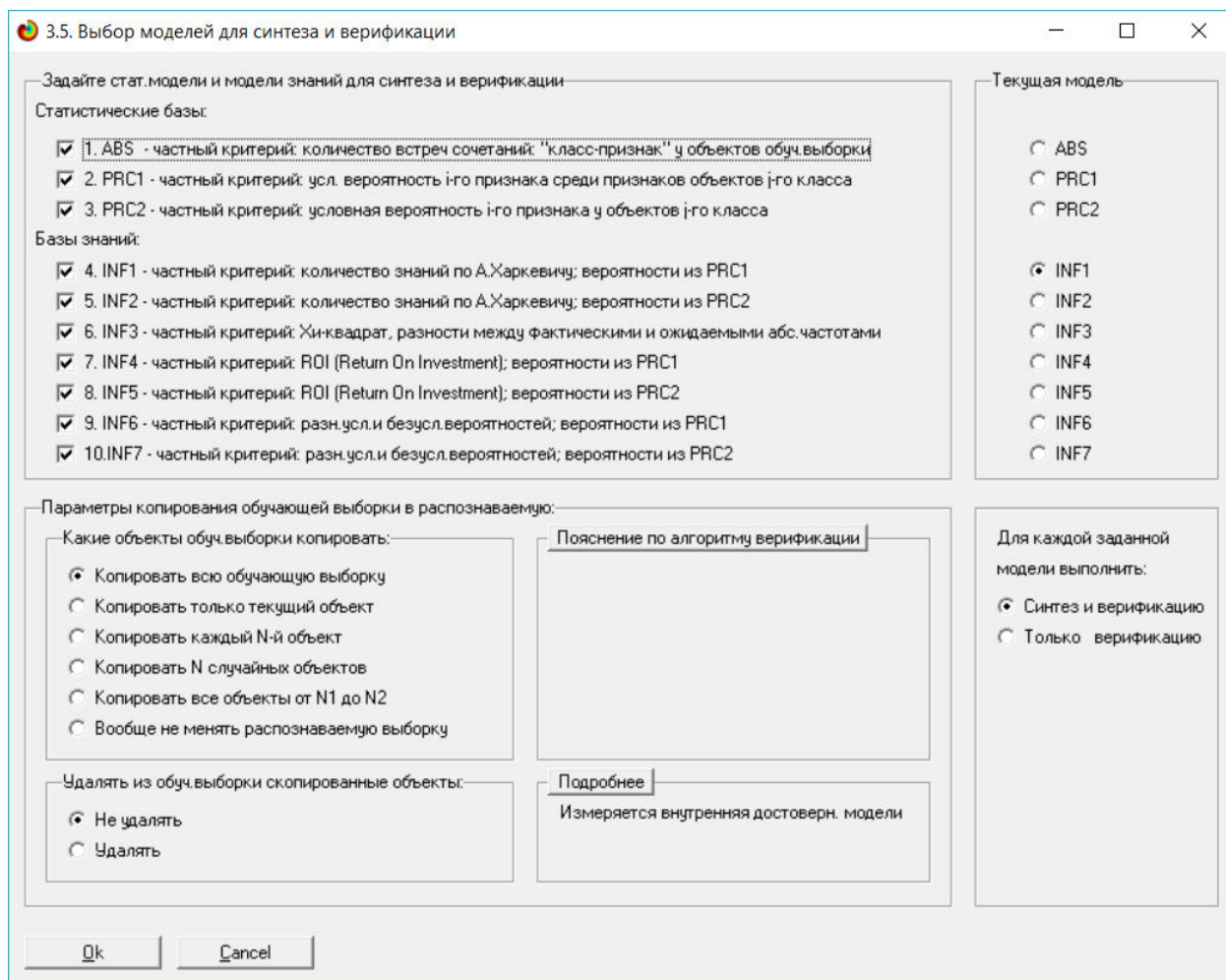


Рисунок 7. Выбор моделей для синтеза и верификации, а также текущей модели

В данном режиме имеется много различных методов верификации моделей, в том числе и поддерживающие бутстрепный метод. Но мы используем параметры по умолчанию, приведенные на рисунке 10. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 11.

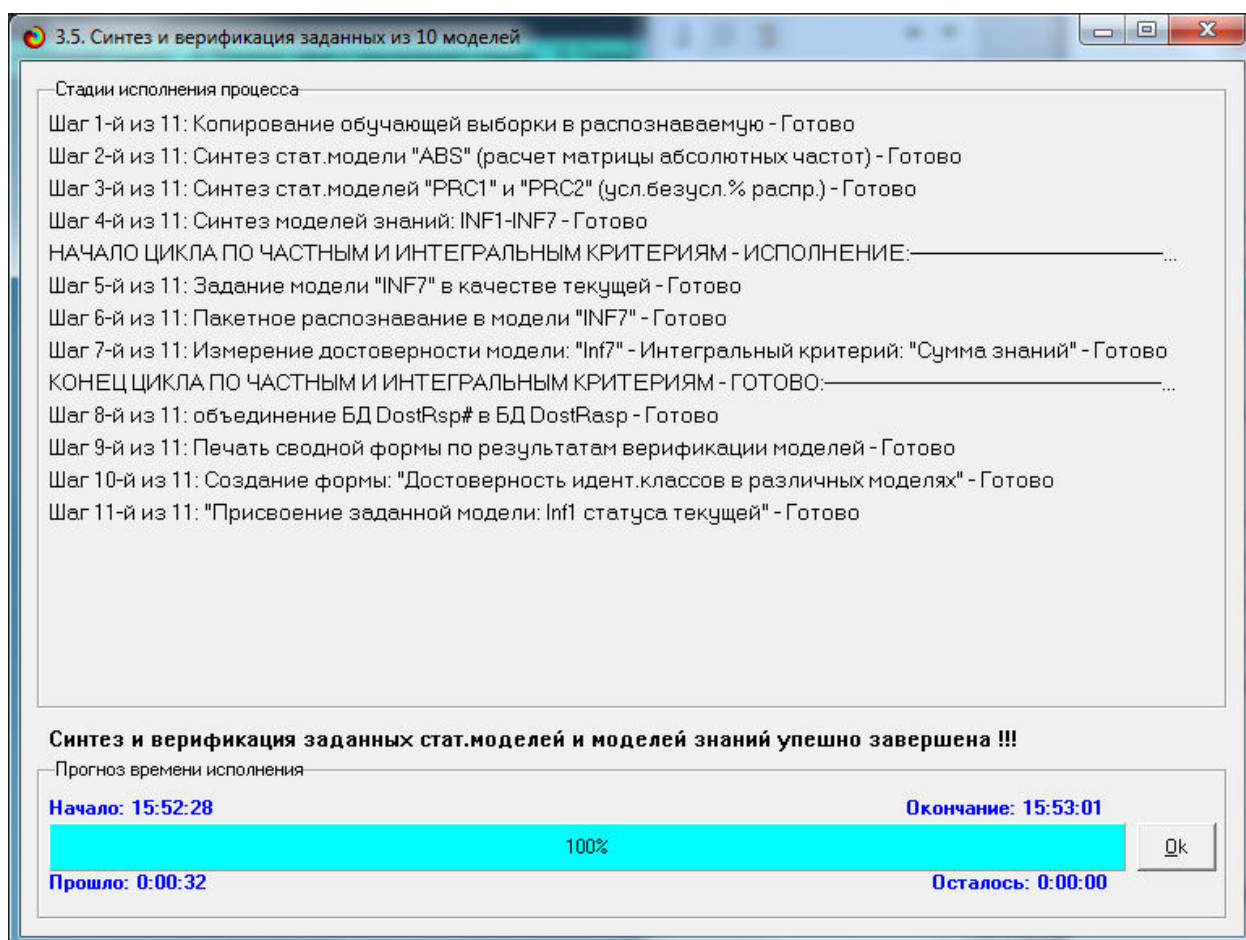


Рисунок 8. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 32 секунды. При этом верификация (оценка достоверности моделей) проводилась на всех 260 примерах наблюдения из обучающей выборки. В результате выполнения режима 3.5 созданы все модели, со всеми частными критериями, перечисленные на рисунке 10, но ниже мы приведем лишь некоторые из них (таблицы 2, 3, 4).

1.4. Виды моделей системы «Эйдос»

Рассмотрим решение задачи идентификации на примере модели INF3, 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: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. ДЕНЬ НЕДЕЛИ ВОСКРЕСЕ.	2. ДЕНЬ НЕДЕЛИ ВТОРНИК	3. ДЕНЬ НЕДЕЛИ ПОНЕДЕЛ.	4. ДЕНЬ НЕДЕЛИ ПЯТНИЦА	5. ДЕНЬ НЕДЕЛИ СРЕДА	6. ДЕНЬ НЕДЕЛИ СУББОТА	7. ДЕНЬ НЕДЕЛИ ЧЕТВЕРГ	Сумма	Среднее	Средн. квадрат. откл.
1	DESCRIPTION-ARPT TITLE 16		0.373			0.326		0.310	1.010	0.144	0.181
2	DESCRIPTION-ASSAULT	0.126	-0.060	0.043	-0.037	-0.107	0.035	-0.176	-0.175	-0.025	0.102
3	DESCRIPTION-BURGLARY	0.124	-0.239	0.058	-0.163	-0.046	0.075	-0.062	-0.254	-0.036	0.132
4	DESCRIPTION-CONSERVATION	0.078	0.273			0.226		0.210	0.787	0.112	0.121
5	DESCRIPTION-DAMAGED PROP	0.187	0.167	0.093	-0.192	-0.557	-0.071	-0.190	-0.562	-0.080	0.263
6	DESCRIPTION-DRUGS	-0.121	-0.229	-0.328	0.188	-0.003	0.075	0.112	-0.307	-0.044	0.189
7	DESCRIPTION-DUI ALCOHOL	-0.332	-0.136		0.181	0.057	0.302	-0.199	-0.127	-0.018	0.221
8	DESCRIPTION-DUI DRUGS						0.634		0.634	0.091	0.240
9	DESCRIPTION-ESCAPE	-0.045	-0.253	-0.033	0.014	0.125	-0.070	0.129	-0.133	-0.019	0.130
10	DESCRIPTION-EXTORTION		0.755						0.755	0.108	0.285
11	DESCRIPTION-FLEEING							0.692	0.692	0.099	0.262
12	DESCRIPTION-FORGERY	-0.117	0.320	0.058	0.014	0.031		0.016	0.323	0.046	0.133
13	DESCRIPTION-FRAUD	-0.078	0.196	-0.384	-0.428	0.148	-0.003	0.133	-0.416	-0.059	0.255
14	DESCRIPTION-GAMBLING	0.559							0.559	0.080	0.211
15	DESCRIPTION-HEALTH/SAFETY				0.691				0.691	0.099	0.261
16	DESCRIPTION-HIT AND RUN	-0.063		-0.129	0.309	-0.156	0.012	0.070	0.041	0.006	0.156
17	DESCRIPTION-IMP/ABAND VEH	0.178	0.514					0.070	0.761	0.109	0.191
18	DESCRIPTION-INSP COMM VEH					0.708			0.708	0.101	0.267
19	DESCRIPTION-INV OF PRIVACY	-0.204	0.169	-0.170	-0.073	0.044	0.048	0.106	-0.080	-0.011	0.141
20	DESCRIPTION-KIDNAP					0.708			0.708	0.101	0.267
21	DESCRIPTION-LARCENY	0.054	-0.007	-0.012	-0.056	0.004	0.062	-0.121	-0.075	-0.011	0.063
22	DESCRIPTION-LIQUOR		0.033	0.393	-0.032	0.226		-0.030	0.589	0.084	0.163
23	DESCRIPTION-MORALS-DECENCY				0.309		0.252	0.310	0.872	0.125	0.156
24	DESCRIPTION-MOV TRAF VIOL				0.550			0.310	0.860	0.123	0.221

Таблица 3 – Матрица информативностей (модель INF1) в битах (фрагмент)

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

Код признака	Наименование описательной шкалы и градации	1. ДЕНЬ НЕДЕЛИ ВОСКРЕСЕНЬЕ	2. ДЕНЬ НЕДЕЛИ ВТОРНИК	3. ДЕНЬ НЕДЕЛИ ПОНЕДЕЛ	4. ДЕНЬ НЕДЕЛИ ПЯТНИЦА	5. ДЕНЬ НЕДЕЛИ СРЕДА	6. ДЕНЬ НЕДЕЛИ СУББОТА	7. ДЕНЬ НЕДЕЛИ ЧЕТВЕРГ	Сумма	Среднее	Средн. квадрат. откл.
1	DESCRIPTION-ARPR TITLE 16		0.373			0.326		0.310	1.010	0.144	0.181
2	DESCRIPTION-ASSAULT	0.126	-0.060	0.043	-0.037	-0.107	0.035	-0.176	-0.175	-0.025	0.102
3	DESCRIPTION-BURGLARY	0.124	-0.239	0.058	-0.163	-0.046	0.075	-0.060	-0.254	-0.036	0.132
4	DESCRIPTION-CONSERVATION	0.078	0.273			0.226		0.210	0.787	0.112	0.121
5	DESCRIPTION-DAMAGED PROP	0.187	0.167	0.093	-0.192	-0.557	-0.071	-0.190	-0.562	-0.080	0.263
6	DESCRIPTION-DRUGS	-0.121	-0.229	-0.328	0.188	-0.003	0.075	0.112	-0.307	-0.044	0.189
7	DESCRIPTION-DUI ALCOHOL	-0.332	-0.136		0.181	0.057	0.302	-0.199	-0.127	-0.018	0.221
8	DESCRIPTION-DUI DRUGS						0.634		0.634	0.091	0.240
9	DESCRIPTION-ESCAPE	-0.045	-0.253	-0.033	0.014	0.125	-0.070	0.129	-0.133	-0.019	0.130
10	DESCRIPTION-EXTORTION		0.755						0.755	0.108	0.285
11	DESCRIPTION-FLEEING							0.692	0.692	0.099	0.262
12	DESCRIPTION-FORGERY	-0.117	0.320	0.058	0.014	0.031		0.016	0.323	0.046	0.133
13	DESCRIPTION-FRAUD	-0.078	0.196	-0.384	-0.428	0.148	-0.003	0.133	-0.416	-0.059	0.255
14	DESCRIPTION-GAMBLING	0.559							0.559	0.080	0.211
15	DESCRIPTION-HEALTH/SAFETY				0.691				0.691	0.099	0.261
16	DESCRIPTION-HIT AND RUN	-0.063		-0.129	0.309	-0.156	0.012	0.070	0.041	0.006	0.156
17	DESCRIPTION-IMP/ABAND VEH	0.178	0.514					0.070	0.761	0.109	0.191
18	DESCRIPTION-INSP COMM VEH					0.708			0.708	0.101	0.267
19	DESCRIPTION-INV OF PRIVACY	-0.204	0.169	-0.170	-0.073	0.044	0.048	0.106	-0.080	-0.011	0.141
20	DESCRIPTION-KIDNAP					0.708			0.708	0.101	0.267
21	DESCRIPTION-LARCENY	0.054	-0.007	-0.012	-0.056	0.004	0.062	-0.121	-0.075	-0.011	0.063
22	DESCRIPTION-LIQUOR		0.033	0.393	-0.032	0.226		-0.030	0.589	0.084	0.163
23	DESCRIPTION-MORALS-DECENCY				0.309		0.252	0.310	0.872	0.125	0.156
24	DESCRIPTION-MOV TRAF VIOL				0.550			0.310	0.860	0.123	0.221

Таблица 4 – Матрица знаний (модель INF3) (фрагмент)

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

Код признака	Наименование описательной шкалы и градации	1. ДЕНЬ НЕДЕЛИ ВОСКРЕСЕНЬЕ	2. ДЕНЬ НЕДЕЛИ ВТОРНИК	3. ДЕНЬ НЕДЕЛИ ПОНЕДЕЛ	4. ДЕНЬ НЕДЕЛИ ПЯТНИЦА	5. ДЕНЬ НЕДЕЛИ СРЕДА	6. ДЕНЬ НЕДЕЛИ СУББОТА	7. ДЕНЬ НЕДЕЛИ ЧЕТВЕРГ	Сумма	Среднее	Средн. квадрат. откл.
1	DESCRIPTION-ARPR TITLE 16	-0.600	0.659	-0.363	-0.411	0.608	-0.484	0.591			0.584
2	DESCRIPTION-ASSAULT	6.400	-1.307	1.172	-1.005	-2.530	1.229	-3.960			3.382
3	DESCRIPTION-BURGLARY	3.000	-1.983	0.767	-1.797	-0.569	1.357	-0.775			1.803
4	DESCRIPTION-CONSERVATION	0.200	0.545	-0.484	-0.548	0.478	-0.645	0.454			0.536
5	DESCRIPTION-DAMAGED PROP	5.400	2.676	1.405	-2.208	-3.961	-1.127	-2.184			3.289
6	DESCRIPTION-DRUGS	-5.000	-4.673	-6.279	8.350	-0.096	3.295	4.403			5.571
7	DESCRIPTION-DUI ALCOHOL	-1.600	-0.479	-1.572	1.218	0.303	2.904	-0.774			1.626
8	DESCRIPTION-DUI DRUGS	-0.200	-0.114	-0.121	-0.137	-0.131	0.839	-0.136			0.371
9	DESCRIPTION-ESCAPE	-2.200	-5.356	-1.005	0.528	5.121	-2.673	5.584			4.069
10	DESCRIPTION-EXTORTION	-0.200	0.886	-0.121	-0.137	-0.131	-0.161	-0.136			0.392
11	DESCRIPTION-FLEEING	-0.200	-0.114	-0.121	-0.137	-0.131	-0.161	0.864			0.382
12	DESCRIPTION-FORGERY	-0.400	1.203	0.153	0.041	0.086	-1.129	0.045			0.697
13	DESCRIPTION-FRAUD	-1.000	2.155	-2.023	-2.426	1.736	-0.031	1.589			1.878
14	DESCRIPTION-GAMBLING	0.800	-0.114	-0.121	-0.137	-0.131	-0.161	-0.136			0.353
15	DESCRIPTION-HEALTH/SAFETY	-0.200	-0.114	-0.121	0.863	-0.131	-0.161	-0.136			0.382
16	DESCRIPTION-HIT AND RUN	-0.400	-1.366	-0.451	2.355	-0.567	0.065	0.363			1.171
17	DESCRIPTION-IMP/ABAND VEH	0.800	2.317	-0.726	-0.822	-0.783	-0.967	0.181			1.211
18	DESCRIPTION-INSP COMM VEH	-0.200	-0.114	-0.121	-0.137	0.869	-0.161	-0.136			0.384
19	DESCRIPTION-INV OF PRIVACY	-2.400	1.927	-1.265	-0.700	0.475	0.647	1.316			1.524
20	DESCRIPTION-KIDNAP	-0.200	-0.114	-0.121	-0.137	0.869	-0.161	-0.136			0.384
21	DESCRIPTION-LARCENY	6.600	-0.418	-0.823	-4.000	0.283	6.236	-7.878			5.187
22	DESCRIPTION-LIQUOR	-1.600	0.090	2.033	-0.096	0.956	-1.290	-0.091			1.245
23	DESCRIPTION-MORALS-DECENCY	-0.600	-0.341	-0.363	0.589	-0.392	0.516	0.591			0.536
24	DESCRIPTION-MOV TRAF VIOL	-0.600	-0.341	-0.363	1.589	-0.392	-0.484	0.591			0.803

1.5. Результаты верификации моделей

Результаты верификации (оценки достоверности) моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями приведены на рисунке 9.

4.1.3.6. Обобщенная форма по доств.моделям при разн.критер. Текущая модель: "INF1"													
Наименование модели и частного критерия	Интегральный критерий	Личный модуль сходств. истинно-отриц. решений	Средняя модель уровней сходств. ложно-положит. решений	Средняя модель уровней сходств. ложно-отриц. решений	А-Точность модели A-Precision = ATR/ATR	А-Полнота модели A-Recall = ATR/ATR	L2-мера проф. Е.В.Луценко	Процент правильных идентификаций	Процент правильных не идентиф.	Процент ошибочных идентификаций	Процент ошибочных не идентиф.	Процент правильных результатов	Дата получения результата
1. ABS - частный критерий: количество встреч сочетаний "Клас.	Корреляция абс. частот с обр...		0.351	0.016	0.546	1.000	0.706	100.000	3.992	96.008		51.996	06.06.201
1. ABS - частный критерий: количество встреч сочетаний "Клас.	Средн. абс. частот по призна...		0.255		0.557	1.000	0.715	100.000	3.242	96.758		51.621	06.06.201
2. PRC1 - частный критерий: усл. вероятность н-го признака сред.	Корреляция усл. отн. частот с о...		0.351	0.016	0.546	1.000	0.706	100.000	3.992	96.008		51.996	06.06.201
2. PRC1 - частный критерий: усл. вероятность н-го признака сред.	Средн. усл. отн. частот по при...		0.330		0.546	1.000	0.706	100.000	3.242	96.758		51.621	06.06.201
3. PRC2 - частный критерий: условная вероятность н-го признака.	Корреляция усл. отн. частот с о...		0.351	0.016	0.546	1.000	0.706	100.000	3.992	96.008		51.996	06.06.201
3. PRC2 - частный критерий: условная вероятность н-го признака.	Средн. усл. отн. частот по при...		0.332		0.546	1.000	0.706	100.000	3.242	96.758		51.621	06.06.201
4. INF1 - частный критерий: количество знаний по АХаркевичу; в.	Семантический резонанс зна...	0.011	0.074	0.169	0.862	0.977	0.916	99.278	94.882	5.118	0.722	97.080	06.06.201
4. INF1 - частный критерий: количество знаний по АХаркевичу; в.	Средн. знаний		0.061	0.059	0.896	1.000	0.945	100.000	51.668	48.332		75.834	06.06.201
5. INF2 - частный критерий: количество знаний по АХаркевичу; в.	Семантический резонанс зна...	0.012	0.073	0.170	0.863	0.975	0.916	99.278	94.812	5.188	0.722	97.045	06.06.201
5. INF2 - частный критерий: количество знаний по АХаркевичу; в.	Средн. знаний		0.061	0.058	0.896	1.000	0.945	100.000	51.705	48.295		75.852	06.06.201
6. INF3 - частный критерий: Хинкватер, разности между фактис...	Семантический резонанс зна...	0.240	0.241	0.230	0.568	0.569	0.569	81.137	57.972	42.028	18.863	69.554	06.06.201
6. INF3 - частный критерий: Хинкватер, разности между фактис...	Средн. знаний	0.251	0.243	0.229	0.572	0.564	0.568	81.137	57.972	42.028	18.863	69.554	06.06.201
7. INF4 - частный критерий: ROI (Return On Investment); веро...	Семантический резонанс зна...		0.042	0.143	0.910	1.000	0.953	100.000	98.336	1.664		99.168	06.06.201
7. INF4 - частный критерий: ROI (Return On Investment); веро...	Средн. знаний		0.028	0.014	0.944	1.000	0.971	100.000	50.979	49.021		75.490	06.06.201
8. INF5 - частный критерий: ROI (Return On Investment); веро...	Семантический резонанс зна...		0.042	0.143	0.910	1.000	0.953	100.000	98.336	1.664		99.168	06.06.201
8. INF5 - частный критерий: ROI (Return On Investment); веро...	Средн. знаний		0.028	0.013	0.944	1.000	0.971	100.000	50.990	49.010		75.495	06.06.201
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер.	Семантический резонанс зна...	0.244	0.256	0.224	0.546	0.558	0.552	78.791	59.926	40.074	21.209	69.358	06.06.201
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер.	Средн. знаний	0.241	0.230	0.221	0.573	0.561	0.567	81.227	52.989	47.011	18.773	67.108	06.06.201
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	0.238	0.239	0.238	0.567	0.568	0.568	78.159	57.271	42.729	21.841	67.715	06.06.201
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Средн. знаний	0.237	0.235	0.224	0.573	0.571	0.572	80.776	53.185	46.815	19.224	66.981	06.06.201

Рисунок 9. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели ABS, INF4 при интегральном критерии «Семантический резонанс знаний». При этом точность модели составляет 0,971. Таким образом, уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 69%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием (рисунок 10).

Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++"													
Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++".													
положительный псевдопрогноз. Предположим, модель дает такой прогноз: выпадет 1, 2, 3, 4, 5 или 6. В этом случае у нее будет 100% достоверность идентификации, т.е. не будет ни одного объекта, не отнесенного к тому классу, к которому он действительно относится, но при этом будет очень большая ошибка ложной идентификации, т.к. огромное количество объектов будет отнесено к классам, к которым они не относятся (и именно за счет этого у модели и будет очень высокая достоверность идентификации). Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.													
отрицательный псевдопрогноз. Представим себе, что мы выбираем кубик с 6 гранями, и модель предсказывает, что не выпадет: 1, 2, 3, 4, 5 и 6, а что-то из этого естественно выпало. Конечно, модель дает ошибку в прогнозе в том плане, что не предсказала, что выпадет, зато она очень хорошо угадала, что не выпадет. Но ясно, что выпадет что-то одно, а не все, что предсказано, поэтому такого рода предсказания хорошо оправдываются в том, что не произошло и плохо в том, что произошло, т.е. в этом случае у модели будет 100% достоверность не идентификации, но очень низкая достоверность идентификации.													
идеальный прогноз. Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.													
реальный прогноз. На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представьте себе, что у вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.													
Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, к которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В.Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {-1, +1} и {0, 1}: $L2 = \frac{TP + TN - FP - FN}{TP + TN + FP + FN}$ (нормировка: {-1, +1}) $L2 = \frac{1 + (TP + TN - FP - FN)}{TP + TN + FP + FN}$ / 2 (нормировка: {0, 1}) где: TP – истинно-положительное решение; TN – истинно-отрицательное решение; FP – ложно-положительное решение; FN – ложно-отрицательное решение;													
Классическая F-мера достоверности моделей Ван Ризбергена: $Precision = TP / (TP + FP)$ – точность модели; $Recall = TP / (TP + FN)$ – полнота модели; $F\text{-мера} = 2 * (Precision * Recall) / (Precision + Recall).$													
В АСК-анализе и системе "Эйдос" предлагается L-мера, представляющая собой нечеткое мультиклассовое обобщение классической F-меры достоверности моделей Ван Ризбергена: $SPrecision = STP / (STP + SFP)$ – нечеткая мультиклассовая точность модели; $SRecall = STP / (STP + SFN)$ – нечеткая мультиклассовая полнота модели; $L\text{-мера} = 2 * (SPrecision * SRecall) / (SPrecision + SRecall)$ – нечеткая мультиклассовая достоверность модели.													
Где: STP – Сумма модулей сходства истинно-положительных решений; STN – Сумма модулей сходства истинно-отрицательных решений; SFP – Сумма модулей сходства ложно-положительных решений; SFN – Сумма модулей сходства ложно-отрицательных решений.													

Рисунок 10. Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

Также обращает на себя внимание, что статистические модели, как правило, дают более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, и практически никогда – более высокую. Этим и оправдано применение моделей знаний и интеллектуальных технологий. На рисунке 11 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и неидентифицированных ситуаций в наиболее достоверной модели INF4.

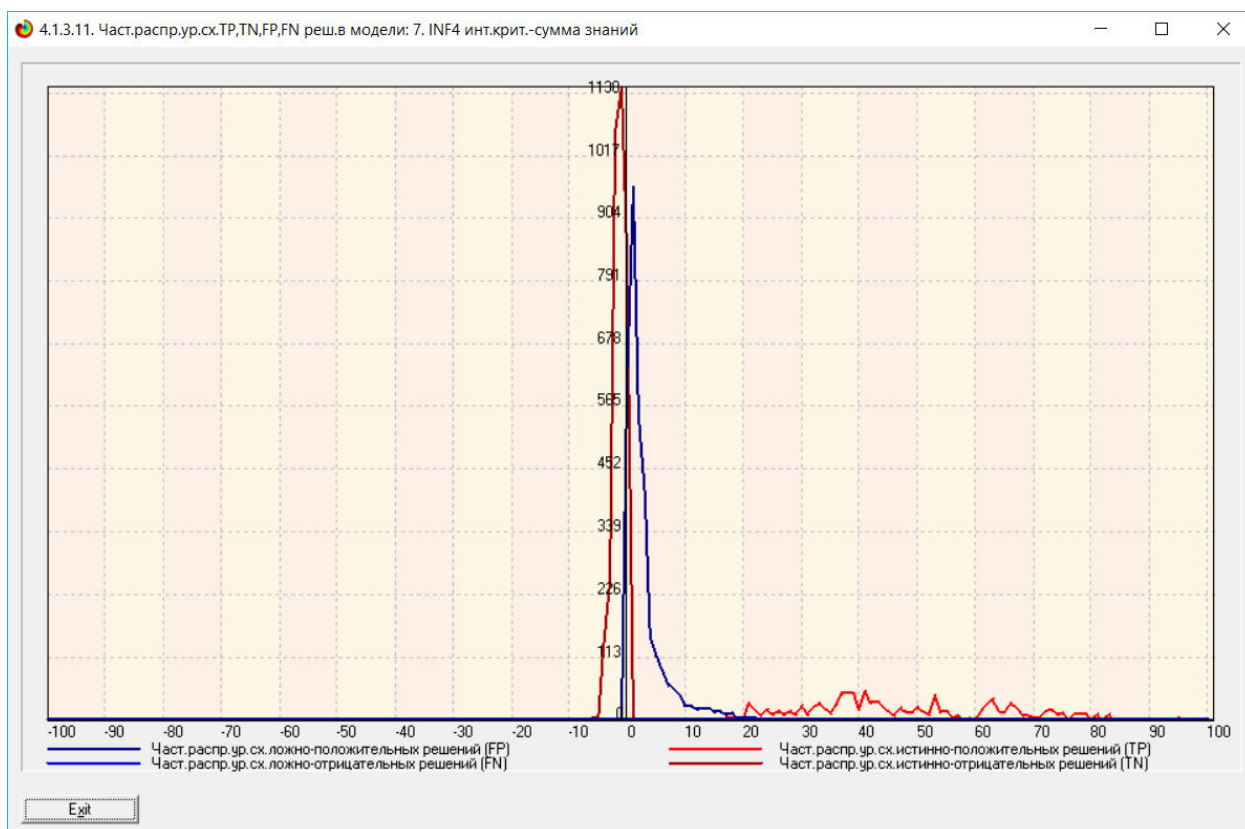


Рисунок 11. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF4.

Из рисунка 11 видно, что:

- наиболее достоверная модель INF4 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 9);

– модуль уровня сходства-различия в наиболее достоверной модели INF4 для верно идентифицированных и верно неидентифицированных объектов значительно выше, чем для ошибочно идентифицированных и ошибочно неидентифицированных. Это верно практически для всего диапазона уровней сходства-различия, кроме небольших по модулю значений в диапазоне от 0 до 15% уровня сходства. Для уровня сходства выше 70% все положительные решения являются истинными. Для очень больших значений уровней сходства-различия (более 70%) также различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями практически отсутствует.

Любые данные о наблюдениях можно считать суммой истинного значения и шума, причем ни первое, ни второе неизвестны. Поэтому имеет смысл сравнить созданные модели с чисто случайными моделями, совпадающими по основным характеристикам. В системе «Эйдос» есть лабораторная работа № 2.01: «Исследование RND-модели при различных объемах выборки». Если данная работа устанавливается при отсутствии текущего приложения, то все параметры создаваемых моделей задаются вручную, если же текущая модель существует, как в нашем случае, то все основные ее параметры определяются автоматически.

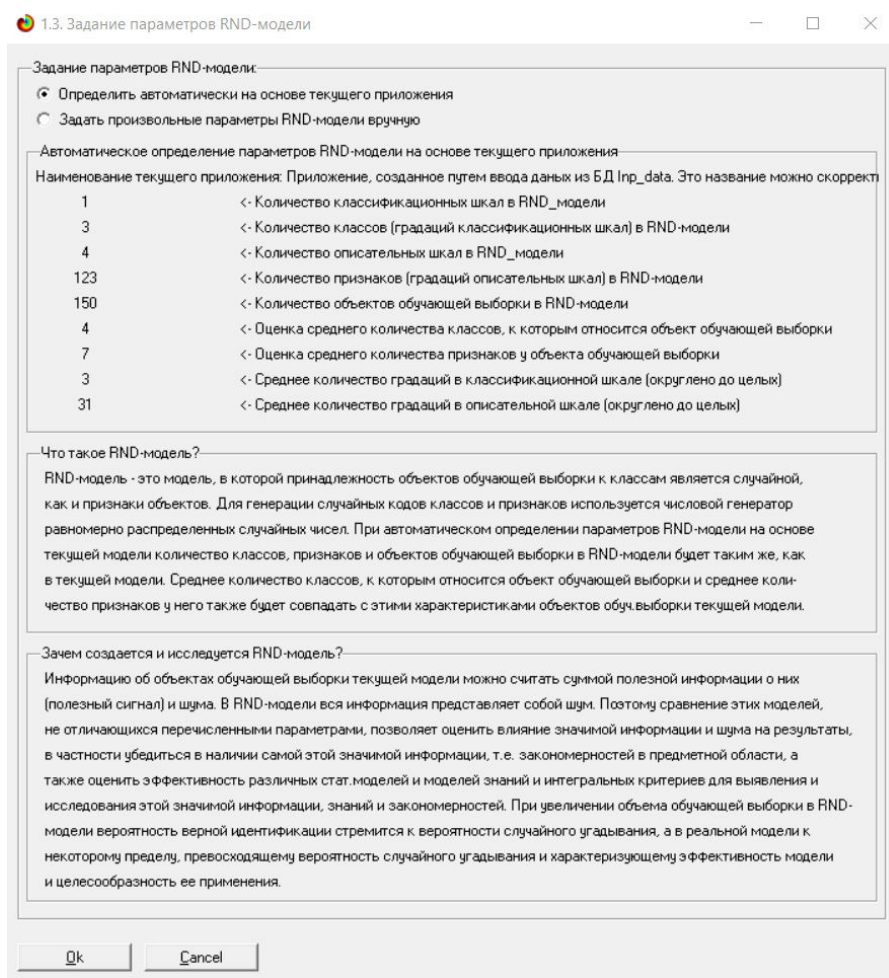


Рисунок 12. Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

На рисунке 13 показано частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF3.

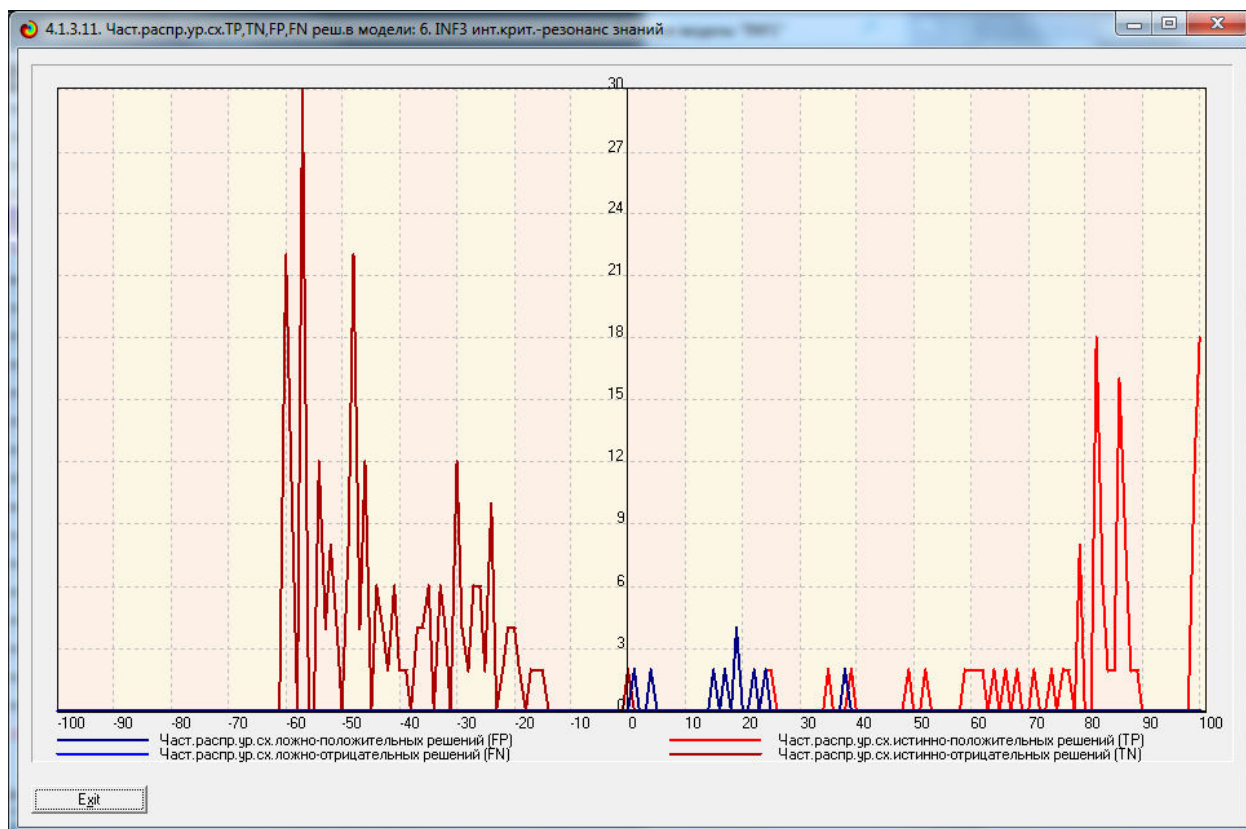


Рисунок 13. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF3

Совершенно очевидное различие частотных распределений уровней сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования и случайной модели (рисунки 11 и 13) объясняется тем, что в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны. Если же такой информации в модели нет, то и распределение получается типа, приведенного на рисунке 13.

На рисунке 14 приведены данные по достоверности статистических и когнитивных моделей, созданных на основе случайной выборки.

413.6. Обобщенная форма по достов. моделям при разн. крит. Тестуемая модель: "INF1"													
Наименование модели и частного критерия	Интегральный критерий	Средний модуль условной соглас. ложно-положит. решений	Средний модуль условной соглас. ложно-отрицат. решений	A-Точность модели A/Precision = ATR/ATR	A-Полнота модели A/Rcall = ATR/ATR	L2-мера проф. Е.В. Пыцкого	Процент правильных идентификаций	Процент правильных не идентификаций	Процент ошибочной идентификации	Процент ошибочной не идентификации	Процент правильных результатов	Дата получения результата	Время получения результата
1. ABS - частный критерий: количество встреч сопоставлений "х/лас"	Корреляция абс частот с абс частот по признакам	0.173	0.419	0.819	1.000	0.900	100.000	92.424	7.576		96.212	29.05.2017	15:52:38
1. ABS - частный критерий: количество встреч сопоставлений "х/лас"	Сумма абс частот по признакам	0.180		0.820	1.000	0.901	100.000			100.000	50.000	29.05.2017	15:52:38
2. PRIC1 - частный критерий: усл. вероятность его признака сред.	Корреляция усл.отн частот с о...	0.173	0.419	0.819	1.000	0.900	100.000	92.424	7.576		96.212	29.05.2017	15:52:40
2. PRIC1 - частный критерий: усл. вероятность его признака сред.	Сумма усл.отн частот по признакам	0.180		0.820	1.000	0.901	100.000			100.000	50.000	29.05.2017	15:52:40
3. PRIC2 - частный критерий: условная вероятность его признака	Корреляция усл.отн частот с о...	0.173	0.419	0.819	1.000	0.900	100.000	92.424	7.576		96.212	29.05.2017	15:52:42
3. PRIC2 - частный критерий: условная вероятность его признака	Сумма усл.отн частот по признакам	0.180		0.820	1.000	0.901	100.000			100.000	50.000	29.05.2017	15:52:42
4. INF1 - частный критерий: количество знаний по АХаркевичу, в.	Семантический резонанс зна...	0.179	0.404	0.807	0.892	0.847	92.424	90.909	9.091	7.576	91.667	29.05.2017	15:52:45
4. INF1 - частный критерий: количество знаний по АХаркевичу, в.	Сумма знаний	0.099	0.355	0.825	0.681	0.746	93.939	94.697	5.303	6.061	94.318	29.05.2017	15:52:45
5. INF2 - частный критерий: количество знаний по АХаркевичу, в.	Семантический резонанс зна...	0.179	0.404	0.807	0.892	0.847	92.424	90.909	9.091	7.576	91.667	29.05.2017	15:52:47
5. INF2 - частный критерий: количество знаний по АХаркевичу, в.	Сумма знаний	0.099	0.355	0.825	0.681	0.746	93.939	94.697	5.303	6.061	94.318	29.05.2017	15:52:48
6. INF3 - частный критерий: Хинкватер, разности между фактич.	Семантический резонанс зна...	0.174	0.432	0.818	1.000	0.900	100.000	93.182	6.818		96.591	29.05.2017	15:52:50
6. INF3 - частный критерий: Хинкватер, разности между фактич.	Сумма знаний	0.141	0.396	0.834	1.000	0.909	100.000	92.424	7.576		96.212	29.05.2017	15:52:50
7. INF4 - частный критерий: ROI (Return On Investment): веротно...	Семантический резонанс зна...	0.180	0.410	0.807	0.856	0.831	98.485	93.939	6.061	1.515	96.212	29.05.2017	15:52:52
7. INF4 - частный критерий: ROI (Return On Investment): веротно...	Сумма знаний	0.168	0.146	0.812	1.000	0.896	100.000	86.364	13.636		93.182	29.05.2017	15:52:53
8. INF5 - частный критерий: ROI (Return On Investment): веротно...	Семантический резонанс зна...	0.180	0.410	0.807	0.856	0.831	98.485	93.939	6.061	1.515	96.212	29.05.2017	15:52:55
8. INF5 - частный критерий: ROI (Return On Investment): веротно...	Сумма знаний	0.168	0.146	0.812	1.000	0.896	100.000	86.364	13.636		93.182	29.05.2017	15:52:55
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер.	Семантический резонанс зна...	0.170	0.413	0.819	0.918	0.866	96.970	93.182	6.818	3.030	95.076	29.05.2017	15:52:57
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер.	Сумма знаний	0.169	0.144	0.808	1.000	0.894	100.000	86.364	13.636		93.182	29.05.2017	15:52:58
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве.	Семантический резонанс зна...	0.170	0.413	0.819	0.918	0.866	96.970	93.182	6.818	3.030	95.076	29.05.2017	15:53:00
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве.	Сумма знаний	0.169	0.144	0.808	1.000	0.894	100.000	86.364	13.636		93.182	29.05.2017	15:53:00

Рисунок 14. Достоверность статистических и когнитивных моделей, созданных на основе случайной выборки

На основе его сравнения с рисунком 9 можно сделать следующие выводы:

- достоверность лучшей модели INF4, отражающей реальный объект моделирования, примерно на 20% выше, чем аналогичной случайной модели;
- различие между достоверностью статистических моделей и моделей знаний, созданных на основе случайной выборки, значительно меньше, чем у моделей, отражающих реальный объект моделирования;
- в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны, причем примерно 1/3 достоверности обусловлена отражением в реальных моделях закономерностей предметной области, а 2/3 достоверности обусловлено наличием шума в исходных данных. На основании этого можно предположить, что в исходных данных уровень сигнала о реальных причинно-следственных связях в моделируемой предметной области примерно в два раза ниже уровня шума.

2. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи идентификации

В соответствии с технологией АСК-анализа зададим текущей модель INF5 (режим 5.6) (рисунок 15) и проведем пакетное распознавание в режиме 4.1.2.

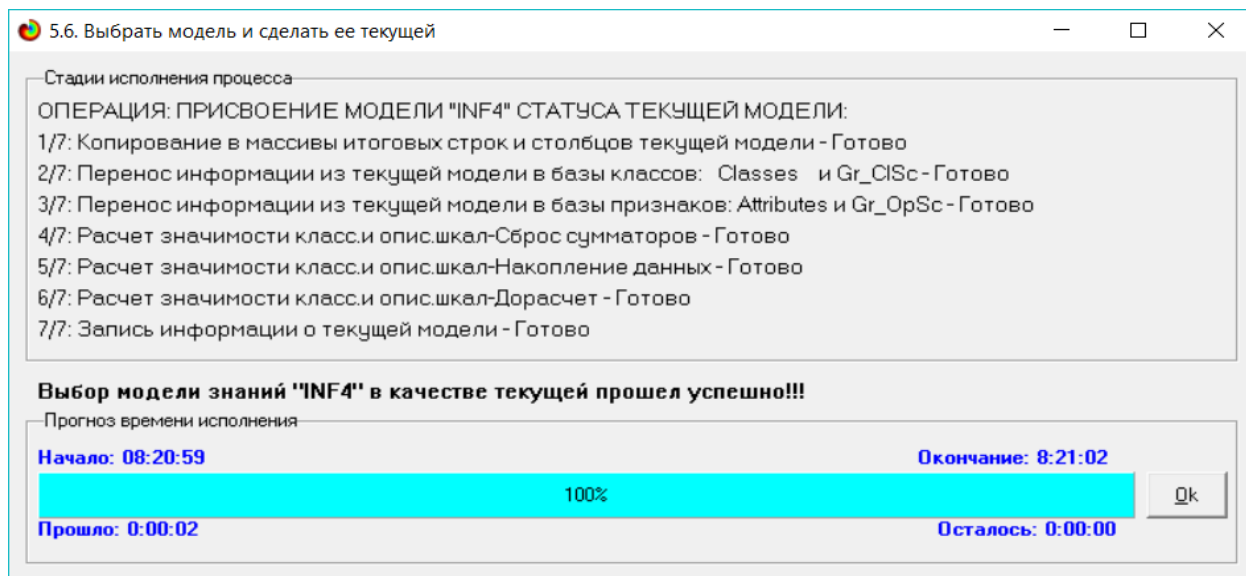


Рисунок 14. Стадии исполнения процесса выбора модели.

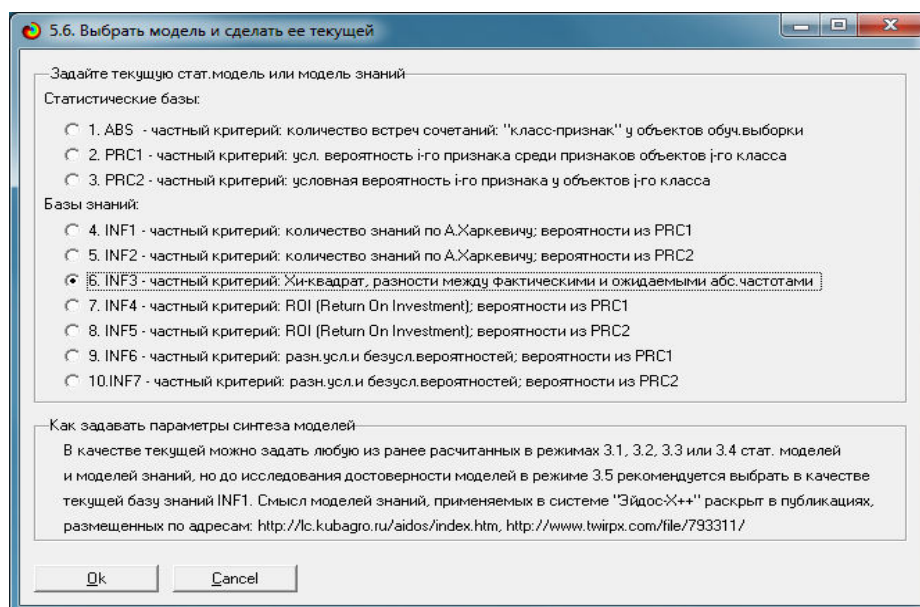


Рисунок 15. Экранные формы режима задания модели в качестве текущей

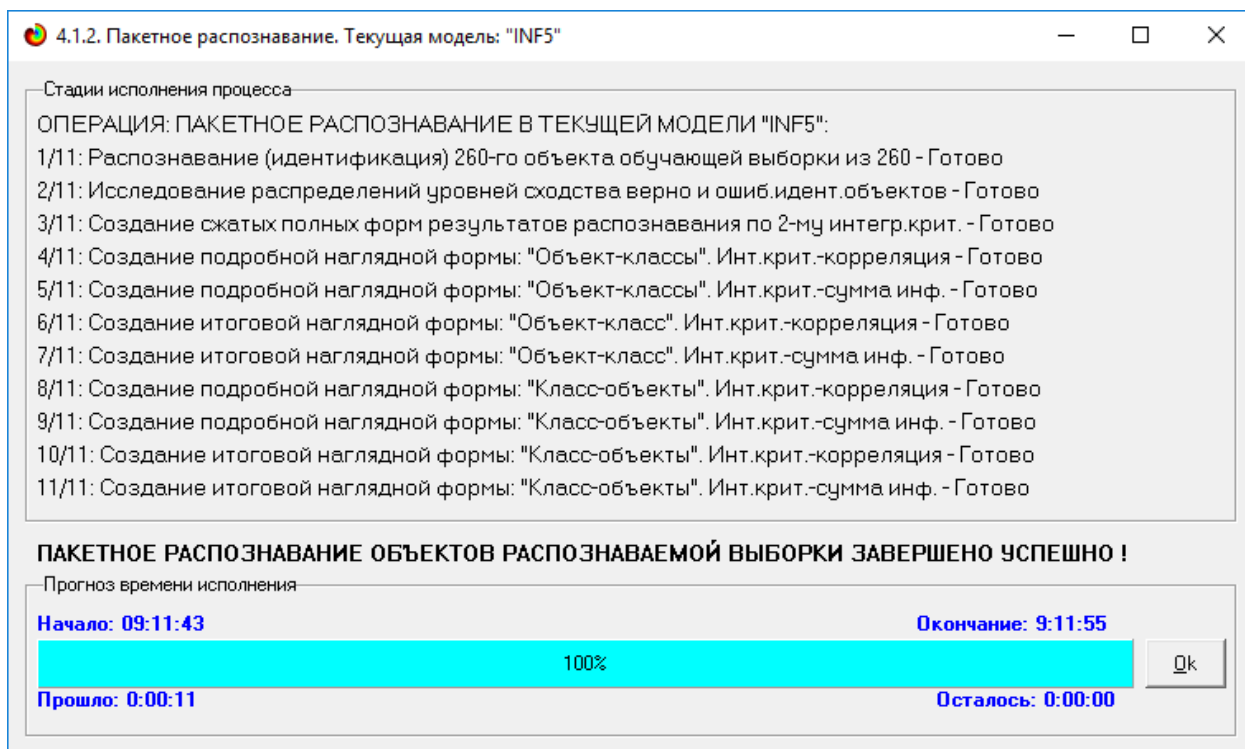


Рисунок 16. Экранная форма режима пакетного распознавания в текущей модели

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 «Эйдос» 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 17 и 18 приведены примеры прогнозов высокой и низкой достоверности частоты и классов ирисов в наиболее достоверной модели INF4 на основе наблюдения предыстории их развития:

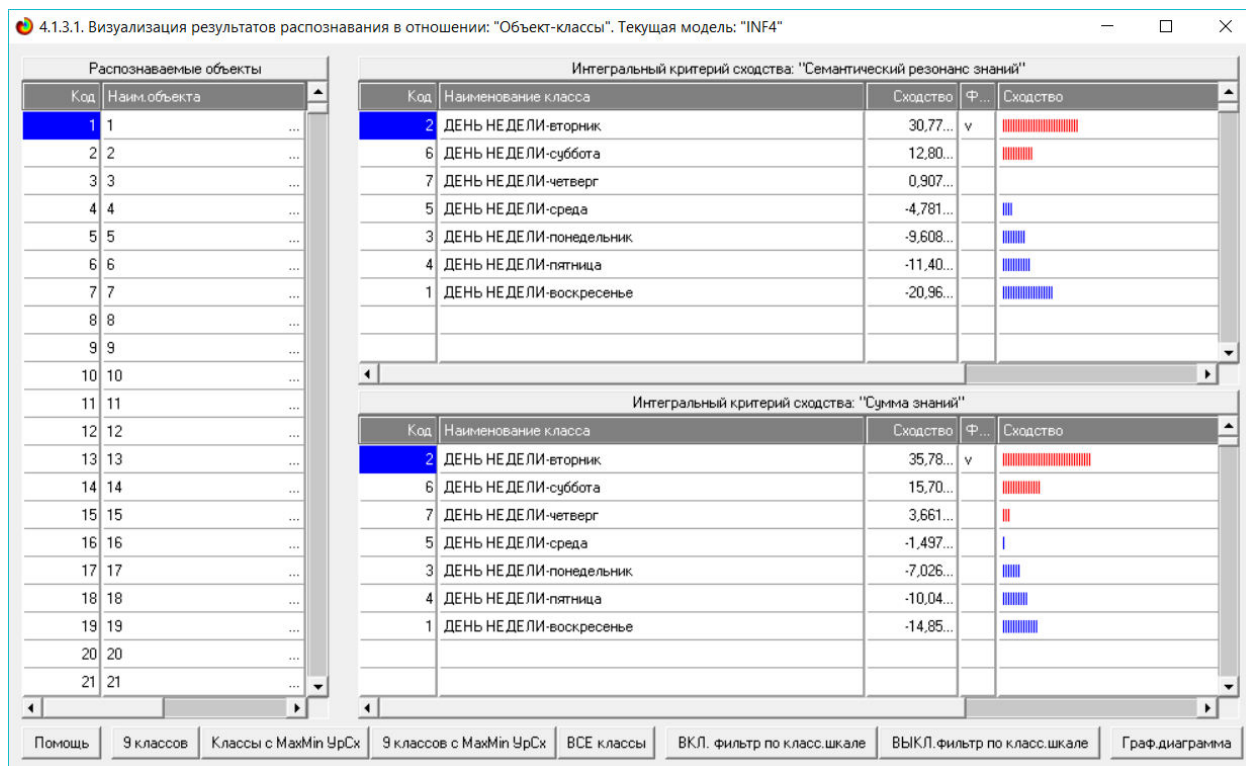


Рисунок 17. Пример идентификации классов в модели INF4

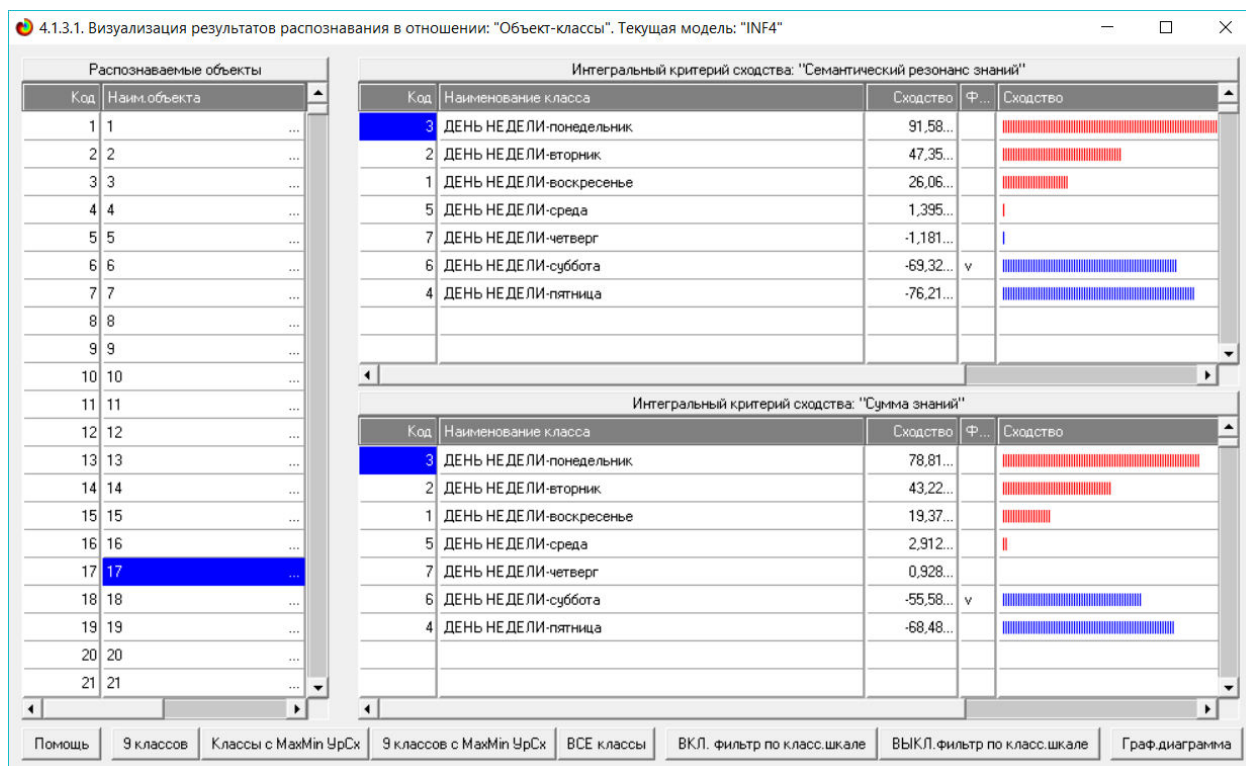


Рисунок 18. Пример идентификации классов в модели INF4

2.2. Когнитивные функции

Рассмотрим режим 4.5, в котором реализована возможность визуализации когнитивных функций для любых моделей и любых сочетаний классификационных и описательных шкал (рисунок 19)

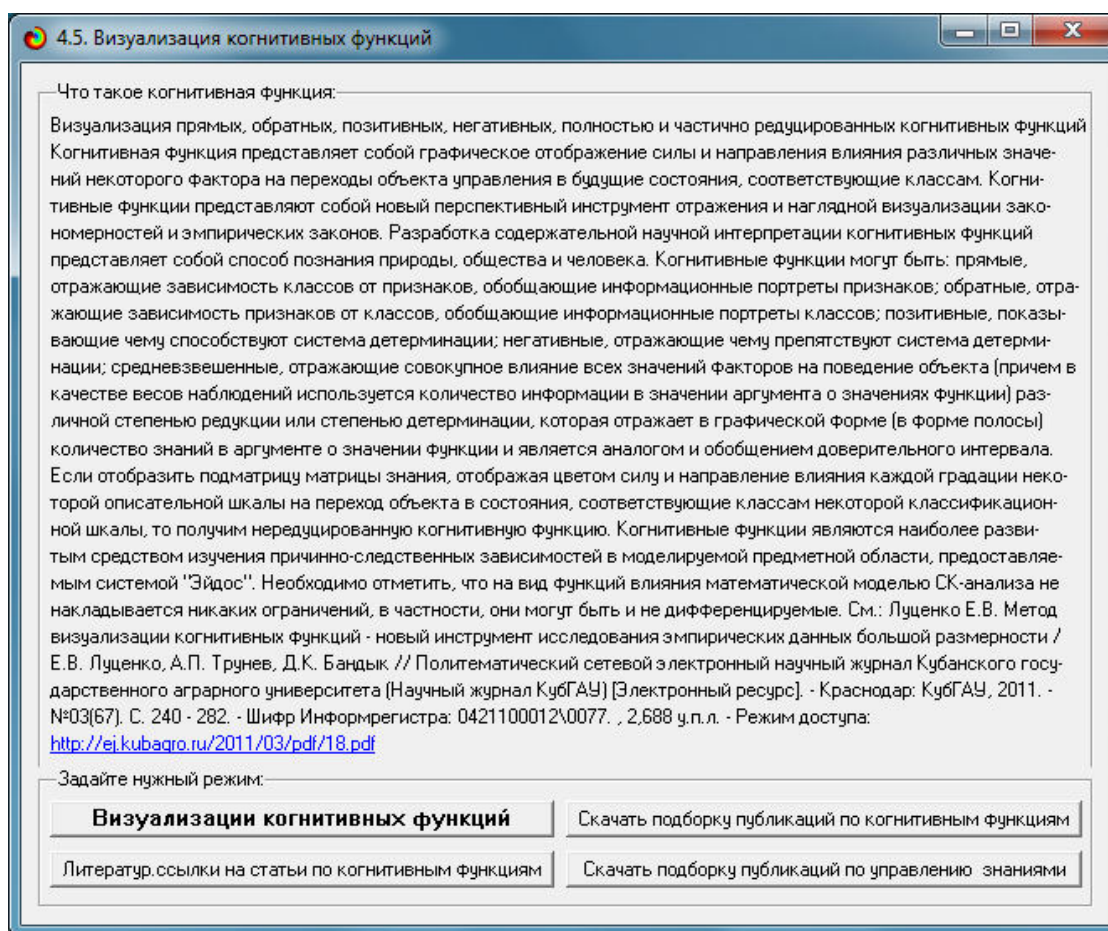


Рисунок 19. Экранная форма режима 4.5 системы «Эйдос-X++» «Визуализация когнитивных функций»

Применительно к задаче, рассматриваемой 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. Когнитивным функциям посвящено много работ автора 9, но наиболее новой и обобщающей из них является работа. Поэтому здесь не будем останавливаться на описании того, что представляют собой когнитивные функции в АСК-анализе. На рисунке 23. приведены визуализации всех когнитивных функций данного приложения для модели INF5.

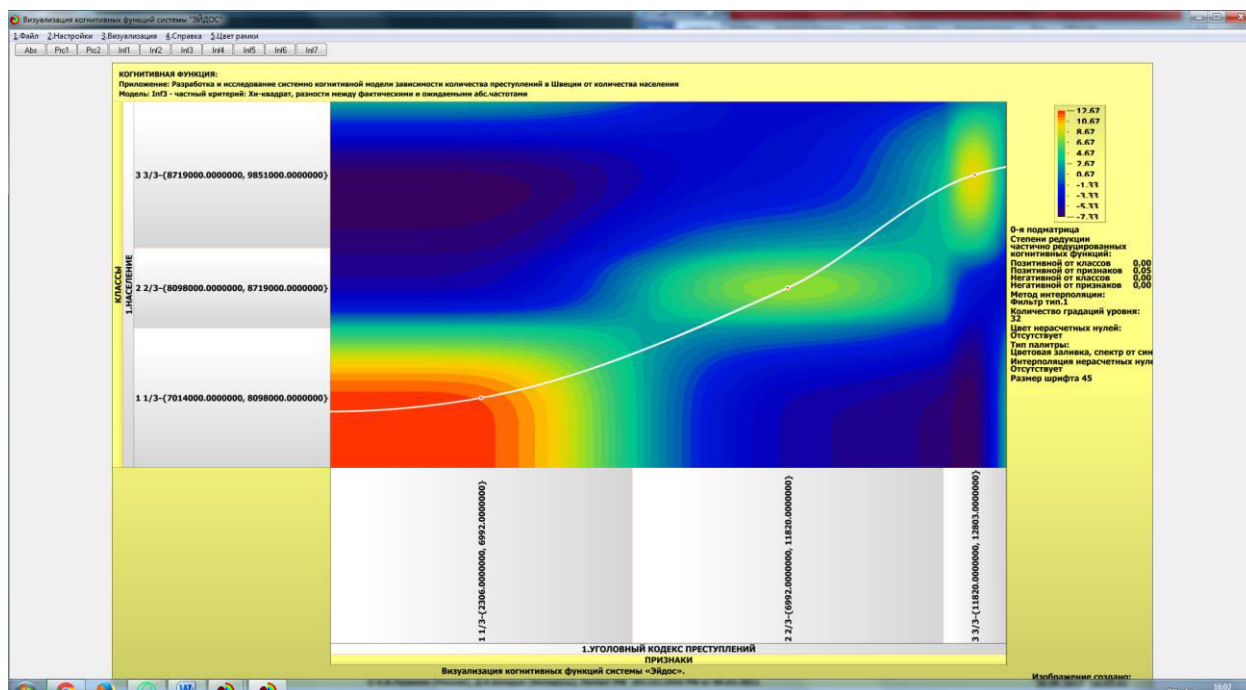


Рисунок 20. Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF3

2.3. 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. Подобная технология разработана давно, ей уже около 30 лет, но она малоизвестна – это интеллектуальная система «Эйдос». Данная система всегда обеспечивала возможность проведения количественного автоматизированного SWOT-анализа без использования экспертных оценок непосредственно на основе эмпирических данных. Результаты SWOT-анализа выводились в форме информационных портретов. В версии системы под MS Windows: «Эйдос-X++» предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (рисунок 21).

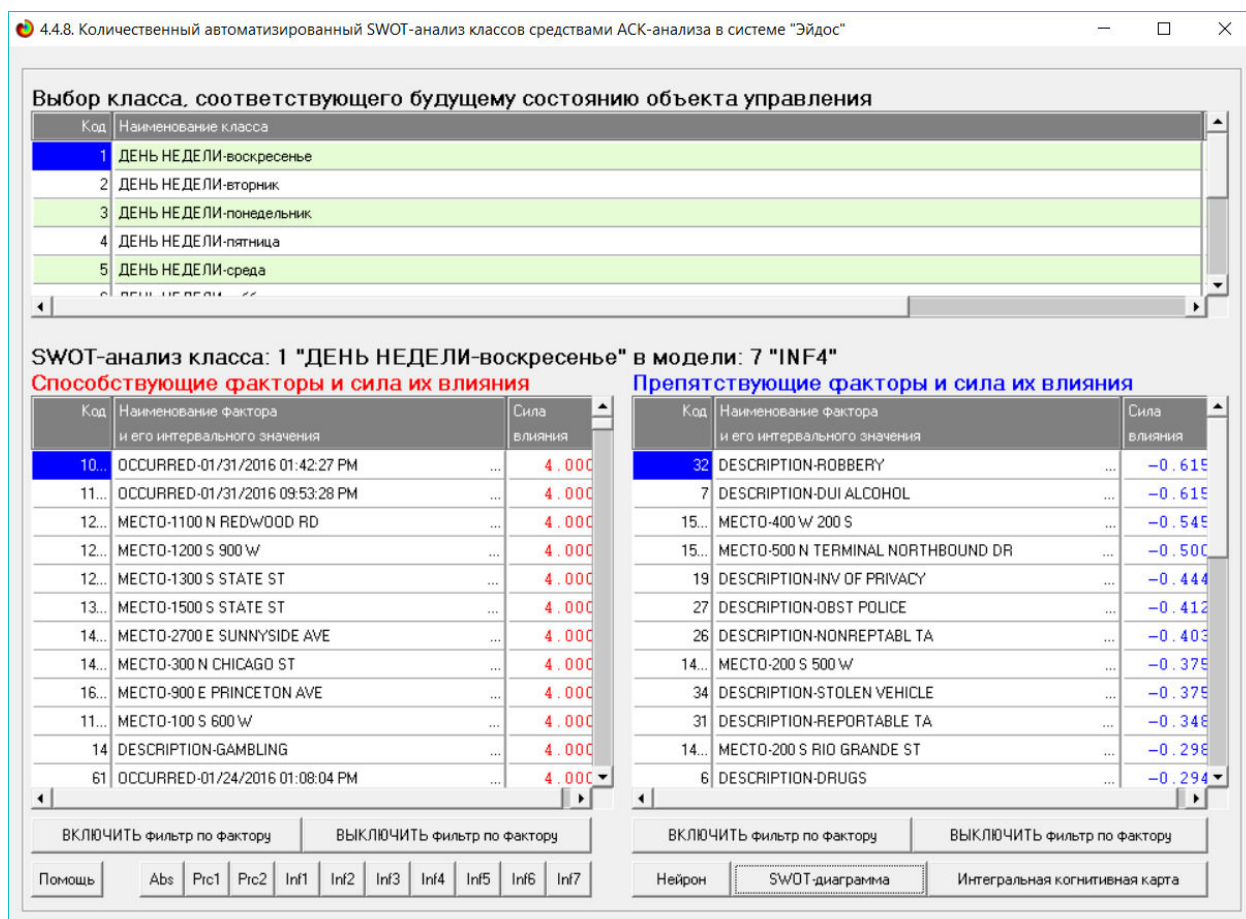


Рисунок 21. Пример SWOT-матрицы в модели INF4

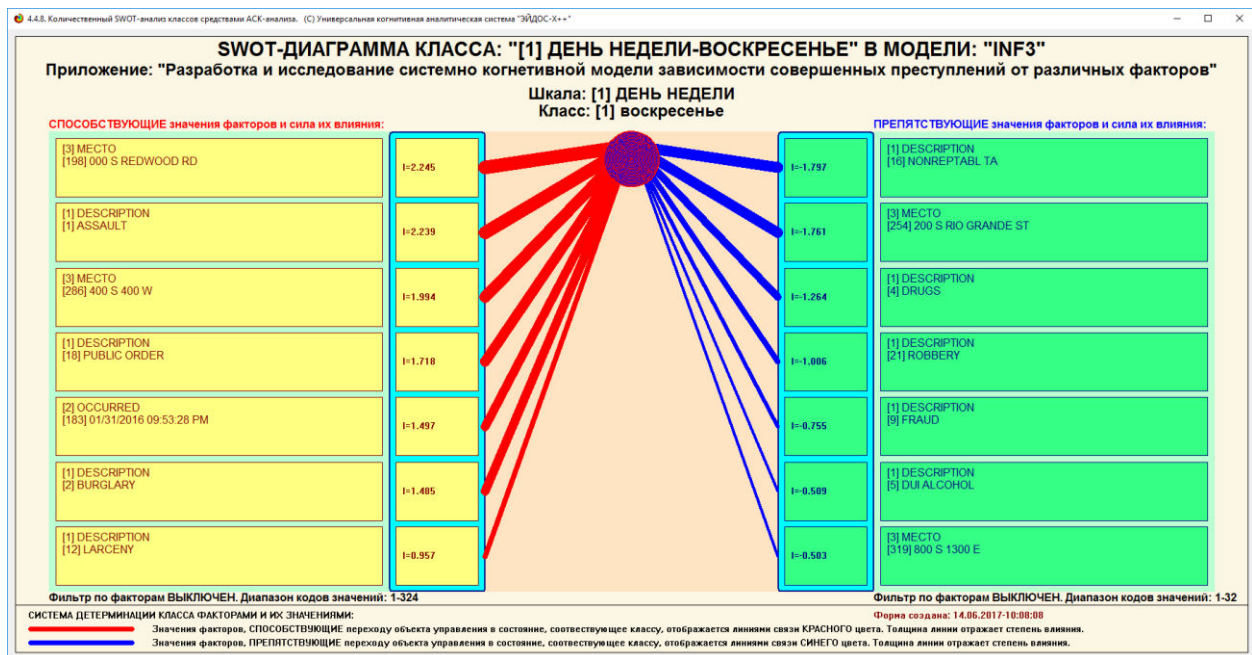


Рисунок 22 - Пример SWOT-матрицы в модели INF3

2.4. НЕЛОКАЛЬНЫЕ НЕЙРОННЫЕ СЕТИ И НЕЙРОНЫ.

Каждому классу системно-когнитивной модели соответствует не локальный нейрон совокупность, которых образует не локальную нейронную сеть [9] В качестве примеров приведены рисунок 23-24.

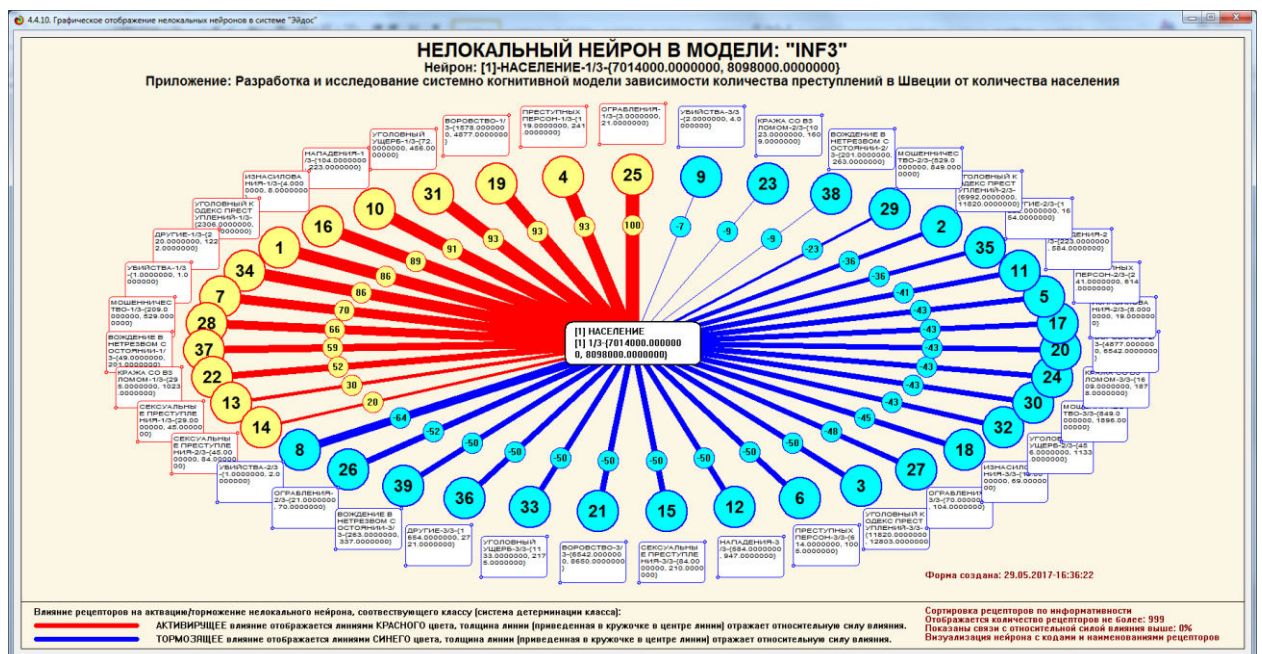


Рисунок 26. Пример нелокальные нейронные сети и нейроны модели INF3

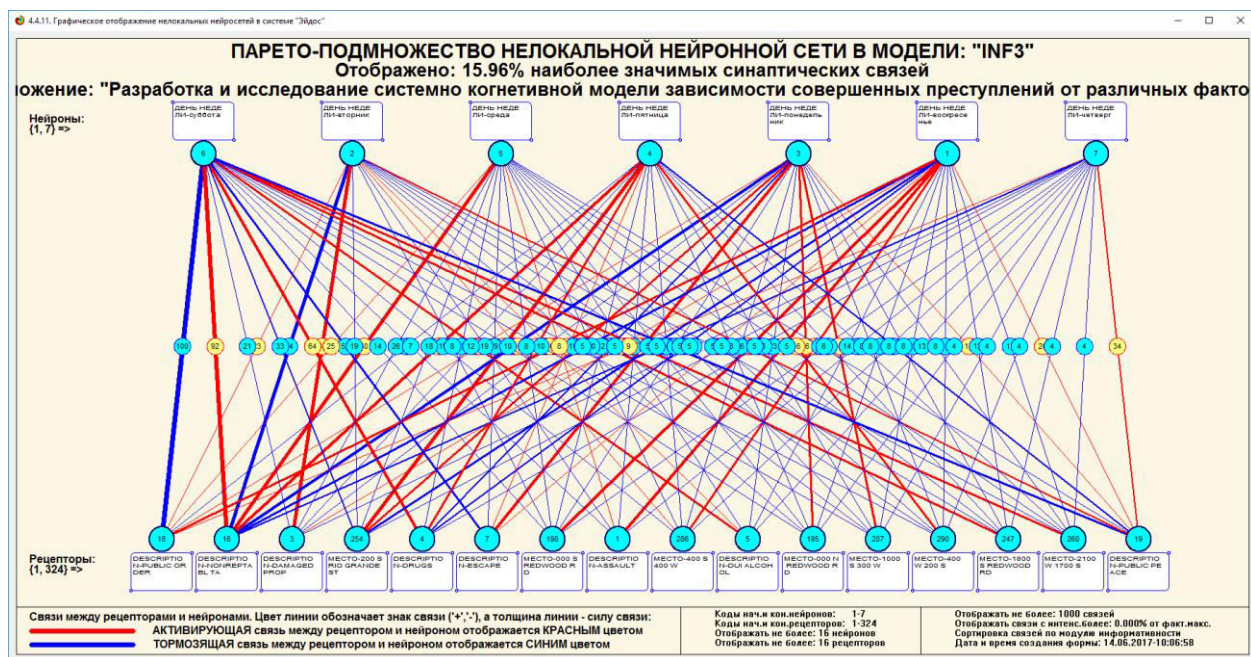


Рисунок 26. Пример нелокальные нейронные сети и нейроны модели INF3

2.5 КЛАСТЕРНЫЙ И КОНСТРУКТИВНЫЙ АНАЛИЗ

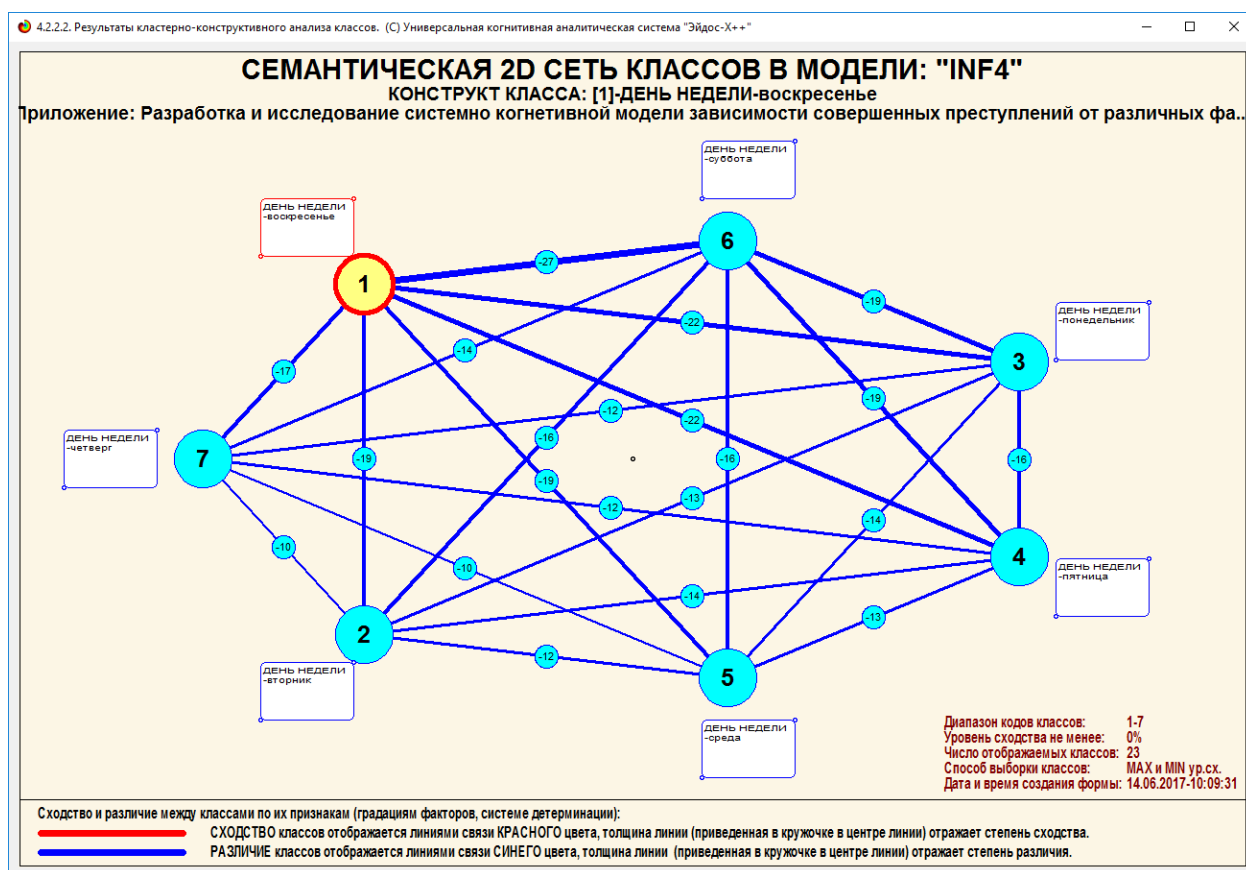


Рисунок 27. Пример кластерный и конструктивный анализ модели INF4

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 приводится развернутый пример использования базы данных Kaggle для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». При этом наиболее достоверной в данном приложении оказались модели INF4, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,971, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 91%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием. Также обращает на себя внимание, что статистические модели в данном приложении дают примерно на 21% более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, что, как правило, наблюдается и в других приложениях. Этим и оправдано применение моделей знаний.

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

СПИСОК ИСПОЛЬЗОВАННЫХ 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.

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

3. Луценко Е.В. 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.

4. 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. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: <http://lc.kubagro.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 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>