

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:

АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ ПРОЦЕНТОЙ ДОЛИ СЧАСТЛИВОГО НАСЕЛЕНИЯ НА ОСНОВЕ СТАТИСТИЧЕСКИХ ДАННЫХ

выполнил студент группы ИТ1421 Серeda Егор Алексеевич

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)

Середа Егору Алексеевичу
(Ф.И.О.)

Тема проекта: **Автоматизированный системно-когнитивный анализ
процентной доли счастливого населения на основе статистических
данных**

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 _____ 37 _____ листа формата
А4

б) графическая часть _____ 27 _____ лист формата А4

Рекомендуемая литература: Луценко Е.В. Универсальная когнитивная
аналитическая система «Эйдос- Х++» 2017 г.

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)

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	9
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	10
1.1. Описание решения.....	10
1.2. Преобразование исходных данных из HTML-формата в файл исходных данных MS Excel	10
1.3. Синтез и верификация статистических и интеллектуальных моделей .	16
1.4. Виды моделей системы «Эйдос».....	18
1.5. Результаты верификации моделей	19
2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	27
2.1. Решение задачи	27
2.2. Когнитивные функции	30
ЗАКЛЮЧЕНИЕ	40
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	41

РЕФЕРАТ

37 страниц, 28 рисунков, 4 таблицы, 8 литературных источников

ЭЙДОС X, СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА,
ГРАДАЦИИ, КЛАССИФИКАЦИОННЫЕ ШКАЛЫ, ОПИСАТЕЛЬНЫЕ
ШКАЛЫ

Целью работы является анализ чрезвычайных ситуации по их типам.

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) Середя Егор Алексеевича
курса 3 очной (заочной) формы обучения
направления подготовки Информационные системы и технологии
направленность(профиль)Интеллектуальные информационные системы и технологии.
Наименование темы «Автоматизированный системно-когнитивный анализ
процентной доли счастливого населения на основе статистических данных».
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 _____
«_____» _____ 201__ г.

Е.В. Луценко

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 идентификации вызовов на номер 911.

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 база данных тестовых задач для систем искусственного интеллекта репозитория UCI.

В данной курсовой работе использована база данных «911_1» из банка исходных данных по задачам искусственного интеллекта – репозитория UCI.

Для решения задачи используем стандартные возможности Microsoft Office Word и Excel, блокнот, 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 HTML-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 HTML-формата в файл исходных данных MS Excel

Из электронного ресурса баз данных allexcel возьмем базу данных сотрудников – «911_1.xls», которую оставим без изменений.

Общее описание задачи:

1. Lat-широта
2. Ing-долгота
3. Desc-период принятия вызова, станция и штат
4. Zip-индекс
5. timeStamp-дата и время вызова
6. twp-населенный пункт
7. addr-адрес
8. title-тип чрезвычайной ситуации

Столбцы 1-7 описательные шкалы.

Столбец 8 является классификационной шкалой. Этот столбец тип чрезвычайной ситуации.

Обучающая выборка:

Таблица 1 – 911_1.xls

nam	title	desc	zip	lat	lng	timeStamp	twp	addr
1	EMS: BACK PAINS/INJURY	REINDEER CT & D	19525	40,29788	-75,5812935	2015-12-10 17:10:52	NEW HANOVER	REINDEER CT & DEAD END
2	EMS: DIABETIC EMERGEN	BRIAR PATH & W	19446	40,25806	-75,2646799	2015-12-10 17:29:21	HATFIELD TOWNSH	BRIAR PATH & WHITEMARSH LN
3	Fire: GAS-ODOR/LEAK	HAWS AVE; NOR	19401	40,12118	-75,3519752	2015-12-10 14:39:21	NORRISTOWN	HAWS AVE
4	EMS: CARDIAC EMERGEN	AIRY ST & SWEDE	19401	40,11615	-75,343513	2015-12-10 16:47:36	NORRISTOWN	AIRY ST & SWEDE ST
5	EMS: DIZZINESS	CHERRYWOOD CT & DEA	40,25149	-75,6033497	2015-12-10 16:56:52	LOWER POTTS GROV	CHERRYWOOD CT & DEAD END	
6	EMS: HEAD INJURY	CANNON AVE & I	19446	40,25347	-75,283245	2015-12-10 15:39:04	LANSDALE	CANNON AVE & W 9TH ST
7	EMS: NAUSEA/VOMITIN	LAUREL AVE & O	19044	40,18211	-75,1277951	2015-12-10 16:46:48	HORSHAM	LAUREL AVE & OAKDALE AVE
8	EMS: RESPIRATORY EME	COLLEGEVILLE RD	19426	40,21729	-75,405182	2015-12-10 16:17:05	SKIPPAK	COLLEGEVILLE RD & LYWISKI RD
9	EMS: SYNCOPAL EPISOD	MAIN ST & OLD S	19438	40,28903	-75,3995896	2015-12-10 16:51:42	LOWER SALFORD	MAIN ST & OLD SUMNEYTOWN F
10	Traffic: VEHICLE ACCIDE	BLUEROUTE & R	19462	40,1024	-75,2914577	2015-12-10 17:35:41	PLYMOUTH	BLUEROUTE & RAMP I476 NB TO
11	Traffic: VEHICLE ACCIDE	RT202 PKWY & KNAPP R	40,23199	-75,2518915	2015-12-10 17:33:50	MONTGOMERY	RT202 PKWY & KNAPP RD	
12	Traffic: VEHICLE ACCIDE	BROOK RD & COL	19428	40,08416	-75,3083857	2015-12-10 16:32:10	PLYMOUTH	BROOK RD & COLWELL LN
13	Traffic: VEHICLE ACCIDE	BYBERRY AVE & S	19040	40,17413	-75,0984907	2015-12-10 17:15:49	UPPER MORELAND	BYBERRY AVE & S WARMINSTER
14	Traffic: VEHICLE ACCIDE	OLD YORK RD & V	19027	40,06297	-75,135914	2015-12-10 17:12:47	CHELTENHAM	OLD YORK RD & VALLEY RD
15	Traffic: VEHICLE ACCIDE	SCHUYLKILL EXPY & CRO	40,09722	-75,3761952	2015-12-10 17:09:49	UPPER MERION	SCHUYLKILL EXPY & CROTON RD	
16	Traffic: VEHICLE ACCIDE	STUMP RD & WIT	18936	40,22378	-75,2353993	2015-12-10 17:31:00	MONTGOMERY	STUMP RD & WITCHWOOD DR
17	EMS: RESPIRATORY EME	SUSQUEHANNA	19446	40,24326	-75,2865516	2015-12-10 17:42:44	LANSDALE	SUSQUEHANNA AVE & W MAIN
18	EMS: DIZZINESS	CHARLOTTE ST &	19525	40,31218	-75,5742598	2015-12-10 17:41:54	NEW HANOVER	CHARLOTTE ST & MILES RD
19	EMS: VEHICLE ACCIDENT	PENN ST & ARCH	19401	40,11424	-75,338508	2015-12-10 17:43:29	NORRISTOWN	PENN ST & ARCH ST
20	Traffic: DISABLED VEHIC	COUNTY LINE RD	18974	40,20934	-75,1352655	2015-12-10 17:45:23	HORSHAM	COUNTY LINE RD & WILLOW DR
21	Traffic: VEHICLE ACCIDE	PENN ST & ARCH	19401	40,11424	-75,338508	2015-12-10 17:43:45	NORRISTOWN	PENN ST & ARCH ST
22	Traffic: DISABLED VEHIC	CHURCH RD & RE	19031	40,11795	-75,2098476	2015-12-10 17:53:22	WHITEMARSH	CHURCH RD & REDCOAT DR
23	Fire: APPLIANCE FIRE	LILAC CT & PRIM	19446	40,19901	-75,3000584	2015-12-10 17:59:24	UPPER GWYNEDD	LILAC CT & PRIMROSE DR
24	Traffic: DISABLED VEHIC	RT422 & PAWLINGS RD	40,14333	-75,422819	2015-12-10 18:00:38	LOWER PROVIDENC	RT422 & PAWLINGS RD OVERPA	
25	Traffic: VEHICLE ACCIDE	SUMMIT AVE & RT309 U	40,15327	-75,1895576	2015-12-10 17:58:22	UPPER DUBLIN	SUMMIT AVE & RT309 UNDERPA	
26	EMS: GENERAL WEAKNE	SHANNONDELL C	19403	40,13304	-75,4084631	2015-12-10 18:02:02	LOWER PROVIDENC	SHANNONDELL DR & SHANNONI
27	EMS: HEAD INJURY	PENLLYN BLUE BE	19422	40,15528	-75,2642296	2015-12-10 18:02:38	WHITPAIN	PENLLYN BLUE BELL PIKE & VILLA
28	Fire: CARBON MONOXID	EDENTON PL & D	19085	40,0289	-75,3518224	2015-12-10 18:05:19	DELAWARE COUNT	EDENTON PL & DURHAM DR
29	Traffic: VEHICLE ACCIDE	SCHUYLKILL EXPY & WE	40,09722	-75,3761952	2015-12-10 18:05:39	UPPER MERION	SCHUYLKILL EXPY & WEADLEY RD	
30	Traffic: DISABLED VEHIC	COUNTY LINE RD	18974	40,20934	-75,1352655	2015-12-10 18:10:40	HORSHAM	COUNTY LINE RD & WILLOW DR

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 html-файла в xls-файл, для чего выполним следующие операции.

Скопируем получившуюся таблицу из MS Word в MS Excel и запишем ее с именем: Inp_data.xls в папку: c:\Aidos-X\AID_DATA\Inp_data\. В файле Inp_data.xls добавим пустую колонку на позиции «A» и автоматически пронумеруем все строки. 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.

The screenshot displays a software interface for configuring data import. It is organized into several sections with radio buttons, checkboxes, and input fields.

- Задайте тип файла исходных данных: "Inp_data":**
 - ☒ XLS - MS Excel-2003 (with button: Стандарт XLS-файла)
 - ☐ XLSX- MS Excel-2007(2010)
 - ☐ DBF - DBASE IV (DBF/NTX) (with button: Стандарт DBF-файла)
 - ☐ CSV - Comma-Separated Values (with button: 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: 2
 - Конечный столбец классификационных шкал: 2
- Задайте диапазон столбцов описательных шкал:**
 - Начальный столбец описательных шкал: 3
 - Конечный столбец описательных шкал: 9
- 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

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

В экранной форме, приведенной на рисунке 1, 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(второй столбец в таблице);

– "Задайте диапазон столбцов описательных шкал": "Начальный столбец описательных шкал" – 3, "Конечный столбец описательных шкал" – 9;

– "Задание параметров формирования сценариев или способа интерпретации текстовых полей": "Не применять сценарный метод АСК-анализа и спец.интерпретацию ТХТ-полей".

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".

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	1	6	6,00	5	75	15,00
Текстовые	0	0	0,00	3	19	6,33
ВСЕГО:	1	6	6,00	8	94	11,75

Задайте число интервалов (градаций) в шкале:

В классификационных шкалах: В описательных шкалах:

Пересчитать шкалы и градации Параметры числ. шкал и градаций 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".

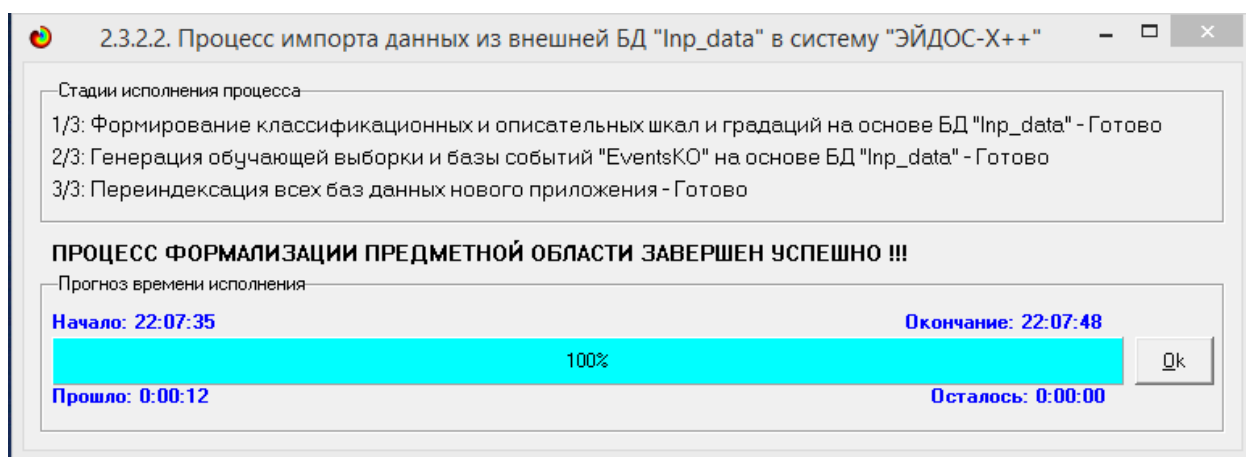


Рисунок 3. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

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 4).

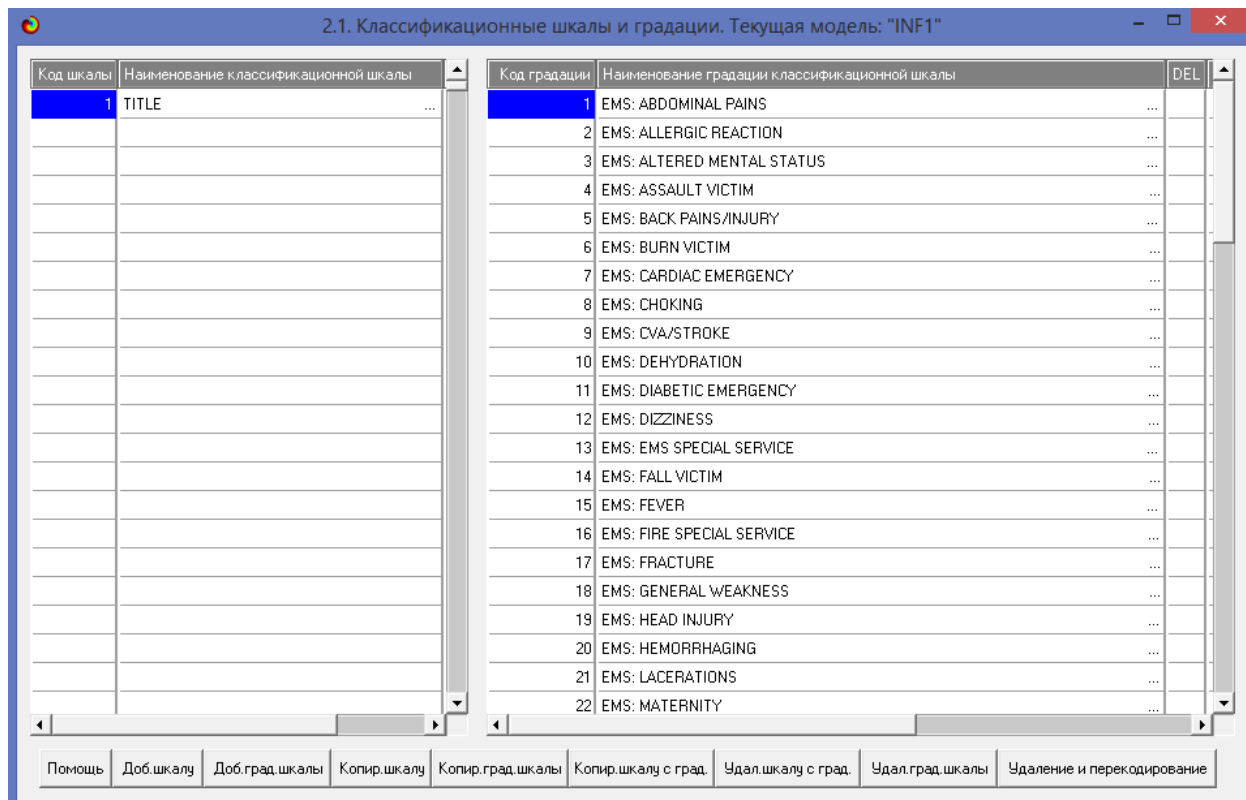


Рисунок 4. Классификационные шкалы и градации (фрагмент)

[illegible]

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	5	0	0	0
2	11	0	0	0
3	45	0	0	0
4	7	0	0	0
5	12	0	0	0
6	19	0	0	0
7	24	0	0	0
8	27	0	0	0
9	30	0	0	0
10	59	0	0	0

Код объекта	Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
1	74	81	784	1094	2070	2661	0
2	73	88	189	1097	2056	2180	0
3	68	87	481	1085	2071	2417	0
4	68	87	116	1090	2071	2119	0
5	72	81	250	1092	2065	2230	0
6	72	88	224	1086	2058	2209	0
7	70	92	565	1089	2057	2494	0
8	71	85	266	1087	2081	2245	0
9	74	85	619	1091	2067	2533	0
10	67	88	183	1100	2074	2174	0

Помощь

Скопировать 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 [8]).

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).

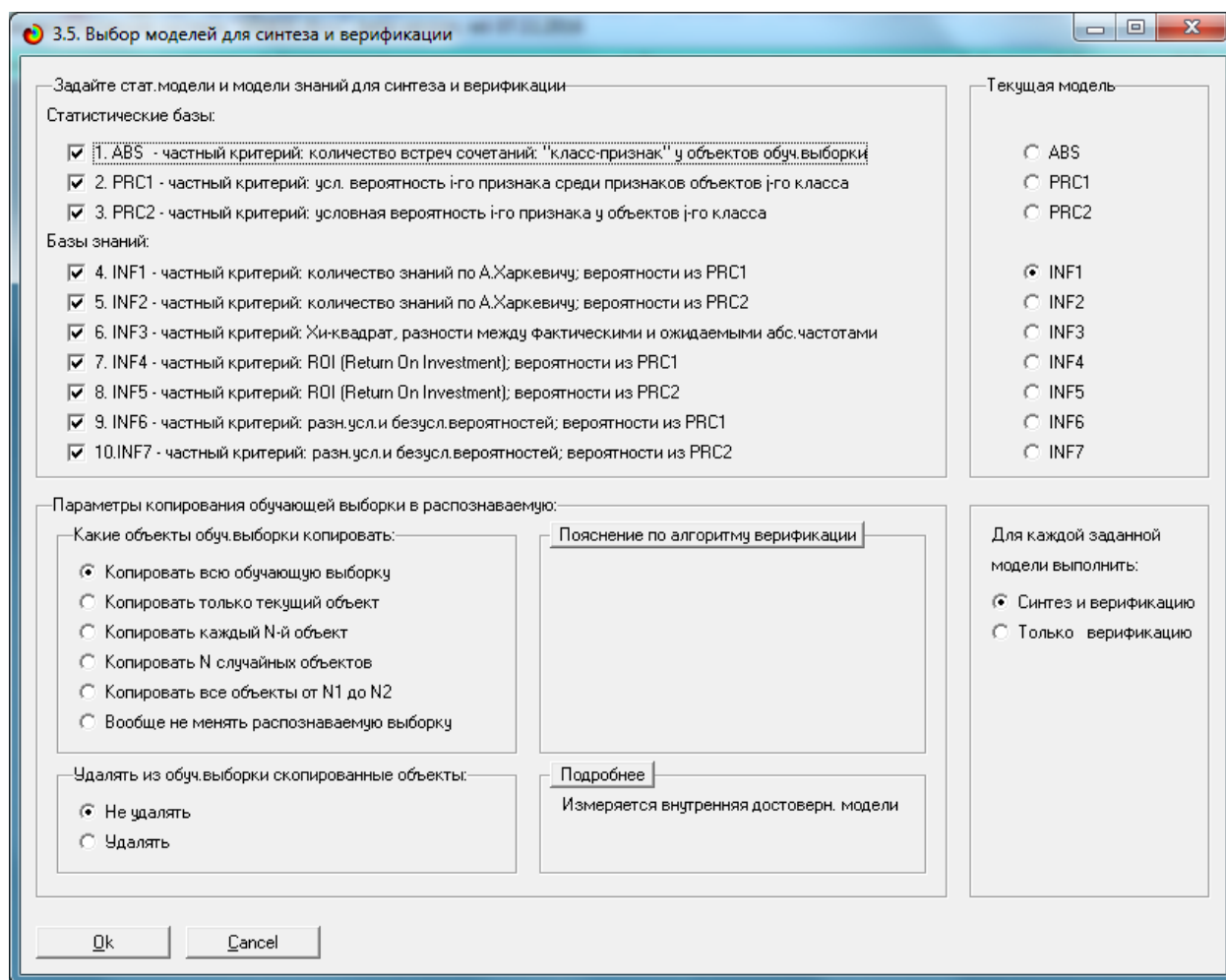


Рисунок 7. Выбор моделей для синтеза и верификации, а также текущей модели

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 8.

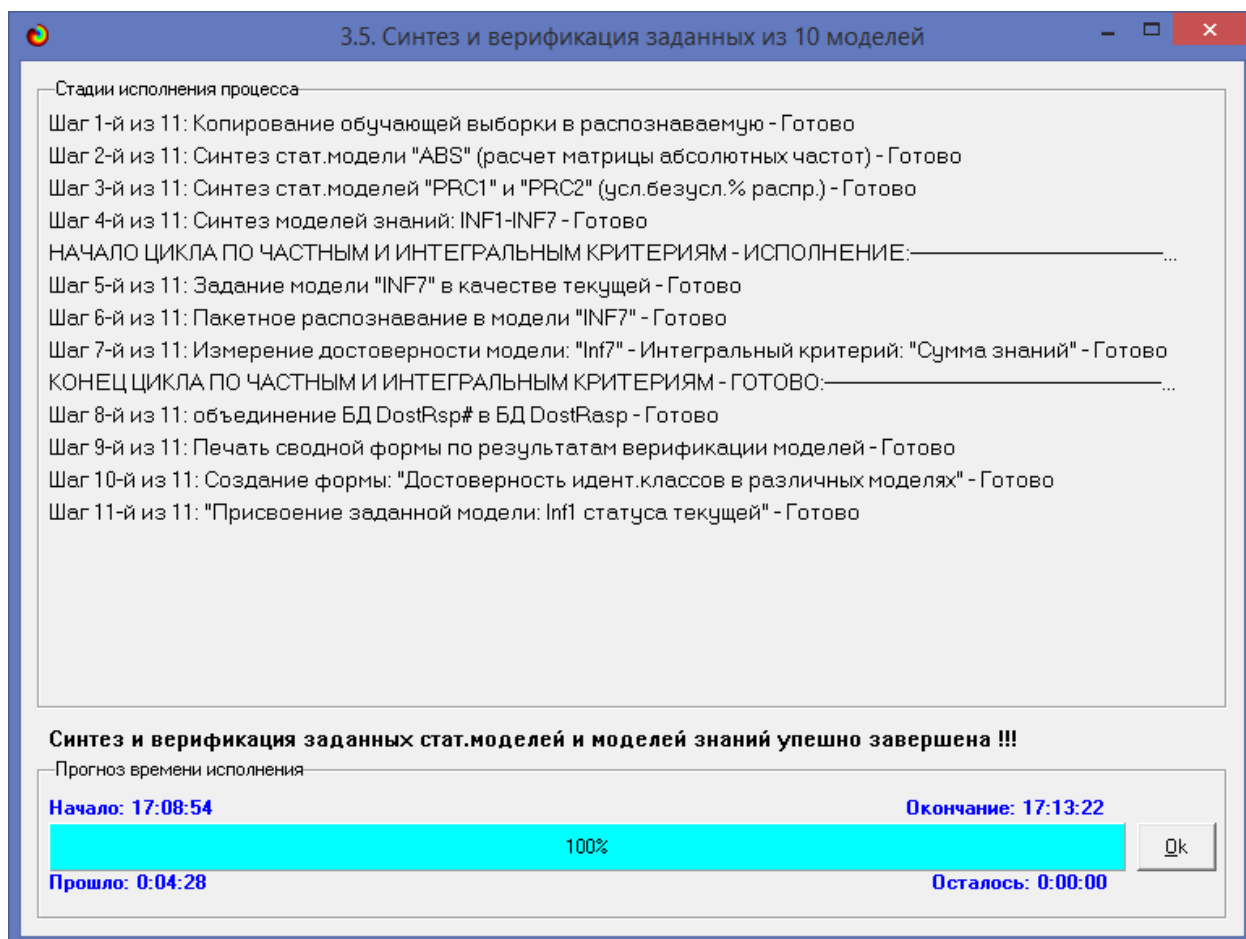


Рисунок 8. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 42 минут, так как обрабатывается большой объем данных (500 строк). При этом верификация (оценка достоверности моделей) проводилась на всех 131 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 INF1, в которой рассчитано количество информации по А.Харкевичу, которое мы получаем о принадлежности идентифицируемого объекта к каждому из классов, если знаем, что у этого объекта есть некоторый признак.

По сути, частные критерии представляют собой просто формулы для преобразования матрицы абсолютных частот (таблица 2) в матрицы условных и безусловных процентных распределений, и матрицы знаний (таблицы 3 и 4) (проф. В.И.Лойко).

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

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обучающей выборки"															
Код признака	Наименование описательной шкалы и градации	1. LAT 1/6 (39.9866944, 40.0612338)	2. LAT 2/6 (40.0612338, 40.1357732)	3. LAT 3/6 (40.1357732, 40.2103127)	4. LAT 4/6 (40.2103127, 40.2848521)	5. LAT 5/6 (40.2848521, 40.3593915)	6. LAT 6/6 (40.3593915, 40.4339309)	7. LNG 1/6 { 75.6744381, 75.5689601}	8. LNG 2/6 { 75.5689601, 75.4634822}	9. LNG 3/6 { 75.4634822, 75.3580042}	10. LNG 4/6 { 75.3580042, 75.2525262}	11. LNG 5/6 { 75.2525262, 75.1470483}	12. LNG 6/6 { 75.1470483, 75.0415703}	Сумма	
1	TIMESTAMP-2015-12-10 14:39:21	...	1								1			2	
2	TIMESTAMP-2015-12-10 15:39:04	...			1						1			2	
3	TIMESTAMP-2015-12-10 16:17:05	...			1					1				2	
4	TIMESTAMP-2015-12-10 16:32:10	...	1								1			2	
5	TIMESTAMP-2015-12-10 16:46:48	...		1									1	2	
6	TIMESTAMP-2015-12-10 16:47:36	...	1								1			2	
7	TIMESTAMP-2015-12-10 16:51:42	...				1				1				2	
8	TIMESTAMP-2015-12-10 16:56:52	...			1			1						2	
9	TIMESTAMP-2015-12-10 17:09:49	...	1							1				2	
10	TIMESTAMP-2015-12-10 17:10:52	...				1		1						2	
11	TIMESTAMP-2015-12-10 17:12:47	...	1										1	2	
12	TIMESTAMP-2015-12-10 17:15:49	...		1									1	2	
13	TIMESTAMP-2015-12-10 17:29:21	...			1						1			2	
14	TIMESTAMP-2015-12-10 17:31:00	...			1							1		2	
15	TIMESTAMP-2015-12-10 17:33:50	...			1							1		2	
16	TIMESTAMP-2015-12-10 17:35:41	...	1							1				2	
17	TIMESTAMP-2015-12-10 17:41:54	...				1		1						2	
18	TIMESTAMP-2015-12-10 17:42:44	...			1						1			2	
19	TIMESTAMP-2015-12-10 17:43:29	...	1								1			2	
20	TIMESTAMP-2015-12-10 17:43:45	...	1								1			2	
21	TIMESTAMP-2015-12-10 17:45:23	...		1									1	2	
22	TIMESTAMP-2015-12-10 17:53:22	...	1									1		2	
23	TIMESTAMP-2015-12-10 17:58:22	...		1								1		2	

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

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"														
Код признака	Наименование описательной шкалы и градации	1. TITLE EMS: ABDOMINAL PAINS	2. TITLE EMS: ALTERED MENTAL STATUS	3. TITLE EMS: ASSAULT VICTIM	4. TITLE EMS: BACK PAINS/INJURY	5. TITLE EMS: CARDIAC EMERGENCY	6. TITLE EMS: CVA/STROKE	7. TITLE EMS: DIABETIC EMERGENCY	8. TITLE EMS: DIZZINESS	9. TITLE EMS: FALL VICTIM	10. TITLE EMS: GENERAL WEAKNESS	11. TITLE EMS: HEAD INJURY	12. TITLE EMS: HEMORRHAGE	13. TITLE EMS: MEDICAL ALERT ALARM
44	DESC-4TH ST & DEPOT ST; BRIDGEPORT; Stati...			3.572										
45	DESC-AIRY ST & SWEDE ST; NORRISTOWN; St...					2.682								
46	DESC-ALLENDALE RD & WILLS BLVD; UPPER M...													
47	DESC-AUTO PARK BLVD & LINFIELD TRAPPE R...													
48	DESC-AVIGNON DR & DEAD END; LOWER MERI...													
49	DESC-BELMONT AVE & E CITY AVE; LOWER ME...													
50	DESC-BLOMMER DR & GRAVEL PIKE; UPPER H...													
51	DESC-BLUEROUTE & RAMP 1476 NB TO CHEMI...													
52	DESC-BRIAR PATH & WHITEMARSH LN; HATFI...							3.037						
53	DESC-BROAD ST & E 6TH ST; LANSDALE; Stati...									2.330				
54	DESC-BROOK RD & COLWELL LN; PLYMOUTH; ...													
55	DESC-BUSTARD RD & KRIEBEL RD; TOWAMEN...													
56	DESC-BUSTARD RD & OLD MORRIS RD; TOWA...													
57	DESC-BUTLER AVE & N RIDGE AVE; AMBLER; ...													
58	DESC-BYBERRY AVE & S WARMINSTER RD; UP...													
59	DESC-CANNON AVE & W 9TH ST; LANSDALE; S...											2.724		
60	DESC-CEGAR LN & WILLOW LN; NEW HANOVE...													
61	DESC-CEGAR ST & E HANCOCK ST; LANSDALE; ...													
62	DESC-CHARLOTTE ST & MILES RD; NEW HAN...							2.975						
63	DESC-CHELTENHAM AVE & DELPHINE RD; SPR...													
64	DESC-CHERRYWOOD CT & DEAD END; LOWE...							2.975						

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

5.5. Модель: "7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1"														
Код признака	Наименование описательной шкалы и градации	1. TITLE EMS: ABDOMINAL PAINS	2. TITLE EMS: ALTERED MENTAL STATUS	3. TITLE EMS: ASSAULT VICTIM	4. TITLE EMS: BACK PAINS/INJURY	5. TITLE EMS: CARDIAC EMERGENCY	6. TITLE EMS: CVA/STROKE	7. TITLE EMS: DIABETIC EMERGENCY	8. TITLE EMS: DIZZINESS	9. TITLE EMS: FALL VICTIM	10. TITLE EMS: GENERAL WEAKNESS	11. TITLE EMS: HEAD INJURY	12. TITLE EMS: HEMORRHAGE	13. TITLE EMS: MEDICAL ALERT ALARM
1	LAT-Малое: 1/3-(39.9867698, 40.1292628) ...	1.439		1.439		1.439		0.220		-0.024			1.439	
2	LAT-Среднее: 2/3-(40.1292628, 40.2717558) ...		1.083				0.389	0.042	0.042	-0.167	1.083	0.389		1.083
3	LAT-Большое: 3/3-(40.2717558, 40.4142488) ...				8.091		2.030		3.545	0.818		2.030		
4	LNG-Малое: 1/3-(75.6545599, -75.4689427) ...				4.263		0.754		4.263	0.053				4.263
5	LNG-Среднее: 2/3-(75.4689427, -75.2833256) ...		1.273	1.273		0.515	0.515			0.364	1.273	-0.242		
6	LNG-Большое: 3/3-(75.2833256, -75.0977084) ...	1.703				-0.099		1.703		-0.459		0.802	1.703	
7	DESC-4TH ST & DEPOT ST; BRIDGEPORT; Stati...			99.000										
8	DESC-AIRY ST & SWEDE ST; NORRISTOWN; St...					32.333								
9	DESC-ALLENDALE RD & WILLS BLVD; UPPER M...													
10	DESC-AUTO PARK BLVD & LINFIELD TRAPPE R...													
11	DESC-AVIGNON DR & DEAD END; LOWER MERI...													
12	DESC-BELMONT AVE & E CITY AVE; LOWER ME...													
13	DESC-BLOMMER DR & GRAVEL PIKE; UPPER H...													
14	DESC-BLUEROUTE & RAMP 1476 NB TO CHEMI...													
15	DESC-BRIAR PATH & WHITEMARSH LN; HATFI...							49.000						
16	DESC-BROAD ST & E 6TH ST; LANSDALE; Stati...									19.000				
17	DESC-BROOK RD & COLWELL LN; PLYMOUTH; ...													
18	DESC-BUSTARD RD & KRIEBEL RD; TOWAMEN...													
19	DESC-BUSTARD RD & OLD MORRIS RD; TOWA...													
20	DESC-BUTLER AVE & N RIDGE AVE; AMBLER; ...													
21	DESC-BYBERRY AVE & S WARMINSTER RD; UP...													

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 достов. моделям при разн. интегр. крит. Текущая модель: "INF4"										
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 Ризбергера	Сумма уровней с истинно-положительными решениями
1. ABS - частный критерий: количество встреч сочетаний "клас...	Корреляция абс. частот с обр...	100	100	687	2313		0.041	1.000	0.080	60.8
1. ABS - частный критерий: количество встреч сочетаний "клас...	Сумма абс. частот по признак...	100	100	680	2320		0.041	1.000	0.079	37.8
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл. отн. частот с о...	100	100	687	2313		0.041	1.000	0.080	60.8
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл. отн. частот по признак...	100	100	680	2320		0.041	1.000	0.079	40.9
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл. отн. частот с о...	100	100	687	2313		0.041	1.000	0.080	60.8
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл. отн. частот по признак...	100	100	680	2320		0.041	1.000	0.079	35.2
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	100	100	1731	1269		0.073	1.000	0.136	42.2
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	100	100	1082	1918		0.050	1.000	0.094	54.9
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	100	100	1735	1265		0.073	1.000	0.137	42.2
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	100	100	1040	1960		0.049	1.000	0.093	54.8
6. INF3 - частный критерий: Хинквадрат, разности между факти...	Семантический резонанс зна...	100	100	1730	1270		0.073	1.000	0.136	49.2
6. INF3 - частный критерий: Хинквадрат, разности между факти...	Сумма знаний	100	100	1730	1270		0.073	1.000	0.136	54.9
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	100	100	2853	147		0.405	1.000	0.576	43.8
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	100	100	1032	1968		0.048	1.000	0.092	24.2
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	100	100	2853	147		0.405	1.000	0.576	43.8
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	100	100	1014	1986		0.048	1.000	0.091	25.2
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	100	98	1326	1674	2	0.055	0.980	0.105	48.9
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	100	100	1092	1908		0.050	1.000	0.095	28.4
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	100	98	1311	1689	2	0.055	0.980	0.104	49.0
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	100	100	1056	1944		0.049	1.000	0.093	27.8

Рисунок 9. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели INF4 и INF5 при интегральном критерии «Семантический резонанс знаний». При этом точность модели (F-мера Ван Ризбергера) составляет 0,576, а точность модели (L1-мера профессора Луценко 0,994). L1-мера профессора Луценко является более достоверной, по сравнению с F-мерой Ван Ризбергера. Таким образом, уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 100%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется F-мера Ван Ризбергера и L-мера, представляющая собой ее нечеткое мультиклассовое обобщение, предложенное профессором Е.В.Луценко [7] (рисунок 10).

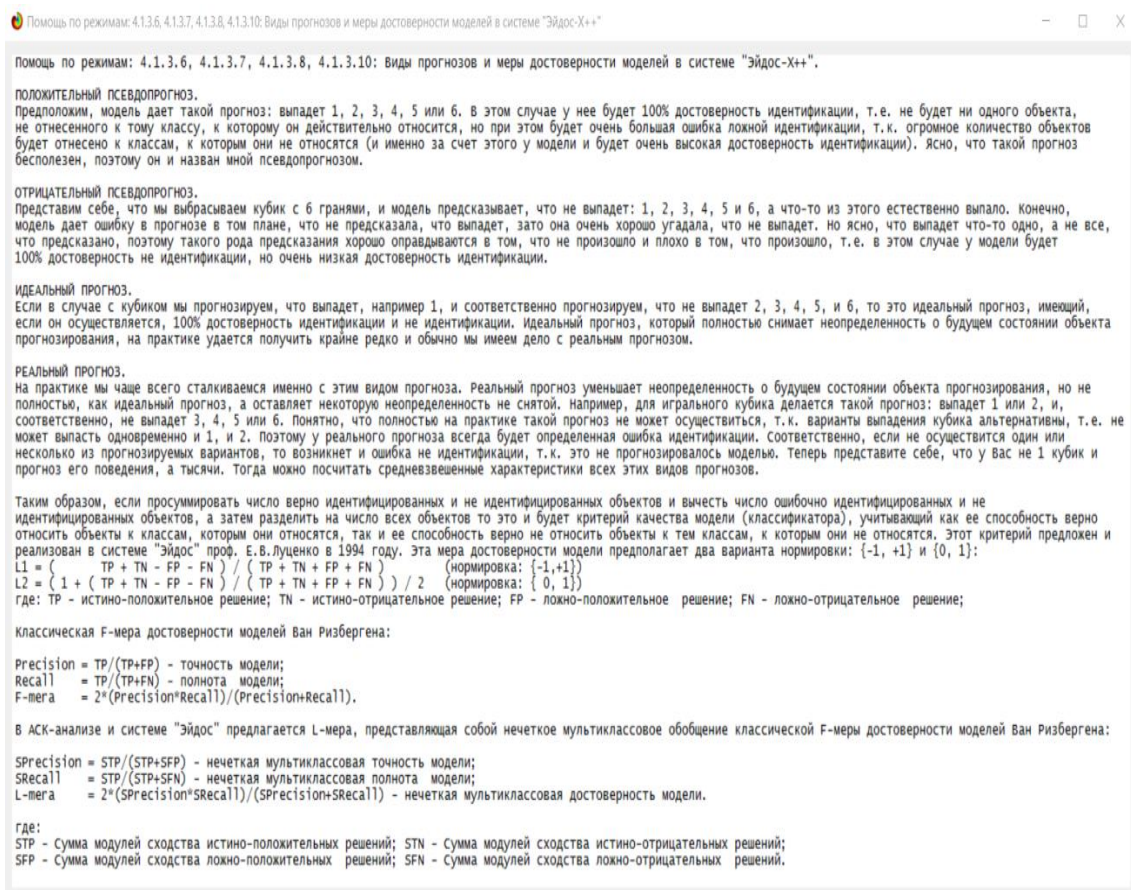


Рисунок 10. Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

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.

4.1.3.11. Част.распр.уровн.сход.ТР,TN,FP,FN решений при разных моделях и инт.критериях. Текущая модель: "INF4"													
Наименование частного и интегрального критерия	-100	-99	-98	-97	-96	-95	-94	-93	-92	-91	-90	-89	-88
1. АБС - ЧАСТНЫЙ КРИТЕРИЙ: КОЛИЧЕСТВО ВСТРЕЧ СОБЪЕКТОВ "ЧЛАСД.ПРИЗНАК" У ОБ...													
Интегральный критерий: КОРРЕЛЯЦИЯ АБС.ЧАСТОТ С ОБР.ОБЪЕКТА													
Уровни сходства (Ур.Сх.) (%):	-100	-99	-98	-97	-96	-95	-94	-93	-92	-91	-90	-89	-88
Част.распр.Уровней Сходства истинных решений (TP+TN)													
Част.распр.Уровней Сходства ложных решений (FP+FN)													
Част.распр.Уровней Сходства истинно-положительных решений (TP)													
Част.распр.Уровней Сходства истинно-отрицательных решений (TN)													
Част.распр.Уровней Сходства ложно-положительных решений (FP)													
Част.распр.Уровней Сходства ложно-отрицательных решений (FN)													
Интегральный критерий: СУММА АБС.ЧАСТОТ ПО ПРИЗНАКАМ ОБЪЕКТА													
Уровни сходства (Ур.Сх.) (%):	-100	-99	-98	-97	-96	-95	-94	-93	-92	-91	-90	-89	-88
Част.распр.Уровней Сходства истинных решений (TP+TN)													
Част.распр.Уровней Сходства ложных решений (FP+FN)													
Част.распр.Уровней Сходства истинно-положительных решений (TP)													
Част.распр.Уровней Сходства истинно-отрицательных решений (TN)													
Част.распр.Уровней Сходства ложно-положительных решений (FP)													
Част.распр.Уровней Сходства ложно-отрицательных решений (FN)													
2. РАС1 - ЧАСТНЫЙ КРИТЕРИЙ: УСЛ. ВЕРОЯТНОСТЬ I-ГО ПРИЗНАКА СРЕДИ ПРИЗНАКОВ ОБ...													
Интегральный критерий: КОРРЕЛЯЦИЯ УСЛ.ОТН.ЧАСТОТ С ОБР.ОБЪЕКТА													
Уровни сходства (Ур.Сх.) (%):	-100	-99	-98	-97	-96	-95	-94	-93	-92	-91	-90	-89	-88
Част.распр.Уровней Сходства истинных решений (TP+TN)													
Част.распр.Уровней Сходства ложных решений (FP+FN)													
Част.распр.Уровней Сходства истинно-положительных решений (TP)													

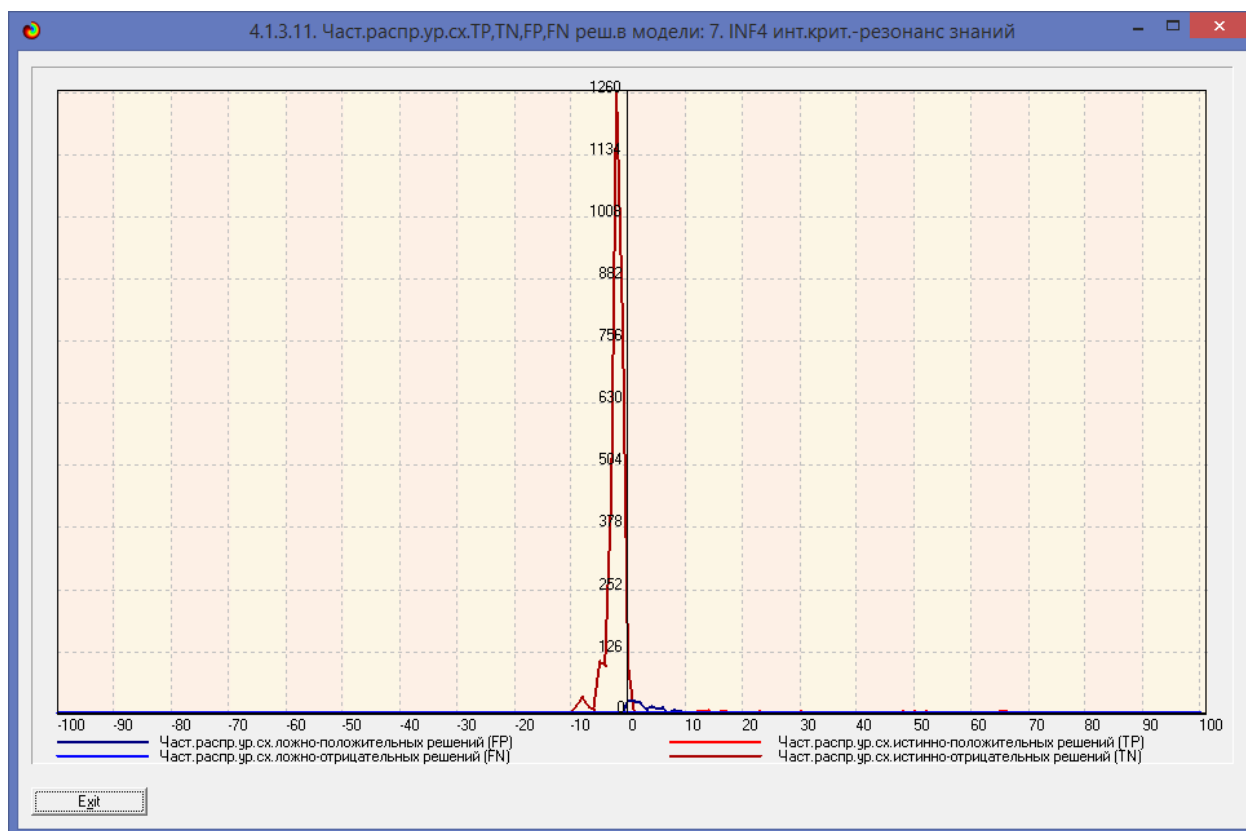


Рисунок 11. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF4

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 INF3 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 очень больших значений уровней сходства-различия (более 90%) также различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями практически отсутствует[1].

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):

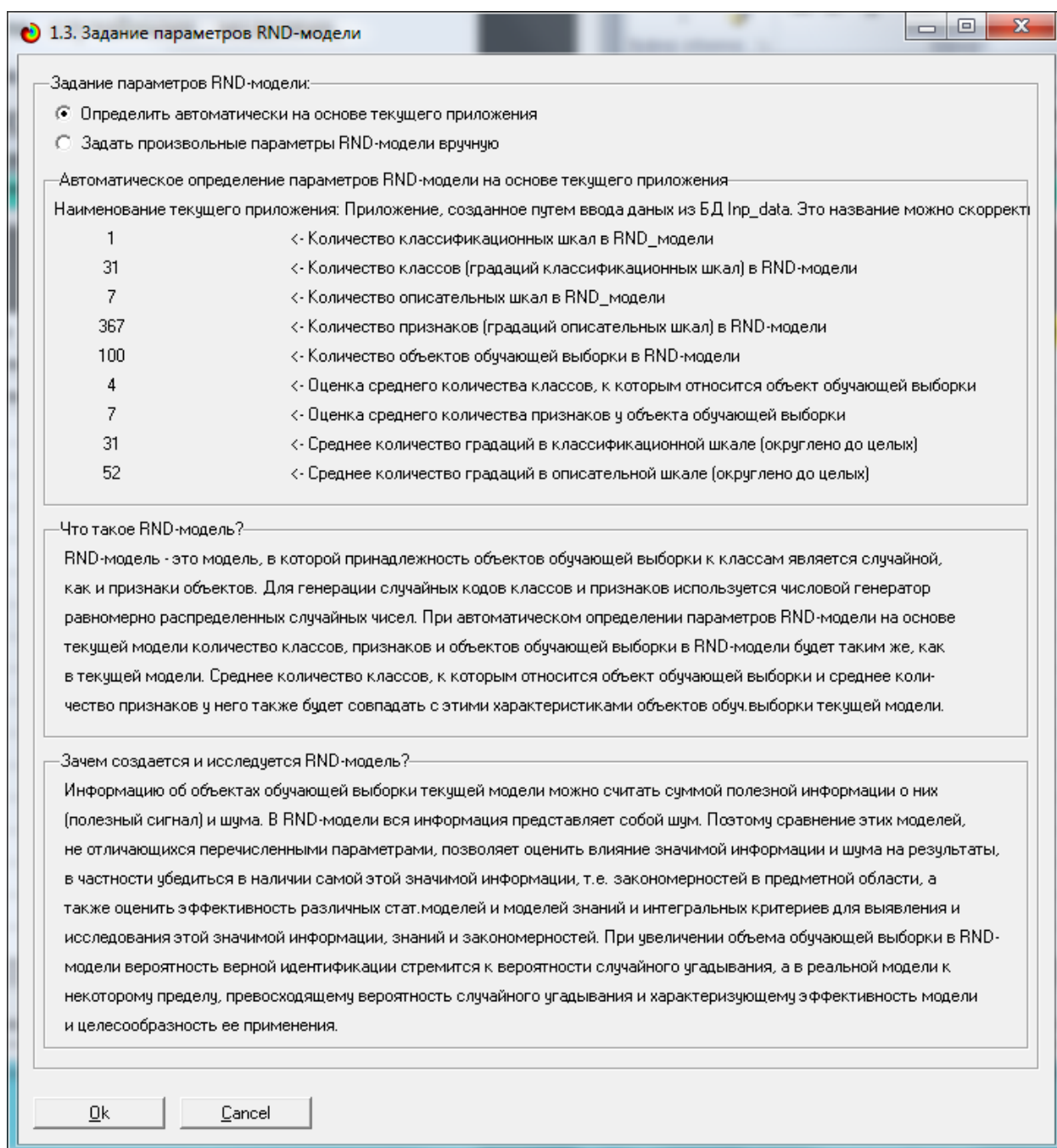
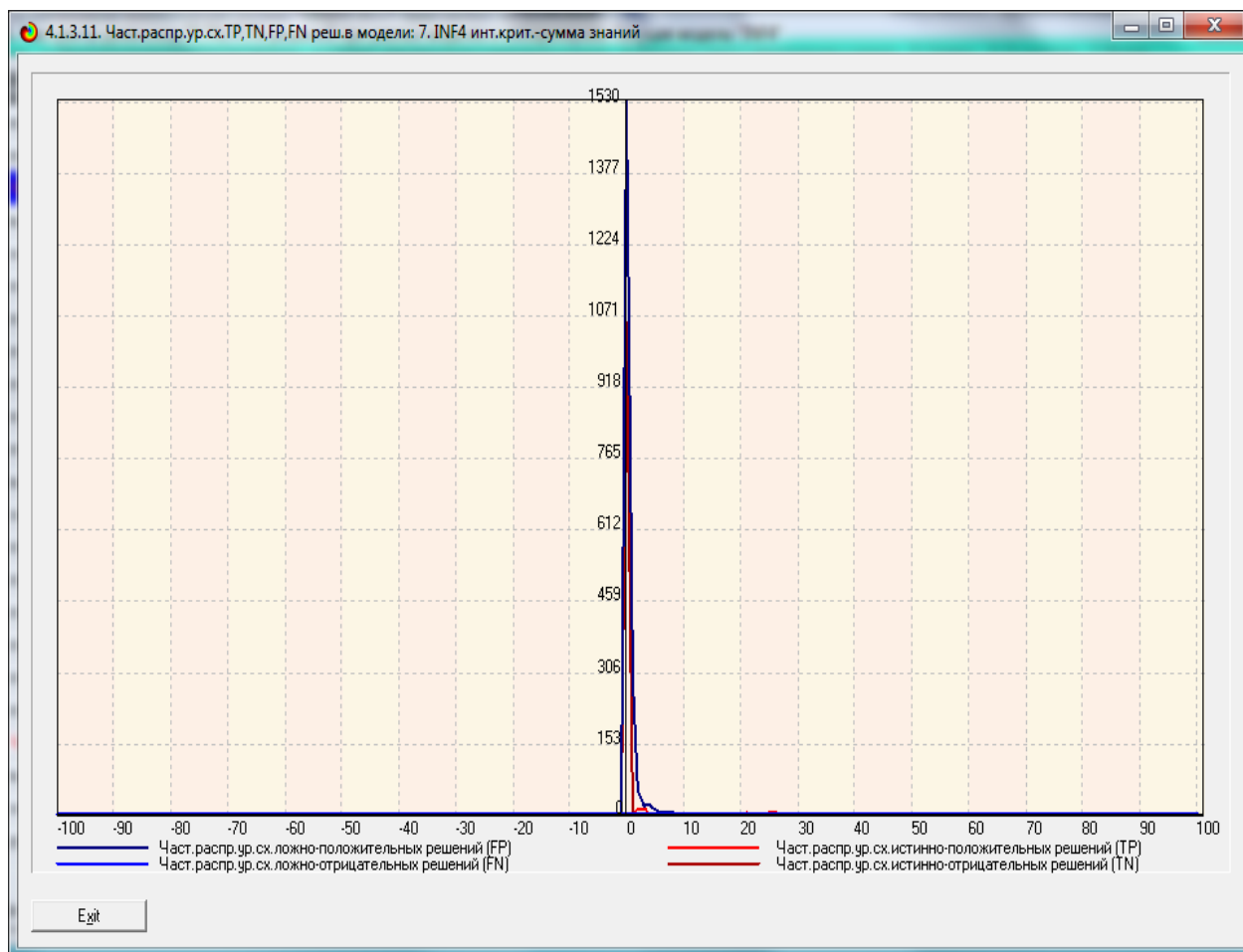


Рисунок 12. Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

На рисунке 13 показано частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF4



На рисунке 13 приведены данные по достоверности статистических и когнитивных моделей, созданных на основе случайной выборки.

Совершенно очевидное различие частотных распределений уровней сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования и случайной модели (рисунки 12 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 14[4].

4.13.6. Обобщенная форма по достов. моделям при разн. крит. Титульная модель: "INF4"

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложн.-положительных решений (FP)	Число ложн.-отрицательных решений (FN)	Точность модели	Полнота модели	Фигура Бен-Розенберга	Сумма между-уровней со- истинно-поло- решений (ST)	Сумма между-уровней со- истинно-отри- цательных (ST)	Сумма между-уровней со- ложно-поло- жительных (SFP)	Сумма между-уровней со- ложно-отриц- ательных (SFP)
1. ABS - частный критерий: количество встреч сочетаний "х/лас"	Корреляция абс частот с абс...	100	100	697	2303		0.042	1.000	0.080	60.425	12.821	478.920	
1. ABS - частный критерий: количество встреч сочетаний "х/лас"	Сигма абс частот по призна...	100	100	690	2310		0.041	1.000	0.080	42.625		224.125	
2. PRIC1 - частный критерий: усл. вероятность его признака сред.	Корреляция усл.отн частот с о...	100	100	697	2303		0.042	1.000	0.080	60.425	12.821	478.920	
2. PRIC1 - частный критерий: усл. вероятность его признака сред.	Сигма усл.отн частот по при...	100	100	690	2310		0.041	1.000	0.080	41.013		384.694	
3. PRIC2 - частный критерий: условная вероятность его признака.	Корреляция усл.отн частот с о...	100	100	697	2303		0.042	1.000	0.080	60.424	12.821	478.916	
3. PRIC2 - частный критерий: условная вероятность его признака.	Сигма усл.отн частот по при...	100	100	690	2310		0.041	1.000	0.080	41.013		384.694	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	100	100	1773	1227		0.075	1.000	0.140	42.907	44.422	55.652	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сигма знаний	100	100	1082	1918		0.050	1.000	0.094	54.418	8.088	93.151	
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	100	100	1773	1227		0.075	1.000	0.140	42.907	44.422	55.652	
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сигма знаний	100	100	1082	1918		0.050	1.000	0.094	54.418	8.088	93.151	
6. INF3 - частный критерий: Хинкавадет, разности между факти...	Семантический резонанс зна...	100	100	1669	1331		0.070	1.000	0.131	50.002	219.946	155.131	
6. INF3 - частный критерий: Хинкавадет, разности между факти...	Сигма знаний	100	100	1669	1331		0.070	1.000	0.131	58.148	198.721	140.573	
7. INF4 - частный критерий: ROI (Return On Investment): верно...	Семантический резонанс зна...	100	100	2862	138		0.420	1.000	0.592	44.266	61.441	5.234	
7. INF4 - частный критерий: ROI (Return On Investment): верно...	Сигма знаний	100	100	1054	1946		0.049	1.000	0.093	25.128	0.312	10.989	
8. INF5 - частный критерий: ROI (Return On Investment): верно...	Семантический резонанс зна...	100	100	2862	138		0.420	1.000	0.592	44.266	61.441	5.234	
8. INF5 - частный критерий: ROI (Return On Investment): верно...	Сигма знаний	100	100	1054	1946		0.049	1.000	0.093	25.128	0.312	10.989	
9. INF6 - частный критерий: разн. усл.и безуслов. вероятностей, вер...	Семантический резонанс зна...	100	98	1343	1657	2	0.056	0.980	0.106	48.974	51.861	207.628	0.06
9. INF6 - частный критерий: разн. усл.и безуслов. вероятностей, вер...	Сигма знаний	100	100	1092	1908		0.050	1.000	0.095	28.598	9.301	170.884	
10. INF7 - частный критерий: разн. усл.и безуслов. вероятностей, ве...	Семантический резонанс зна...	100	98	1343	1657	2	0.056	0.980	0.106	48.974	51.861	207.628	0.06
10. INF7 - частный критерий: разн. усл.и безуслов. вероятностей, ве...	Сигма знаний	100	100	1092	1908		0.050	1.000	0.095	28.598	9.301	170.884	

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[3].

2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи

В соответствии с технологией АСК-анализа зададим текущей модель INF4 (режим 5.6) (рисунок 15) и проведем пакетное распознавание в режиме 4.2.1.

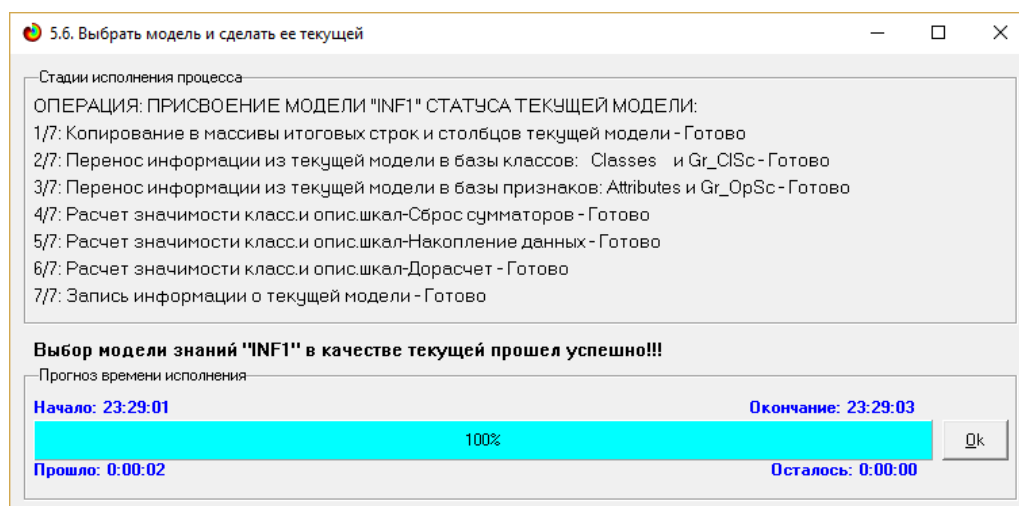
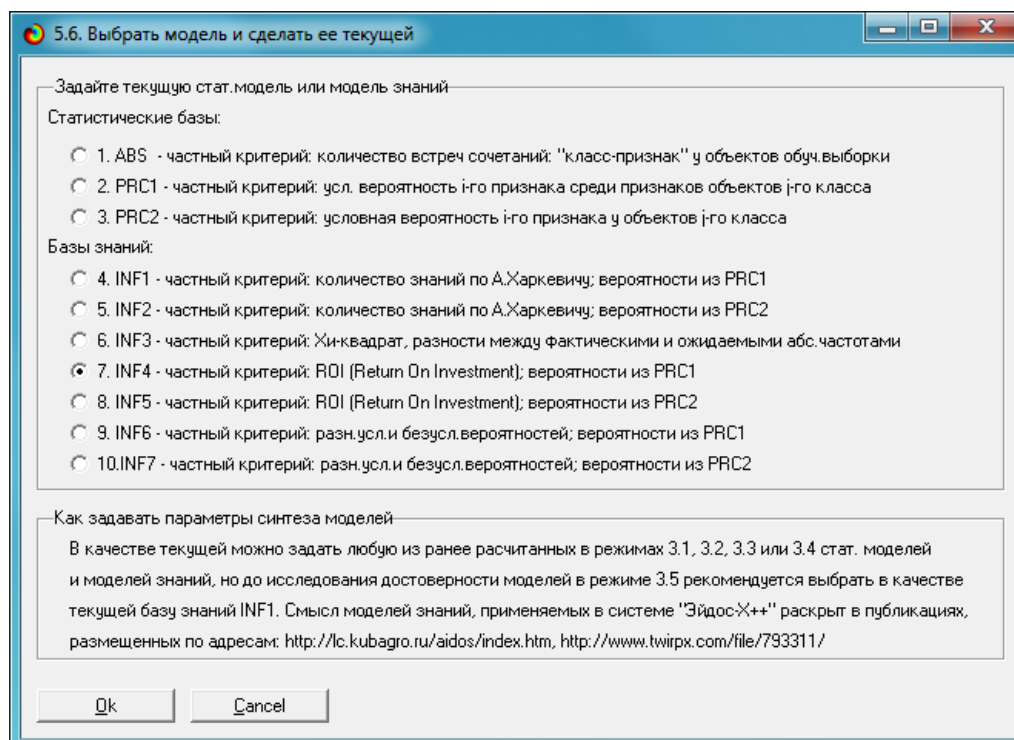


Рисунок 15. Экранные формы режима задания модели в качестве текущей

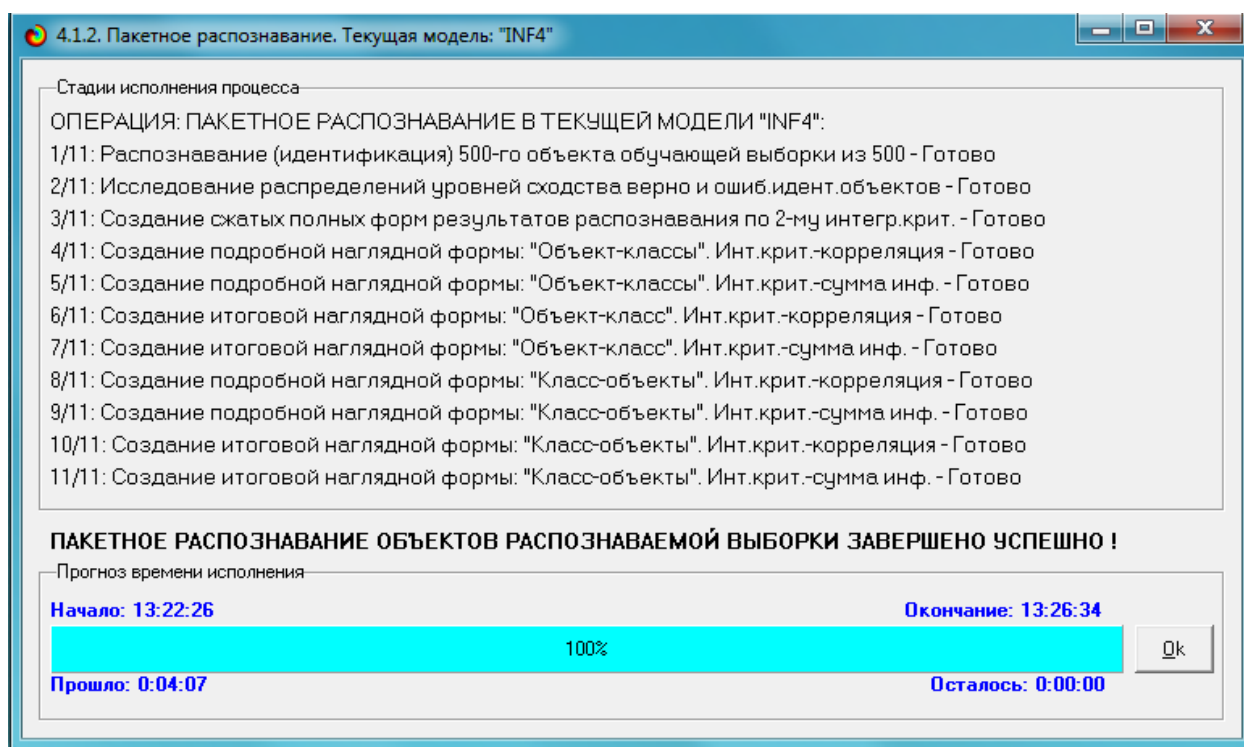


Рисунок 16. Экранная форма режима пакетного распознавания в текущей модели

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 15 и 16 приведены примеры прогнозов высокой и низкой достоверности частоты и классов ирисов в наиболее достоверной модели INF4 на основе наблюдения предыстории их развития:

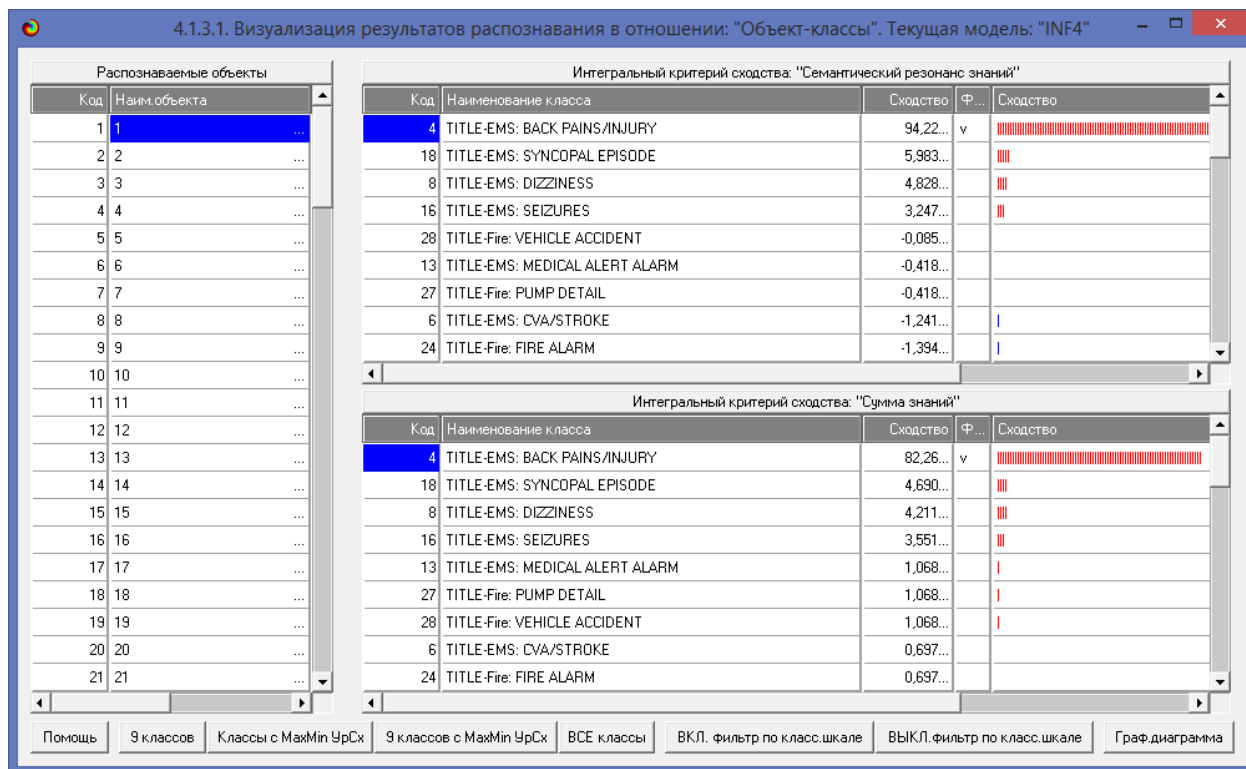


Рисунок 17. Пример идентификации классов в модели INF4

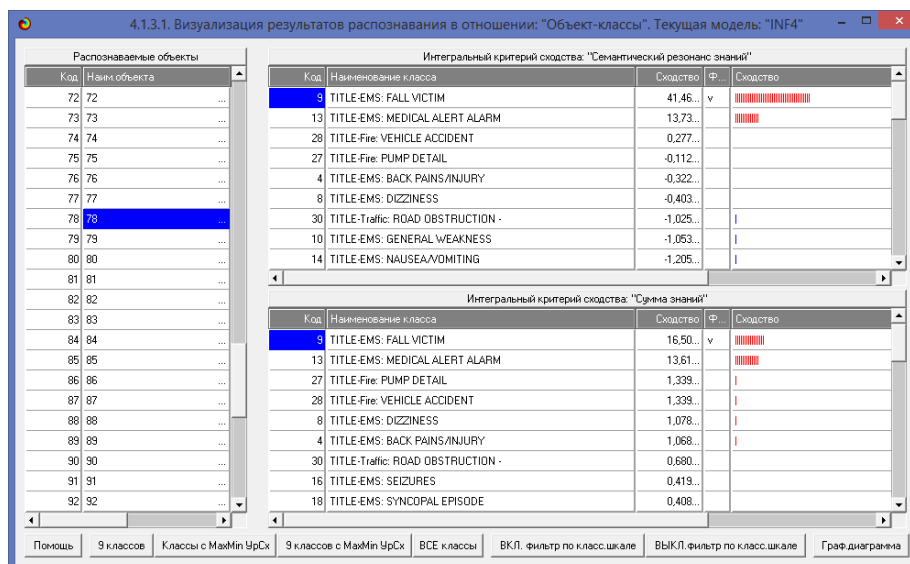


Рисунок 18. Пример идентификации классов в модели INF4

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)

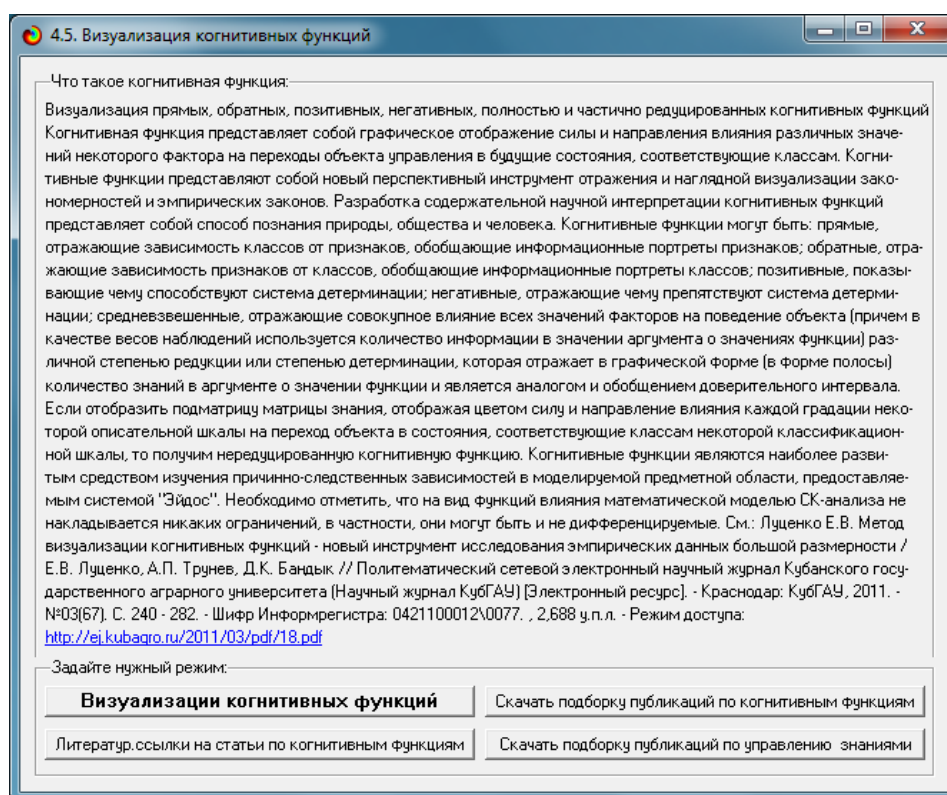


Рисунок 19. Экранная форма режима визуализации когнитивных функций

В когнитивных функциях количество информации в значениях аргумента о значениях функции отображается цветом (красным максимальное, синим минимальное), линией соединены значения функции о которых в значении аргумента содержится максимальное количество информации, ширина линии (аналог доверительного интервала) отражает степень неопределенности значения функции, которое тем ниже, чем больше информации о нем в значении функции (рис. 20–24):

На рисунке 20 представлена когнитивная функция, показывающая, что в меньшей широте и долготе преобладают вызовы по чрезвычайные ситуации, связанных с травмами людей. В средней широте и долготе вызовы связанные с кровоизлиянием, а в большей долготе вызовы, связанные с тошнотой [5].

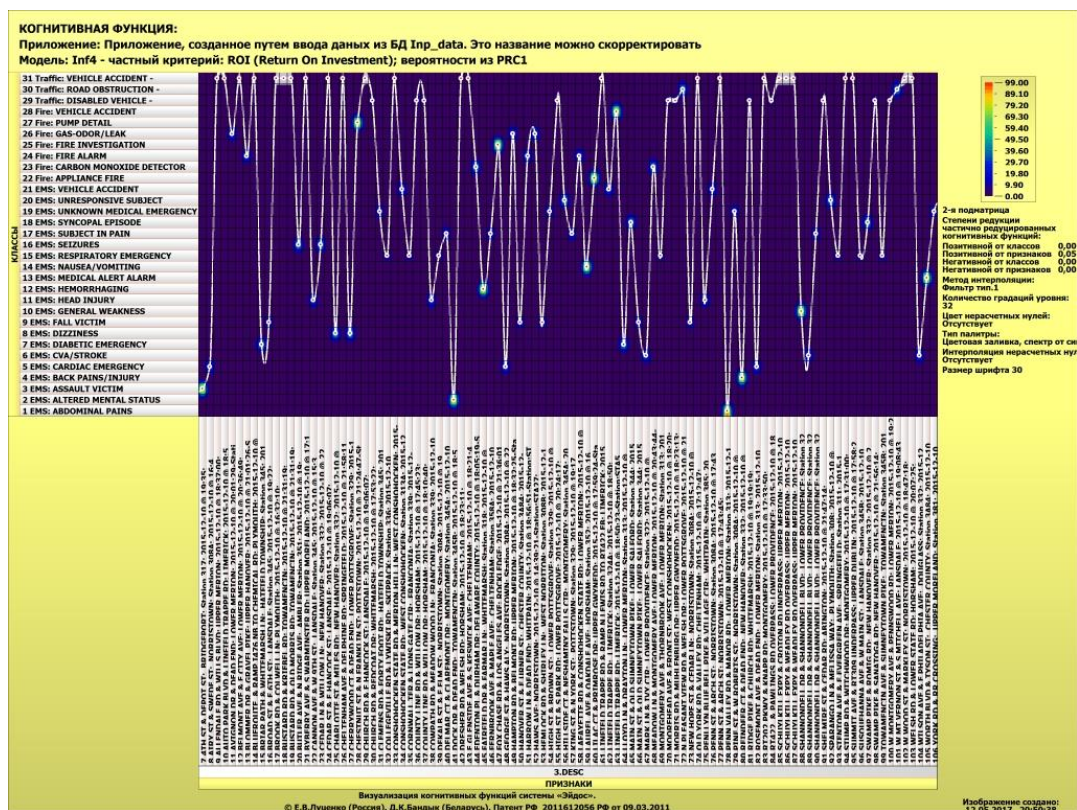


Рисунок 21. Когнитивная функция, отражающая взаимосвязь времени принятия вызова, станции принятия и типа чрезвычайной ситуации

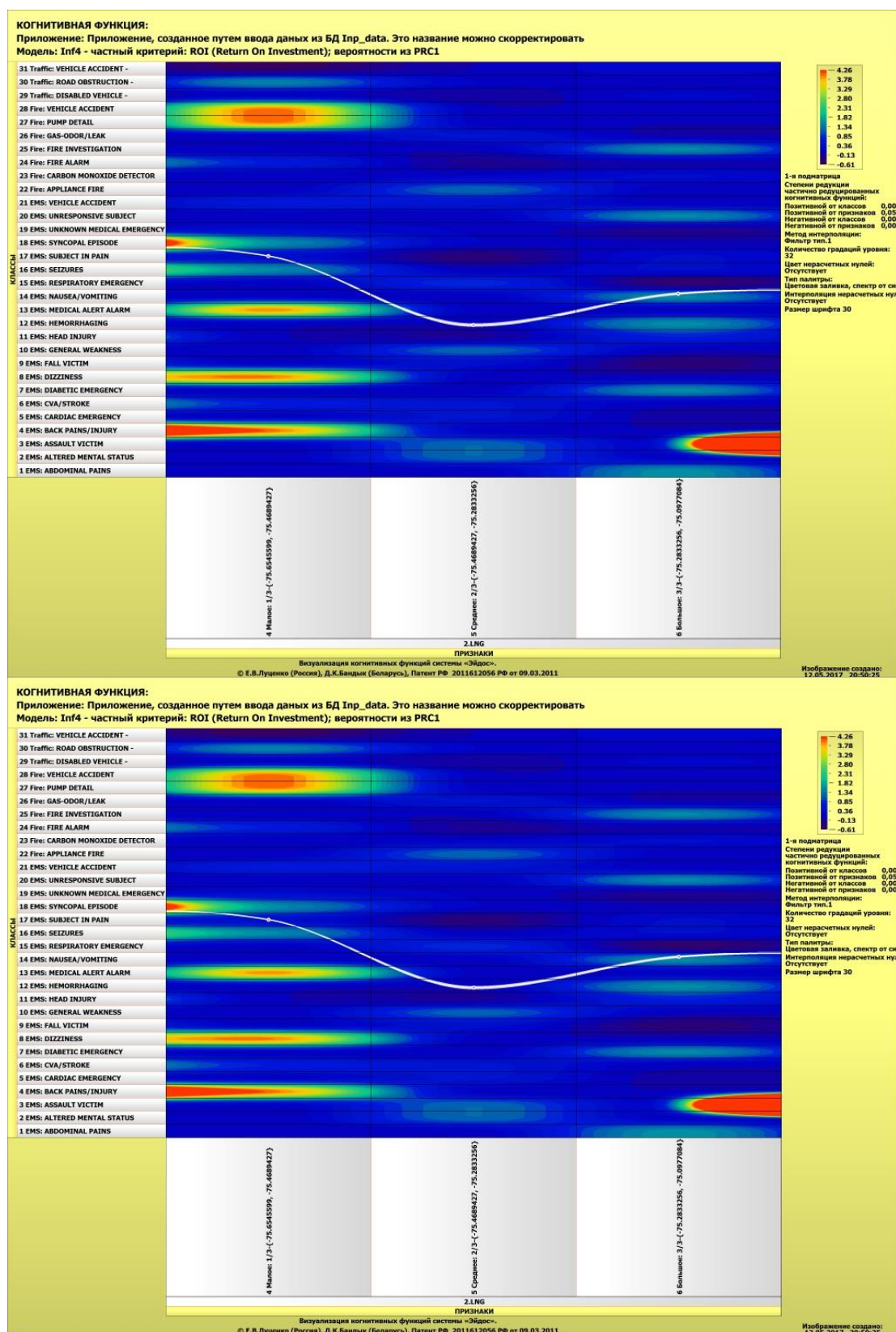


Рисунок 20. Когнитивная функция, отражающая взаимосвязь широты и долготы с типом чрезвычайной ситуации

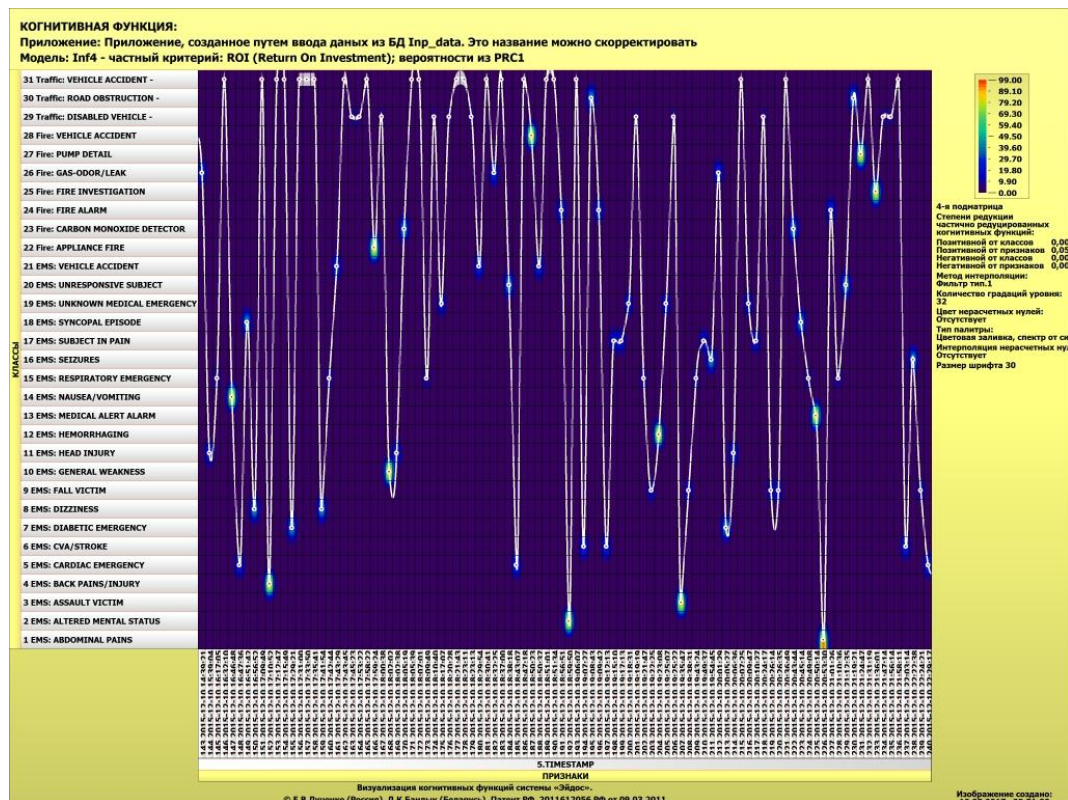


Рисунок 22. Когнитивная функция, отражающая взаимосвязь даты образования чрезвычайной ситуации и типа чрезвычайной ситуации

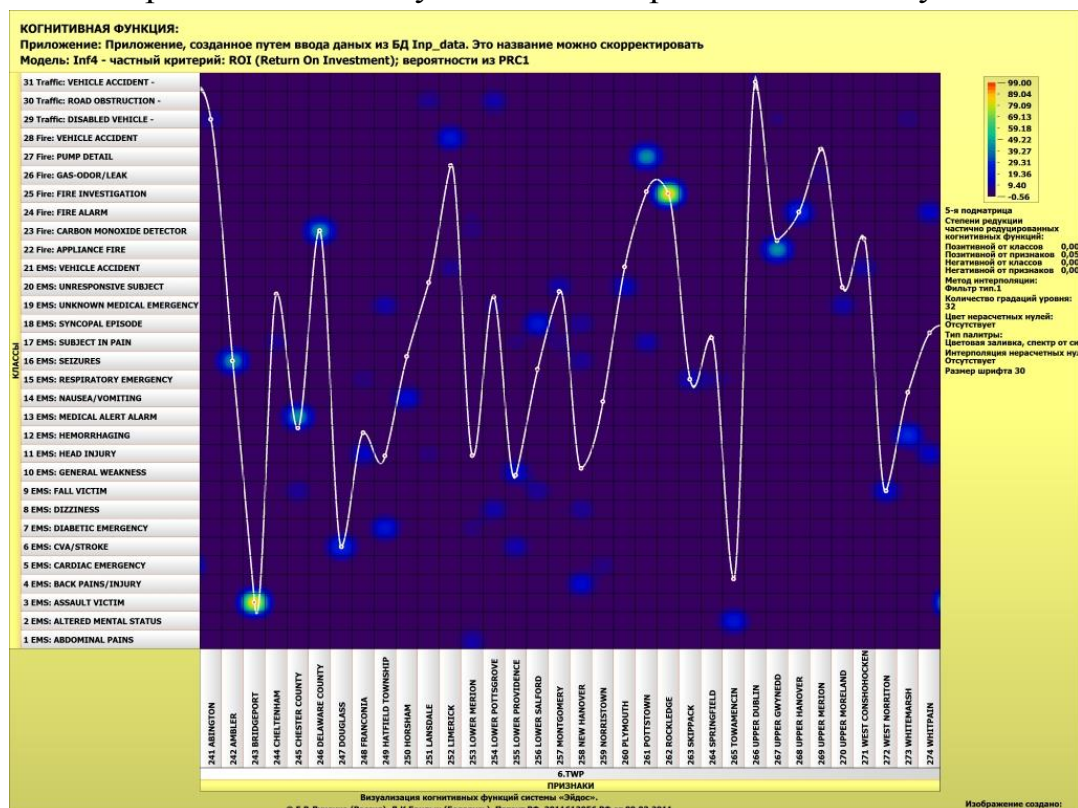


Рисунок 23. Когнитивная функция, отражающая взаимосвязь штата и типа чрезвычайной ситуации

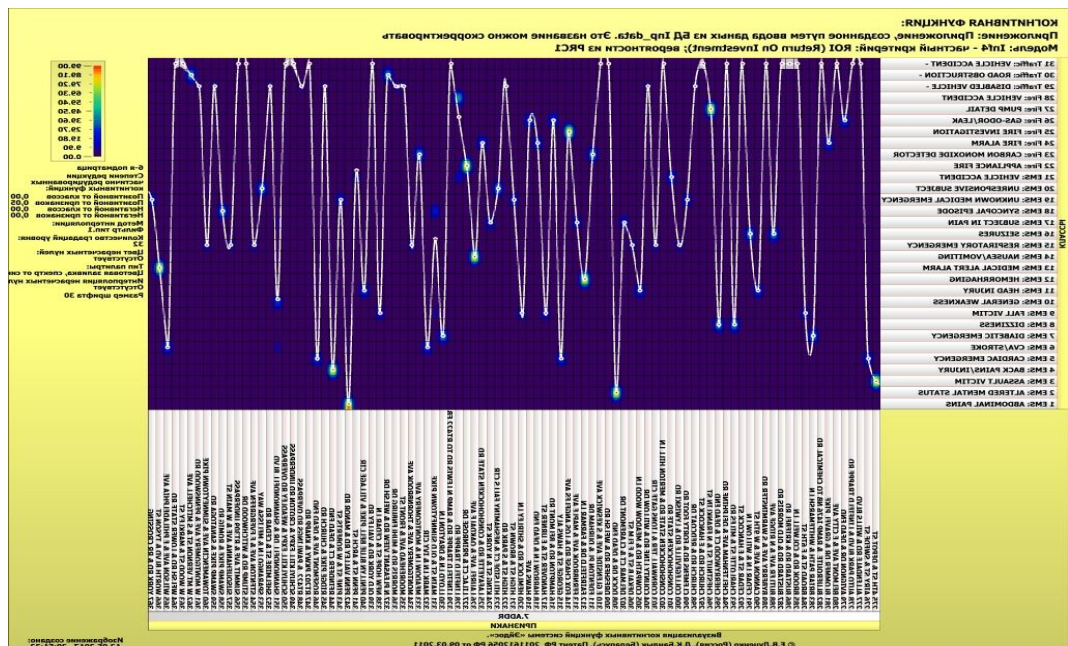


Рисунок 24. Когнитивная функция, отражающая взаимосвязь адреса и типа чрезвычайной ситуации

Сходство-различие обобщенных образов различных результатов научной деятельности по характерным для них системам значений показателей. Результаты сравнения классов по системе характерных для типов чрезвычайных ситуаций приведены на рисунок 25:

4.2.2.2. Результаты кластерно-конструктивного анализа классов

Конструкт класса: 1 "TITLE-EMS: ABDOMINAL PAINS" в модели: 4 "INF1"

Код	Наименование класса	№	Код класса	Наименование класса	Сходство
1	TITLE-EMS: ABDOMINAL PAINS	1	1	TITLE-EMS: ABDOMINAL PAINS	100.000
2	TITLE-EMS: ALTERED MENTAL STATUS	2	7	TITLE-EMS: DIABETIC EMERGENCY	4.486
3	TITLE-EMS: ASSAULT VICTIM	3	23	TITLE-Fire: CARBON MONOXIDE DETECTOR	4.263
4	TITLE-EMS: BACK PAINS/INJURY	4	5	TITLE-EMS: CARDIAC EMERGENCY	2.307
5	TITLE-EMS: CARDIAC EMERGENCY	5	26	TITLE-Fire: GAS-ODOR/LEAK	2.230
6	TITLE-EMS: CVA/STROKE	6	12	TITLE-EMS: HEMORRHAGING	1.138
7	TITLE-EMS: DIABETIC EMERGENCY	7	25	TITLE-Fire: FIRE INVESTIGATION	1.008
8	TITLE-EMS: DIZZINESS	8	24	TITLE-Fire: FIRE ALARM	0.936
9	TITLE-EMS: FALL VICTIM	9	14	TITLE-EMS: NAUSEA/VOMITING	0.103
10	TITLE-EMS: GENERAL WEAKNESS	10	3	TITLE-EMS: ASSAULT VICTIM	-0.256
11	TITLE-EMS: HEAD INJURY	11	20	TITLE-EMS: UNRESPONSIVE SUBJECT	-0.370
12	TITLE-EMS: HEMORRHAGING	12	10	TITLE-EMS: GENERAL WEAKNESS	-1.269
13	TITLE-EMS: MEDICAL ALERT ALARM	13	2	TITLE-EMS: ALTERED MENTAL STATUS	-1.269
14	TITLE-EMS: NAUSEA/VOMITING	14	22	TITLE-Fire: APPLIANCE FIRE	-1.276
15	TITLE-EMS: RESPIRATORY EMERGENCY	15	28	TITLE-Fire: VEHICLE ACCIDENT	-1.314
16	TITLE-EMS: SEIZURES	16	13	TITLE-EMS: MEDICAL ALERT ALARM	-1.315
17	TITLE-EMS: SUBJECT IN PAIN	17	27	TITLE-Fire: PUMP DETAIL	-1.315
18	TITLE-EMS: SYNCOPAL EPISODE	18	4	TITLE-EMS: BACK PAINS/INJURY	-1.369
19	TITLE-EMS: UNKNOWN MEDICAL EMERGENCY	19	16	TITLE-EMS: SEIZURES	-1.465
20	TITLE-EMS: UNRESPONSIVE SUBJECT	20	11	TITLE-EMS: HEAD INJURY	-1.505
21	TITLE-EMS: VEHICLE ACCIDENT	21	21	TITLE-EMS: VEHICLE ACCIDENT	-1.582
22	TITLE-Fire: APPLIANCE FIRE	22	30	TITLE-Traffic: ROAD OBSTRUCTION	-1.747
23	TITLE-Fire: CARBON MONOXIDE DETECTOR	23	8	TITLE-EMS: DIZZINESS	-1.803

Рисунок 25 - Результаты сравнения классов по системе характерных для типов чрезвычайных ситуаций

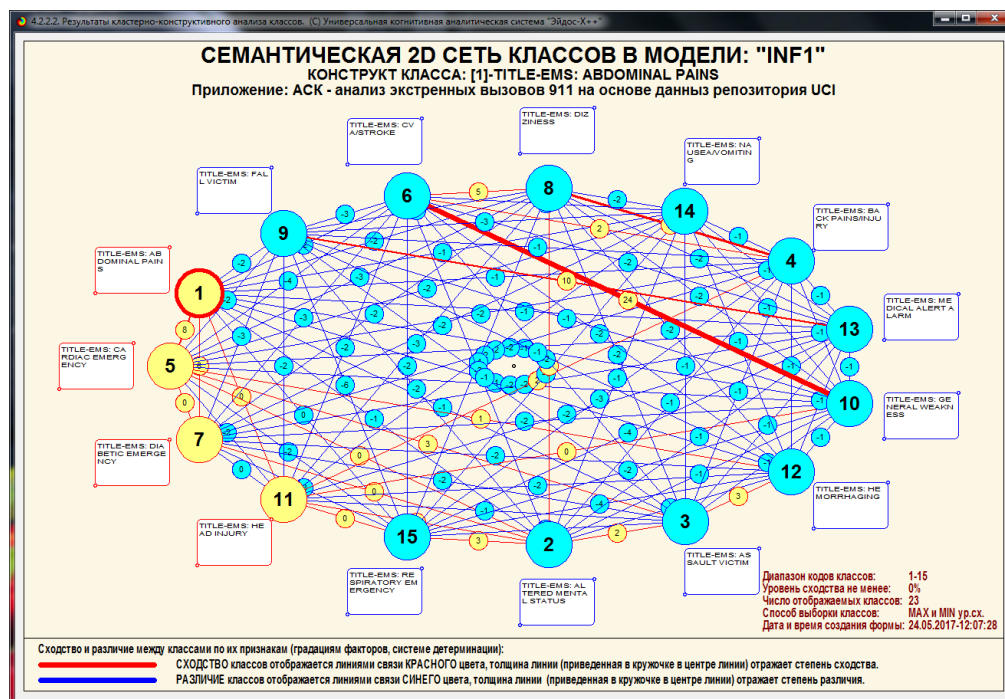


Рисунок 26. Результаты сравнения классов по системе характерных для значений чрезвычайных ситуаций

Из когнитивной диаграммы, приведенной на рисунке 26, мы видим, что, наибольшее количество вызовов по чрезвычайным ситуациям происходит из за болей в животе, диабета, угарного газа, пожар, утечки газа, кровоизлияний, пожарной сигнализации, средние количество вызовов из за тошноты, головокружений, травм головы, повреждений дыхательных путей, транспортных происшествий, а низкое количество вызовов из за отклонения психического состояния, головных болей, судорог и инсультов. Мы видим также, что пожары и срабатывания сигнализаций образуют один кластер, с низкой вариабельностью внутри него, а боли в животе и кровоизлияния образуют противоположный кластер с более высокой вариабельностью объектов, внутри него. Отметим также, что приведенная когнитивная диаграмма формируется системой «Эйдос» автоматически на основе созданных моделей.

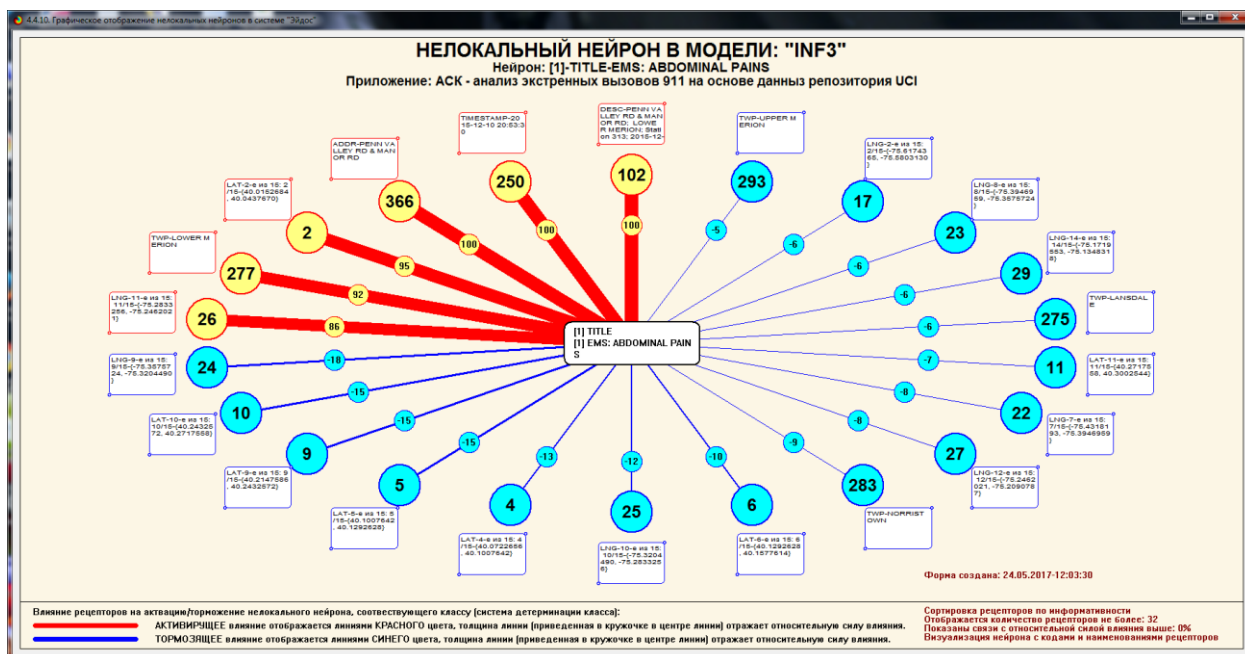


Рисунок 27. Графическое отображение нелокальных нейронов

Для каждого технологического фактора в соответствии с предложенной моделью определяется величина и направление его влияния на осуществление всех желаемых и не желаемых хозяйственных ситуаций. Для каждой ситуации эта информация отображается в различных текстовых и графических формах, в частности в форме нелокального нейрона (рисунок 27). На данной диаграмме цвет линии означает знак связи (красный – положительная, синий – отрицательная), а толщина – ее модуль.

Дополнение модели нейрона связями факторов позволяет построить классическую когнитивную карту ситуации (будущего состояния АОУ). Детальная внутренняя структура любой связи отображается в форме инвертированной когнитивной диаграммы (рисунок 28). Необходимо отметить, что все указанные графические формы генерируются системой "Эйдос" автоматически в соответствии с созданной моделью[9].

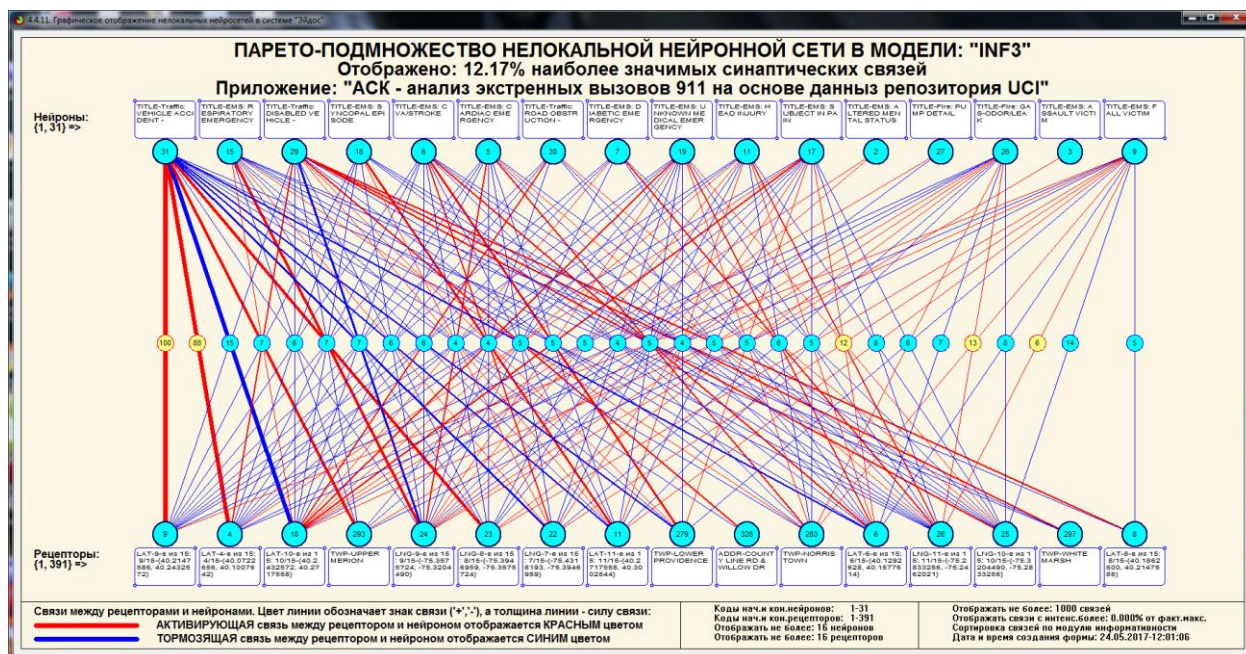


Рисунок 28 – Паретто – подмножеств нелокадной нейронной сети в системе «Эйдос - X»

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 29).

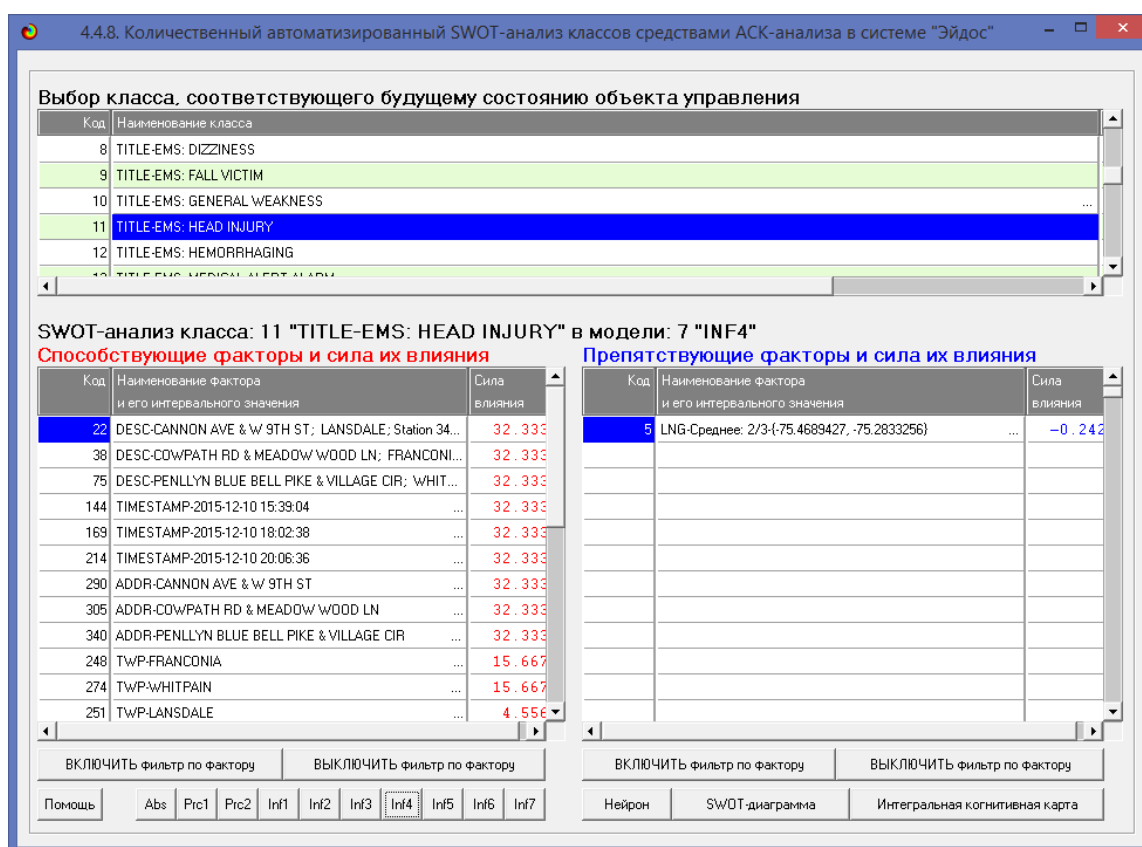


Рисунок 29. Пример SWOT-матрицы в модели INF4

На рисунке 30 приведены примеры инвертированной SWOT- матрицы и инвертированной SWOT-диаграммы в модели INF4.

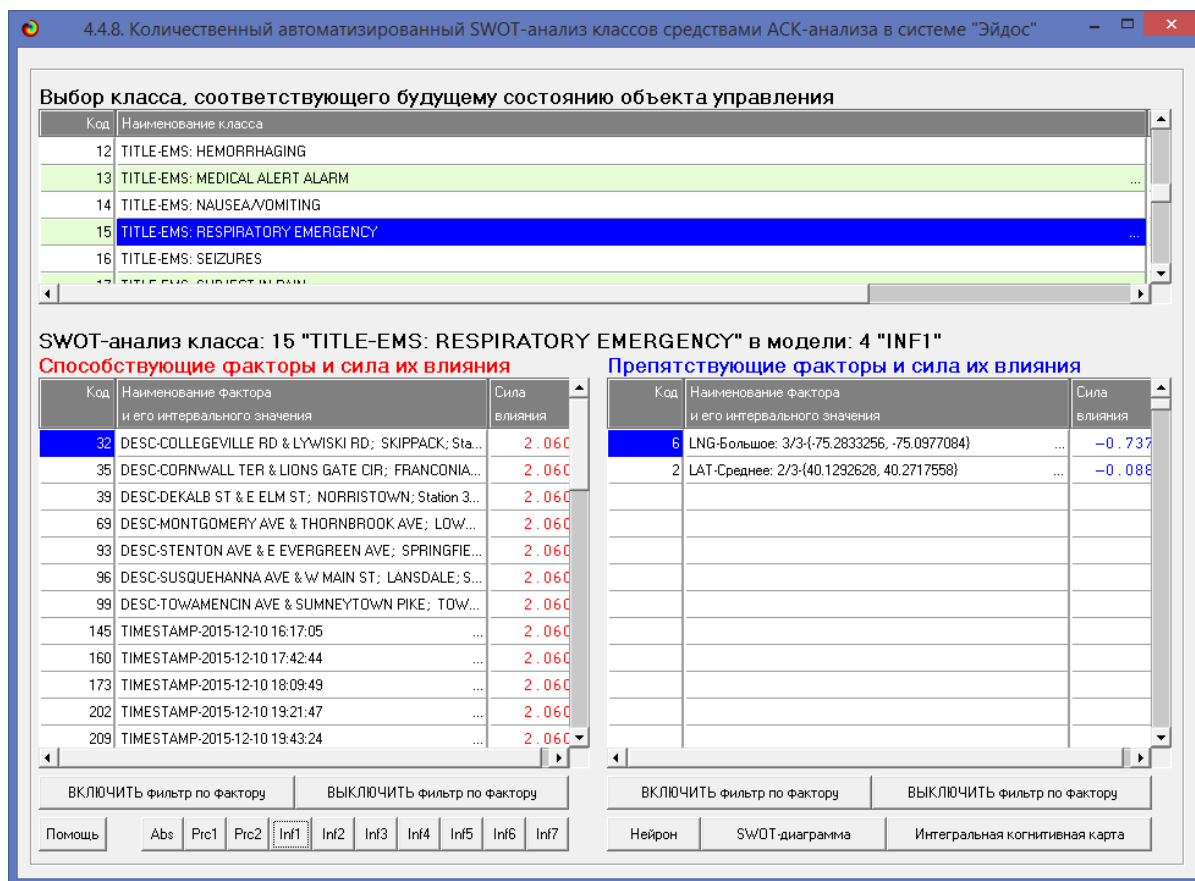


Рисунок 30. Табличная выходная форма количественного автоматизированного SWOT- и PEST-анализа средствами системы «Эйдос»

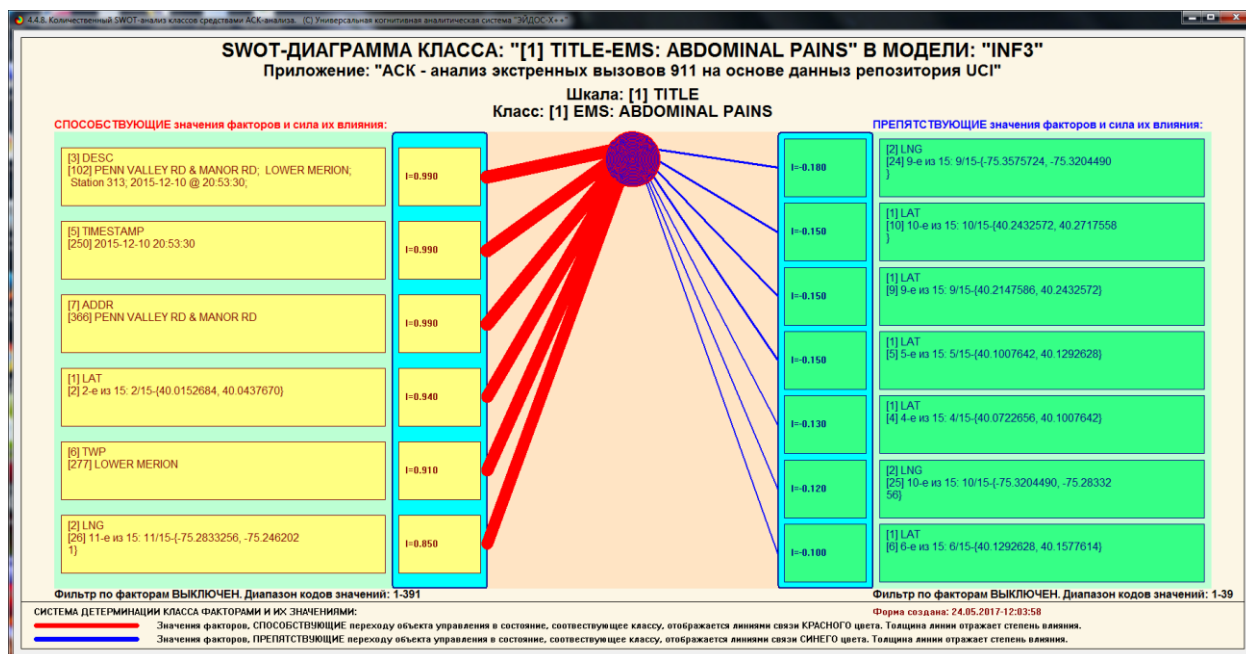


Рисунок 31. Графическая выходная форма количественного автоматизированного SWOT- и PEST-анализа средствами системы «Эйдос»

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 репозитория UCI. В данной курсовой работе приводится развернутый пример использования базы данных репозитория UCI 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 0,980, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 75%. 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 UCI, 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. 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. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос» и системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №02(116). С. 1 – 60. – IDA [article ID]: 1161602001. – Режим доступа: <http://ej.kubagro.ru/2016/02/pdf/01.pdf>, 3,75 у.п.л. 3.

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.

4. [Электронный ресурс]. Статья "Emergency – 911 Calls": <https://www.kaggle.com/mchirico/montcoalert>, свободный. - Загл. с экрана. Яз. англ.

5. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: <http://lc.kubagro.ru/>, свободный. - Загл. с экрана. Яз. рус.

6. Луценко Е.В. Количественная оценка степени манипулирования индексом Хирша и его модификация, устойчивая к манипулированию / Е.В.

Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №07(121). С. 202 – 234. – IDA [article ID]: 1211607005. – Режим доступа: <http://ej.kubagro.ru/2016/07/pdf/05.pdf>, 2,062 у.п.л., <http://dx.doi.org/10.21515/1990-4665-121-0057>.

7. 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.

8. 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>

9. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>, 0,812 у.п.л.