

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) Журавлева Евгения Сергеевича
курса 3 очной (заочной) формы обучения
направления подготовки Информационные системы и технологии
направленность(профиль)Интеллектуальные информационные системы и технологии.
Наименование темы «Употребление табака 1995-2010».
Рецензент: Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор
(Ф.И.О., ученое звание и степень, должность)

Оценка качества выполнения курсового проекта (работы)

№ п/п	Показатель	Оценка соответствия заданию (по 5-и балльной шкале)
1.	Актуальность тематики работы	5
2.	Степень полноты обзора состояния проблемы и корректность постановки цели и задач исследования	5
3.	Уровень и корректность использования в работе различных методов исследований	4
4.	Степень комплексности работы, применения в ней знаний общепрофессиональных и специальных дисциплин	5
5.	Ясность, четкость, последовательность и обоснованность изложения	4
6.	Применение современных технологий обработки информации	5
7.	Качество оформления работы (общий уровень грамотности, стиль изложения, качество иллюстраций, соответствие требованиям по оформлению)	4
8.	Ответы на вопросы при защите	5

Достоинства работы

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

Недостатки работы

Отдельные погрешности в оформлении

Итоговая оценка при защите 5 (отлично)

Рецензент 

Е.В. Луценко

« » 201 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:
**Употребление табака 1995-2010 с помощью АСК-анализа
с помощью «Эйдос» на основе данных репозитория UCI**
выполнил студент группы ИТ1421 **Журавлев Евгений Сергеевич**

Допущен к защите _____

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

(подпись, расшифровка подписи)

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

Защищена _____

Оценка

(дата)

Члены комиссии _____ В.И. Лойко

_____ Е.В. Луценко

_____ **И. В. Николаева (подпись, дата,**
расшифровка подписи)

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)

Журавлеву Евгению Сергеевичу
(Ф.И.О.)

Тема проекта: **Употребление табака 1995-2010 с помощью АСК-анализа с помощью «Эйдос» на основе репозитория UC1**

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.....	4
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ.....	5
1.1. Описание решения.....	5
1.2. Преобразование исходных данных из HTML-формата в файл исходных данных MS Excel.....	5
1.3. Синтез и верификация статистических и интеллектуальных моделей.	11
1.4. Виды моделей системы «Эйдос».....	13
1.5. Результаты верификации моделей.....	15
2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	23
2.1. Решение задачи.....	23
2.2. Когнитивные функции.....	26
ЗАКЛЮЧЕНИЕ.....	39
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	40

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. lng-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

	A	B	C	D	E	F	G
1	nomer	Smoke everyday	Smoke some days	Former smoker	Never smoked	Year	State
2	1	22,50	4,50	22,00	51,00	1999	Indiana
3	2	15,10	6,20	23,40	55,30	2008	Illinois
4	3	13,20	4,90	30,00	51,90	2004	Connecticut
5	4	19,10	4,80	24,90	51,20	1999	Wyoming
6	5	19,60	5,90	23,00	51,50	2003	South Carolina
7	6	19,10	5,00	21,90	54,00	1998	Mississippi
8	7	21,90	2,90	20,70	54,40	1999	Tennessee
9	8	19,90	6,20	22,50	51,40	2004	Oklahoma
10	9	14,30	5,40	24,50	55,80	2010	North Carolina
11	10	21,50	2,50	18,40	57,50	1996	Oklahoma
12	11	10,60	7,30	22,10	60,00	2006	District of Columbia
13	12	16,10	4,30	24,80	54,80	2005	Iowa
14	13	17,00	4,20	31,00	47,80	2003	New Hampshire
15	14	11,40	4,60	24,70	59,30	2010	Colorado
16	15	15,80	5,40	24,80	54,10	2007	Michigan
17	16	13,70	4,20	23,30	58,80	2007	Kansas
18	17	6,40	2,70	14,30	76,60	2010	Utah
19	18	22,60	4,60	24,70	48,10	2000	Missouri
20	19	20,40	6,20	20,70	52,70	2003	Louisiana
21	20	12,80	4,30	27,60	55,30	2009	Florida
22	21	17,20	4,40	29,10	49,30	1997	Connecticut
23	22	18,50	3,60	22,70	55,20	1997	Nebraska
24	23	18,80	5,60	25,80	49,90	1997	South Dakota
25	24	14,10	6,40	23,70	55,80	2006	Illinois

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.

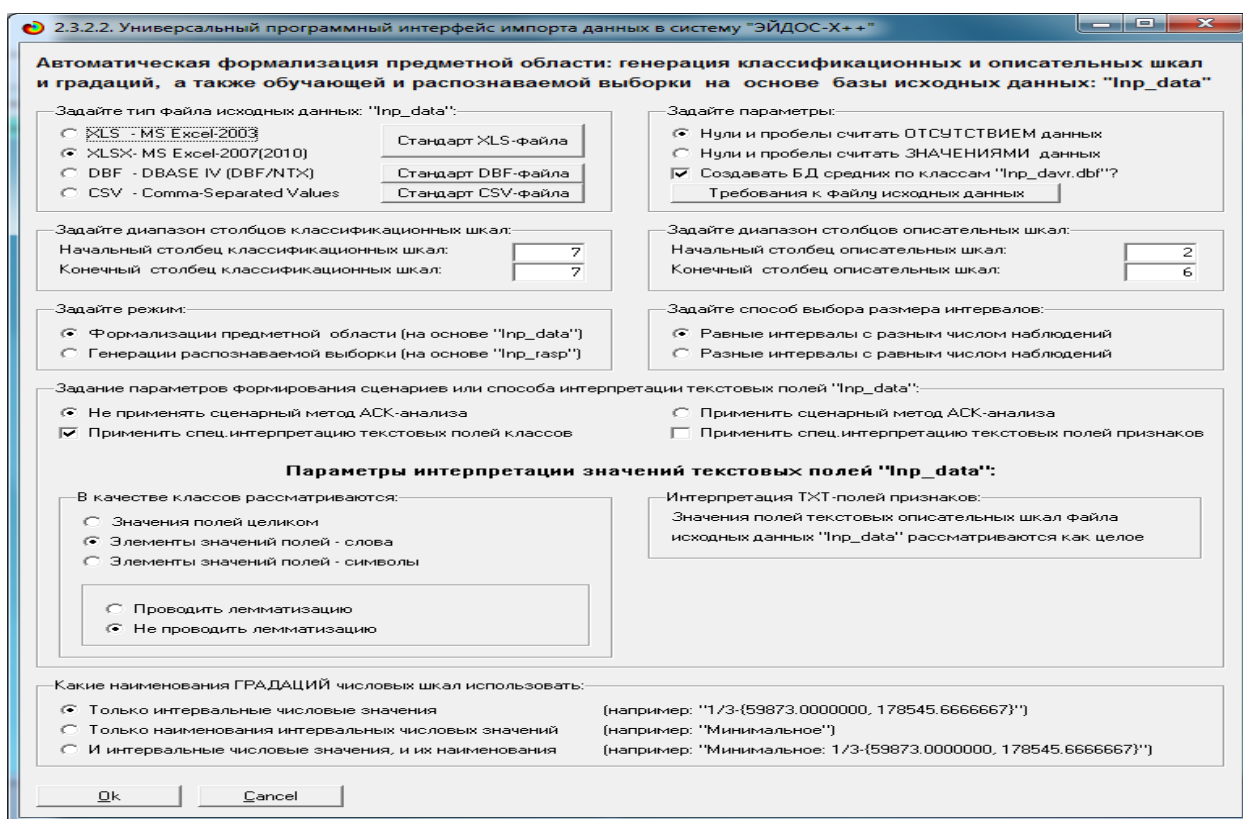


Рисунок 1 Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

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".

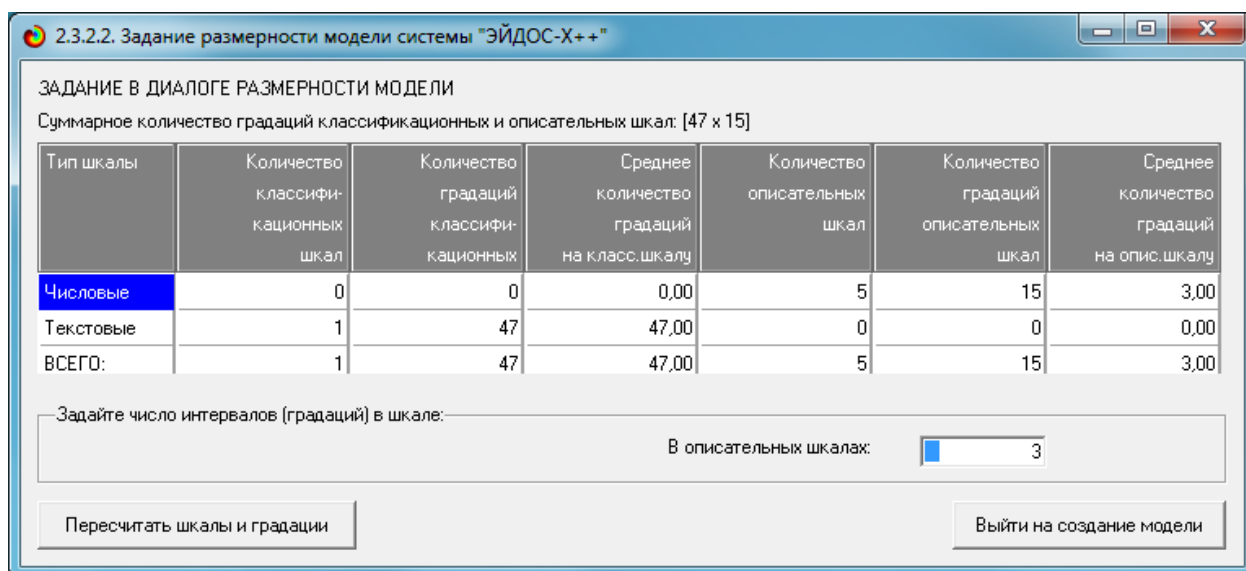


Рисунок 2. Задание размерности модели системы "Эйдос"

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".

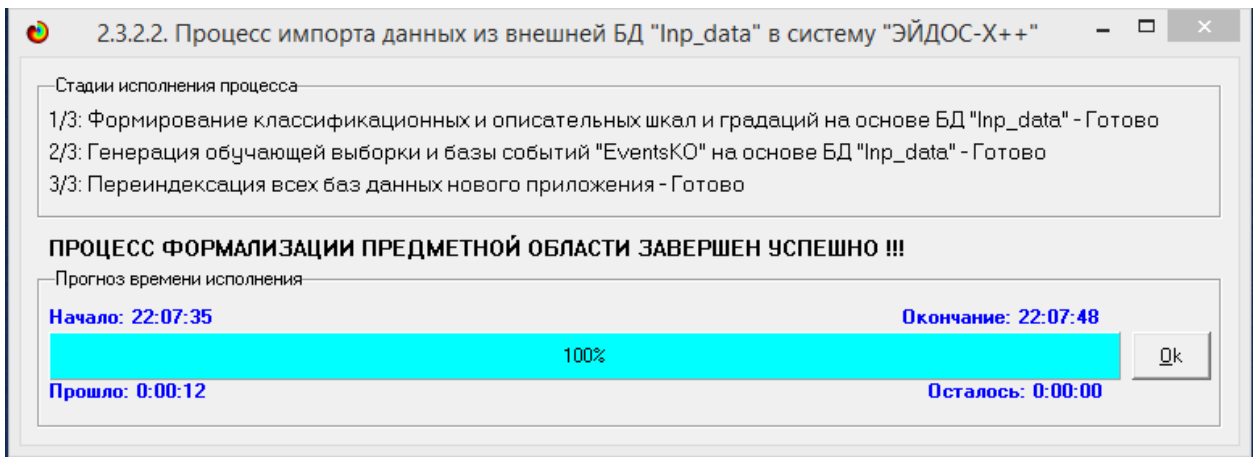


Рисунок 3. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

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).

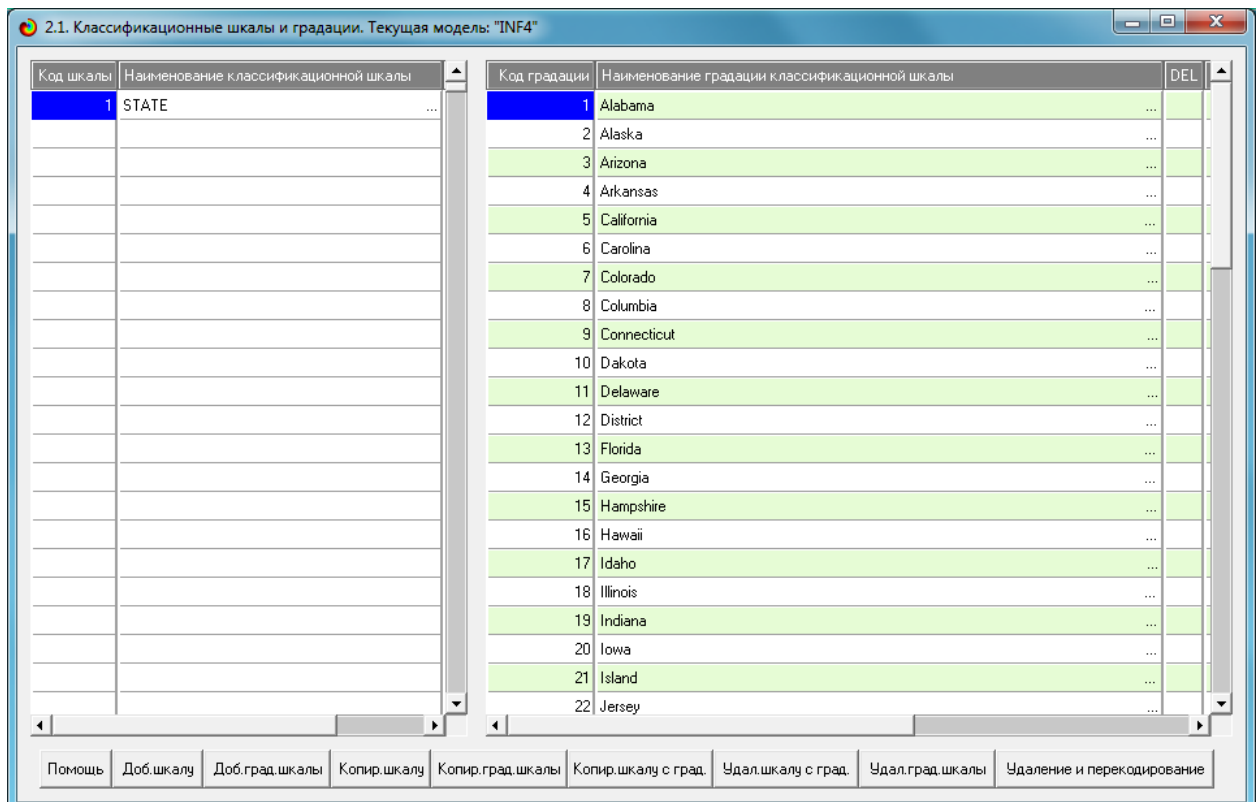


Рисунок 4. Классификационные шкалы и градации (фрагмент)

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

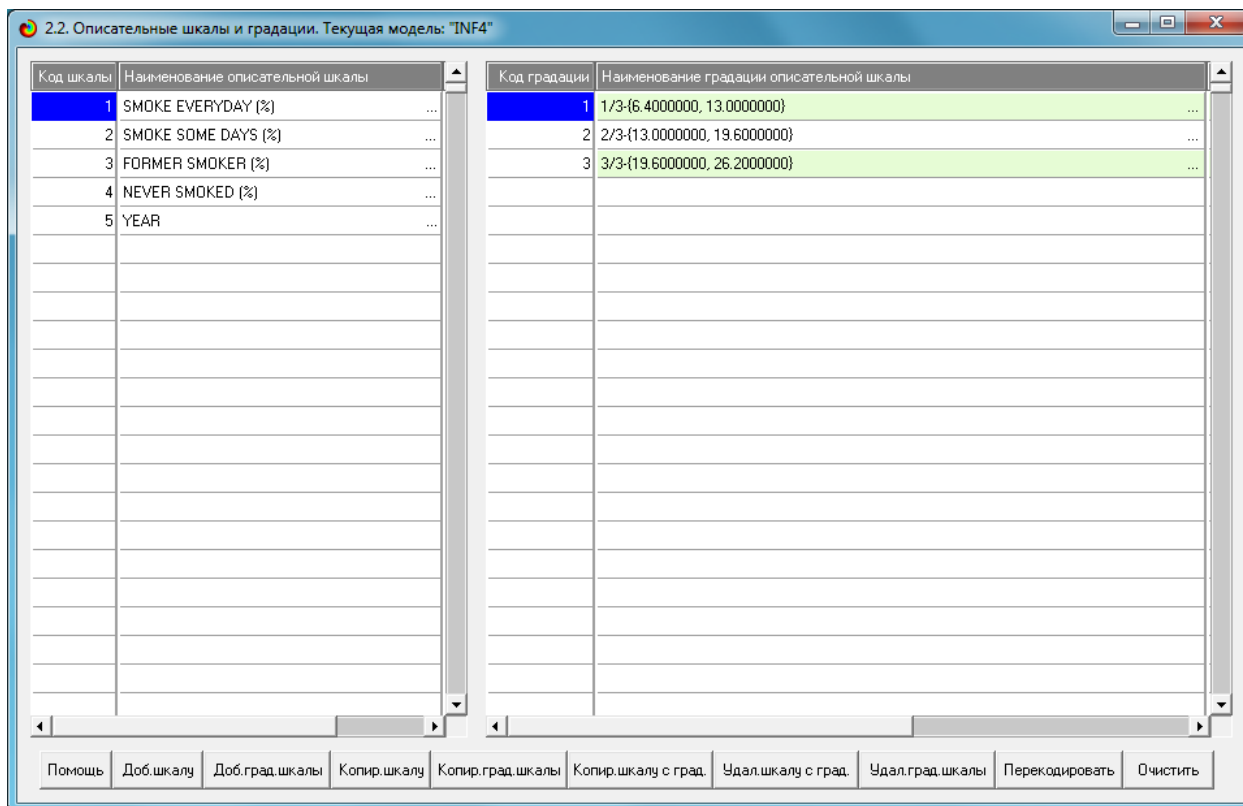


Рисунок 5. Описательные шкалы и градации (фрагмент)

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	Код объекта	Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
1	19	0	0	0	1	3	5	8	10	13	0	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).

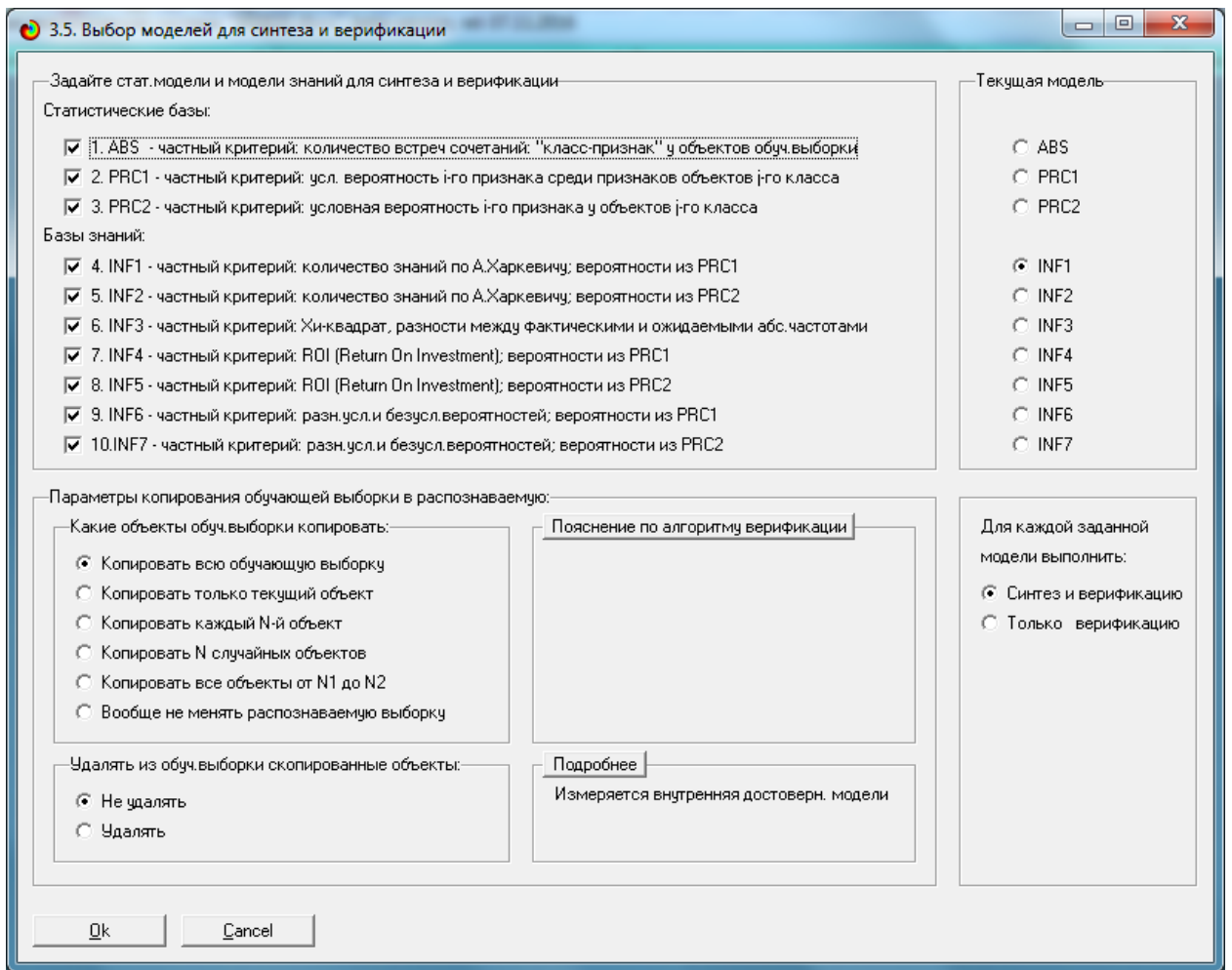


Рисунок 7. Выбор моделей для синтеза и верификации, а также текущей модели

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.

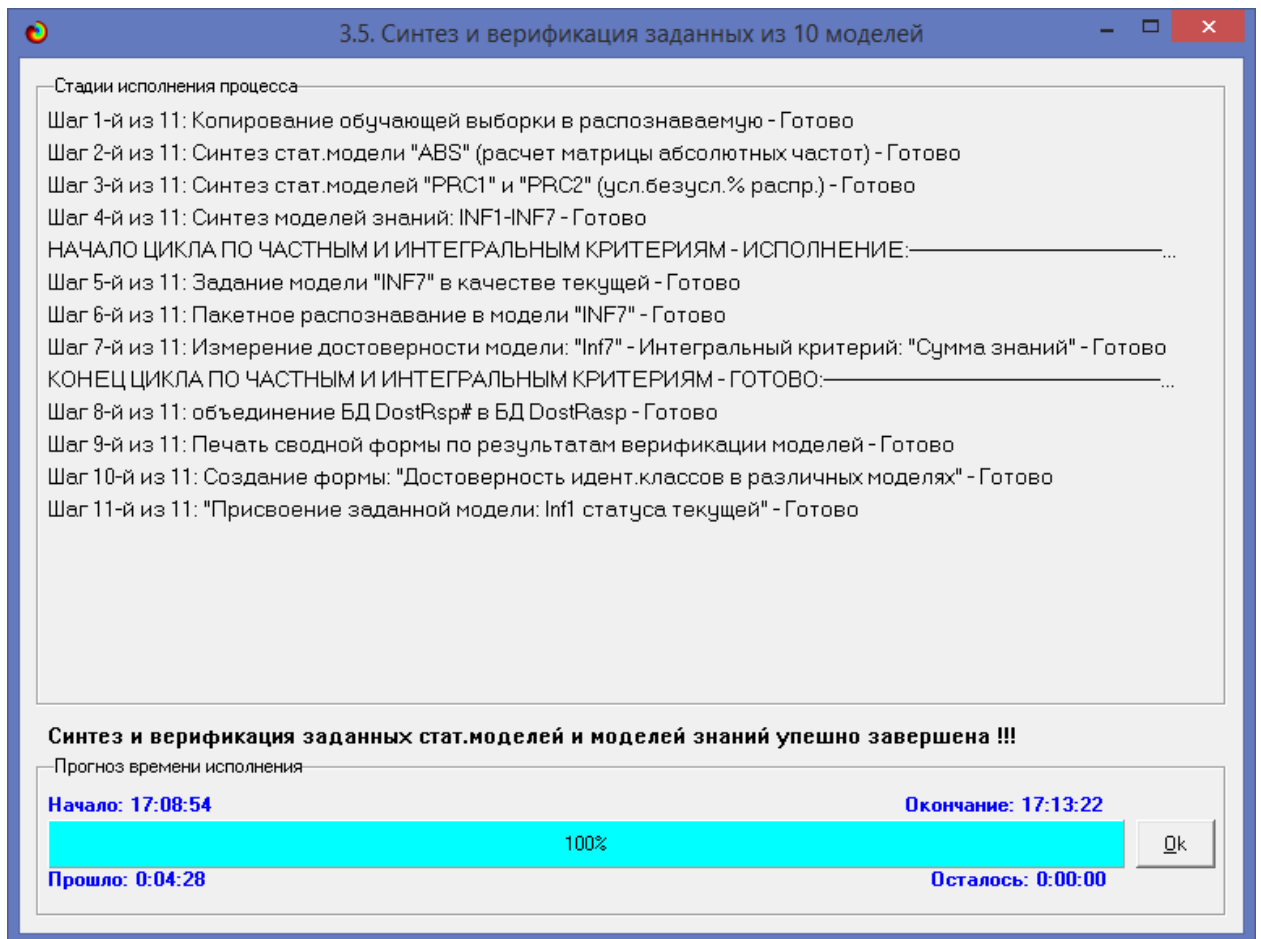


Рисунок 8. Синтез и верификация статистических моделей и моделей знаний

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)

Код признака	Наименование описательной шкалы и градации	1. STATE ALABAMA	2. STATE ALASKA	3. STATE ARIZONA	4. STATE ARKANSAS	5. STATE CALIFORNIA	6. STATE CAROLINA	7. STATE COLORADO	8. STATE COLUMBIA	9. STATE CONNECTI...	10. STATE DAKOTA	11. STATE DELAWARE	12. STATE DISTRICT	13. STATE FLORIDA
1	SMOKE EVERYDAY (5)1/3-6.4000000, 13.0000000							2	1					
2	SMOKE EVERYDAY (3)2/3-13.0000000, 19.6000000	1	2	1	1	1	4	1	1	3	6	3	1	1
3	SMOKE EVERYDAY (3)3/3-19.6000000, 26.2000000			1			3				1			
4	SMOKE SOME DAYS (2)1/3-1.9000000, 3.8000000			1				1			2			
5	SMOKE SOME DAYS (2)2/3-3.8000000, 5.7000000			1		1	5	2		3	2	4		2
6	SMOKE SOME DAYS (2)3/3-5.7000000, 7.6000000	1	2		1		2		2		3		2	
7	FORMER SMOKER (2)1/3-14.3000000, 20.1666667			1					1				1	
8	FORMER SMOKER (2)2/3-20.1666667, 26.0333333	1		1	1		7	2	1		5		1	1
9	FORMER SMOKER (2)3/3-26.0333333, 31.9000000		2			1		1		3	2	4		1
10	NEVER SMOKED (1)1/3-39.5000000, 51.8666667		2	1	1		2	1		2	2	1		
11	NEVER SMOKED (1)2/3-51.8666667, 64.2333333	1		1		1	5	2	2	1	5	3	2	2
12	NEVER SMOKED (1)3/3-64.2333333, 76.6000000													
13	YEAR-1/3-1995.0000000, 2000.0000000		1	2		1	3	1	1	1	4		1	
14	YEAR-2/3-2000.0000000, 2005.0000000	1	1		1		2			2	3			1
15	YEAR-3/3-2005.0000000, 2010.0000000						2	2	1				1	1
	Сумма числа признаков	5	10	10	5	5	35	15	10	15	35	20	10	10
	Среднее	0	1	1	0	0	2	1	1	1	2	1	1	1
	Среднеквадратичное отклонение	1	1	1	1	1	2	1	1	1	2	2	1	1
	Сумма числа объектов обучающей выборки	1	2	2	1	1	7	3	2	3	7	4	2	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. STATE ALABAMA	2. STATE ALASKA	3. STATE ARIZONA	4. STATE ARKANSAS	5. STATE CALIFORNIA	6. STATE CAROLINA	7. STATE COLORADO	8. STATE COLUMBIA	9. STATE CONNECTI...	10. STATE DAKOTA	11. STATE DELAWARE	12. STATE DISTRICT	13. STATE FLORIDA	14. STATE GE
1	SMOKE EVERYDAY (%)1/3-16.4000000, 13.0000...							1.422	1.168						
2	SMOKE EVERYDAY (%)2/3-113.0000000, 19.600...	0.424	0.424	-0.187	0.424	0.424	-0.069	-0.545	-0.187	0.424	0.288	0.171	-0.187	-0.187	
3	SMOKE EVERYDAY (%)3/3-119.6000000, 26.200...			0.638			0.502				-0.467				
4	SMOKE SOME DAYS (%)1/3-11.9000000, 3.80000...			0.794				0.436			0.300				
5	SMOKE SOME DAYS (%)2/3-13.8000000, 5.70000...			-0.105		0.506	0.209	0.149		0.506	-0.599	0.506		0.506	
6	SMOKE SOME DAYS (%)3/3-15.7000000, 7.60000...	1.279	1.279		1.279		0.173		1.279		0.531		1.279		
7	FORMER SMOKER (%)1/3-14.3000000, 20.1666...			1.476					1.476				1.476		
8	FORMER SMOKER (%)2/3-120.1666667, 26.0333...	0.494		-0.117	0.494		0.494	0.136	-0.117		0.197		-0.117	-0.117	
9	FORMER SMOKER (%)3/3-126.0333333, 31.9000...		0.961			0.961		-0.008		0.961	-0.144	0.961		0.350	
10	NEVER SMOKED (%)1/3-139.5000000, 51.866666...		0.864	0.252	0.864		-0.241	-0.105		0.506	-0.241	-0.359			
11	NEVER SMOKED (%)2/3-151.8666667, 64.233333...	0.459		-0.153		0.459	0.162	0.101	0.459	-0.511	0.162	0.205	0.459	0.459	
12	NEVER SMOKED (%)3/3-164.2333333, 76.60000...														
13	YEAR-1/3-11995.0000000, 2000.0000000		0.165	0.777		0.777	0.029	-0.193	0.165	-0.193	0.283		0.165		
14	YEAR-2/3-12000.0000000, 2005.0000000	1.118	0.506		1.118		0.013			0.760	0.370			0.506	
15	YEAR-3/3-12005.0000000, 2010.0000000						-0.058	0.690	0.436			1.047	0.436	0.436	
	Сумма	3.774	4.200	3.374	4.180	3.127	1.214	2.082	4.678	2.455	0.680	3.088	4.678	3.120	
	Среднее	0.252	0.280	0.225	0.279	0.208	0.081	0.139	0.312	0.164	0.045	0.206	0.312	0.208	
	Среднеквадратичное отклонение	0.428	0.430	0.479	0.454	0.330	0.201	0.447	0.547	0.386	0.312	0.392	0.547	0.359	

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. STATE ALABAMA	2. STATE ALASKA	3. STATE ARIZONA	4. STATE ARKANSAS	5. STATE CALIFORNIA	6. STATE CAROLINA	7. STATE COLORADO	8. STATE COLUMBIA	9. STATE CONNECTI...	10. STATE DAKOTA	11. STATE DELAWARE	12. STATE DISTRICT	13. STATE FLORIDA	14. STATE GE
1	SMOKE EVERYDAY (%)1/3-16.4000000, 13.0000...							4.012	2.759				0.879	2.759	2.759
2	SMOKE EVERYDAY (%)2/3-113.0000000, 19.600...	0.618	0.618	-0.191	0.618	0.618	-0.076	-0.461	-0.191	0.618	0.387	0.213	-0.191	-0.191	
3	SMOKE EVERYDAY (%)3/3-119.6000000, 26.200...			1.061			0.767				-0.411				
4	SMOKE SOME DAYS (%)1/3-11.9000000, 3.80000...			1.458				0.638			0.404				
5	SMOKE SOME DAYS (%)2/3-13.8000000, 5.70000...			-0.113		0.775	0.268	0.183		0.775	-0.493	0.775		0.775	
6	SMOKE SOME DAYS (%)3/3-15.7000000, 7.60000...	3.260	3.260		3.260		0.217		3.260		0.826		3.260		
7	FORMER SMOKER (%)1/3-14.3000000, 20.1666...			4.325					4.325				4.325		
8	FORMER SMOKER (%)2/3-120.1666667, 26.0333...	0.751		-0.125	0.751		0.751	0.167	-0.125		0.250		-0.125	-0.125	
9	FORMER SMOKER (%)3/3-126.0333333, 31.9000...		1.972			1.972		-0.009		1.972	-0.151	1.972		0.486	
10	NEVER SMOKED (%)1/3-139.5000000, 51.866666...		1.663	0.331	1.663		-0.239	-0.113		0.775	-0.239	-0.334			
11	NEVER SMOKED (%)2/3-151.8666667, 64.233333...	0.682		-0.159		0.682	0.201	0.121	0.682	-0.439	0.201	0.261	0.682	0.682	
12	NEVER SMOKED (%)3/3-164.2333333, 76.60000...														
13	YEAR-1/3-11995.0000000, 2000.0000000		0.206	1.411		1.411	0.033	-0.196	0.206	-0.196	0.378		0.206		
14	YEAR-2/3-12000.0000000, 2005.0000000	2.550	0.775		2.550		0.014			1.367	0.521			0.775	
15	YEAR-3/3-12005.0000000, 2010.0000000						-0.064	1.185	0.638			2.277	0.638	0.638	
	Сумма	7.860	8.493	7.999	8.841	5.458	1.873	5.528	11.554	4.871	1.674	6.044	11.554	5.799	
	Среднее	0.524	0.566	0.533	0.589	0.364	0.125	0.369	0.770	0.325	0.112	0.403	0.770	0.387	
	Среднеквадратичное отклонение	1.015	0.982	1.192	1.057	0.618	0.286	1.076	1.440	0.654	0.360	0.767	1.440	0.743	

1.5. Результаты верификации моделей

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

4.13.6. Обобщенная форма по достов. моделям при разн. 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 ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризбергера	Сумма междууровневой схожести истинно-положительных решений (ST)	Сумма междууровневой схожести истинно-отрицательных решений (ST)	Сумма междууровневой схожести ложно-положительных решений (SFP)	Сумма междууровневой схожести ложно-отрицательных решений (SFP)
1. ABS - частный критерий: количество встреч соизганий "у/лас"	Корреляция абс частот с обр...	128	127	1762	3310		1	0.037	0.992	0.071	100.257	421.955	1241.189
1. ABS - частный критерий: количество встреч соизганий "у/лас"	Сумма абс частот по призна...	128	128	167	4905			0.025	1.000	0.050	60.185		956.037
2. PRIC1 - частный критерий: усл. вероятность его признака сред...	Корреляция усл.отн частот с о...	128	128	1732	3340			0.037	1.000	0.071	100.257	421.955	1241.189
2. PRIC1 - частный критерий: усл. вероятность его признака сред...	Сумма усл.отн частот по при...	128	128	167	4905			0.025	1.000	0.050	97.000		2148.978
3. PRIC2 - частный критерий: условная вероятность его признака...	Корреляция усл.отн частот с о...	128	127	1762	3310		1	0.037	0.992	0.071	100.253	421.944	1241.150
3. PRIC2 - частный критерий: условная вероятность его признака...	Сумма усл.отн частот по при...	128	128	167	4905			0.025	1.000	0.050	96.824		2147.025
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	128	121	2733	2339		7	0.049	0.945	0.094	77.589	719.470	736.016
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	128	124	764	4308		4	0.028	0.969	0.054	37.109	28.374	627.068
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	128	121	2725	2347		7	0.049	0.945	0.093	77.610	719.005	736.600
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	128	124	761	4311		4	0.028	0.969	0.054	37.079	28.069	627.404
6. INF3 - частный критерий: Хинвадат, разности между фактич...	Семантический резонанс зна...	128	126	2674	2398		2	0.050	0.984	0.095	94.199	817.319	849.225
6. INF3 - частный критерий: Хинвадат, разности между фактич...	Сумма знаний	128	126	2674	2398		2	0.050	0.984	0.095	39.829	303.850	264.230
7. INF4 - частный критерий: ROI (Return On Investment): вероатно...	Семантический резонанс зна...	128	119	2956	2116		9	0.053	0.930	0.101	76.265	740.236	684.524
7. INF4 - частный критерий: ROI (Return On Investment): вероатно...	Сумма знаний	128	126	601	4471		2	0.027	0.984	0.053	15.200	3.007	201.561
8. INF5 - частный критерий: ROI (Return On Investment): вероатно...	Семантический резонанс зна...	128	119	2956	2116		9	0.053	0.930	0.101	76.271	739.779	684.817
8. INF5 - частный критерий: ROI (Return On Investment): вероатно...	Сумма знаний	128	126	601	4471		2	0.027	0.984	0.053	15.174	2.969	201.475
9. INF6 - частный критерий: разн усл.и безуслов вероятностей, вер...	Семантический резонанс зна...	128	123	2429	2643		5	0.044	0.961	0.085	84.253	613.441	849.869
9. INF6 - частный критерий: разн усл.и безуслов вероятностей, вер...	Сумма знаний	128	126	673	4399		2	0.028	0.984	0.054	56.100	21.662	1068.689
10. INF7 - частный критерий: разн усл.и безуслов вероятностей, ве...	Семантический резонанс зна...	128	123	2428	2644		5	0.044	0.961	0.085	84.309	612.841	850.583
10. INF7 - частный критерий: разн усл.и безуслов вероятностей, ве...	Сумма знаний	128	126	672	4400		2	0.028	0.984	0.054	55.960	21.362	1068.299

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 мультимассовое обобщение, предложенное профессором Е.В.Луценко [7] (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++".

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 бесполозен, поэтому он и назван мной псевдопрогнозом.

ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.

Представим себе, что мы выбрасываем 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, 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 $\{0, 1\}$:

$$L1 = \frac{TP + TN - FP - FN}{TP + TN + FP + FN} \quad (\text{нормировка: } \{-1, +1\})$$

$$L2 = \frac{1 + (TP + TN - FP - FN)}{TP + TN + FP + FN} / 2 \quad (\text{нормировка: } \{0, 1\})$$

где: TP - истинно-положительное решение; TN - истинно-отрицательное решение; FP - ложно-положительное решение; FN - ложно-отрицательное решение;

Классическая F-мера достоверности моделей Ван Ризбергена:

$$\text{Precision} = TP / (TP + FP) \text{ - точность модели;}$$

$$\text{Recall} = TP / (TP + FN) \text{ - полнота модели;}$$

$$F\text{-мера} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}).$$

В АСК-анализе и системе "Эйдос" предлагается L-мера, представляющая собой нечеткое мультиклассовое обобщение классической F-меры достоверности моделей Ван Ризбергена:

$$SPrecision = STP / (STP + SFP) \text{ - нечеткая мультиклассовая точность модели;}$$

$$SRecall = STP / (STP + SFN) \text{ - нечеткая мультиклассовая полнота модели;}$$

$$L\text{-мера} = 2 * (SPrecision * SRecall) / (SPrecision + SRecall) \text{ - нечеткая мультиклассовая достоверность модели.}$$

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.

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
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. РРС1 - ЧАСТНЫЙ КРИТЕРИЙ: УСЛ. ВЕРОЯТНОСТЬ ИГО 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)													

Помощь Графики TP,TN,FP,FN, инт. крит.-резонанс знаний Графики TP,TN,FP,FN, инт. крит.-сумма знаний Excel

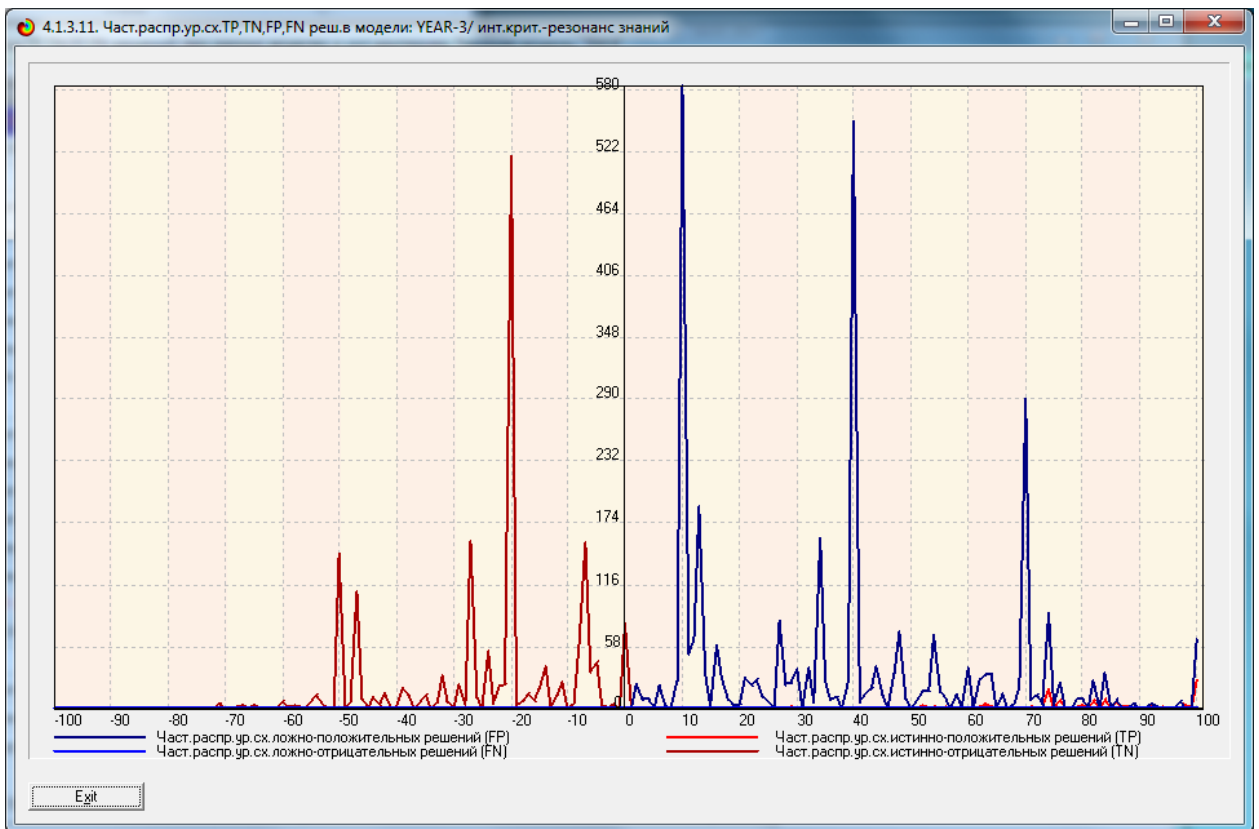


Рисунок 11. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF4

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):

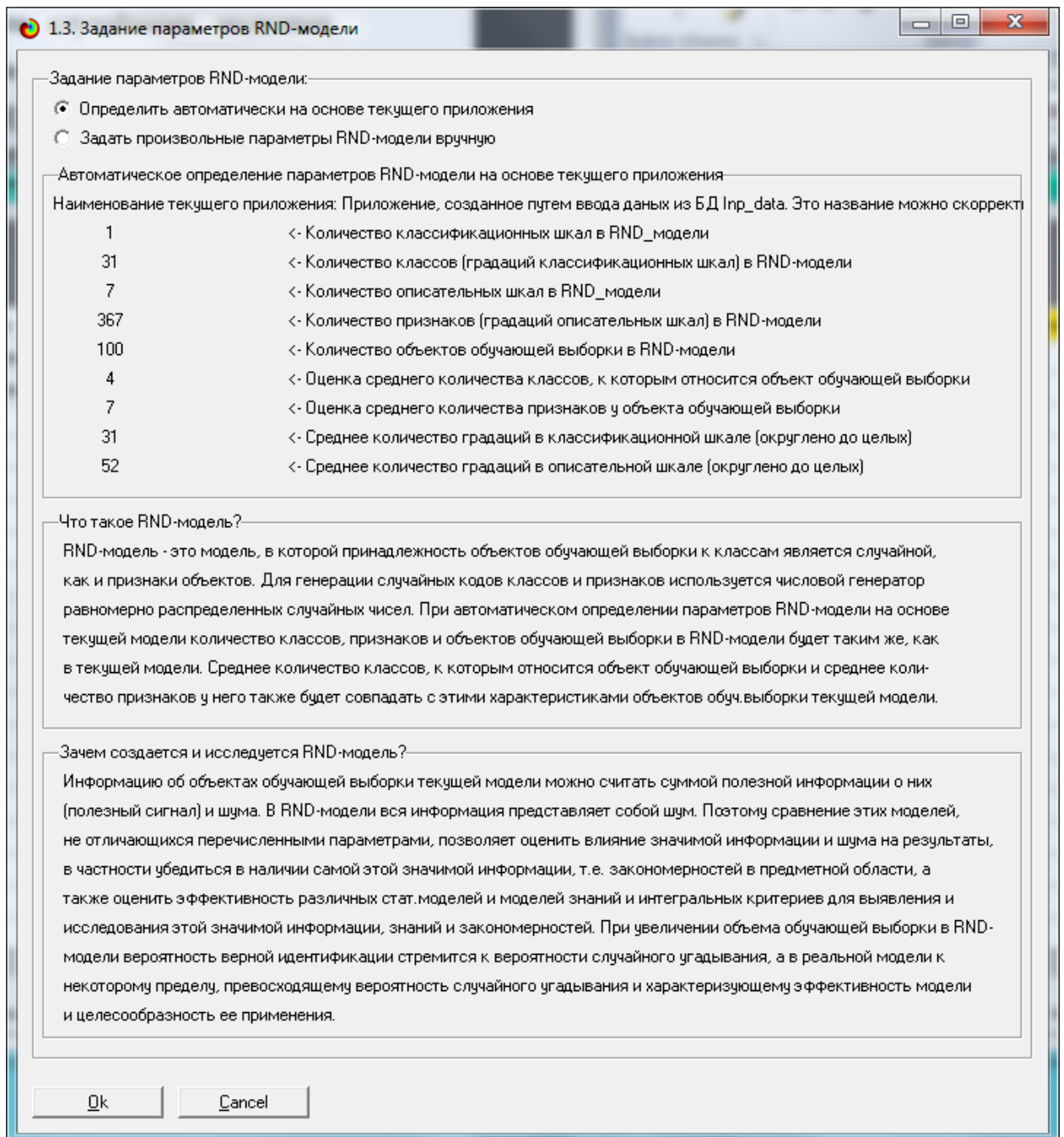
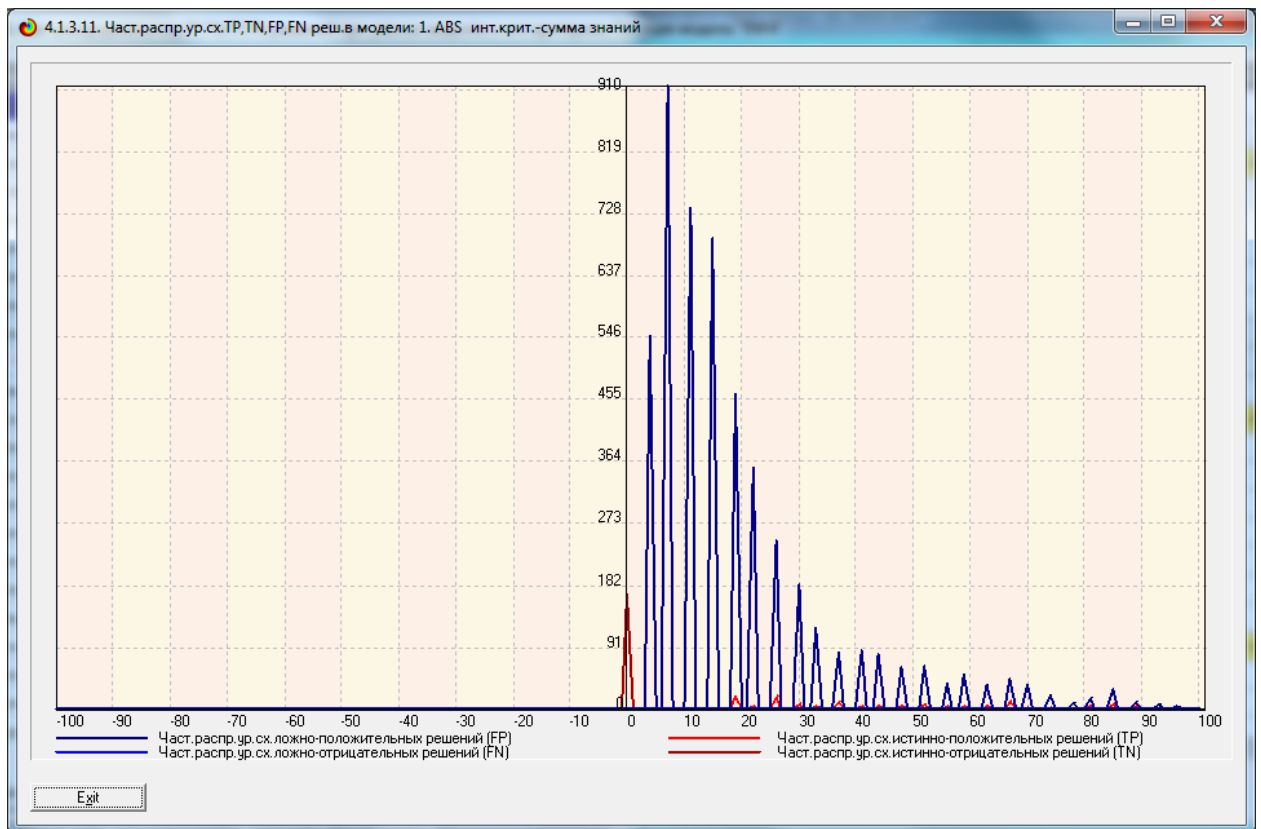


Рисунок 12. Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

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].

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	Фигура Бэн-Риббергено	Средняя между-уровневая сигма истинно-положительных решений (ST)	Средняя между-уровневая сигма истинно-отрицательных решений (ST)	Средняя между-уровневая сигма ложно-положительных решений (SFP)	Средняя между-уровневая сигма ложно-отрицательных решений (SFP)
1. ABS - частный критерий: количество встреч согласия... "Клас...	Корреляция абс частот с обр...	128	127	1762	3310	1	0.037	0.992	0.071	100.257	421.955	1241.189	
1. ABS - частный критерий: количество встреч согласия... "Клас...	Сигма абс частот по призна...	128	128	167	4905		0.025	1.000	0.050	60.185		956.037	
2. PRC1 - частный критерий: усл. вероятность его признака сред...	Корреляция усл.отн частот с о...	128	128	1732	3340		0.037	1.000	0.071	100.257	421.955	1241.189	
2. PRC1 - частный критерий: усл. вероятность его признака сред...	Сигма усл.отн частот по при...	128	128	167	4905		0.025	1.000	0.050	97.000		2148.978	
3. PRC2 - частный критерий: условная вероятность его признака...	Корреляция усл.отн частот с о...	128	127	1762	3310	1	0.037	0.992	0.071	100.253	421.944	1241.150	
3. PRC2 - частный критерий: условная вероятность его признака...	Сигма усл.отн частот по при...	128	128	167	4905		0.025	1.000	0.050	96.824		2147.025	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	128	121	2733	2339	7	0.049	0.945	0.094	77.589	719.470	736.016	1.301
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сигма знаний	128	124	764	4308	4	0.028	0.969	0.054	37.109	28.374	627.068	0.211
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	128	121	2725	2347	7	0.049	0.945	0.093	77.610	719.005	736.600	1.294
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сигма знаний	128	124	761	4311	4	0.028	0.969	0.054	37.079	28.069	627.404	0.211
6. INF3 - частный критерий: Хинкавадат, разности между факти...	Семантический резонанс зна...	128	126	2674	2398	2	0.050	0.984	0.095	94.199	817.319	849.225	0.411
6. INF3 - частный критерий: Хинкавадат, разности между факти...	Сигма знаний	128	126	2674	2398	2	0.050	0.984	0.095	39.829	303.850	264.230	0.201
7. INF4 - частный критерий: ROI (Return On Investment): вероятно...	Семантический резонанс зна...	128	119	2956	2116	9	0.053	0.930	0.101	76.265	740.236	684.524	1.031
7. INF4 - частный критерий: ROI (Return On Investment): вероятно...	Сигма знаний	128	126	601	4471	2	0.027	0.984	0.053	15.200	3.007	201.561	0.011
8. INF5 - частный критерий: ROI (Return On Investment): вероятно...	Семантический резонанс зна...	128	119	2956	2116	9	0.053	0.930	0.101	76.271	739.779	684.817	1.021
8. INF5 - частный критерий: ROI (Return On Investment): вероятно...	Сигма знаний	128	126	601	4471	2	0.027	0.984	0.053	15.174	2.969	201.475	0.011
9. INF6 - частный критерий: разн. усл.и безуслов. вероятностей, вер...	Семантический резонанс зна...	128	123	2429	2643	5	0.044	0.961	0.085	84.253	613.441	849.869	0.801
9. INF6 - частный критерий: разн. усл.и безуслов. вероятностей, вер...	Сигма знаний	128	126	673	4399	2	0.028	0.984	0.054	56.100	21.662	1068.689	0.161
10. INF7 - частный критерий: разн. усл.и безуслов. вероятностей, ве...	Семантический резонанс зна...	128	123	2428	2644	5	0.044	0.961	0.085	84.309	612.841	850.583	0.801
10. INF7 - частный критерий: разн. усл.и безуслов. вероятностей, ве...	Сигма знаний	128	126	672	4400	2	0.028	0.984	0.054	55.960	21.362	1068.299	0.161

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 В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи идентификации

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

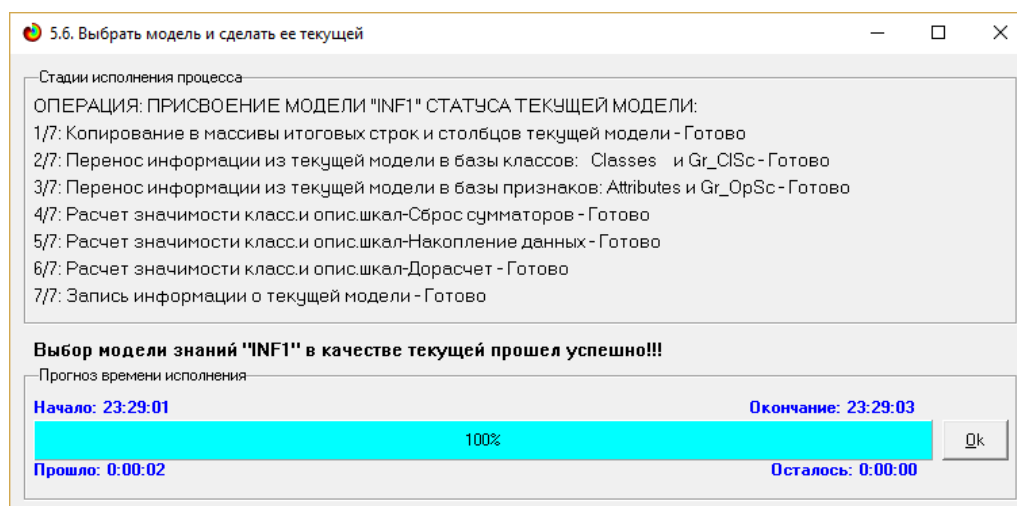
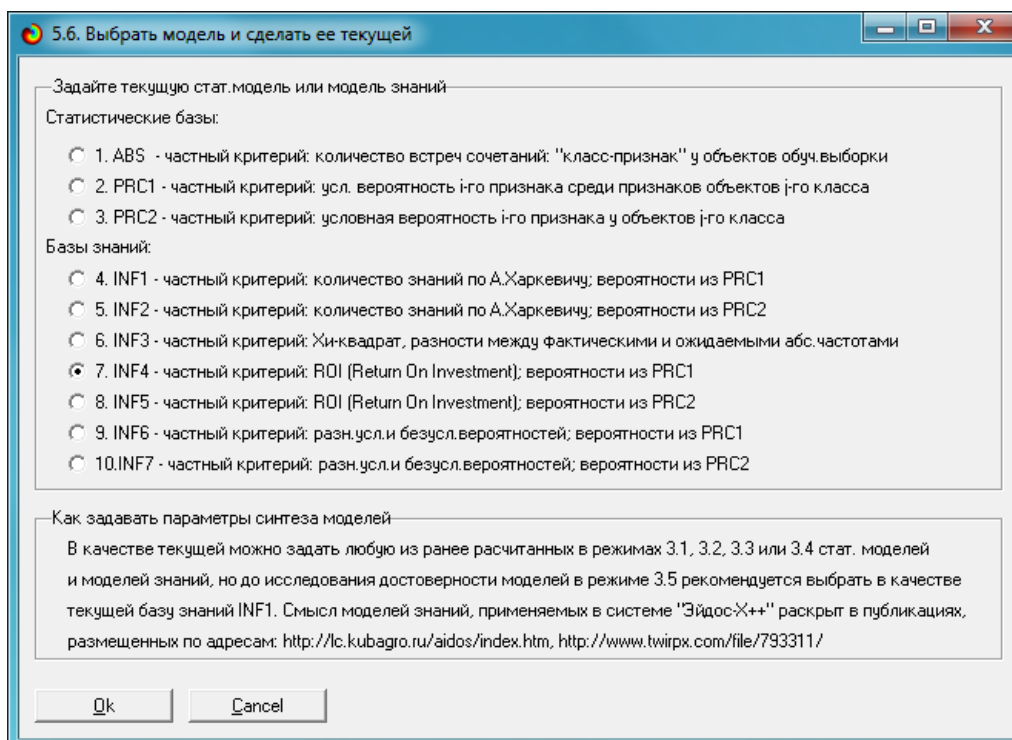


Рисунок 15. Экранные формы режима задания модели в качестве текущей

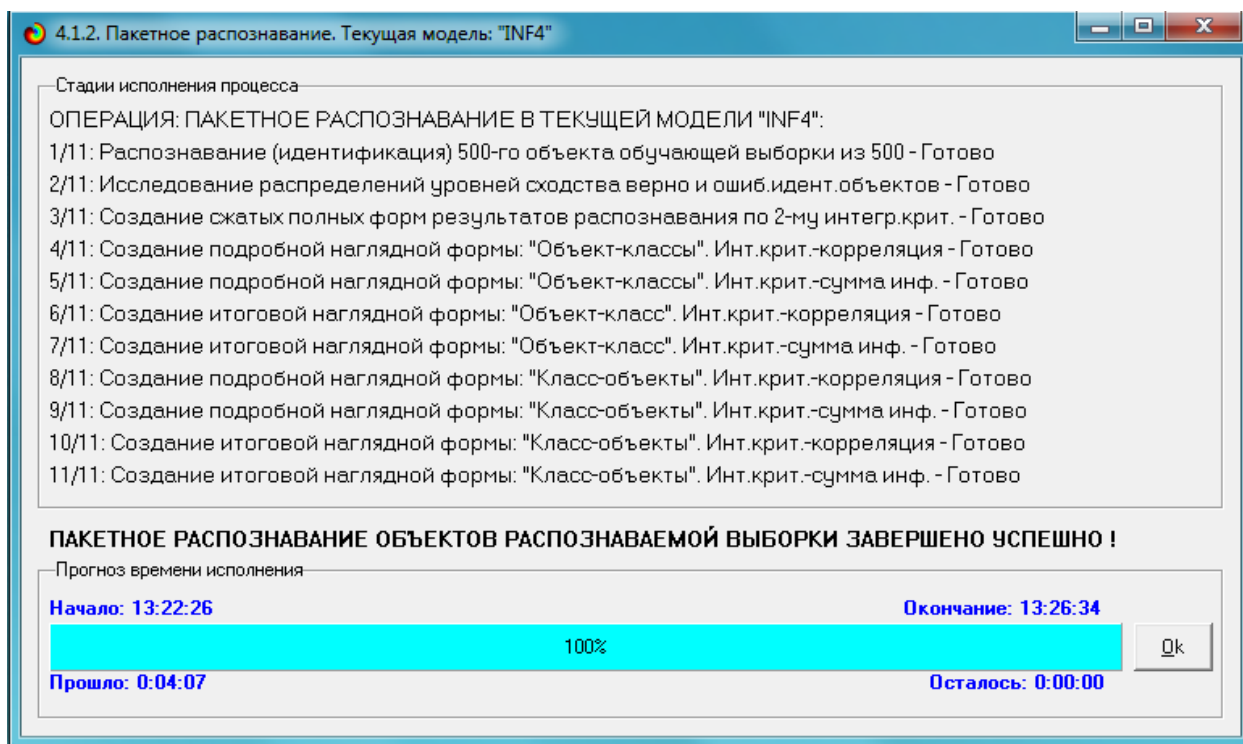


Рисунок 16. Экранная форма режима пакетного распознавания в текущей модели

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:

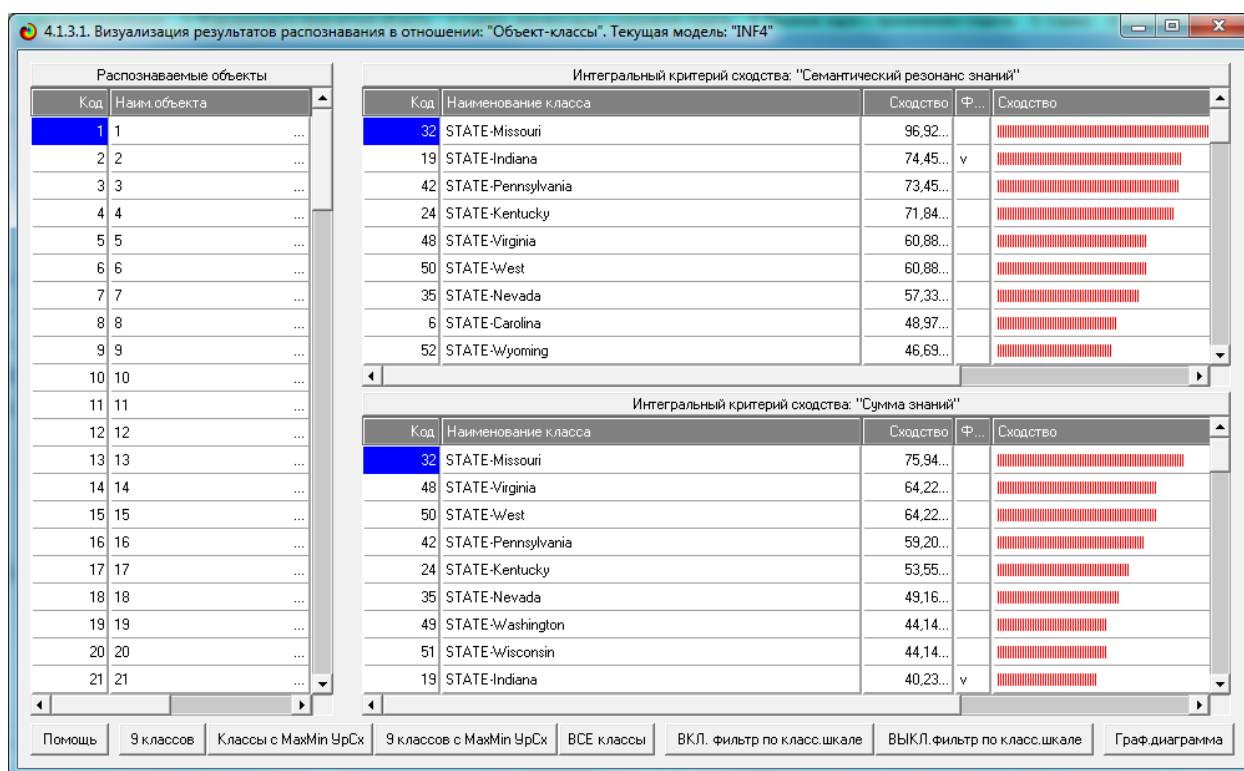


Рисунок 17. Пример идентификации классов в модели INF4

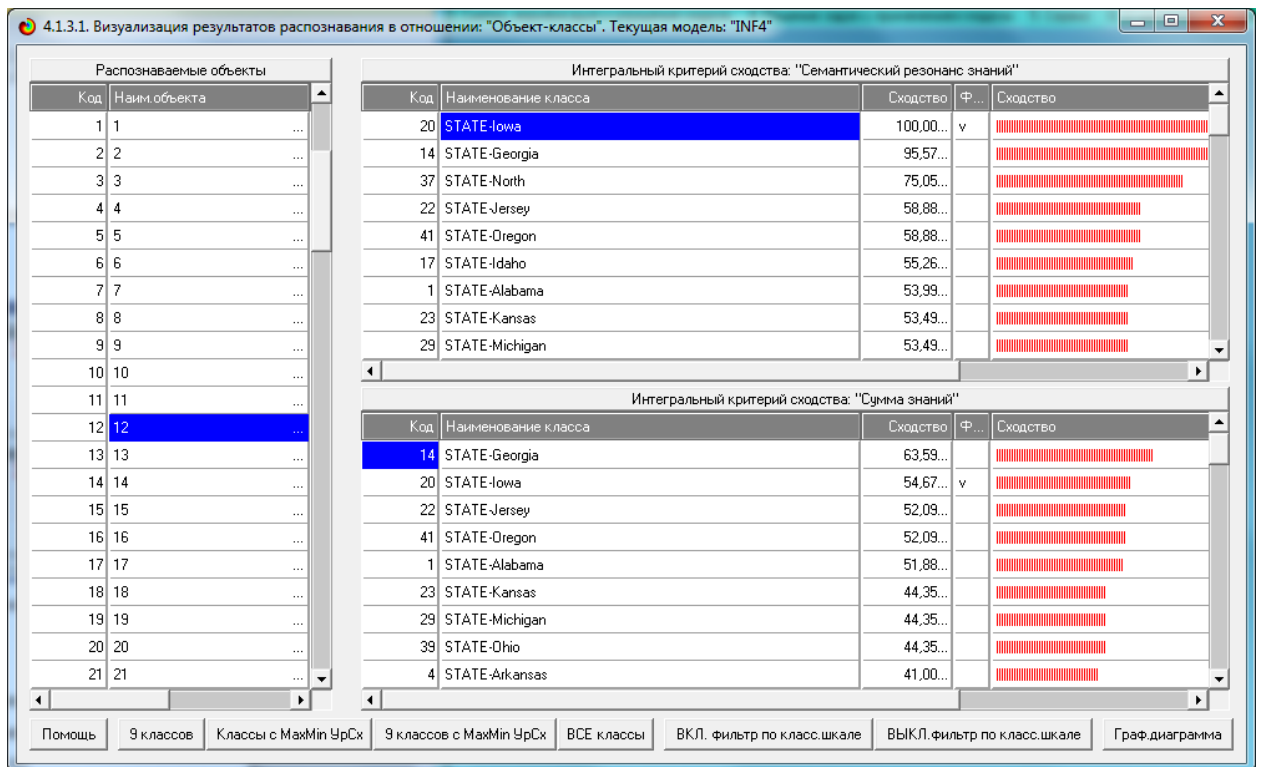


Рисунок 18. Пример идентификации классов в модели INF4

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)

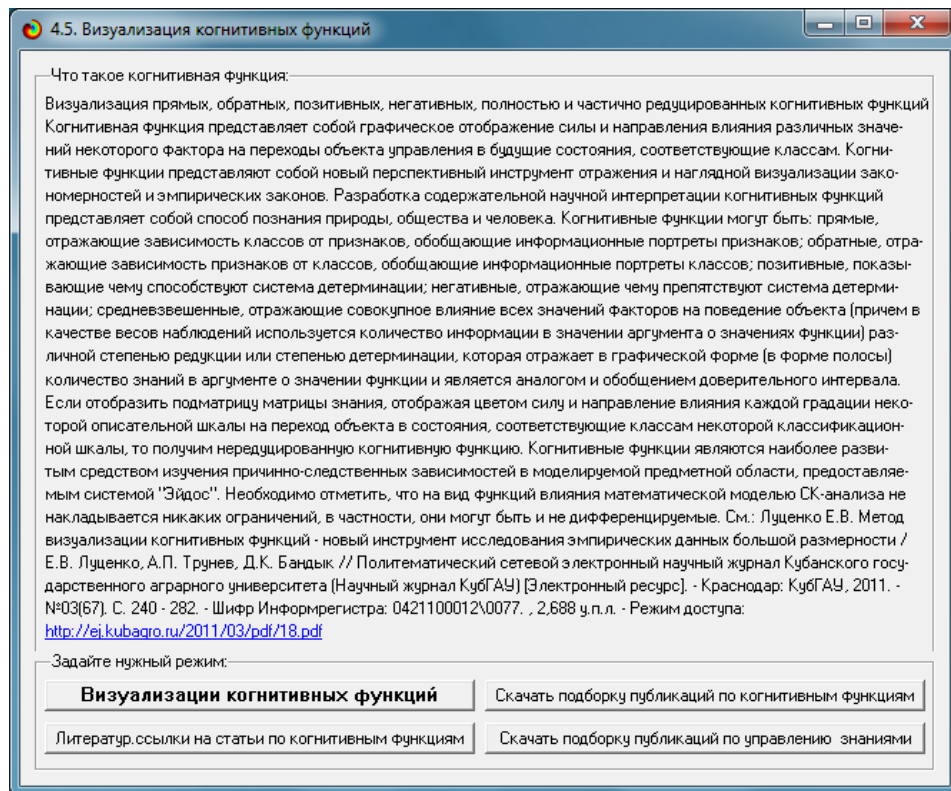


Рисунок 19. Экранная форма режима визуализации когнитивных функций

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].

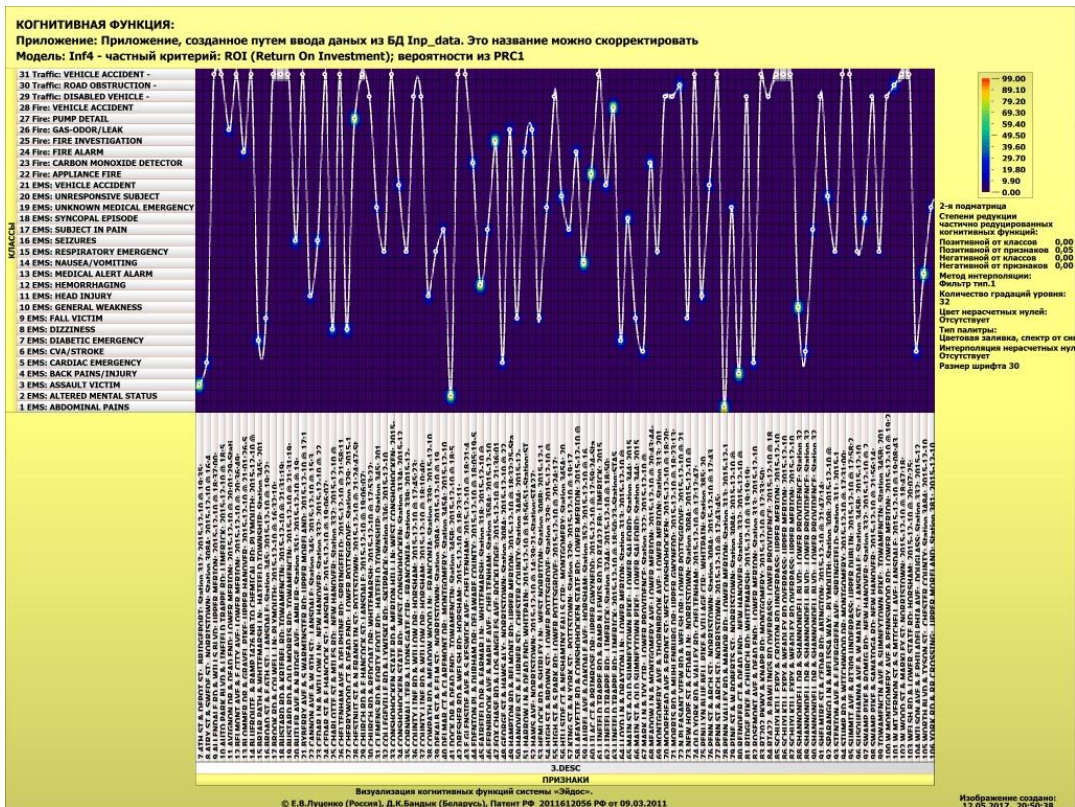
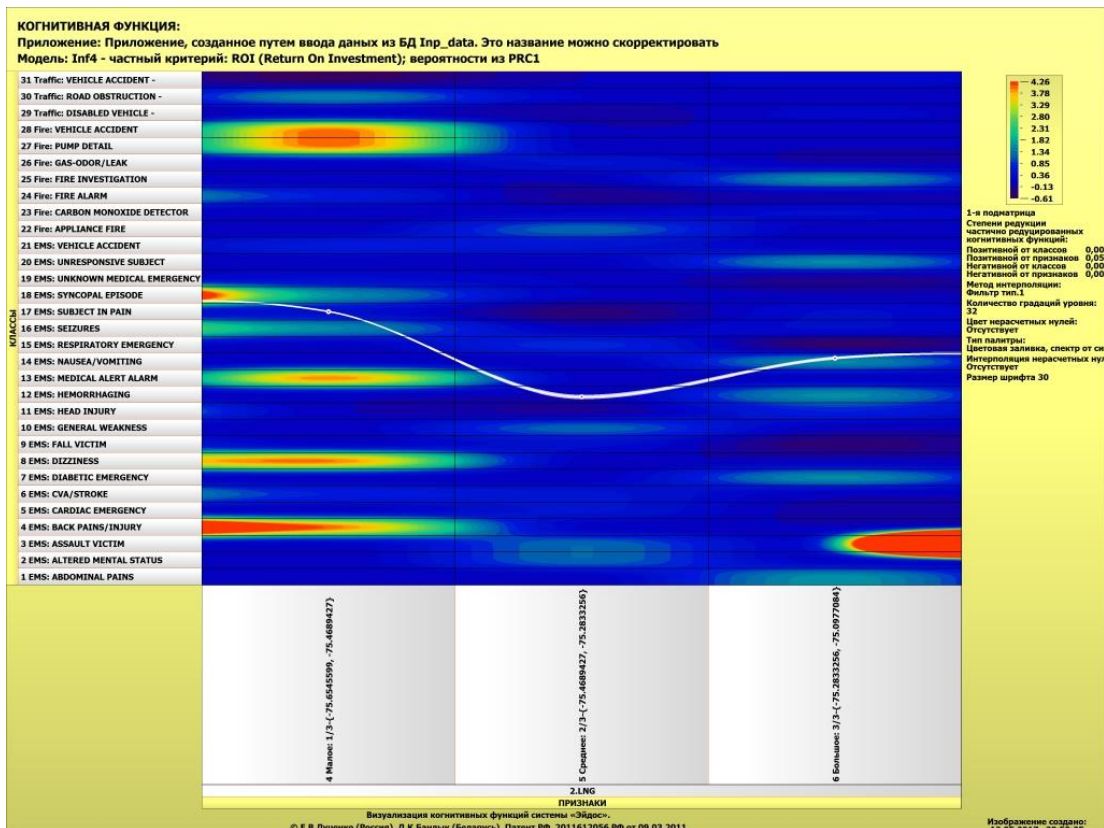


Рисунок 21. Когнитивная функция, отражающая взаимосвязь времени принятия вызова, станции принятия и типа чрезвычайной ситуации



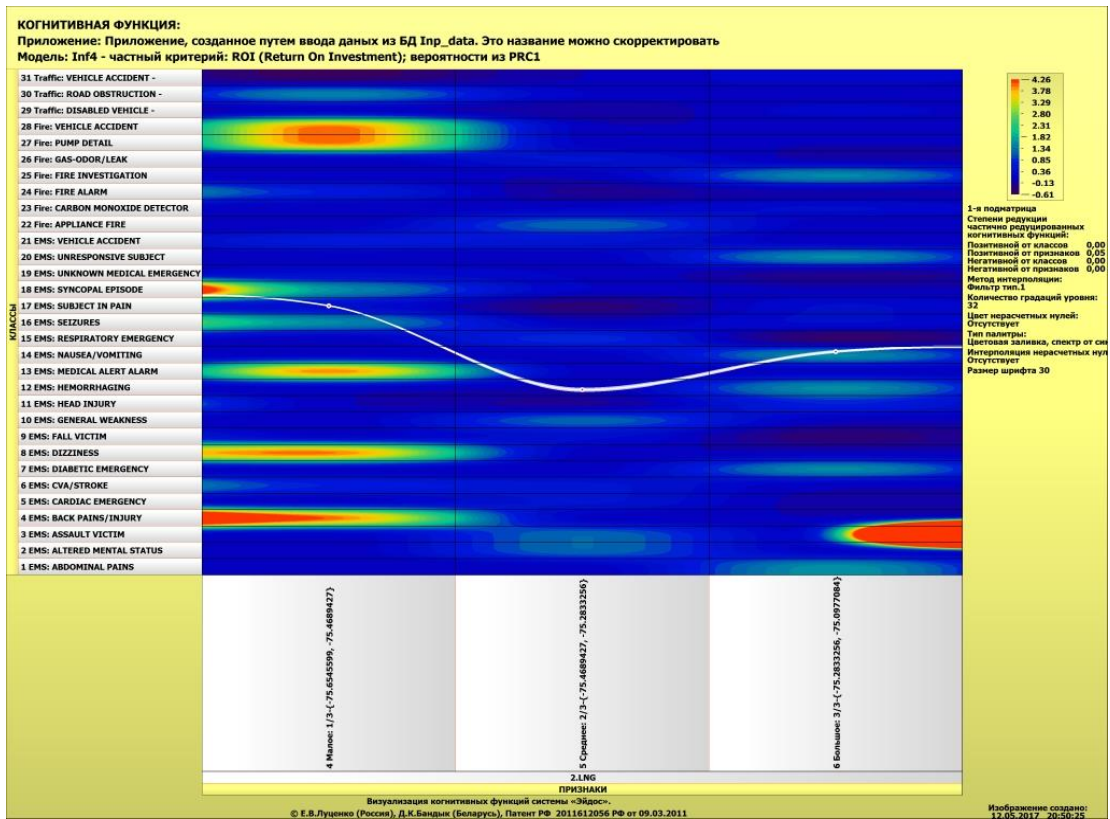


Рисунок 20. Когнитивная функция, отражающая взаимосвязь широты и долготы с типом чрезвычайной ситуации

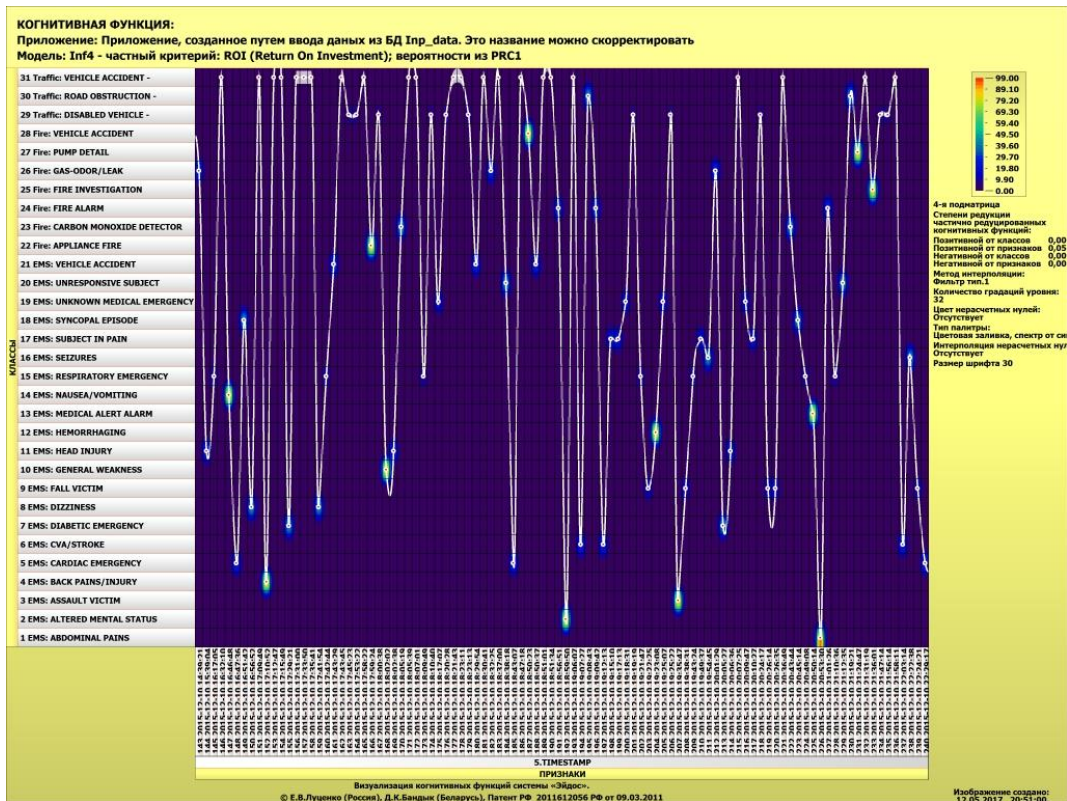


Рисунок 22. Когнитивная функция, отражающая взаимосвязь даты образования чрезвычайной ситуации и типа чрезвычайной ситуации

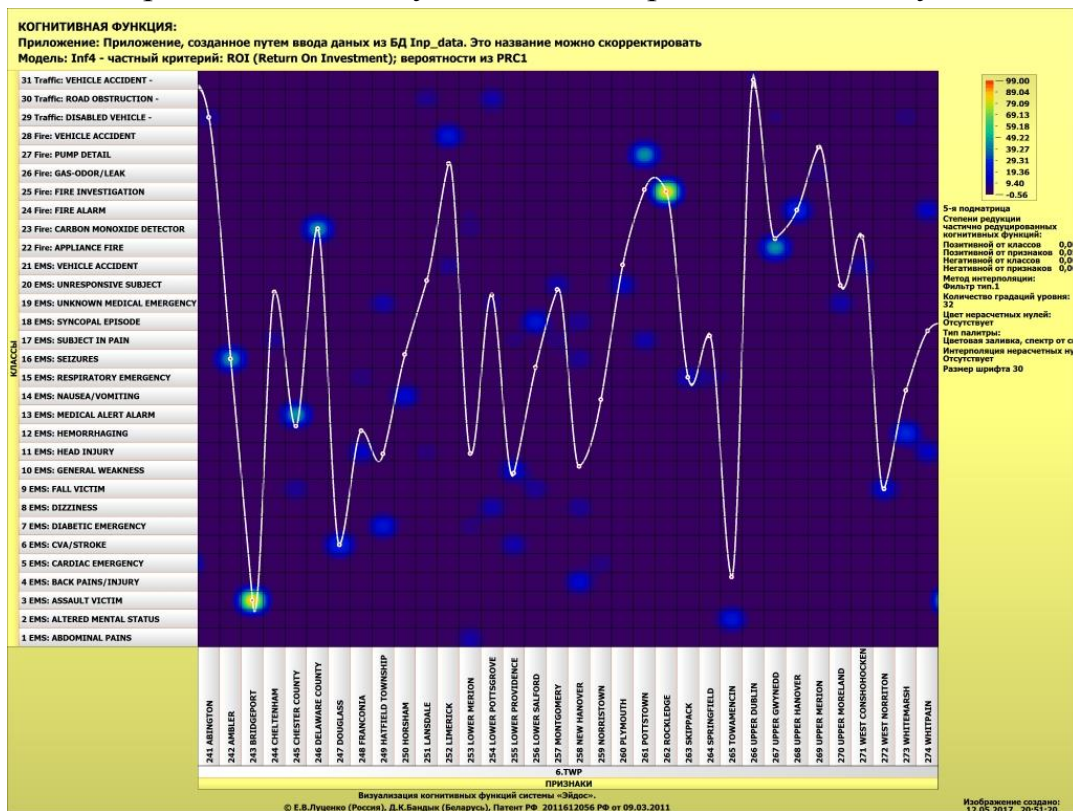


Рисунок 23. Когнитивная функция, отражающая взаимосвязь штата и типа чрезвычайной ситуации

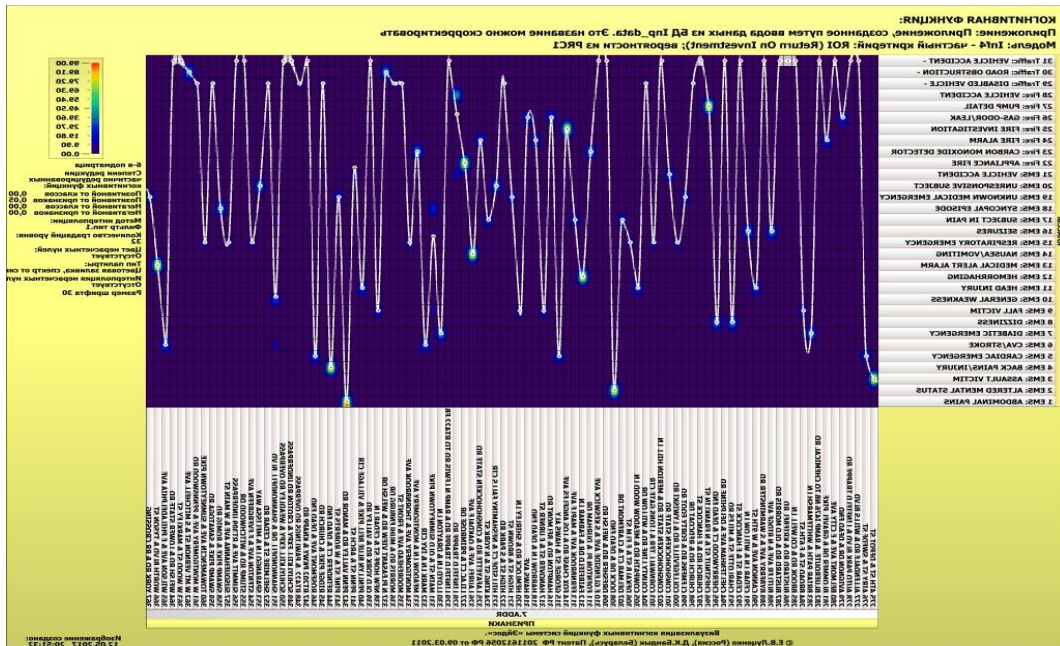


Рисунок 24. Когнитивная функция, отражающая взаимосвязь адреса и типа чрезвычайной ситуации

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 "STATE-Alabama" в модели: 4 "INF1"

Код	Наименование класса	№	Код класса	Наименование класса	Сходство
1	STATE-Alabama	1	1	STATE-Alabama	100.000
2	STATE-Alaska	2	4	STATE-Arkansas	82.823
3	STATE-Arizona	3	10	STATE-Dakota	64.990
4	STATE-Arkansas	4	44	STATE-South	62.037
5	STATE-California	5	14	STATE-Georgia	56.757
6	STATE-Carolina	6	2	STATE-Alaska	51.313
7	STATE-Colorado	7	20	STATE-Iowa	50.360
8	STATE-Columbia	8	18	STATE-Illinois	40.737
9	STATE-Connecticut	9	25	STATE-Louisiana	39.365
10	STATE-Dakota	10	22	STATE-Jersey	33.701
11	STATE-Delaware	11	41	STATE-Oregon	33.701
12	STATE-District	12	6	STATE-Carolina	18.185
13	STATE-Florida	13	8	STATE-Columbia	16.180
14	STATE-Georgia	14	12	STATE-District	16.180
15	STATE-Hampshire	15	38	STATE-of	16.180
16	STATE-Hawaii	16	28	STATE-Mexico	11.340
17	STATE-Idaho	17	37	STATE-North	9.770
18	STATE-Illinois	18	26	STATE-Maine	9.755
19	STATE-Indiana	19	9	STATE-Connecticut	7.682
20	STATE-Iowa	20	17	STATE-Idaho	4.196
21	STATE-Island	21	34	STATE-Nebraska	-4.331
22	STATE-Jersey	22	35	STATE-Nevada	-5.102
23	STATE-Kansas	23	23	STATE-Kansas	-5.387

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 График Вкл. фильтр по кл.шкале Выкл. фильтр по кл.шкале Вписать в окно Показать ВСЕ

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

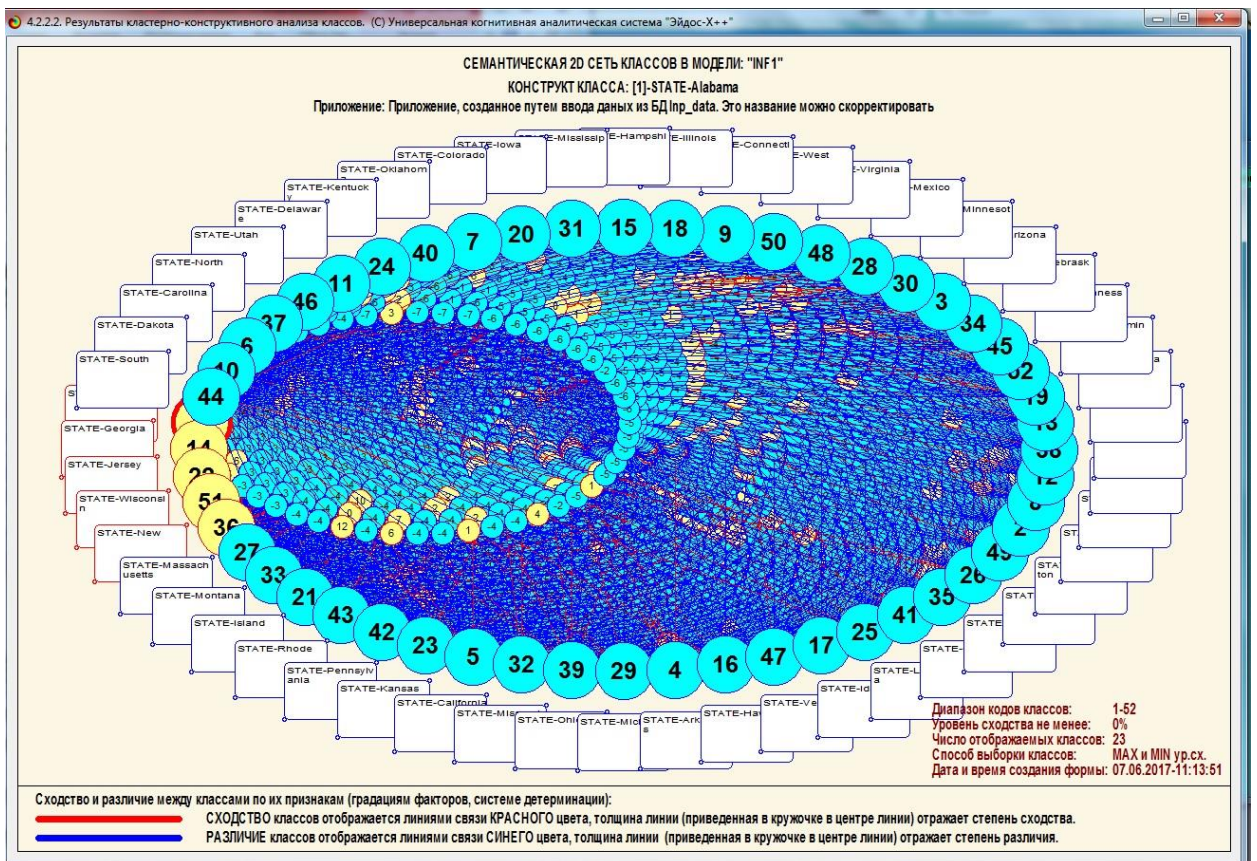


Рисунок 26. Результаты сравнения классов по системе характерных для значений чрезвычайных ситуаций

Из когнитивной диаграммы, приведенной на рисунке 26, мы видим, что, наибольшее количество вызовов по чрезвычайным ситуациям происходит из за более 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.

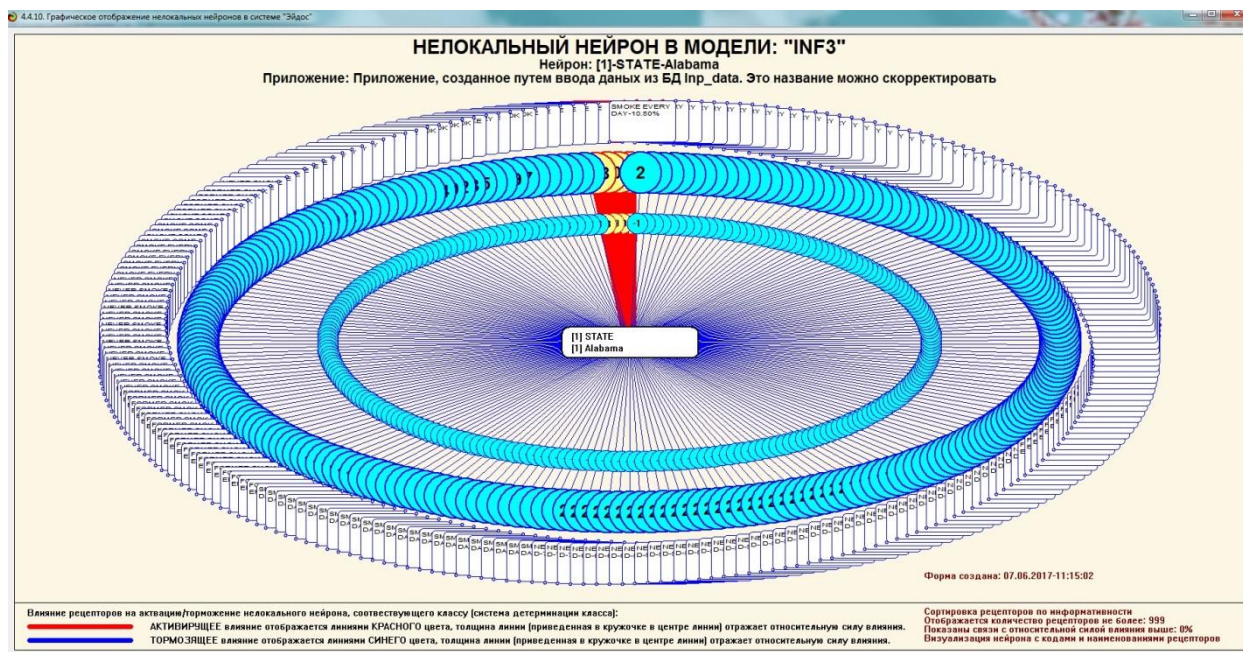


Рисунок 27. Графическое отображение нелокальных нейронов

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].

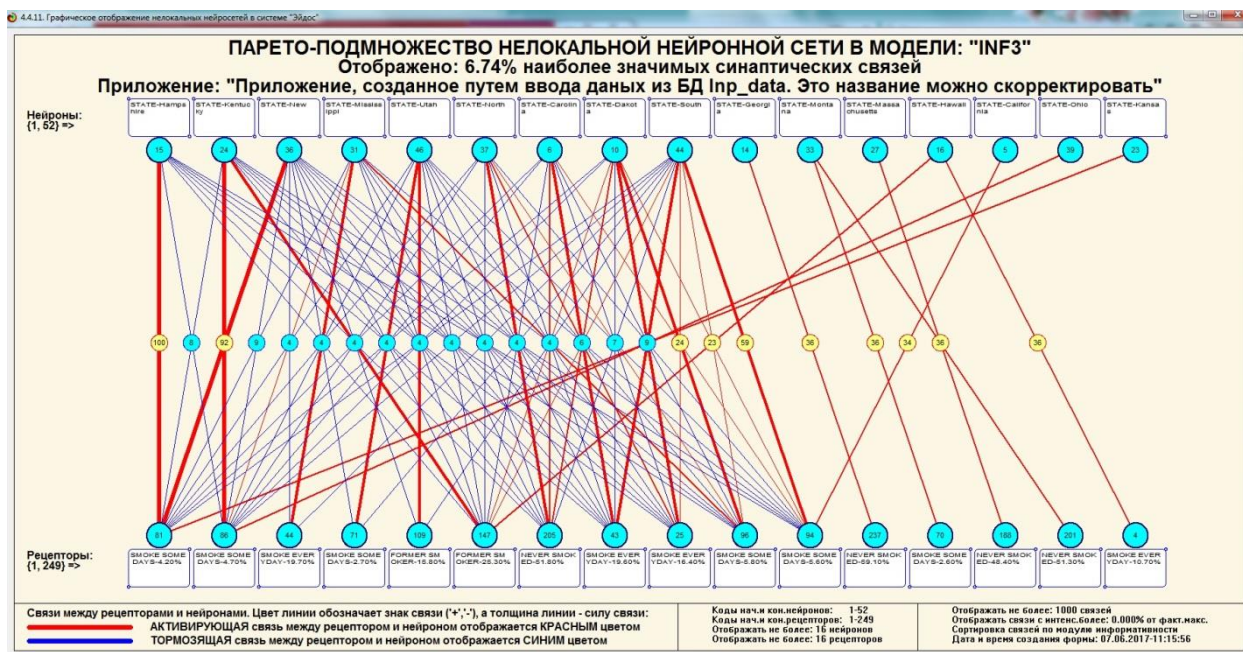


Рисунок 28 – Паретто – подмножеств нелокадной нейронной сети в системе «Эйдос - X»

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).

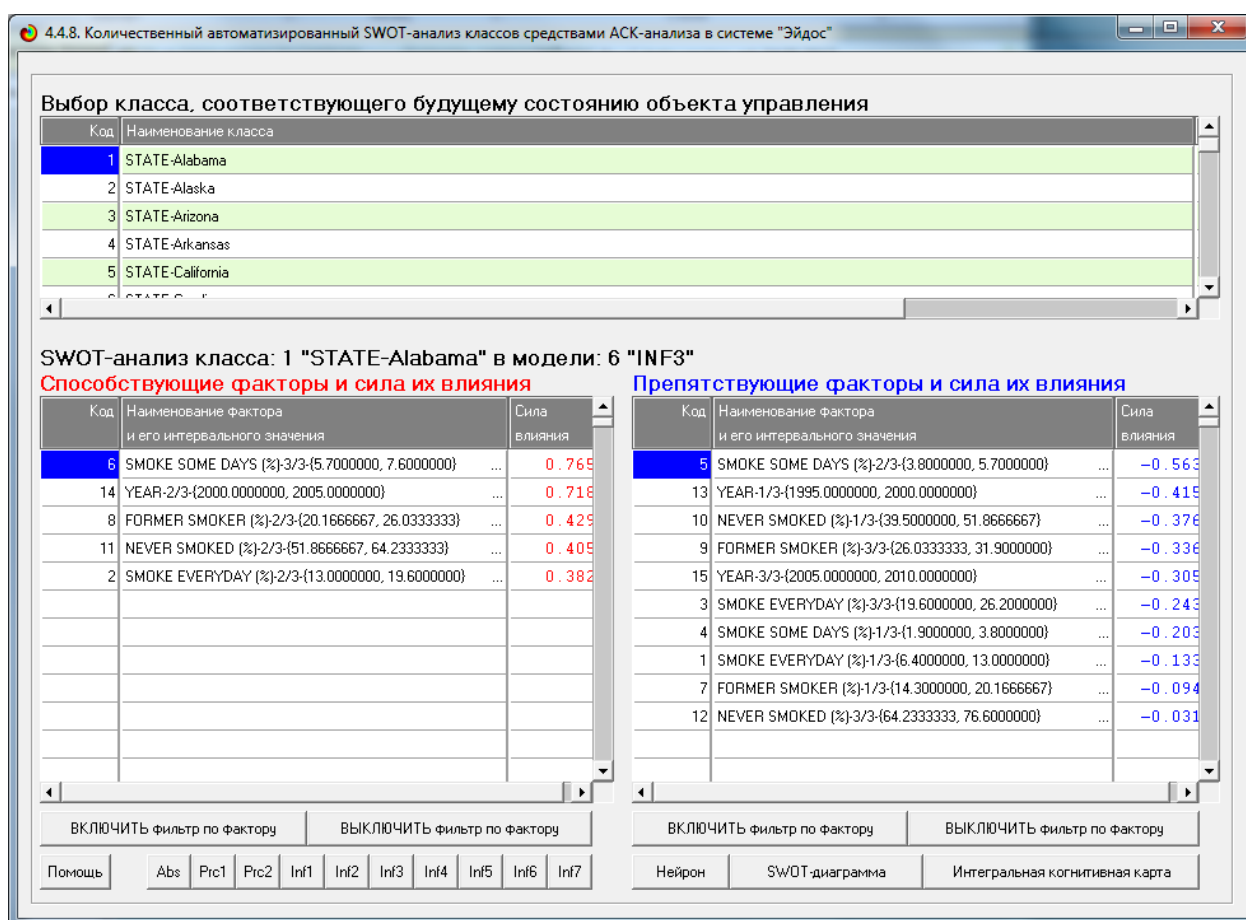


Рисунок 29. Пример SWOT-матрицы в модели INF4

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

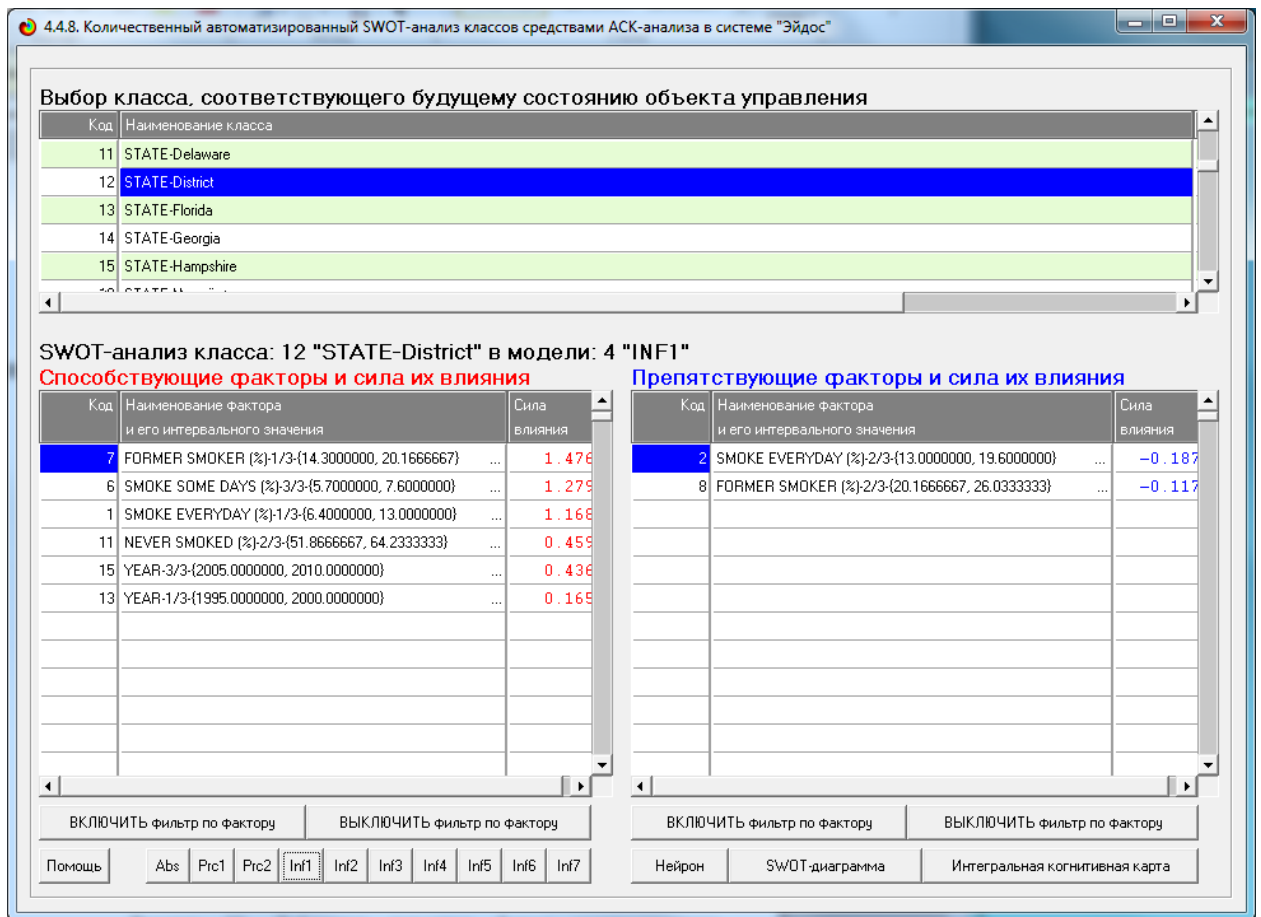


Рисунок 30. Табличная выходная форма количественного автоматизированного SWOT- и PEST-анализа средствами системы «Эйдос»

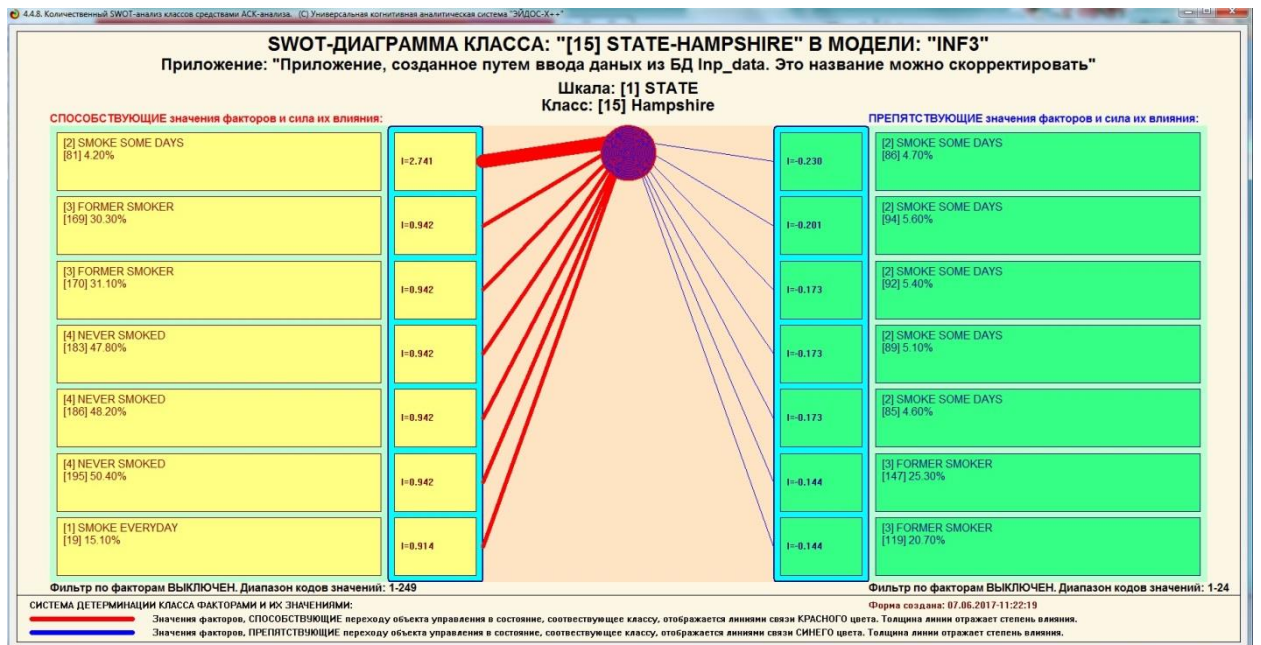


Рисунок 31. Графическая выходная форма количественного автоматизированного SWOT- и PEST-анализа средствами системы «Эйдос»

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.