

Министерство сельского хозяйства Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ И. Т. ТРУБИЛИНА»
Факультет прикладной информатики
Кафедра компьютерных технологий и систем

ЛАБОРАТОРНАЯ РАБОТА

по дисциплине: **Интеллектуальные информационные системы**

на тему: АСК-анализ зависимости персональных затрат на
медицинские услуги от личностных характеристик

Выполнил: студент 3 курса,
группы ПИ1501
Губанов В. А.
Руководитель:
профессор, д.э.н., к.т.н., Луценко Е.В.

Краснодар, 2018

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ.....	4
1.1 Описание решения	4
1.2 Преобразование исходных данных в файл MS Excel.....	4
1.3 Синтез и верификация статистических и интеллектуальных моделей .	9
1.4 Виды моделей системы «Эйдос»	10
1.5 Результаты верификации моделей	12
РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	16
2.1 Решение задачи идентификации	16
2.2 Когнитивные функции.....	19
2.3 SWOT и PEST матрицы и диаграммы	21
2.4 Нелокальные нейронные сети	23
2.5 Кластерный и конструктивный анализ	24
ЗАКЛЮЧЕНИЕ	25
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	26

ВВЕДЕНИЕ

Одним из важнейших и перспективных направлений в развитии современных информационных технологий, на сегодняшний день, является создание систем искусственного интеллекта. В результате развития технологий в данном направлении, возникла необходимость оценки качества математических моделей этих систем.

В данной работе рассмотрено решение задачи выявления зависимости состояния здоровья от персональных затрат на медицинские услуги. При выполнении работы нужно учесть, что необходимы свободный доступ к тестовым исходным данным и методика, которая необходима для преобразования этих данных в форму, которая необходима для работы в системе искусственного интеллекта. Среди баз данных текстовых задач для систем искусственного интеллекта удачным выбором является репозиторий «Kaggle». В данной лабораторной работе использована база данных «Medical Cost Personal Datasets» из банка исходных данных по задачам искусственного интеллекта – репозитория «Kaggle». Для решения задачи используем стандартные возможности Microsoft Office Excel, блокнот, а также универсальную когнитивную аналитическую систему искусственного интеллекта «Aidos- X++».

СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

1.1 Описание решения

Порядок решения задачи в соответствии с методологией АСК-анализа:

1. Преобразование базы исходных данных в необходимый для системы формат файла MS Excel.
2. Преобразование исходных данных из файла 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 в файл MS Excel

Из электронного ресурса баз данных «Kaggle» возьмем базу данных «Medical Cost Personal Datasets», которую приведем к виду на рисунке 1.

	A	B	C	D	E	F	G	H
1	Номер	Возраст	Траты	Пол	Индекс массы тела	Дети	Курение	Регион
2	1	19	16884,92	female	27,90	0	yes	southwest
3	2	18	1725,55	male	33,77	1	no	southeast
4	3	28	4449,46	male	33	3	no	southeast
5	4	33	21984,47	male	22,71	0	no	northwest
6	5	32	3866,86	male	28,88	0	no	northwest
7	6	31	3756,62	female	25,74	0	no	southeast
8	7	46	8240,5896	female	33,44	1	no	southeast
9	8	37	7281,5056	female	27,74	3	no	northwest
10	9	37	6406,4107	male	29,83	2	no	northeast
11	10	60	28923,13692	female	25,84	0	no	northwest
12	11	25	2721,3208	male	26,22	0	no	northeast
13	12	62	27808,7251	female	26,29	0	yes	southeast
14	13	23	1826,843	male	34,40	0	no	southwest
15	14	56	11090,7178	female	39,82	0	no	southeast
16	15	27	39611,7577	male	42,13	0	yes	southeast
17	16	19	1837,237	male	24,60	1	no	southwest
18	17	52	10797,3362	female	30,78	1	no	northeast
19	18	23	2395,17155	male	23,85	0	no	northeast
20	19	56	10602,385	male	40,30	0	no	southwest
21	20	30	36837,467	male	35,30	0	yes	southwest
22	21	60	13228,84695	female	36,01	0	no	northeast
23	22	30	4149,736	female	32,40	1	no	southwest
24	23	18	1137,011	male	34,10	0	no	southeast

Рисунок 1 - Исходные данные в файле MS Excel

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 таблицу 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 «А» 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.

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

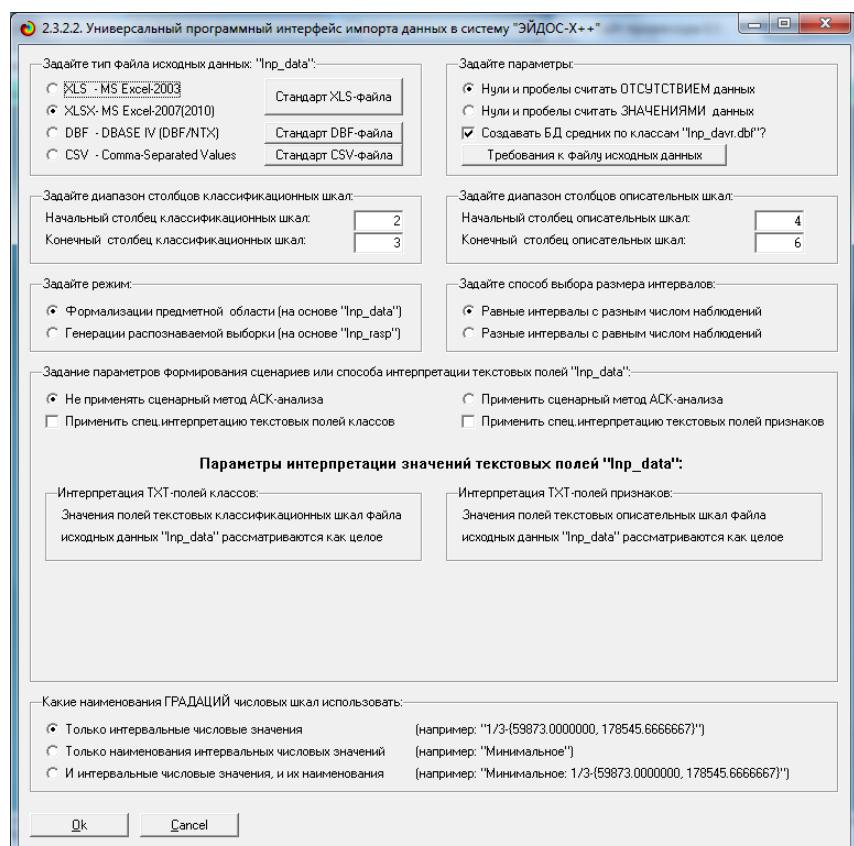


Рисунок 2 - Экранная форма Универсального программного интерфейса импорта данных в систему “Эйдос” (режим 2.3.2.2)

В экранной форме, приведенной на рисунке 2, необходимо задать настройки:

- «Тип файла исходных данных Inp_data»: «XLS – MS Excel2003»;
- «Диапазон столбцов классификационных шкал»: «Начальный столбец классификационных шкал» - 2, «Конечный столбец классификационных шкал» - 3;
- «Диапазон столбцов описательных шкал»: «Начальный столбец описательных шкал» - 4, «Конечный столбец описательных шкал» - 8;
- «Параметры формирования сценариев или способа интерпретации текстовых полей»: «Не применять сценарный метод АСК - анализа и спец.интерпретацию TXT-полей».

После нажать кнопку «OK». После нажатия кнопки открывается окно, где размещена информация о размерности модели. В этом окне необходимо нажать кнопку «Выйти на создание модели», что показано на рисунке 3.

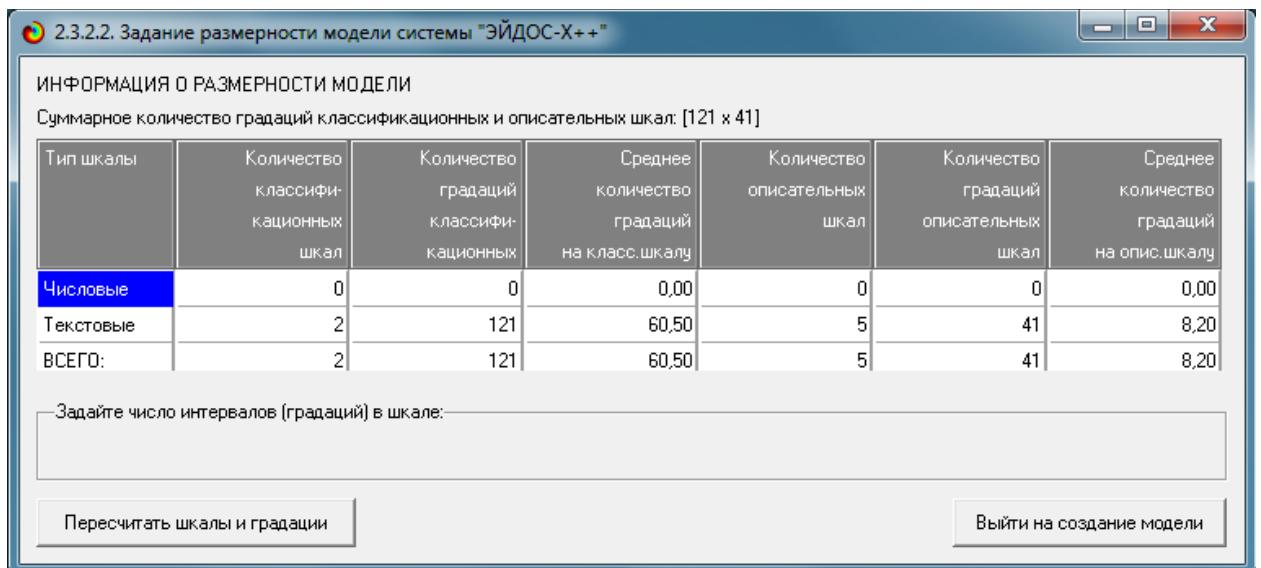


Рисунок 1- Задание размерности модели системы «Эйдос»

Далее следует открытие окна, которое отображает стадию процесса импорта данных из внешней БД «Inp_data.xls» в систему «Эйдос», а также прогноз времени завершения этого процесса. В данном окне необходимо дождаться завершения формализации предметной области и нажать кнопку «OK», как показано на рисунке 4.

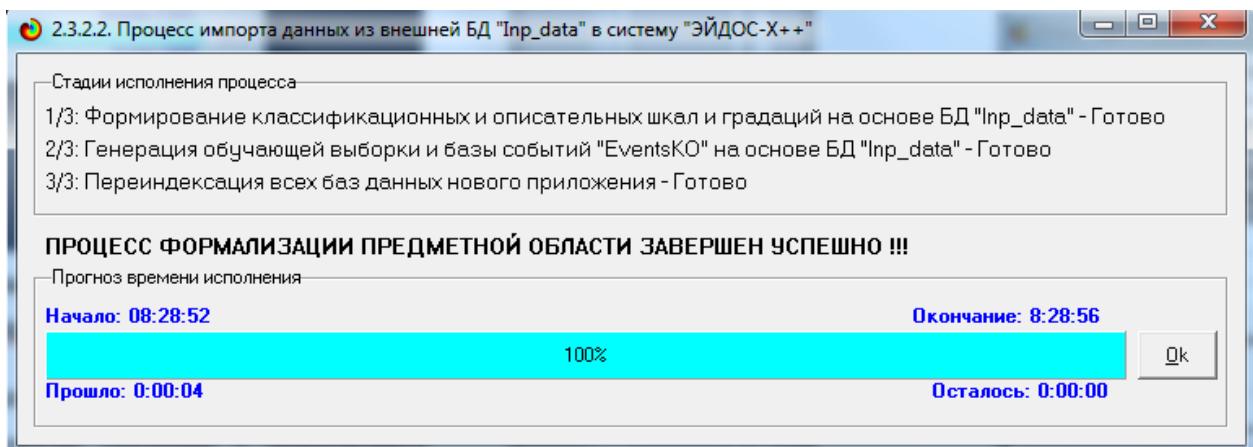


Рисунок 2- Процесс импорта данных из внешней БД «Inp_data.xls» в систему «Эйдос»

В результате выполненных действий формируются классификационные и описательные шкалы, градации, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Эти действия полностью автоматизированы и выполняются на 2-м этапе АСК - анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1, рисунок 5.

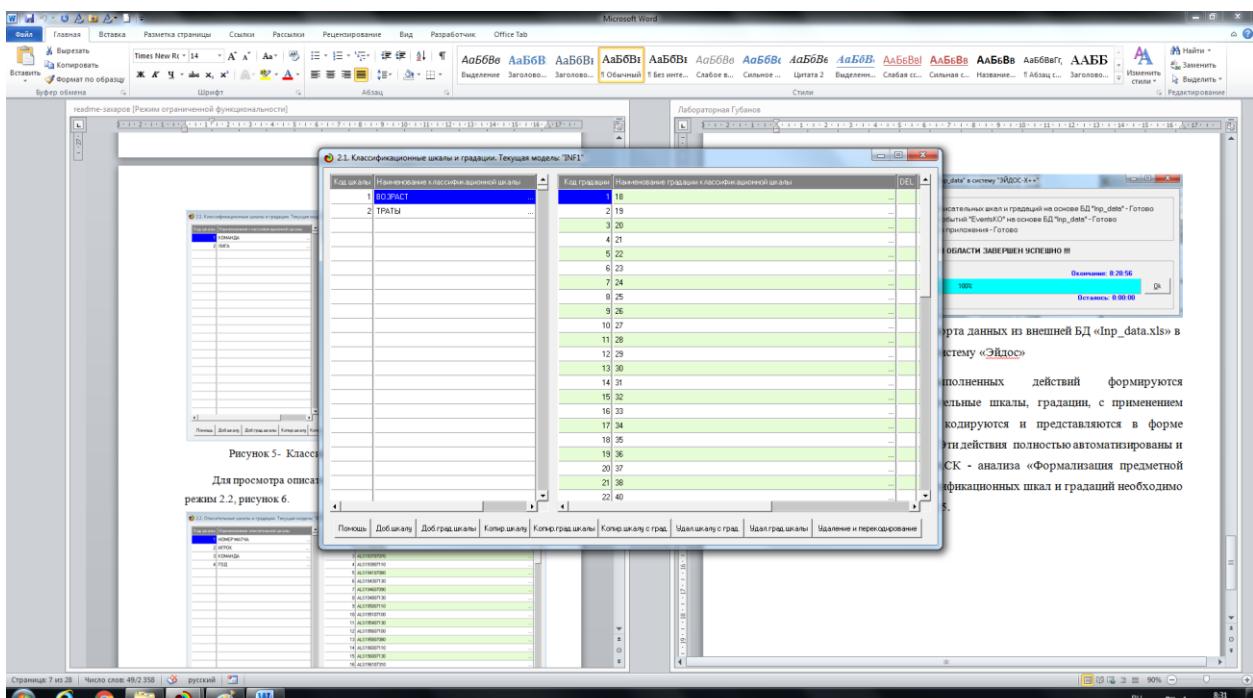


Рисунок 3- Классификационные шкалы и градации (фрагмент)

Для просмотра описательных шкал и градаций необходимо запустить режим 2.2, рисунок 6.

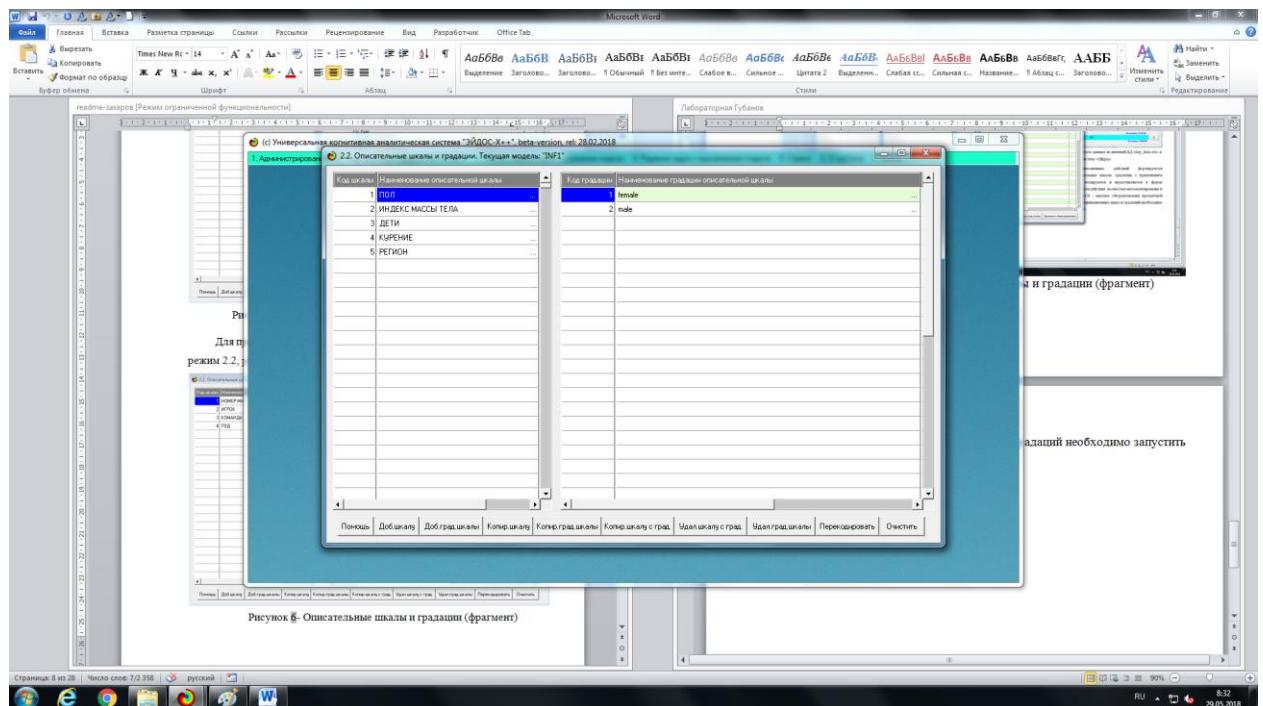


Рисунок 4- Описательные шкалы и градации (фрагмент)

Для просмотра обучающей выборки необходимо запустить режим 2.3.1., рисунок 7.

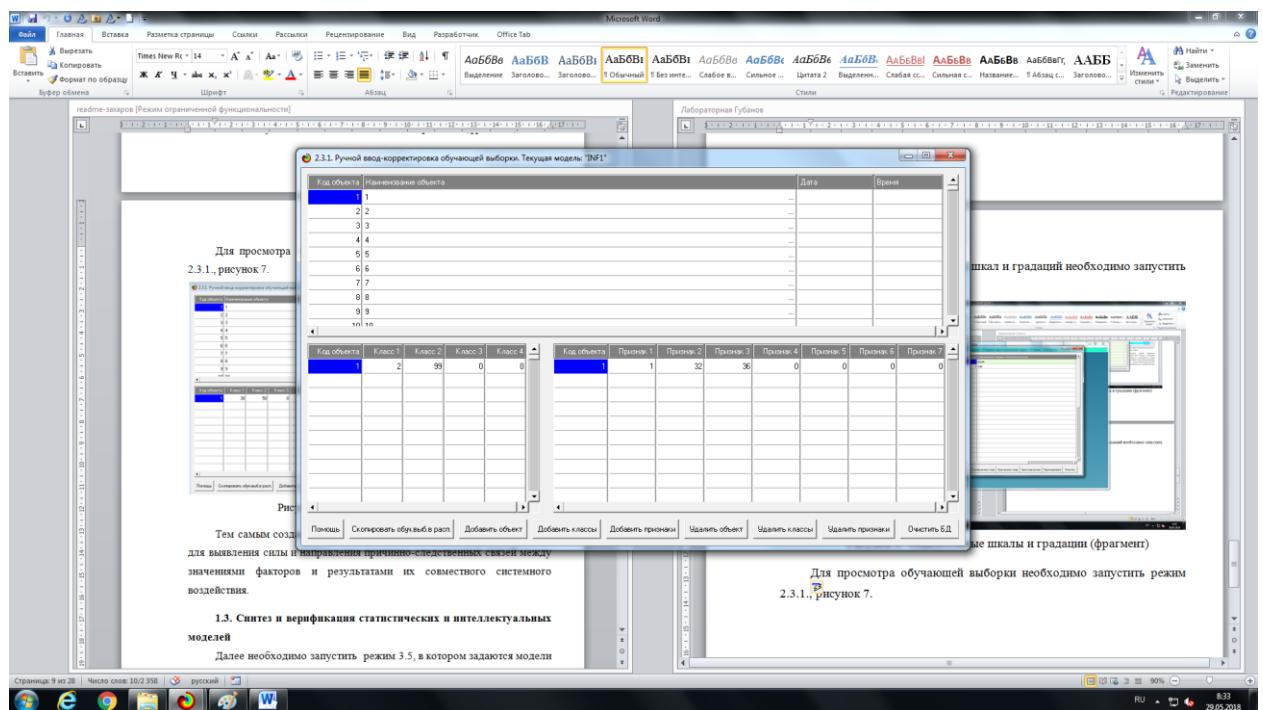


Рисунок 5- Обучающая выборка (фрагмент)

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.

1.3 Синтез и верификация статистических и интеллектуальных моделей

Далее необходимо запустить режим 3.5, в котором задаются модели для синтеза и верификации, а также задается модель, которой по окончании режима присваивается статус текущей, что показано на рисунке 8.

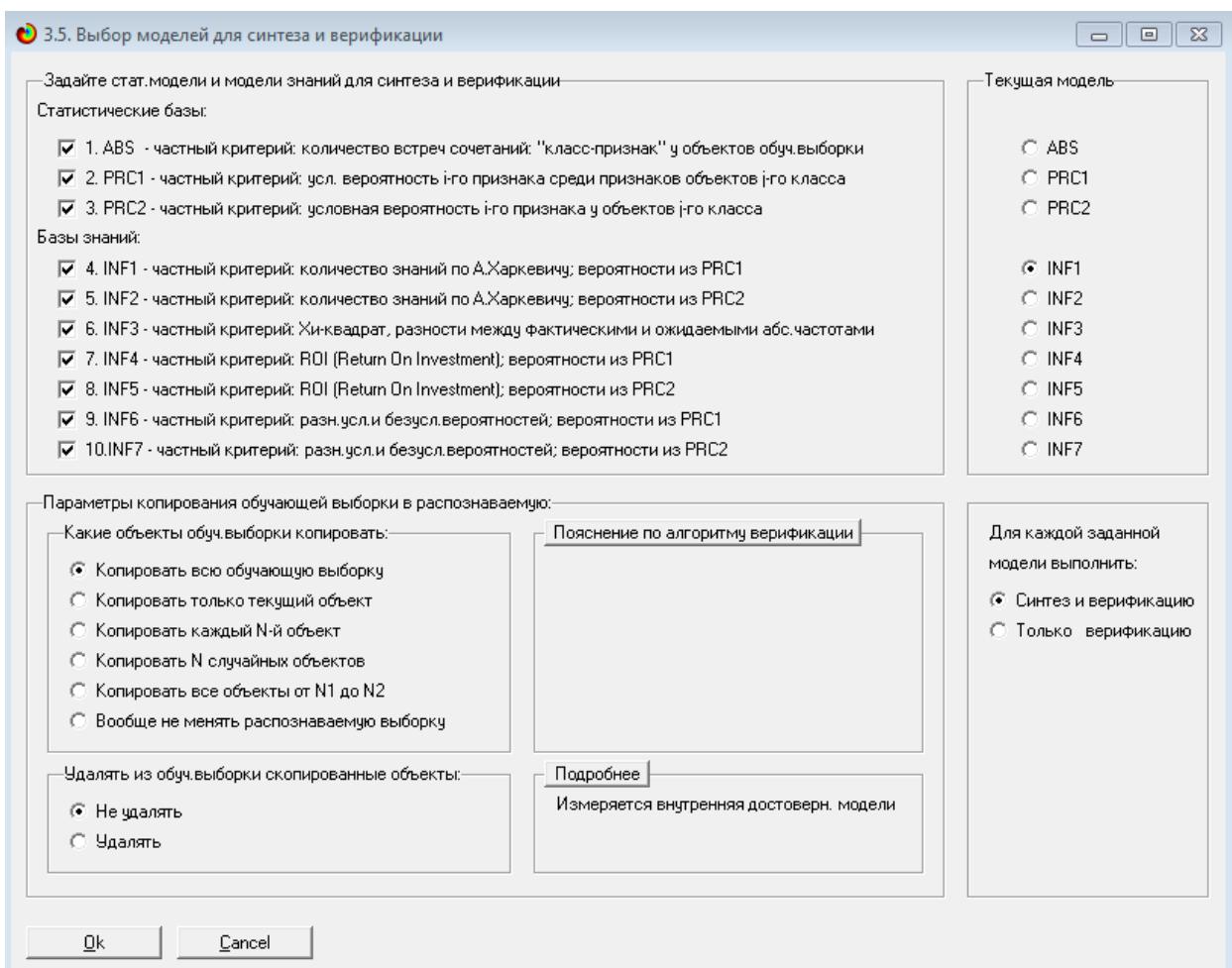


Рисунок 6 - Выбор моделей для синтеза и верификации, а также текущей модели

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 8. Стадия

процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 9.

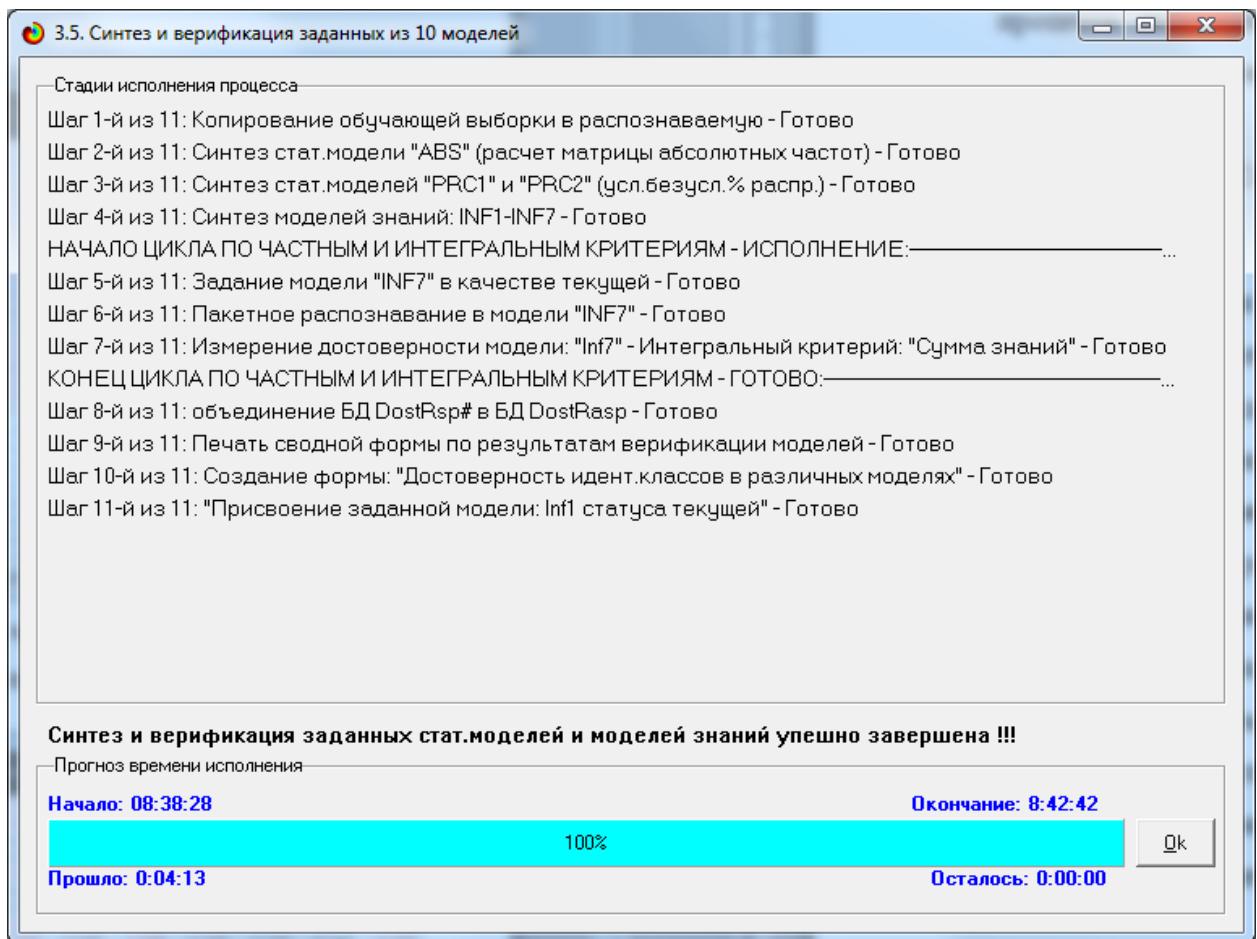


Рисунок 7- Синтез и верификация статистических моделей и моделей знаний

Верификация (оценка достоверности моделей) проводилась на всех примерах наблюдения из обучающей выборки. В результате выполнения режима 3.5 созданы все модели, со всеми частными критериями, перечисленными на рисунке 8.

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 в матрицы условных и безусловных процентных распределений, и матрицы знаний.

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обучивборки"												
Код признака	Наименование описательной шкалы и градации	1. ВОЗРАСТ 18	2. ВОЗРАСТ 19	3. ВОЗРАСТ 20	4. ВОЗРАСТ 21	5. ВОЗРАСТ 22	6. ВОЗРАСТ 23	7. ВОЗРАСТ 24	8. ВОЗРАСТ 25	9. ВОЗРАСТ 26	10. ВОЗРАСТ 27	11. ВОЗРАСТ 28
1	ПОЛ-female	3	3	1	1	1			1		1	2
2	ПОЛ-male	3	2		1	1	3		2	1	1	3
3	ИНДЕКС МАССЫ ТЕЛА-17											
4	ИНДЕКС МАССЫ ТЕЛА-20											
5	ИНДЕКС МАССЫ ТЕЛА-21											
6	ИНДЕКС МАССЫ ТЕЛА-22											
7	ИНДЕКС МАССЫ ТЕЛА-23											
8	ИНДЕКС МАССЫ ТЕЛА-24											
9	ИНДЕКС МАССЫ ТЕЛА-25											
10	ИНДЕКС МАССЫ ТЕЛА-26											
11	ИНДЕКС МАССЫ ТЕЛА-27											
12	ИНДЕКС МАССЫ ТЕЛА-28											
13	ИНДЕКС МАССЫ ТЕЛА-29											
14	ИНДЕКС МАССЫ ТЕЛА-30											
15	ИНДЕКС МАССЫ ТЕЛА-31											
16	ИНДЕКС МАССЫ ТЕЛА-32											
17	ИНДЕКС МАССЫ ТЕЛА-33											
18	ИНДЕКС МАССЫ ТЕЛА-34											
19	ИНДЕКС МАССЫ ТЕЛА-35											
20	ИНДЕКС МАССЫ ТЕЛА-36											
21	ИНДЕКС МАССЫ ТЕЛА-37											
22	ИНДЕКС МАССЫ ТЕЛА-39											
23	ИНДЕКС МАССЫ ТЕЛА-40											
24	ИНДЕКС МАССЫ ТЕЛА-42											

Рисунок 8- Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений (фрагмент)

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"												
Код признака	Наименование описательной шкалы и градации	1. ВОЗРАСТ 18	2. ВОЗРАСТ 19	3. ВОЗРАСТ 20	4. ВОЗРАСТ 21	5. ВОЗРАСТ 22	6. ВОЗРАСТ 23	7. ВОЗРАСТ 24	8. ВОЗРАСТ 25	9. ВОЗРАСТ 26	10. ВОЗРАСТ 27	11. ВОЗРАСТ 28
1	ПОЛ-female	0.028	0.232	0.802	0.028	0.028			0.802			0.028
2	ПОЛ-male	-0.028	-0.277		-0.028	-0.028	0.747			0.747	0.747	-0.028
3	ИНДЕКС МАССЫ ТЕЛА-17											
4	ИНДЕКС МАССЫ ТЕЛА-20											
5	ИНДЕКС МАССЫ ТЕЛА-21											
6	ИНДЕКС МАССЫ ТЕЛА-22											
7	ИНДЕКС МАССЫ ТЕЛА-23											
8	ИНДЕКС МАССЫ ТЕЛА-24											
9	ИНДЕКС МАССЫ ТЕЛА-25											
10	ИНДЕКС МАССЫ ТЕЛА-26											
11	ИНДЕКС МАССЫ ТЕЛА-27											
12	ИНДЕКС МАССЫ ТЕЛА-28											
13	ИНДЕКС МАССЫ ТЕЛА-29											
14	ИНДЕКС МАССЫ ТЕЛА-30											
15	ИНДЕКС МАССЫ ТЕЛА-31											
16	ИНДЕКС МАССЫ ТЕЛА-32											
17	ИНДЕКС МАССЫ ТЕЛА-33											
18	ИНДЕКС МАССЫ ТЕЛА-34											
19	ИНДЕКС МАССЫ ТЕЛА-35											
20	ИНДЕКС МАССЫ ТЕЛА-36											
21	ИНДЕКС МАССЫ ТЕЛА-37											
22	ИНДЕКС МАССЫ ТЕЛА-39											
23	ИНДЕКС МАССЫ ТЕЛА-40											
24	ИНДЕКС МАССЫ ТЕЛА-42											

Рисунок 9- Матрица информативности (модель INF1) в битах (фрагмент)

Код признака	Наименование описательной шкалы и градации	1. ВОЗРАСТ 18	2. ВОЗРАСТ 19	3. ВОЗРАСТ 20	4. ВОЗРАСТ 21	5. ВОЗРАСТ 22	6. ВОЗРАСТ 23	7. ВОЗРАСТ 24	8. ВОЗРАСТ 25	9. ВОЗРАСТ 26	10. ВОЗРАСТ 27	11. ВОЗРАСТ 28
1	ПОЛ-female	0.075	0.563	0.513	0.025	0.025	-1.463	0.513	-0.975	-0.488	0.025	
2	ПОЛ-male	-0.075	-0.563	-0.513	-0.025	-0.025	1.463	-0.513	0.975	0.488	-0.025	
3	ИНДЕКС МАССЫ ТЕЛА-17											
4	ИНДЕКС МАССЫ ТЕЛА-20											
5	ИНДЕКС МАССЫ ТЕЛА-21											
6	ИНДЕКС МАССЫ ТЕЛА-22											
7	ИНДЕКС МАССЫ ТЕЛА-23											
8	ИНДЕКС МАССЫ ТЕЛА-24											
9	ИНДЕКС МАССЫ ТЕЛА-25											
10	ИНДЕКС МАССЫ ТЕЛА-26											
11	ИНДЕКС МАССЫ ТЕЛА-27											
12	ИНДЕКС МАССЫ ТЕЛА-28											
13	ИНДЕКС МАССЫ ТЕЛА-29											
14	ИНДЕКС МАССЫ ТЕЛА-30											
15	ИНДЕКС МАССЫ ТЕЛА-31											
16	ИНДЕКС МАССЫ ТЕЛА-32											
17	ИНДЕКС МАССЫ ТЕЛА-33											
18	ИНДЕКС МАССЫ ТЕЛА-34											
19	ИНДЕКС МАССЫ ТЕЛА-35											
20	ИНДЕКС МАССЫ ТЕЛА-36											
21	ИНДЕКС МАССЫ ТЕЛА-37											
22	ИНДЕКС МАССЫ ТЕЛА-39											
23	ИНДЕКС МАССЫ ТЕЛА-40											
24	ИНДЕКС МАССЫ ТЕЛА-42											

Рисунок 10- Матрица знаний (модель INF3) (фрагмент)

Код признака	Наименование описательной шкалы и градации	1. ВОЗРАСТ 18	2. ВОЗРАСТ 19	3. ВОЗРАСТ 20	4. ВОЗРАСТ 21	5. ВОЗРАСТ 22	6. ВОЗРАСТ 23	7. ВОЗРАСТ 24	8. ВОЗРАСТ 25	9. ВОЗРАСТ 26	10. ВОЗРАСТ 27	11. ВОЗРАСТ 28
1	ПОЛ-female	0.026	0.231	1.051	0.026	0.026		1.051				0.026
2	ПОЛ-male	-0.024	-0.220		-0.024	-0.024	0.951		0.951	0.951	0.951	-0.024
3	ИНДЕКС МАССЫ ТЕЛА-17											
4	ИНДЕКС МАССЫ ТЕЛА-20											
5	ИНДЕКС МАССЫ ТЕЛА-21											
6	ИНДЕКС МАССЫ ТЕЛА-22											
7	ИНДЕКС МАССЫ ТЕЛА-23											
8	ИНДЕКС МАССЫ ТЕЛА-24											
9	ИНДЕКС МАССЫ ТЕЛА-25											
10	ИНДЕКС МАССЫ ТЕЛА-26											
11	ИНДЕКС МАССЫ ТЕЛА-27											
12	ИНДЕКС МАССЫ ТЕЛА-28											
13	ИНДЕКС МАССЫ ТЕЛА-29											
14	ИНДЕКС МАССЫ ТЕЛА-30											
15	ИНДЕКС МАССЫ ТЕЛА-31											
16	ИНДЕКС МАССЫ ТЕЛА-32											
17	ИНДЕКС МАССЫ ТЕЛА-33											
18	ИНДЕКС МАССЫ ТЕЛА-34											
19	ИНДЕКС МАССЫ ТЕЛА-35											
20	ИНДЕКС МАССЫ ТЕЛА-36											
21	ИНДЕКС МАССЫ ТЕЛА-37											
22	ИНДЕКС МАССЫ ТЕЛА-39											
23	ИНДЕКС МАССЫ ТЕЛА-40											
24	ИНДЕКС МАССЫ ТЕЛА-42											

Рисунок 11 -Матрица знаний (модель INF4) (фрагмент)

1.5 Результаты верификации моделей

Результаты верификации моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями, приведены на рисунке 14.

4.1.3.6. Обобщ форма по достоверности моделей при разн. инт. крит. Текущая модель: "INF1"													
Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризбергена	Сумма модулей сходства истинно-положительных решений (STP)	Сумма модулей сходства истинно-отрицательных решений (STN)	Сумма модулей сходства истинно-положительных и отрицательных решений (SFP)	Цифра
1. ABS - частный критерий: количество встреч союзений: "клас..."	Корреляция abs частот с общ... Сумма abs частот по признак...	160	160	1117	8243		0.019	1.000	0.037	145.323	101.605	4122.152	
2. PRC1 - частный критерий: усл. вероятность iго признака сред...	Корреляция усл.отн.частот по признак...	160	160	1117	8243		0.019	1.000	0.037	145.323	101.605	4122.153	
3. PRC2 - частный критерий: условная вероятность iго признака...	Корреляция усл.отн.частот с о...	160	160	1117	8243		0.019	1.000	0.037	136.633	4283.108		
4. INF1 - частный критерий: количество знаний по АХаркевичу, в...	Сумма усл.отн.частот по признак...	160	160	1117	8243		0.019	1.000	0.037	136.633	4283.978		
5. INF2 - частный критерий: количество знаний по АХаркевичу, в...	Семантический резонанс зна...	160	155	1613	7747	5	0.020	0.969	0.038	118.404	193.051	2530.830	
6. INF2 - частный критерий: количество знаний по АХаркевичу, в...	Сумма знаний	160	155	1514	7846	5	0.019	0.969	0.038	85.388	33.592	2103.561	
6. INF3 - частный критерий: "Линкадрат": разности между фактами...	Семантический резонанс зна...	160	155	4707	4653	5	0.032	0.969	0.062	129.027	1848.189	1874.713	
6. INF3 - частный критерий: "Линкадрат": разности между фактами...	Сумма знаний	160	155	4707	4653	5	0.032	0.969	0.062	63.183	881.426	819.038	
7. INF4 - частный критерий ROI (Return On Investment): вероятно...	Семантический резонанс зна...	160	158	2868	6492	2	0.024	0.988	0.046	116.117	184.780	2229.576	
7. INF4 - частный критерий ROI (Return On Investment): вероятно...	Сумма знаний	160	158	1480	7880	2	0.020	0.988	0.039	73.711	12.098	1663.951	
8. INF5 - частный критерий ROI (Return On Investment): вероятно...	Семантический резонанс зна...	160	158	2868	6492	2	0.024	0.988	0.046	116.117	184.780	2229.576	
8. INF5 - частный критерий ROI (Return On Investment): вероятно...	Сумма знаний	160	158	1480	7880	2	0.020	0.988	0.039	73.711	12.098	1663.951	
9. INF6 - частный критерий: разные и безраз.вероятностях: вер...	Семантический резонанс зна...	160	153	1621	7739	7	0.019	0.956	0.038	124.212	207.810	2902.732	
9. INF6 - частный критерий: разные и безраз.вероятностях: вер...	Сумма знаний	160	155	1536	7824	5	0.019	0.969	0.038	91.038	31.644	2469.891	
10. INF7 - частный критерий: разные и безраз.вероятностях: вер...	Семантический резонанс зна...	160	153	1621	7739	7	0.019	0.956	0.038	124.212	207.810	2902.732	
10. INF7 - частный критерий: разные и безраз.вероятностях: вер...	Сумма знаний	160	155	1542	7818	5	0.019	0.969	0.038	91.038	31.644	2469.891	

Рисунок 12- Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели INF3 при интегральном критерии «Семантический резонанс знаний». При этом точность модели составляет 0,134. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием, описанным на рисунке 15.

Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: виды прогнозов и меры достоверности моделей в системе "Эйдос-Х+"

Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: виды прогнозов и меры достоверности моделей в системе "Эйдос-Х+".

ПОДПОЛЮТЕЛЬНЫЙ 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, ни отнесенного к тому классу, к которому он действительно относится, но при этом будет очень большая ошибка ложной идентификации, т.к. огромное количество объектов будет отнесено к классам, к которым они не относятся (и именно за счет этого у модели и будет очень высокая достоверность идентификации). Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.

ОПТИМАЛЬНЫЙ 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 осуществляться, 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 осуществляется 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, 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}:

$$L_1 = \left(\frac{TP + FN}{TP + TN + FP + FN} \right) / \left(\frac{TP + FN}{TP + TN + FP + FN} \right) \quad (\text{нормировка: } \{-1, +1\})$$

$$L_2 = \left(\frac{1 + (TP + TN - FP - FN)}{TP + TN + FP + FN} \right) / 2 \quad (\text{нормировка: } \{0, 1\})$$

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

Классическая F-мера достоверности моделей Ван Ризбергена:

```

Precision = TP/(TP+FP) - точность модели;
Recall = TP/(TP+FN) - полнота модели;
F-mera = 2*(Precision*Recall)/(Precision+Recall).

```

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

```

SPrecision = STP/(STP+SFP) - нечеткая мультиклассовая точность модели;
SRecall = STP/(STP+SFN) - нечеткая мультиклассовая полнота модели;
L-мера = 2*(SPrecision*SRecall)/(SPrecision+SRecall) - нечеткая мультиклассовая достоверность модели.

```

где:

STP – Сумма модулей сходства истинно-положительных решений; STN – Сумма модулей сходства истинно-отрицательных решений; SFP – Сумма модулей сходства ложно-положительных решений; SFN – Сумма модулей сходства ложно-отрицательных решений.

Рисунок 13- Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

Стоит отметить, 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 16 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и неидентифицированных ситуаций в наиболее достоверной модели INF3.

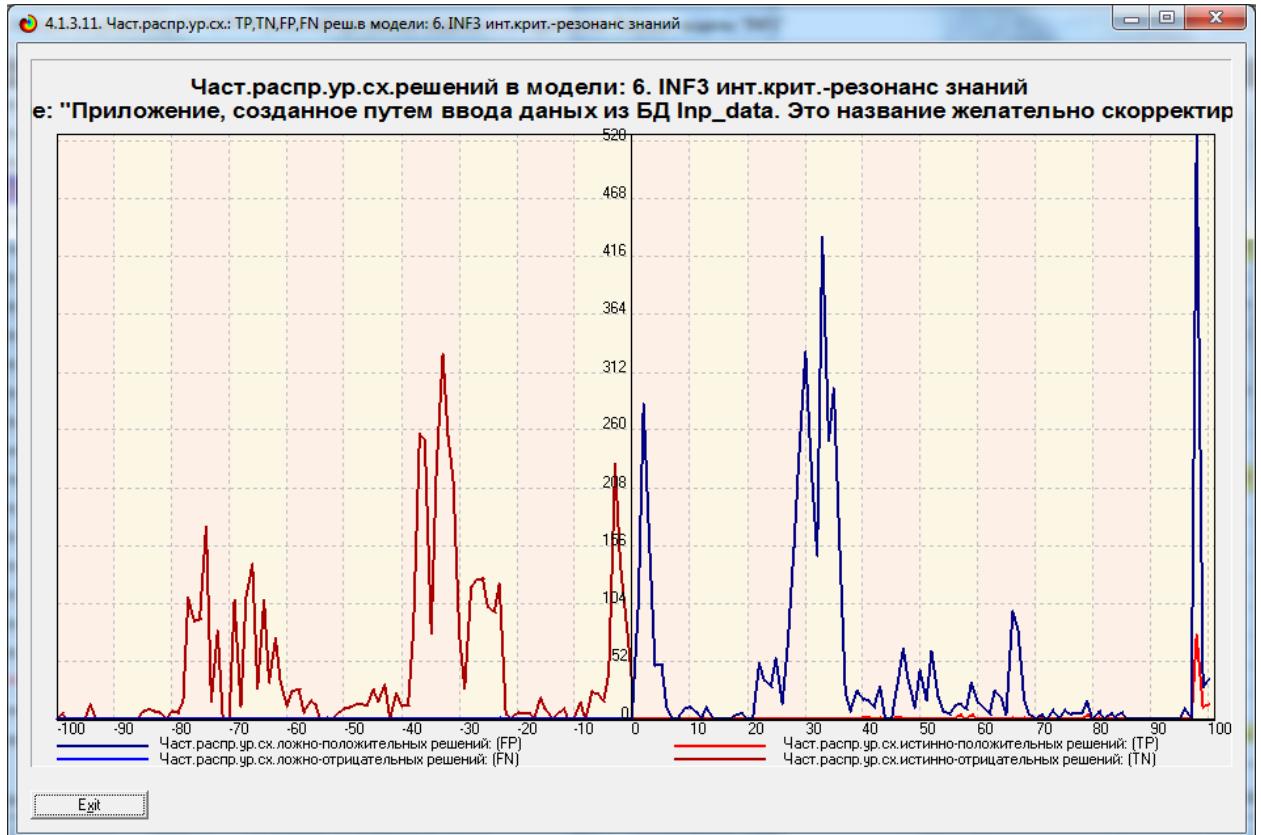


Рисунок 14- Частное распределение ложно-положительных решений, ложно-отрицательных решений, истинно-положительных решений и истинно-отрицательных решений объекта моделирования в модели INF4

Из рисунка 16 видно, что:

- наиболее достоверная модель INF3 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 14);
- 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. Для очень больших значений уровней сходства-различия (более 95%) также различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями практически отсутствует.

РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1 Решение задачи идентификации

В соответствии с технологией АСК-анализа зададим текущей модель INF3 (режим 5.6) и проведем пакетное распознавание в режиме 4.2.1, как показано на рисунке 18.

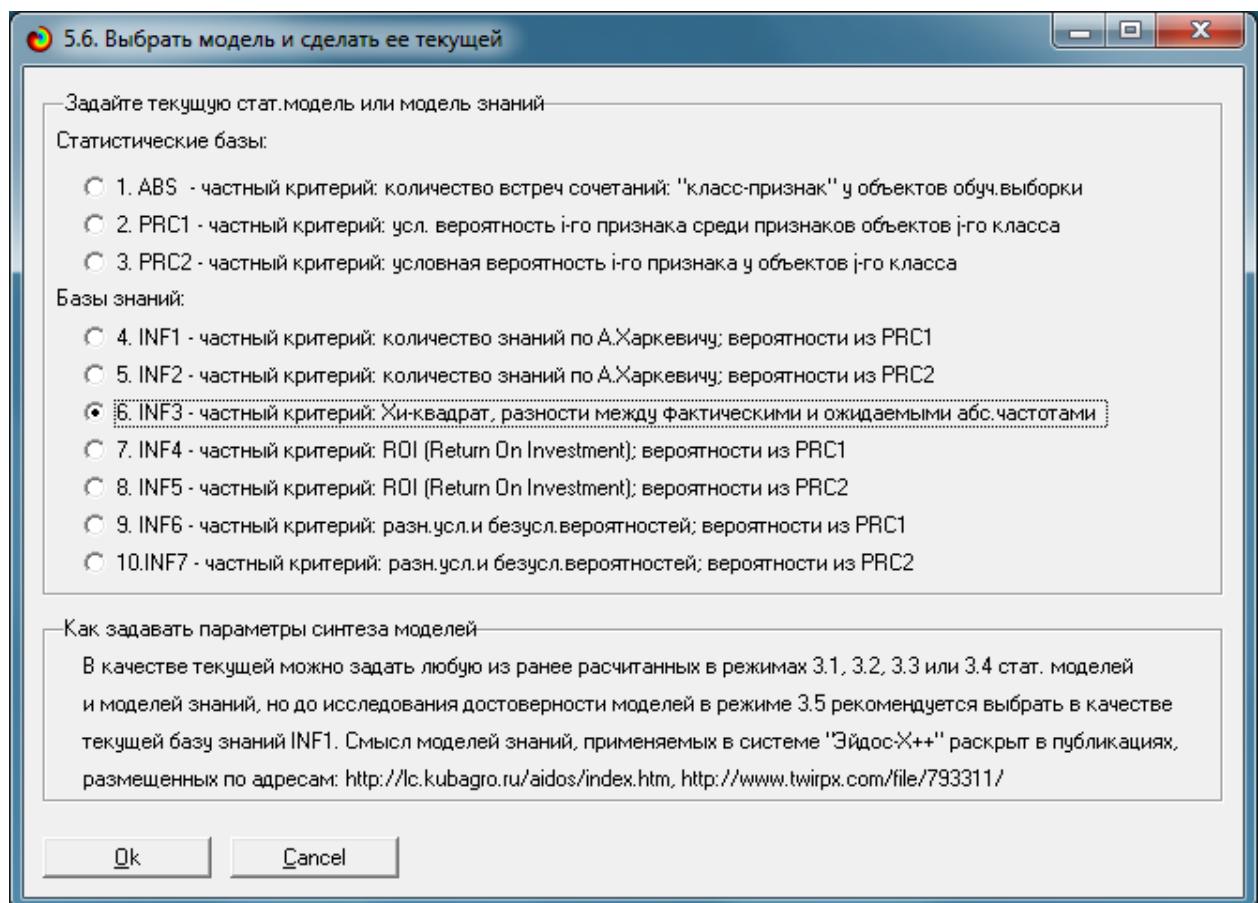


Рисунок 15- Экранная форма режима задания модели в качестве текущей

Следующим шагом необходимо произвести пакетное распознавание в текущей модели. Если распознавание производится впервые, то вы увидите сообщение, уведомляющее об этом.

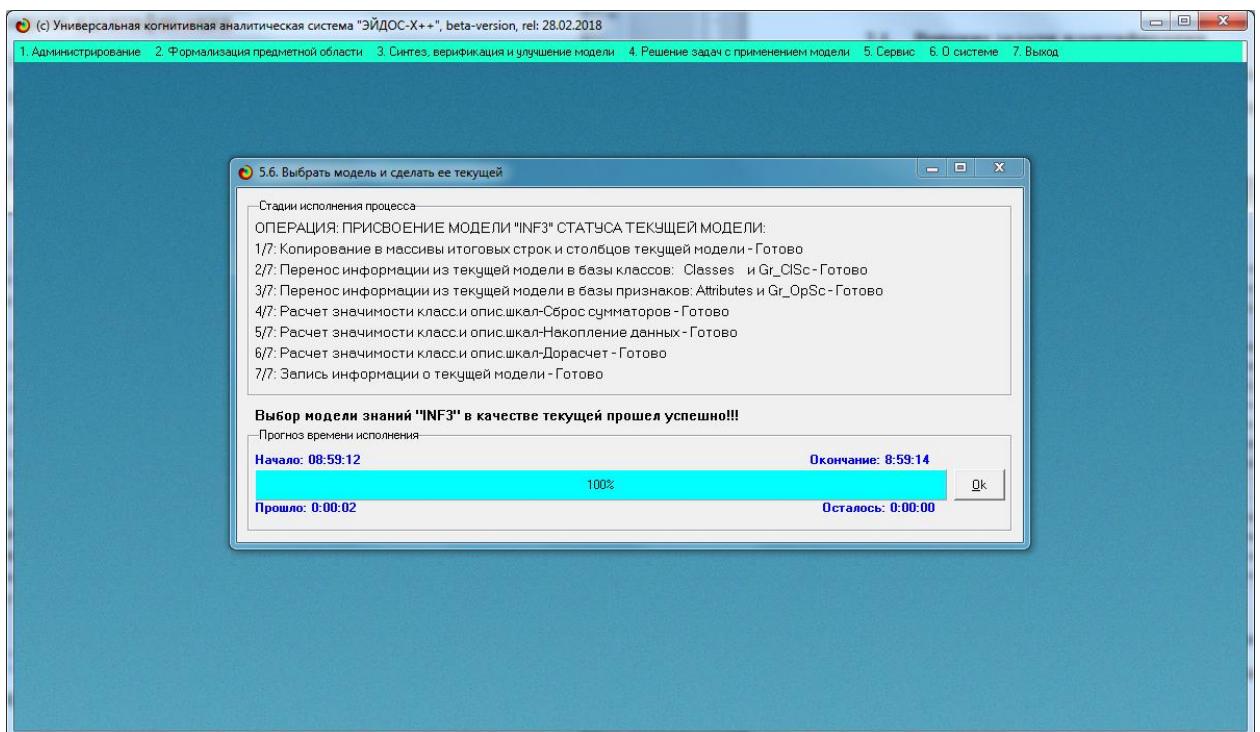


Рисунок 16- Экранная форма режима пакетного распознавания в текущей модели

В результате пакетного распознавания в текущей модели создается ряд баз данных, которые визуализируются в выходных экранах 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.

Для наглядного примера, кратко рассмотрим некоторые из них.

На рисунках 19 и 20 приведены примеры прогнозов высокой и низкой достоверности частоты и классов бейсбольных команд в наиболее достоверной модели INF3 на основе описательных шкал.

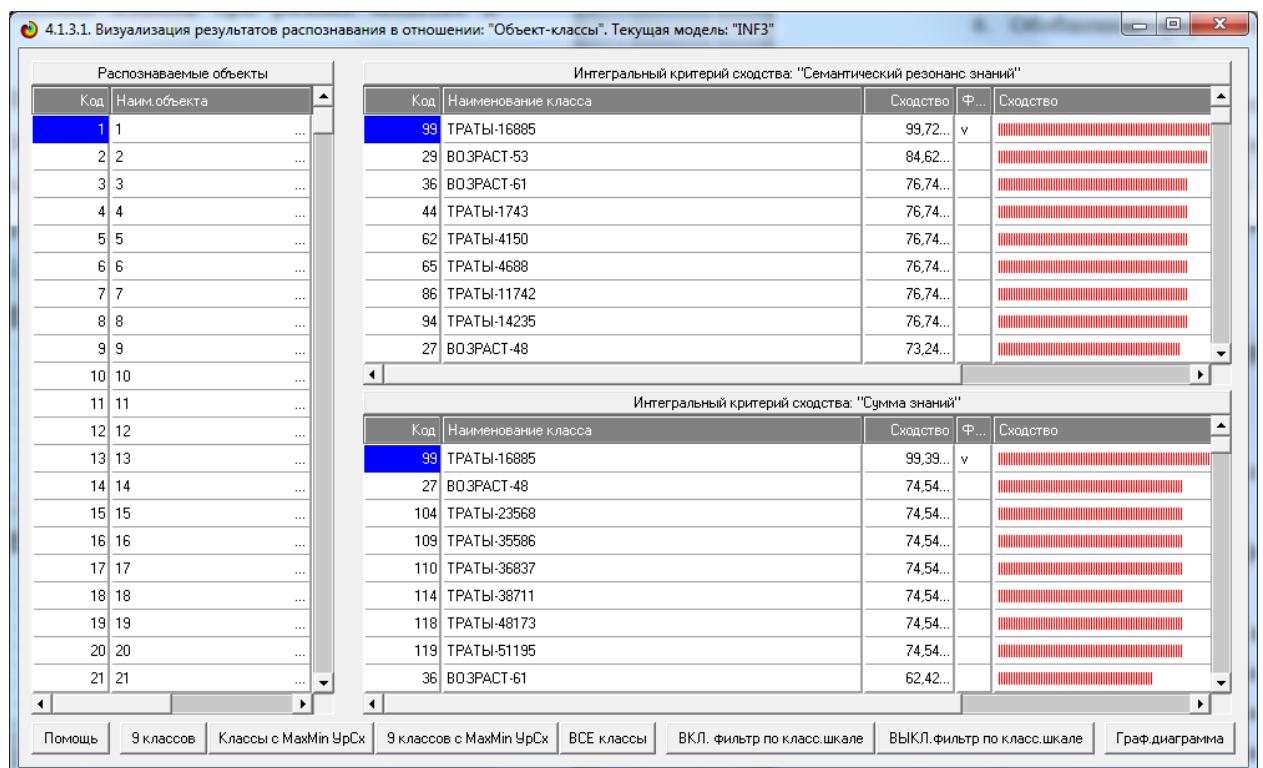


Рисунок 17 - Пример идентификации классов в модели INF3

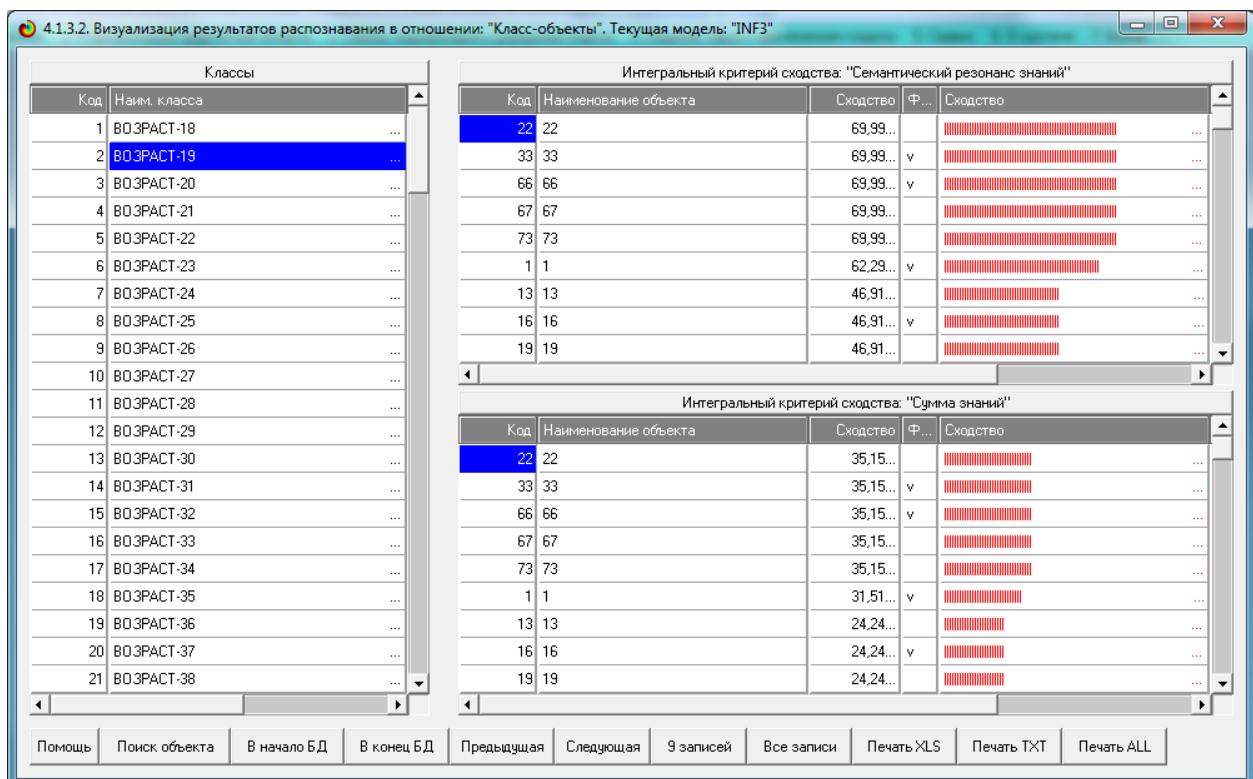


Рисунок 18- Пример идентификации классов в модели INF4

2.2 Когнитивные функции

Далее стоит рассмотреть режим работы 4.5, в котором реализована возможность визуализации когнитивных функций для любых моделей и любых сочетаний классификационных и описательных шкал. Начало работы в этом режиме начинается с понятия когнитивных функций, а также предоставления возможности перехода на статьи по данной проблематике и скачивания публикаций, как это продемонстрировано на рисунке 21.

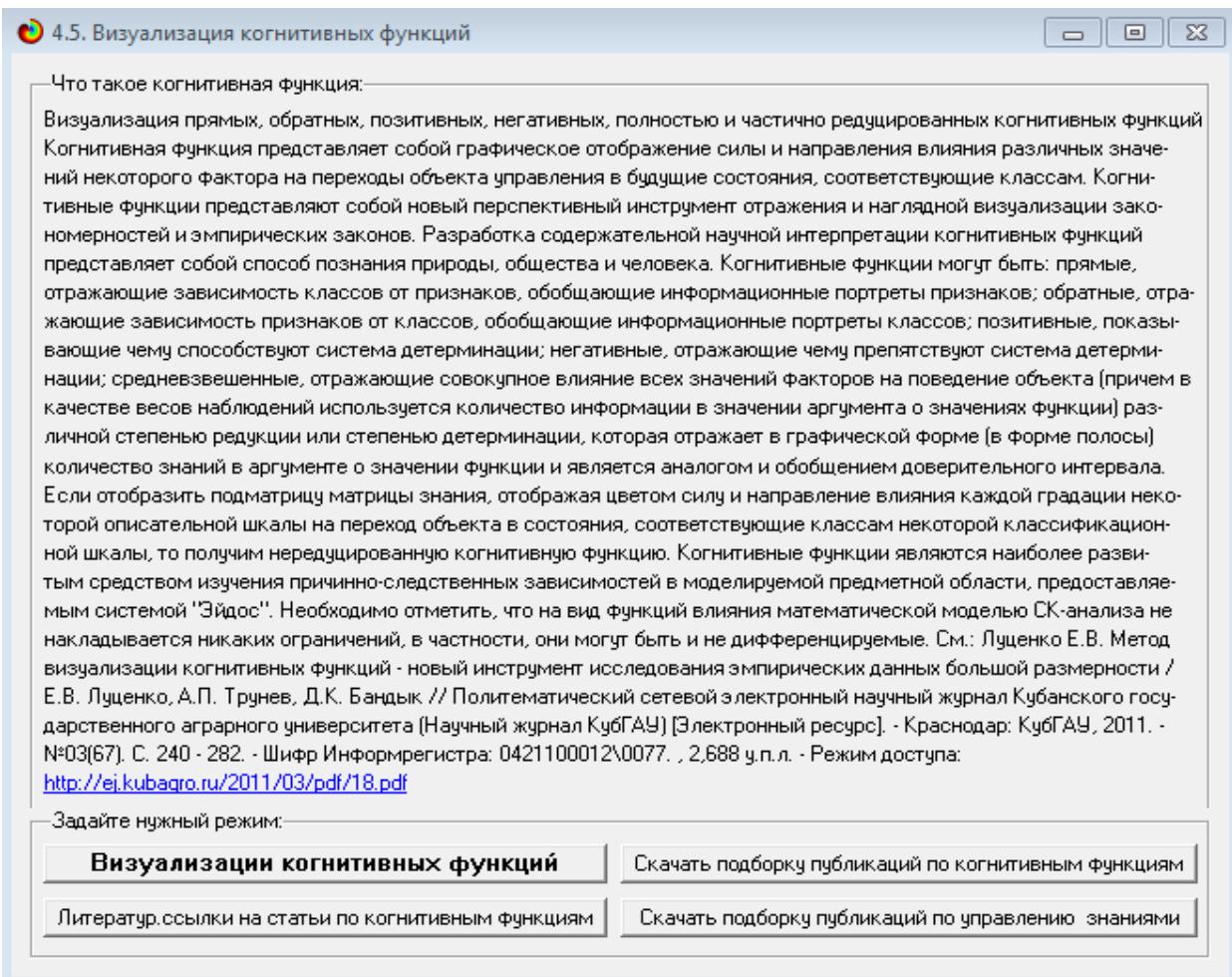


Рисунок 19- Экранная форма режима 4.5 системы «Эйдос-Х++»
«Визуализация когнитивных функций»

Применительно к задаче, рассматриваемой в данной работе, когнитивная функция показывает, какое количество информации содержится в различных значениях факторов о том, что объект моделирования перейдет в те или иные будущие состояния. Когнитивным функциям посвящено множество работ, поэтому здесь не будем останавливаться на описании того, что представляют собой когнитивные функции в АСК-анализе. На рисунке 22 приведены визуализации всех когнитивных функций данного приложения для модели INF3.

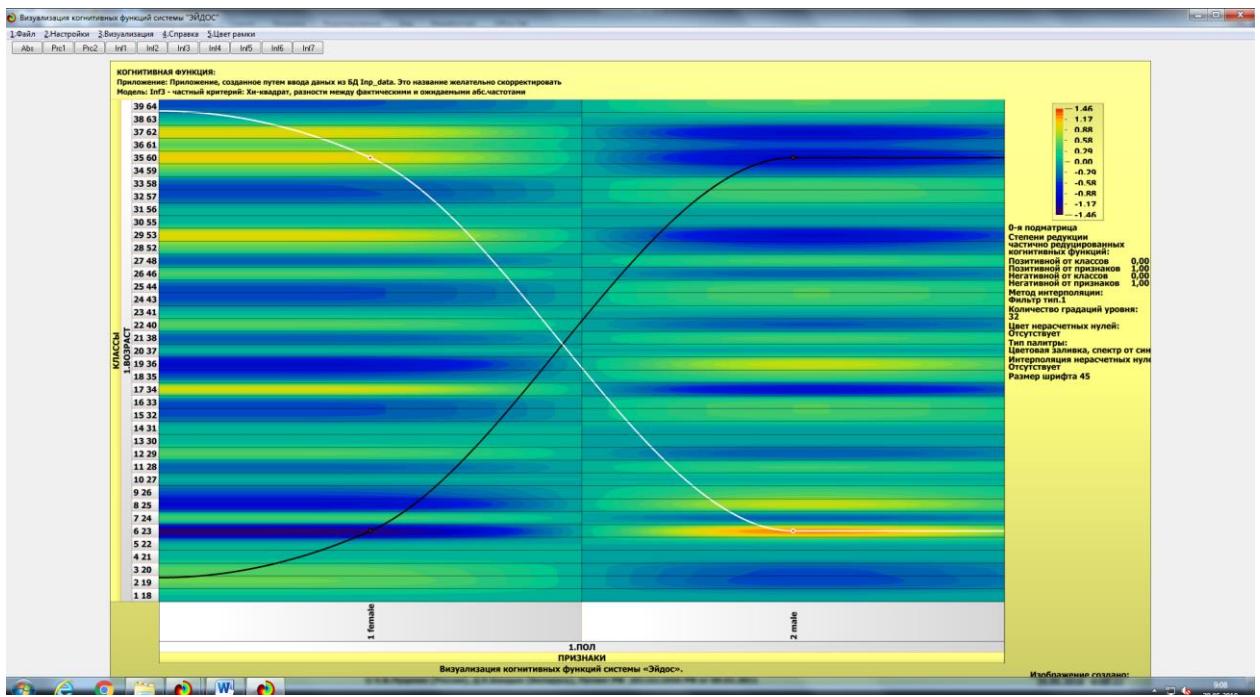


Рисунок 20 - Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF3

2.3 SWOT и PEST матрицы и диаграммы

SWOT-анализ является широко известным и общепризнанным методом стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего профессионального опыта и компетенции. Но возможности экспертов имеют свои ограничения и часто по различным причинам они не могут и не хотят это сделать. Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов. Эта проблема может решаться путем автоматизации функций экспертов, т.е. путем измерения силы и направления влияния факторов непосредственно на основе эмпирических данных.

Подобная технология разработана давно, ей уже около 30 лет, но она малоизвестна – это интеллектуальная система «Эйдос». Данная система всегда обеспечивала возможность проведения количественного автоматизированного SWOT-анализа без использования экспертных оценок непосредственно на основе эмпирических данных. Результаты SWOT-анализа выводились в форме информационных портретов. В версии системы под MS Windows: «Эйдос-X++» предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (рисунок 23).

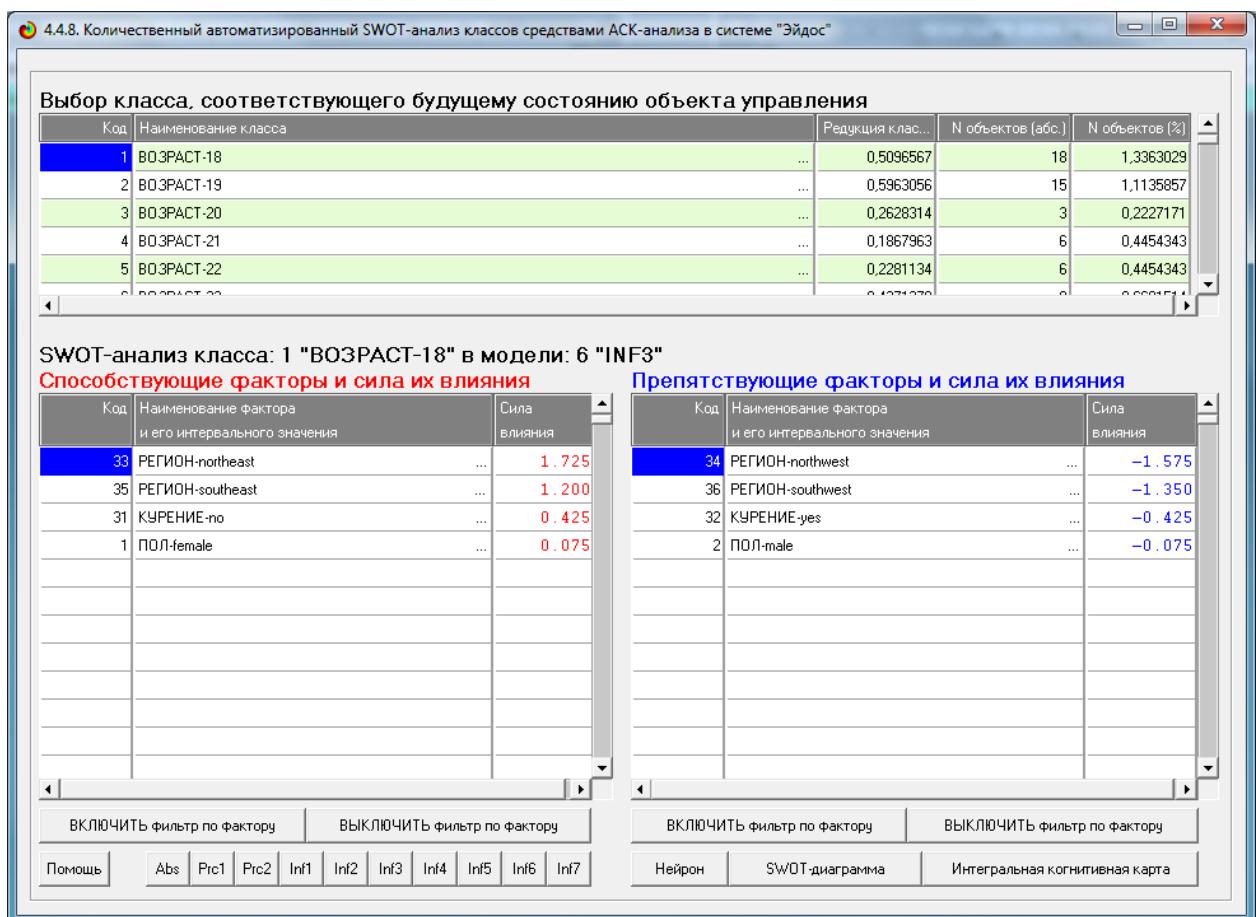


Рисунок 21 - Пример SWOT-Матрицы в модели INF3

Следующим шагом является построение SWOT-диаграммы, для наглядной демонстрации работы, продемонстрированной на рисунке 24.

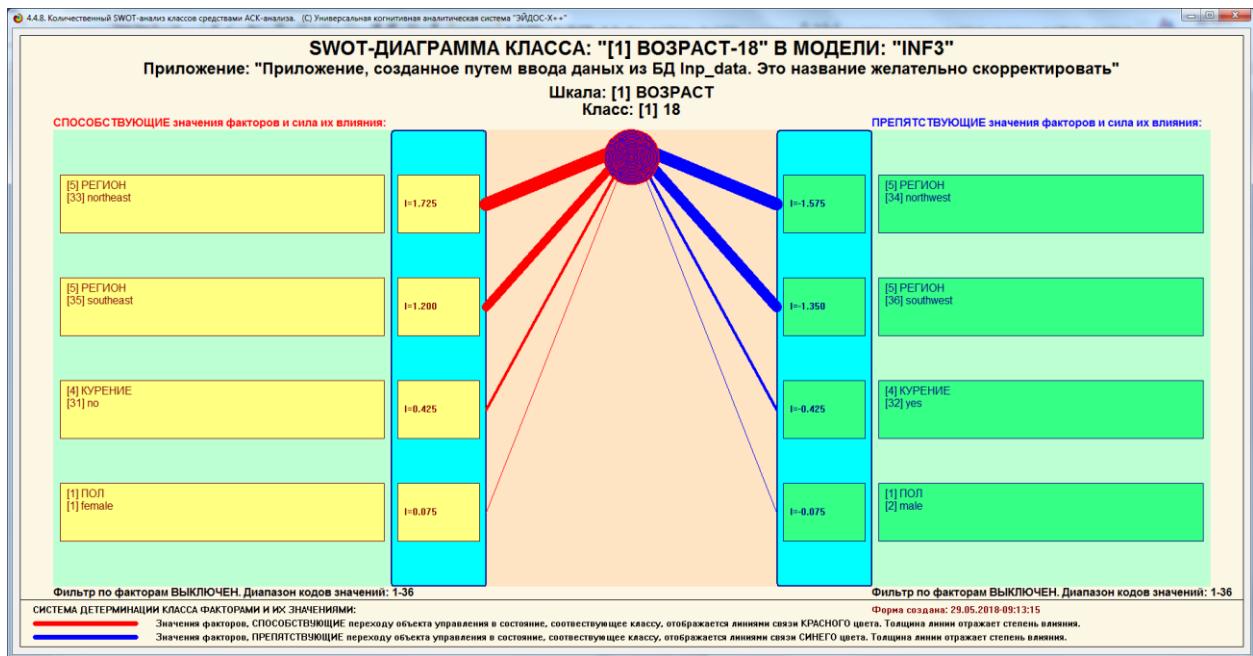


Рисунок 22– SWOT-матрица, построенная для 18 лет

2.4 Нелокальные нейронные сети

Каждому классу системно-когнитивной модели соответствует нелокальный нейрон, совокупность которых образует не локальную нейронную сеть. Рассмотрим пример на рисунке 25.

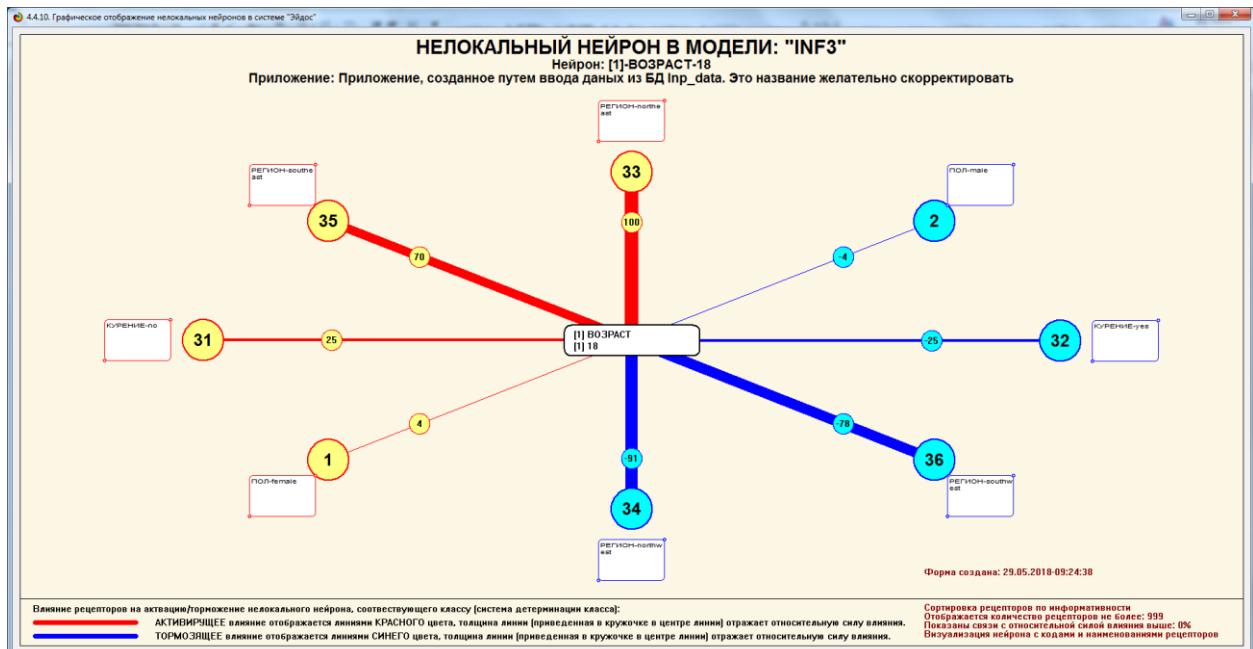


Рисунок 23 - Нейрон для INF3

2.5 Кластерный и конструктивный анализ

В режиме 4.2.2.2, после расчета матриц сходства, кластеров и конструкторов, строим 2D сеть классов в выбранной модели знаний, для наглядного представления сети классов.

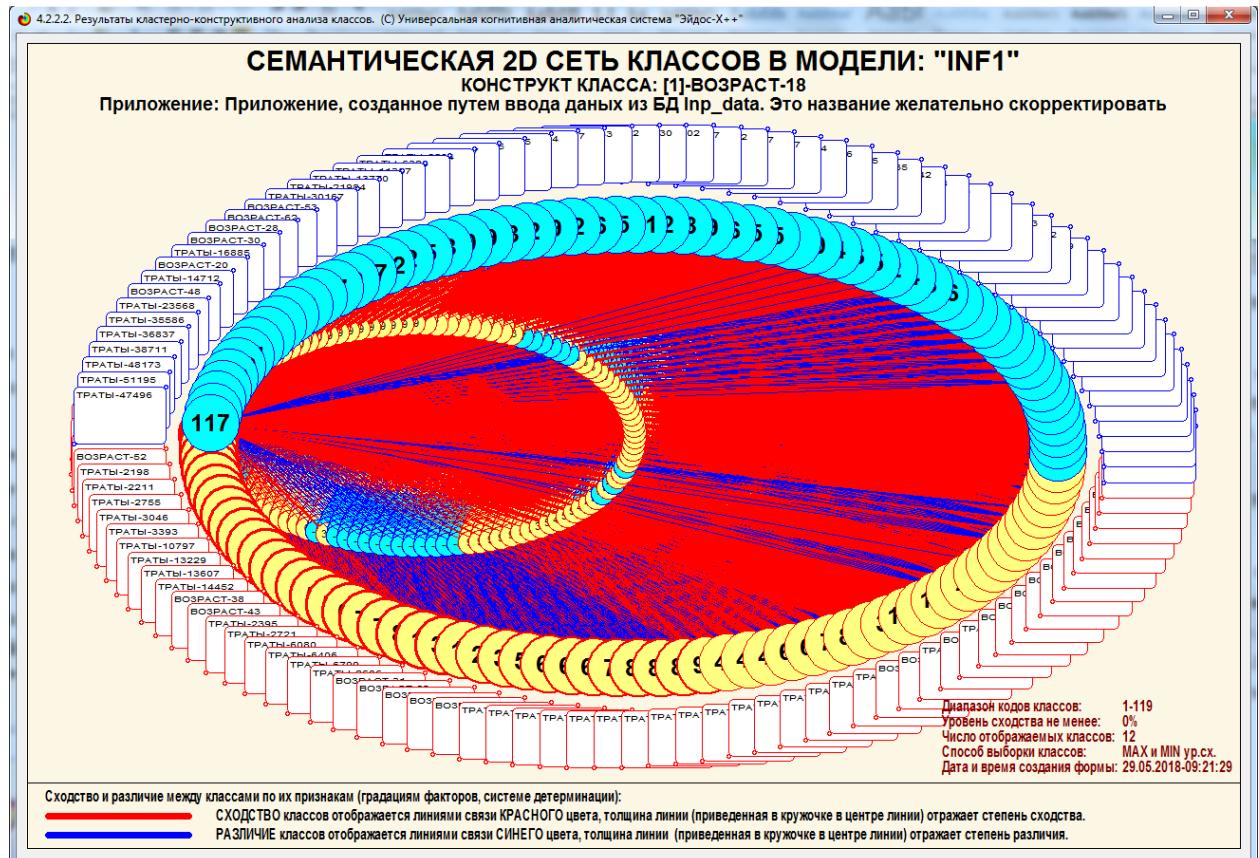


Рисунок 24– Семантическая 2D сеть классов

ЗАКЛЮЧЕНИЕ

В связи с существованием множества 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 «Kaggle».

В данной лабораторной приводится развернутый пример использования базы данных репозитория «Kaggle» для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». При этом наиболее достоверной в данном приложении оказались модели INF3, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,134, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 10%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием.

На основе базы данных «Kaggle», рассмотренной в данной лабораторной работе, построен модели прогнозирования с помощью АСК-анализа и реализующей его системы «Эйдос».

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Луценко Е.В. Методика использования репозитория UCI для оценки качества математических моделей систем искусственного интеллекта / Е.В. Луценко // Полitemатический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: http://lc.kubagro.ru/My_training_schedule.doc КубГАУ, 2003. – №02(002). С. 120 – 145. – IDA [article ID]: 0020302012. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 у.п.л.
2. Луценко Е.В. ACK-анализ, моделирование и идентификация живых существ на основе их фенотипических признаков / Е.В. Луценко, Ю.Н. Пенкина // Полitemатический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2014. – №06(100). С. 1346 – 1395. – IDA [article ID]: 1001406090. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/90.pdf>, 3,125 у.п.л.
3. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Полitemатический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.
4. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

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

7. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее время / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 48 – 77. – Шифр Информрегистра: 0420900012\0110, IDA [article ID]: 0540910004. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>, 1,875 у.п.л.