

Министерство сельского хозяйства российской федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
им. И.Т. Трубилина

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

Лабораторная работа

по дисциплине: Интеллектуальные информационные системы

на тему:

АСК- анализ футбола

выполнила студентка группы: ПИ1501

Бабенко А.В.

Руководитель работы:
профессор Луценко Е.В.

Краснодар 2018

Содержание

ВВЕДЕНИЕ.....	3
1.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.....	10
1.4. Виды моделей системы «Эйдос».....	12
1.5. Результаты верификации моделей	14
2.РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	17
2.1.Решениезадачи	17
2.2 Когнитивные функции.....	20
2.3. SWOT и PEST матрицы и диаграммы	22
2.4. Нелокальные нейронные сети	24
2.5 Кластерный и конструктивный анализ.....	25
ЗАКЛЮЧЕНИЕ	26
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	27

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 идентификации классов волейбольных команд в лигах.

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 «The History of Soccer» из банка исходных данных по задачам искусственного интеллект-репозитория «Kaggle».

Для решения задачи используем стандартные возможности Microsoft Office Excel, блокнот, а также систему искусственного интеллекта «Aidos-X++».

1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

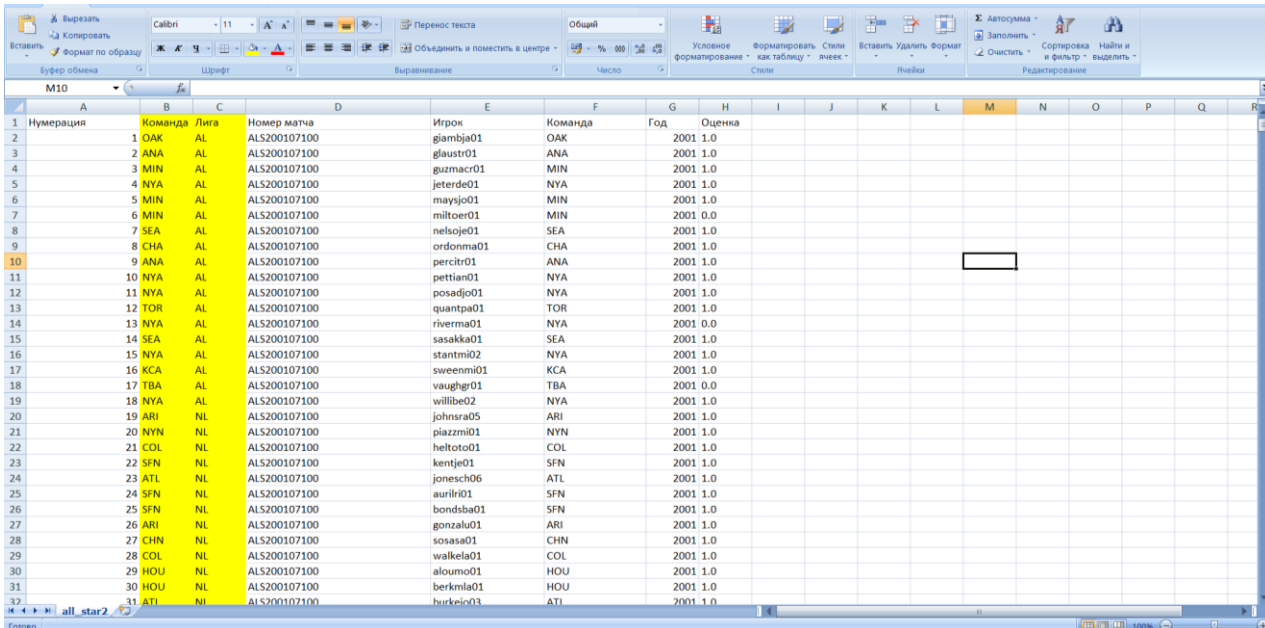
1.1. Описание решения

В соответствии с методологией АСК-анализа решение поставленной задачи проведем в четыре этапа:

1. Преобразование исходных данных 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 в файл MS Excel

Из электронного ресурса баз данных «Kaggle» возьмем базу данных «The History of Soccer», которую приведем к виду на рисунке 1.



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Нумерация	Команда	Лига	Номер матча	Игрок	Команда	Год	Оценка																								
1	OAK	AL	ALS200107100	giambja01	OAK	2001	1.0																								
2	ANA	AL	ALS200107100	glaustr01	ANA	2001	1.0																								
3	MIN	AL	ALS200107100	guzmacr01	MIN	2001	1.0																								
4	NYA	AL	ALS200107100	jeterde01	NYA	2001	1.0																								
5	MIN	AL	ALS200107100	maysjo01	MIN	2001	1.0																								
6	MIN	AL	ALS200107100	mittoer01	MIN	2001	0.0																								
7	SEA	AL	ALS200107100	nelsoje01	SEA	2001	1.0																								
8	CHA	AL	ALS200107100	ordonma01	CHA	2001	1.0																								
9	ANA	AL	ALS200107100	percitr01	ANA	2001	1.0																								
10	NYA	AL	ALS200107100	pettjan01	NYA	2001	1.0																								
11	NYA	AL	ALS200107100	posadjo01	NYA	2001	1.0																								
12	TOR	AL	ALS200107100	quantpa01	TOR	2001	1.0																								
13	NYA	AL	ALS200107100	riverma01	NYA	2001	0.0																								
14	SEA	AL	ALS200107100	sasakka01	SEA	2001	1.0																								
15	NYA	AL	ALS200107100	stantmi02	NYA	2001	1.0																								
16	KCA	AL	ALS200107100	sweenmi01	KCA	2001	1.0																								
17	TBA	AL	ALS200107100	vaughgr01	TBA	2001	0.0																								
18	NYA	AL	ALS200107100	willibe02	NYA	2001	1.0																								
19	ARI	NL	ALS200107100	johnsra05	ARI	2001	1.0																								
20	NYN	NL	ALS200107100	piazmi01	NYN	2001	1.0																								
21	COL	NL	ALS200107100	heltoto01	COL	2001	1.0																								
22	SFN	NL	ALS200107100	kentje01	SFN	2001	1.0																								
23	ATL	NL	ALS200107100	jonesch06	ATL	2001	1.0																								
24	SFN	NL	ALS200107100	aurinri01	SFN	2001	1.0																								
25	SFN	NL	ALS200107100	bondsba01	SFN	2001	1.0																								
26	ARI	NL	ALS200107100	gonzalu01	ARI	2001	1.0																								
27	CHN	NL	ALS200107100	soxasa01	CHN	2001	1.0																								
28	COL	NL	ALS200107100	walkela01	COL	2001	1.0																								
29	HOU	NL	ALS200107100	aloumo01	HOU	2001	1.0																								
30	HOU	NL	ALS200107100	berkmla01	HOU	2001	1.0																								
31	ATL	NL	ALS200107100	burkejo03	ATL	2001	1.0																								

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.

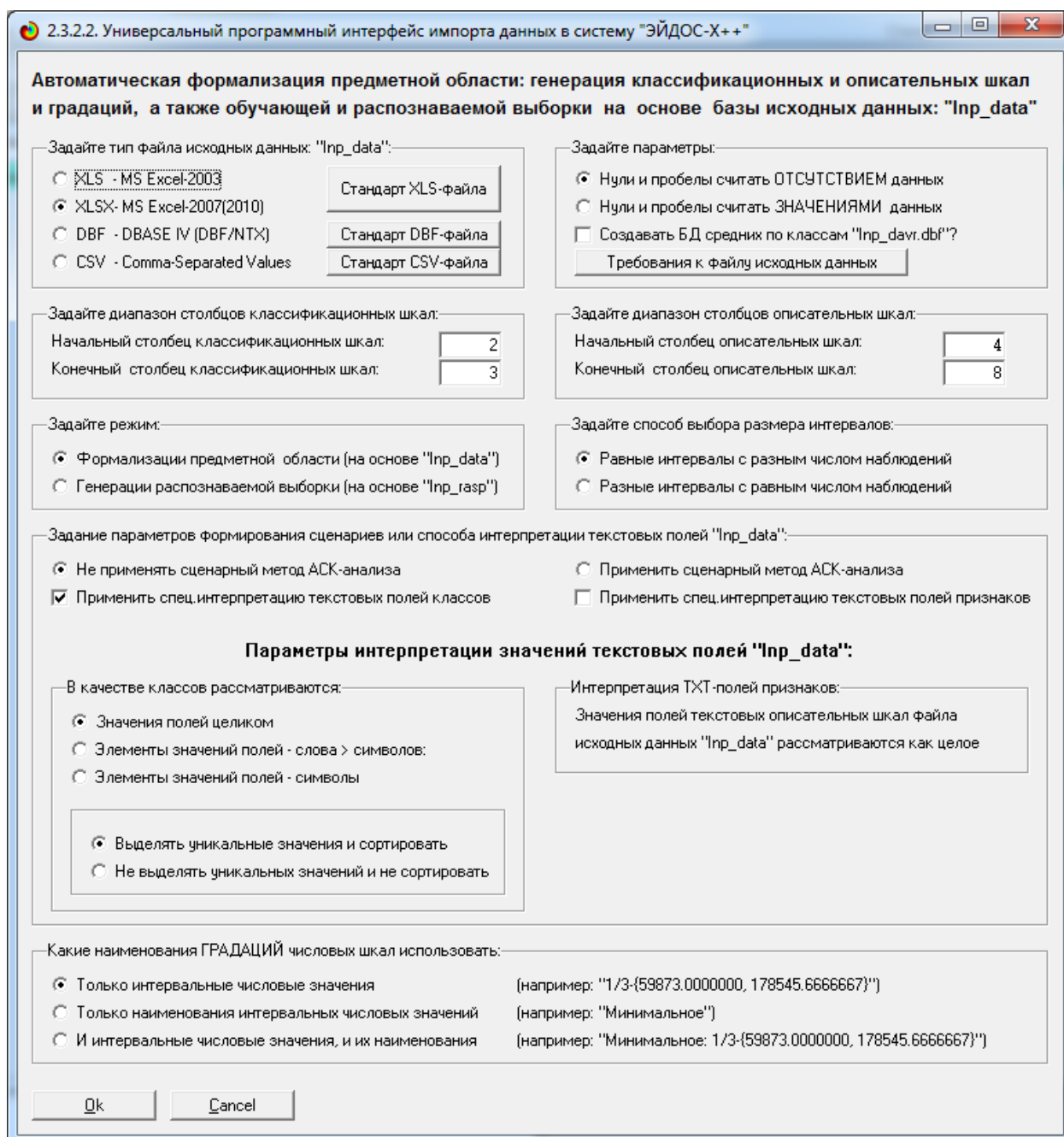


Рисунок 2- Экранная форма универсального программного интерфейса импорта данных в систему «Эйдос»

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

— «Тип файла исходных данных Inp_data»: «XLS – MS Excel-2003»;

— «Диапазон столбцов классификационных шкал»: «Начальный столбец классификационных шкал» – 1, «Конечный столбец классификационных шкал» – 3;

— «Диапазон столбцов описательных шкал»: «Начальный столбец описательных шкал»-6, «Конечный столбец описательных шкал»-13;

— «Параметры формирования сценариев или способа интерпретации текстовых полей»: «Не применять сценарный метод АСК-анализа и спец.интерпретацию ТХТ-полей».

После нажать кнопку «ОК». 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.

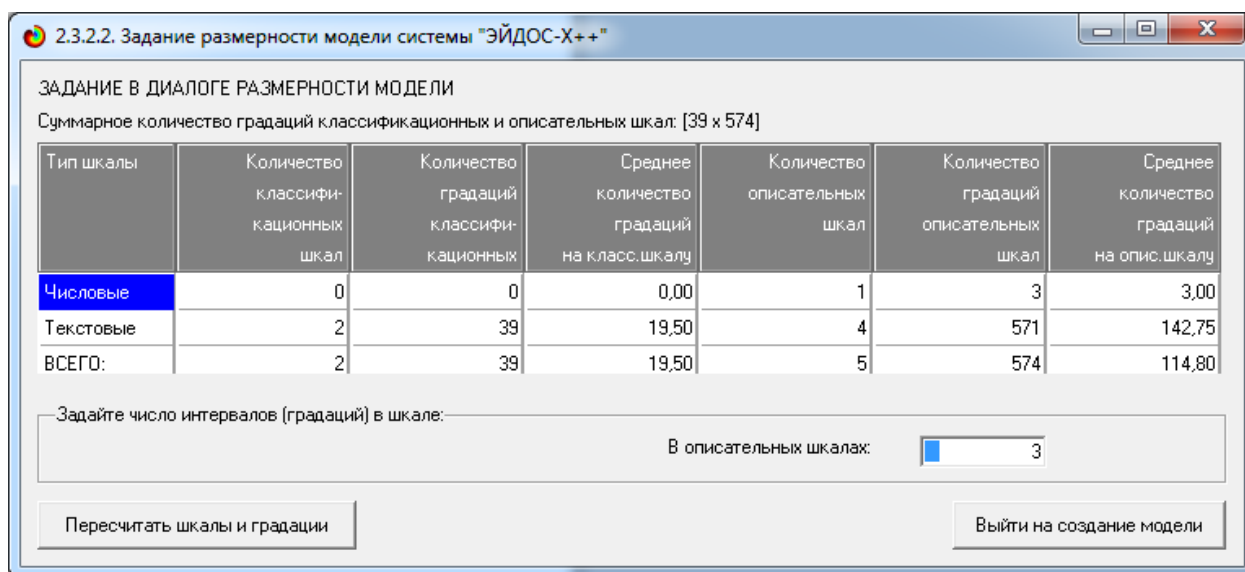


Рисунок 3- Задание размерности модели системы «Эйдос»

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 «ОК», как показано на рисунке 4.

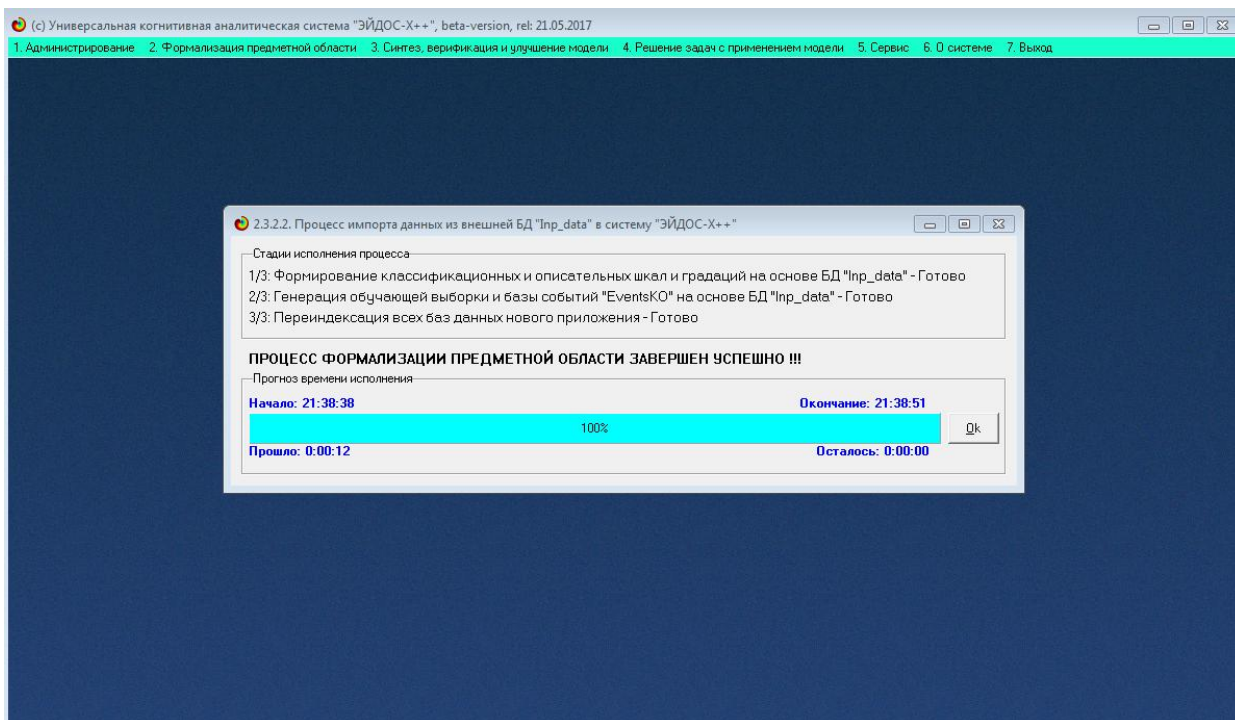


Рисунок 4- Процесс импорта данных из внешней БД «Inp_data.xls» в систему «Эйдос»

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-й этапе АСК- анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1, рисунок 5.

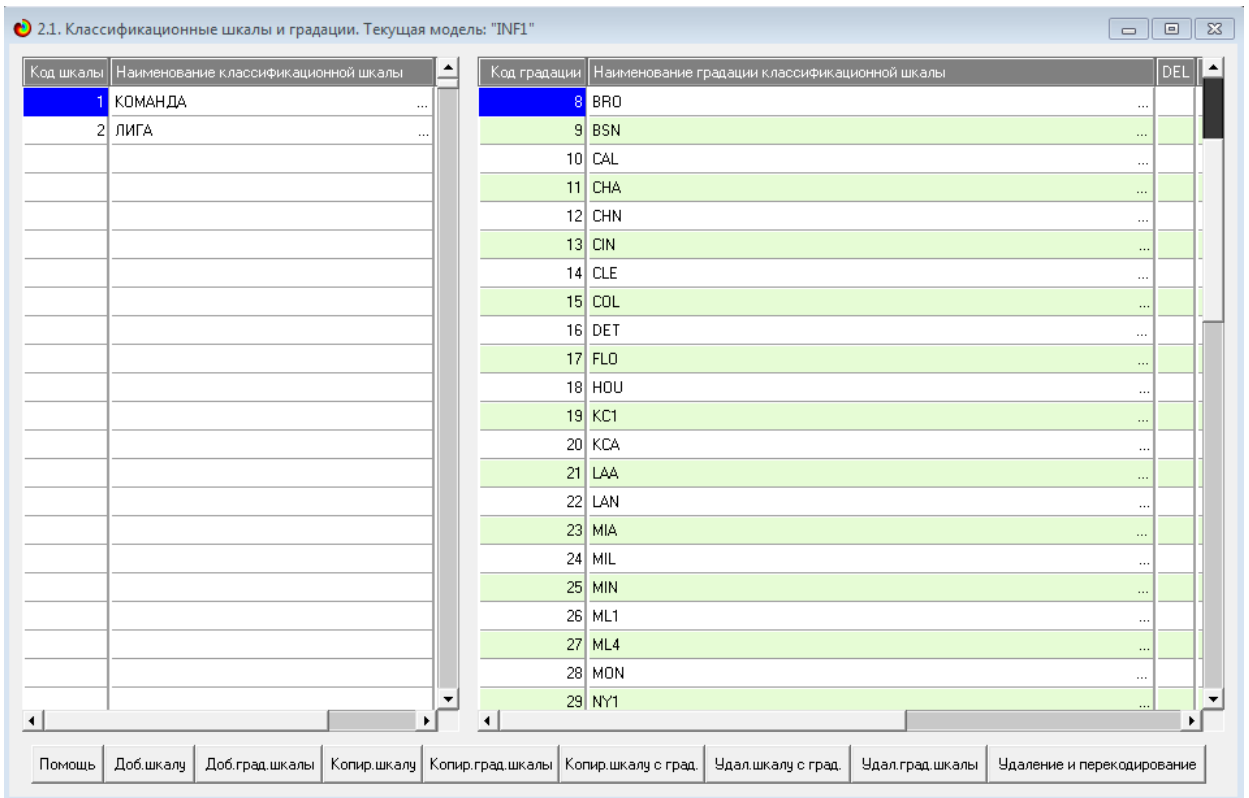


Рисунок 5- Классификационные шкалы и градации (фрагмент)

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

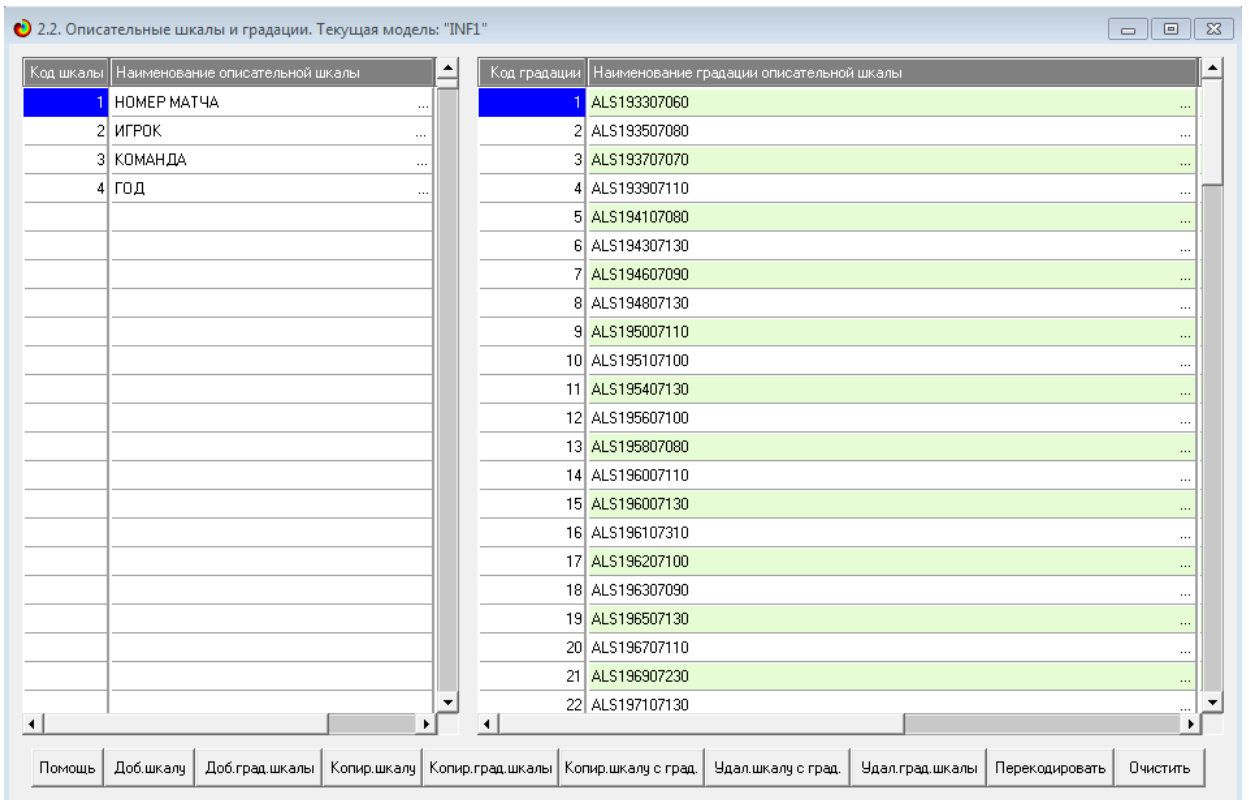


Рисунок 6- Описательные шкалы и градации (фрагмент)

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

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	25	38	0	0	1	1	181	556	569	573	0	0

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

Рисунок 7- 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.

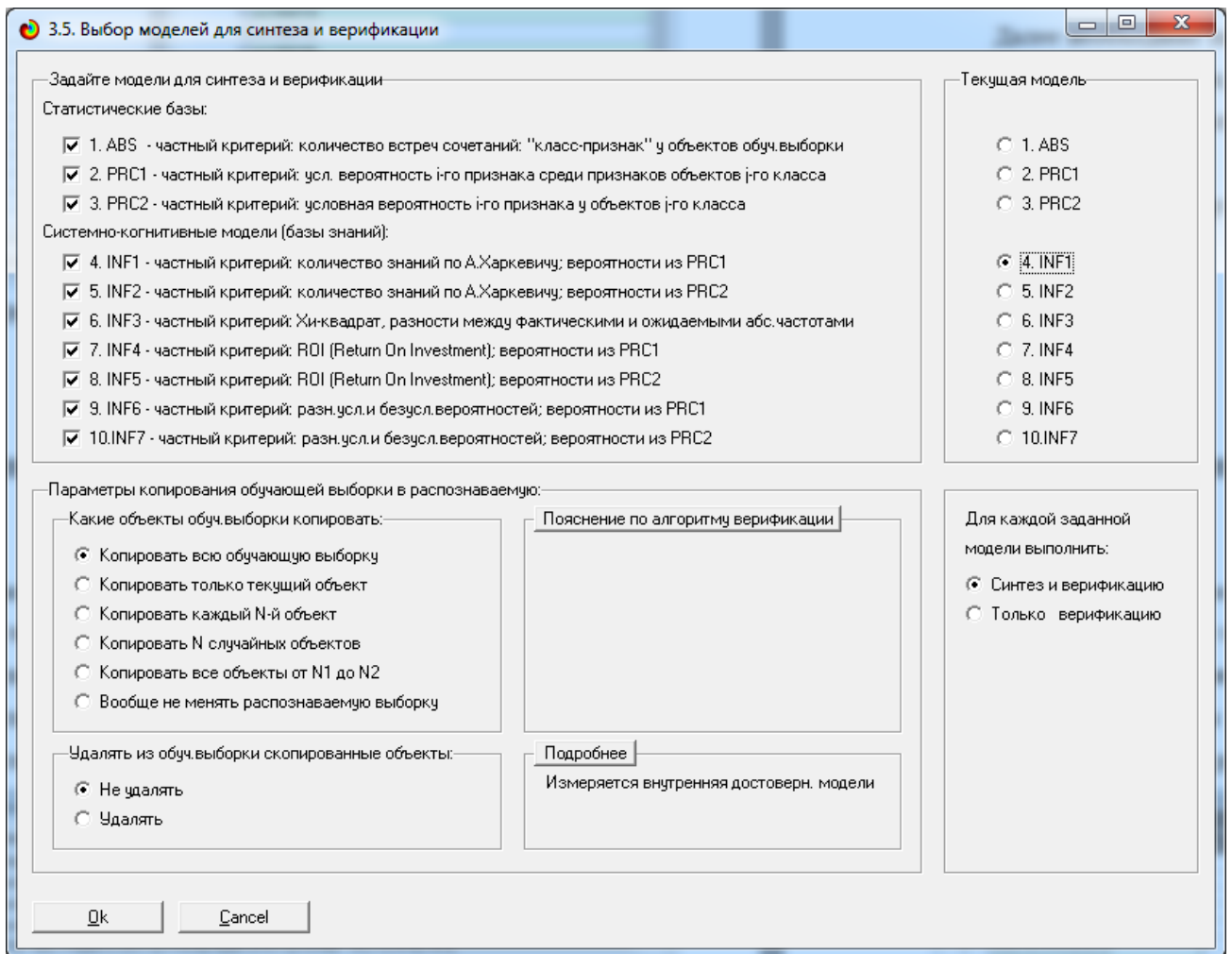


Рисунок 8- Выбор моделей для синтеза и верификации, а также текущей модели

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.

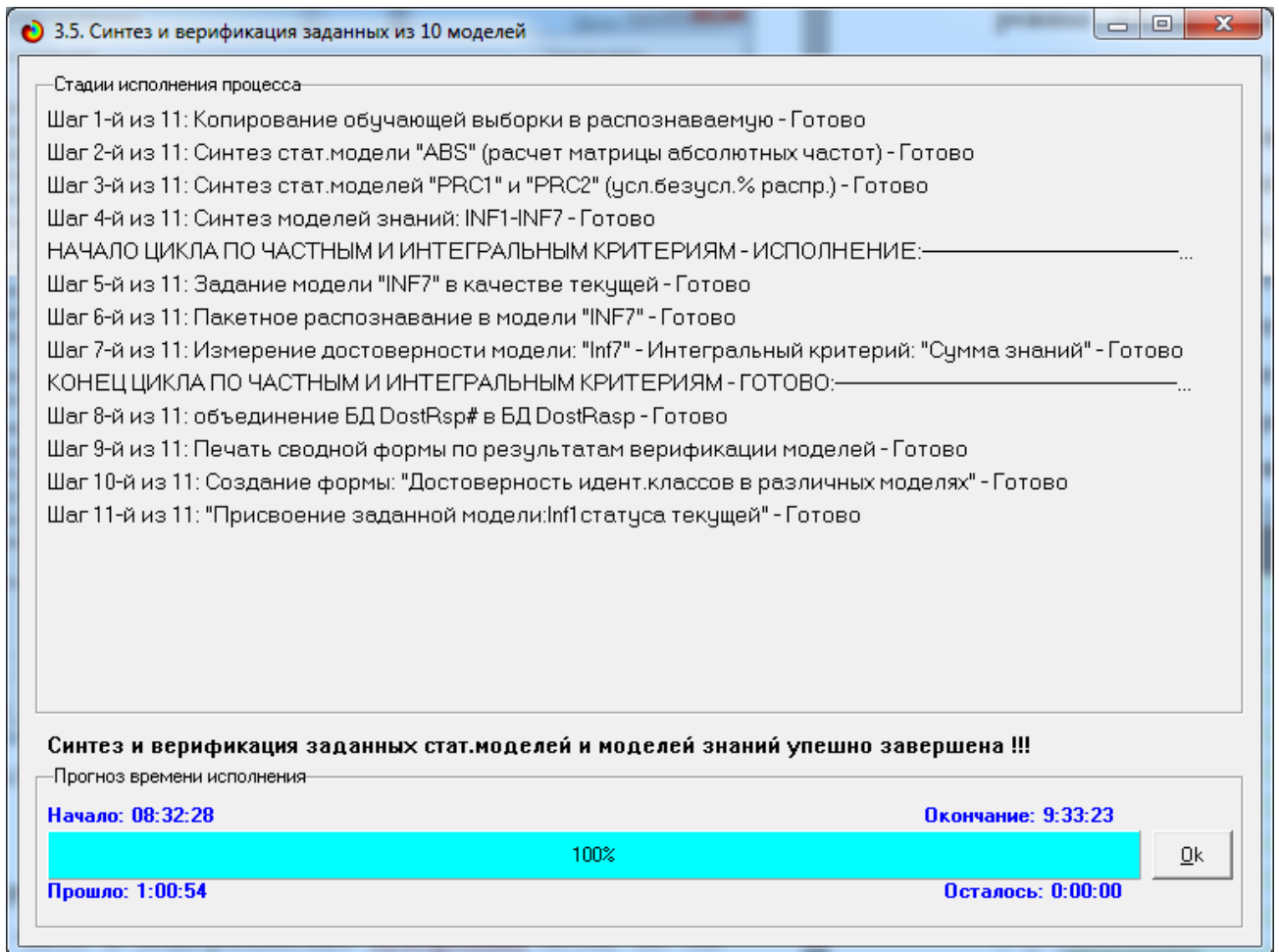


Рисунок 9- Синтез и верификация статистических моделей и моделей знаний

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, перечисленные 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 обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. КОМАНДА ANA	2. КОМАНДА ARI	3. КОМАНДА ARL	4. КОМАНДА ATL	5. КОМАНДА BAL	6. КОМАНДА BIS	7. КОМАНДА BOS	8. КОМАНДА CHA	9. КОМАНДА CHN	10. КОМАНДА CIN	11. КОМАНДА CLE
1	НОМЕР МАТЧА-АЛS200107100	2	3		2				1	2	1	
2	НОМЕР МАТЧА-АЛS200307150	3	1		7	1		3	3	2	1	1
3	НОМЕР МАТЧА-АЛS200507120		1		2	4		5	4	2	1	1
4	НОМЕР МАТЧА-АЛS200807150		2		2	1		7	2	8	1	2
5	НОМЕР МАТЧА-АЛS201007130		1		6	1		6	2	1	4	1
6	НОМЕР МАТЧА-АЛS201207100		1		4	3		1	4	2	3	2
7	НОМЕР МАТЧА-АЛS201407150		2		3	3		2	3	3	5	1
8	НОМЕР МАТЧА-АЛS200207090	1	6		4	1		7	2	1	1	1
9	НОМЕР МАТЧА-АЛS200407130	2	1		1	1		3	1	3	4	5
10	НОМЕР МАТЧА-АЛS200607110		1		3	1		4	7	1	1	1
11	НОМЕР МАТЧА-АЛS200707100		3		2	1		6	1	2	1	3
12	НОМЕР МАТЧА-АЛS200907140		2		1	1		6	1	1	1	1
13	НОМЕР МАТЧА-АЛS201107120		2	1	4	1	1	5	2	1	4	2
14	НОМЕР МАТЧА-АЛS201307160		2	2	3	5	1	3	2	1	3	2
15	НОМЕР МАТЧА-АЛS201507140		2			3		1		2	2	1
16	НОМЕР МАТЧА-АЛS201507141											
17	НОМЕР МАТЧА-АЛS201507142											
18	НОМЕР МАТЧА-АЛS201507143											
19	НОМЕР МАТЧА-АЛS201507144				1							
20	НОМЕР МАТЧА-АЛS201507145											
21	НОМЕР МАТЧА-АЛS201507146											
22	НОМЕР МАТЧА-АЛS201507147											
23	НОМЕР МАТЧА-АЛS201507148											
24	НОМЕР МАТЧА-АЛS201507149											

Рисунок 10- 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. КОМАНДА ANA	2. КОМАНДА ARI	3. КОМАНДА ARL	4. КОМАНДА ATL	5. КОМАНДА BAL	6. КОМАНДА BIS	7. КОМАНДА BOS	8. КОМАНДА CHA	9. КОМАНДА CHN	10. КОМАНДА CIN	11. КОМАНДА CLE
1	НОМЕР МАТЧА-АЛS200107100	0.965	0.443		-0.017				-0.284	0.175	-0.237	
2	НОМЕР МАТЧА-АЛS200307150	0.993	-0.386		0.494	-0.343		-0.146	0.139	-0.028	-0.441	
3	НОМЕР МАТЧА-АЛS200507120		-0.370		-0.203	0.464		0.162	0.320	-0.012	-0.424	
4	НОМЕР МАТЧА-АЛS200807150		0.034		-0.195	-0.318		0.362	-0.067	0.787	-0.416	
5	НОМЕР МАТЧА-АЛS201007130		-0.476		0.316	-0.433		0.159	-0.182	-0.513	0.259	
6	НОМЕР МАТЧА-АЛS201207100		-0.418		0.144	0.252		-0.803	0.271	-0.060	0.154	
7	НОМЕР МАТЧА-АЛS201407150		-0.074		-0.072	0.200		-0.460	0.056	0.120	0.393	
8	НОМЕР МАТЧА-АЛS200207090	0.418	0.686		0.226	-0.292		0.388	-0.041	-0.372	-0.390	
9	НОМЕР МАТЧА-АЛS200407130	0.778	-0.370		-0.598	-0.326		-0.129	-0.471	0.220	0.366	
10	НОМЕР МАТЧА-АЛS200607110		-0.394		0.003	-0.351		0.010	0.614	-0.431	-0.449	
11	НОМЕР МАТЧА-АЛS200707100		0.273		-0.186	-0.309		0.283	-0.453	0.005	-0.407	
12	НОМЕР МАТЧА-АЛS200907140		-0.007		-0.631	-0.359		0.233	-0.503	-0.439	-0.457	
13	НОМЕР МАТЧА-АЛS201107120		-0.095	1.448	0.071	-0.447	1.448	0.042	-0.196	-0.527	0.246	
14	НОМЕР МАТЧА-АЛS201307160		-0.053		-0.050	0.513		-0.207	-0.154	-0.485	0.124	
15	НОМЕР МАТЧА-АЛS201507140		0.126			0.400		-0.655		0.089	0.071	
16	НОМЕР МАТЧА-АЛS201507141											
17	НОМЕР МАТЧА-АЛS201507142											
18	НОМЕР МАТЧА-АЛS201507143											
19	НОМЕР МАТЧА-АЛS201507144				1.806							
20	НОМЕР МАТЧА-АЛS201507145											
21	НОМЕР МАТЧА-АЛS201507146											
22	НОМЕР МАТЧА-АЛS201507147											
23	НОМЕР МАТЧА-АЛS201507148											
24	НОМЕР МАТЧА-АЛS201507149											

Рисунок 11- Матрица информативности (модель INF1) в битах (фрагмент)

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

Код признака	Наименование описательной шкалы и градации	1. КОМАНДА ANA	2. КОМАНДА ARI	3. КОМАНДА ARL	4. КОМАНДА ATL	5. КОМАНДА BAL	6. КОМАНДА BIS	7. КОМАНДА BOS	8. КОМАНДА CHA	9. КОМАНДА CHN	10. КОМАНДА CIN	11. КОМАНДА CL
1	НОМЕР МАТЧА-АЛS200107100	1.632	1.621	-0.046	-0.059	-1.278	-0.046	-2.711	-0.645	0.529	-0.517	
2	НОМЕР МАТЧА-АЛS200307150	2.475	-0.970	-0.066	4.059	-0.825	-0.066	-0.874	0.650	-0.101	-1.167	
3	НОМЕР МАТЧА-АЛS200507120	-0.510	-0.913	-0.064	-0.857	2.227	-0.064	1.237	1.717	-0.041	-1.105	
4	НОМЕР МАТЧА-АЛS200807150	-0.503	0.115	-0.063	-0.815	-0.747	-0.063	3.292	-0.250	5.989	-1.074	
5	НОМЕР МАТЧА-АЛS201007130	-0.615	-1.307	-0.077	2.554	-1.138	-0.077	1.462	-0.753	-1.461	1.462	
6	НОМЕР МАТЧА-АЛS201207100	-0.555	-1.082	-0.069	0.891	1.071	-0.069	-3.095	1.515	-0.221	0.710	
7	НОМЕР МАТЧА-АЛS201407150	-0.608	-0.279	-0.076	-0.403	0.888	-0.076	-2.482	0.280	0.569	2.493	
8	НОМЕР МАТЧА-АЛS200207090	0.520	4.199	-0.060	1.311	-0.669	-0.060	3.458	-0.149	-0.921	-0.981	
9	НОМЕР МАТЧА-АЛS200407130	1.490	-0.913	-0.064	-1.857	-0.773	-0.064	-0.763	-1.283	0.959	1.895	
10	НОМЕР МАТЧА-АЛS200607110	-0.533	-0.998	-0.067	0.017	-0.851	-0.067	0.071	4.616	-1.131	-1.198	
11	НОМЕР МАТЧА-АЛS200707100	-0.495	1.143	-0.062	-0.773	-0.721	-0.062	2.348	-1.216	0.019	-1.043	
12	НОМЕР МАТЧА-АЛS200907140	-0.540	-0.026	-0.068	-2.025	-0.877	-0.068	2.016	-1.418	-1.161	-1.228	
13	НОМЕР МАТЧА-АЛS201107120	-0.630	-0.364	0.921	0.470	-1.190	0.921	0.352	-0.820	-1.521	1.400	
14	НОМЕР МАТЧА-АЛS201307160	-0.585	-0.195	-0.073	-0.277	2.966	-0.073	-1.316	-0.619	-1.341	0.586	
15	НОМЕР МАТЧА-АЛS201507140	-0.428	0.396	-0.053	-2.395	1.514	-0.053	-2.154	-1.914	0.289	0.236	
16	НОМЕР МАТЧА-АЛS201507141	-0.008	-0.028	-0.001	-0.042	-0.026	-0.001	-0.055	-0.034	-0.030	-0.031	
17	НОМЕР МАТЧА-АЛS201507142	-0.008	-0.028	-0.001	-0.042	-0.026	-0.001	-0.055	-0.034	-0.030	-0.031	
18	НОМЕР МАТЧА-АЛS201507143	-0.008	-0.028	-0.001	-0.042	-0.026	-0.001	-0.055	-0.034	-0.030	-0.031	
19	НОМЕР МАТЧА-АЛS201507144	-0.008	-0.028	-0.001	0.958	-0.026	-0.001	-0.055	-0.034	-0.030	-0.031	
20	НОМЕР МАТЧА-АЛS201507145	-0.008	-0.028	-0.001	-0.042	-0.026	-0.001	-0.055	-0.034	-0.030	-0.031	
21	НОМЕР МАТЧА-АЛS201507146	-0.008	-0.028	-0.001	-0.042	-0.026	-0.001	-0.055	-0.034	-0.030	-0.031	
22	НОМЕР МАТЧА-АЛS201507147	-0.008	-0.028	-0.001	-0.042	-0.026	-0.001	-0.055	-0.034	-0.030	-0.031	
23	НОМЕР МАТЧА-АЛS201507148	-0.008	-0.028	-0.001	-0.042	-0.026	-0.001	-0.055	-0.034	-0.030	-0.031	
24	НОМЕР МАТЧА-АЛS201507149	-0.008	-0.028	-0.001	-0.042	-0.026	-0.001	-0.055	-0.034	-0.030	-0.031	

Рисунок 12- Матрица знаний (модель INF3) (фрагмент)

5.5. Модель: "7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRCI"

Код признака	Наименование описательной шкалы и градации	1. КОМАНДА ANA	2. КОМАНДА ARI	3. КОМАНДА ARL	4. КОМАНДА ATL	5. КОМАНДА BAL	6. КОМАНДА BIS	7. КОМАНДА BOS	8. КОМАНДА CHA	9. КОМАНДА CHN	10. КОМАНДА CIN	11. КОМАНДА CL
1	НОМЕР МАТЧА-АЛS200107100	4.440	1.176		-0.029				-0.392	0.360	-0.341	
2	НОМЕР МАТЧА-АЛS200307150	4.712	-0.492		1.380	-0.452		-0.226	0.276	-0.048	-0.538	
3	НОМЕР МАТЧА-АЛS200507120		-0.477		-0.300	1.256		0.329	0.752	-0.020	-0.525	
4	НОМЕР МАТЧА-АЛS200807150		0.061		-0.290	-0.428		0.888	-0.111	2.978	-0.518	
5	НОМЕР МАТЧА-АЛS201007130		-0.567		0.741	-0.532		0.322	-0.274	-0.594	0.576	
6	НОМЕР МАТЧА-АЛS201207100		-0.520		0.286	0.555		-0.756	0.610	-0.099	0.310	
7	НОМЕР МАТЧА-АЛS201407150		-0.122		-0.119	0.420		-0.554	0.103	0.234	0.994	
8	НОМЕР МАТЧА-АЛS200207090	1.082	2.332		0.487	-0.401		0.977	-0.069	-0.479	-0.495	
9	НОМЕР МАТЧА-АЛS200407130	2.920	-0.477		-0.650	-0.436		-0.203	-0.562	0.470	0.901	
10	НОМЕР МАТЧА-АЛS200607110		-0.499		0.006	-0.460		0.018	1.936	-0.531	-0.545	
11	НОМЕР МАТЧА-АЛS200707100		0.615		-0.279	-0.419		0.643	-0.549	0.010	-0.510	
12	НОМЕР МАТЧА-АЛS200907140		-0.013		-0.669	-0.467		0.506	-0.586	-0.537	-0.551	
13	НОМЕР МАТЧА-АЛS201107120		-0.154	11.693	0.133	-0.543	11.693	0.076	-0.291	-0.603	0.539	
14	НОМЕР МАТЧА-АЛS201307160		-0.089		-0.085	1.458		-0.305	-0.236	-0.573	0.243	
15	НОМЕР МАТЧА-АЛS201507140		0.247			1.019		-0.683		0.169	0.134	
16	НОМЕР МАТЧА-АЛS201507141											
17	НОМЕР МАТЧА-АЛS201507142											
18	НОМЕР МАТЧА-АЛS201507143											
19	НОМЕР МАТЧА-АЛS201507144				22.799							
20	НОМЕР МАТЧА-АЛS201507145											
21	НОМЕР МАТЧА-АЛS201507146											
22	НОМЕР МАТЧА-АЛS201507147											
23	НОМЕР МАТЧА-АЛS201507148											
24	НОМЕР МАТЧА-АЛS201507149											

Рисунок 13 -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.

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложноположительных решений (FP)	Число ложноотрицательных решений (FN)	Точность модели	Полнота модели	Фильтрация Ван Ризбергена	Сумма модулей сходства истинно-положительных решений (STR)	Сумма модулей сходства истинно-отрицательных решений (STN)	Сумма модулей сходства ложноположительных решений (SFP)
1. ABS - частный критерий: количество встреч соименных "класс"	Корреляция абс. частот с обр...	2140	2140	1172	38418		0.053	1.000	0.100	1214.286	10.712	11754.036
1. ABS - частный критерий: количество встреч соименных "класс"	Сумма абс. частот по признакам...	2140	2140	1172	38418		0.053	1.000	0.100	1006.011		2416.504
2. PRC1 - частный критерий: усл. вероятность i-го признака сред.	Корреляция усл.отн. частот с о...	2140	2140	1172	38418		0.053	1.000	0.100	1214.286	10.712	11754.036
2. PRC1 - частный критерий: усл. вероятность i-го признака сред.	Сумма усл.отн. частот по призна...	2140	2140	1172	38418		0.053	1.000	0.100	723.826		8366.896
3. PRC2 - частный критерий: условная вероятность i-го признака.	Корреляция усл.отн. частот с о...	2140	2140	1172	38418		0.053	1.000	0.100	1214.268	10.712	11753.927
3. PRC2 - частный критерий: условная вероятность i-го признака.	Сумма усл.отн. частот по призна...	2140	2140	1172	38418		0.053	1.000	0.100	721.252		8332.689
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в.	Семантический резонанс зна...	2140	1230	30182	9408	910	0.116	0.575	0.193	261.250	1093.344	328.933
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в.	Сумма знаний	2140	2131	20637	18953	9	0.101	0.996	0.184	457.617	682.595	688.140
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в.	Семантический резонанс зна...	2140	1230	30161	9429	910	0.115	0.575	0.192	261.214	1094.112	327.937
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в.	Сумма знаний	2140	2131	20534	19056	9	0.101	0.996	0.183	457.203	683.983	685.901
6. INF3 - частный критерий: Хи-квадрат: разности между факти...	Семантический резонанс зна...	2140	2126	25358	14232	14	0.130	0.993	0.230	679.274	1974.050	1174.157
6. INF3 - частный критерий: Хи-квадрат: разности между факти...	Сумма знаний	2140	2126	25358	14232	14	0.130	0.993	0.230	755.458	1324.488	569.614
7. INF4 - частный критерий: ROI (Return On Investment): веро...	Семантический резонанс зна...	2140	1087	38847	743	1053	0.594	0.508	0.548	340.268	1042.007	20.325
7. INF4 - частный критерий: ROI (Return On Investment): веро...	Сумма знаний	2140	2138	19547	20043	2	0.096	0.999	0.176	34.704	4.263	13.002
8. INF5 - частный критерий: ROI (Return On Investment): веро...	Семантический резонанс зна...	2140	1087	38847	743	1053	0.594	0.508	0.548	340.278	1041.889	20.345
8. INF5 - частный критерий: ROI (Return On Investment): веро...	Сумма знаний	2140	2138	19549	20041	2	0.096	0.999	0.176	34.525	4.255	12.924
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер.	Семантический резонанс зна...	2140	2099	20251	19339	41	0.098	0.981	0.178	630.744	1012.911	1443.777
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер.	Сумма знаний	2140	2126	17743	21847	14	0.089	0.993	0.163	287.141	454.899	1076.408
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	2140	2099	20199	19391	41	0.098	0.981	0.178	630.725	1024.431	1444.067
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	2140	2127	17604	21986	13	0.088	0.994	0.162	285.979	457.435	1071.332

Рисунок 14- Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели INF4 и INF5 при интегральном критерии «Семантический резонанс знаний». При этом точность модели составляет 0,960. Таким образом, уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 70%. 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++".

ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Предположим, модель дает такой прогноз: выпадет 1, 2, 3, 4, 5 или 6. В этом случае у нее будет 100% достоверность идентификации, т.е. не будет ни одного объекта, не отнесенного к тому классу, к которому он действительно относится, но при этом будет очень большая ошибка ложной идентификации, т.к. огромное количество объектов будет отнесено к классам, к которым они не относятся (и именно за счет этого у модели и будет очень высокая достоверность идентификации). Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.

ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Представим себе, что мы выбираем кубик с 6 гранями, и модель предсказывает, что не выпадет: 1, 2, 3, 4, 5 и 6, а что-то из этого естественно выпало. Конечно, модель дает ошибку в прогнозе в том плане, что не предсказала, что выпадет, зато она очень хорошо угадала, что не выпадет. Но ясно, что выпадет что-то одно, а не все, что предсказано, поэтому такого рода предсказания хорошо оправдываются в том, что не произошло и плохо в том, что произошло, т.е. в этом случае у модели будет 100% достоверность не идентификации, но очень низкая достоверность идентификации.

ИДЕАЛЬНЫЙ ПРОГНОЗ.
Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.

РЕАЛЬНЫЙ ПРОГНОЗ.
На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2, поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представьте себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, 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}:

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

$$L2 = \frac{(1 + (TP + TN - FP - FN) / (TP + TN + FP + FN))}{2}$$
 (нормировка: {0, 1})
 где: TP - истинно-положительное решение; TN - истинно-отрицательное решение; FP - ложноположительное решение; FN - ложноотрицательное 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:

$$SPrecision = STR / (STR + SFP)$$
 - нечеткая мультиклассовая точность модели;

$$SRecall = STN / (STN + SFN)$$
 - нечеткая мультиклассовая полнота модели;

$$L-mera = 2 * (SPrecision * SRecall) / (SPrecision + SRecall)$$
 - нечеткая мультиклассовая достоверность модели.

Где:
 STR - Сумма модулей сходства истинно-положительных решений; STN - Сумма модулей сходства истинно-отрицательных решений;
 SFP - Сумма модулей сходства ложноположительных решений; SFN - Сумма модулей сходства ложноотрицательных решений.

Рисунок 15- 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 INF4.

Рисунок 16- 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 INF4 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 14);

- модуль уровня сходства-различия в наиболее достоверной модели INF4 для верно идентифицированных и верно неидентифицированных объектов значительно выше, чем для ошибочно идентифицированных и ошибочно неидентифицированных. Это верно практически для всего диапазона уровней сходства-различия. Для очень больших значений уровней сходства-различия (более 95%) также различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями практически отсутствует.

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) и проведем пакетное распознавание в режиме 4.2.1, как показано на рисунке 18.

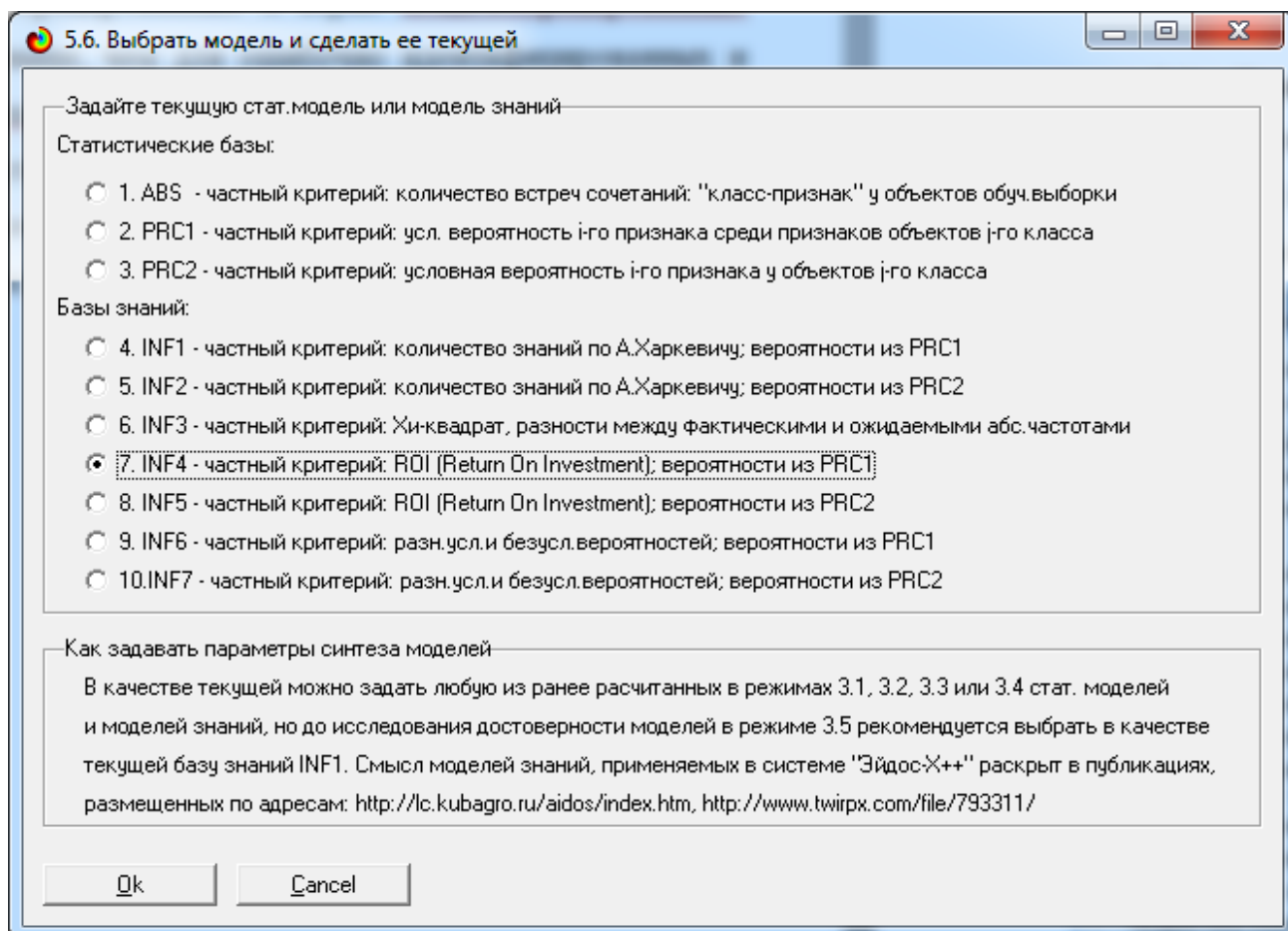


Рисунок 17- Экранная форма режима задания модели в качестве текущей

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

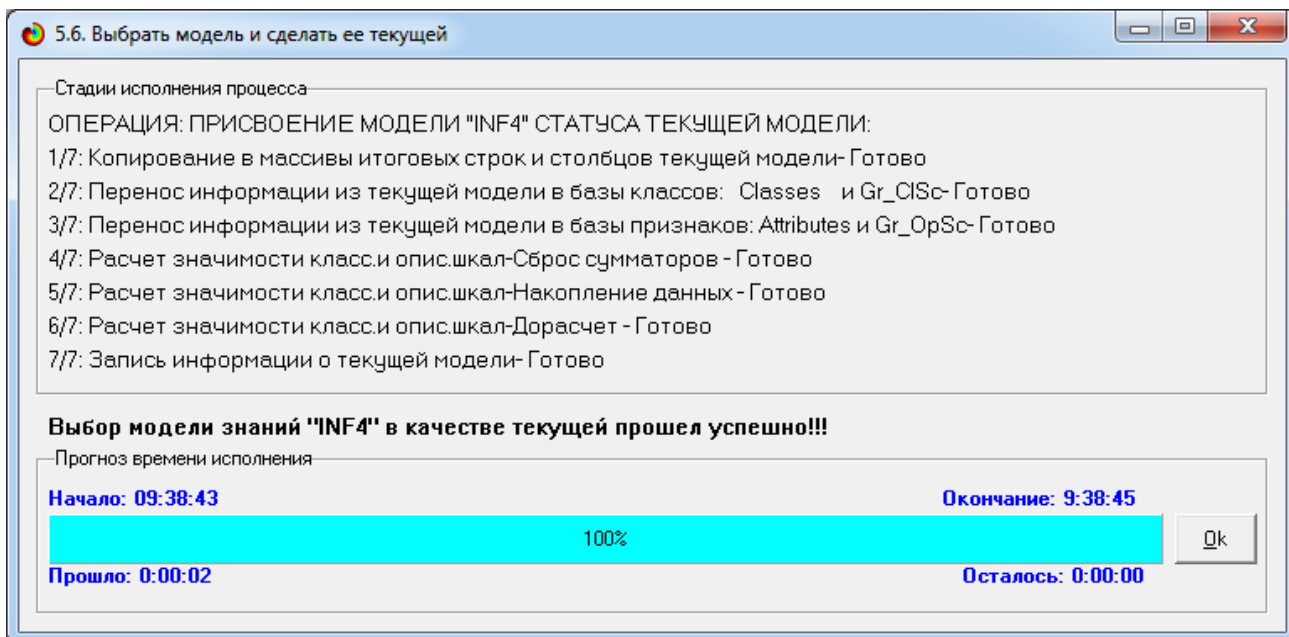


Рисунок 18- Экранная форма режима пакетного распознавания в текущей модели

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.

Для наглядного примера, кратко рассмотрим некоторые из них.

На рисунках 20 и 21 приведены примеры прогнозов высокой и низкой достоверности частоты и классов бейсбольных команд в наиболее достоверной модели INF4 на основе описательных шкал.

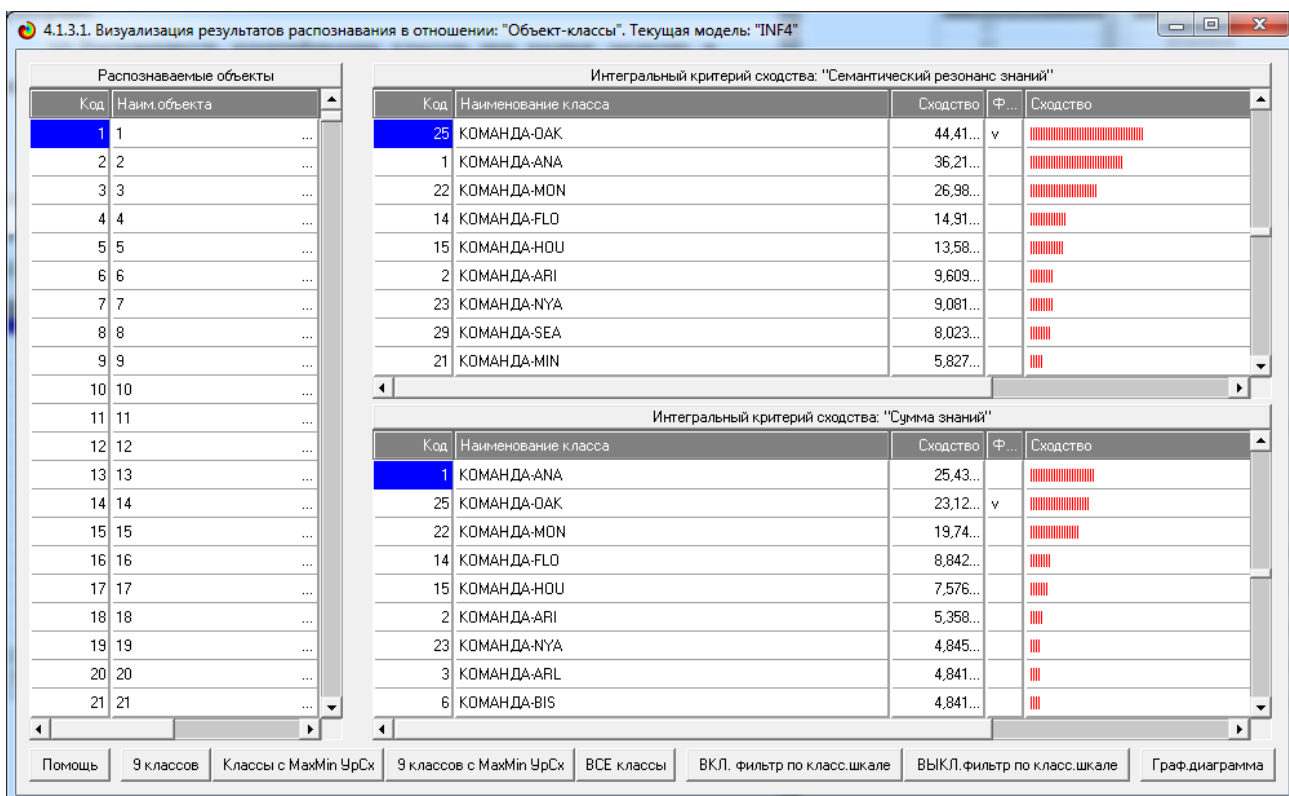


Рисунок 19- Пример идентификации классов в модели INF4

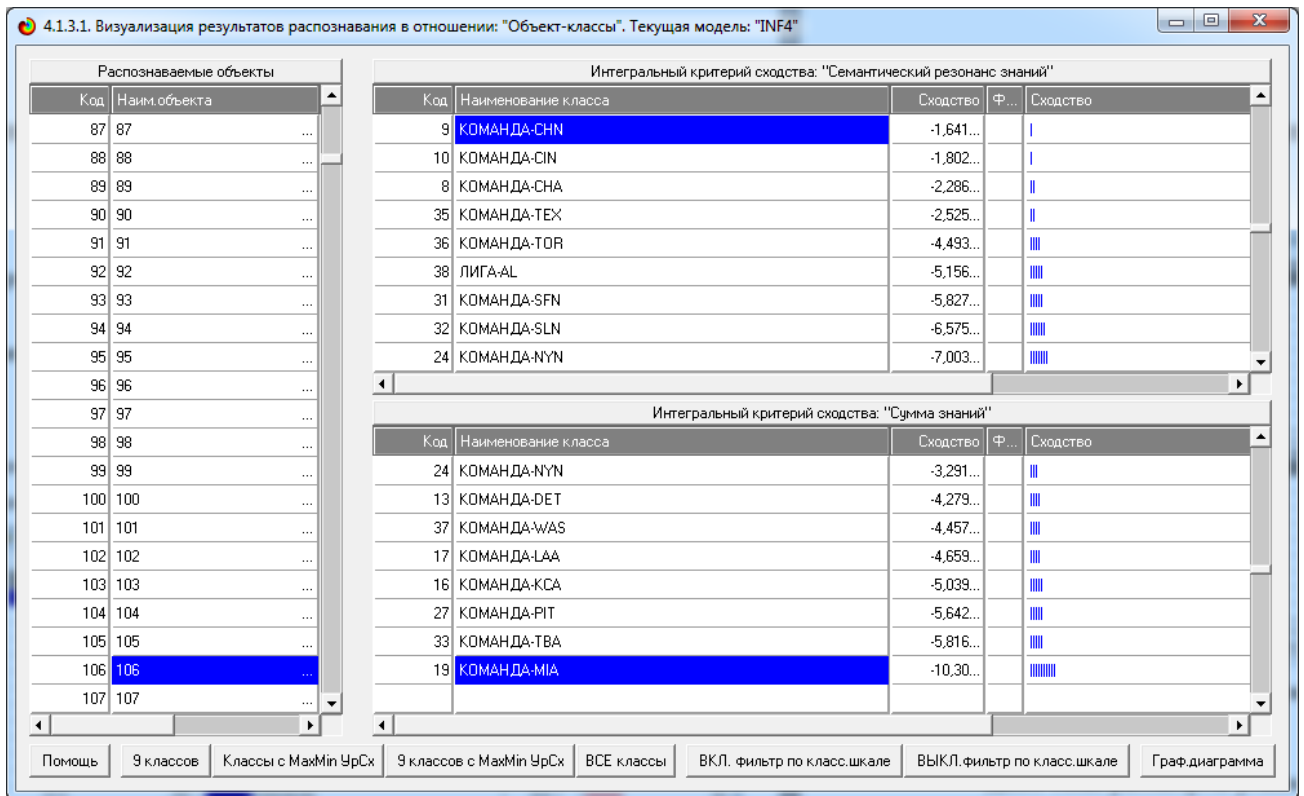


Рисунок 20- Пример идентификации классов в модели INF4

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 22.

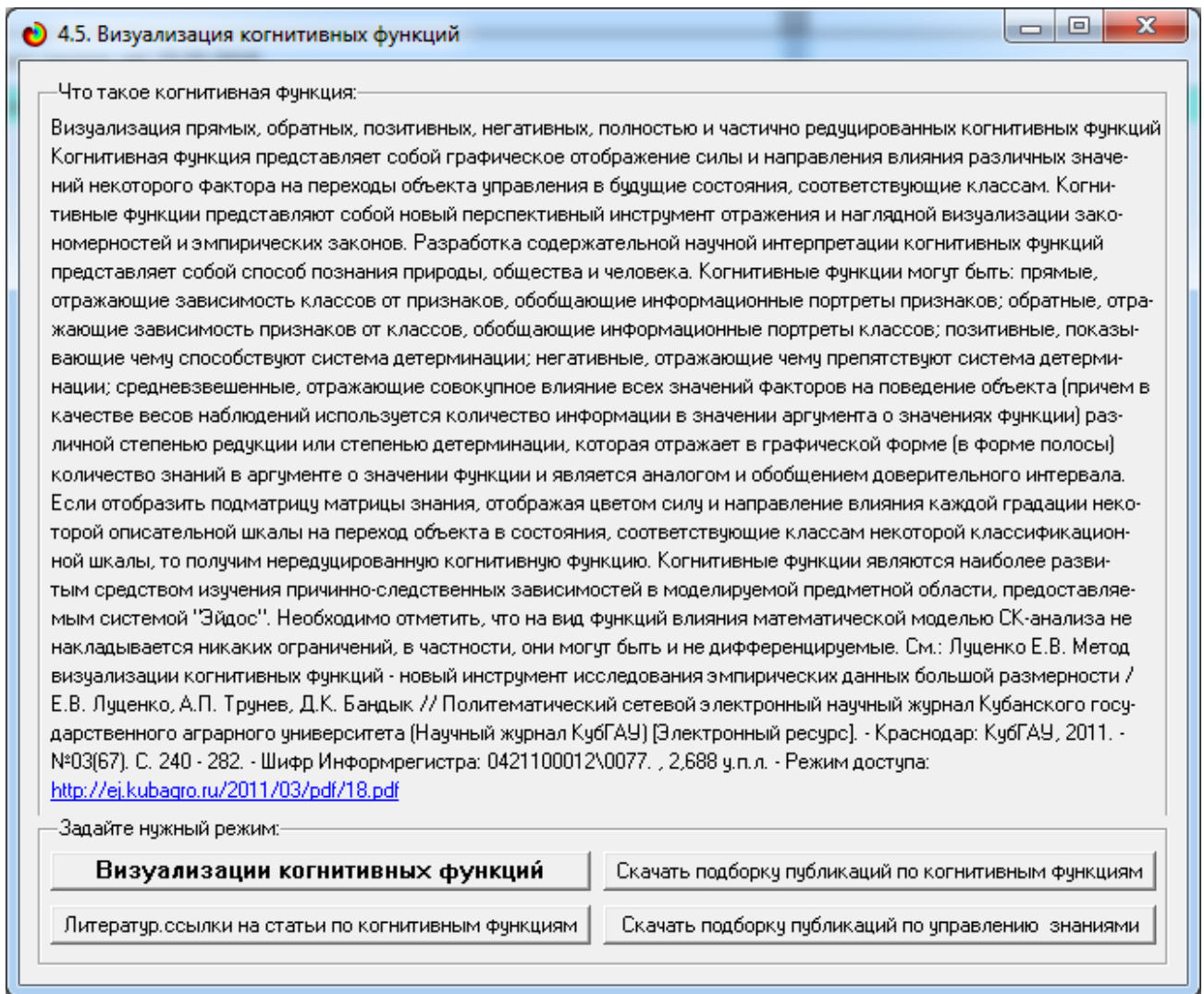


Рисунок 21- Экранная форма режима 4.5 системы «Эйдос-Х++»
«Визуализация когнитивных функций»

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 23 приведены визуализации всех когнитивных функций данного приложения для модели INF4.

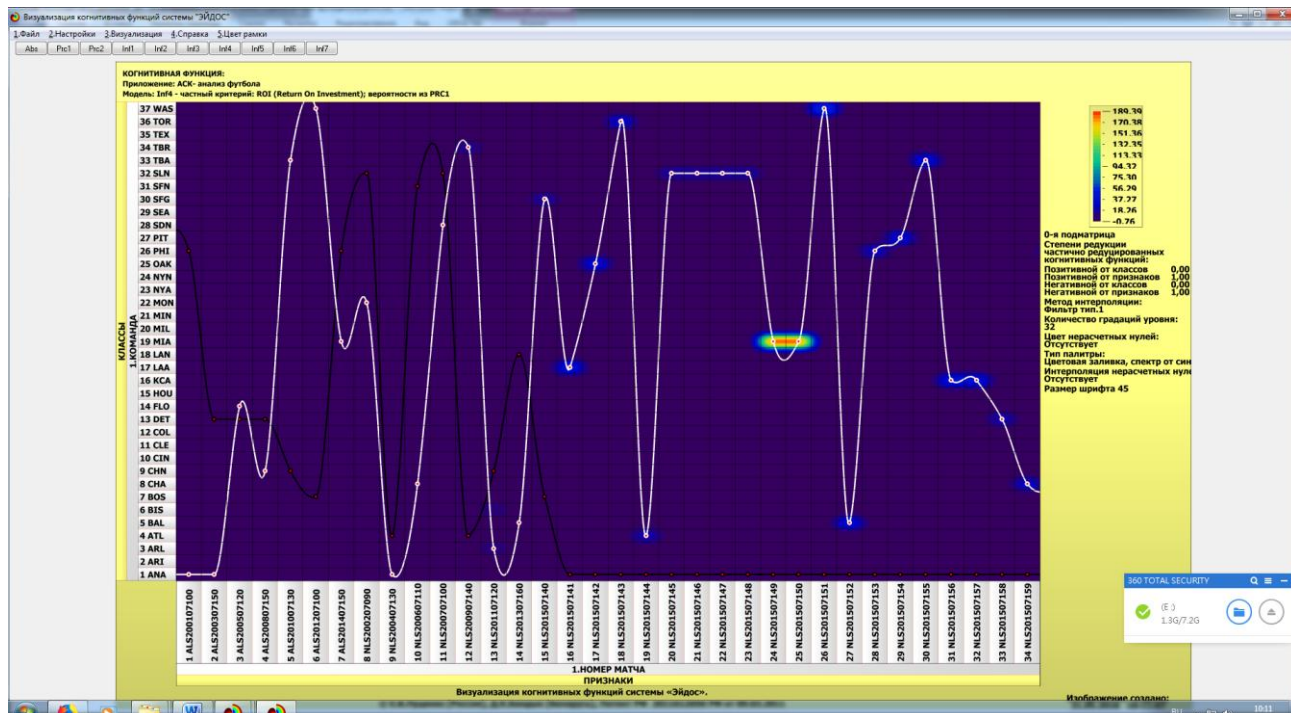


Рисунок 22-Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF4

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 24).

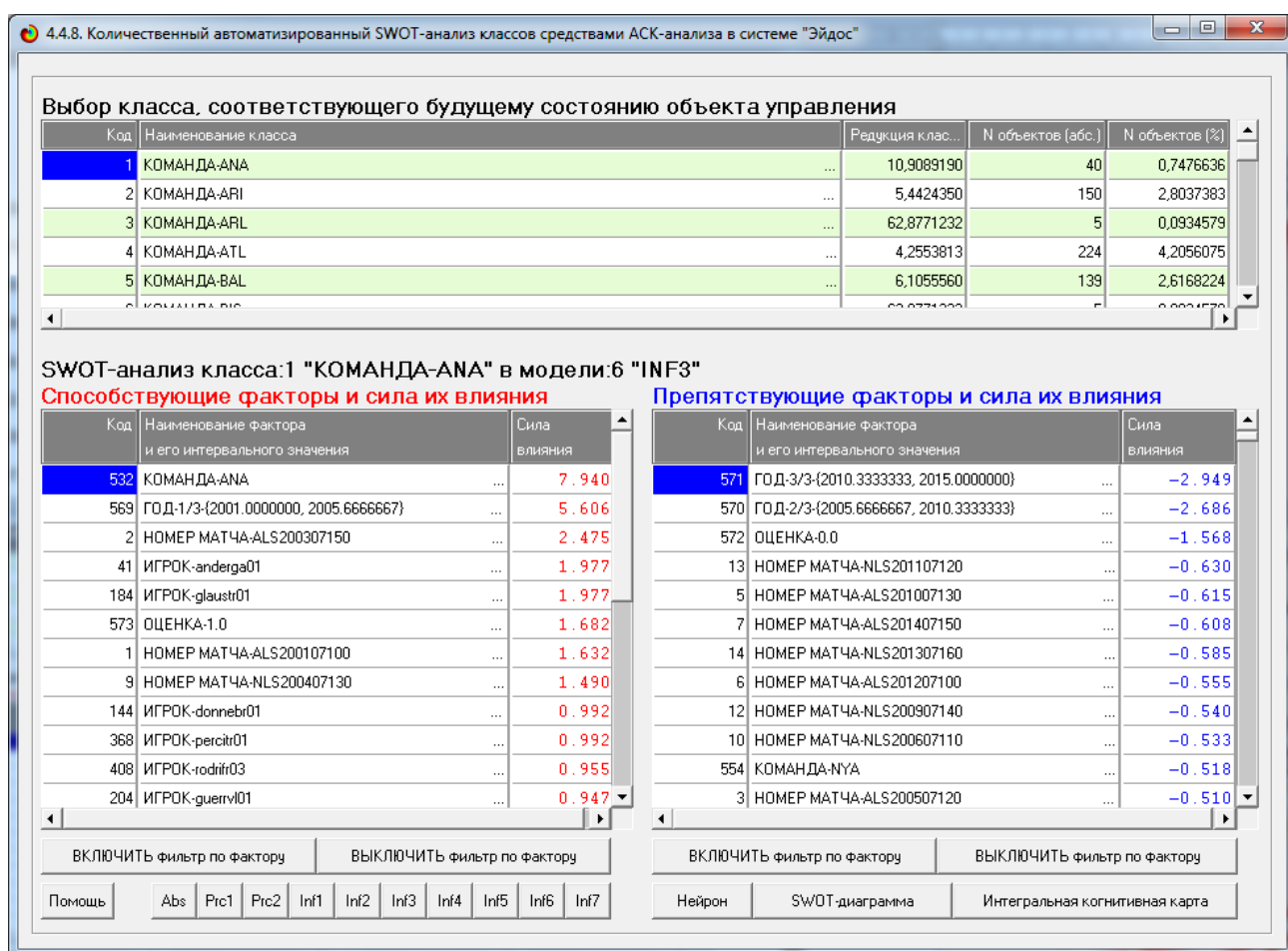


Рисунок 23-Пример SWOT-Матрицы в модели INF4

Следующим шагом является построение SWOT-диаграммы, для наглядной демонстрации работы, продемонстрированной на рисунке 25.

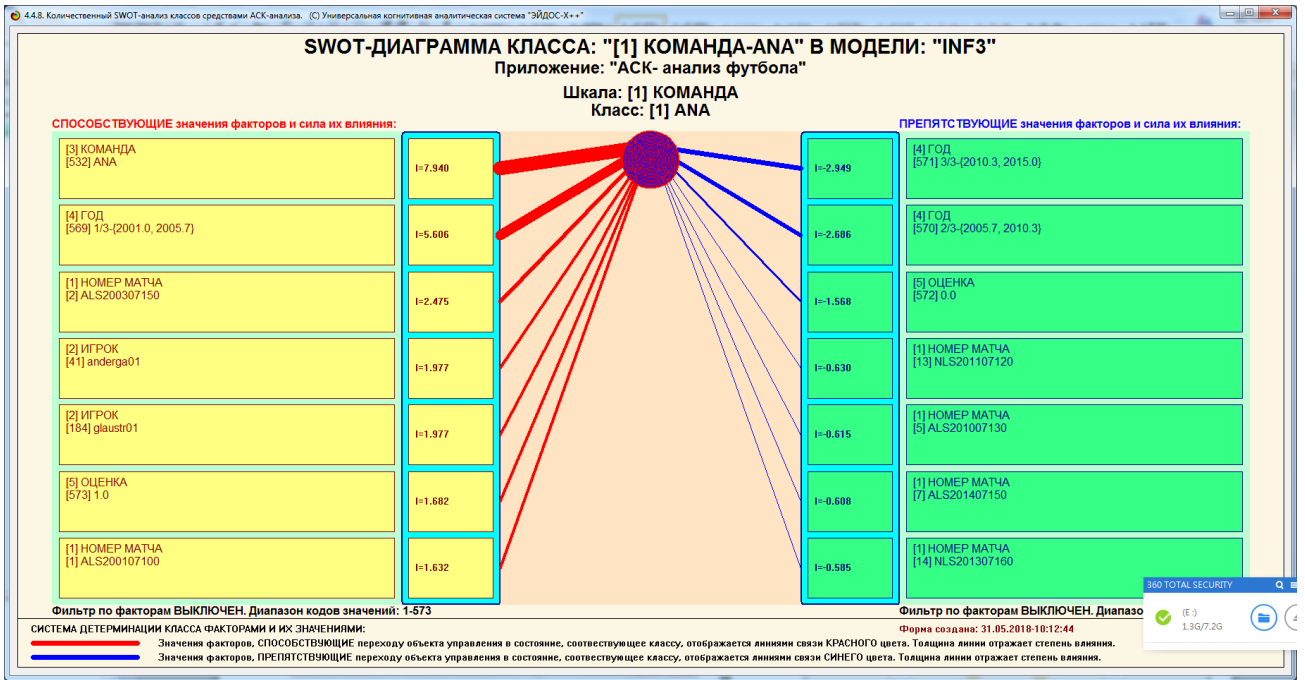


Рисунок 24– SWOT-матрица, построенная для 2001 года

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.

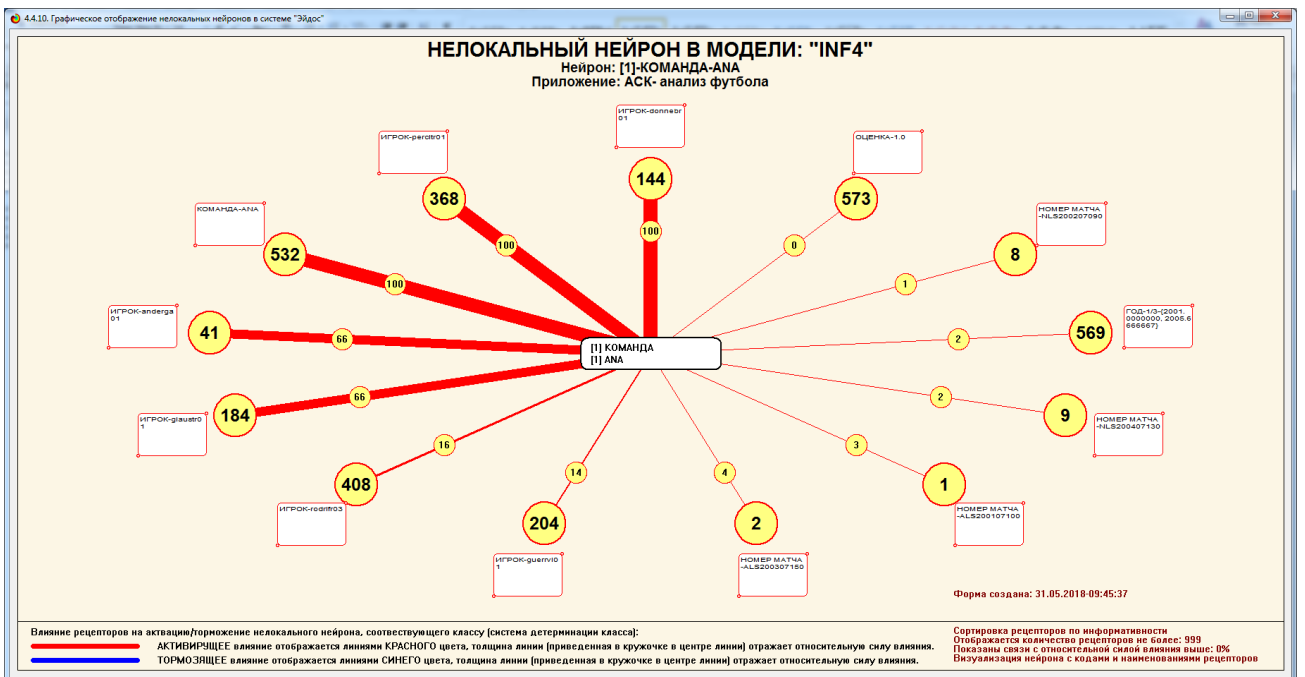


Рисунок 25- Нейрон для лиги NL

Благодаря данному нейрону видно, как игроки различных команд влияют на состояние лиги, какие оказывают активирующее влияние, а какие тормозящее.

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. На рисунке 26 наглядно видно, что наибольшее сходство наблюдается в классе «Лига AL», когда как наибольшее различие наблюдается в классе «Лига NL».

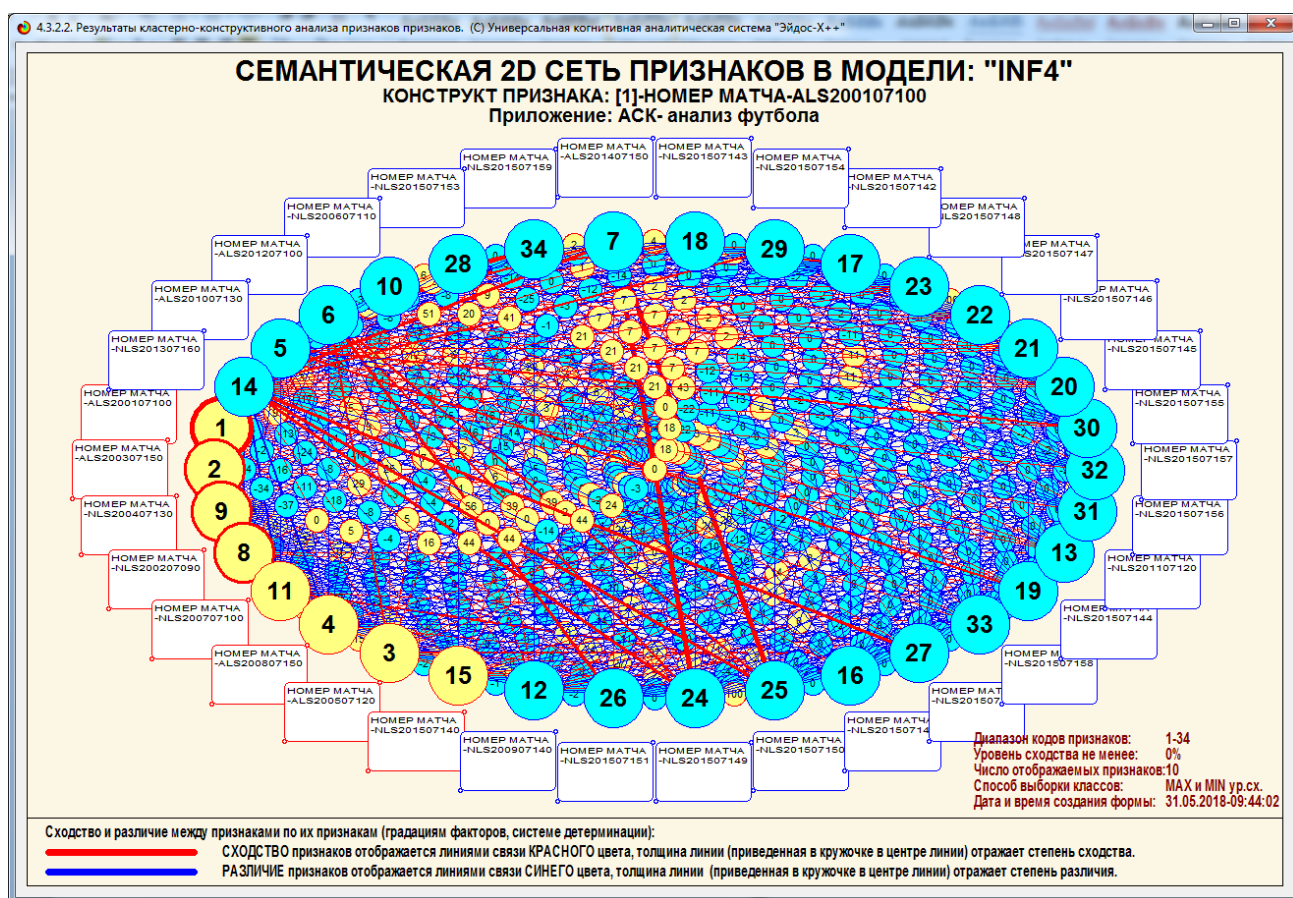


Рисунок 26– Семантическая 2D сеть классов

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 INF4, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,930, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 70%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием.

На основе базы данных «Kaggle», рассмотренной в данной 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 // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: http://lc.kubagro.ru/My_training_schedule.doc КубГАУ, 2003. – №02(002). С. 120 – 145. – IDA [article ID]: 0020302012. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 у.п.л.

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

3. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.

4. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических,

технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.

6. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: <http://ic.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.