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:

Исследование истории бейсбола с помощью АСК-анализа

с помощью «Эйдос » на основе данных репозитория UCI

выполнил студент группы ИТ1421 **Цепилов Даниил Олегович**

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)

Краснодар
2017 г.

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ ИМЕНИ И.Т. ТРУБИЛИНА»

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

УТВЕРЖДАЮ:
Зав. кафедрой КТС_____ В. И. Лойко

ЗАДАНИЕ на
курсовую работу

Студенту: __ИТ1421____ группы _3__ курса
Факультета прикладной информатики_____
Специальности: _09.03.02 Информационные системы и технологии_____ _
(шифр)

Цепилову Даниилу Олеговичу
(Ф.И.О.)

Тема проекта: **Исследовать историю бейсбола с помощью АСК-анализа с помощью «Эйдос» на основе репозитория UCI**

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 в файл MS Excel 5

1.3. Синтез и верификация статистических и интеллектуальных моделей 12

1.4. Виды моделей системы «Эйдос» 14

1.5. Результаты верификации моделей 16

2.РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ 22

2.1 Решение задачи идентификации 22

2.2 Когнитивные функции 25

2.3. SWOT и PEST матрицы и диаграммы 27

2.4. Нелокальные нейронные сети 29

2.5 Кластерный и конструктивный анализ 30

ЗАКЛЮЧЕНИЕ 31

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 32

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 бейсбольных 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 курсовой работе использована база данных «The History of Baseball» 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 Baseball»: https://www.kaggle.com/seanlahman/the-history-of-baseball , которую приведем к виду на рисунке 1.

	A	B	C	D	E	F	G	H
1	Нумерация	Команда	Лига	Номер матча	Игрок	Команда	Год	Оценка
2	1	OAK	AL	ALS200107100	giambja01	OAK	2001	1.0
3	2	ANA	AL	ALS200107100	glaustr01	ANA	2001	1.0
4	3	MIN	AL	ALS200107100	guzmacr0	MIN	2001	1.0
5	4	NYA	AL	ALS200107100	jeterde01	NYA	2001	1.0
6	5	MIN	AL	ALS200107100	maysjo01	MIN	2001	1.0
7	6	MIN	AL	ALS200107100	miltoer01	MIN	2001	0.0
8	7	SEA	AL	ALS200107100	nelsoje01	SEA	2001	1.0
9	8	CHA	AL	ALS200107100	ordonma(	CHA	2001	1.0
10	9	ANA	AL	ALS200107100	percitr01	ANA	2001	1.0
11	10	NYA	AL	ALS200107100	pettian01	NYA	2001	1.0
12	11	NYA	AL	ALS200107100	posadjo01	NYA	2001	1.0
13	12	TOR	AL	ALS200107100	quantpa0	TOR	2001	1.0
14	13	NYA	AL	ALS200107100	riverma01	NYA	2001	0.0
15	14	SEA	AL	ALS200107100	sasakka01	SEA	2001	1.0
16	15	NYA	AL	ALS200107100	stantmi02	NYA	2001	1.0
17	16	KCA	AL	ALS200107100	sweenmi(	KCA	2001	1.0
18	17	TBA	AL	ALS200107100	vaughgr01	TBA	2001	0.0
19	18	NYA	AL	ALS200107100	willibe02	NYA	2001	1.0
20	19	ARI	NL	ALS200107100	johnsra05	ARI	2001	1.0
21	20	NYN	NL	ALS200107100	piazzmi01	NYN	2001	1.0
22	21	COL	NL	ALS200107100	heltoto01	COL	2001	1.0
23	22	SFN	NL	ALS200107100	kentje01	SFN	2001	1.0
24	23	ATL	NL	ALS200107100	jonesch06	ATL	2001	1.0
25	24	SFN	NL	ALS200107100	aurilri01	SFN	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 «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.

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.

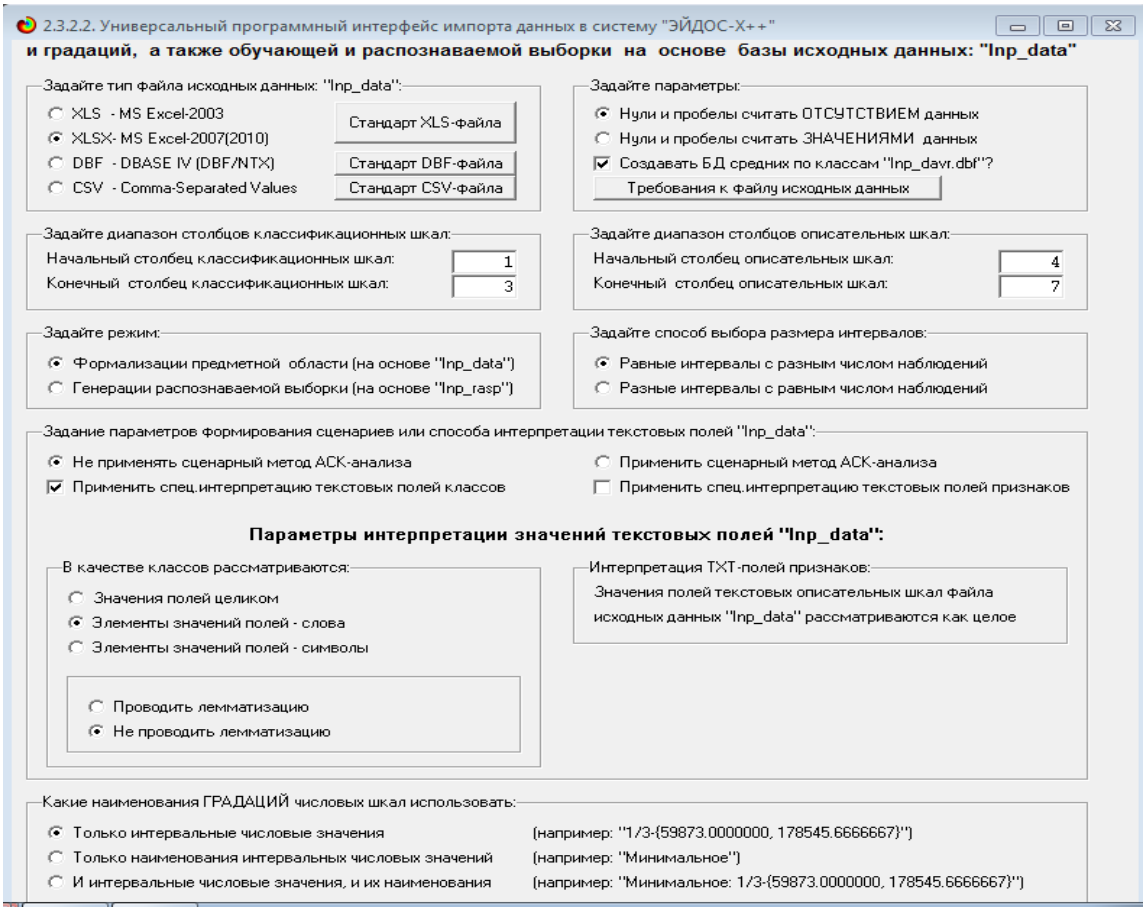


Рисунок 2- Экранная форма универсального программного интерфейса импорта данных в систему «Эйдос»

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 TXT-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.

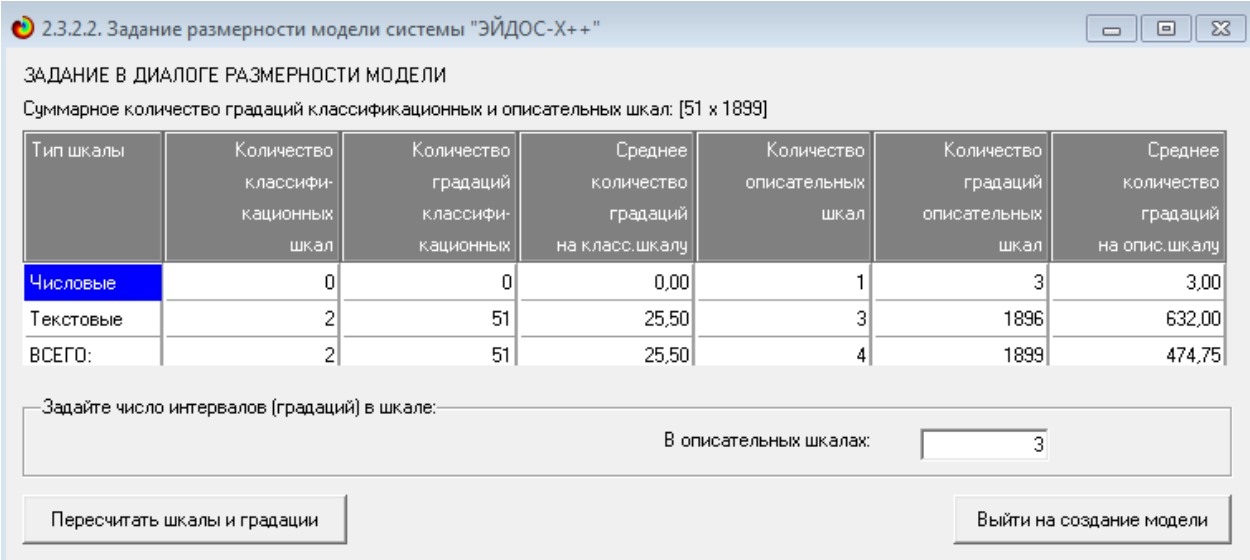


Рисунок 3- Задание размерности модели системы «Эйдос»

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.

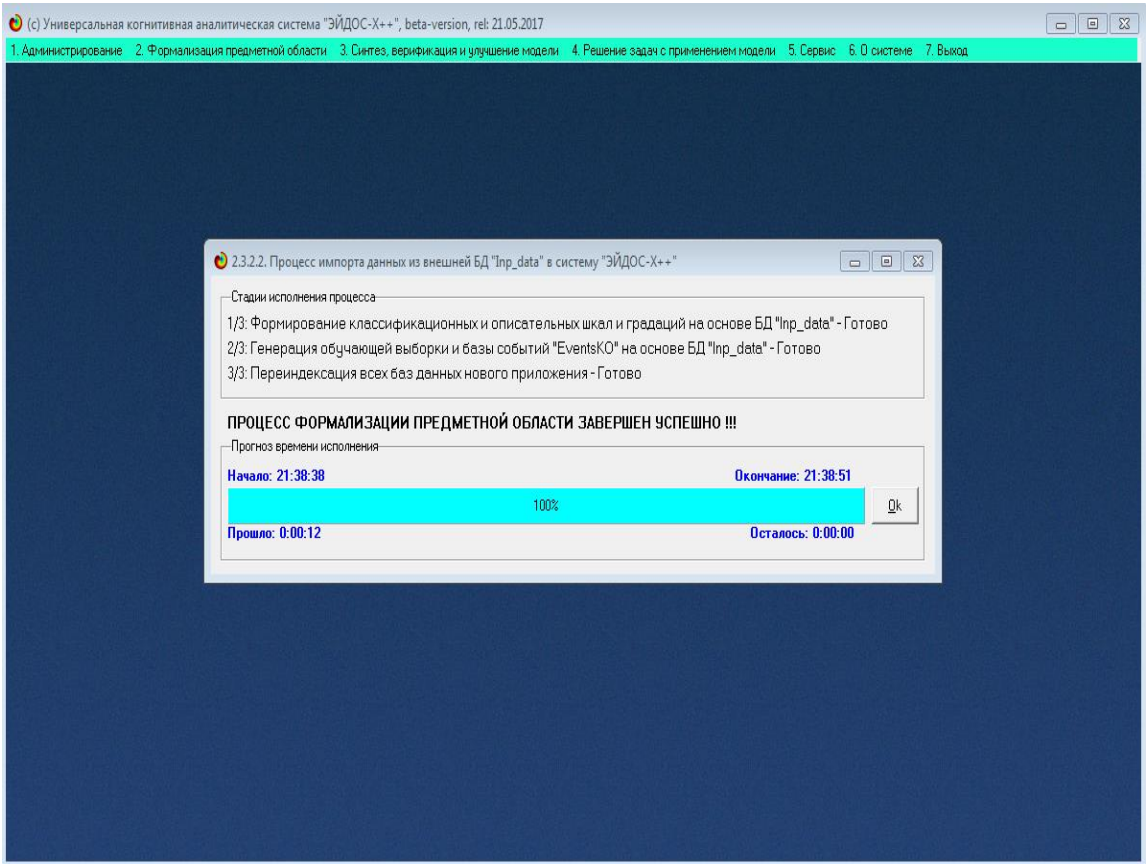


Рисунок 4- Процесс импорта данных из внешней БД «Inp_data.xls» в систему «Эйдос»

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.

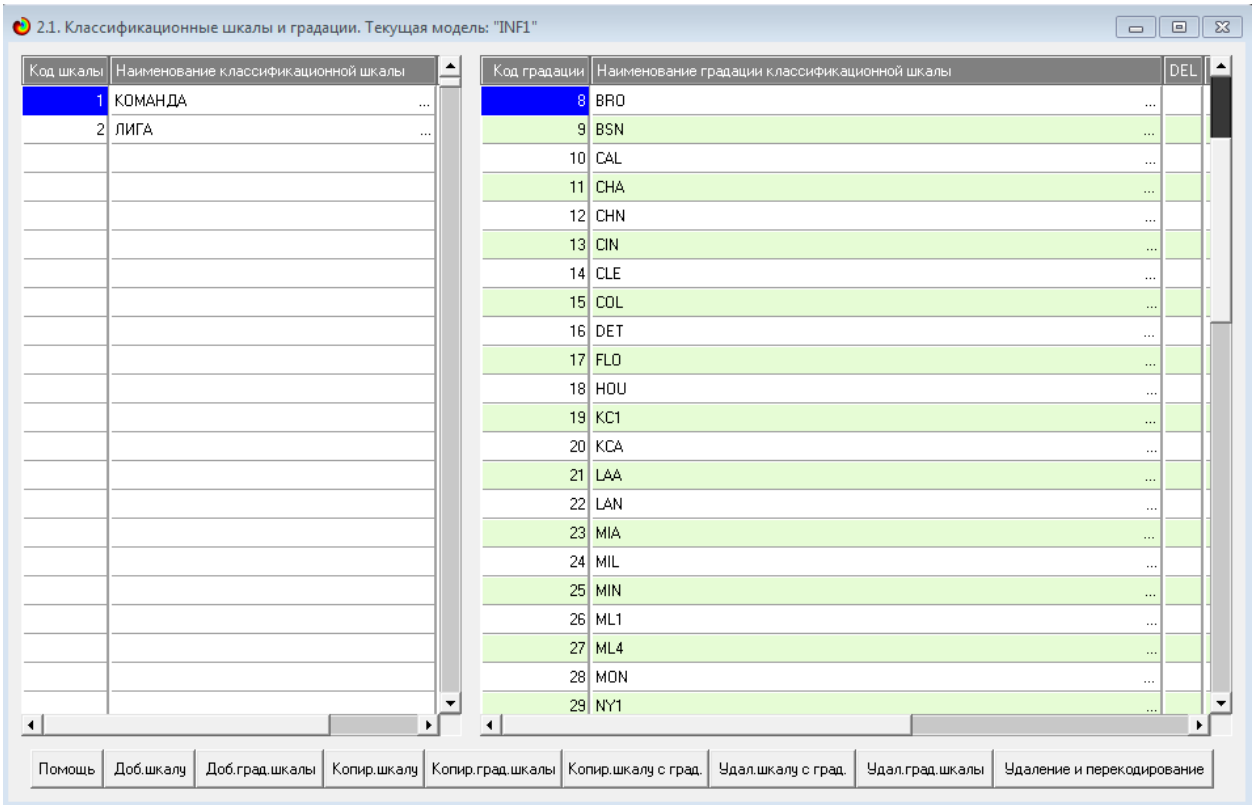


Рисунок 5- Классификационные шкалы и градации (фрагмент)

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

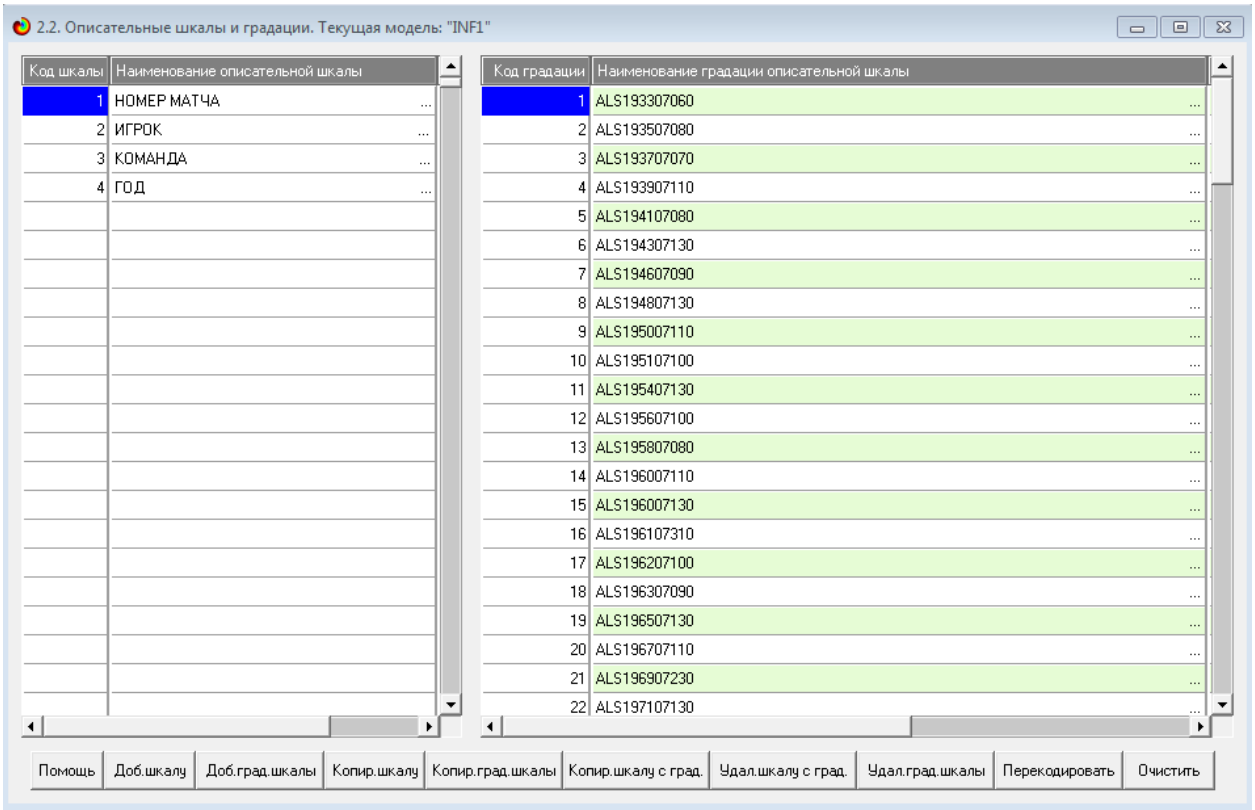


Рисунок 6- Описательные шкалы и градации (фрагмент)

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

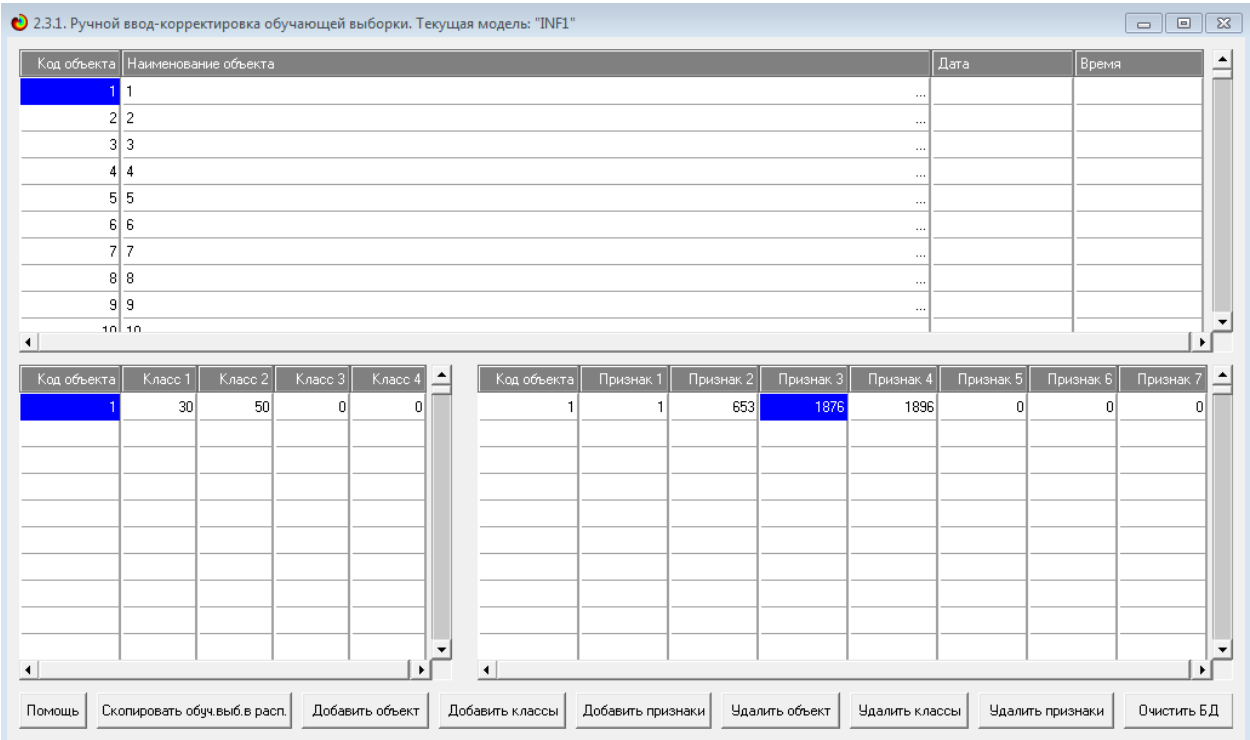


Рисунок 7- Обучающая выборка (фрагмент)

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.

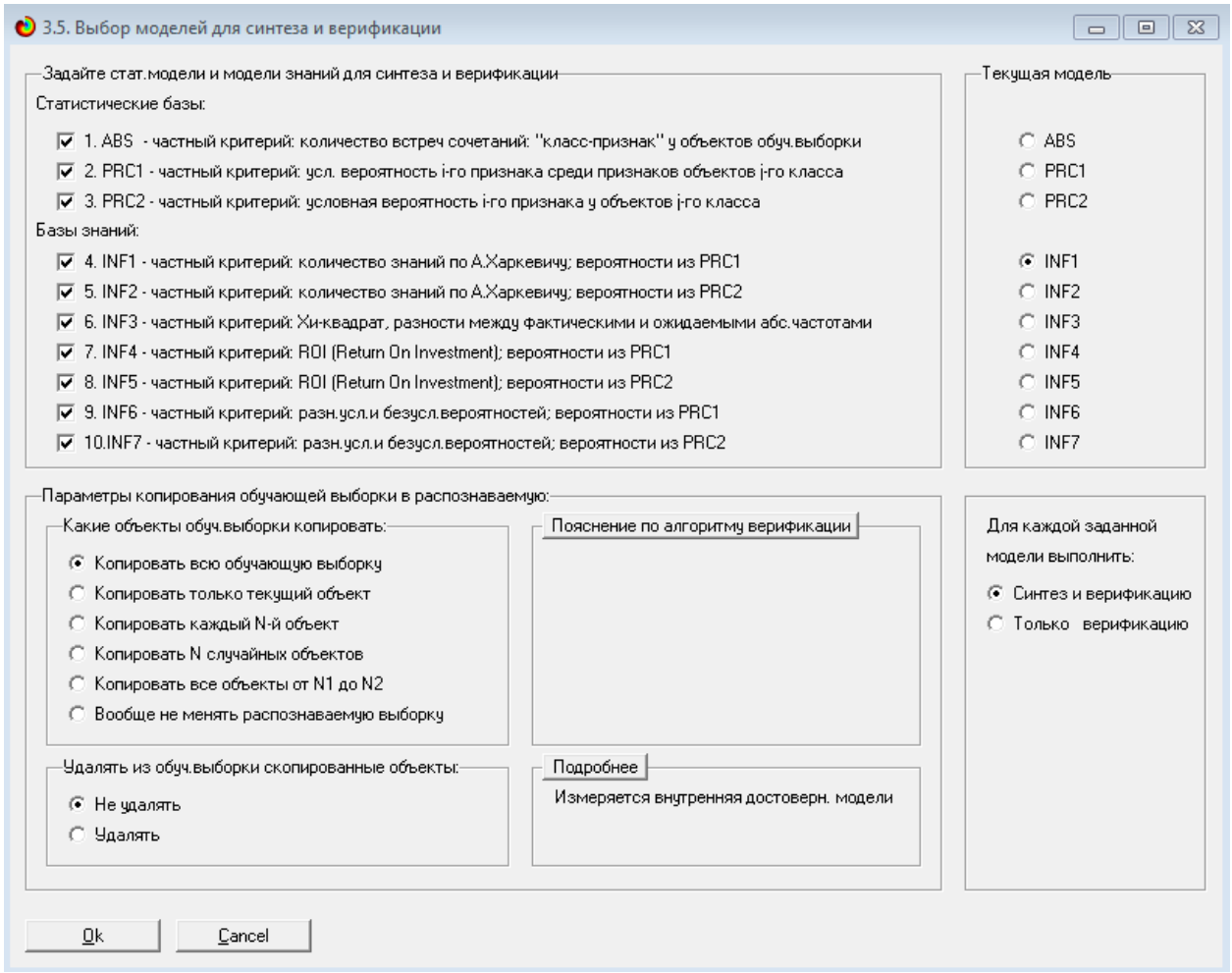


Рисунок 8- Выбор моделей для синтеза и верификации, а также текущей модели

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.

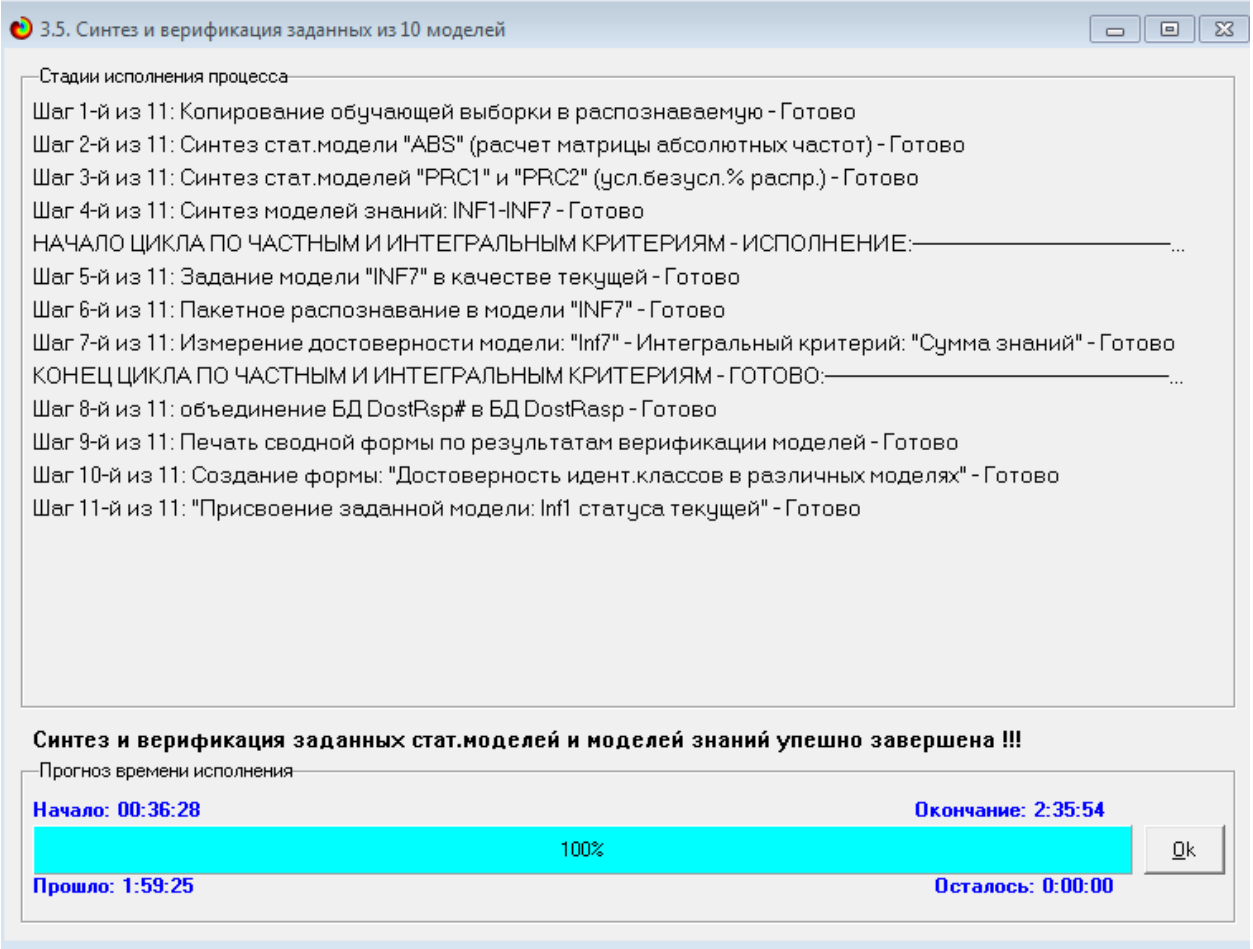


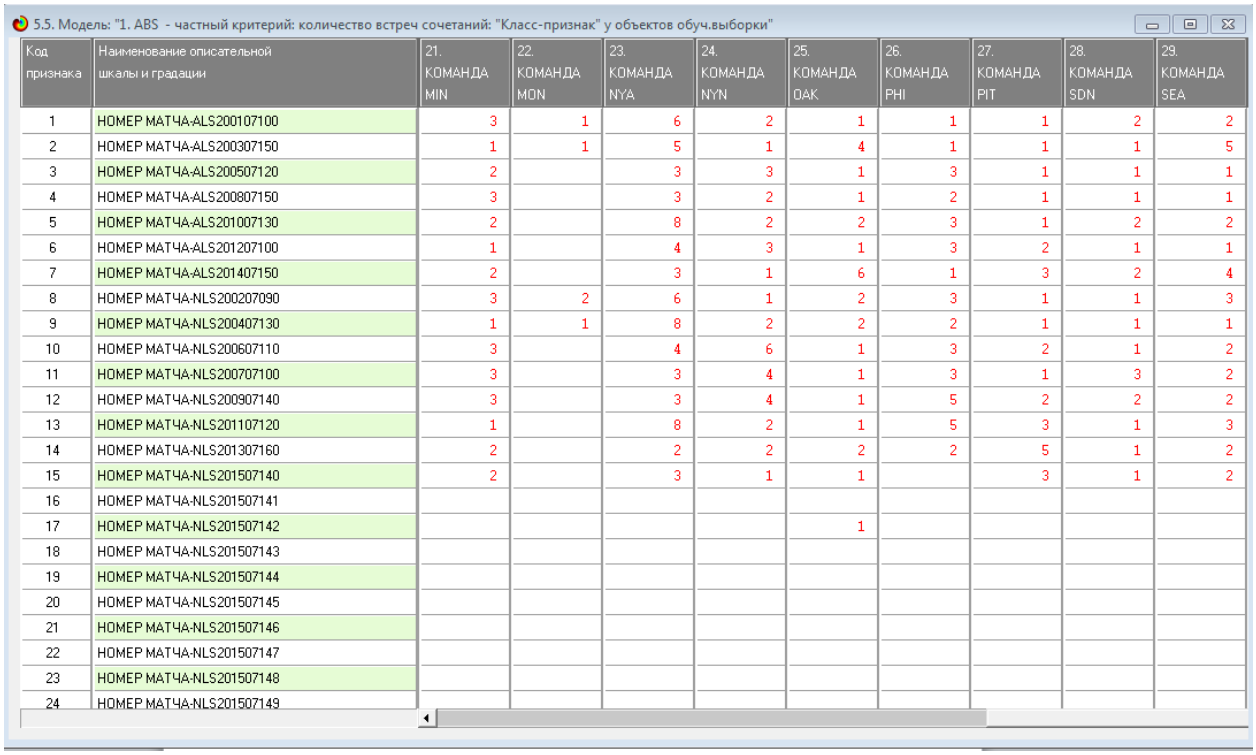
Рисунок 9- Синтез и верификация статистических моделей и моделей знаний

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	21. КОМАНДА MIN	22. КОМАНДА MON	23. КОМАНДА NYA	24. КОМАНДА NYN	25. КОМАНДА OAK	26. КОМАНДА PHI	27. КОМАНДА PIT	28. КОМАНДА SDN	29. КОМАНДА SEA
1	НОМЕР МАТЧА-ALS200107100	3	1	6	2	1	1	1	2	2
2	НОМЕР МАТЧА-ALS200307150	1	1	5	1	4	1	1	1	5
3	НОМЕР МАТЧА-ALS200507120	2		3	3	1	3	1	1	1
4	НОМЕР МАТЧА-ALS200807150	3		3	2	1	2	1	1	1
5	НОМЕР МАТЧА-ALS201007130	2		8	2	2	3	1	2	2
6	НОМЕР МАТЧА-ALS201207100	1		4	3	1	3	2	1	1
7	НОМЕР МАТЧА-ALS201407150	2		3	1	6	1	3	2	4
8	НОМЕР МАТЧА-NLS200207090	3	2	6	1	2	3	1	1	3
9	НОМЕР МАТЧА-NLS200407130	1	1	8	2	2	2	1	1	1
10	НОМЕР МАТЧА-NLS200607110	3		4	6	1	3	2	1	2
11	НОМЕР МАТЧА-NLS200707100	3		3	4	1	3	1	3	2
12	НОМЕР МАТЧА-NLS200907140	3		3	4	1	5	2	2	2
13	НОМЕР МАТЧА-NLS201107120	1		8	2	1	5	3	1	3
14	НОМЕР МАТЧА-NLS201307160	2		2	2	2	2	5	1	2
15	НОМЕР МАТЧА-NLS201507140	2		3	1	1		3	1	2
16	НОМЕР МАТЧА-NLS201507141									
17	НОМЕР МАТЧА-NLS201507142					1				
18	НОМЕР МАТЧА-NLS201507143									
19	НОМЕР МАТЧА-NLS201507144									
20	НОМЕР МАТЧА-NLS201507145									
21	НОМЕР МАТЧА-NLS201507146									
22	НОМЕР МАТЧА-NLS201507147									
23	НОМЕР МАТЧА-NLS201507148									
24	НОМЕР МАТЧА-NLS201507149									

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"

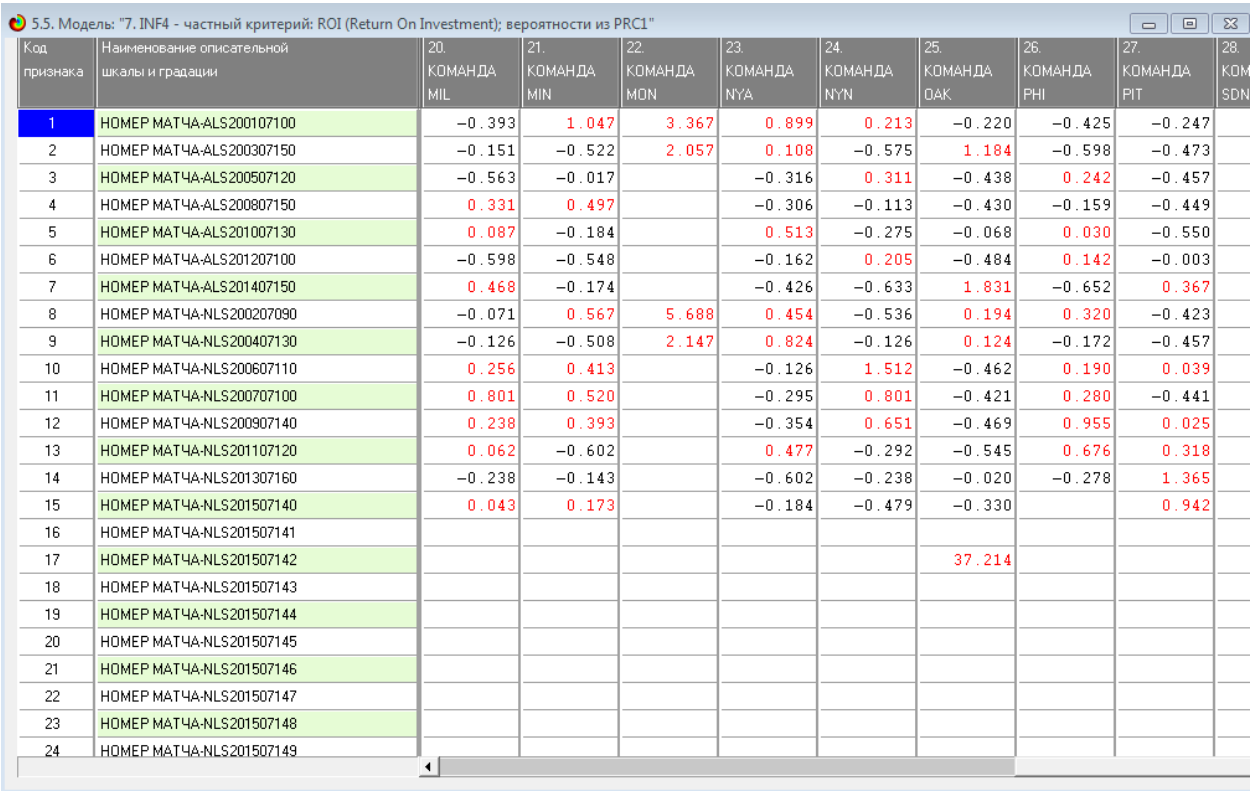
Код признака	Наименование описательной шкалы и градации	20. КОМАНДА MIL	21. КОМАНДА MIN	22. КОМАНДА MON	23. КОМАНДА NYA	24. КОМАНДА NYN	25. КОМАНДА OAK	26. КОМАНДА PHI	27. КОМАНДА PIT	28. КОМАНДА SDN	29. КОМАНДА SEA
1	НОМЕР МАТЧА-ALS200107100	-0.292	0.418	0.860	0.374	0.113	-0.145	-0.323	-0.166	0.427	0.164
2	НОМЕР МАТЧА-ALS200307150	-0.095	-0.431	0.652	0.060	-0.500	0.456	-0.532	-0.374	-0.185	0.490
3	НОМЕР МАТЧА-ALS200507120	-0.483	-0.010		-0.222	0.158	-0.336	0.127	-0.357	-0.168	-0.432
4	НОМЕР МАТЧА-ALS200807150	0.167	0.236		-0.213	-0.070	-0.328	-0.101	-0.348	-0.160	-0.424
5	НОМЕР МАТЧА-ALS201007130	0.049	-0.119		0.242	-0.188	-0.041	0.017	-0.466	0.127	-0.137
6	НОМЕР МАТЧА-ALS201207100	-0.532	-0.464		-0.103	0.109	-0.386	0.077	-0.002	-0.218	-0.482
7	НОМЕР МАТЧА-ALS201407150	0.224	-0.112		-0.324	-0.585	0.607	-0.617	0.182	0.134	0.275
8	НОМЕР МАТЧА-NLS200207090	-0.043	0.262	1.109	0.218	-0.448	0.104	0.162	-0.321	-0.133	0.244
9	НОМЕР МАТЧА-NLS200407130	-0.078	-0.414	0.669	0.351	-0.078	0.068	-0.110	-0.357	-0.168	-0.432
10	НОМЕР МАТЧА-NLS200607110	0.133	0.202		-0.079	0.538	-0.362	0.101	0.023	-0.194	-0.053
11	НОМЕР МАТЧА-NLS200707100	0.344	0.244		-0.204	0.344	-0.319	0.144	-0.339	0.490	-0.010
12	НОМЕР МАТЧА-NLS200907140	0.125	0.194		-0.255	0.293	-0.370	0.391	0.014	0.203	-0.061
13	НОМЕР МАТЧА-NLS201107120	0.035	-0.538		0.228	-0.202	-0.460	0.301	0.161	-0.292	0.086
14	НОМЕР МАТЧА-NLS201307160	-0.159	-0.090		-0.538	-0.159	-0.012	-0.190	0.502	-0.249	-0.108
15	НОМЕР МАТЧА-NLS201507140	0.025	0.093		-0.119	-0.380	-0.233		0.387	-0.065	0.075
16	НОМЕР МАТЧА-NLS201507141										
17	НОМЕР МАТЧА-NLS201507142						2.127				
18	НОМЕР МАТЧА-NLS201507143										
19	НОМЕР МАТЧА-NLS201507144										
20	НОМЕР МАТЧА-NLS201507145										
21	НОМЕР МАТЧА-NLS201507146										
22	НОМЕР МАТЧА-NLS201507147										
23	НОМЕР МАТЧА-NLS201507148										
24	НОМЕР МАТЧА-NLS201507149										

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	19. КОМАНДА MIA	20. КОМАНДА MIL	21. КОМАНДА MIN	22. КОМАНДА MON	23. КОМАНДА NYA	24. КОМАНДА NYN	25. КОМАНДА OAK	26. КОМАНДА PHI	27. КОМАНДА PIT
1	НОМЕР МАТЧА-ALS200107100	-0.275	-0.649	1.535	0.771	2.840	0.351	-0.282	-0.740	-0.328
2	НОМЕР МАТЧА-ALS200307150	-0.393	-0.355	-1.093	0.673	0.486	-1.355	2.168	-1.486	-0.897
3	НОМЕР МАТЧА-ALS200507120	-0.381	-1.288	-0.034	-0.318	-1.385	0.712	-0.779	0.585	-0.843
4	НОМЕР МАТЧА-ALS200807150	-0.376	0.746	0.996	-0.313	-1.321	-0.254	-0.753	-0.379	-0.816
5	НОМЕР МАТЧА-ALS201007130	-0.460	0.241	-0.452	-0.383	2.712	-0.759	-0.146	0.088	-1.222
6	НОМЕР МАТЧА-ALS201207100	0.585	-1.490	-1.213	-0.346	-0.772	0.510	-0.936	0.372	-0.006
7	НОМЕР МАТЧА-ALS201407150	1.546	1.275	-0.422	-0.379	-2.223	-1.725	3.880	-1.877	0.805
8	НОМЕР МАТЧА-NLS200207090	-0.359	-0.153	1.086	1.701	1.873	-1.153	0.325	0.727	-0.735
9	НОМЕР МАТЧА-NLS200407130	-0.381	-0.288	-1.034	0.682	3.615	-0.288	0.221	-0.415	-0.843
10	НОМЕР МАТЧА-NLS200607110	-0.398	0.611	0.877	-0.332	-0.579	3.611	-0.858	0.479	0.076
11	НОМЕР МАТЧА-NLS200707100	-0.370	1.779	1.026	-0.308	-1.256	1.779	-0.727	0.656	-0.789
12	НОМЕР МАТЧА-NLS200907140	-0.404	0.578	0.847	-0.336	-1.643	1.578	-0.884	2.443	0.049
13	НОМЕР МАТЧА-NLS201107120	-0.471	0.174	-1.512	-0.393	2.583	-0.826	-1.198	2.017	0.723
14	НОМЕР МАТЧА-NLS201307160	0.563	-0.624	-0.333	-0.364	-3.030	-0.624	-0.041	-0.770	2.886
15	НОМЕР МАТЧА-NLS201507140	-0.320	0.082	0.295	-0.266	-0.676	-0.918	-0.492	-2.024	1.455
16	НОМЕР МАТЧА-NLS201507141	-0.006	-0.034	-0.030	-0.005	-0.064	-0.034	-0.026	-0.036	-0.027
17	НОМЕР МАТЧА-NLS201507142	-0.006	-0.034	-0.030	-0.005	-0.064	-0.034	0.974	-0.036	-0.027
18	НОМЕР МАТЧА-NLS201507143	-0.006	-0.034	-0.030	-0.005	-0.064	-0.034	-0.026	-0.036	-0.027
19	НОМЕР МАТЧА-NLS201507144	-0.006	-0.034	-0.030	-0.005	-0.064	-0.034	-0.026	-0.036	-0.027
20	НОМЕР МАТЧА-NLS201507145	-0.006	-0.034	-0.030	-0.005	-0.064	-0.034	-0.026	-0.036	-0.027
21	НОМЕР МАТЧА-NLS201507146	-0.006	-0.034	-0.030	-0.005	-0.064	-0.034	-0.026	-0.036	-0.027
22	НОМЕР МАТЧА-NLS201507147	-0.006	-0.034	-0.030	-0.005	-0.064	-0.034	-0.026	-0.036	-0.027
23	НОМЕР МАТЧА-NLS201507148	-0.006	-0.034	-0.030	-0.005	-0.064	-0.034	-0.026	-0.036	-0.027
24	НОМЕР МАТЧА-NLS201507149	0.994	-0.034	-0.030	-0.005	-0.064	-0.034	-0.026	-0.036	-0.027

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



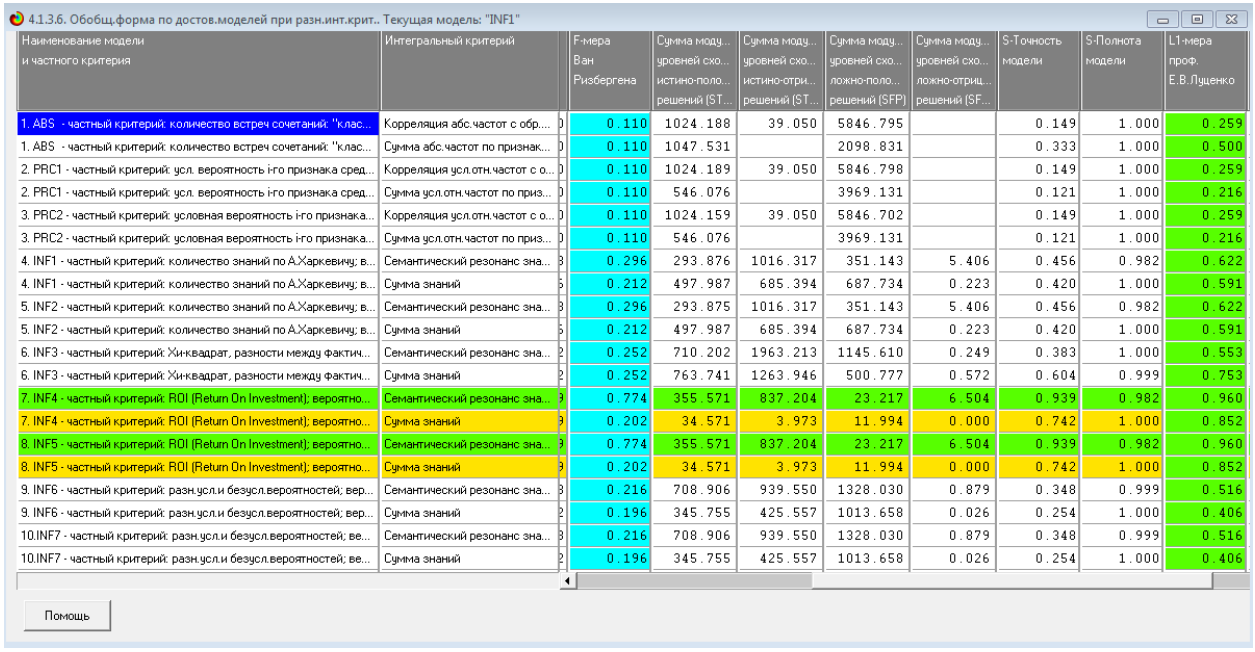
5.5. Модель: "7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	20. КОМАНДА MIL	21. КОМАНДА MIN	22. КОМАНДА MON	23. КОМАНДА NYA	24. КОМАНДА NYN	25. КОМАНДА OAK	26. КОМАНДА PHI	27. КОМАНДА PIT	28. KOM SDN
1	НОМЕР МАТЧА-ALS200107100	-0.393	1.047	3.367	0.899	0.213	-0.220	-0.425	-0.247	
2	НОМЕР МАТЧА-ALS200307150	-0.151	-0.522	2.057	0.108	-0.575	1.184	-0.598	-0.473	
3	НОМЕР МАТЧА-ALS200507120	-0.563	-0.017		-0.316	0.311	-0.438	0.242	-0.457	
4	НОМЕР МАТЧА-ALS200807150	0.331	0.497		-0.306	-0.113	-0.430	-0.159	-0.449	
5	НОМЕР МАТЧА-ALS201007130	0.087	-0.184		0.513	-0.275	-0.068	0.030	-0.550	
6	НОМЕР МАТЧА-ALS201207100	-0.598	-0.548		-0.162	0.205	-0.484	0.142	-0.003	
7	НОМЕР МАТЧА-ALS201407150	0.468	-0.174		-0.426	-0.633	1.831	-0.652	0.367	
8	НОМЕР МАТЧА-NLS200207090	-0.071	0.567	5.688	0.454	-0.536	0.194	0.320	-0.423	
9	НОМЕР МАТЧА-NLS200407130	-0.126	-0.508	2.147	0.824	-0.126	0.124	-0.172	-0.457	
10	НОМЕР МАТЧА-NLS200607110	0.256	0.413		-0.126	1.512	-0.462	0.190	0.039	
11	НОМЕР МАТЧА-NLS200707100	0.801	0.520		-0.295	0.801	-0.421	0.280	-0.441	
12	НОМЕР МАТЧА-NLS200907140	0.238	0.393		-0.354	0.651	-0.469	0.955	0.025	
13	НОМЕР МАТЧА-NLS201107120	0.062	-0.602		0.477	-0.292	-0.545	0.676	0.318	
14	НОМЕР МАТЧА-NLS201307160	-0.238	-0.143		-0.602	-0.238	-0.020	-0.278	1.365	
15	НОМЕР МАТЧА-NLS201507140	0.043	0.173		-0.184	-0.479	-0.330		0.942	
16	НОМЕР МАТЧА-NLS201507141									
17	НОМЕР МАТЧА-NLS201507142						37.214			
18	НОМЕР МАТЧА-NLS201507143									
19	НОМЕР МАТЧА-NLS201507144									
20	НОМЕР МАТЧА-NLS201507145									
21	НОМЕР МАТЧА-NLS201507146									
22	НОМЕР МАТЧА-NLS201507147									
23	НОМЕР МАТЧА-NLS201507148									
24	НОМЕР МАТЧА-NLS201507149									

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		F-мера Ван Ризбергена	Сумма моду... уровней схо... истино-поло... решений (ST...	Сумма моду... уровней схо... истино-отри... решений (ST...	Сумма моду... уровней схо... ложно-поло... решений (SFP)	Сумма моду... уровней схо... ложно-отриц... решений (SF...	S-Точность модели	S-Полнота модели	L1-мера проф. Е.В.Луценко
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс.частот с обр....	0	0.110	1024.188	39.050	5846.795		0.149	1.000	0.259
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс.частот по признак...	0	0.110	1047.531		2098.831		0.333	1.000	0.500
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн.частот с о...	0	0.110	1024.189	39.050	5846.798		0.149	1.000	0.259
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн.частот по приз...	0	0.110	546.076		3969.131		0.121	1.000	0.216
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн.частот с о...	0	0.110	1024.159	39.050	5846.702		0.149	1.000	0.259
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн.частот по приз...	0	0.110	546.076		3969.131		0.121	1.000	0.216
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	3	0.296	293.876	1016.317	351.143	5.406	0.456	0.982	0.622
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	5	0.212	497.987	685.394	687.734	0.223	0.420	1.000	0.591
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	3	0.296	293.875	1016.317	351.143	5.406	0.456	0.982	0.622
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	5	0.212	497.987	685.394	687.734	0.223	0.420	1.000	0.591
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Семантический резонанс зна...	2	0.252	710.202	1963.213	1145.610	0.249	0.383	1.000	0.553
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Сумма знаний	2	0.252	763.741	1263.946	500.777	0.572	0.604	0.999	0.753
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	8	0.774	355.571	837.204	23.217	6.504	0.939	0.982	0.960
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	9	0.202	34.571	3.973	11.994	0.000	0.742	1.000	0.852
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	8	0.774	355.571	837.204	23.217	6.504	0.939	0.982	0.960
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	9	0.202	34.571	3.973	11.994	0.000	0.742	1.000	0.852
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Семантический резонанс зна...	3	0.216	708.906	939.550	1328.030	0.879	0.348	0.999	0.516
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Сумма знаний	2	0.196	345.755	425.557	1013.658	0.026	0.254	1.000	0.406
10.INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Семантический резонанс зна...	3	0.216	708.906	939.550	1328.030	0.879	0.348	0.999	0.516
10.INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Сумма знаний	2	0.196	345.755	425.557	1013.658	0.026	0.254	1.000	0.406

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.

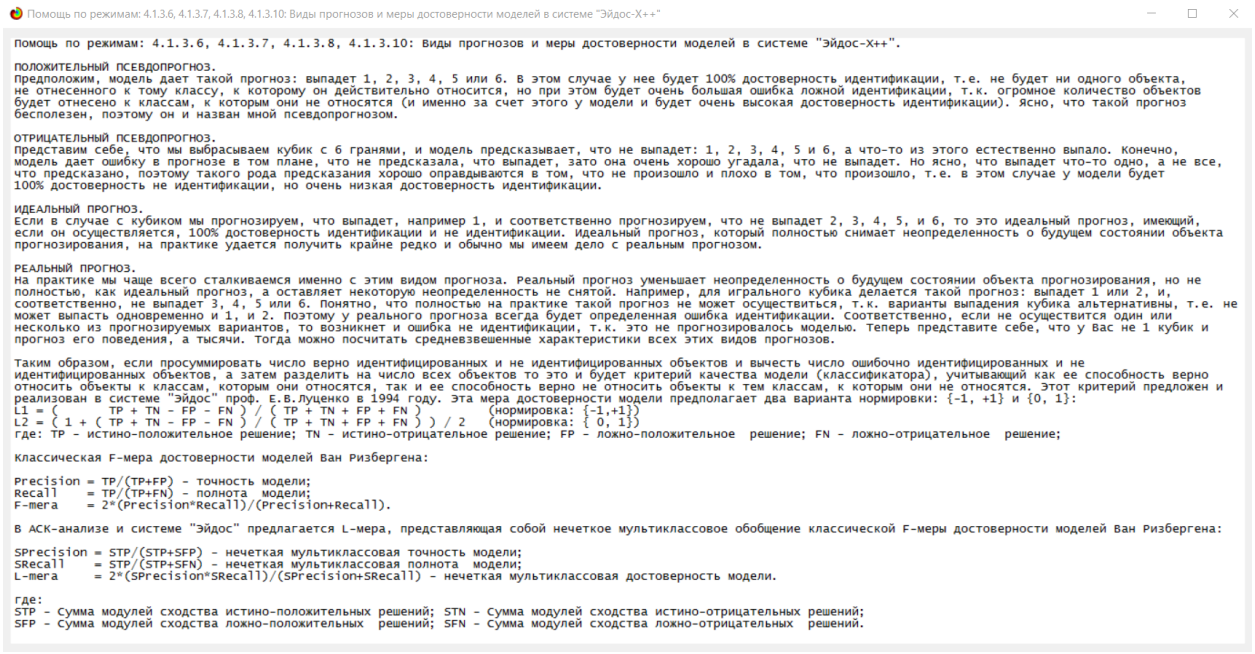


Рисунок 15- Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

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.

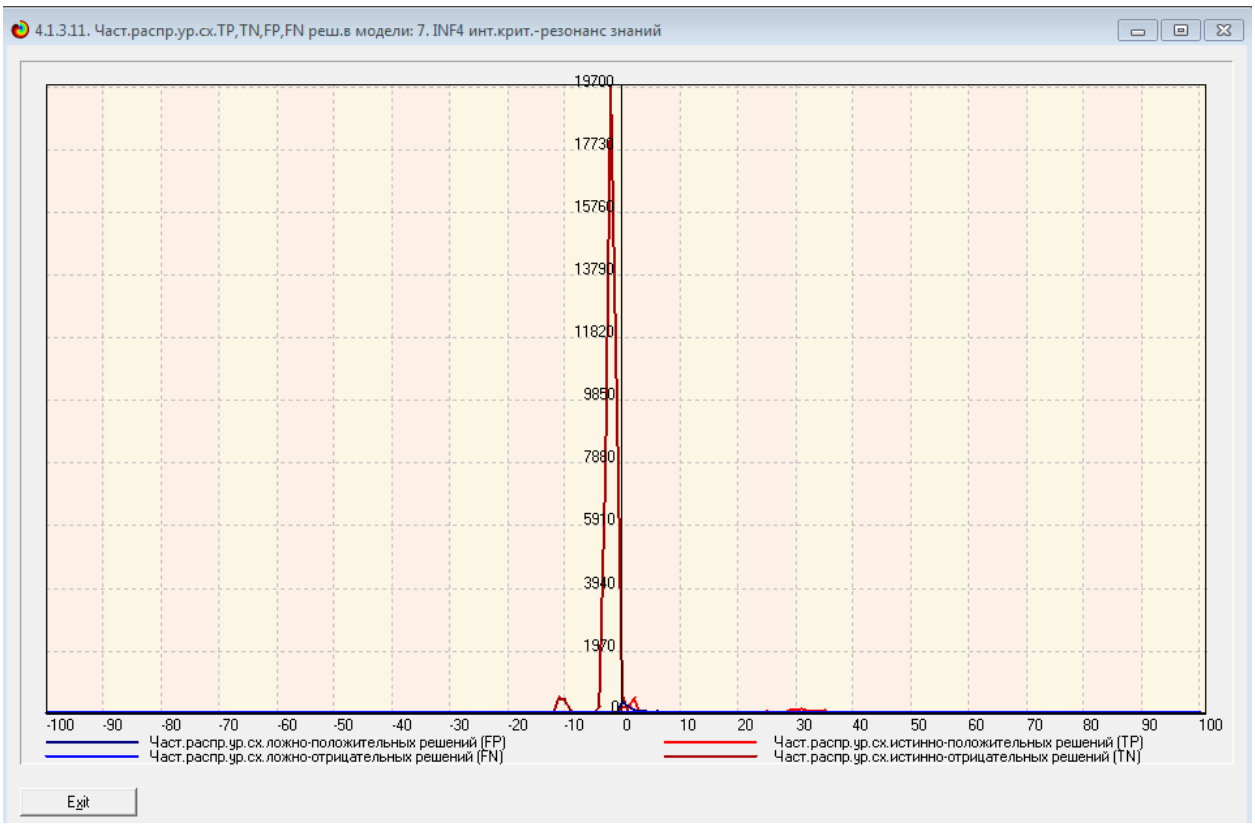


Рисунок 16- Частное распределение ложно-положительных решений, ложно-отрицательных решений, истинно-положительных решений и истинно-отрицательных решений объекта моделирования в модели INF4

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.

Точно известно, что 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. Для этого можно использовать встроенную в «Эйдос» работу «Исследование RND-модели при различных объемах выборки».

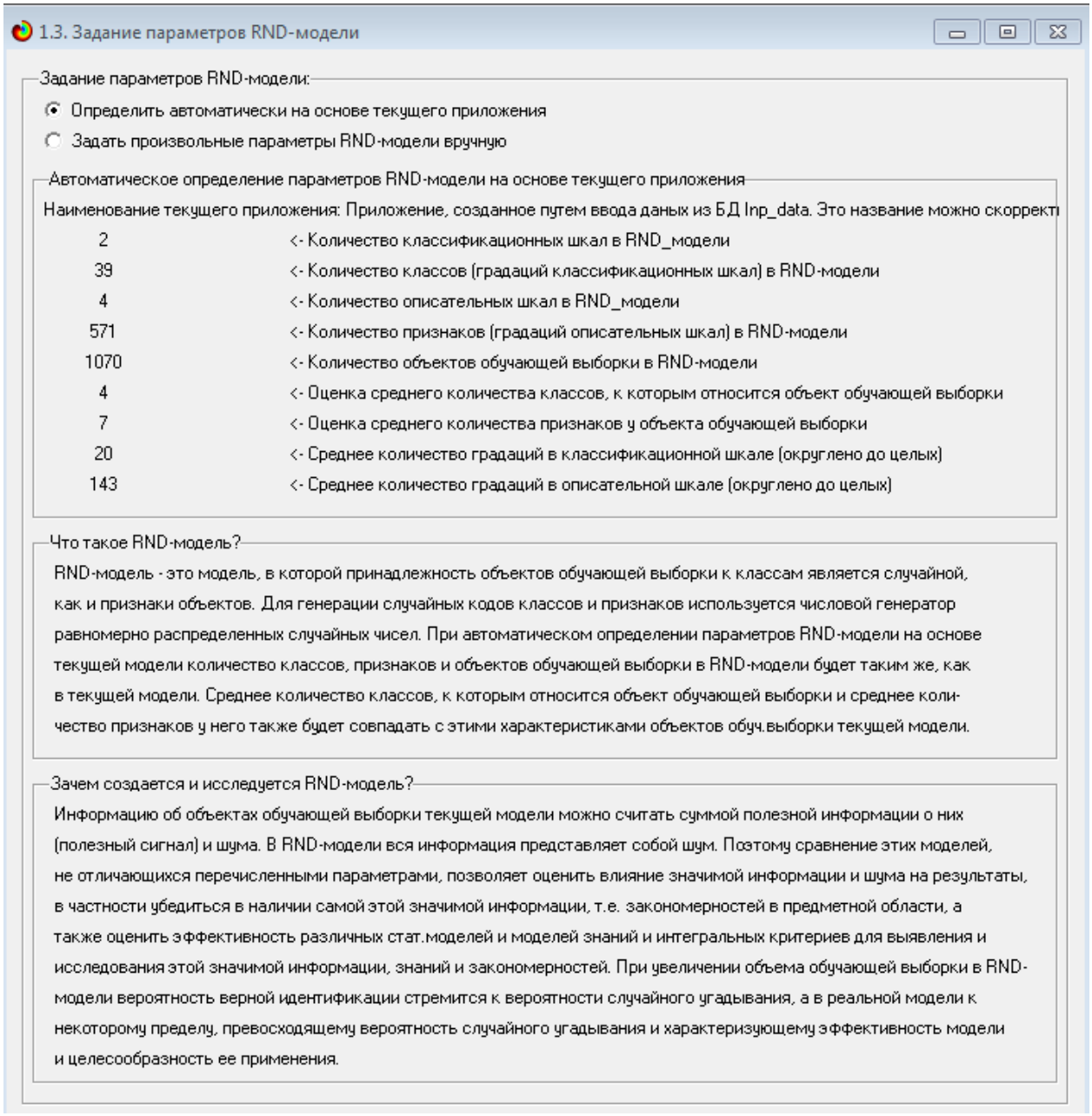


Рисунок 17- Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

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.

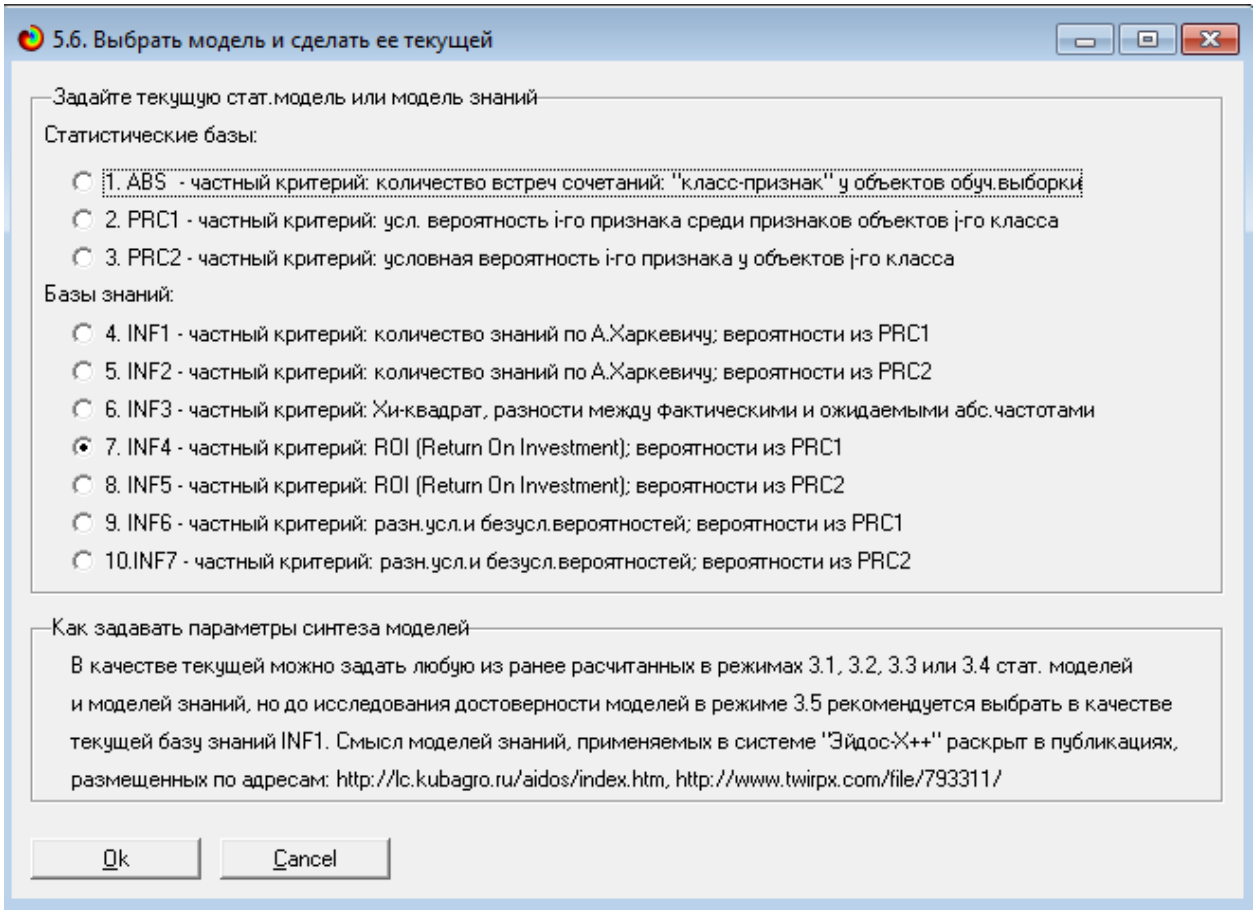


Рисунок 18- Экранная форма режима задания модели в качестве текущей

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

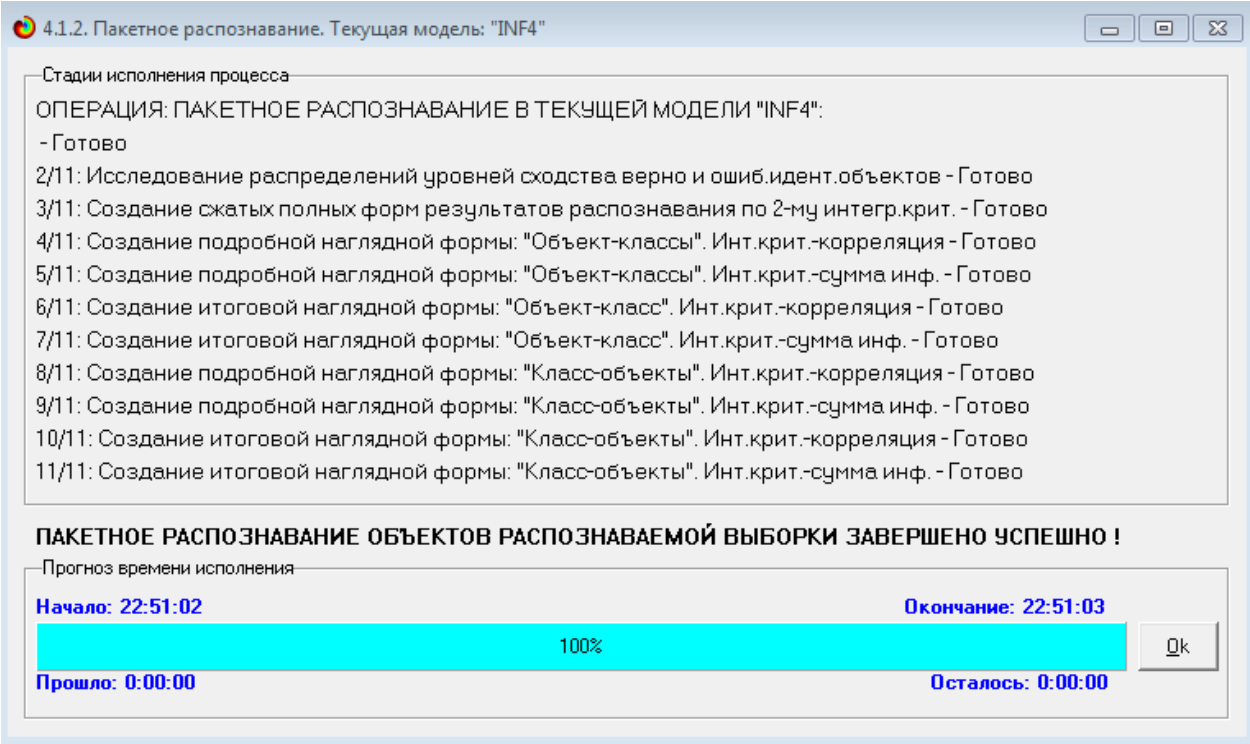


Рисунок 19- Экранная форма режима пакетного распознавания в текущей модели

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.

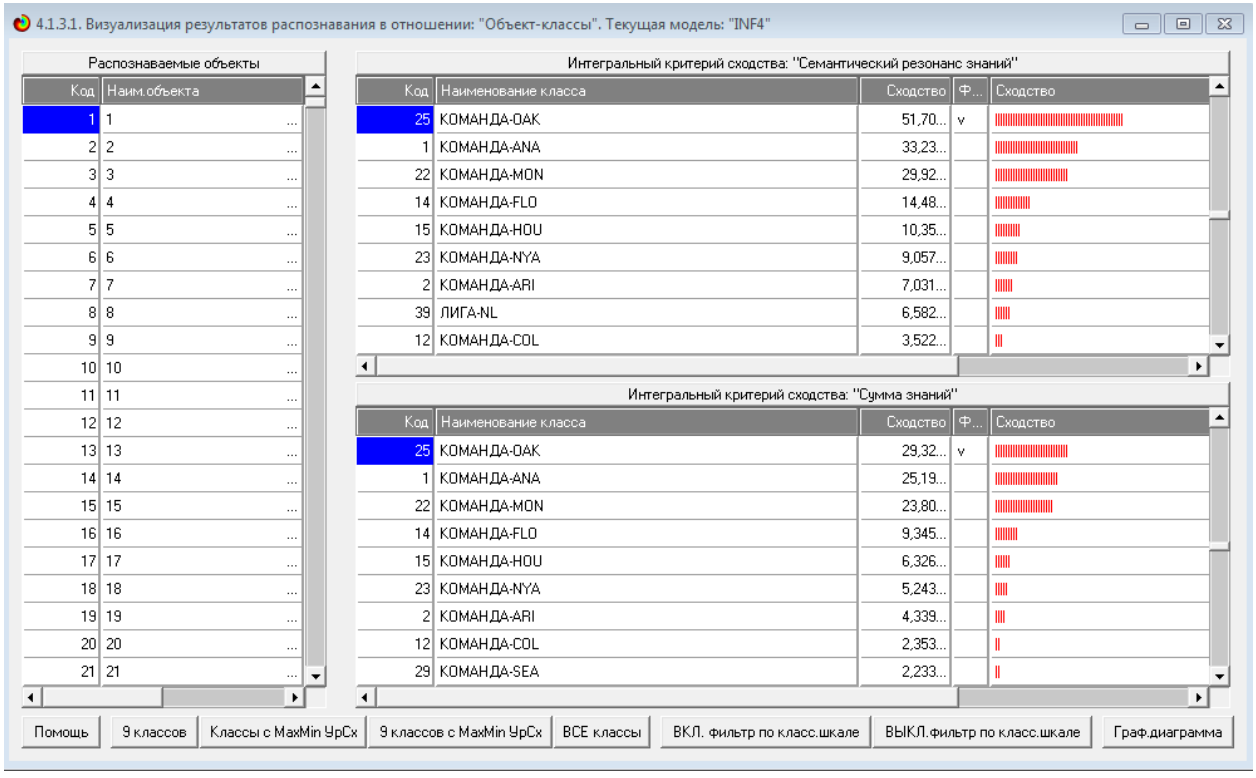


Рисунок 20- Пример идентификации классов в модели INF4

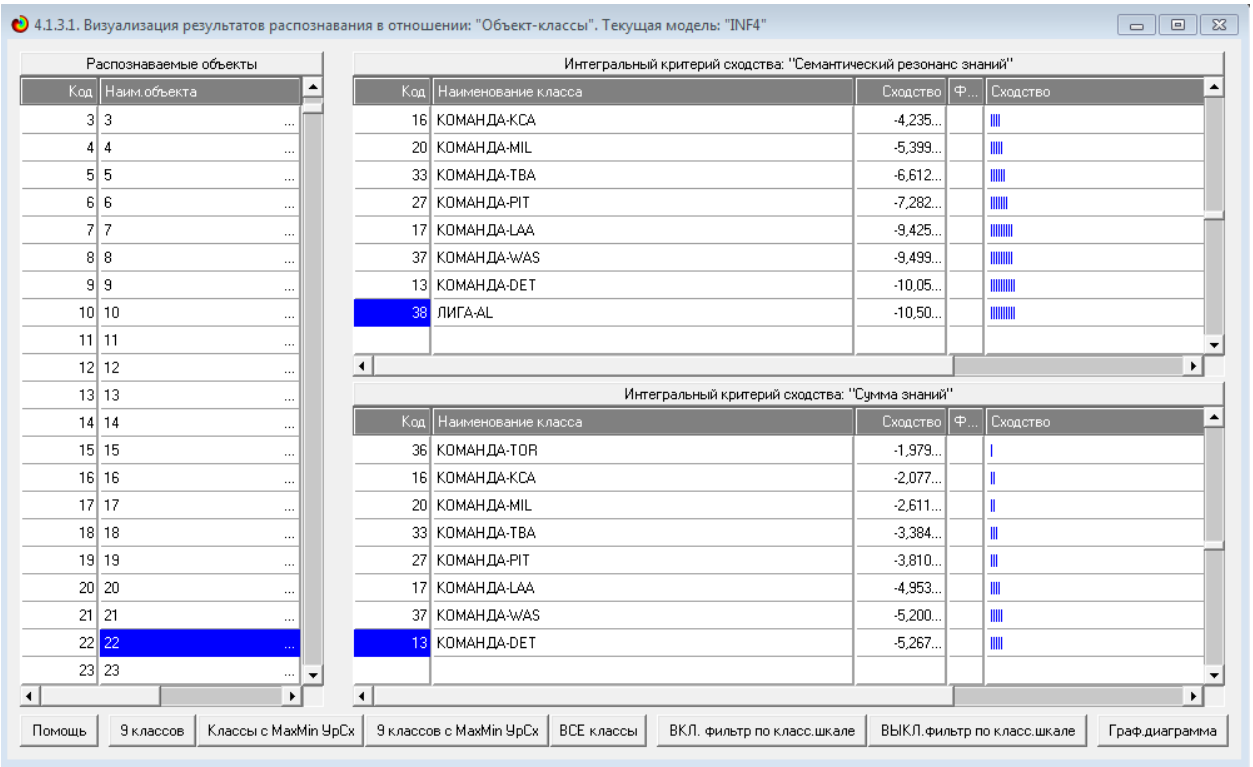


Рисунок 21- Пример идентификации классов в модели INF4

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.

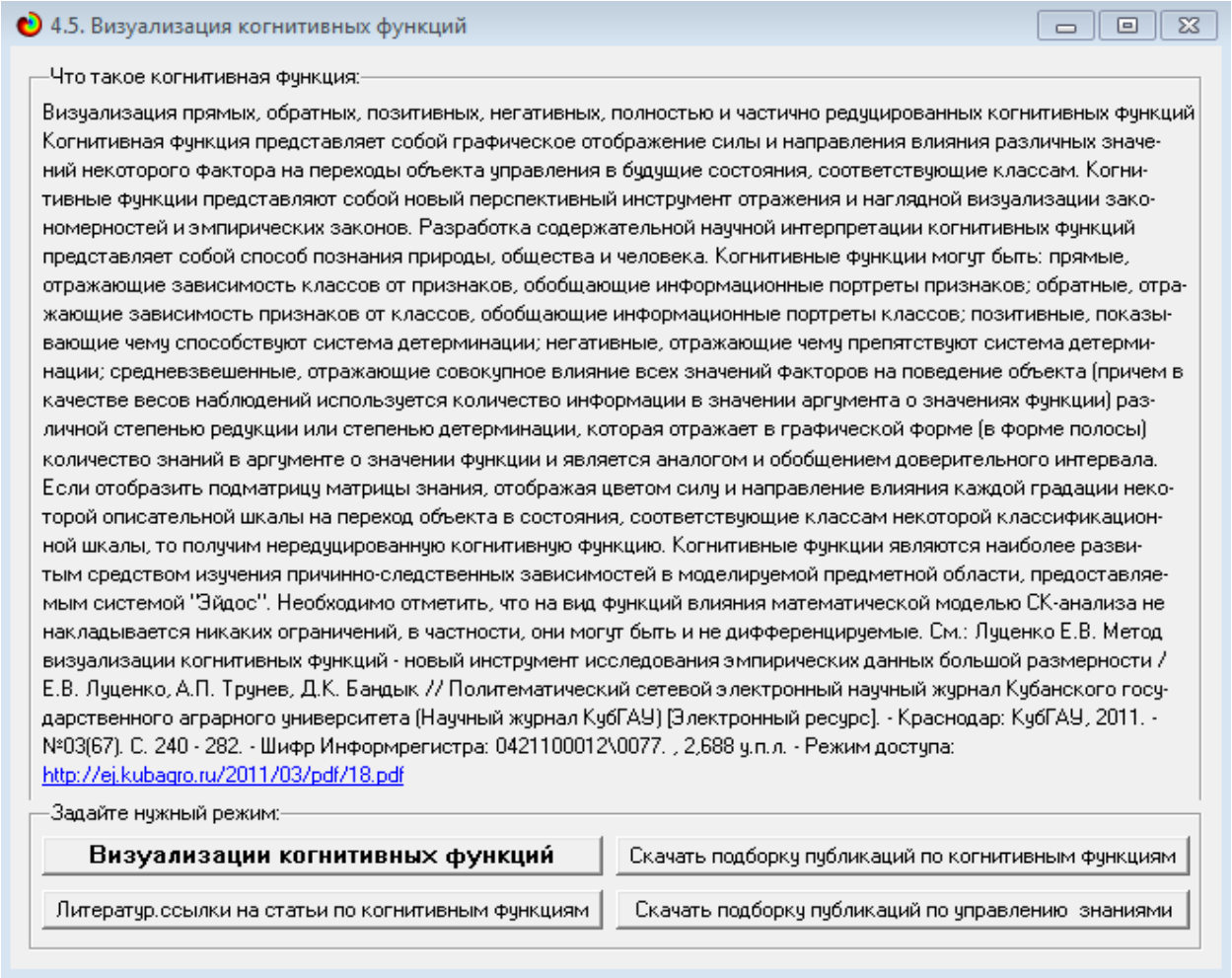


Рисунок 22- Экранная форма режима 4.5 системы «Эйдос-Х++»
«Визуализация когнитивных функций»

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.

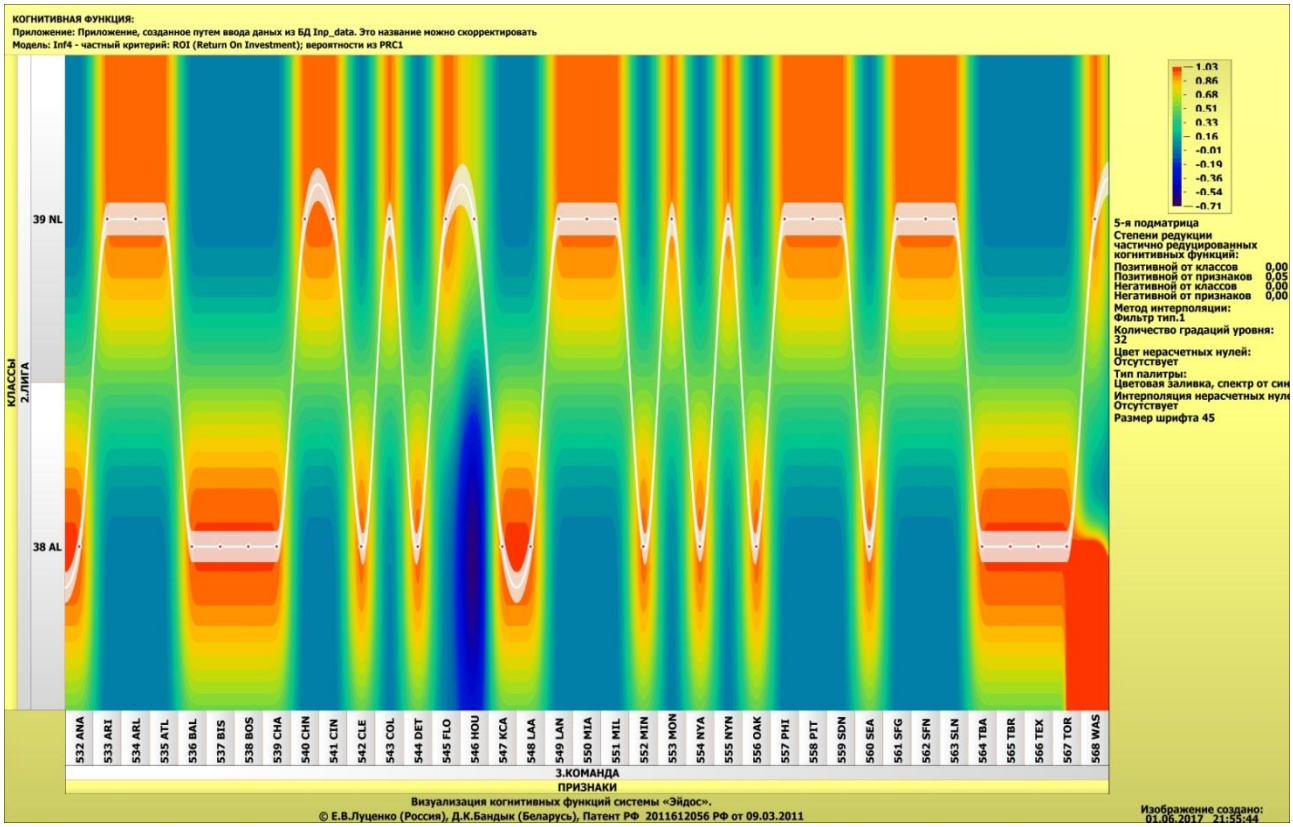


Рисунок 23-Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF4

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).

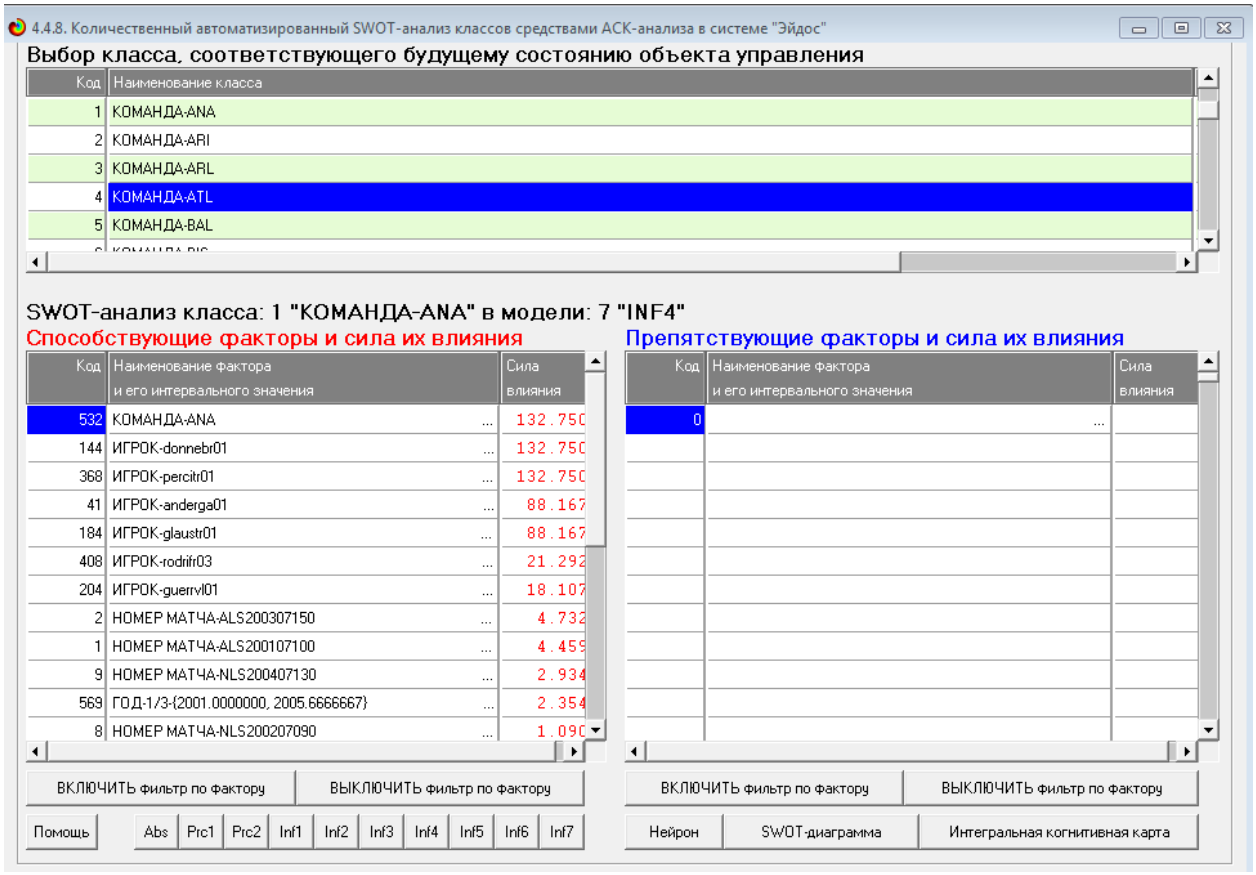


Рисунок 24-Пример SWOT-Матрицы в модели INF4

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

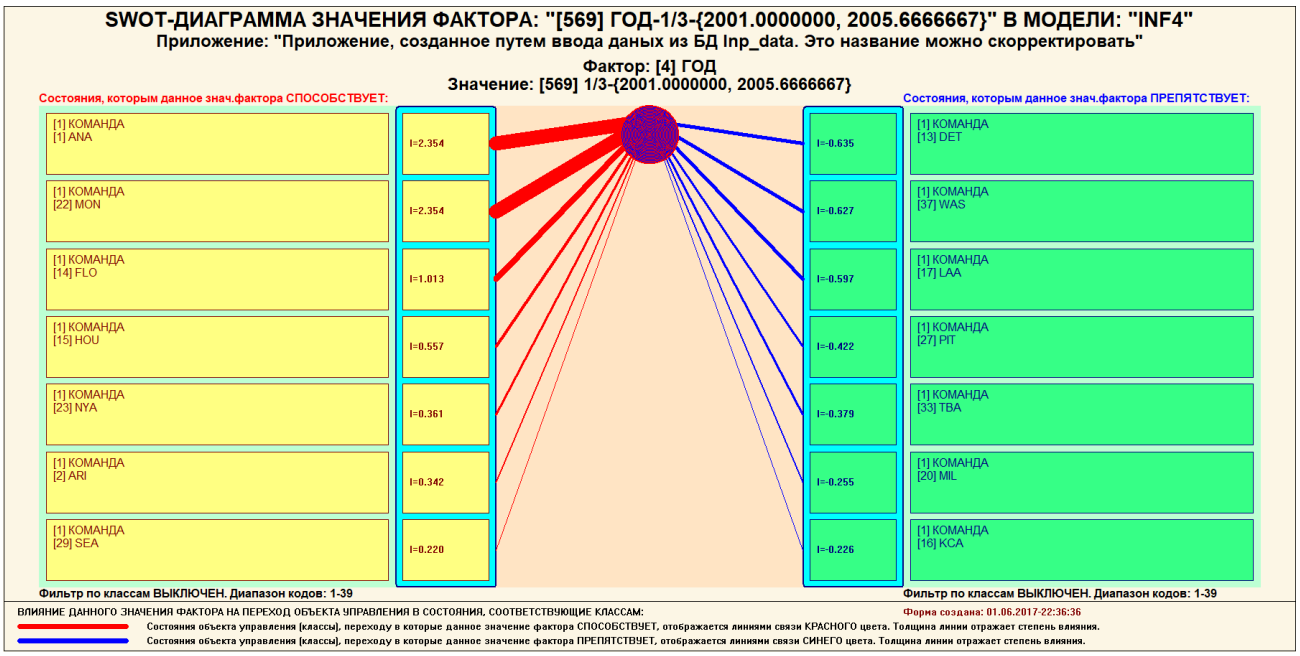


Рисунок 25– SWOT-матрица, построенная для 2001 года

2.4. Нелокальные нейронные сети

Каждому классу системно-когнитивной модели соответствует нелокальный нейрон, совокупность которых образует не локальную нейронную сеть. Рассмотрим пример на рисунке 26.

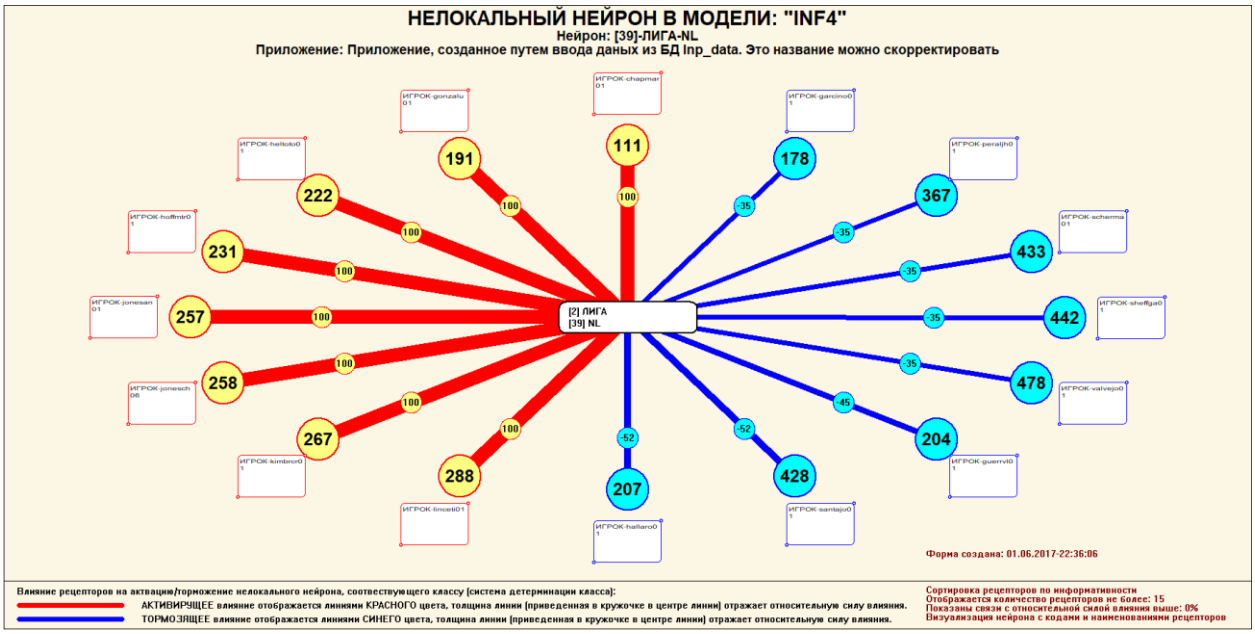


Рисунок 26- Нейрон для лиги NL

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

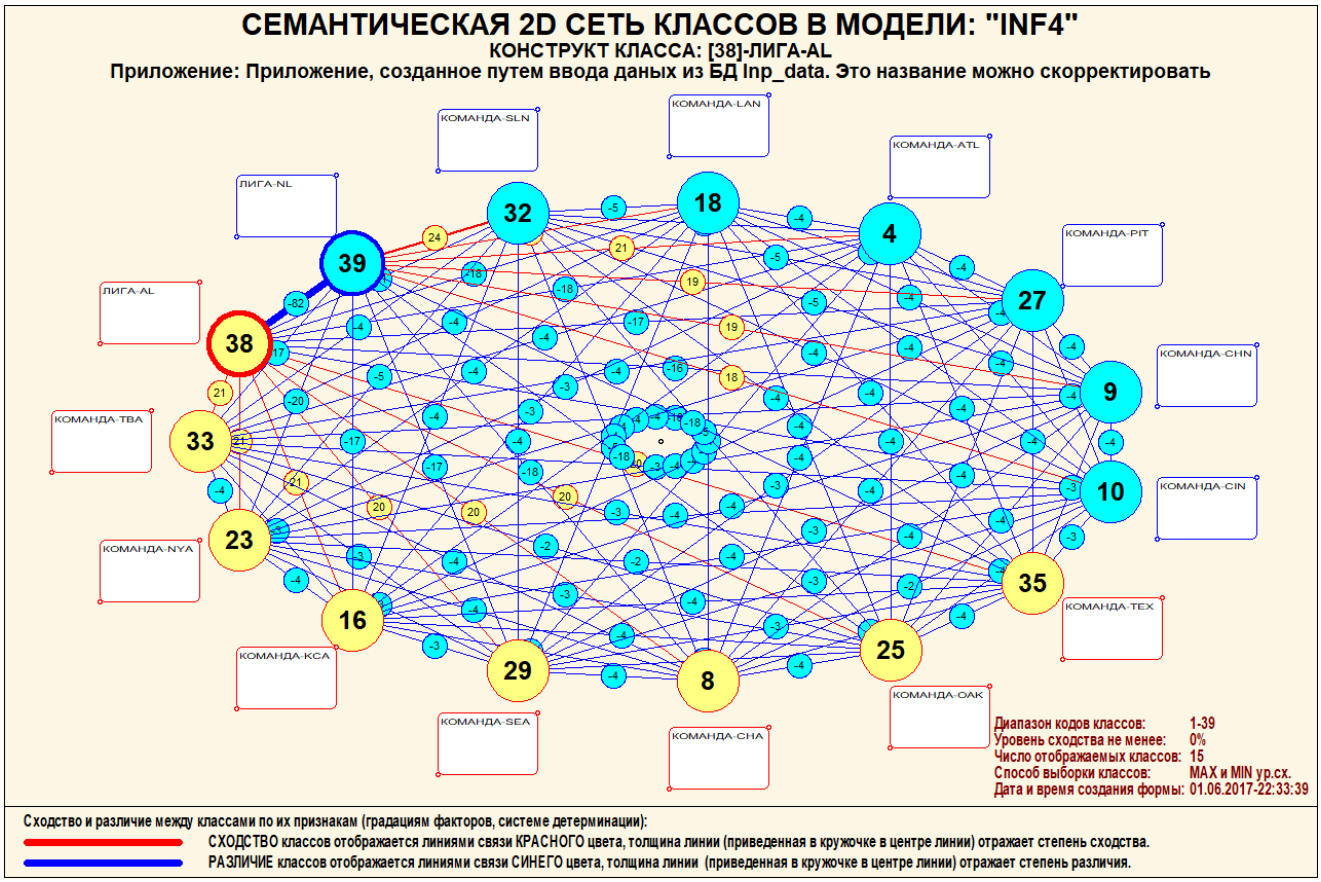


Рисунок 27– Семантическая 2D сеть классов

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 «UCI».

В данной курсовой работе 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 курсовой 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 автомати- зированного 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 аграрно- го 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.