

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:

**АСК-анализ данных опроса учеников разных школ на основе данных
репозитория UCI**

выполнил студент группы ИТ1401 Толстов Андрей Романович

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 работуСтудентке: ИТ1401 группы 3 курса
Факультета прикладной информатики
Специальности: 09.03.02 Информационные системы и технологии
(шифр)**Толстов Андрей Романович**
(Ф.И.О.)Тема проекта: **АСК-анализ данных опроса учеников разных школ на основе данных
репозитория UCI**Содержание задания: Провести исследование и разработать системно-когнитивную
модель «АСК-анализ данных опроса учеников разных школ на основе данных
репозитория UCI»

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.

РЕФЕРАТ

Курсовая работа содержит 33 страницы, 25 рисунков, 3 таблицы, 4 используемых источника.

Ключевые слова: оценка успеваемости, интеллектуальные системы, системно-когнитивная модель, АСК-анализ.

Цель работы – Провести исследование и разработать системно-когнитивную модель «АСК-анализ данных опроса учеников разных школ на основе данных репозитория UCI»

Результаты работы: На основе базы данных UCI построены модели прогнозирования с помощью АСК-анализа и реализующей его системы «Эйдос».

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	5
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	6
1.1 Исследование предметной области	6
1.2 Описание решения	6
1.3. Преобразование исходных данных из HTML-формата в файл.....	7
исходных данных MS Excel	7
1.3. Синтез и верификация статистических и интеллектуальных моделей.....	14
1.4. Виды моделей системы «Эйдос»	16
1.5. Результаты верификации моделей	18
2 РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ	20
ДОСТОВЕРНОЙ МОДЕЛИ.....	20
2.1. Решение задачи.....	20
2.2. Когнитивные функции.....	23
2.3. SWOT и PEST матрицы и диаграммы Error! Bookmark not defined.	25
2.3. SWOT и PEST матрицы и диаграммы.....	26
2.4. Нелокальные нейроны и нейронная сеть	28
2.5. Кластерный и конструктивный анализ классов.....	30
Заключение	32
Список использованных источников	33

ВВЕДЕНИЕ

Интеллектуальные информационные системы – естественный результат развития обычных информационных систем, сосредоточили в себе наиболее наукоемкие технологии с высоким уровнем автоматизации не только процессов подготовки информации для принятия решений, но и самих процессов выработки вариантов решений, опирающихся на полученные информационной системой данные. Так как существует множество альтернатив систем искусственного интеллекта, то возникает необходимость оценки качества математических моделей этих систем. В данной работе рассмотрено решение задачи идентификации городов по рейтингу качества жизни.

Для достижения поставленной цели необходимы свободный доступ к тестовым исходным данным и методика, которая поможет преобразовать эти данные в форму, которая необходима для работы в системе искусственного интеллекта.

Данные были взяты из портала UCI. Он является совместной онлайн-базой данных, которая позволяет пользователям обмениваться и сравнивать различную аналитическую информацию.

Для решения задачи используем стандартные возможности Microsoft Office Word и Excel, блокнот, а также систему искусственного интеллекта "Эйдос-Х++".

1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

1.1 Исследование предметной области

Данные для разработки системно-когнитивной модели были взяты из открытого ресурса UCI. База данных была получена по результатам опроса учеников. Она содержит много интересной социальной, гендерной и учебной информации об учениках.

В соответствии с темой, была выбрана таблица, которая содержит 395 записей в виде рейтингового списка успеваемости учеников с параметрами, влияющими на данный рейтинг, такими как:

- Пол (м/ж);
- Возраст (от 15 до 22);
- Размер семьи («до 3-х»/«более 3-х»);
- Работа матери;
- Работа отца;
- Увлечения вне школы (да/нет);
- Имеет доступ в интернет дома (да/нет);
- В романтических отношениях (да/нет);
- Свободное время после школы (от 1 – «очень мало» до 5 – «очень много»);
- Пропуски занятий (количество).

1.2 Описание решения

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

- Преобразование исходных данных из HTML-формата в промежуточные файлы MS Excel.
- Преобразование исходных данных из промежуточных файлов

- MS Excel в базы данных системы "Эйдос".
- Синтез и верификация моделей предметной области.
- Применение моделей для решения задач идентификации, прогнозирования и исследования предметной области.

1.3. Преобразование исходных данных из HTML-формата в файл исходных данных 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 данные из html-файла были преобразованы в xls-файл. Данные, представленные на интернет-ресурсе UCI, показаны на рисунке 1.

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob
GP	F	18	U	GT3	A	4	4	at_home
GP	F	17	U	GT3	T	1	1	at_home
GP	F	15	U	LE3	T	1	1	at_home
GP	F	15	U	GT3	T	4	2	health
GP	F	16	U	GT3	T	3	3	other
GP	M	16	U	LE3	T	4	3	services
GP	M	16	U	LE3	T	2	2	other
GP	F	17	U	GT3	A	4	4	other
GP	M	15	U	LE3	A	3	2	services
GP	M	15	U	GT3	T	3	4	other
GP	F	15	U	GT3	T	4	4	teacher
GP	F	15	U	GT3	T	2	1	services
GP	M	15	U	LE3	T	4	4	health
GP	M	15	U	GT3	T	4	3	teacher
GP	M	15	U	GT3	A	2	2	other

Рисунок 1 – Успеваемость учеников UCI

Данные, полученные после преобразования в файл MS Excel и изменены сокращения на более полные наименования на русском языке, представлены на таблице 1.

Затем данный файл был назван «Inp_data» и помещен в папку: c:\Aidos-X\AID_DATA\Inp_data\ для дальнейшего анализа в системе.

Следующим шагом необходимо выполнить автоматизированную формализацию предметной области путем импорта исходных данных из внешних баз данных в систему "Эйдос".

Таблица 1 – Исходные данные в MS Excel

Номер п/п	Успеваемость	Пол	Возраст	Размер семьи	Работа матери	Работа отца	Увлечения	Имеет доступ в интернет дома	В романтических отношениях	Свободное время	Пропуски занятий
1	6	ж	18	более 3-х	на дому	учитель	нет	нет	нет	3	6
2	6	ж	17	более 3-х	на дому	другое	нет	да	нет	3	4
3	10	ж	15	до 3-х	на дому	другое	нет	да	нет	3	10
4	15	ж	15	более 3-х	врач	сервис	да	да	да	2	2
5	10	ж	16	более 3-х	другое	другое	нет	нет	нет	3	4
6	15	м	16	до 3-х	сервис	другое	да	да	нет	4	10
7	11	м	16	до 3-х	другое	другое	нет	да	нет	4	0
8	6	ж	17	более 3-х	другое	учитель	нет	нет	нет	1	6
9	19	м	15	до 3-х	сервис	другое	нет	да	нет	2	0
10	15	м	15	более 3-х	другое	другое	да	да	нет	5	0
11	9	ж	15	более 3-х	учитель	врач	нет	да	нет	3	0
12	12	ж	15	более 3-х	сервис	другое	да	да	нет	2	4
13	14	м	15	до 3-х	врач	сервис	да	да	нет	3	2
14	11	м	15	более 3-х	учитель	другое	нет	да	нет	4	2
15	16	м	15	более 3-х	другое	другое	нет	да	да	5	0
16	14	ж	16	более 3-х	врач	другое	нет	да	нет	4	4
17	14	ж	16	более 3-х	сервис	сервис	да	да	нет	2	6
18	10	ж	16	более 3-х	другое	другое	да	нет	нет	3	4
19	5	м	17	более 3-х	сервис	сервис	да	да	нет	5	16
20	10	м	16	до 3-х	врач	другое	да	да	нет	1	4
21	15	м	15	более 3-х	учитель	другое	нет	да	нет	4	0
22	15	м	15	более 3-х	врач	врач	нет	да	нет	4	0
23	16	м	16	до 3-х	учитель	другое	да	да	нет	5	2
24	12	м	16	до 3-х	другое	другое	да	да	нет	4	0
25	8	ж	15	более 3-х	сервис	врач	да	да	нет	3	2
26	8	ж	16	более 3-х	сервис	сервис	нет	да	нет	2	14
27	11	м	15	более 3-х	другое	другое	нет	да	нет	2	2
28	15	м	15	более 3-х	врач	сервис	нет	да	нет	2	4
29	11	м	16	до 3-х	сервис	другое	да	да	нет	3	4
30	11	м	16	более 3-х	учитель	учитель	да	да	да	4	16

Для входа в систему «Эйдос» задаем имя:1 пароль:1 на рисунке 2.

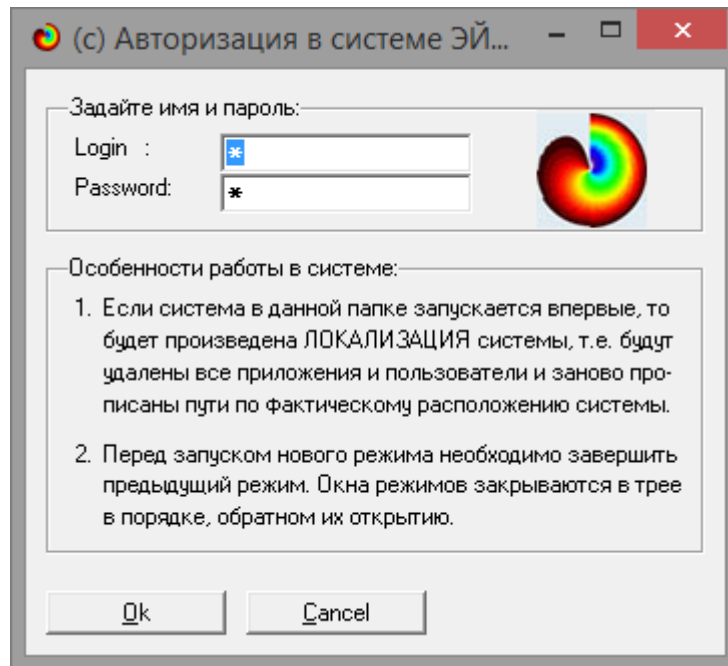


Рисунок 2 – Авторизация в «Эйдос»

Для загрузки базы исходных данных в систему «Эйдос» необходимо воспользоваться универсальным программным интерфейсом для ввода данных из внешних баз данных табличного вида, т.е. режимом 2.3.2.2 (Рисунок 3).

2.3.2.2. Универсальный программный интерфейс импорта данных в систему "ЭЙДОС-Х+..."

Автоматическая формализация предметной области: генерация классификационных и описательных шкал и градаций, а также обучающей и распознаваемой выборки на основе базы исходных данных: "Inp_data"

Задайте тип файла исходных данных: "Inp_data":

☐ XLS - MS Excel-2003 Стандарт XLS-файла
☒ XLSX- MS Excel-2007(2010)
☐ DBF - DBASE IV (DBF/NTX) Стандарт DBF-файла
☐ CSV - Comma-Separated Values Стандарт CSV-файла

Задайте параметры:

☐ Нули и пробелы считать ОТСУТСТВИЕМ данных
☒ Нули и пробелы считать ЗНАЧЕНИЯМИ данных
☒ Создавать БД средних по классам "Inp_davr.dbf"?
 Требования к файлу исходных данных

Задайте диапазон столбцов классификационных шкал:

Начальный столбец классификационных шкал: 2
 Конечный столбец классификационных шкал: 2

Задайте диапазон столбцов описательных шкал:

Начальный столбец описательных шкал: 3
 Конечный столбец описательных шкал: 12

Задайте режим:

☒ Формализации предметной области (на основе "Inp_data")
☐ Генерации распознаваемой выборки (на основе "Inp_rasp")

Задайте способ выбора размера интервалов:

☒ Равные интервалы с разным числом наблюдений
☐ Разные интервалы с равным числом наблюдений

Задание параметров формирования сценариев или способа интерпретации текстовых полей "Inp_data":

☒ Не применять сценарный метод АСК-анализа
☐ Применить спец. интерпретацию текстовых полей классов
☐ Применить сценарный метод АСК-анализа
☐ Применить спец. интерпретацию текстовых полей признаков

Параметры интерпретации значений текстовых полей "Inp_data":

Интерпретация TXT-полей классов:

Значения полей текстовых классификационных шкал файла исходных данных "Inp_data" рассматриваются как целое

Интерпретация TXT-полей признаков:

Значения полей текстовых описательных шкал файла исходных данных "Inp_data" рассматриваются как целое

Какие наименования ГРАДАЦИЙ числовых шкал использовать:

☐ Только интервальные числовые значения (например: "1/3-{59873.0000000, 178545.6666667}")
☐ Только наименования интервальных числовых значений (например: "Минимальное")
☒ И интервальные числовые значения, и их наименования (например: "Минимальное: 1/3-{59873.0000000, 178545.6666667}")

Ok Cancel

Рисунок 3 – Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

В экранной форме, приведенной на рисунке 3, задать настройки, показанные на рисунке:

– «Задайте тип файла исходных данных Inp_data»: «XLS – MS Excel-2007(2010)»;

– «Задайте диапазон столбцов классификационных шкал»: «Начальный столбец классификационных шкал» – 2, «Конечный столбец классификационных шкал» – 2;

– «Задайте диапазон столбцов описательных шкал»: «Начальный столбец классификационных шкал» – 3, «Конечный столбец классификационных шкал» – 12 (последний столбец в таблице);

Далее было задано число интервалов в шкале, а именно в классификационных шкалах и в остальных шкалах, равное 5 (Рисунок 4)

2.3.2.2. Задание размерности модели системы "Эйдос-Х++"

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [5 x 40]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс. шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис. шкалу
Числовые	1	5	5,00	4	20	5,00
Текстовые	0	0	0,00	7	20	2,86
ВСЕГО:	1	5	5,00	11	40	3,64

Задайте число интервалов (градаций) в шкале:

В классификационных шкалах: В описательных шкалах:

Пересчитать шкалы и градации

Выйти на создание модели

Рисунок 4 – Экранная форма задания размерности модели.

Далее открывается окно, отображающее стадию процесса импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос" (рисунок 5) и прогноз времени завершения этого процесса. Здесь необходимо дождаться завершения формализации предметной области и нажать кнопку "ОК".

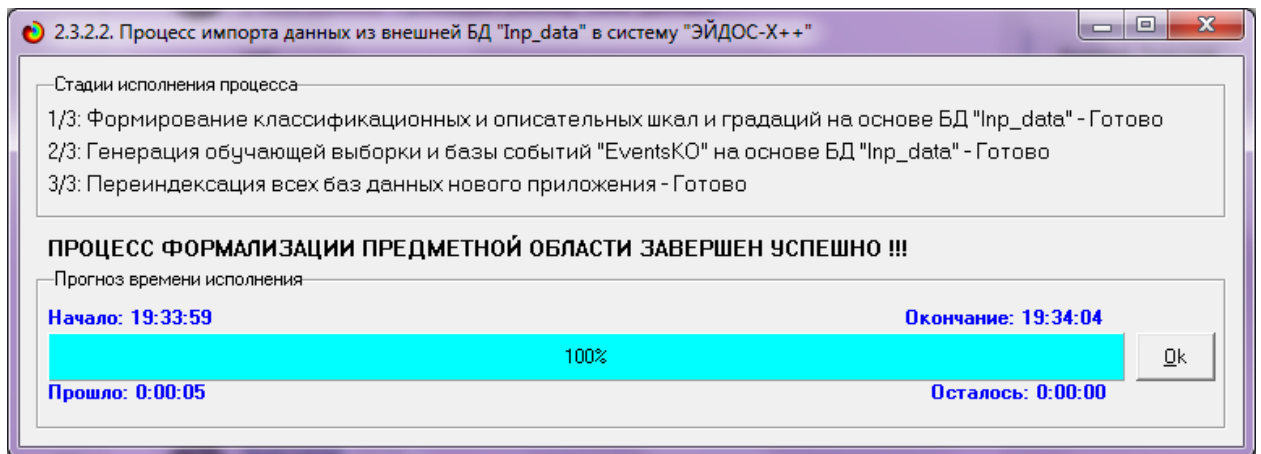


Рисунок 5 – Процесс импорта данных

В результате формируются классификационные и описательные шкалы и градации, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Этим самым полностью автоматизировано выполняется 2-й этап АСК-анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1 (рисунок 6).

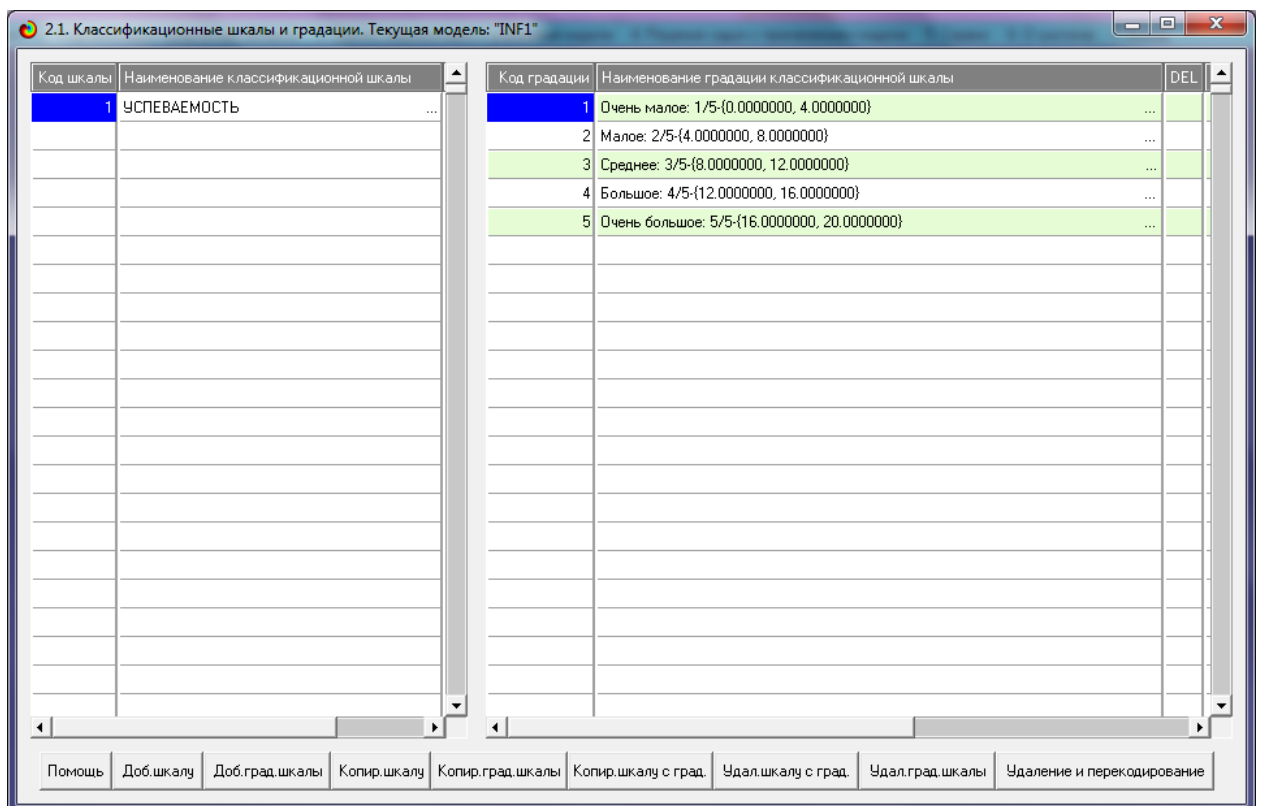


Рисунок 6 – Классификационные шкалы и градации

Для просмотра описательных шкал и градаций необходимо запустить режим 2.2 (рисунок 7), а обучающей выборки – режим 2.3.1. (рисунок 8).

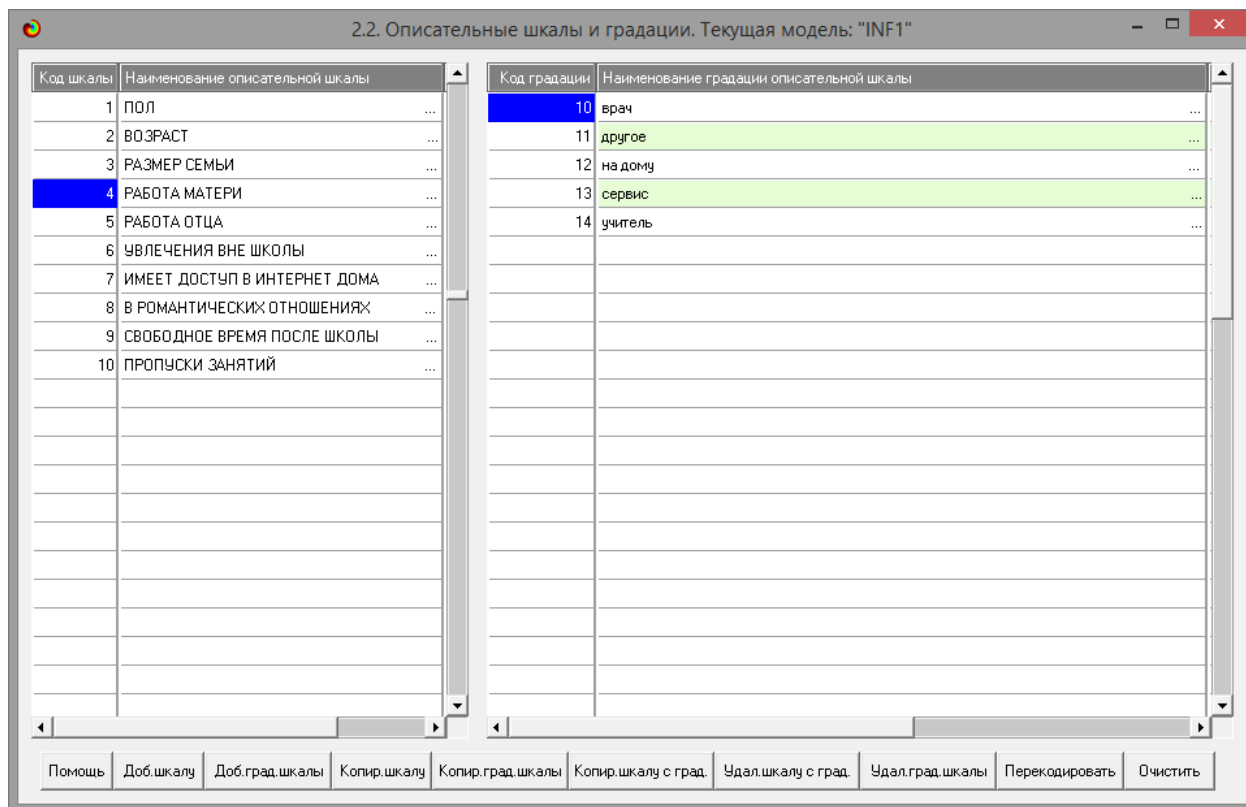


Рисунок 7 – Описательные шкалы

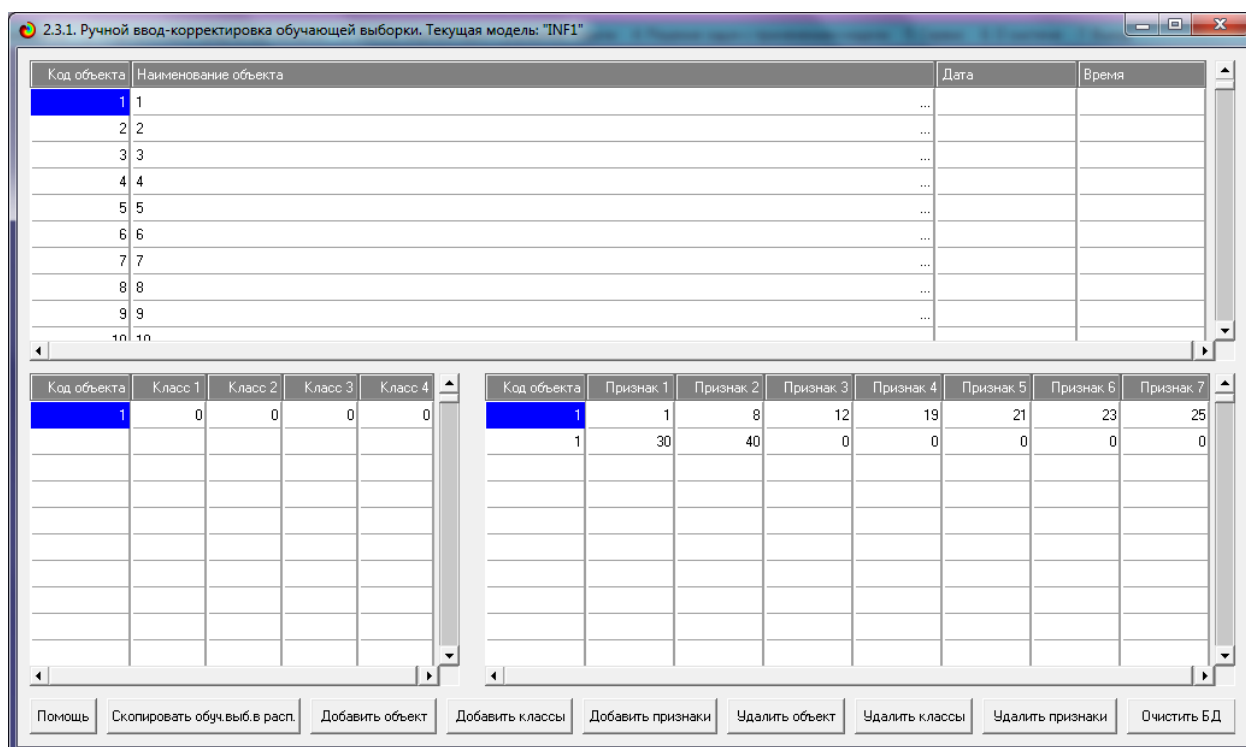


Рисунок 8 – Корректировка обучающей выборки

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 запускаем режим 3.5, в котором задаются модели для синтеза и верификации, а также задается модель, которой по окончании режима присваивается статус текущей (рисунок 9).

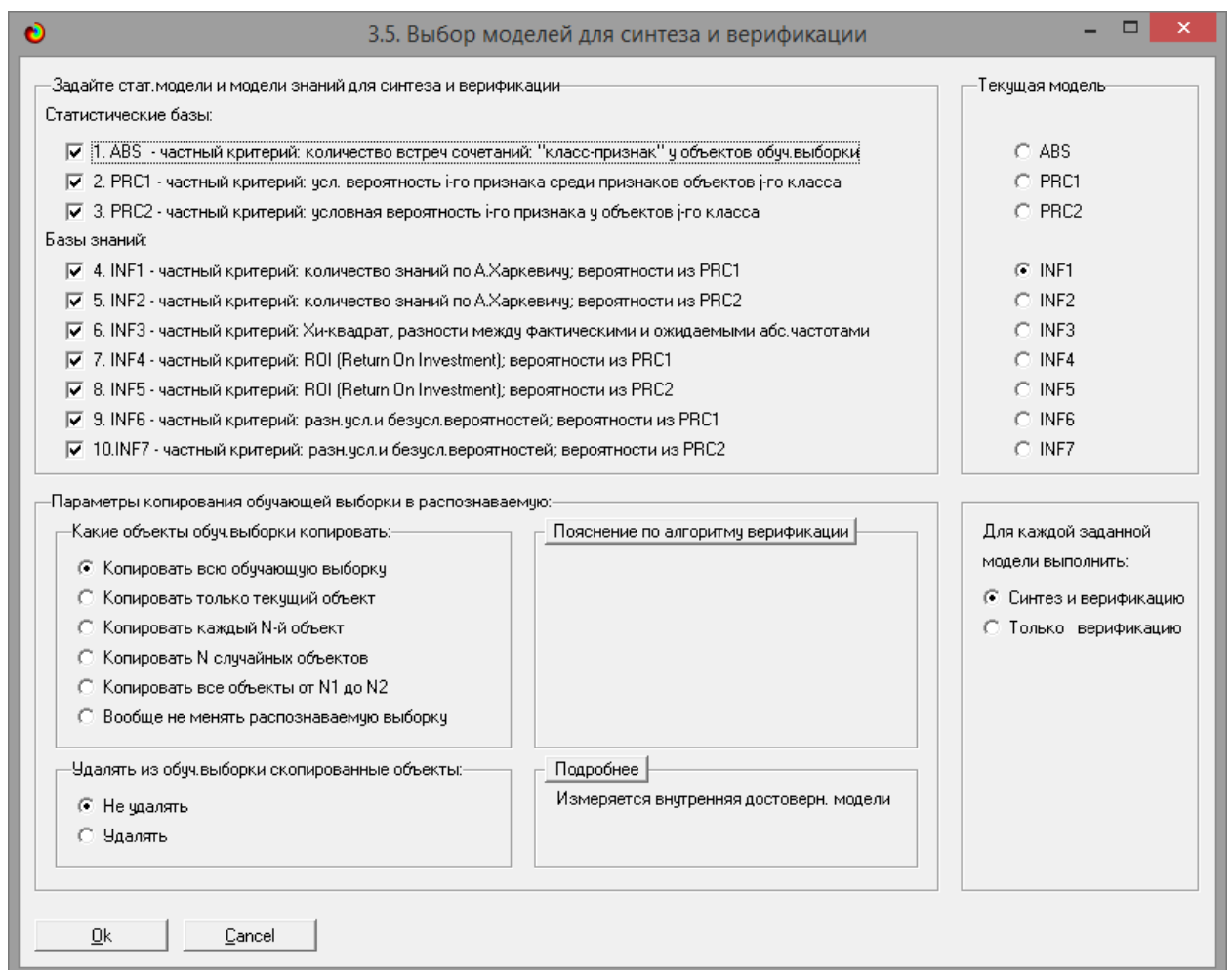


Рисунок 9 – Выбор моделей для синтеза и верификации

В данном режиме имеется много различных методов верификации моделей, в том числе и поддерживающие бутстрепный метод. Оставляем здесь всё без изменений. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 10.

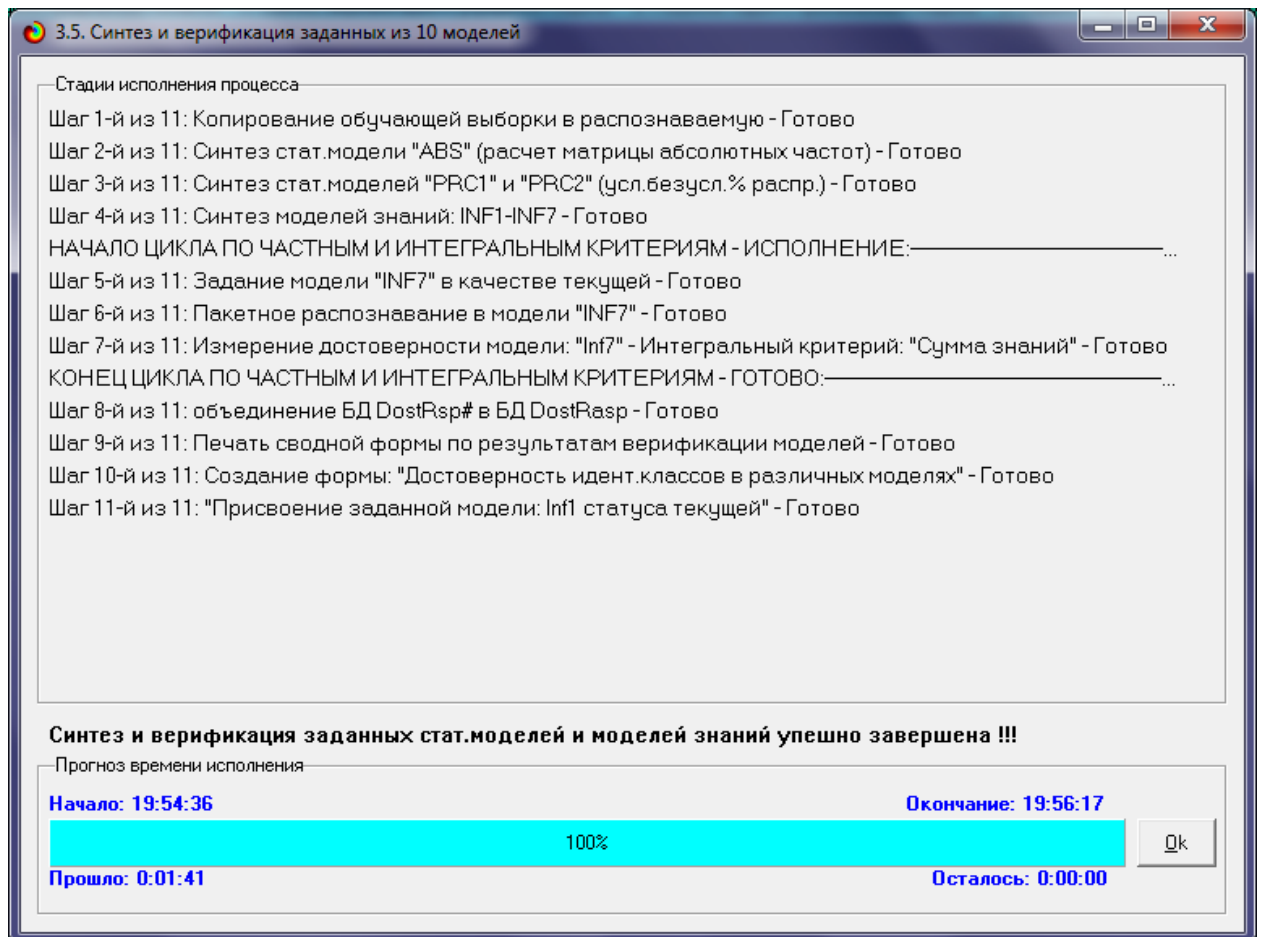


Рисунок 10 – Синтез и верификация заданных моделей

Синтез и верификация всех 10 моделей на данной задаче заняли 1.41 минуты. При этом верификация (оценка достоверности моделей) проводилась на всех примерах наблюдения из обучающей выборки.

1.4. Виды моделей системы «Эйдос»

Рассмотрим решение задачи идентификации на примере модели INF1, в которой рассчитано количество информации по А.Харкевичу, которое мы получаем о принадлежности идентифицируемого объекта к каждому из классов, если знаем, что у этого объекта есть некоторый признак.

По сути, частные критерии представляют собой просто формулы для преобразования матрицы абсолютных частот (таблица 1) в матрицы условных и безусловных процентных распределений, и матрицы знаний (таблицы 2 и 3).

Таблица 1 – Матрица абсолютных частот (модель ABS (фрагмент))

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"									
Код признака	Наименование описательной шкалы и градации	1. УСПЕВАЕМ... ОЧЕНЬ МАЛОЕ: 1/5 {0.0000000, 4.0000000}	2. УСПЕВАЕМ... МАЛОЕ: 2/5 {4.0000000, 8.0000000}	3. УСПЕВАЕМ... СРЕДНЕЕ: 3/5 {8.0000000, 12.0000000}	4. УСПЕВАЕМ... БОЛЬШОЕ: 4/5 {12.0000000, 16.0000000}	5. УСПЕВАЕМ... ОЧЕНЬ БОЛЬШОЕ: 5/5 {16.0000000, 20.0000000}	Сумма	Среднее	Средн. квадр. откл.
1	ПОЛ-ж	24	34	87	53	10	208	41.600	29.821
2	ПОЛ-м	15	29	75	54	14	187	37.400	26.501
3	ВОЗРАСТ-Очень малое: 1/5-{15.0000000, 16.4000000...}	13	26	74	59	14	186	37.200	27.743
4	ВОЗРАСТ-Малое: 2/5-{16.4000000, 17.8000000} ...	9	15	45	25	4	98	19.600	16.211
5	ВОЗРАСТ-Среднее: 3/5-{17.8000000, 19.2000000} ...	17	20	42	22	5	106	21.200	13.368
6	ВОЗРАСТ-Большое: 4/5-{19.2000000, 20.6000000} ...			1	1	1	3	0.600	0.548
7	ВОЗРАСТ-Очень большое: 5/5-{20.6000000, 22.0000...}		2				2	0.400	0.894
8	РАЗМЕР СЕМЬИ-более 3-х ...	32	46	113	74	16	281	56.200	38.239
9	РАЗМЕР СЕМЬИ-до 3-х	7	17	49	33	8	114	22.800	17.978
10	РАБОТА МАТЕРИ-врач	2	3	10	17	2	34	6.800	6.611
11	РАБОТА МАТЕРИ-другое ...	15	28	60	35	3	141	28.200	21.603
12	РАБОТА МАТЕРИ-на дому ...	9	9	26	14	1	59	11.800	9.203
13	РАБОТА МАТЕРИ-сервис ...	9	16	41	23	14	103	20.600	12.462
14	РАБОТА МАТЕРИ-учитель ...	4	7	25	18	4	58	11.600	9.450
15	РАБОТА ОТЦА-врач		4	6	7	1	18	3.600	3.050
16	РАБОТА ОТЦА-другое	22	38	90	56	11	217	43.400	31.093
17	РАБОТА ОТЦА-на дому	3	2	7	7	1	20	4.000	2.828
18	РАБОТА ОТЦА-сервис	11	16	51	28	5	111	22.200	18.185
19	РАБОТА ОТЦА-учитель	3	3	8	9	6	29	5.800	2.775
20	АКТИВИСТ ВНЕ ШКОЛЫ-да ...	21	26	86	54	14	201	40.200	29.769

Таблица 2 – Матрица знаний (модель INF1 (фрагмент))

Код признака	Наименование описательной шкалы и градации	1. УСПЕВАЕМ... ОЧЕНЬ МАЛОЕ: 1/5 {0.0000000, 4.0000000}	2. УСПЕВАЕМ... МАЛОЕ: 2/5 {4.0000000, 8.0000000}	3. УСПЕВАЕМ... СРЕДНЕЕ: 3/5 {8.0000000, 12.0000000}	4. УСПЕВАЕМ... БОЛЬШОЕ: 4/5 {12.0000000, 16.0000000}	5. УСПЕВАЕМ... ОЧЕНЬ БОЛЬШОЕ: 5/5 {16.0000000, 20.0000000}	Сумма	Среднее	Средн. квадр. откл.
1	ПОЛ-ж	0.043	0.007	0.005	-0.017	-0.065	-0.026	-0.005	0.040
2	ПОЛ-м	-0.058	-0.008	-0.006	0.018	0.058	0.004	0.001	0.042
3	ВОЗРАСТ-Очень малое: 1/5-{15.0000000, 16.40...}	-0.096	-0.037	-0.008	0.044	0.059	-0.038	-0.008	0.063
4	ВОЗРАСТ-Малое: 2/5-{16.4000000, 17.8000000} ...	-0.020	-0.011	0.031	-0.017	-0.110	-0.127	-0.025	0.052
5	ВОЗРАСТ-Среднее: 3/5-{17.8000000, 19.200000...}	0.134	0.047	-0.010	-0.074	-0.070	0.027	0.005	0.087
6	ВОЗРАСТ-Большое: 4/5-{19.2000000, 20.600000...}			-0.057	0.058	0.472	0.472	0.094	0.215
7	ВОЗРАСТ-Очень большое: 5/5-{20.6000000, 22....}		0.509				0.509	0.102	0.228
8	РАЗМЕР СЕМЬИ-более 3-х ...	0.040	0.007	-0.005	-0.008	-0.018	0.015	0.003	0.022
9	РАЗМЕР СЕМЬИ-до 3-х ...	-0.132	-0.019	0.013	0.018	0.040	-0.079	-0.016	0.068
10	РАБОТА МАТЕРИ-врач ...	-0.144	-0.164	-0.092	0.170	-0.009	-0.239	-0.048	0.136
11	РАБОТА МАТЕРИ-другое ...	0.021	0.061	0.010	-0.024	-0.291	-0.223	-0.045	0.141
12	РАБОТА МАТЕРИ-на дому ...	0.121	-0.012	0.020	-0.037	-0.354	-0.262	-0.052	0.179
13	РАБОТА МАТЕРИ-сервис ...	-0.034	-0.007	-0.008	-0.054	0.223	0.120	0.024	0.113
14	РАБОТА МАТЕРИ-учитель ...	-0.099	-0.077	0.014	0.038	0.035	-0.090	-0.018	0.065
15	РАБОТА ОТЦА-врач ...		0.092	-0.057	0.100	-0.025	0.110	0.022	0.071
16	РАБОТА ОТЦА-другое ...	0.007	0.026	0.003	-0.013	-0.050	-0.027	-0.005	0.029
17	РАБОТА ОТЦА-на дому ...	0.116	-0.129	-0.044	0.071	-0.054	-0.040	-0.008	0.100
18	РАБОТА ОТЦА-сервис ...	0.001	-0.028	0.031	-0.020	-0.083	-0.098	-0.020	0.042
19	РАБОТА ОТЦА-учитель ...	0.013	-0.120	-0.110	0.038	0.340	0.160	0.032	0.186
20	АКТИВИСТ ВНЕ ШКОЛЫ-да ...	0.016	-0.058	0.012	-0.002	0.038	0.005	0.001	0.036

Таблица 3 – Матрица знаний (модель INF3) (фрагмент)

Код признака	Наименование описательной шкалы и градации	1. УСПЕВАЕМ... ОЧЕНЬ МАЛОЕ: 1/5 {0.0000000, 4.0000000}	2. УСПЕВАЕМ... МАЛОЕ: 2/5 {4.0000000, 8.0000000}	3. УСПЕВАЕМ... СРЕДНЕЕ: 3/5 {8.0000000, 12.0000000}	4. УСПЕВАЕМ... БОЛЬШОЕ: 4/5 {12.0000000, 16.0000000}	5. УСПЕВАЕМ... ОЧЕНЬ БОЛЬШОЕ: 5/5 {16.0000000, 20.0000000}	Сумма	Среднее	Средн. квадр. откл.
1	ПОЛ-ж	3.463	0.825	1.694	-3.344	-2.638			2.902
2	ПОЛ-м	-3.463	-0.825	-1.694	3.344	2.638			2.902
3	ВОЗРАСТ-Очень малое: 1/5-{15.0000000, 16.40...}	-5.365	-3.666	-2.284	8.615	2.699			5.677
4	ВОЗРАСТ-Малое: 2/5-{16.4000000, 17.8000000} ...	-0.676	-0.630	4.808	-1.547	-1.954			2.747
5	ВОЗРАСТ-Среднее: 3/5-{17.8000000, 19.200000...}	6.534	3.094	-1.473	-6.714	-1.441			5.040
6	ВОЗРАСТ-Большое: 4/5-{19.2000000, 20.600000...}	-0.296	-0.478	-0.230	0.187	0.818			0.518
7	ВОЗРАСТ-Очень большое: 5/5-{20.6000000, 22....}	-0.197	1.681	-0.820	-0.542	-0.122			0.981
8	РАЗМЕР СЕМЬИ-более 3-х ...	4.256	1.182	-2.246	-2.119	-1.073			2.747
9	РАЗМЕР СЕМЬИ-до 3-х ...	-4.256	-1.182	2.246	2.119	1.073			2.747
10	РАБОТА МАТЕРИ-врач ...	-1.357	-2.423	-3.944	7.790	-0.066			4.581
11	РАБОТА МАТЕРИ-другое ...	1.078	5.511	2.172	-3.195	-5.567			4.400
12	РАБОТА МАТЕРИ-на дому ...	3.175	-0.410	1.803	-1.982	-2.585			2.455
13	РАБОТА МАТЕРИ-сервис ...	-1.170	-0.428	-1.243	-4.901	7.742			4.665
14	РАБОТА МАТЕРИ-учитель ...	-1.727	-2.251	1.213	2.289	0.476			1.935
15	РАБОТА ОТЦА-врач ...	-1.777	1.129	-1.382	2.124	-0.094			1.648
16	РАБОТА ОТЦА-другое ...	0.575	3.390	1.003	-2.782	-2.185			2.517
17	РАБОТА ОТЦА-на дому ...	1.025	-1.190	-1.203	1.582	-0.215			1.271
18	РАБОТА ОТЦА-сервис ...	0.041	-1.704	5.476	-2.068	-1.744			3.171
19	РАБОТА ОТЦА-учитель ...	0.137	-1.625	-3.894	1.144	4.238			3.045
20	АКТИВИСТ ВНЕ ШКОЛЫ-да ...	1.154	-6.058	3.565	-0.448	1.787			3.679

1.5. Результаты верификации моделей

Результаты верификации (оценки достоверности) моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями приведены на рисунке 12.

Наименование модели и частного критерия	Интегральный критерий	га	L1-мера проф. Е.В.Луценко	Средний модуль уровней сходств истинно-положит. решений	Средний модуль уровней сходств истинно-отрицат. решений	Средний модуль уровней сходств ложно-положит. решений	Средний модуль уровней сходств ложно-отрицат. решений	A-Точность модели APrecision = ATP/ATP...	A-Полнота модели APRecall = ATP/ATP...	L2-мера проф. Е.В.Луценко	Процент правильности
1. ABS - частный критерий: количество встреч сочтаний: "клас...	Корреляция абс частот с абс...	0.000	0.356	0.570	0.004	0.517	0.090	0.524	0.992	0.686	91
1. ABS - частный критерий: количество встреч сочтаний: "клас...	Сумма абс частот по призна...	0.000	0.441	0.575		0.365		0.612	1.000	0.759	100
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн. частот с о...	0.000	0.356	0.570	0.004	0.517	0.090	0.524	0.992	0.686	91
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн. частот по приз...	0.000	0.339	0.733		0.716		0.506	1.000	0.672	100
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн. частот с о...	0.000	0.356	0.570	0.004	0.517	0.090	0.524	0.992	0.686	91
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн. частот по приз...	0.000	0.339	0.733		0.716		0.506	1.000	0.672	100
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	0.850	0.522	0.309	0.143	0.216	0.246	0.588	0.683	0.632	72
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	0.709	0.469	0.103	0.064	0.091	0.179	0.530	0.616	0.570	68
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	0.850	0.522	0.309	0.143	0.216	0.246	0.588	0.683	0.632	72
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	0.709	0.469	0.103	0.064	0.091	0.179	0.530	0.616	0.570	68
6. INF3 - частный критерий: Хинквадрат, разности между фактич...	Семантический резонанс зна...	0.730	0.426	0.310	0.214	0.262	0.267	0.543	0.592	0.566	65
6. INF3 - частный критерий: Хинквадрат, разности между фактич...	Сумма знаний	0.722	0.438	0.334	0.240	0.265	0.276	0.558	0.582	0.570	68
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.723	0.533	0.299	0.156	0.218	0.236	0.578	0.657	0.615	81
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.846	0.443	0.104	0.038	0.092	0.093	0.529	0.730	0.613	61
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.723	0.533	0.299	0.156	0.218	0.236	0.578	0.657	0.615	81
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.846	0.443	0.104	0.038	0.092	0.093	0.529	0.730	0.613	61
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	0.721	0.425	0.310	0.216	0.263	0.270	0.541	0.589	0.564	64
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	0.815	0.385	0.149	0.063	0.159	0.157	0.483	0.702	0.572	61
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	0.721	0.425	0.310	0.216	0.263	0.270	0.541	0.589	0.564	64
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	0.815	0.385	0.149	0.063	0.159	0.157	0.483	0.702	0.572	64

Рисунок 12 – Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели INF4, INF5 и ABS при интегральном критерии «Резонанс знаний». При этом достоверность модели в соответствии с критерием согласия Пирсона составляет 0,759, что является очень хорошим результатом. 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 F-критерий Ван Ризбергена, а также L-критерий, предложенный проф.Е.В.Луценко [11] и являющийся нечетким мультиклассовым обобщением F-критерия.

На рисунке 12 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и неидентифицированных ситуаций в наиболее достоверной модели INF4.

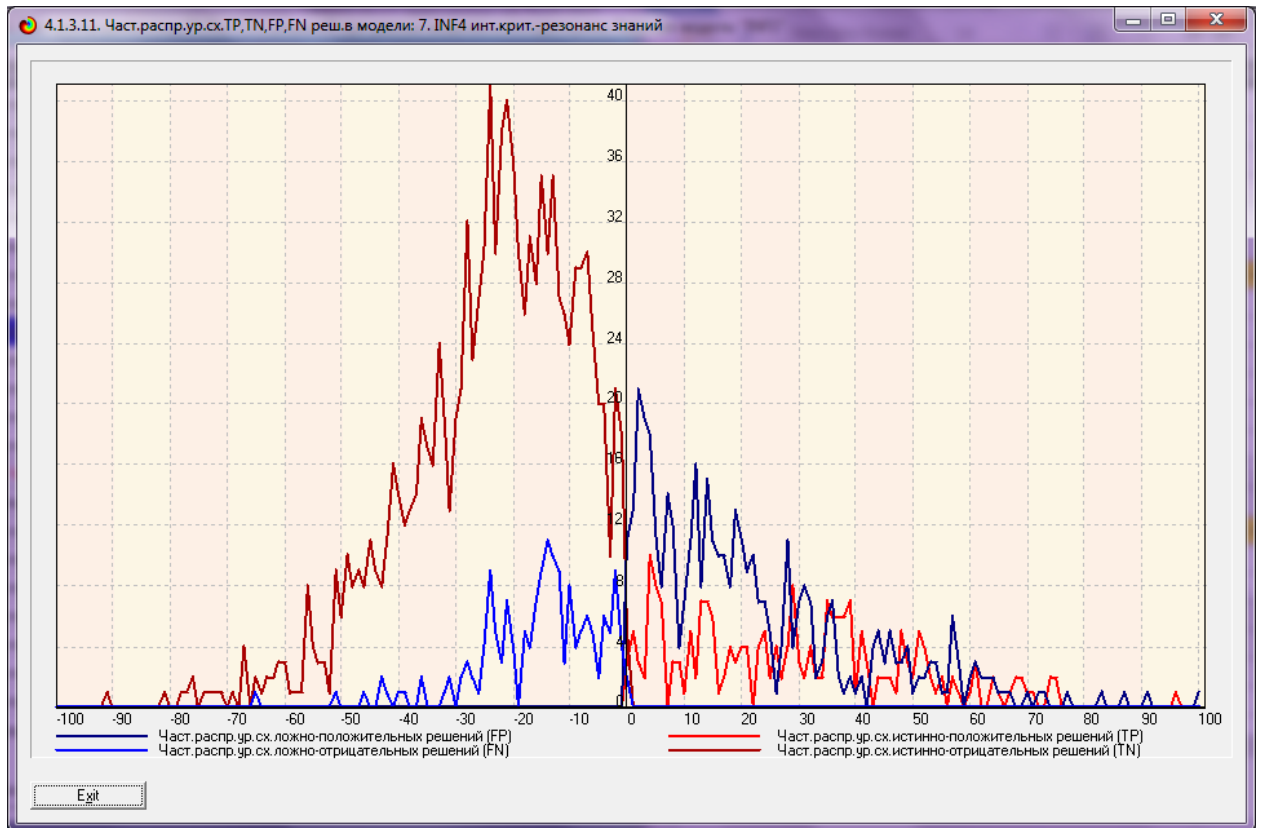


Рисунок 12 – Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF4.

Из рисунка 13 видно, что:

- наиболее достоверная модель INF4 лучше определяет непринадлежность объекта к классу, чем принадлежность.

- модуль уровня сходства-различия в наиболее достоверной модели INF2 для верно идентифицированных и верно неидентифицированных объектов значительно выше, чем для ошибочно идентифицированных и ошибочно неидентифицированных. Это верно практически для всего диапазона уровней сходства-различия, кроме небольших по значений в диапазоне от 0 до 60. Для очень больших значений уровней сходства-различия (более 50%) также различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями практически отсутствует.

2 РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи

В соответствии с технологией АСК-анализа зададим текущей модель INF4 (режим 5.6) (рисунок 13) и проведем пакетное распознавание в режиме 4.2.1 (рисунок 14).

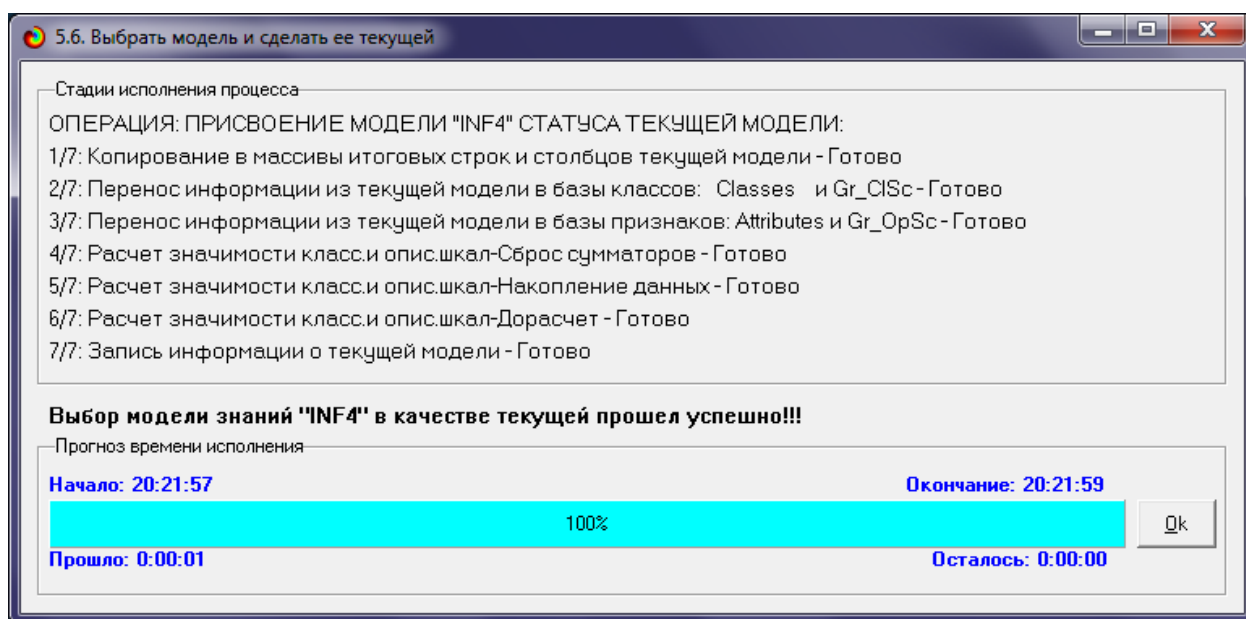


Рисунок 13 – Экранные формы режима задания модели в качестве текущей.

В результате пакетного распознавания в текущей модели создается ряд баз данных, которые визуализируются в выходных экранных формах, отражающих результаты решения задачи идентификации и прогнозирования.

4.2.1. Информационные портреты классов

Инф.портрет класса: 1 "УСПЕВАЕМОСТЬ-Очень малое: 1/5-{0.0000000, 4.0000000}" в модели: 7 "INF4"

Код	Наименование признака	Значимость
1	УСПЕВАЕМОСТЬ-Очень малое: 1/5-{0.0000000, 4.0000000,...	
2	УСПЕВАЕМОСТЬ-Малое: 2/5-{4.0000000, 8.0000000,...	
3	УСПЕВАЕМОСТЬ-Среднее: 3/5-{8.0000000, 12.0000000,...	
4	УСПЕВАЕМОСТЬ-Большое: 4/5-{12.0000000, 16.0000000,...	
5	УСПЕВАЕМОСТЬ-Очень большое: 5/5-{16.0000000, 20.0000000,...	
5	ВОЗРАСТ-Среднее: 3/5-{17.8000000, 19.2000000}	0.624
24	В РОМАНТИЧЕСКИХ ОТНОШЕНИЯХ-да	0.611
37	ЗДОРОВЬЕ-Малое: 2/5-{1.8000000, 2.6000000}	0.575
12	РАБОТА МАТЕРИ-на дому	0.545
17	РАБОТА ОТЦА-на дому	0.519
32	УРОВЕНЬ ПОТРЕБЛЕНИЯ АЛКОГОЛЯ-Малое: 2/5-{1.8000000, 2.6000000}	0.430
28	СВОБОДНОЕ ВРЕМЯ ПОСЛЕ ШКОЛЫ-Среднее: 3/5-{2.6000000, 3.4000000}	0.355
23	ИМЕЕТ ДОСТУП В ИНТЕРНЕТ ДОМА-нет	0.228
39	ЗДОРОВЬЕ-Большое: 4/5-{3.4000000, 4.2000000}	0.228
31	УРОВЕНЬ ПОТРЕБЛЕНИЯ АЛКОГОЛЯ-Очень малое: 1/5-{1.0000000, 1.8000000,...	0.207
1	ПОЛ-ж	0.169
8	РАЗМЕР СЕМЬИ-более 3-х	0.153
11	РАБОТА МАТЕРИ-другое	0.077
26	СВОБОДНОЕ ВРЕМЯ ПОСЛЕ ШКОЛЫ-Очень малое: 1/5-{1.0000000, 1.8000000,...	0.066
20	АКТИВИСТ ВНЕ ШКОЛЫ-да	0.058
19	РАБОТА ОТЦА-учитель	0.048
16	РАБОТА ОТЦА-другое	0.027
18	РАБОТА ОТЦА-сервис	0.004
22	ИМЕЕТ ДОСТУП В ИНТЕРНЕТ ДОМА-да	-0.046
21	АКТИВИСТ ВНЕ ШКОЛЫ-нет	-0.060
4	ВОЗРАСТ-Малое: 2/5-{16.4000000, 17.8000000}	-0.070
40	ЗДОРОВЬЕ-Очень большое: 5/5-{4.2000000, 5.0000000}	-0.098
38	ЗДОРОВЬЕ-Среднее: 3/5-{2.6000000, 3.4000000}	-0.110

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 MS Excel ВКЛ. фильтр по фактору ВЫКЛ. фильтр по фактору Вписать в окно Показать ВСЕ

Рисунок 15 – Экранная форма информационного портрета класса в модели INF4

Режим 4.1.3 системы «Эйдос» обеспечивает отображение результатов идентификации и прогнозирования в различных формах:

1. Подробно наглядно: "Объект – классы".
2. Подробно наглядно: "Класс – объекты".
3. Итоги наглядно: "Объект – классы".
4. Итоги наглядно: "Класс – объекты".
5. Подробно сжато: "Объект – классы".
6. Обобщенная форма по достоверности моделей при разных интегральных критериях.
7. Обобщенный статистический анализ результатов идентификации по моделям и интегральным критериям.
8. Статистический анализ результатов идентификации по классам, моделям и интегральным критериям.

9. Распознавание уровня сходства при разных моделях и интегральных критериях.

10. Достоверность идентификации классов при разных моделях и интегральных критериях.

Ниже кратко рассмотрим некоторые из них. На рисунке 16 и 17 приведены примеры прогнозов высокой и низкой достоверности индекса качества жизни в наиболее достоверной модели INF4 на основе рейтинга UCI.

Следует заметить, что при данном вычислении были выявлены погрешности в рейтинге UCI.

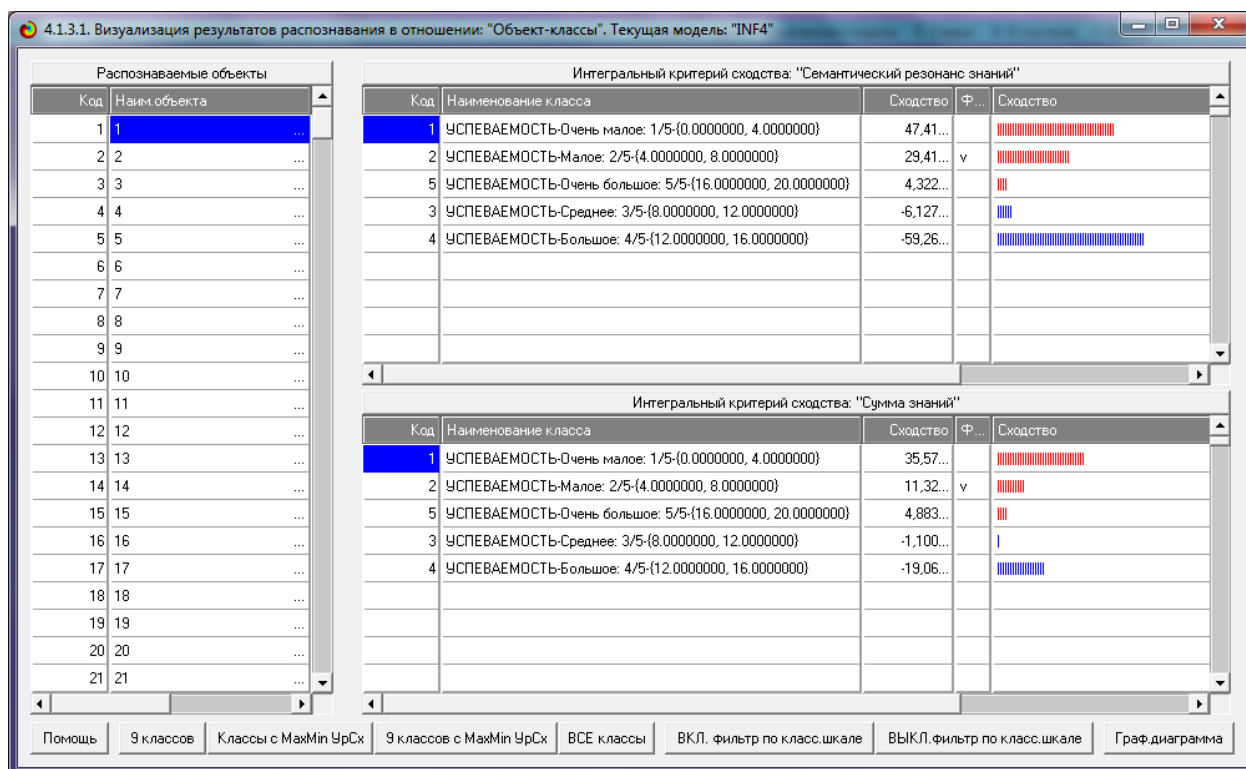


Рисунок 15 – Пример оценки успеваемости учеников в модели INF4

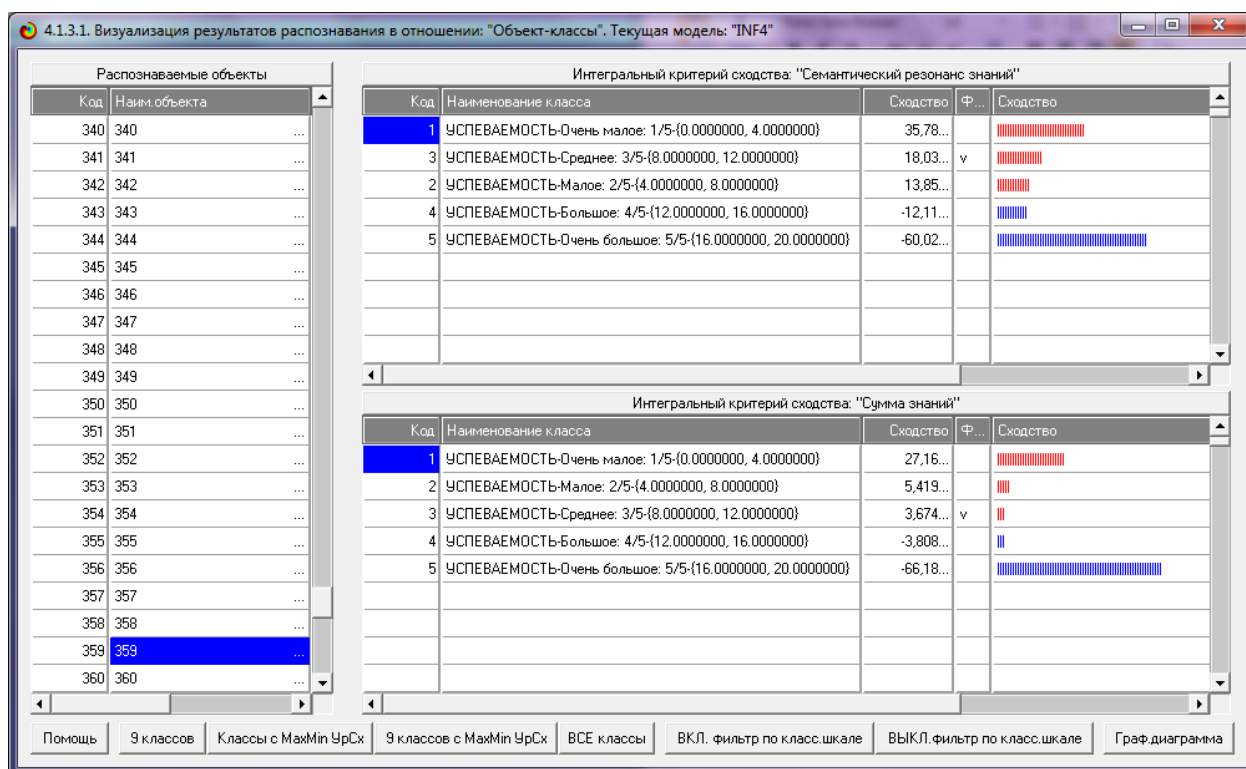


Рисунок 17 – Пример оценки успеваемости учеников в модели INF4

2.2. Когнитивные функции

Рассмотрим режим 4.5, в котором реализована возможность визуализации когнитивных функций для любых моделей и любых сочетаний классификационных и описательных шкал (рисунок 18).

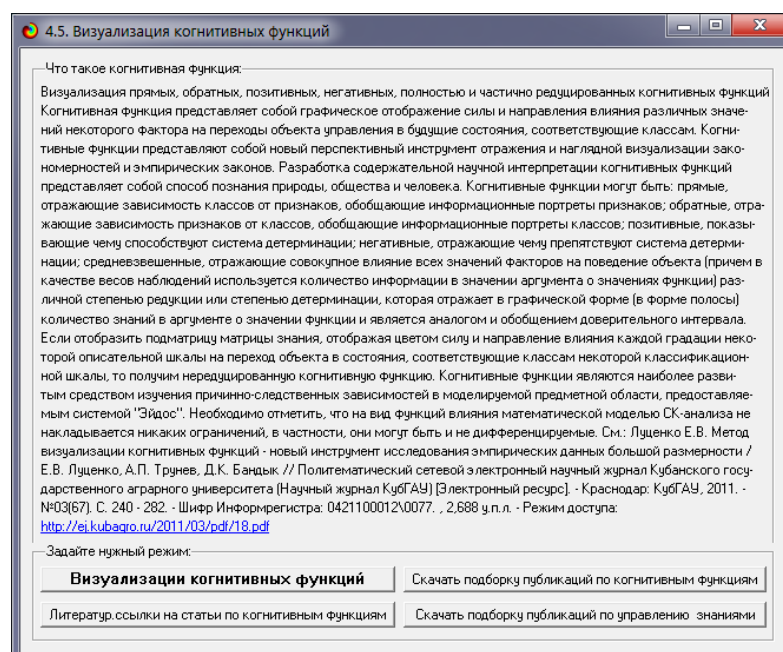


Рисунок 18 – Визуализация когнитивных функций

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 много работ автора, но наиболее новой и обобщающей из них является работа. Поэтому здесь не будем останавливаться на описании того, что представляют собой когнитивные функции в АСК-анализе. На рисунках 18 и 19 приведены визуализации всех когнитивных функций данного приложения для модели INF4.

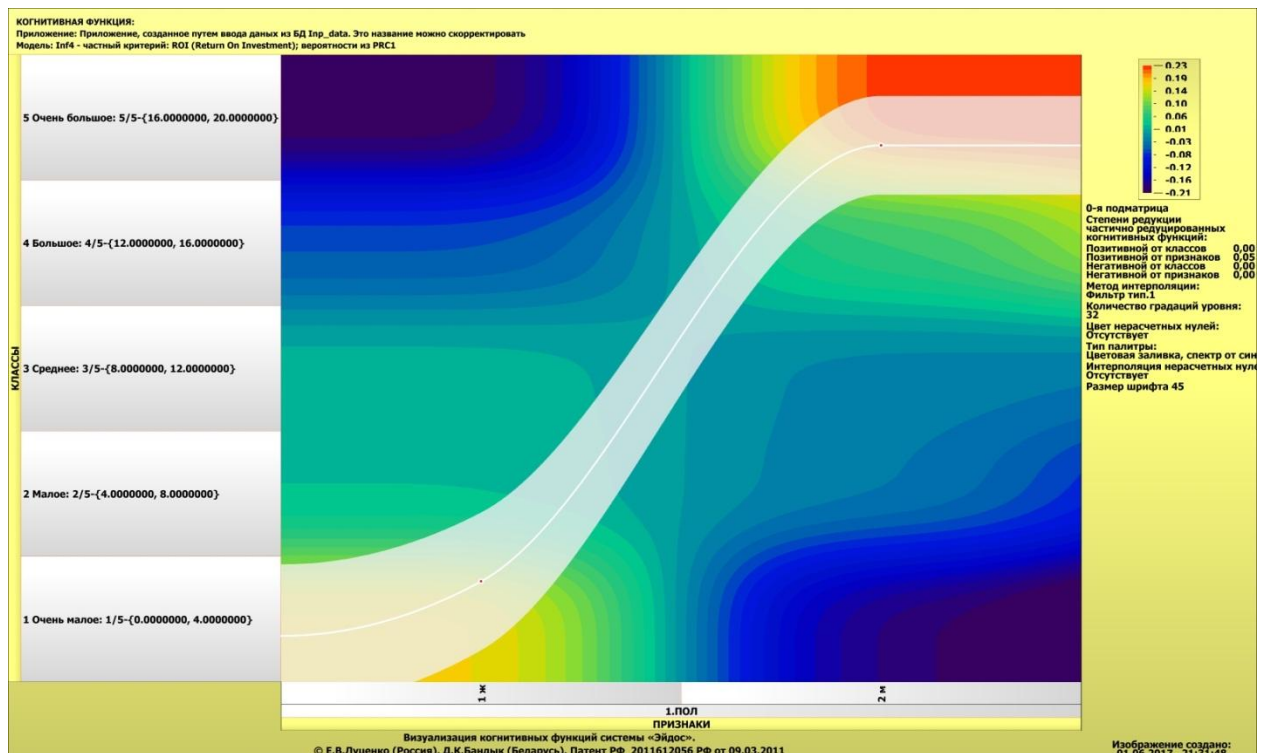


Рисунок 18 – Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF4.

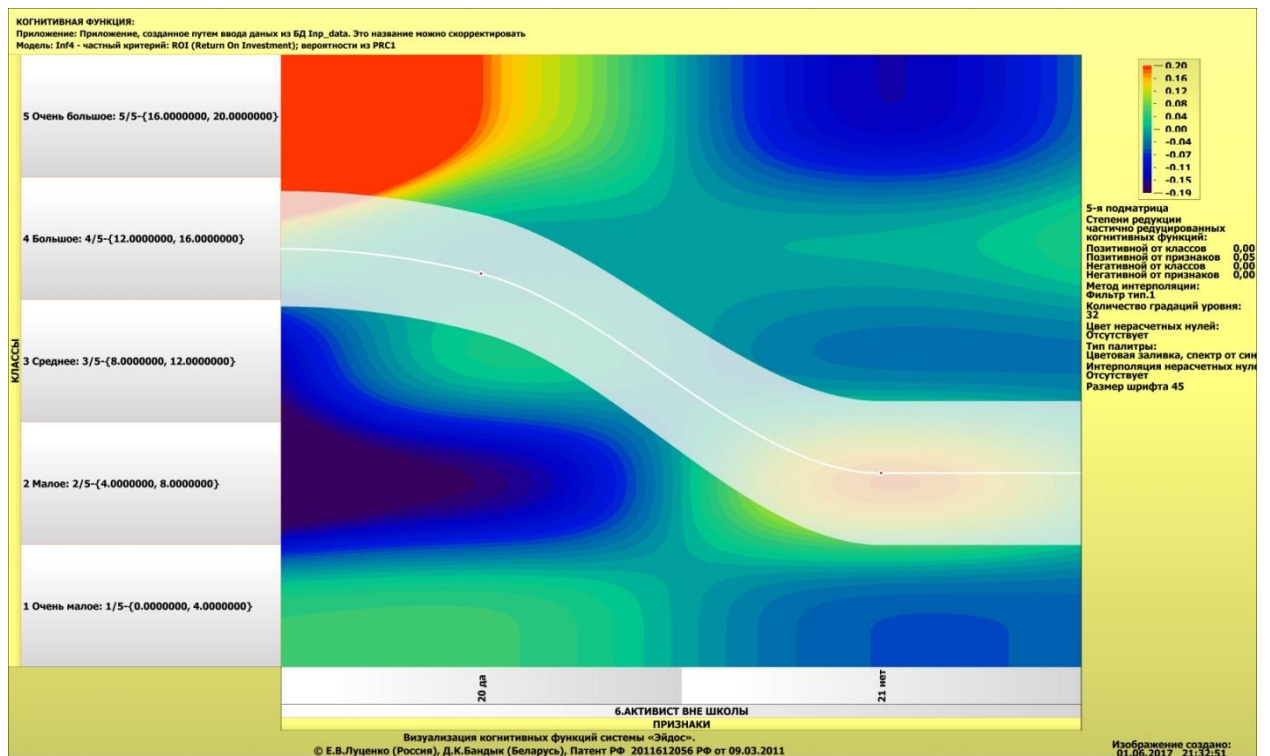
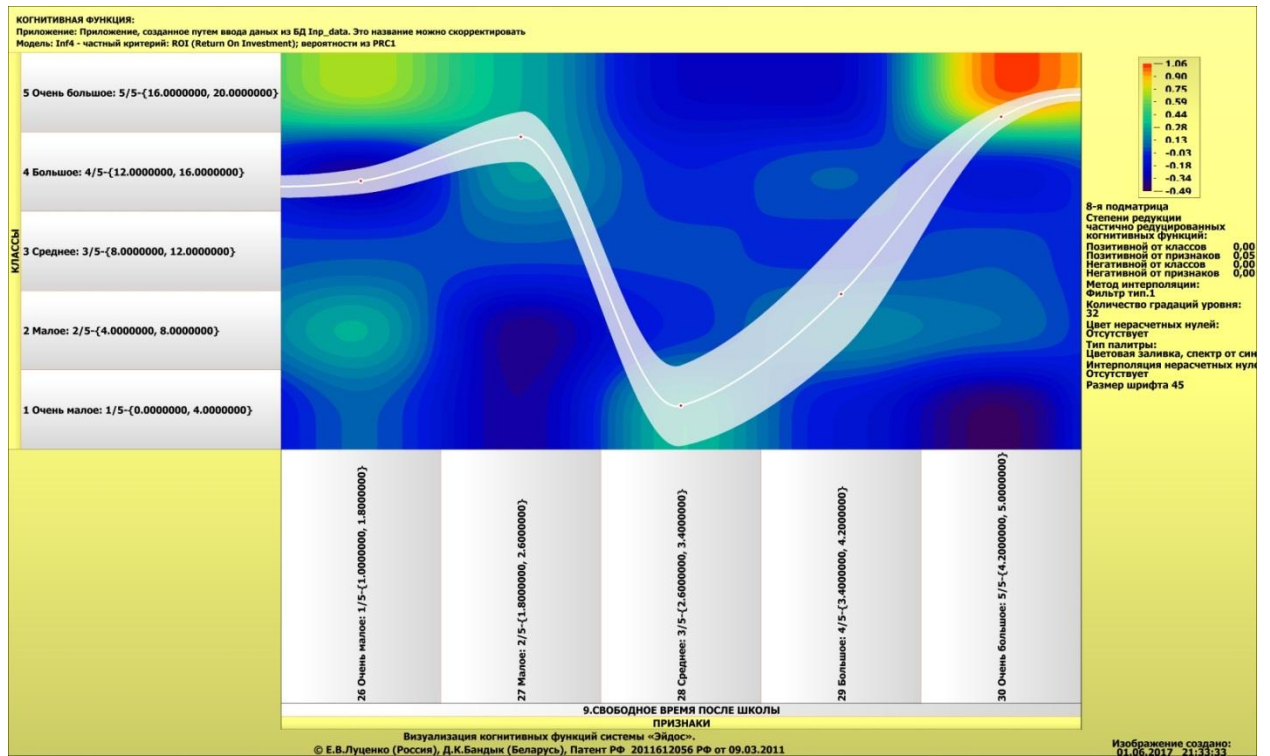


Рисунок 19 – Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF4

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 20).

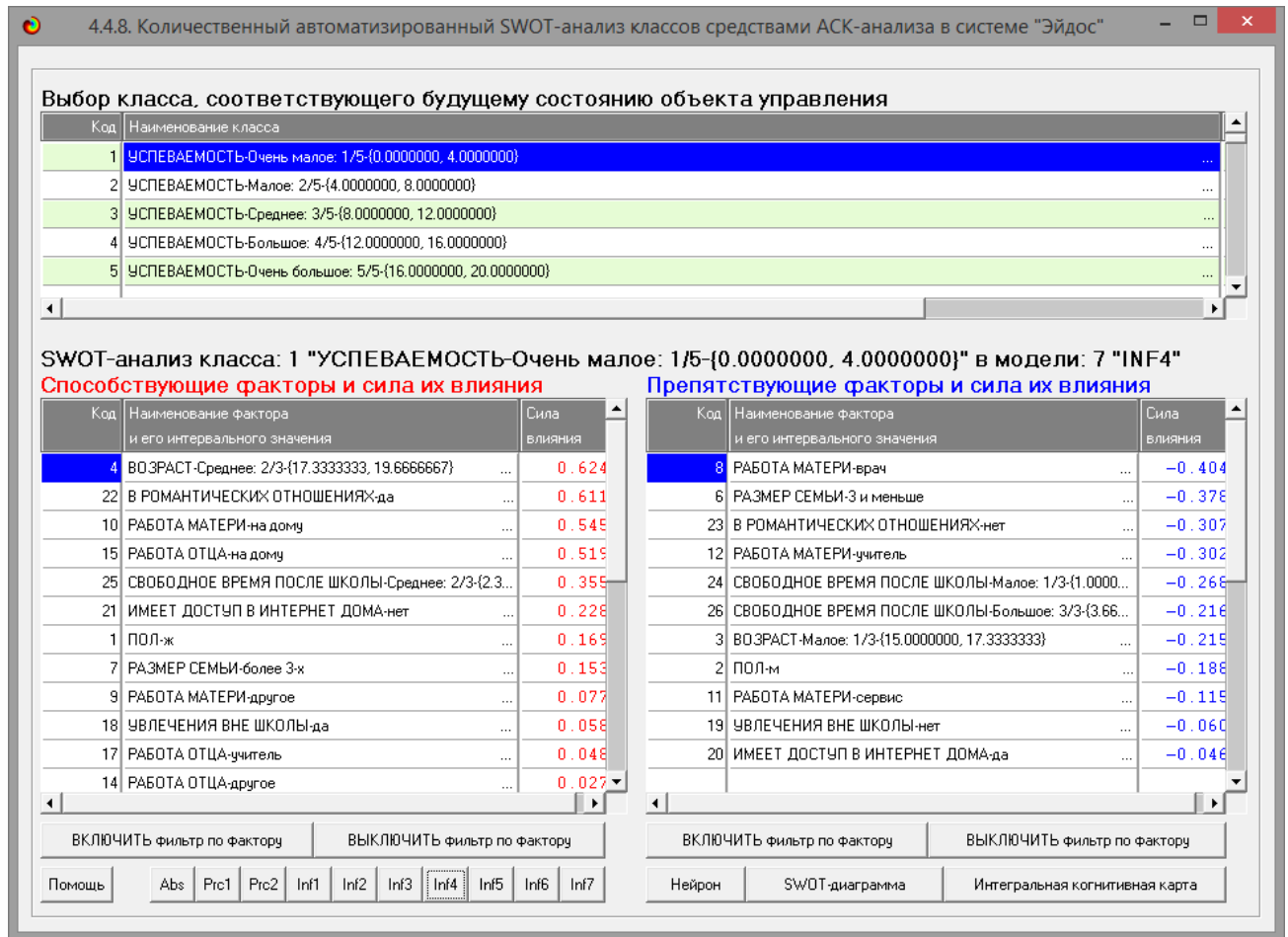


Рисунок 20 – Пример SWOT-матрицы в модели INF4

На рисунке 21 приведены примеры инвертированной SWOT- матрицы и инвертированной SWOT-диаграммы в модели INF4.

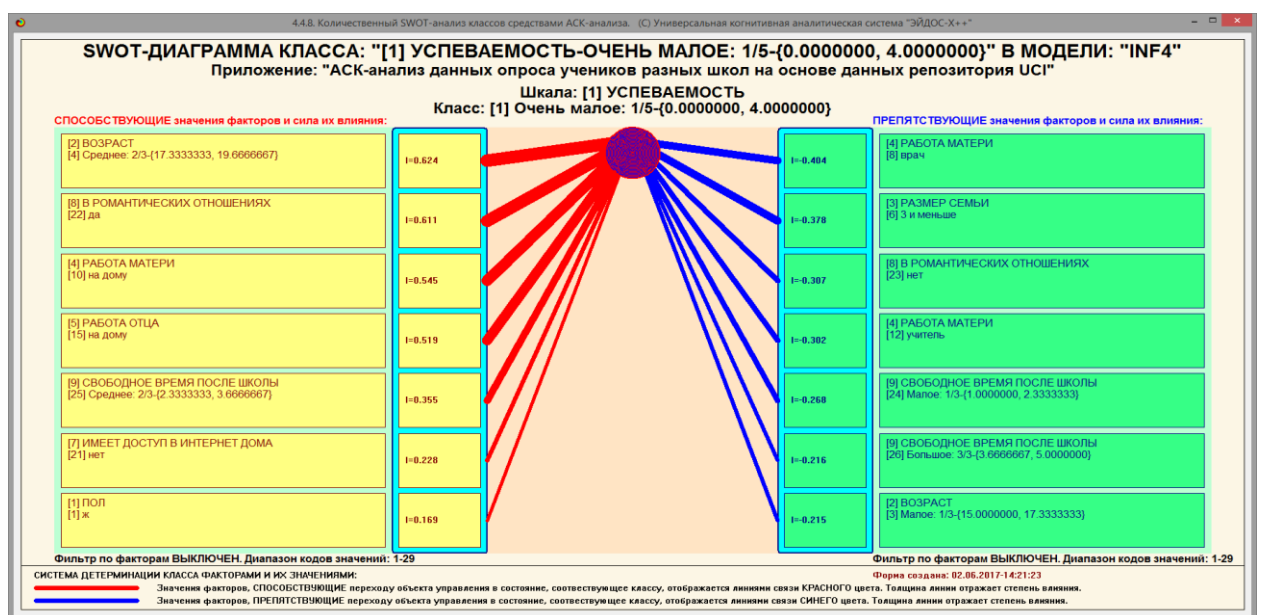


Рисунок 21. Пример SWOT-матрицы в модели INF4

НЕЛОКАЛЬНЫЙ НЕЙРОН В МОДЕЛИ: "INF4"
 Нейрон: [1]-УСПЕШАЕМОСТЬ-Очень малое: 1/5-(0.000000, 4.000000)
 Приложение: АСК-анализ данных опроса учеников разных школ на основе данных репозитория UCI

Формы созданы: 02.06.2017-14:23:44

Влияние рецепторов на активацию/торможение нелокального нейрона, соответствующего классу [система детерминации класса]:
 АКТИВИРУЮЩЕЕ влияние отображается линией КРАСНОГО цвета, толщина линии [приведенная в кружочке в центре линии] отражает относительную силу влияния.
 ТОРМОЗЯЩЕЕ влияние отображается линией СИНЕГО цвета, толщина линии [приведенная в кружочке в центре линии] отражает относительную силу влияния.

Сортировка рецепторов по информативности
 Отображается количество репозитория не более: 999
 Показаны связи с относительной силой влияния выше: 0%
 Визуализация нейрона с кодами и наименованиями рецепторов

В качестве примера приведено графическое отображение нелокальных нейронов в модели INF4 на основе признака «Успеваемость-Очень малое». Каждому классу соответствует нейрон, совокупность которых образует нелокальную нейронную сеть. Отсюда можно сделать такой же вывод.

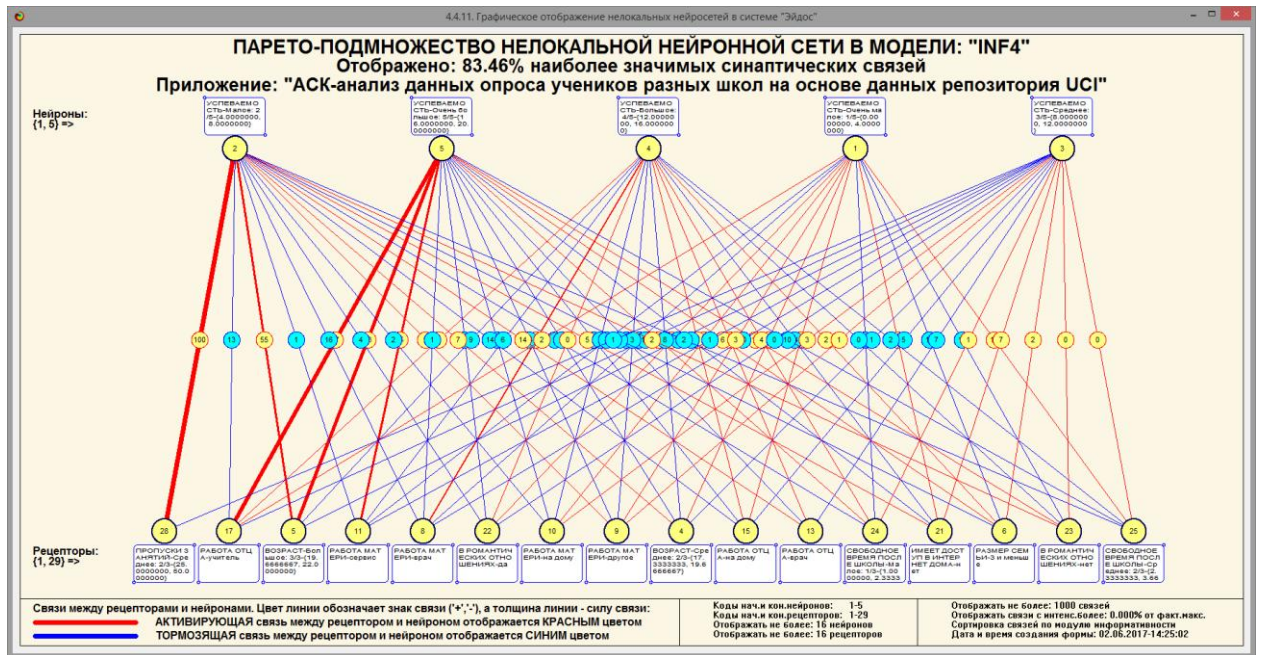


Рисунок 23. Пример графического отображения парето-подмножества нелокальной нейронной сети

На рисунке 23 в графическом отображения парето-подмножества нелокальной нейронной сети в модели INF4 у классов слева сильнее связи между значениями факторов и результатом [9].

Из рисунка 23 видно, что, например, 5 и 17 (успеваемость и работа отца) признаки имеют сходство.

График данного анализа представлен на рисунке 25.

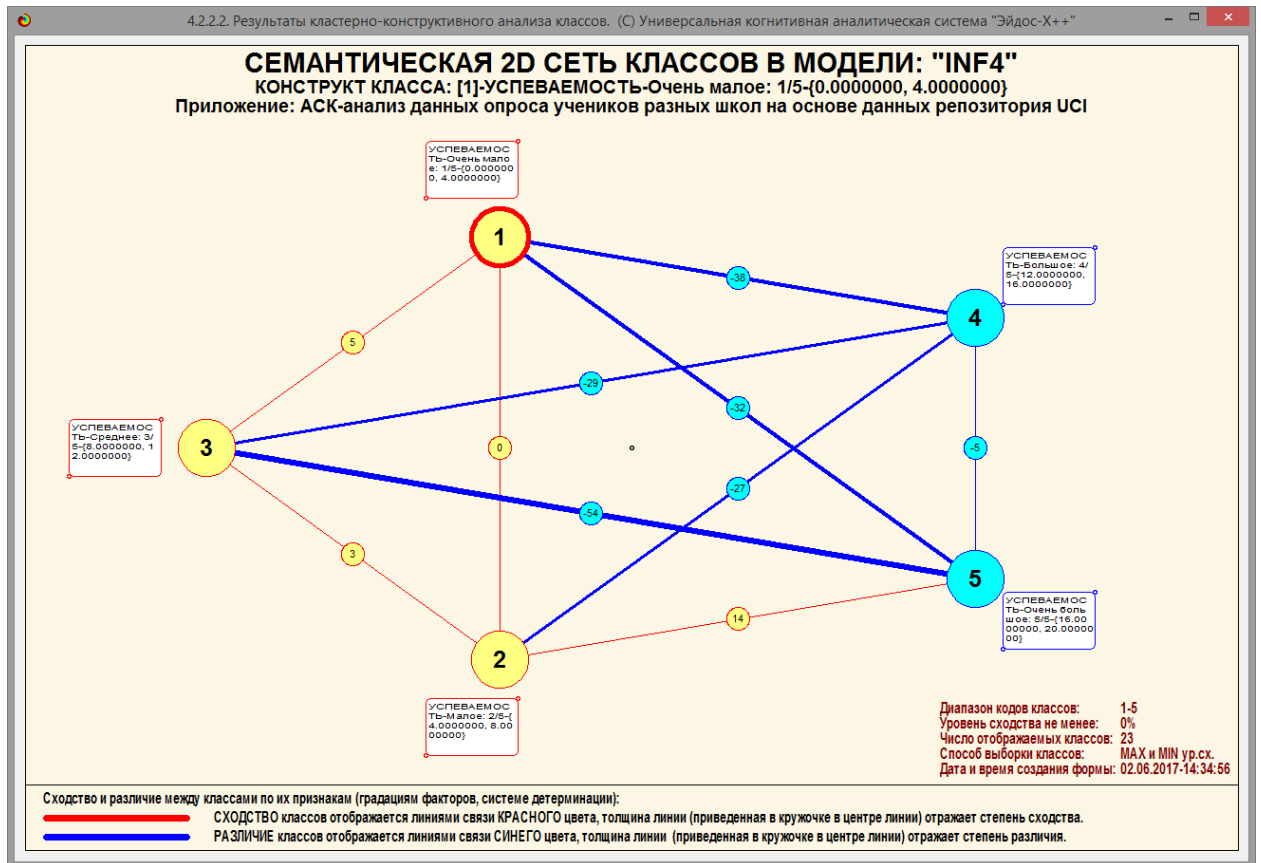


Рисунок 25 – Графические результаты кластерного и конструктивного анализа классов

ЗАКЛЮЧЕНИЕ

Так как существует множество систем искусственного интеллекта, то возникает необходимость сопоставимой оценки качества их математических моделей. Одним из вариантов решения этой задачи является тестирование различных системы на общей базе исходных данных, для чего очень удобно использовать общедоступные базы. В данной работе приводится развернутый пример использования базы данных UCI для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». При этом наиболее достоверной в данном приложении оказалась модель INF4 при интегральном критерии «Резонанс знаний». Точность модели составляет 0,759, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 70%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется F-мера Ван Ризбергена и ее нечеткое мультиклассовое обобщение L-мера, предложенная проф.Е.В.Луценко. Если на основе базы данных UCI, рассмотренной в данной работе, построить модели прогнозирования не с помощью АСК-анализа и реализующей его системы «Эйдос», а с применением других математических методов и реализующих их программных систем, то можно сопоставимо сравнить их качество.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. База данных numbeo.com [Электронный ресурс],-Режим доступа: <https://www.numbeo.com/quality-of-life/rankings.jsp>, свободный, Загл. с экрана, -Яз. англ.
2. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе «Эйдос» // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 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.
3. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.
4. 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: КубГАУ, 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.