

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 зависимости сброса загрязненных сточных вод в поверхностные
водные объекты по субъектам Российской Федерации

Выполнил студент
Группы ПИ1401
Огульчанский Дмитрий
Сергеевич

Руководитель работы:
профессор Луценко Е.В.

Краснодар

2017

Оглавление

Введение	3
1 СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	4
1.1 Описание решения	4
1.2 Преобразование исходных данных из HTML-формата в файл исходных данных MS Excel	4
1.3 Синтез и верификация статистических и интеллектуальных моделей.....	11
1.4 Виды моделей системы «Эйдос»	13
1.5 Результаты верификации моделей.....	16
2 РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	19
2.1 Решение задачи идентификации	19
2.2 SWOT и PEST Матрицы и диаграммы.....	22
2.3 Кластерно-конструктивный анализ признаков.....	25
2.4 Нелокальный нейрон и нелокальные нейронные сети.....	26
Заключение	29
Использованная литература.....	30

ВВЕДЕНИЕ

Создание систем искусственного интеллекта является одним из важных и перспективных направлений развития современных информационных технологий. Так как существует множество альтернатив систем искусственного интеллекта, то возникает необходимость оценки качества математических моделей этих систем. В данной работе проведён АСК-анализ зависимости сброса загрязнённых сточных вод в поверхностные водные объекты по субъектам Российской Федерации.

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

В данной работе использована база данных «mesta sbrosa».

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

1 СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

1.1 Описание решения

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

1. Преобразование исходных данных из HTML-формата в промежуточные файлы MS Excel.
2. Преобразование исходных данных из промежуточных файлов MS Excel в базы данных системы «Эйдос».
3. Синтез и верификация моделей предметной области.
4. Применение моделей для решения задач идентификации, прогнозирования и исследования предметной области.

1.2 Преобразование исходных данных из HTML-формата в файл исходных данных MS Excel

Из электронного ресурса данных http://statistika.ru/stat/stat4/2007/12/21/stat4_10136.html возьмем данные и переведем в Excel, где получим следующие колонки:

1. Регион
2. Всего за 1995-2005гг.
3. 1995 г.
4. 2000 г.
5. 2001 г.
6. 2002 г.
7. 2003 г.
8. 2004 г.
9. 2005 г.

Столбец 2 является классификационной шкалой, а столбцы 3-9 описательными шкалами.

Рисунок 3 - Таблица «Сброс загрязненных сточных вод в поверхностные водные объекты по субъектам Российской Федерации» (продолжение)

Сохраним получившуюся таблицу с именем Inp_data.xlsx в папку:
D:\Aidos-X\AID_DATA\Inp_data\.

Автоматизированная формализация предметной области путем импорта исходных данных из внешних данных в систему «Эйдос».

Для загрузки базы 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 4)

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

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) Стандарт XLSX-файла
- ☐ 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: 9

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 сценарный метод АСК-анализа
- ☐ Применить спец. интерпретацию текстовых полей классов
- ☐ 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 4 – Экранная форма Универсального программного интерфейса импорта данных в систему «Эйдос» (режим 2.3.2.2)

В данной экранной форме, приведённой выше, задать настройки: — «Задайте тип файла исходных данных Inp_data»: «XLSX – MS Excel- 2007»; — «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» – 9; — «Задание параметров формирования сценариев или способа интерпретации текстовых полей»: «Не применять сценарный метод АСК - анализа и спец. интерпретацию ТХТ-полей».

После нажать кнопку «ОК». Далее открывается окно, где размещена информация о размерности модели (Рисунок 5). В данном окне необходимо нажать кнопку «Выйти на создание модели»

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

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [4 x 28]

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	4	4,00	7	28	4,00
Текстовые	0	0	0,00	0	0	0,00
ВСЕГО:	1	4	4,00	7	28	4,00

Задайте число интервалов (градаций) в шкале:

В классификационных шкалах: В описательных шкалах:

Пересчитать шкалы и градации Параметры числ. шкал и градаций Выйти на создание модели

Рисунок 5 – Задание размерности модели системы «Эйдос»

Далее открывается окно, отображающее стадию процесса импорта данных из внешней БД «Inp_data.xlsx» в систему «Эйдос» (Рисунок 6), а так же прогноз времени завершения этого процесса. В этом окне необходимо дождаться завершения формализации предметной области и нажать кнопку «ОК».

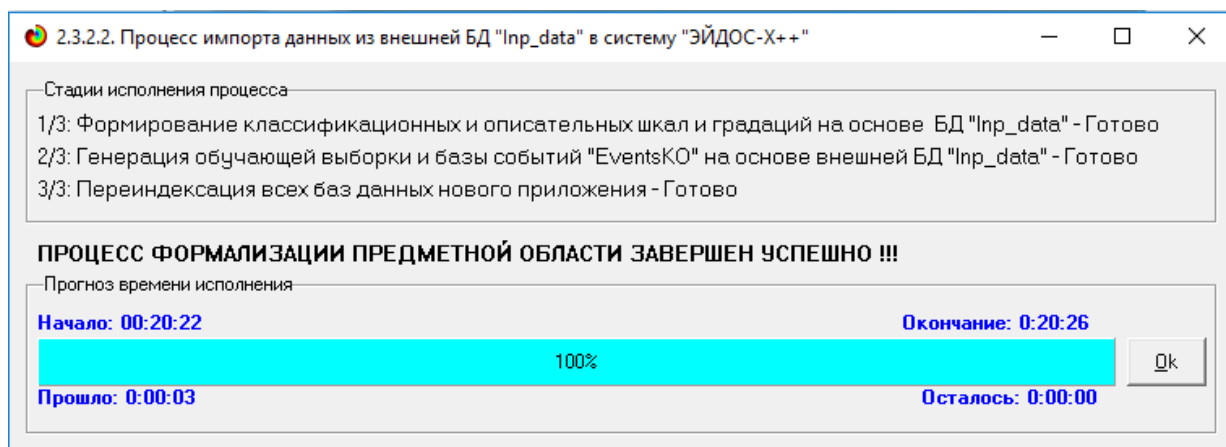


Рисунок 6 – Процесс импорта данных из внешней БД «Inp_data.xlsx» в систему «Эйдос»

В результате формируются классификационные и описательные шкалы и градации, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Этим самым полностью автоматизировано выполняется 2-й этап АСК-анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1 (рисунок 7)

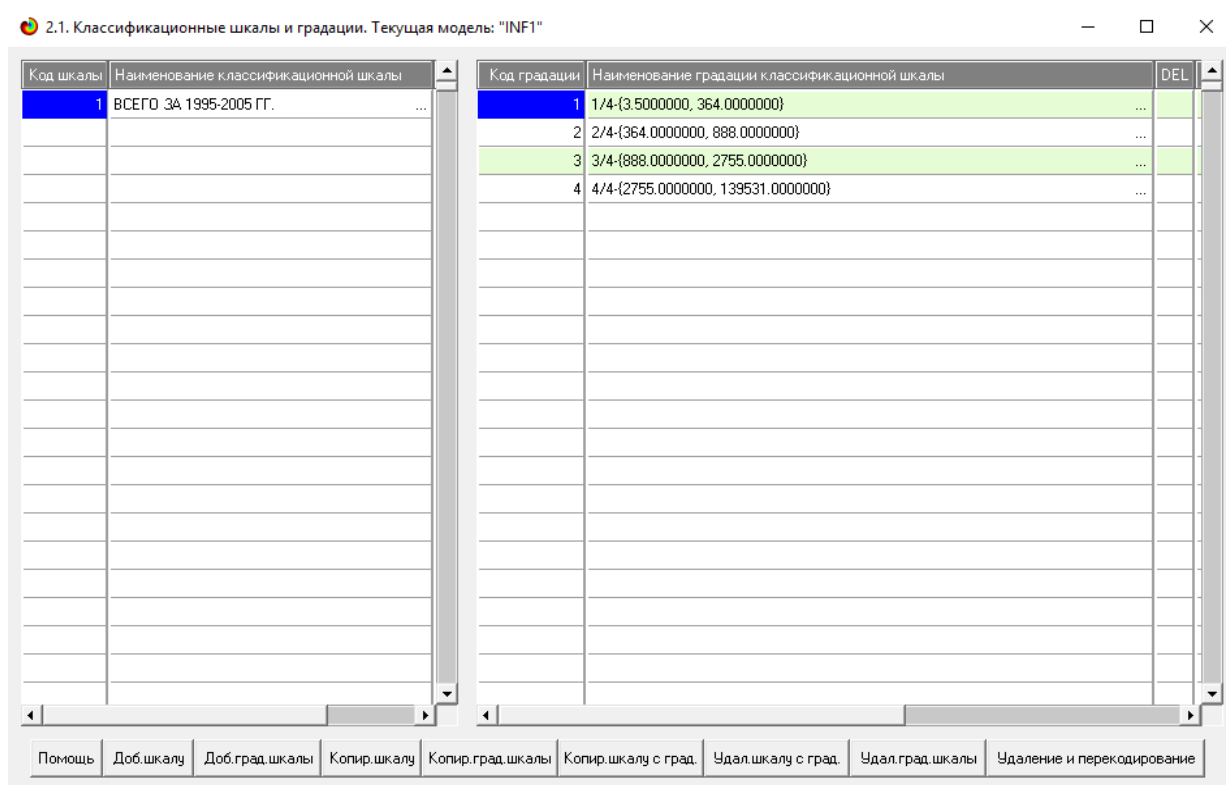


Рисунок 7 – Классификационные шкалы и градации

Для просмотра описательных шкал и градаций необходимо запустить режим 2.2 (Рисунок 8), а обучающей выборки режим 2.3.1 (Рисунок 9):

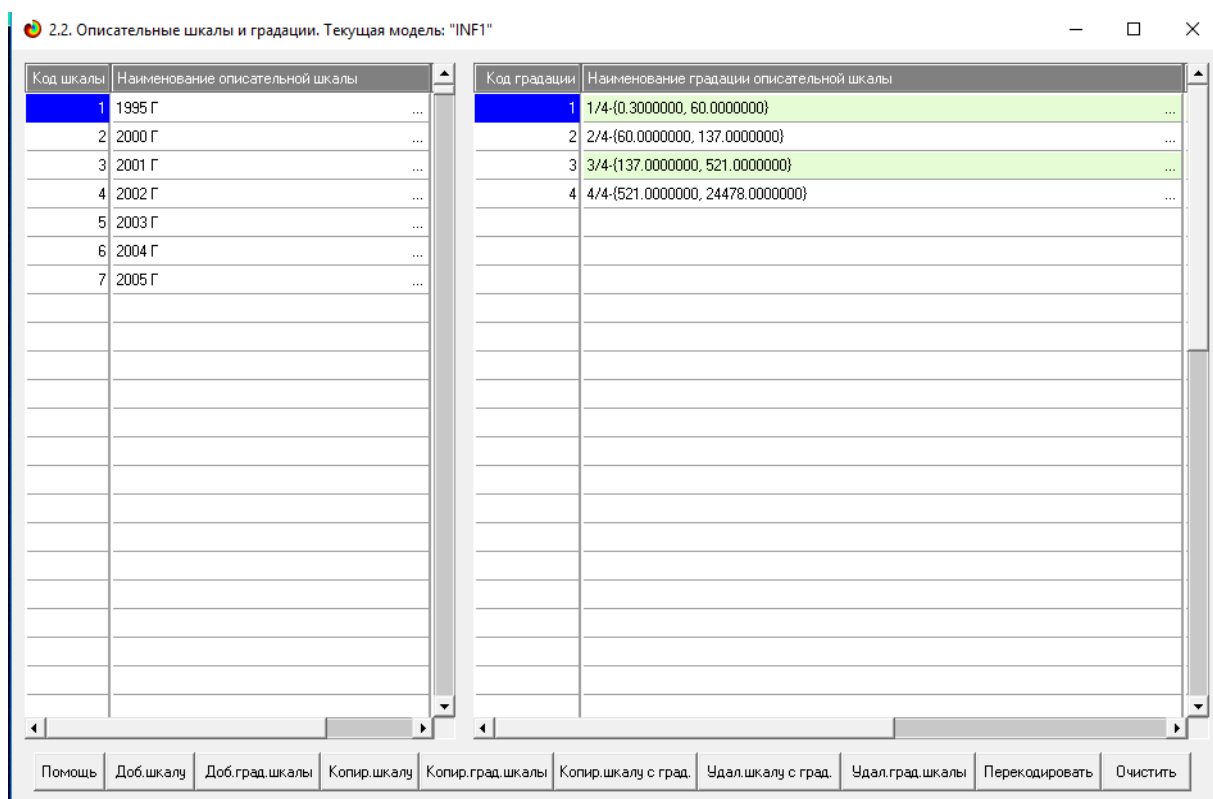


Рисунок 8 – Описательные шкалы и градации

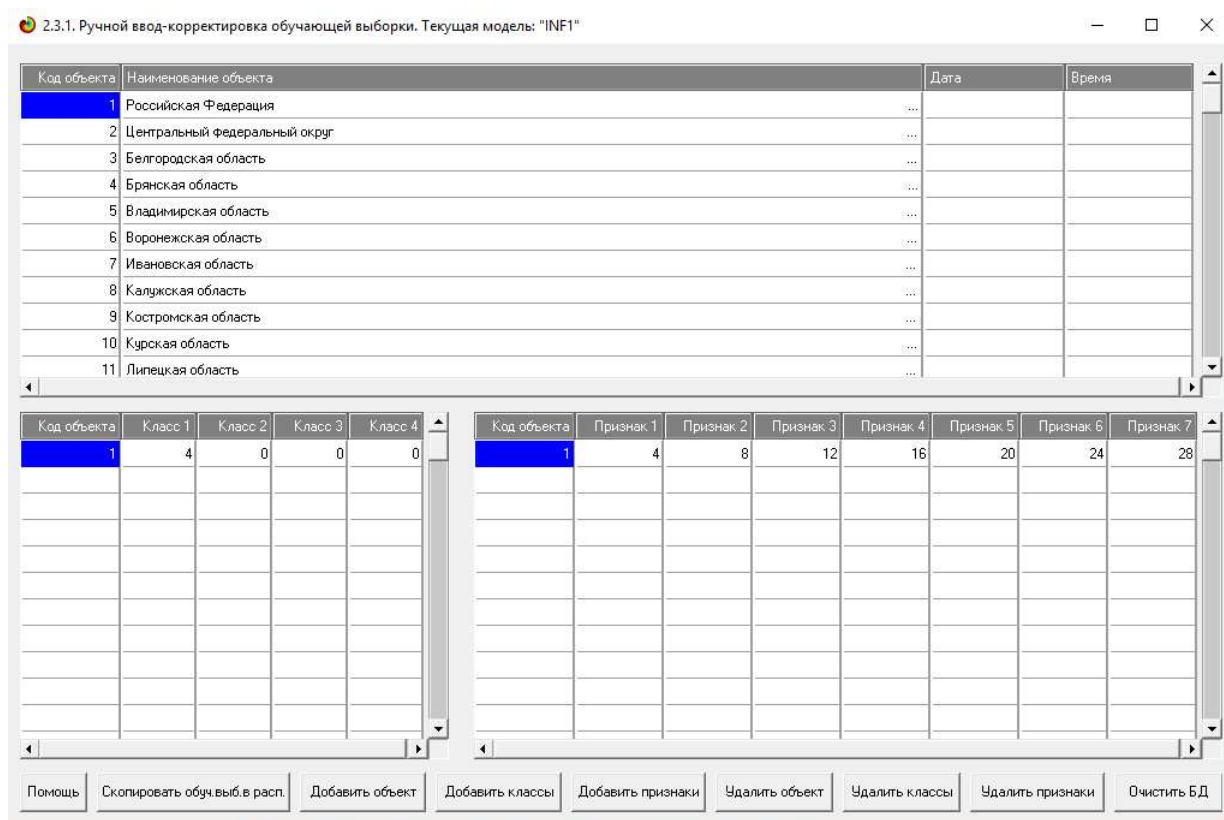


Рисунок 9 – Обучающая выборка (фрагмент)

Ниже представлены параметры классификационных и описательных шкал и градаций.

1	ПАРАМЕТРЫ ШКАЛ И ГРАДАЦИЙ С АДАПТИВНЫМИ ГРАНИЦАМИ И ПРИМЕРНО РАВНЫМ КОЛИЧЕСТВОМ НАБЛЮДЕНИЙ ПО ГРАДАЦИЯМ			
2	с коррекцией соотношения числа наблюдений по интервалу градаций при переходе к следующей градации			
3				
4	КЛАССИФИКАЦИОННАЯ ШКАЛА: код: [1], наим.: "ВСЕГО ЗА 1995-2005 ГГ.", набл.на шкалу (всего): 94, тип/число градаций в шкале: "Равное число событий в интервалах"/4			
5	1	Наим.градации: 1/4-(	3.5000000,	364.0000000), размер интервала= 360.5000000, расч./факт.число наблюдений на градацию: 23/23
6	2	Наим.градации: 2/4-(	364.0000000,	888.0000000), размер интервала= 524.0000000, расч./факт.число наблюдений на градацию: 23/23
7	3	Наим.градации: 3/4-(	888.0000000,	2755.0000000), размер интервала= 1867.0000000, расч./факт.число наблюдений на градацию: 24/24
8	4	Наим.градации: 4/4-(	2755.0000000,	139531.0000000), размер интервала=136776.0000000, расч./факт.число наблюдений на градацию: 24/24
9				
10				
11				
12	ОПИСАТЕЛЬНАЯ ШКАЛА: код: [1], наим.: "1995 Г.", набл.на шкалу (всего): 94, тип/число градаций в шкале: "Равное число событий в интервалах"/4			
13	1	Наим.градации: 1/4-(	0.3000000,	60.0000000), размер интервала= 59.7000000, расч./факт.число наблюдений на градацию: 23/23
14	2	Наим.градации: 2/4-(	60.0000000,	137.0000000), размер интервала= 77.0000000, расч./факт.число наблюдений на градацию: 23/23
15	3	Наим.градации: 3/4-(	137.0000000,	521.0000000), размер интервала= 384.0000000, расч./факт.число наблюдений на градацию: 24/24
16	4	Наим.градации: 4/4-(	521.0000000,	24478.0000000), размер интервала=23957.0000000, расч./факт.число наблюдений на градацию: 24/24
17				
18	ОПИСАТЕЛЬНАЯ ШКАЛА: код: [2], наим.: "2000 Г.", набл.на шкалу (всего): 94, тип/число градаций в шкале: "Равное число событий в интервалах"/4			
19	5	Наим.градации: 1/4-(	0.2000000,	50.0000000), размер интервала= 49.8000000, расч./факт.число наблюдений на градацию: 23/23
20	6	Наим.градации: 2/4-(	50.0000000,	132.0000000), размер интервала= 82.0000000, расч./факт.число наблюдений на градацию: 23/23
21	7	Наим.градации: 3/4-(	132.0000000,	424.0000000), размер интервала= 292.0000000, расч./факт.число наблюдений на градацию: 24/24
22	8	Наим.градации: 4/4-(	424.0000000,	20291.0000000), размер интервала=19867.0000000, расч./факт.число наблюдений на градацию: 24/24
23				
24	ОПИСАТЕЛЬНАЯ ШКАЛА: код: [3], наим.: "2001 Г.", набл.на шкалу (всего): 94, тип/число градаций в шкале: "Равное число событий в интервалах"/4			
25	9	Наим.градации: 1/4-(	0.5000000,	49.0000000), размер интервала= 48.5000000, расч./факт.число наблюдений на градацию: 23/23
26	10	Наим.градации: 2/4-(	49.0000000,	130.0000000), размер интервала= 81.0000000, расч./факт.число наблюдений на градацию: 23/23
27	11	Наим.градации: 3/4-(	130.0000000,	390.0000000), размер интервала= 260.0000000, расч./факт.число наблюдений на градацию: 24/24
28	12	Наим.градации: 4/4-(	390.0000000,	19773.0000000), размер интервала=19383.0000000, расч./факт.число наблюдений на градацию: 24/24
29				
30	ОПИСАТЕЛЬНАЯ ШКАЛА: код: [4], наим.: "2002 Г.", набл.на шкалу (всего): 93, тип/число градаций в шкале: "Равное число событий в интервалах"/4			
31	13	Наим.градации: 1/4-(	0.2000000,	54.0000000), размер интервала= 53.8000000, расч./факт.число наблюдений на градацию: 23/23
32	14	Наим.градации: 2/4-(	54.0000000,	123.0000000), размер интервала= 69.0000000, расч./факт.число наблюдений на градацию: 23/23
33	15	Наим.градации: 3/4-(	123.0000000,	398.0000000), размер интервала= 275.0000000, расч./факт.число наблюдений на градацию: 23/23
34	16	Наим.градации: 4/4-(	398.0000000,	19767.0000000), размер интервала=19369.0000000, расч./факт.число наблюдений на градацию: 24/24
35				
36	ОПИСАТЕЛЬНАЯ ШКАЛА: код: [5], наим.: "2003 Г.", набл.на шкалу (всего): 93, тип/число градаций в шкале: "Равное число событий в интервалах"/4			
37	17	Наим.градации: 1/4-(	0.6000000,	55.0000000), размер интервала= 54.4000000, расч./факт.число наблюдений на градацию: 23/23
38	18	Наим.градации: 2/4-(	55.0000000,	108.0000000), размер интервала= 80.0000000, расч./факт.число наблюдений на градацию: 23/23
39	19	Наим.градации: 3/4-(	108.0000000,	380.0000000), размер интервала= 272.0000000, расч./факт.число наблюдений на градацию: 23/23
40	20	Наим.градации: 4/4-(	380.0000000,	18961.0000000), размер интервала=18581.0000000, расч./факт.число наблюдений на градацию: 24/24
41				
42	ОПИСАТЕЛЬНАЯ ШКАЛА: код: [6], наим.: "2004 Г.", набл.на шкалу (всего): 93, тип/число градаций в шкале: "Равное число событий в интервалах"/4			
43	21	Наим.градации: 1/4-(	0.4000000,	53.0000000), размер интервала= 52.6000000, расч./факт.число наблюдений на градацию: 23/23
44	22	Наим.градации: 2/4-(	53.0000000,	105.0000000), размер интервала= 52.0000000, расч./факт.число наблюдений на градацию: 23/23
45	23	Наим.градации: 3/4-(	105.0000000,	346.0000000), размер интервала= 241.0000000, расч./факт.число наблюдений на градацию: 23/23
46	24	Наим.градации: 4/4-(	346.0000000,	18534.0000000), размер интервала=18188.0000000, расч./факт.число наблюдений на градацию: 24/24
47				
48	ОПИСАТЕЛЬНАЯ ШКАЛА: код: [7], наим.: "2005 Г.", набл.на шкалу (всего): 93, тип/число градаций в шкале: "Равное число событий в интервалах"/4			
49	25	Наим.градации: 1/4-(	0.5000000,	50.0000000), размер интервала= 49.5000000, расч./факт.число наблюдений на градацию: 23/23
50	26	Наим.градации: 2/4-(	50.0000000,	115.0000000), размер интервала= 65.0000000, расч./факт.число наблюдений на градацию: 23/23
51	27	Наим.градации: 3/4-(	115.0000000,	348.0000000), размер интервала= 233.0000000, расч./факт.число наблюдений на градацию: 23/23
52	28	Наим.градации: 4/4-(	348.0000000,	17727.0000000), размер интервала=17379.0000000, расч./факт.число наблюдений на градацию: 24/24
53				
54				

Normal text file length: 9 111 lines: 54 Ln: 11 Col: 1 Sel: 0 | 0 Windows (CR LF) OEM 862 INS

Рисунок 10 - Параметры классификационных и описательных шкал и градаций

Тем самым создаются все необходимые и достаточные предпосылки для выявления силы и направления причинно-следственных связей между значениями факторов и результатами их совместного системного воздействия (с учетом нелинейности системы).

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 11).

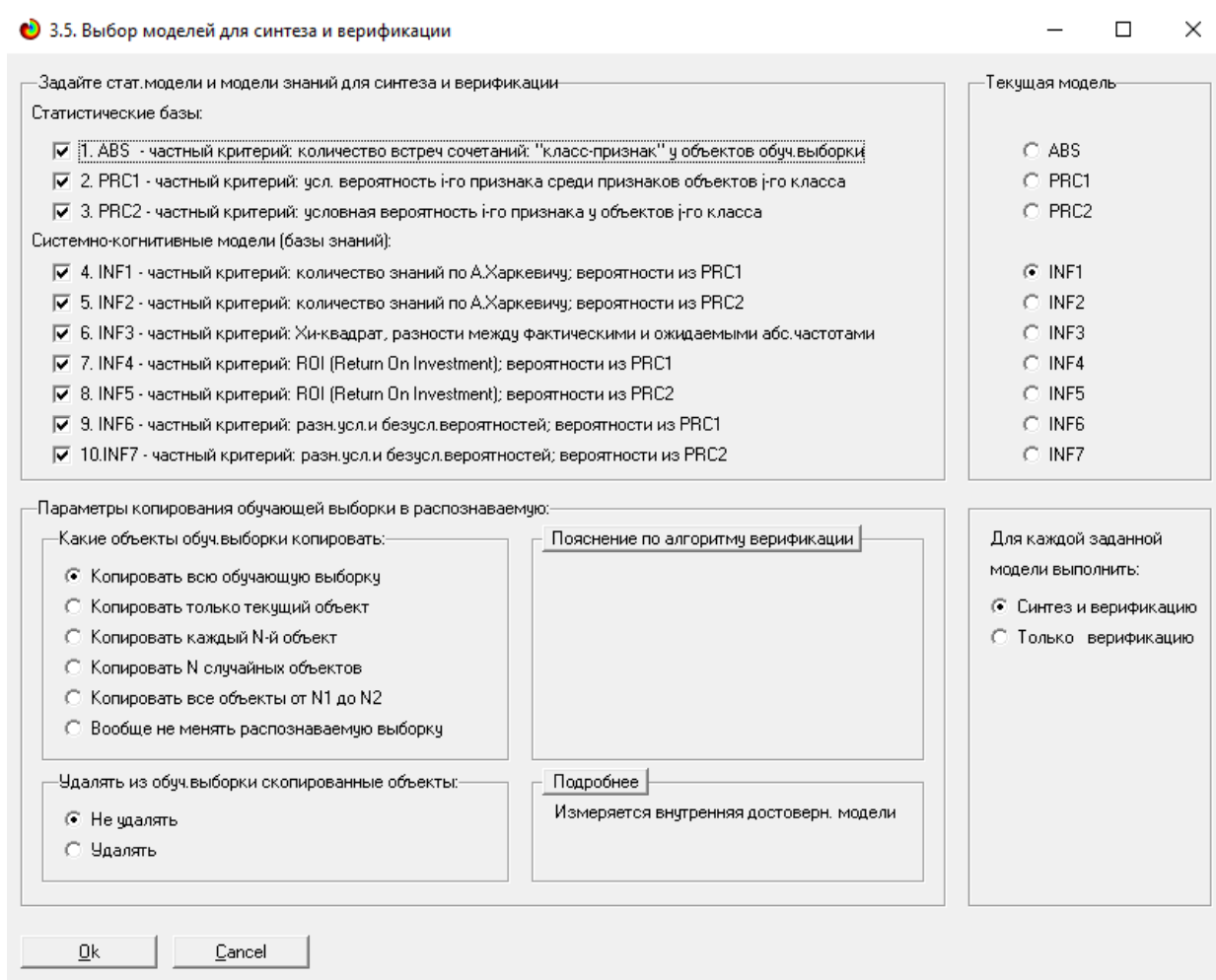


Рисунок 11 - Выбор моделей для синтеза и верификации, а также текущей модели

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 12. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 12.

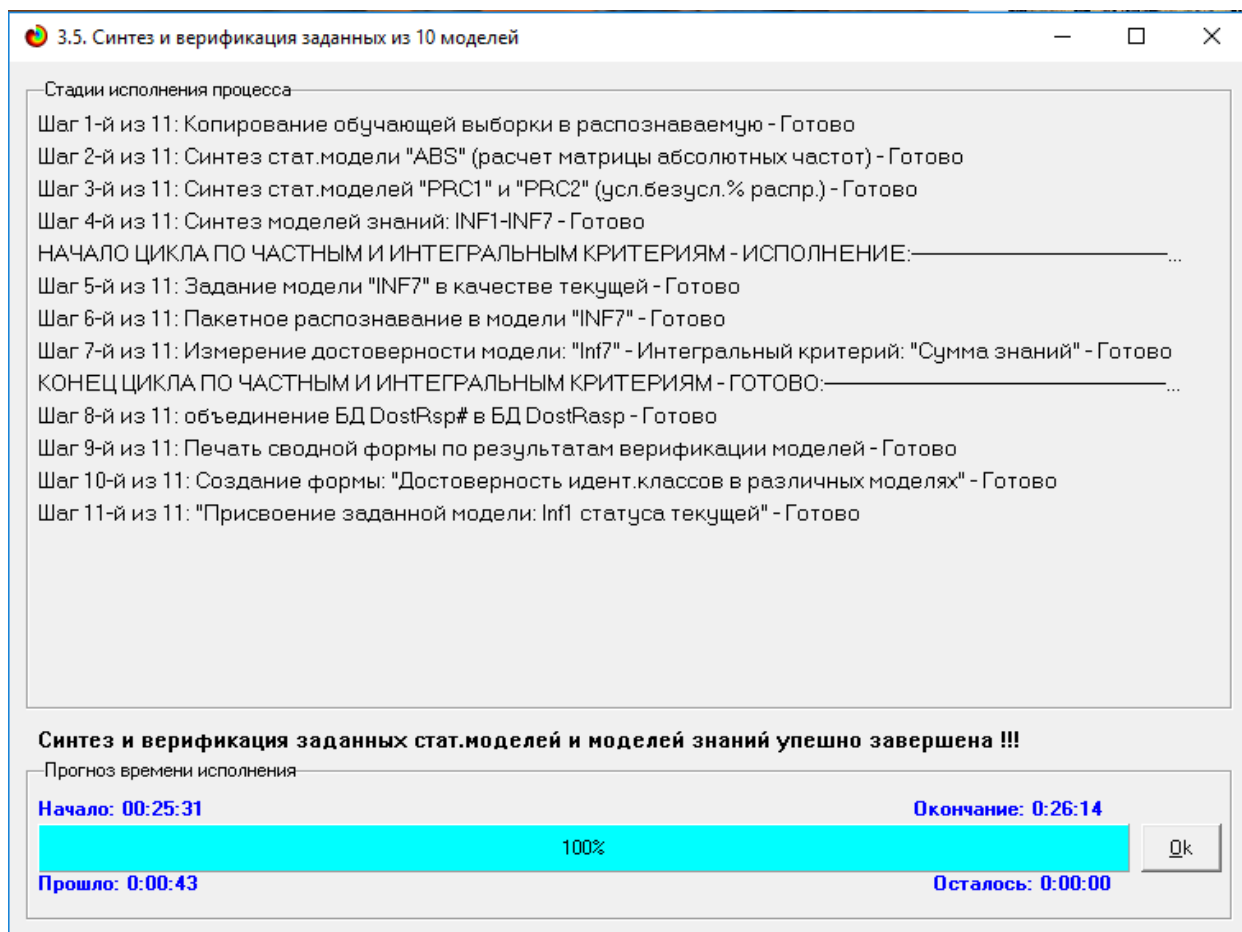


Рисунок 12 – Синтез и верификация статистических моделей и моделей знаний.

Интересно заметить (см. рисунок 10), что синтез и верификация всех 12 моделей на данной задаче занял всего 43 секунд. При этом верификация (оценка достоверности моделей) проводилась на всех 45 примерах наблюдения из обучающей выборки. В результате выполнения режима 3.5 созданы все модели, со всеми частными критериями, перечисленные на рисунке 12, но ниже мы приведем лишь некоторые из них.

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 (рисунки 13 и 14) в матрицы

условных и безусловных процентных распределений, и матрицы знаний
(рисунки 15-
18).

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. ВСЕГО ЗА 1995 2005 ГГ. 1/4 (3.5, 364.0)	2. ВСЕГО ЗА 1995 2005 ГГ. 2/4 (364.0, 888.0)	3. ВСЕГО ЗА 1995 2005 ГГ. 3/4 (888.0, 2755.0)	4. ВСЕГО ЗА 1995 2005 ГГ. 4/4 (2755.0, 139531.0)	Сумма	Среднее	Средн. кв.откл.
1	1995 Г-1/4-{0.30000000, 60.00000000}	20	3			23	5.75	9.60
2	1995 Г-2/4-{60.00000000, 137.00000000}	3	17	3		23	5.75	7.63
3	1995 Г-3/4-{137.00000000, 521.00000000}		3	19	2	24	6.00	8.76
4	1995 Г-4/4-{521.00000000, 24478.00000000}			2	22	24	6.00	10.71
5	2000 Г-1/4-{0.20000000, 50.00000000}	22	1			23	5.75	10.84
6	2000 Г-2/4-{50.00000000, 132.00000000}	1	21	1		23	5.75	10.18
7	2000 Г-3/4-{132.00000000, 424.00000000}		1	22	1	24	6.00	10.68
8	2000 Г-4/4-{424.00000000, 20291.00000000}			1	23	24	6.00	11.34
9	2001 Г-1/4-{0.50000000, 49.00000000}	22	1			23	5.75	10.84
10	2001 Г-2/4-{49.00000000, 130.00000000}	1	22	1		24	6.00	10.68
11	2001 Г-3/4-{130.00000000, 390.00000000}			22	1	23	5.75	10.84
12	2001 Г-4/4-{390.00000000, 19773.00000000}			1	23	24	6.00	11.34
13	2002 Г-1/4-{0.20000000, 54.00000000}	21	2			23	5.75	10.21
14	2002 Г-2/4-{54.00000000, 123.00000000}	1	20	2		23	5.75	9.54
15	2002 Г-3/4-{123.00000000, 398.00000000}		1	21	1	23	5.75	10.18
16	2002 Г-4/4-{398.00000000, 19767.00000000}			1	23	24	6.00	11.34
17	2003 Г-1/4-{0.60000000, 55.00000000}	21	2			23	5.75	10.21
18	2003 Г-2/4-{55.00000000, 108.00000000}	1	21	1		23	5.75	10.18
19	2003 Г-3/4-{108.00000000, 380.00000000}			22	1	23	5.75	10.84
20	2003 Г-4/4-{380.00000000, 18961.00000000}			1	23	24	6.00	11.34
21	2004 Г-1/4-{0.40000000, 53.00000000}	21	2			23	5.75	10.21

Рисунок 13 - Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений (начало)

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. ВСЕГО ЗА 1995 2005 ГГ. 1/4 (3.5, 364.0)	2. ВСЕГО ЗА 1995 2005 ГГ. 2/4 (364.0, 888.0)	3. ВСЕГО ЗА 1995 2005 ГГ. 3/4 (888.0, 2755.0)	4. ВСЕГО ЗА 1995 2005 ГГ. 4/4 (2755.0, 139531.0)	Сумма	Среднее	Средн. кв.откл.
12	2001 Г-4/4-{390.00000000, 19773.00000000}			1	23	24	6.00	11.34
13	2002 Г-1/4-{0.20000000, 54.00000000}	21	2			23	5.75	10.21
14	2002 Г-2/4-{54.00000000, 123.00000000}	1	20	2		23	5.75	9.54
15	2002 Г-3/4-{123.00000000, 398.00000000}		1	21	1	23	5.75	10.18
16	2002 Г-4/4-{398.00000000, 19767.00000000}			1	23	24	6.00	11.34
17	2003 Г-1/4-{0.60000000, 55.00000000}	21	2			23	5.75	10.21
18	2003 Г-2/4-{55.00000000, 108.00000000}	1	21	1		23	5.75	10.18
19	2003 Г-3/4-{108.00000000, 380.00000000}			22	1	23	5.75	10.84
20	2003 Г-4/4-{380.00000000, 18961.00000000}			1	23	24	6.00	11.34
21	2004 Г-1/4-{0.40000000, 53.00000000}	21	2			23	5.75	10.21
22	2004 Г-2/4-{53.00000000, 105.00000000}	1	21	1		23	5.75	10.18
23	2004 Г-3/4-{105.00000000, 346.00000000}			21	2	23	5.75	10.21
24	2004 Г-4/4-{346.00000000, 18534.00000000}			2	22	24	6.00	10.71
25	2005 Г-1/4-{0.50000000, 50.00000000}	21	2			23	5.75	10.21
26	2005 Г-2/4-{50.00000000, 115.00000000}	1	21	1		23	5.75	10.18
27	2005 Г-3/4-{115.00000000, 348.00000000}			21	2	23	5.75	10.21
28	2005 Г-4/4-{348.00000000, 17727.00000000}			2	22	24	6.00	10.71
Сумма числа признаков		157	161	168	168	654		
Среднее		6	6	6	6		5.84	
Среднеквадратичное отклонение		9	9	9	10			9.03
Сумма числа объектов обуч.выборки		23	23	24	24	94		

Рисунок 14 - Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений (продолжение)

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"								
Код признака	Наименование описательной шкалы и градации	1. ВСЕГО ЗА 1995 2005 ГГ. 1/4 (3.5, 364.0)	2. ВСЕГО ЗА 1995 2005 ГГ. 2/4 (364.0, 888.0)	3. ВСЕГО ЗА 1995 2005 ГГ. 3/4 (888.0, 2755.0)	4. ВСЕГО ЗА 1995 2005 ГГ. 4/4 (2755.0, 139531.0)	Сумма	Среднее	Средн. квадр. откл.
1	1995 Г-1/4-(0.3000000, 60.0000000)	0.397	-0.196			0.201	0.050	0.249
2	1995 Г-2/4-(60.0000000, 137.0000000)	-0.188	0.339	-0.209		-0.058	-0.015	0.254
3	1995 Г-3/4-(137.0000000, 521.0000000)		-0.209	0.347	-0.347	-0.209	-0.052	0.302
4	1995 Г-4/4-(521.0000000, 24478.0000000)			-0.347	0.392	0.045	0.011	0.302
5	2000 Г-1/4-(0.2000000, 50.0000000)	0.426	-0.535			-0.108	-0.027	0.394
6	2000 Г-2/4-(50.0000000, 132.0000000)	-0.527	0.404	-0.548		-0.671	-0.168	0.458
7	2000 Г-3/4-(132.0000000, 424.0000000)		-0.548	0.392	-0.561	-0.717	-0.179	0.462
8	2000 Г-4/4-(424.0000000, 20291.0000000)			-0.561	0.406	-0.155	-0.039	0.397
9	2001 Г-1/4-(0.5000000, 49.0000000)	0.426	-0.535			-0.108	-0.027	0.394
10	2001 Г-2/4-(49.0000000, 130.0000000)	-0.540	0.406	-0.561		-0.696	-0.174	0.466
11	2001 Г-3/4-(130.0000000, 390.0000000)			0.406	-0.548	-0.142	-0.036	0.391
12	2001 Г-4/4-(390.0000000, 19773.0000000)			-0.561	0.406	-0.155	-0.039	0.397
13	2002 Г-1/4-(0.2000000, 54.0000000)	0.412	-0.321			0.091	0.023	0.300
14	2002 Г-2/4-(54.0000000, 123.0000000)	-0.527	0.389	-0.334		-0.472	-0.118	0.402
15	2002 Г-3/4-(123.0000000, 398.0000000)		-0.535	0.391	-0.548	-0.692	-0.173	0.455
16	2002 Г-4/4-(398.0000000, 19767.0000000)			-0.561	0.406	-0.155	-0.039	0.397
17	2003 Г-1/4-(0.6000000, 55.0000000)	0.412	-0.321			0.091	0.023	0.300
18	2003 Г-2/4-(55.0000000, 108.0000000)	-0.527	0.404	-0.548		-0.671	-0.168	0.458
19	2003 Г-3/4-(108.0000000, 380.0000000)			0.406	-0.548	-0.142	-0.036	0.391
20	2003 Г-4/4-(380.0000000, 18961.0000000)			-0.561	0.406	-0.155	-0.039	0.397
21	2004 Г-1/4-(0.4000000, 53.0000000)	0.412	-0.321			0.091	0.023	0.300

Рисунок 15 - Матрица информативностей (модель INF1) в битах (начало)

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"								
Код признака	Наименование описательной шкалы и градации	1. ВСЕГО ЗА 1995 2005 ГГ. 1/4 (3.5, 364.0)	2. ВСЕГО ЗА 1995 2005 ГГ. 2/4 (364.0, 888.0)	3. ВСЕГО ЗА 1995 2005 ГГ. 3/4 (888.0, 2755.0)	4. ВСЕГО ЗА 1995 2005 ГГ. 4/4 (2755.0, 139531.0)	Сумма	Среднее	Средн. квадр. откл.
11	2001 Г-3/4-(130.0000000, 390.0000000)			0.406	-0.548	-0.142	-0.036	0.391
12	2001 Г-4/4-(390.0000000, 19773.0000000)			-0.561	0.406	-0.155	-0.039	0.397
13	2002 Г-1/4-(0.2000000, 54.0000000)	0.412	-0.321			0.091	0.023	0.300
14	2002 Г-2/4-(54.0000000, 123.0000000)	-0.527	0.389	-0.334		-0.472	-0.118	0.402
15	2002 Г-3/4-(123.0000000, 398.0000000)		-0.535	0.391	-0.548	-0.692	-0.173	0.455
16	2002 Г-4/4-(398.0000000, 19767.0000000)			-0.561	0.406	-0.155	-0.039	0.397
17	2003 Г-1/4-(0.6000000, 55.0000000)	0.412	-0.321			0.091	0.023	0.300
18	2003 Г-2/4-(55.0000000, 108.0000000)	-0.527	0.404	-0.548		-0.671	-0.168	0.458
19	2003 Г-3/4-(108.0000000, 380.0000000)			0.406	-0.548	-0.142	-0.036	0.391
20	2003 Г-4/4-(380.0000000, 18961.0000000)			-0.561	0.406	-0.155	-0.039	0.397
21	2004 Г-1/4-(0.4000000, 53.0000000)	0.412	-0.321			0.091	0.023	0.300
22	2004 Г-2/4-(53.0000000, 105.0000000)	-0.527	0.404	-0.548		-0.671	-0.168	0.458
23	2004 Г-3/4-(105.0000000, 346.0000000)			0.391	-0.334	0.057	0.014	0.297
24	2004 Г-4/4-(346.0000000, 18534.0000000)			-0.347	0.392	0.045	0.011	0.302
25	2005 Г-1/4-(0.5000000, 50.0000000)	0.412	-0.321			0.091	0.023	0.300
26	2005 Г-2/4-(50.0000000, 115.0000000)	-0.527	0.404	-0.548		-0.671	-0.168	0.458
27	2005 Г-3/4-(115.0000000, 348.0000000)			0.391	-0.334	0.057	0.014	0.297
28	2005 Г-4/4-(348.0000000, 17727.0000000)			-0.347	0.392	0.045	0.011	0.302
Сумма		-0.465	-1.090	-3.858	-0.419	-5.833		
Среднее		-0.017	-0.039	-0.138	-0.015		-0.052	
Среднеквадратичное отклонение		0.328	0.316	0.376	0.315			0.334

Рисунок 16 - Матрица информативностей (модель INF1) в битах
(продолжение)

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

Код признака	Наименование описательной шкалы и градации	1. ВСЕГО за 1995 2005 гг. 1/4 (35, 364.0)	2. ВСЕГО за 1995 2005 гг. 2/4 (364.0, 888.0)	3. ВСЕГО за 1995 2005 гг. 3/4 (888.0, 2755.0)	4. ВСЕГО за 1995 2005 гг. 4/4 (2755.0, 139531.0)	Сумма	Среднее	Средн. кв.откл.
1	1995 Г-1/4-(0.3000000, 60.0000000)	14.479	-2.662	-5.908	-5.908			9.773
2	1995 Г-2/4-(60.0000000, 137.0000000)	-2.521	11.338	-2.908	-5.908			7.709
3	1995 Г-3/4-(137.0000000, 521.0000000)	-5.761	-2.908	12.835	-4.165			8.636
4	1995 Г-4/4-(521.0000000, 24478.0000000)	-5.761	-5.908	-4.165	15.835			10.586
5	2000 Г-1/4-(0.2000000, 50.0000000)	16.479	-4.662	-5.908	-5.908			11.001
6	2000 Г-2/4-(50.0000000, 132.0000000)	-4.521	15.338	-4.908	-5.908			10.242
7	2000 Г-3/4-(132.0000000, 424.0000000)	-5.761	-4.908	15.835	-5.165			10.563
8	2000 Г-4/4-(424.0000000, 20291.0000000)	-5.761	-5.908	-5.165	16.835			11.228
9	2001 Г-1/4-(0.5000000, 49.0000000)	16.479	-4.662	-5.908	-5.908			11.001
10	2001 Г-2/4-(49.0000000, 130.0000000)	-4.761	16.092	-5.165	-6.165			10.744
11	2001 Г-3/4-(130.0000000, 390.0000000)	-5.521	-5.662	16.092	-4.908			10.733
12	2001 Г-4/4-(390.0000000, 19773.0000000)	-5.761	-5.908	-5.165	16.835			11.228
13	2002 Г-1/4-(0.2000000, 54.0000000)	15.479	-3.662	-5.908	-5.908			10.373
14	2002 Г-2/4-(54.0000000, 123.0000000)	-4.521	14.338	-3.908	-5.908			9.595
15	2002 Г-3/4-(123.0000000, 398.0000000)	-5.521	-4.662	15.092	-4.908			10.068
16	2002 Г-4/4-(398.0000000, 19767.0000000)	-5.761	-5.908	-5.165	16.835			11.228
17	2003 Г-1/4-(0.6000000, 55.0000000)	15.479	-3.662	-5.908	-5.908			10.373
18	2003 Г-2/4-(55.0000000, 108.0000000)	-4.521	15.338	-4.908	-5.908			10.242
19	2003 Г-3/4-(108.0000000, 380.0000000)	-5.521	-5.662	16.092	-4.908			10.733
20	2003 Г-4/4-(380.0000000, 18961.0000000)	-5.761	-5.908	-5.165	16.835			11.228
21	2004 Г-1/4-(0.4000000, 53.0000000)	15.479	-3.662	-5.908	-5.908			10.373

Рисунок 17 – Матрица знаний (модель INF3) (начало)

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

Код признака	Наименование описательной шкалы и градации	1. ВСЕГО за 1995 2005 гг. 1/4 (35, 364.0)	2. ВСЕГО за 1995 2005 гг. 2/4 (364.0, 888.0)	3. ВСЕГО за 1995 2005 гг. 3/4 (888.0, 2755.0)	4. ВСЕГО за 1995 2005 гг. 4/4 (2755.0, 139531.0)	Сумма	Среднее	Средн. кв.откл.
11	2001 Г-3/4-(130.0000000, 390.0000000)	-5.521	-5.662	16.092	-4.908			10.733
12	2001 Г-4/4-(390.0000000, 19773.0000000)	-5.761	-5.908	-5.165	16.835			11.228
13	2002 Г-1/4-(0.2000000, 54.0000000)	15.479	-3.662	-5.908	-5.908			10.373
14	2002 Г-2/4-(54.0000000, 123.0000000)	-4.521	14.338	-3.908	-5.908			9.595
15	2002 Г-3/4-(123.0000000, 398.0000000)	-5.521	-4.662	15.092	-4.908			10.068
16	2002 Г-4/4-(398.0000000, 19767.0000000)	-5.761	-5.908	-5.165	16.835			11.228
17	2003 Г-1/4-(0.6000000, 55.0000000)	15.479	-3.662	-5.908	-5.908			10.373
18	2003 Г-2/4-(55.0000000, 108.0000000)	-4.521	15.338	-4.908	-5.908			10.242
19	2003 Г-3/4-(108.0000000, 380.0000000)	-5.521	-5.662	16.092	-4.908			10.733
20	2003 Г-4/4-(380.0000000, 18961.0000000)	-5.761	-5.908	-5.165	16.835			11.228
21	2004 Г-1/4-(0.4000000, 53.0000000)	15.479	-3.662	-5.908	-5.908			10.373
22	2004 Г-2/4-(53.0000000, 105.0000000)	-4.521	15.338	-4.908	-5.908			10.242
23	2004 Г-3/4-(105.0000000, 346.0000000)	-5.521	-5.662	15.092	-3.908			10.093
24	2004 Г-4/4-(346.0000000, 18534.0000000)	-5.761	-5.908	-4.165	15.835			10.586
25	2005 Г-1/4-(0.5000000, 50.0000000)	15.479	-3.662	-5.908	-5.908			10.373
26	2005 Г-2/4-(50.0000000, 115.0000000)	-4.521	15.338	-4.908	-5.908			10.242
27	2005 Г-3/4-(115.0000000, 348.0000000)	-5.521	-5.662	15.092	-3.908			10.093
28	2005 Г-4/4-(348.0000000, 17727.0000000)	-5.761	-5.908	-4.165	15.835			10.586
Сумма								
Среднее								
Среднеквадратичное отклонение		9.217	8.746	8.958	9.670			9.029

Рисунок 18 – Матрица знаний (модель INF3) (продолжение)

1.5 Результаты верификации моделей

Результаты верификации (оценки достоверности) моделей, отличающихся частными критериями с двумя приведенными ниже интегральными критериями, приведены на рисунке 19.

4.1.3.6. Обобщенная форма для достов.моделей при разн.крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризбергера	Сумма модулей уровней сходства истинно-положительных решений (STP)	Сумма модулей уровней сходства истинно-отрицательных решений (STN)	Сумма модулей уровней сходства ложно-положительных решений (SFP)	Сумма модулей уровней сходства ложно-отрицательных решений (SFN)
1. ABS - частный критерий: количество встреч сочетаний: "клас..."	Корреляция абс частот с обр...	94	93	268	14	1	0.869	0.989	0.925	83.466	87.704	4.323	
1. ABS - частный критерий: количество встреч сочетаний: "клас..."	Сумма абс частот по признакам...	94	94	127	155		0.378	1.000	0.548	81.494		15.228	
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн.частот с о...	94	93	268	14	1	0.869	0.989	0.925	83.466	87.704	4.323	
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн.частот по признакам...	94	94	127	155		0.378	1.000	0.548	83.513		15.605	
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн.частот с о...	94	93	268	14	1	0.869	0.989	0.925	83.465	87.704	4.323	
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн.частот по признакам...	94	94	127	155		0.378	1.000	0.548	83.178		15.563	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	94	91	145	137	3	0.399	0.968	0.565	77.482	88.427	12.532	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	94	89	277	5	5	0.947	0.947	0.947	66.421	106.650	1.795	
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	94	91	145	137	3	0.399	0.968	0.565	77.058	88.019	12.551	
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	94	89	277	5	5	0.947	0.947	0.947	65.618	105.266	1.790	
6. INF3 - частный критерий: Хинквардт, разности между факти...	Семантический резонанс зна...	94	93	268	14	1	0.869	0.989	0.925	83.454	87.742	4.329	
6. INF3 - частный критерий: Хинквардт, разности между факти...	Сумма знаний	94	93	268	14	1	0.869	0.989	0.925	78.915	82.870	4.059	
7. INF4 - частный критерий: ROI (Return On Investment), веро...	Семантический резонанс зна...	94	91	272	10	3	0.901	0.968	0.933	82.193	84.881	3.233	
7. INF4 - частный критерий: ROI (Return On Investment), веро...	Сумма знаний	94	93	265	17	1	0.845	0.989	0.912	78.350	28.347	4.382	
8. INF5 - частный критерий: ROI (Return On Investment), веро...	Семантический резонанс зна...	94	91	272	10	3	0.901	0.968	0.933	82.285	84.962	3.229	
8. INF5 - частный критерий: ROI (Return On Investment), веро...	Сумма знаний	94	93	265	17	1	0.845	0.989	0.912	80.363	29.065	4.498	
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Семантический резонанс зна...	94	91	272	10	3	0.901	0.968	0.933	82.208	84.905	3.231	
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Сумма знаний	94	93	265	17	1	0.845	0.989	0.912	79.582	28.788	4.436	
10. INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Семантический резонанс зна...	94	91	272	10	3	0.901	0.968	0.933	82.168	84.848	3.222	
10. INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Сумма знаний	94	93	265	17	1	0.845	0.989	0.912	80.378	29.065	4.484	

Рисунок 19 - Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели INF1 при интегральном критерии «Сумма знаний». При этом точность модели составляет 0,947, при полноте 0,947.

Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется F-критерий Ван Ризбергера, а также его нечеткое мультиклассовое обобщение, предложенное проф.Е.В.Луценко (рисунок 20).

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

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

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

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В.Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {-1, +1} и {0, 1}:

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

$$L2 = \frac{1 + (TP + TN - FP - FN)}{(TP + TN + FP + FN)} \quad \text{(нормировка: \{0, 1\})}$$
где количество: TP - истинно-положительных решений; TN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;

классическая F-мера достоверности моделей Ван Ризбергера (колонка выделена ярко-голубым фоном):

$$F\text{-мера} = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$
Precision = TP / (TP + FP) - точность модели;
Recall = TP / (TP + FN) - полнота модели;

L1-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом сумм уровней сходства (колонка выделена ярко-зеленым фоном):

$$L1\text{-мера} = \frac{2 * (SPrecision * SRecall)}{(SPrecision + SRecall)}$$
SPrecision = STP / (STP + SFP) - точность с учетом сумм уровней сходства;
SRecall = STP / (STP + SFN) - полнота с учетом сумм уровней сходства;
STP - сумма модулей сходства истинно-положительных решений; STN - сумма модулей сходства истинно-отрицательных решений;
SFP - сумма модулей сходства ложно-положительных решений; SFN - сумма модулей сходства ложно-отрицательных решений.

L2-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желтым фоном):

$$L2\text{-мера} = \frac{2 * (APrecision * ARecall)}{(APrecision + ARecall)}$$
APrecision = ATP / (ATP + AFP) - точность с учетом средних уровней сходства;
ARecall = ATP / (ATP + AFN) - полнота с учетом средних уровней сходства;
ATP = STP / TP - среднее модулей сходства истинно-положительных решений; AFN = SFN / FN - среднее модулей сходства ложно-отрицательных решений;
AFP = SFP / FP - среднее модулей сходства ложно-положительных решений; AFN = SFN / FN - среднее модулей сходства ложно-отрицательных решений.

Строки с максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке.

Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергера в АСК-анализе и системе "Эйдос" // Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №02(126). - С. 1 - 32. - IDA [article ID]: 1261702001. - Режим доступа: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 у.п.н.л.

Рисунок 20 - Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием.

Также обращает на себя внимание, что статистические модели, как правило, дают значительно 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 14 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и неидентифицированных ситуаций в наиболее достоверной модели INF1.

Из рисунка 21 видно, что:

- Можно выделить 3 интервала;
- На интервале от -100% до -83% представлены только достоверные решения;
- С -83% и до 80% представлены ошибочные и достоверные решения
- На промежутке 80-94% представлены только достоверные решения

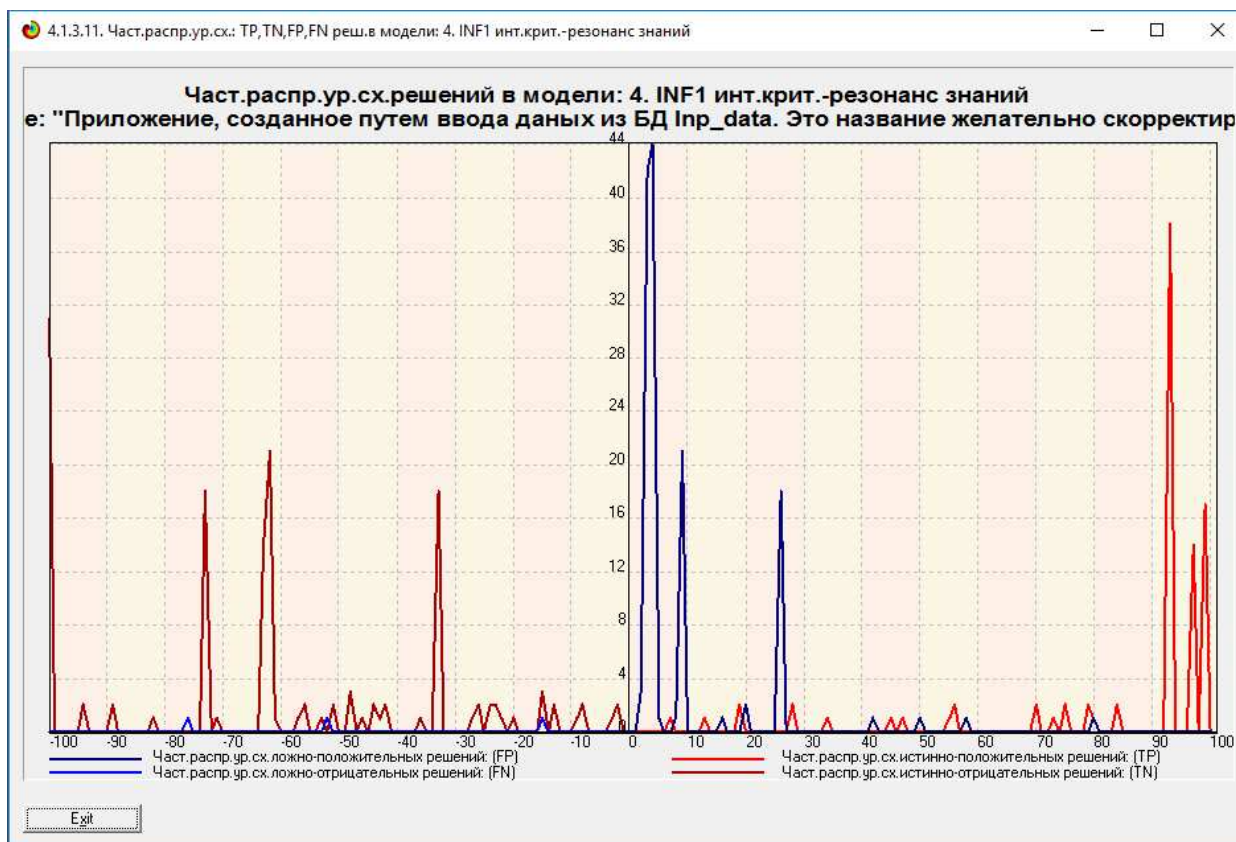
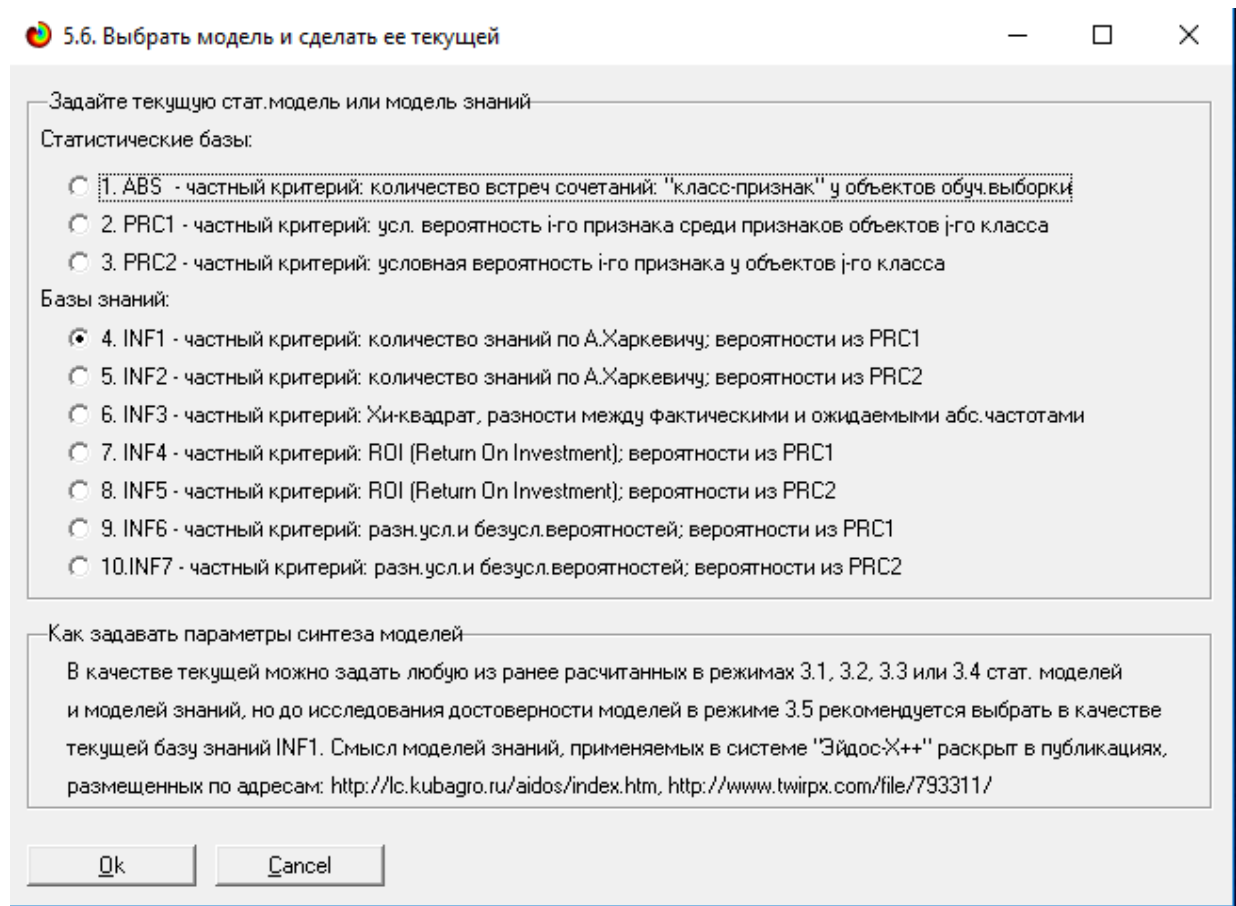


Рисунок 21 - Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF1

2 РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1 Решение задачи идентификации

В соответствии с технологией АСК-анализа зададим текущей модели INF1 (режим 5.6) (рисунок 22).



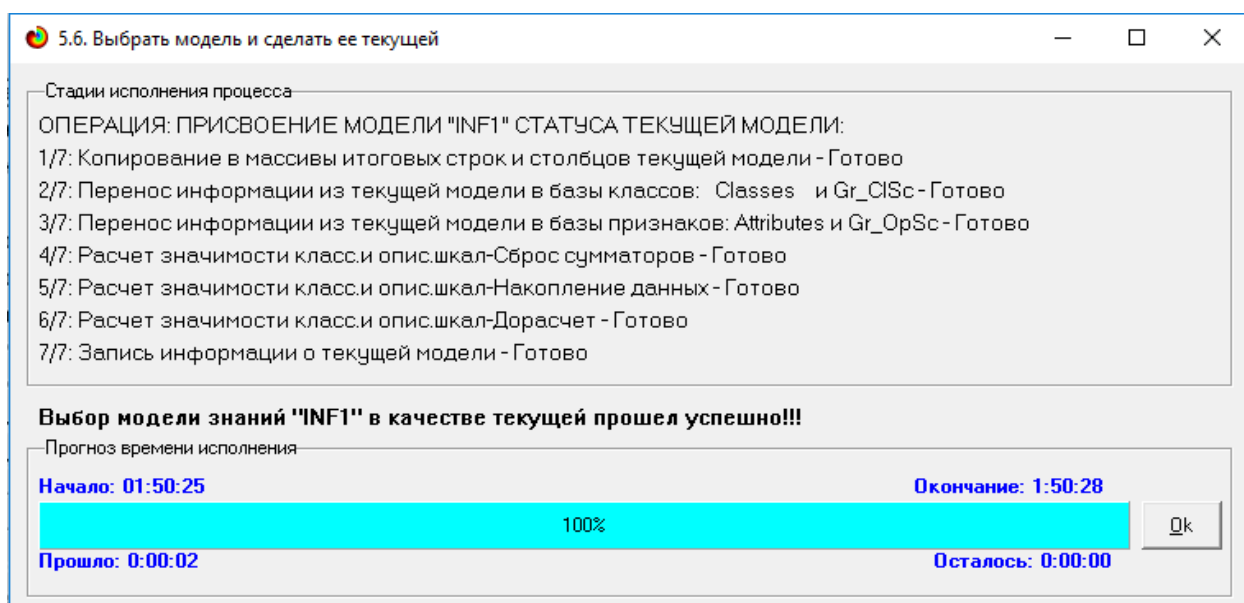


Рисунок 22 – Экранные формы режима задания модели в качестве текущей

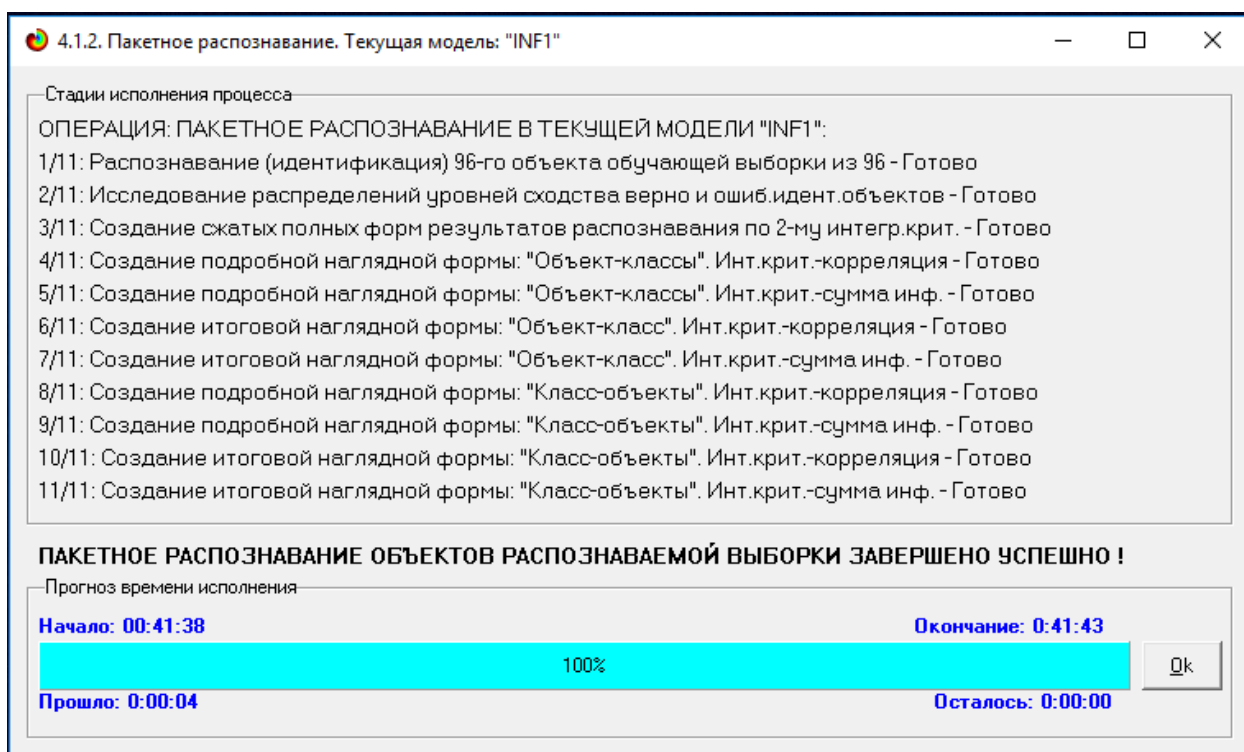


Рисунок 23 – Экранная форма режима пакетного распознавания в текущей модели INF1

В результате пакетного распознавания в текущей модели создается ряд баз данных, которые визуализируются в выходных экранных формах, отражающих результаты решения задачи идентификации и прогнозирования. Режим 4.1.3 системы «Эйдос» обеспечивает отображение результатов идентификации и прогнозирования в различных формах:

1. Подробно наглядно: "Объект – классы".

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

Ниже кратко рассмотрим некоторые из них.

На рисунках 24 и 25 приведены примеры прогнозов в наиболее достоверной модели

INF1:

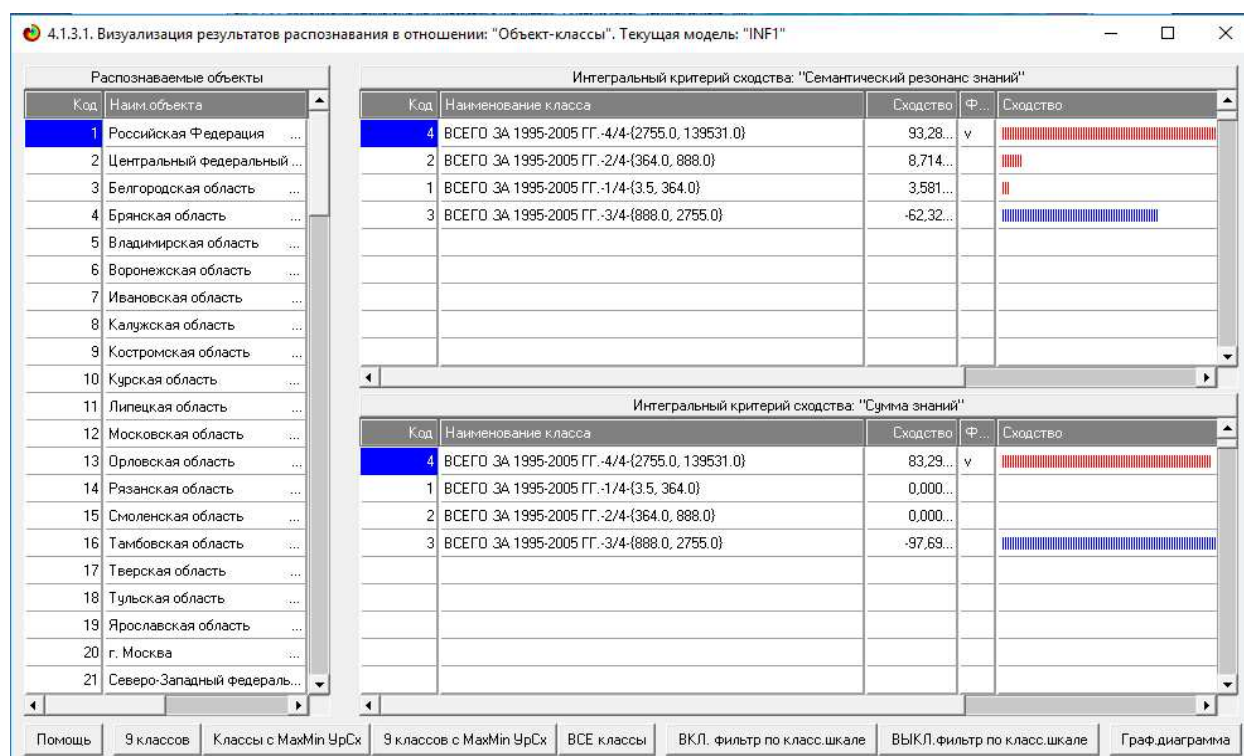


Рисунок 24 – Пример идентификации классов в модели INF1

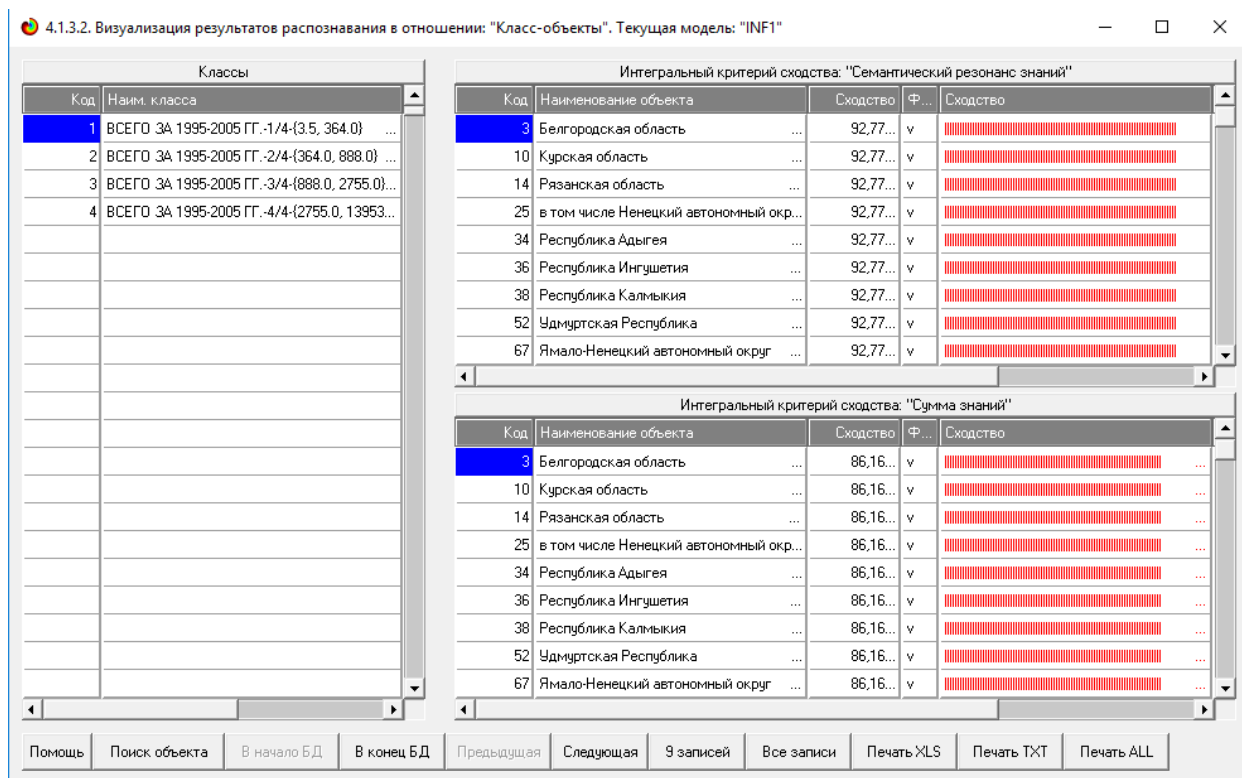


Рисунок 24 – Пример идентификации классов в модели INF1

2.2 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 OS Windows: «Эйдос-X++» предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (Рисунок 25)

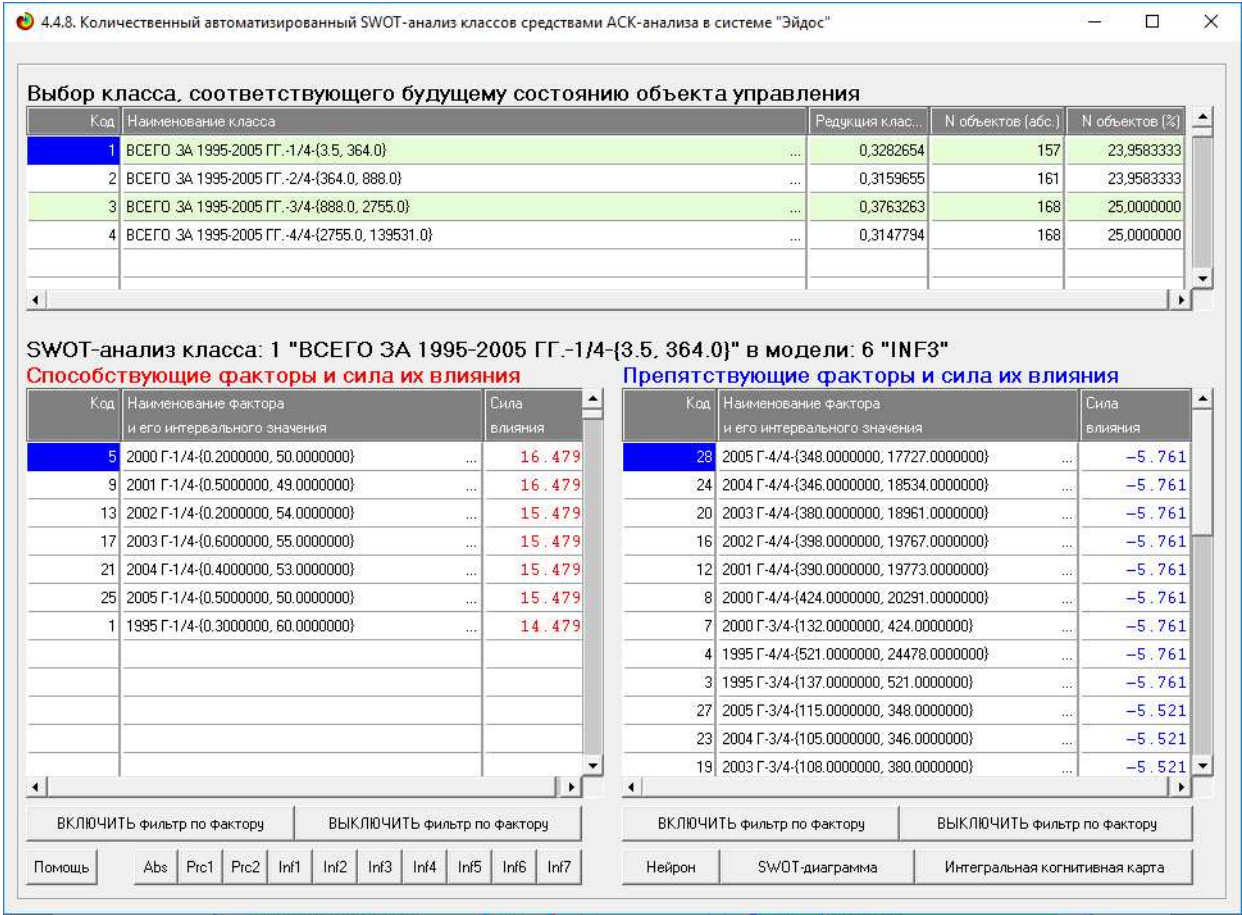


Рисунок 25 – Пример SWOT –матрицы в модели INF1

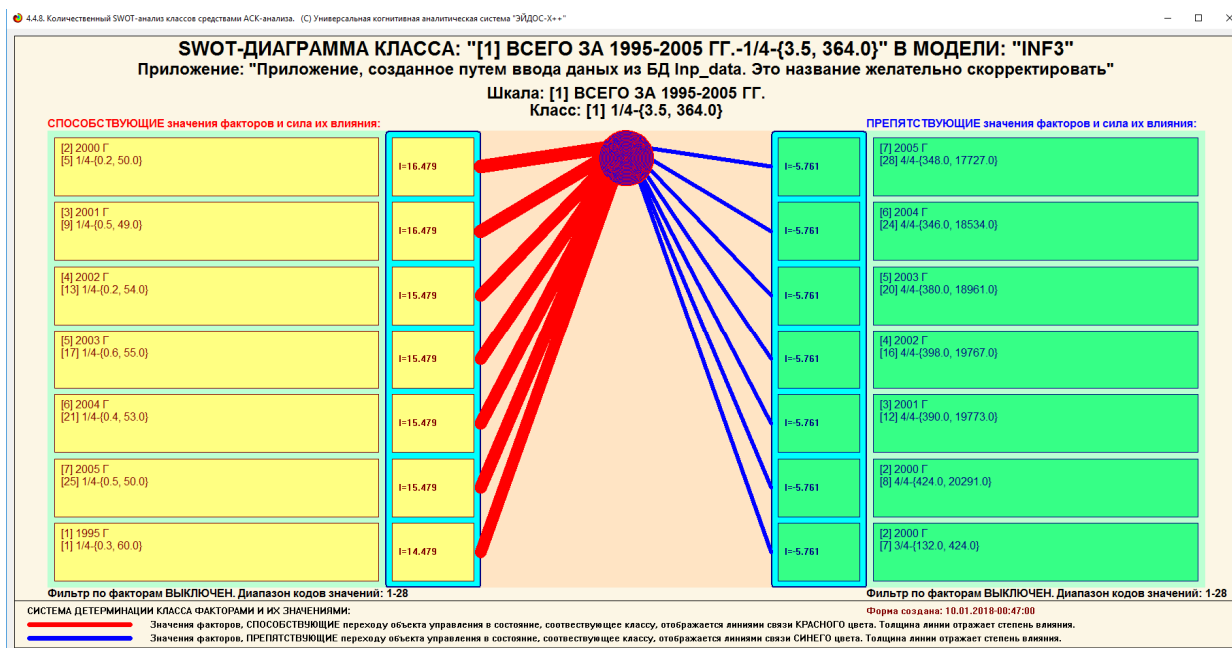


Рисунок 26 – Пример SWOT –матрицы в модели INF1

На рисунке 27, 28 приведены примеры инвертированной SWOT-матрицы и инвертированной SWOT-диаграммы в модели INF1.

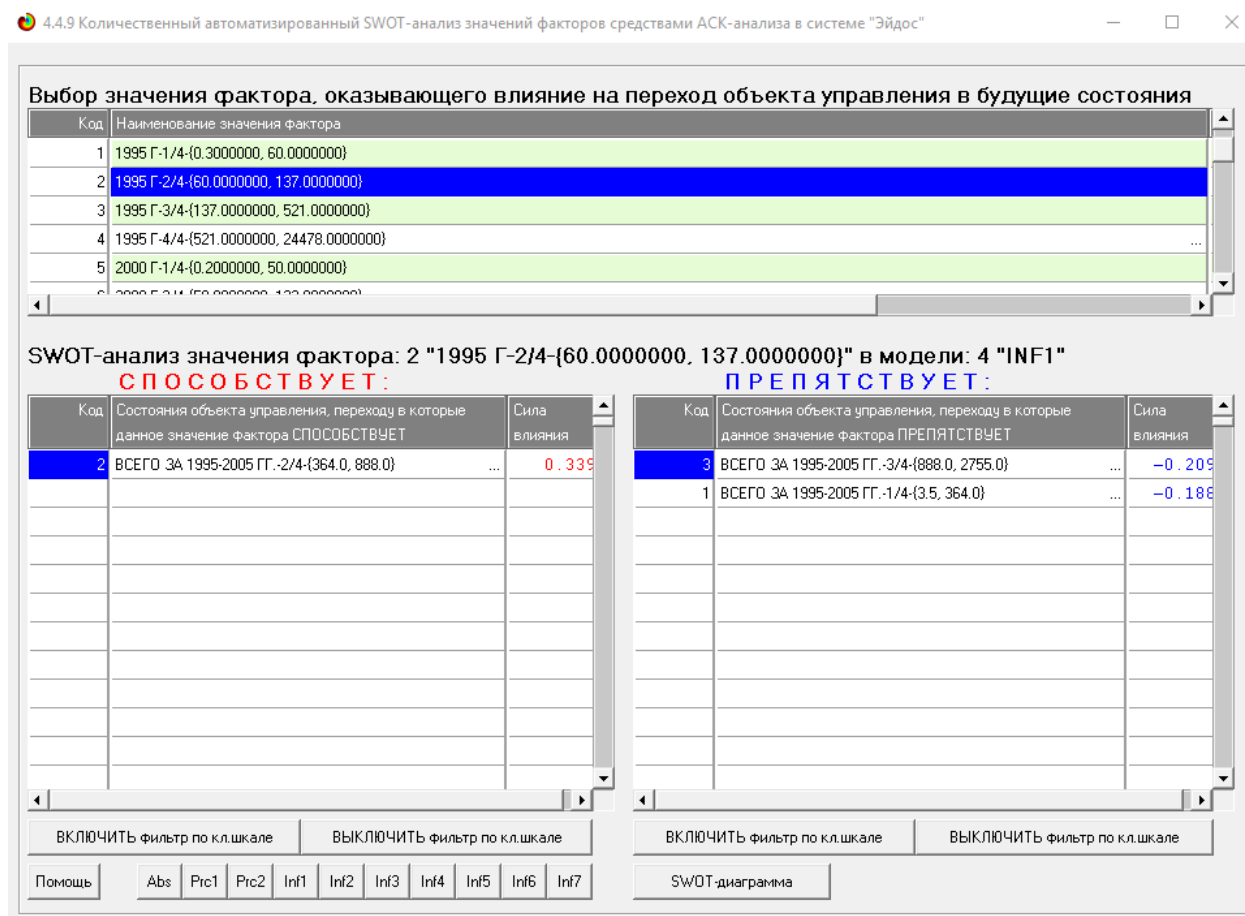


Рисунок 27 – Пример Swot-матрицы в модели

INF1

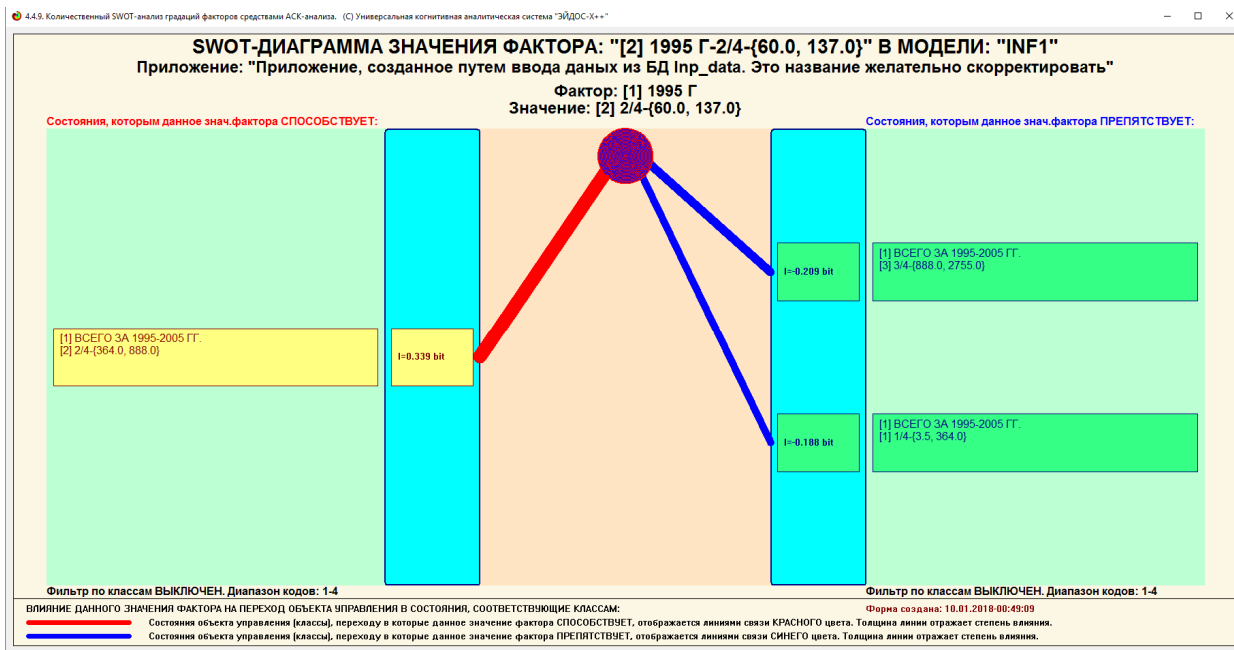


Рисунок 29 – Пример Swot-матрицы в модели INF1

2.3 Кластерно-конструктивный анализ признаков

На рисунке 30,31 приведены результаты кластерно-конструктивного анализа признаков

Конструкт признака: 1 "1995 Г-1/4-{0.3, 60.0}" в модели: 4 "INF1"

Код	Наименование признака	№	Код призна...	Наименование признака	Сходство
1	1995 Г-1/4-{0.3000000, 60.0000000}	1	1	1995 Г-1/4-{0.3000000, 60.0000000}	100.000
2	1995 Г-2/4-{60.0000000, 137.0000000}	2	13	2002 Г-1/4-{0.2000000, 54.0000000}	98.186
3	1995 Г-3/4-{137.0000000, 521.0000000}	3	17	2003 Г-1/4-{0.6000000, 55.0000000}	98.186
4	1995 Г-4/4-{521.0000000, 24478.0000000}	4	21	2004 Г-1/4-{0.4000000, 53.0000000}	98.186
5	2000 Г-1/4-{0.2000000, 50.0000000}	5	25	2005 Г-1/4-{0.5000000, 50.0000000}	98.186
6	2000 Г-2/4-{50.0000000, 132.0000000}	6	5	2000 Г-1/4-{0.2000000, 50.0000000}	91.690
7	2000 Г-3/4-{132.0000000, 424.0000000}	7	9	2001 Г-1/4-{0.5000000, 49.0000000}	91.690
8	2000 Г-4/4-{424.0000000, 20291.0000000}	8	7	2000 Г-3/4-{132.0000000, 424.0000000}	31.016
9	2001 Г-1/4-{0.5000000, 49.0000000}	9	15	2002 Г-3/4-{123.0000000, 398.0000000}	30.825
10	2001 Г-2/4-{49.0000000, 130.0000000}	10	3	1995 Г-3/4-{137.0000000, 521.0000000}	18.739
11	2001 Г-3/4-{130.0000000, 390.0000000}	11	8	2000 Г-4/4-{424.0000000, 20291.0000000}	0.680
12	2001 Г-4/4-{390.0000000, 19773.0000000}	12	12	2001 Г-4/4-{390.0000000, 19773.0000000}	0.680
13	2002 Г-1/4-{0.2000000, 54.0000000}	13	16	2002 Г-4/4-{398.0000000, 19767.0000000}	0.680
14	2002 Г-2/4-{54.0000000, 123.0000000}	14	20	2003 Г-4/4-{380.0000000, 18961.0000000}	0.680
15	2002 Г-3/4-{123.0000000, 398.0000000}	15	11	2001 Г-3/4-{130.0000000, 390.0000000}	0.635
16	2002 Г-4/4-{398.0000000, 19767.0000000}	16	19	2003 Г-3/4-{108.0000000, 380.0000000}	0.635
17	2003 Г-1/4-{0.6000000, 55.0000000}	17	4	1995 Г-4/4-{521.0000000, 24478.0000000}	-0.261
18	2003 Г-2/4-{55.0000000, 108.0000000}	18	24	2004 Г-4/4-{346.0000000, 18534.0000000}	-0.261
19	2003 Г-3/4-{108.0000000, 380.0000000}	19	28	2005 Г-4/4-{348.0000000, 17727.0000000}	-0.261
20	2003 Г-4/4-{380.0000000, 18961.0000000}	20	23	2004 Г-3/4-{105.0000000, 346.0000000}	-0.337
21	2004 Г-1/4-{0.4000000, 53.0000000}	21	27	2005 Г-3/4-{115.0000000, 348.0000000}	-0.337
22	2004 Г-2/4-{53.0000000, 105.0000000}	22	2	1995 Г-2/4-{60.0000000, 137.0000000}	-72.504
23	2004 Г-3/4-{105.0000000, 346.0000000}	23	10	2001 Г-2/4-{49.0000000, 130.0000000}	-75.307

Рисунок 30 – Результаты кластерно-конструктивного анализа признаков

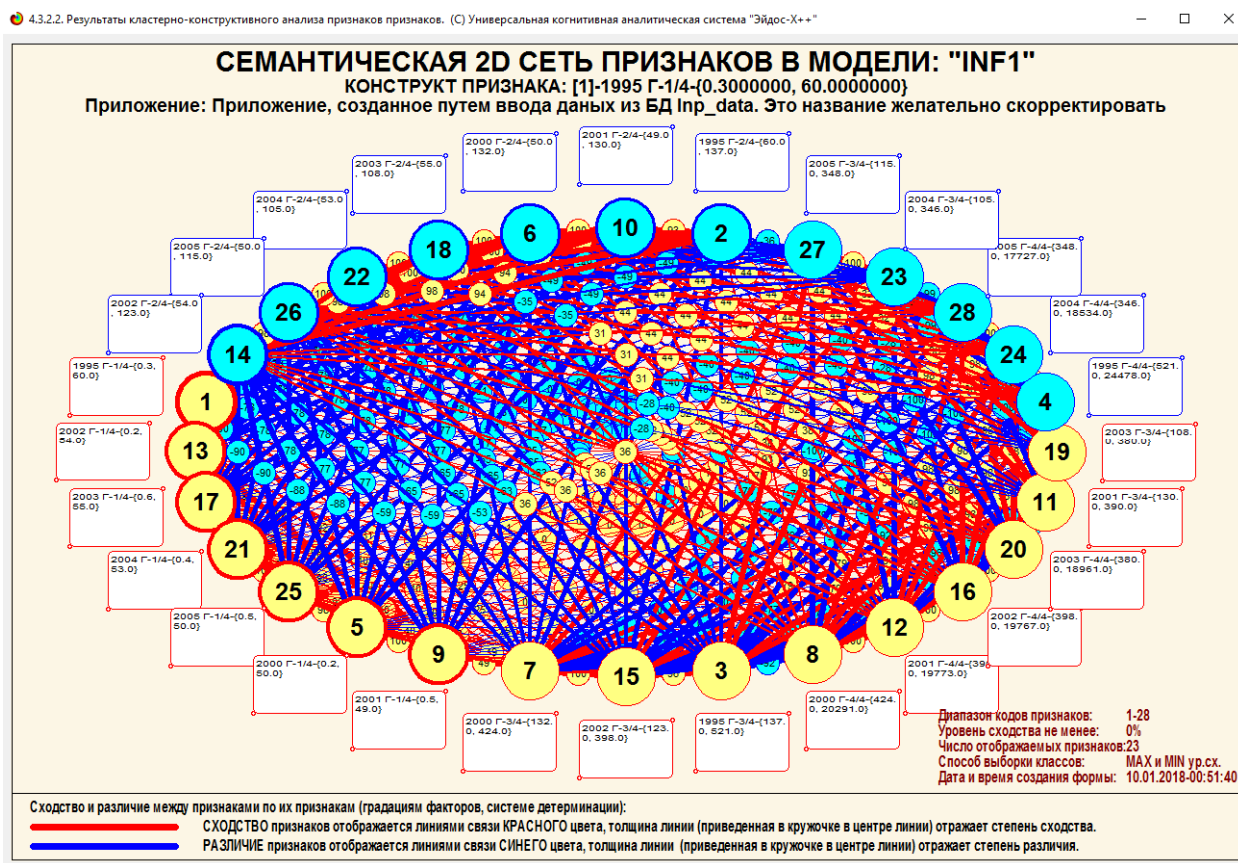


Рисунок 31 – Результаты кластерно-конструктивного анализа признаков

2.4 Нелокальный нейрон и нелокальные нейронные сети

Система «Эйдос» обеспечивает вывод системно-когнитивных моделей в форме нелокальных нейронов и нелокальных нейронных сетей.

4.4.10. Графическое отображение нелокальных нейронов в системе "Эйдос"

Выбор нелокального нейрона (класса) для визуализации

Код	Наименование нелокального нейрона (класса)
1	ВСЕГО ЗА 1995-2005 ГГ.-1/4-{3.5, 364.0}
2	ВСЕГО ЗА 1995-2005 ГГ.-2/4-{364.0, 888.0}
3	ВСЕГО ЗА 1995-2005 ГГ.-3/4-{888.0, 2755.0}
4	ВСЕГО ЗА 1995-2005 ГГ.-4/4-{2755.0, 139531.0}

Подготовка визуализации нейрона: 1 "ВСЕГО ЗА 1995-2005 ГГ.-1/4-{3.5, 364.0}" в модели: 6 "INF3"

АКТИВИРУЮЩИЕ рецепторы и сила их влияния

Код	Наименование фактора и его интервального значения	Сила влияния
5	2000 Г-1/4-{0.20000000, 50.00000000}	16.479
9	2001 Г-1/4-{0.50000000, 49.00000000}	16.479
13	2002 Г-1/4-{0.20000000, 54.00000000}	15.479
17	2003 Г-1/4-{0.60000000, 55.00000000}	15.479
21	2004 Г-1/4-{0.40000000, 53.00000000}	15.479
25	2005 Г-1/4-{0.50000000, 50.00000000}	15.479
1	1995 Г-1/4-{0.30000000, 60.00000000}	14.479

ТОРМОЗЯЩИЕ рецепторы и сила их влияния

Код	Наименование фактора и его интервального значения	Сила влияния
28	2005 Г-4/4-{348.00000000, 17727.00000000}	-5.761
24	2004 Г-4/4-{346.00000000, 18534.00000000}	-5.761
20	2003 Г-4/4-{380.00000000, 18961.00000000}	-5.761
16	2002 Г-4/4-{398.00000000, 19767.00000000}	-5.761
12	2001 Г-4/4-{390.00000000, 19773.00000000}	-5.761
8	2000 Г-4/4-{424.00000000, 20291.00000000}	-5.761
7	2000 Г-3/4-{132.00000000, 424.00000000}	-5.761
4	1995 Г-4/4-{521.00000000, 24478.00000000}	-5.761

ВКЛЮЧИТЬ фильтр по фактору

ВЫКЛЮЧИТЬ фильтр по фактору

Помощь

Аbs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7

НЕЙРОН

Максимальное количество отображаемых рецепторов: 999

Минимальный вес.коэф.отображаемых рецепторов: 0.000

Сортировать рецепторы:

☒ по информативности

☐ по модулю информативности

Отображать рецепторы:

☒ с наименованиями

☐ только с кодами

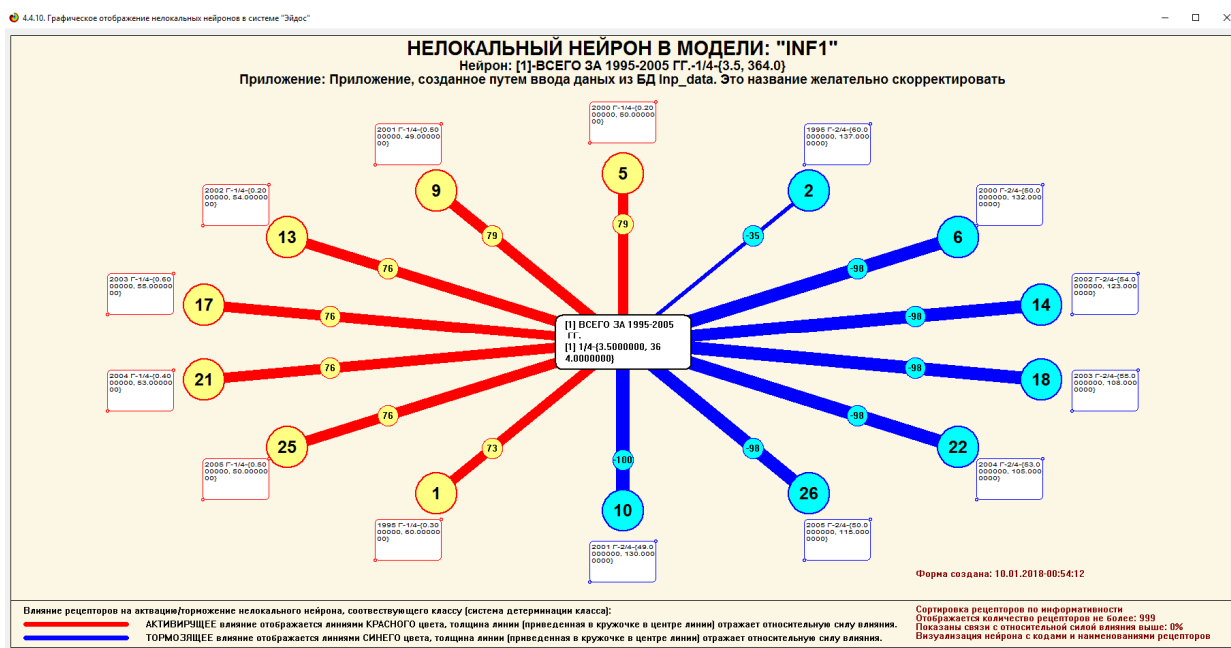


Рисунок 32 – Пример нелокальный нейрон в модели INF1

4.4.11. Отображение Паретто-подмножеств нелокальной нейронной сети в системе "Эйдос"

Выбор нелокальных нейронов (классов) для визуализации в нейросети

Sel	Код	Наименование нелокального нейрона (класса)
☒	1	ВСЕГО ЗА 1995-2005 ГГ.-1/4-{3.5, 364.0}
☐	2	ВСЕГО ЗА 1995-2005 ГГ.-2/4-{364.0, 888.0}
☒	3	ВСЕГО ЗА 1995-2005 ГГ.-3/4-{888.0, 2755.0}
☐	4	ВСЕГО ЗА 1995-2005 ГГ.-4/4-{2755.0, 139531.0}

Помощь | Максимальное количество отображаемых нейронов: 16 | ClearSet | Диапазон кодов отображаемых нейронов: 1 - 4
 Максимальное количество отображаемых связей: 1000 | Диапазон кодов отображаемых рецепторов: 1 - 28

Подготовка визуализации нейрона: 1 "ВСЕГО ЗА 1995-2005 ГГ.-1/4-{3.5, 364.0}" в модели: 6 "INF3"

АКТИВИРУЮЩИЕ рецепторы и сила их влияния | **ТОРМОЗЯЩИЕ рецепторы и сила их влияния**

Код	Наименование фактора и его интервального значения	Сила влияния
5	2000 Г-1/4-{0.20000000, 50.00000000}	16.479
9	2001 Г-1/4-{0.50000000, 49.00000000}	16.479
13	2002 Г-1/4-{0.20000000, 54.00000000}	15.479
17	2003 Г-1/4-{0.60000000, 55.00000000}	15.479
21	2004 Г-1/4-{0.40000000, 53.00000000}	15.479
25	2005 Г-1/4-{0.50000000, 50.00000000}	15.479
1	1995 Г-1/4-{0.30000000, 60.00000000}	14.479

Код	Наименование фактора и его интервального значения	Сила влияния
28	2005 Г-4/4-{348.00000000, 17727.00000000}	-5.761
24	2004 Г-4/4-{346.00000000, 18534.00000000}	-5.761
20	2003 Г-4/4-{380.00000000, 18961.00000000}	-5.761
16	2002 Г-4/4-{398.00000000, 19767.00000000}	-5.761
12	2001 Г-4/4-{390.00000000, 19773.00000000}	-5.761
8	2000 Г-4/4-{424.00000000, 20291.00000000}	-5.761
7	2000 Г-3/4-{132.00000000, 424.00000000}	-5.761
4	1995 Г-4/4-{521.00000000, 24478.00000000}	-5.761

ВКЛЮЧИТЬ фильтр по фактору | ВЫКЛЮЧИТЬ фильтр по фактору

НейроСеть | Abs | Prc1 | Prc2 | Inf1 | Inf2 | Inf3 | Inf4 | Inf5 | Inf6 | Inf7
 Максимальное количество отображаемых рецепторов: 16
 Отображать связи с интенсивностью >= % от макс.: 0.000

Сортировать связи:
☒ по модулю информативности
☐ по информативности и знаку

Отображать наименования:
☒ нейронов
☒ рецепторов

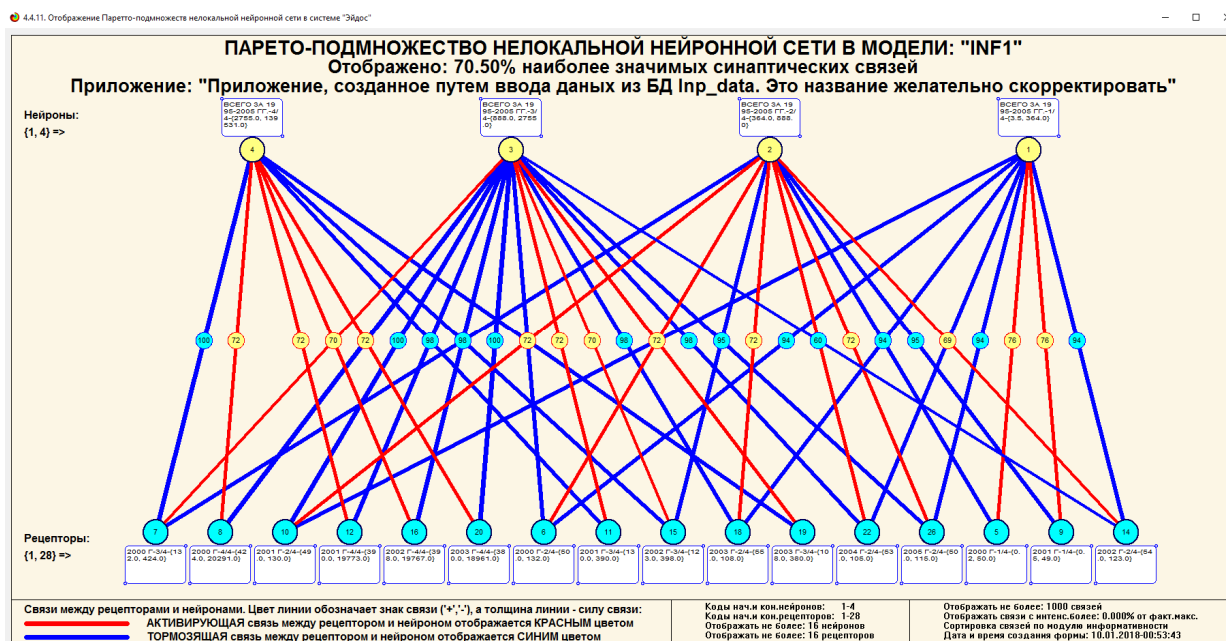


Рисунок 33 – Пример нейронной сети в модели INF1

ЗАКЛЮЧЕНИЕ

Данная работа является примером использования АСК-анализа для оценки качества математических моделей в системе «Эйдос». В результате было выяснено, что достоверной оказалась модель INF1, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели - 0,915, а полнота – 0,986. Это показывает, что уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок. Для оценки достоверности моделей в АСК – анализе в системе «Эйдос» используется L1-критерий профессора Е.В.Луценко, а также его нечеткое мультиклассовое обобщение, предложенное профессором Е.В.Луценко.

Данная лабораторная работа будет загружена в облако средствами системы «Эйдос-онлайн»[6]

Использованная литература

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: 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 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.

5. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>, 0,812 у.п.л.

6. Луценко Е.В. Открытая масштабируемая интерактивная интеллектуальная on-line среда для обучения и научных исследований на базе АСК-анализа и системы «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №06(130). С. 1 – 55. – IDA [article ID]: 1301706001. – Режим доступа: <http://ej.kubagro.ru/2017/06/pdf/01.pdf>, 3,438 у.п.л.