

УДК 004.8  
05.00.00. Технические науки

**АСК-АНАЛИЗ ЭФФЕКТИВНОСТИ РАБОТЫ  
ПРЕПОДАВАТЕЛЯ АГРАРНОГО ВУЗА НА  
ОСНОВЕ ДАННЫХ РЕПОЗИТОРИЯ UCI**

Луценко Евгений Вениаминович  
д.э.н., к.т.н., профессор  
Scopus Author ID: 57191193316  
РИНЦ SPIN-код: 9523-7101  
[prof.lutsenko@gmail.com](mailto:prof.lutsenko@gmail.com)

Кострома Дмитрий Сергеевич  
студент группы: ПИ1301  
[dimagelen@mail.ru](mailto:dimagelen@mail.ru)  
*Кубанский государственный аграрный  
университет имени И.Т.Трубилина, Краснодар,  
Россия*

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 рассмотрено решение задачи идентификации классов уровней оплаты сотрудников фирмы по их характеристикам. 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 тестовых задач для систем искусственного интеллекта репозитория UCI. В данной работе использована база данных по эффективности преподавания в течение трех регулярных семестров и двух летних семестров 151 ассистентом преподавателя (ТА) назначений в департаменте статистики Университета Висконсин-Мэдисон. При этом наиболее достоверной в данном приложении оказались модели INF4. Достоверность модели в соответствии с L-мерой составила 0,809, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 70%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется F-критерий Ван Ризбергера и ее нечеткое мультиклассовое обобщение, предложенное проф.Е.В.Луценко

Ключевые слова: ИССЛЕДОВАНИЕ  
ЭФФЕКТИВНОСТИ РАБОТЫ  
ПРЕПОДАВАТЕЛЯ АГРАРНОГО ВУЗА НА  
ОСНОВЕ МНОГОКРИТЕРИАЛЬНОГО  
ПОДХОДА

**Doi: 10.21515/1990-4665-124-003**

UDC 004.8  
Engineering

**ASC-ANALYSIS OF THE EFFICIENCY OF  
WORK OF TEACHERS OF AN AGRARIAN  
UNIVERSITY ON THE BASIS OF THE UCI  
REPOSITORY DATA**

Lutsenko Eugeny Veniaminovich  
Dr.Sci.Econ., Cand.Tech.Sci., professor  
Scopus Author ID: 57191193316  
RSCI SPIN-code: 9523-7101  
[prof.lutsenko@gmail.com](mailto:prof.lutsenko@gmail.com)

Kostroma Dmitry Sergeevich  
Student groups: PI1301  
[dimagelen@mail.ru](mailto:dimagelen@mail.ru)  
*Kuban State Agrarian University, Krasnodar,  
Russia*

The creation of artificial intelligence systems is one of important and perspective directions of development of modern information technology. As there are many alternatives to artificial intelligence systems, there is a need to evaluate mathematical models of these systems. In this article, we consider a solution of the problem of identifying classes of levels of pay to employees on their characteristics. To achieve this goal it requires free access to test the source data and methodology, which will help to convert the data into the form needed for work in artificial intelligence systems. A good choice is a database of test problems for systems of UCI artificial intelligence repository. In this work we have used data base on teaching effectiveness for three regular semesters and two summer semesters of 151 teaching assistant (TA) assignments at the statistics Department of the University of Wisconsin-Madison. The most reliable in this application was the model of the INF4. The accuracy of the model in accordance with L-measure made up 0,809, which is much higher than the reliability of expert evaluations, which is equal to about 70%. To assess the reliability of the models in the ASC-analysis and in the system of "Eidos" we use F-criterion of van Ritbergen and its fuzzy multiclass generalization proposed by Professor E. V. Lutsenko

Keywords: STUDY OF THE EFFECTIVENESS  
OF THE WORK OF TEACHERS OF AN  
AGRARIAN UNIVERSITY ON THE BASIS OF  
MULTI-CRITERIA APPROACH

## СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ .....</b>	<b>2</b>
<b>1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ.....</b>	<b>3</b>
<b>1.1. ОПИСАНИЕ РЕШЕНИЯ.....</b>	<b>3</b>
<b>1.2. ПРЕОБРАЗОВАНИЕ ИСХОДНЫХ ДАННЫХ ИЗ HTML-ФОРМАТА В ФАЙЛ ИСХОДНЫХ ДАННЫХ MS EXCEL.....</b>	<b>3</b>
<b>1.3. СИНТЕЗ И ВЕРИФИКАЦИЯ СТАТИСТИЧЕСКИХ И ИНТЕЛЛЕКТУАЛЬНЫХ МОДЕЛЕЙ.....</b>	<b>13</b>
<b>1.4. ВИДЫ МОДЕЛЕЙ СИСТЕМЫ «ЭЙДОС».....</b>	<b>15</b>
<b>1.5. РЕЗУЛЬТАТЫ ВЕРИФИКАЦИИ МОДЕЛЕЙ .....</b>	<b>17</b>
<b>2. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....</b>	<b>23</b>
<b>2.1. РЕШЕНИЕ ЗАДАЧИ ИДЕНТИФИКАЦИИ.....</b>	<b>23</b>
<b>2.2. КОГНИТИВНЫЕ ФУНКЦИИ .....</b>	<b>26</b>
<b>2.3. SWOT и PEST МАТРИЦЫ И ДИАГРАММЫ.....</b>	<b>28</b>
<b>ЗАКЛЮЧЕНИЕ.....</b>	<b>31</b>
<b>ЛИТЕРАТУРА.....</b>	<b>31</b>

---

## 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 работах [1, 2] рассматривалось решение этой задачи для оценки качества модели, реализованной в универсальной автоматизированной системе «Эйдос», представляющей собой программный инструментарий автоматизированного системно- когнитивного анализа (АСК-анализ) [3, 4]. 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

тестовых задач для систем искусственного интеллекта репозитория UCI: <http://archive.ics.uci.edu/ml/datasets.html>. «Репозиторий UCI (UCI Machine Learning Repository) — крупнейший репозиторий реальных и модельных задач машинного обучения. Содержит реальные данные по прикладным задачам в области биологии, медицины, физики, техники, социологии, и др. Задачи (наборы данных, data set) именно этого репозитория чаще всего используются научным сообществом для эмпирического анализа алгоритмов машинного обучения. Репозиторий UCI создан в университете г.Ирвин (Калифорния, США).

## **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 банка исходных данных по задачам искусственного интеллекта — репозитория UCI получаем исходную информацию по базе данных «Teaching Assistant Evaluation Data Set», которую оставим без изменений.

#### **Общее описание задачи:**

(файл: «tae.names» по адресу: <http://archive.ics.uci.edu/ml/machine-learning-databases/tae/tae.names>)

1. Title: Teaching Assistant Evaluation
2. Sources:
  - (a) Collector: Wei-Yin Loh (Department of Statistics, UW-Madison)
  - (b) Donor: Tjen-Sien Lim (limt@stat.wisc.edu)
  - (b) Date: June 7, 1997

3. Past Usage:
  1. Loh, W.-Y. & Shih, Y.-S. (1997). Split Selection Methods for Classification Trees, *Statistica Sinica* 7: 815-840.
  2. Lim, T.-S., Loh, W.-Y. & Shih, Y.-S. (1999). A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-three Old and New Classification Algorithms. *Machine Learning*. Forthcoming.

(ftp://ftp.stat.wisc.edu/pub/loh/treeprogs/quest1.7/mach1317.pdf or  
<http://www.stat.wisc.edu/~limt/mach1317.pdf>)

4. Relevant Information:
 

The data consist of evaluations of teaching performance over three regular semesters and two summer semesters of 151 teaching assistant (TA) assignments at the Statistics Department of the University of Wisconsin-Madison. The scores were divided into 3 roughly equal-sized categories ("low", "medium", and "high") to form the class variable.

5. Number of Instances: 151
6. Number of Attributes: 6 (including the class attribute)
7. Attribute Information:
  1. Whether of not the TA is a native English speaker (binary) 1=English speaker, 2=non-English speaker
  2. Course instructor (categorical, 25 categories)
  3. Course (categorical, 26 categories)
  4. Summer or regular semester (binary) 1=Summer, 2=Regular
  5. Class size (numerical)
  6. Class attribute (categorical) 1=Low, 2=Medium, 3=High
8. Missing Attribute Values: None

### Обучающая выборка

На сайте UCI по адресу: <http://archive.ics.uci.edu/ml/machine-learning-databases/tae/tae.data> представлена обучающая выборка.

Фрагмент исходных данных «flare.data1» приведен в таблице 1:

Таблица 1 – tae.data; (фрагмент)

```
1, 23, 3, 1, 19, 3
2, 15, 3, 1, 17, 3
1, 23, 3, 2, 49, 3
1, 5, 2, 2, 33, 3
2, 7, 11, 2, 55, 3
2, 23, 3, 1, 20, 3
```

2, 9, 5, 2, 19, 3  
 2, 10, 3, 2, 27, 3  
 1, 22, 3, 1, 58, 3  
 2, 15, 3, 1, 20, 3  
 2, 10, 22, 2, 9, 3  
 2, 13, 1, 2, 30, 3  
 2, 18, 21, 2, 29, 3  
 2, 6, 17, 2, 39, 3  
 2, 6, 17, 2, 42, 2  
 2, 6, 17, 2, 43, 2  
 2, 7, 11, 2, 10, 2  
 2, 22, 3, 2, 46, 2  
 2, 13, 3, 1, 10, 2

### **Data Set Information:**

The data consist of evaluations of teaching performance over three regular semesters and two summer semesters of 151 teaching assistant (TA) assignments at the Statistics Department of the University of Wisconsin-Madison. The scores were divided into 3 roughly equal-sized categories ("low", "medium", and "high") to form the class variable.

### **Attribute Information:**

1. Whether or not the TA is a native English speaker (binary); 1=English speaker, 2=non-English speaker
2. Course instructor (categorical, 25 categories)
3. Course (categorical, 26 categories)
4. Summer or regular semester (binary) 1=Summer, 2=Regular
5. Class size (numerical)
6. Class attribute (categorical) 1=Low, 2=Medium, 3=High

### **Информация набора данных:**

Данные состоят из оценок эффективности преподавания в течение трех регулярных семестров и двух летних семестров 151 ассистентом преподавателя (ТА) назначений в департаменте статистики Университета Висконсин-Мэдисон. Счет были разделены на 3 примерно одинакового размера категории ("низкий", "средний" и "высокий"), чтобы сформировать переменную класса.

### **Описательная информация:**

1. Знание английского (бинарный) 1 = английского языка, 2 = без английского языка
2. Преподаватель курса (категоричны, 25 категорий)
3. Курс (категоричны, 26 категорий)
4. Летний семестр или обычный (бинарный) 1 = Лето, 2 = Регулярное
5. Размер класса (числовой)
6. Атрибуты класса (категоричны) 1 = низкий, 2 = средний, 3 = высокий

Поскольку ввод исходных данных в систему «Эйдос» планируется осуществить с помощью ее универсального программного интерфейса импорта данных из внешних баз данных, который работает с файлами MS Excel, то преобразуем данные из html-файла в xls-файл, для чего выполним следующие операции.

Скопируем получившуюся таблицу из MS Word в MS Excel и запишем ее с именем: Inp\_data.xls в папку: c:\Aidos-X\AID\_DATA\Inp\_data\. В файле Inp\_data.xls добавим пустую колонку на позиции «А» и автоматически пронумеруем все строки. В результате получим таблицу исходных данных, полностью подготовленную для обработки в системе «Эйдос» и записанную в нужную папку в виде файла нужного типа с нужным именем (таблица 1):

Таблица 1 – Исходные данные

n	Class attribute	English	Course instructor	Course	Summer or regular	Class size
1	3/3-High	yes	23	3	Summer	19
2	3/3-High	no	15	3	Summer	17
3	3/3-High	yes	23	3	Regular	49
4	3/3-High	yes	5	2	Regular	33
5	3/3-High	no	7	11	Regular	55
6	3/3-High	no	23	3	Summer	20
7	3/3-High	no	9	5	Regular	19
8	3/3-High	no	10	3	Regular	27
9	3/3-High	yes	22	3	Summer	58
10	3/3-High	no	15	3	Summer	20
11	3/3-High	no	10	22	Regular	9
12	3/3-High	no	13	1	Regular	30
13	3/3-High	no	18	21	Regular	29
14	3/3-High	no	6	17	Regular	39
15	2/3-Medium	no	6	17	Regular	42
16	2/3-Medium	no	6	17	Regular	43
17	2/3-Medium	no	7	11	Regular	10
18	2/3-Medium	no	22	3	Regular	46
19	2/3-Medium	no	13	3	Summer	10
20	2/3-Medium	no	7	25	Regular	42
21	2/3-Medium	no	25	7	Regular	27
22	2/3-Medium	no	25	7	Regular	23
23	2/3-Medium	no	2	9	Regular	31
24	2/3-Medium	no	1	15	Summer	22
25	2/3-Medium	no	15	13	Regular	37
26	2/3-Medium	no	7	11	Regular	13
27	2/3-Medium	no	8	3	Regular	24
28	2/3-Medium	no	14	15	Regular	38
29	1/3-Low	no	21	2	Regular	42
30	1/3-Low	no	22	3	Regular	28
31	1/3-Low	no	11	1	Regular	51
32	1/3-Low	no	18	5	Regular	19
33	1/3-Low	no	13	1	Regular	31
34	1/3-Low	yes	13	3	Summer	13
35	1/3-Low	no	5	2	Regular	37

36	1/3-Low	no	16	8	Regular	36
37	1/3-Low	no	4	16	Regular	21
38	1/3-Low	no	5	2	Regular	48
39	1/3-Low	no	14	15	Regular	38
40	3/3-High	yes	23	3	Summer	19
41	3/3-High	no	15	3	Summer	17
42	3/3-High	yes	23	3	Regular	49
43	3/3-High	yes	5	2	Regular	33
44	3/3-High	no	7	11	Regular	55
45	3/3-High	no	23	3	Summer	20
46	3/3-High	no	9	5	Regular	19
47	3/3-High	no	10	3	Regular	27
48	3/3-High	yes	22	3	Regular	58
49	3/3-High	no	15	3	Summer	20
50	3/3-High	no	10	22	Regular	9
51	3/3-High	no	13	1	Regular	30
52	3/3-High	no	18	21	Regular	29
53	3/3-High	no	6	17	Regular	39
54	2/3-Medium	no	6	17	Regular	42
55	2/3-Medium	no	6	17	Regular	43
56	2/3-Medium	no	7	11	Regular	10
57	2/3-Medium	no	22	3	Regular	46
58	2/3-Medium	no	13	3	Summer	10
59	2/3-Medium	no	7	25	Regular	42
60	2/3-Medium	no	25	7	Regular	27
61	2/3-Medium	no	25	7	Regular	23
62	2/3-Medium	no	2	9	Regular	31
63	2/3-Medium	no	1	15	Summer	22
64	2/3-Medium	no	15	13	Regular	37
65	2/3-Medium	no	7	11	Regular	13
66	2/3-Medium	no	8	3	Regular	24
67	2/3-Medium	no	14	15	Regular	38
68	1/3-Low	no	21	2	Regular	42
69	1/3-Low	no	22	3	Regular	28
70	1/3-Low	no	11	1	Regular	51
71	1/3-Low	no	18	5	Regular	19
72	1/3-Low	no	13	1	Regular	31
73	1/3-Low	yes	13	3	Summer	13
74	1/3-Low	no	5	2	Regular	37
75	1/3-Low	no	16	8	Regular	36
76	1/3-Low	no	4	16	Regular	21
77	1/3-Low	no	5	2	Regular	48
78	1/3-Low	no	14	15	Regular	38
79	3/3-High	yes	23	3	Summer	25
80	3/3-High	yes	13	3	Summer	17
81	3/3-High	no	16	19	Regular	11
82	3/3-High	no	9	2	Regular	39
83	3/3-High	no	13	3	Summer	11
84	3/3-High	no	18	21	Regular	19
85	3/3-High	yes	22	3	Regular	45
86	3/3-High	no	7	11	Summer	20
87	3/3-High	no	23	3	Summer	20
88	3/3-High	yes	23	3	Summer	20
89	3/3-High	yes	23	3	Regular	38
90	3/3-High	no	14	22	Regular	17
91	3/3-High	yes	17	17	Regular	19
92	3/3-High	no	9	5	Regular	24
93	3/3-High	no	18	25	Regular	25
94	3/3-High	yes	17	17	Regular	31
95	3/3-High	no	1	15	Regular	31
96	2/3-Medium	no	1	8	Regular	18
97	2/3-Medium	yes	11	16	Regular	22
98	2/3-Medium	yes	22	13	Regular	27
99	2/3-Medium	no	9	2	Regular	14
100	2/3-Medium	no	13	1	Regular	20
101	2/3-Medium	yes	6	17	Regular	35
102	2/3-Medium	no	23	3	Summer	20
103	2/3-Medium	yes	23	3	Summer	20
104	2/3-Medium	no	6	17	Regular	37
105	2/3-Medium	yes	22	3	Regular	15
106	2/3-Medium	no	20	2	Regular	25
107	2/3-Medium	no	23	3	Regular	10

108	1/3-Low	no	20	2	Regular	14
109	1/3-Low	yes	23	3	Regular	38
110	1/3-Low	no	13	1	Regular	29
111	1/3-Low	no	10	3	Regular	19
112	1/3-Low	no	7	11	Regular	30
113	1/3-Low	yes	14	15	Regular	32
114	1/3-Low	no	8	3	Regular	27
115	1/3-Low	no	12	7	Regular	34
116	1/3-Low	no	8	7	Regular	23
117	1/3-Low	no	15	1	Regular	66
118	1/3-Low	no	23	3	Regular	12
119	1/3-Low	no	2	9	Regular	29
120	1/3-Low	no	15	1	Regular	19
121	1/3-Low	no	20	2	Regular	3
122	3/3-High	no	13	14	Regular	17
123	3/3-High	no	9	6	Regular	7
124	3/3-High	yes	10	3	Regular	21
125	3/3-High	no	14	15	Regular	36
126	3/3-High	yes	13	1	Regular	54
127	3/3-High	yes	8	3	Regular	29
128	3/3-High	no	20	2	Regular	45
129	2/3-Medium	no	22	1	Regular	11
130	2/3-Medium	no	18	12	Regular	16
131	2/3-Medium	no	20	15	Regular	18
132	2/3-Medium	yes	17	18	Regular	44
133	2/3-Medium	no	14	23	Regular	17
134	2/3-Medium	no	24	26	Regular	21
135	2/3-Medium	no	9	24	Regular	20
136	2/3-Medium	no	12	8	Regular	24
137	2/3-Medium	no	9	6	Regular	5
138	2/3-Medium	no	22	1	Regular	42
139	1/3-Low	no	7	11	Regular	30
140	1/3-Low	no	10	3	Regular	19
141	1/3-Low	no	23	3	Regular	11
142	1/3-Low	no	17	18	Regular	29
143	1/3-Low	no	16	20	Regular	15
144	1/3-Low	no	3	2	Regular	37
145	1/3-Low	no	19	4	Regular	10
146	1/3-Low	no	23	3	Regular	24
147	1/3-Low	no	3	2	Regular	26
148	1/3-Low	no	10	3	Regular	12
149	1/3-Low	yes	18	7	Regular	48
150	1/3-Low	no	22	1	Regular	51
151	1/3-Low	no	2	10	Regular	27

**Автоматизированная формализация предметной области путем импорта исходных данных из внешних баз данных в систему "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 1):

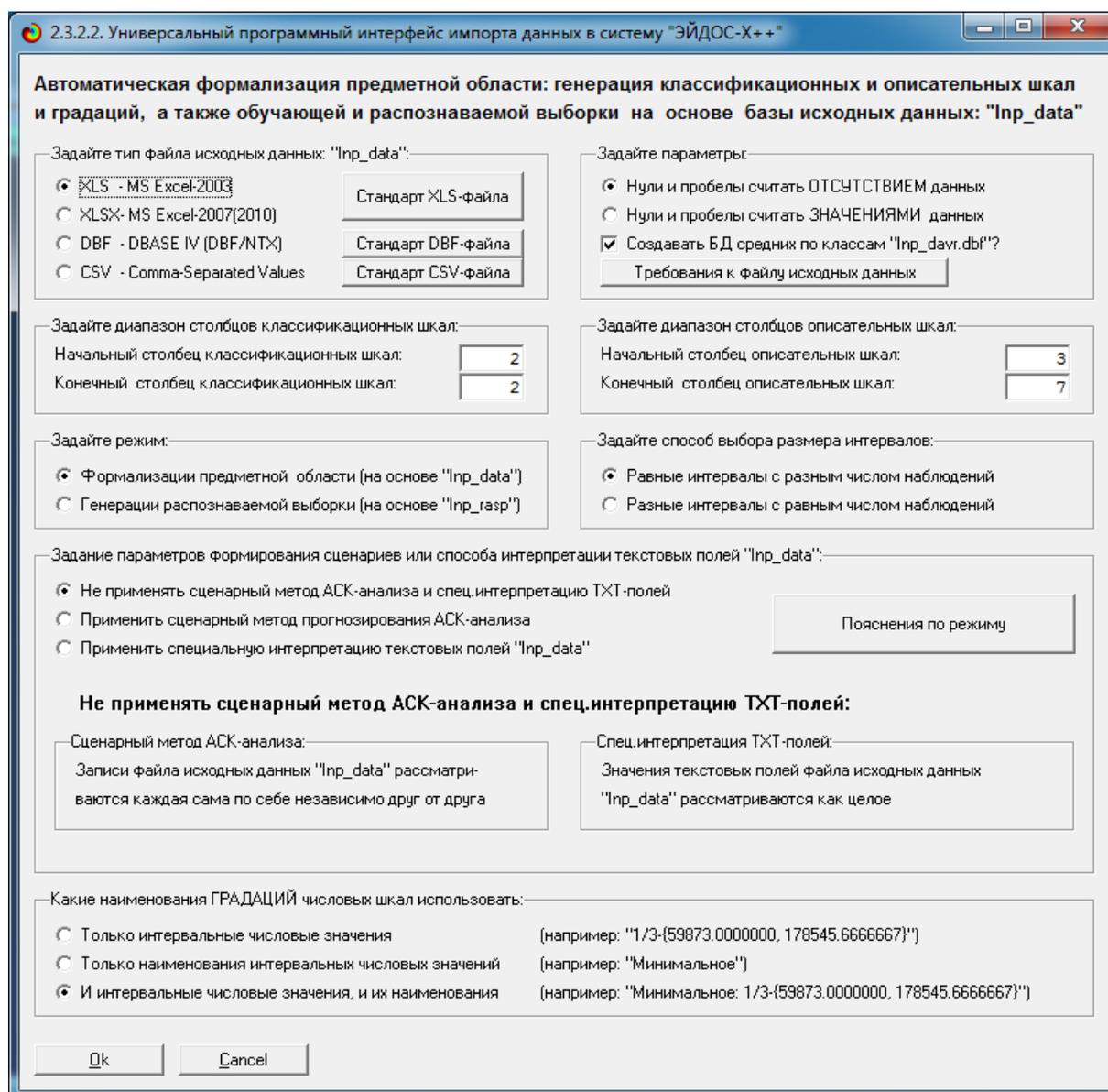


Рисунок 1. Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

В экранной форме, приведенной на рисунке 1, задать настройки, показанные на рисунке:

- "Задайте тип файла исходных данных Inp\_data": "XLS – MS Excel-2003";
- "Задайте диапазон столбцов классификационных шкал": "Начальный столбец классификационных шкал" – 2, "Конечный столбец классификационных шкал" – 2 (последний столбец в таблице);

– "Задайте диапазон столбцов описательных шкал": "Начальный столбец описательных шкал" – 3, "Конечный столбец описательных шкал" – 7;

– "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 2). В этом окне необходимо нажать кнопку "Выйти на создание модели".

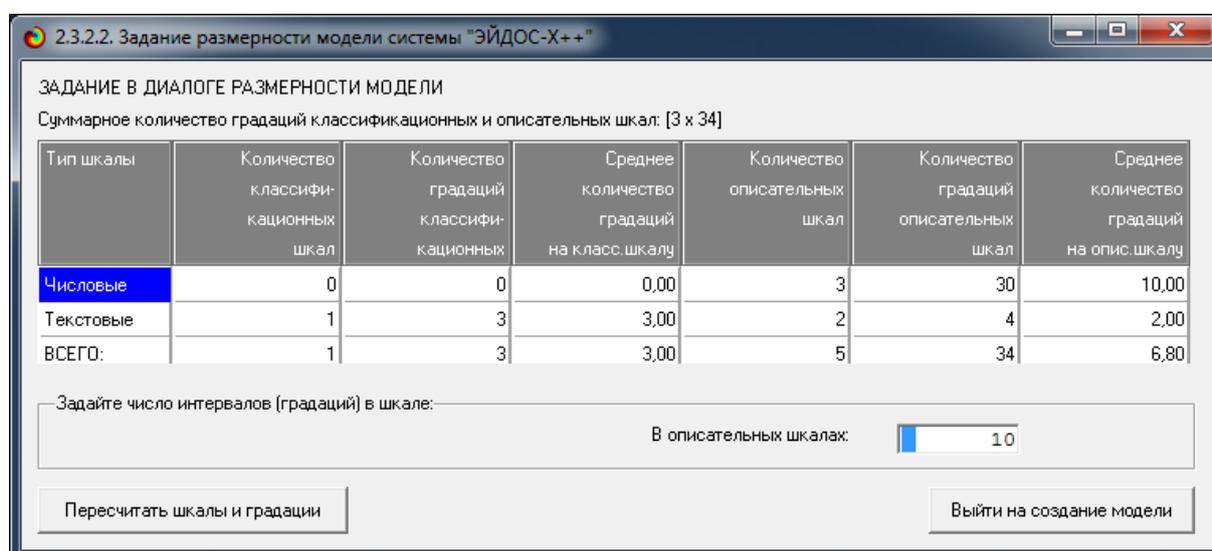


Рисунок 2. Задание размерности модели системы "Эйдос"

Далее открывается окно, отображающее стадию процесса импорта данных из внешней БД "Inp\_data.xls" в систему "Эйдос" (рисунок 3), а также прогноз времени завершения этого процесса. В том окне необходимо дождаться завершения формализации предметной области и нажать кнопку "ОК".

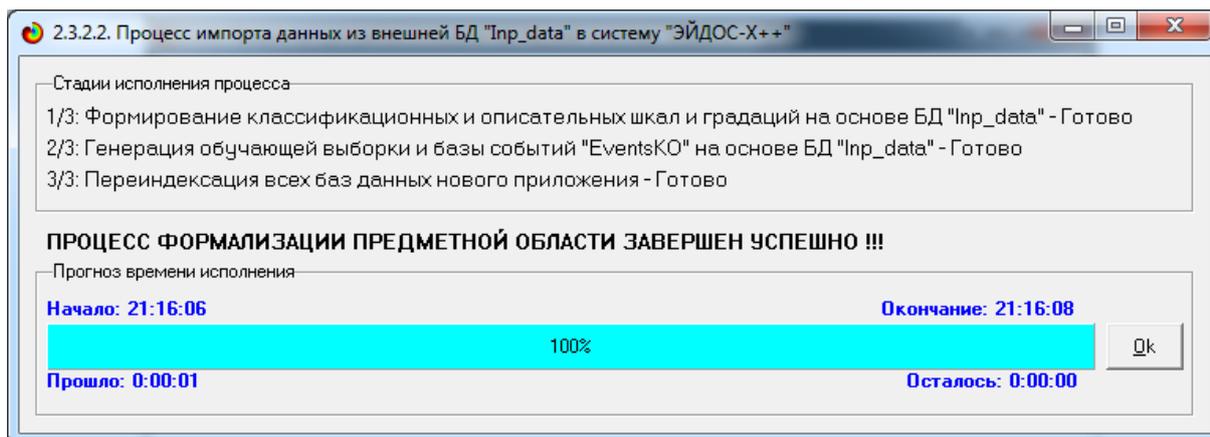


Рисунок 3. Процесс импорта данных из внешней БД "Inp\_data.xls" в систему "Эйдос"

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 4).

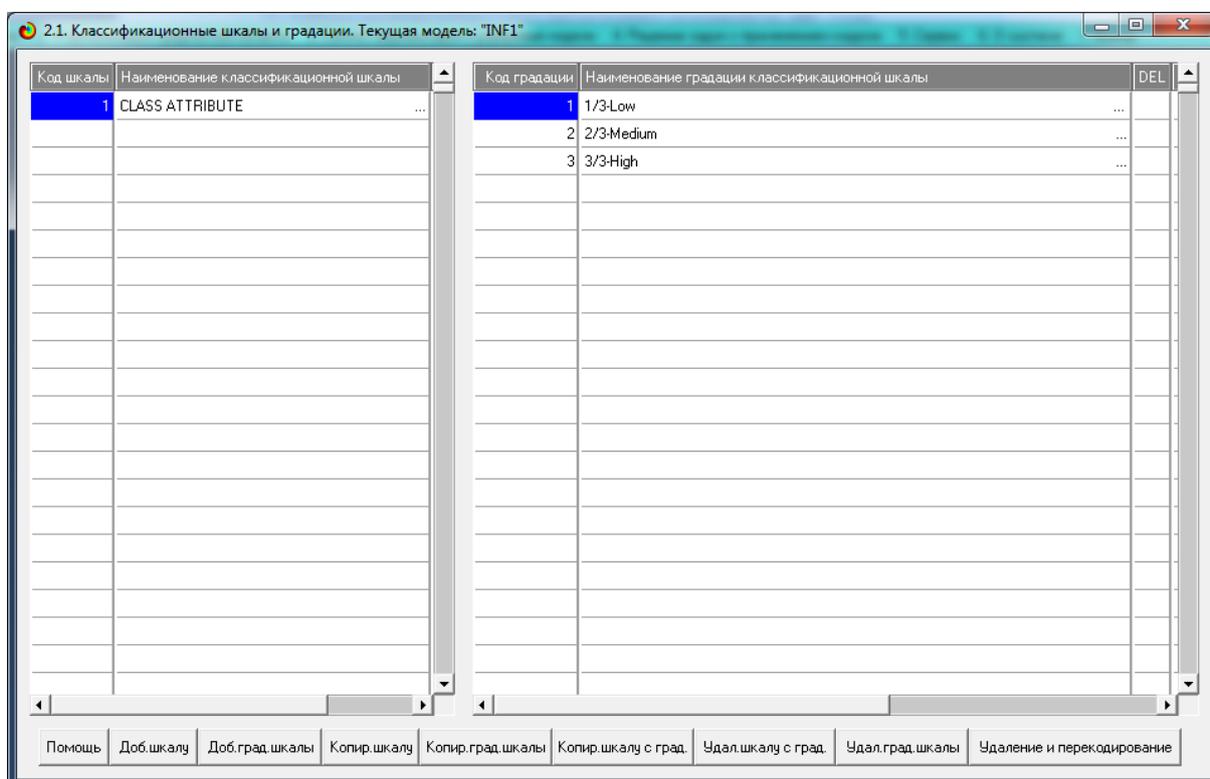


Рисунок 4. Классификационные шкалы и градации

Для просмотра описательных шкал и градаций необходимо запустить режим 2.2 (рисунок 5):

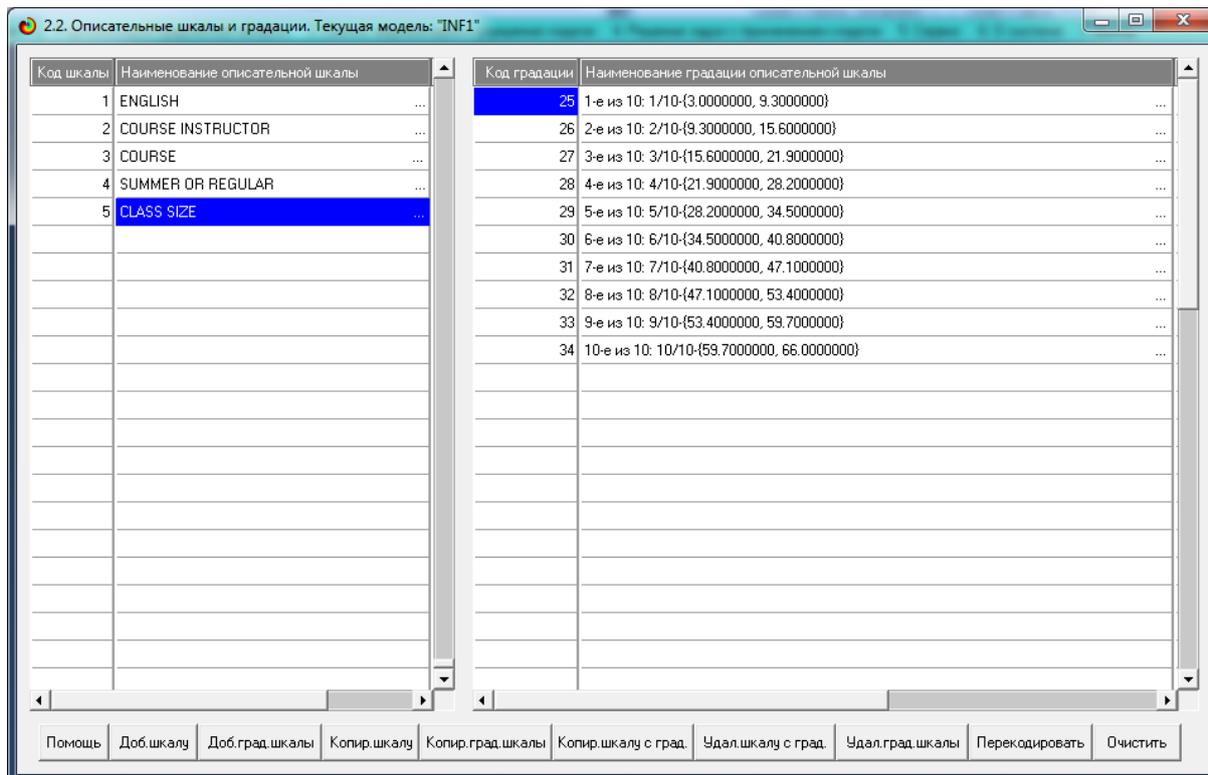


Рисунок 5. Описательные шкалы и градации (фрагмент)

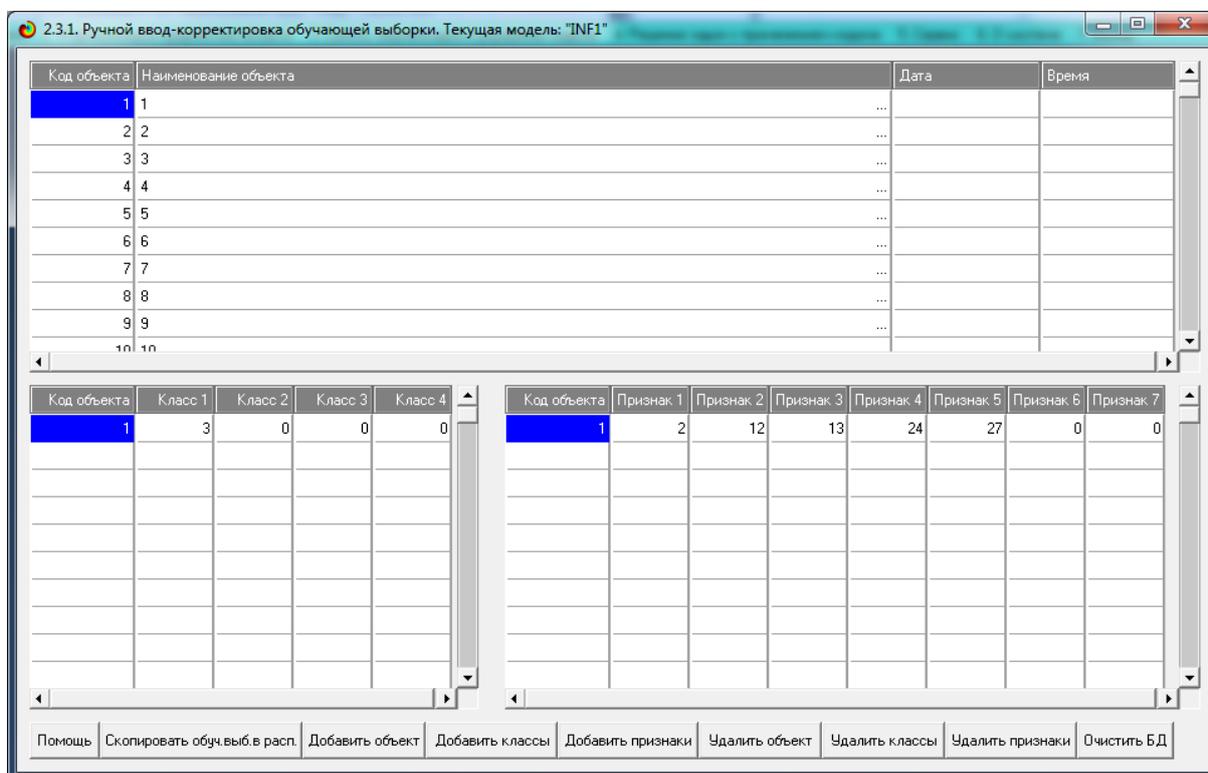


Рисунок 6. Обучающая выборка (фрагмент)

Тем самым создаются все необходимые и достаточные предпосылки для выявления силы и направления причинно-следственных связей между значениями факторов и результатами их совместного системного воздействия (с учетом нелинейности системы [10]).

### 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 7).

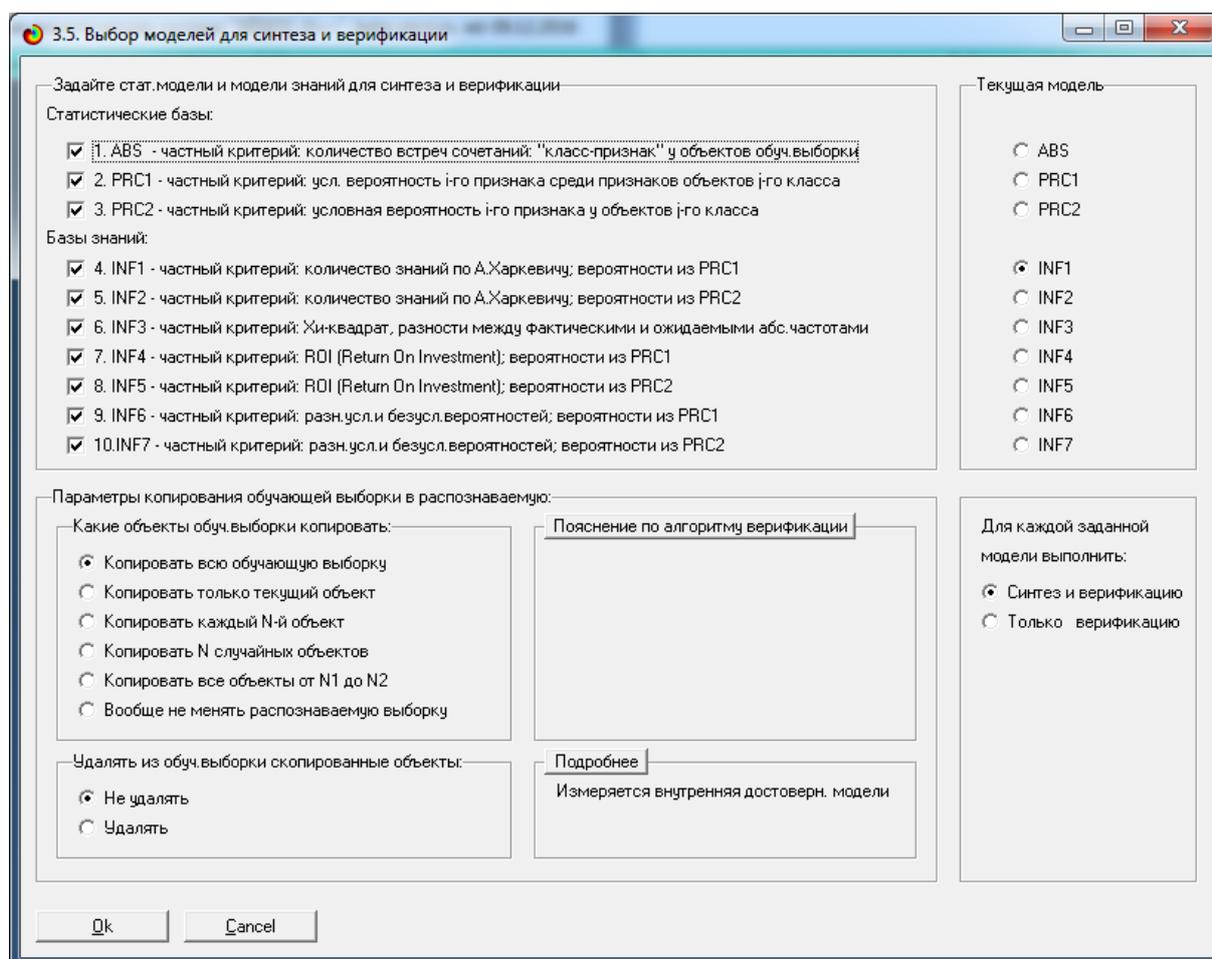


Рисунок 7. Выбор моделей для синтеза и верификации, а также текущей модели

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 7. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 8.

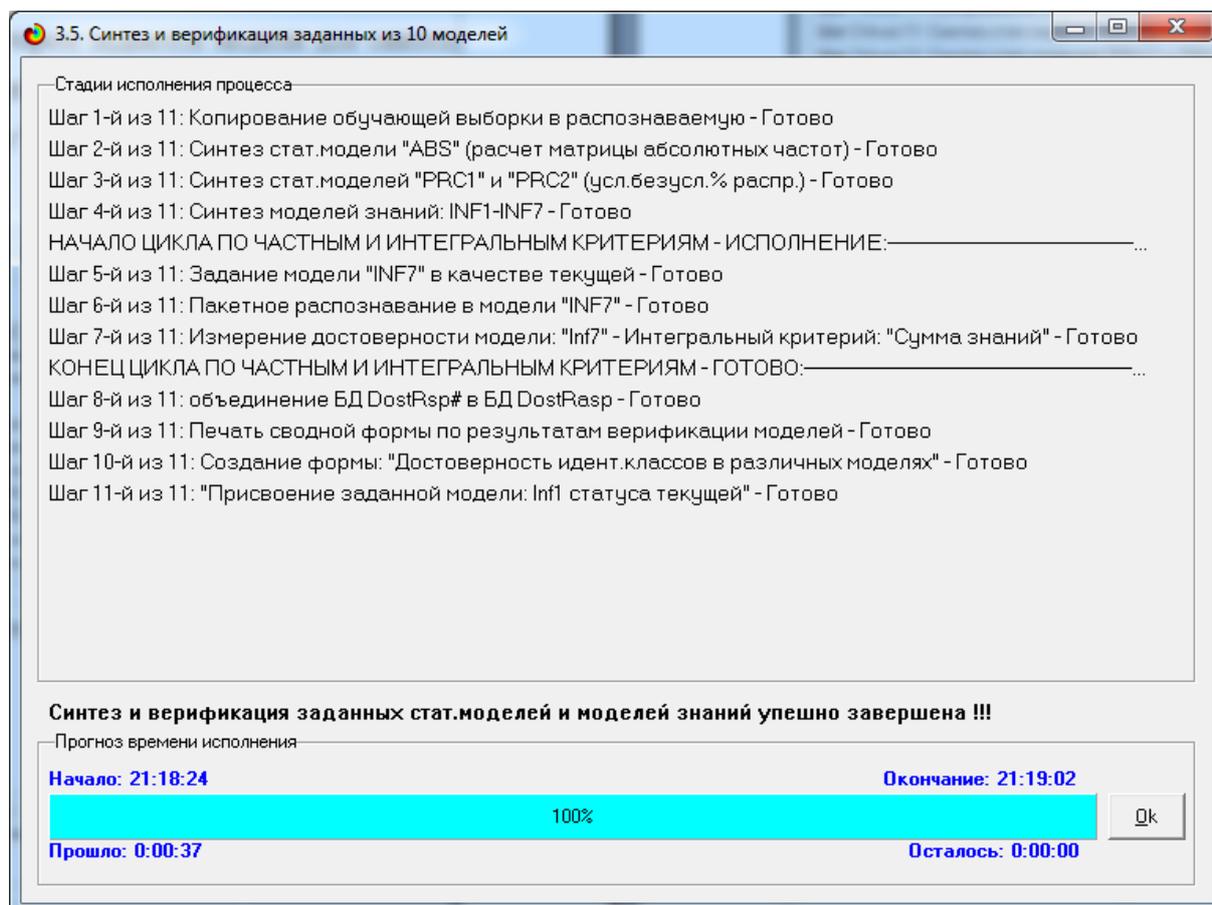


Рисунок 8. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 37 секунд. При этом верификация (оценка достоверности моделей) проводилась на всех 151 примерах наблюдения из обучающей выборки. В результате выполнения режима 3.5 созданы все модели, со всеми частными критериями, перечисленные на рисунке 7, но ниже мы приведем лишь некоторые из них (таблицы 2, 3, 4).

### 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 2) в матрицы условных и безусловных процентных распределений, и матрицы знаний (таблицы 3 и 4).

Таблица 2 – Матрица абсолютных частот (модель ABS)

Код признака	Наименование описательной шкалы и градации	1. CLASS ATTRIBUTE 1/3 LOW	2. CLASS ATTRIBUTE 2/3 MEDIUM	3. CLASS ATTRIBUTE 3/3 HIGH	Сумма	Среднее	Средн. квадрат. откл.
1	ENGLISH-no	44	44	34	122	40.667	5.774
2	ENGLISH-yes	5	6	18	29	9.667	7.234
3	COURSE INSTRUCTOR-1-е из 10: 1/10-{1.0000000, 3...	4	5	1	10	3.333	2.082
4	COURSE INSTRUCTOR-2-е из 10: 2/10-{3.4000000, 5...	6		2	8	2.667	3.055
5	COURSE INSTRUCTOR-3-е из 10: 3/10-{5.8000000, 8...	4	14	6	24	8.000	5.292
6	COURSE INSTRUCTOR-4-е из 10: 4/10-{8.2000000, 1...	3	3	10	16	5.333	4.041
7	COURSE INSTRUCTOR-5-е из 10: 5/10-{10.6000000, ...	8	5	6	19	6.333	1.528
8	COURSE INSTRUCTOR-6-е из 10: 6/10-{13.0000000, ...	5	5	6	16	5.333	0.577
9	COURSE INSTRUCTOR-7-е из 10: 7/10-{15.4000000, ...	4	1	3	8	2.667	1.528
10	COURSE INSTRUCTOR-8-е из 10: 8/10-{17.8000000, ...	6	3	5	14	4.667	1.528
11	COURSE INSTRUCTOR-9-е из 10: 9/10-{20.2000000, ...	5	6	3	14	4.667	1.528
12	COURSE INSTRUCTOR-10-е из 10: 10/10-{22.6000000, ...	4	8	10	22	7.333	3.055
13	COURSE-1-е из 10: 1/10-{1.0000000, 3.5000000} ...	30	15	30	75	25.000	8.660
14	COURSE-2-е из 10: 2/10-{3.5000000, 6.0000000} ...	3	1	4	8	2.667	1.528
15	COURSE-3-е из 10: 3/10-{6.0000000, 8.5000000} ...	5	6		11	3.667	3.215
16	COURSE-4-е из 10: 4/10-{8.5000000, 11.0000000} ...	4	6	3	13	4.333	1.528
17	COURSE-5-е из 10: 5/10-{11.0000000, 13.5000000} ...		4	4	4	1.333	2.309
18	COURSE-6-е из 10: 6/10-{13.5000000, 16.0000000} ...	5	6	3	14	4.667	1.528

Таблица 3 – Матрица информативностей (модель INF1) в битах (фрагмент)

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу, вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. CLASS ATTRIBUTE 1/3 LOW	2. CLASS ATTRIBUTE 2/3 MEDIUM	3. CLASS ATTRIBUTE 3/3 HIGH	Сумма	Среднее	Средн. квадр. откл.
1	ENGLISH-no	0.025	0.020	-0.051	-0.005	-0.002	0.042
2	ENGLISH-yes	-0.151	-0.112	0.141	-0.123	-0.041	0.159
3	COURSE INSTRUCTOR-1-е из 10: 1/10-{1.00000...}	0.050	0.099	-0.296	-0.147	-0.049	0.215
4	COURSE INSTRUCTOR-2-е из 10: 2/10-{3.40000...}	0.200		-0.077	0.124	0.041	0.143
5	COURSE INSTRUCTOR-3-е из 10: 3/10-{5.80000...}	-0.159	0.135	-0.077	-0.101	-0.034	0.152
6	COURSE INSTRUCTOR-4-е из 10: 4/10-{8.20000...}	-0.131	-0.136	0.143	-0.125	-0.042	0.159
7	COURSE INSTRUCTOR-5-е из 10: 5/10-{10.60000...}	0.062	-0.055	-0.021	-0.013	-0.004	0.060
8	COURSE INSTRUCTOR-6-е из 10: 6/10-{13.00000...}	-0.009	-0.014	0.020	-0.002	-0.001	0.019
9	COURSE INSTRUCTOR-7-е из 10: 7/10-{15.40000...}	0.103	-0.233	0.020	-0.109	-0.036	0.175
10	COURSE INSTRUCTOR-8-е из 10: 8/10-{17.80000...}	0.067	-0.104	0.009	-0.029	-0.010	0.087
11	COURSE INSTRUCTOR-9-е из 10: 9/10-{20.20000...}	0.023	0.062	-0.113	-0.029	-0.010	0.092
12	COURSE INSTRUCTOR-10-е из 10: 10/10-{22.60000...}	-0.139	0.022	0.066	-0.050	-0.017	0.108
13	COURSE-1-е из 10: 1/10-{1.0000000, 3.5000000} ...	0.050	-0.121	0.036	-0.035	-0.012	0.095
14	COURSE-2-е из 10: 2/10-{3.5000000, 6.0000000} ...	0.035	-0.233	0.089	-0.109	-0.036	0.172
15	COURSE-3-е из 10: 3/10-{6.0000000, 8.5000000} ...	0.081	0.119		0.200	0.067	0.061
16	COURSE-4-е из 10: 4/10-{8.5000000, 11.0000000} ...	-0.013	0.079	-0.096	-0.029	-0.010	0.088
17	COURSE-5-е из 10: 5/10-{11.0000000, 13.5000000} ...		0.264		0.264	0.088	0.153
18	COURSE-6-е из 10: 6/10-{13.5000000, 16.0000000} ...	0.023	0.062	-0.113	-0.029	-0.010	0.092

Помощь MS Excel MS Word

Таблица 4 – Матрица знаний (модель INF3) (фрагмент)

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

Код признака	Наименование описательной шкалы и градации	1. CLASS ATTRIBUTE 1/3 LOW	2. CLASS ATTRIBUTE 2/3 MEDIUM	3. CLASS ATTRIBUTE 3/3 HIGH	Сумма	Среднее	Средн. квадр. откл.
1	ENGLISH-no	4.411	3.603	-8.013			6.951
2	ENGLISH-yes	-4.411	-3.603	8.013			6.951
3	COURSE INSTRUCTOR-1-е из 10: 1/10-{1.00000...}	0.755	1.689	-2.444			2.167
4	COURSE INSTRUCTOR-2-е из 10: 2/10-{3.40000...}	3.404	-2.649	-0.755			3.096
5	COURSE INSTRUCTOR-3-е из 10: 3/10-{5.80000...}	-3.788	6.053	-2.265			5.297
6	COURSE INSTRUCTOR-4-е из 10: 4/10-{8.20000...}	-2.192	-2.298	4.490			3.889
7	COURSE INSTRUCTOR-5-е из 10: 5/10-{10.60000...}	1.834	-1.291	-0.543			1.632
8	COURSE INSTRUCTOR-6-е из 10: 6/10-{13.00000...}	-0.192	-0.298	0.490			0.428
9	COURSE INSTRUCTOR-7-е из 10: 7/10-{15.40000...}	1.404	-1.649	0.245			1.541
10	COURSE INSTRUCTOR-8-е из 10: 8/10-{17.80000...}	1.457	-1.636	0.179			1.554
11	COURSE INSTRUCTOR-9-е из 10: 9/10-{20.20000...}	0.457	1.364	-1.821			1.641
12	COURSE INSTRUCTOR-10-е из 10: 10/10-{22.60000...}	-3.139	0.715	2.424			2.850
13	COURSE-1-е из 10: 1/10-{1.0000000, 3.5000000} ...	5.662	-9.834	4.172			8.549
14	COURSE-2-е из 10: 2/10-{3.5000000, 6.0000000} ...	0.404	-1.649	1.245			1.489
15	COURSE-3-е из 10: 3/10-{6.0000000, 8.5000000} ...	1.430	2.358	-3.788			3.313
16	COURSE-4-е из 10: 4/10-{8.5000000, 11.0000000} ...	-0.219	1.695	-1.477			1.597
17	COURSE-5-е из 10: 5/10-{11.0000000, 13.5000000} ...	-1.298	2.675	-1.377			2.317
18	COURSE-6-е из 10: 6/10-{13.5000000, 16.0000000} ...	0.457	1.364	-1.821			1.641

Помощь MS Excel MS Word

### 1.5. Результаты верификации моделей

Результаты верификации (оценки достоверности) моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями приведены на рисунке 9.

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 истинно-положительных решений (ST...)	Сумма модулей истинно-отрицательных решений (ST...)	Сумма модулей ложно-положительных решений (SFP)	Сумма модулей ложно-отрицательных решений (SF...)	S-точность модели
1. ABS - частный критерий: количество встреч сочетаний "класс"	Корреляция абс частот с обр...	151	151	1	301		0.334	1.000	0.501	104.083	0.038	179.928		
1. ABS - частный критерий: количество встреч сочетаний "класс"	Сумма абс частот по признак...	151	151		302		0.333	1.000	0.500	115.355		209.754		
2. PRC1 - частный критерий: усл. вероятность иго признака сред.	Корреляция усл.отн частот с о...	151	151	1	301		0.334	1.000	0.501	104.083	0.038	179.928		
2. PRC1 - частный критерий: усл. вероятность иго признака сред.	Сумма усл.отн частот по при...	151	151		302		0.333	1.000	0.500	112.532		204.459		
3. PRC2 - частный критерий: условная вероятность иго признака.	Корреляция усл.отн частот с о...	151	151	1	301		0.334	1.000	0.501	104.085	0.038	179.931		
3. PRC2 - частный критерий: условная вероятность иго признака.	Сумма усл.отн частот по при...	151	151		302		0.333	1.000	0.500	112.532		204.459		
4. INF1 - частный критерий: количество знаний по А\Харкевичу. в.	Семантический резонанс зна...	151	115	218	84	36	0.578	0.762	0.657	47.202	86.633	18.678	8.403	
4. INF1 - частный критерий: количество знаний по А\Харкевичу. в.	Сумма знаний	151	106	233	69	45	0.606	0.702	0.650	37.817	86.935	13.094	9.215	
5. INF2 - частный критерий: количество знаний по А\Харкевичу. в.	Семантический резонанс зна...	151	115	218	84	36	0.578	0.762	0.657	47.202	86.633	18.678	8.403	
5. INF2 - частный критерий: количество знаний по А\Харкевичу. в.	Сумма знаний	151	106	233	69	45	0.606	0.702	0.650	37.817	86.935	13.094	9.215	
6. INF3 - частный критерий: Хинквадрат, разности между фактил...	Семантический резонанс зна...	151	114	173	129	37	0.469	0.755	0.579	54.078	69.673	40.487	12.226	
6. INF3 - частный критерий: Хинквадрат, разности между фактил...	Сумма знаний	151	114	173	129	37	0.469	0.755	0.579	46.336	67.237	33.072	12.171	
7. INF4 - частный критерий: ROI (Return On Investment); веротно...	Семантический резонанс зна...	151	95	255	47	56	0.669	0.629	0.648	34.556	102.548	8.336	13.541	
7. INF4 - частный критерий: ROI (Return On Investment); веротно...	Сумма знаний	151	126	198	104	25	0.548	0.834	0.661	46.244	42.967	18.119	3.777	
8. INF5 - частный критерий: ROI (Return On Investment); веротно...	Семантический резонанс зна...	151	95	255	47	56	0.669	0.629	0.648	34.556	102.548	8.336	13.541	
8. INF5 - частный критерий: ROI (Return On Investment); веротно...	Сумма знаний	151	126	198	104	25	0.548	0.834	0.661	46.244	42.967	18.119	3.777	
9. INF6 - частный критерий: разн.усли безул.вероятностей; вер...	Семантический резонанс зна...	151	111	189	113	40	0.496	0.735	0.592	51.707	76.267	38.592	14.174	
9. INF6 - частный критерий: разн.усли безул.вероятностей; вер...	Сумма знаний	151	114	170	132	37	0.463	0.755	0.574	48.228	64.808	36.168	12.258	
10. INF7 - частный критерий: разн.усли безул.вероятностей; ве...	Семантический резонанс зна...	151	111	189	113	40	0.496	0.735	0.592	51.707	76.267	38.592	14.174	
10. INF7 - частный критерий: разн.усли безул.вероятностей; ве...	Сумма знаний	151	114	170	132	37	0.463	0.755	0.574	48.228	64.808	36.168	12.258	

а

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 истинно-положительных решений (ST...)	Сумма модулей истинно-отрицательных решений (ST...)	Сумма модулей ложно-положительных решений (SFP)	Сумма модулей ложно-отрицательных решений (SF...)	S-точность модели	S-Полнота модели	L-мера проф. Е.В.Луценко	Процент правильной идентификации...	Процент ошибочной идентификации...	Процент правильной идентификации...	Процент ошибочной идентификации...	Процент правильного результата	Дата получения результата	Время получения результ...
1	301		0.334	1.000	0.501	104.083	0.038	179.928		0.366	1.000	0.536	100.000	0.328	99.672		50.164	10.12.2016	21:18:34
	302		0.333	1.000	0.500	115.355		209.754		0.355	1.000	0.524	100.000		100.000		50.000	10.12.2016	21:18:34
1	301		0.334	1.000	0.501	104.083	0.038	179.928		0.366	1.000	0.536	100.000	0.328	99.672		50.164	10.12.2016	21:18:37
	302		0.333	1.000	0.500	112.532		204.459		0.355	1.000	0.524	100.000		100.000		50.000	10.12.2016	21:18:37
1	301		0.334	1.000	0.501	104.085	0.038	179.931		0.366	1.000	0.536	100.000	0.328	99.672		50.164	10.12.2016	21:18:40
	302		0.333	1.000	0.500	112.532		204.459		0.355	1.000	0.524	100.000		100.000		50.000	10.12.2016	21:18:40
218	84	36	0.578	0.762	0.657	47.202	86.633	18.678	8.403	0.716	0.849	0.777	76.159	72.399	27.601	23.841	74.279	10.12.2016	21:18:43
233	69	45	0.606	0.702	0.650	37.817	86.935	13.094	9.215	0.743	0.804	0.772	70.199	77.290	22.710	29.801	73.744	10.12.2016	21:18:43
218	84	36	0.578	0.762	0.657	47.202	86.633	18.678	8.403	0.716	0.849	0.777	76.159	72.399	27.601	23.841	74.279	10.12.2016	21:18:46
233	69	45	0.606	0.702	0.650	37.817	86.935	13.094	9.215	0.743	0.804	0.772	70.199	77.290	22.710	29.801	73.744	10.12.2016	21:18:46
173	129	37	0.469	0.755	0.579	54.078	69.673	40.487	12.226	0.572	0.816	0.672	75.497	58.363	41.637	24.503	66.930	10.12.2016	21:18:49
173	129	37	0.469	0.755	0.579	46.336	67.237	33.072	12.171	0.584	0.792	0.672	75.497	58.363	41.637	24.503	66.930	10.12.2016	21:18:49
255	47	56	0.669	0.629	0.648	34.556	102.548	8.336	13.541	0.806	0.718	0.760	62.914	84.565	15.435	37.086	73.739	10.12.2016	21:18:52
198	104	25	0.548	0.834	0.661	46.244	42.967	18.119	3.777	0.718	0.924	0.809	83.444	65.752	34.248	16.556	74.595	10.12.2016	21:18:52
255	47	56	0.669	0.629	0.648	34.556	102.548	8.336	13.541	0.806	0.718	0.760	62.914	84.565	15.435	37.086	73.739	10.12.2016	21:18:55
198	104	25	0.548	0.834	0.661	46.244	42.967	18.119	3.777	0.718	0.924	0.809	83.444	65.752	34.248	16.556	74.595	10.12.2016	21:18:55
189	113	40	0.496	0.735	0.592	51.707	76.267	38.592	14.174	0.573	0.785	0.662	73.510	63.561	36.439	26.490	68.536	10.12.2016	21:18:58
170	132	37	0.463	0.755	0.574	48.228	64.808	36.168	12.258	0.571	0.797	0.666	75.497	57.369	42.631	24.503	66.433	10.12.2016	21:18:58
189	113	40	0.496	0.735	0.592	51.707	76.267	38.592	14.174	0.573	0.785	0.662	73.510	63.561	36.439	26.490	68.536	10.12.2016	21:19:01
170	132	37	0.463	0.755	0.574	48.228	64.808	36.168	12.258	0.571	0.797	0.666	75.497	57.369	42.631	24.503	66.433	10.12.2016	21:19:01

б

Рисунок 9. Оценки достоверности моделей помощью F-меры Ван Ризбергера (а) и L-меры проф.Е.В.Луценко (б)

Наиболее достоверной в данном приложении оказались модели INF4 и INF5 при интегральном критерии «Семантический резонанс знаний». При этом достоверность модели по F-критерию Ван Ризбергена составляет 0,661. Таким образом, уровень достоверности прогнозирования с применением модели примерно такой же, как экспертных оценках, достоверность которых считается равной примерно 70%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используются не только F-мера Ван Ризбергена, но и ее нечеткое мультиклассовое обобщение, предложенное проф.Е.В.Луценко (L-мера) [11] (рисунок 10).

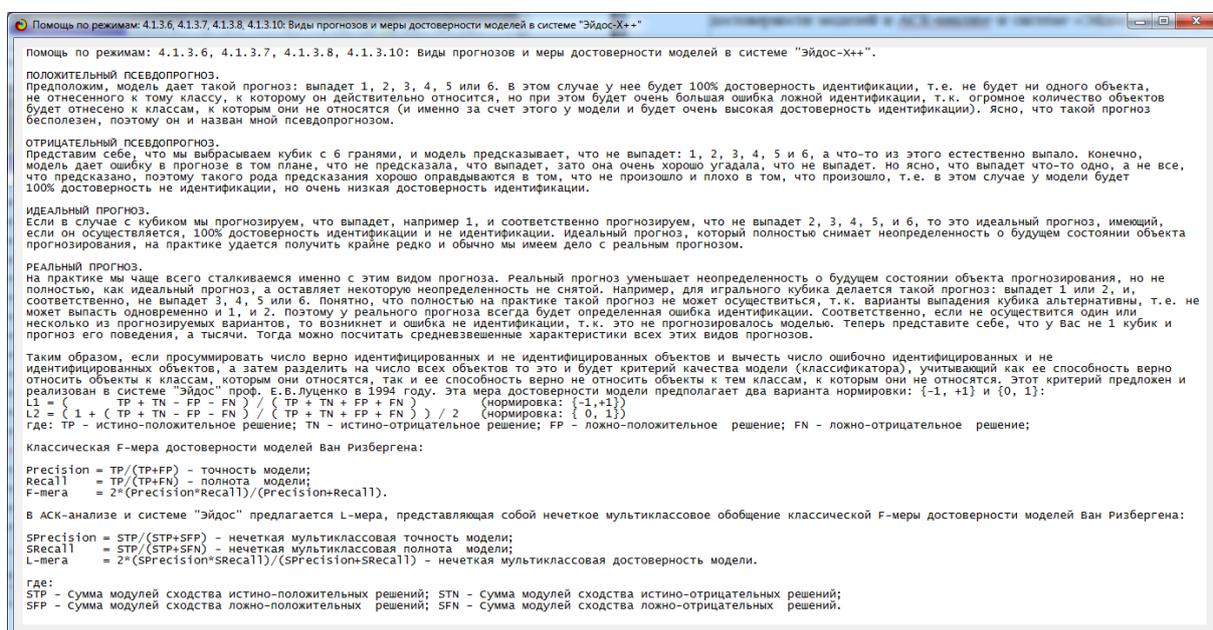


Рисунок 10. Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

Также обращает на себя внимание, что статистические модели, как правило, дают 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 11 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и неидентифицированных ситуаций в наиболее достоверной модели INF4.

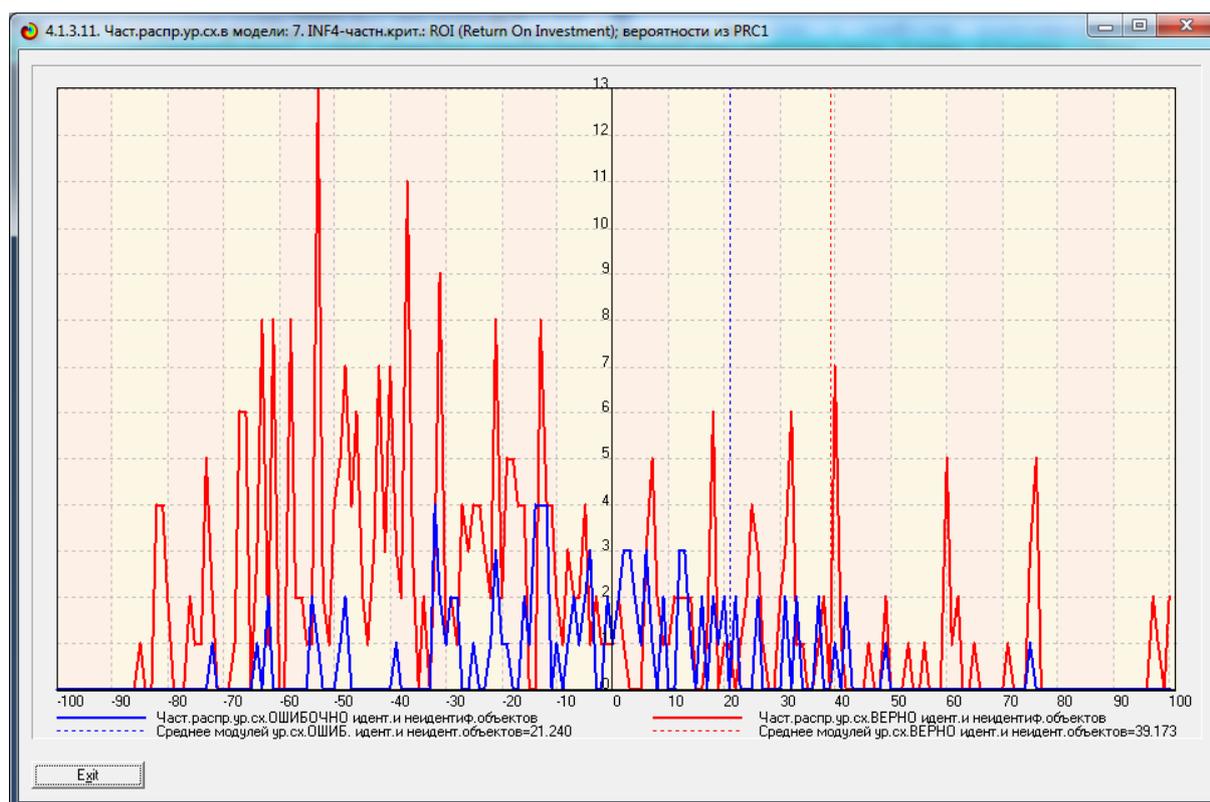


Рисунок 11. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF4

Из рисунка 11 видно, что:

- наиболее достоверная модель INF4 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 9);

- модуль уровня сходства-различия в наиболее достоверной модели INF4 для верно идентифицированных и верно неидентифицированных объектов значительно выше, чем для ошибочно идентифицированных и ошибочно неидентифицированных. Это верно практически для всего диапазона уровней сходства-различия, кроме небольших по модулю значений в диапазоне от 0 до 15% уровня сходства. Для больших значений уровней сходства-различия (более 50%) различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями на столько велико, что учитывая уровень сходства-различия не представляет никакой

проблемы разделить истинные положительные и отрицательные решения от ложных положительных и отрицательных решений. Это и стало основой для нечеткого обобщения F-меры Ван Ризбергена, т.е. такого ее обобщения, которое учитывает не только сам факт ложно или истинно положительного или отрицательного решения, но и его надежность [11].

Любые данные о наблюдениях можно считать суммой истинного значения и шума, причем ни первое, ни второе неизвестны. Поэтому имеет смысл сравнить созданные модели с чисто случайными моделями, совпадающими по основным характеристикам. В системе «Эйдос» есть лабораторная работа № 2.01: «Исследование RND-модели при различных объемах выборки». Если данная работа устанавливается при отсутствии текущего приложения, то все параметры создаваемых моделей задаются вручную, если же текущая модель существует, как в нашем случае, то все основные ее параметры определяются автоматически.

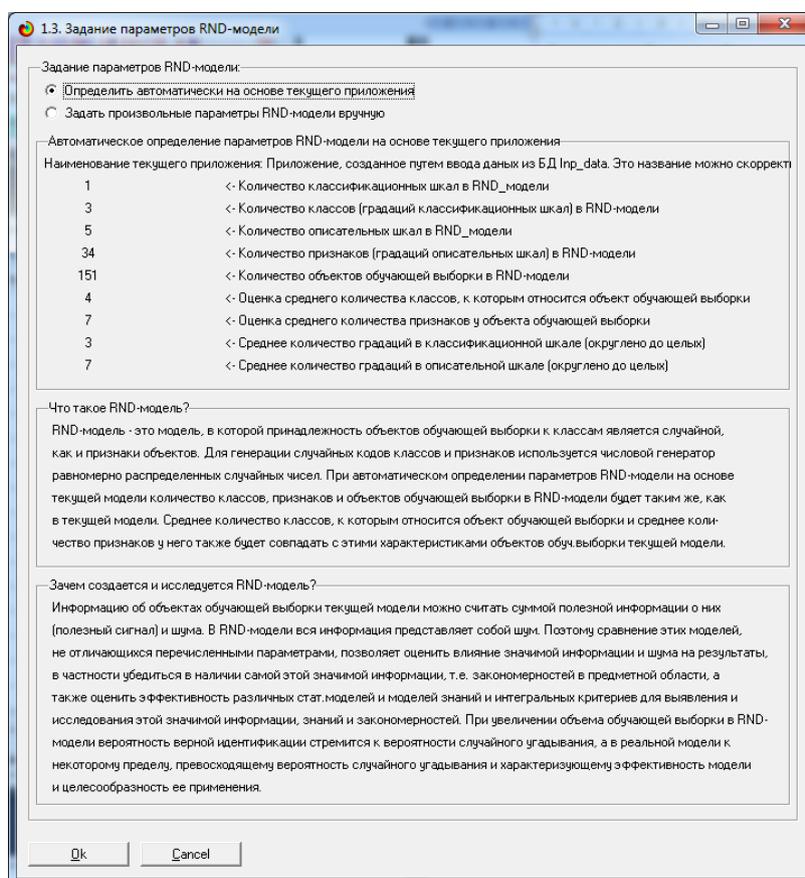


Рисунок 12. Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

На рисунке 13 показано частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF4.

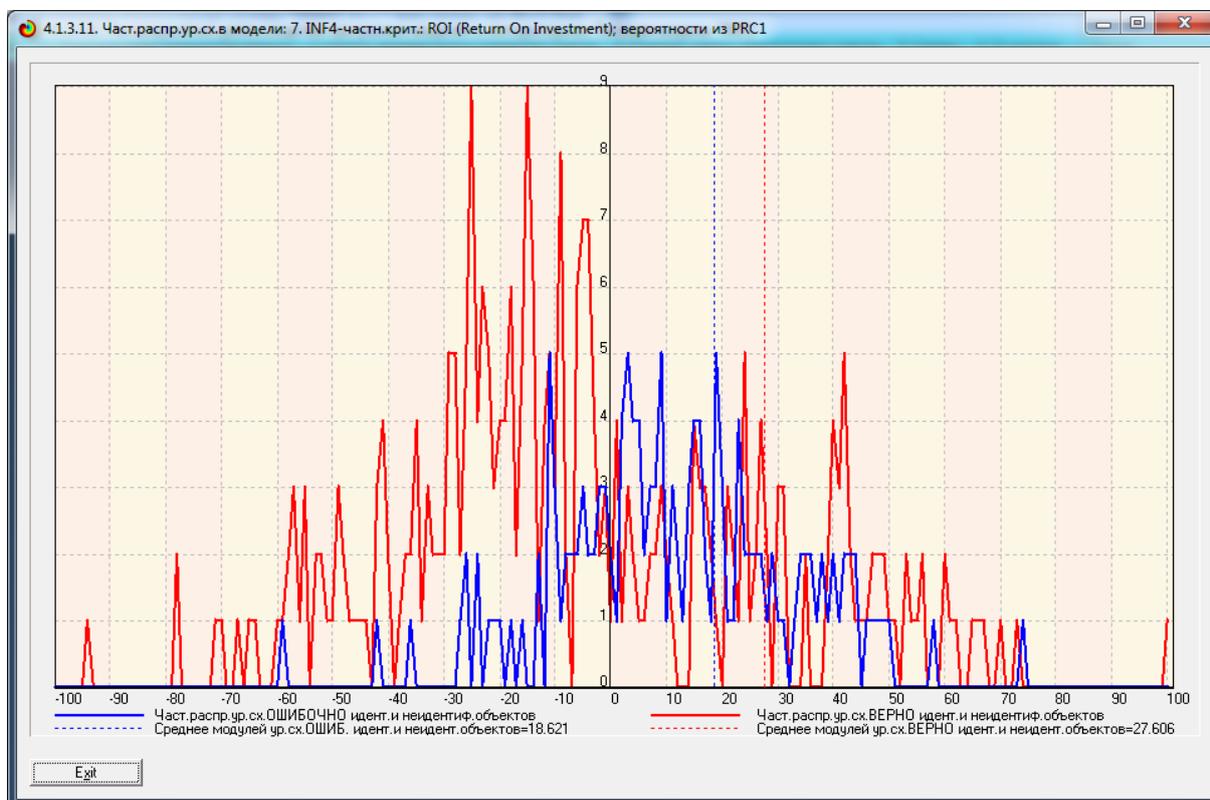


Рисунок 13. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF4

Совершенно очевидное различие частотных распределений уровней сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования и случайной модели (рисунки 11 и 13) объясняется тем, что в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны. Если же такой информации в модели нет, то и распределение получается типа, приведенного на рисунке 13.

На рисунке 14 приведены данные по достоверности статистических и когнитивных моделей, созданных на основе случайной выборки.

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризбергена	Средня мод. истинно-полож. решений (ST)	Средня мод. истинно-отриц. решений (ST)	Средня мод. ложно-полож. решений (SFP)	Средня мод. ложно-отриц. решений (SFN)	S-T мод.
1. ABS - частный критерий: количество встреч сочетаний "клас...	Корреляция абс частот с абс...	151	121	163	139	30	0.465	0.801	0.589	42.030	35.115	35.357	5.405	
2. PRC1 - частный критерий: усл. вероятность его признака сред...	Средня абс частот по признак...	151	151		302		0.333	1.000	0.500	90.571		158.444		
3. PRC2 - частный критерий: усл. вероятность его признака сред...	Корреляция усл.отн частот с о...	151	121	163	139	30	0.465	0.801	0.589	42.030	35.115	35.357	5.405	
4. INF1 - частный критерий: количество знаний по А\Заркевичу; в...	Средня усл.отн частот по при...	151	151		302		0.333	1.000	0.500	88.768		155.395		
5. INF2 - частный критерий: усл. вероятность его признака...	Корреляция усл.отн частот с о...	151	121	163	139	30	0.465	0.801	0.589	42.028	35.114	35.355	5.405	
6. INF3 - частный критерий: X\квадрат; разности между факти...	Средня абс частот по признак...	151	151		302		0.333	1.000	0.500	91.717		160.638		
7. INF4 - частный критерий: ROI (Return On Investment); веро...	Семантический резонанс зна...	151	116	183	119	35	0.494	0.768	0.601	29.473	44.834	21.200	4.574	
8. INF5 - частный критерий: X\квадрат; разности между факти...	Средня знаний	151	96	218	84	55	0.533	0.636	0.580	12.148	37.768	7.509	5.156	
9. INF6 - частный критерий: разн усл и без усл вероятностей; вер...	Семантический резонанс зна...	151	116	183	119	35	0.494	0.768	0.601	29.473	44.834	21.200	4.574	
10. INF7 - частный критерий: разн усл и без усл вероятностей; ве...	Средня знаний	151	96	217	85	55	0.530	0.636	0.578	12.314	37.990	7.667	5.251	
11. INF8 - частный критерий: разн усл и без усл вероятностей; вер...	Семантический резонанс зна...	151	107	194	108	44	0.498	0.709	0.585	35.736	54.214	24.262	5.887	
12. INF9 - частный критерий: разн усл и без усл вероятностей; ве...	Средня знаний	151	107	194	108	44	0.498	0.709	0.585	30.526	46.783	21.318	5.061	
13. INF10 - частный критерий: разн усл и без усл вероятностей; ве...	Семантический резонанс зна...	151	108	194	108	43	0.500	0.715	0.589	33.313	50.057	22.510	5.608	
14. INF11 - частный критерий: разн усл и без усл вероятностей; ве...	Средня знаний	151	111	191	111	40	0.500	0.735	0.595	31.090	47.144	21.277	5.222	
15. INF12 - частный критерий: разн усл и без усл вероятностей; ве...	Семантический резонанс зна...	151	108	194	108	43	0.500	0.715	0.589	33.313	50.057	22.510	5.608	
16. INF13 - частный критерий: разн усл и без усл вероятностей; ве...	Средня знаний	151	113	192	110	38	0.507	0.748	0.604	31.768	47.272	21.645	5.695	
17. INF14 - частный критерий: разн усл и без усл вероятностей; ве...	Семантический резонанс зна...	151	107	194	108	44	0.498	0.709	0.585	35.736	54.214	24.262	5.887	
18. INF15 - частный критерий: разн усл и без усл вероятностей; ве...	Средня знаний	151	107	194	108	44	0.498	0.709	0.585	29.978	46.047	20.930	4.933	
19. INF16 - частный критерий: разн усл и без усл вероятностей; ве...	Семантический резонанс зна...	151	107	191	111	44	0.491	0.709	0.580	35.746	54.168	24.263	5.910	
20. INF17 - частный критерий: разн усл и без усл вероятностей; ве...	Средня знаний	151	109	195	107	42	0.505	0.722	0.594	33.920	51.484	23.534	5.818	

Рисунок 14. Достоверность статистических и когнитивных моделей, созданных на основе случайной выборки

На основе сравнения рисунка 14 с рисунком 9 можно сделать следующие выводы:

- достоверность лучшей модели INF4, отражающей реальный объект моделирования, примерно на 12% выше, чем аналогичной случайной модели;

- различие между достоверностью статистических моделей и моделей знаний, созданных на основе случайной выборки, значительно меньше, чем у моделей, отражающих реальный объект моделирования;

- в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны, причем примерно 1/3 достоверности обусловлена отражением в реальных моделях закономерностей предметной области, а 2/3 достоверности обусловлено наличием шума в исходных данных. На основании этого можно предположить, что в исходных данных уровень сигнала о реальных причинно-следственных

связях в моделируемой предметной области примерно в два раза ниже уровня шума.

## 2. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

### 2.1. Решение задачи идентификации

В соответствии с технологией АСК-анализа зададим текущей модель INF4 (режим 5.6) (рисунок 15) и проведем пакетное распознавание в режиме 4.2.1.

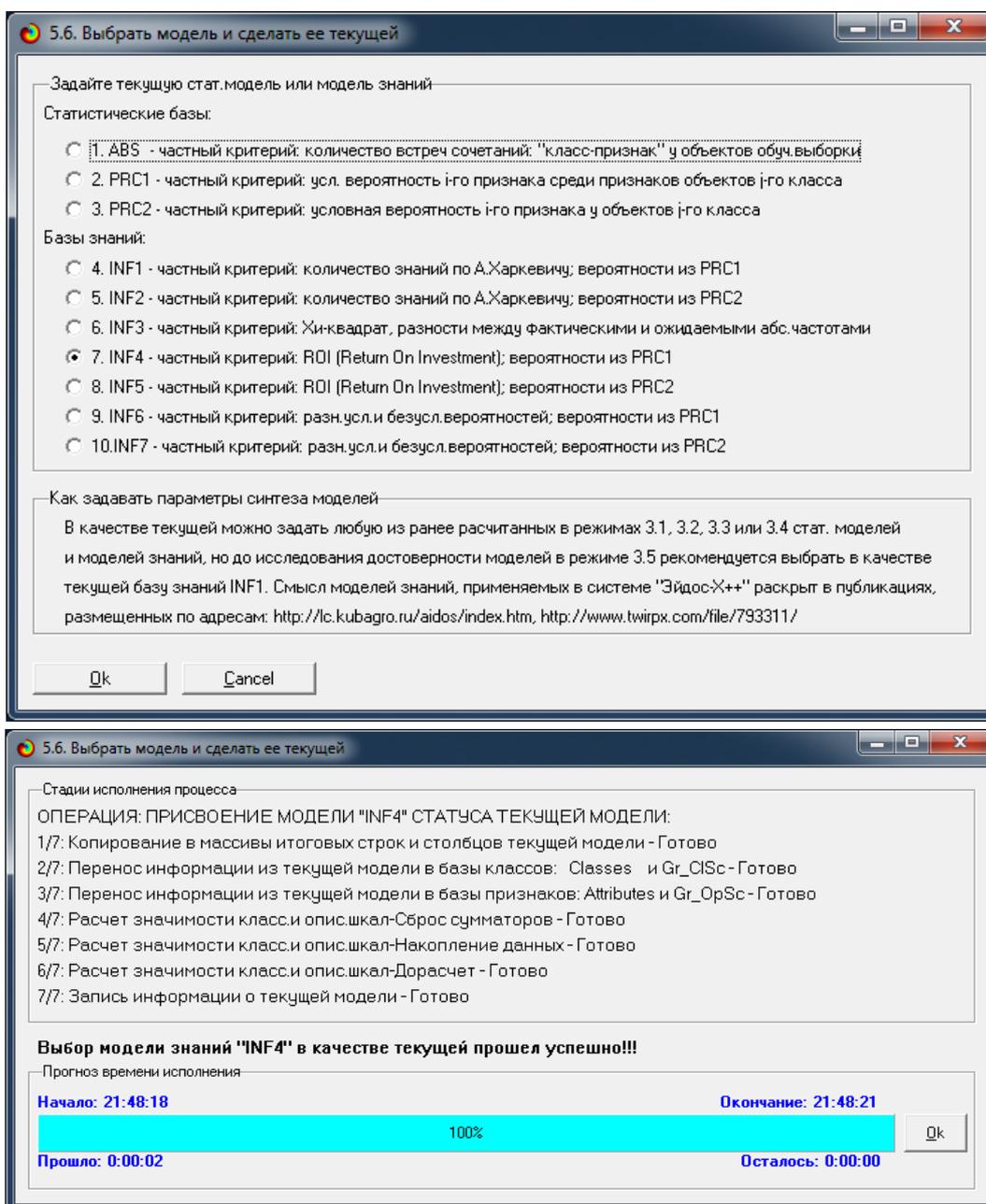


Рисунок 15. Экранные формы режима задания модели в качестве текущей

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

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 приведен пример прогнозов достоверности частоты и классов преподавателей в наиболее достоверной модели INF4 на основе наблюдения предыстории их развития:

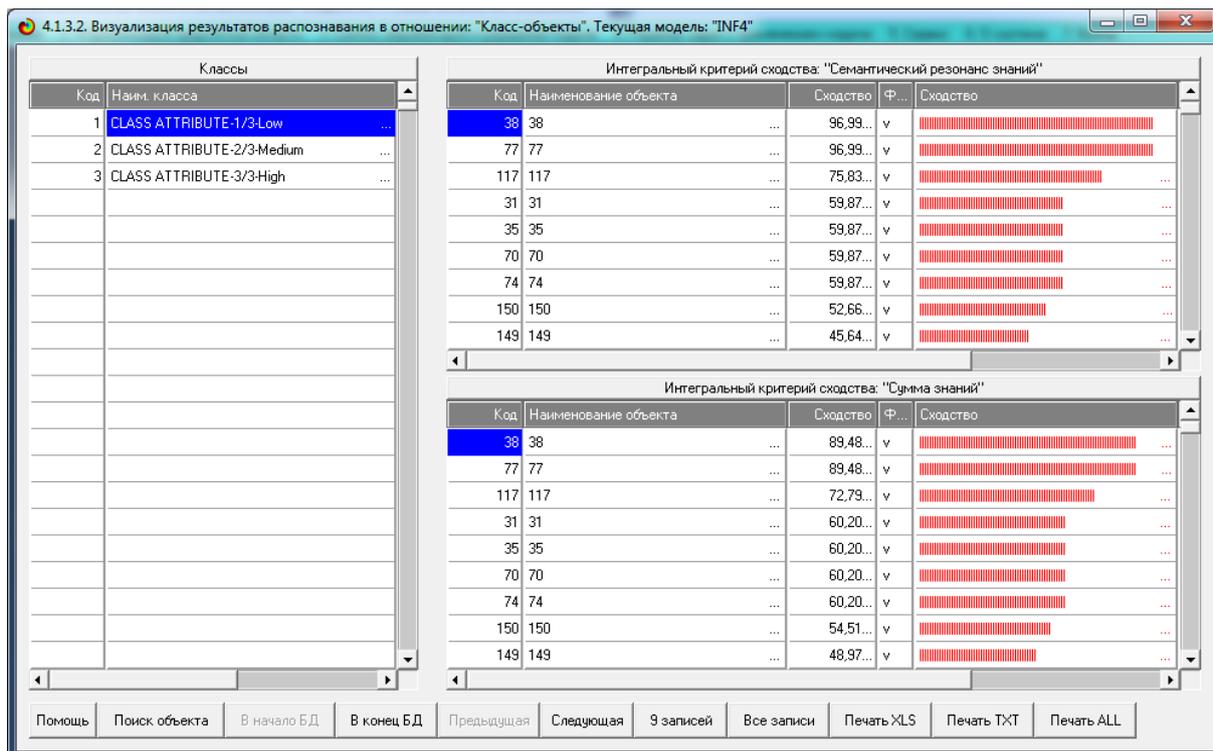
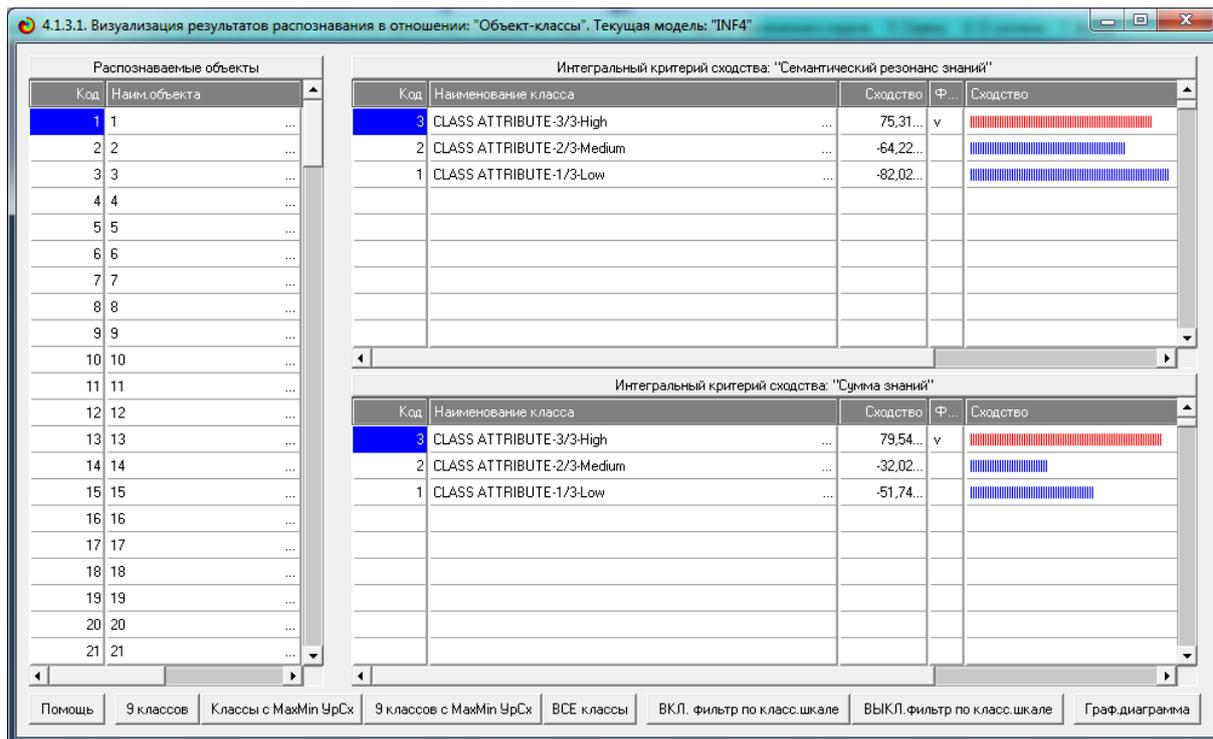


Рисунок 16. Пример идентификации классов преподавания в модели INF4

## 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 17)

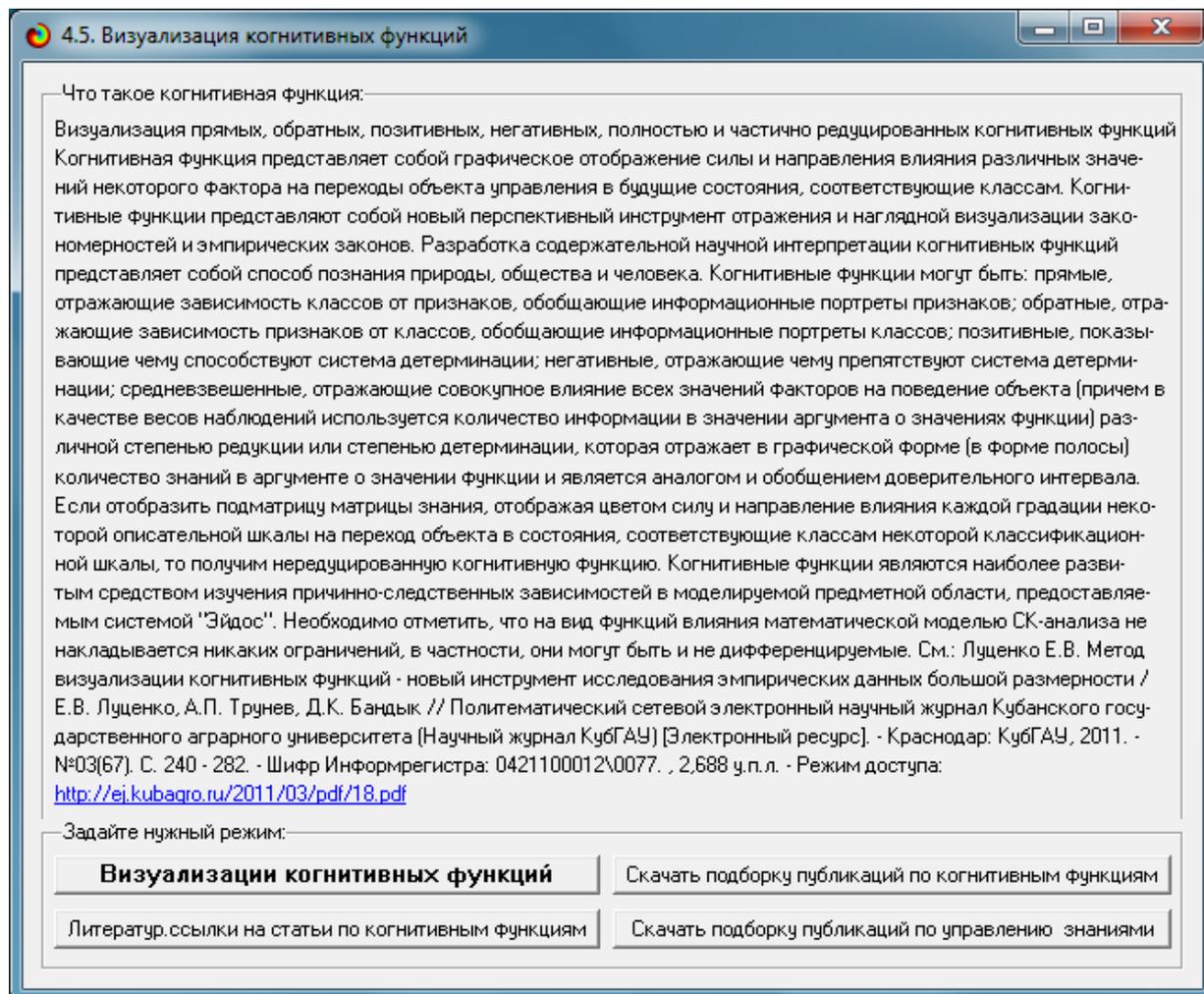


Рисунок 17. Экранная форма режима 4.5 системы «Эйдос-X++» «Визуализация когнитивных функций»

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. Поэтому здесь не будем останавливаться на описании того, что представляют собой когнитивные

функции в АСК-анализе. На рисунке 18 приведены визуализации двух когнитивных функций данного приложения для модели INF4.

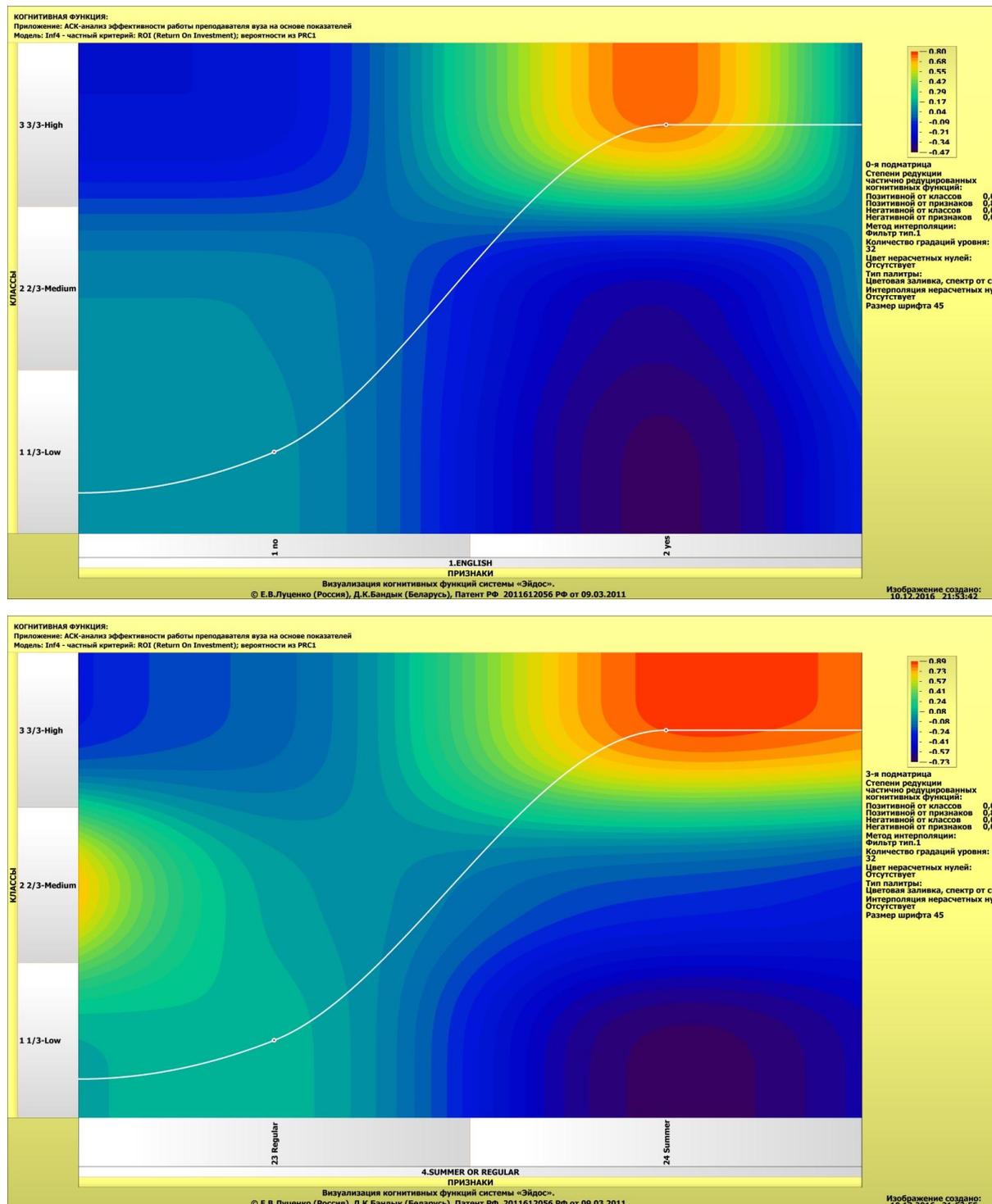


Рисунок 18. Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF4

### 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 [12] (рисунок 19).

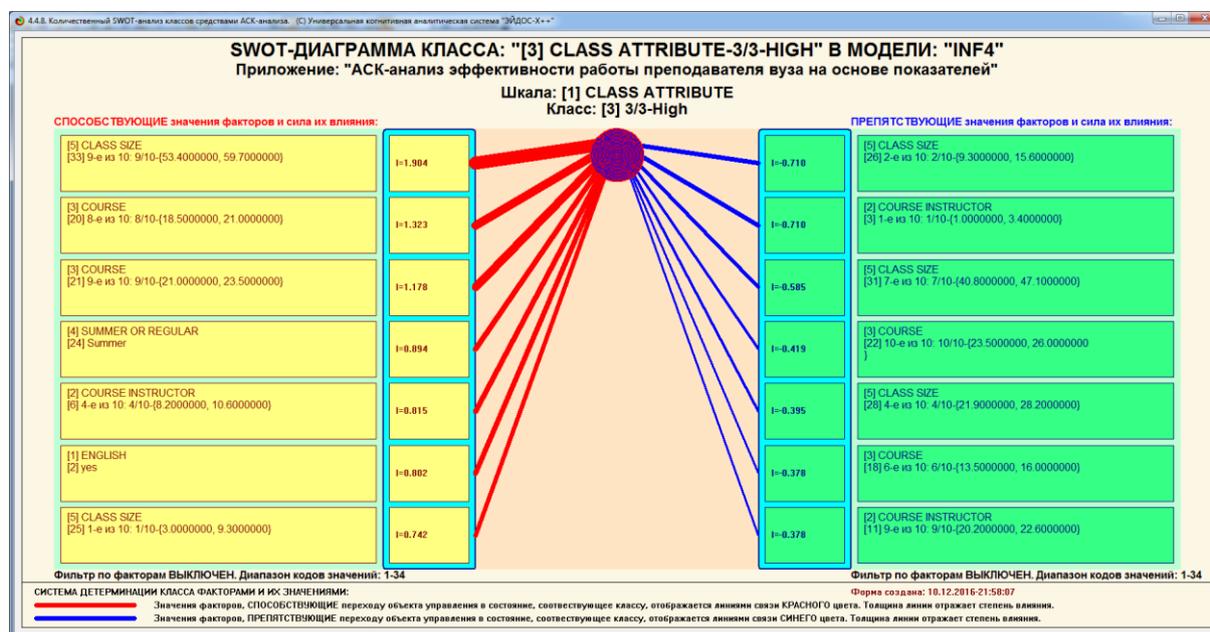
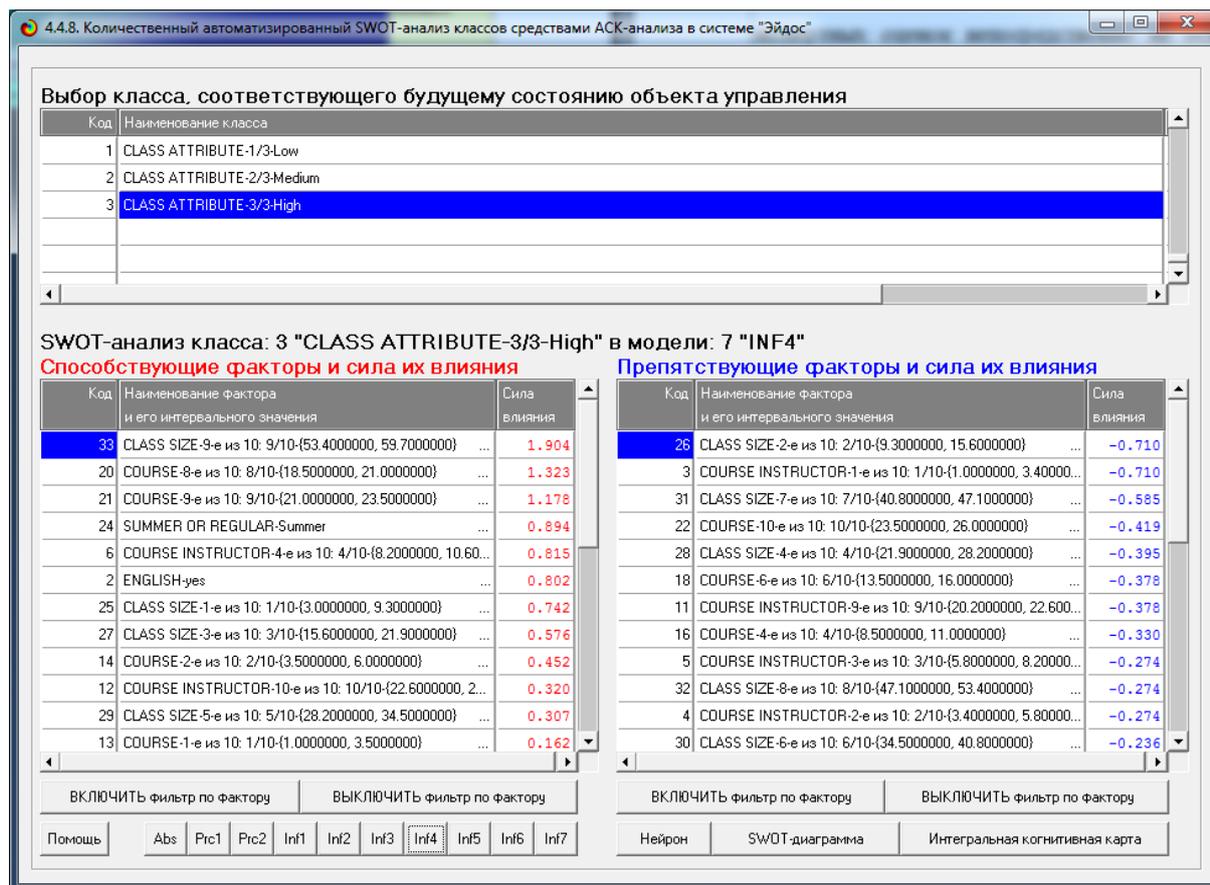


Рисунок 19. Пример SWOT-матрицы в модели INF4

На рисунке 20 приведены примеры инвертированной SWOT-матрицы и инвертированной SWOT-диаграммы в модели INF4.

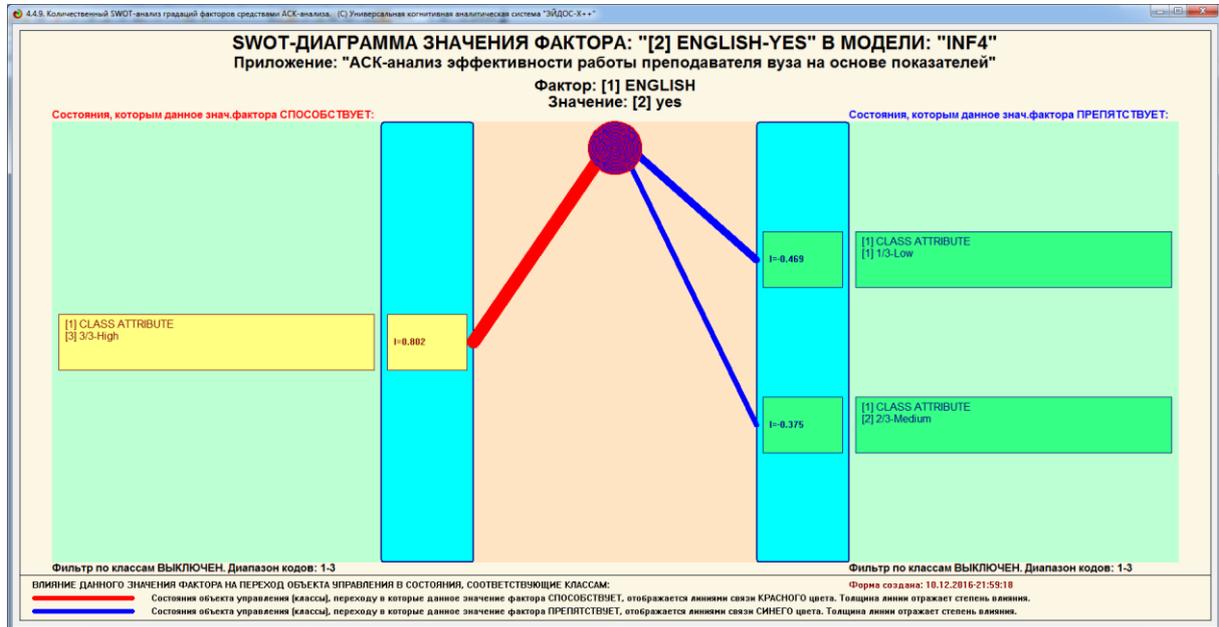
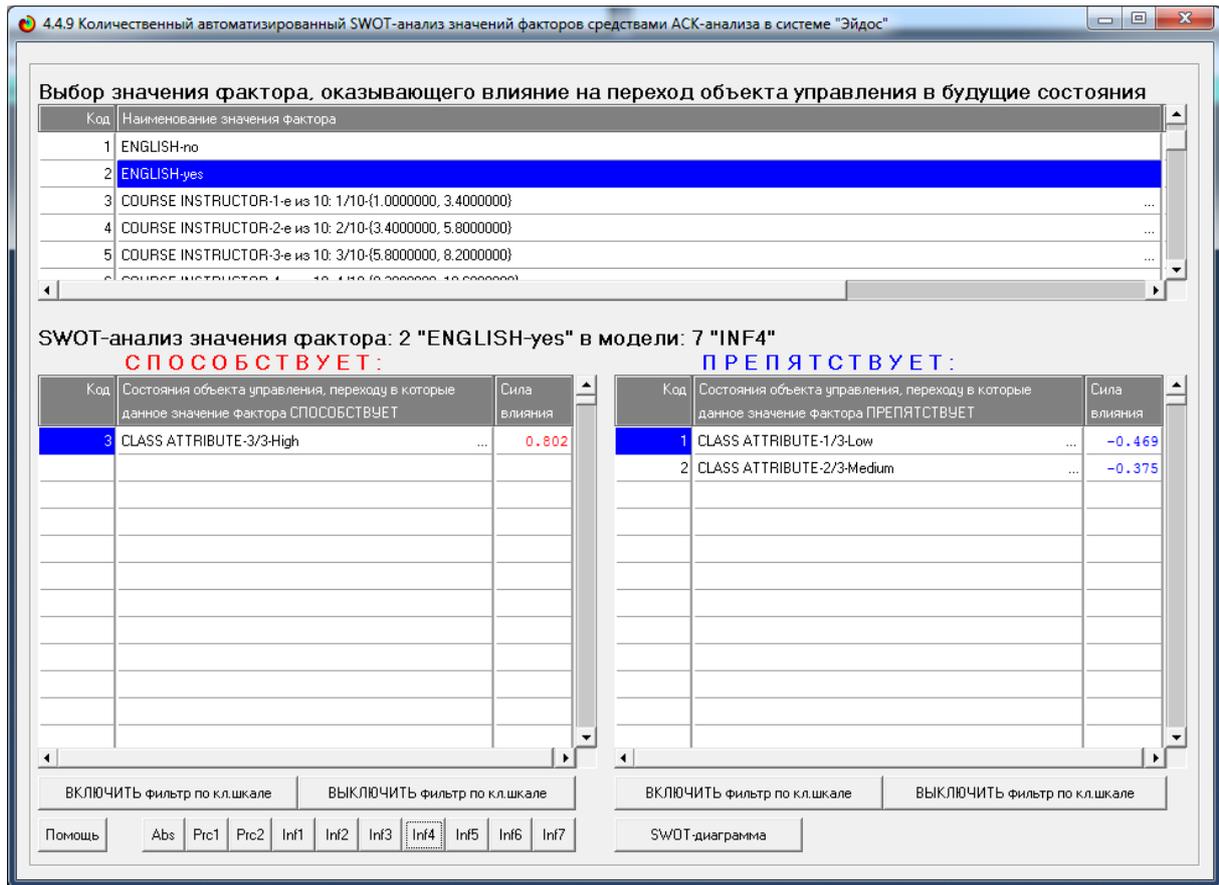


Рисунок 20. Пример SWOT-матрицы в модели INF4

## 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 общедоступную базу репозитория UCI. В данной курсовой работе приводится развернутый пример использования базы данных репозитория 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 оказались модели INF4, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели по L-критерию [11] составляет 0,809, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 70%. Также обращает на себя внимание, что статистические модели в данном приложении дают примерно на 21% более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, что, как правило, наблюдается и в других приложениях. Этим и оправдано применение моделей знаний.

На основе базы данных UCI, рассмотренной в данной курсовой 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.

### 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. Репозиторий UCI [Электронный ресурс]. Статья " Teaching Assistant Evaluation Data Set": <http://archive.ics.uci.edu/ml/machine-learning-databases/>, свободный. - Загл. с экрана. Яз. англ.

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

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

8. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 328 – 356. – IDA [article ID]: 0831209025. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/25.pdf>, 1,812 у.п.л.

9. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.

10. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-Х++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой

электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 у.п.л.

11. Луценко Е.В. Нечеткое мультиклассовое обобщение классической F-меры достоверности моделей Ван Ризбергера в АСК-анализе и системе «Эйдос» / Луценко Е.В. // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №09(123). – Режим доступа: <http://ej.kubagro.ru/2016/09/pdf/01.pdf>, 1,813 у.п.л. – IDA [article ID]: 1231609001. <http://dx.doi.org/10.21515/1990-4665-123-001>

12. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.

### References

1. Lucenko E.V. Metodika ispol'zovaniya repozitorija UCI dlja ocenki kachestva matematicheskikh modelej sistem iskusstvennogo intellekta / E.V. Lucenko // Politematicheskij setevoy jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2003. – №02(002). S. 120 – 145. – IDA [article ID]: 0020302012. – Rezhim dostupa: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 u.p.l.

2. Lucenko E.V. ASK-analiz, modelirovanie i identifikacija zhivyh sushhestv na osnove ih fenotipicheskikh priznakov / E.V. Lucenko, Ju.N. Penkina // Politematicheskij setevoy jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: Kub- GAU, 2014. – №06(100). S. 1346 – 1395. – IDA [article ID]: 1001406090. – Rezhim dostupa: <http://ej.kubagro.ru/2014/06/pdf/90.pdf>, 3,125 u.p.l.

3. Lucenko E.V. Teoreticheskie osnovy, tehnologija i instrumentarij avtomatizirovannogo sistemno-kognitivnogo analiza i vozmozhnosti ego primeneniya dlja sopostavimoj ocenki jeffektivnosti vuzov / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setevoy jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarno- go universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: Kub- GAU, 2013. – №04(088). S. 340 – 359. – IDA [article ID]: 0881304022. – Rezhim dostupa: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 u.p.l.

4. Lucenko E.V. Avtomatizirovannyj sistemno-kognitivnyj analiz v upravlenii aktivnymi ob#ektami (sistemnaja teorija informacii i ee primeneniye v issledovanii jekonomicheskikh, social'no-psihologicheskikh, tehnologicheskikh i organizacionno-tehnicheskikh sistem): Monografija (nauchnoe izdanie). – Krasnodar: KubGAU. 2002. – 605 s.

5. Repozitorij UCI [Jelektronnyj resurs]. Stat'ja " Teaching Assistant Evaluation Data Set": <http://archive.ics.uci.edu/ml/machine-learning-databases/>, svobodnyj. - Zagl. s jekrana. Jaz. ang.

6. Sajt professora E.V.Lucenko [Jelektronnyj resurs]. Rezhim dostupa: <http://lc.kubagro.ru/>, svobodnyj. - Zagl. s jekrana. Jaz. rus.

7. Lucenko E.V. 30 let sisteme «Jejdos» – odnoj iz starejsih otechestvennyh universal'nyh sistem iskusstvennogo intellekta, shiroko primenjaemyh i razvivajushhihsja i v nastojashhee vremja / E.V. Lucenko // Politematicheskij setevoy jelektronnyj nauchnyj

zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2009. – №10(054). S. 48 – 77. – Shifr Informregistra: 0420900012\0110, IDA [article ID]: 0540910004. – Rezhim dostupa: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>, 1,875 u.p.l.

8. Lucenko E.V. Universal'naja kognitivnaja analiticheskaja sistema «Jejdos- H++» / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2012. – №09(083). S. 328 – 356. – IDA [article ID]: 0831209025. – Rezhim dostupa: <http://ej.kubagro.ru/2012/09/pdf/25.pdf>, 1,812 u.p.l.

9. Lucenko E.V. Metrizacija izmeritel'nyh shkal razlichnyh tipov i sovmestnaja sopostavimaja kolichestvennaja obrabotka raznorodnyh faktorov v sistemno-kognitivnom analize i sisteme «Jejdos» / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №08(092). S. 859 – 883. – IDA [article ID]: 0921308058. – Rezhim dostupa: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 u.p.l.

10. Lucenko E.V. Modelirovanie slozhnyh mnogofaktornyh nelinejnyh ob#ektov upravlenija na osnove fragmentirovannyh zashumlennyh jempiricheskikh dannyh bol'shoj razmernosti v sistemno-kognitivnom analize i intellektual'noj sisteme «Jejdos-H++» / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №07(091). S. 164 – 188. – IDA [article ID]: 0911307012. – Rezhim dostupa: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 u.p.l.

11. Lucenko E.V. Nechetkoe mul'tiklassovoe obobshhenie klassicheskoj F-mery dostovernosti modelej Van Rizbergena v ASK-analize i sisteme «Jejdos» / Lucenko E.V. // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2016. – №09(123). – Rezhim dostupa: <http://ej.kubagro.ru/2016/09/pdf/01.pdf>, 1,813 u.p.l. – IDA [article ID]: 1231609001. <http://dx.doi.org/10.21515/1990-4665-123-001>

12. Lucenko E.V. Kolichestvennyj avtomatizirovannyj SWOT- i PEST-analiz sredstvami ASK-analiza i intellektual'noj sistemy «Jejdos-H++» / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №07(101). S. 1367 – 1409. – IDA [article ID]: 1011407090. – Rezhim dostupa: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 u.p.l.