

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ
ФЕДЕРАЦИИ

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

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

**ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
к курсовой работе**

по дисциплине: ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ
СИСТЕМЫ

на тему:

**АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ
ДАННЫХ О СМЕРТЯХ ЗНАМЕНИТОСТЕЙ НА ОСНОВЕ ДАННЫХ
РЕПОЗИТОРИЯ UCI**

выполнил студент группы ИТ1421 Ильин Владислав Сергеевич

Допущена к защите _____

Руководитель проекта _____ Луценко Евгений Вениаминович,
д.э.н., к.т.н., профессор
(подпись, расшифровка подписи)

Нормоконтролер _____ Николаева Ирина Валентиновна, к.т.н.,
доцент

(подпись, расшифровка подписи)

Защищена _____

(дата)

Оценка

Члены комиссии _____ В.И. Лойко

Е.В. Луценко

И. В. Николаева
(подпись, дата, расшифровка подписи)

Краснодар

2017 г.

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

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

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

УТВЕРЖДАЮ:

Зав. кафедрой КТС _____ В. И. Лойко

**ЗАДАНИЕ
на курсовую работу**

Студенту: ИТ1421 группы 3 курса

Факультета прикладной информатики

Специальности: 09.03.02 Информационные системы и технологии
(шифр)

Ильин Владислав Сергеевич
(Ф.И.О.)

Тема проекта: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ
АНАЛИЗ ДАННЫХ О СМЕРТЯХ ЗНАМЕНИОСТЕЙ НА ОСНОВЕ ДАННЫХ
РЕПОЗИТОРИЯ UCI

Содержание задания: Проанализировать методы формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами, принятия решений и исследования моделируемой предметной области путем исследования ее модели

Объем работы:

- а) пояснительная записка к работе _____ листа формата А4
б) графическая часть _____ лист формата А4

Рекомендуемая литература: Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для бакалавриата. 7-е изд., перераб. и доп.- Краснодар: КубГАУ – 2016, – 615 с., в электронном виде на сайте автора: <http://lc.kubagro.ru/aidos/p14.htm>

Срок выполнения проекта: с “ ” _____ по “ ” _____ 2017 г.

Срок защиты: “ ” _____ 2017 г.

Дата выдачи задания: “ ” _____ 2017 г.

Дата сдачи проекта на кафедру: “ ” _____ 2017 г.

Руководитель проекта: Луценко Евгений Вениаминович,
д.э.н., к.т.н., профессор

(подпись, Ф.И.О., звание, степень)

Задание принял студент _____ “ ” _____ 2017 г.
(подпись, дата)

Краснодар
2017 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	6
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	7
1.1. Описание решения	7
1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel	7
1.3. Синтез и верификация статистических и интеллектуальных моделей	14
1.4. Виды моделей системы «Эйдос»	16
1.5. Результаты верификации моделей.....	17
2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	24
2.1. Решение задачи	24
2.2. Когнитивные функции	27
2.3. SWOT и PEST матрицы и диаграммы.....	29
2.4. Нелокальные нейронные сети и нейроны.....	31
2.5 Кластерный и конструктивный анализ	33
ЗАКЛЮЧЕНИЕ	34
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	35

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 рассмотрено решение задачи по поиску зависимости смерти знаменитости от профессии.

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 Kaggle.

В данной курсовой работе использована база данных «celebrity_deaths_4» из банка исходных данных по задачам искусственного интеллекта – репозитория Kaggle.

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 csv-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 CSV-формата в файл исходных данных MS Excel

Из электронного ресурса баз данных Kaggle возьмем базу данных сотрудников – «celebrity_deaths_4.csv», которую оставим без изменений.

Общее описание задачи:

1. Знаменит благодаря
2. Возраст
3. Дата смерти

Столбцы 1-2 описательные шкалы.

Столбец 2 является классификационной шкалой. Этот столбец показывает причину смерти знаменитости.

Обучающая выборка:

Таблица 1 – celebrity_deaths_4.xls

cause_of_death	cause_of_death	famous_for	age	death_year
cancer	cancer	academic and author	59	2012г
Alzheimer's disease	Alzheimer's disease	academic founder of NYIT Chancellor	89	2009г
cancer	cancer	academic President of Rice Universit	74	2015г
cancer	cancer	academic Rector of the University of	78	2013г
cancer	cancer	academic scientific skeptic and write	71	2016г
cancer	cancer	academic secretary-general of the W	64	2015г
cancer	cancer	academic Vice-Chancellor of the Univ	78	2009г
cancer	cancer	Academy Award-winning film compc	84	2009г
bone cancer	bone cancer	accordion player	79	2011г
cancer	cancer	accountant and chairman	78	2009г
bowel cancer	bowel cancer	activist	62	2013г
after long illness	after long illness	actor	87	2010г
after long illness	after long illness	actor	84	2010г
after short illness	after short illness	actor	66	2011г
Alzheimer's disease	Alzheimer's disease	actor	86	2010г
cancer	cancer	actor	65	2008г
cancer	cancer	actor	65	2009г
cancer	cancer	actor	81	2010г
cancer	cancer	actor	75	2010г
cancer	cancer	actor	41	2011г
cancer	cancer	actor	63	2010г
car accident	car accident	actor	38	2008г
car accident	car accident	actor	72	2009г
car accident	car accident	actor	30	2009г
cardiac arrest	cardiac arrest	actor	82	2009г
cardiac arrest	cardiac arrest	actor	70	2012г
cardiac arrest	cardiac arrest	actor	75	2009г
cardiac arrest	cardiac arrest	actor	71	2010г
accidental heroin overdose	accidental heroin overdose	actor	25	2008г
accidental prescription drug o	accidental prescription drug o	actor	28	2008г
acute leukemia	acute leukemia	actor	83	2016г
acute myeloid leukemia	acute myeloid leukemia	actor	75	2016г
after brief illness	after brief illness	actor	83	2012г
after long illness	after long illness	actor	67	2007г
after long illness	after long illness	actor	77	2012г
after long illness	after long illness	actor	70	2012г
after short illness	after short illness	actor	73	2009г
after short illness	after short illness	actor	92	2010г
after short illness	after short illness	actor	94	2011г
Alzheimer's disease	Alzheimer's disease	actor	88	2014г
Alzheimer's disease	Alzheimer's disease	actor	85	2015г
amyloidosis	amyloidosis	actor	61	2013г
amyotrophic lateral sclerosis	amyotrophic lateral sclerosis	actor	80	2014г
and author	and author	actor	79	2016г
and director	and director	actor	71	2015г
and radio host	and radio host	actor	82	2006г
and World War II veteran	and World War II veteran	actor	80	2006г
anemia	anemia	actor	73	2014г
apparent heart attack	apparent heart attack	actor	63	2012г
apparent heart failure	apparent heart failure	actor	52	2013г
apparent suicide	apparent suicide	actor	44	2006г

Всего в базе 5 столбцов, 1 классификационная и 3 описательных.

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 csv-файла в xlsx-файл, для удобства дальнейшего редактирования и выполним следующие операции.

Конвертируем исходный файл из csv формата в xlsx и запишем его с именем: Inp_data.xlsx в папку: c:\Aidos-X\AID_DATA\Inp_data\. 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.

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.

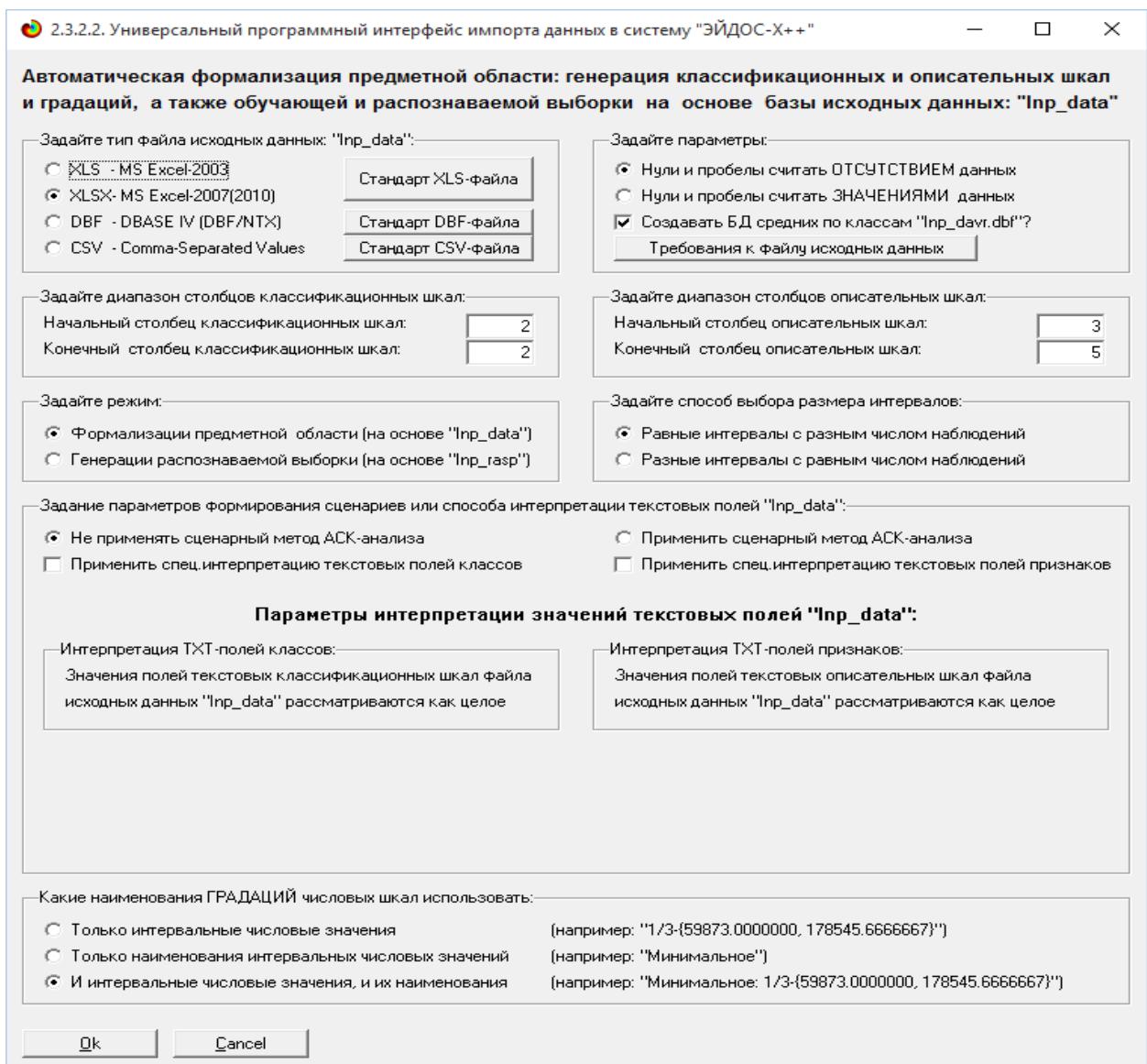


Рисунок 4. Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

В экранной форме, приведенной на рисунке 5, задать настройки, показанные на рисунке:

- "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(последний столбец в таблице);

- "Задайте диапазон столбцов описательных шкал": "Начальный столбец описательных шкал" – 3, "Конечный столбец описательных шкал" – 5;
- "Задание параметров формирования сценариев или способа интерпретации текстовых полей": "Не применять сценарный метод АСК-анализа".

После нажать кнопку "OK". Далее открывается окно, где размещена информация о размерности модели (рисунок 2). В этом окне необходимо нажать кнопку "Выйти на создание модели".

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	0	0	0,00	1	10	10,00
Текстовые	1	21	21,00	2	31	15,50
ВСЕГО:	1	21	21,00	3	41	13,67

Задайте число интервалов (градаций) в шкале: В описательных шкалах:

Рисунок 5. 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 "OK".

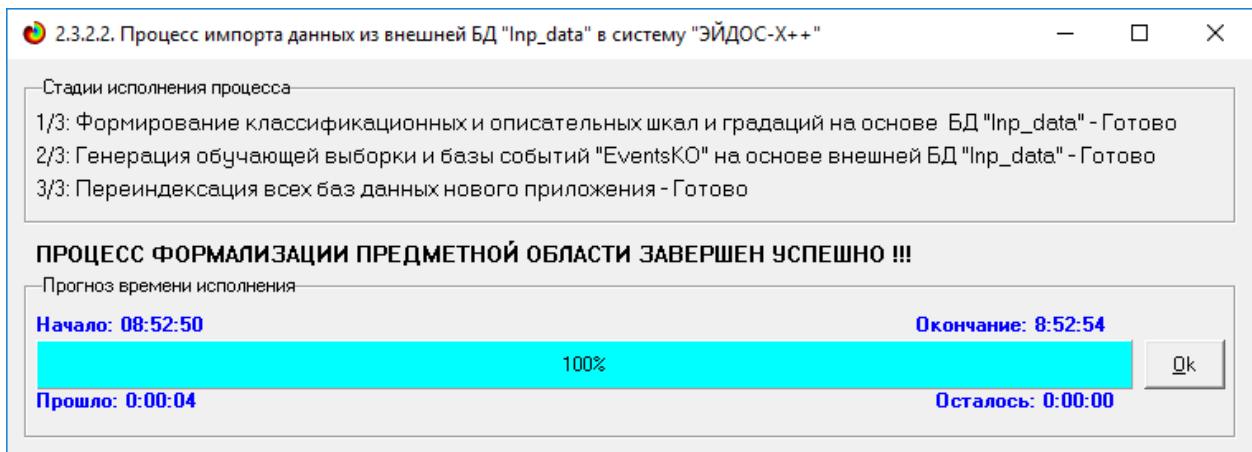


Рисунок 6. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

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 7).

2.1. Классификационные шкалы и градации. Текущая модель: "INF1"	
Код шкалы	Наименование классификационной шкалы
1	CAUSE_OF_DEATH
	...
Код градации	Наименование градации классификационной шкалы
1	accidental drug overdose
2	after long illness
3	accidental drug overdose
4	after brief illness
5	after long illness
6	Alzheimer's disease
7	apparent heart attack
8	bladder cancer
9	blood cancer
10	blood infection
11	bone cancer
12	bowel cancer
13	brain aneurysm
14	brain cancer
15	brain tumor
16	brain tumour
17	breast cancer
18	cancer
19	car accident
20	suicide
21	leukaemia

DEL

Помощь Добав.шкалам Добав.град.шкалам Копир.шкалам Копир.град.шкалам Копир.шкалам с град. Удал.шкалам с град. Удал.град.шкалам Удаление и перекодирование

Рисунок 7. Классификационные шкалы и градации (фрагмент)

Для просмотра описательных шкал и градаций необходимо запустить режим 2.2 (рисунок 8):

Код шкалы	Наименование описательной шкалы	...	
1	FAMOUS_FOR	...	
2	AGE	...	
3	DEATH_YEAR	...	
Код градации	Наименование градации описательной шкалы	...	
1	actor	...	
2	actress	...	
3	author	...	
4	baseball player	...	
5	basketball player	...	
6	bass guitarist	...	
7	bassist	...	
8	film producer	...	
9	football coach	...	
10	football player	...	
11	guitarist	...	
12	ice hockey player	...	
13	journalist	...	
14	musician	...	
15	politician	...	
16	politician MP	...	
17	politician Prime Minister	...	
18	singer-songwriter	...	
19	singer	...	
20	writer	...	

Рисунок 8. Описательные шкалы и градации (фрагмент)

Рисунок 9. 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 [11]).

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 10).

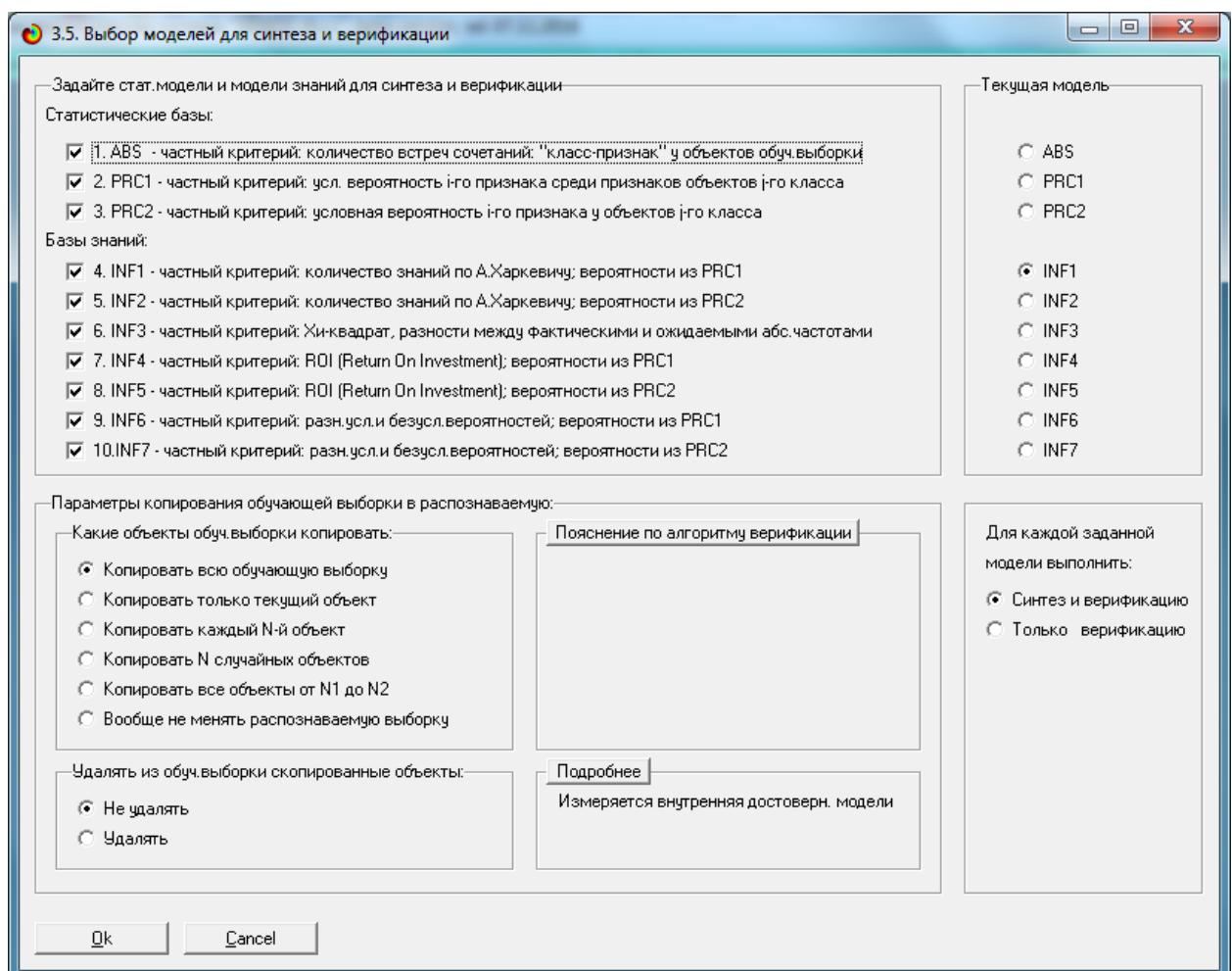


Рисунок 10. Выбор моделей для синтеза и верификации, а также текущей модели

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 10. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 11.

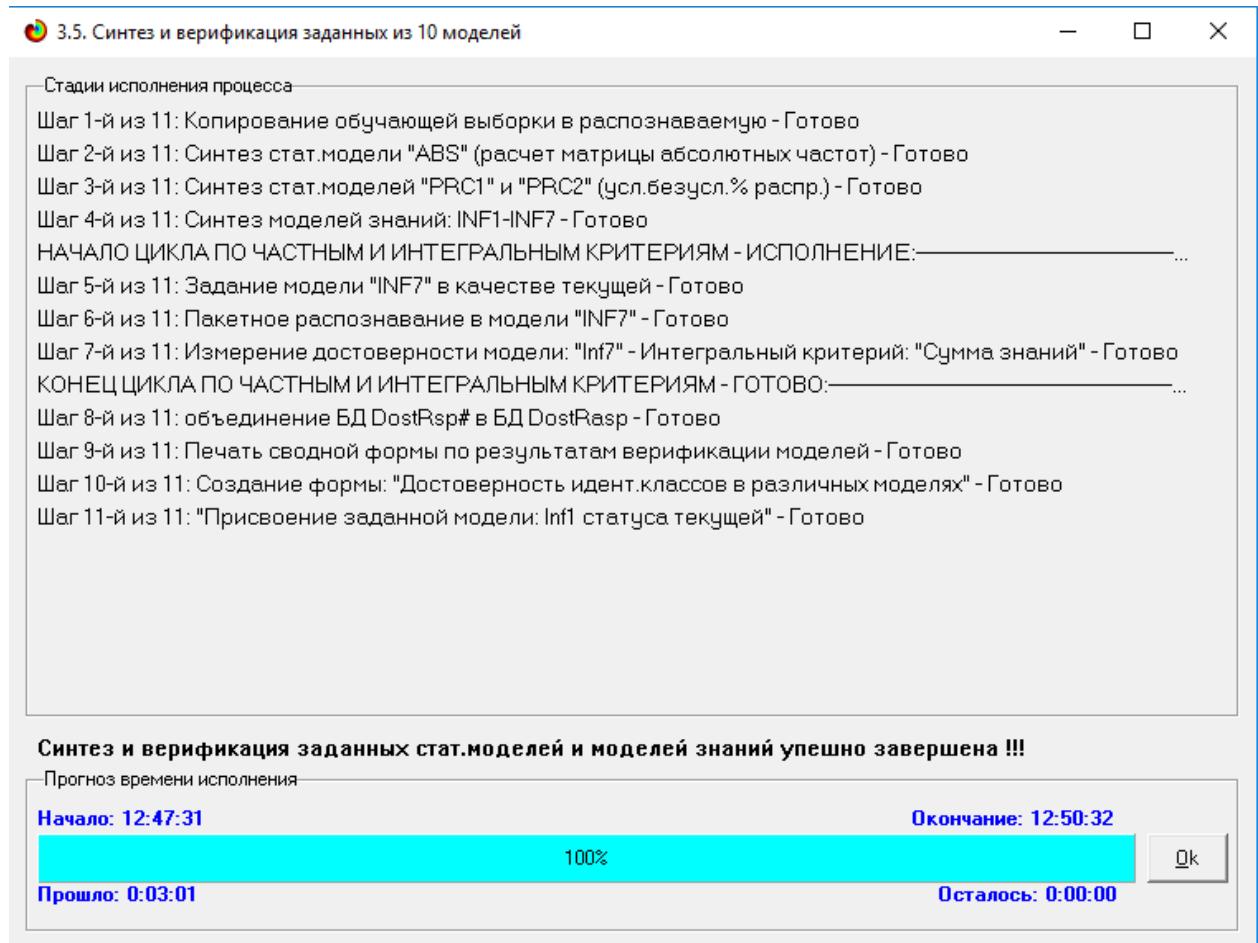


Рисунок 11. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 181 секунду. При этом верификация (оценка достоверности моделей) проводилась на всех 456 примерах наблюдения из обучающей выборки. В результате выполнения режима 3.5 созданы все модели, со всеми частными критериями, перечисленные на рисунке 10, 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) (проф. В.И.Лойко).

Таблица 2 – Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений (фрагменты)

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"														
Код признака	Наименование описательной шкалы и градации	1. CAUSE_OF... ACCIDENTAL DRUG OVERDOSE	2. CAUSE_OF... AFTER LONG ILLNESS	3. CAUSE_OF... ACCIDENTAL DRUG OVERDOSE	4. CAUSE_OF... AFTER BRIEF ILLNESS	5. CAUSE_OF... AFTER LONG ILLNESS	6. CAUSE_OF... ALZHEIME... DISEASE	7. CAUSE_OF... APPARENT HEART ATTACK	8. CAUSE_OF... BLADDER CANCER	9. CAUSE_OF... BLOOD CANCER	10. CAUSE_OF... BLOOD INFECTION	11. CAUSE_OF... BONE CANCER	12. CAUSE_OF... BOWEL CANCER	13. CAUSE_OF... BRAIN ANEURYSM
21	AGE_1-е из 10: 1/10-[21.0000000, 28.7000000]	6	1											
22	AGE_2-е из 10: 2/10-[28.7000000, 36.4000000)													
23	AGE_3-е из 10: 3/10-[36.4000000, 44.1000000)	6	1								2		1	
24	AGE_4-е из 10: 4/10-[44.1000000, 51.8000000)	6	2											1
25	AGE_5-е из 10: 5/10-[51.8000000, 59.5000000)	5	3						1	1	1	1	1	1
26	AGE_6-е из 10: 6/10-[59.5000000, 67.2000000)	4	3					6	1	1	1	4	1	1
27	AGE_7-е из 10: 7/10-[67.2000000, 74.9000000)	5	8				3				1	1		
28	AGE_8-е из 10: 8/10-[74.9000000, 92.6000000)	7	18				8	1		1		1		1
29	AGE_9-е из 10: 9/10-[82.6000000, 90.3000000)	1	13		1		11			1				1
30	AGE_10-е из 10: 10/10-[90.3000000, 98.0000000)	2	4				9					1		
31	DEATH_YEAR-2006г	2	2				1							
32	DEATH_YEAR-2007г	2	2				1	1				1		
33	DEATH_YEAR-2008г	2	8				2	1						
34	DEATH_YEAR-2009г	8	12				4			1		1		
35	DEATH_YEAR-2010г	2	12				5	1	1		1	1	1	2
36	DEATH_YEAR-2011г	10	6				2	3		1	2			
37	DEATH_YEAR-2012г	2	3		1		9				1			
38	DEATH_YEAR-2013г	4	1				3				1		1	
39	DEATH_YEAR-2014г	3	1				6			3		1		1
40	DEATH_YEAR-2015г	5	3				3				1	1		
41	DEATH_YEAR-2016г	2	3				4			2		1		
Сумма числа признаков		84	106		2		80	12	2	14	14	8	6	6

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

Код признака	Наименование описательной шкалы и градации	1. CAUSE_OF... ACCIDENTAL DRUG OVERDOSE	2. CAUSE_OF... AFTER LONG ILLNESS	3. CAUSE_OF... ACCIDENTAL DRUG OVERDOSE	4. CAUSE_OF... AFTER BRIEF ILLNESS	5. CAUSE_OF... AFTER LONG ILLNESS	6. CAUSE_OF... ALZHEIME... DISEASE	7. CAUSE_OF... APPARENT HEART ATTACK	8. CAUSE_OF... BLADDER CANCER	9. CAUSE_OF... BLOOD CANCER	10. CAUSE_OF... BLOOD INFECTION	11. CAUSE_OF... BONE CANCER	12. CAUSE_OF... BOWEL CANCER	13. CAUSE_OF... BRAIN ANEURYSM	14. CA BRA CA
22	AGE_2-е из 10: 2/10-[28.700000, 36.400000)							-0.508	1.424		1.536		1.450		
23	AGE_3-е из 10: 3/10-[36.400000, 44.100000)	0.616	-0.689											1.162	
24	AGE_4-е из 10: 4/10-[44.100000, 51.800000)	0.566	-0.292				-0.557							1.113	
25	AGE_5-е из 10: 5/10-[51.800000, 59.500000)	0.052	-0.428				-0.954	0.269	1.424	0.169	0.169	0.530	0.715		
26	AGE_6-е из 10: 6/10-[59.500000, 67.200000)	-0.395	-0.731				-0.102	-0.034		-0.134	0.760	0.227	0.412		
27	AGE_7-е из 10: 7/10-[67.200000, 74.900000)	-0.320	-0.167				-0.618			-0.203	-0.203				
28	AGE_8-е из 10: 8/10-[74.900000, 82.600000)	-0.125	0.334				-0.007	-0.125		-0.224		0.137		0.322	
29	AGE_9-е из 10: 9/10-[82.600000, 90.300000)	-0.973	0.531		1.437		0.604			0.182			0.728		
30	AGE_10-е из 10: 10/10-[90.300000, 98.000000)	0.052	0.348			1.053				0.760					
31	DEATH_YEAR-2006r	0.566	0.416				0.151								
32	DEATH_YEAR-2007r	0.237	0.087				-0.178	1.045			0.945				
33	DEATH_YEAR-2008r	-0.271	0.472				-0.240	0.536							
34	DEATH_YEAR-2009r	0.282	0.393				-0.134			0.096		0.457			
35	DEATH_YEAR-2010r	-0.688	0.317				-0.066	0.120	1.275		0.020	0.381	0.566	1.013	
36	DEATH_YEAR-2011r	0.414	-0.065				-0.592	0.892		0.085	0.532				
37	DEATH_YEAR-2012r	-0.612	-0.501		1.351		0.389				0.096				
38	DEATH_YEAR-2013r	0.195	-0.848				0.041				0.457		1.003		
39	DEATH_YEAR-2014r	-0.101	-0.959				0.378			1.054			0.892	0.892	
40	DEATH_YEAR-2015r	0.134	-0.345				-0.164				0.252	0.612			
41	DEATH_YEAR-2016r	-0.564	-0.453				-0.086			0.591		0.505			
	Сумма	0.210	-2.738		2.787			-1.591	4.126	2.698	3.153	3.788	4.298	4.318	4.502
	Среднее	0.005	-0.067		0.068		-0.039	0.101	0.066	0.077	0.092	0.105	0.105	0.110	

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

Код признака	Наименование описательной шкалы и градации	1. CAUSE_OF... ACCIDENTAL DRUG OVERDOSE	2. CAUSE_OF... AFTER LONG ILLNESS	3. CAUSE_OF... ACCIDENTAL DRUG OVERDOSE	4. CAUSE_OF... AFTER BRIEF ILLNESS	5. CAUSE_OF... AFTER LONG ILLNESS	6. CAUSE_OF... ALZHEIME... DISEASE	7. CAUSE_OF... APPARENT HEART ATTACK	8. CAUSE_OF... BLADDER CANCER	9. CAUSE_OF... BLOOD CANCER	10. CAUSE_OF... BLOOD INFECTION	11. CAUSE_OF... BONE CANCER	12. CAUSE_OF... BOWEL CANCER	13. CAUSE_OF... BRAIN ANEURYSM	14. CA BRA CA
21	AGE_1-е из 10: 1/10-[21.000000, 28.700000)	4.985	-0.281		-0.024		-0.967	-0.145	-0.024	-0.169	-0.169	-0.097	-0.073	-0.073	
22	AGE_2-е из 10: 2/10-[28.700000, 36.400000)	-1.108	-1.398		-0.026		-1.055	-0.158	-0.026	1.815	-0.185	0.895	-0.079	-0.079	
23	AGE_3-е из 10: 3/10-[36.400000, 44.100000)	3.692	-1.912		-0.055		-1.198	2.670	-0.055	-0.385	-0.385	-0.220	-0.165	0.835	
24	AGE_4-е из 10: 4/10-[44.100000, 51.800000)	3.508	-1.145		-0.059		-1.374	-0.356	-0.059	-0.415	-0.415	-0.237	-0.178	0.822	
25	AGE_5-е из 10: 5/10-[51.800000, 59.500000)	0.385	-2.824		-0.110		-3.396	0.341	0.890	0.231	0.231	0.560	0.670	-0.330	
26	AGE_6-е из 10: 6/10-[59.500000, 67.200000)	-3.385	-6.319		-0.176		-1.033	-0.055	-0.176	-0.231	2.769	0.297	0.473	-0.527	
27	AGE_7-е из 10: 7/10-[67.200000, 74.900000)	-3.215	-2.367		-0.196		-4.824	-1.174	-0.196	-0.369	-0.369	-0.782	-0.587	-0.587	
28	AGE_8-е из 10: 8/10-[74.900000, 82.600000)	-1.492	7.284		-0.202		-0.088	-0.213	-0.202	-0.415	-1.415	0.191	-0.607	0.393	
29	AGE_9-е из 10: 9/10-[82.600000, 90.300000)	-3.523	7.292	0.892			6.692	-0.646	-0.108	0.246	-0.754	-0.431	0.677	-0.323	
30	AGE_10-е из 10: 10/10-[90.300000, 98.000000)	0.154	1.670		-0.044		7.242	-0.264	-0.044	-0.308	0.692	-0.176	-0.132	-0.132	
31	DEATH_YEAR-2006r	1.169	0.952		-0.020		0.209	-0.119	-0.020	-0.138	-0.138	-0.079	-0.059	-0.059	
32	DEATH_YEAR-2007r	0.615	0.253		-0.033		-0.319	0.802	-0.033	-0.231	0.769	-0.132	-0.099	-0.099	
33	DEATH_YEAR-2008r	-1.046	4.156		-0.073		-0.901	0.565	-0.073	-0.508	-0.508	-0.290	-0.218	-0.218	
34	DEATH_YEAR-2009r	2.831	5.477		-0.123		-0.923	-0.738	-0.123	0.138	-0.862	0.508	-0.369	-0.369	
35	DEATH_YEAR-2010r	-3.815	4.662		-0.138		-0.538	0.169	0.862	-0.969	0.031	0.446	0.585	1.585	
36	DEATH_YEAR-2011r	4.738	-0.640		-0.125		-3.011	2.248	-0.125	0.123	1.123	-0.501	-0.376	-0.376	
37	DEATH_YEAR-2012r	-3.169	-3.523	0.877	4.077		-0.738	-0.123	-0.862	0.138	-0.492	-0.369	-0.369		
38	DEATH_YEAR-2013r	1.046	-2.727		-0.070		0.187	-0.422	-0.070	-0.492	0.508	-0.281	0.789	-0.211	
39	DEATH_YEAR-2014r	-0.508	-3.426		-0.084		2.659	-0.501	-0.084	2.415	-0.585	-0.334	0.749	0.749	
40	DEATH_YEAR-2015r	0.938	-2.125		-0.097		-0.868	-0.580	-0.097	-0.677	0.323	0.613	-0.290	-0.290	
41	DEATH_YEAR-2016r	-2.800	-3.057		-0.114		-0.571	-0.686	-0.114	1.200	-0.800	0.543	-0.343	-0.343	
	Сумма														

1.5. Результаты верификации моделей

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

4.1.3.6. Обобщ.форма по достоверн.моделей при разн.инт.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 Ризбергена	Сумма модулей сходства истинно-положительных решений (STP)	Сумма модулей сходства истинно-отрицательных решений (SFN)	Сумма модулей сходства ложно-положительных решений (SFP)	Сумма модулей сходства ложно-отрицательных решений (SFF)	Сумма модулей сходства истинно-положительных и отрицательных решений (S1)	Сумма модулей сходства истинно-отрицательных и положительных решений (S2)	Сумма модулей сходства истинно-положительных и отрицательных решений (S3)	
1. ABS - частный критерий: количество встреч с сочетанием "Х/л...	Корреляция abs.частот с обр...	455	444	2778	5412	11	0.076	0.976	0.141	165.296	189.894	1355.873	0.25.				
1. ABS - частный критерий: количество встреч с сочетанием "Х/л...	Сумма abs.частот по признакам...	455	455	2321	5869		0.072	1.000	0.134	169.467		404.956					
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн.частот с о...	455	444	2778	5412	11	0.076	0.976	0.141	165.296	189.894	1355.873	0.25.				
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн.частот по приз...	455	455	2321	5869		0.072	1.000	0.134	169.467		404.956					
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн.частот с о...	455	444	2778	5412	11	0.076	0.976	0.141	165.213	189.855	1355.375	0.25.				
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн.частот по приз...	455	455	2321	5869		0.072	1.000	0.134	165.213	189.855	1355.375	0.25.				
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	455	356	4983	3207	99	0.100	0.782	0.177	85.044	602.902	686.767	12.60.				
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	455	334	4747	3443	121	0.088	0.734	0.159	65.496	386.032	778.813	14.00.				
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	455	356	4983	3207	99	0.100	0.782	0.177	85.044	602.902	686.767	12.60.				
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	455	334	4747	3443	121	0.088	0.734	0.159	65.496	386.032	778.813	14.00.				
6. INF3 - частный критерий Хинкарадзе: различия между фактами...	Семантический резонанс зна...	455	357	5263	2927	98	0.109	0.785	0.191	126.460	1023.161	721.348	14.43.				
6. INF3 - частный критерий Хинкарадзе: различия между фактами...	Сумма знаний	455	357	5264	2926	98	0.109	0.785	0.191	110.386	317.573	221.192	14.00.				
7. INF4 - частный критерий ROI (Return On Investment); вероятно...	Семантический резонанс зна...	455	353	5267	2923	102	0.108	0.776	0.189	96.668	460.805	614.585	11.55.				
7. INF4 - частный критерий ROI (Return On Investment); вероятно...	Сумма знаний	455	357	4495	3695	98	0.088	0.785	0.158	34.311	60.980	443.382	2.25.				
8. INF5 - частный критерий ROI (Return On Investment); вероятно...	Семантический резонанс зна...	455	353	5267	2923	102	0.108	0.776	0.189	96.668	460.805	614.585	11.55.				
8. INF5 - частный критерий ROI (Return On Investment); вероятно...	Сумма знаний	455	357	4495	3695	98	0.088	0.785	0.158	34.311	60.980	443.382	2.25.				
9. INF6 - частный критерий: различия и безразл.вероятностей; вер...	Семантический резонанс зна...	455	356	4988	3202	99	0.100	0.782	0.177	123.835	586.952	845.904	15.13.				
9. INF6 - частный критерий: различия и безразл.вероятностей; вер...	Сумма знаний	455	357	4629	3561	98	0.091	0.785	0.163	30.735	87.850	452.549	2.68.				
10.INF7 - частный критерий: различия и безразл.вероятностей; вер...	Семантический резонанс зна...	455	356	4988	3202	99	0.100	0.782	0.177	123.835	586.952	845.904	15.13.				
10.INF7 - частный критерий: различия и безразл.вероятностей; вер...	Сумма знаний	455	357	4629	3561	98	0.091	0.785	0.163	30.735	87.850	452.549	2.68.				

Рисунок 12. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказалась модели ABS, INF5 при интегральном критерии «Семантический резонанс знаний». При этом точность модели составляет 0,687. Таким образом, уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 69%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием (рисунок 13).

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-X++"																
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-X++".																
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 отнесеного 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 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, 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 осуществляется 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, то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, которым они относятся, так как 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}:																
$L_1 = \frac{TP + TN}{TP + TN + FP + FN}$ / { $TP + TN + FP + FN$ } (нормировка: {-1,+1})																
$L_2 = \frac{1}{2} + \frac{(TP + TN - FP - FN)}{(TP + TN + FP + FN)}$ / 2 (нормировка: {0, 1})																
где: TP – истинно-положительное решение; TN – истинно-отрицательное решение; FP – ложно-положительное решение; FN – ложно-отрицательное решение;																
Классическая F-мера достоверности моделей Ван Ризбергена:																
$Precision = \frac{TP}{TP+FP}$ – точность модели;																
$Recall = \frac{TP}{TP+FN}$ – полнота модели;																
$F\text{-мера} = 2 * (\frac{Precision * Recall}{Precision + Recall})$ – нечеткая мультиклассовая достоверность модели.																
Где:																
STP – Сума модулей сходства истинно-положительных решений; STN – Сума модулей сходства истинно-отрицательных решений;																
SFP – Сума модулей сходства ложно-положительных решений; SFN – Сума модулей сходства ложно-отрицательных решений.																

Рисунок 13. 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 INF3.

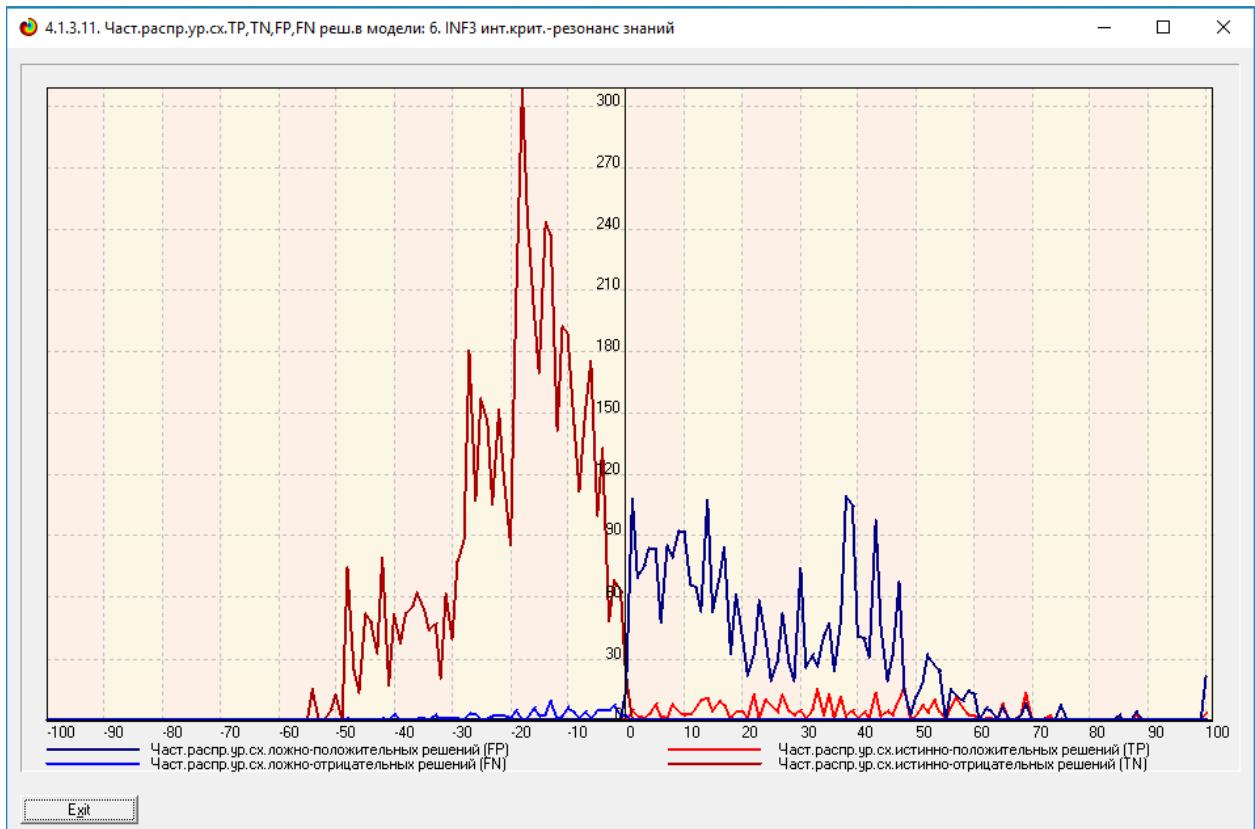


Рисунок 14. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF3

Из рисунка 11 видно, что:

- наиболее достоверная модель INF5 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 9);

– модуль уровня сходства-различия в наиболее достоверной модели INF5 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 уровня сходства выше 70% все положительные решения являются истинными. Для очень больших значений уровней сходства-различия (более 70%) также различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями практически отсутствует.

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 15. 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 INF5.

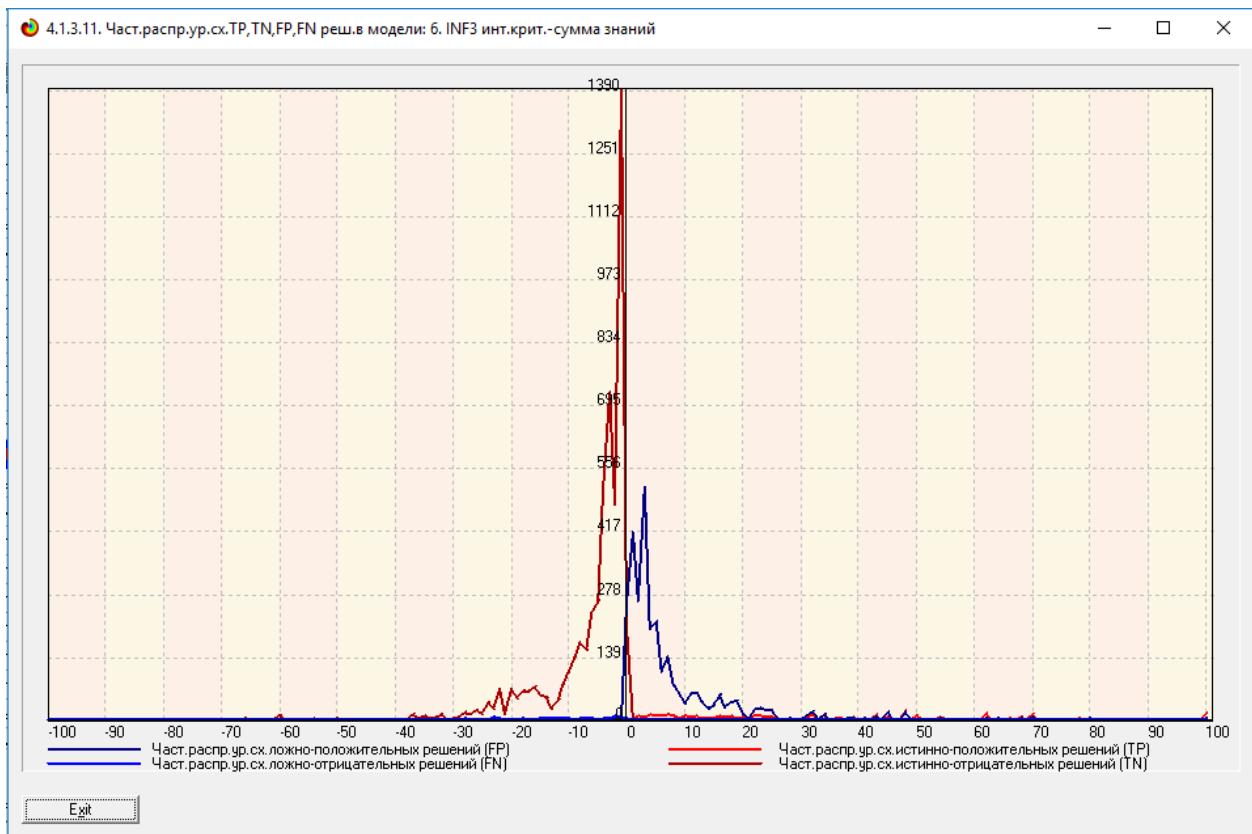


Рисунок 16. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF3

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.

Нанесение модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложных положительных решений (FP)	Число ложных отрицательных решений (FN)	Точность модели	Полнота модели	Фильтрация Ризбергена	Сумма модулей уровнями скопления	Сумма модулей уровнями скопления	Сумма модулей уровнями скопления
1. ABS - частный критерий: количество встреч соединений "Ход...".	Корреляция abs.частот с обр...	455	444	2778	5412	11	0.076	0.976	0.141	165.296	189.894	1355.873
1. ABS - частный критерий: количество встреч соединений "Ход...".	Сумма abs.частот по признакам...	455	455	2321	5869		0.072	1.000	0.134	169.467		404.956
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн.частот с о...	455	444	2778	5412	11	0.076	0.976	0.141	165.296	189.894	1355.873
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн.частот по приз...	455	455	2321	5869		0.072	1.000	0.134	81.628		920.322
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн.частот с о...	455	444	2778	5412	11	0.076	0.976	0.141	165.213	189.855	1355.375
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн.частот по приз...	455	455	2321	5869		0.072	1.000	0.134	81.628		920.322
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	455	356	4983	3207	99	0.100	0.782	0.177	85.044	602.902	686.767
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	455	334	4747	3443	121	0.088	0.734	0.158	65.496	386.032	778.813
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	455	356	4983	3207	99	0.100	0.782	0.177	85.044	602.902	686.767
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	455	334	4747	3443	121	0.088	0.734	0.158	65.496	386.032	778.813
6. INF3 - частный критерий Хинкалада: разности между фактами...	Семантический резонанс зна...	455	357	5263	2927	98	0.109	0.785	0.191	126.460	1023.161	721.348
6. INF3 - частный критерий Хинкалада: разности между фактами...	Сумма знаний	455	357	5264	2926	98	0.109	0.785	0.191	110.386	317.573	221.192
7. INF4 - частный критерий ROI (Return On Investment); вероятно...	Семантический резонанс зна...	455	353	5267	2923	102	0.108	0.776	0.189	96.668	460.805	614.595
7. INF4 - частный критерий ROI (Return On Investment); вероятно...	Сумма знаний	455	357	4495	3695	98	0.088	0.785	0.158	34.311	60.980	443.382
8. INF5 - частный критерий ROI (Return On Investment); вероятно...	Семантический резонанс зна...	455	353	5267	2923	102	0.108	0.776	0.189	96.668	460.805	614.595
8. INF5 - частный критерий ROI (Return On Investment); вероятно...	Сумма знаний	455	357	4495	3695	98	0.088	0.785	0.158	34.311	60.980	443.382
9. INF6 - частный критерий: разн.усл и безузл.вероятностей; вер...	Семантический резонанс зна...	455	356	4988	3202	99	0.100	0.782	0.177	123.835	586.952	845.904
9. INF6 - частный критерий: разн.усл и безузл.вероятностей; вер...	Сумма знаний	455	357	4629	3561	98	0.091	0.785	0.163	30.735	87.850	452.549
10.INF7 - частный критерий: разн.усл и безузл.вероятностей; вер...	Семантический резонанс зна...	455	356	4988	3202	99	0.100	0.782	0.177	123.835	586.952	845.904
10.INF7 - частный критерий: разн.усл и безузл.вероятностей; вер...	Сумма знаний	455	357	4629	3561	98	0.091	0.785	0.163	30.735	87.850	452.549

Рисунок 17. Достоверность статистических и когнитивных моделей, созданных на основе случайной выборки

На основе его сравнения с рисунком 9 можно сделать следующие выводы:

- достоверность лучшей модели INF3, отражающей реальный объект моделирования, примерно на 20% 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 ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи

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

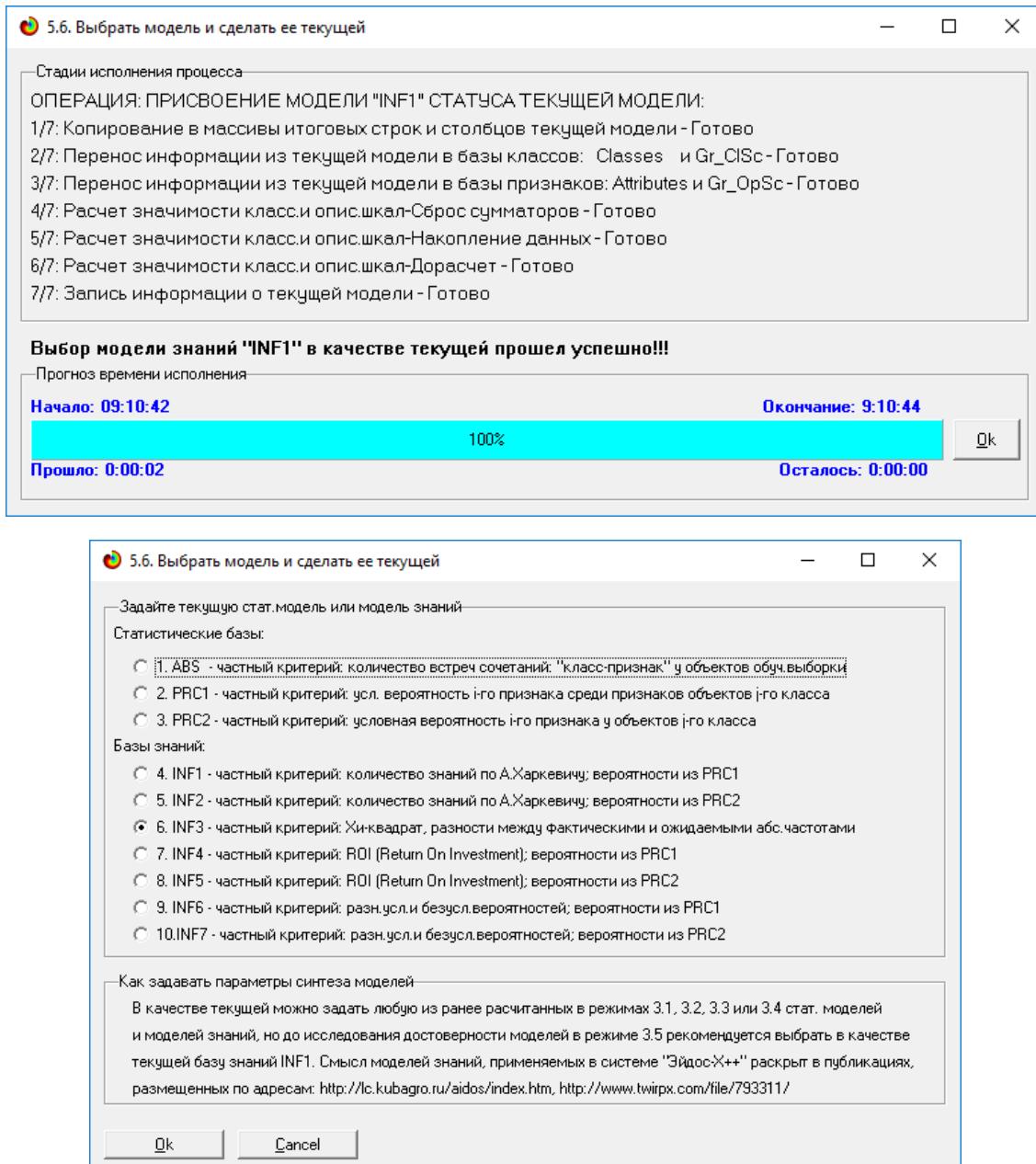


Рисунок 18. Экранные формы режима задания модели в качестве текущей

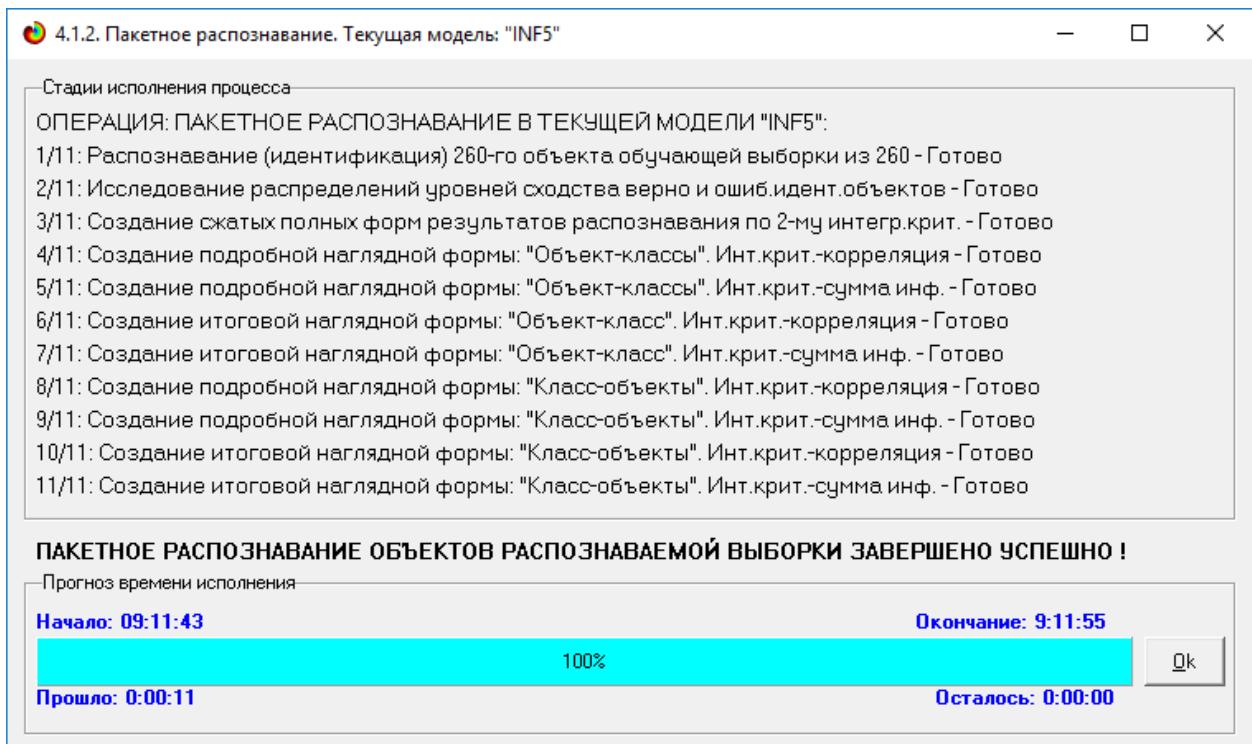


Рисунок 19. Экранная форма режима пакетного распознавания в текущей модели

В результате пакетного распознавания в текущей модели создается ряд баз данных, которые визуализируются в выходных экранах 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 17 и 18 приведены примеры прогнозов высокой и низкой достоверности частоты и классов ирисов в наиболее достоверной модели INF3 на основе наблюдения предыстории их развития:

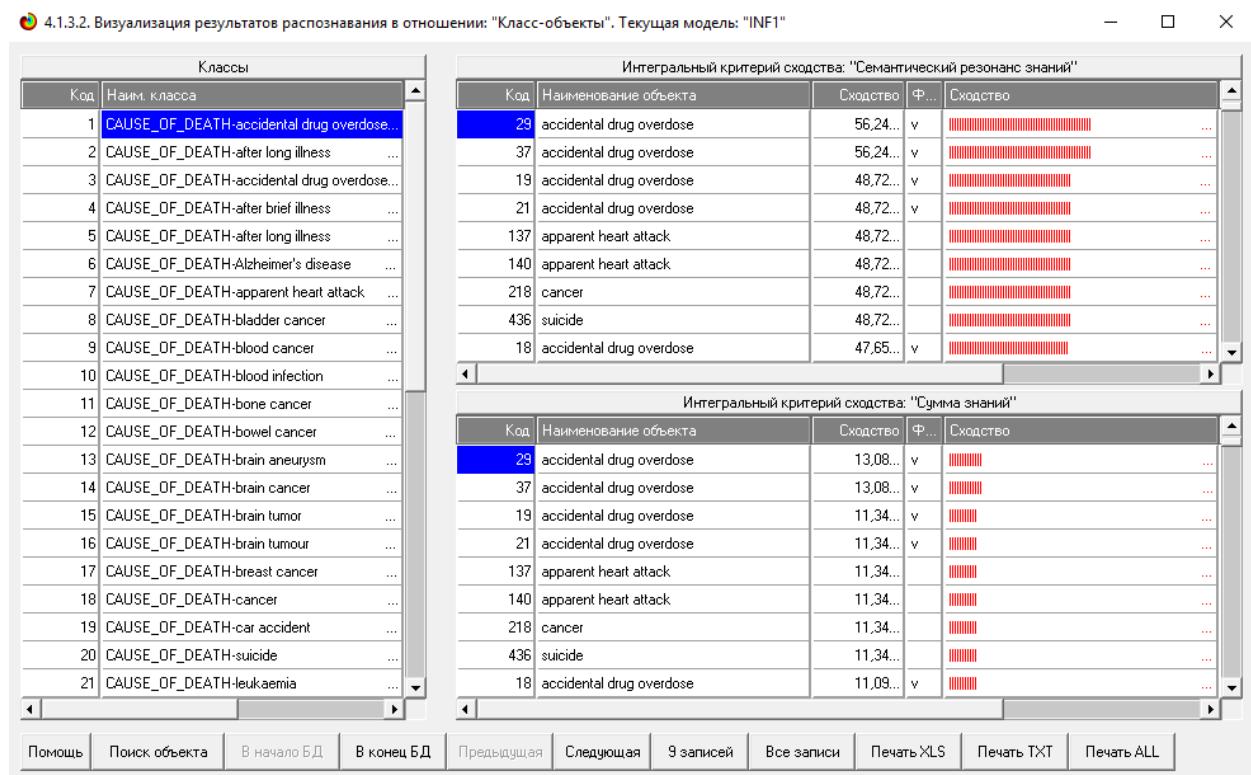


Рисунок 20. Пример идентификации классов в модели INF1

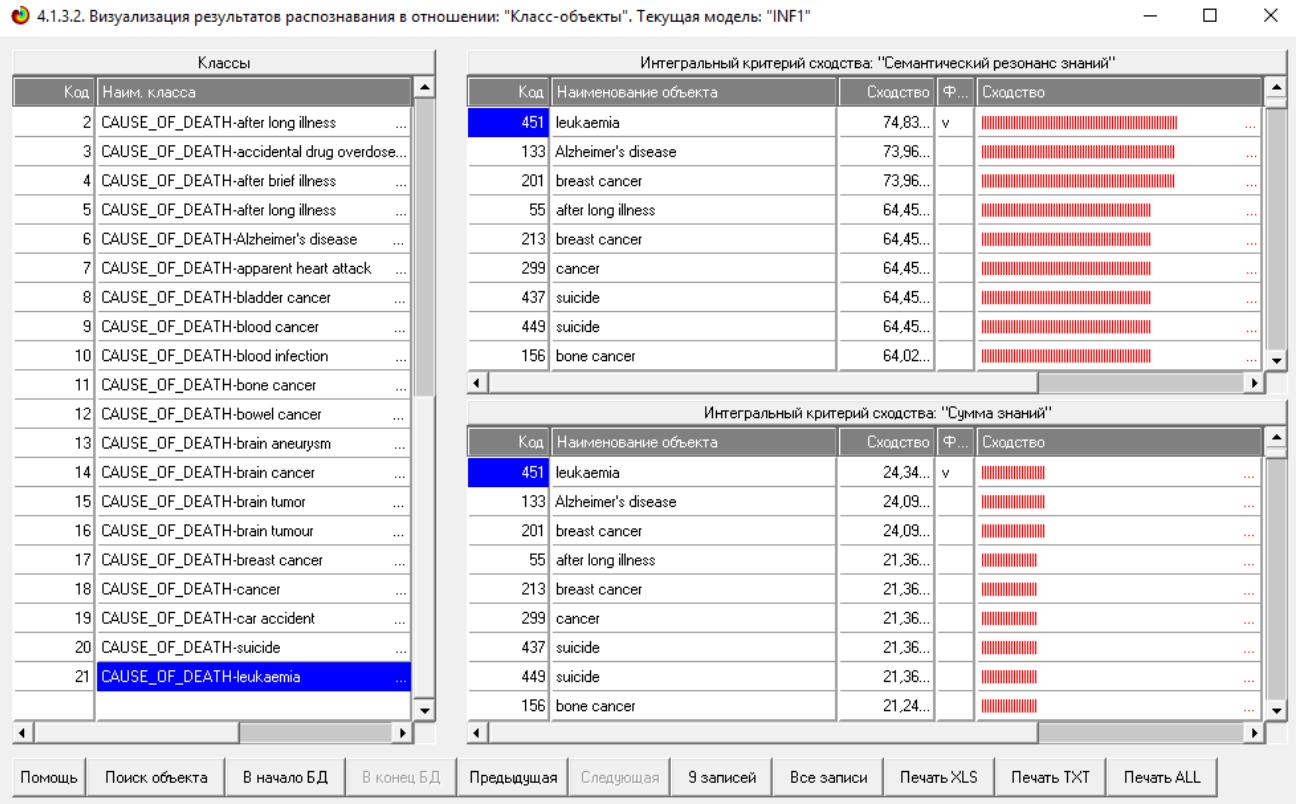


Рисунок 21. Пример идентификации классов в модели INF1

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 22)

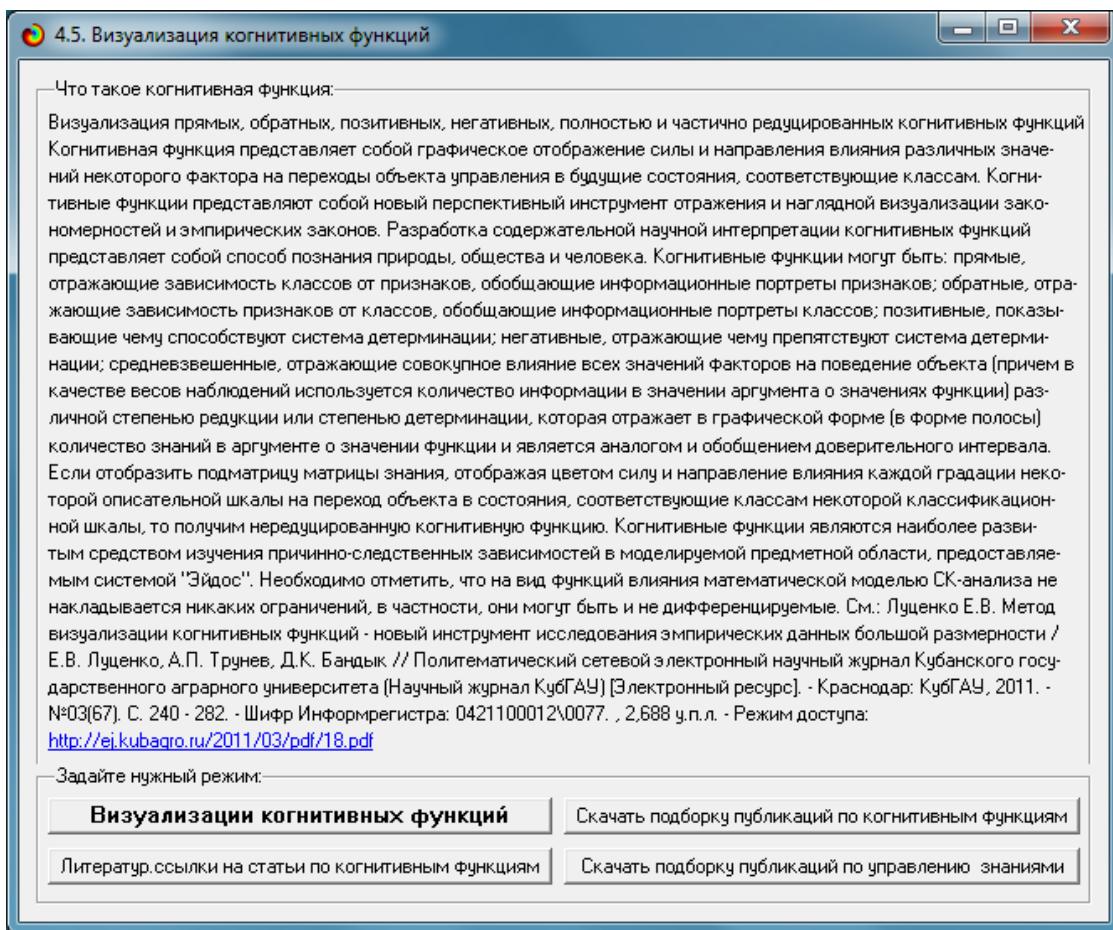


Рисунок 22. Экранная форма режима 4.5 системы «Эйдос-Х++» «Визуализация когнитивных функций»

Применительно к задаче, рассматриваемой в данной работе, когнитивная функция показывает, какое количество информации содержится в различных значениях факторов о том, что объект моделирования перейдет в те или иные будущие состояния. Когнитивным функциям посвящено много работ автора 9 , но наиболее новой и обобщающей из них является работа. Поэтому здесь не будем останавливаться на описании того, что представляют собой когнитивные функции в АСК-анализе. На рисунке 23. приведены визуализации всех когнитивных функций данного приложения для модели INF3.

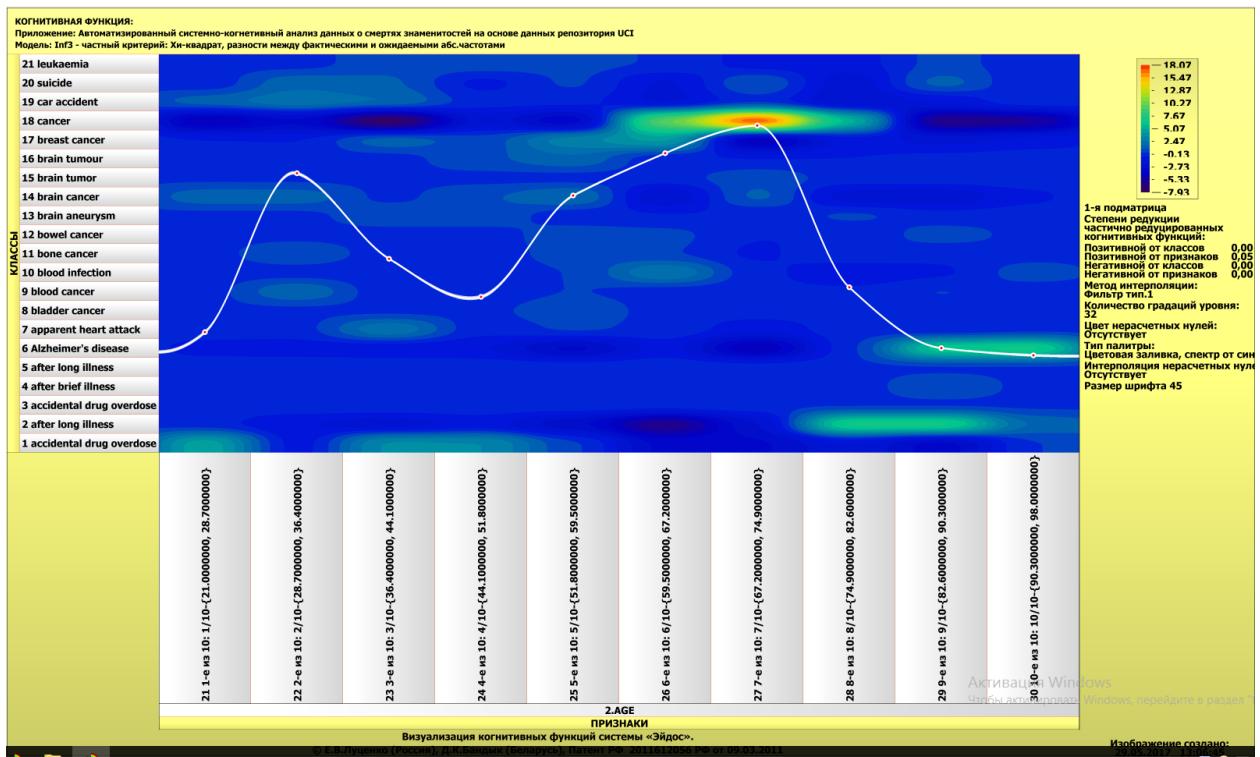


Рисунок 23. Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF3

2.3. SWOT и PEST матрицы и диаграммы

SWOT-анализ является широко известным и общепризнанным методом стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего профессионального опыта и компетенции. Но возможности экспертов имеют свои ограничения и часто по различным причинам они не могут и не

хотят это сделать. Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов. Эта проблема может решаться путем автоматизации функций экспертов, т.е. путем измерения силы и направления влияния факторов непосредственно на основе эмпирических данных. Подобная технология разработана давно, ей уже около 30 лет, но она малоизвестна – это интеллектуальная система «Эйдос». Данная система всегда обеспечивала возможность проведения количественного автоматизированного SWOT-анализа без использования экспертных оценок непосредственно на основе эмпирических данных. Результаты SWOT-анализа выводились в форме информационных портретов. В версии системы под MS Windows: «Эйдос-X++» предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (рисунок 24).

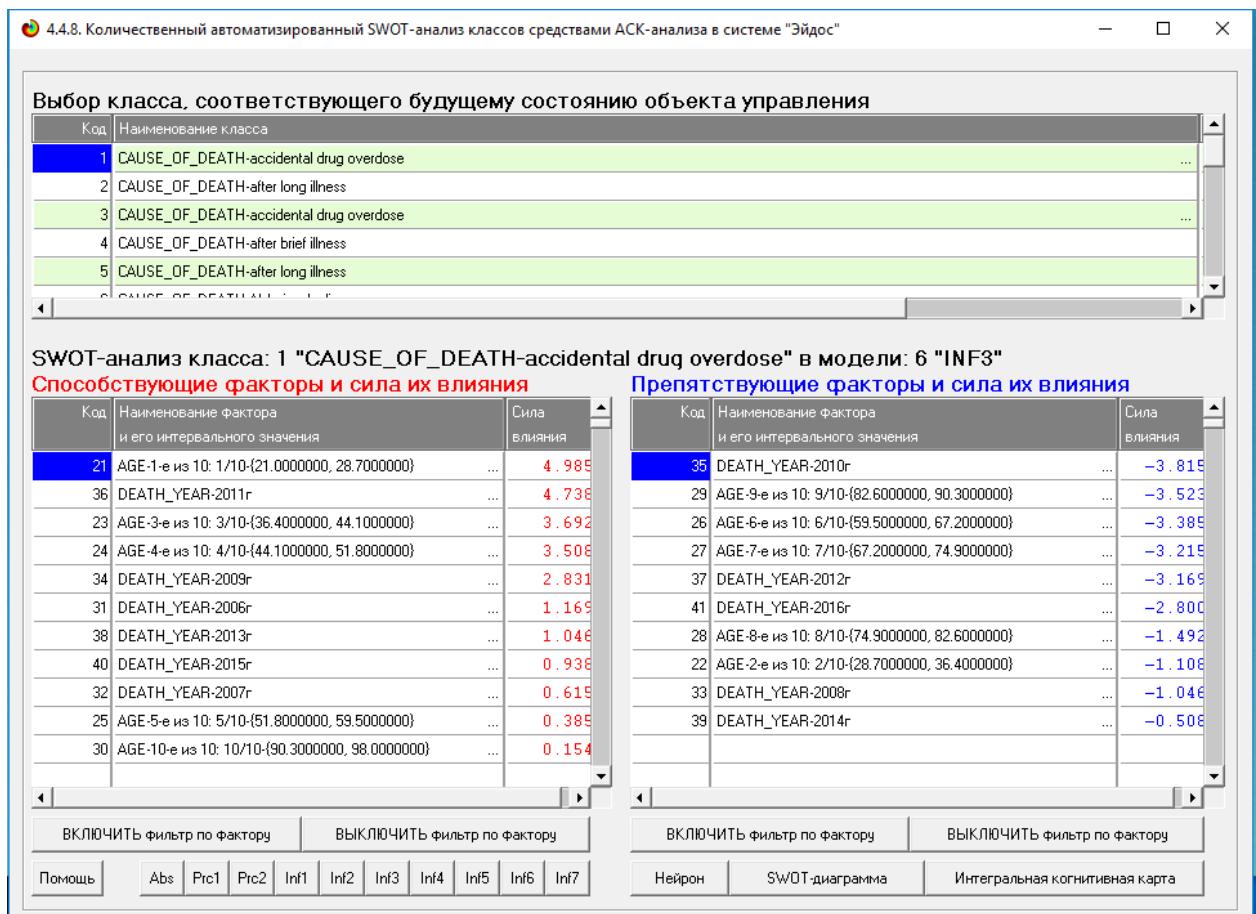


Рисунок 24. Пример SWOT-матрицы в модели INF3

На рисунке 25 приведены примеры инвертированной SWOT-матрицы и инвертированной SWOT-диаграммы в модели INF5.

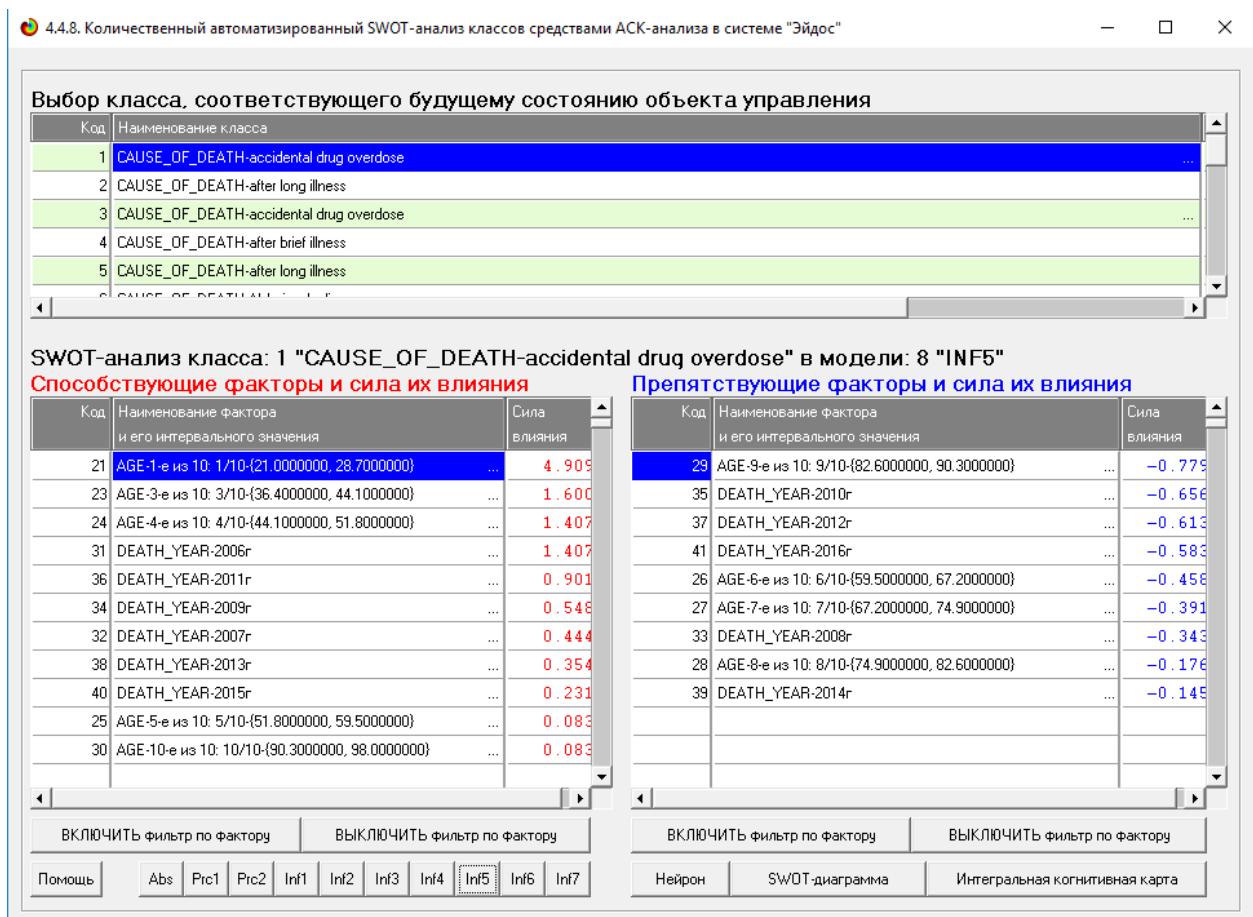


Рисунок 25. Пример SWOT-матрицы в модели INF5

2.4. НЕЛОКАЛЬНЫЕ НЕЙРОННЫЕ СЕТИ И НЕЙРОНЫ.

Каждому классу системно-когнитивной модели соответствует не локальный нейрон совокупность, которых образует не локальную нейронную сеть [9] В качестве примеров приведены рисунок 26-27.

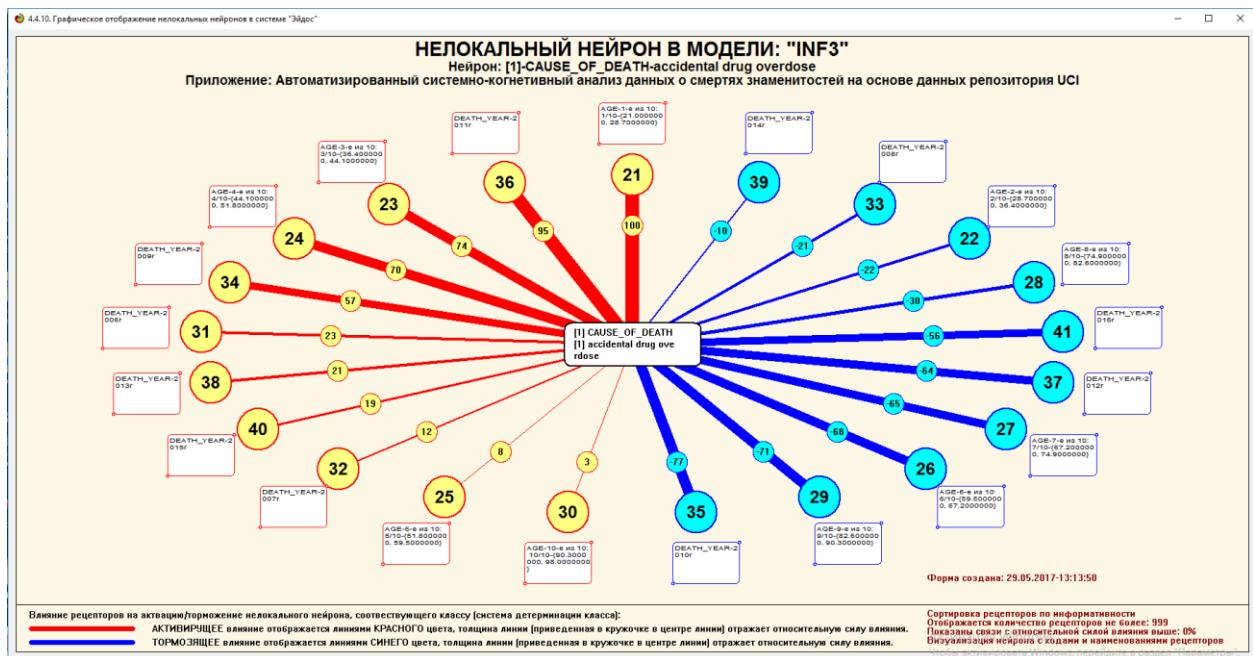


Рисунок 26. Пример нелокальные нейронные сети и нейроны модели INF3

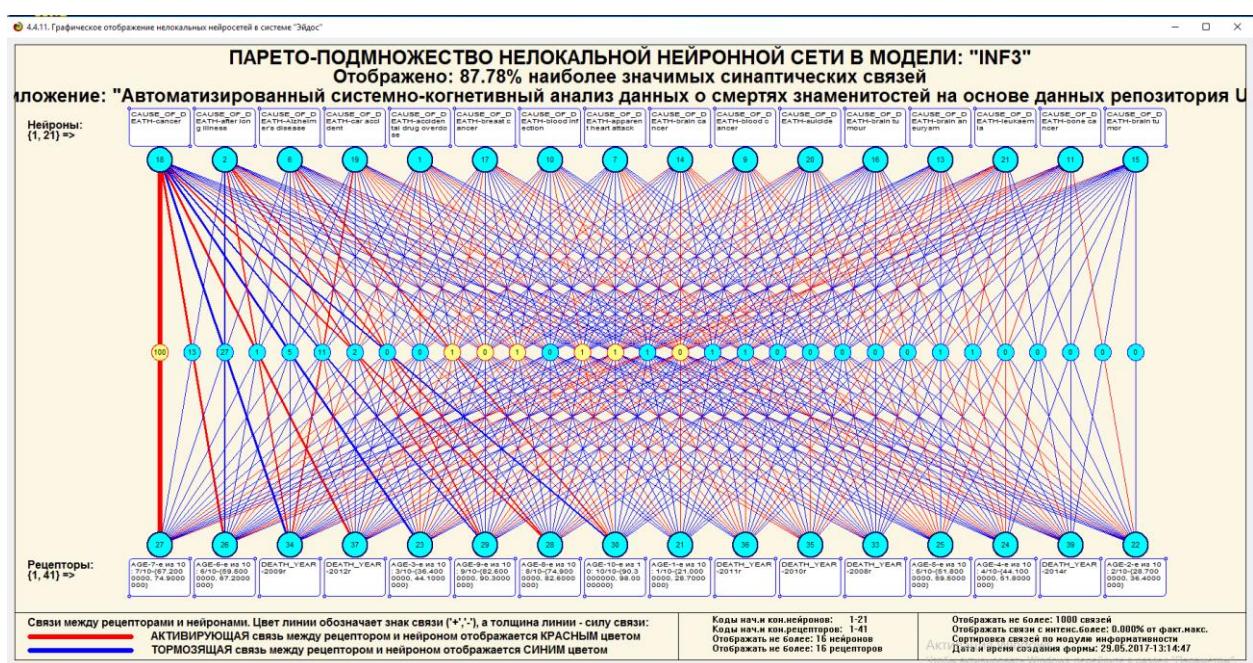


Рисунок 26. Пример нелокальные нейронные сети и нейроны модели INF3

2.5 КЛАСТЕРНЫЙ И КОНСТРУКТИВНЫЙ АНАЛИЗ

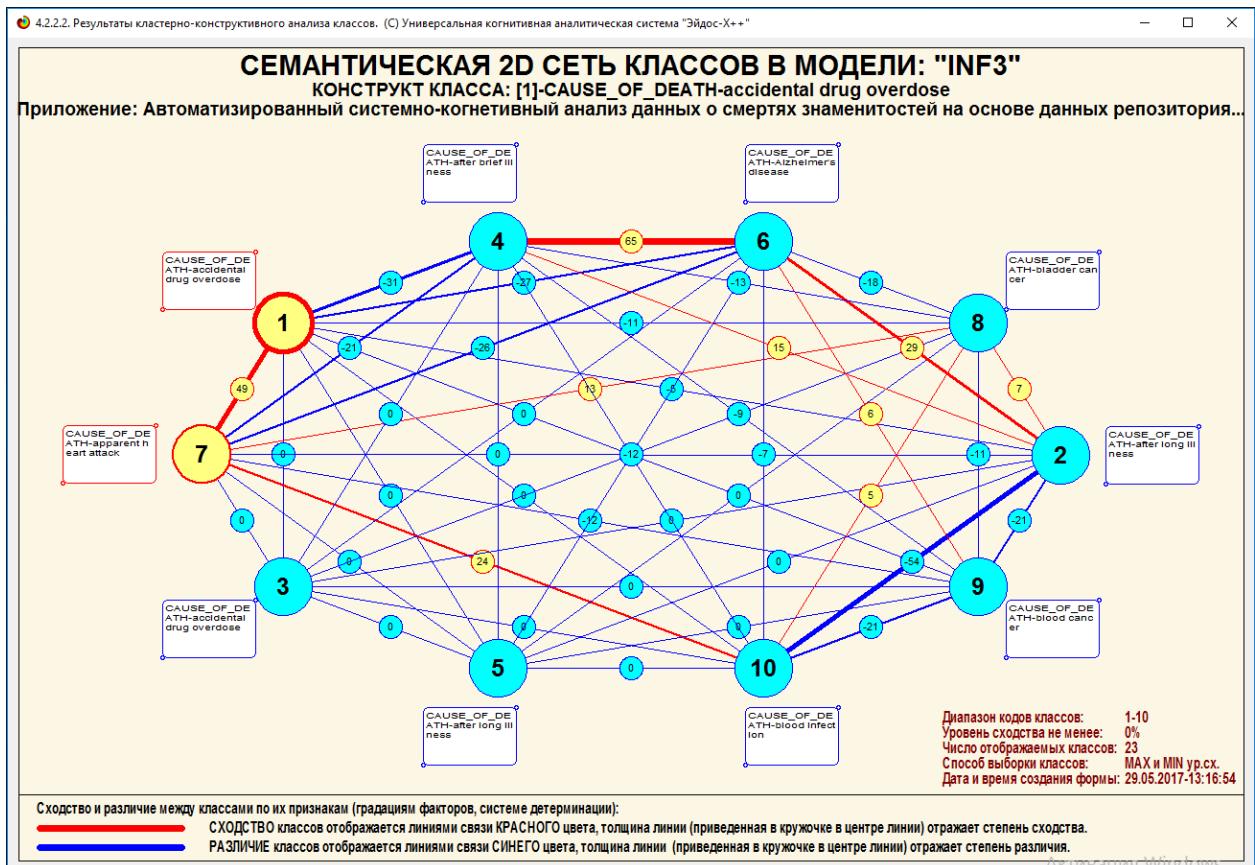


Рисунок 27. Пример кластерный и конструктивный анализ модели INF3

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 Kaggle. В данной курсовой работе приводится развернутый пример использования базы данных Kaggle для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». При этом наиболее достоверной в данном приложении оказались модели INF3, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,191, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 19%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием. 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 Kaggle, 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.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Луценко Е.В. Методика использования репозитория UCI для оценки качества математических моделей систем искусственного интеллекта / Е.В. Луценко // Политетматический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: http://lc.kubagro.ru/My_training_schedule.doc КубГАУ, 2003. – №02(002). С. 120 – 145. – IDA [article ID]: 0020302012. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 у.п.л.
2. Луценко Е.В. АСК-анализ, моделирование и идентификация живых существ на основе их фенотипических признаков / Е.В. Луценко, Ю.Н. Пенкина // Политетматический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №06(100). С. 1346 – 1395. – IDA [article ID]: 1001406090. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/90.pdf>, 3,125 у.п.л.
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 // Политетматический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.
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. [Электронный ресурс]. Статья "baza dannix sotrudnikov": <http://allexcel.ru/gotovye-tablitsy-excel-besplatno>, свободный. - Загл. с экрана. Яз. анг.
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- X++» / 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., Боровко А.Ю. ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА И КЛАССОВ СОЛНЕЧНЫХ ВСПЫШЕК НА ОСНОВЕ ИХ ПРЕДЫСТОРИИ ПО ДАННЫМ РЕПОЗИТОРИЯ UCI С ПРИМЕНЕНИЕМ АСК- АНАЛИЗА И ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ «ЭЙДОС» // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – IDA [article ID]: 0831209025. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/99.pdf>