

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ И.Т.ТРУБИЛИНА»

Факультет прикладной информатики

РЕЦЕНЗИЯ

на курсовой проект (работу)

Студента Филоненко Максим Викторович курса 3 очной формы обучения
направления подготовки Информационные системы и технологии.

направленность (профиль) академический бакалавриат.

Наименование темы «Разработка и исследование системно когнитивной модели
зависимости количества преступлений в Швеции от количества населения».

Рецензент:

Луценко Евгений Вениаминович, д. э. н.

(Ф.И.О., ученое звание и степень, должность)

Оценка качества выполнения курсового проекта (работы)

№ п/п	Показатель	Оценка соответствия заданию (по 5-и балльной шкале)
1.	Актуальность тематики работы	5
2.	Степень полноты обзора состояния проблемы и корректность постановки цели и задач исследования	5
3.	Уровень и корректность использования в работе различных методов исследований	5
4.	Степень комплексности работы, применения в ней знаний общепрофессиональных и специальных дисциплин	5
5.	Ясность, четкость, последовательность и обоснованность изложения	5
6.	Применение современных технологий обработки информации	5
7.	Качество оформления работы (общий уровень грамотности, стиль изложения, качество иллюстраций, соответствие требованиям по оформлению)	4
8.	Ответы на вопросы при защите	5

Достоинства работы

Применение современных математических моделей и программного
инструмента

Недостатки работы

Недостаточно качественное оформление

Итоговая оценка при защите _____

Рецензент _____

(Е.В. Луценко)

« _____ » _____ 201 ____ г.

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:

**«Разработка и исследование системно когнитивной модели влияния
посещений образовательных ресурсов студентами на их активность в
дискуссиях на основе данных репозитория UCI»**

выполнил студент группы ИТ1401

Филоненко Максим Викторович

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 _____ Оценка _____
(дата)

Члены комиссии _____ В.И. Лойко
_____ Е.В. Луценко
_____ И. В. Николаева
(подпись, дата, расшифровка подписи)

Краснодар
2017 г.

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

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

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

УТВЕРЖДАЮ:

Зав. кафедрой КТС _____ В. И. Лойко

ЗАДАНИЕ
на курсовую работу

Студенту: ИТ1401 группы 3 курса
Факультета прикладной информатики
Специальности: 09.03.02 Информационные системы и технологии
(шифр)

Филоненко Максиму Викторовичу
(Ф.И.О.)

Тема проекта: **Разработка и исследование системно когнитивной модели влияния посещений образовательных ресурсов студентами на их активность в дискуссиях 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	6
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 состава продуктов быстрого питания от выбранной категории.

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 сборник баз данных Kaggle.

В данной курсовой работе использована база данных «menu» 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 преступлений – «students.csv», которую оставим без изменений.

Общее описание задачи:

1. Пол
2. Национальность
3. Место рождения
4. Уровень
5. Ответственный родитель
6. Поднятые руки
7. VisITed ресурсы
8. Просмотр объявлений
9. Уведомление родителей
10. Класс
11. Пропуски
12. Кол-во дискуссий

Столбцы 1-11 описательные шкалы.

Столбцы 12-13 является классификационными шкалами. Эти столбцы показывают количество пропусков и дискуссий.

Обучающая выборка:

Таблица 1 – students.xls

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Пол	Национал	Место ро.	Уровень	Ответств	Подняты	VisiTed	р	Просмотр	объявлен	Уведомл	Класс	Пропуски	Кол-во дискуссий
2		M	KW	Kuwait	0	Father	15	16	2	Yes	M		2	20	
3		M	KW	Kuwait	0	Father	20	20	3	Yes	M		2	25	
4		M	KW	Kuwait	0	Father	10	7	0	No	L		5	30	
5		M	KW	Kuwait	0	Father	30	25	5	No	L		8	35	
6		M	KW	Kuwait	0	Father	40	50	12	No	M		4	50	
7		F	KW	Kuwait	1	Father	42	30	13	Yes	M		6	70	
8		M	KW	Kuwait	1	Father	35	12	0	No	L		6	17	
9		M	KW	Kuwait	1	Father	50	10	15	Yes	M		9	22	
10		F	KW	Kuwait	1	Father	12	21	16	Yes	M		7	50	
11		F	KW	Kuwait	1	Father	70	80	25	Yes	M		6	70	
12		M	KW	Kuwait	1	Father	50	88	30	Yes	H		8	80	
13		M	KW	Kuwait	0	Father	19	6	19	Yes	M		2	12	
14		M	KW	Kuwait	1	Father	5	1	0	No	L		8	11	
15		M	Lebanon	Lebanon	1	Father	20	14	12	No	L		9	19	
16		F	KW	Kuwait	1	Mum	62	70	44	No	H		5	60	
17		F	KW	Kuwait	1	Father	30	40	22	Yes	M		6	66	
18		M	KW	Kuwait	1	Father	36	30	20	No	M		8	80	
19		M	KW	Kuwait	1	Father	55	13	35	No	M		8	90	
20		F	KW	Kuwait	1	Mum	69	15	36	Yes	M		9	96	
21		M	KW	Kuwait	1	Mum	70	50	40	Yes	H		8	99	
22		F	KW	Kuwait	1	Father	60	60	33	No	M		3	90	
23		F	KW	Kuwait	1	Father	10	12	4	No	M		3	80	
24		M	KW	Kuwait	1	Father	15	21	2	No	M		3	90	
25		M	KW	Kuwait	1	Father	2	0	2	No	L		5	50	
26		M	KW	Kuwait	1	Father	0	2	3	Yes	L		3	70	
27		M	KW	Kuwait	1	Father	8	7	30	Yes	L		4	40	
28		M	KW	Kuwait	1	Father	19	19	25	Yes	M		6	40	
29		M	KW	Kuwait	1	Father	25	15	12	No	L		4	33	
30		M	KW	Kuwait	1	Father	75	85	52	Yes	M		2	12	

Всего в базе 16 столбцов, 2 классификационные и 14 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 xls-файл, для удобства дальнейшего редактирования и выполним следующие операции.

Конвертируем исходный файл из csv формата в xls и запишем его с именем: Inp_data.xls 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.

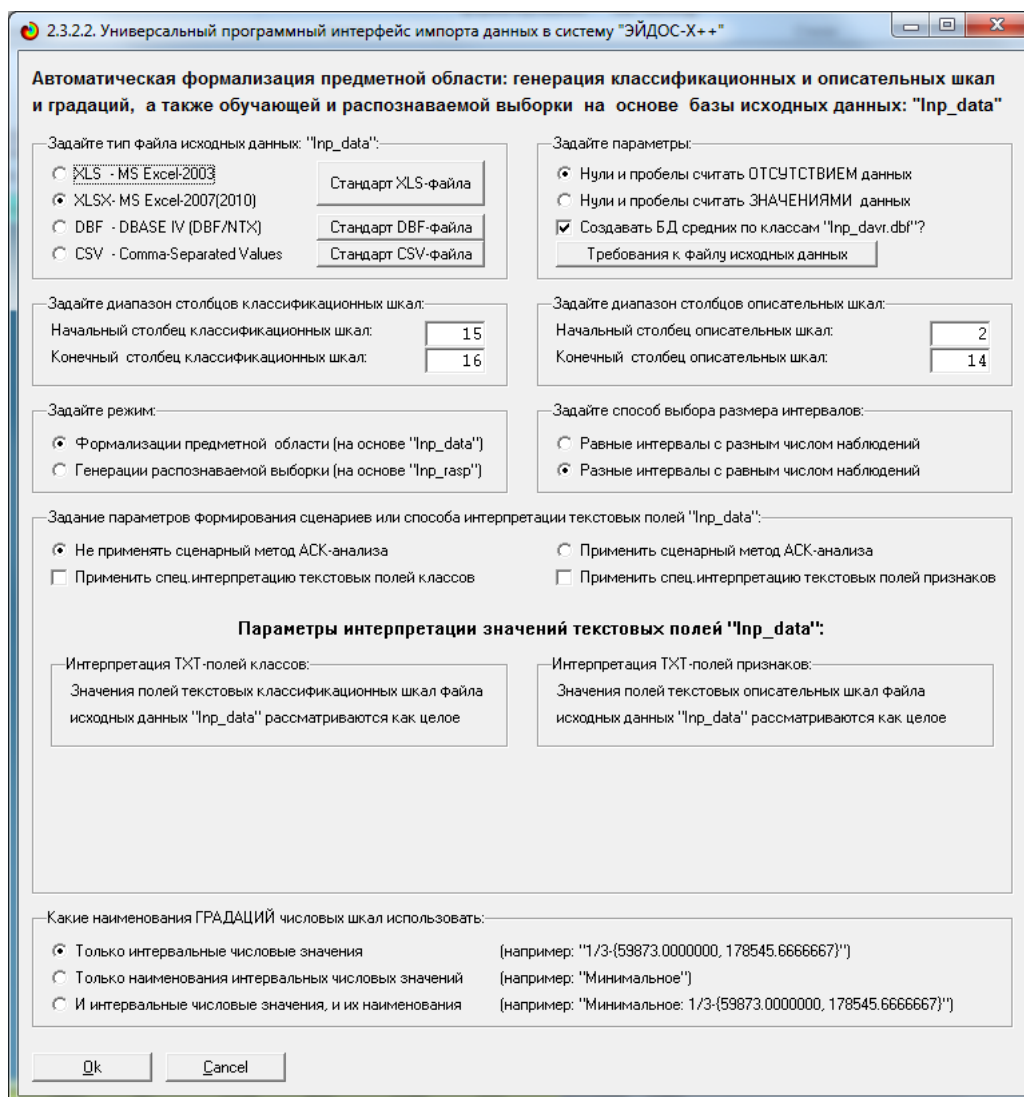


Рисунок 4. Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

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

– "Задайте тип файла исходных данных Inp_data": "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" – 24;
- "Задание параметров формирования сценариев или способа интерпретации текстовых полей": "Не применять сценарный метод АСК-анализа".

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".

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс. шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис. шкалу
Числовые	1	3	3,00	4	12	3,00
Текстовые	0	0	0,00	11	49	4,45
ВСЕГО:	1	3	3,00	15	61	4,07

Задайте число интервалов (градаций) в шкале:

В классификационных шкалах: В описательных шкалах:

Пересчитать шкалы и градации Выйти на создание модели

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 "ОК".

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

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [3 x 61]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс. шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис. шкалу
Числовые	1	3	3,00	4	12	3,00
Текстовые	0	0	0,00	11	49	4,45
ВСЕГО:	1	3	3,00	15	61	4,07

Задайте число интервалов (градаций) в шкале:

В классификационных шкалах: В описательных шкалах:

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	Код градации	Наименование градаций классификационной шкалы	DEL
1	ПРОПЫСКИ	1	1/3-(1.0000000, 4.0000000)	
2	КОЛ-ВО ДИСКУССИЙ	2	2/3-(4.0000000, 6.0000000)	
		3	3/3-(6.0000000, 9.0000000)	

Помощь Доб.шкалу Доб.град.шкалы Копир.шкалу Копир.град.шкалы Копир.шкалу с град. Удал.шкалу с град. Удал.град.шкалы Удаление и перекодирование

Рисунок 7. Классификационные шкалы и градации (фрагмент)

2.1. Классификационные шкалы и градации. Текущая модель: "INF1"

Код шкалы	Наименование классификационной шкалы
1	ПРОПЬСКИ
2	КОЛ-ВО ДИСКУССИЙ

Код градации	Наименование градации классификационной шкалы	DEL
1	1/3 (1.0000000, 4.0000000)	
2	2/3 (4.0000000, 6.0000000)	
3	3/3 (6.0000000, 9.0000000)	

Помощь Доб. шкалу Доб. град. шкалы Копир. шкалу Копир. град. шкалы Копир. шкалу с град. Удал. шкалу с град. Удал. град. шкалы Удаление и перекодирование

2.3.1. Ручной ввод-корректировка обучающей выборки. Текущая модель: "INF1"

Код объекта	Наименование объекта	Дата	Время
1	М	...	
2	М	...	
3	М	...	
4	М	...	
5	М	...	
6	М	...	
7	М	...	
8	М	...	
9	М	...	
10	М	...	

Код объекта	Класс 1	Класс 2	Класс 3	Класс 4
3	2	4	0	0

Код объекта	Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
3	4	12	17	21	23	26	32
3	35	0	0	0	0	0	0

Помощь Склонировать обуч.выб.в расп. Добавить объект Добавить классы Добавить признаки Удалить объект Удалить классы Удалить признаки Очистить БД

11

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).

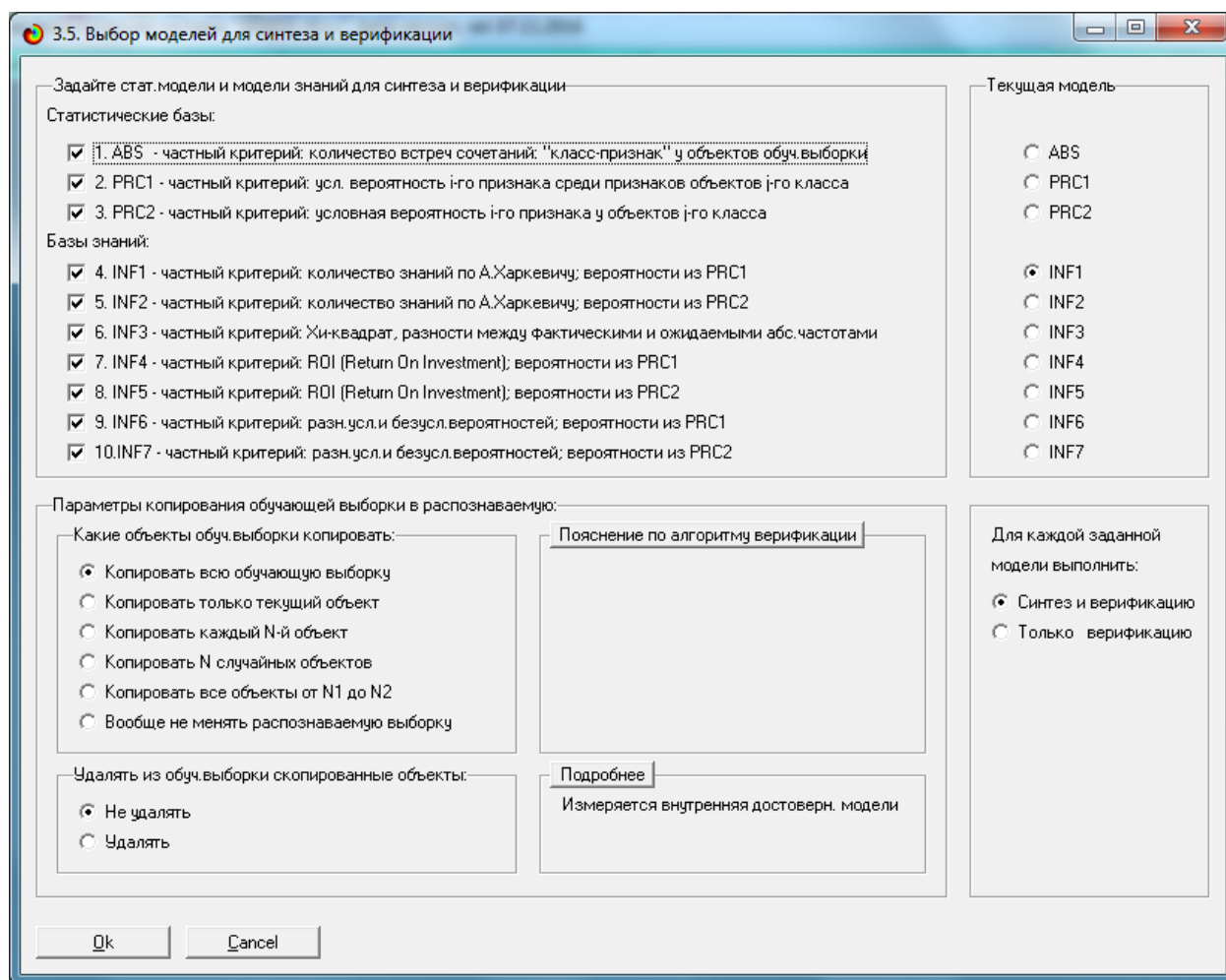


Рисунок 10. Выбор моделей для синтеза и верификации, а также текущей модели

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.

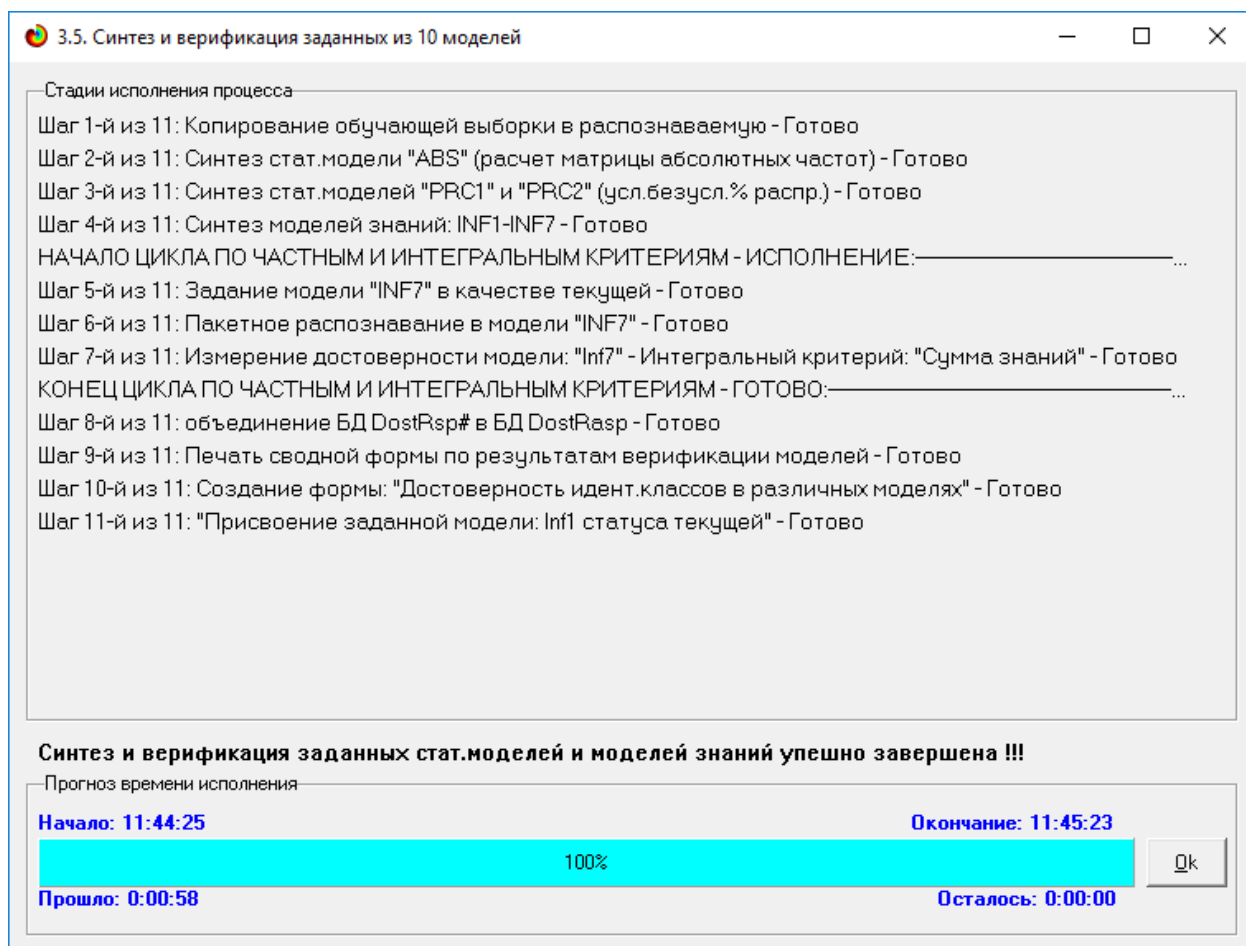


Рисунок 11. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 58 секунд. При этом верификация (оценка достоверности моделей) проводилась на всех 260 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 INF3, 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)

[illegible]

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

[illegible]

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

5.5. Модель "6, INF) - частный критерий Хи-квадрат: равенство между фактическими и ожидаемыми абсолютными частотами"										
Ряд, Рядовый №	Наименование элементарных событий (исходов)	1. ПРЕДПОСЫЛКА 1/1 P(1 00000) 0,0000000	2. ПРЕДПОСЫЛКА 2/1 P(2 00000) 0,0000000	3. ПРЕДПОСЫЛКА 3/1 P(3 00000) 0,0000000	4. ЕДП 4/1 P(4 00000) 0,0000000	5. ЕДП 5/1 P(5 00000) 0,0000000	6. ЕДП 6/1 P(6 00000) 0,0000000	Сумма	Среднее	Средн. квадратичн.
1	НАЦИОНАЛЬНОСТЬ-6/Egypt	-0,482	0,749	-0,266	-0,363	-0,302	0,665			0,553
2	НАЦИОНАЛЬНОСТЬ-2/EIran	-0,482	-0,261	0,754	-0,363	0,699	-0,335			0,568
3	НАЦИОНАЛЬНОСТЬ-3/EJordan	-1,411	0,743	0,668	0,183	0,490	-0,673			0,862
4	НАЦИОНАЛЬНОСТЬ-4/EKu	-1,289	0,957	0,432	1,137	-0,990	-0,127			0,903
5	НАЦИОНАЛЬНОСТЬ-5/EKuwait	0,036	-0,583	0,467	0,273	-0,604	0,321			0,452
6	НАЦИОНАЛЬНОСТЬ-6/ESaudiArabia	1,553	-0,754	-0,799	-0,090	0,094	-0,004			0,853
7	НАЦИОНАЛЬНОСТЬ-7/EUSA	2,589	-1,257	-1,332	0,183	0,490	-0,673			1,469
8	НАЦИОНАЛЬНОСТЬ-8/Evenzuela	-0,482	0,749	-0,266	-0,363	-0,302	0,665			0,553
9	МЕСТО РОЖДЕНИЯ-1/Egypt	-0,482	0,749	-0,266	-0,363	-0,302	0,665			0,553
10	МЕСТО РОЖДЕНИЯ-2/EIran	-0,482	-0,261	0,754	-0,363	0,699	-0,335			0,568
11	МЕСТО РОЖДЕНИЯ-3/EJordan	-1,411	0,743	0,668	0,183	0,490	-0,673			0,862
12	МЕСТО РОЖДЕНИЯ-4/EKuwait	-1,289	0,957	0,432	1,137	-0,990	-0,127			0,903
13	МЕСТО РОЖДЕНИЯ-5/EKuwait	0,036	-0,583	0,467	0,273	-0,604	0,321			0,452
14	МЕСТО РОЖДЕНИЯ-6/SaudiArabia	1,553	-0,754	-0,799	-0,090	0,094	-0,004			0,853
15	МЕСТО РОЖДЕНИЯ-7/EUSA	2,589	-1,257	-1,332	0,183	0,490	-0,673			1,469
16	МЕСТО РОЖДЕНИЯ-8/Evenzuela	-0,482	0,749	-0,266	-0,363	-0,302	0,665			0,553
17	УРОВЕНЬ-1/A0	5,268	-2,838	-2,452	2,375	3,015	-6,390			8,134
18	УРОВЕНЬ-2/A1	-7,145	2,929	8,224	-0,558	-1,495	2,307			4,448
19	УРОВЕНЬ-3/A2	1,392	0,467	-1,859	-1,994	0,356	1,538			1,575
20	УРОВЕНЬ-4/Control	0,510	-0,261	-0,266	-0,363	-0,302	0,665			0,442
21	ОТВЕТСТВЕННЫЙ РОДИТЕЛЬ-1/Father	1,229	-0,394	-0,835	1,783	2,709	-4,461			2,556

1.5. Результаты верификации моделей

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

[illegible]

Рисунок 12. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели ABS, INF4 при интегральном критерии «Семантический резонанс знаний». При этом точность модели составляет 0,676. 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).



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 INF4.

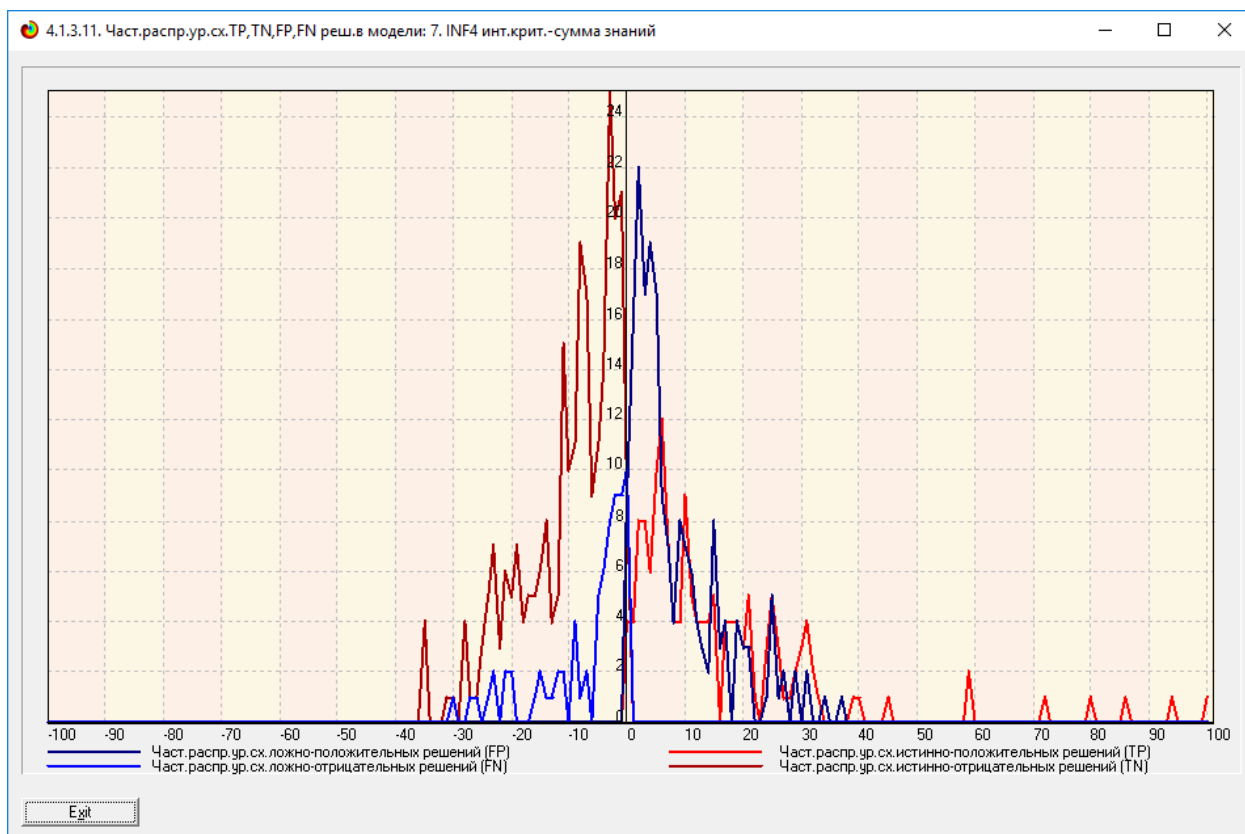


Рисунок 14. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF4

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 INF3 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.

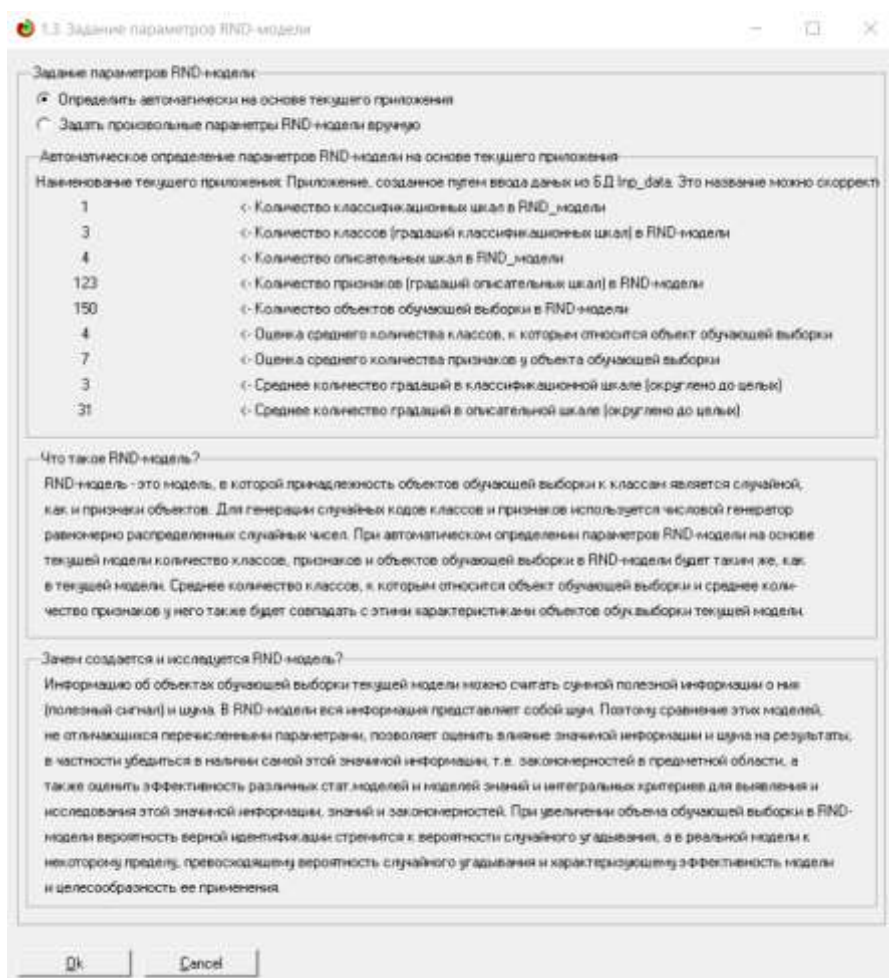


Рисунок 15. Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

На рисунке 13 показано частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF3.

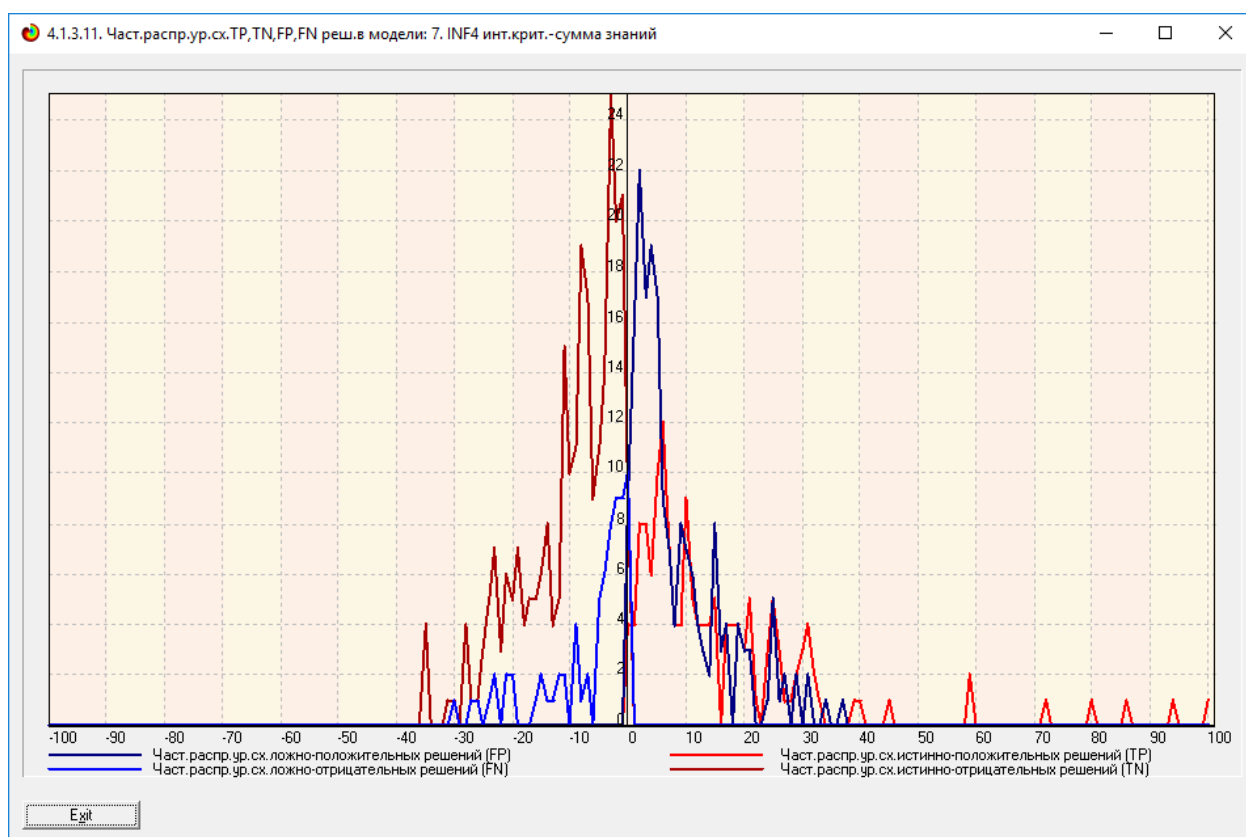


Рисунок 16. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF4

Совершенно очевидное различие частотных распределений уровней сходства-различия верно и ошибочно идентифицированных и не идентифицированных 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.

Идентификатор модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели	Достоверность модели
1. INF1 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2. INF2 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3. INF3 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4. INF4 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5. INF5 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6. INF6 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7. INF7 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8. INF8 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9. INF9 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10. INF10 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11. INF11 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12. INF12 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13. INF13 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14. INF14 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15. INF15 - частный случай: достоверность модели	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

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 INF4, 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 INF5 (режим 5.6) (рисунок 15) и проведем пакетное распознавание в режиме 4.1.2.

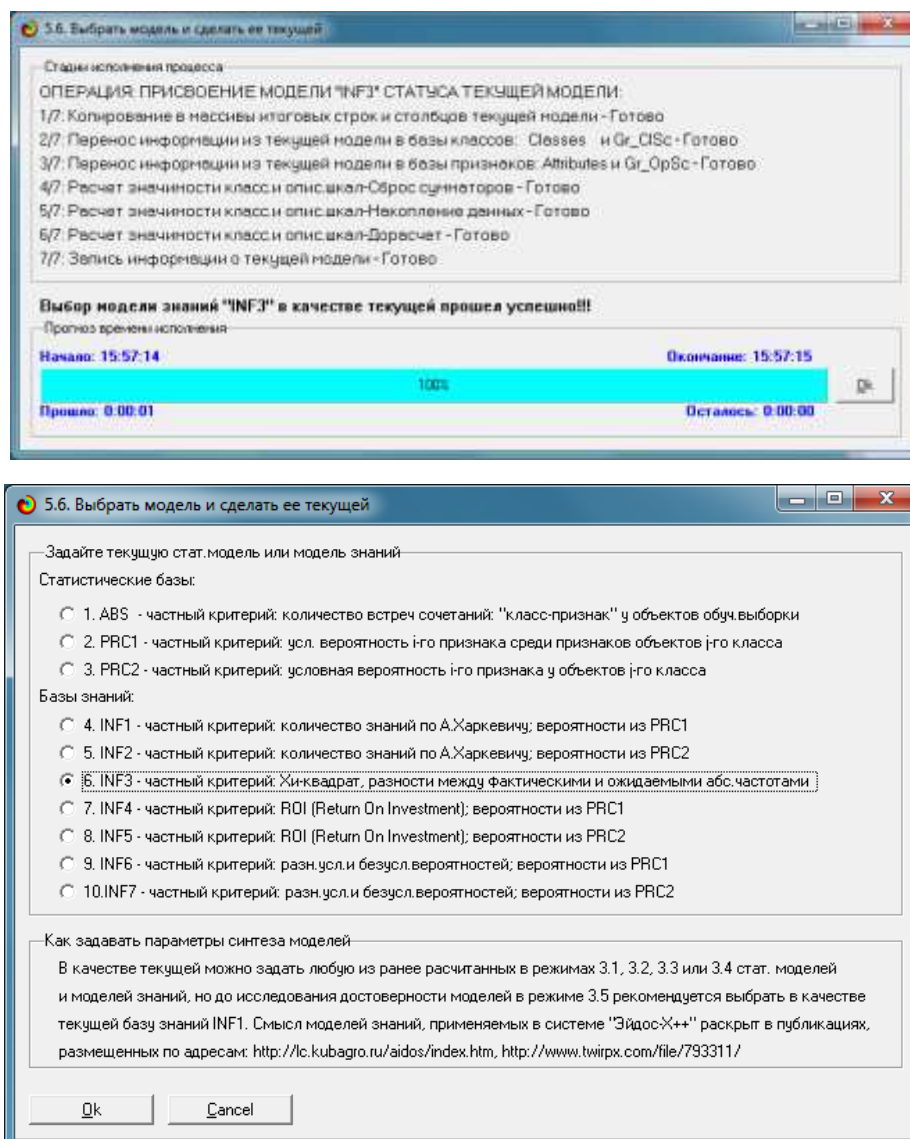


Рисунок 18. Экранные формы режима задания модели в качестве текущей

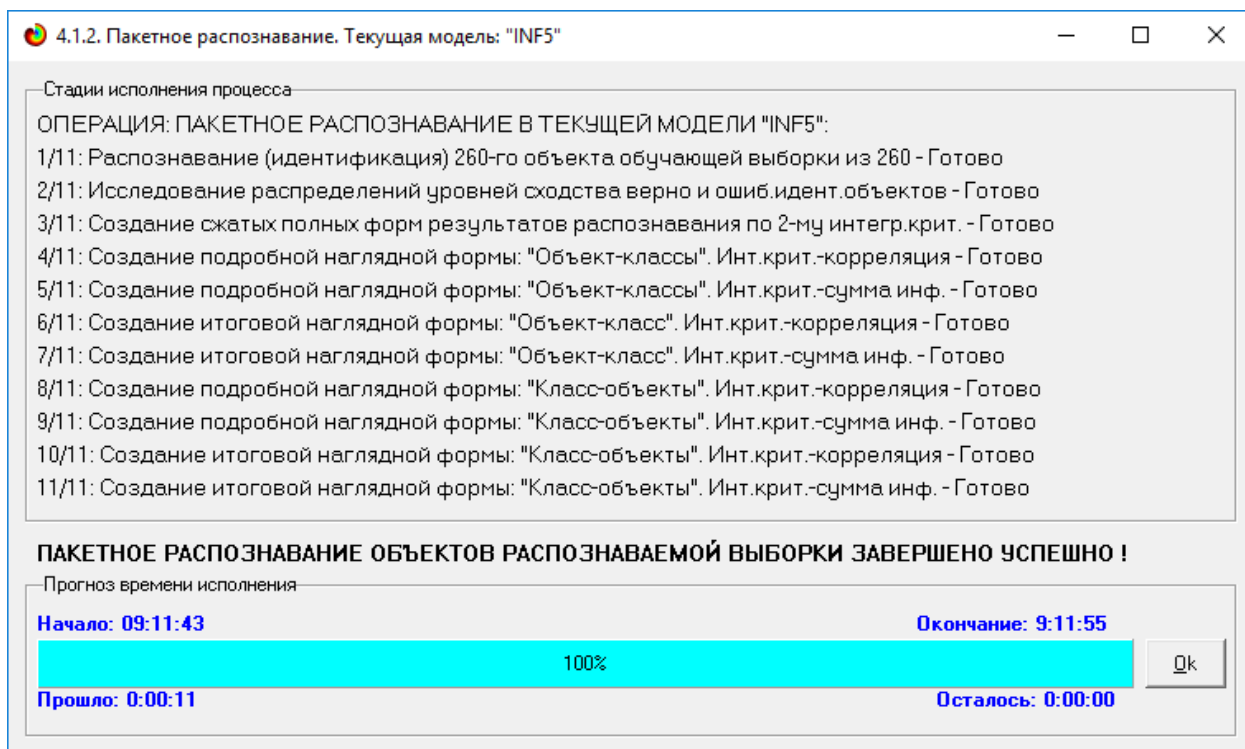


Рисунок 19. Экранная форма режима пакетного распознавания в текущей модели

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

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

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

8. Статистический анализ результатов идентификации по классам, моделям и интегральным критериям.

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

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

Ниже кратко рассмотрим некоторые из них.

На рисунках 17 и 18 приведены примеры прогнозов высокой и низкой достоверности частоты и классов ирисов в наиболее достоверной модели INF4 на основе наблюдения предыстории их развития:

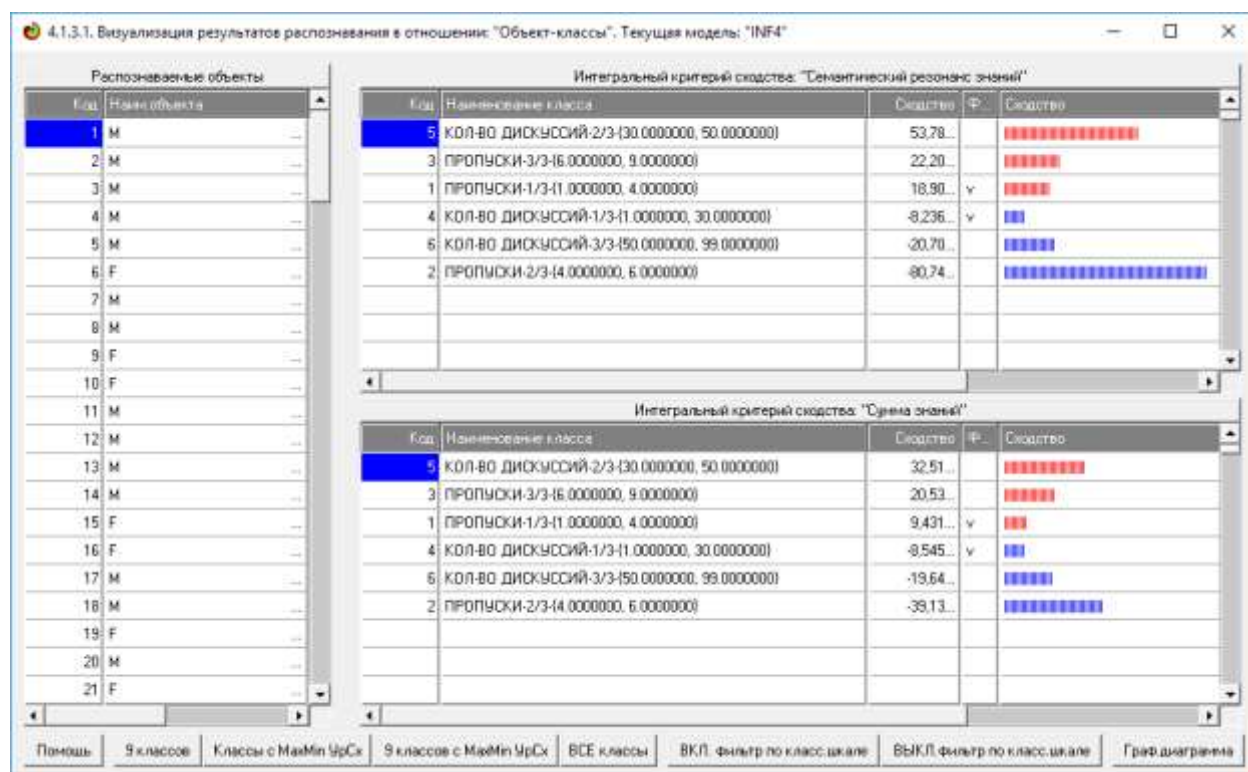


Рисунок 20. Пример идентификации классов в модели INF4

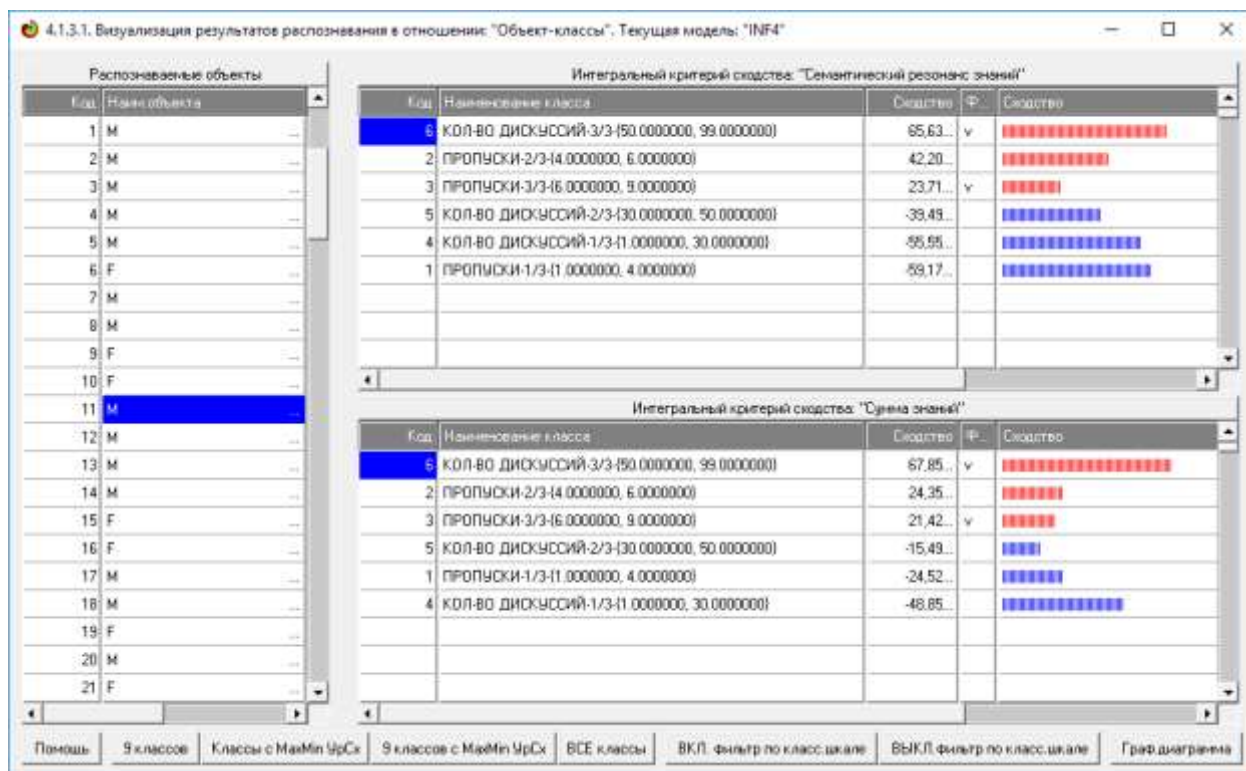


Рисунок 21. Пример идентификации классов в модели INF4

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 22)

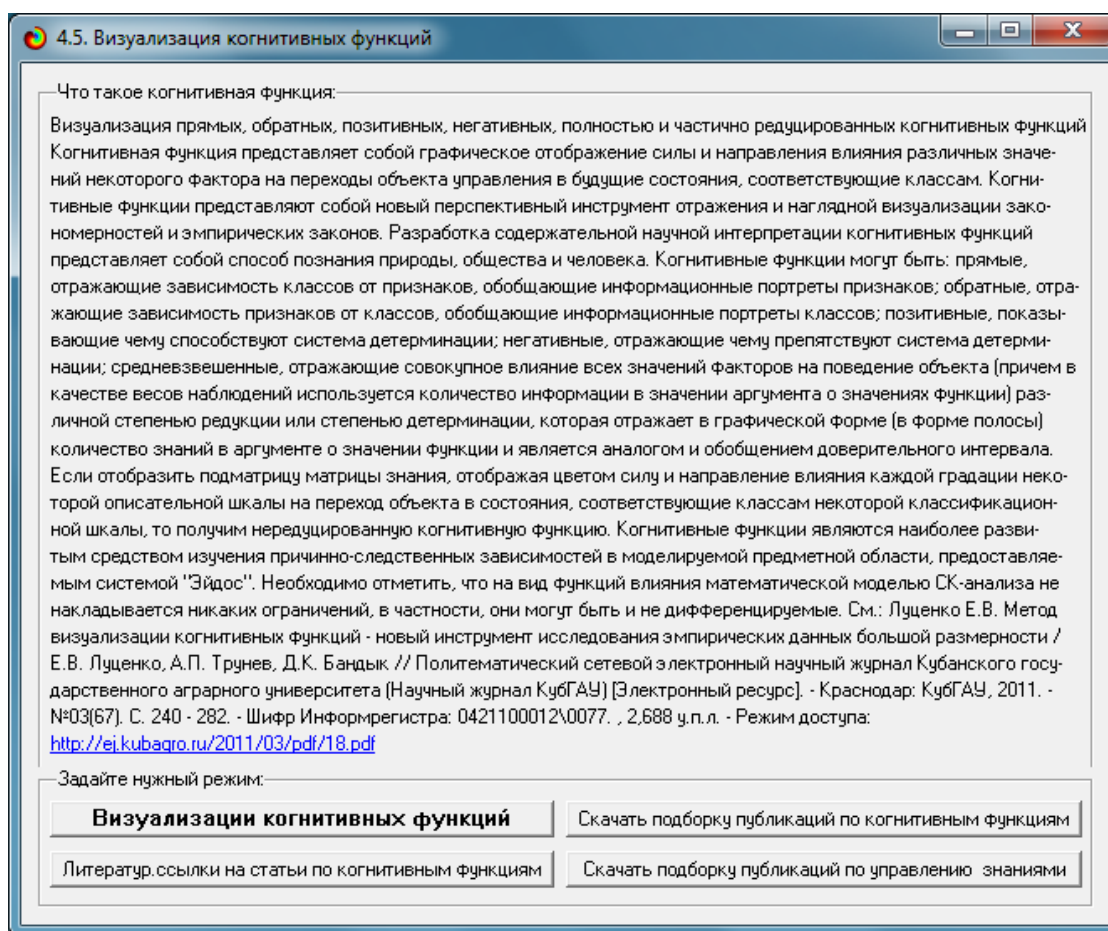


Рисунок 22. Экранная форма режима 4.5 системы «Эйдос-X++» «Визуализация когнитивных функций»

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 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 INF4.

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).

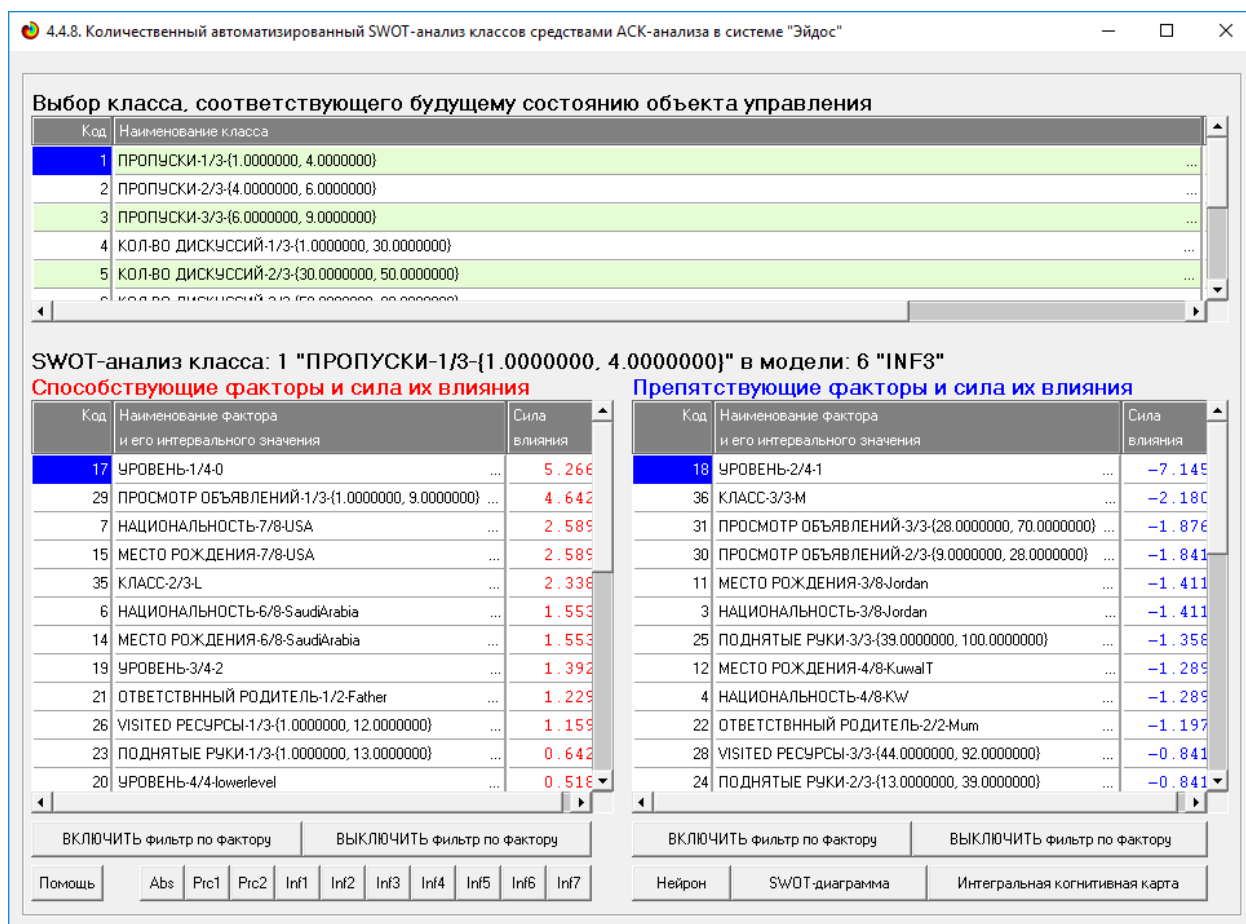


Рисунок 24. Пример SWOT-матрицы в модели INF3

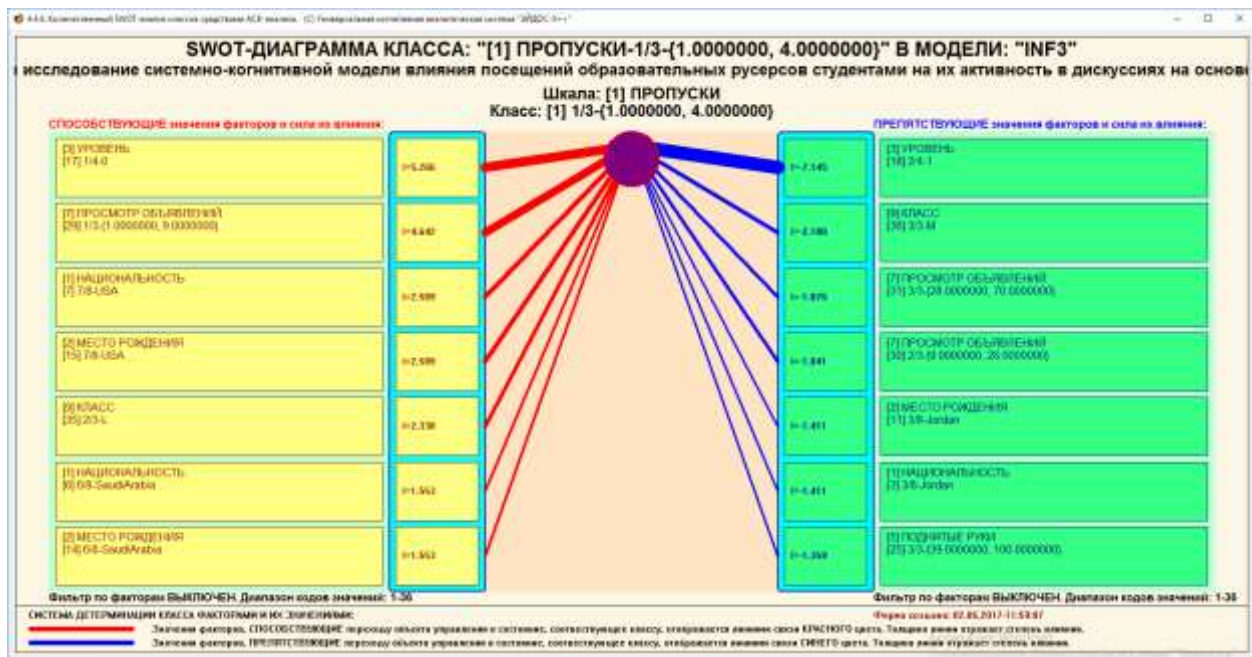


Рисунок 25 - Пример SWOT-матрицы в модели INF3

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.

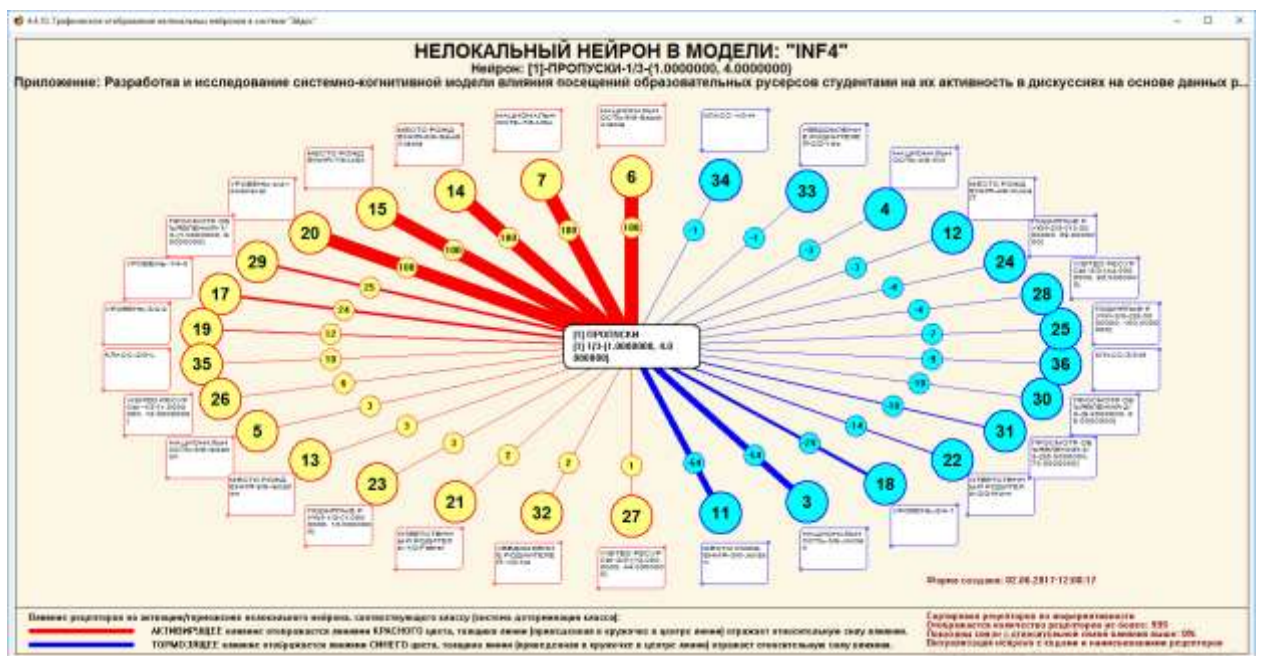


Рисунок 26. Пример нелокальные нейронные сети и нейроны модели INF4

2.5 КЛАСТЕРНЫЙ И КОНСТРУКТИВНЫЙ АНАЛИЗ

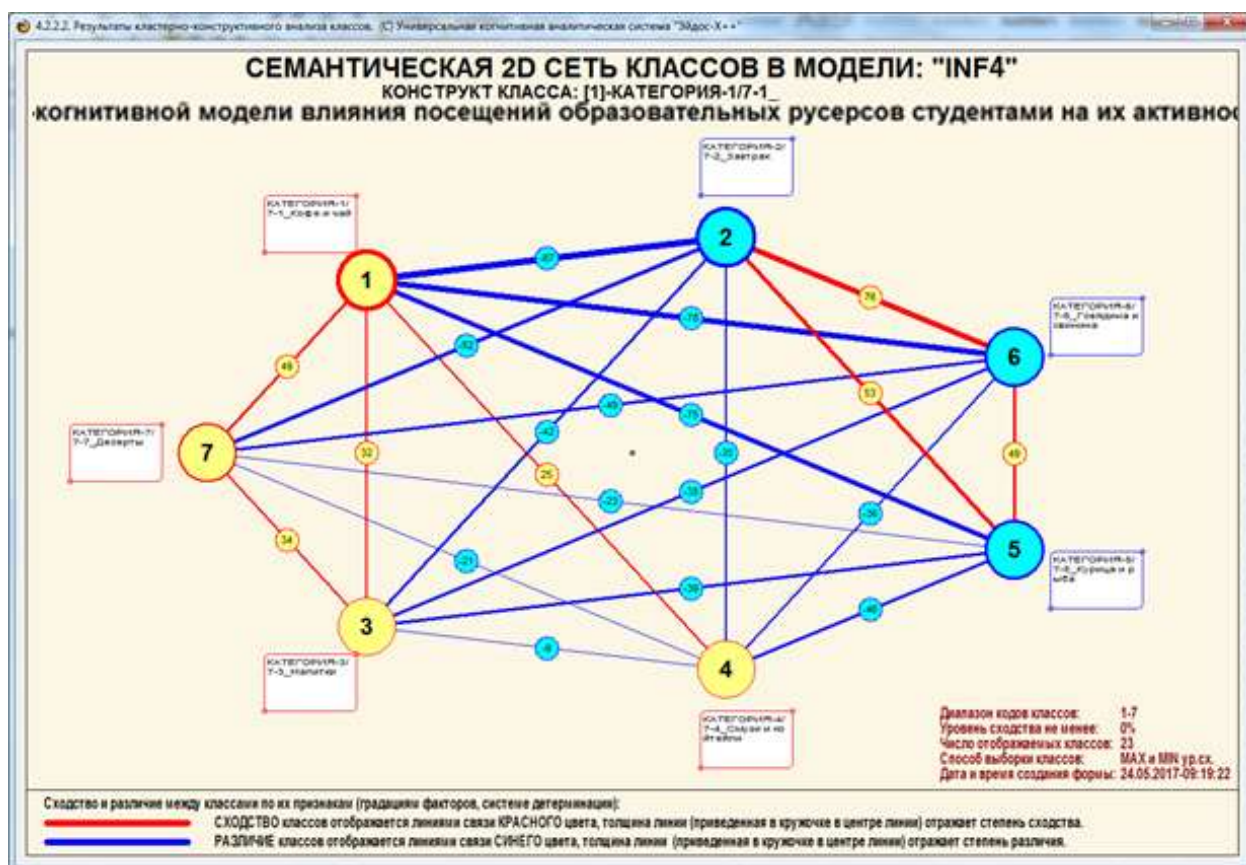


Рисунок 27. Пример кластерный и конструктивный анализ модели INF4

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 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 0,909, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 91%. 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 // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: http://lc.kubagro.ru/My_training_schedule.doc

КубГАУ, 2003. – №02(002). С. 120 – 145. – IDA [article ID]: 0020302012. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 у.п.л.

2. Луценко Е.В. АСК-анализ, моделирование и идентификация живых существ на основе их фенотипических признаков / Е.В. Луценко, Ю.Н. Пенкина // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2014. – №06(100). С. 1346 – 1395. – IDA [article ID]: 1001406090. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/90.pdf>, 3,125 у.п.л.

3. Луценко Е.В. Теоретические основы, технология и инструментарий 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 аграрного 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. [Электронный ресурс]. Статья "baza dannix sotrudnikov": <http://allexcel.ru/gotovye-tablitsy-excel-besplatno>, 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., 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>