

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
классификации б/у машин по марке и пробегу автомобиля на основе
данных репозитория UCI**

выполнила студентка группы ИТ1401 Шевченко Алексей 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

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

Шевченко Алексей Викторович
(Ф.И.О.)

Тема проекта: **Разработка и исследование системно-когнитивной модели
классификации б/у машин по марке и пробегу автомобиля 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.

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

Факультет Прикладная информатика

РЕЦЕНЗИЯ

на курсовой проект (работу)

Студента (ки) Шевченко Алексея Викторовича

курса 3 очной (заочной) формы обучения

направления подготовки 09.03.02

направленность (профиль) Информационные системы и технологии

Наименование темы « Разработка и исследование системно-когнитивной модели классификации б/у машин по марке и пробегу автомобиля на основе данных репозитория UCI ».

Рецензент: Луценко Е.В., д.э.н., к.т.н., профессор

(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	4
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.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	7
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	8
1.1. Описание решения	8
1.2. Преобразование исходных данных из HTML-формата в файл исходных данных MS Excel	8
1.3. Синтез и верификация статистических и интеллектуальных моделей	16
1.4. Виды моделей системы «Эйдос»	17
1.5. Результаты верификации моделей	19
2. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	26
2.1. Решение задачи идентификации	26
2.2. Когнитивные функции	29
2.3. SWOT и PEST матрицы и диаграммы	31
2.4. НЕЛОКАЛЬНЫЕ НЕЙРОННЫЕ СЕТИ И НЕЙРОНЫ	33
2.5. КЛАСТЕРНЫЙ И КОНСТРУКТИВНЫЙ АНАЛИЗ	35
ЗАКЛЮЧЕНИЕ	35
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	37

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 MicrosoftOfficeWord 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 HTML-формата в файл исходных данных MS Excel

Из электронного ресурса баз данных рассматриваемого предприятия возьмем базу данных преступлений – «MUP car.xls», которую оставим без изменений. Ссылка на базу - <https://www.kaggle.com/toramky/automobile-dataset>

Общее описание задачи:

1. Марка автомобиля
2. Цена
3. Вид кузова
4. Пробег
5. Калории
6. Вид топлива
7. Регистрация
8. Год
9. Модель
10. Привод

Столбцы 3-10 описательные шкалы.

Столбец 2 является классификационной шкалой. Этот столбец показывает цену автомобиля.

Обучающая выборка:

Таблица 1 – menu.xls

	A	B	C	D	E	F	G	H	I	J
1	car	price	body	mileage	engV	engType	registratic	year	model	drive
2	Ford	15500	crossover	68	2,5	Gas	yes	2010	Kuga	full
3	Mercedes	20500	sedan	173	1,8	Gas	yes	2011	E-Class	rear
4	Mercedes	35000	other	135	5,5	Petrol	yes	2008	CL 550	rear
5	Mercedes	17800	van	162	1,8	Diesel	yes	2012	B 180	front
6	Mercedes	33000	vagon	91	NA	Other	yes	2013	E-Class	
7	Nissan	16600	crossover	83	2	Petrol	yes	2013	X-Trail	full
8	Honda	6500	sedan	199	2	Petrol	yes	2003	Accord	front
9	Renault	10500	vagon	185	1,5	Diesel	yes	2011	Megane	front
10	Mercedes	21500	sedan	146	1,8	Gas	yes	2012	E-Class	rear
11	Mercedes	22700	sedan	125	2,2	Diesel	yes	2010	E-Class	rear
12	Nissan	20447,15	crossover	0	1,2	Petrol	yes	2016	Qashqai	front
13	Mercedes	20400	sedan	190	1,8	Gas	yes	2011	E-Class	rear
14	Mercedes	22500	sedan	164	1,8	Gas	yes	2012	E-Class	rear
15	BMW	4700	sedan	200	NA	Petrol	yes	1996	316	rear
16	Mercedes	21500	sedan	159	1,8	Gas	yes	2012	E-Class	rear
17	BMW	19999	sedan	290	4,8	Petrol	yes	2006	750	rear
18	BMW	129222	sedan	2	5	Petrol	yes	2016	750	full
19	Mercedes	99999	crossover	0	3	Petrol	yes	2016	GLE-Class	full
20	Nissan	16600	crossover	83	2	Petrol	yes	2013	X-Trail	full
21	BMW	73900	sedan	57	4,4	Petrol	yes	2013	M5	rear
22	Land Rove	0	crossover	0	4,4	Diesel	yes	2016	Range Rov	full
23	Nissan	26033,55	crossover	0	1,6	Diesel	yes	2016	X-Trail	full
24	BMW	104999	crossover	2	3	Diesel	yes	2016	X5	full
25	BMW	66500	crossover	1	2	Diesel	yes	2016	X5	full
26	BMW	65099	crossover	0	2	Diesel	yes	2016	X5	full
27	BMW	23900	crossover	235	3	Diesel	yes	2007	X5	full
28	Mercedes	69999	crossover	0	2,2	Diesel	yes	2016	GLE-Class	full
29	BMW	66200	crossover	70	3	Diesel	yes	2014	X5	full
30	BMW	63000	crossover	0	2	Diesel	yes	2015	X5	full
31	Mercedes	105999	crossover	0	2,98	Diesel	yes	2016	GLE-Class	full
32	Nissan	13980	hatch	31	NA	Other	yes	2013	Leaf	front
33	Nissan	17300	hatch	24	NA	Other	yes	2013	Leaf	front
34	Volkswag	10700	sedan	147	2	Petrol	yes	2007	Passat B6	front
35	Audi	2600	vagon	273	2,5	Diesel	no	1999	A6	front
36	Chrysler	13700	sedan	70	2,4	Petrol	yes	2008	Sebring	front

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.

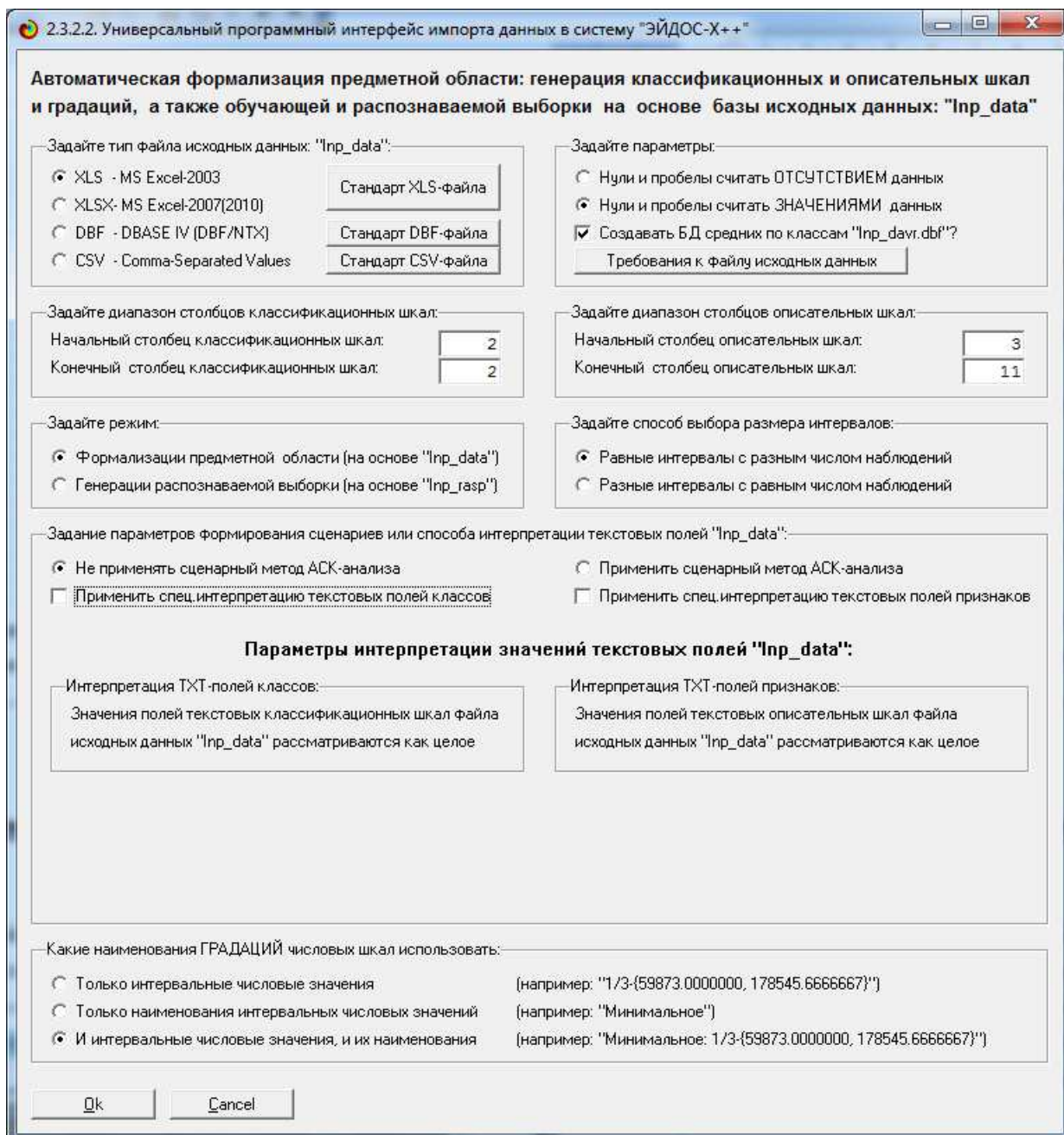


Рисунок 4. Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

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".

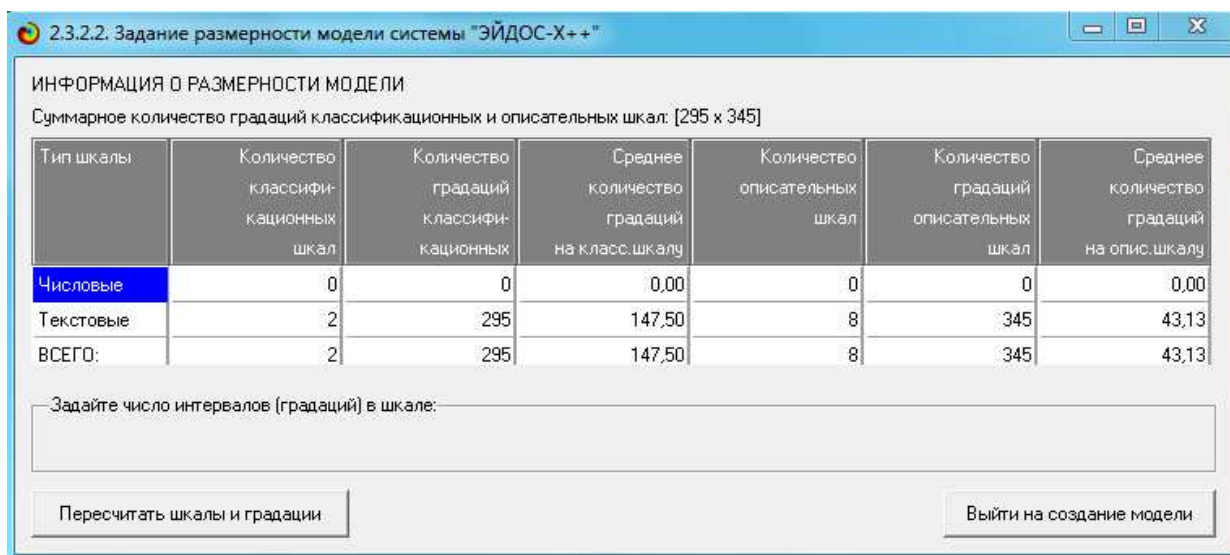


Рисунок 5. Задание размерности модели системы "Эйдос"

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".

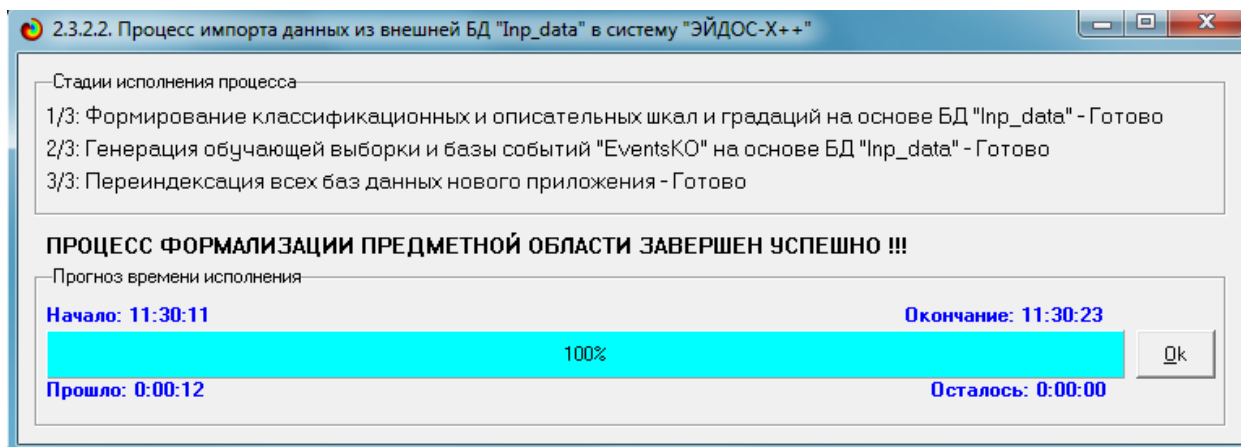


Рисунок 6. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

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).

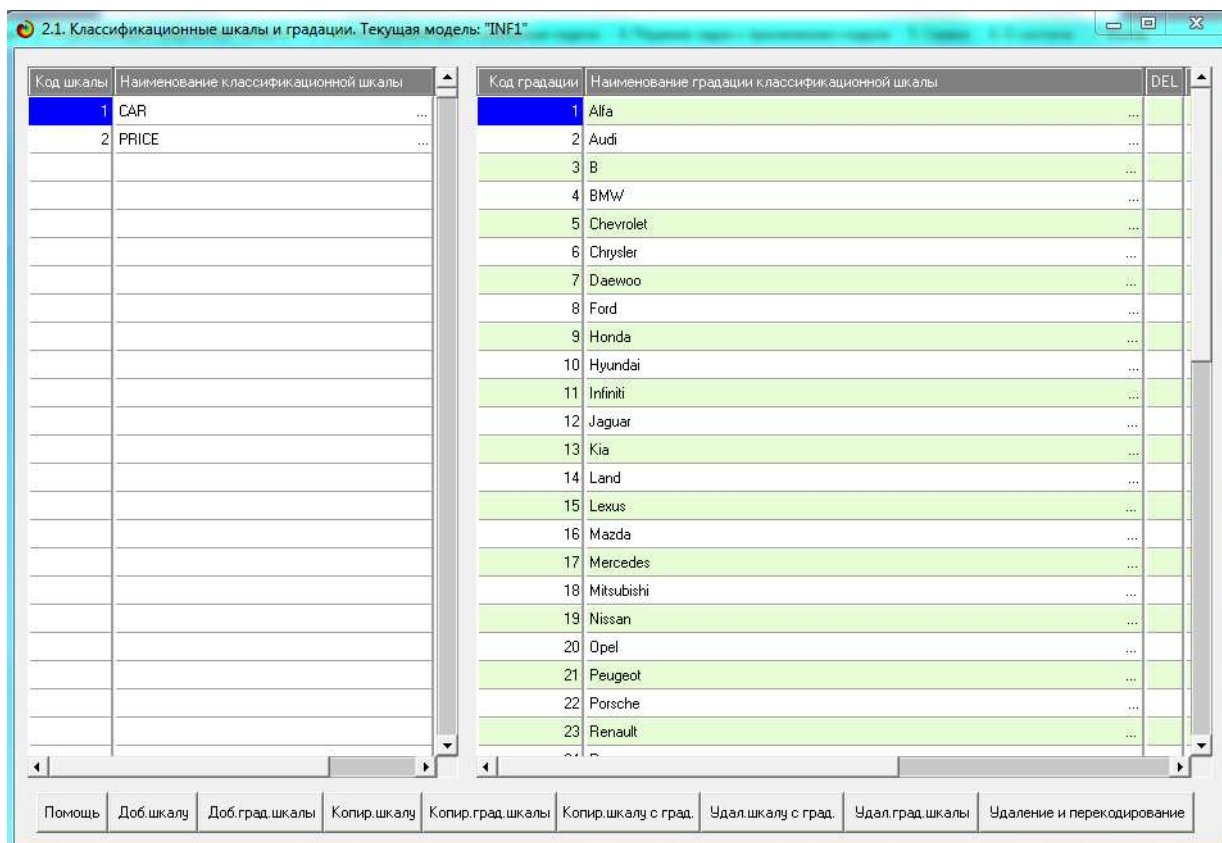


Рисунок 7. Классификационные шкалы и градации (фрагмент)

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

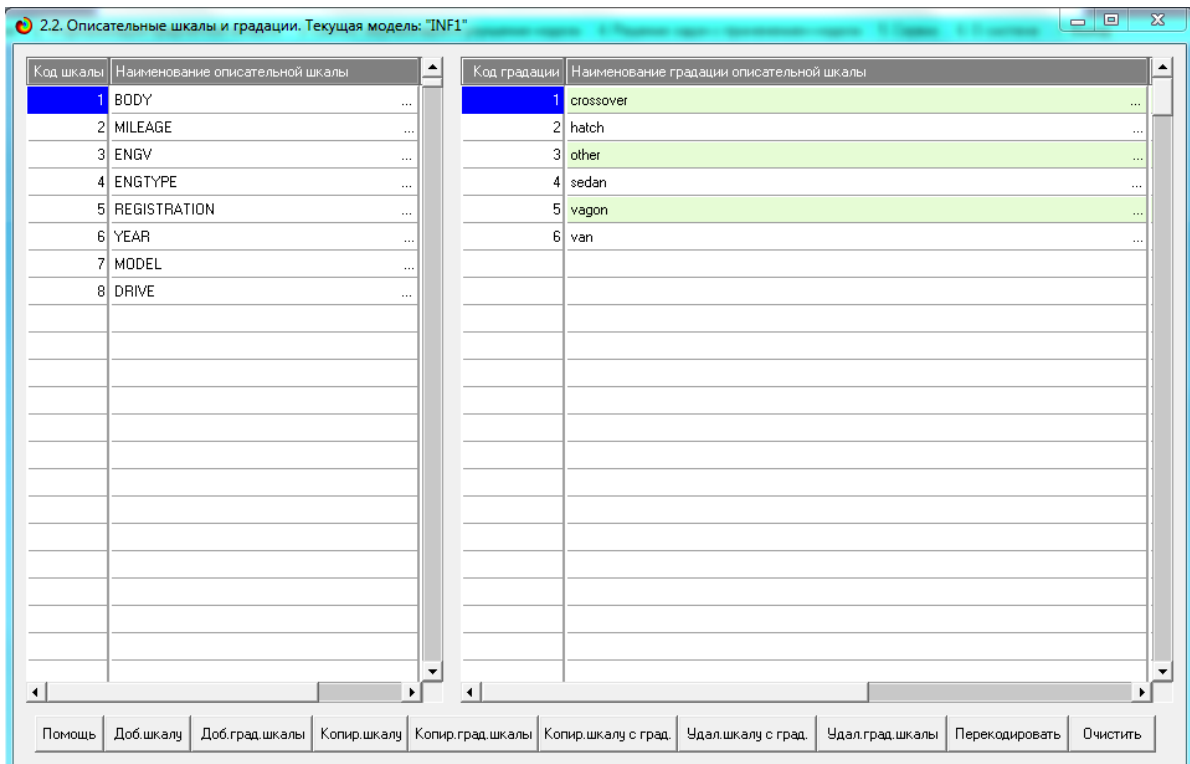


Рисунок 8. Описательные шкалы и градации (фрагмент)

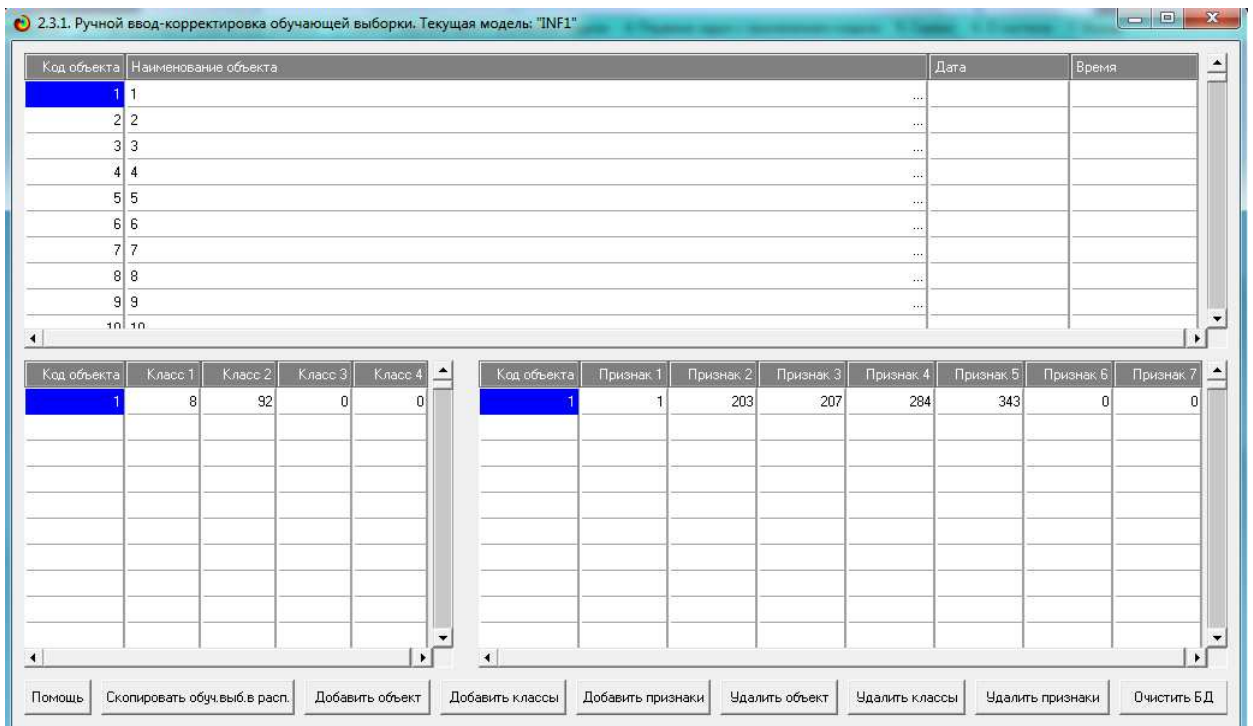


Рисунок 9. Обучающая выборка (фрагмент)

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).

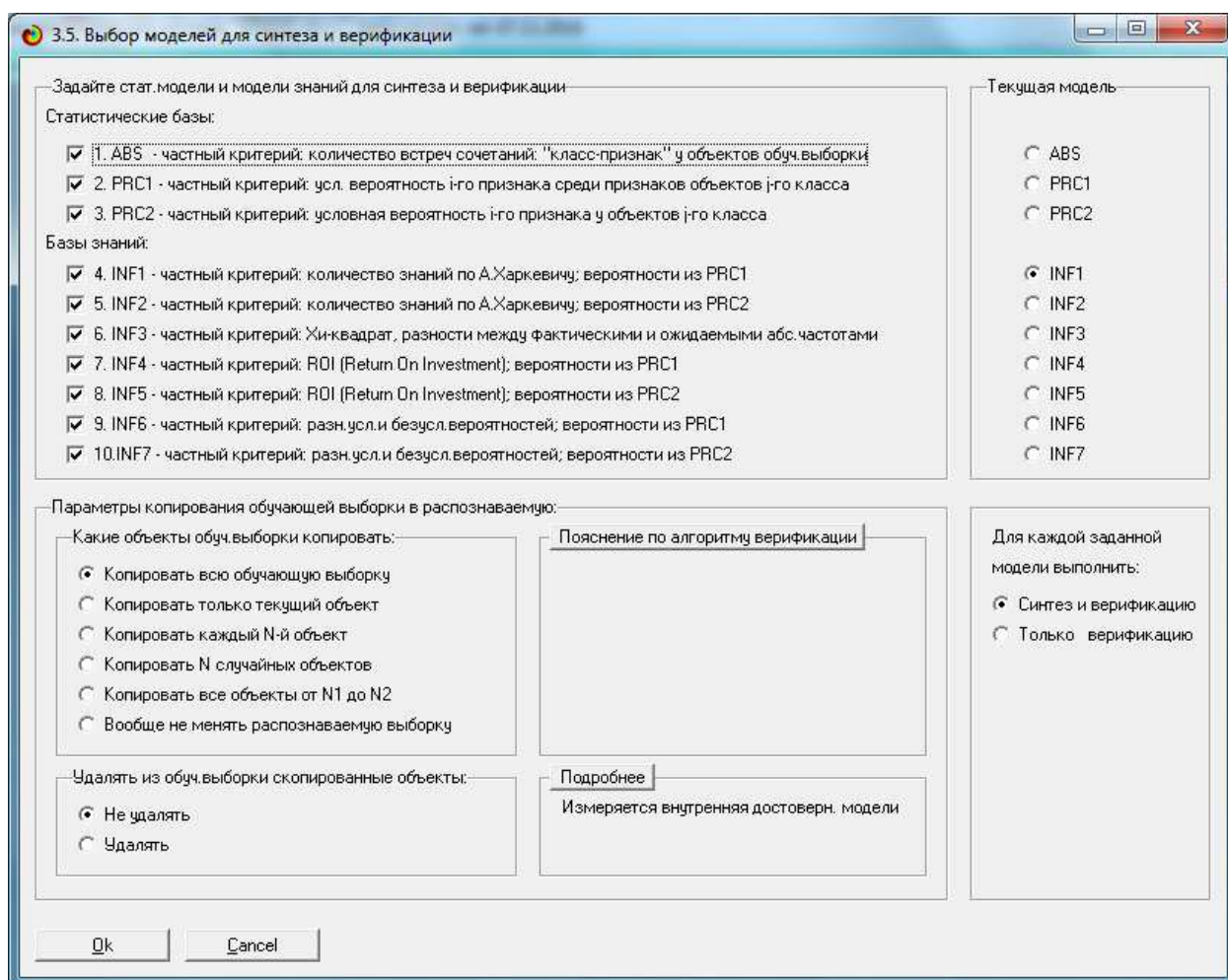


Рисунок 10. Выбор моделей для синтеза и верификации, а также текущей модели

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.

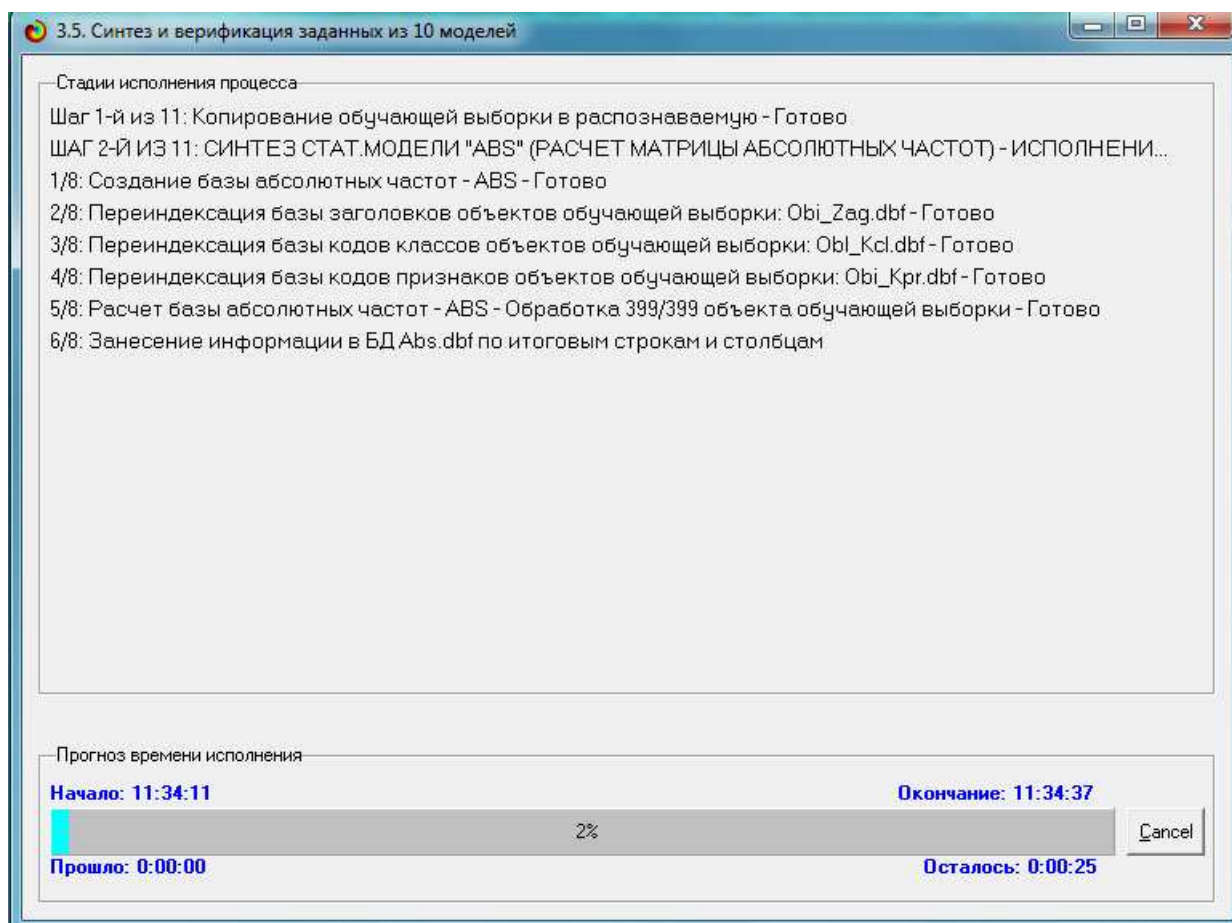


Рисунок 11. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 141 секунд¹. 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 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)

Код признака	Наименование описательной шкалы и градации	1. CAR ALFA	2. CAR AUDI	3. CAR B	4. CAR BMW	5. CAR CHEVROLET	6. CAR CHRYSLER	7. CAR DAEWOOD	8. CAR FORD	9. CAR HONDA	10. CAR HYUNDAI	11. CAR INFINITI	12. CAR JAGUAR	13. CAR KIA	14. LA
1	BODY-crossover		-0.568	-0.227	0.344				0.185		0.048	0.065			0.515
2	BODY-hatch	2.898						1.570	1.320	0.422	0.504				1.102
3	BODY-other		-0.695	0.678	-0.225										0.389
4	BODY-sedan		0.409	0.047	0.199	1.095	1.095	0.665	-0.264	0.076	0.278		1.095		
5	BODY-vagon		1.229	-0.382	-0.888					0.405					
6	BODY-van			0.842											
7	MILEAGE-0														
8	MILEAGE-1														
9	MILEAGE-2														
10	MILEAGE-3														
11	MILEAGE-6														
12	MILEAGE-8														
13	MILEAGE-9														
14	MILEAGE-11														
15	MILEAGE-12														
16	MILEAGE-13														
17	MILEAGE-14														
18	MILEAGE-15														
19	MILEAGE-16														
20	MILEAGE-17														
21	MILEAGE-18														
22	MILEAGE-19														
23	MILEAGE-20														
24	MILEAGE-21														

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

Код признака	Наименование описательной шкалы и градации	1. CAR ALFA	2. CAR AUDI	3. CAR B	4. CAR BMW	5. CAR CHEVROLET	6. CAR CHRYSLER	7. CAR DAEWOOD	8. CAR FORD	9. CAR HONDA	10. CAR HYUNDAI	11. CAR INFINITI	12. CAR JAGUAR	13. CAR KIA	14. LA
1	BODY-crossover	-0.331	-5.500	-5.732	6.520	-2.483	-0.414	-2.566	0.689	-4.139	0.192	0.596	-0.414	2.861	
2	BODY-hatch	0.948	-1.963	-4.355	-2.431	-0.390	-0.065	1.597	1.480	0.350	0.402	-0.065	-0.065	1.350	
3	BODY-other	-0.054	-1.032	4.493	-0.516	-0.404	-0.067	-0.417	-0.538	-0.673	-0.619	-0.067	-0.067	0.327	
4	BODY-sedan	-0.262	5.117	1.074	2.761	4.036	0.673	1.971	-0.618	4.727	0.989	-0.327	0.673	-3.273	
5	BODY-vagon	-0.053	5.003	-1.431	-1.473	-0.397	-0.066	-0.410	-0.529	0.339	-0.608	-0.066	-0.066	-0.661	
6	BODY-van	-0.056	-2.101	6.340	-2.601	-0.417	-0.070	-0.431	-0.556	-0.696	-0.640	-0.070	-0.070	-0.696	
7	MILEAGE-0														
8	MILEAGE-1														
9	MILEAGE-2														
10	MILEAGE-3														
11	MILEAGE-6														
12	MILEAGE-8														
13	MILEAGE-9														
14	MILEAGE-11														
15	MILEAGE-12														
16	MILEAGE-13														
17	MILEAGE-14														
18	MILEAGE-15														
19	MILEAGE-16														
20	MILEAGE-17														
21	MILEAGE-18														
22	MILEAGE-19														
23	MILEAGE-20														
24	MILEAGE-21														

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

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	Средний индекс уровня сходства: логно/логноит. решений	Средний индекс уровня сходства: логно/логноит. решений	А.Точность модели = АР/АПР	А.Полнота модели = АР/АПР	L2 мера проф. Е.В. Лазарева	Процент правильной идентификации	Процент правильной не идентификации	Процент ошибочной идентификации	Процент ошибочной не идентификации	Процент правильных результатов	Дата получения результата	Время получения результата
1. ABS - частный критерий: количество встреч сочетаний "Клас"	Корреляция абс. частот с обр...	0.314	0.047	0.682	0.999	0.810	99.615	8.261	91.739	0.385	53.938	20.05.2017	08:57:49
1. ABS - частный критерий: количество встреч сочетаний "Клас"	Сумма абс. частот по призна...	0.173		0.735	1.000	0.847	100.000	1.314	98.686		50.657	20.05.2017	08:57:49
2. PRC1 - частный критерий: усл. вероятность n-го признака сред.	Корреляция усл.отн. частот с о...	0.314	0.047	0.682	0.999	0.810	99.615	8.261	91.739	0.385	53.938	20.05.2017	08:58:00
2. PRC1 - частный критерий: усл. вероятность n-го признака сред.	Сумма усл.отн. частот по приз...	0.352		0.628	1.000	0.771	100.000	1.314	98.686		50.657	20.05.2017	08:58:00
3. PRC2 - частный критерий: условная вероятность n-го признака...	Корреляция усл.отн. частот с о...	0.314	0.047	0.682	0.999	0.810	99.615	8.261	91.739	0.385	53.938	20.05.2017	08:58:11
3. PRC2 - частный критерий: условная вероятность n-го признака...	Сумма усл.отн. частот по приз...	0.284		0.633	1.000	0.775	100.000	1.314	98.686		50.657	20.05.2017	08:58:11
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в.	Семантический резонанс зна...	0.184	0.328	0.654	0.673	0.663	79.615	76.371	23.629	20.385	77.993	20.05.2017	08:58:22
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в.	Сумма знаний	0.178	0.259	0.642	0.645	0.644	91.538	67.086	32.914	8.462	79.312	20.05.2017	08:58:22
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в.	Семантический резонанс зна...	0.188	0.327	0.612	0.625	0.618	82.692	77.762	22.238	17.308	80.227	20.05.2017	08:58:33
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в.	Сумма знаний	0.196	0.253	0.617	0.648	0.632	91.923	68.067	31.933	8.077	79.995	20.05.2017	08:58:34
6. INF3 - частный критерий: Хинкватат, разности между факти...	Семантический резонанс зна...	0.358	0.332	0.609	0.787	0.687	95.385	65.090	34.910	4.615	80.237	20.05.2017	08:58:45
6. INF3 - частный критерий: Хинкватат, разности между факти...	Сумма знаний	0.174	0.196	0.705	0.789	0.744	95.385	65.090	34.910	4.615	80.237	20.05.2017	08:58:45
7. INF4 - частный критерий: ROI (Return On Investment), вероятн...	Семантический резонанс зна...	0.184	0.241	0.695	0.726	0.710	68.077	79.824	20.176	31.923	73.950	20.05.2017	08:58:56
7. INF4 - частный критерий: ROI (Return On Investment), вероятн...	Сумма знаний	0.126	0.057	0.668	0.963	0.789	98.846	54.125	45.875	1.154	76.485	20.05.2017	08:58:56
8. INF5 - частный критерий: ROI (Return On Investment), вероятн...	Семантический резонанс зна...	0.208	0.248	0.655	0.723	0.687	70.000	82.085	17.915	30.000	76.043	20.05.2017	08:59:07
8. INF5 - частный критерий: ROI (Return On Investment), вероятн...	Сумма знаний	0.122	0.048	0.656	0.927	0.768	98.846	55.176	44.824	1.154	77.011	20.05.2017	08:59:07
9. INF6 - частный критерий: разн.усли безул.вероятностей, вер...	Семантический резонанс зна...	0.330	0.217	0.615	0.744	0.673	95.000	61.781	38.219	5.000	78.390	20.05.2017	08:59:18
9. INF6 - частный критерий: разн.усли безул.вероятностей, вер...	Сумма знаний	0.278	0.130	0.587	0.786	0.672	95.385	55.336	44.664	4.615	75.360	20.05.2017	08:59:18
10. INF7 - частный критерий: разн.усли безул.вероятностей: ве...	Семантический резонанс зна...	0.329	0.201	0.609	0.800	0.691	94.615	63.221	36.779	5.385	78.918	20.05.2017	08:59:29
10. INF7 - частный критерий: разн.усли безул.вероятностей: ве...	Сумма знаний	0.216	0.091	0.582	0.852	0.692	95.385	56.588	43.412	4.615	75.986	20.05.2017	08:59:29

Рисунок 12. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались 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).

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 INF5.

Рисунок 14. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF5

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 очень больших значений уровней сходства-различия (более 90%) 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 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 INF5

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	Средний модельный уровень сходства/ложно-положит. решений	Средний модельный уровень сходства/ложно-отрицат. решений	A.T. точность модели ATR/ATR	A.T. погрешность модели ATR/ATR	L2-мера проф. Е.В. Падеева	Процент правильной идентификации...	Процент правильной не-идентификации...	Процент ошибочной идентификации...	Процент ошибочной не-идентификации...	Процент правильных результатов	Дата получения результата	Время получения результата
1. ABS - частный критерий: количество встреч сонетаний "Клас...	Корреляция абс. частот с обр...	0.314	0.047	0.682	0.999	0.810	99.615	8.261	91.739	0.385	53.938	20.05.2017	08:57:49
1. ABS - частный критерий: количество встреч сонетаний "Клас...	Сумма абс. частот по признак...	0.173		0.735	1.000	0.847	100.000	1.314	98.686		50.657	20.05.2017	08:57:49
2. PRC1 - частный критерий: усл. вероятность n-го признака сред...	Корреляция усл.отн. частот с о...	0.314	0.047	0.682	0.999	0.810	99.615	8.261	91.739	0.385	53.938	20.05.2017	08:58:00
2. PRC1 - частный критерий: усл. вероятность n-го признака сред...	Сумма усл.отн. частот по приз...	0.352		0.628	1.000	0.771	100.000	1.314	98.686		50.657	20.05.2017	08:58:00
3. PRC2 - частный критерий: условная вероятность n-го признака...	Корреляция усл.отн. частот с о...	0.314	0.047	0.682	0.999	0.810	99.615	8.261	91.739	0.385	53.938	20.05.2017	08:58:11
3. PRC2 - частный критерий: условная вероятность n-го признака...	Сумма усл.отн. частот по приз...	0.284		0.633	1.000	0.775	100.000	1.314	98.686		50.657	20.05.2017	08:58:11
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	0.184	0.328	0.654	0.673	0.663	79.615	76.371	23.629	20.385	77.993	20.05.2017	08:58:22
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	0.178	0.259	0.642	0.645	0.644	91.538	67.086	32.914	8.462	79.312	20.05.2017	08:58:22
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	0.188	0.327	0.612	0.625	0.618	82.692	77.762	22.238	17.308	80.227	20.05.2017	08:58:33
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	0.195	0.253	0.617	0.648	0.632	91.923	68.067	31.933	8.077	79.995	20.05.2017	08:58:34
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Семантический резонанс зна...	0.358	0.332	0.609	0.787	0.687	95.385	65.090	34.910	4.615	80.237	20.05.2017	08:58:45
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Сумма знаний	0.174	0.196	0.705	0.789	0.744	95.385	65.090	34.910	4.615	80.237	20.05.2017	08:58:45
7. INF4 - частный критерий: ROI (Return On Investment), вероятн...	Семантический резонанс зна...	0.184	0.241	0.695	0.726	0.710	68.077	79.824	20.176	31.923	73.950	20.05.2017	08:58:56
7. INF4 - частный критерий: ROI (Return On Investment), вероятн...	Сумма знаний	0.126	0.057	0.668	0.963	0.789	98.846	54.125	45.875	1.154	76.405	20.05.2017	08:58:56
8. INF5 - частный критерий: ROI (Return On Investment), вероятн...	Семантический резонанс зна...	0.208	0.248	0.655	0.723	0.687	70.000	82.085	17.915	30.000	76.043	20.05.2017	08:59:07
8. INF5 - частный критерий: ROI (Return On Investment), вероятн...	Сумма знаний	0.122	0.048	0.656	0.927	0.768	98.846	55.176	44.824	1.154	77.011	20.05.2017	08:59:07
9. INF6 - частный критерий: разн. усл.и безуслов. вероятностей, вер...	Семантический резонанс зна...	0.330	0.217	0.615	0.744	0.673	95.000	61.781	38.219	5.000	78.390	20.05.2017	08:59:18
9. INF6 - частный критерий: разн. усл.и безуслов. вероятностей, вер...	Сумма знаний	0.278	0.130	0.587	0.786	0.672	95.385	55.336	44.664	4.615	75.360	20.05.2017	08:59:18
10. INF7 - частный критерий: разн. усл.и безуслов. вероятностей: ве...	Семантический резонанс зна...	0.329	0.201	0.609	0.800	0.691	94.615	63.221	36.779	5.385	78.918	20.05.2017	08:59:29
10. INF7 - частный критерий: разн. усл.и безуслов. вероятностей: ве...	Сумма знаний	0.216	0.091	0.582	0.852	0.692	95.385	56.588	43.412	4.615	75.986	20.05.2017	08:59:29

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 INF5, 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 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 INF3 на основе наблюдения предыстории их развития:

Рисунок 20. Пример идентификации классов в модели INF5

Рисунок 21. Пример идентификации классов в модели INF5

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 INF5.

Рисунок 23. Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF5

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 INF5

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
INF5

Рисунок 26. Пример нелокальные нейронные сети и нейроны модели
INF5

2.5 КЛАСТЕРНЫЙ И КОНСТРУКТИВНЫЙ АНАЛИЗ

Рисунок 27. Пример кластерный и конструктивный анализ модели INF5

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 INF5, основанная на семантической мере целесообразности информации

А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,687, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 69%. 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 "bazadannixsotrudnikov": <http://allexcel.ru/gotovye-tablitsy-excel-besplatno>, свободный. - Загл. сэкрана. 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>

