

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

Факультет _____

РЕЦЕНЗИЯ

на курсовой проект (работу)

Студента (ки) Концевой Евгений Александрович
курса 3 очной (заочной) формы обучения
направления подготовки 09.03.02
направленность (профиль) Информационные системы и технологии.
Наименование темы «Исследование и разработка системно-когнитивной модели
некоторых сортов крафтового пива на основе данных web- сервера Kaggle».
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	
2.	Степень полноты обзора состояния проблемы и корректность постановки цели и задач исследования	
3.	Уровень и корректность использования в работе различных методов исследований	
4.	Степень комплексности работы, применения в ней знаний общепрофессиональных и специальных дисциплин	
5.	Ясность, четкость, последовательность и обоснованность изложения	
6.	Применение современных технологий обработки информации	
7.	Качество оформления работы (общий уровень грамотности, стиль изложения, качество иллюстраций, соответствие требованиям по оформлению)	
8.	Ответы на вопросы при защите	

Достоинства работы

Применение современных математических моделей и программного инструментария

Недостатки работы

Оформление оставляет желать лучшего

Итоговая оценка при защите _____

Рецензент _____

(Е.В. Луценко)

«7» июня 2017г.

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:

**Исследование и разработка системно-когнитивной модели некоторых
сортов крафтового пива на основе данных web- сервера Kaggle**

выполнил студент группы ИТ1401 **Концевой Евгений Александрович**

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)

Концевой Евгений Александрович
(Ф.И.О.)

Тема проекта: **Исследование и разработка системно-когнитивной модели некоторых сортов крафтового пива на основе данных web- сервера Kaggle**

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.....	7
1.СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	8
1.1. Описание решения.....	8
1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel	8
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	28
2.3. SWOT и PEST матрицы и диаграммы	29
ЗАКЛЮЧЕНИЕ	32
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	33

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 HTML-формата в файл исходных данных MS Excel

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

<http://kaggle.com/nickhould/craft-cans>

Общее описание задачи:

1. Id
2. abv
3. ibu
4. ounces
5. name
6. style
7. breweries
8. city
9. state

Столбцы 4-9 описательные шкалы.

Столбцы 2,3 является классификационными шкалами. Эти столбцы показывают градус и горечь.

Обучающая выборка:

Таблица 1 – beers.xls

id	abv	ibu	ou nc es	name	style	breweries	city	state
1	0,061	60	12	Bitter Bitch Hell or High Watermelon	American Pale Ale (APA)	18th Street Brewery	Gary	IN
2	0,049	17	12	Wheat	Fruit / Vegetable Beer	21st Amendment Brewery	San Francisco	CA
3	0,07	82	12	Bimini Twist	American IPA	3 Daughters Brewing	St Petersburg	FL
4	0,035	11	12	Passion Fruit Prussia	Berliner Weissbier	4 Hands Brewing Company	Saint Louis	MO
5	0,046	15	16	Honey Kolsch	Kölsch	450 North Brewing Company	Columbus	IN
6	0,054	30	16	British Pale Ale (2010)	English Pale Ale	7 Seas Brewing Company	Gig Harbor	WA
7	0,055	26	12	Colorado Native	American Amber / Red Lager	AC Golden Brewing Company	Golden	CO
8	0,042	13	12	Purple Haze	Fruit / Vegetable Beer	Abita Brewing Company	Abita Springs	LA
9	0,05	20	16	The Brown Note	American Brown Ale	Against The Grain Brewery	Louisville	KY
10	0,04	13	16	Sho'nuff	Belgian Pale Ale	Against the Grain Brewery	Louisville	KY
11	0,052	33	16	Pre Flight Pilsner	American Pilsner	Airways Brewing Company	Kent	WA
12	0,04	20	12	P-Town Pilsner	American Pilsner	Alameda Brewing	Portland	OR
13	0,053	18	12	Alaskan Amber	Altbier	Alaskan Brewing Company	Juneau	AK
14	0,062	43	16	Desolation IPA	American IPA	Amnesia Brewing Company	Washougal	WA
15	0,056	4	12	Summer Solstice	Cream Ale	Anderson Valley Brewing Company	Boonville	CA
16	0,047	19	16	Honey Badger Blonde	American Blonde Ale	Appalachian Mountain Brewery	Boone	NC
17	0,062	17	12	Whitsun	American Pale Wheat Ale	Arcadia Brewing Company	Battle Creek	MI
18	0,051	20	12	Greenbelt Farmhouse Ale	Saison / Farmhouse Ale	Armadillo Ale Works	Denton	TX
19	0,052	10	16	Mo's Gose Green Bullet Organic India	Gose	Armstrong Brewing Company	South San Francisco	CA
20	0,07	45	16	Pale Ale	American IPA	Asher Brewing Company	Boulder	CO
21	0,053	26	12	Ninja Porter	American Porter	Asheville Brewing Company	Asheville	NC
22	0,056	27	12	This Season's Blonde	American Blonde Ale	Aspen Brewing Company	Aspen	CO
23	0,062	35	12	Poop Deck Porter	American Porter	Astoria Brewing Company	Astoria	OR
24	0,045	8	12	Dirty Blonde Ale	American Blonde Ale	Atwater Brewery	Detroit	MI
25	0,056	10	12	White Rascal Devils Tramping Ground	Witbier	Avery Brewing Company	Boulder	CO
26	0,092	5	12	Tripel	Tripel	Aviator Brewing Company	Fuquay-Varina	NC
27	0,06	43	12	Naked Pig Pale Ale	American Pale Ale (APA)	Back Forty Beer Company	Gadsden	AL
28	0,044	38	12	Field 41 Pale Ale	American Pale Ale (APA)	Bale Breaker Brewing Company	Yakima	WA
29	0,052	23	12	Ballast Point Pale Ale	Kölsch	Ballast Point Brewing Company	San Diego	CA
30	0,045	20	12	Banner American Rye	Rye Beer	Banner Beer Company	Williamsburg	MA
31	0,07	46	16	Thai.p.a	American IPA	Bare Hands Brewery	Granger	IN
32	0,06	60	12	Barrio Blanco	American IPA	Barrio Brewing Company	Tucson	AZ
33	0,05	20	12	Lost Meridian Wit	Witbier	Base Camp Brewing Co.	Portland	OR
34	0,05	28	12	Stargrazer	Schwarzbier	Bauhaus Brew Labs	Minneapolis	MN
35	0,049	28	12	Pamola Xtra Pale Ale	American Pale Ale (APA)	Baxter Brewing Company	Lewiston	ME
36	0,088	8	16	Hoptopus Double IPA	American Double / Imperial IPA	Beach Brewing Company	Virginia Beach	VA
37	0,05	10	12	Watermelon Ale	Fruit / Vegetable Beer	Beer Works Brewery	Lowell	

									MA
38	0,057	27	12	Nordic Blonde	American Blonde Ale	Bent Brewstillery	Roseville	MN	
39	0,056	32	12	14° ESB	Extra Special / Strong Bitter	Bent Paddle Brewing Company	Duluth	MN	
40	0,053	20	12	Steel Rail Extra Pale Ale	American Pale Ale (APA)	Berkshire Brewing Company	South Deerfield	MA	
41	0,081	17	16	Aprè Shred	American Strong Ale	Big Choice Brewing	Broomfield	CO	

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.

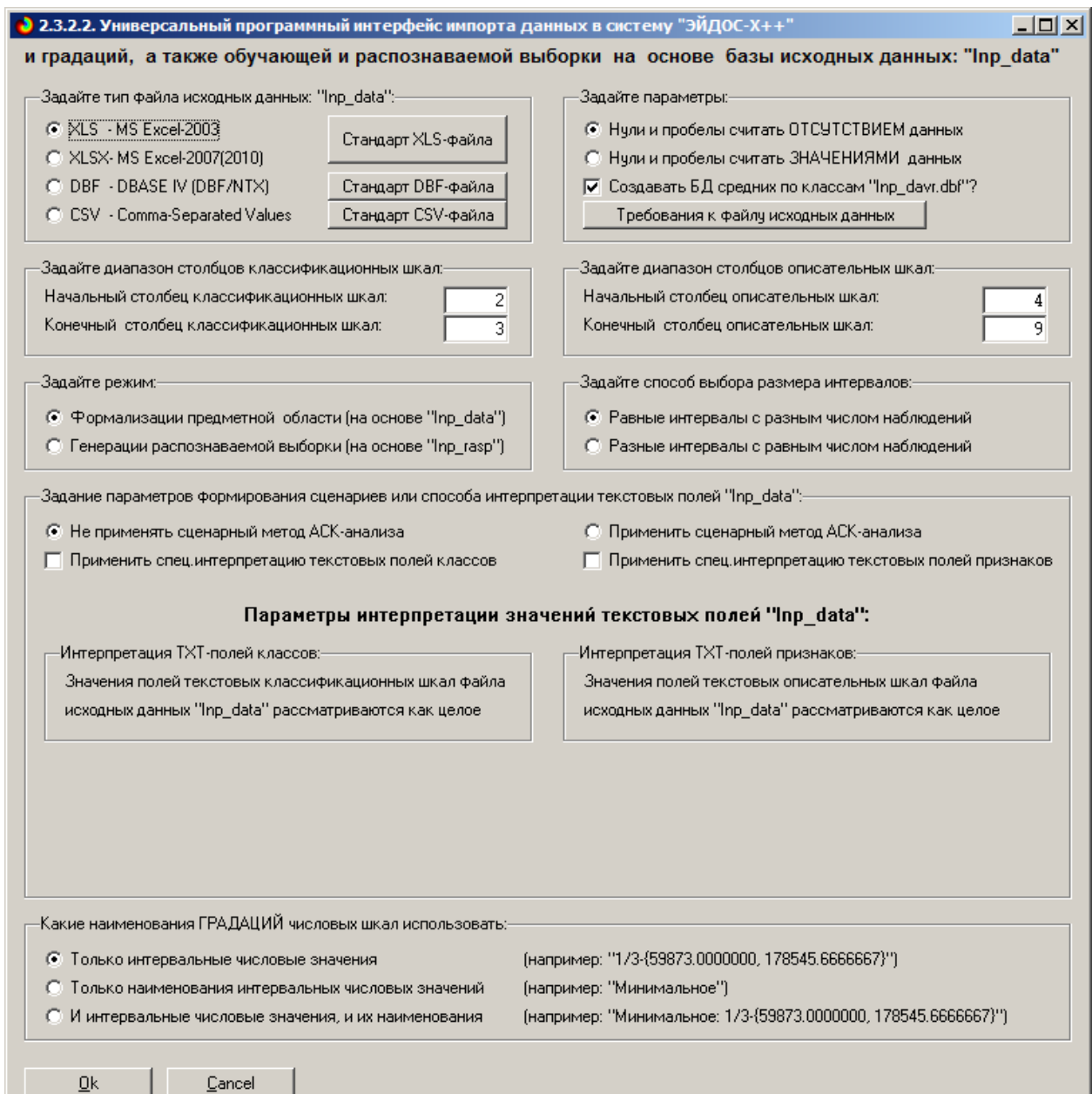


Рисунок 4. Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

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" – 3(последний столбец в таблице);

– "Задайте диапазон столбцов описательных шкал": "Начальный столбец описательных шкал" –4, "Конечный столбец описательных шкал" – 9;

– "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".

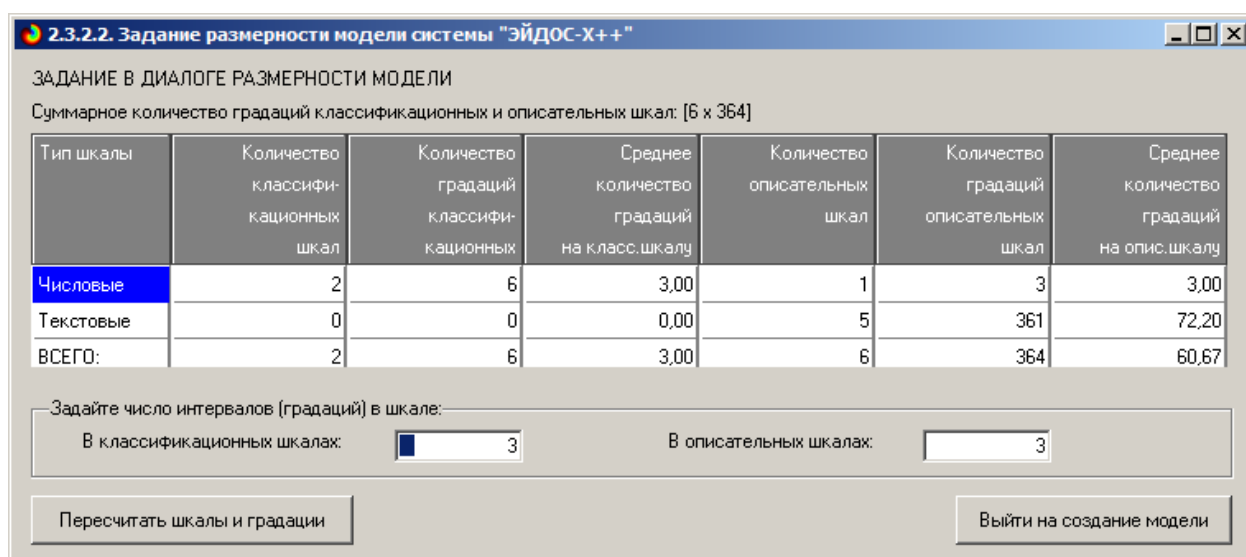


Рисунок 5. Задание размерности модели системы "Эйдос"

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".

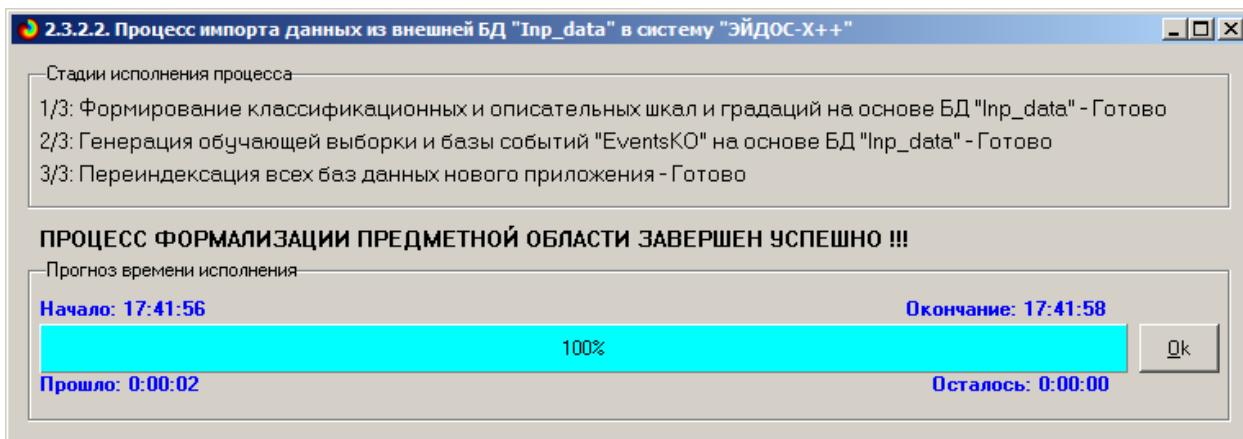


Рисунок 6. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

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).

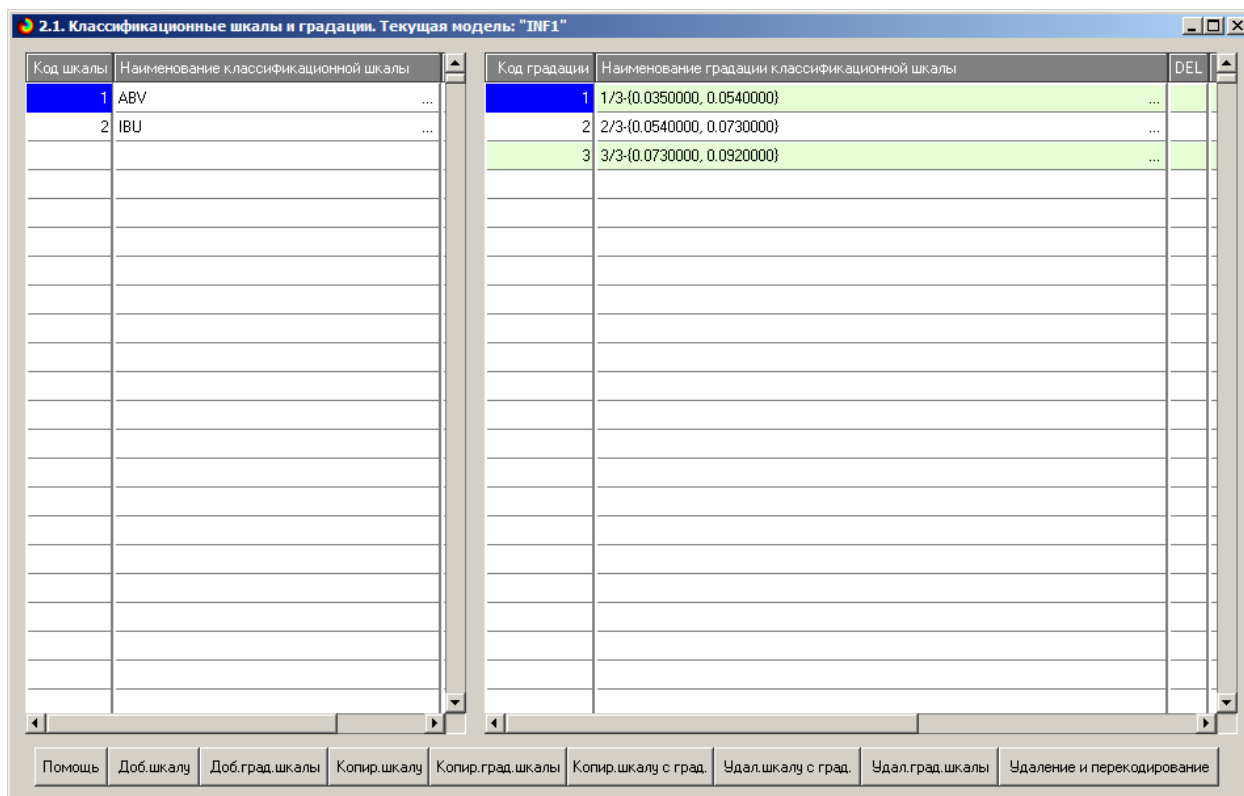


Рисунок 7. Классификационные шкалы и градации (фрагмент)

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

Код шкалы	Наименование описательной шкалы
1	OUNCES
2	NAME
3	STYLE
4	BREWERIES
5	CITY
6	STATE

Код градации	Наименование градации описательной шкалы
1	1/3-{12.0000000, 13.3333333}
2	2/3-{13.3333333, 14.6666667}
3	3/3-{14.6666667, 16.0000000}

Рисунок 8. Описательные шкалы и градации (фрагмент)

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

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

Код объекта	Класс 1	Класс 2	Класс 3	Класс 4
1	2	5	0	0

Код объекта	Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
1	1	17	112	144	279	0	0

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

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).

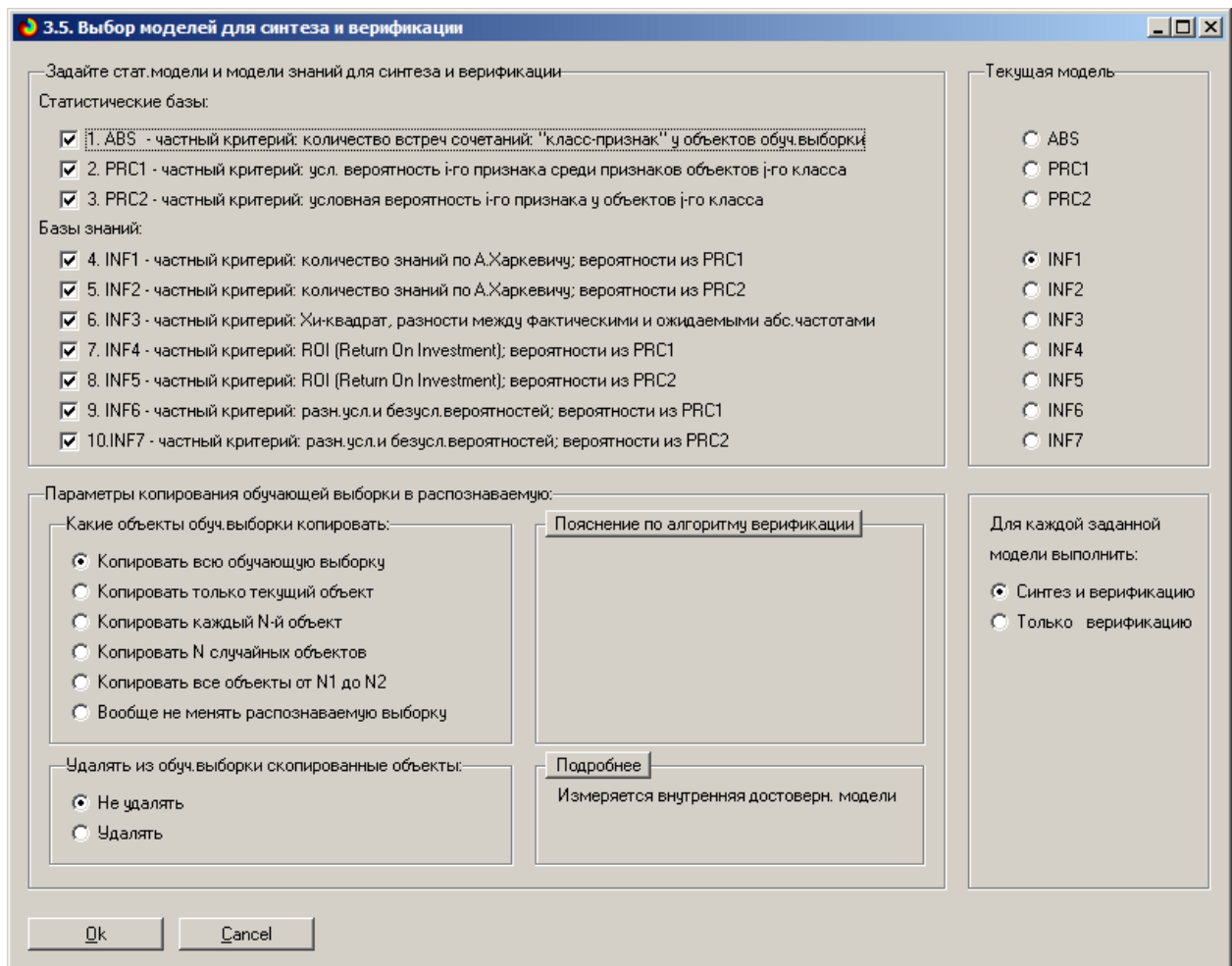


Рисунок 10. Выбор моделей для синтеза и верификации, а также текущей модели

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.

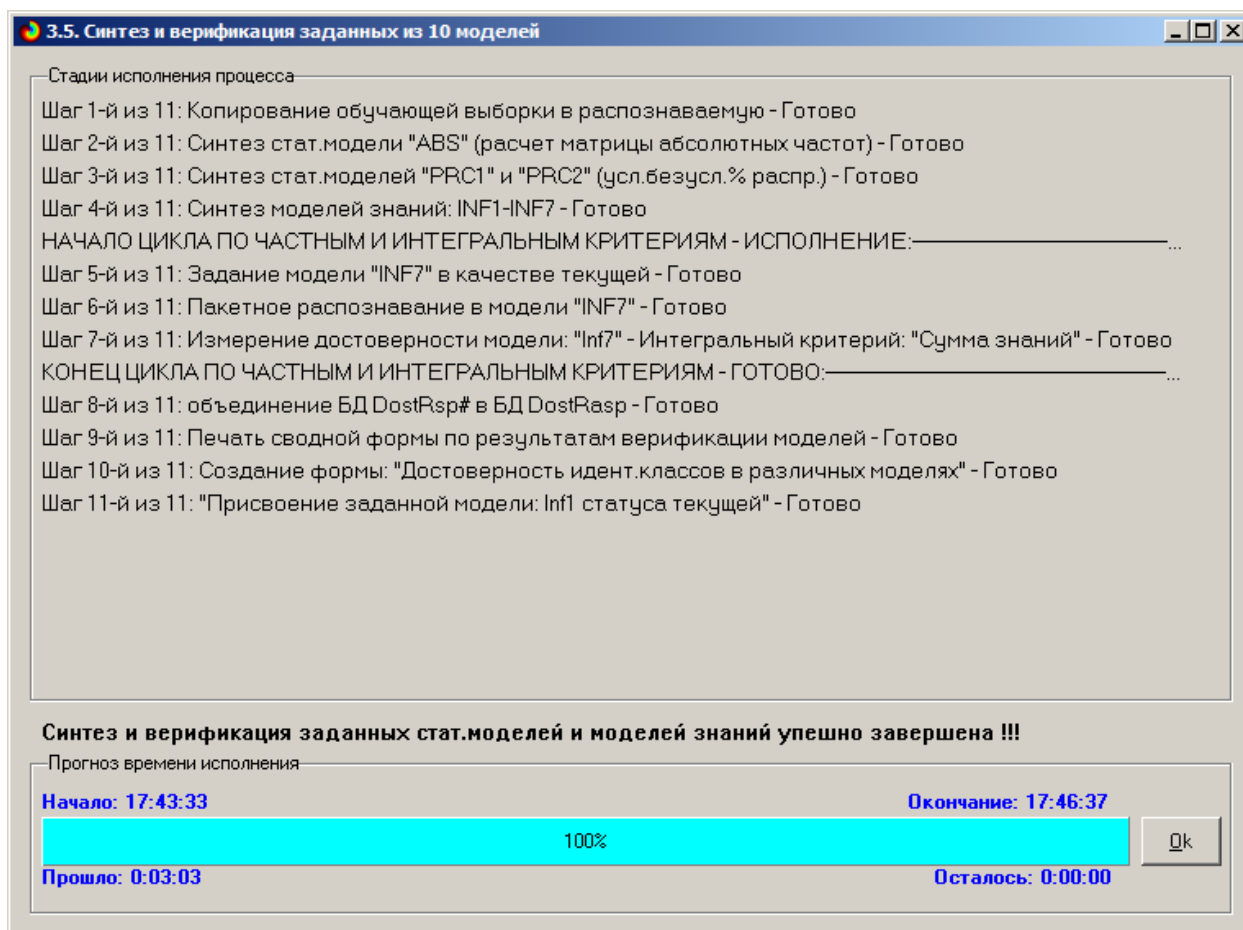


Рисунок 11. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 183 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 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. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"										
д. знака	Наименование описательной шкалы и градации	1. ABV 1/3 (0.0350000, 0.0540000)	2. ABV 2/3 (0.0540000, 0.0730000)	3. ABV 3/3 (0.0730000, 0.0920000)	4. IBU 1/3 (4.0000000, 38.6666667)	5. IBU 2/3 (38.6666667, 73.3333333)	6. IBU 3/3 (73.3333333, 108.0000000)	Сумма	Среднее	Средн. квадрат. откл.
1	DUNCES-1/3-{12.0000000, 13.3333333}	41	19	3	56	4	3	126	21.000	22.654
2	DUNCES-2/3-{13.3333333, 14.6666667}									
3	DUNCES-3/3-{14.6666667, 16.0000000}	15	16	6	26	9	2	74	12.333	8.548
4	NAME-14" ESB		1		1			2	0.333	0.516
5	NAME-B34 Happy As Ale	1			1			2	0.333	0.516
6	NAME-Agave Wheat	1			1			2	0.333	0.516
7	NAME-Alaskan Amber	1			1			2	0.333	0.516
8	NAME-Ale Cider		1		1			2	0.333	0.516
9	NAME-American Red Saison			1	1			2	0.333	0.516
10	NAME-Apre Shred			1	1			2	0.333	0.516
11	NAME-Ballast Point Pale Ale	1			1			2	0.333	0.516
12	NAME-Banner American Rye	1			1			2	0.333	0.516
13	NAME-Barrio Blanco		1			1		2	0.333	0.516
14	NAME-Beach Blonde	1			1			2	0.333	0.516
15	NAME-Big River Pilsner	1			1			2	0.333	0.516
16	NAME-Bimini Twist		1				1	2	0.333	0.516
17	NAME-Bitter Bitch		1			1		2	0.333	0.516
18	NAME-Black Market Hefeweizen	1			1			2	0.333	0.516
19	NAME-Blood Orange Honey		1		1			2	0.333	0.516
20	NAME-Blue Mountain Classic Lager	1			1			2	0.333	0.516
21	NAME-Blue Point Summer Ale	1			1			2	0.333	0.516
22	NAME-Bomb Lager (Old Recipe)	1			1			2	0.333	0.516
23	NAME-Bozone Hefe Weizen		1		1			2	0.333	0.516

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

5. Модель "4. INF1 - частный критерий количество знаний по А.Харкевичу, вероятности из PRC1"										
д изменя	Наименование описательной шкалы и градации	1. ABV 1/3 (0.0350000, 0.0540000)	2. ABV 2/3 (0.0540000, 0.0730000)	3. ABV 3/3 (0.0730000, 0.0920000)	4. IBU 1/3 (4.0000000, 38.6666667)	5. IBU 2/3 (38.6666667, 73.3333333)	6. IBU 3/3 (73.3333333, 108.0000000)	Сумма	Среднее	Средн. кв.др. откл.
1	DUNCES-1/3(12.0000000, 13.3333333)	0.055	-0.054	-0.241	0.030	-0.271	-0.005	-0.485	-0.081	0.141
2	DUNCES-2/3(13.3333333, 14.6666667)									
3	DUNCES-3/3(14.6666667, 16.0000000)	-0.122	0.081	0.218	-0.058	0.232	0.042	0.394	0.066	0.143
4	NAME-14' ESB		0.395		0.074			0.469	0.078	0.158
5	NAME-934 Happy As Ale	0.216			0.074			0.290	0.048	0.087
6	NAME-Agave Wheat	0.216			0.074			0.290	0.048	0.087
7	NAME-Alaskan Amber	0.216			0.074			0.290	0.048	0.087
8	NAME-Ale Cider		0.395		0.074			0.469	0.078	0.158
9	NAME-American Red Saison			0.900	0.074			0.974	0.162	0.362
10	NAME-Apre Shred			0.900	0.074			0.974	0.162	0.362
11	NAME-Ballast Point Pale Ale	0.216			0.074			0.290	0.048	0.087
12	NAME-Banner American Rye	0.216			0.074			0.290	0.048	0.087
13	NAME-Barrio Blanco		0.395			0.762		1.157	0.193	0.320
14	NAME-Beach Blonde	0.216			0.074			0.290	0.048	0.087
15	NAME-Big River Pilsner	0.216			0.074			0.290	0.048	0.087
16	NAME-Bimini Twist		0.395				1.135	1.530	0.255	0.459
17	NAME-Bitter Bitch		0.395			0.762		1.157	0.193	0.320
18	NAME-Black Market Hefeweizen	0.216			0.074			0.290	0.048	0.087
19	NAME-Blood Orange Honey		0.395		0.074			0.469	0.078	0.158
20	NAME-Blue Mountain Classic Lager	0.216			0.074			0.290	0.048	0.087
21	NAME-Blue Point Summer Ale	0.216			0.074			0.290	0.048	0.087
22	NAME-Bomb Lager (Old Recipe)	0.216			0.074			0.290	0.048	0.087
23	NAME-Bozone Hefe Weizen		0.395		0.074			0.469	0.078	0.158

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

5. Модель "6. INF3 - частный критерий Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами"										
д изменя	Наименование описательной шкалы и градации	1. ABV 1/3 (0.0350000, 0.0540000)	2. ABV 2/3 (0.0540000, 0.0730000)	3. ABV 3/3 (0.0730000, 0.0920000)	4. IBU 1/3 (4.0000000, 38.6666667)	5. IBU 2/3 (38.6666667, 73.3333333)	6. IBU 3/3 (73.3333333, 108.0000000)	Сумма	Среднее	Средн. кв.др. откл.
1	DUNCES-1/3(12.0000000, 13.3333333)	5.634	-2.930	-2.704	4.282	-4.239	-0.042			4.097
2	DUNCES-2/3(13.3333333, 14.6666667)									
3	DUNCES-3/3(14.6666667, 16.0000000)	-5.771	3.121	2.650	-4.374	4.161	0.213			4.161
4	NAME-14' ESB	-0.561	0.652	-0.091	0.179	-0.131	-0.048			0.400
5	NAME-934 Happy As Ale	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
6	NAME-Agave Wheat	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
7	NAME-Alaskan Amber	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
8	NAME-Ale Cider	-0.561	0.652	-0.091	0.179	-0.131	-0.048			0.400
9	NAME-American Red Saison	-0.561	-0.348	0.909	0.179	-0.131	-0.048			0.513
10	NAME-Apre Shred	-0.561	-0.348	0.909	0.179	-0.131	-0.048			0.513
11	NAME-Ballast Point Pale Ale	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
12	NAME-Banner American Rye	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
13	NAME-Barrio Blanco	-0.561	0.652	-0.091	-0.821	0.869	-0.048			0.660
14	NAME-Beach Blonde	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
15	NAME-Big River Pilsner	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
16	NAME-Bimini Twist	-0.561	0.652	-0.091	-0.821	-0.131	0.952			0.685
17	NAME-Bitter Bitch	-0.561	0.652	-0.091	-0.821	0.869	-0.048			0.660
18	NAME-Black Market Hefeweizen	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
19	NAME-Blood Orange Honey	-0.561	0.652	-0.091	0.179	-0.131	-0.048			0.400
20	NAME-Blue Mountain Classic Lager	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
21	NAME-Blue Point Summer Ale	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
22	NAME-Bomb Lager (Old Recipe)	0.439	-0.348	-0.091	0.179	-0.131	-0.048			0.273
23	NAME-Bozone Hefe Weizen	-0.561	0.652	-0.091	0.179	-0.131	-0.048			0.400

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 модуль уровней сходства ложно-положит. решений	Средний модуль уровней сходства ложно-отрицат. решений	A-точность модели AРесол = AТР/AТР	A-Полнота модели AРесол = AТР/AТР	L2-мера проф. Е.В. Луценко	Процент правильной идентификац.	Процент правильной не идентифи...	Процент ошибочной идентификац.	Процент ошибочной не идентифи...	Процент правильных результатов	Дата получения результата	Время получения результата
1. ABS - частый 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
2. PRC1 - частый критерий: усл. вероятность i-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
3. PRC2 - частый критерий: условная вероятность i-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
4. INF1 - частый критерий: количество знаний по АХаркевичу 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
5. INF2 - частый критерий: количество знаний по АХаркевичу в...	Семантический резонанс зна...	0.184	0.259	0.642	0.645	0.644	91.538	67.086	32.914	8.462	79.312	20.05.2017	08:58:34
6. INF3 - частый критерий: Хинквартд, разности между факти...	Семантический резонанс зна...	0.196	0.253	0.617	0.648	0.632	91.923	68.067	31.933	8.077	79.995	20.05.2017	08:58:45
7. INF4 - частый критерий: ROI (Return On Investment): верою...	Семантический резонанс зна...	0.174	0.196	0.705	0.789	0.744	95.385	65.090	34.910	4.615	80.237	20.05.2017	08:58:56
8. INF5 - частый критерий: ROI (Return On Investment): верою...	Семантический резонанс зна...	0.126	0.057	0.668	0.963	0.789	98.846	54.125	45.875	1.154	76.485	20.05.2017	08:59:07
9. INF6 - частый критерий: разн услд без усл вероятностей; вер...	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	08:59:18
10. INF7 - частый критерий: разн услд без усл вероятностей; е...	Семантический резонанс зна...	0.122	0.048	0.656	0.927	0.768	98.846	55.176	44.824	1.154	77.011	20.05.2017	08:59:29
11. INF8 - частый критерий: разн услд без усл вероятностей; е...	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	08:59:40
12. INF9 - частый критерий: разн услд без усл вероятностей; е...	Семантический резонанс зна...	0.278	0.130	0.587	0.786	0.672	95.385	55.336	44.664	4.615	75.360	20.05.2017	08:59:51
13. INF10 - частый критерий: разн услд без усл вероятностей; е...	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	08:59:52
14. INF11 - частый критерий: разн услд без усл вероятностей; е...	Семантический резонанс зна...	0.216	0.091	0.682	0.852	0.692	95.385	56.588	43.412	4.615	75.986	20.05.2017	08:59:53

Рисунок 12. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказалась модель INF5 при интегральном критерии «Семантический резонанс знаний». При этом точность модели составляет 0,967. 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).

ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Предположим, модель дает такой прогноз: выпадет 1, 2, 3, 4, 5 или 6. В этом случае у нее будет 100% достоверность идентификации, т.е. не будет ни одного объекта, не отнесенного к тому классу, к которому он действительно относится, но при этом будет очень большая ошибка ложной идентификации, т.к. огромное количество объектов будет отнесено к классам, к которым они не относятся (и именно за счет этого у модели и будет очень высокая достоверность идентификации). Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.

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

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

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

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

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

$$L2 = \frac{(1 + (TP + TN - FP - FN))}{(TP + TN + FP + FN)} \quad (\text{нормировка: } \{0, 1\})$$

где количество: TP - истинно-положительных решений; TN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;

Классическая F-мера достоверности моделей ван Ризбергера (колонка выделена ярко-голубым фоном):

$$F\text{-мера} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

$$\text{Precision} = TP / (TP + FP) \quad \text{точность модели};$$

$$\text{Recall} = TP / (TP + FN) \quad \text{полнота модели};$$

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

$$L1\text{-мера} = 2 * (\text{SPrecision} * \text{SRecall}) / (\text{SPrecision} + \text{SRecall})$$

$$\text{SPrecision} = \text{STP} / (\text{STP} + \text{SFP}) \quad \text{точность с учетом сумм уровней сходства};$$

$$\text{SRecall} = \text{STP} / (\text{STP} + \text{SFN}) \quad \text{полнота с учетом сумм уровней сходства};$$

STP - Сумма модулей сходства истинно-положительных решений; SFN - Сумма модулей сходства истинно-отрицательных решений; SFP - Сумма модулей сходства ложно-положительных решений; SFN - Сумма модулей сходства ложно-отрицательных решений.

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

$$L2\text{-мера} = 2 * (\text{APrecision} * \text{AREcall}) / (\text{APrecision} + \text{AREcall})$$

$$\text{APrecision} = \text{ATP} / (\text{ATP} + \text{AFP}) \quad \text{точность с учетом средних уровней сходства};$$

$$\text{AREcall} = \text{ATP} / (\text{ATP} + \text{AFN}) \quad \text{полнота с учетом средних уровней сходства};$$

$$\text{ATP} = \text{STP} / \text{TP} \quad \text{Среднее модулей сходства истинно-положительных решений}; \text{AFP} = \text{SFP} / \text{FP} \quad \text{Среднее модулей сходства ложно-положительных решений};$$

$$\text{AFN} = \text{SFN} / \text{FN} \quad \text{Среднее модулей сходства истинно-отрицательных решений}; \text{AFN} = \text{SFN} / \text{FN} \quad \text{Среднее модулей сходства ложно-отрицательных решений}.$$

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

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

Рисунок 13. Види 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.

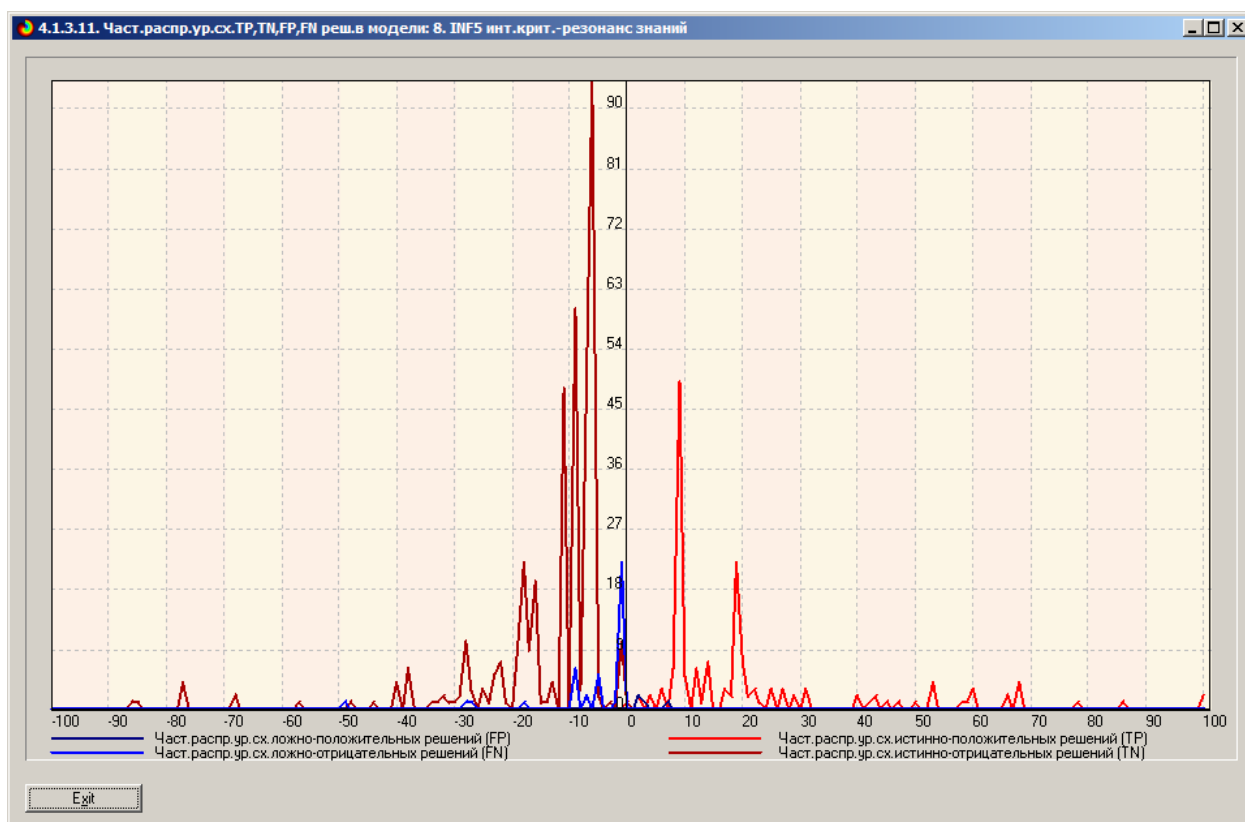


Рисунок 14. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF5

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 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.

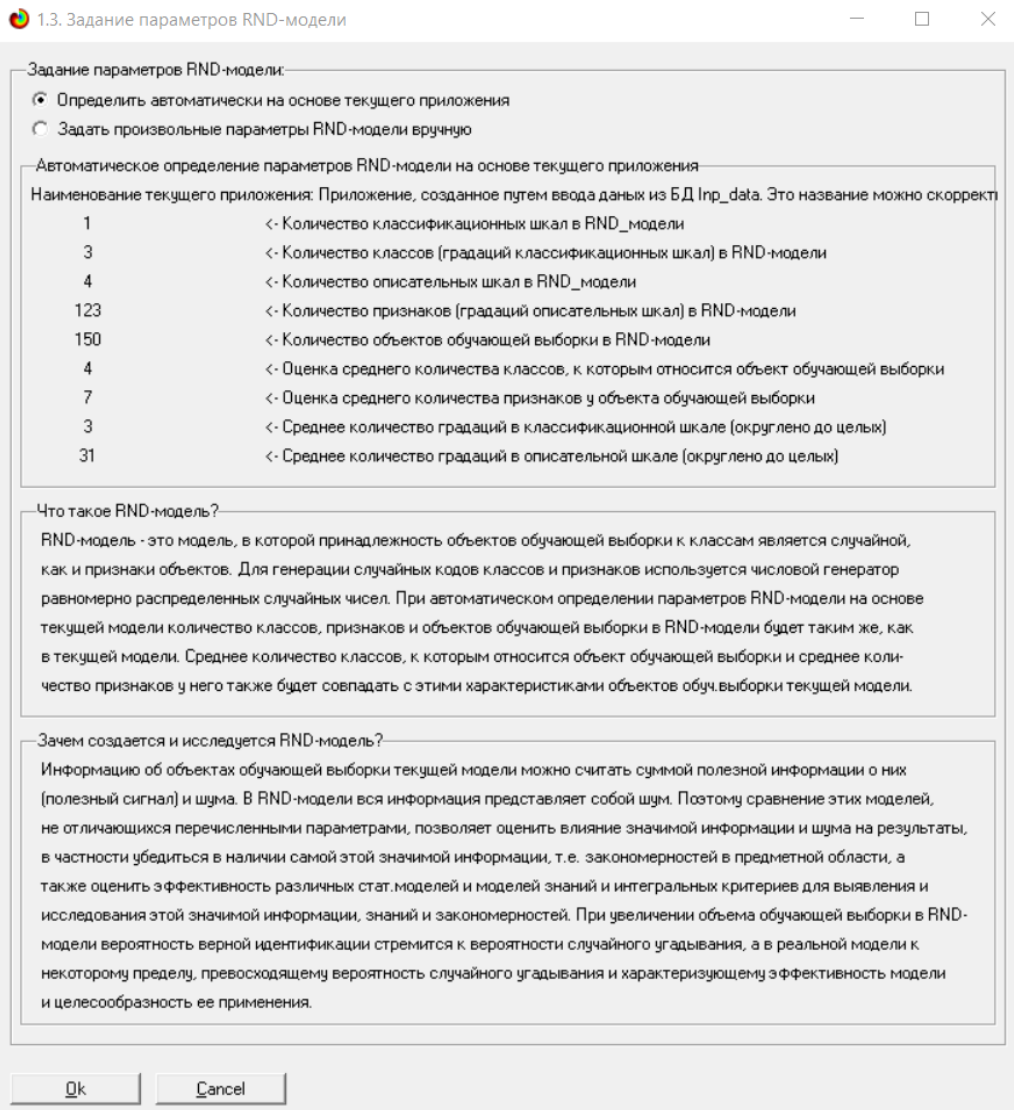


Рисунок 15. Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

На рисунке 13 показано частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF5.

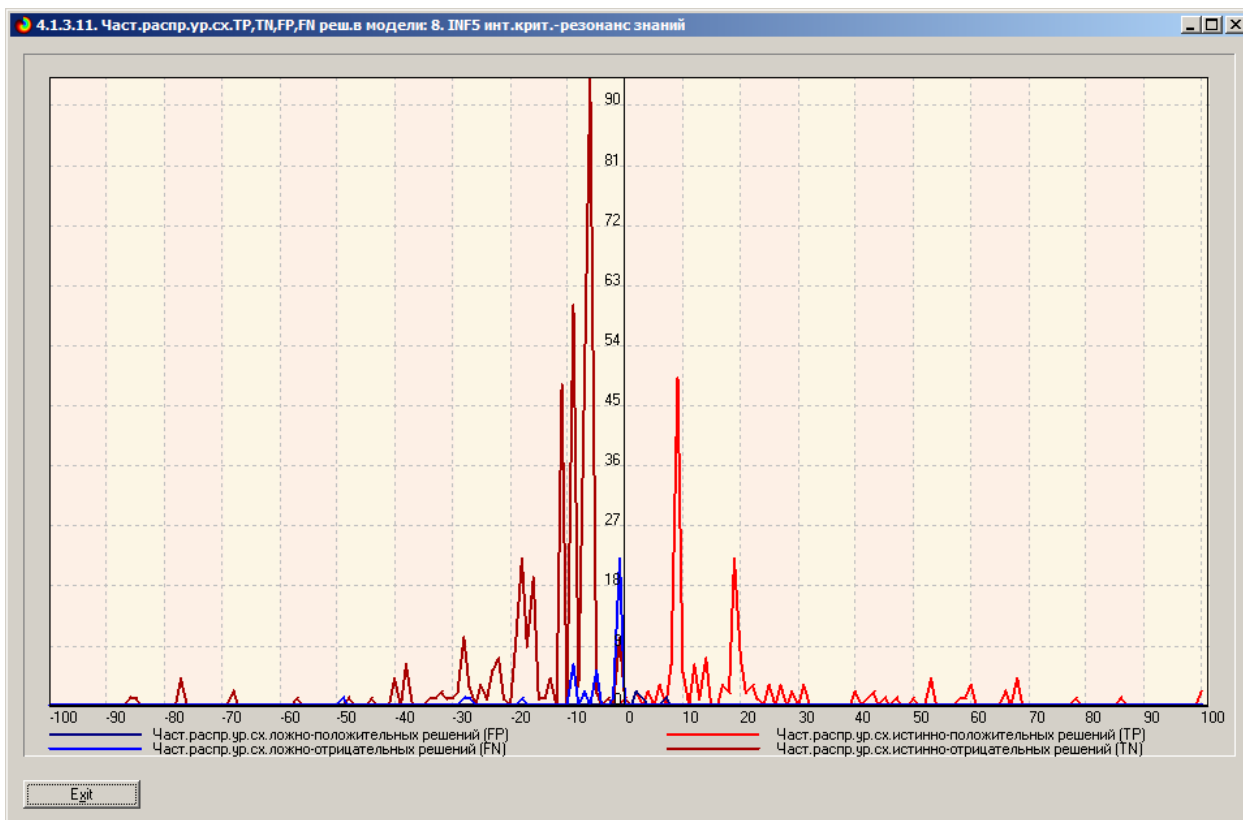


Рисунок 16. Частное распределение сходства-различия верно и ошибочно идентифицированных и не идентифицированных состояний в случайной модели INF5

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.

4.1.3.6. Обобщенная форма по достов. моделей при разн. крит. - Текущая модель: "INF1"											
Наименование модели и частного критерия	Интегральный критерий	Средний модуль уровня сходст. ложно-положит. решений	Средний модуль уровня сходст. ложно-отрицат. решений	A-Точность модели APrecision = ATR/AATP...	A-Полнота модели ARecall = ATR/AATP...	L2-мера проц. Е.В. Ляженко	Процент правильной идентифика...	Процент правильной не идентиф...	Процент ошибочной идентифика...	Процент ошибочной не идентиф...	Процент правильного результата
1. ABS - частный критерий: количество встреч соеетаний: "к.лас...	Корреляция абс частот с обр...	0.358		0.636	1.000	0.778	100.000		100.000		50
1. ABS - частный критерий: количество встреч соеетаний: "к.лас...	Сумма абс частот по признак...	0.164		0.778	1.000	0.875	100.000		100.000		50
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл отн частот с о...	0.358		0.636	1.000	0.778	100.000		100.000		50
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл отн частот по призна...	0.310		0.591	1.000	0.743	100.000		100.000		50
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл отн частот с о...	0.358		0.636	1.000	0.778	100.000		100.000		50
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл отн частот по призна...	0.318		0.593	1.000	0.744	100.000		100.000		50
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	0.080	0.138	0.718	0.782	0.748	78.500	99.591	0.409	21.500	89
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	0.040	0.039	0.845	0.815	0.830	98.500	73.555	26.445	1.500	86
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	0.078	0.139	0.725	0.784	0.753	78.500	99.591	0.409	21.500	89
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	0.040	0.040	0.848	0.817	0.832	98.500	73.555	26.445	1.500	86
6. INF3 - частный критерий: Хи-квадрат, разности между фактик...	Семантический резонанс зна...	0.226	0.329	0.682	0.713	0.697	71.000	68.312	31.688	29.000	69
6. INF3 - частный критерий: Хи-квадрат, разности между фактик...	Сумма знаний	0.183	0.275	0.717	0.695	0.706	71.000	68.312	31.688	29.000	69
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.032	0.134	0.870	0.775	0.820	80.000	99.775	0.225	20.000	89
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.011	0.004	0.892	0.952	0.921	99.500	71.985	28.015	0.500	85
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.032	0.134	0.870	0.777	0.821	80.000	99.775	0.225	20.000	89
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.011	0.004	0.893	0.953	0.922	99.500	71.985	28.015	0.500	85
9. INF6 - частный критерий: разн усл и безуслов вероятностей; вер...	Семантический резонанс зна...	0.283	0.309	0.629	0.624	0.627	70.000	67.272	32.728	30.000	68
9. INF6 - частный критерий: разн усл и безуслов вероятностей; вер...	Сумма знаний	0.181	0.163	0.518	0.826	0.637	71.000	66.487	33.513	29.000	68
10. INF7 - частный критерий: разн усл и безуслов вероятностей; ве...	Семантический резонанс зна...	0.287	0.314	0.628	0.625	0.626	70.000	67.250	32.750	30.000	68
10. INF7 - частный критерий: разн усл и безуслов вероятностей; ве...	Сумма знаний	0.187	0.176	0.522	0.824	0.639	71.000	66.487	33.513	29.000	68

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

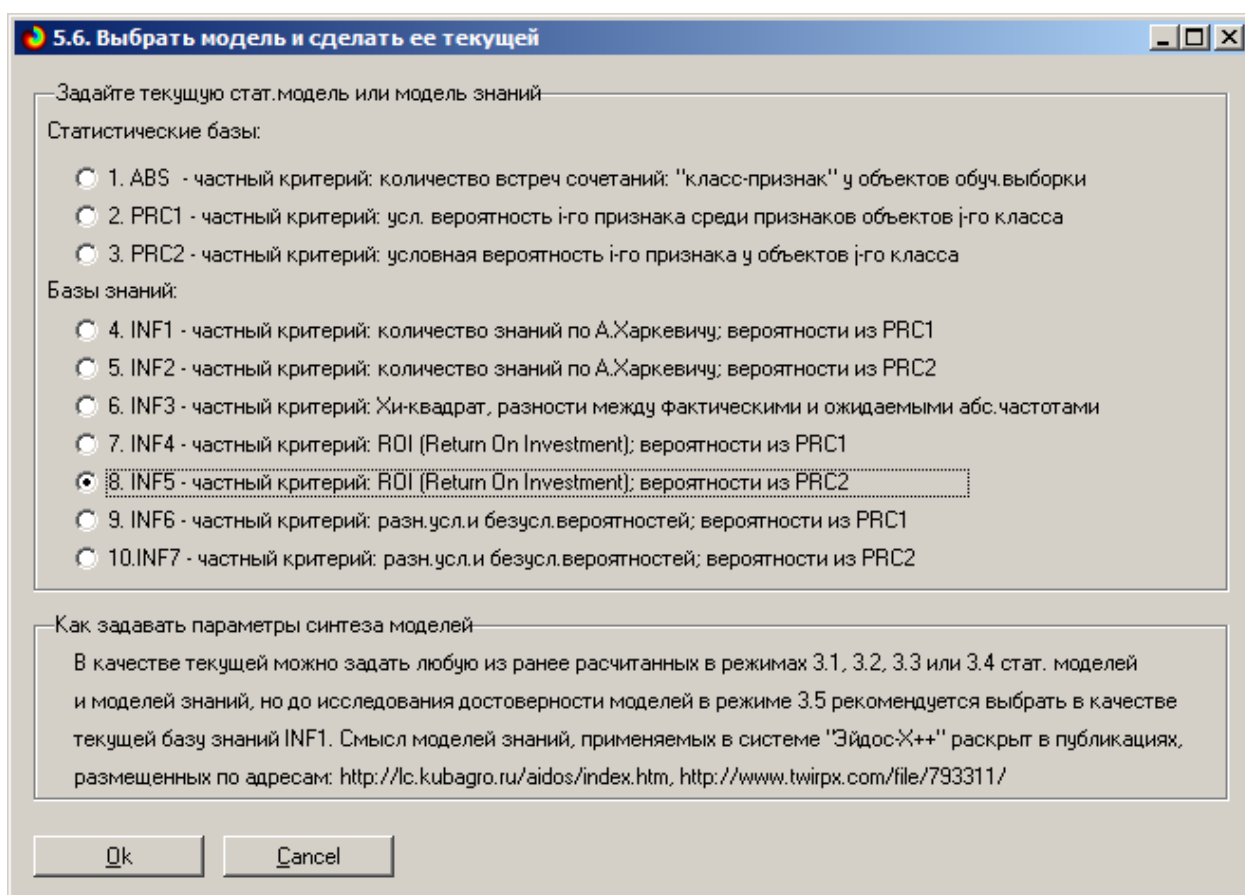
На основе его сравнения с рисунком 9 можно сделать следующие **ВЫВОДЫ**:

- 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 INF5 (режим 5.6) (рисунок 15) и проведем пакетное распознавание в режиме 4.2.1.



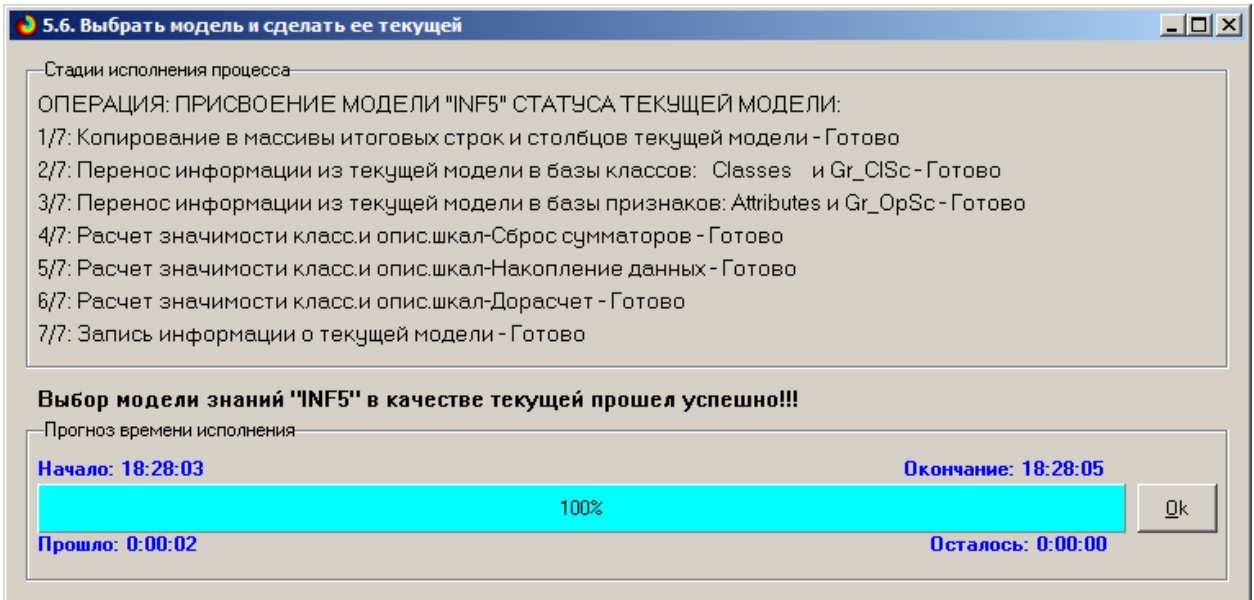


Рисунок 18. Экранные формы режима задания модели в качестве текущей

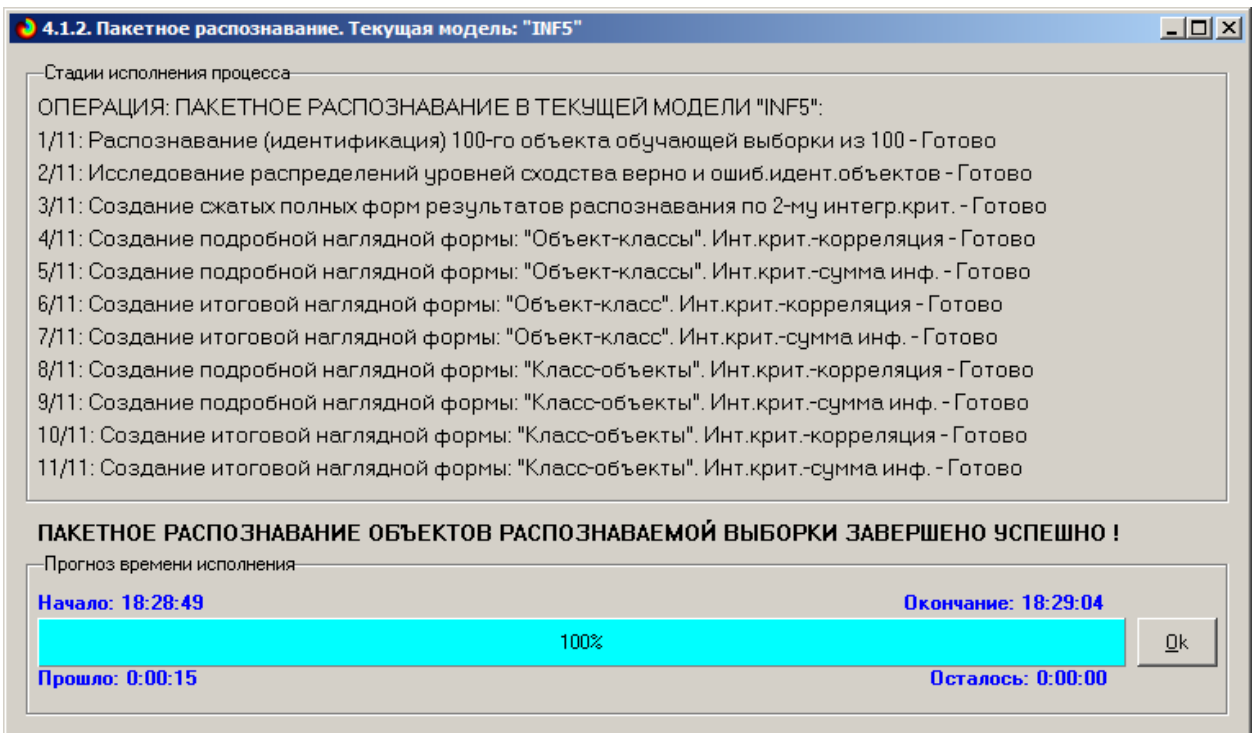


Рисунок 19. Экранные формы режима пакетного распознавания в текущей модели

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:

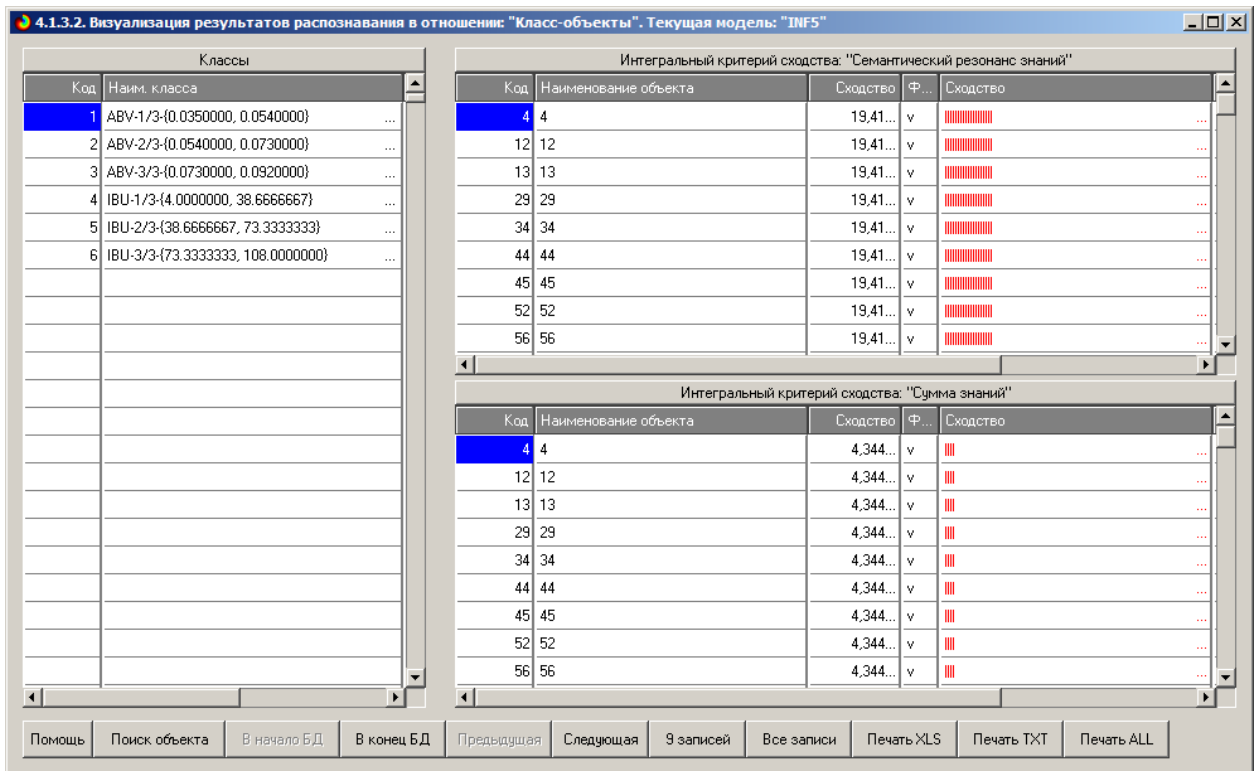


Рисунок 20. Пример идентификации классов в модели INF5

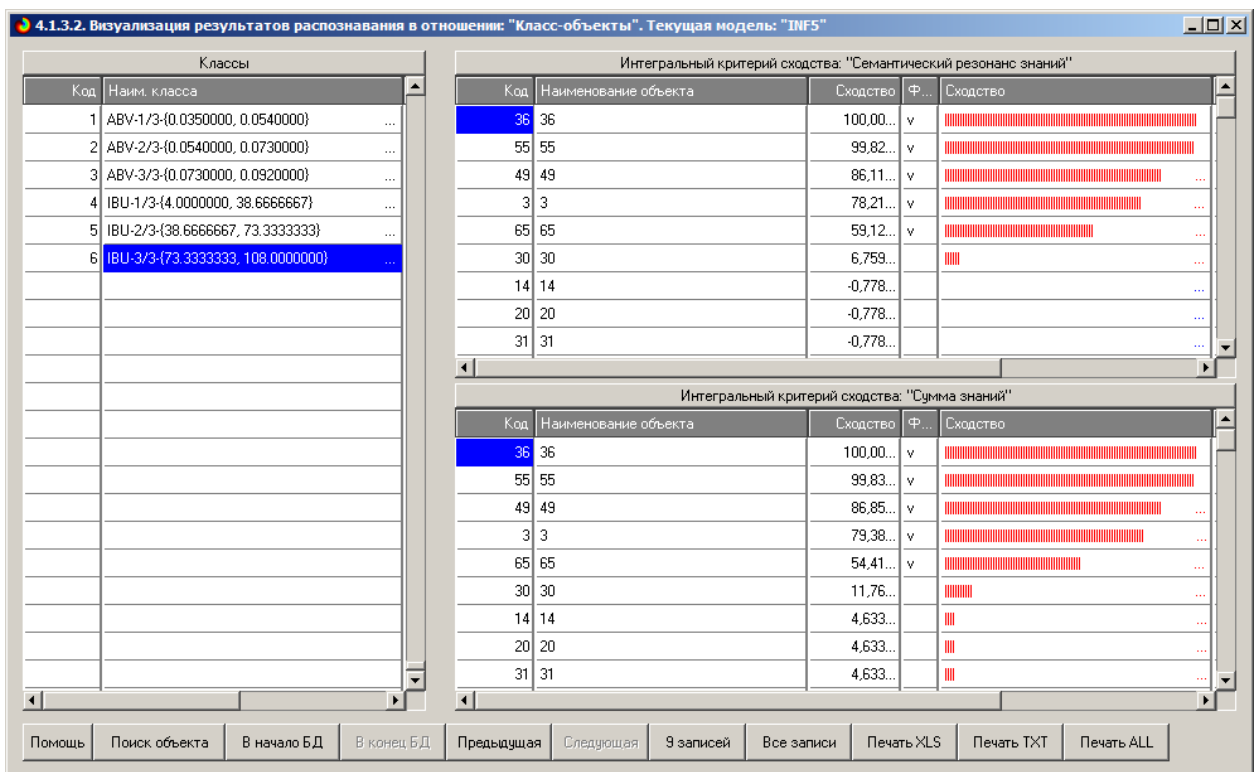


Рисунок 21. Пример идентификации классов в модели INF5

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)

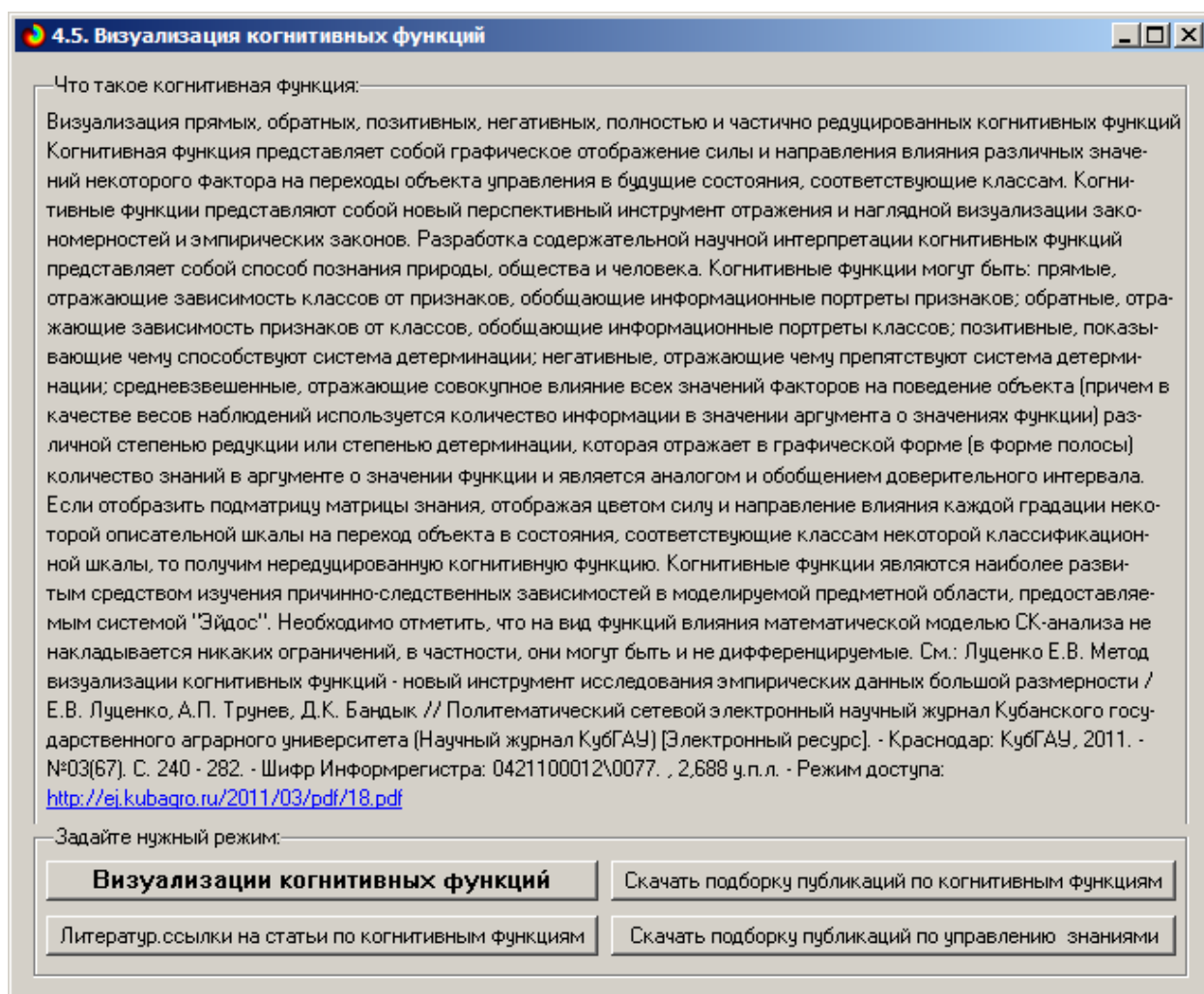


Рисунок 22. Экранная форма режима 4.5 системы «Эйдос-X++» «Визуализация когнитивных функций»

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 INF5.

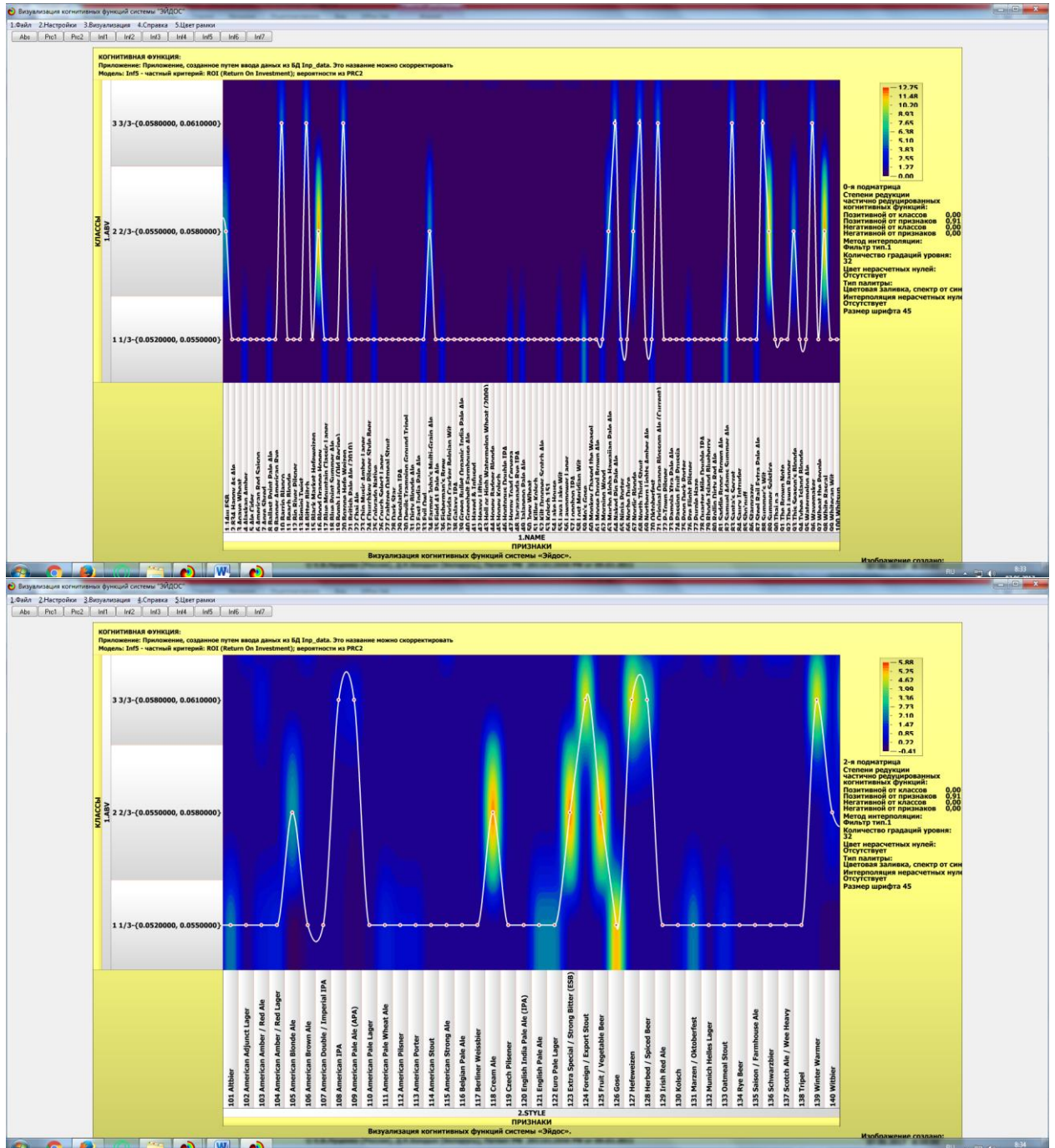


Рисунок 23. Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF5

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).

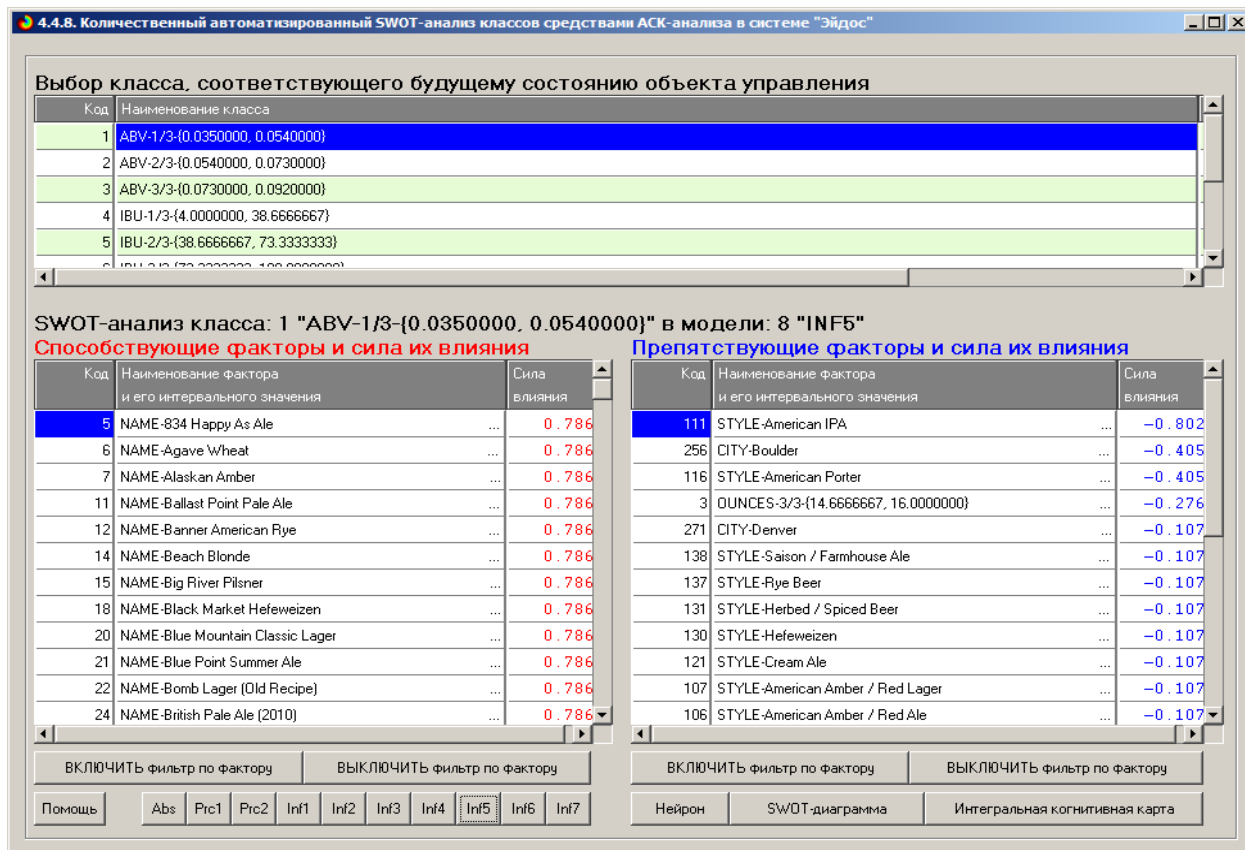


Рисунок 24. Пример SWOT-матрицы в модели INF5

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

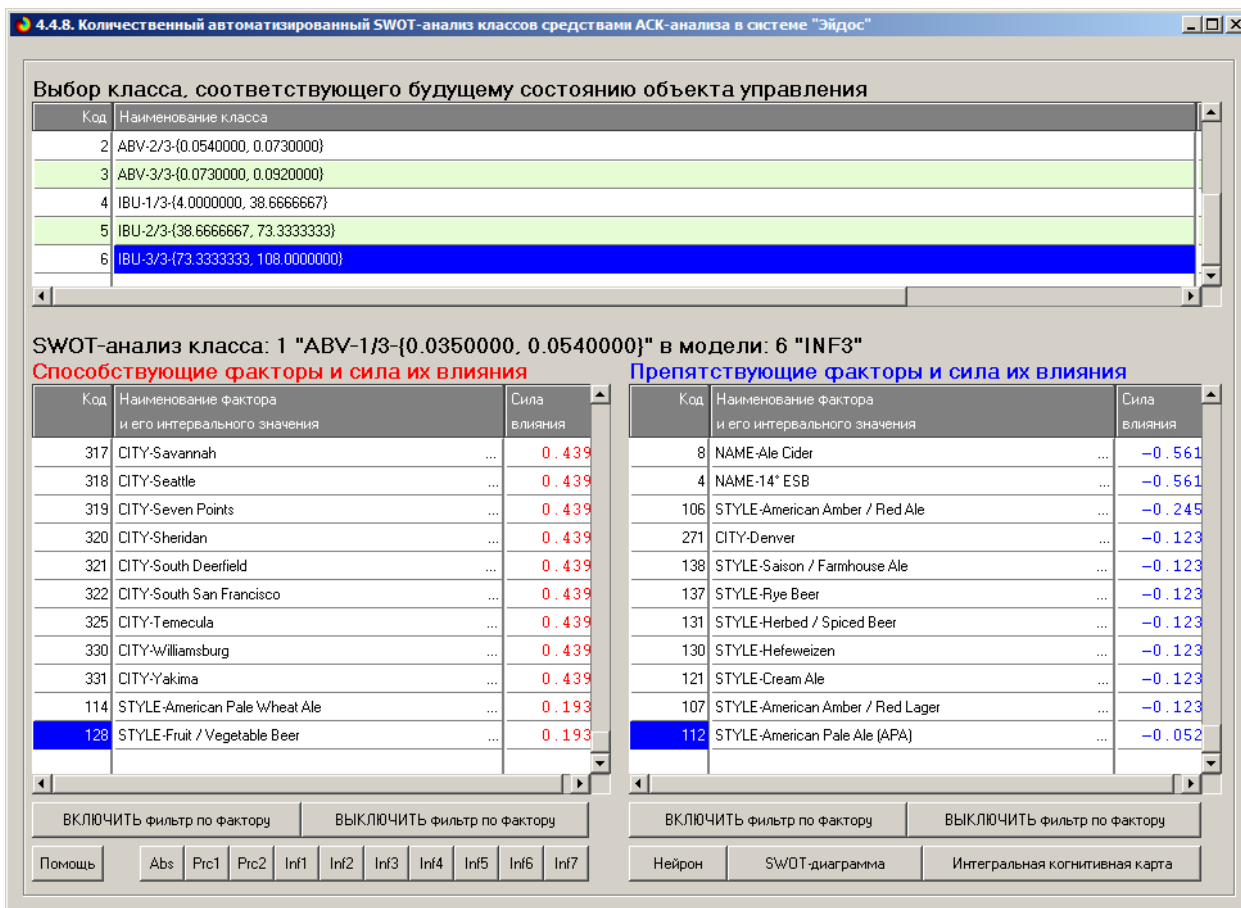


Рисунок 25. Пример SWOT-матрицы в модели INF5

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.

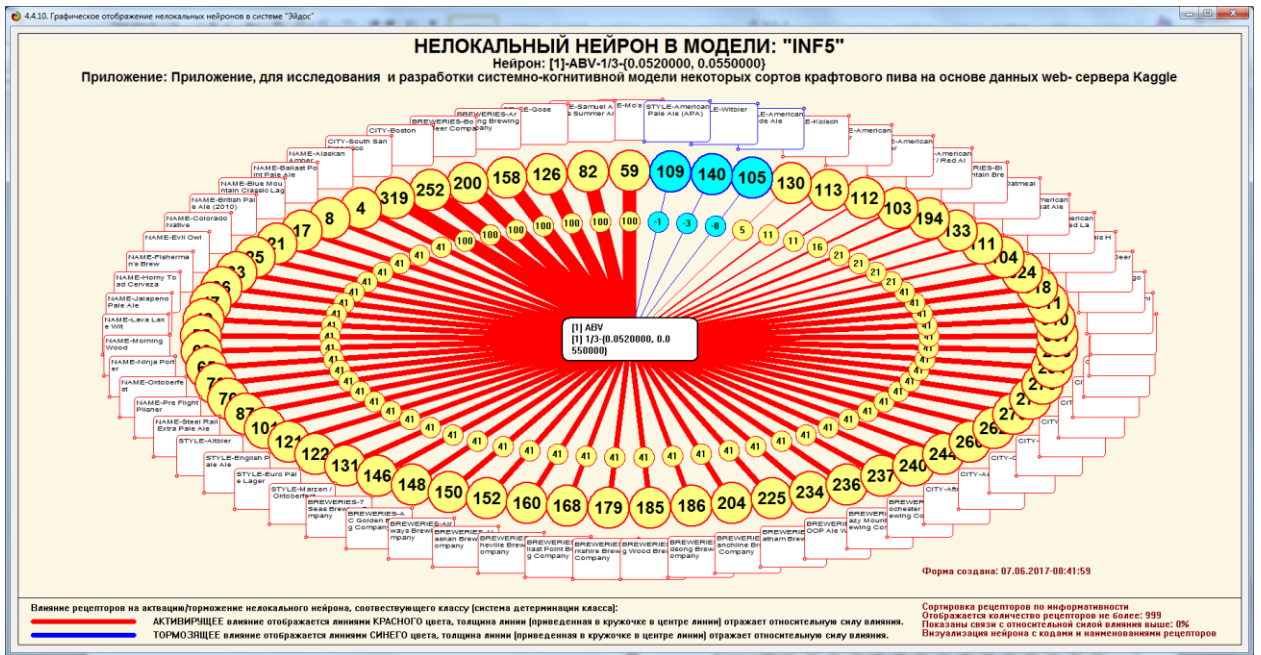


Рисунок 26. Пример нелокальные нейронные сети и нейроны модели INF5

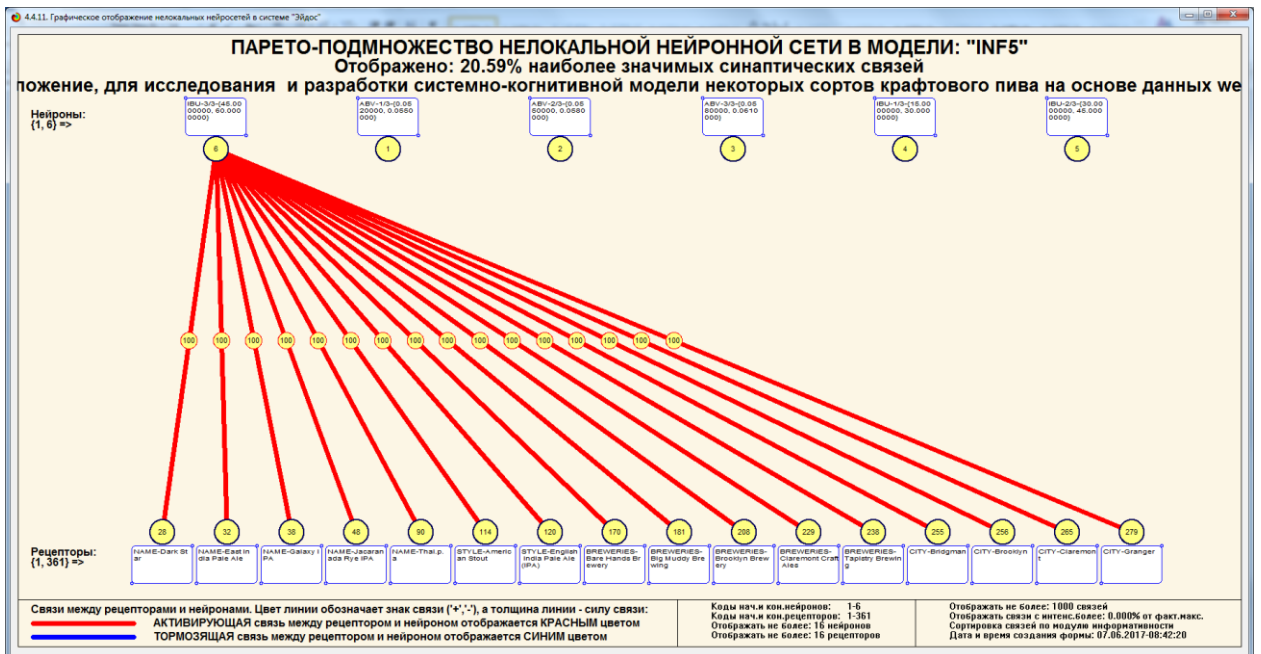


Рисунок 26. Пример нелокальные нейронные сети и нейроны модели INF5

2.5 КЛАСТЕРНЫЙ И КОНСТРУКТИВНЫЙ АНАЛИЗ

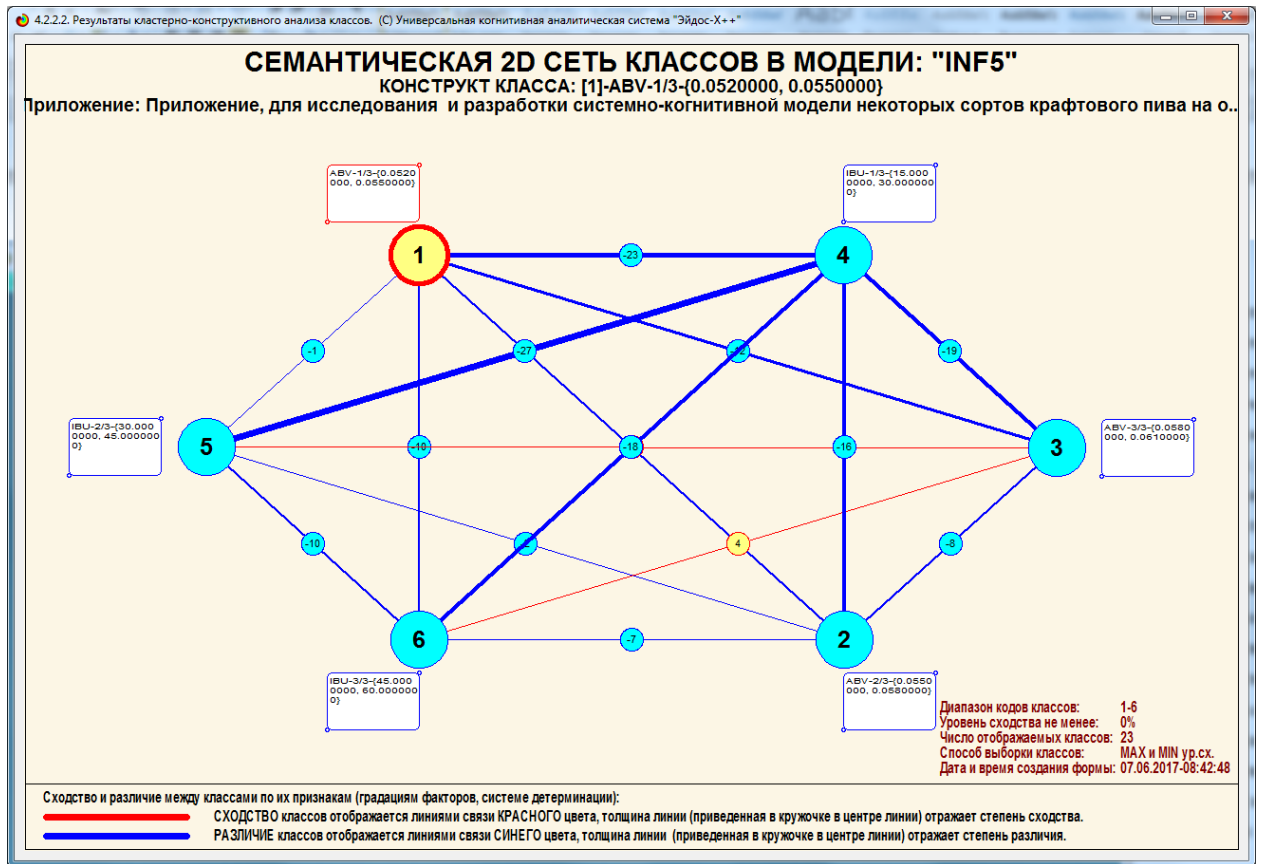


Рисунок 27. Пример кластерный и конструктивный анализ модели INF5

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 инструментарию 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 "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- 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>