

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

Факультет заочного обучения

РЕЦЕНЗИЯ

на курсовой проект (работу)

Студента (ки) Елисеева Николая Геннадьевича

курса 2 заочной формы обучения

направления подготовки 09.04.02 «Информационные системы и технологии»

направленность(профиль) «Информационные системы и технологии в науке и управлении» (академическая магистратура)

Наименование темы «Разработка системно-когнитивной модели изменения цен на недвижимость на основе данных web-сервера Kaggle с использованием системы Aidos».

Рецензент: Луценко Евгений Вениаминович, 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	5
3.	Уровень и корректность использования в работе различных методов исследований	4
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

Актуальность темы применения современных математических моделей и программного инструментария для проведения АСК-анализа на основе базы данных

Недостатки работы

Отдельные погрешности в оформлении

Итоговая оценка при защите _____ 5 _____

Рецензент _____  _____ Е.В. Луценко

«6» декабря 2019

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ И.Т. ТРУБИЛИНА»

Факультет заочного обучения
Кафедра компьютерных технологий и систем

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

к курсовой работе

по дисциплине: Системы представления и приобретения знаний
на тему:

Разработка системно-когнитивной модели изменения цен на недвижимость на основе данных web-сервера Kaggle с использованием системы Aidos

выполнил студент группы ИТ31841

Елисеев Николай Геннадьевич

Допущен к защите

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

(подпись, расшифровка подписи)

Нормоконтролер Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор

(подпись, расшифровка подписи)

Защищена _____

Оценка _____

(дата)

Члены комиссии _____

В.И. Лойко

Е.В. Луценко

В.Н. Лаптев

(подпись, дата, расшифровка подписи)

Краснодар

2019 г.

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ И.Т. ТРУБИЛИНА»

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

УТВЕРЖДАЮ:

Зав. кафедрой КТС _____ В. И. Лойко

ЗАДАНИЕ

на курсовую работу

Студенту: ИТ31841 группы 2 курса
Факультета прикладной информатики
Специальности: 09.04.02 «Информационные системы и технологии»
(шифр)

Елисееву Николаю Геннадьевичу
(Ф.И.О.)

Тема проекта: **Разработка системно-когнитивной модели изменения цен на недвижимость на основе данных web-сервера Kaggle с использованием системы Aidos**

Содержание задания: Проанализировать современные методы и средства формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами

Объем работы:

а) пояснительная записка к работе _____ 31 _____ листов формата А4

б) графическая часть _____ лист формата А4

Рекомендуемая литература: Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос- X++» 2017 г.

Срок выполнения проекта: с “ _____ ” _____ по “ _____ ” _____ 20__ г.

Срок защиты: “ _____ ” _____ 20__ г.

Дата выдачи задания: “ _____ ” _____ 20__ г.

Дата сдачи проекта на кафедру: “ _____ ” _____ 20__ г.

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

(подпись, Ф.И.О., звание, степень)

Задание принял студент

(подпись, дата)

Реферат

Пояснительная записка курсового проекта (работы) 31 с., 21 рис., 1 табл.

Ключевые слова: ДОМ, ЭЙДОС X, СИСТЕМА, ДАННЫЕ, ГРАДАЦИИ, КЛАССИФИКАЦИОННЫЕ ШКАЛЫ, ОПИСАТЕЛЬНЫЕ ШКАЛЫ

В курсовом проекте рассматриваются вопросы прогнозирования повышения цен на недвижимость на основе АСК-анализа показаний датчиков и системы Эйдос на основе данных репозитория Kaggle.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	6
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ.....	7
1.1. Описание решения.....	7
1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel	7
1.3. Синтез и верификация статистических и интеллектуальных моделей.....	13
1.4. Виды моделей системы «Эйдос».....	14
1.5. Результаты верификации моделей	16
2. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ .	20
2.1. Решение задачи идентификации	21
2.2. Когнитивные функции	23
2.3. SWOT матрицы и диаграммы.....	25
2.4. НЕЛОКАЛЬНЫЕ НЕЙРОННЫЕ СЕТИ И НЕЙРОНЫ.....	27
ЗАКЛЮЧЕНИЕ	29
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	30

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 «houses» из банка исходных данных по задачам искусственного интеллекта – репозитория Kaggle.

Для решения задачи используем стандартные возможности Microsoft Office Word и Excel, блокнот, а также систему искусственного интеллекта "Эйдос- X++".

1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

1.1. Описание решения

В соответствии с методологией АСК-анализа решение поставленной задачи проведем в четыре этапа:

1. Преобразование исходных данных из csv-формата в промежуточные файлы MS Excel.
2. Преобразование исходных данных из промежуточных файлов MS Excel в базы данных системы "Эйдос".
3. Синтез и верификация моделей предметной области.
4. Применение моделей для решения задач идентификации, прогнозирования и исследования предметной области.

1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel

Из электронного ресурса баз данных Kaggle возьмем базу данных– <https://www.kaggle.com/shree1992/housedata> , которую оставим без изменений.

Общее описание задачи:

1. Дата
2. Спальня
3. Ванная комната
4. Площадь дома
5. Число этажей
6. Улица
7. Город
8. Индекс

Столбцы 2-13 описательные шкалы.

Столбец 2 является классификационной шкалой. Этот столбец показывает категорию блюда.

Обучающая выборка:

Таблица 1 – houses.xls

№	А	В	С	Д	Е	Г	Н	И	Т	К	Л	М	О	Р	Q	Т	U	V	W	
1	№	дата	bedrooms	bathrooms	sqft_bsvg	sqft_tot	floor	sqft_bsvg	sqft_basement	sqft_tot	yr_built	yr_renovated	street	city	state	price	view	condition		
2	1	02.05.2014 1.0	2.0	1340	7912 2.0	1340	0	1340	0	1340	2005	18822	22000	Everett	WA	98201	0	3		
3	2	02.05.2014 1.0	2.0	3650	9656 2.0	3370	280	3370	0	1921	0	1921	0	709 W Blaine St	Seattle	WA	98113	234800.0	4	5
4	3	02.05.2014 1.0	2.0	1330	11947 1.0	3330	0	1948	0	1948	0	1948	0	26296 26234 134th Av Kent	WA	98042	342000.0	0	4	
5	4	02.05.2014 1.0	2.0	2000	8010 1.0	8000	1000	1843	0	1843	0	1843	0	0 217 320th Pl NE	Bellevue	WA	98008	420000.0	0	4
6	5	02.05.2014 4.0	2.0	1940	16506 1.0	1140	833	1976	0	1976	1992	9105 1305th Ave NE	Redmond	WA	98052	530000.0	0	4		
7	6	02.05.2014 2.0	1.0	880	4188 1.0	880	0	1918	0	1918	1994	532 NE 80th St	Seattle	WA	98115	490000.0	0	3		
8	7	02.05.2014 1.0	2.0	1350	2568 1.0	1350	0	1978	0	1978	0	2614 194th Ave NE	Redmond	WA	98052	315000.0	0	3		
9	8	02.05.2014 4.0	2.0	8750	35864 2.0	2730	0	1989	0	1989	0	2376 SE 213rd Pl	Maple Vall	WA	98039	432000.0	0	3		
10	9	02.05.2014 1.0	2.0	2450	88426 1.0	3730	880	1985	0	1985	0	46611 46625 SE 128th North Ben	WA	98045	412500.0	0	4			
11	10	02.05.2014 4.0	2.0	1320	8290 1.0	3520	0	1945	0	1945	2010	4611 509th Ave NE	Seattle	WA	98112	463000.0	0	4		
12	11	02.05.2014 1.0	2.0	1750	7139 1.0	1750	0	1948	0	1948	1994	4090 14th Ave NE	Seattle	WA	98105	463000.0	0	4		
13	12	02.05.2014 4.0	2.0	7920	4080 1.0	7910	1010	1909	0	1909	1908	1838 4098 44th Ave N Seattle	WA	98107	1400000.0	0	3			
14	13	02.05.2014 1.0	1.0	1090	1482 1.0	1090	300	1988	0	1988	2003	2048 SW Fordland Ct	Seattle	WA	98136	265000.0	0	4		
15	14	02.05.2014 1.0	2.0	2910	9489 1.0	2910	0	1919	0	1919	1969	3334 48th Ave NE	Seattle	WA	98125	1200000.0	0	3		
16	15	02.05.2014 1.0	2.0	1300	9728 1.0	1300	0	1965	0	1965	0	14814 SE 201st St	Kent	WA	98042	242600.0	0	4		
17	16	02.05.2014 1.0	2.1	1370	6761 1.0	1370	0	1946	0	1946	0	0 16424 NE 107 St	Bellevue	WA	98007	413000.0	0	4		
18	17	02.05.2014 4.0	3.0	5110	7211 2.0	3110	0	1997	0	1997	0	0 11224 SE 106th Pl	Auburn	WA	98002	387500.0	0	3		
19	18	02.05.2014 1.0	3.1	1370	3658 1.0	1370	0	1987	0	1987	2000	14925 S 242nd Pl	Des Moines	WA	98139	257500.0	0	3		
20	19	02.05.2014 1.0	3.2	1180	10277 1.0	1180	0	1983	0	1983	2009	12421 421st Ave SE	North Ben	WA	98046	275000.0	0	3		
21	20	02.05.2014 1.0	3.3	2340	10578 2.0	1550	890	1923	0	1923	0	0 3225 NE 52nd St	Seattle	WA	98115	750000.0	0	5		
22	21	02.05.2014 1.0	3.0	1450	8806 1.0	1450	0	1964	0	1964	1979	1912 1547th Ave SE	Bellevue	WA	98008	413000.0	0	4		
23	22	02.05.2014 1.0	2.0	1790	1571 1.1	1690	280	2005	0	2005	0	0 1348 Franklin Ave NE	Seattle	WA	98105	415000.0	0	3		
24	23	02.05.2014 4.0	2.1	8750	12281 2.0	2730	0	1991	0	1991	0	0 14212 NE 155th Pl	Bellevue	WA	98011	812500.0	0	3		
25	24	02.05.2014 4.0	2.2	1800	6188 1.0	1130	470	1959	0	1959	1989	2021 NE 130th St	Seattle	WA	98122	435000.0	0	3		
26	25	02.05.2014 1.0	3.3	2090	10834 1.0	1560	750	1987	0	1987	0	0 27734 29th Avenue S Federal W	WA	98023	285000.0	0	4			
27	26	02.05.2014 1.0	2.4	2960	7291 1.0	1890	1300	1948	0	1948	0	0 1816 16th 41st Ave S Federal W	WA	98026	413000.0	0	3			
28	27	02.05.2014 4.0	2.5	2080	11798 1.1	1900	833	1936	0	1936	0	0 10164th St	Kirkland	WA	98034	448000.0	0	3		
29	28	02.05.2014 1.0	2.4	2820	8791 2.0	2820	0	1979	0	1979	2014	1335 S 12nd Way	Hoquiam	WA	98529	475000.0	0	3		
30	29	02.05.2014 1.0	2.7	2660	4764 1.0	1700	800	1951	0	1951	1999	314 19th 76th St	Seattle	WA	98112	790000.0	0	4		
31	30	02.05.2014 4.0	2.8	1560	8790 1.0	1560	0	1967	0	1967	0	0 14104 112th Ave NE	Kirkland	WA	98034	302500.0	0	4		
32	31	02.05.2014 4.0	2.9	2860	1343 2.0	2290	870	2004	0	2004	2003	20120 137th Ave NE	Woodinville	WA	98072	499500.0	0	3		
33	32	02.05.2014 4.0	2.0	1620	9690 1.1	1640	180	1945	0	1945	2012	7905 7905 10th Ave S Seattle	WA	98116	410000.0	0	3			
34	33	02.05.2014 4.0	2.1	2620	8486 2.0	2620	0	2014	0	2014	0	0 17624 4th Ave NE	Shoreline	WA	98153	425000.0	0	3		
35	34	02.05.2014 4.0	2.2	3630	42864 2.1	2300	1330	1979	0	1979	2014	5172 5172 21st Heather At Auburn	WA	98002	420000.0	0	3			
36	35	02.05.2014 1.0	2.3	3240	3321 2.0	3240	0	1995	0	1995	0	0 36022 36th Ct SW	Federal W	WA	98023	604000.0	2	3		

Всего в базе 17 столбцов, 2 классификационных и 13 описательных.

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

Конвертируем исходный файл из csv формата в xls и запишем его с именем: Inp_data.xls в папку: c:\Aidos-X\AID_DATA\Inp_data\. В результате получим таблицу исходных данных, полностью подготовленную для обработки в системе «Эйдос» и записанную в нужную папку в виде файла нужного типа с нужным именем.

Автоматизированная формализация предметной области путем импорта исходных данных из внешних баз данных в систему "Эйдос".

Для загрузки базы исходных данных в систему «Эйдос» необходимо воспользоваться универсальным программным интерфейсом для ввода

данных из внешних баз данных табличного вида, т.е. режимом 2.3.2.2.

Импорт данных

Задайте тип файла исходных данных: "Inp_data":

- XLS - MS Excel-2003 (Стандарт XLS-файла)
- XLSX- MS Excel-2007(2010)
- DBF - DBASE IV (DBF/NTX) (Стандарт DBF-файла)
- CSV - Comma-Separated Values (Стандарт CSV-файла)

Задайте параметры:

- Нули и пробелы считать ОТСУТСТВИЕМ данных
- Нули и пробелы считать ЗНАЧЕНИЯМИ данных
- Создавать БД средних по классам "Inp_davr.dbf"?

Требования к файлу исходных данных

Задайте диапазон столбцов классификационных шкал:

Начальный столбец классификационных шкал: 15

Конечный столбец классификационных шкал: 17

Задайте диапазон столбцов описательных шкал:

Начальный столбец описательных шкал: 2

Конечный столбец описательных шкал: 14

Задайте режим:

- Формализации предметной области (на основе "Inp_data")
- Генерации распознаваемой выборки (на основе "Inp_gasp")

Задайте способ выбора размера интервалов:

- Равные интервалы с разным числом наблюдений
- Разные интервалы с равным числом наблюдений

Задание параметров формирования сценариев или способа интерпретации текстовых полей "Inp_data":

- Не применять сценарный метод АСК-анализа
- Применить сценарный метод АСК-анализа
- Применить спец. интерпретацию текстовых полей классов
- Применить спец. интерпретацию текстовых полей признаков

Параметры интерпретации значений текстовых полей "Inp_data":

Интерпретация TXT-полей классов:

Значения полей текстовых классификационных шкал файла исходных данных "Inp_data" рассматриваются как целое

Интерпретация TXT-полей признаков:

Значения полей текстовых описательных шкал файла исходных данных "Inp_data" рассматриваются как целое

Какие наименования ГРАДАЦИЙ числовых шкал использовать:

- Только интервальные числовые значения (например: "1/3-{59873.0000000, 178545.6666667}")
- Только наименования интервальных числовых значений (например: "Минимальное")
- И интервальные числовые значения, и их наименования (например: "Минимальное: 1/3-{59873.0000000, 178545.6666667}")

Ok Cancel

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

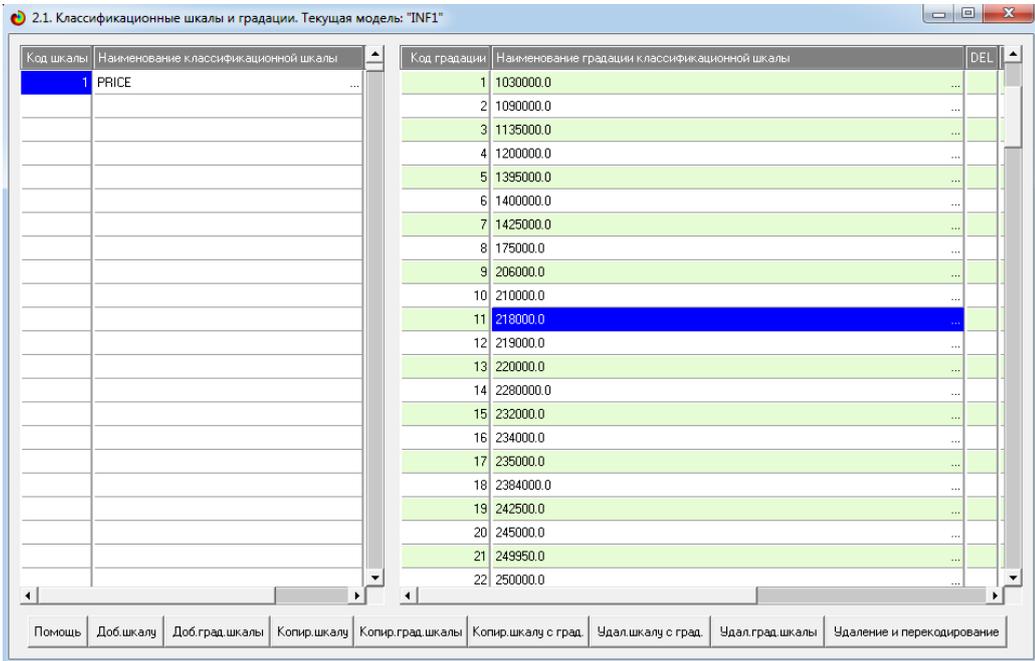
В экранной форме, приведенной на рисунке 5, задать настройки, показанные на рисунке:

– "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);

Рисунок 3. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

В результате формируются классификационные и описательные шкалы и градации, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Этим самым полностью автоматизировано выполняется 2-й этап АСК- анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1 (рисунок 7).



The screenshot shows a software window with two data tables. The left table lists classification scales, and the right table lists gradations for a specific scale.

Код шкалы	Наименование классификационной шкалы
1	PRICE

Код градации	Наименование градации классификационной шкалы	DEL
1	1030000.0	...
2	1090000.0	...
3	1135000.0	...
4	1200000.0	...
5	1395000.0	...
6	1400000.0	...
7	1425000.0	...
8	175000.0	...
9	206000.0	...
10	210000.0	...
11	218000.0	...
12	219000.0	...
13	220000.0	...
14	2280000.0	...
15	232000.0	...
16	234000.0	...
17	235000.0	...
18	2384000.0	...
19	242500.0	...
20	245000.0	...
21	249950.0	...
22	250000.0	...

At the bottom of the window, there is a toolbar with the following buttons: Помощь, Доб. шкалу, Доб. град. шкалы, Копир. шкалу, Копир. град. шкалы, Копир. шкалу с град., Удал. шкалу с град., Удал. град. шкалы, Удаление и перекодирование.

Рисунок 4. Классификационные шкалы и градации (фрагмент)

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

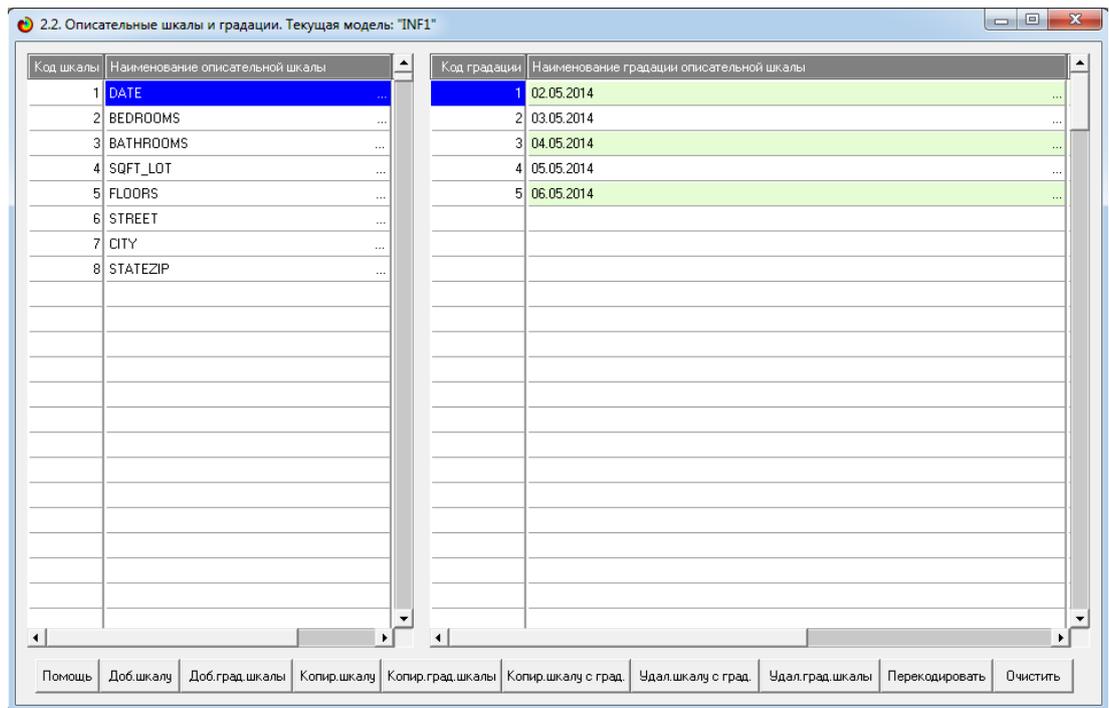


Рисунок 5. Описательные шкалы и градации (фрагмент)

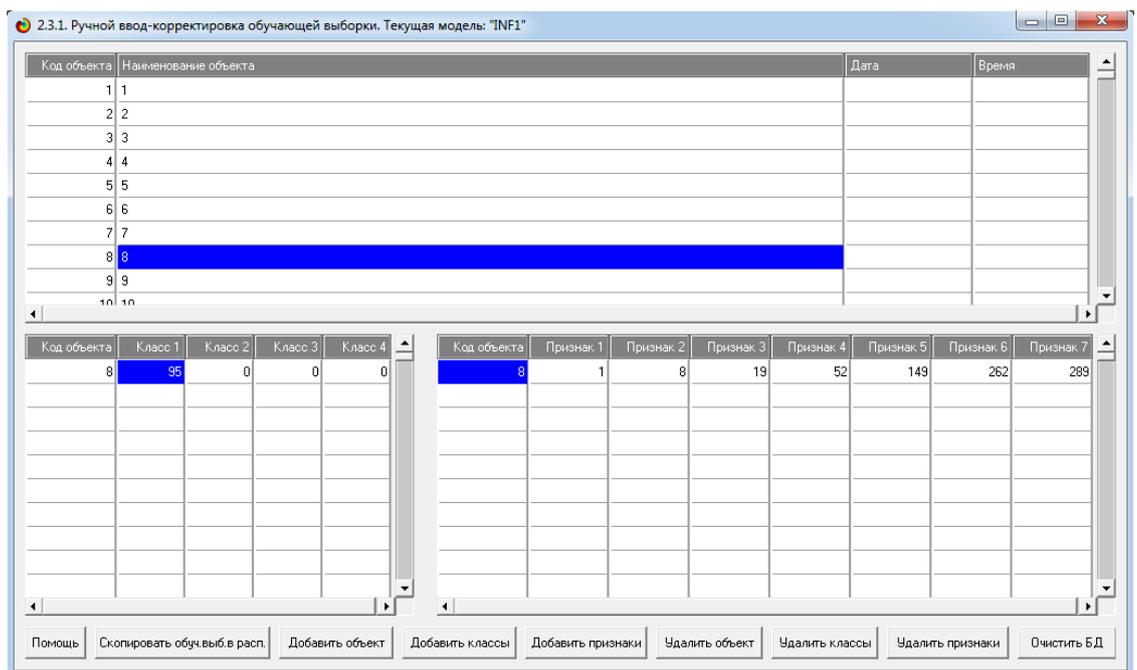


Рисунок 6. Обучающая выборка (фрагмент)

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

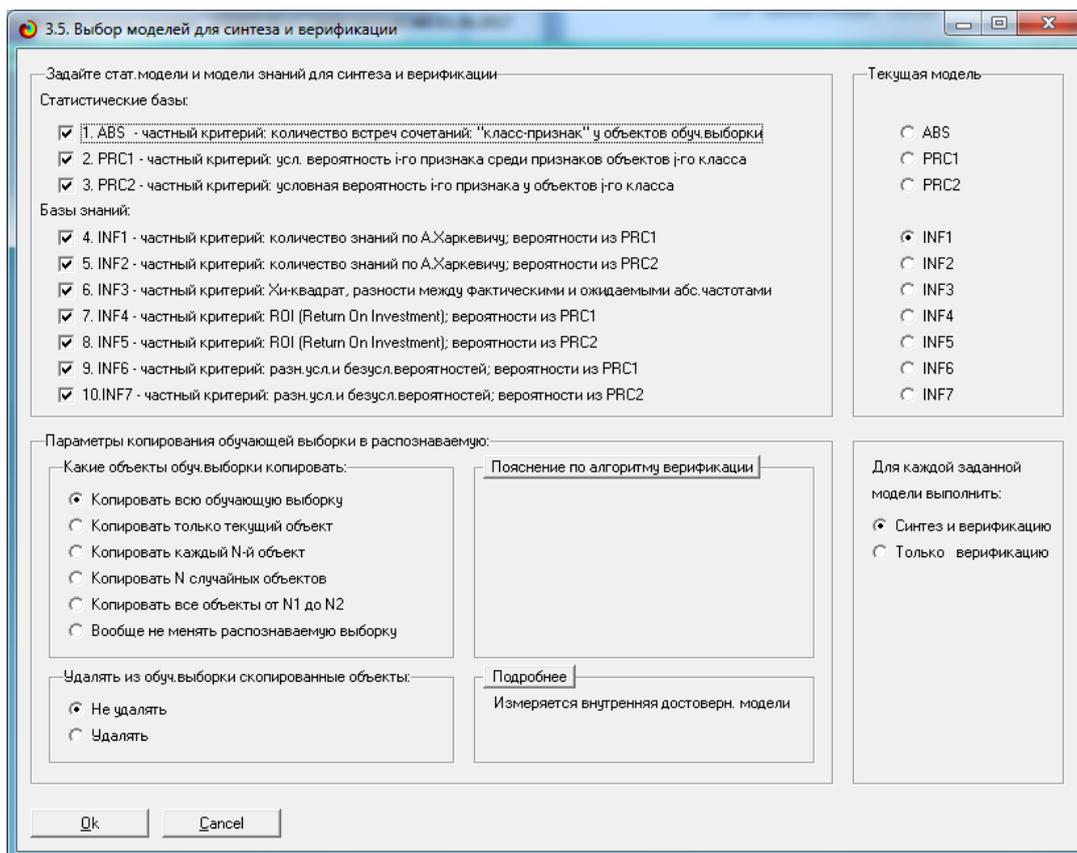


Рисунок 7. Выбор моделей для синтеза и верификации, а также текущей модели

В данном режиме имеется много различных методов верификации моделей, в том числе и поддерживающие бутстрепный метод. Но мы используем параметры по умолчанию, приведенные на рисунке 10. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 11.

условных и безусловных процентных распределений, и матрицы знаний (таблицы 3 и 4) (проф. В.И.Лойко).

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

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний "Класс-признак" у объектов обучающей выборки"

Код признака	Наименование описательной шкалы и градации	1. PRICE 1030000.0	2. PRICE 1090000.0	3. PRICE 1135000.0	4. PRICE 1200000.0	5. PRICE 1395000.0	6. PRICE 1400000.0	7. PRICE 1425000.0	8. PRICE 175000.0	9. PRICE 206000.0
1	DATE-02.05.2014				1		1			
2	DATE-03.05.2014									
3	DATE-04.05.2014									
4	DATE-05.05.2014	1	1			1		1		
5	DATE-06.05.2014			1						1
6	BEDROOMS-2.0							1		
7	BEDROOMS-3.0									1
8	BEDROOMS-4.0		1	1			1			
9	BEDROOMS-5.0	1			1	1				
10	BEDROOMS-6.0									
11	BEDROOMS-7.0									
12	BATHROOMS-1.0								1	
13	BATHROOMS-1.1	1	1							
14	BATHROOMS-1.2									
15	BATHROOMS-1.3									
16	BATHROOMS-1.4									
17	BATHROOMS-1.5									
18	BATHROOMS-1.6									
19	BATHROOMS-2.0				1		1			
20	BATHROOMS-2.1							1		
21	BATHROOMS-2.10									
22	BATHROOMS-2.11									
23	BATHROOMS-2.12									
24	BATHROOMS-2.2			1						

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

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу: вероятности из PRCL1"

Код признака	Наименование описательной шкалы и градации	1. PRICE 1030000.0	2. PRICE 1090000.0	3. PRICE 1135000.0	4. PRICE 1200000.0	5. PRICE 1395000.0	6. PRICE 1400000.0	7. PRICE 1425000.0	8. PRICE 175000.0	9. PRICE 206000.0	10. PRICE 210000.0
1	DATE-02.05.2014				1.112		1.112				
2	DATE-03.05.2014										
3	DATE-04.05.2014										
4	DATE-05.05.2014	0.900	0.900			0.900		0.900		0.900	0.90
5	DATE-06.05.2014			1.647					1.647		
6	BEDROOMS-2.0							2.037		2.037	
7	BEDROOMS-3.0								0.895		
8	BEDROOMS-4.0		1.112	1.112			1.112				
9	BEDROOMS-5.0	2.474			2.474	2.474					2.47
10	BEDROOMS-6.0										
11	BEDROOMS-7.0										
12	BATHROOMS-1.0								1.998	1.998	
13	BATHROOMS-1.1	2.474	2.474								
14	BATHROOMS-1.2										
15	BATHROOMS-1.3										
16	BATHROOMS-1.4										
17	BATHROOMS-1.5										
18	BATHROOMS-1.6										
19	BATHROOMS-2.0				1.732		1.732				1.73
20	BATHROOMS-2.1							2.675			
21	BATHROOMS-2.10										
22	BATHROOMS-2.11										
23	BATHROOMS-2.12										
24	BATHROOMS-2.2			3.019							

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

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

Код признака	Наименование отдельной ячейки и графы	1. PRICE 1030000.0	2. PRICE 1090000.0	3. PRICE 1150000.0	4. PRICE 1200000.0	5. PRICE 1365000.0	6. PRICE 1400000.0	7. PRICE 1425000.0	8. PRICE 1750000.0	9. PRICE 2060000.0	10. PRICE 2100000.0
1	DATE 02.05.2014	-0.339	-0.339	-0.339	0.661	-0.339	0.661	-0.339	-0.339	-0.339	-0.339
2	DATE 03.05.2014	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021	-0.021
3	DATE 04.05.2014	-0.026	-0.026	-0.026	-0.026	-0.026	-0.026	-0.026	-0.026	-0.026	-0.026
4	DATE 05.05.2014	0.587	0.587	-0.413	-0.413	0.587	-0.413	0.587	-0.413	0.587	0
5	DATE 06.05.2014	-0.201	-0.201	0.799	-0.201	-0.201	-0.201	-0.201	0.799	-0.201	-0.201
6	BEDROOMS-2.0	-0.138	-0.138	-0.138	-0.138	-0.138	-0.138	0.862	-0.138	0.862	-0.138
7	BEDROOMS-3.0	-0.418	-0.418	-0.418	-0.418	-0.418	-0.418	-0.418	0.799	0.592	-0.418
8	BEDROOMS-4.0	-0.339	0.661	0.661	-0.339	-0.339	0.661	-0.339	-0.339	-0.339	-0.339
9	BEDROOMS-5.0	0.910	-0.090	-0.090	0.910	0.910	-0.090	-0.090	-0.090	-0.090	0
10	BEDROOMS-6.0	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011
11	BEDROOMS-7.0	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005
12	BATHROOMS-1.0	-0.143	-0.143	-0.143	-0.143	-0.143	-0.143	-0.143	0.857	0.857	-0.143
13	BATHROOMS-1.1	0.910	0.910	-0.090	-0.090	-0.090	-0.090	-0.090	-0.090	-0.090	-0.090
14	BATHROOMS-1.2	-0.037	-0.037	-0.037	-0.037	-0.037	-0.037	-0.037	-0.037	-0.037	-0.037
15	BATHROOMS-1.3	-0.032	-0.032	-0.032	-0.032	-0.032	-0.032	-0.032	-0.032	-0.032	-0.032
16	BATHROOMS-1.4	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011
17	BATHROOMS-1.5	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005
18	BATHROOMS-1.6	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005
19	BATHROOMS-2.0	-0.105	-0.105	-0.105	0.015	-0.105	0.015	-0.105	-0.105	-0.105	0
20	BATHROOMS-2.1	-0.074	-0.074	-0.074	-0.074	-0.074	-0.074	0.926	-0.074	-0.074	-0.074
21	BATHROOMS-2.10	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011
22	BATHROOMS-2.11	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005
23	BATHROOMS-2.12	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005
24	BATHROOMS-2.2	-0.053	-0.053	0.947	-0.053	-0.053	-0.053	-0.053	-0.053	-0.053	-0.053

1.5. Результаты верификации моделей

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

4.1.3.6. Обобщенная форма по достов.моделям при разн.крит. Текущая модель: "INF2"

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Рибзегена	Средний уровень истинности реше
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс. частот с обр...	18	18	2941	13574		0.001	1.000	0.003	
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс. частот по признак...	18	18	14119	2396		0.007	1.000	0.015	
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	18	18	2984	13531		0.001	1.000	0.003	
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	18	18	14137	2378		0.008	1.000	0.015	
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	18	18	2984	13531		0.001	1.000	0.003	
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	18	18	14136	2379		0.008	1.000	0.015	
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Семантический резонанс зна...	18	18	2980	13535		0.001	1.000	0.003	
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Сумма знаний	18	18	15297	1210		0.015	1.000	0.029	
7. INF4 - частный критерий: ROI (Return On Investment), вероятно...	Семантический резонанс зна...	18	18	2988	13527		0.001	1.000	0.003	
7. INF4 - частный критерий: ROI (Return On Investment), вероятно...	Сумма знаний	18	18	14136	2379		0.008	1.000	0.015	

Помощь по мерам достоверности | Помощь по частотам распределения | TP, TN, FP, FN | (TP-FP), (TN-FN) | (T-F)/(T+F)*100 | Задать интервал сглаживания

Рисунок 9. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели ABS, INF3 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 70%.

Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием (рисунок 13).

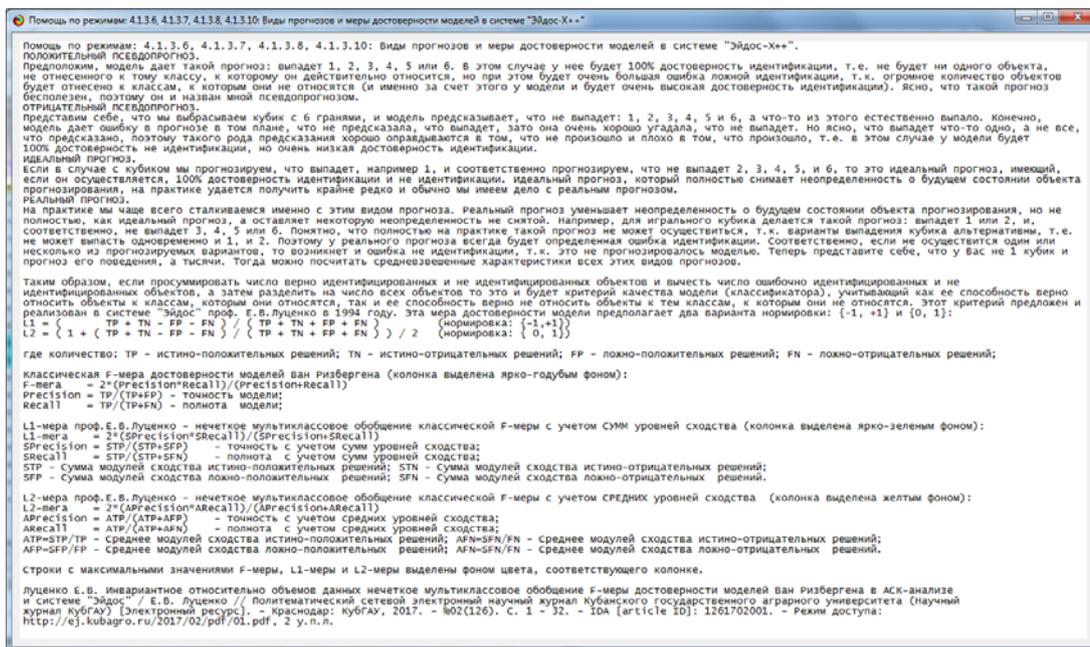


Рисунок 10. Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

Также обращает на себя внимание, что статистические модели, как правило, дают более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, и практически никогда – более высокую. Этим и оправдано применение моделей знаний и интеллектуальных технологий. На рисунке 11 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и неидентифицированных ситуаций в наиболее достоверной модели INF3.

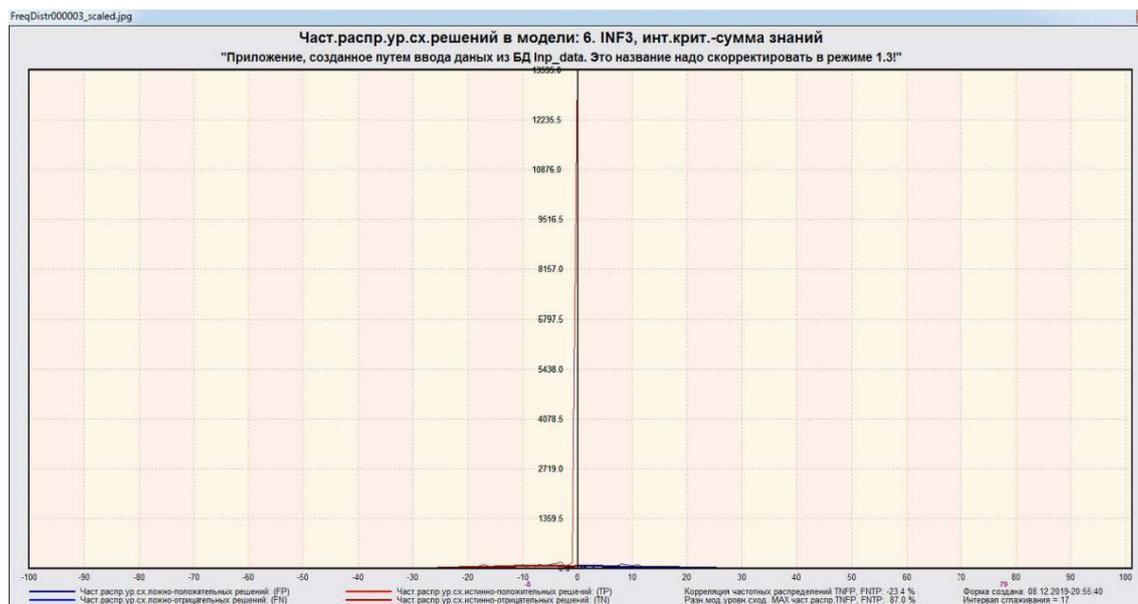


Рисунок 11. Частотное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF3

Из рисунка 11 видно, что:

- наиболее достоверная модель INF3 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 9);

- модуль уровня сходства-различия в наиболее достоверной модели INF5 для верно идентифицированных и верно неидентифицированных объектов значительно выше, чем для ошибочно идентифицированных и ошибочно неидентифицированных. Это верно практически для всего диапазона уровней сходства-различия, кроме небольших по модулю значений в диапазоне от 0 до 15% уровня сходства. Для уровня сходства выше 70% все положительные решения являются истинными. Для очень больших значений уровней сходства-различия (более 70%) также различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями практически отсутствует.

Любые данные о наблюдениях можно считать суммой истинного значения и шума, причем ни первое, ни второе неизвестны. Поэтому имеет смысл сравнить созданные модели с чисто случайными моделями,

совпадающими по основным характеристикам. В системе «Эйдос» есть лабораторная работа № 2.01: «Исследование RND-модели при различных объемах выборки». Если данная работа устанавливается при отсутствии текущего приложения, то все параметры создаваемых моделей задаются вручную, если же текущая модель существует, как в нашем случае, то все основные ее параметры определяются автоматически.



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

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

Совершенно очевидное различие частотных распределений уровней сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования и случайной

модели (рисунки 9 и 11) 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 1.

На рисунке 12 приведены данные по достоверности статистических и когнитивных моделей, созданных на основе случайной выборки.

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

- достоверность лучшей модели INF3, отражающей реальный объект моделирования, примерно на 20% выше, чем аналогичной случайной модели;

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

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

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:

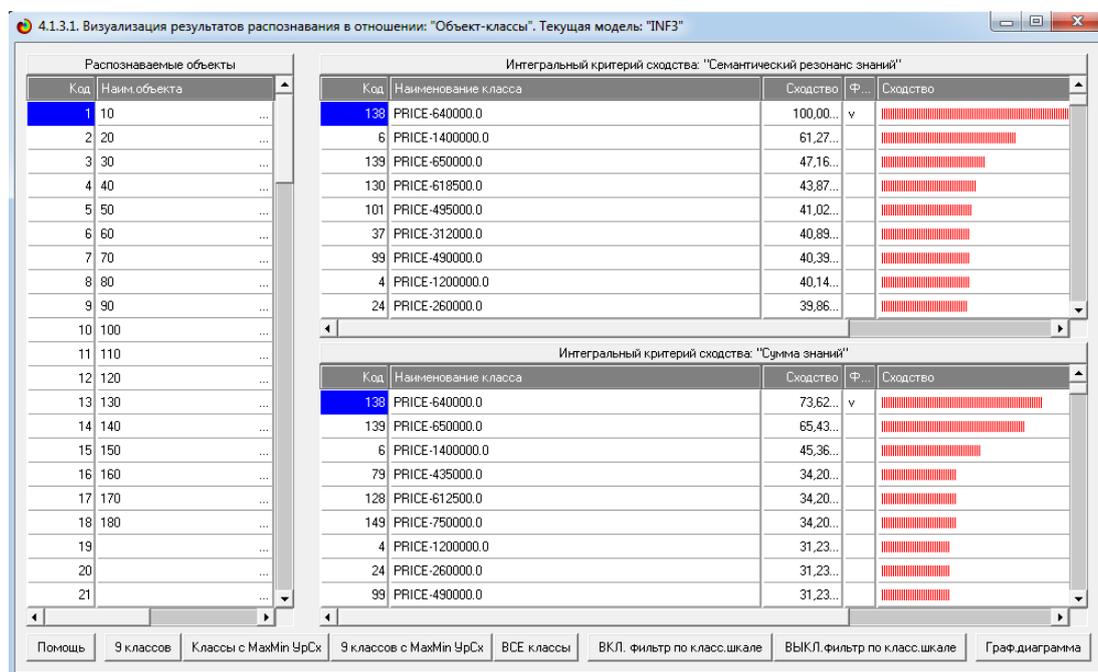


Рисунок 14. Пример идентификации классов в модели INF3

4.1.3.3. Итоговая наглядная форма результатов распознавания: "Объект-класс". Текущая модель: "INF3"

Код: Интегральный критерий

1 Семантический резонанс знаний

2 Сумма знаний

Пояснения по смыслу частных и интегральных критериев

Частн.крит.: 7 моделей знаний Инт.крит.: "Сумма знаний" Инт.крит.: "Резонанс знаний"

Код объекта распозн. выборки	Наименование объекта распознаваемой выборки	Код класса с МАХ ур. сход.	Наименование класса с МАХ уровнем сходства	МАХ уровень сходства	Код класса с MIN ур. сход.	Наименование класса с MIN уровнем сходства	MIN уровень сходства
1	10	138	PRICE-640000.0	100.000	105	PRICE-509900.0	-31.600
2	20	26	PRICE-275000.0	99.566	46	PRICE-340000.0	-31.025
3	30	157	PRICE-790000.0	99.851	162	PRICE-830000.0	-29.335
4	40	149	PRICE-750000.0	78.025	166	PRICE-890000.0	-33.623
5	50	164	PRICE-838000.0	99.608	123	PRICE-600000.0	-26.813
6	60	32	PRICE-295000.0	99.577	21	PRICE-249950.0	-22.548
7	70	51	PRICE-365000.0	83.357	46	PRICE-340000.0	-25.284
8	80	144	PRICE-705380.0	99.564	138	PRICE-640000.0	-24.797
9	90	15	PRICE-232000.0	99.564	81	PRICE-439950.0	-23.392
10	100	5	PRICE-1395000.0	99.572	113	PRICE-550000.0	-29.797
11	110	13	PRICE-220000.0	99.587	117	PRICE-561000.0	-25.871
12	120	123	PRICE-600000.0	84.126	132	PRICE-625000.0	-27.168
13	130	112	PRICE-549000.0	73.415	45	PRICE-335000.0	-27.988
14	140	137	PRICE-631625.0	99.637	45	PRICE-335000.0	-34.355
15	150	75	PRICE-419190.0	99.691	51	PRICE-365000.0	-24.507
16	160	98	PRICE-489000.0	99.755	115	PRICE-555000.0	-22.701

Помощь В начало БД В конец БД Предыдущая Следующая

Рисунок 15. Пример идентификации классов в модели INF3

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)

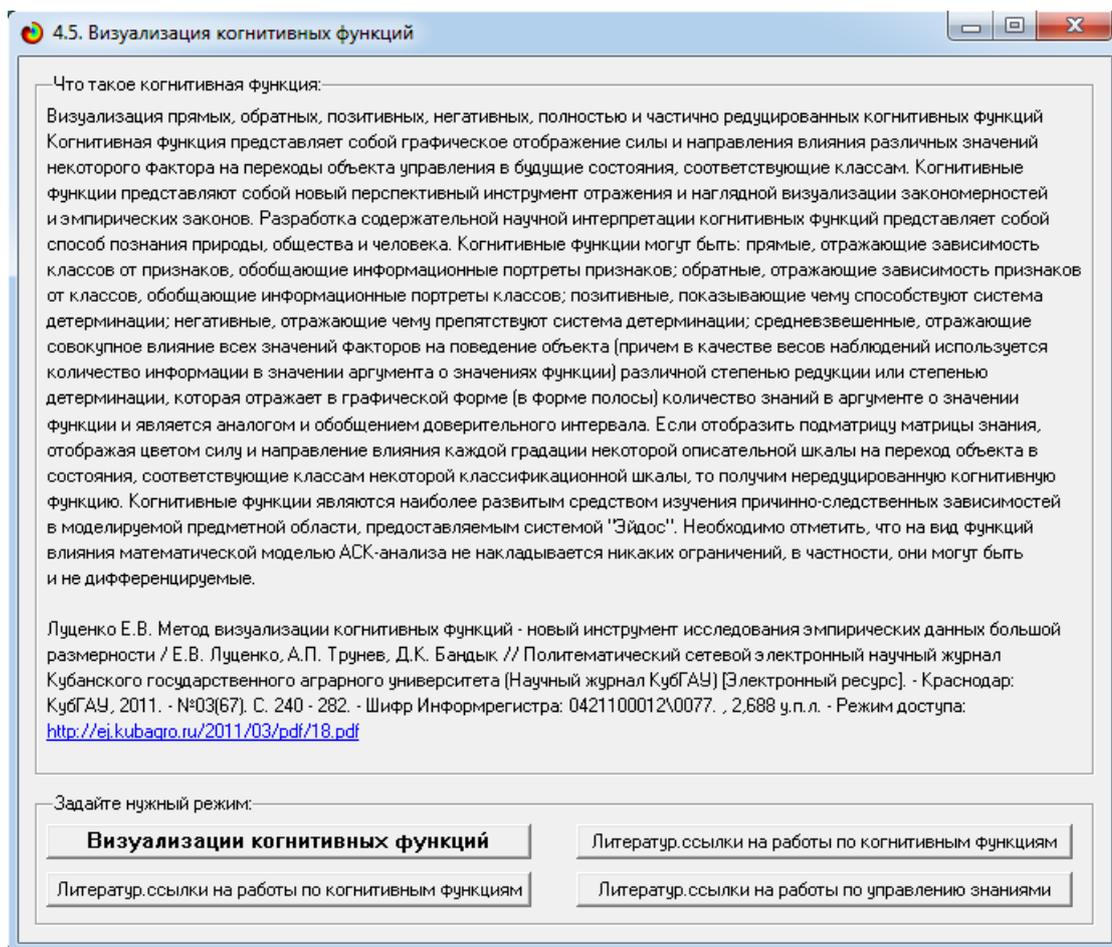


Рисунок 16. Экранная форма режима 4.5 системы «Эйдос-X++» «Визуализация когнитивных функций»

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 INF3.

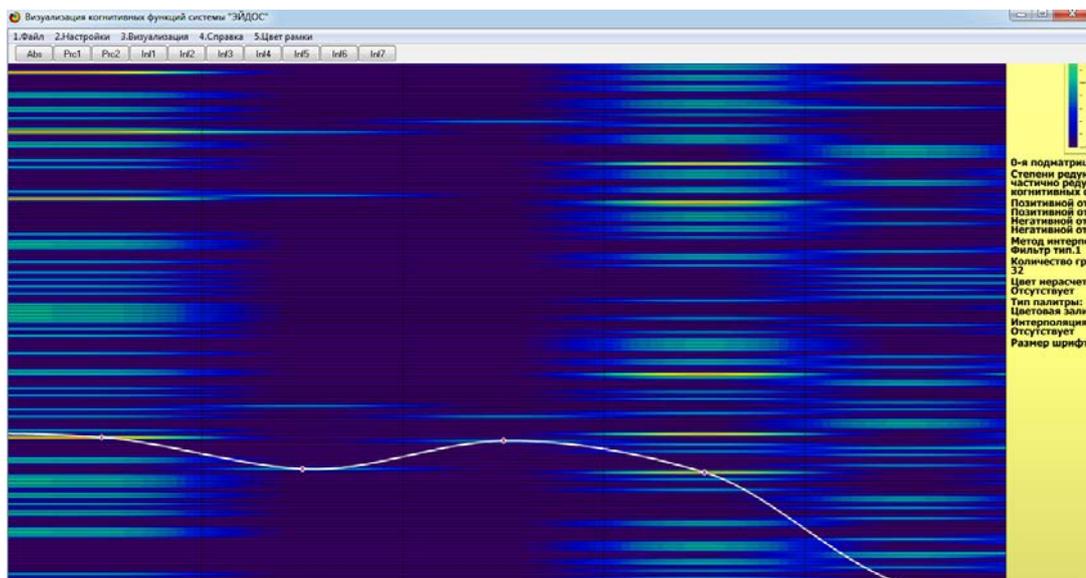


Рисунок 17. Визуализация когнитивных функций для обобщенных классов и всех описательных шкал для модели INF3

2.3. SWOT 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).

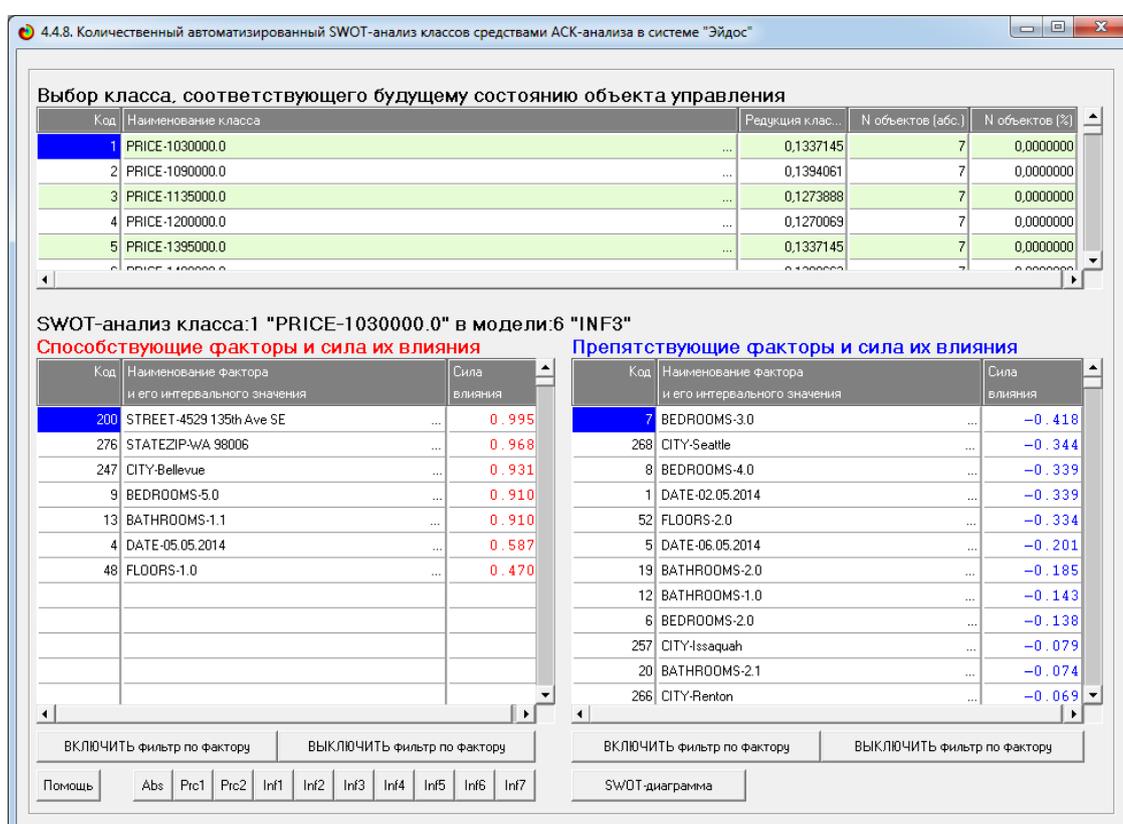


Рисунок 18. Пример SWOT-матрицы в модели INF3

На рисунке 25 приведены примеры инвертированной SWOT- матрицы и инвертированной SWOT-диаграммы в модели INF3.

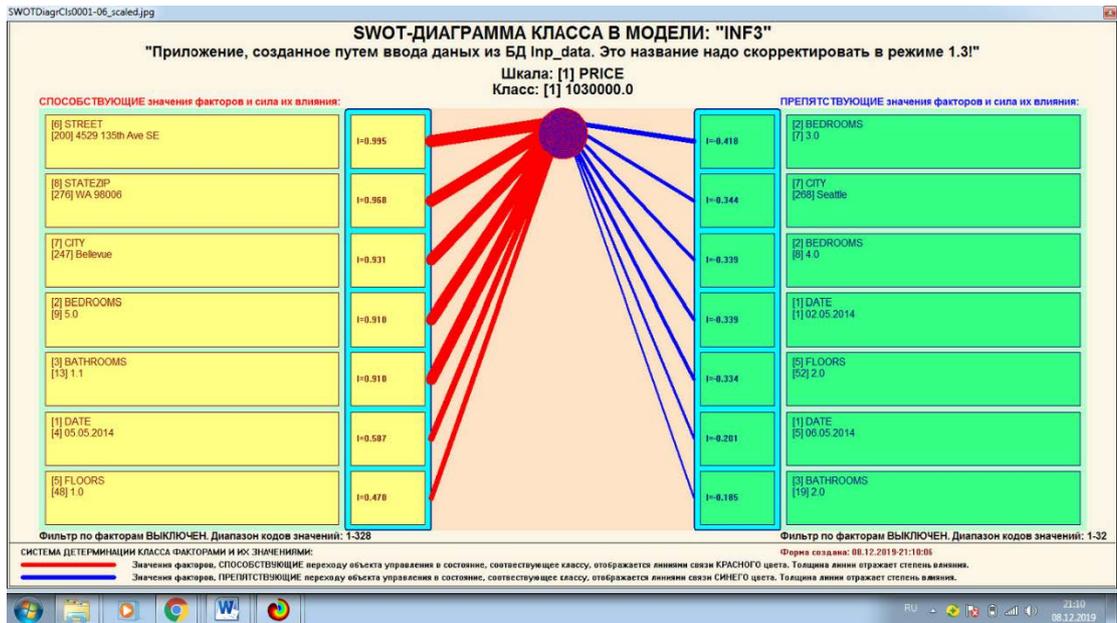


Рисунок 19. Пример SWOT-матрицы в модели INF3

2.4. НЕЛОКАЛЬНЫЕ НЕЙРОННЫЕ СЕТИ И НЕЙРОНЫ.

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

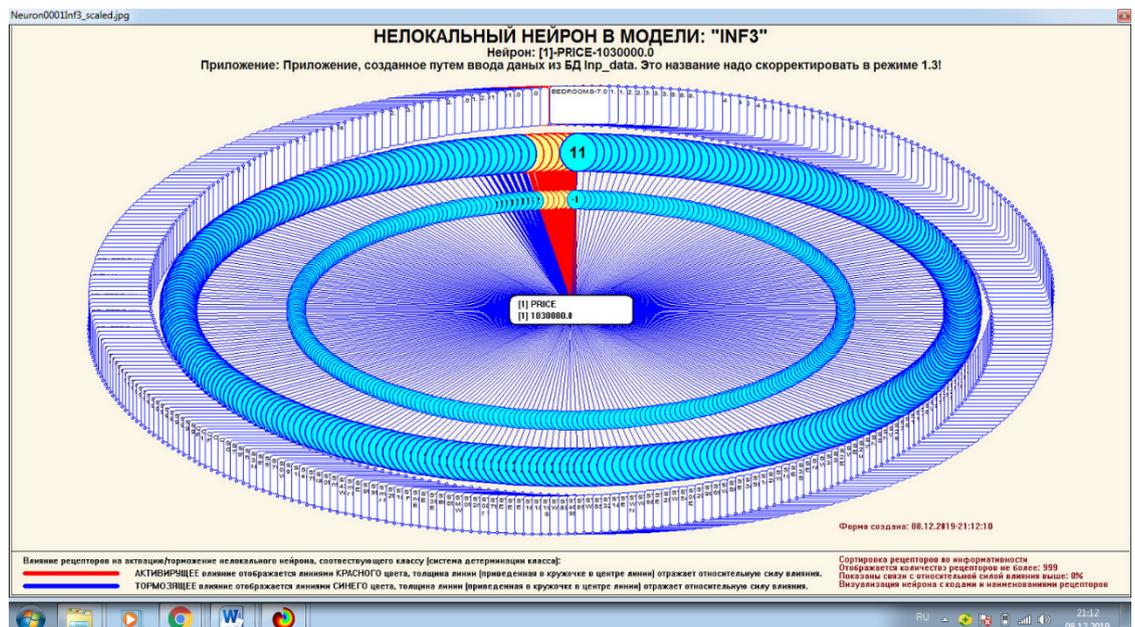


Рисунок 20. Пример нелокальные нейронные сети и нейроны модели INF3



Рисунок 21. Пример нелокальные нейронные сети и нейроны модели INF3

ЗАКЛЮЧЕНИЕ

Так как существует множество систем искусственного интеллекта, то возникает необходимость сопоставимой оценки качества их математических моделей. Одним из вариантов решения этой задачи является тестирование различных системы на общей базе исходных данных, для чего очень удобно использовать общедоступную базу Kaggle. В данной курсовой работе приводится развернутый пример использования базы данных Kaggle для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». При этом наиболее достоверной в данном приложении оказались модели 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>