

**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ
ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ И.Т. ТРУБИЛИНА»**

Факультет прикладной информатики

Кафедра Компьютерных систем и технологий

КУРСОВАЯ РАБОТА

по дисциплине: Интеллектуальные системы и технологии
на тему: «АСК-анализ и прогнозирование цен на жилье по основным
характеристикам на основе данных с портала Kaggle»
Выполнил студент группы: ИТз2041 Тышченко Алексей Алексеевич

Допущен к защите _____

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

Защищен 11.04.2022
(дата)

Оценка отлично

Краснодар

2022

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

Факультет прикладной информатики

**РЕЦЕНЗИЯ
на курсовую работу**

Студента: Тыщенко Алексея Алексеевич

курса 2 заочной формы обучения группы ИТз2041

направления подготовки: 09.04.02 «Информационные системы и технологии»

Наименование темы: «АСК-анализ и прогнозирование цен на жилье по основным характеристикам на основе данных с портала Kaggle»

Рецензент: Луценко Евгений Вениаминович, доктор экономических наук, кандидат технических наук, профессор

Оценка качества выполнения курсовой работы

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

Достоинства работы:

Недостатки работы:

Итоговая оценка при защите: отлично

Рецензент Луценко Е.В..)
«11» 04 2022 г.

РЕФЕРАТ

Курсовая работа содержит: 35 страницы, 30 рисунка, 10 литературных источников.

Ключевые слова: СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, AIDOS-X, КОГНИТИВНЫЙ АНАЛИЗ, МОДЕЛИ, ШКАЛЫ, КЛАССЫ.

Целью работы является анализ и прогнозирование цен на жилье по основным характеристикам на основе данных на основе АСК-анализа данных портала Kaggle.

В данной курсовой работе необходимо 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.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1 СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	6
1.1 Описание решения	6
1.2 Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel	8
1.3 Ввод выборки в систему Aidos-X	9
1.4 Синтез и верификация статистических и интеллектуальных моделей	11
1.5 Виды моделей Aidos-X	13
1.6 Результаты верификации моделей	15
2 РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ..	19
2.1 Решение задач идентификации.....	19
2.2 SWOT и PEST матрицы и диаграммы	22
2.3 Исследование моделируемой предметной области путем исследования ее модели	24
2.3.1 Когнитивные диаграммы классов	24
2.3.2 Агломеративная когнитивная кластеризация классов.....	25
2.3.3 Когнитивные диаграммы значений факторов	26
2.3.4 Агломеративная когнитивная кластеризация значений факторов.....	27
2.3.5 Нелокальные нейроны и нелокальные нейронные сети	29
2.3.6 3D-интегральные когнитивные карты	30
2.3.7 Когнитивные функции	30
ЗАКЛЮЧЕНИЕ	33
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ.....	34

ВВЕДЕНИЕ

Одним из важных и перспективных направлений решения современных информационных технологий является создание систем искусственного интеллекта. Существует множество альтернатив систем искусственного интеллекта но возникает необходимость оценки качества математических моделей этих систем. В данной работе рассмотрены решения задачи исследования сведений о ценах на жилье по основным характеристикам на основе анализа данных с портала Kaggle.

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 "Housing Prices Dataset" из банка исходных данных по задачам искусственного интеллекта Kaggle.

Для решения задачи использованы стандартные возможности MS Office Word и Excel, а также системы искусственного интеллекта "Aidos-X++".

1 СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

1.1 Описание решения

В качестве метода исследования, решения проблемы и достижения цели было решено использовать новый метод искусственного интеллекта: Автоматизированный системно-когнитивный анализ или АСК-анализ.

Основной причиной выбора АСК-анализа является то, что он включает теорию и метод количественного выявления в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал и единицах измерения.

Очень важным является также то, что АСК-анализ имеет свой развитый и доступный программный инструментарий, в качестве которого в настоящее время выступает Универсальная когнитивная аналитическая система Aidos-X.

Преимущества данной системы:

- универсальность;
- доступность, то есть данная система находится в полном открытом бесплатном доступе, причем с актуальными исходными текстами;
- одна из первых отечественных систем искусственного интеллекта персонального уровня, т.е. она не требует от пользователя специальной подготовки в области технологий искусственного интеллекта (акт внедрения системы Aidos-X 1987 года);
- обеспечивает устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных взаимозависимых (нелинейных) данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения;

- содержит большое количество локальных и облачных учебных и научных приложений (в настоящее время их 31 и 270, соответственно);
- мультиязычная поддержка интерфейса (больше 50 языков);
- поддерживает online среду накопления знаний и широко используется во всем мире;
- наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решения этих задач в несколько тысяч раз;
- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение задач классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе);
- хорошо имитирует человеческий стиль мышления: дает результаты анализа, понятные экспертам на основе их опыта, интуиции и профессиональной компетенции.

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 Aidos-X.
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

С электронного ресурса kaggle.com возьмем набор данных "Housing Prices Dataset": <https://www.kaggle.com/datasets/yasserh/housing-prices-dataset>.

Csv файл содержит следующие данные:

- N – порядковый номер;
- area – площадь;
- bedrooms – количество спальных комнат;
- bathrooms – количество ванных комнат;
- stories – количество этажей дома;
- mainroad – наличие выхода на главную дорогу;
- guestroom – наличие гостевой комнаты;
- basement – наличие подвала;
- hotwaterheating – наличие водонагревателя;
- airconditioning – наличие кондиционера;
- parking – количество парковочных мест;
- prefarea – наличие придомовой территории;
- furnishingstatus – внутренняя отделка;
- price – цена.

Данный набор данных позволяет прогнозировать стоимость жилья в зависимости от необходимых критериев, заданных пользователем.

Для конвертации csv-файла в xls был использован онлайн конвертер: <https://document.online-convert.com/ru/convert/csv-to-excel>.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Nº	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus	price				
2	1	7420	4	2	3	yes	no	no	yes	2 yes	furnished	1330000						
3	2	8960	4	4	4	yes	no	no	yes	3 no	furnished	1225000						
4	3	9960	3	2	2	yes	no	yes	no	2 yes	semi-furnished	12215000						
5	4	7500	4	2	2	yes	no	yes	yes	3 yes	furnished	11410000						
6	5	7420	4	1	2	yes	yes	yes	no	2 no	furnished	10850000						
7	6	7500	3	3	1	yes	no	yes	yes	2 yes	semi-furnished	10150000						
8	7	8580	4	3	4	yes	no	no	yes	2 yes	semi-furnished	9870000						
9	8	16200	5	3	2	yes	no	no	no	0 no	unfurnished	10150000						
10	9	8100	4	1	2	yes	yes	yes	no	2 yes	furnished	9240000						
11	10	5750	3	2	4	yes	no	no	yes	1 yes	unfurnished	9800000						
12	11	13200	3	1	2	yes	no	yes	yes	2 yes	furnished	9800000						
13	12	6000	4	3	2	yes	yes	yes	yes	2 no	semi-furnished	9681000						
14	13	6550	4	2	2	yes	no	no	yes	1 yes	semi-furnished	9310000						
15	14	3500	4	2	2	yes	no	no	yes	2 no	furnished	9240000						
16	15	7800	3	2	2	yes	no	no	no	0 yes	semi-furnished	9240000						
17	16	6000	4	1	2	yes	no	yes	no	2 no	semi-furnished	9100000						
18	17	6600	4	2	2	yes	yes	yes	no	1 yes	unfurnished	9100000						
19	18	8500	3	2	4	yes	no	no	yes	2 no	furnished	8960000						
20	19	4600	3	2	2	yes	yes	no	yes	2 no	furnished	8890000						
21	20	6420	3	2	2	yes	no	no	yes	1 yes	semi-furnished	8855000						
22	21	4320	3	1	2	yes	no	yes	yes	2 no	semi-furnished	8750000						
23	22	7155	3	2	1	yes	yes	yes	no	2 no	unfurnished	8680000						
24	23	8050	3	1	1	yes	yes	yes	no	1 no	furnished	8645000						
25	24	4560	3	2	2	yes	yes	yes	no	1 no	furnished	8645000						
26	25	8800	3	2	2	yes	no	no	yes	2 no	furnished	8575000						
27	26	6540	4	2	2	yes	yes	yes	no	2 yes	furnished	8540000						
28	27	6000	3	2	4	yes	yes	yes	no	0 yes	semi-furnished	8463000						
29	28	8875	3	1	1	yes	no	no	no	1 no	semi-furnished	8400000						
30	29	7950	5	2	2	yes	no	yes	yes	2 no	unfurnished	8400000						
31	30	6500	4	2	2	yes	no	yes	no	1 yes	semi-furnished	8400000						

Рисунок 1 – Фрагмент обучающей выборки

1.3 Ввод выборки в систему Aidos-X

Далее обучающая выборка в формате *.xlsx, импортирована в систему Aidos-X: скопирована в папку Inp_data и переименована в Inp_data.

Далее был использован универсальный программный интерфейс импорта данных в систему Aidos-X (режим 2.3.2.2).

В форме задания размерности модели системы оставляем всё без изменений (Рисунок 2).

Задайте тип файла исходных данных: "Inp_data":

<input checked="" type="radio"/> XLS - MS Excel-2003	<input type="radio"/> XLSX- MS Excel-2007(2010)	<input type="radio"/> DBF - DBASE IV (DBF/NTX)	<input type="radio"/> CSV - CSV => DBF конвертер	<input type="checkbox"/> Стандарт XLS-файла	<input type="checkbox"/> Стандарт DBF-файла	<input type="checkbox"/> Стандарт CSV-файла
--	---	--	--	---	---	---

Задайте параметры:

<input checked="" type="radio"/> Нули и пробелы считать ОТСУТСТВИЕМ данных	<input type="radio"/> Нули и пробелы считать ЗНАЧЕНИЯМИ данных	<input checked="" type="checkbox"/> Создавать БД средник по классам "Inp_davr.dbf"?
--	--	---

Требования к файлу исходных данных

Задайте диапазон столбцов классификационных шкал:

Начальный столбец классификационных шкал:	<input type="text" value="14"/>
Конечный столбец классификационных шкал:	<input type="text" value="14"/>

Задайте диапазон столбцов описательных шкал:

Начальный столбец описательных шкал:	<input type="text" value="2"/>
Конечный столбец описательных шкал:	<input type="text" value="13"/>

Задайте режим:

<input checked="" type="radio"/> Формализации предметной области (на основе "Inp_data")	<input type="radio"/> Генерации распознаваемой выборки (на основе "Inp_rasp")
---	---

Задайте способ выбора размера интервалов:

<input checked="" type="radio"/> Равные интервалы с разным числом наблюдений	<input type="radio"/> Разные интервалы с равным числом наблюдений
--	---

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

<input checked="" type="radio"/> Не применять сценарный метод АСК-анализа	<input type="radio"/> Применить сценарный метод АСК-анализа
<input checked="" type="checkbox"/> Применить спец.интерпретацию текстовых полей классов	<input type="checkbox"/> Применить спец.интерпретацию текстовых полей признаков

Параметры интерпретации значений текстовых полей "Inp_data":

В качестве классов рассматриваются:

<input checked="" type="radio"/> Значения полей целиком	<input type="radio"/> Элементы значений полей - слова > символы:
<input type="radio"/> Элементы значений полей - символы	<input type="checkbox"/> Выделять уникальные значения и сортировать
<input type="checkbox"/> Не выделять уникальных значений и не сортировать	<input type="checkbox"/> Проводить лемматизацию

В качестве признаков рассматриваются:

<input type="radio"/> Значения полей целиком	<input checked="" type="radio"/> Элементы значений полей - слова > символы:
<input type="radio"/> Элементы значений полей - символы	<input type="checkbox"/> Не проводить лемматизацию

Какие наименования ГРАДАЦИЙ числовых шкал использовать:

<input checked="" type="radio"/> Только интервальные числовые значения	(например: "1/3-(59873.0000000, 178545.6666667)")
<input type="radio"/> Только наименования интервальных числовых значений	(например: "Минимальное")
<input type="radio"/> И интервальные числовые значения, и их наименования	(например: "Минимальное: 1/3-(59873.0000000, 178545.6666667)")

Ok **Cancel**

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

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ: (равные интервалы)

Количество градаций классификационных и описательных шкал в модели, т.е.: [3 классов x 45 признаков]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс.шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис.шкалу
Числовые	1	3	3,00	5	30	6,00
Текстовые	0	0	0,00	7	15	2,14
ВСЕГО:	1	3	3,00	12	45	3,75

Задайте количество числовых диапазонов (интервалов, градаций) в шкале:

В классификационных шкалах:	<input type="text" value="3"/>	В описательных шкалах:	<input type="text" value="6"/>
-----------------------------	--------------------------------	------------------------	--------------------------------

Пересчитать шкалы и градации **Выйти на создание модели**

Рисунок 2 - Задание размерности модели системы Aidos-X

Процесс импорта данных из внешней БД "Inp_data.xls" в систему Aidos-X представлен на рисунке 3.

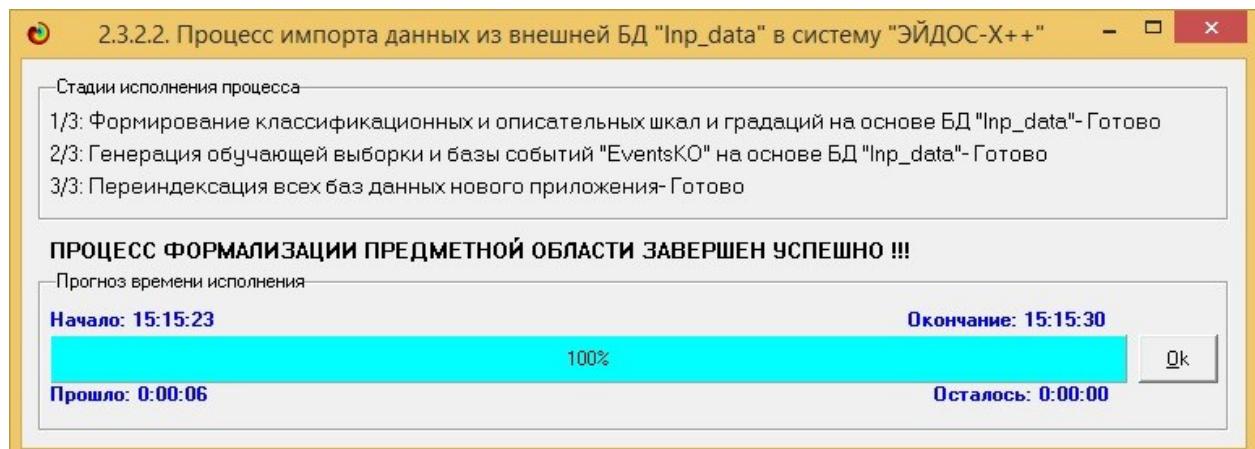


Рисунок 3 – Процесс импорта данных

Так как предварительно числовые шкалы были разбиты на интервалы, то пересчет шкал после ввода выборки производить не надо. После импорта автоматически формируются классификационные и описательные шкалы, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Тем самым этап формализации предметной области выполняется полностью автоматизировано.

1.4 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 4).

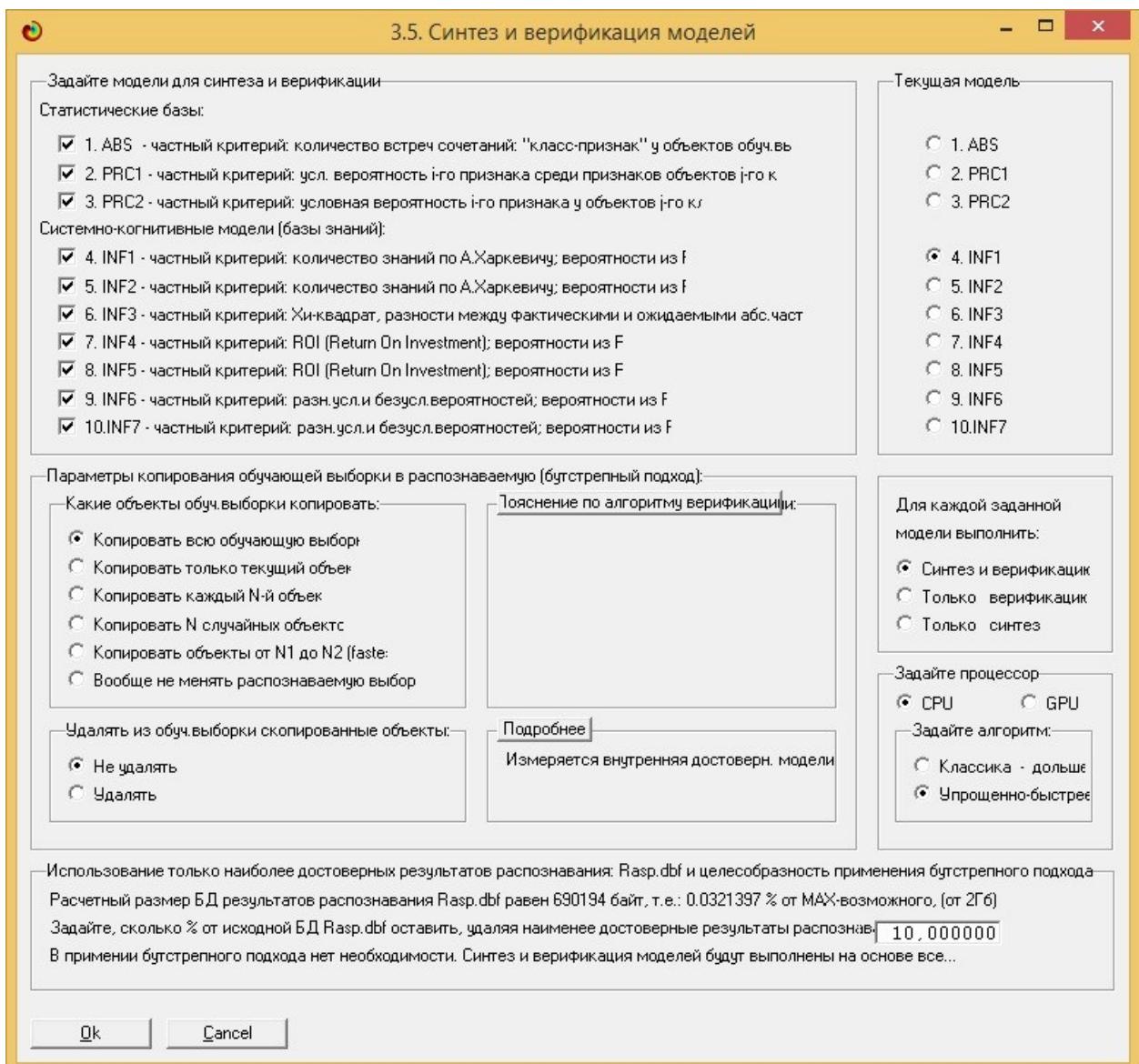


Рисунок 4 – Выбор моделей для синтеза и верификации

В данном режиме имеется много различных методов верификации моделей. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 5.

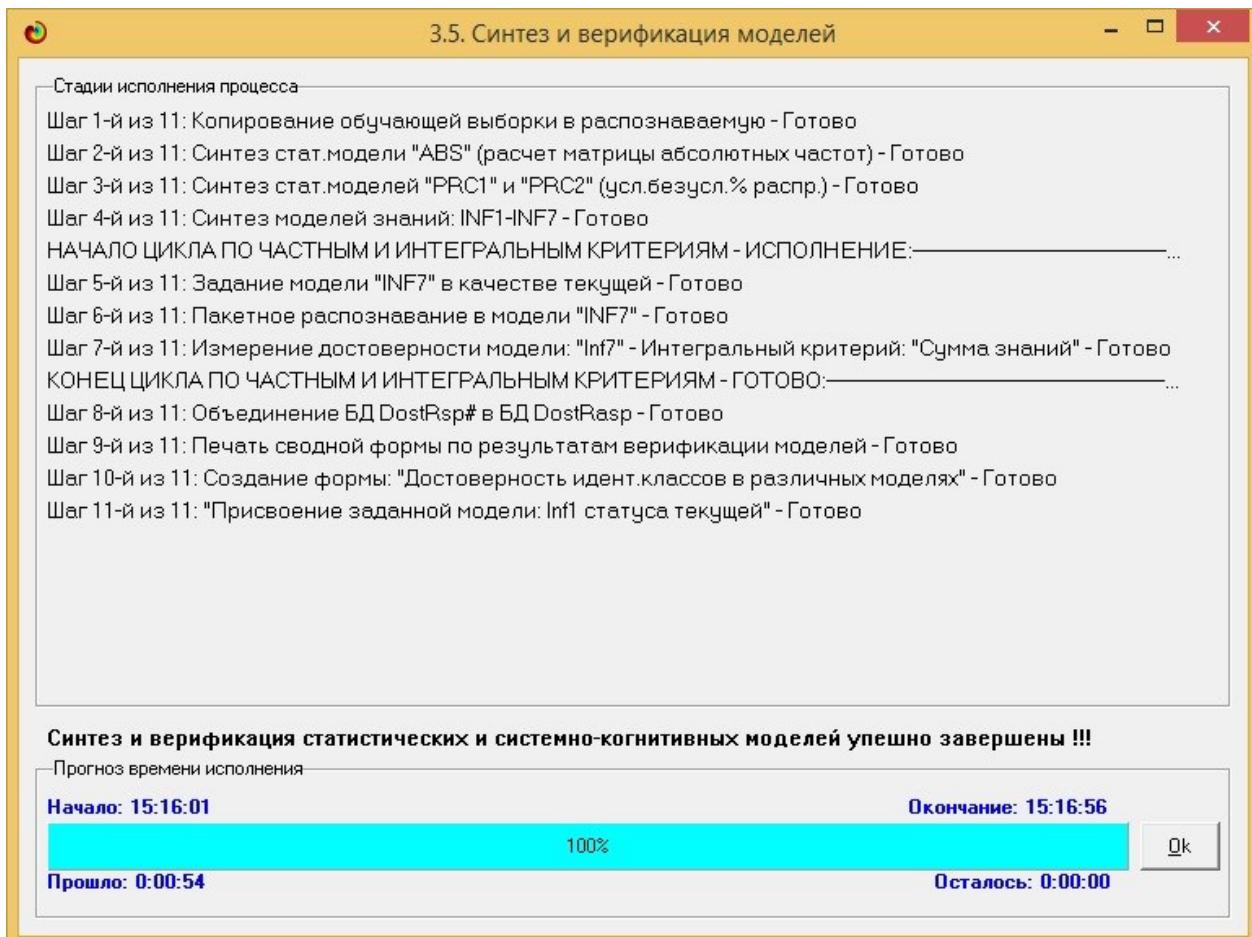


Рисунок 5 – Процесс исполнения режима 3.5

Интересно заметить (см. рисунок 5), что синтез и верификация всех 10 моделей на данной задаче заняли 54 секунды. Далее перейдем непосредственно к выбору наиболее достоверной модели.

1.5 Виды моделей Aidos-X

Фрагменты созданных статистических и системно-когнитивных моделей (СК-модели) приведены на рисунках 6 и 7.

Рассмотрим решение задачи идентификации на примере нескольких моделей, в которой рассчитано количество информации по А. Харкевичу, которое было получено по принадлежности идентифицируемого объекта к каждому из классов, если известно, что у этого объекта есть некоторый признак.

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.вы... - □ ×

д изнака	Наименование описательной шкалы и градации	1. PRICE 1/3 (1750000.000000,... 5600000.000000...)	2. PRICE 2/3 (5600000.000000,... 9450000.000000...)	3. PRICE 3/3 (9450000.000000,... 13300000.000000...)	Сумма	Среднее	Средн. квадр. откл.
1	AREA-1/6-(1650.0000000, 4075.0000000)	211	8		219	73.00	119.58
2	AREA-2/6-(4075.0000000, 6500.0000000)	136	65	2	203	67.67	67.04
3	AREA-3/6-(6500.0000000, 8925.0000000)	41	46	6	93	31.00	21.79
4	AREA-4/6-(8925.0000000, 11350.0000000)	12	6	2	20	6.67	5.03
5	AREA-5/6-(11350.0000000, 13775.0000000)	3	4	1	8	2.67	1.53
6	AREA-6/6-(13775.0000000, 16200.0000000)		1	1	2	0.67	0.58
7	BEDROOMS-1/6-{1.0000000, 1.8333333}	2			2	0.67	1.15
8	BEDROOMS-2/6-{1.8333333, 2.6666667}	131	5		136	45.33	74.23
9	BEDROOMS-3/6-{2.6666667, 3.5000000}	207	89	4	300	100.00	101.95
10	BEDROOMS-4/6-{3.5000000, 4.3333333}	56	32	7	95	31.67	24.50
11	BEDROOMS-5/6-{4.3333333, 5.1666667}	6	3	1	10	3.33	2.52
12	BEDROOMS-6/6-{5.1666667, 6.0000000}	1	1		2	0.67	0.58
13	BATHROOMS-1/6-{1.0000000, 1.5000000}	343	55	3	401	133.67	183.14
14	BATHROOMS-2/6-{1.5000000, 2.0000000}	56	73	4	133	44.33	35.95
15	BATHROOMS-3/6-{2.0000000, 2.5000000}						
16	BATHROOMS-4/6-{2.5000000, 3.0000000}	4	2	4	10	3.33	1.15
17	BATHROOMS-5/6-{3.0000000, 3.5000000}						
18	BATHROOMS-6/6-{3.5000000, 4.0000000}			1	1	0.33	0.58
19	STORIES-1/6-{1.0000000, 1.5000000}	192	34	1	227	75.67	102.09
20	STORIES-2/6-{1.5000000, 2.0000000}	190	41	7	238	79.33	97.34
21	STORIES-3/6-{2.0000000, 2.5000000}						
22	STORIES-4/6-{2.5000000, 3.0000000}	16	22	1	39	13.00	10.82
23	STORIES-5/6-{3.0000000, 3.5000000}						
24	STORIES-6/6-{3.5000000, 4.0000000}	5	33	3	41	13.67	16.77
25	PARKING-1/6-{1.0000000, 1.3333333}	78	47	1	126	42.00	38.74

Рисунок 6 – Матрица абсолютных частот (фрагмент)

5.5. Модель: "6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами" - □ ×

д изнака	Наименование описательной шкалы и градации	1. PRICE 1/3 (1750000.000000,... 5600000.000000...)	2. PRICE 2/3 (5600000.000000,... 9450000.000000...)	3. PRICE 3/3 (9450000.000000,... 13300000.000000...)	Сумма	Среднее	Средн. квадр. откл.
1	AREA-1/6-(1650.0000000, 4075.0000000)	51.098	-45.962	-5.136			48.734
2	AREA-2/6-(4075.0000000, 6500.0000000)	-12.219	14.980	-2.761			13.808
3	AREA-3/6-(6500.0000000, 8925.0000000)	-26.903	23.084	3.819			25.212
4	AREA-4/6-(8925.0000000, 11350.0000000)	-2.603	1.072	1.531			2.266
5	AREA-5/6-(11350.0000000, 13775.0000000)	-2.841	2.029	0.812			2.535
6	AREA-6/6-(13775.0000000, 16200.0000000)	-1.460	0.507	0.953			1.284
7	BEDROOMS-1/6-{1.0000000, 1.8333333}	0.540	-0.493	-0.047			0.518
8	BEDROOMS-2/6-{1.8333333, 2.6666667}	31.700	-28.511	-3.189			30.232
9	BEDROOMS-3/6-{2.6666667, 3.5000000}	-12.043	15.079	-3.036			13.813
10	BEDROOMS-4/6-{3.5000000, 4.3333333}	-13.364	8.592	4.772			11.730
11	BEDROOMS-5/6-{4.3333333, 5.1666667}	-1.301	0.536	0.765			1.133
12	BEDROOMS-6/6-{5.1666667, 6.0000000}	-0.460	0.507	-0.047			0.485
13	BATHROOMS-1/6-{1.0000000, 1.5000000}	50.212	-43.808	-6.404			47.336
14	BATHROOMS-2/6-{1.5000000, 2.0000000}	-41.109	40.228	0.881			40.676
15	BATHROOMS-3/6-{2.0000000, 2.5000000}	-3.301	-0.464	3.765			3.556
16	BATHROOMS-4/6-{2.5000000, 3.0000000}						
17	BATHROOMS-5/6-{3.0000000, 3.5000000}	-0.730	-0.246	0.977			0.880
18	BATHROOMS-6/6-{3.5000000, 4.0000000}	-26.257	-21.934	-4.324			24.385
19	STORIES-1/6-{1.0000000, 1.5000000}	16.226	-17.644	1.418			16.979
20	STORIES-2/6-{1.5000000, 2.0000000}	-12.476	12.390	0.085			12.433
21	STORIES-3/6-{2.0000000, 2.5000000}	-24.936	22.897	2.038			23.982
22	STORIES-4/6-{2.5000000, 3.0000000}	-13.998	15.953	-1.955			15.071

Рисунок 7 – Модель INF3 (фрагмент)

1.6 Результаты верификации моделей

Результаты верификации моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями представлены на рисунке 8.

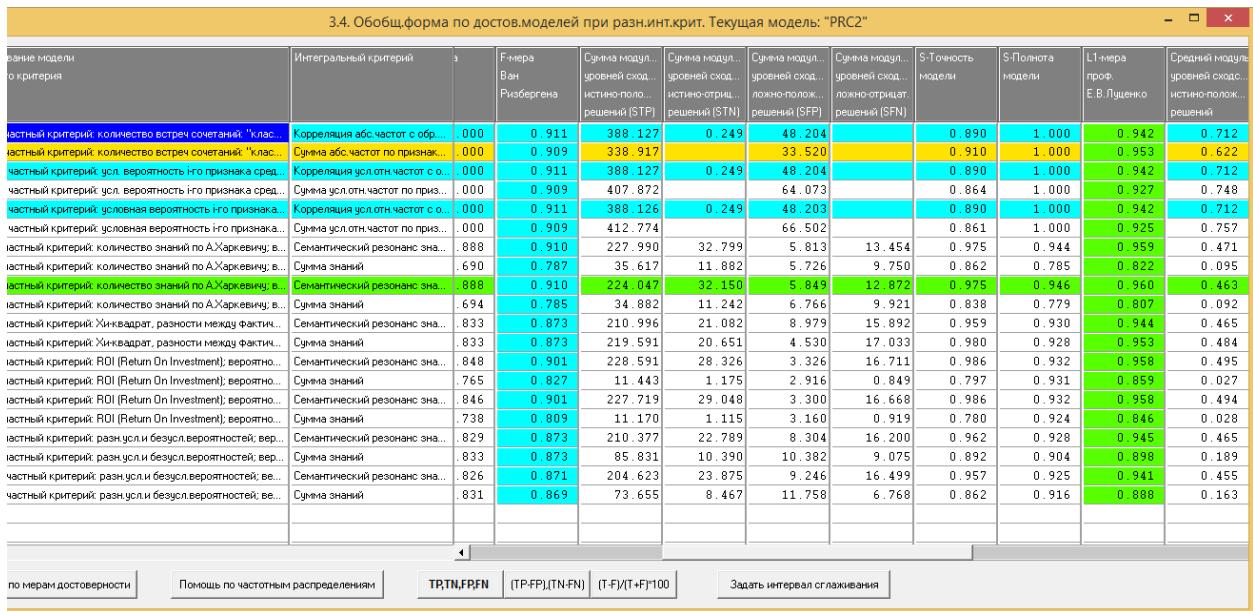


Рисунок 8 – Оценка достоверности моделей

Наиболее достоверной в данном приложении оказалась модель ABS при интегральном критерии «Семантический резонанс знаний». При этом точность модели (F- мера Ван Ризбергена) составляет 0,909 при максимуме 1,000, а точность модели (L1-мера профессора Луценко) - 0,953 при максимуме 1,000. L1-мера профессора Луценко является более достоверной, по сравнению с F-мерой Ван Ризбергена.

Таким образом, уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 100%. Для оценки достоверности моделей в АСК-анализе и системе Aidos-X используется F-мера Ван Ризбергена и L-мера, представляющая собой ее нечеткое мультиклассовое обобщение, предложенное профессором Е.В.Луценко (рисунок 9).

Также необходимо обратить внимание на то, что статистические модели, как правило, дают более низкую средневзвешенную достоверность

идентификации и не идентификации, чем модели знаний, и практически никогда – более высокую.

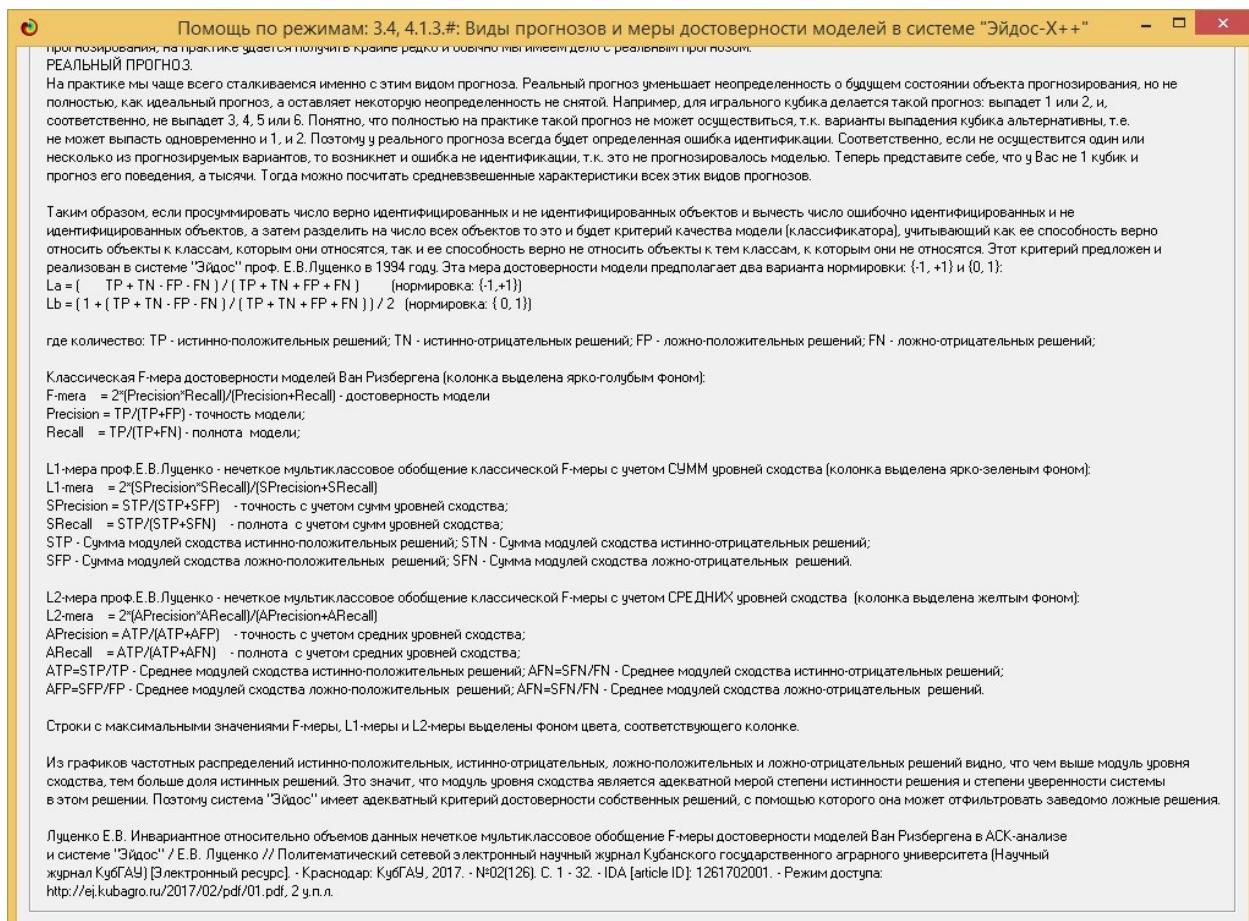


Рисунок 9 – Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

На рисунках ниже приведено частотное распределение числа истинных и ложных положительных и отрицательных решений в СК-модели по данным обучающей выборки.

Левое распределение включает истинно-отрицательные и ложно-положительные решения, а правое включает ложные отрицательные и истинно-положительные решения.

Из рисунка 10 видно, что:

- модель INF3 лучше определяет принадлежность объекта к классу, чем непринадлежность;
- истинных решений гораздо больше, чем ложных;
- при решении о непринадлежности больше истинных решений.

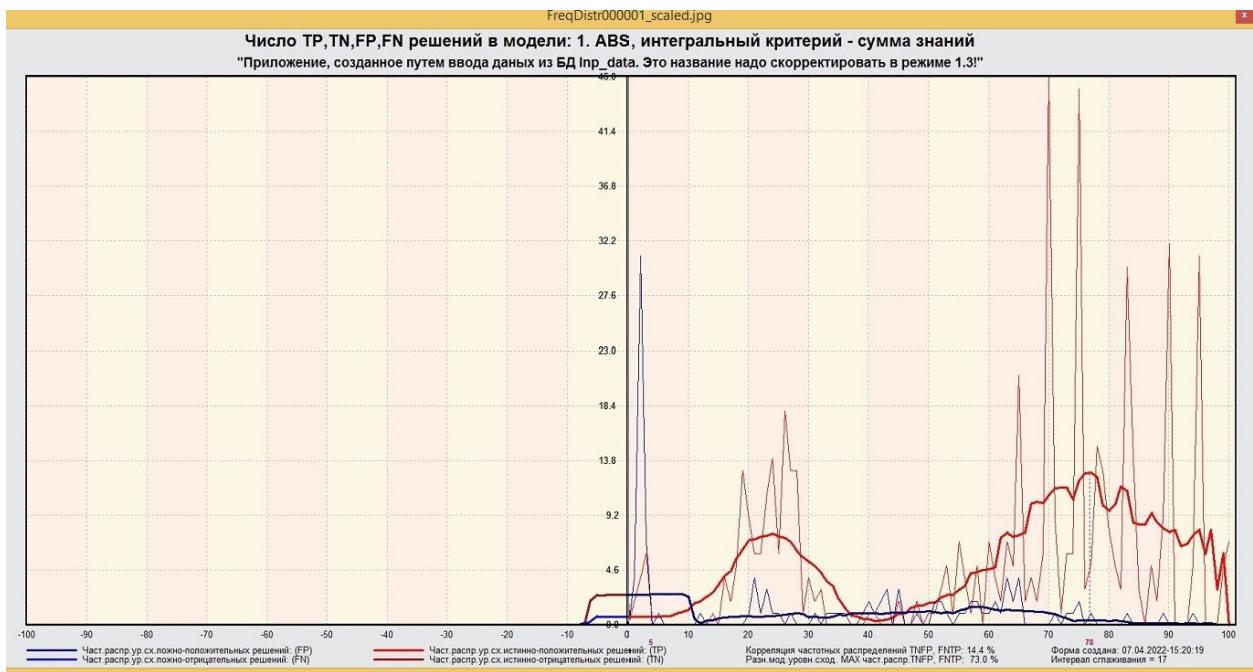


Рисунок 10 – Частотные распределения числа верно и ошибочно идентифицированных и не идентифицированных состояний объекта моделирования в зависимости от уровня сходства в модели ABS
На рисунке 11 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и не идентифицированных ситуаций в наиболее достоверной модели INF3.

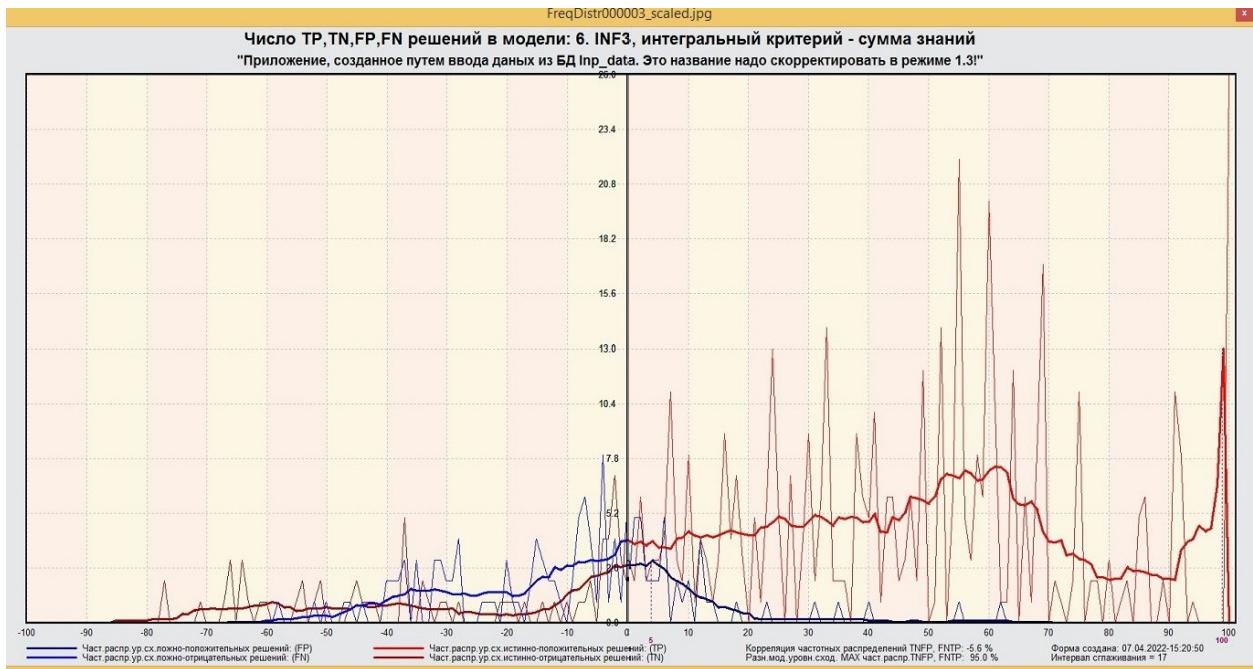


Рисунок 11 – Частотные распределения модели INF3

Из рисунка 11 видно, что по сравнению с моделью INF3 в модели ABS больше истинных решений.

Из рисунка 12 видно, что модель PRC2 абсолютно четко показывает, что доля истинных решений гораздо выше, чем ложных начиная с 0%, т.е. подавляющее число решений является истинным.

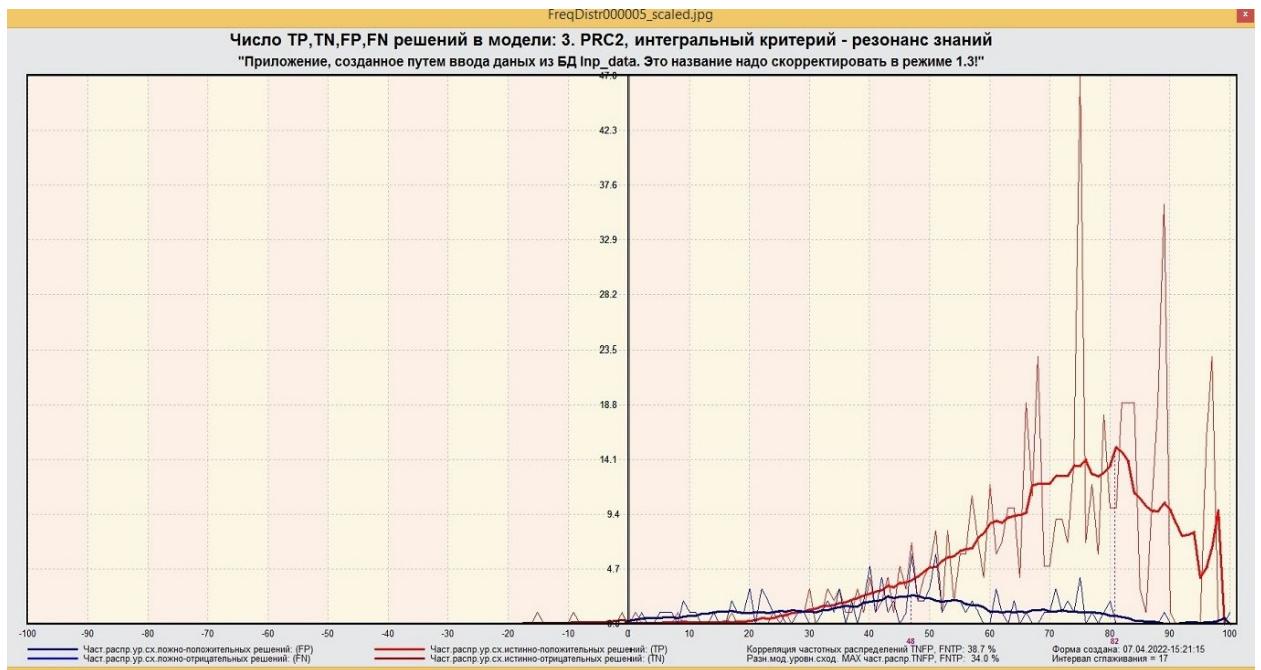


Рисунок 12 – Частотные распределения числа верно и ошибочно идентифицированных и не идентифицированных состояний объекта моделирования в зависимости от уровня сходства в модели PRC2

2 РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1 Решение задач идентификации

Решим задачу системной идентификации, т.е. определим к каким респондентам к каким категориям относятся на основе обучающей выборки в наиболее достоверной СК-модели PRC2 (режим 5.6) (рисунок 13, 14).

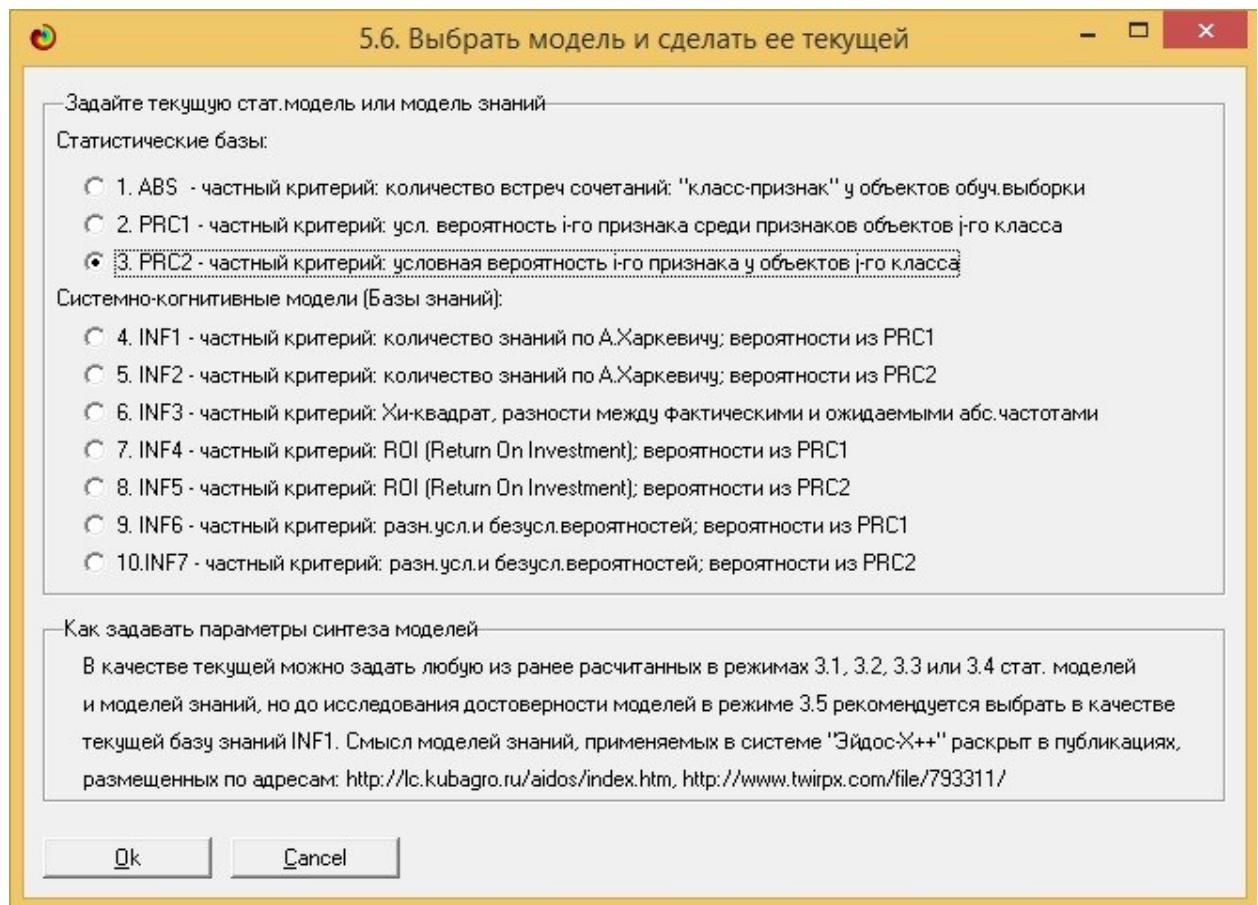


Рисунок 13 – Экранная форма режима задания модели в качестве текущей

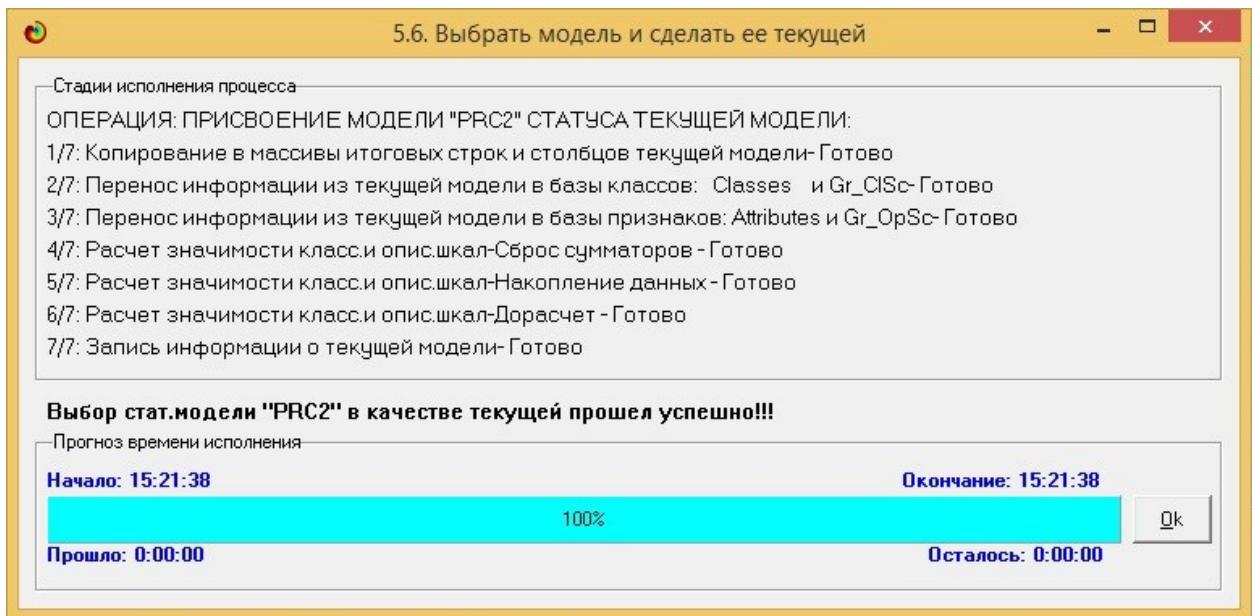


Рисунок 14 – Экранная форма режима задания модели в качестве текущей

Проведем пакетное распознавание в режиме 4.1.2. (рисунок 15).

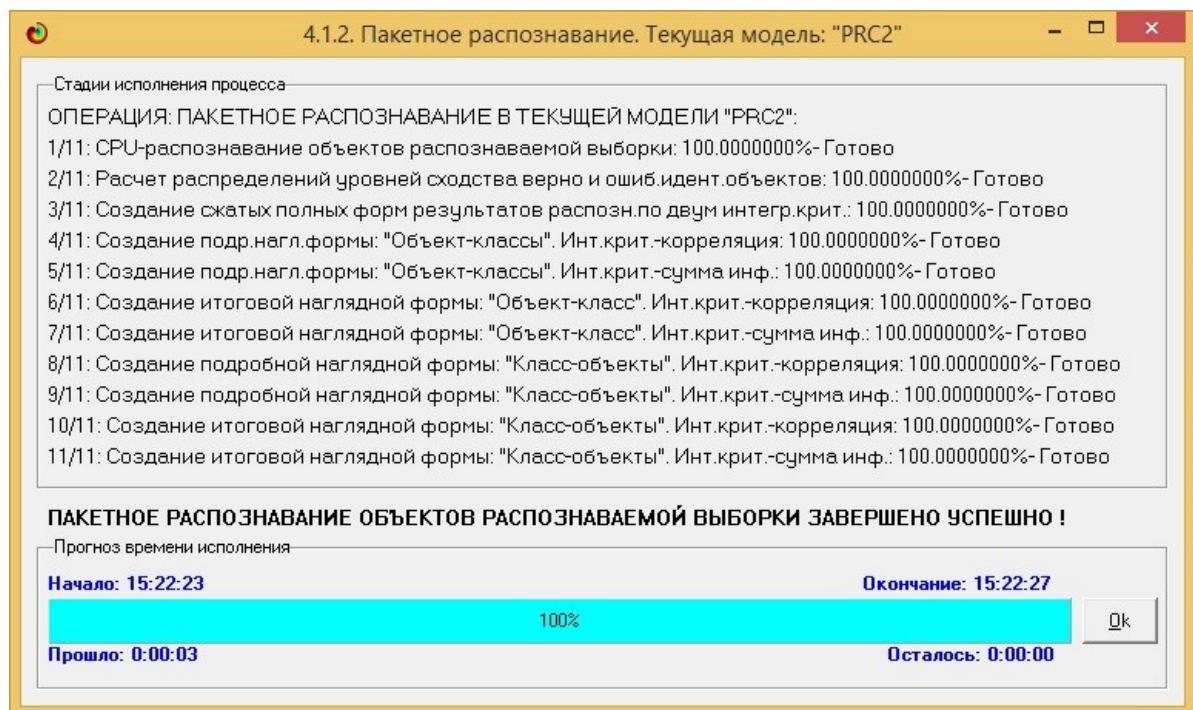


Рисунок 15 – Экранная форма режима пакетного распознавания в текущей модели

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

Режим 4.1.3 системы Aidos-X обеспечивает отображение результатов идентификации и прогнозирования в различных формах:

- Подробно наглядно: «Объект – классы».
- Подробно наглядно: «Класс – объекты».
- Итоги наглядно: «Объект – классы».
- Итоги наглядно: «Класс – объекты».
- Подробно сжато: «Объект – классы».

Обобщенная форма по достоверности моделей при разных интегральных критериях.

Обобщенный статистический анализ результатов идентификации по моделям и интегральным критериям.

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

На рисунке 16 и 17 приведен пример прогноза высокой и низкой достоверности частоты объектов и классов в модели PRC2 на основе наблюдения предыстории их развития. Визуализация результатов распознавания представлена в соотношении: «Объект-классы» и «Класс-объекты».

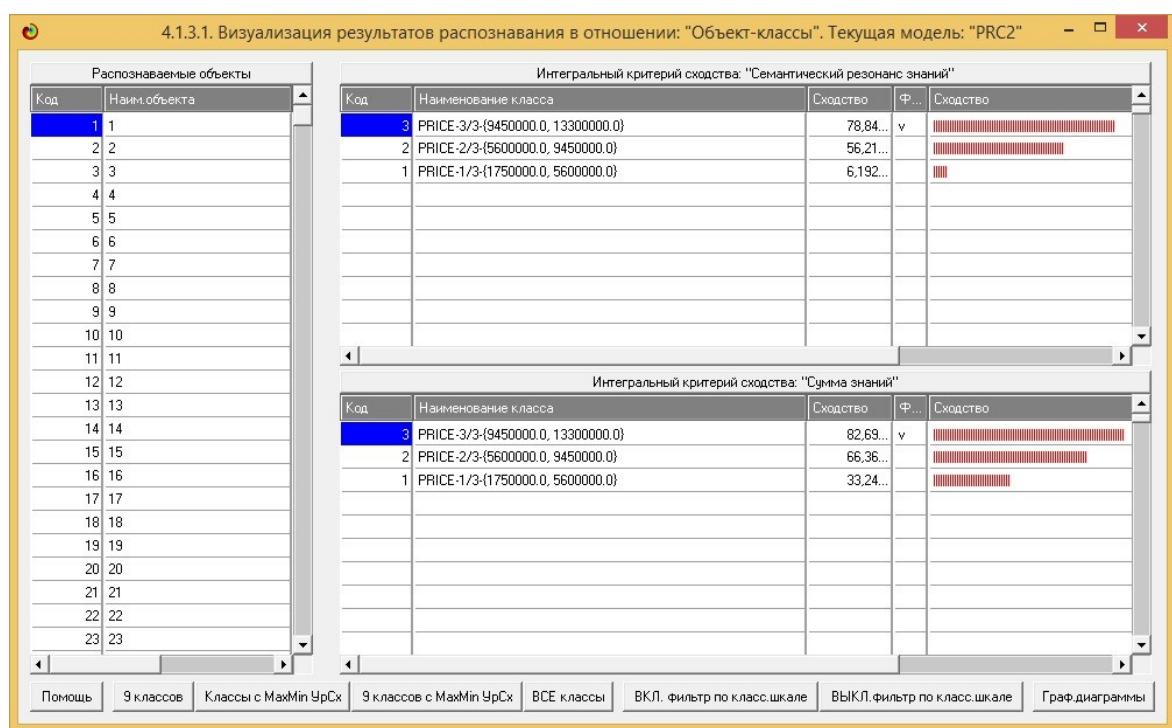


Рисунок 16. Пример идентификации объектов в модели PRC2

4.1.3.2. Визуализация результатов распознавания в отношении: "Класс-объекты". Текущая модель: "PRC2"

Классы	
Код	Наим. класса
1	PRICE-1/3(1750000.000000, 5600000.0000...
2	PRICE-2/3-{5600000.000000, 9450000.0000...
3	PRICE-3/3(9450000.000000, 13300000.000...

Интегральный критерий сходства: "Семантический резонанс знаний"				
Код	Наименование объекта	Сходство	Ф...	Сходство
247	247	96,87...	v	
457	457	96,87...	v	
504	504	96,87...	v	
542	542	96,87...	v	
183	183	96,67...	v	
208	208	96,67...	v	
262	262	96,67...	v	
264	264	96,67...	v	
266	266	96,67...	v	
296	296	96,67...	v	

Интегральный критерий сходства: "Сумма знаний"				
Код	Наименование объекта	Сходство	Ф...	Сходство
241	241	100,00...	v	
252	252	100,00...	v	
259	259	100,00...	v	
306	306	100,00...	v	
309	309	100,00...	v	
481	481	100,00...	v	
490	490	100,00...	v	
242	242	98,83...	v	
273	273	98,83...	v	
229	229	98,70...	v	

Помощь | Поиск объекта | В начало БД | В конец БД | Предыдущая | Следующая | 9 записей | Все записи | Печать XLS | Печать TXT | Печать ALL

Рисунок 17. Пример идентификации объектов в модели PRC2

2.2 SWOT и PEST матрицы и диаграммы

При принятии решений определяется сила и направление влияния значений факторов на принадлежность состояний объекта моделирования к тем или иным классам, соответствующим различным будущим состояниям. По сути, это решение задачи SWOT-анализа.

Применительно к задаче, решаемой в данной работе, SWOT-анализ показывает, как различные критерии влияют на статус кредита (рисунок 18).

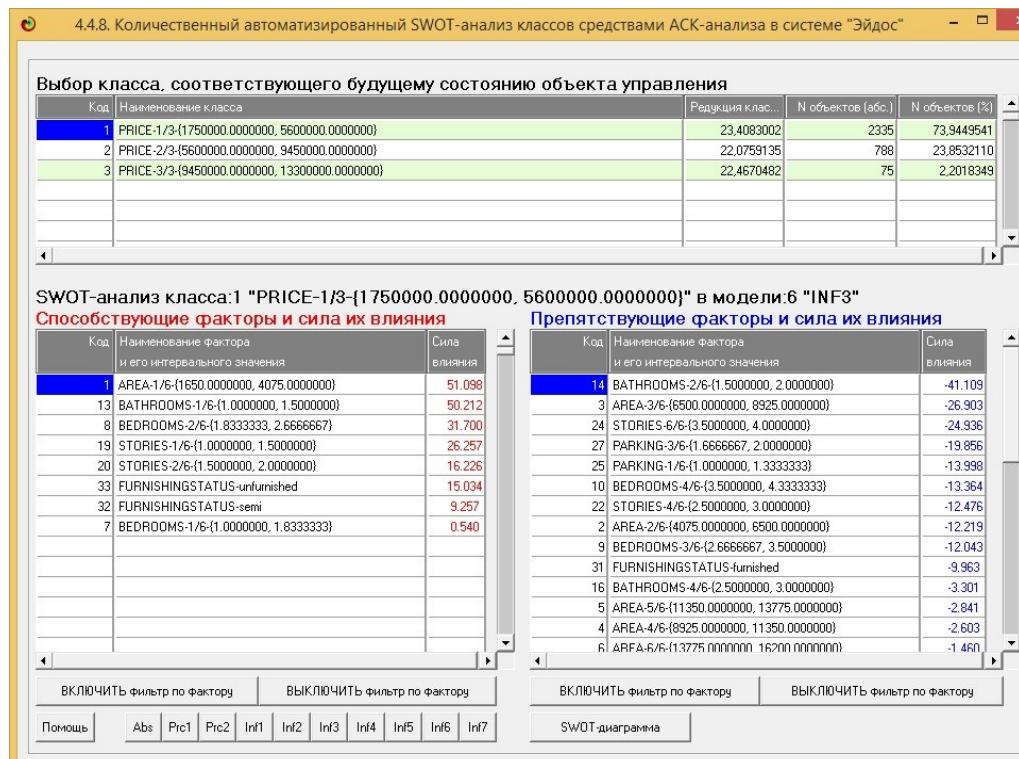


Рисунок 18 – Автоматизированное построение SWOT-матрицы

На рисунке 19 приведен пример табличной выходной формы количественного автоматизированного SWOT-анализа и PEST-анализа средствами системы Aidos.

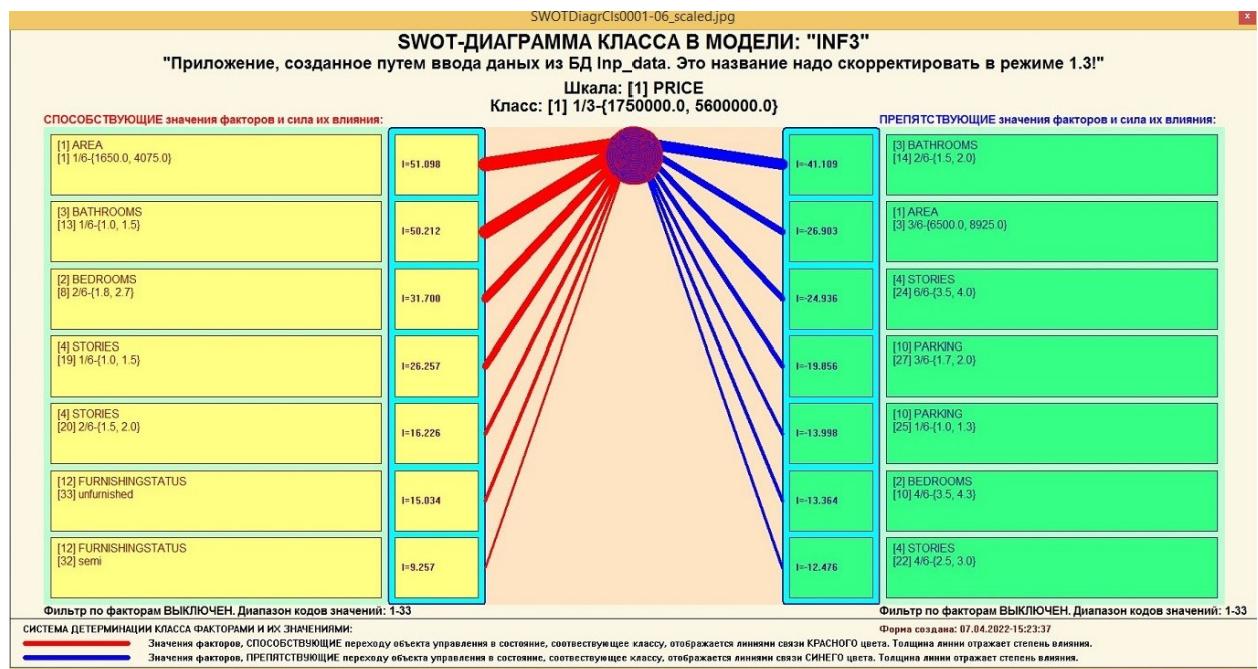


Рисунок 19 - Пример SWOT-матрицы в модели INF3

В левой стороне представлены факторы характерные для класса, а справа – нехарактерные.

2.3 Исследование моделируемой предметной области путем исследования ее модели

Если модель предметной области достоверна, то исследование модели можно считать исследованием самого моделируемого объекта, т.е. результаты исследования модели корректно относить к самому объекту моделирования, «переносить на него».

В системе «Aidos-X» есть довольно много возможностей для такого исследования, но в данной работе из-за ограничений на ее объем мы рассмотрим лишь результаты кластерно-конструктивного анализа классов и признаков (когнитивные диаграммы и дендрограммы), а также нелокальные нейроны, нелокальные нейронные сети, 3d-интегральные когнитивные карты и когнитивные функции.

2.3.1 Когнитивные диаграммы классов

На когнитивной диаграмме, приведенной на рисунке 20, показаны количественные оценки кредитной истории. Однако так как он может сделать только два выбора, то диаграмма максимально проста.

Данную диаграмму мы получаем в режимах 4.2.2.1 и 4.2.2.2 (рисунок 20). Важно, что эти результаты получены с применением системно-когнитивной модели, созданной непосредственно на основе эмпирических данных, а не как традиционно делается на основе экспертных оценок неформализуемым путем на основе опыта, интуиции и профессиональной компетенции.

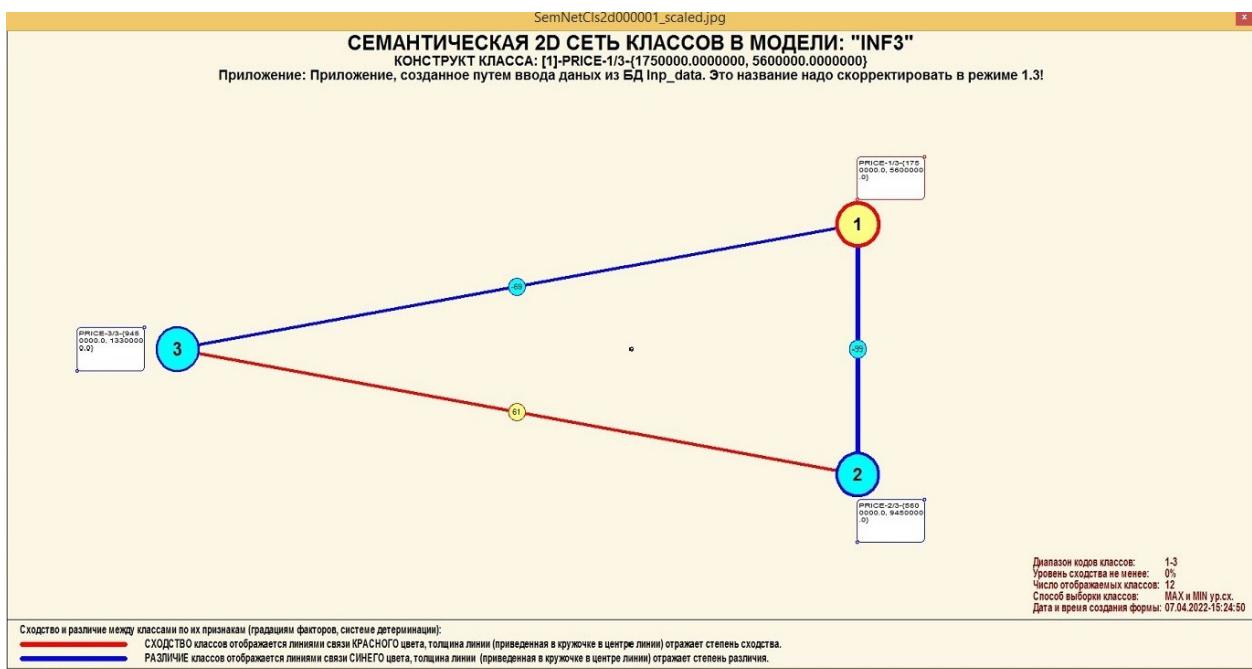
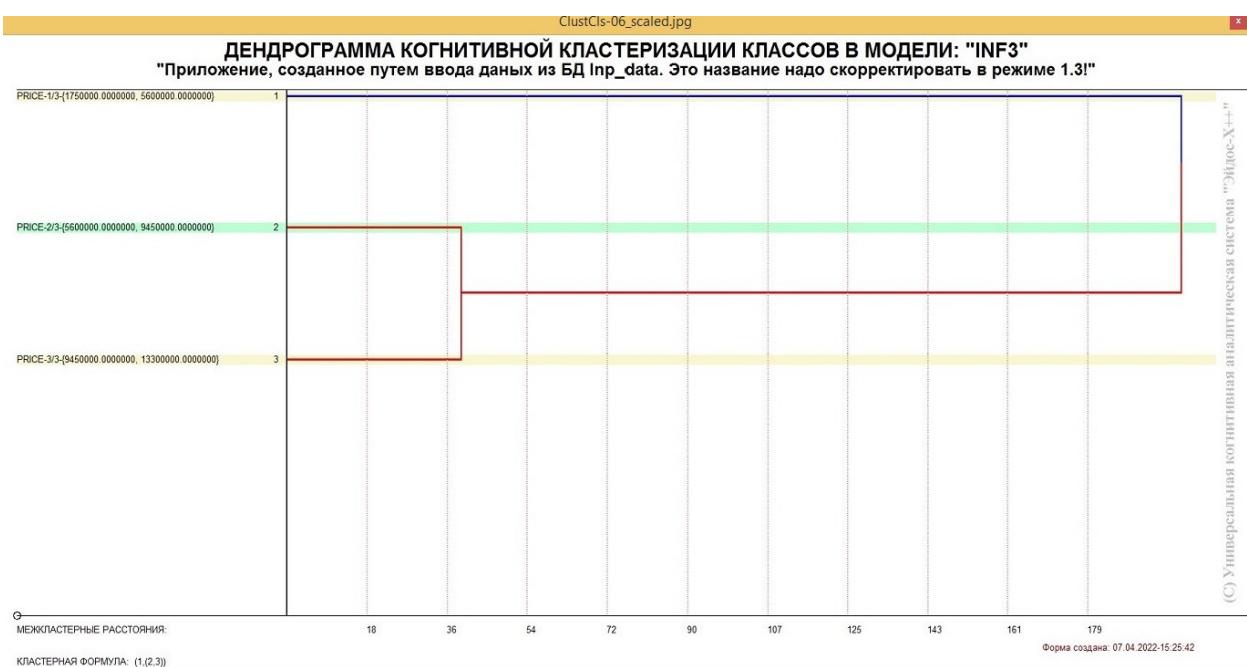


Рисунок 20 – Когнитивная диаграмма классов

2.3.2 Агломеративная когнитивная кластеризация классов

Информация о сходстве/различии классов, содержащаяся в матрице сходства, может быть визуализирована не только в форме, когнитивных диаграмм, но и в форме агломеративных дендрограмм, полученных в результате когнитивной кластеризации (рисунок 21):



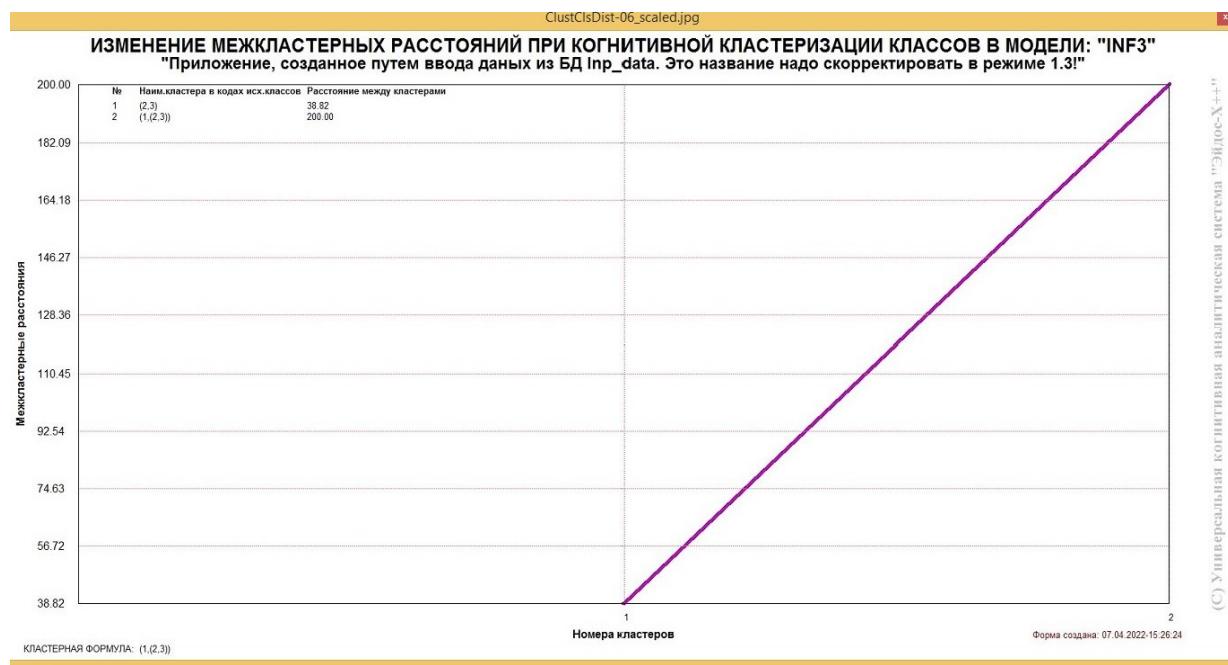


Рисунок 21. Дендрограмма когнитивной агломеративной кластеризации классов и график изменения межкластерных расстояний

Из рисунка 21 мы видим, что некоторые характеристики сходны по системе детерминирующих их факторов, и, следовательно, могут наблюдаться одновременно, а другие по этой системе обуславливающих факторов сильно отличаются, и, следовательно, являются взаимоисключающими, т.е. являются альтернативными. Из дендрограммы когнитивной агломеративной кластеризации классов, приведенной на рисунке 21, мы видим, что все приведенные характеристики образуют два противоположных по системе детерминации кластера, являющихся полюсами конструкта: в верхнем кластере собраны явления, связанные с высокой стоимостью, а в нижнем – с низкой.

2.3.3 Когнитивные диаграммы значений факторов

Такие диаграммы отражают зависимость параметров. Эти диаграммы мы получаем в режимах 4.3.2.1 и 4.3.2.2 (рисунок 22).

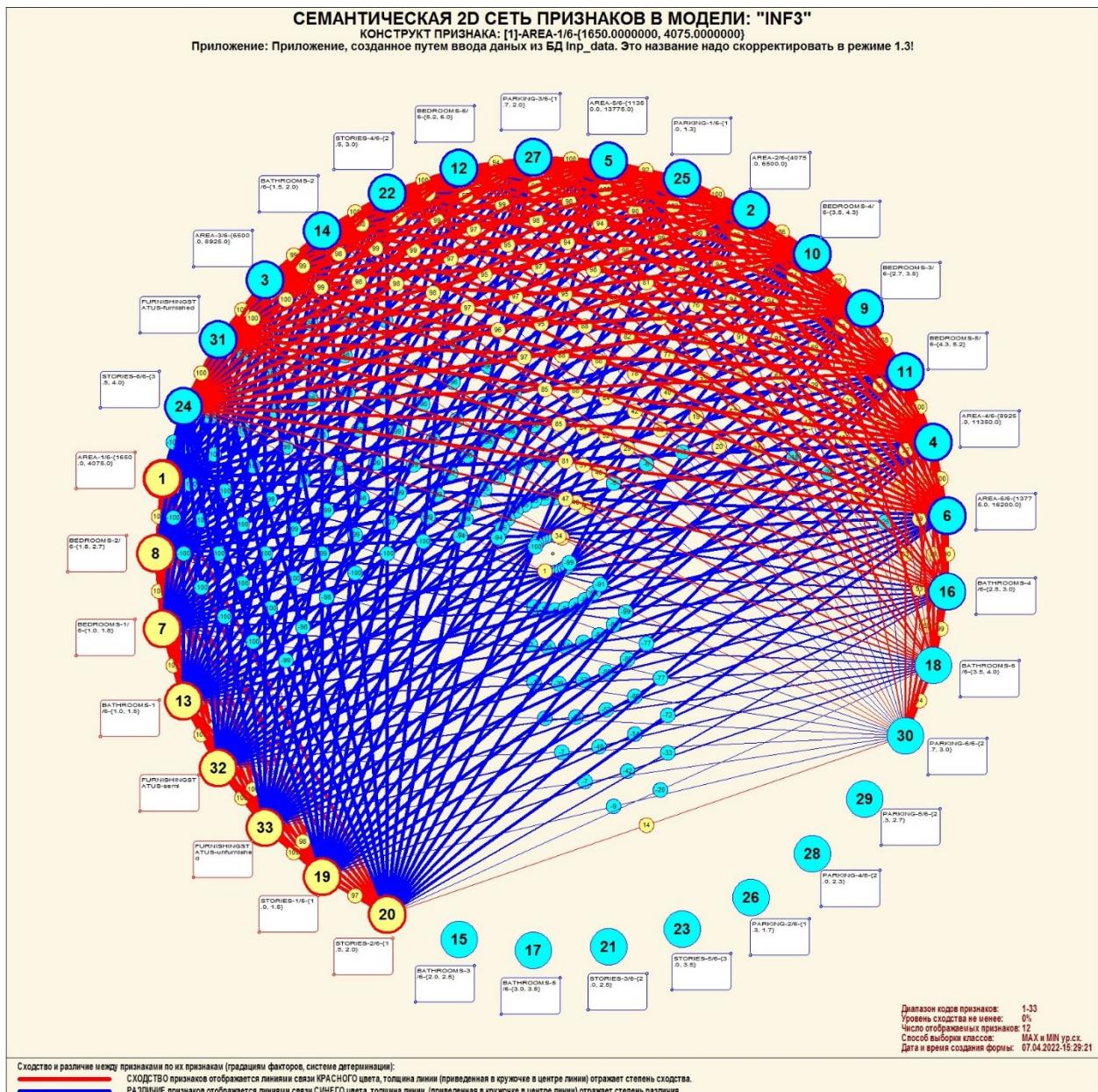


Рисунок 22 – Критерии, от которых зависит стоимость жилья

2.3.4 Агломеративная когнитивная кластеризация значений факторов

На рисунке 23 приведена агломеративная дендрограмма когнитивной кластеризации значений факторов и график изменения межкластерных расстояний, полученные на основе той же матрицы сходства признаков по их смыслу, что и в когнитивных диаграммах, пример приведен на рисунке 22.

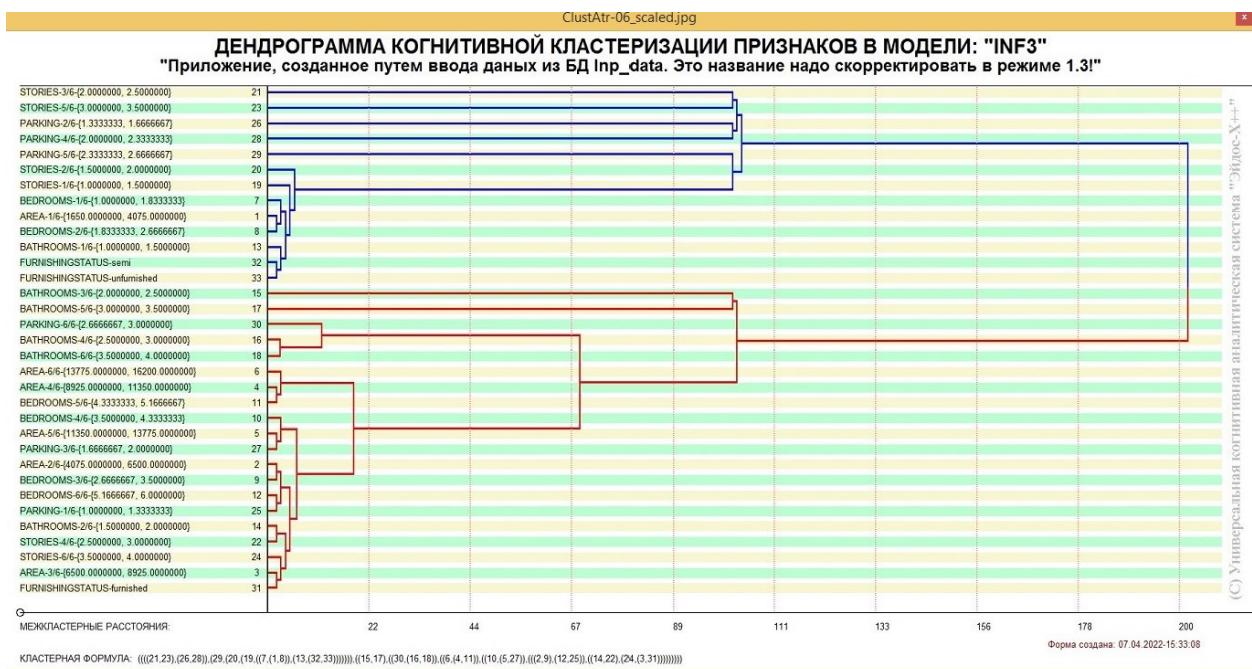


Рисунок 23 – Дендрограмма когнитивной кластеризации признаков

Из дендрограммы на рисунке 23 мы видим, что все значения факторов образуют 2 четко выраженных кластера, объединенных в полюса конструкта (показаны синими и красным цветами). Хорошо видна группировка признаков по их смыслу.

На рисунке 24 приведен график межклластерных расстояний значений признаков.



Рисунок 24 – График изменения межклластерных расстояний при когнитивной кластеризации значений факторов

2.3.5 Нелокальные нейроны и нелокальные нейронные сети

На рисунке 25 приведён пример нелокального нейрона, а на рисунке 26 – фрагмент одного слоя нелокальной нейронной сети, отражающий 93,93% наиболее значимых синаптических связей.

В приведенном фрагменте слоя нейронной сети нейроны соответствуют статусу кредита (оплачен или нет), а рецепторы – параметрам, влияющим на его выбор. Нейроны расположены слева направо в порядке убывания силы детерминации, т.е. слева находятся результаты, наиболее жестко обусловленные обуславливающими их значениями факторами, а справа – менее жестко обусловленные.

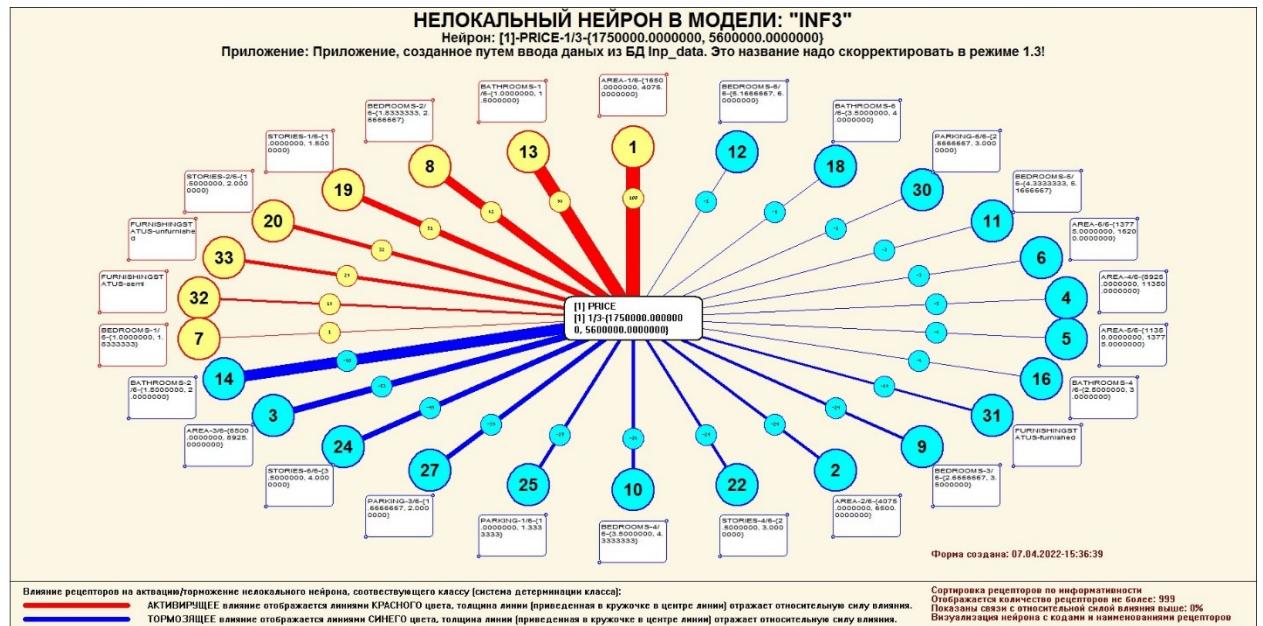


Рисунок 25 – Пример нелокального нейрона

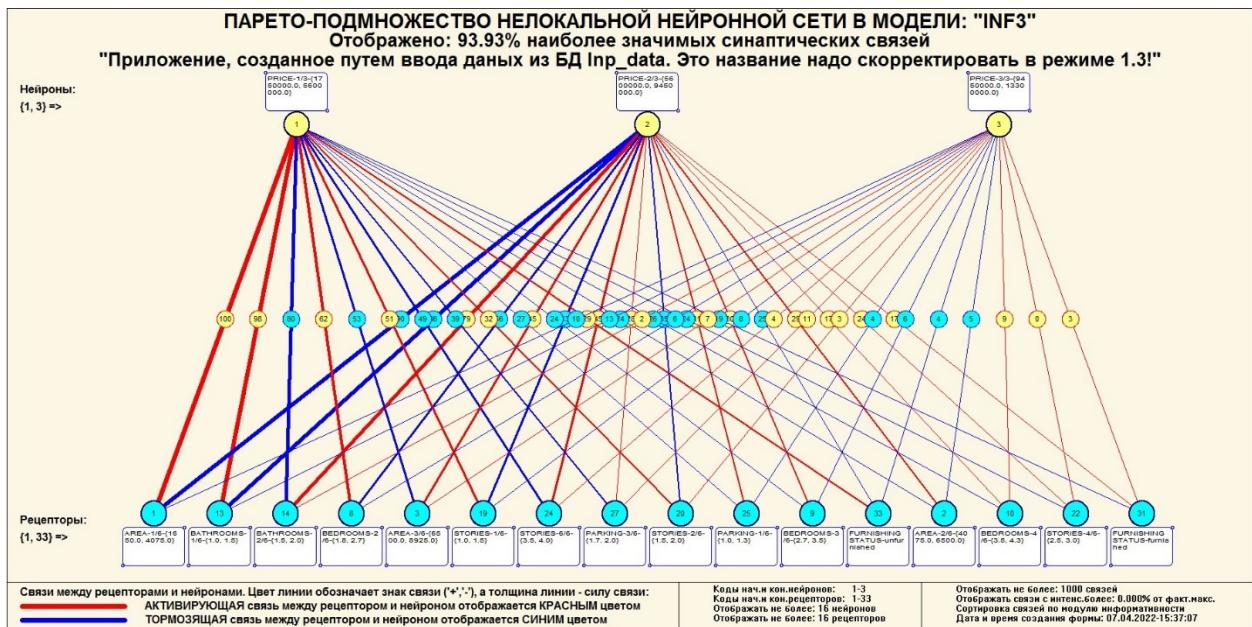


Рисунок 26 – Один слой нелокальной нейронной сети

2.3.6 3D-интегральные когнитивные карты

На рисунке 27 приведен фрагмент 3D-интегральной когнитивной карты, отражающий фрагмент СК-модели Inf3.

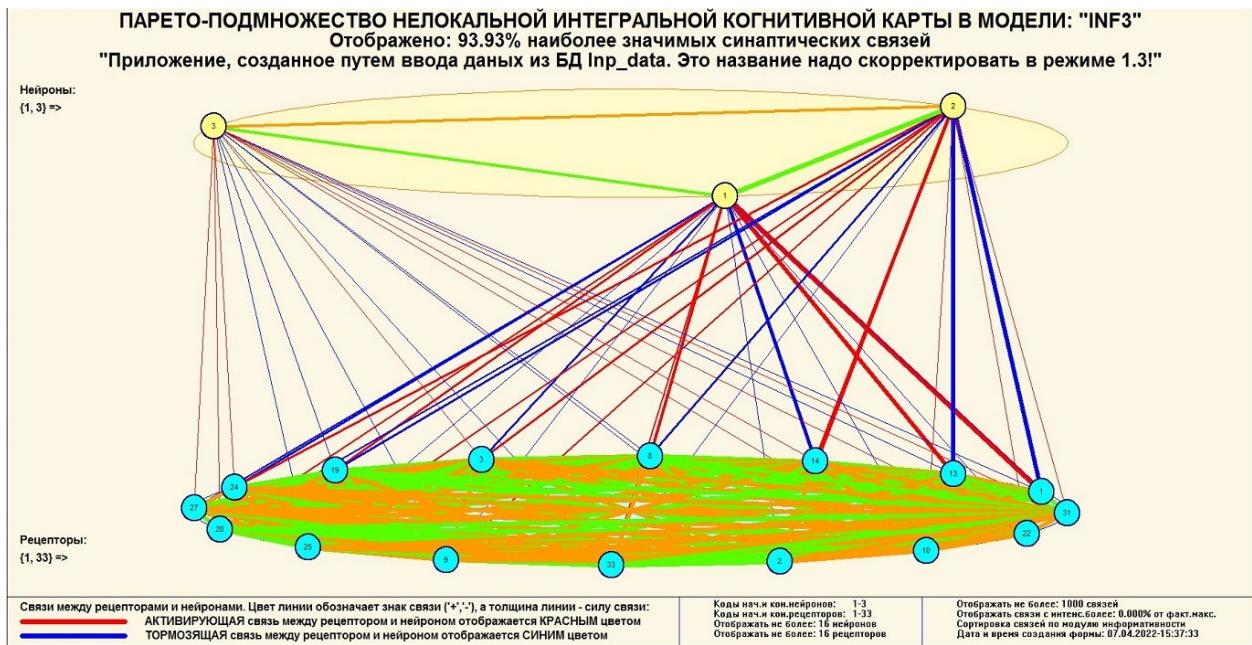


Рисунок 27 – 3D-интегральная когнитивная карта в СК-модели INF3

2.3.7 Когнитивные функции

Такие диаграммы отражают то, как влияют градации различных факторов на прогноз. Эти диаграммы мы получаем в режиме 4.5

«Визуализация когнитивных функций» и некоторые из них представлены на рисунках 28, 29, 30.

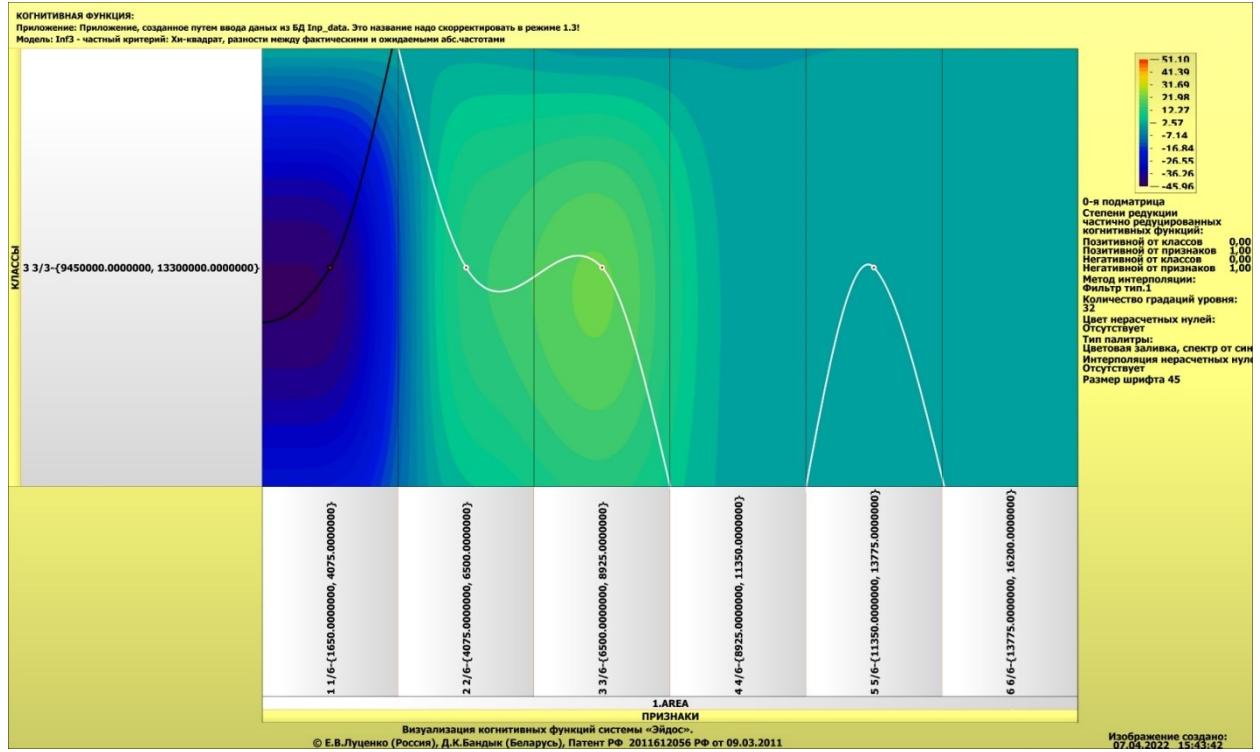


Рисунок 28 – Пример визуализации когнитивной функции модели INF3

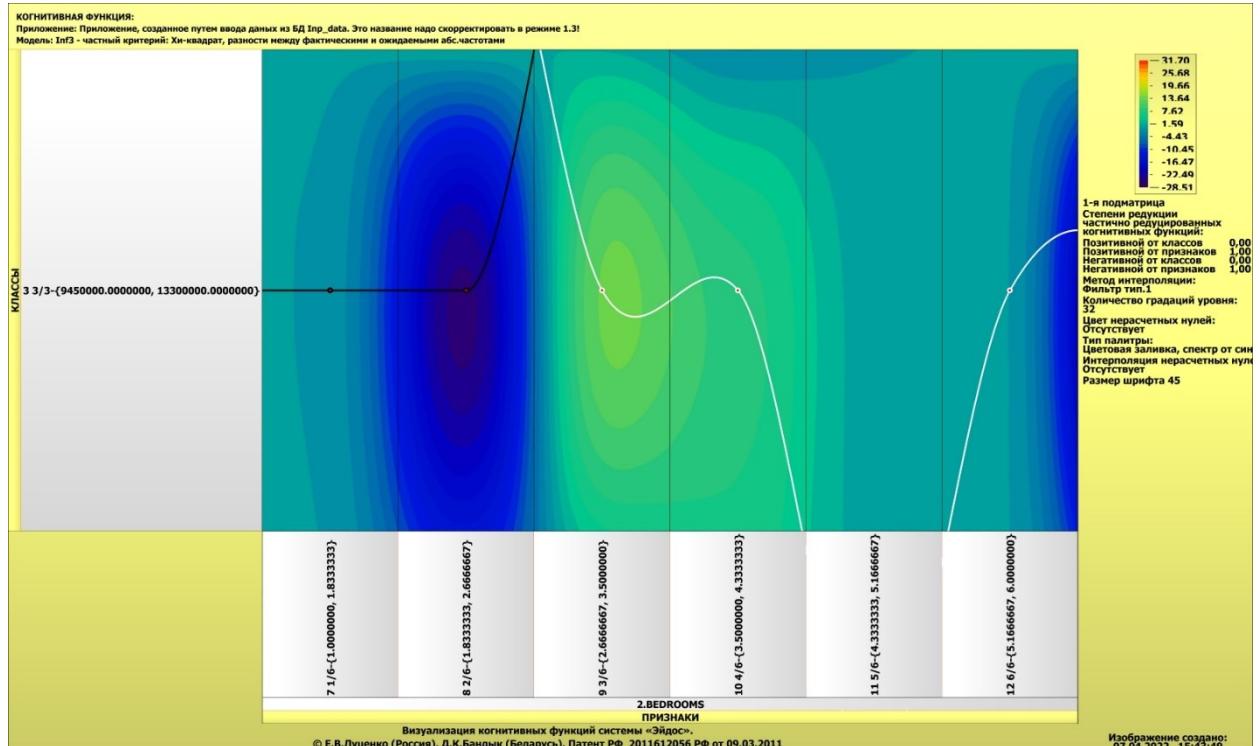


Рисунок 29 – Пример визуализации когнитивной функции модели INF3

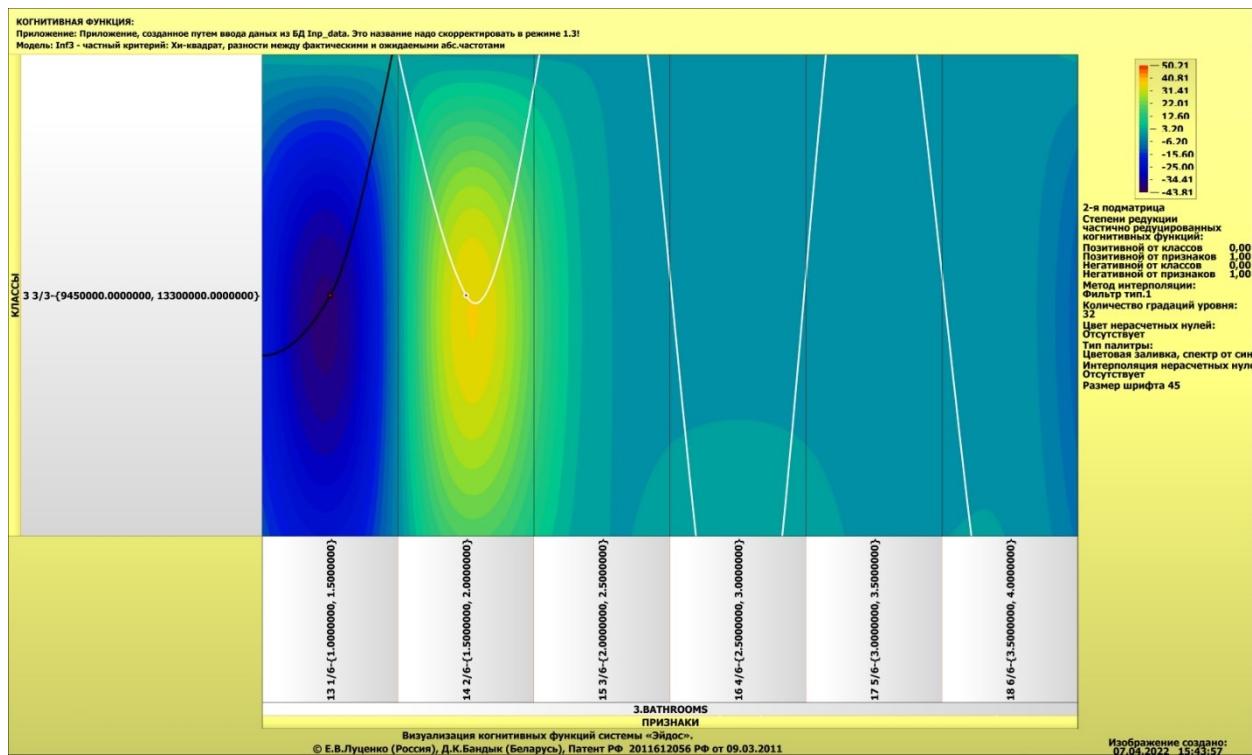


Рисунок 30 – Пример визуализации когнитивной функции модели INF3

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 систем на общей базе исходных данных, для чего очень удобно использовать общедоступную электронную базу Kaggle.

Как показывает анализ результатов численного эксперимента предложенное и реализованное в системе «Aidos-X» решение поставленных задач является вполне эффективным, что позволяет обоснованно утверждать, что цель работы достигнута, поставленная проблема решена.

В результате проделанной работы, с помощью системы «Aidos-X» были созданы статистические и системно-когнитивные модели, в которых непосредственно на основе эмпирических данных сформированы обобщенные образы классов показывающие предполагаемую стоимость жилья, и на основе этого, решены задачи идентификации, классификации и исследования моделируемой предметной области путем исследования ее модели.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Луценко Е.В. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос» / Е.В. Луценко, Е.К. Печурина, А.Э. Сергеев // Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №06(160). С. 95 – 114. – IDA [article ID]: 1602006009. – Режим доступа: <http://ej.kubagro.ru/2020/06/pdf/09.pdf>, 1,25 у.п.л.
2. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 у.п.л.
3. Луценко Е.В., Подсистема агломеративной когнитивной кластеризации классов системы «Эйдос» ("Эйдос-кластер"). Пат. № 2012610135 РФ. Заяв. № 2011617962 РФ 26.10.2011. Опубл. От 10.01.2012. – Режим доступа: <http://lc.kubagro.ru/aidos/2012610135.jpg>, 3,125 у.п.л.
4. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 у.п.л.
5. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе «Эйдос» / Е.В. Луценко // Политехнический 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.п.л.
6. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политехнический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал

КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.

7. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>, 0,812 у.п.л.

8. Луценко Е.В. Метод визуализации когнитивных функций – новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №03(067). С. 240 – 282. – Шифр Информрегистра: 0421100012\0077, IDA [article ID]: 0671103018. – Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>, 2,688 у.п.л.

9. Луценко Е.В. Сценарный АСК-анализ как метод разработки на основе эмпирических данных базисных функций и весовых коэффициентов для разложения в ряд функции состояния объекта или ситуации по теореме А.Н.Колмогорова (1957) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №07(161). С. 76 – 120. – IDA [article ID]: 1612007009. – Режим доступа: <http://ej.kubagro.ru/2020/07/pdf/09.pdf>, 2,812 у.п.л.

10. Луценко Е.В. Детальный численный пример сценарного Автоматизированного системно-когнитивного анализа в интеллектуальной системе "Эйдос" / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №08(162). С. 273 – 355. – IDA [article ID]: 1622008020. – Режим доступа: <http://ej.kubagro.ru/2020/08/pdf/20.pdf>, 5,188 у.п.л.