

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 технологий и систем

КУРСОВАЯ РАБОТА

по дисциплине: Интеллектуальные системы и технологии

на тему: ACK-анализ прогноза ипотечного кредита для домов Zillow

Выполнил студент группы: ИТ2041 Верещагин Константин Геннадьевич

Допущен к защите: _____

Руководитель проекта: д.э.н., к. т. н., профессор Луценко Е. В. ()

(подпись, расшифровка подписи)

Защищен _____ 20.02.2022 _____

(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 Верещагина Константина Геннадьевича
курса 2 очной формы обучения группы ИТ2041

Направление подготовки 09.04.02 «Информационные системы и технологии»

Наименование темы «АСК-анализ прогноза ипотечного кредита для домов Zillow на основе данных портала Kaggle»

Рецензент: Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор
(*Ф.И.О., ученое звание и степень, должность*)

Оценка качества выполнения курсовой работы

№ п/п	Показатель	<i>Оценка соответствия (<u>неудовлетворительно</u>, <u>удовлетворительно</u>, <u>хорошо</u>, <u>отлично</u>)</i>
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

(Е. В. Луценко)

Рецензент _____

«20» февраля 2022 г.

РЕФЕРАТ

Курсовая работа содержит: 32 страниц, 25 рисунков, 8 литературных источников.

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

Целью работы является провести автоматизированный системно-когнитивный анализ прогноза ипотечного кредита на один год для домов Zillow на основе данных портала Kaggle. Добиться этого можно анализом методов формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами, принятия решений и исследования моделируемой предметной области путем исследования модели.

СОДЕРЖАНИЕ

Содержание	4
ВВЕДЕНИЕ	5
1 СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	6
1.1 Описание решения	6
1.2 Преобразование исходных данных из CSV-формата в файл формата XLSX	8
1.3 Ввод выборки в систему AIDOS-X	9
1.4 Синтез и верификация статистических и интеллектуальных моделей ..	13
1.5 Виды моделей системы AIDOS-X	15
1.6 Результаты верификации моделей	18
2 РЕШЕНИЕ ЗАДАЧ НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	21
2.1 Решение задачи идентификации	21
2.2 Кластерно-конструктивный анализ	24
2.3 Нелокальные нейронные сети и нейтроны	25
2.4 SWOT и REST-матрицы и диаграммы	27
ЗАКЛЮЧЕНИЕ	31
СПИСОК ЛИТЕРАТУРЫ	32

ВВЕДЕНИЕ

Современные информационные системы получают все большее развитие благодаря технологиям искусственного интеллекта. Оценка качества математических моделей некоторых из них не выносит критики. В данной курсовой работе рассмотрено решение задачи АСК-анализа прогноза ипотечного кредита на один год для домов Zillow на основе данных портала Kaggle.

Целью данной курсовой работы является.

Задачами, поставленными в данной курсовой работе, являются:

- подготовка исходных данных и формализация предметной области;
- синтез и верификация статистических и системно-когнитивных моделей и выбор наиболее достоверной модели;
- решение различных задач в наиболее достоверной модели: прогнозирование, поддержка принятия решений, исследование полученных моделей.

Объектом исследования данной работы является выборка данных об ипотечных кредитах на один год для домов Zillow.

Результатом данной работы можно считать получение теоретических и практических знаний в области анализа работы систем искусственного интеллекта и анализа результата их работы.

Курсовая работа состоит из введения, двух глав, заключения и списка литературы. Общий объем работы составляет 32 страница.

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 доступный программный инструментарий, в качестве которого в настоящее время выступает Универсальная когнитивная аналитическая система Aidos-X. Система «Эйдос» выгодно отличается от других интеллектуальных систем следующими параметрами:

- разработана в универсальной постановке, не зависящей от предметной области. Поэтому она является универсальной и может быть применена во многих предметных областях, в которых не требуется автоматического, т. е. без непосредственного участия человека в реальном времени решения задач идентификации, прогнозирования, принятия решений и исследования предметной области;
- находится в полном открытом бесплатном доступе причем с актуальными исходными текстами;
- является одной из первых отечественных систем искусственного интеллекта персонального уровня, т. е. не требует от пользователя специальной подготовки в области технологий искусственного интеллекта: «имеет нулевой порог входа»;

- 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 (т.е. не предъявляет жестких требований к данным, которые невозможно выполнить, а обрабатывает те данные, которые есть);
- содержит большое количество локальных (поставляемых с инсталляцией) и более 300 учебных и научных интеллектуальных облачных Эйдос-приложений;
- поддерживает on-line среду накопления знаний и широко используется во всем мире;
- обеспечивает мультиязычную поддержку интерфейса на 51 языке. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;
- 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, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA);
- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний 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;
- вместо того, чтобы предъявлять к исходным данным практически неосуществимые требования (вроде нормальности распределения, абсолютной точности и полных повторности всех сочетаний значений факторов и их полной независимости и аддитивности) автоматизированный системно-когнитивный анализ (АСК-анализ) предлагает без какой-либо предварительной обработки осмыслить эти данные и тем самым преобразовать их в информацию, а затем преобразовать эту информацию в знания путем ее применения для достижения целей (т.е. для управления) и решения задач классификации, поддержки принятия решений и содержательного эмпирического исследования моделируемой предметной области.

1.2 Преобразование исходных данных из CSV-формата в файл формата XLSX

С электронного ресурса kaggle.com возьмем набор данных «One Year Mortgage Loan Prediction for Zillow Homes»,
<https://www.kaggle.com/yamqwe/zillow-home-value-forecast?select=AllRegionsForePublic.csv>

CSV-файл содержит 7 столбцов с данными:

- Region – тип региона;
- RegionName – название региона;
- State – код штата;
- ForecastYoYPctChange – прогноз изменений в годовом исчислении в процентах;
- MSAName – место проведения;
- CountyName – название страны;

- CityName – название города.

Для загрузки модели в систему AIDOS-X необходимо конвертировать CSV-файл в файл формата XLSX. Для конвертации был использован онлайн конвертор: <https://convertio.co/ru/csv-xlsx/>

После конвертации необходимо добавить еще один столбец, который будет называться классифицирующим, таким столбцом было решено выбрать ForecastYoYPctChange, итоговую таблицу можно увидеть на рисунке 1.

A	B	C	D	E	F	G	H
Object	ForecastYoYPctChange	Region	RegionName	State	MSAName	CountyName	CityName
1	3,0958312	National	United States	US			NA
2	2,973574534	MSA	Akron, OH	OH	Akron, OH		NA
3	3,775382633	MSA	Albany, GA	GA	Albany, GA		NA
4	1,498370635	MSA	Albany, NY	NY	Albany, NY		NA
5	1,176320195	MSA	Albuquerque, NM	NM	Albuquerque, NM		NA
6	2,156294351	MSA	Alexandria, LA	LA	Alexandria, LA		NA
7	3,066000632	MSA	Allentown, PA	PA	Allentown, PA		NA
8	4,574864656	MSA	Altoona, PA	PA	Altoona, PA		NA
9	3,047663605	MSA	Amarillo, TX	TX	Amarillo, TX		NA
10	3,721190901	MSA	Ames, IA	IA	Ames, IA		NA
11	4,454014544	MSA	Anchorage, AK	AK	Anchorage, AK		NA
12	5,175313616	MSA	Ann Arbor, MI	MI	Ann Arbor, MI		NA
13	4,066802757	MSA	Appleton, WI	WI	Appleton, WI		NA
14	2,148950872	MSA	Asheville, NC	NC	Asheville, NC		NA
15	4,283114052	MSA	Athens, GA	GA	Athens, GA		NA
16	5,239899702	MSA	Atlanta, GA	GA	Atlanta, GA		NA
17	2,090909859	MSA	Atlantic City, NJ	NJ	Atlantic City, NJ		NA
18	2,932023314	MSA	Auburn, AL	AL	Auburn, AL		NA
19	3,623974132	MSA	Augusta, GA	GA	Augusta, GA		NA
20	3,338547884	MSA	Austin, TX	TX	Austin, TX		NA
21	2,926440132	MSA	Bakersfield, CA	CA	Bakersfield, CA		NA
22	2,915608112	MSA	Baltimore, MD	MD	Baltimore, MD		NA
23	2,464469365	MSA	Bangor, ME	ME	Bangor, ME		NA
24	4,109969356	MSA	Cape Cod, MA	MA	Cape Cod, MA		NA
25	3,003938522	MSA	Baton Rouge, LA	LA	Baton Rouge, LA		NA
26	2,087209143	MSA	Battle Creek, MI	MI	Battle Creek, MI		NA

Рисунок 1 – Фрагмент обучающей выборки

1.3 Ввод выборки в систему AIDOS-X

Для импорта обучающей выборки в систему AIDOS-X необходимо скопировать ее в папку Inp_data и переименовать в Inp_data.xlsx, после этого можно запустить саму программу и универсальный программный интерфейс импорта данных в систему (режим 2.3.2.2), результат заполнения которого представлен на рисунке 2.

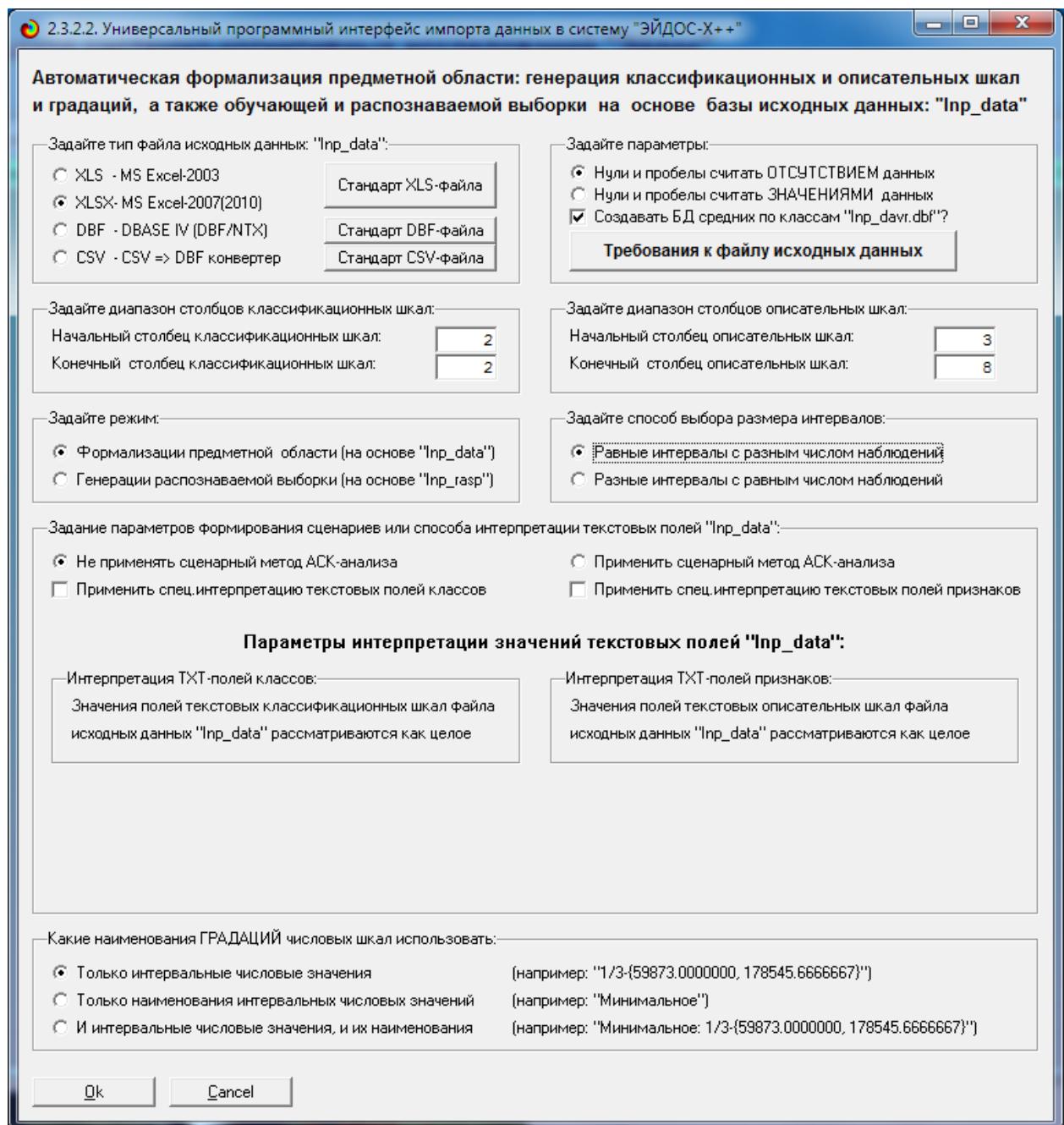


Рисунок 2 – Интерфейс импорта

Следует выделить следующие настройки:

- Тип файла – XLSX;
- Классификационная шкала – 2;
- Описательные шкалы – 3-8.

После этого приложение просит задать размерности модели системы, оставляем всё как есть (рисунок 3).

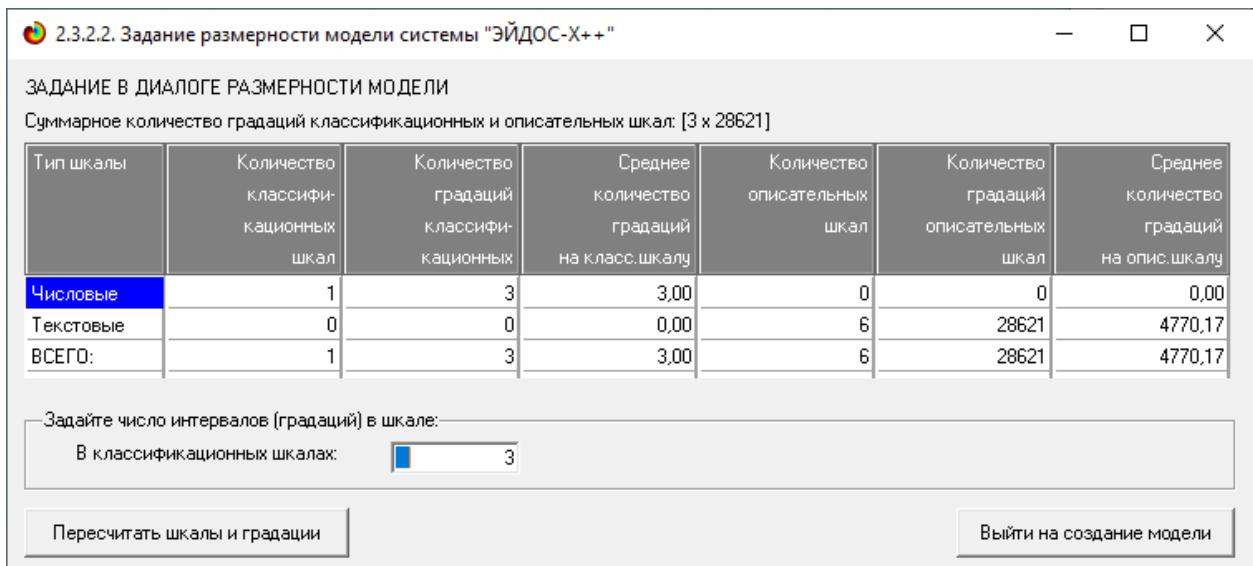


Рисунок 3 – Задание размерностей системы

Процесс импорта данных из внешнего файла в систему представлен на рисунке 4.

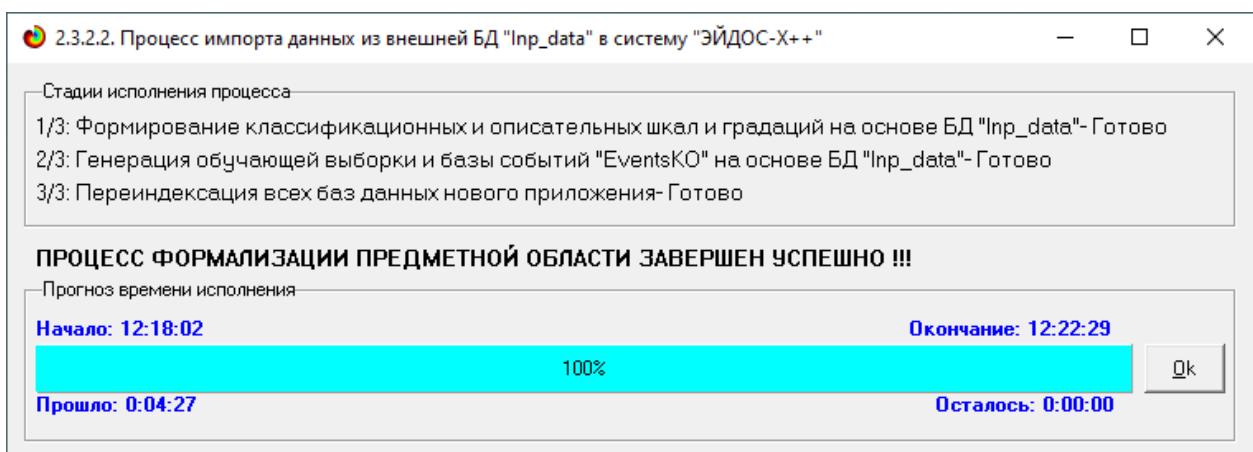


Рисунок 4 – Импорт данных

После загрузки данных система автоматически нашла классификационные шкалы, которые можно посмотреть в режиме 2.1 (рисунок 5) и описательные шкалы, которые можно посмотреть в режиме 2.2 (рисунок 6).

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

Код шкалы	Наименование классификационной шкалы	Код градации	Наименование градации классификационной шкалы
1	FORECASTYDYPCTCHANGE	1	1/3(-4.5155579, 2.6781946)
		2	2/3(2.6781946, 9.8719470)
		3	3/3(9.8719470, 17.0656995)

Помощь | Доб.шкулу | Доб.град.шкулы | Копир.шкулу | Копир.град.шкулы | Копир.шкулу с град. | Удал.шкулу с град. | Удал.град.шкулы | Удаление и перекодирование | Графики будущих сценариев

Рисунок 5 – Классификационные шкалы

2.2. Описательные шкалы и градации. Текущая модель: "INF1"

Код шкалы	Наименование описательной шкалы	Код градации	Наименование градации описательной шкалы
1	REGION	1	City
2	REGIONNAME	2	County
3	STATE	3	National
4	MSANAME	4	ZipCode
5	COUNTYNAME		
6	CITYNAME		

Помощь | Доб.шкулу | Доб.град.шкулы | Копир.шкулу | Копир.град.шкулы | Копир.шкулу с град. | Удал.шкулу с град. | Удал.град.шкулы | Перекодировать | Очистить | Графики прошлых сценариев

Рисунок 6 – Описательные шкалы

Так же существует возможность ручной корректировки выгруженных данных и добавление новых объектов, которая открывается с помощью режима 2.3.1 (рисунок 7).

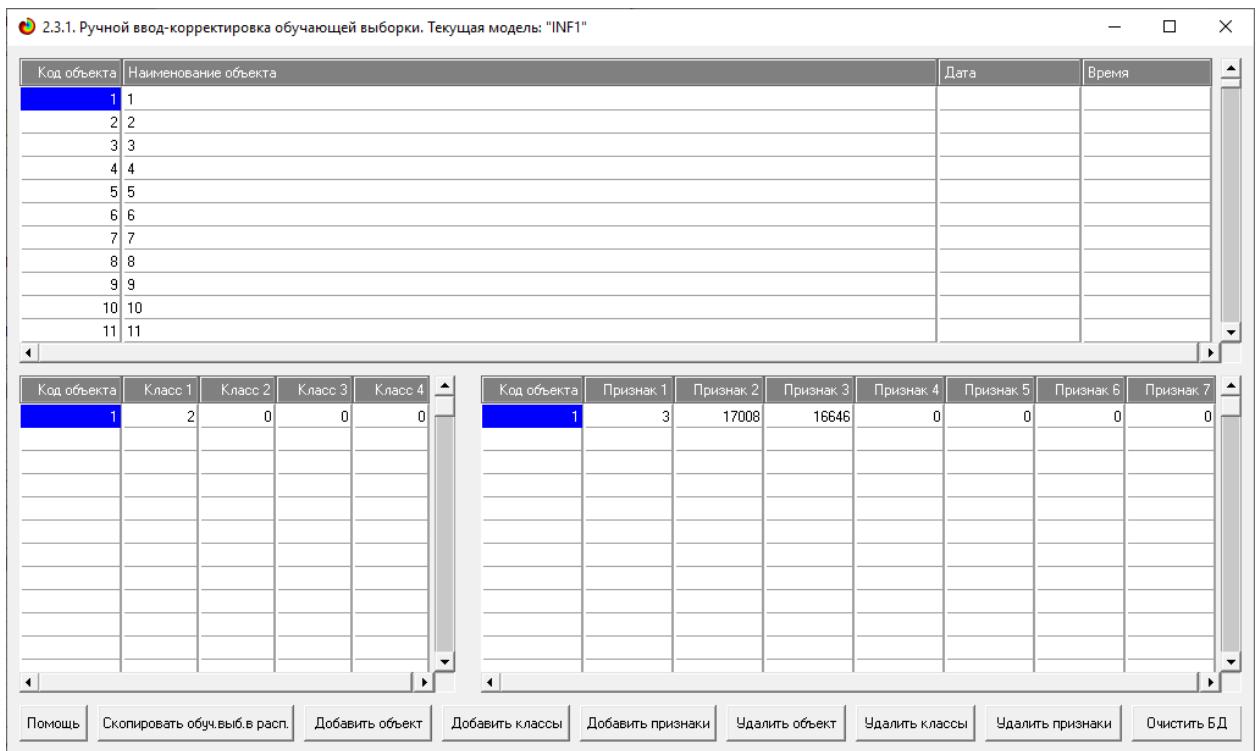


Рисунок 7 – Ручная корректировка

1.4 Синтез и верификация статистических и интеллектуальных моделей

Для синтеза и верификации моделей создан режим 3.5, после его запуска задается модель, которая помечается текущей (рисунок 8).

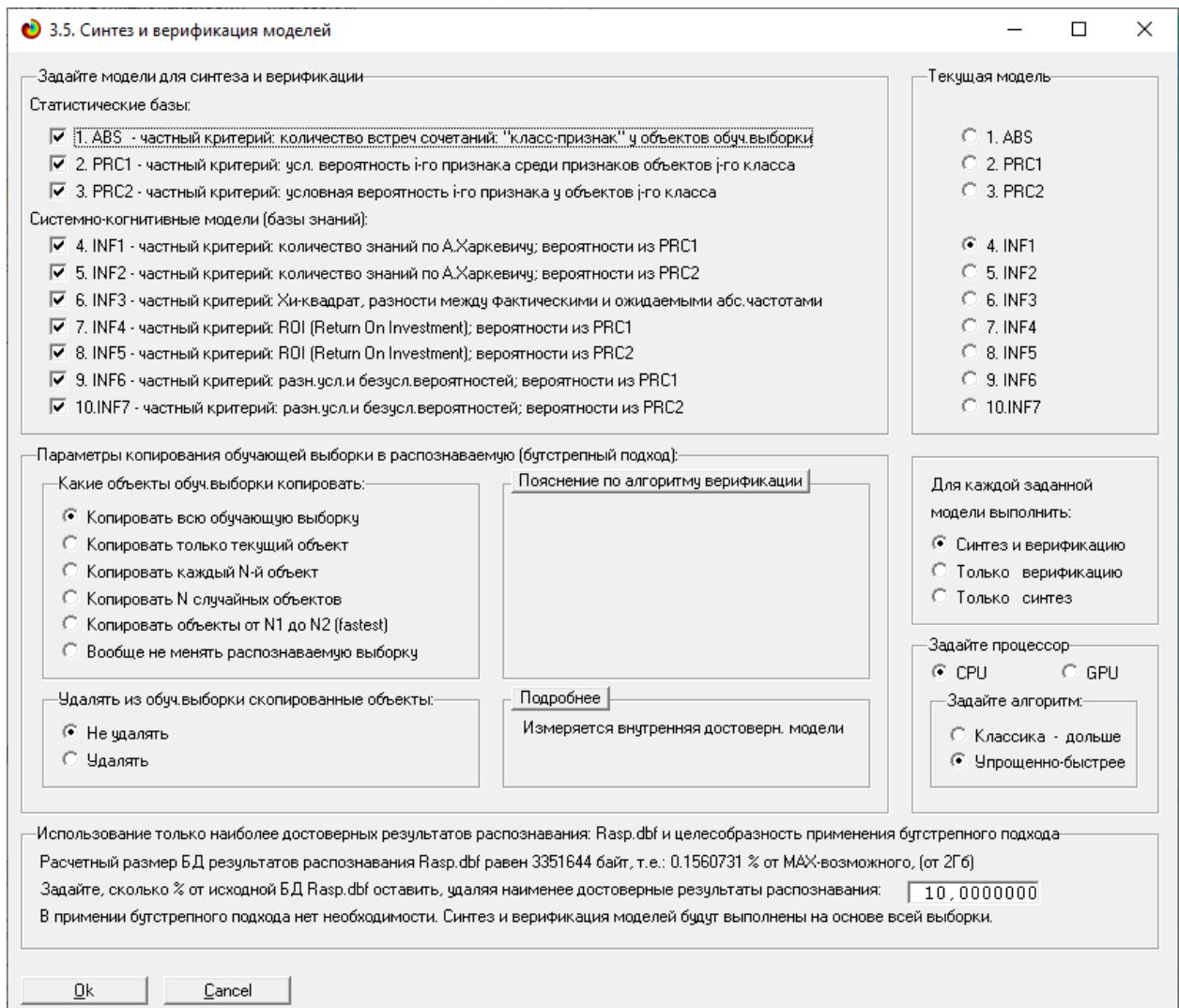


Рисунок 8 – Режим синтеза моделей

Данный режим содержит множество различных методов верификации, но мы используем параметры, которые система предлагает по умолчанию. Стадия процесса исполнения режима 3.5 и прогноз времени исполнения показаны на рисунке 9.

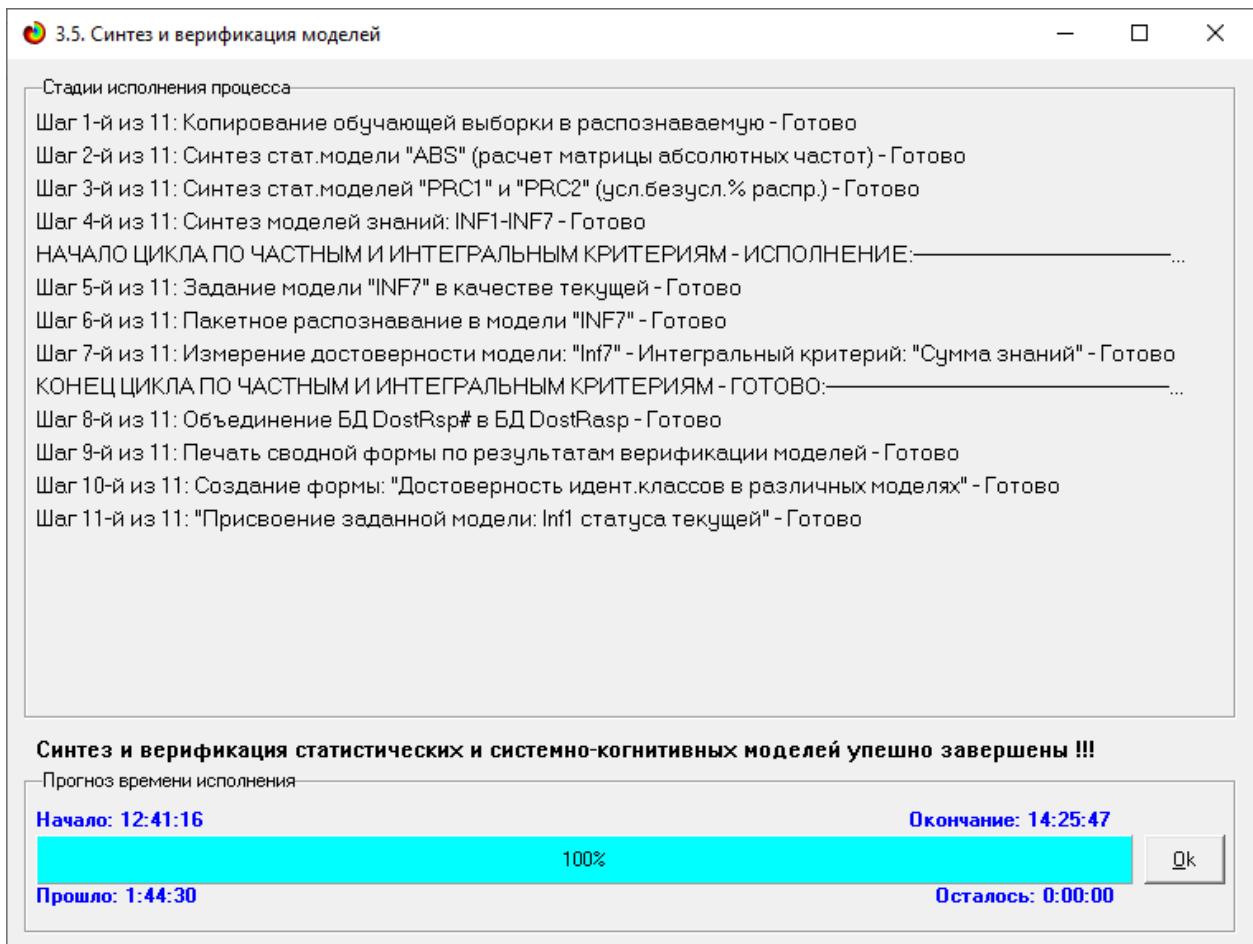


Рисунок 9 – Исполнение режима синтеза моделей

Следует заметить, что синтез и верификация всех моделей занял 29 секунд. После данного этапа можно приступить к выбору наиболее достоверной модели.

1.5 Виды моделей системы AIDOS-X

Решение задачи идентификации может быть рассмотрено на нескольких моделях, в которых рассчитано количество информации по А. Харкевичу, которое было сгенерировано по принадлежности идентифицируемого объекта к каждому из классов, учитывая признаки объекта.

Таким образом, частные критерии представляют собой формулы для преобразования матрицы абсолютных частот, которая представлена на

рисунке 10 и открытой с помощью режима 5.5.1, в матрицы условных и безусловных процентных распределений и матрицы знаний.

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"							
Код признака	Наименование описательной шкалы и градации	1. FORECASTYO... 1/3 { 4.5, 2.7})	2. FORECASTYO... 2/3 {2.7, 9.9})	3. FORECASTYO... 3/3 {9.9, 17.1})	Сумма	Среднее	Средн. квадр. откл.
1	REGION-City	4001	5499	5	9505	3168.33	2840.07
2	REGION-County	257	542	2	801	267.00	270.14
3	REGION-National		1		1	0.33	0.58
4	REGION-ZipCode	4603	7084	19	11706	3902.00	3584.29
5	REGIONNAME-0000000	3926	6371	18	10315	3438.33	3204.45
6	REGIONNAME-01001		1		1	0.33	0.58
7	REGIONNAME-01002	1			1	0.33	0.58
8	REGIONNAME-01005		1		1	0.33	0.58
9	REGIONNAME-01007	1			1	0.33	0.58
10	REGIONNAME-01008	1			1	0.33	0.58
11	REGIONNAME-01010		1		1	0.33	0.58
12	REGIONNAME-01011	1			1	0.33	0.58
13	REGIONNAME-01013	1			1	0.33	0.58
14	REGIONNAME-01020	1			1	0.33	0.58
15	REGIONNAME-01022	1			1	0.33	0.58
16	REGIONNAME-01026	1			1	0.33	0.58
17	REGIONNAME-01027		1		1	0.33	0.58
18	REGIONNAME-01028	1			1	0.33	0.58
19	REGIONNAME-01030	1			1	0.33	0.58
20	REGIONNAME-01033		1		1	0.33	0.58
21	REGIONNAME-01034	1			1	0.33	0.58
22	REGIONNAME-01035	1			1	0.33	0.58
23	REGIONNAME-01036	1			1	0.33	0.58
24	REGIONNAME-01038	1			1	0.33	0.58

Рисунок 10 – Матрица абсолютных частот

5.5. Модель: "7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. FORECASTYOY... 1/3 { 4.5. 2.7}	2. FORECASTYOY... 2/3 (2.7, 9.9)	3. FORECASTYOY... 3/3 (9.9, 17.1)	Сумма	Среднее	Средн. квадр. откл.
1	REGION-City	0.018	-0.011	-0.554	-0.548	-0.183	0.322
2	REGION-County	-0.224	0.156	1.118	1.050	0.350	0.692
3	REGION-National		0.709		0.709	0.236	0.409
4	REGION-ZipCode	-0.049	0.034	0.377	0.362	0.121	0.226
5	REGIONNAME-0000000	-0.080	0.055	0.480	0.456	0.152	0.292
6	REGIONNAME-01001		0.709		0.709	0.236	0.409
7	REGIONNAME-01002	1.418			1.418	0.473	0.818
8	REGIONNAME-01005		0.709		0.709	0.236	0.409
9	REGIONNAME-01007	1.418			1.418	0.473	0.818
10	REGIONNAME-01008	1.418			1.418	0.473	0.818
11	REGIONNAME-01010		0.709		0.709	0.236	0.409
12	REGIONNAME-01011	1.418			1.418	0.473	0.818
13	REGIONNAME-01013	1.418			1.418	0.473	0.818
14	REGIONNAME-01020	1.418			1.418	0.473	0.818
15	REGIONNAME-01022	1.418			1.418	0.473	0.818
16	REGIONNAME-01026	1.418			1.418	0.473	0.818
17	REGIONNAME-01027		0.709		0.709	0.236	0.409
18	REGIONNAME-01028	1.418			1.418	0.473	0.818
19	REGIONNAME-01030	1.418			1.418	0.473	0.818
20	REGIONNAME-01033		0.709		0.709	0.236	0.409
21	REGIONNAME-01034	1.418			1.418	0.473	0.818
22	REGIONNAME-01035	1.418			1.418	0.473	0.818
23	REGIONNAME-01036	1.418			1.418	0.473	0.818
24	REGIONNAME-01038	1.418			1.418	0.473	0.818

Рисунок 11 – Фрагмент модели INF4

5.5. Модель: "8. INF5 - частный критерий: ROI (Return On Investment); вероятности из PRC2"

Код признака	Наименование описательной шкалы и градации	1. FORECASTYOY... 1/3 { 4.5. 2.7}	2. FORECASTYOY... 2/3 (2.7, 9.9)	3. FORECASTYOY... 3/3 (9.9, 17.1)	Сумма	Среднее	Средн. квадр. откл.
1	REGION-City	0.050	-0.032	-0.548	-0.531	-0.177	0.324
2	REGION-County	-0.200	0.132	1.146	1.078	0.359	0.701
3	REGION-National		0.673		0.673	0.224	0.388
4	REGION-ZipCode	-0.019	0.012	0.395	0.388	0.129	0.231
5	REGIONNAME-0000000	-0.051	0.033	0.500	0.482	0.161	0.297
6	REGIONNAME-01001		0.673		0.673	0.224	0.388
7	REGIONNAME-01002	1.494			1.494	0.498	0.862
8	REGIONNAME-01005		0.673		0.673	0.224	0.388
9	REGIONNAME-01007	1.494			1.494	0.498	0.862
10	REGIONNAME-01008	1.494			1.494	0.498	0.862
11	REGIONNAME-01010		0.673		0.673	0.224	0.388
12	REGIONNAME-01011	1.494			1.494	0.498	0.862
13	REGIONNAME-01013	1.494			1.494	0.498	0.862
14	REGIONNAME-01020	1.494			1.494	0.498	0.862
15	REGIONNAME-01022	1.494			1.494	0.498	0.862
16	REGIONNAME-01026	1.494			1.494	0.498	0.862
17	REGIONNAME-01027		0.673		0.673	0.224	0.388
18	REGIONNAME-01028	1.494			1.494	0.498	0.862
19	REGIONNAME-01030	1.494			1.494	0.498	0.862
20	REGIONNAME-01033		0.673		0.673	0.224	0.388
21	REGIONNAME-01034	1.494			1.494	0.498	0.862
22	REGIONNAME-01035	1.494			1.494	0.498	0.862
23	REGIONNAME-01036	1.494			1.494	0.498	0.862
24	REGIONNAME-01038	1.494			1.494	0.498	0.862

Рисунок 12 – Фрагмент модели INF5

1.6 Результаты верификации моделей

С результатами верификации моделей, отличающихся частными критериями, можно ознакомиться в режиме 3.4 системы «Эйдос», они представлены на рисунке 13.

3.4. Обобщ форма по достов.моделей при разн.инт.крит. Текущая модель: "INF1"											—	□	×
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 Ризбергена	Сумма модул. уровней сход. истинно-поло...	Сумма н. уровней сход. истинно-решения (STP)		
1. ABS - частный критерий: количество встреч сопоставл. "Усл...	Корреляция abs частот с обр...	22343	22340	438	4030	3	0.847	1.000	0.917	10766.585			
1. ABS - частный критерий: количество встреч сопоставл. "Усл...	Сумма abs частот по признак...	22343	22343	349	4119		0.844	1.000	0.916	12296.331			
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн частот с о...	22343	22340	438	4030	3	0.847	1.000	0.917	10766.591			
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн частот по при...	22343	22343	349	4119		0.844	1.000	0.916	7886.972			
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн частот с о...	22343	22340	438	4030	3	0.847	1.000	0.917	10766.565			
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн частот по при...	22343	22343	349	4119		0.844	1.000	0.916	7781.813			
4. INF1 - частный критерий: количество знаний по АХаркевичиу...	Семантический резонанс зна...	22343	11615	3419	1049	10728	0.917	0.520	0.664	267.900	24		
4. INF1 - частный критерий: количество знаний по АХаркевичиу...	Сумма знаний	22343	19616	3086	1382	2727	0.934	0.878	0.905	1340.199	18		
5. INF2 - частный критерий: количество знаний по АХаркевичиу...	Семантический резонанс зна...	22343	11500	3416	1052	10843	0.916	0.515	0.659	270.912	24		
5. INF2 - частный критерий: количество знаний по АХаркевичиу...	Сумма знаний	22343	19497	3065	1403	2846	0.933	0.873	0.902	1339.519	17		
6. INF3 - частный критерий: Хиквадрат, различия между фактами...	Семантический резонанс зна...	22343	13842	2937	1531	8501	0.900	0.620	0.734	3250.130	39		
6. INF3 - частный критерий: Хиквадрат, различия между фактами...	Сумма знаний	22343	13842	2937	1531	8501	0.900	0.620	0.734	3880.646	14		
7. INF4 - частный критерий ROI (Return On Investment); вероятно...	Семантический резонанс зна...	22343	10155	4240	228	12188	0.978	0.455	0.621	342.779	23		
7. INF4 - частный критерий ROI (Return On Investment); вероятно...	Сумма знаний	22343	20621	3016	1452	1722	0.934	0.923	0.929	50.883			
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	22343	10091	4239	229	12252	0.978	0.452	0.618	344.689	23		
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	22343	20522	2993	1475	1821	0.933	0.918	0.926	50.876			
9. INF6 - частный критерий: разн.усл и безузл.вероятностей; вер...	Семантический резонанс зна...	22343	13773	2868	1600	8570	0.896	0.616	0.730	3189.184	36		
9. INF6 - частный критерий: разн.усл и безузл.вероятностей; вер...	Сумма знаний	22343	13842	2859	1609	8501	0.896	0.620	0.732	462.024	25		
10. INF7 - частный критерий: разн.усл и безузл.вероятностей; вер...	Семантический резонанс зна...	22343	13015	2774	1694	9328	0.885	0.583	0.703	2478.348	38		
10. INF7 - частный критерий: разн.усл и безузл.вероятностей; вер...	Сумма знаний	22343	13032	2767	1701	9311	0.885	0.583	0.703	385.431	24		

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

Рисунок 13 – Оценки достоверности моделей

Для оценки достоверности моделей в АСК-анализе и систему Aidos-X используется F-мера Ван Ризбергена и L-мера, представляющая собой ее нечеткое мультиклассовое обобщение, предложенное профессором Луценко (рисунок 14). Наиболее достоверной оказалась модель INF4. Точность данной модели по F-мере Ван Ризбергена составляет 0.929, а по L1-мере профессора Луценко – 0.941. L1-мера, предложенная профессором Луценко является более достоверной, в сравнении с F-мерой Ван Ризбергена. Таким образом, уровень достоверности прогнозирования с применением модели выше экспертных оценок, достоверность которых считается равной примерно 100%.

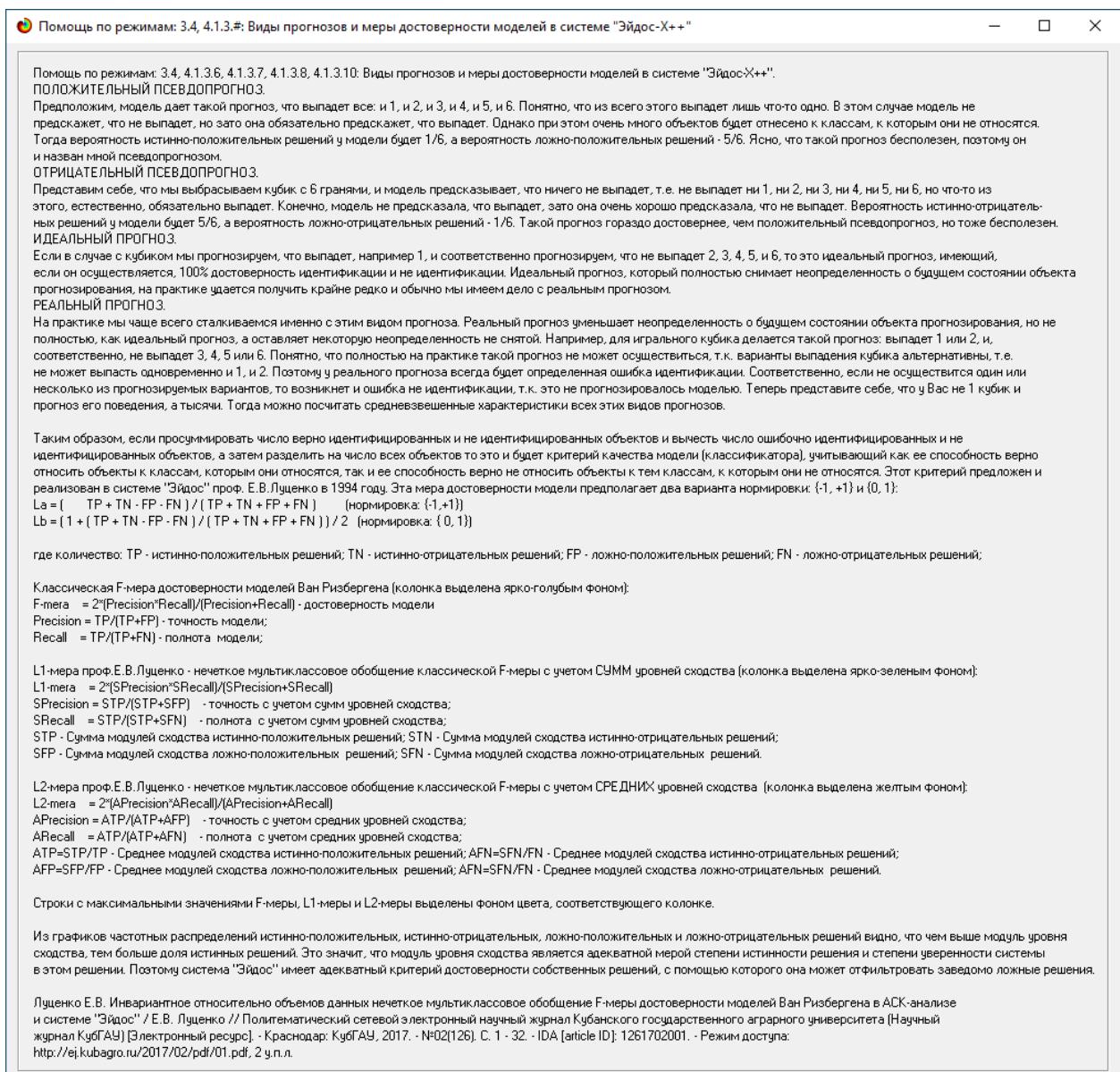


Рисунок 14 – Виды прогнозов и меры достоверностей в системе «Эйдос»

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. Этим и оправдано применение моделей знаний из 23 интеллектуальных технологий. На рисунке 15 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и не идентифицированных ситуаций в наиболее достоверной модели INF4, однако далее при анализе мы будем использовать модель INF4.

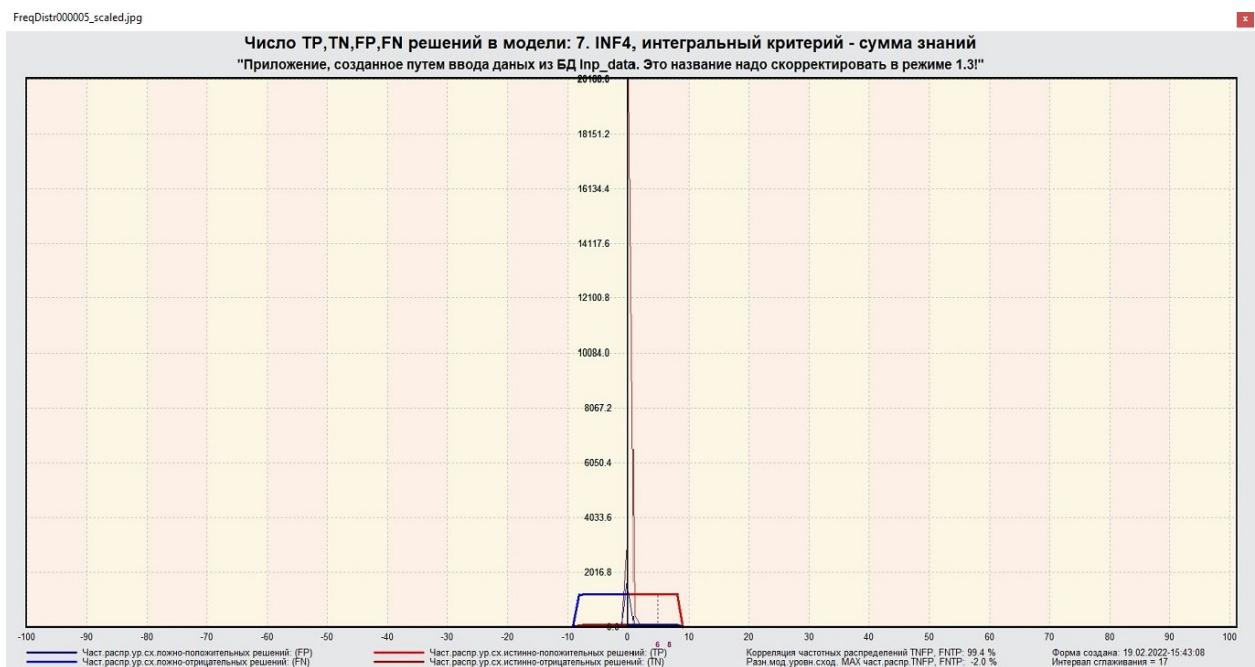


Рисунок 15 – Частотные распределения верно и ошибочно идентифицированных и не идентифицированных состояний объекта моделирования в зависимости от уровня сходства в модели INF4

Левое распределение, включает только истинно-отрицательные значения, а правое включает ложные отрицательные и истинно-положительные решения. Сдвиг этих распределений относительно друг друга и другие различия между ними и позволяют решать задачу идентификации видов героев по его характеристикам и другие задачи.

Для положительных решений от 0% до примерно 10% количество ложных решений больше числа истинных, но далее идет на спад и ложные решения перестают встречаться после 10%.

2 РЕШЕНИЕ ЗАДАЧ НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1 Решение задачи идентификации

В соответствии с технологией АСК-анализа необходимо задать текущую модель в режиме 5.6, в ее качестве выберем наиболее достоверную модель INF4, что показано на рисунке 16.

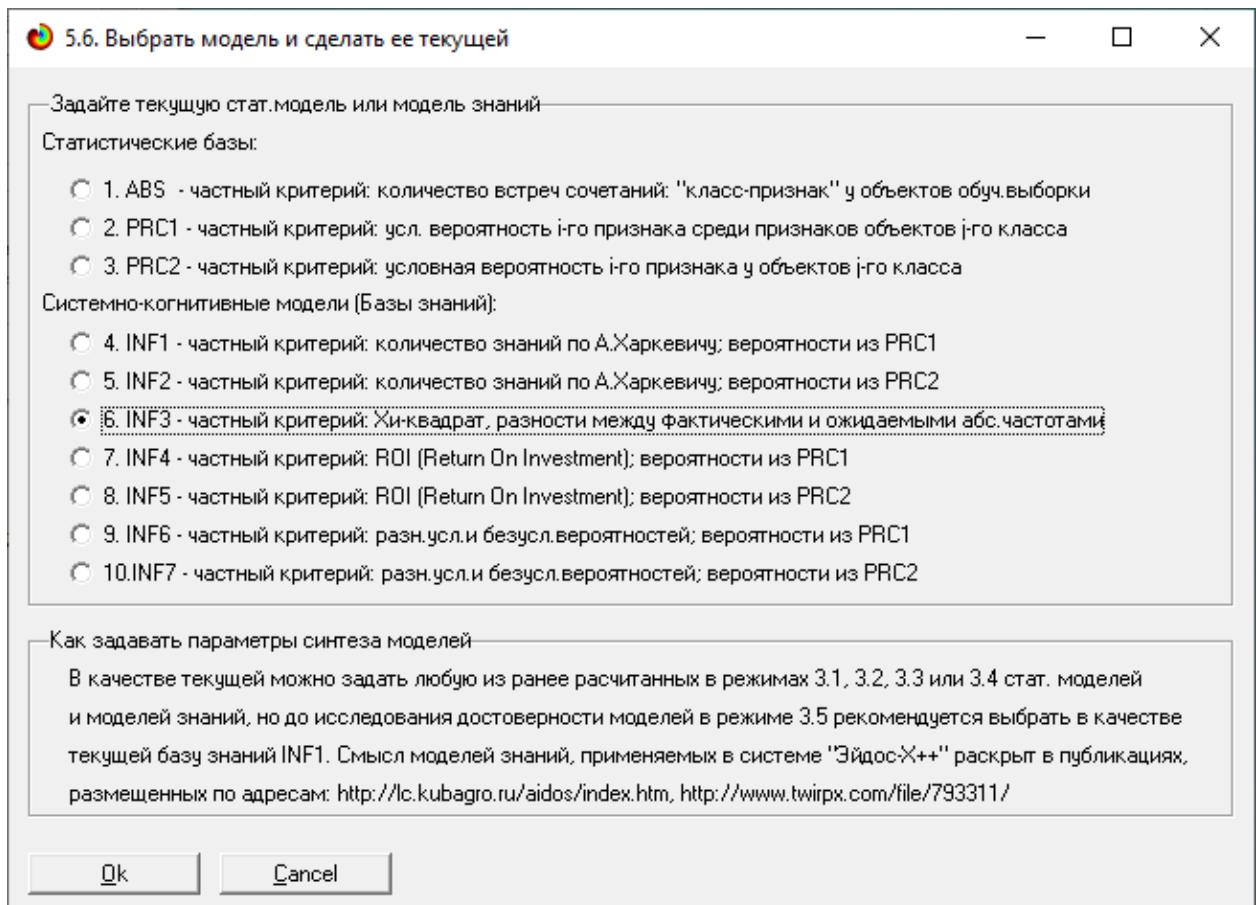


Рисунок 16 – Выбор текущей модели

После этого необходимо провести пакетное распознавание текущей модели в режиме 4.2.1 (рисунок 17)

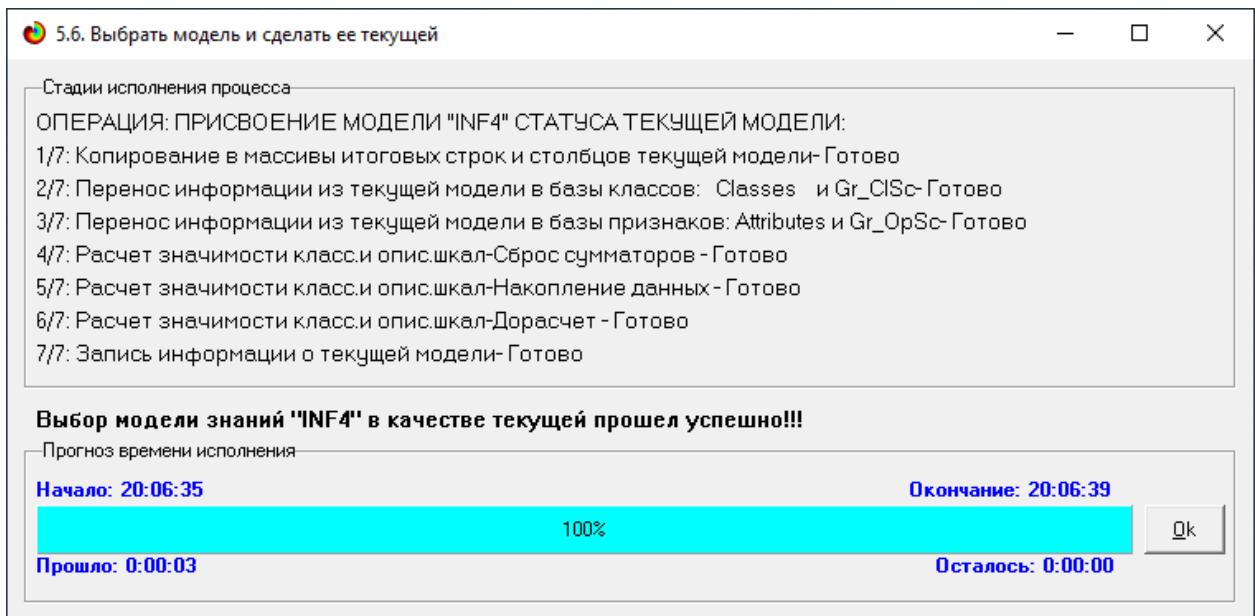


Рисунок 17 – Пакетное распознавание текущей модели

Режим 4.1.3 системы «Эйдос» обеспечивает отображение результатов идентификации и прогнозирования в различный 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;
- Распознавание уровня сходства при разных моделях и интегральных критериях;
 - Достоверность идентификации классов при разных моделях и интегральных критериях.

Рассмотрим подробнее режимы работы 4.1.3.1 и 4.1.3.2, результаты которых представлены на рисунке 18 и 19.

Рисунок 18 – Результат режима работы 4.1.3.1, идентификация объектов в модели INF4

Классы		Интегральный критерий сходства: "Семантический резонанс знаний"			
Код	Наим. класса	Код	Наименование объекта	Сходство	Ф...
1	FORECASTYOUPCTCHANGE-1/3(-4.5, 2.7)	25...	2545	73,53...	v
2	FORECASTYOUPCTCHANGE-2/3(2.7, 9.9)	80...	8039	71,51...	v
3	FORECASTYOUPCTCHANGE-3/3(9.9, 17.1)	83...	8307	71,50...	v
		28...	2865	71,46...	v
		41...	4128	71,40...	v
		50...	5030	71,31...	v
		24...	2407	71,30...	v
		63...	6329	71,27...	v
		84...	8499	71,27...	v
		35...	3585	71,26...	v

Интегральный критерий сходства: "Сумма знаний"					
Код	Наименование объекта	Сходство	Ф...	Сходство	
73...	7318	100,00...	v		
32...	3284	99,79...	v		
34...	3455	99,60...	v		
55...	5543	99,52...	v		
49...	4945	99,48...	v		
88...	8861	98,99...	v		
95...	9580	98,38...	v		
25...	2545	98,38...	v		
28...	2849	98,21...	v		
76...	7601	97,95...	v		

Рисунок 19 – Результат режима работы 4.1.3.2, идентификация классов в модели INF4

Из рисунков выше видно, что результаты идентификации являются динамичными.

2.2 Кластерно-конструктивный анализ

Для выявления сходства-различия обобщенных образов различных результатов научной деятельности по характерных для них системам значений показателей можно осуществить с помощью режимов 4.2.2.1 и 4.2.2.2, результаты выполнения этих режимов показаны на рисунке 20.

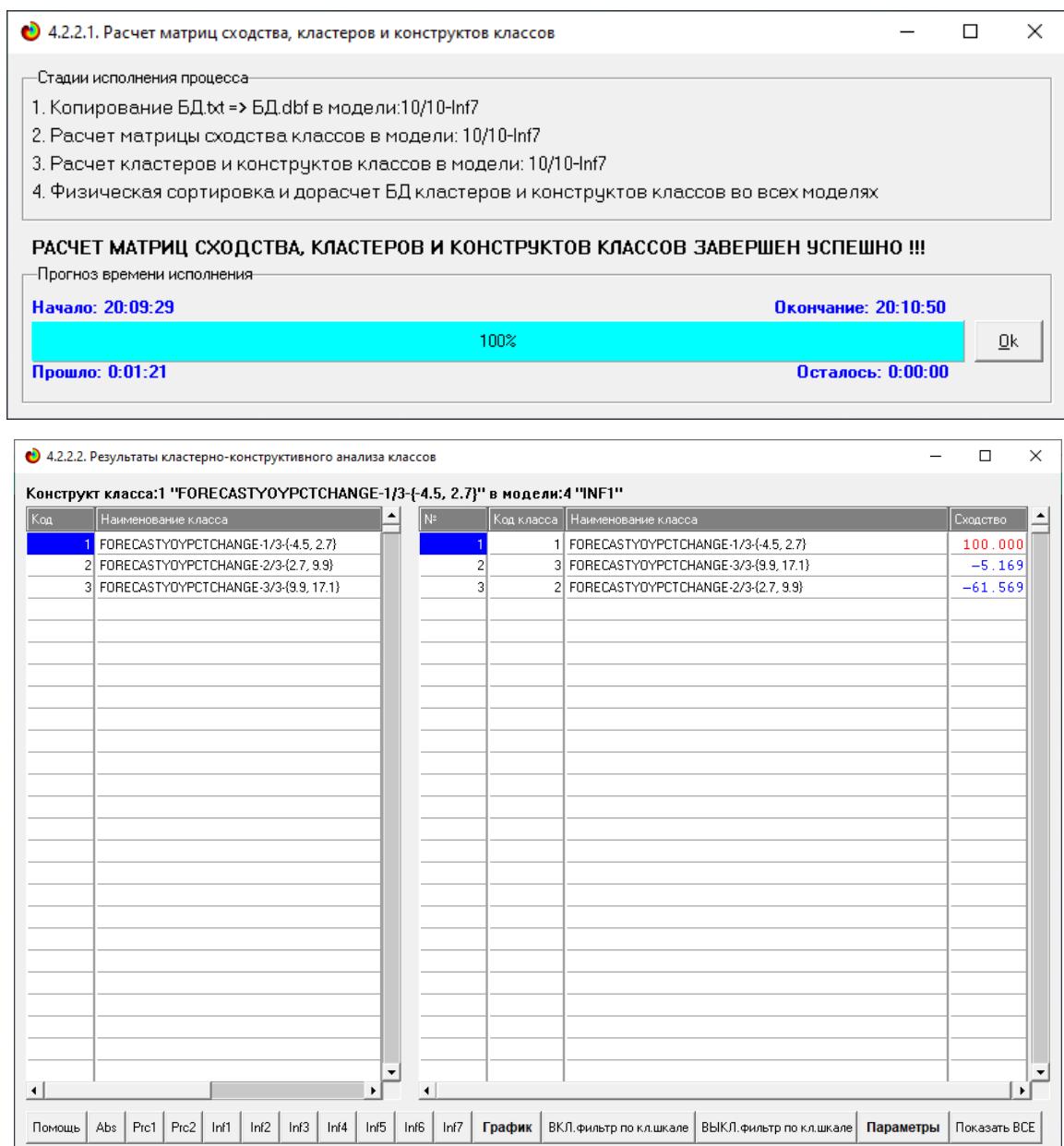


Рисунок 20 – Кластерно-конструктивный анализ модели INF4

2.3 Нелокальные нейронные сети и нейтроны

После Каждому классу системно-когнитивной модели соответствует нелокальный нейрон, совокупность которых образует нелокальную нейронную сеть.

На рисунке 21 изображено графическое отображение нелокальных нейронов в системе Aidos-X.

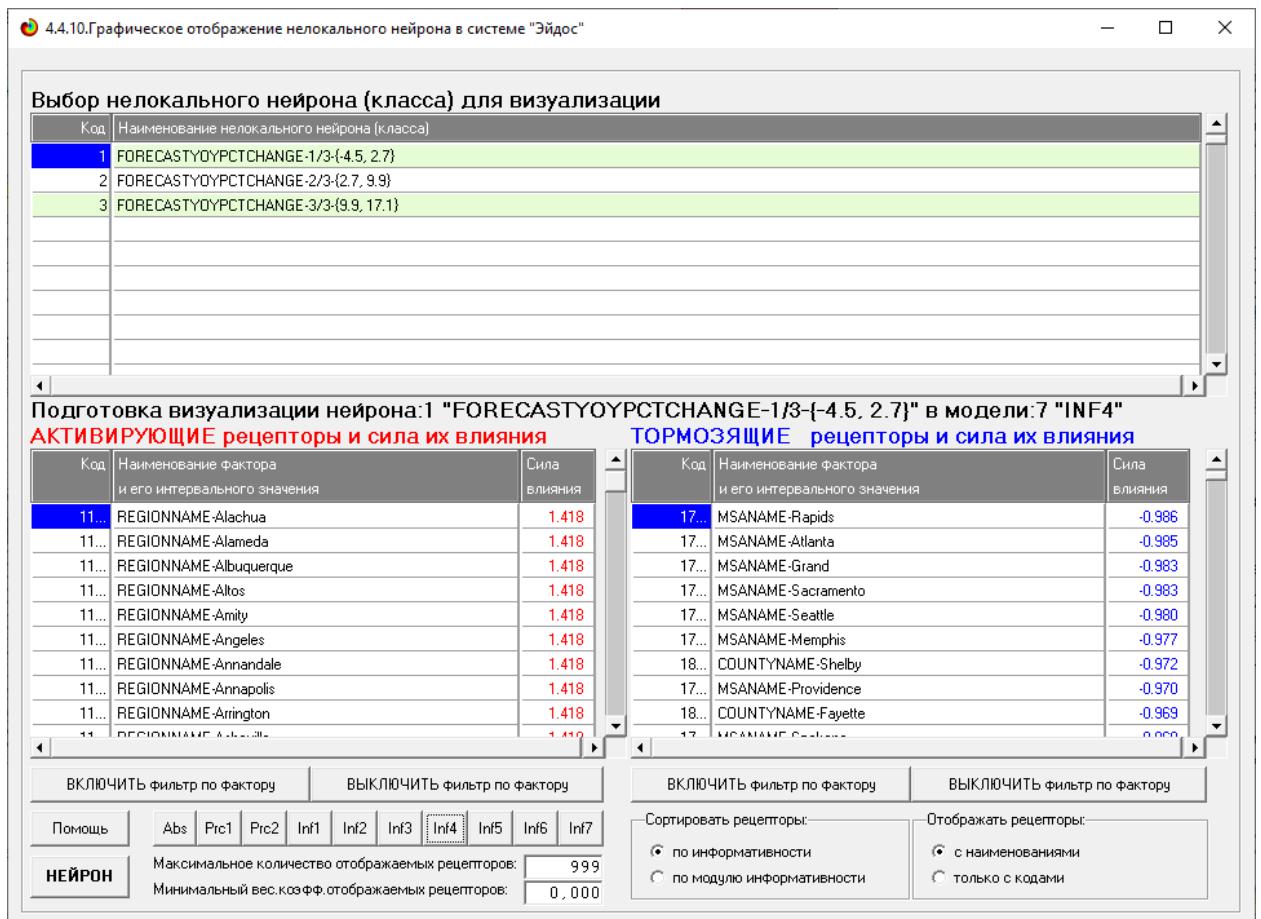


Рисунок 21 – Нелокальные нейроны

Для каждого технологического фактора в соответствии с предложенной моделью определяется величина и направление его влияния на осуществление всех желаемых и не желаемых хозяйственных ситуаций. Для каждой ситуации эта информация отображается в различных текстовых и графических формах, в частности в форме нелокального нейрона (рисунок 22). На данной диаграмме цвет линии означает знак связи (красный – положительная, синий – отрицательная), а толщина – её модуль

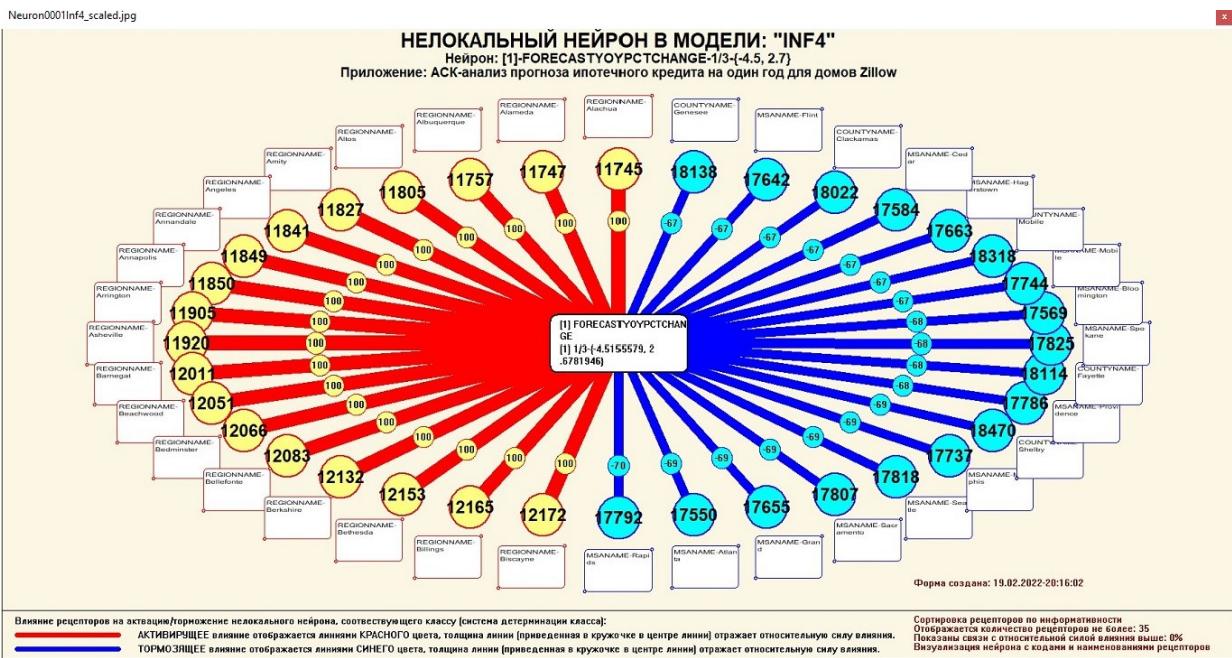


Рисунок 22 – Нелокальный нейрон в модели «INF4»

Дополнение модели нейрона связями факторов позволяет построить классическую когнитивную карту ситуации (будущего состояния АОУ). Детальная внутренняя структура любой связи отображается в форме инвертированной когнитивной диаграммы (рисунок 23). Необходимо отметить, что все указанные графические формы генерируются системой Aidos-X автоматически в соответствии с созданной моделью.

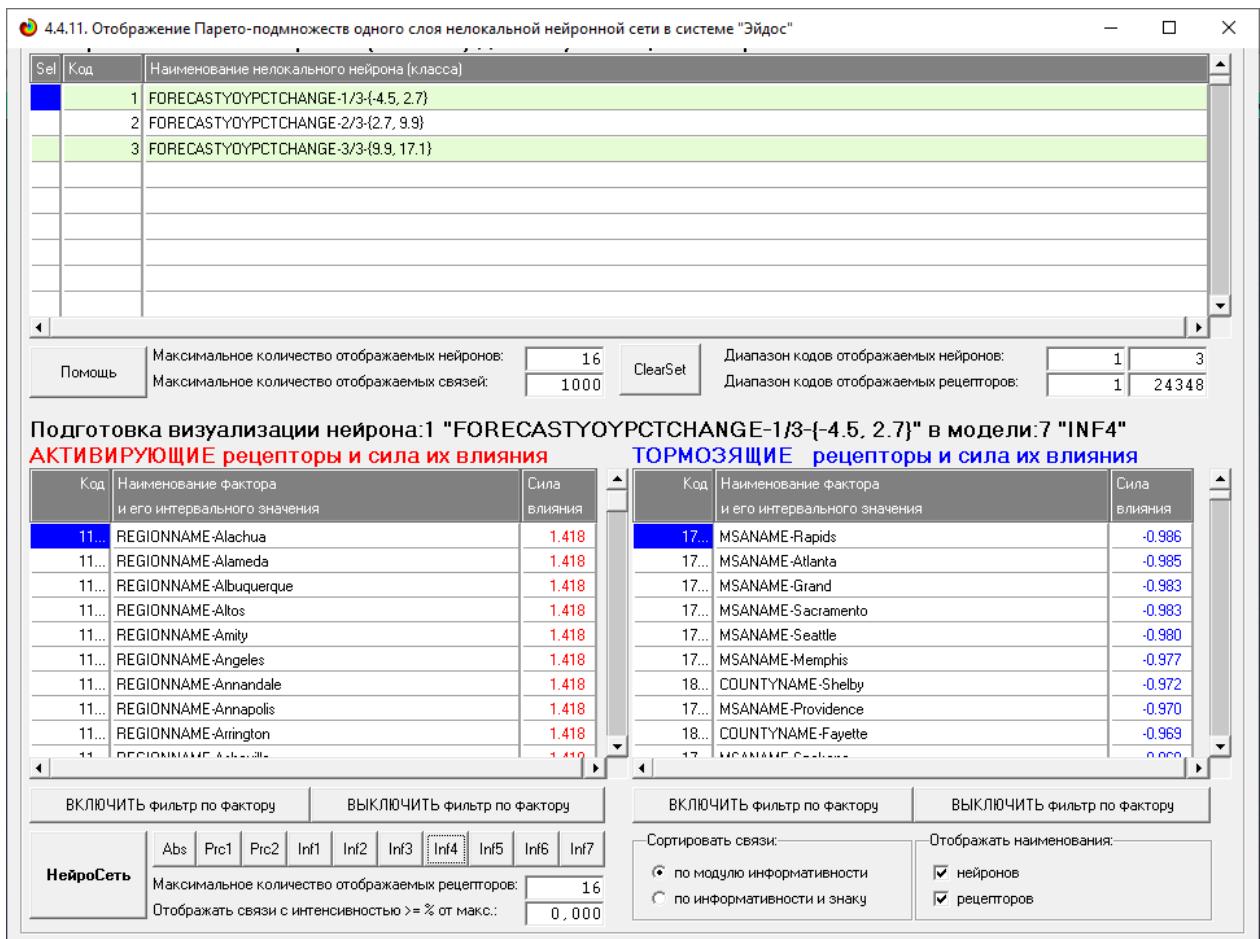


Рисунок 23 – Паретто – подмножество нелокальной нейронной сети

2.4 SWOT и REST-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 не формализуемым 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 Aidos-X. 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: Aidos-X++ предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (рисунки 24 и 25).

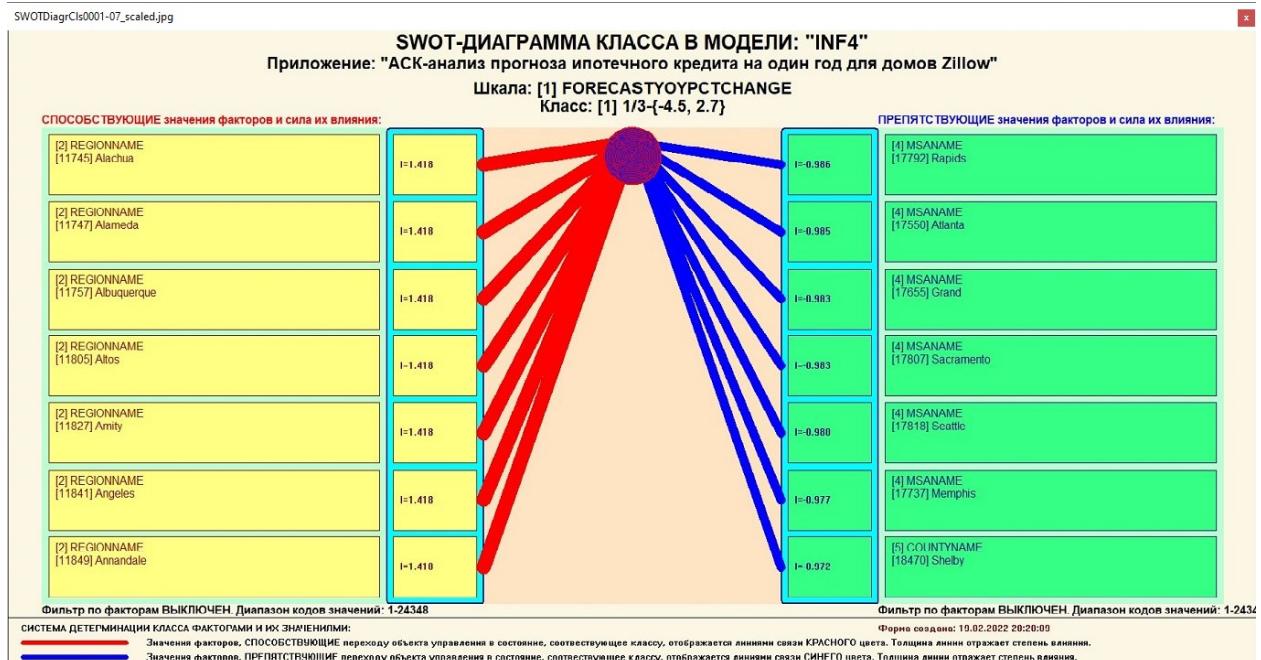
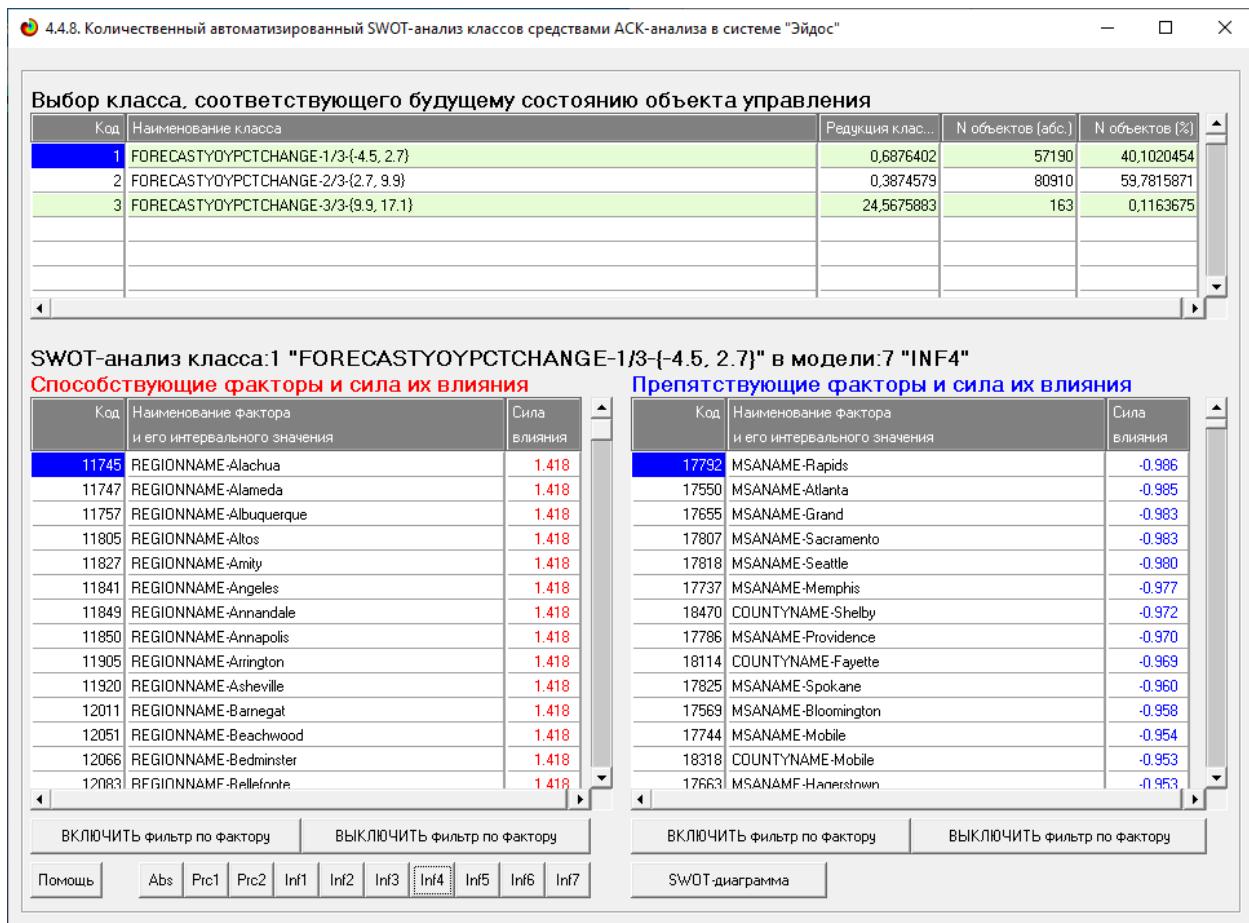


Рисунок 24 – Количественный SWOT-анализ классов в модели INF4

4.4.9 Количественный автоматизированный SWOT-анализ значений факторов средствами АСК-анализа в системе "Эйдос"

Выбор значения фактора, оказывающего влияние на переход объекта управления в будущие состояния

Код	Наменование значения фактора
1	REGION-City
2	REGION-County
3	REGION-National
4	REGION-ZipCode
5	REGIONNAME-0000000
6	REGIONNAME-01001

SWOT-анализ значения фактора: 1 "REGION-City" в модели: 7 "INF4"

СПОСОБСТВУЕТ: **ПРЕПЯТСТВУЕТ:**

Код	Состояния объекта управления, переходу в которые данное значение фактора СПОСОБСТВУЕТ	Сила влияния
1	FORECASTYOYPCTCHANGE-1/3(-4.5, 2.7)	0.018

Код	Состояния объекта управления, переходу в которые данное значение фактора ПРЕПЯТСТВУЕТ	Сила влияния
3	FORECASTYOYPCTCHANGE-3/3(9.9, 17.1)	-0.554
2	FORECASTYOYPCTCHANGE-2/3(2.7, 9.9)	-0.011

ВКЛЮЧИТЬ фильтр по кл.шкале ВЫКЛЮЧИТЬ фильтр по кл.шкале

ВКЛЮЧИТЬ фильтр по кл.шкале ВЫКЛЮЧИТЬ фильтр по кл.шкале

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 SWOT-диаграмма

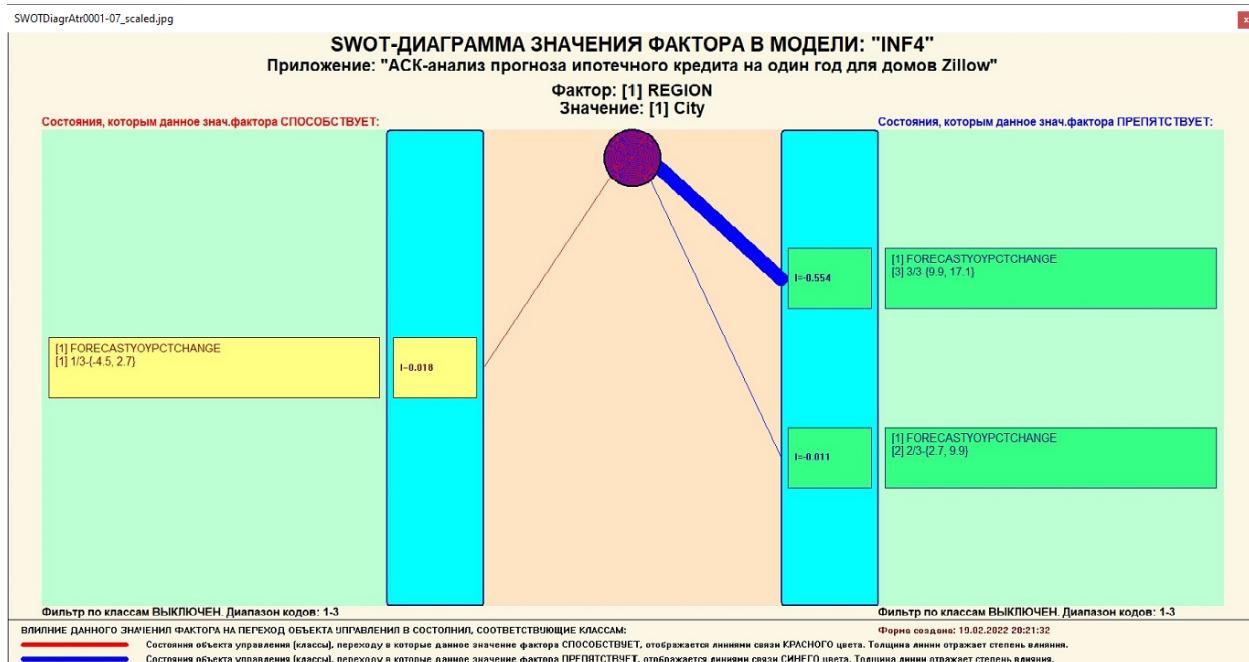


Рисунок 25 – Количественный SWOT-анализ значений в модели PRC2

ЗАКЛЮЧЕНИЕ

Целью работы было провести автоматизированный системно-когнитивный анализ прогнозов ипотечного кредита на один год для домов Zillow на основе данных портала Kaggle.

Для этого были изучены методы формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами. Построение моделей было осуществлено с помощью системы искусственного интеллекта Aidos-X, наиболее достоверной моделью оказалась модель INF4, точность модели составила 0.929.

ACK-анализ, использованный в данной работе, позволяет:

- сформировать обобщенные лингвистические образы классов (семантические ядра) на основе фрагментов или примеров относящихся к ним текстов на любом языке;
- количественно сравнить лингвистический образ конкретного героя, или описание объекта, процесса с обобщенными лингвистическими образами групп (классов);
- сравнить обобщенные лингвистические образы классов друг с другом и создавать их кластеры и конструкты;
- исследовать моделируемую предметную область путем исследования лингвистической системно-когнитивной модели.

Все это можно делать для любого естественного или искусственного языка или системы кодирования.

СПИСОК ЛИТЕРАТУРЫ

1. Луценко Е.В. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос» и системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №02(116). С. 1 – 60. – IDA [article ID]: 1161602001. – Режим доступа: <http://ej.kubagro.ru/2016/02/pdf/01.pdf>, 3,75 у.п.л.3.
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: КубГАУ, 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. 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.
4. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: <http://lc.kubagro.ru/>, свободный. - Загл. с экрана. Яз.рус.
5. Луценко Е.В. Количественная оценка степени манипулирования индексом Хирша и его модификация, устойчивая к манипулированию / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №07(121). С. 202 – 234. – IDA [article ID]: 1211607005. – Режим доступа: <http://ej.kubagro.ru/2016/07/pdf/05.pdf>, 2,062 у.п.л.. <http://dx.doi.org/10.21515/1990-4665-121-0057>.
6. 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.
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. Синтез семантических ядер научных специальностей ВАК РФ и автоматическая классификации статей по научным специальностям с применением АСК-анализа и интеллектуальной системы «Эйдос» (на примере Научного журнала КубГАУ и его научных специальностей: механизации, агрономии и ветеринарии) / Е.В. Луценко, Н.В. Андрафанова, Н.В. Потапова // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2019. – №01(145). С. 31 – 102. – IDA [article ID]: 1451901033. – Режим доступа: <http://ej.kubagro.ru/2019/01/pdf/33.pdf>, 4,5 у.п.л.