

**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: кадровая аналитика – смена профессии специалистов
на основе данных с портала Kaggle»

Выполнила студентка группы: ИТз1941 Позднякова Екатерина Геннадьевна

Допущен к защите _____

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

Защищен 10.04.2021
(дата)

Оценка отлично

Краснодар

2021

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**

Студентки: Поздняковой Екатерины Геннадьевны
курса 2 заочной формы обучения группы ИТз1941
направления подготовки: 09.04.02 «Информационные системы и технологии»
Наименование темы: «АСК-анализ: кадровая аналитика – смена профессии
специалистов 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.)

« » 2021 г.

РЕФЕРАТ

Курсовая работа содержит: 31 страница, 30 рисунков, 10 литературных источников.

Ключевые слова: СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, AIDOS-X, КОГНИТИВНЫЙ АНАЛИЗ, МОДЕЛИ, НЕЛОКАЛЬНЫЕ НЕЙРОНОВЫЕ СЕТИ, НЕЙРОНЫ.

Целью работы является проведение АСК-анализа кадровой аналитики по смене профессии специалистов на основе данных с 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	7
1.3 Ввод выборки в систему Aidos-X	9
1.4 Синтез и верификация статистических и интеллектуальных моделей	11
1.5 Виды моделей системы Aidos-X.....	13
1.6 Результаты верификации моделей	15
2 РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	19
2.1 Решение задач прогнозирования.....	19
2.2 SWOT и PEST матрицы и диаграммы.....	22
2.3 Исследование моделируемой предметной области путем исследования ее модели ...	23
2.3.1 Когнитивные диаграммы классов.....	23
2.3.2 Агломеративная когнитивная класстеризация классов.....	24
2.3.3 Когнитивные диаграммы значений факторов.....	24
2.3.4 Агломеративная когнитивная кластеризация значений факторов.....	25
2.3.5 Нелокальные нейроны и нелокальные нейронные сети.....	28
2.3.6 3D-интегральные когнитивные карты.....	29
2.3.7 Когнитивные функции.....	30
ЗАКЛЮЧЕНИЕ	34
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ	36

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 мировом счастье 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 "HR Analytics: Job Change of Data Scientists" 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 актуальным исходным текстом;
- 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;

- мультязычная поддержка интерфейса;
- 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 её в знания и решение задач классификации, поддержки принятия решений и исследования предметной области, путем исследования её системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходов, у многих из которых нет никаких аналогов в других системах;
- 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 www.kaggle.com возьмем набор данных «HR Analytics: Job Change of Data Scientists» <https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists>

Файл данных в формате csv содержит следующие данные:

- enrollee_id: Уникальный идентификатор кандидата
- city: Код города
- city_development_index: Индекс развития города
- gender: Пол кандидата
- relevent_experience: Соответствующий опыт кандидата
- enrolled_university: Тип зачисленного на университетский курс, если таковой имеется
- education_level: Уровень образования кандидата
- major_discipline: Обучение основной дисциплине кандидата
- experience: Общий стаж кандидата в годах
- company_size: Количество сотрудников в компании текущего работодателя
- company_type: Тип текущего работодателя
- lastnewjob: Разница в годах между предыдущей работой и текущей работой
- training_hours: Часы обучения завершены
- target: No – Не рассматривает смену работы, Yes – Рассматривает возможность работы в компании

Данный набор данных позволяет прогнозировать вероятность того, что кандидат, пришедший в компанию, будет работать в ней в дальнейшем после прохождения практики, интерпретировать модель таким образом, чтобы показать, какие особенности влияют на решение кандидата.

Компания хочет знать, кто из этих кандидатов действительно хочет работать в компании после обучения или ищет новую работу, потому что это помогает снизить затраты и время, а также качество обучения или планирование курсов и категоризацию кандидатов. Информация, связанная с демографическими данными, образованием, опытом, поступает от кандидатов при регистрации и зачислении.

Далее файл был конвертирован из формата csv в xls и отредактирован. Результат представлен на рисунке 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	enrollee_id	city	city_develop ment_index	gender	relevent_experience	enrolled_univ ersity	education_le vel	major_disci pline	experience	company_size	company_typ e	last_ne w_job	training_ hours	target
1	8949	city_103	0,920	Male	Has relevent experience	no_enrollmen	Graduate	STEM	>20			1	36	Yes
2	29725	city_40	0,776	Male	No relevent experience	no_enrollmen	Graduate	STEM	15	50-99	Pvt Ltd	>4	47	No
3	11561	city_21	0,624		No relevent experience	Full time cours	Graduate	STEM	5			never	83	No
4	33241	city_115	0,789		No relevent experience		Graduate	Business Deg	<1		Pvt Ltd	never	52	Yes
5	666	city_162	0,767	Male	Has relevent experience	no_enrollmen	Masters	STEM	>20	50-99	funded Startu	4	8	No
6	21651	city_176	0,764		Has relevent experience	Part time cours	Graduate	STEM	11			1	24	Yes
7	28806	city_160	0,920	Male	Has relevent experience	no_enrollmen	High School		5	50-99	funded Startu	1	24	No
8	402	city_46	0,762	Male	Has relevent experience	no_enrollmen	Graduate	STEM	13	<10	Pvt Ltd	>4	18	Yes
9	27107	city_103	0,920	Male	Has relevent experience	no_enrollmen	Graduate	STEM	7	50-99	Pvt Ltd	1	46	Yes
10	699	city_103	0,920		Has relevent experience	no_enrollmen	Graduate	STEM	17	10000+	Pvt Ltd	>4	123	No
11	29452	city_21	0,624		No relevent experience	Full time cours	High School		2			never	32	Yes
12	23853	city_103	0,920	Male	Has relevent experience	no_enrollmen	Graduate	STEM	5	5000-9999	Pvt Ltd	1	108	No
13	25619	city_61	0,913	Male	Has relevent experience	no_enrollmen	Graduate	STEM	>20	1000-4999	Pvt Ltd	3	23	No
14	5826	city_21	0,624	Male	No relevent experience				2			never	24	No
15	8722	city_21	0,624		No relevent experience	Full time cours	High School		5			never	26	No
16	6588	city_114	0,926	Male	Has relevent experience	no_enrollmen	Graduate	STEM	16	10/49	Pvt Ltd	>4	18	No
17	4167	city_103	0,920		Has relevent experience	no_enrollmen	Graduate	STEM	1	50-99	Pvt Ltd	never	106	No
18	5764	city_21	0,624		Has relevent experience	no_enrollmen	Graduate	STEM	2	5000-9999	Pvt Ltd	2	7	No
19	2156	city_21	0,624		Has relevent experience	no_enrollmen	Graduate	STEM	7	10000+	Pvt Ltd	never	23	Yes
20	11399	city_13	0,827	Female	Has relevent experience	no_enrollmen	Graduate	Arts	4			1	132	Yes
21	31972	city_159	0,843	Male	Has relevent experience	no_enrollmen	Masters	STEM	11	100-500	Pvt Ltd	1	68	No
22	19061	city_114	0,926	Male	Has relevent experience	no_enrollmen	Masters	STEM	11	100-500	Pvt Ltd	2	50	No
23	6491	city_102	0,804		Has relevent experience	no_enrollmen	Masters	STEM	10			1	48	No
24	7041	city_40	0,776	Male	Has relevent experience	no_enrollmen	Graduate	Humanities	<1	1000-4999	Pvt Ltd	1	65	No
25	22767	city_21	0,624		Has relevent experience	no_enrollmen	Graduate	STEM	5	1000-4999	Pvt Ltd	1	13	No
26	14505	city_67	0,855		No relevent experience	no_enrollmen	High School		4			never	22	No
27	17139	city_21	0,624		Has relevent experience	Part time cours	Graduate	STEM	14	500-999	Pvt Ltd	1	148	Yes

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 переименуем ее 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).

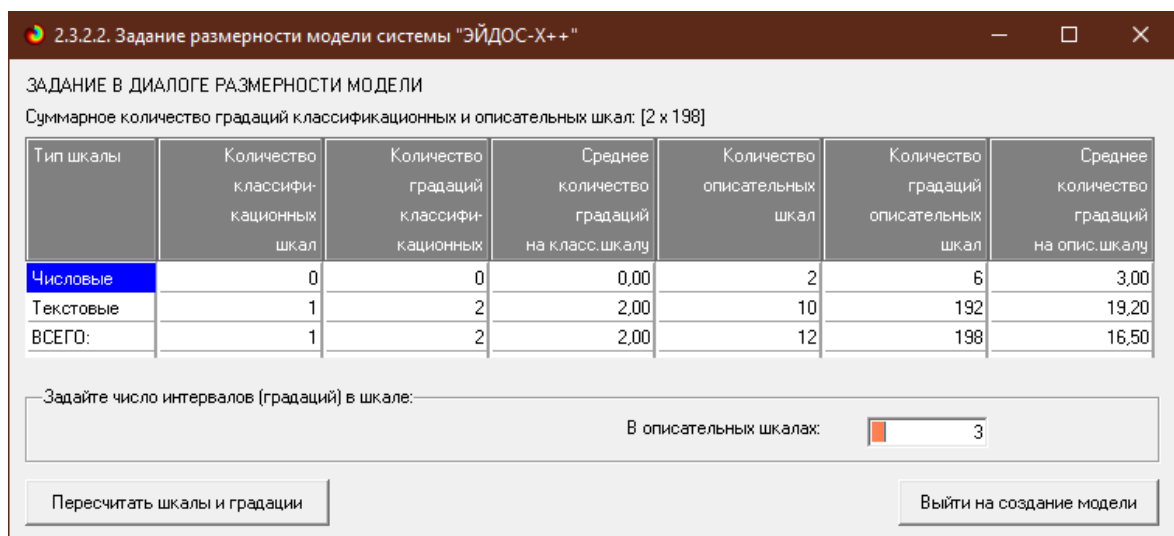
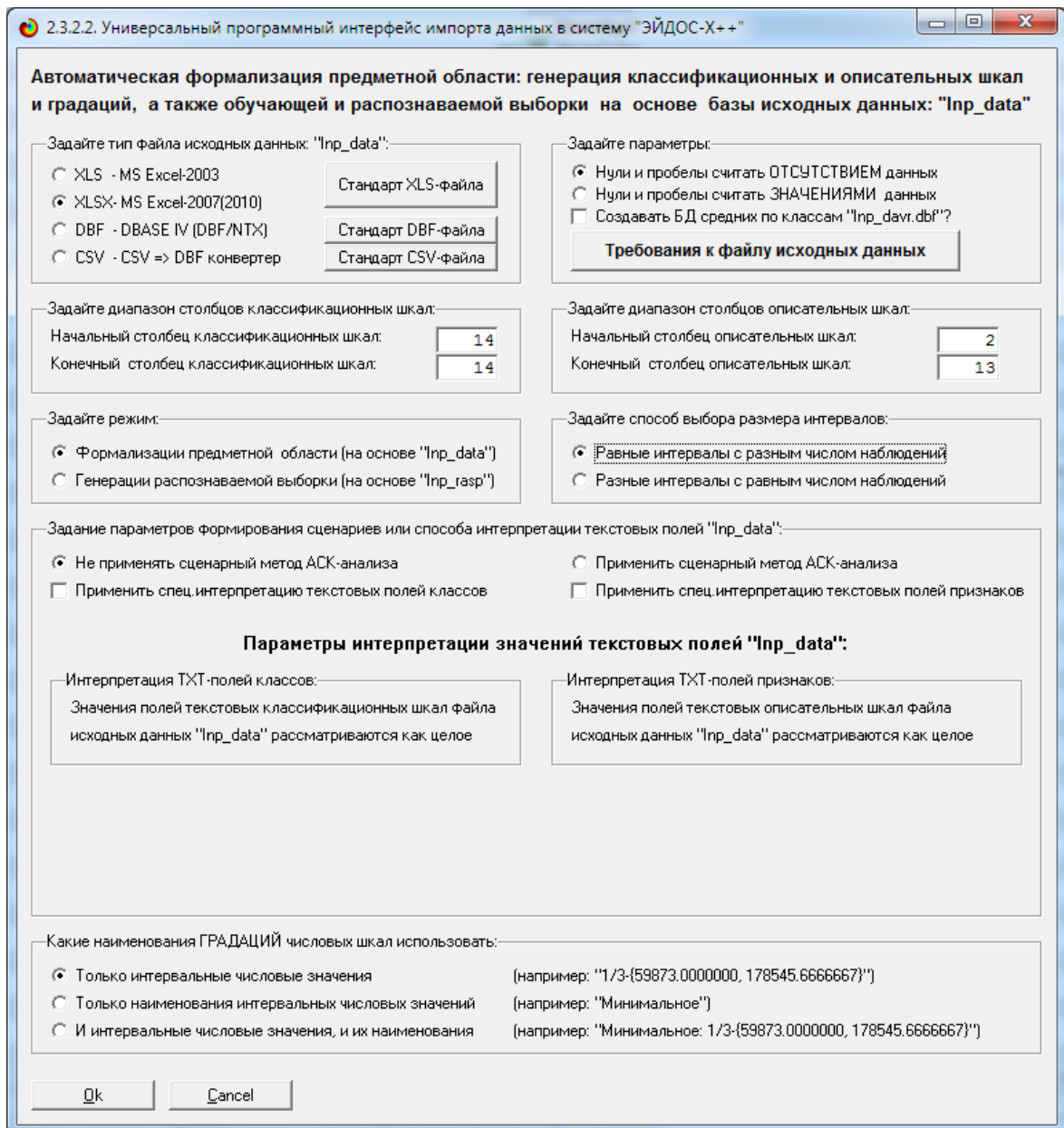


Рисунок 2 - Задание размерности модели системы Aidos-X

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

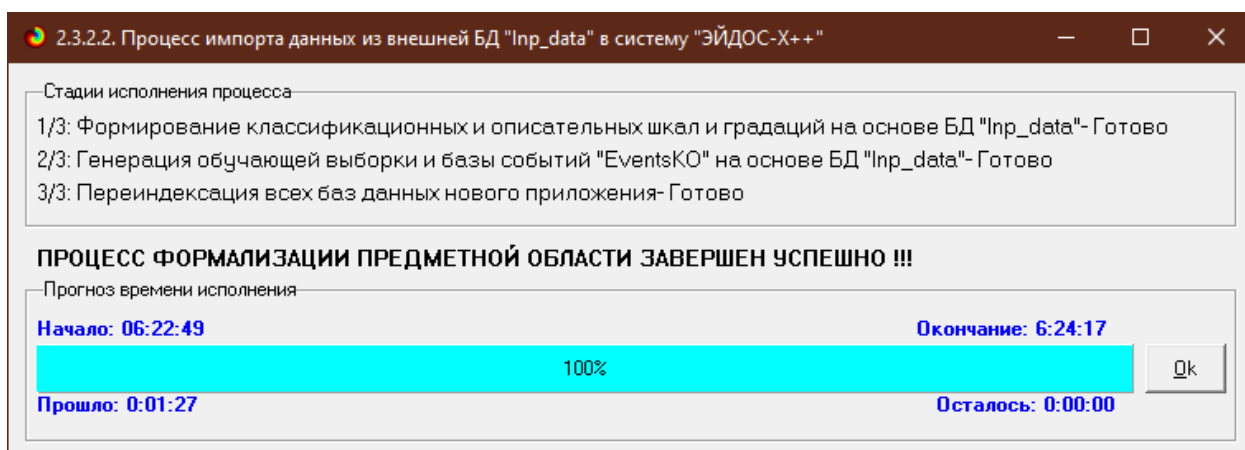


Рисунок 3 – Процесс импорта данных

Так как предварительно числовые шкалы были разбиты на интервалы, то перерасчет 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).

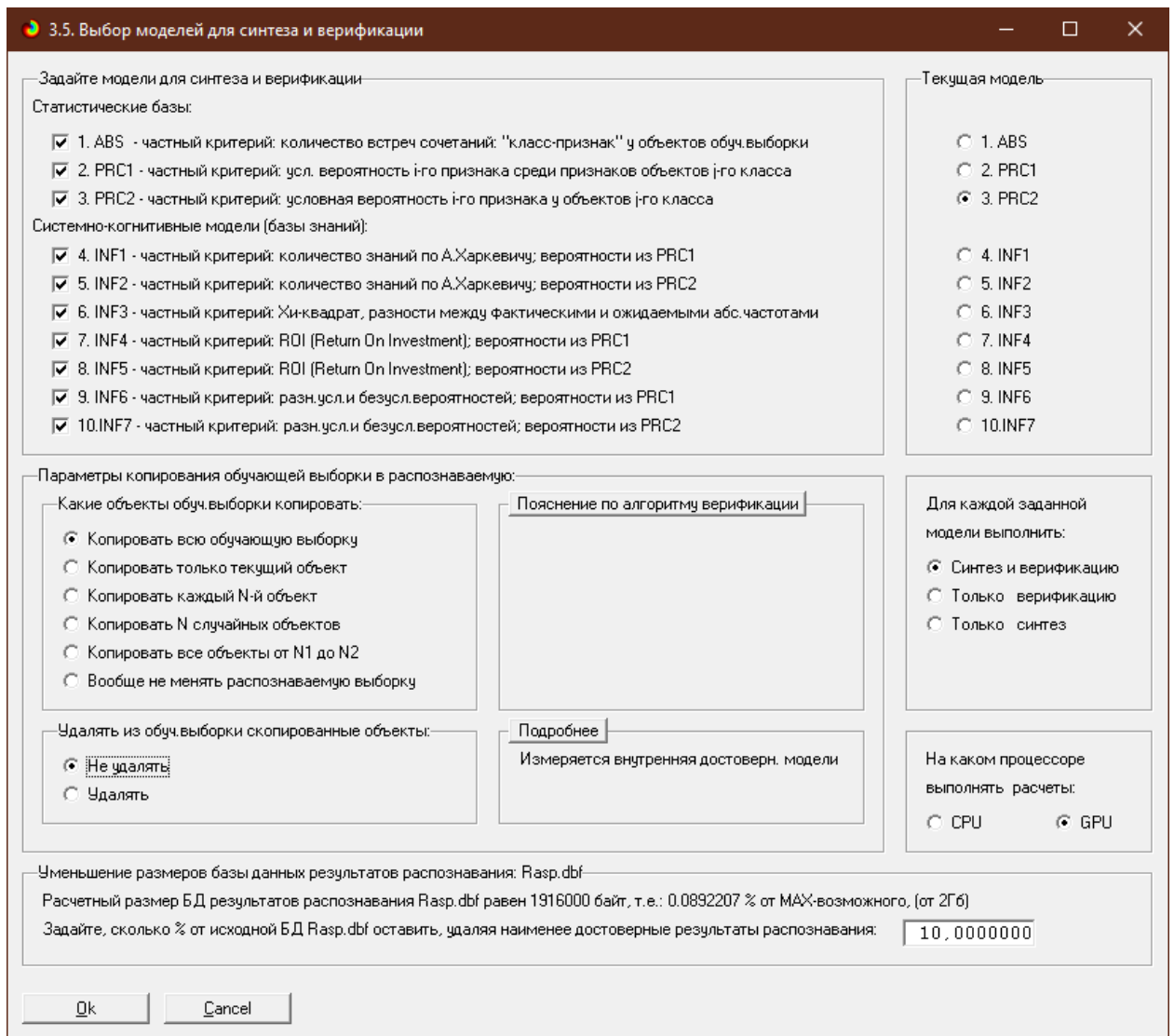


Рисунок 4 – Выбор моделей для синтеза и верификации

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.

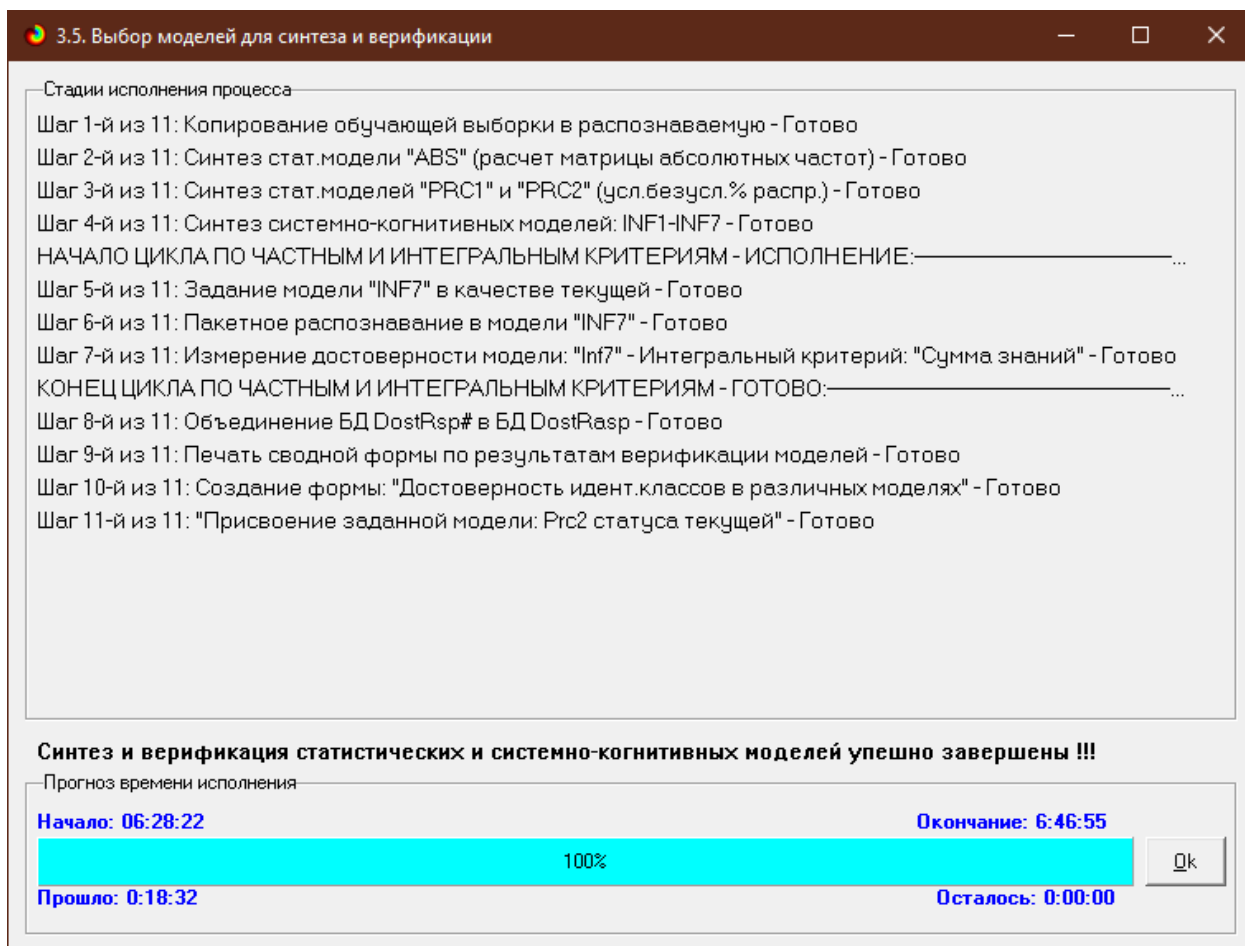


Рисунок 5 – Процесс исполнения режима 3.5

Интересно заметить (см. рисунок 5), что синтез и верификация всех 10 моделей на данной задаче заняли 18 минут 32 секунды. Далее перейдем непосредственно к выбору наиболее достоверной модели.

1.5 Виды моделей системы 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.выборки"

Код признака	Наименование описательной шкалы и градации	1. TARGET NO	2. TARGET YES	Сумма	Среднее	Средн. квадрат. откл.
121	city_97					
122	city_98					
123	city_99					
124	CITY_DEVELOPMENT_INDEX-1/3-{0.4480000, 0.6150000...}	305	420	725	362.50	81.32
125	CITY_DEVELOPMENT_INDEX-2/3-{0.6150000, 0.7820000...}	2823	2215	5038	2519.00	429.92
126	CITY_DEVELOPMENT_INDEX-3/3-{0.7820000, 0.9490000...}	11182	2134	13316	6658.00	6397.90
127	GENDER-Female	912	326	1238	619.00	414.36
128	GENDER-Male	10209	3012	13221	6610.50	5089.05
129	GENDER-Other	141	50	191	95.50	64.35
130	RELEVANT_EXPERIENCE-Has relevant experience	10831	2961	13792	6896.00	5564.93
131	RELEVANT_EXPERIENCE-No relevant experience	3550	1816	5366	2683.00	1226.12
132	ENROLLED_UNIVERSITY-Full time course	2326	1431	3757	1878.50	632.86
133	ENROLLED_UNIVERSITY-no_enrollment	10896	2921	13817	6908.50	5639.18
134	ENROLLED_UNIVERSITY-Part time course	896	302	1198	599.00	420.02
135	EDUCATION_LEVEL-Graduate	8353	3245	11598	5799.00	3611.90
136	EDUCATION_LEVEL-High School	1623	394	2017	1008.50	869.03
137	EDUCATION_LEVEL-Masters	3426	935	4361	2180.50	1761.40
138	EDUCATION_LEVEL-Phd	356	58	414	207.00	210.72
139	EDUCATION_LEVEL-Primary School	267	41	308	154.00	159.81
140	MAJOR_DISCIPLINE-Arts	200	53	253	126.50	103.94
141	MAJOR_DISCIPLINE-Business Degree	241	86	327	163.50	109.60
142	MAJOR_DISCIPLINE-Humanities	528	141	669	334.50	273.65
143	MAJOR_DISCIPLINE-No Major	168	55	223	111.50	79.90
144	MAJOR_DISCIPLINE-Other	279	102	381	190.50	125.16
145	MAJOR_DISCIPLINE-STEM	10701	3791	14492	7246.00	4886.11
146	EXPERIENCE- 1.0000000					

Рисунок 6 – Матрица абсолютных частот (фрагмент)

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

Код признака	Наименование описательной шкалы и градации	1. TARGET NO	2. TARGET YES	Сумма	Среднее	Средн. квадрат. откл.
169	COMPANY_SIZE-10/49	2.756	-2.756	0.000	0.000	3.898
170	COMPANY_SIZE-100-500	191.057	-191.057			270.196
171	COMPANY_SIZE-1000-4999	113.047	-113.047			159.873
172	COMPANY_SIZE-10000+	90.935	-90.935	0.000	0.000	128.602
173	COMPANY_SIZE-50-99	181.750	-181.750			257.034
174	COMPANY_SIZE-500-999	54.734	-54.734	0.000	0.000	77.405
175	COMPANY_SIZE-5000-9999	30.715	-30.715	0.000	0.000	43.438
176	COMPANY_TYPE-Early Stage Startup	0.144	-0.144	0.000	0.000	0.204
177	COMPANY_TYPE-Funded Startup	95.964	-95.964			135.714
178	COMPANY_TYPE-NGO	25.814	-25.814			36.507
179	COMPANY_TYPE-Other	-0.477	0.477			0.674
180	COMPANY_TYPE-Public Sector	15.120	-15.120	0.000	0.000	21.384
181	COMPANY_TYPE-Pvt Ltd	539.144	-539.144	0.000	0.000	762.465
182	LAST_NEW_JOB- 1.0000000					
183	LAST_NEW_JOB- 2.0000000					
184	LAST_NEW_JOB- 3.0000000					
185	LAST_NEW_JOB- 4.0000000					
186	LAST_NEW_JOB->4	175.546	-175.546	0.000	0.000	248.259
187	LAST_NEW_JOB-never	-160.994	160.994	0.000	0.000	227.680
188	TRAINING_HOURS-1/3-{1.0000000, 112.6666667}	-274.331	274.331	0.000	0.000	387.963
189	TRAINING_HOURS-2/3-{112.6666667, 224.3333333}	-5.466	5.466			7.730
190	TRAINING_HOURS-3/3-{224.3333333, 336.0000000}	18.879	-18.879	0.000	0.000	26.699
	Сумма	0.000	0.000	0.000		
	Среднее	0.000	0.000			
	Среднеквадратичное отклонение	145.113	145.113			144.921

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.

Наименование модели и частного критерия	Интегральный критерий	Полнота модели	F-мера Ван Ризбергера	Сумма модул. уровней сход. истинно-полож. решений (STP)	Сумма модул. уровней сход. истинно-отриц. решений (STN)	Сумма модул. ложно-полож. решений (SFP)	Сумма модул. ложно-отрицат. решений (SFN)	S-Точность модели	S-Полнота модели	L1-мера проф. Е.В.Луценко
1. ABS - частный критерий: количество встреч сочетаний "х.лас...	Корреляция абс. частот с обр...	1.000	0.952	14255.460		1368.800		0.912	1.000	0.954
1. ABS - частный критерий: количество встреч сочетаний "х.лас...	Сумма абс. частот по признаку...	1.000	0.952	11456.011		623.583		0.948	1.000	0.974
2. PRC1 - частный критерий: усл. вероятность n-го признака сред...	Корреляция усл.отн. частот с о...	1.000	0.952	14255.458		1368.799		0.912	1.000	0.954
2. PRC1 - частный критерий: усл. вероятность n-го признака сред...	Сумма усл.отн. частот по призна...	1.000	0.952	13688.159		1316.546		0.912	1.000	0.954
3. PRC2 - частный критерий: усл. вероятность n-го признака...	Корреляция усл.отн. частот с о...	1.000	0.952	14255.460		1368.799		0.912	1.000	0.954
3. PRC2 - частный критерий: усл. вероятность n-го признака...	Сумма усл.отн. частот по призна...	1.000	0.952	13457.489		1244.935		0.915	1.000	0.956
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	0.656	0.779	2288.691	322.720	105.198	1321.062	0.956	0.634	0.762
4. INF1 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	0.640	0.768	1558.498	363.062	94.948	918.987	0.943	0.629	0.755
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Семантический резонанс зна...	0.677	0.796	3000.437	433.490	87.896	1256.053	0.972	0.705	0.817
5. INF2 - частный критерий: количество знаний по А.Харкевичу; в...	Сумма знаний	0.664	0.788	1662.831	518.784	57.973	1017.449	0.966	0.620	0.756
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Семантический резонанс зна...	0.702	0.811	4407.190	445.460	167.537	1636.990	0.963	0.729	0.830
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Сумма знаний	0.702	0.811	4816.247	485.756	175.933	1721.409	0.965	0.737	0.835
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.674	0.790	2698.817	268.800	128.559	1259.104	0.955	0.682	0.795
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.674	0.790	1519.166	206.391	111.378	525.125	0.932	0.743	0.827
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.716	0.821	3678.697	366.955	116.763	1126.266	0.969	0.766	0.855
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.725	0.827	1853.302	368.303	83.410	665.194	0.957	0.736	0.832
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	0.702	0.811	4407.191	445.460	167.537	1636.990	0.963	0.729	0.830
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	0.702	0.811	2303.620	398.331	142.740	840.432	0.942	0.733	0.824
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	0.743	0.840	7255.684	726.597	161.967	1585.166	0.978	0.821	0.893
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	0.746	0.842	2642.206	673.602	76.853	1177.807	0.972	0.692	0.808

Рисунок 8 – Оценка достоверности моделей

Наиболее достоверной в данном приложении оказалась модель ABS при интегральном критерии «Сумма абс. частот по признаку». При этом точность модели (F- мера Ван Ризбергера) составляет 0,952 при максимуме 1,000, а точность модели (L1-мера профессора Луценко) - 0,974 при максимуме 1,000. L1-мера профессора Луценко является более достоверной, по сравнению с F-мерой Ван Ризбергера.

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 Ризбергера и L-мера, представляющая собой ее нечеткое мультимодальное 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.

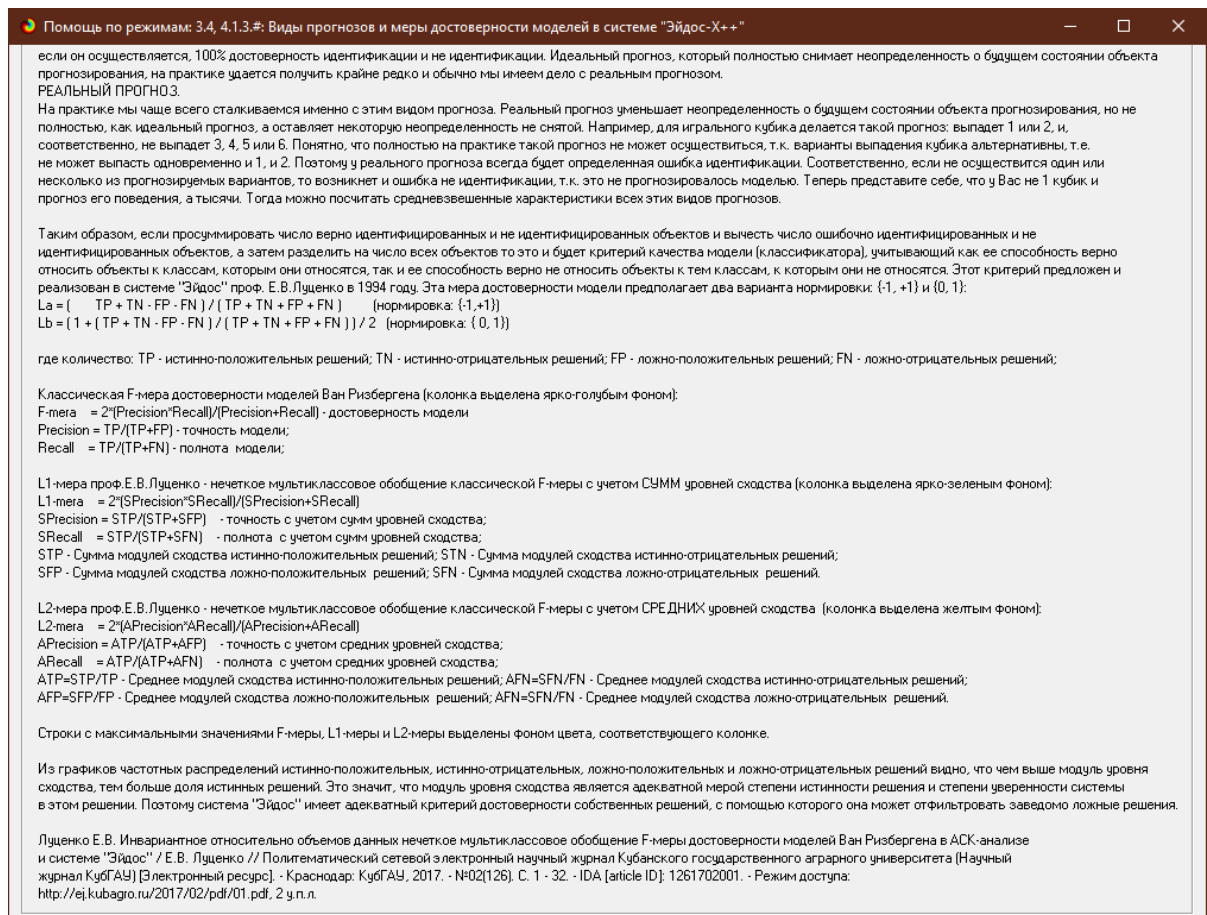


Рисунок 9 – Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

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

Левое распределение, большее по амплитуде включает истинно-отрицательные и ложно-положительные решения, а правое, меньшее по амплитуде, 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 положительных;
- при решении о непринадлежности больше ложных решений.

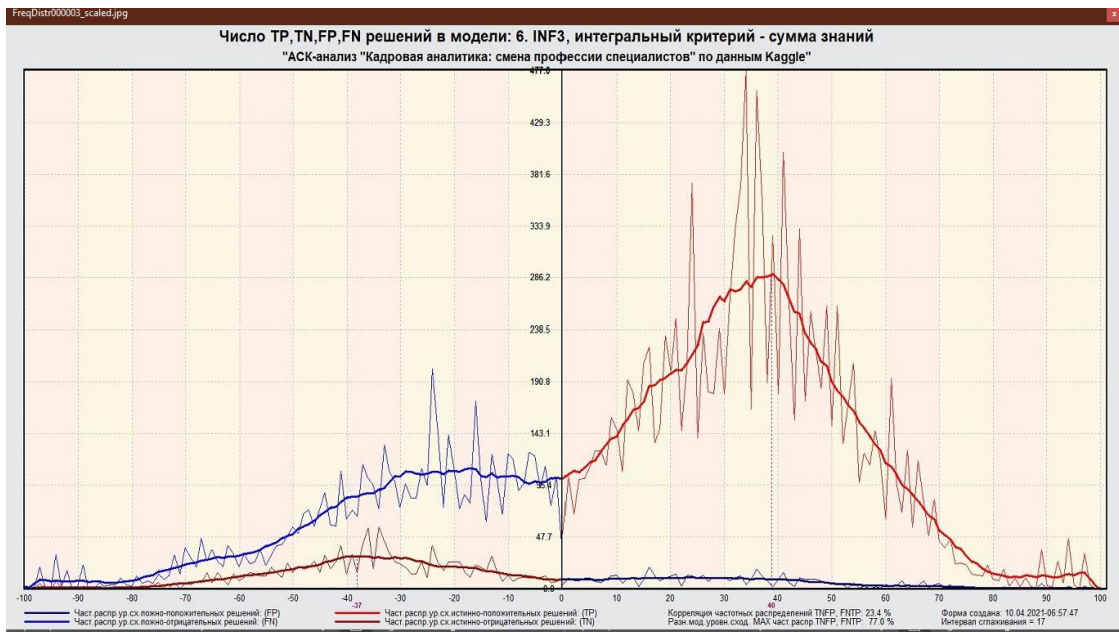


Рисунок 10 – Частотные распределения числа верно и ошибочно идентифицированных и не идентифицированных состояний объекта моделирования в зависимости от уровня сходства в модели INF3

На рисунке 11 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и не идентифицированных ситуаций в наиболее достоверной модели ABS.

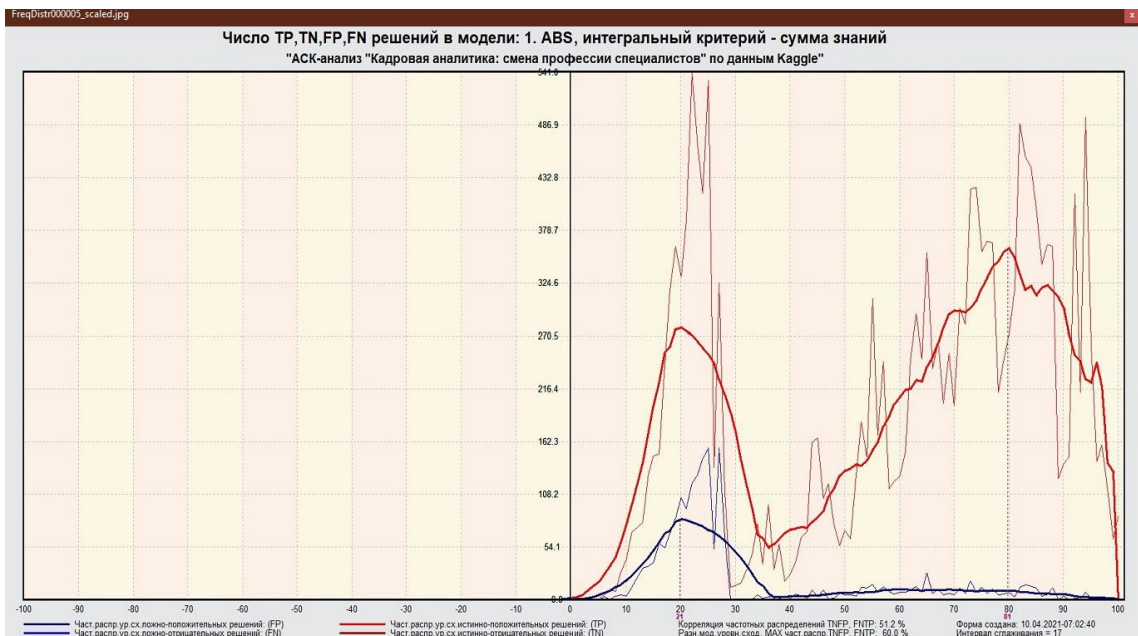


Рисунок 11 – Частотные распределения модели ABS

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

Из рисунка 12 видно, что модель PRC2 абсолютно четко показывает, что доля истинных решений гораздо выше, чем ложных начиная с 8%, т.е. подавляющее число решений является истинным.

Модель показывает высочайший уровень достоверности.

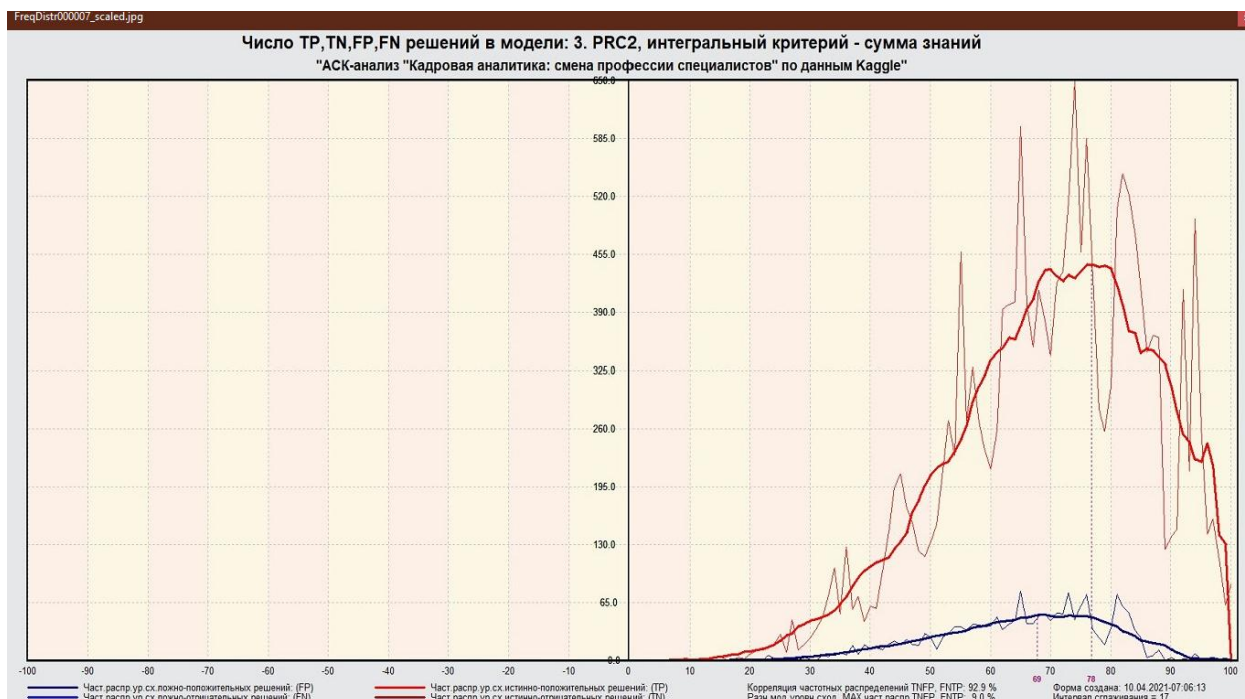


Рисунок 12 – Частотные распределения числа верно и ошибочно идентифицированных и не идентифицированных состояний объекта моделирования в зависимости от уровня сходства в модели PRC2

2 РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1 Решение задач прогнозирования

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

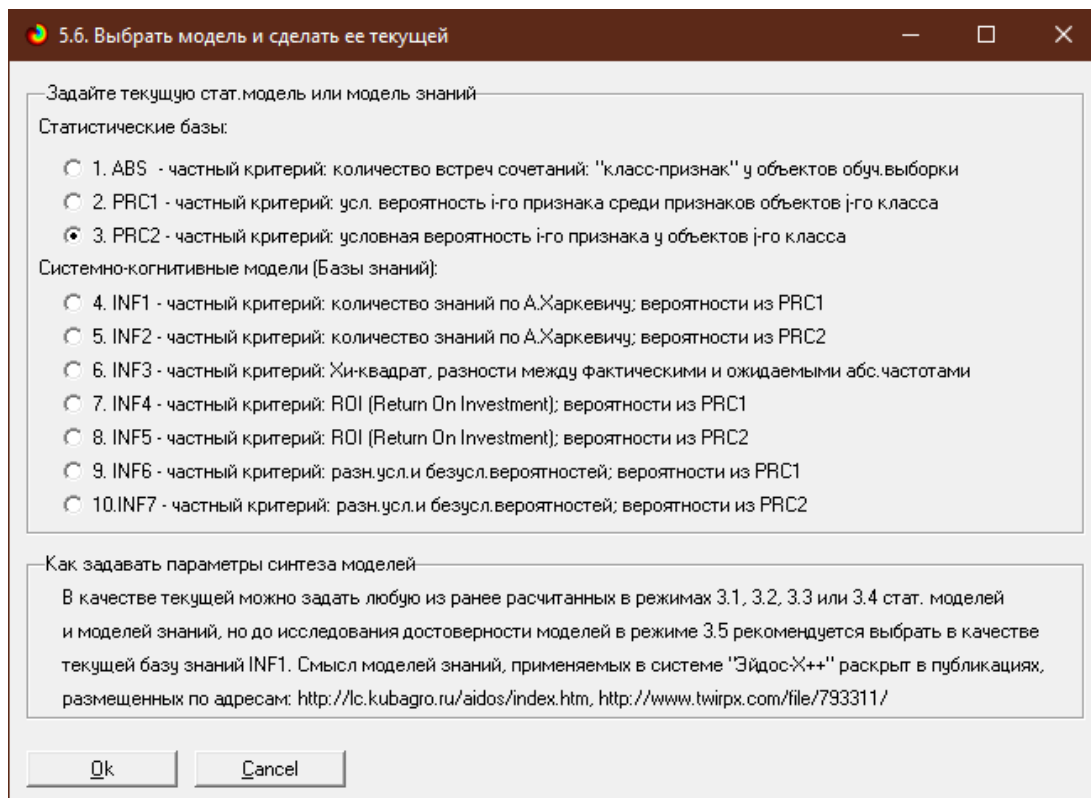


Рисунок 13 – Экранная форма режима задания модели в качестве текущей

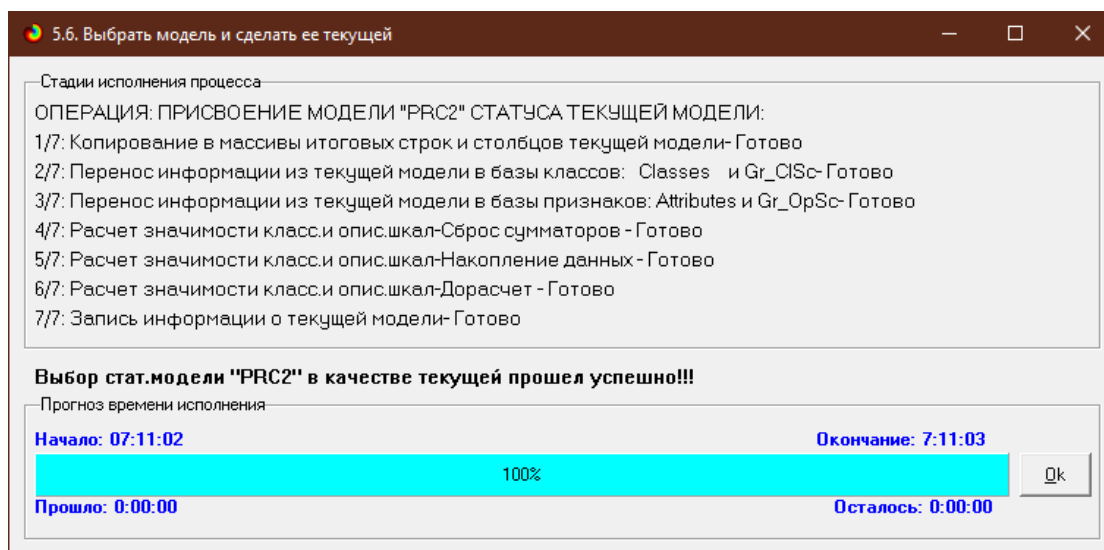


Рисунок 14 – Экранная форма режима задания модели в качестве текущей

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

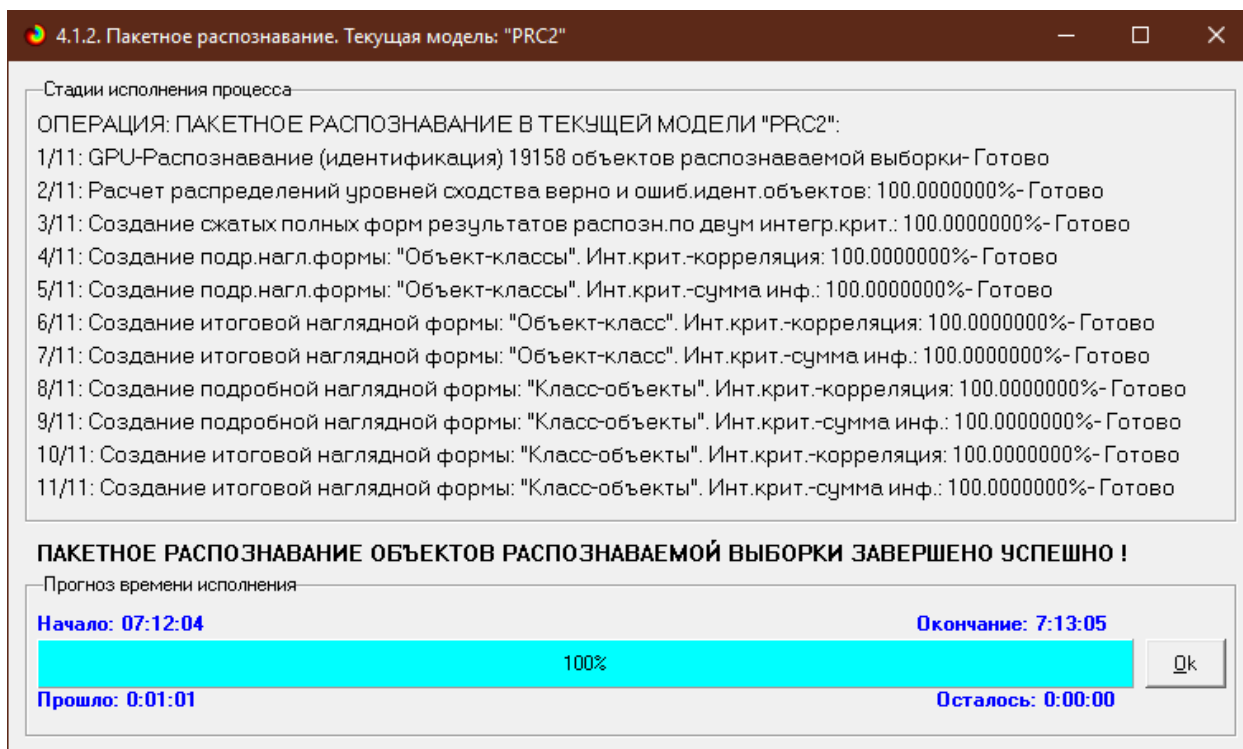


Рисунок 15 – Экранная форма режима пакетного распознавания в текущей модели

В результате пакетного распознавания в текущей модели создается ряд баз данных, которые визуализируются в выходных экранных 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».

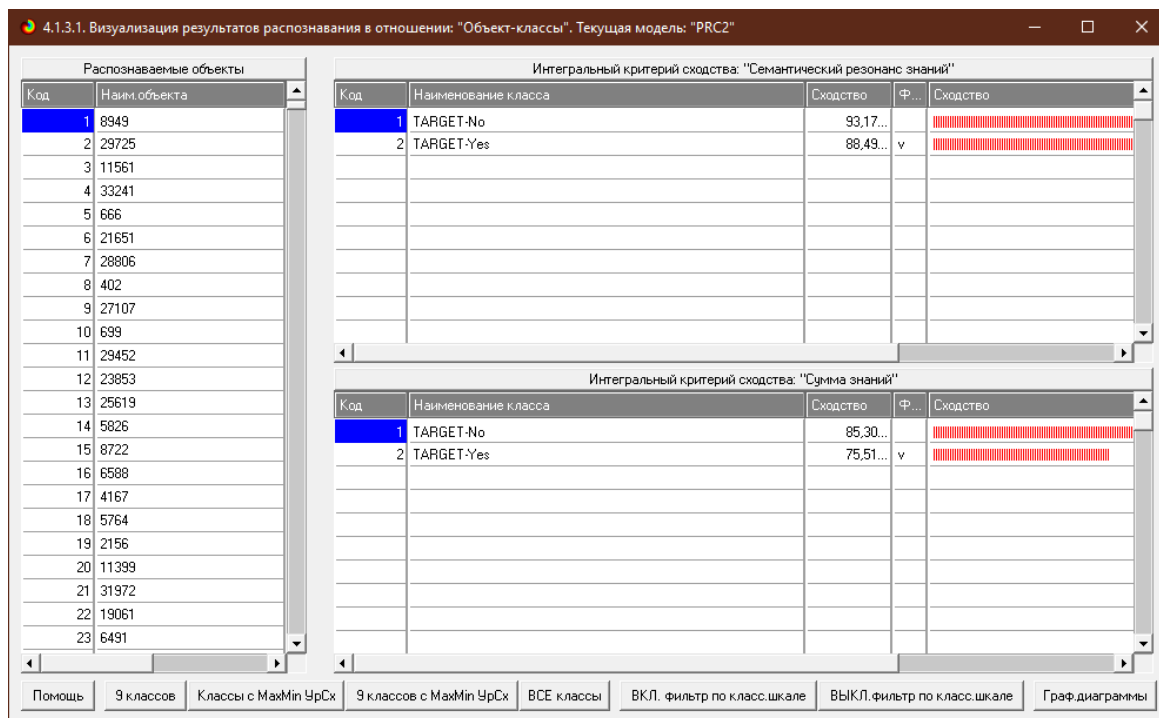


Рисунок 16. Пример идентификации объектов в модели PRC2

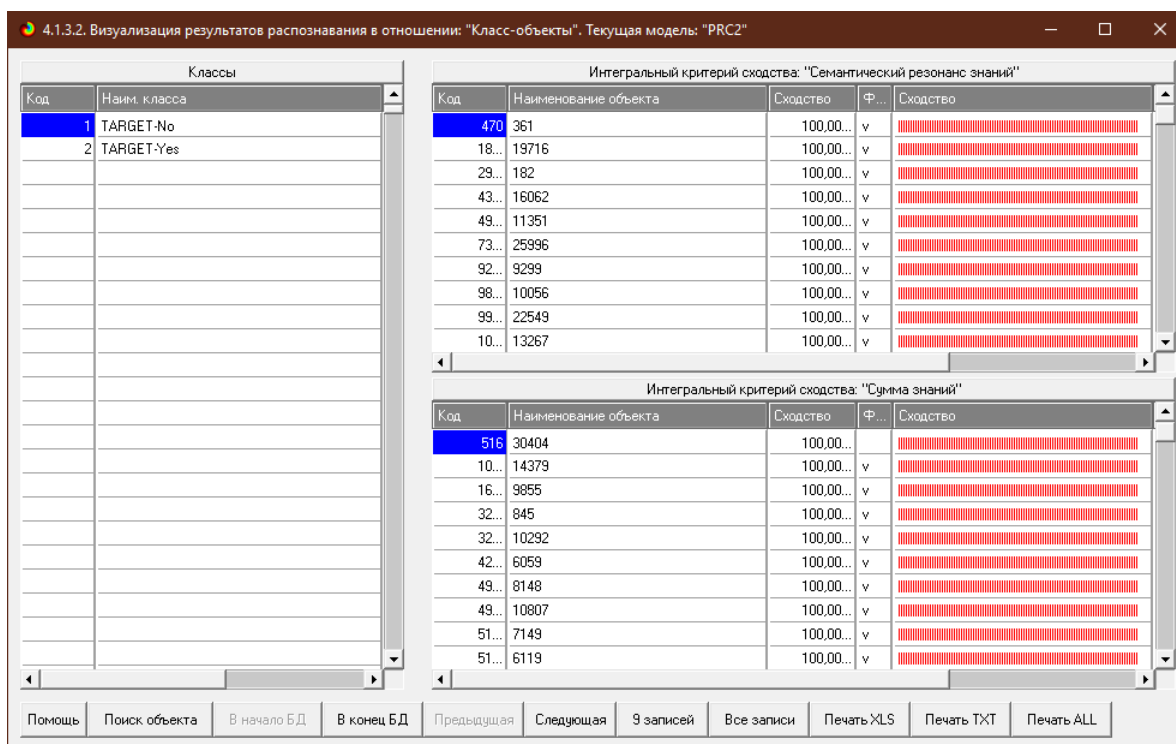


Рисунок 17. Пример идентификации объектов в модели PRC2

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 выбор кандидата (рисунок 18).

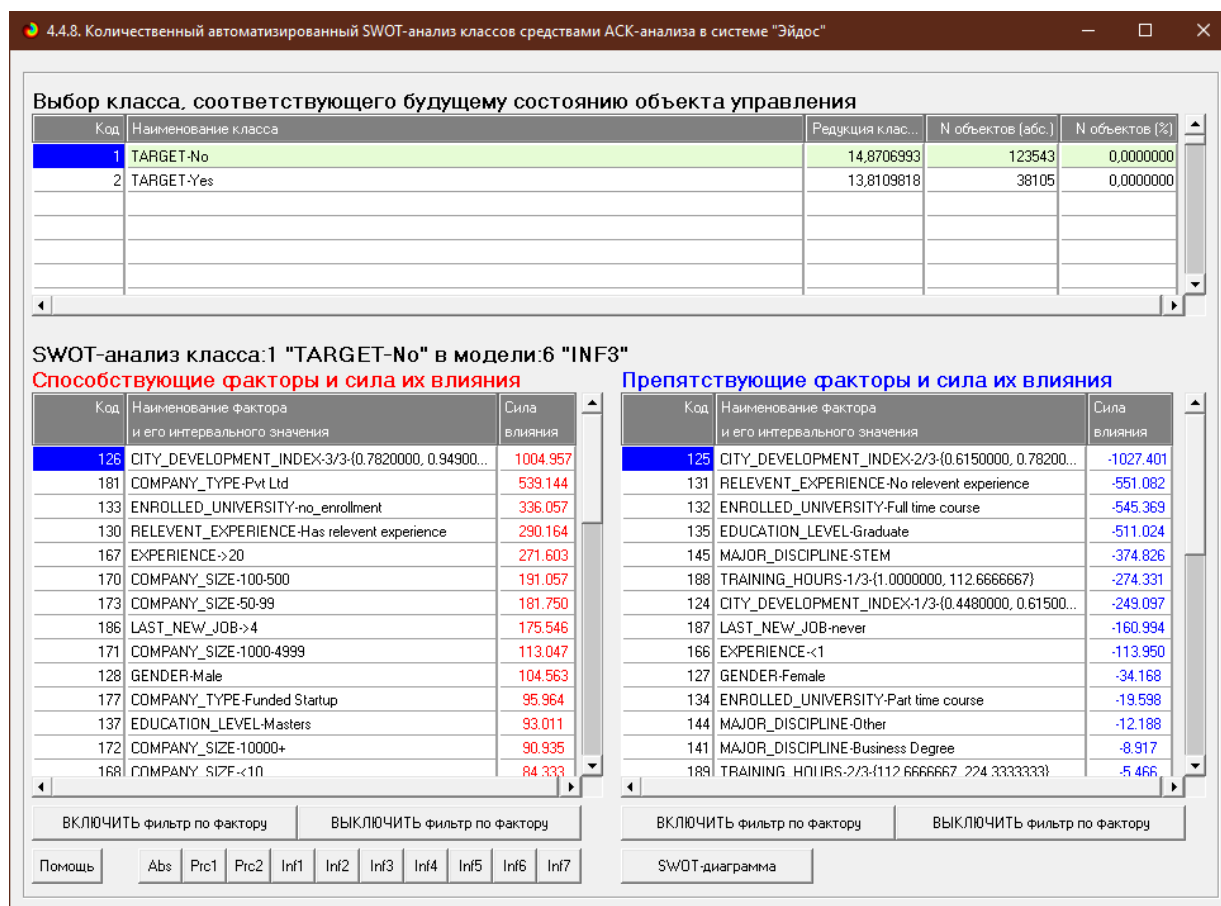


Рисунок 18 – Автоматизированное построение SWOT-матрицы

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

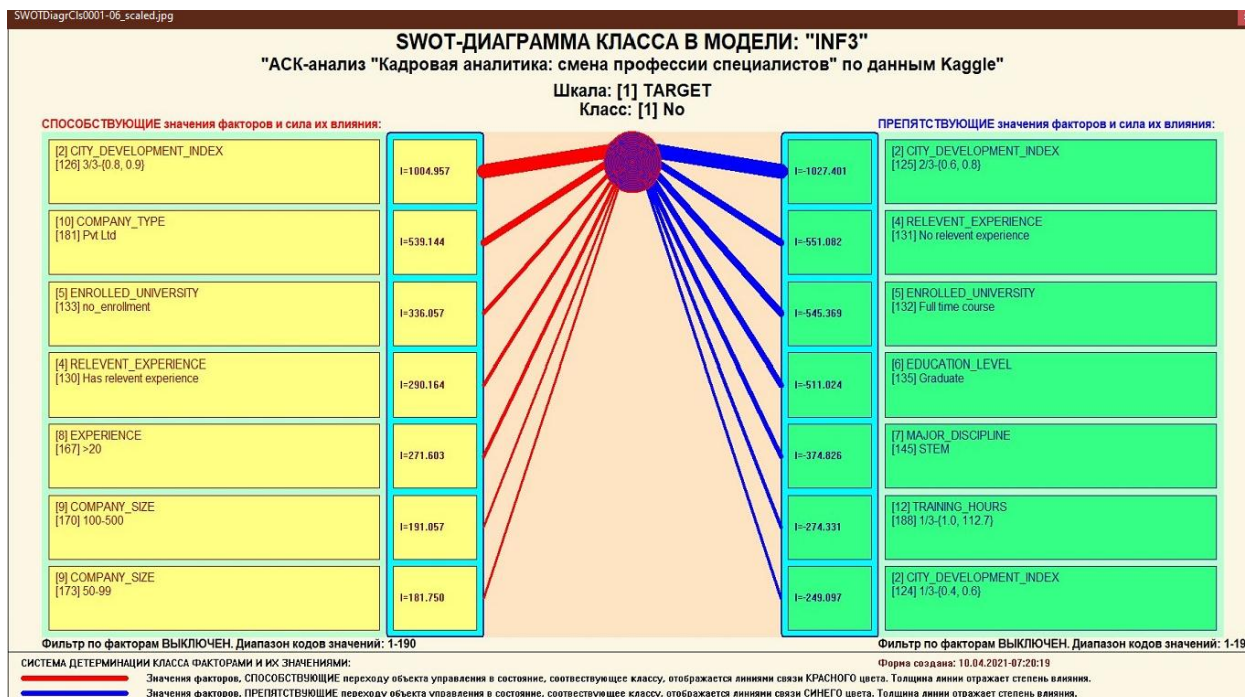


Рисунок 19 - Пример SWOT-матрицы в модели PRC2

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 17, показаны количественные оценки принятого кандидатом решения. 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 17). 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.

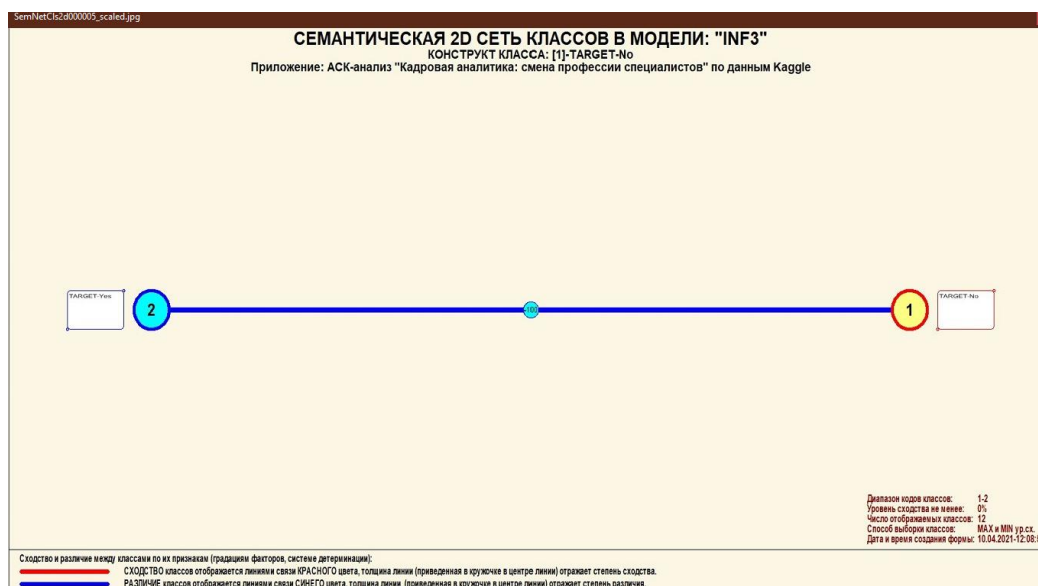


Рисунок 20 – Когнитивная диаграмма классов

2.3.2 Агломеративная когнитивная класстеризация классов

Построить агломеративную дендрограмму и соответственно график изменения межкластерных расстояний не удалось (рисунок 21).

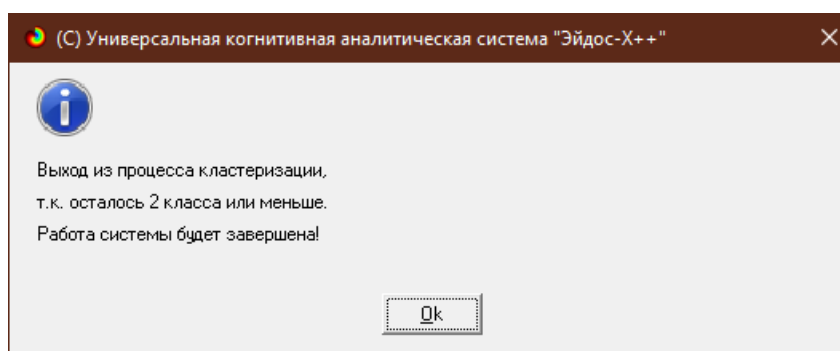


Рисунок 21 – Выход из процесса кластеризации

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).

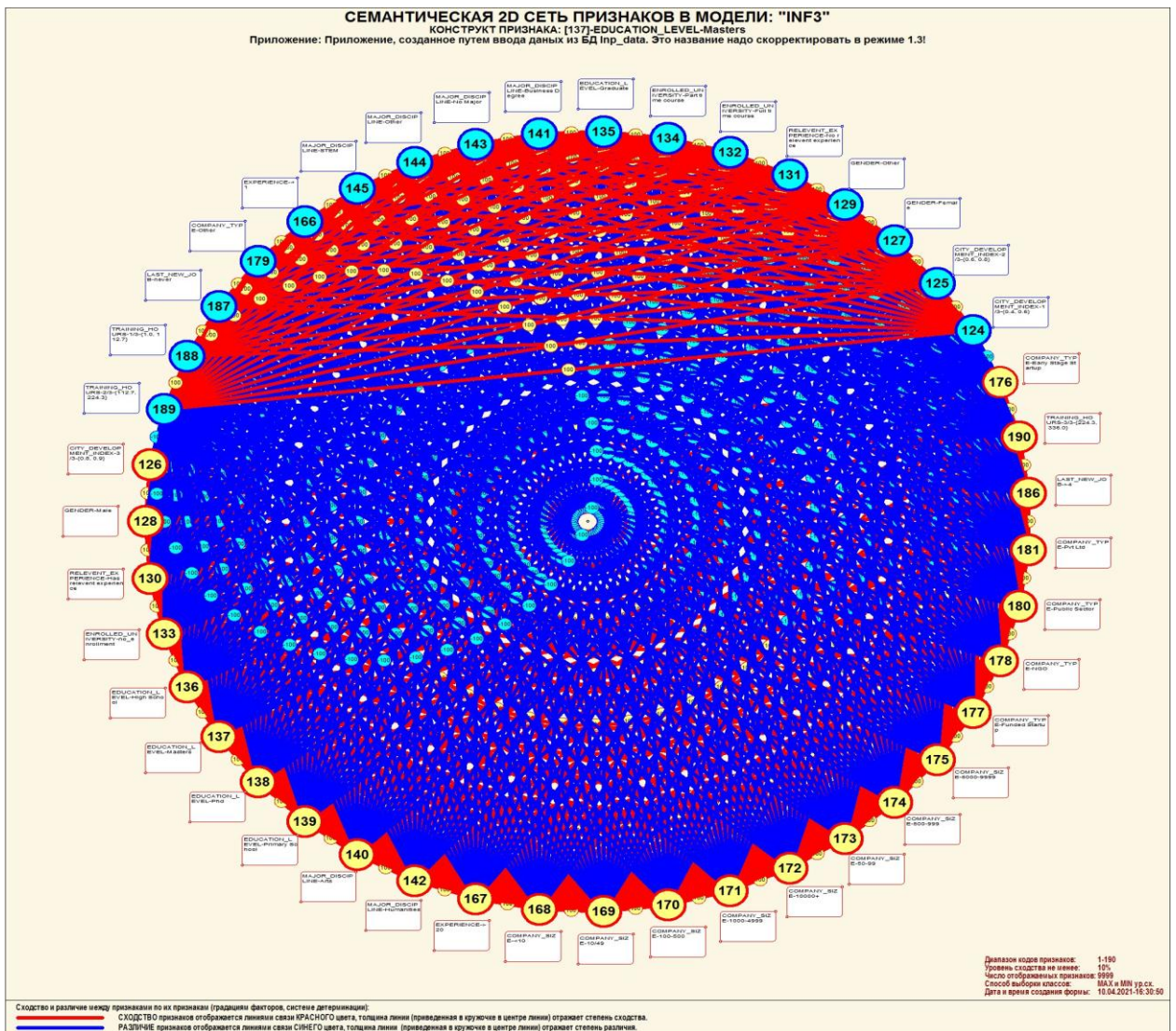


Рисунок 22 – Критерии, влияющие на выбор кандидата

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.

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 межкластерных расстояний значений признаков.

ДЕНДРОГРАММА КОГНИТИВНОЙ КЛАСТЕРИЗАЦИИ ПРИЗНАКОВ В МОДЕЛИ: "INF3"
 "Приложение, созданное путем ввода данных из БД Inp_data. Это название надо скорректировать в режиме 1.3!"

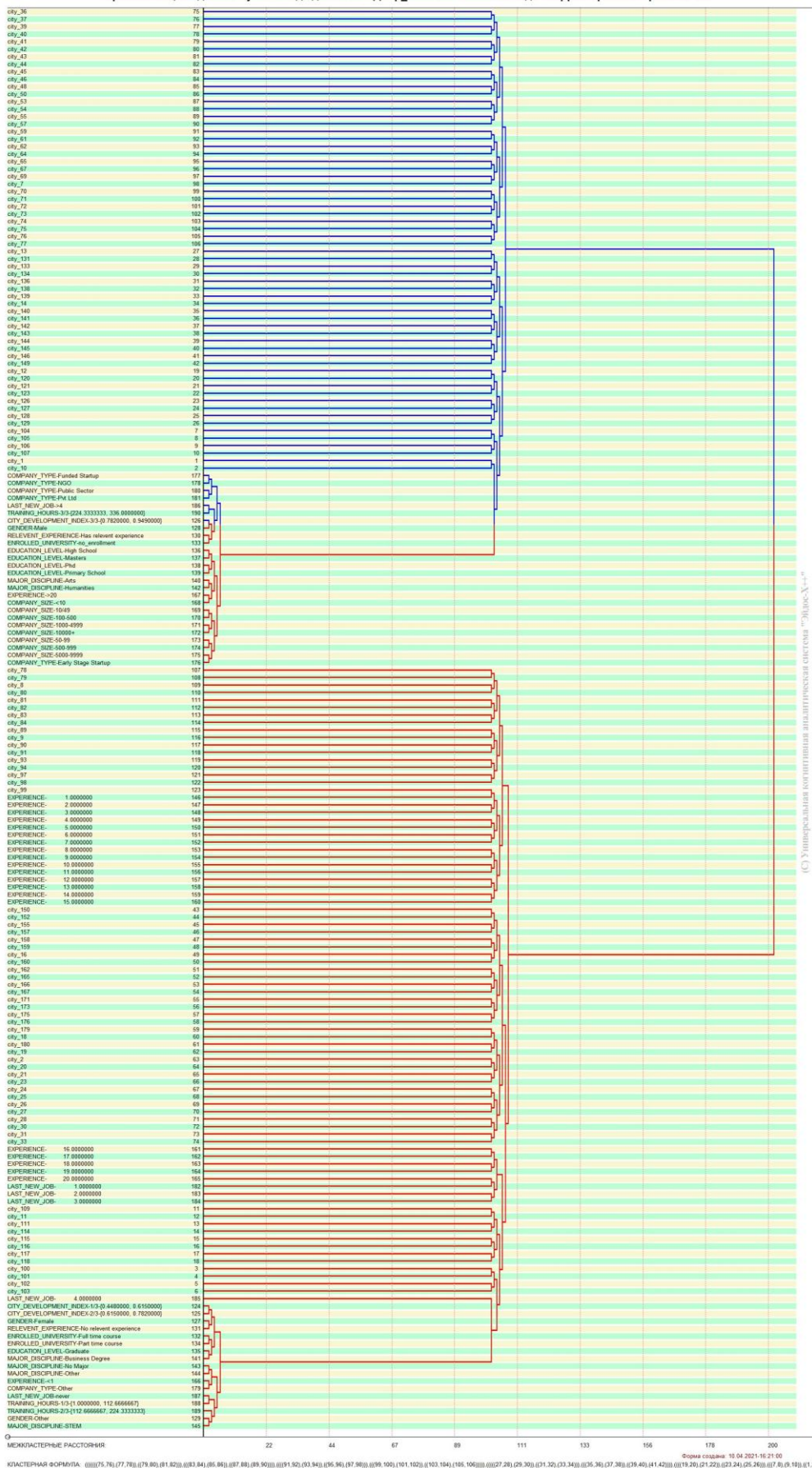


Рисунок 23 – Дендрограмма когнитивной кластеризации признаков

ИЗМЕНЕНИЕ МЕЖКЛАСТЕРНЫХ РАССТОЯНИЙ ПРИ КОГНИТИВНОЙ КЛАСТЕРИЗАЦИИ ПРИЗНАКОВ В МОДЕЛИ: "INF3"
 Приложение, созданное путем ввода данных из БД Inp_data. Это название надо скорректировать в режиме 1.3!

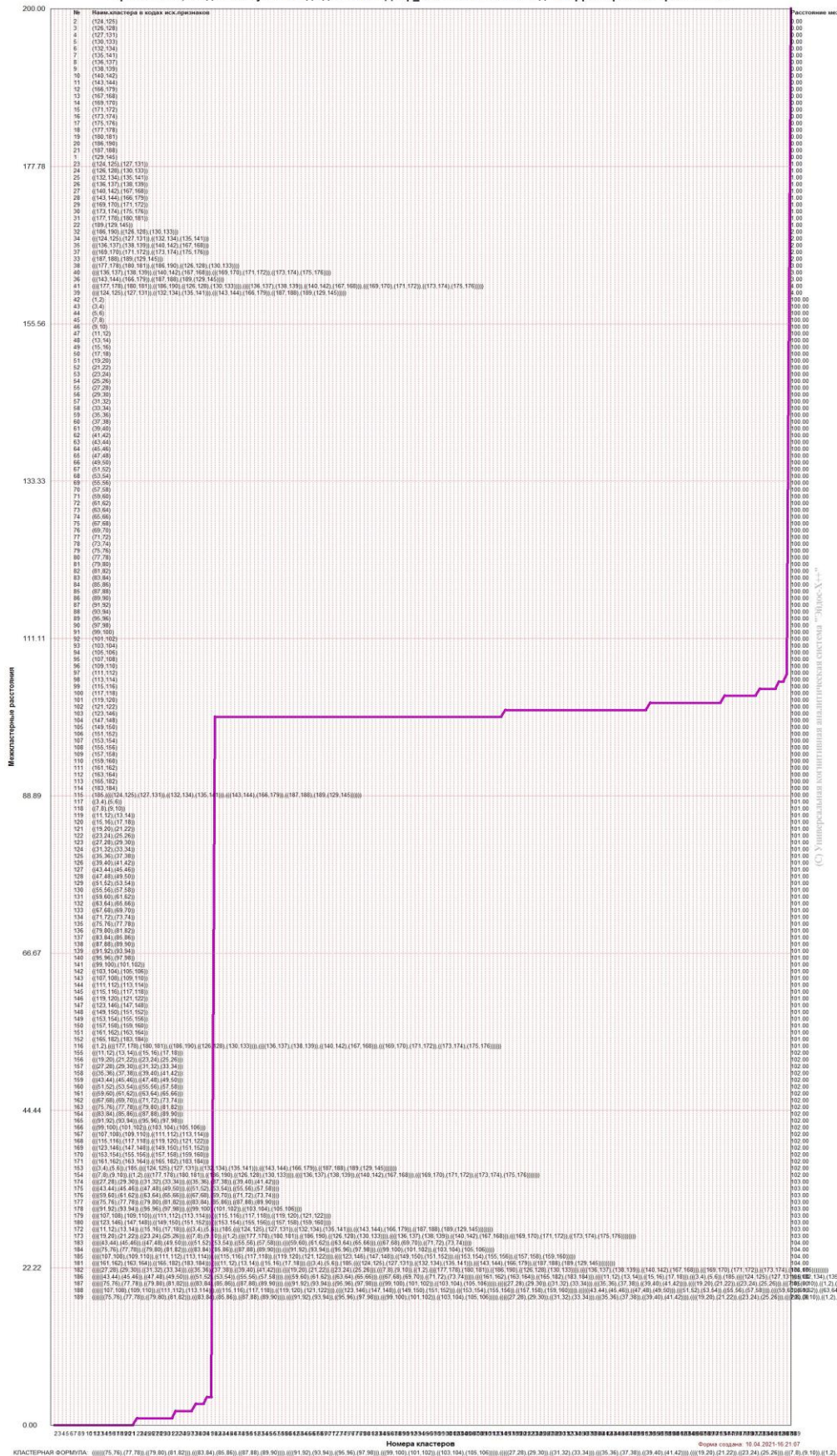


Рисунок 24 – График изменения межкластерных расстояний при когнитивной кластеризации значений факторов

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 85,78% наиболее значимых синаптических связей.

В приведенном фрагменте слоя нейронной сети нейроны соответствуют выбору кандидата (да 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.

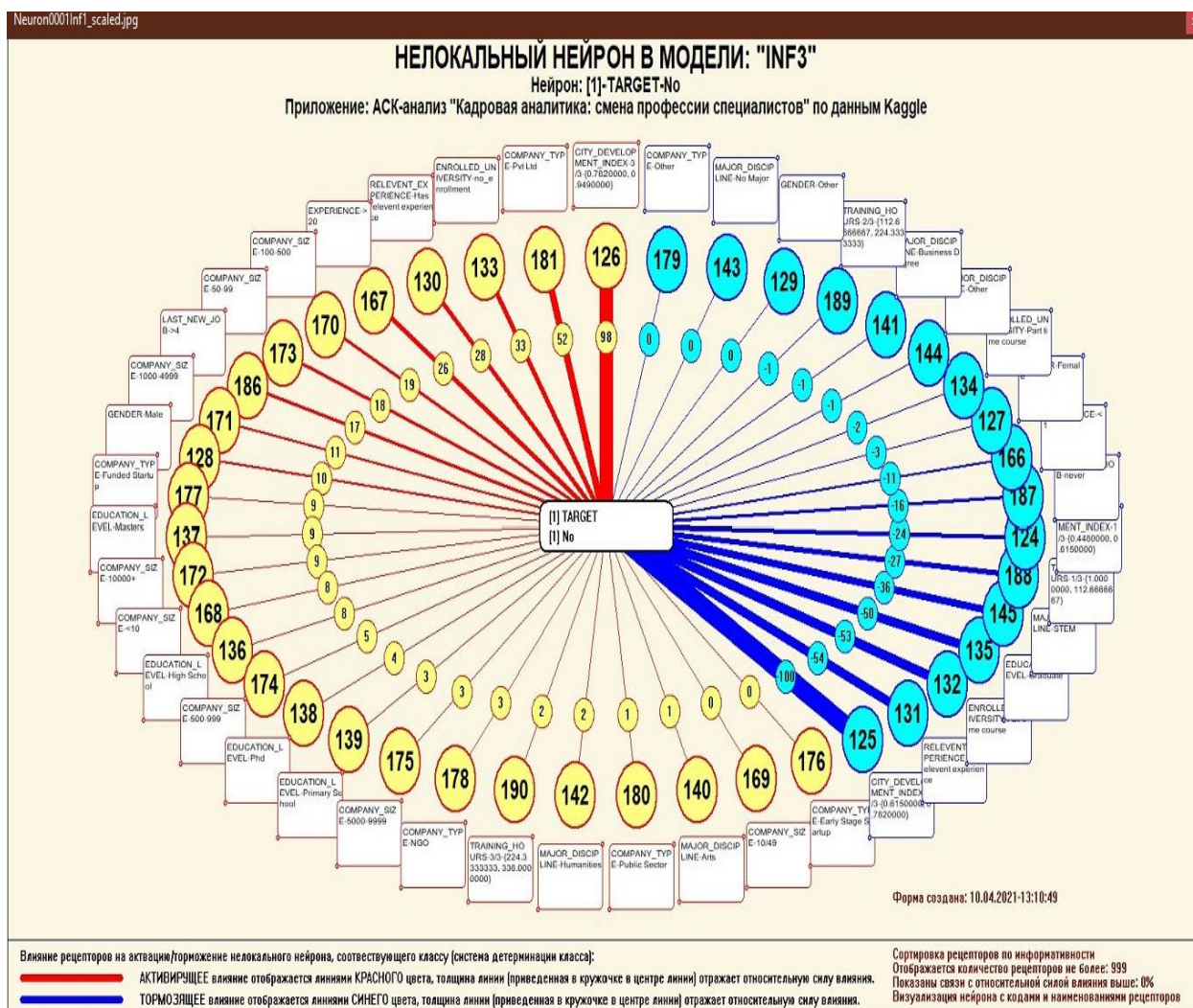


Рисунок 25 – Пример нелокального нейрона

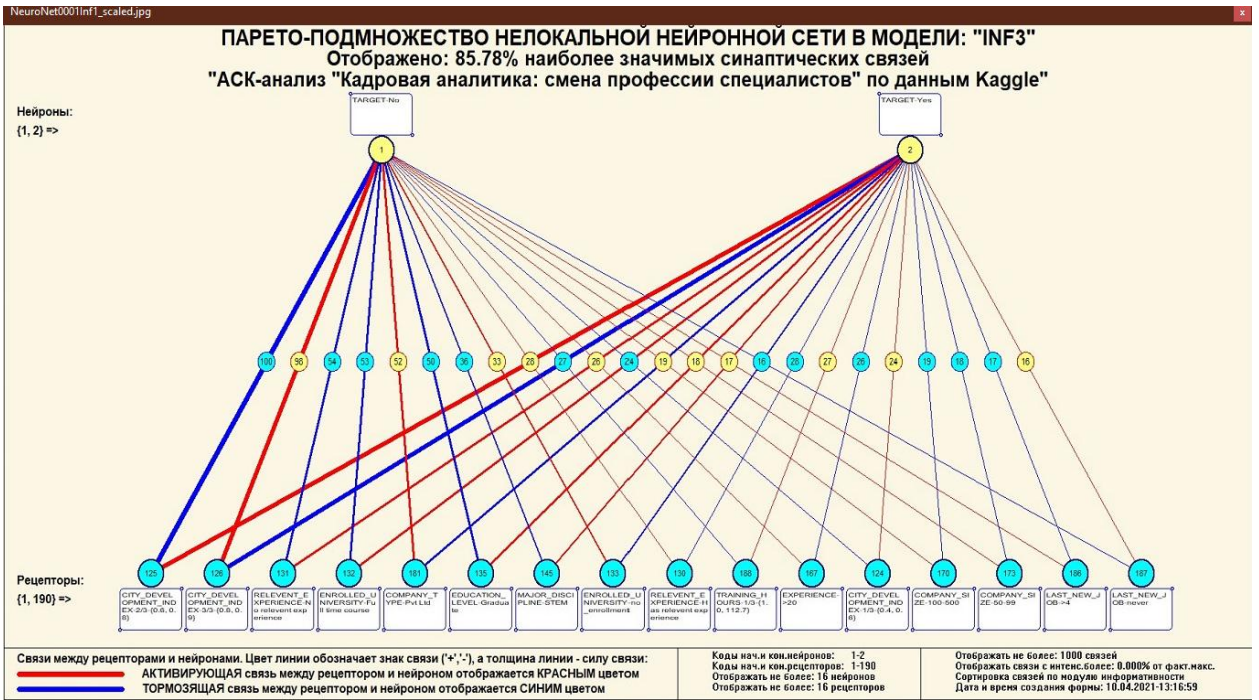


Рисунок 26 – Один слой нелокальной нейронной сети

2.3.6 3D-интегральные когнитивные карты

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

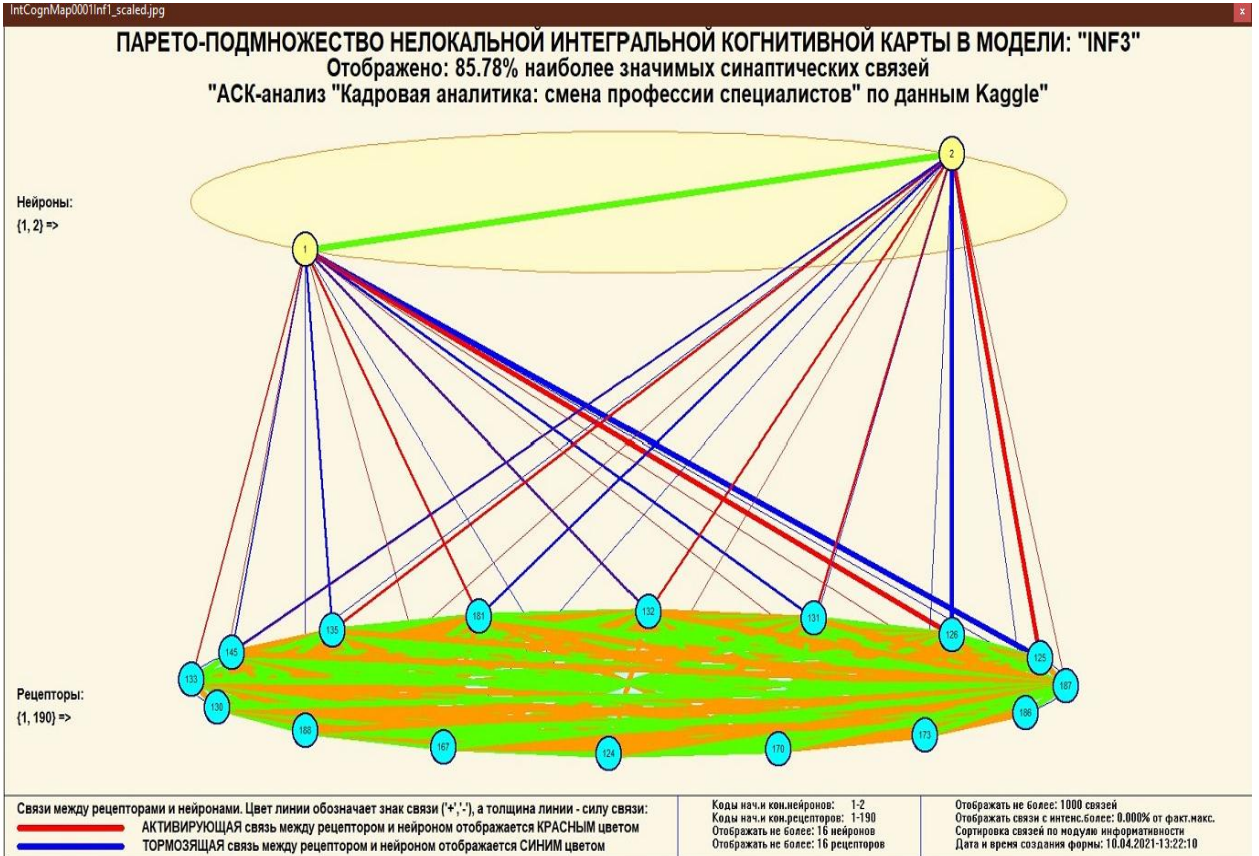


Рисунок 27 – 3D-интегральная когнитивная карта в СК-модели INF3

2.3.7 Когнитивные функции

4.5. Визуализация когнитивных функций

Что такое когнитивная функция:

Визуализация прямых, обратных, позитивных, негативных, полностью и частично редуцированных когнитивных функций Когнитивная функция представляет собой графическое отображение силы и направления влияния различных значений некоторого фактора на переходы объекта управления в будущие состояния, соответствующие классам. Когнитивные функции представляют собой новый перспективный инструмент отражения и наглядной визуализации закономерностей и эмпирических законов. Разработка содержательной научной интерпретации когнитивных функций представляет собой способ познания природы, общества и человека. Когнитивные функции могут быть: прямые, отражающие зависимость классов от признаков, обобщающие информационные портреты признаков; обратные, отражающие зависимость признаков от классов, обобщающие информационные портреты классов; позитивные, показывающие чему способствуют система детерминации; негативные, отражающие чему препятствуют система детерминации; средневзвешенные, отражающие совокупное влияние всех значений факторов на поведение объекта (причем в качестве весов наблюдений используется количество информации в значении аргумента о значениях функции) различной степенью редукции или степенью детерминации, которая отражает в графической форме (в форме полосы) количество знаний в аргументе о значении функции и является аналогом и обобщением доверительного интервала. Если отобразить подматрицу матрицы знания, отображая цветом силу и направление влияния каждой градации некоторой описательной шкалы на переход объекта в состояния, соответствующие классам некоторой классификационной шкалы, то получим нередуцированную когнитивную функцию. Когнитивные функции являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой "Эйдос". Необходимо отметить, что на вид функций влияния математической моделью АСК-анализа не накладывается никаких ограничений, в частности, они могут быть и не дифференцируемые.

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(67). С. 240 - 282. - Шифр Информрегистра: 0421100012\0077. - 2,688 з.п.л. - Режим доступа: <http://ej.kubagro.ru/2011/03/pdf/18.pdf>

Задайте нужный режим:

Визуализации когнитивных функций

Литератур.ссылки на работы по когнитивным функциям

Литератур.ссылки на работы по когнитивным функциям

Литератур.ссылки на работы по управлению знаниями

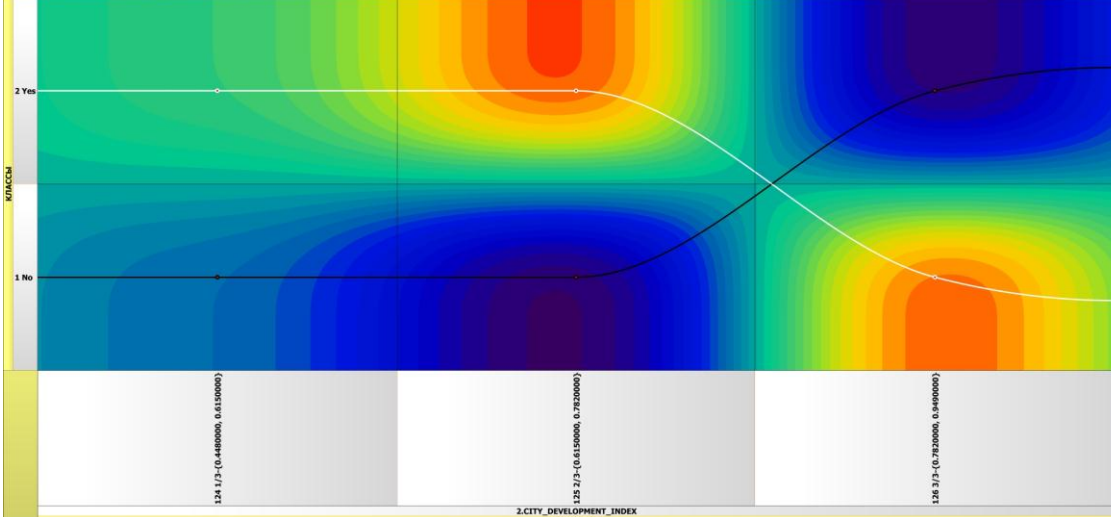
Такие диаграммы отражают то, как влияют градации различных факторов на прогноз. Эти диаграммы мы получаем в режиме 4.5 «Визуализация когнитивных функций» и некоторые из них представлены на рисунках 28.

На рисунке 28 мы видим, как влияет признак «Город» на модель. Можно увидеть, что ярко выделяется 21-й город – содержит большое количество информации о том, что «Yes» и «No», а остальные влияют, но не сильно.

На рисунке 28 мы видим, как влияет признак «Индекс развития города» на модель. При низком значении параметра очень характерно, что «Yes» и нехарактерно что «No», а при среднем значении – наоборот и не ярко выражено.

Из диаграмм видно, что красный цвет отражает содержание большого количества информации в значении фактора «Yes», а синее отражает то, что это не произойдет.

КОГНИТИВНАЯ ФУНКЦИЯ:
Приложение: Приложение, созданное путем ввода данных из БД Esp_data. Это название надо скорректировать в режиме 1.3!
Модель: InT3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами

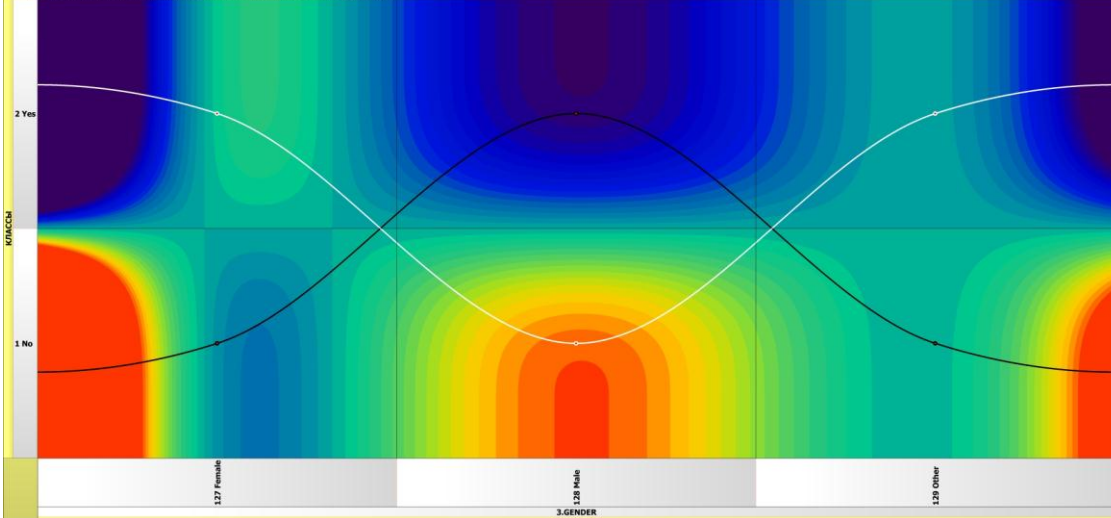


1-я подматрица
Степень редукции: 0,00
Степень редукции частично редукцированных когнитивных функций: 1,00
Позитивной от признаков: 1,00
Позитивной от признаков: 0,00
Негативной от признаков: 1,00
Негативной от признаков: 0,00
Метод интерполяции: Филлер таб.1
Количество градаций уровней: 32
Цвет нерасчетных нулей: Отсутствует
Тип палитры: Отсутствует
Цветовая гамма, спектр от син: Отсутствует
Интерполяция нерасчетных нулей: Отсутствует
Размер шрифта 45

Визуализация когнитивных функций системы «Эйдос».
© Е.В.Луценко (Россия), Д.А.Бандык (Беларусь), Патент РФ 2011612056 РФ от 09.03.2011

Изображение создано:
10.04.2011 16:13:14

КОГНИТИВНАЯ ФУНКЦИЯ:
Приложение: Приложение, созданное путем ввода данных из БД Esp_data. Это название надо скорректировать в режиме 1.3!
Модель: InT3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами

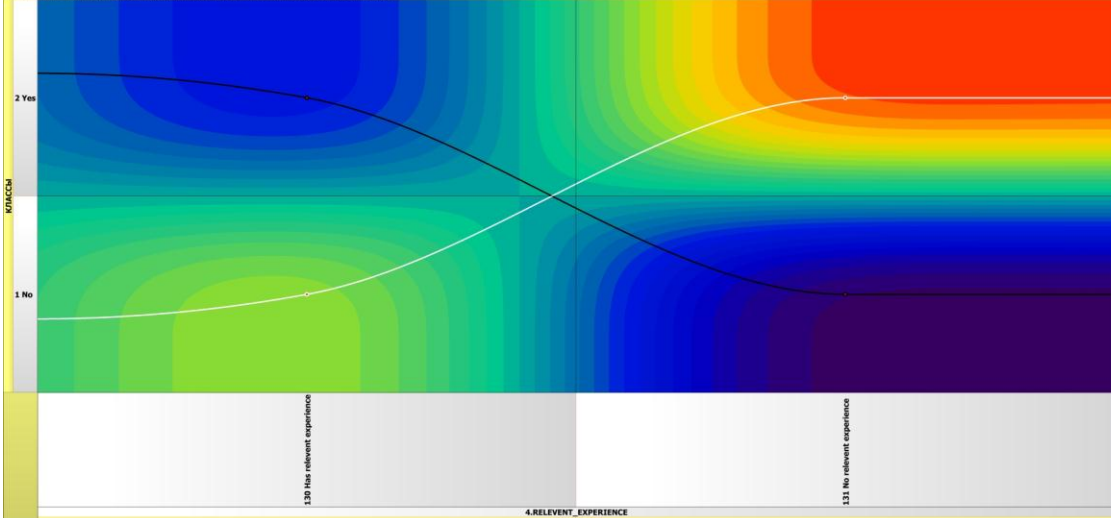


2-я подматрица
Степень редукции: 0,00
Степень редукции частично редукцированных когнитивных функций: 1,00
Позитивной от признаков: 0,00
Позитивной от признаков: 1,00
Негативной от признаков: 0,00
Негативной от признаков: 1,00
Метод интерполяции: Филлер таб.1
Количество градаций уровней: 32
Цвет нерасчетных нулей: Отсутствует
Тип палитры: Отсутствует
Цветовая гамма, спектр от син: Отсутствует
Интерполяция нерасчетных нулей: Отсутствует
Размер шрифта 45

Визуализация когнитивных функций системы «Эйдос».
© Е.В.Луценко (Россия), Д.А.Бандык (Беларусь), Патент РФ 2011612056 РФ от 09.03.2011

Изображение создано:
10.04.2011 16:13:24

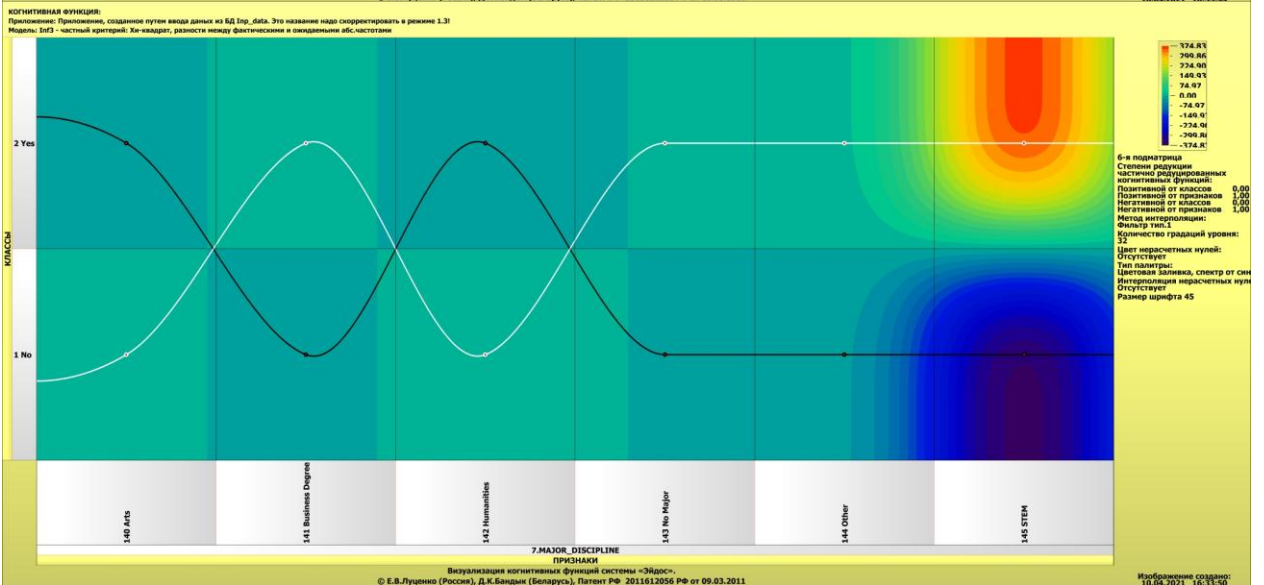
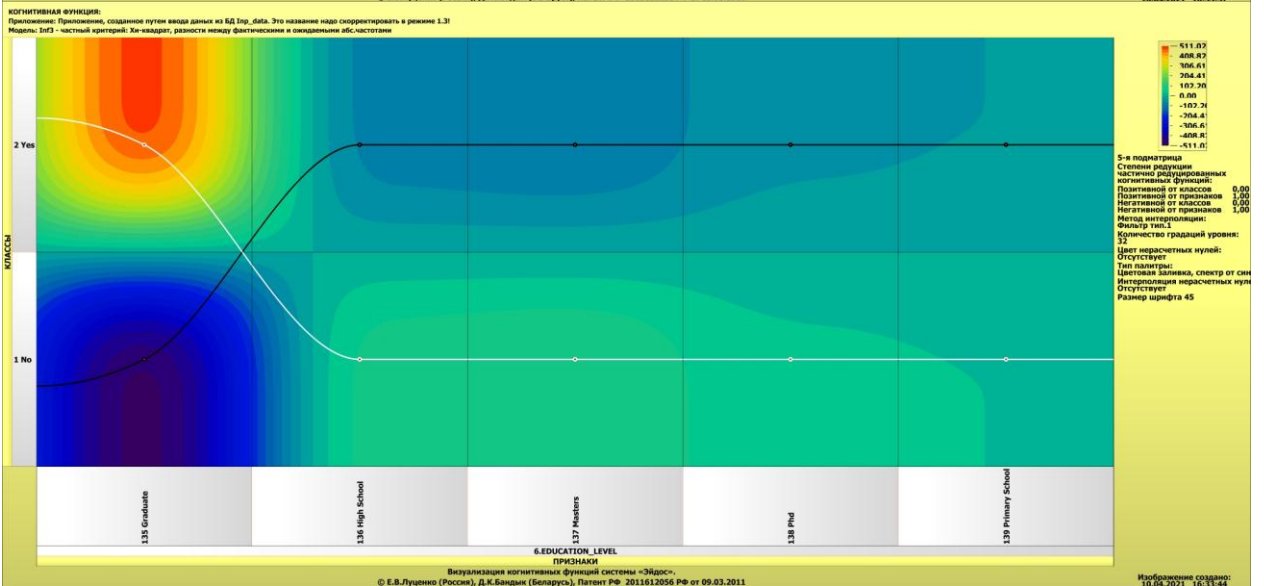
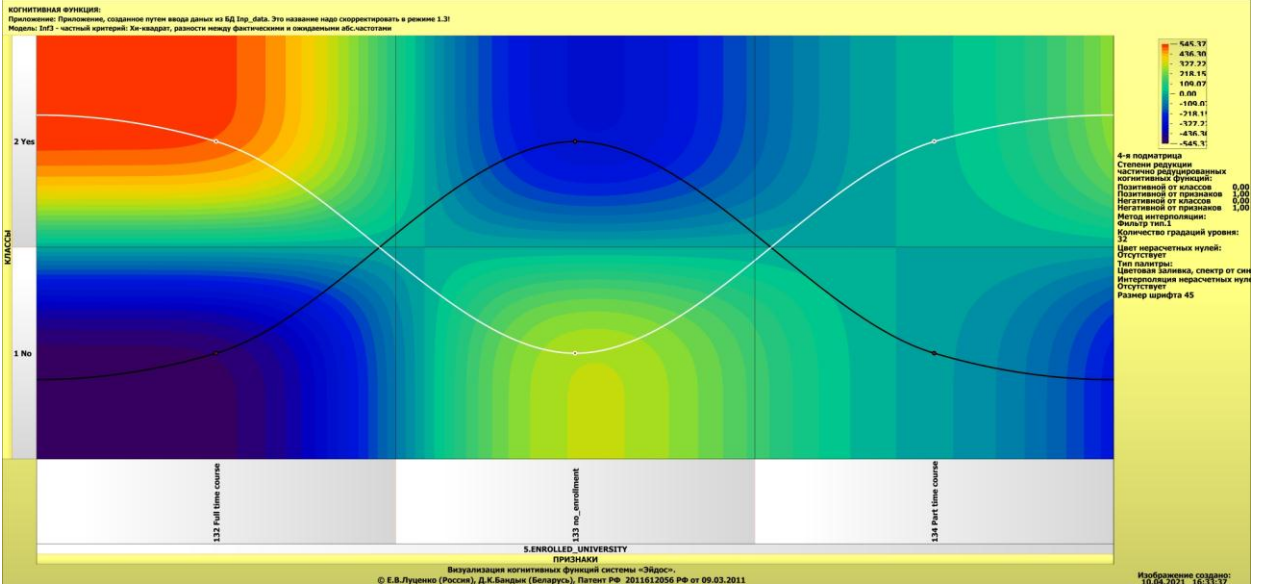
КОГНИТИВНАЯ ФУНКЦИЯ:
Приложение: Приложение, созданное путем ввода данных из БД Esp_data. Это название надо скорректировать в режиме 1.3!
Модель: InT3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами

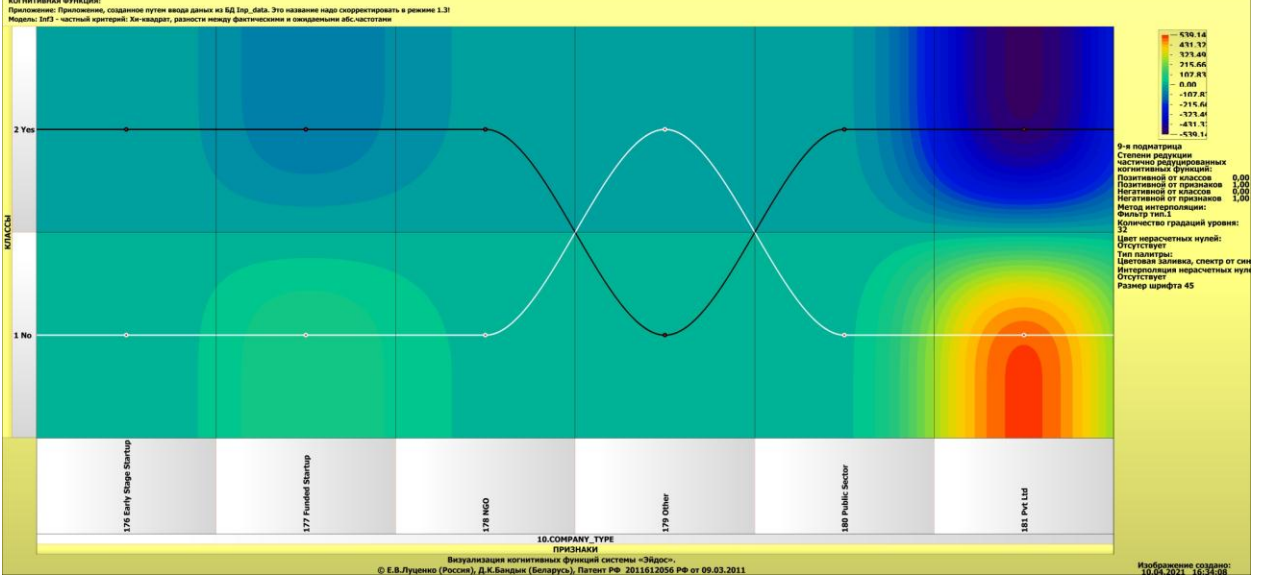
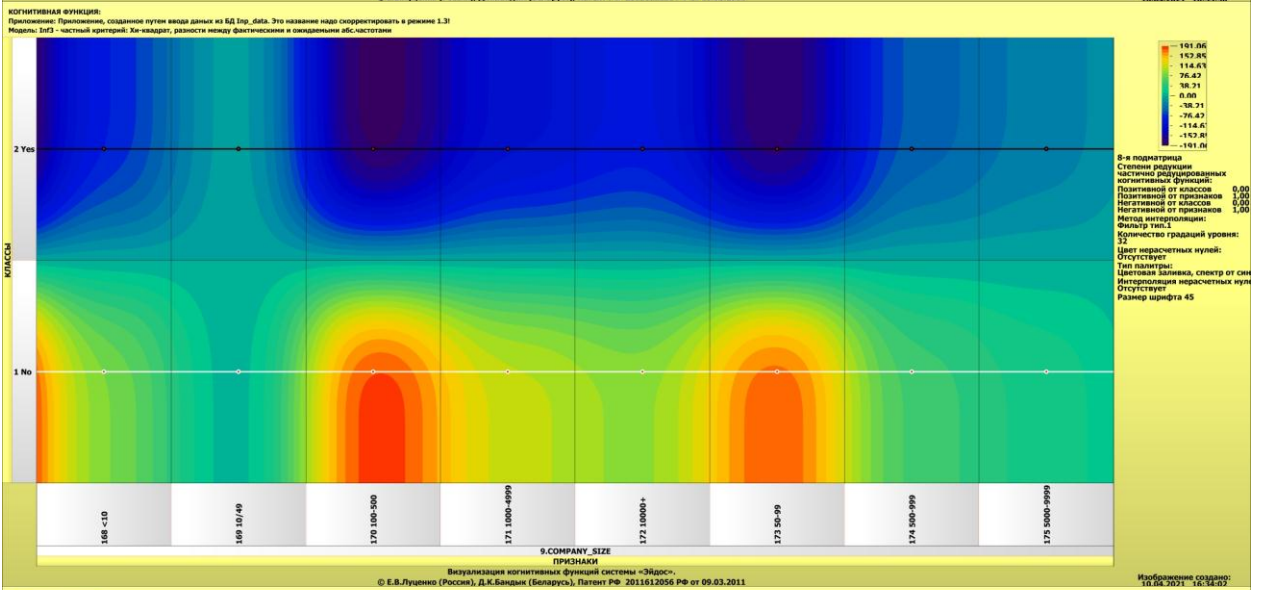
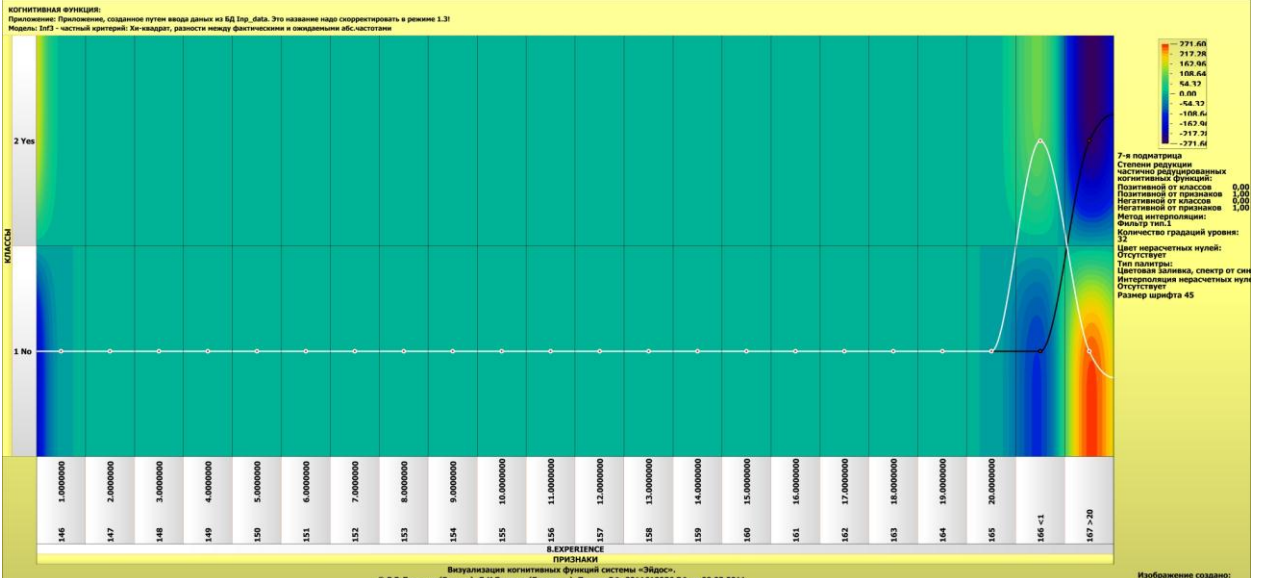


3-я подматрица
Степень редукции: 0,00
Степень редукции частично редукцированных когнитивных функций: 1,00
Позитивной от признаков: 0,00
Позитивной от признаков: 1,00
Негативной от признаков: 0,00
Негативной от признаков: 1,00
Метод интерполяции: Филлер таб.1
Количество градаций уровней: 32
Цвет нерасчетных нулей: Отсутствует
Тип палитры: Отсутствует
Цветовая гамма, спектр от син: Отсутствует
Интерполяция нерасчетных нулей: Отсутствует
Размер шрифта 45

Визуализация когнитивных функций системы «Эйдос».
© Е.В.Луценко (Россия), Д.А.Бандык (Беларусь), Патент РФ 2011612056 РФ от 09.03.2011

Изображение создано:
10.04.2011 16:13:31





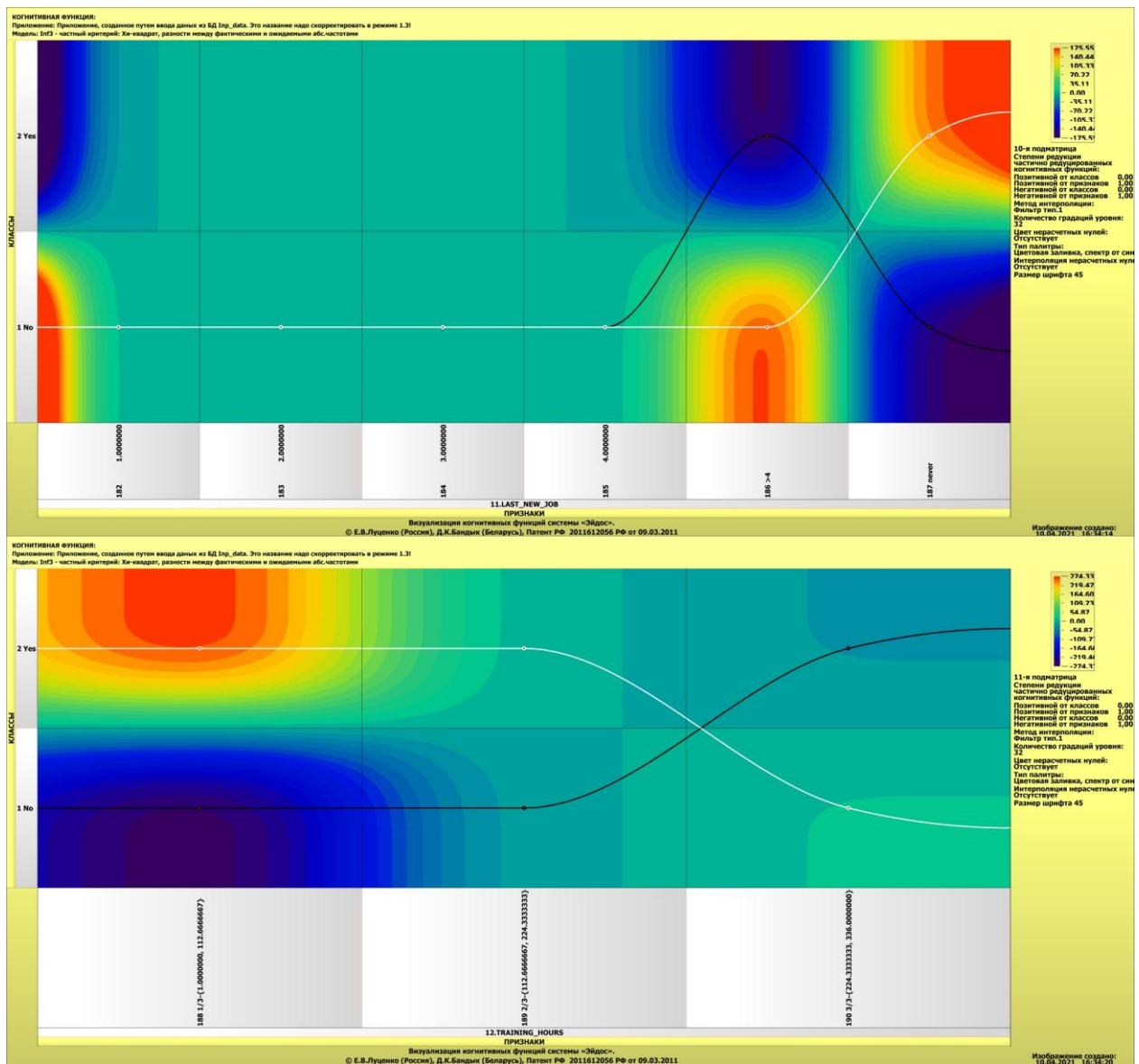


Рисунок 28 – Примеры визуализации когнитивной функции модели INF3

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 системы 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 предполагаемый выбор кандидата по влиянию собранных в анкетах данных, 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 // Политематический 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 // Политематический 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 // Политематический 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 // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №02(126). С. 1 – 32. – IDA [article ID]: 1261702001. – Режим доступа: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 у.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 // Политематический 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.