

**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

Научно исследовательская работа

по дисциплине: Теория информации, данные, знания

на тему: АСК-анализ рейтинга книг по данным портала Kaggle

Выполнили студенты группы: ИТ2102 Перекрестова Анна Андреевна,
Шишкин Назар Евгеньевич

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

(подпись, расшифровка подписи)

Краснодар 2021

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1. ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ.....	4
1.1. Описание решения	4
1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel	6
1.3. Ввод выборки в систему Aidos-X.....	7
2. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	11
2.1 Синтез и верификация статистических и интеллектуальных моделей в системе aidos-x.....	11
2.2 Виды моделей системы Aidos-X.....	12
2.3 Результаты верификации моделей	13
3. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	16
3.1. Решение задачи идентификации.....	16
3.2. Кластерно-конструктивный анализ.....	18
3.3. Нелокальные нейронные сети и нейтроны.....	20
3.4. SWOT и PEST матрицы и диаграммы	23
ЗАКЛЮЧЕНИЕ	26
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ	28

ВВЕДЕНИЕ

Одним из важных и перспективных направлений развития современных информационных технологий является создание систем искусственного интеллекта. Существует множество альтернатив систем искусственного интеллекта, но возникает необходимость оценки качества математических моделей этих систем. Одной из перспективных свободных систем является Aidos-X, с ее помощью будет рассмотрено решение задачи анализа рейтингов книг на основе данных портала Kaggle.

В данной работе использована база данных «Goodreads-books» из банка исходных данных по задачам искусственного интеллекта – Kaggle.

Для решения задачи используем стандартные возможности MS Office Word и Excel, а также систему искусственного интеллекта «Aidos-X++».

1. ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ

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 текстам;
- 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 251, соответственно);

- мультязычная 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 решение 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](https://www.kaggle.com/jealousleopard/goodreadsbooks) возьмем набор данных «Goodreads-books»: <https://www.kaggle.com/jealousleopard/goodreadsbooks>.

Csv файл содержит следующие данные:

- average_rat – рейтинг;
- title – наименование книги;
- authors – автор;
- language_code – язык произведения;
- publisher – издательство;
- isbn13 – универсальный номер книги;
- num_pages – количество страниц;
- ratings_count – количество рейтингов;
- text_reviews_count – количество отзывов.

В качестве разделителей используются запятые.

Для конвертации csv-файла в xls был использован внутренний функционал MS Excel. Получившийся файл представлен на рисунке 1.

	B	C	D	E	F	G	H	I	J
1	average_ra	title	authors	language_code	publisher	isbn13	num_pages	ratings_count	text_reviews_count
2	4.57	Harry Potter and the Half-Blood Prince (Harr	J.K. Rowling/Mary GrandPré	eng	Scholastic Inc.	9.78044E+12	652	2095690	27591
3	4.49	Harry Potter and the Order of the Phoenix (H	J.K. Rowling/Mary GrandPré	eng	Scholastic Inc.	9.78044E+12	870	2153167	29221
4	4.42	Harry Potter and the Chamber of Secrets (Ha	J.K. Rowling	eng	Scholastic	9.78044E+12	352	6333	244
5	4.50	Harry Potter and the Prisoner of Azkaban (Ha	J.K. Rowling/Mary GrandPré	eng	Scholastic Inc.	9.78044E+12	435	2339585	36325
6	4.78	Harry Potter Boxed Set: Books 1-5 (Harry Pot	J.K. Rowling/Mary GrandPré	eng	Scholastic	9.78044E+12	2690	41428	164
7	3.74	Unauthorized Harry Potter Book Seven News	W. Frederick Zimmerman	en-US	Nimble Books	9.78098E+12	152	19	1
8	4.73	Harry Potter Collection (Harry Potter #1-6)	J.K. Rowling	eng	Scholastic	9.78044E+12	3342	28242	808
9	4.38	The Ultimate Hitchhiker's Guide: Five Compl	Douglas Adams	eng	Gramercy Books	9.78052E+12	815	3628	254
10	4.38	The Ultimate Hitchhiker's Guide to the Galax	Douglas Adams	eng	Del Rey Books	9.78035E+12	815	249558	4080
11	4.22	The Hitchhiker's Guide to the Galaxy (Hitchh	Douglas Adams	eng	Crown	9.7814E+12	215	4930	460
12	4.22	The Hitchhiker's Guide to the Galaxy (Hitchh	Douglas Adams/Stephen Fry	eng	Random House Au	9.78074E+12	6	1266	253
13	4.38	The Ultimate Hitchhiker's Guide (Hitchhiker	Douglas Adams	eng	Wings Books	9.78052E+12	815	2877	195
14	4.21	A Short History of Nearly Everything	Bill Bryson	eng	Broadway Books	9.78077E+12	544	248558	9396
15	3.44	Bill Bryson's African Diary	Bill Bryson	eng	Broadway Books	9.78077E+12	55	7270	499
16	3.87	Bryson's Dictionary of Troublesome Words: /	Bill Bryson	eng	Broadway Books	9.78077E+12	256	2088	131
17	4.07	In a Sunburned Country	Bill Bryson	eng	Broadway Books	9.78077E+12	335	72451	4245
18	3.9	I'm a Stranger Here Myself: Notes on Return	Bill Bryson	eng	Broadway Books	9.78077E+12	304	49240	2211
19	3.83	The Lost Continent: Travels in Small Town An	Bill Bryson	eng	William Morrow P	9.78006E+12	299	45712	2257
20	3.80	Neither Here nor There: Travels in Europe	Bill Bryson	eng	William Morrow P	9.78038E+12	254	48701	2238
21	3.91	Notes from a Small Island	Bill Bryson	eng	William Morrow P	9.78038E+12	324	80609	3301
22	3.93	The Mother Tongue: English and How It Got	Bill Bryson	eng	William Morrow P	9.78038E+12	270	28489	2085
23	4.59	J.R.R. Tolkien 4-Book Boxed Set: The Hobbit	J.R.R. Tolkien	eng	Ballantine Books	9.78035E+12	1728	101233	1550
24	4.5	The Lord of the Rings (The Lord of the Rings	J.R.R. Tolkien	eng	Houghton Mifflin I	9.78062E+12	1184	1710	91
25	4.30	The Fellowship of the Ring (The Lord of the	J.R.R. Tolkien	eng	Houghton Mifflin I	9.78062E+12	398	2128944	13670
26	4.5	The Lord of the Rings (The Lord of the Rings	J.R.R. Tolkien/Alan Lee	en-US	Houghton Mifflin I	9.78062E+12	1216	1618	140

Рисунок 1 – Импорт csv в Excel

Таким образом, в качестве классификационной шкалы решено использовать столбец рейтинга. Описательные шкалы – все последующие.

1.3. Ввод выборки в систему Aidos-X

Теперь, когда имеется обучающая выборка в формате *.xls, можно импортировать ее в систему Aidos-X.

Скопировав данную выборку в папку Inp_data, необходимо переименовать ее в Inp_data. Далее был использован универсальный программный интерфейс импорта данных в систему Aidos-X (режим 2.3.2.2), результат заполнения данной формы представлен на рисунке 2.

Рисунок 2 – Ввод обучающей выборки

Стоит отметить следующие настройки:

- тип файла – xls;
- классификационные шкалы – 2;

– описательные шкалы – 3-10;

В форме задания размерности модели системы оставляем всё без изменений (Рисунок 3).

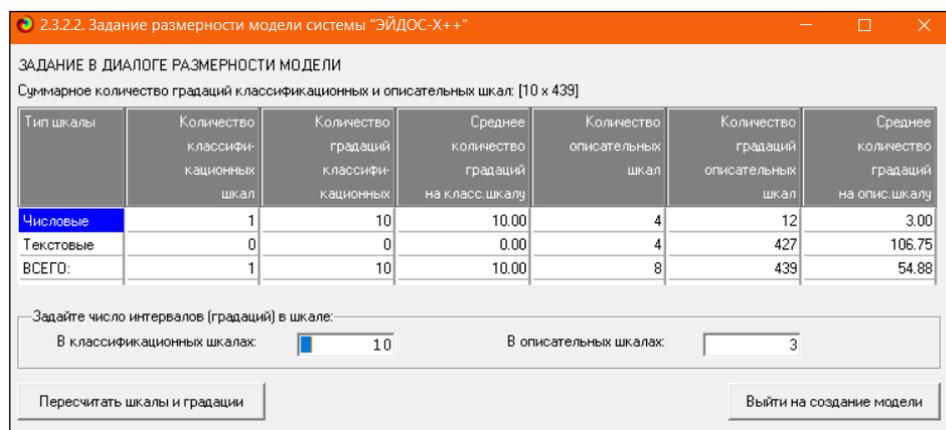


Рисунок 3 – Задание размерности модели системы Aidos-X

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.

Классификационная шкала представлена на рисунке 4, её можно просмотреть в режиме 2.1. Описательные шкалы – в режиме 2.2 (рисунок 5).

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

Код шкалы	Наименование классификационной шкалы	Код градации	Наименование градации классификационной шкалы
1	AVERAGE_RATING	1	1/10-(2.9900000, 3.1690000)
		2	2/10-(3.1690000, 3.3480000)
		3	3/10-(3.3480000, 3.5270000)
		4	4/10-(3.5270000, 3.7060000)
		5	5/10-(3.7060000, 3.8850000)
		6	6/10-(3.8850000, 4.0640000)
		7	7/10-(4.0640000, 4.2430000)
		8	8/10-(4.2430000, 4.4220000)
		9	9/10-(4.4220000, 4.6010000)
		10	10/10-(4.6010000, 4.7800000)

Помощь Доб.шкалу Доб.град.шкалы Копир.шкалу Копир.град.шкалы Копир.шкалу с град. Удал.шкалу с град. Удал.град.шкалы Удаление и перекодирование Графики будущих сценариев

Рисунок 4 – Классификационные шкалы и градации

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

Код шкалы	Наименование описательной шкалы	Код градации	Наименование градации описательной шкалы
1	TITLE	1	"Half"
2	AUTHORS	2	"Hatchet"
3	LANGUAGE_CODE	3	Across
4	PUBLISHER	4	Active
5	NUM_PAGES	5	Actors
6	RATINGS_COUNT	6	Acts
7	TEXT_REVIEWS_COUNT	7	Adams
		8	Adams's
		9	Address
		10	Adventures
		11	African
		12	After
		13	Against
		14	Agile
		15	Album
		16	Always
		17	America
		18	Americana
		19	Among
		20	Analysis
		21	Anni's
		22	Anna
		23	Annals
		24	anos
		25	Another
		26	aruga
		27	Artesia

Помощь Доб.шкалу Доб.град.шкалы Копир.шкалу Копир.град.шкалы Копир.шкалу с град. Удал.шкалу с град. Удал.град.шкалы Перекодировать Очистить Графики прошлых сценариев

Рисунок 5 – Описательные шкалы и градации (фрагмент)

Для ручного ввода-корректировки обучающей выборки существует режим 2.3.1, он представлен на рисунке 6. Установка значений описательных и классификационных шкал объектов осуществляется по их номерам.

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

Код объекта	Наименование объекта	Дата	Время
1	1		
2	2		
3	4		
4	5		
5	8		
6	9		
7	10		
8	12		
9	13		
10	14		
11	16		

Код объекта	Класс 1	Класс 2	Класс 3	Класс 4	Код объекта	Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
1	9	0	0	0	1	170	293	168	47	299	170	293
					1	687	628	557	866	892	897	900

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

Рисунок 6 – Обучающая выборка (фрагмент)

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

2. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

2.1 Синтез и верификация статистических и интеллектуальных моделей в системе aidos-x

Далее необходимо запустить режим 3.5, в котором задаются модели для синтеза и верификации, а также задается модель, которой по окончании режима присваивается статус текущей (рисунок 7).

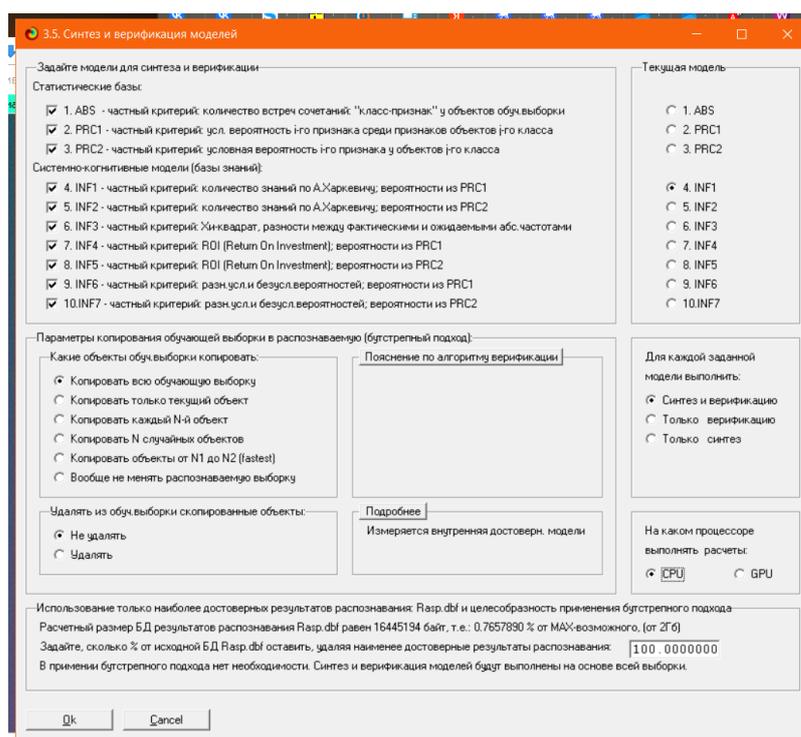


Рисунок 7 – Выбор моделей для синтеза и верификации

В данном режиме имеется много различных методов верификации моделей. Но требуется использовать параметры, приведенные на рисунке 8. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 8.

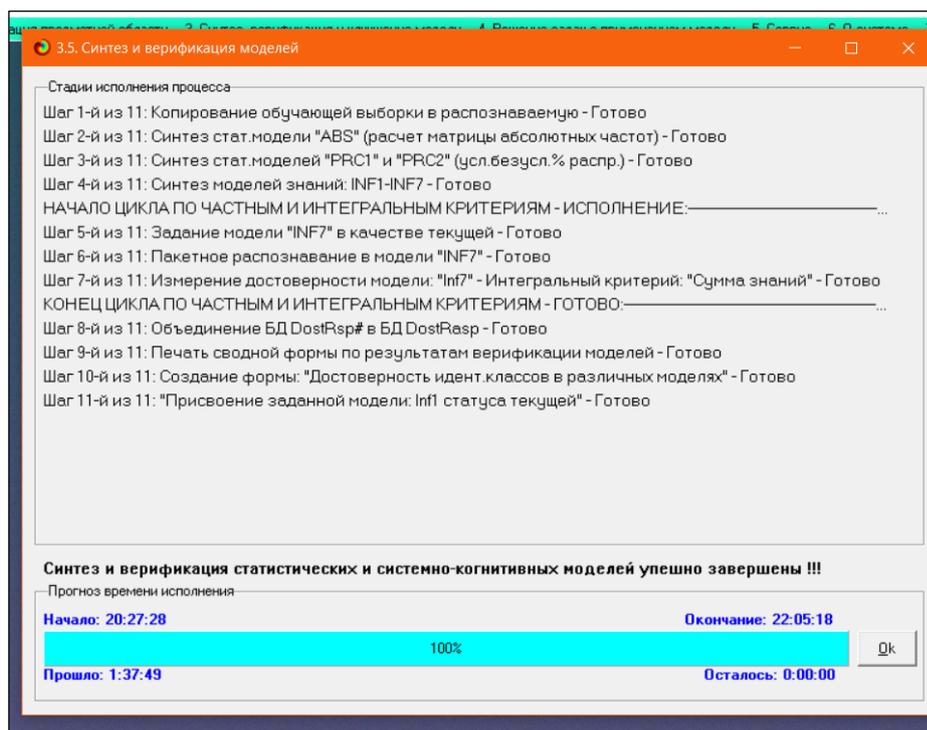


Рисунок 8 – Процесс исполнения режима 3.5

Интересно заметить, что синтез и верификация всех 10 моделей на данной задаче заняли 50 секунд. При этом оценка достоверности моделей проводилась на 10 процентах наблюдения из обучающей выборки. После этого можно перейти непосредственно к выбору наиболее достоверной модели.

2.2 Виды моделей системы Aidos-X

Рассмотрение решения задачи идентификации будет проводиться 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.

То есть частные критерии представляют собой просто формулы для преобразования матрицы абсолютных частот (рисунок 9) в матрицы условных и безусловных процентных распределений, и матрицы знаний.

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

Код признака	Наименование описательных шкалы и градации	1 AVERAGE_RATI... 1/10 (2.2900000; 2.1630000)	2 AVERAGE_RATI... 2/10 (3.1690000; 3.3480000)	3 AVERAGE_RATI... 3/10 (3.3480000; 3.5270000)	4 AVERAGE_RATI... 4/10 (3.5270000; 3.7060000)	5 AVERAGE_RATI... 5/10 (3.7060000; 3.8850000)	6 AVERAGE_RATI... 6/10 (3.8850000; 4.0640000)	7 AVERAGE_RATI... 7/10 (4.0640000; 4.2430000)	8 AVERAGE_RATI... 8/10 (4.2430000; 4.4220000)	9 AVERAGE_RATI... 9/10 (4.4220000; 4.6010000)	10 AVERAGE_RATI... 10/10 (4.6010000; 4.7800000)
1	TITLE:"Hal"	-0.005	-0.018	-0.058	-0.091	0.759	-0.286	-0.147	-0.064	-0.076	
2	TITLE:"Hatchet"	-0.005	-0.018	-0.058	-0.091	-0.241	0.714	-0.147	-0.064	-0.076	
3	TITLE:"Across"	-0.010	-0.035	-0.116	-0.182	-0.482	1.428	-0.295	-0.127	-0.152	
4	TITLE:"Active"	-0.005	-0.018	-0.058	-0.091	-0.241	0.714	-0.147	-0.064	-0.076	
5	TITLE:"Actors"	-0.005	-0.018	-0.058	0.909	-0.241	-0.286	-0.147	-0.064	-0.076	
6	TITLE:"Acts"	-0.005	-0.018	-0.058	-0.091	-0.241	-0.286	0.853	-0.064	-0.076	
7	TITLE:"Adams"	-0.005	-0.018	-0.058	-0.091	-0.241	-0.286	0.853	-0.064	-0.076	
8	TITLE:"Adams's"	-0.005	-0.018	-0.058	0.909	-0.241	-0.286	-0.147	-0.064	-0.076	
9	TITLE:"Address"	-0.005	-0.018	-0.058	-0.091	-0.241	-0.286	-0.147	-0.064	0.924	
10	TITLE:"Adventures"	-0.005	-0.018	-0.058	-0.091	-0.241	-0.286	0.853	-0.064	-0.076	
11	TITLE:"African"	-0.005	-0.018	0.942	-0.091	-0.241	-0.286	-0.147	-0.064	-0.076	
12	TITLE:"Alter"	-0.005	-0.018	-0.058	-0.091	-0.241	0.714	-0.147	-0.064	-0.076	
13	TITLE:"Against"	-0.005	-0.018	-0.058	-0.091	-0.241	-0.286	0.853	-0.064	-0.076	
14	TITLE:"Agle"	-0.005	-0.018	-0.058	-0.091	0.759	-0.286	-0.147	-0.064	-0.076	
15	TITLE:"Album"	-0.005	-0.018	-0.058	-0.091	-0.241	-0.286	0.853	-0.064	-0.076	
16	TITLE:"Always"	-0.010	-0.035	-0.116	-0.182	-0.482	-0.286	-0.295	-0.127	1.848	
17	TITLE:"America"	-0.014	-0.053	-0.174	-0.274	0.276	0.143	0.558	-0.191	-0.228	
18	TITLE:"Americana"	-0.005	-0.018	0.942	-0.091	-0.241	-0.286	-0.147	-0.064	-0.076	
19	TITLE:"Among"	-0.005	-0.018	-0.058	-0.091	-0.241	-0.286	-0.147	-0.064	0.924	
20	TITLE:"Analysis"	-0.005	-0.018	-0.058	-0.091	0.759	-0.286	-0.147	-0.064	-0.076	
21	TITLE:"Ann's"	-0.010	-0.035	-0.116	-0.182	-0.482	-0.572	1.705	-0.127	-0.152	
22	TITLE:"Anna"	0.957	-0.160	-0.521	-0.821	-1.171	3.428	-0.327	-0.572	-0.685	
23	TITLE:"Annals"	-0.005	-0.018	-0.058	-0.091	-0.241	-0.286	-0.147	0.936	-0.076	
24	TITLE:"anos"	-0.005	-0.018	-0.058	-0.091	-0.241	-0.286	0.853	-0.064	-0.076	

Рисунок 9 – Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений

Рисунок 11 - Модель INF4 (фрагмент)

Рисунок 12 - Модель INF5 (фрагмент)

2.3 Результаты верификации моделей

Результаты верификации моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями представлены на рисунке 10.

3.4. Обобщенная по достов. моделям при разн. крит. Текущая модель: "INF3"

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	Фигура Ван Ризбергена	Сумма между-уровней сход. истинно-поло-решений (STP)	Сумма между-уровней сход. истинно-отриц. решений (STN)	Сумма между-уровней сход. ложно-поло-решений	Сумма между-уровней сход. ложно-отриц. решений
1. ABS - частный критерий: количество встреч соцветий "Улос"	Корреляция абс. частот с обр...	200	200	180	180	0.526	1.000	0.690	111.034			60	
1. ABS - частный критерий: количество встреч соцветий "Улос"	Сумма абс. частот по признаку...	200	200	180	180	0.526	1.000	0.690	110.996			42	
2. PR1C1 - частный критерий: усл. вероятность i-го признака сред.	Корреляция усл.отн. частот с о...	200	200	180	180	0.526	1.000	0.690	111.034			60	
2. PR1C1 - частный критерий: усл. вероятность i-го признака сред.	Сумма усл.отн. частот по при...	200	200	180	180	0.526	1.000	0.690	64.165			46	
3. PR1C2 - частный критерий: условная вероятность i-го признака.	Корреляция усл.отн. частот с о...	200	200	180	180	0.526	1.000	0.690	111.034			60	
3. PR1C2 - частный критерий: условная вероятность i-го признака.	Сумма усл.отн. частот по при...	200	200	180	180	0.526	1.000	0.690	67.259			47	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в.	Семантический резонанс зна...	200	188	142	38	12	0.832	0.940	0.883	32.244	7.133	3	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в.	Сумма знаний	200	198	56	124	2	0.615	0.990	0.759	60.539	1.960	8	
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в.	Семантический резонанс зна...	200	188	142	38	12	0.832	0.940	0.883	32.366	7.292	3	
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в.	Сумма знаний	200	198	61	119	2	0.625	0.990	0.766	64.490	1.966	8	
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Семантический резонанс зна...	200	199	104	76	1	0.724	0.995	0.838	51.871	11.962	9	
6. INF3 - частный критерий: Хи-квадрат, разности между факти...	Сумма знаний	200	199	104	76	1	0.724	0.995	0.838	67.213	18.339	9	
7. INF4 - частный критерий: ROI (Return On Investment), верно...	Семантический резонанс зна...	200	185	161	19	15	0.907	0.925	0.916	35.269	7.984	2	
7. INF4 - частный критерий: ROI (Return On Investment), верно...	Сумма знаний	200	200	50	130	0.606	1.000	0.755	10.025	0.050	1		
8. INF5 - частный критерий: ROI (Return On Investment), верно...	Семантический резонанс зна...	200	185	161	19	15	0.907	0.925	0.916	35.291	7.984	2	
8. INF5 - частный критерий: ROI (Return On Investment), верно...	Сумма знаний	200	200	60	120	0.625	1.000	0.769	10.457	0.050	1		
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	200	196	64	116	4	0.628	0.980	0.766	52.805	6.896	12	
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Сумма знаний	200	199	47	133	1	0.599	0.995	0.748	15.276	2.486	8	
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	200	194	102	78	6	0.713	0.970	0.822	52.409	7.374	9	
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Сумма знаний	200	198	61	119	2	0.625	0.990	0.766	15.963	2.270	7	

Рисунок 10 – Оценки достоверности моделей

Наиболее достоверной в данном приложении оказалась модель INF3 при интегральном критерии «Сумма знаний». При этом точность модели (F-

мера Ван Ризбергена) составляет 0,925 а точность модели (L1-мера профессора Луценко) – 0,967. 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.

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 и 21 интеллектуальных технологий. На рисунках 11-12 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и не идентифицированных ситуаций в наиболее достоверной модели ABS.

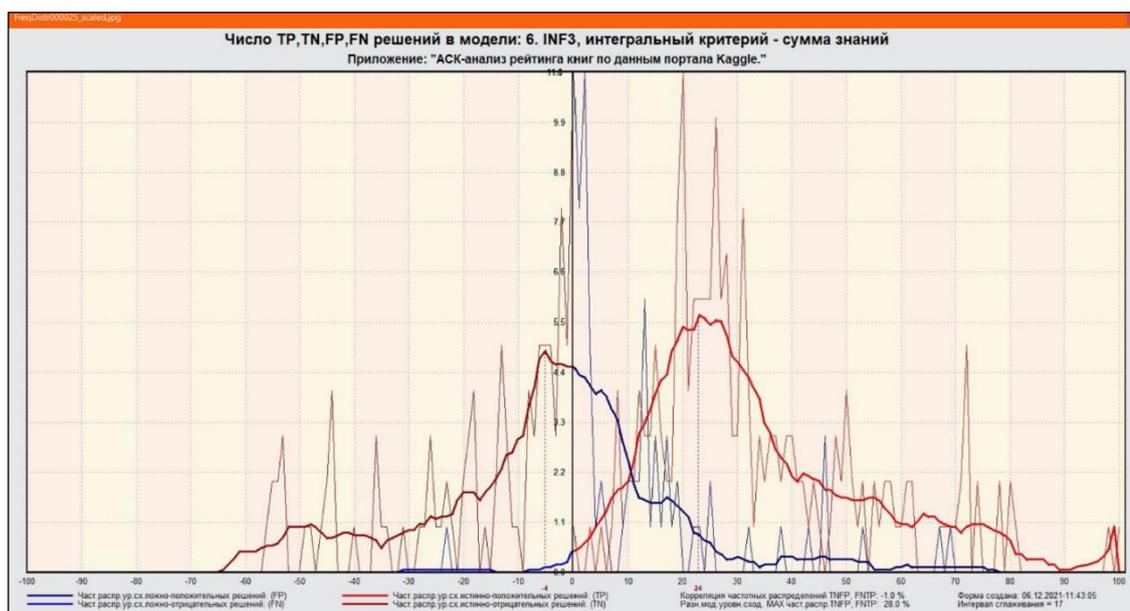


Рисунок 11 – Частотные распределения числа верно и ошибочно идентифицированных и не идентифицированных состояний объекта моделирования в зависимости от уровня сходства в модели INF3

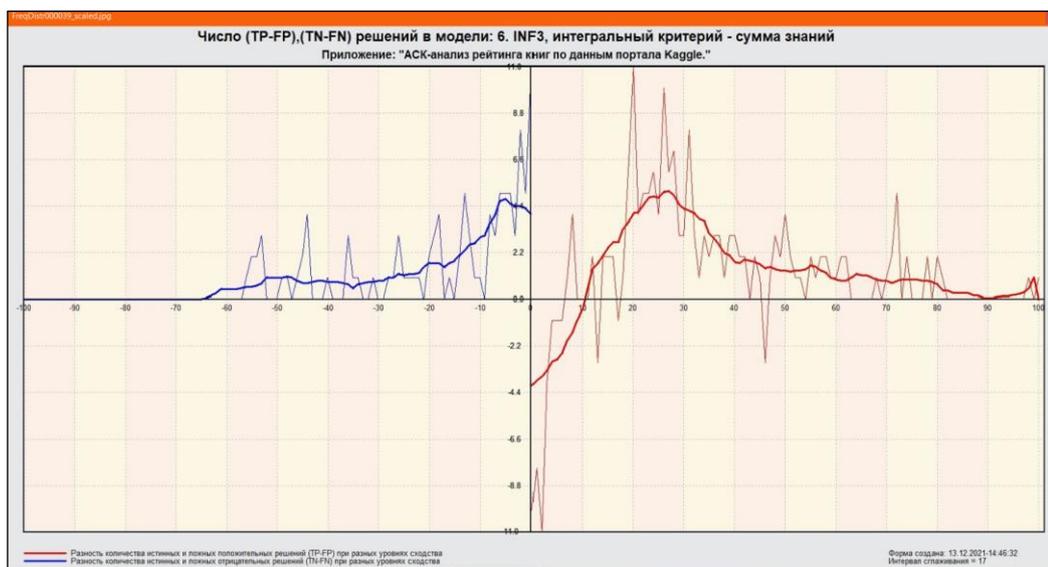


Рисунок 12 – Частотные распределения числа верно и ошибочно идентифицированных и не идентифицированных состояний объекта моделирования в зависимости от уровня сходства в модели INF3

Из рисунка 11 видно, что:

- модель INF3 лучше определяет непринадлежность объекта к классу, чем принадлежность;
- модуль уровня сходства-различия в модели INF3 для верно идентифицированных и верно не идентифицированных объектов значительно выше, чем для ошибочно идентифицированных и ошибочно не идентифицированных. Это верно практически для всего диапазона уровней сходства-различия, кроме небольших по модулю значений в диапазоне от 0 до 50% уровня сходства. Для уровней сходства-различия более 55% ошибочно идентифицированные и не идентифицированными ситуации практически отсутствуют.

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:

1. Подробно наглядно: «Объект – классы»;
2. Подробно наглядно: «Класс – объекты»;
3. Итоги наглядно: «Объект – классы»;
4. Итоги наглядно: «Класс – объекты»;
5. Подробно сжато: «Объект – классы»;
6. Обобщенная форма по достоверности моделей при разных интегральных критериях;
7. Обобщенный статистический анализ результатов идентификации по моделям и интегральным критериям;
8. Статистический анализ результатов идентификации по классам, моделям и интегральным критериям;
9. Распознавание уровня сходства при разных моделях и интегральных критериях;
10. Достоверность идентификации классов при разных моделях и интегральных критериях.

Ниже кратко рассмотрены некоторые из них.

На рисунках 15 и 16 приведены примеры прогнозов высокой и низкой достоверности частоты объектов и классов в модели INF3 на основе наблюдения предыстории их развития:

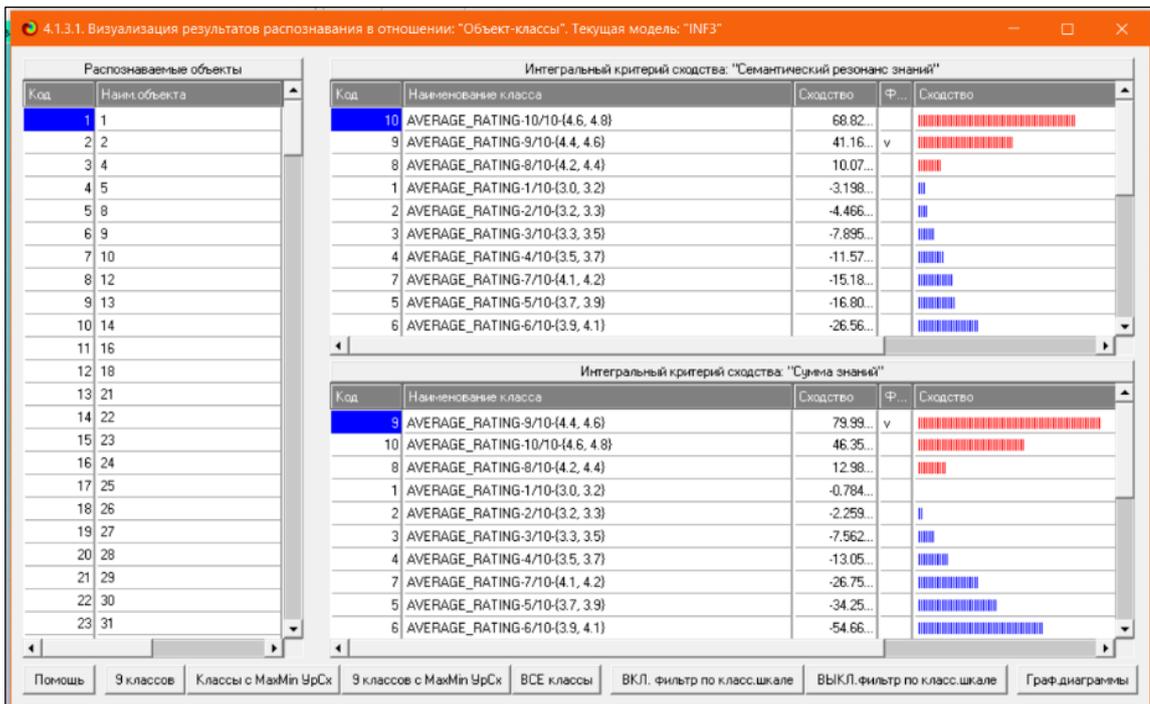


Рисунок 15 – Пример идентификации объектов в модели INF3

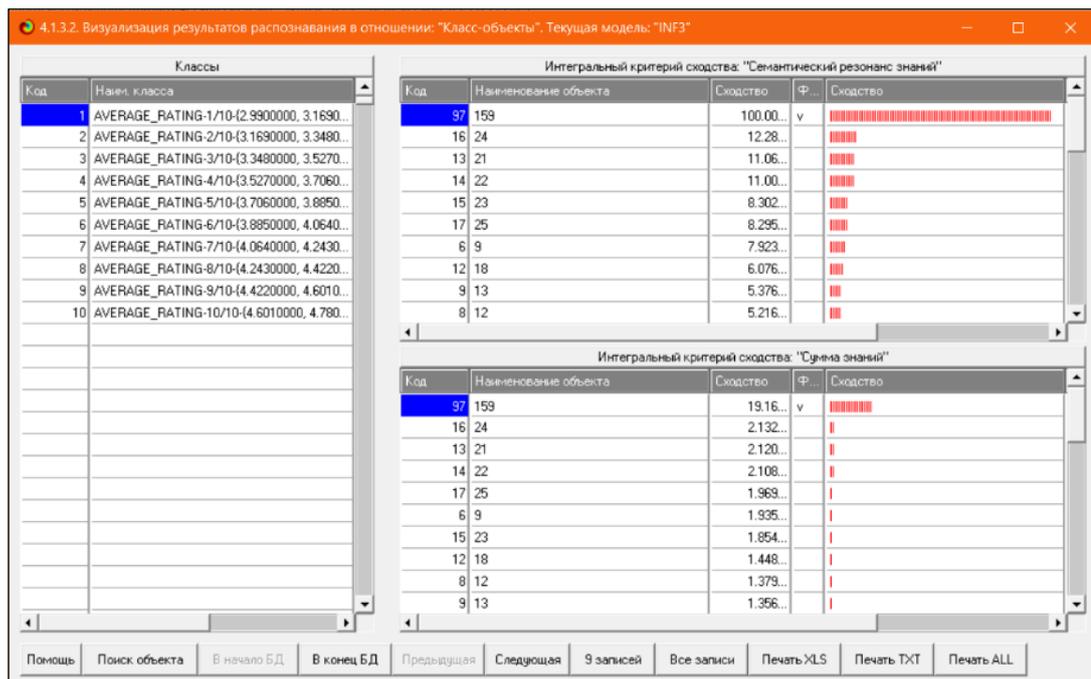


Рисунок 16 – Пример идентификации классов в модели INF3

3.2. Кластерно-конструктивный анализ

Сходство-различие обобщенных образов различных результатов научной деятельности по характерным для них системам значений

показателей. Результаты сравнения классов по системе характерных приведены на рисунке 17.

4.2.2.2. Результаты кластерно-конструктивного анализа классов

Конструкт класса: 1 "AVERAGE_RATING-1/10-{3.0, 3.2}" в модели: 4 "INF1"

Код	Наименование класса	№	Код класса	Наименование класса	Сходство
1	AVERAGE_RATING-1/10-{3.0, 3.2}	1	1	AVERAGE_RATING-1/10-{3.0, 3.2}	100.000
2	AVERAGE_RATING-2/10-{3.2, 3.3}	2	2	AVERAGE_RATING-2/10-{3.2, 3.3}	2.849
3	AVERAGE_RATING-3/10-{3.3, 3.5}	3	3	AVERAGE_RATING-3/10-{3.3, 3.5}	-0.573
4	AVERAGE_RATING-4/10-{3.5, 3.7}	4	10	AVERAGE_RATING-10/10-{4.6, 4.8}	-1.239
5	AVERAGE_RATING-5/10-{3.7, 3.9}	5	8	AVERAGE_RATING-8/10-{4.2, 4.4}	-2.279
6	AVERAGE_RATING-6/10-{3.9, 4.1}	6	9	AVERAGE_RATING-9/10-{4.4, 4.6}	-2.945
7	AVERAGE_RATING-7/10-{4.1, 4.2}	7	6	AVERAGE_RATING-6/10-{3.9, 4.1}	-3.087
8	AVERAGE_RATING-8/10-{4.2, 4.4}	8	4	AVERAGE_RATING-4/10-{3.5, 3.7}	-3.754
9	AVERAGE_RATING-9/10-{4.4, 4.6}	9	7	AVERAGE_RATING-7/10-{4.1, 4.2}	-5.358
10	AVERAGE_RATING-10/10-{4.6, 4.8}	10	5	AVERAGE_RATING-5/10-{3.7, 3.9}	-7.412

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 График ВКЛ. фильтр по кл.шкале ВЫКЛ. фильтр по кл.шкале Параметры Показать ВСЕ

Рисунок 17 – Результаты кластерно-конструктивного анализа классов

На рисунке 18 представлена семантическая сеть классов в модели «INF3».

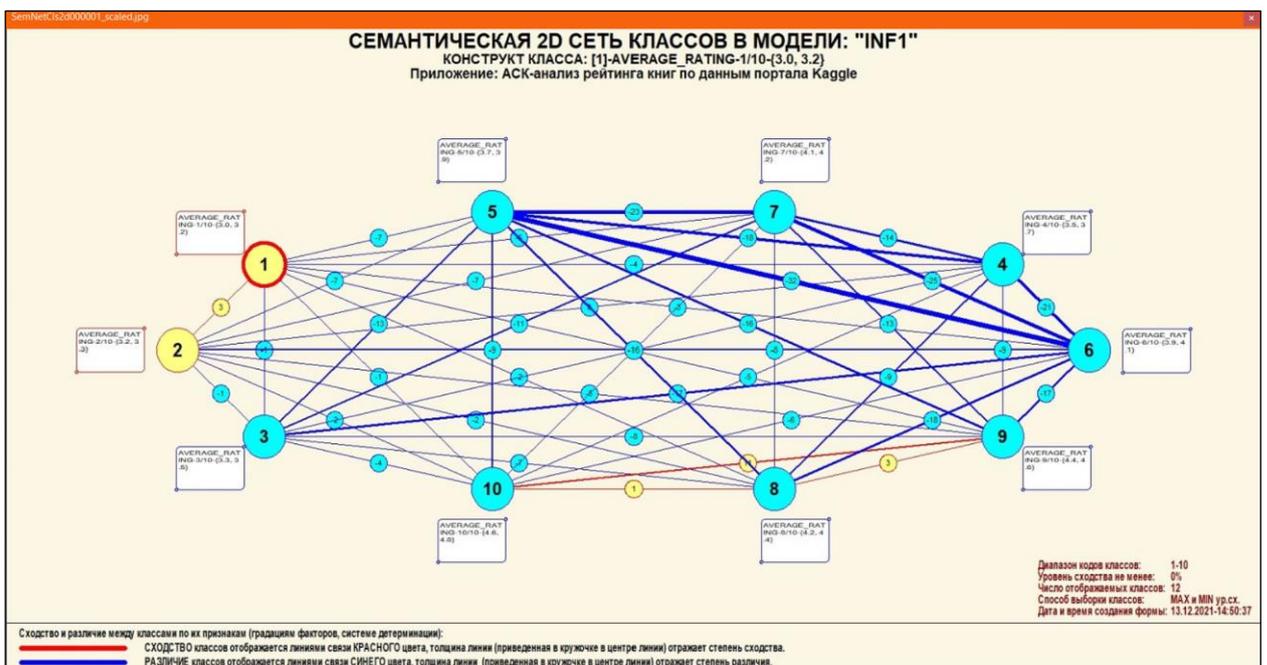


Рисунок 18 – Семантическая 2D сеть классов

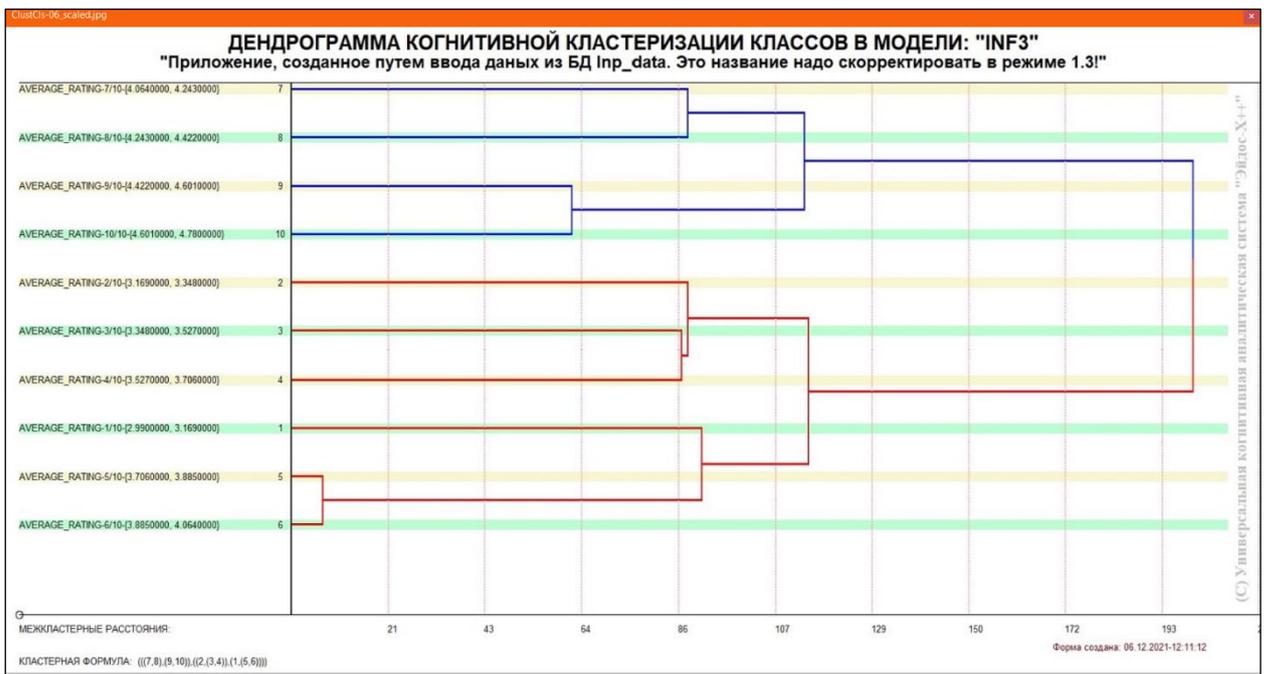


Рисунок 19 – Дендрограмма когнитивной кластеризации

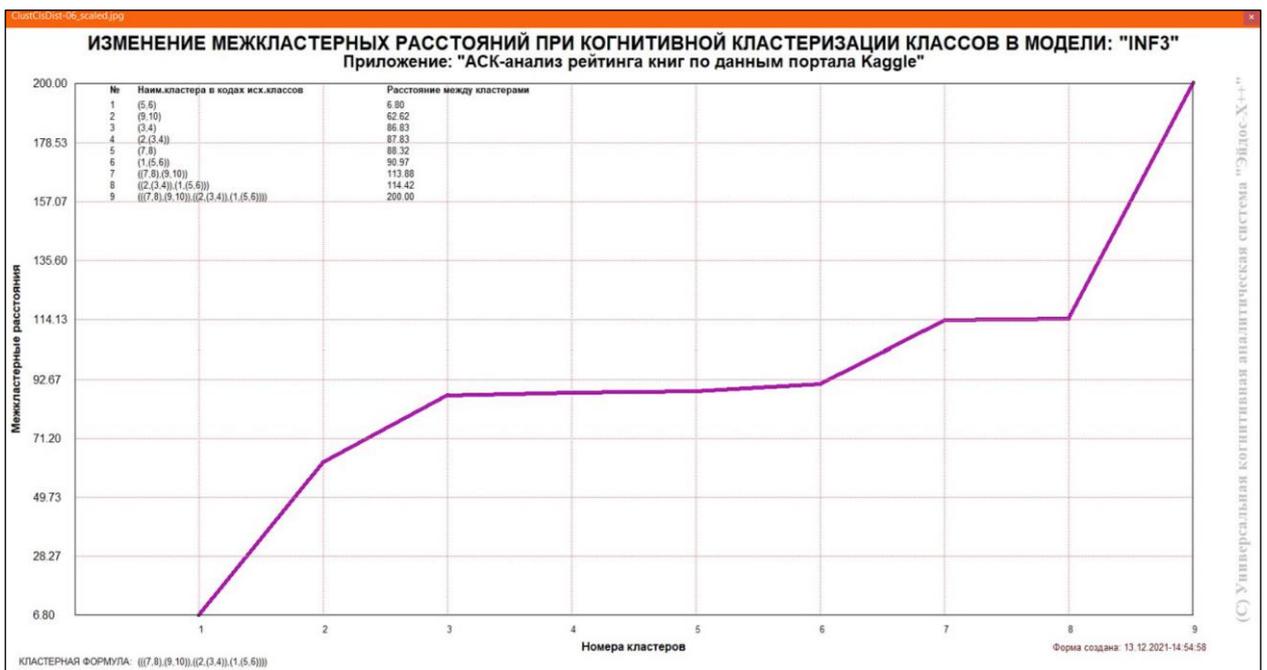


Рисунок 20 – График межкластерных расстояний классов

3.3. Нелокальные нейронные сети и нейтроны

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.

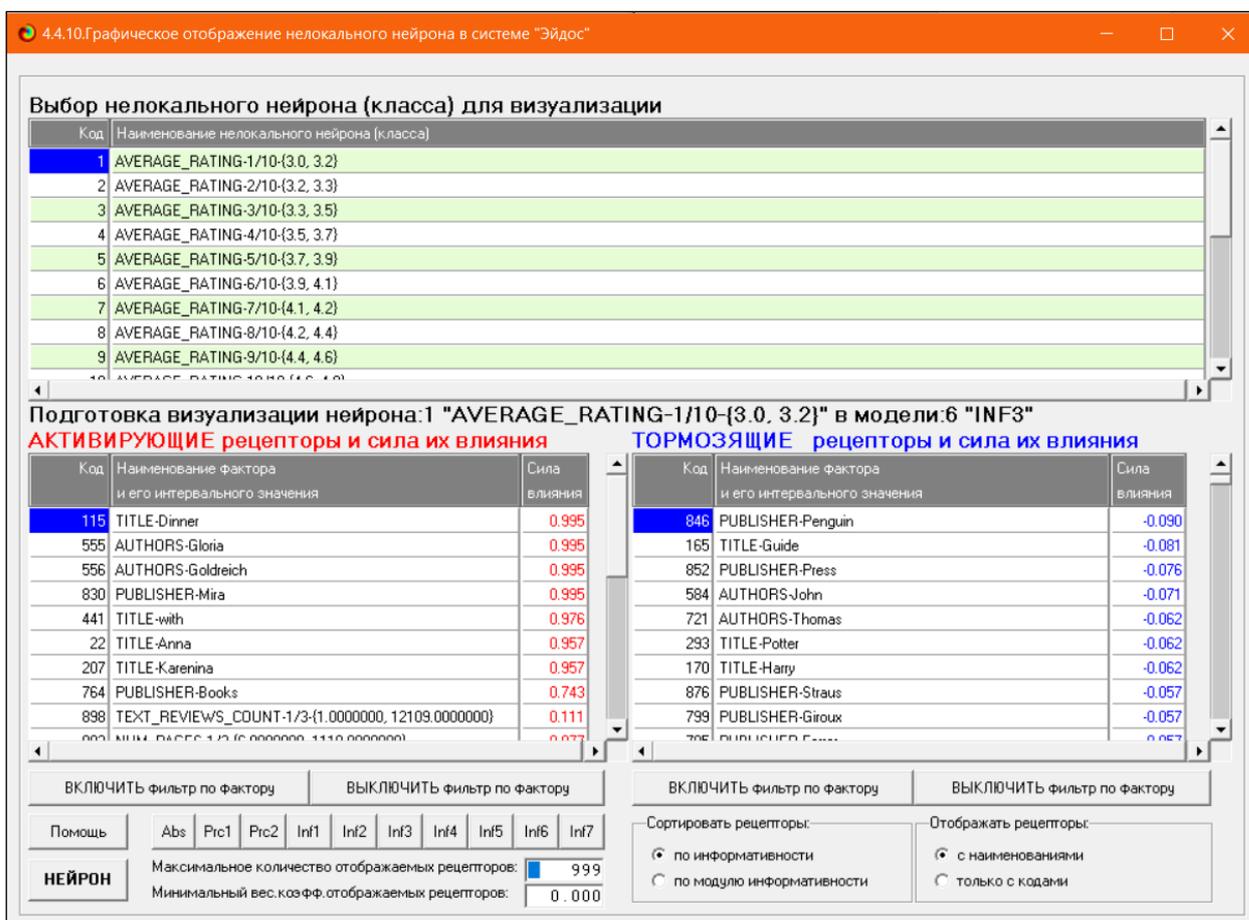


Рисунок 21 – Графическое отображение нелокальных нейронов в системе Aidos-X

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 – ее модуль.

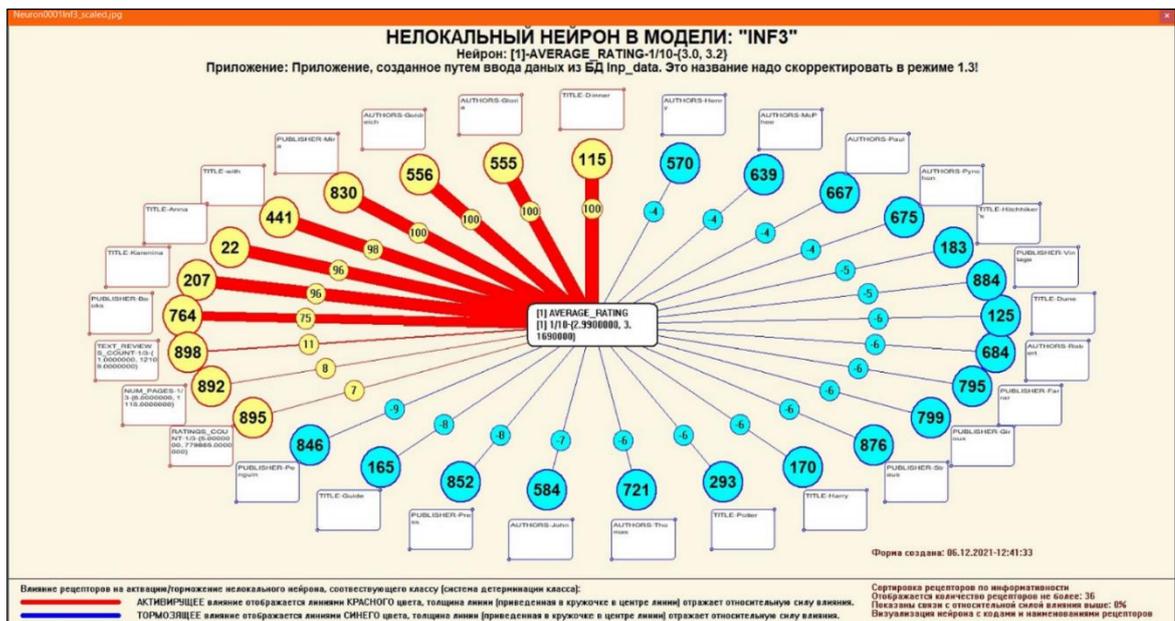


Рисунок 22 – Нелокальный нейрон в модели «INF3»

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[9].

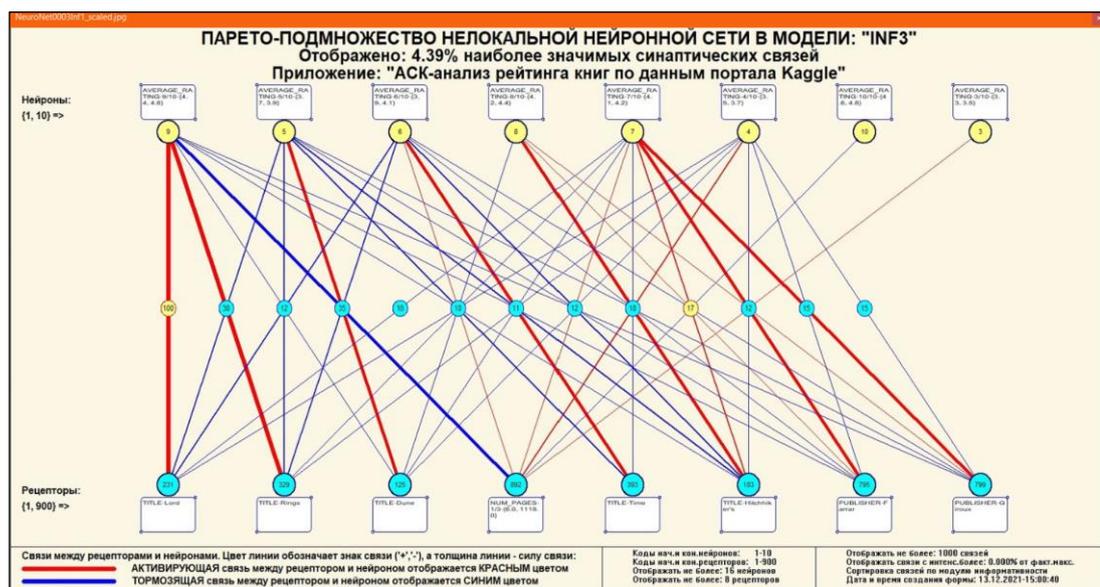


Рисунок 23 – Парето-подмножество нелокальной нейронной сети

3.4. SWOT и PEST матрицы и диаграммы

SWOT-анализ является широко известным и общепризнанным методом стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT анализа выявлено довольно много его слабых сторон (недостатков), источником которых является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают 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 (рисунок 27).

На рисунке 24 приведен пример табличной выходной формы количественного автоматизированного SWOT- и PEST- анализа средствами системы Aidos.

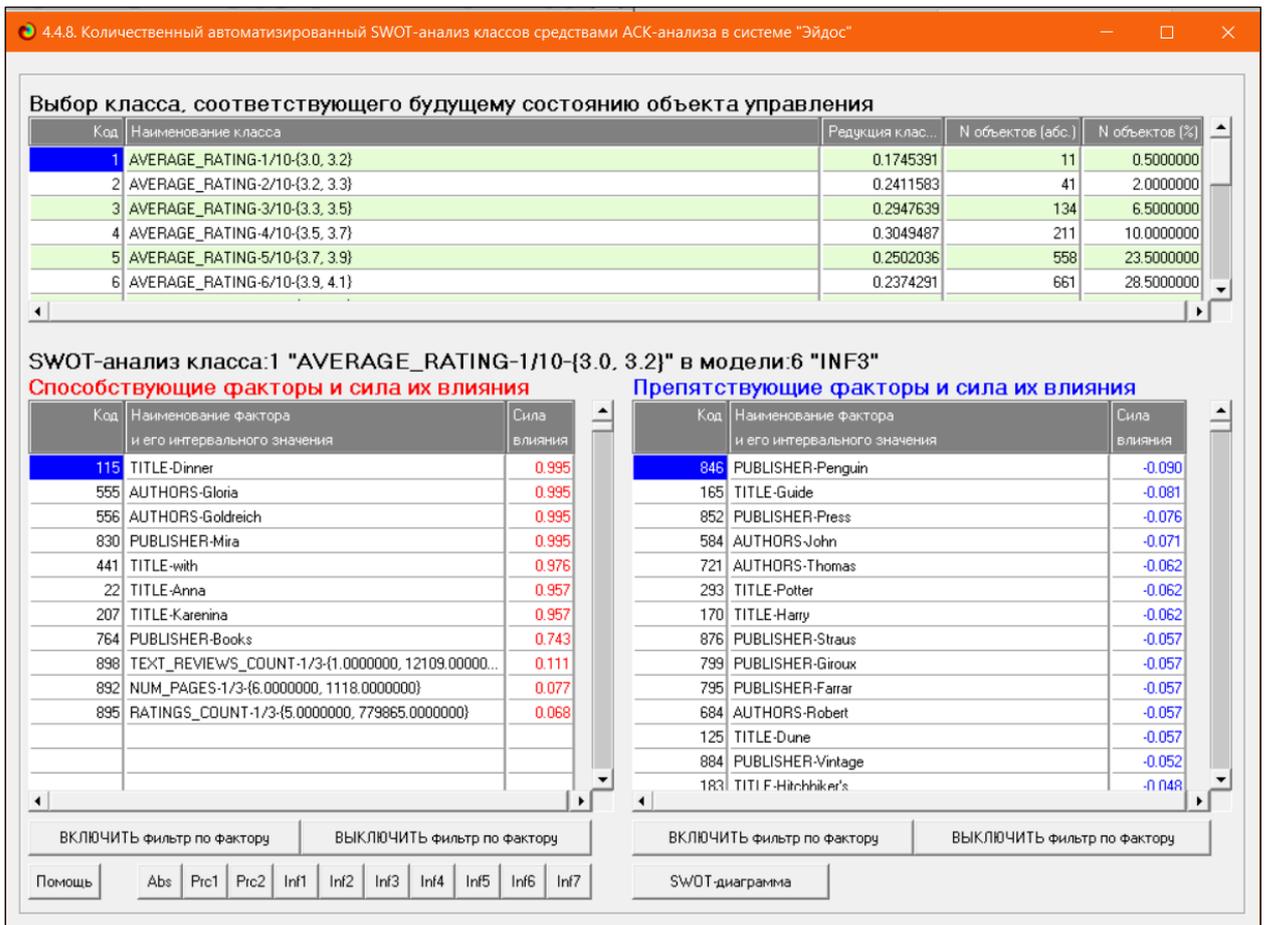


Рисунок 24 – Количественный автоматизированный SWOT-анализ

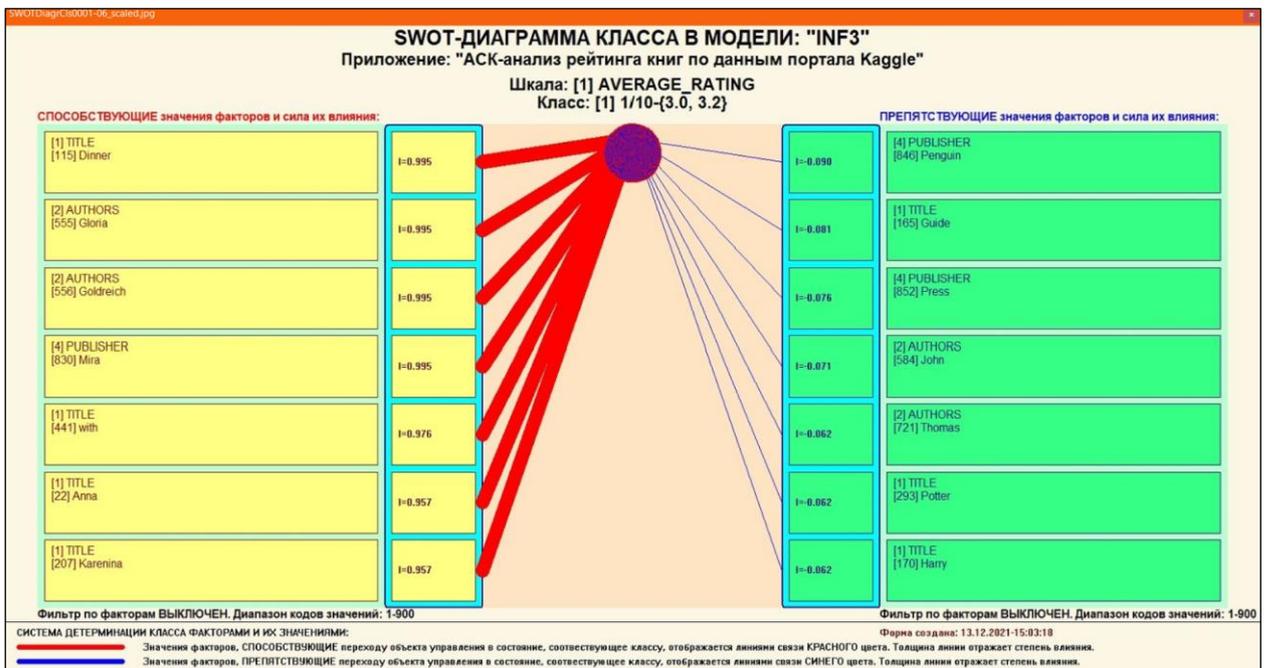


Рисунок 25 – Пример SWOT-матрицы в модели INF3

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 будущее состояние

Код	Наименование значения фактора
1	TITLE-"Half"
2	TITLE-"Hatchel"
3	TITLE-"Across"
4	TITLE-"Active"
5	TITLE-"Actors"
6	TITLE-"Acts"

SWOT-анализ значения фактора: 1 "TITLE-"Half" в модели: 6 "INF3"

СПОСОБСТВУЕТ:

Код	Состояния объекта управления, переходу в которые данное значение фактора СПОСОБСТВУЕТ	Сила влияния
5	AVERAGE_RATING-5/10-(3.7, 3.9)	0.759

ПРЕПЯТСТВУЕТ:

Код	Состояния объекта управления, переходу в которые данное значение фактора ПРЕПЯТСТВУЕТ	Сила влияния
6	AVERAGE_RATING-6/10-(3.9, 4.1)	-0.286
7	AVERAGE_RATING-7/10-(4.1, 4.2)	-0.147
4	AVERAGE_RATING-4/10-(3.5, 3.7)	-0.091
9	AVERAGE_RATING-9/10-(4.4, 4.6)	-0.076
8	AVERAGE_RATING-8/10-(4.2, 4.4)	-0.064
3	AVERAGE_RATING-3/10-(3.3, 3.5)	-0.058
2	AVERAGE_RATING-2/10-(3.2, 3.3)	-0.018
10	AVERAGE_RATING-10/10-(4.6, 4.8)	-0.014
1	AVERAGE_RATING-1/10-(3.0, 3.2)	-0.005

ВКЛЮЧИТЬ фильтр по шкале | ВЫКЛЮЧИТЬ фильтр по шкале

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

Рисунок 26 – Табличная выходная форма количественного автоматизированного SWOT- и PEST- анализа средствами системы Aidos

На рисунке 27 приведен пример графической выходной формы количественного автоматизированного SWOT- и PEST- анализа средствами системы Aidos.

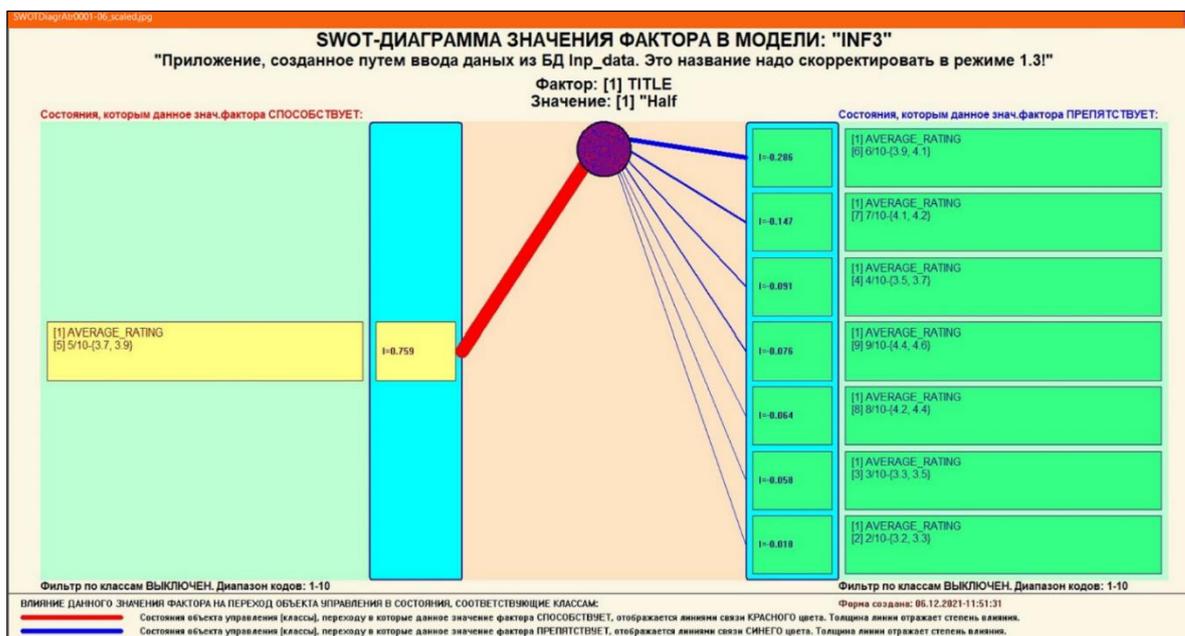


Рисунок 27 – Графическая выходная форма количественного автоматизированного SWOT- и PEST-анализа средствами системы Aidos

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.

Сверхзадачей искусственного интеллекта является построение компьютерной интеллектуальной системы, которая обладала бы уровнем эффективности решений неформализованных задач, сравнимым с человеческим или превосходящим его.

Самую существенную часть систем искусственного интеллекта составляют экспертные системы. Экспертная система обычно определяется как программа ЭВМ, моделирующая действия эксперта человека при решении задач в узкой предметной области: составление базы знаний и накопления их.

В данной работе был проведен анализ рейтинга книг системой искусственного интеллекта «Aidos-X++» с использованием общедоступной базы данных. При этом наиболее достоверной в данном приложении оказалась модель INF3, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,975, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 60%.

АСК-анализ текстов позволяет:

- формировать обобщенные лингвистические образы классов (семантические ядра) на основе фрагментов или примеров, относящихся к ним текстов на любом языке;
- количественно сравнивать лингвистический образ конкретного человека, или описание объекта, процесса с обобщенными лингвистическими образами групп (классов);

- сравнивать обобщенные лингвистические образы классов друг с другом и создавать их кластеры и конструкты;
- исследовать моделируемую предметную область путем исследования ее лингвистической системно-когнитивной модели;
- проводить интеллектуальную атрибуцию текстов, т.е. определять вероятное авторство анонимных и псевдонимных текстов, датировку, жанр и смысловую направленность содержания текстов;
- все это можно делать для любого естественного или искусственного языка, или системы кодирования.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

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: Куб- ГАУ, 2014. – №06(100). С. 1346 – 1395. – IDA [article ID]: 1001406090. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/90.pdf>, 3,125 у.п.л.

3. Луценко Е.В. 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. [Электронный ресурс]. Статья "Emergency – 911 Calls": <https://www.kaggle.com/mchirico/montcoalert>, свободный. - Загл. сэкрана. Яз.анг.

5. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: <http://lc.kubagro.ru/>, свободный. - Загл. с экрана. Яз.рус.

6. 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>.

7. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос- Х++» / 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.

8. 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.

9. 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.

10. Луценко Е.В. Формирование семантического ядра ветеринарии путем Автоматизированного системно-когнитивного анализа паспортов

научных специальностей ВАК РФ и автоматическая классификация текстов по направлениям науки / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №10(144). С. 44 – 102. – IDA [article ID]: 1441810033. – Режим доступа: <http://ej.kubagro.ru/2018/10/pdf/33.pdf>, 3,688 у.п.л.

11. Луценко Е.В. Интеллектуальная привязка некорректных ссылок к литературным источникам в библиографических базах данных с применением АСК-анализа и системы «Эйдос» (на примере Российского индекса научного цитирования – РИНЦ) / Е.В. Луценко, В.А. Глухов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №01(125). С. 1 – 65. – IDA [article ID]: 1251701001. – Режим доступа: <http://ej.kubagro.ru/2017/01/pdf/01.pdf>, 4,062 у.п.л.

12. Луценко Е.В. Применение АСК-анализа и интеллектуальной системы "Эйдос" для решения в общем виде задачи идентификации литературных источников и авторов по стандартным, нестандартным и некорректным библиографическим описаниям / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 498 – 544. – IDA [article ID]: 1031409032. – Режим доступа: <http://ej.kubagro.ru/2014/09/pdf/32.pdf>, 2,938 у.п.л.

13. Луценко Е.В. АСК-анализ проблематики статей Научного журнала КубГАУ в динамике / Е.В. Луценко, В.И. Лойко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №06(100). С. 109 – 145. – IDA [article ID]: 1001406007. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/07.pdf>, 2,312 у.п.л.

14. Луценко Е.В. Атрибуция анонимных и псевдонимных текстов в системно-когнитивном анализе / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 44 – 64. – IDA [article ID]: 0050403003. – Режим доступа: <http://ej.kubagro.ru/2004/03/pdf/03.pdf>, 1,312 у.п.л.

15. Луценко Е.В. Атрибуция текстов, как обобщенная задача идентификации и прогнозирования / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №02(002). С. 146 – 164. – IDA [article ID]: 0020302013. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/13.pdf>, 1,188 у.п.л.

16. Луценко Д.С., Луценко Е.В. Интеллектуальная датировка текста, определение авторства и жанра на примере русской литературы XIX и XX веков, 2020 // Статья в открытом архиве. 38 с. – DOI: 10.13140/RG.2.2.28824.01281, <https://www.elibrary.ru/item.asp?id=43796415>

17. Lutsenko D.S., Lutsenko E.V. Intellectual attribution of literary texts (finding the dates of the text, determining authorship and genre on the example of russian literature of the XIX and XX centuries), 2020 // Статья в открытом архиве. 9 р. – DOI: 10.13140/RG.2.2.15349.81122, <https://www.elibrary.ru/item.asp?id=43794562>