

УДК 004.8

09.03.02 Информационные системы и технологии, «Создание, поддержка и администрирование информационных систем»(программа бакалавриата)

ОБЗОР ТОП-48 БЕСТСЕЛЛЕРОВ AMAZON ЗА 2009–2019 ГГ.

Каширцев Егор Владимирович
egor70985@gmail.com

Мотяш Максим Леонидович
motyash.m@gmail.com

Кубанский государственный аграрный университет имени И.Т.Трубилина, Краснодар, Россия

По данным интернет ресурса Kaggle предложена регрессионная модель измерения топ-48 бестселлеров amazon за 2009–2019 гг. Эта модель отражает объективно существующие взаимосвязи между рейтингом, отзывами и ценой на знаменитые книги разных писателей на платформе amazon. Эти взаимосвязи выявлены авторами способа путем математической обработке 48 самых популярных книг на amazon. Для этого ими был применен регрессионный и корреляционный анализ. Так как количество данных по книгам довольно велико, то данная задача относится к задачам квалиметрии. Однако, данная задача квалиметрии относится также к типичным задачам распознавания образов (многопараметрической типизации и системной идентификации), и, поэтому вполне возможно, что качество ее решения может быть повышено путем применения методов искусственного интеллекта, в частности когнитивных и информационных технологий. Тем более, что эти взаимосвязи имеют довольно сложный характер. Применение интеллектуальных технологий открывает также дополнительные возможности для исследования моделируемой предметной области путем исследования ее модели. Все это представляет большой научный и практический интерес как для ученых исследователей, а так и для практиков. Для решения всех этих задач в работе применен Автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальная система «Эйдос». Подробно рассмотрен численный пример, основанный на реальных данных.

UDC 004.8

03.09.02 Information systems and technologies, "Creation, support and administration of information systems" (bachelor's program)

OVERVIEW OF TOP-48 AMAZON BEST SELLERS FOR 2009–2019.

Egor Kashirtsev
egor70985@gmail.com

Motyash Maxim
motyash.m@gmail.com

Kuban State Agrarian University named after I.T.Trubilin, Krasnodar, Russia

According to the Internet resource Kaggle, a regression model for measuring the top 48 bestsellers of amazon for 2009–2019 has been proposed. This model reflects the objectively existing relationship between the rating, reviews and the price of famous books by different writers on the amazon platform. These relationships were revealed by the authors of the method by mathematical processing of 48 of the most popular books on amazon. For this, they applied regression and correlation analysis. Since the amount of data on books is quite large, this task belongs to the tasks of qualimetry. However, this task of qualimetry also refers to typical tasks of pattern recognition (multi-parameter typification and system identification), and, therefore, it is quite possible that the quality of its solution can be improved by using artificial intelligence methods, in particular, cognitive and information technologies. Moreover, these relationships are rather complex. The use of intelligent technologies also opens up additional opportunities for researching a simulated subject area by examining its model. All this is of great scientific and practical interest both for scientific researchers and for practitioners. To solve all these problems, the work uses the Automated System-Cognitive Analysis (ASK-analysis) and its software toolkit - the intelligent system "Eidos". A numerical example based on real data is considered in detail.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	2
ЗАДАЧА 1: КОГНИТИВНАЯ СТРУКТУРИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ.....	6
ЗАДАЧА 2: ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ И ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ	7
ЗАДАЧА 3: СИНТЕЗ И ВЕРИФИКАЦИЯ СТАТИСТИЧЕСКИХ И СИСТЕМНО-КОГНИТИВНЫХ МОДЕЛЕЙ И ВЫБОР НАИБОЛЕЕ ДОСТОВЕРНОЙ ИЗ НИХ ДЛЯ РЕШЕНИЯ ЗАДАЧ	15
ЗАДАЧА 4: РЕШЕНИЕ РАЗЛИЧНЫХ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	22
ПОДЗАДАЧА 4.1. ПРОГНОЗИРОВАНИЕ (ДИАГНОСТИКА, КЛАССИФИКАЦИЯ, РАСПОЗНАВАНИЕ, ИДЕНТИФИКАЦИЯ).....	22
ПОДЗАДАЧА 4.2. ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ (SWOT-АНАЛИЗ)	24
ПОДЗАДАЧА 4.3. ИССЛЕДОВАНИЕ МОДЕЛИРУЕМОЙ ПРЕДМЕТНОЙ ОБЛАСТИ ПУТЕМ ИССЛЕДОВАНИЯ ЕЕ МОДЕЛИ.....	29
4.3.1. Когнитивные диаграммы классов.....	29
4.3.2. Агломеративная когнитивная кластеризация классов.....	30
4.3.3. Когнитивные диаграммы значений факторов	32
4.3.4. Агломеративная когнитивная кластеризация значений факторов	33
4.3.5. Нелокальные нейроны и нелокальные нейронные сети.....	35
4.3.6. 3d-интегральные когнитивные карты.....	36
4.3.7. Когнитивные функции.....	37
4.3.8. Сила и направление влияния значений параметров популярности книг и сила влияния самых параметров популярности на авторов и года книг	40
4.3.9. Степень детерминированности пола и возраста пострадавших значениями параметров количества.....	42
4.3.10. Устойчивость содержания рейтинга и цен от значений параметров авторов книг.....	43
7. ВЫВОДЫ.....	45
ЛИТЕРАТУРА.....	46

Введение

По данным интернет ресурса Kaggle предложена регрессионная модель измерения топ-48 бестселлеров amazon за 2009–2019 гг. [16]

Эта модель отражает объективно существующие взаимосвязи между рейтингом, отзывами и ценой на знаменитые книги разных писателей на платформе amazon. Эти взаимосвязи выявлены авторами способа путем математической обработке 48 самых популярных книг на amazon.

Так как количество данных по книгам довольно велико, то данная задача относится к задачам квалиметрии. Однако **проблема** состоит в том, что сама зависимость популярности крайне неординарна и ее сложно адекватно отразить в формальной модели. Но данная задача относится не только к квалиметрии, но и к типичным задачам распознавания образов (многопараметрической типизации, т.е. формирования обобщенных

образов классов на основе описаний объектов, и системной идентификации объектов, т.е. определения принадлежности объектов к классам), и, поэтому возможно, что качество ее решения может быть повышено путем применения более адекватных для подобных задач методов искусственного интеллекта, в частности когнитивных и информационных технологий. Применение интеллектуальных технологий открывает также дополнительные возможности для исследования моделируемой предметной области путем исследования ее модели. Все это представляет большой научный и практический интерес как для ученых исследователей, а так и для практиков.

Поэтому **целью** данной работы является разработка интеллектуальных моделей, отражающих реально существующие причинно-следственные взаимосвязи между рейтингом книги, отзывами на нее и годом ее написания автором и применение этих моделей как для косвенного измерения общей оценки книг, так и для исследования моделируемой предметной области.

Для достижения поставленной цели в данной работе предлагается применить Автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальная система «Эйдос».

АСК-анализ предполагает, что для достижения поставленной цели необходимо решить следующие задачи, которые получаются путем декомпозиции цели и являются этапами ее достижения:

Задача 1: когнитивная структуризация предметной области.

Задача 2: подготовка исходных данных и формализация предметной области.

Задача 3: синтез и верификация статистических и системно-когнитивных моделей и выбор наиболее достоверной модели.

Задача 4: решение различных задач в наиболее достоверной модели:

- подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация);

- подзадача 4.2. Поддержка принятия решений;

- подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели (когнитивные диаграммы классов и значений факторов, агломеративная когнитивная кластеризация классов и значений факторов, нелокальные нейроны и нейронные сети, 3d-интегральные когнитивные карты, когнитивные функции и др.).

Эти задачи по сути представляют собой **этапы** Автоматизированного системно-когнитивный анализа (АСК-анализ), который и поэтому и предлагается применить для их решения.

АСК-анализ представляет собой метод искусственного интеллекта, разработанный проф. Е.В. Луценко в 2002 году для решения широкого класса задач идентификации, прогнозирования, классификации,

диагностики, поддержки принятия решений и исследования моделируемой предметной области путем исследования ее модели. АСК-анализ доведен до **инновационного** уровня благодаря тому, что имеет свой программный инструментарий – универсальную когнитивную аналитическую систему «Эйдос-X++» (система «Эйдос»).

Система «Эйдос» выгодно отличается от других интеллектуальных систем следующими параметрами:

- разработана в универсальной постановке, не зависящей от предметной области. Поэтому она является универсальной и может быть применена во многих предметных областях (<http://lc.kubagro.ru/aidos/index.htm>);

- находится в полном открытом бесплатном доступе (http://lc.kubagro.ru/aidos/_Aidos-X.htm), причем с актуальными исходными текстами (http://lc.kubagro.ru/_AIDOS-X.txt);

- является одной из первых отечественных систем искусственного интеллекта персонального уровня, т.е. она не требует от пользователя специальной подготовки в области технологий искусственного интеллекта (есть акт внедрения системы «Эйдос» 1987 года) (<http://lc.kubagro.ru/aidos/aidos02/PR-4.htm>);

- обеспечивает устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных взаимозависимых (нелинейных) данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения (т.е. не предъявляет жестких требований к данным, которые невозможно выполнить, а обрабатывает те данные, которые есть);

- содержит большое количество локальных (поставляемых с инсталляцией) и облачных учебных и научных приложений (в настоящее время их 31 и 300, соответственно) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);

- обеспечивает мультиязычную поддержку интерфейса на 44 языках. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;

- поддерживает on-line среду накопления знаний и широко используется во всем мире (<http://aidos.byethost5.com/map5.php>);

- 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, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний;

- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач

классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе: http://lc.kubagro.ru/aidos/aidos18_LLS/aidos18_LLS.pdf);

- хорошо имитирует человеческий стиль мышления: дает результаты анализа, понятные экспертам на основе их опыта, интуиции и профессиональной компетенции.

Всем этим и обусловлен выбор АСК-анализа и его программного инструментария – интеллектуальной системы «Эйдос» в качестве метода и инструмента решения поставленной проблемы и достижения цели работы (рисунок 1).

Последовательность обработки данных, информации и знаний в системе «Эйдос-X++»

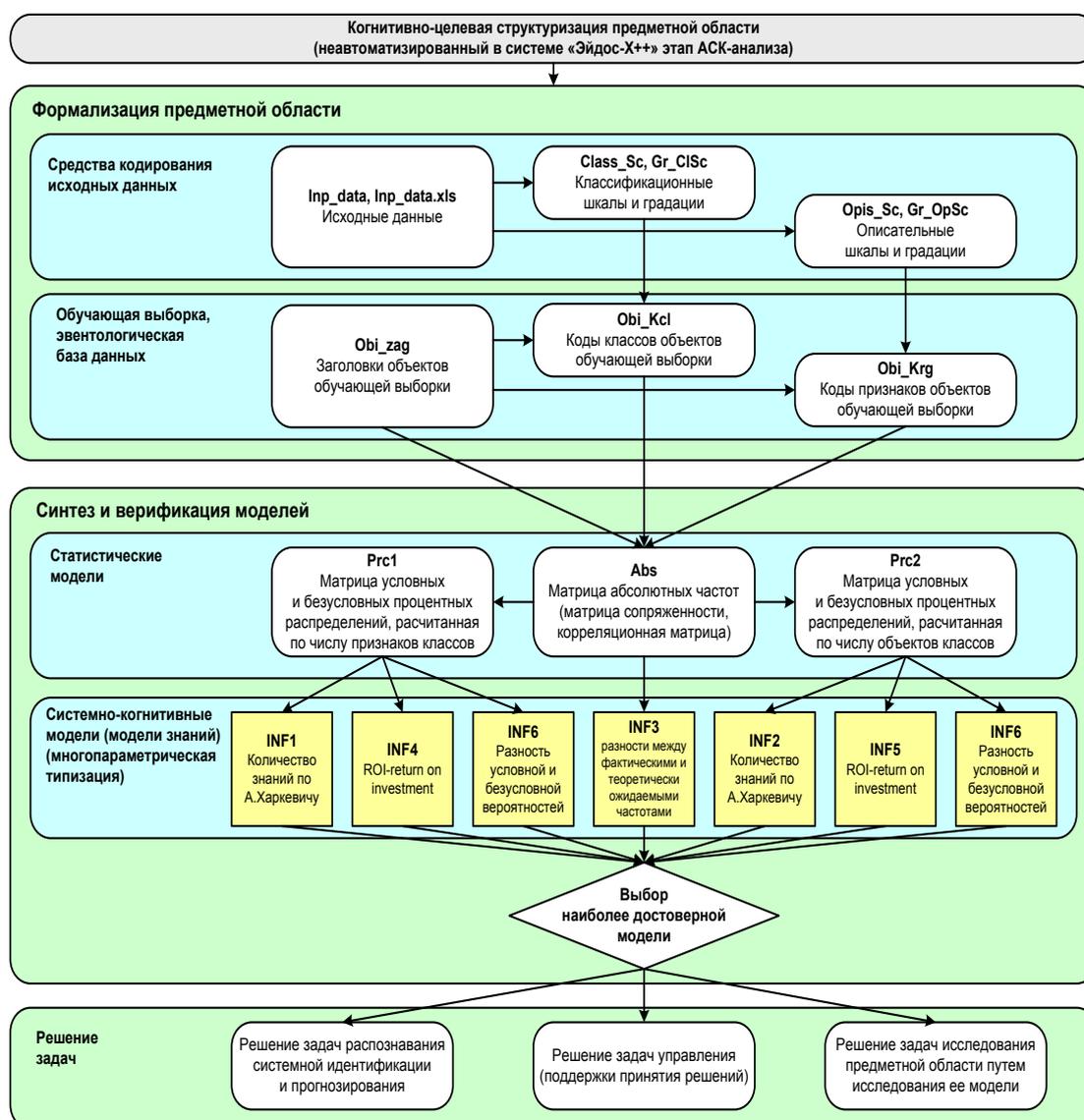


Рисунок 1. Последовательность решения задач в АСК-анализе и системе «Эйдос»

Рассмотрим решение поставленных задач в подробном численном примере.

Задача 1: когнитивная структуризация предметной области

На этапе когнитивно-целевой структуризации предметной области мы неформализуемым путем решаем на качественном уровне, что будем рассматривать в качестве факторов, действующих на моделируемый объект (причин), а что в качестве результатов действия этих факторов (последствий).

При этом необходимо отметить, что системно-когнитивные модели (СК-модели) отражают лишь сам факт наличия зависимостей между значениями факторов и результатами их действия. Но они не отражают причин и механизмов такого влияния.

Это значит:

– *во-первых*, что содержательная интерпретация СК-моделей – это компетенция специалистов-экспертов хорошо разбирающихся в данной предметной области. Иногда встречается ситуация, когда и то, что на первый взгляд является причинами, и то, что казалось бы является их последствиями, на самом деле является последствиями неких глубинных причин, которых мы не видим и никоим образом непосредственно не отражаем в модели;

– *во-вторых*, даже если содержательной интерпретации не разработано, то в принципе это не исключает возможности пользоваться ими на практике для достижения заданных результатов и поставленных целей, т.е. для управления.

В данной работе в качестве классификационных шкал выберем рейтинг читателей, отзывов и цену книги в (\$) (таблица 1), а в качестве факторов, влияющих на эти результаты – од выпуска и автор (таблица 2):

Таблица 1 – Классификационные шкалы

Код	Наименование
1	Рейтинг читателей
2	Отзывов
3	Цена(\$)

Таблица 2 – Описательные шкалы

Код	Наименование
1	Год выпуска
2	Автор

Расшифровка наименований параметров измерения топ-48 бестселлеров amazon за 2009–2019 гг. ничего не скажет специалистам в области специальностей, связанных с литературой, на которых ориентирована данная работа, и здесь не приводится.

Задача 2: подготовка исходных данных **и формализация предметной области**

Исходные данные для данной статьи (таблица 3) взяты с интернет ресурса Kaggle. Исходные данные приведены полностью. Единственное что сделано: введены условные обозначения наименований параметров измерения топ-48 бестселлеров amazon за 2009–2019 гг.

Таблица 3 – Исходные экспериментальные данные
для ввода в систему «Эйдос»¹

Название	Рейтинг читателей	Отзывов	Цена(\$)	Год выпуска	Автор
10-Day Green Smoothie Cleanse	4,7	17350	8	2016	JJ Smith
11/22/63: A Novel	4,6	2052	22	2011	Stephen King
12 Rules for Life: An Antidote to Chaos	4,7	18979	15	2018	Jordan B. Peterson
1984 (Signet Classics)	4,7	21424	6	2017	George Orwell
5,000 Awesome Facts (About Everything!) (National Geographic Kids)	4,8	7665	12	2019	National Geographic Kids
A Dance with Dragons (A Song of Ice and Fire)	4,4	12643	11	2011	George R. R. Martin
A Game of Thrones / A Clash of Kings / A Storm of Swords / A Feast of Crows / A Dance with Dragons	4,7	19735	30	2014	George R. R. Martin
The Very Hungry Caterpillar	4,9	19546	5	2019	Eric Carle
The Whole30: The 30-Day Guide to Total Health and Food Freedom	4,6	7508	16	2015	Melissa Hartwig Urban
The Wonderful Things You Will Be	4,9	8842	10	2016	Emily Winfield Martin
The Wright Brothers	4,7	6169	16	2015	David McCullough
Things That Matter: Three Decades of Passions, Pastimes and Politics [Deckled	4,7	7034	15	2013	Charles Krauthammer

Edge]					
Thinking, Fast and Slow	4,6	11034	19	2011	Daniel Kahneman
Thirteen Reasons Why	4,5	7932	9	2017	Jay Asher
Thomas Jefferson: The Art of Power	4,5	1904	23	2012	Jon Meacham
Three Cups of Tea: One Man's Mission to Promote Peace - One School at a Time	4,3	3319	11	2009	Greg Mortenson
Thug Kitchen: The Official Cookbook: Eat Like You Give a F*ck (Thug Kitchen Cookbooks)	4,6	11128	23	2014	Thug Kitchen
Tina Fey: Bossypants	4,3	5977	12	2011	Tina Fey
To Kill a Mockingbird	4,8	26234	0	2013	Harper Lee
Tools of Titans: The Tactics, Routines, and Habits of Billionaires, Icons, and World-Class Performers	4,6	4360	21	2017	Timothy Ferriss
Towers of Midnight (Wheel of Time, Book Thirteen)	4,8	2282	21	2010	Robert Jordan
True Compass: A Memoir	4,5	438	15	2009	Edward M. Kennedy
Twilight (The Twilight Saga, Book 1)	4,7	11676	9	2009	Stephenie Meyer
Ultimate Sticker Book: Frozen: More Than 60 Reusable Full-Color Stickers	4,5	2586	5	2014	DK
Unbroken: A World War II Story of Survival, Resilience, and Redemption	4,8	29673	16	2010	Laura Hillenbrand
Under the Dome: A Novel	4,3	6740	20	2009	Stephen King
Unfreedom of the Press	4,9	5956	11	2019	Mark R. Levin
Unicorn Coloring Book: For Kids Ages 4-8 (US Edition) (Silly Bear Coloring Books)	4,8	6108	4	2019	Silly Bear
Uninvited: Living Loved When You Feel Less Than, Left Out, and Lonely	4,7	4585	9	2016	Lysa TerKeurst

Watchmen	4,8	3829	42	2009	Alan Moore
Water for Elephants: A Novel	4,5	8958	12	2011	Sara Gruen
What Happened	4,6	5492	18	2017	Hillary Rodham Clinton
What If?: Serious Scientific Answers to Absurd Hypothetical Questions	4,7	9292	17	2014	Randall Munroe
What Pet Should I Get? (Classic Seuss)	4,7	1873	14	2015	Dr. Seuss
What Should Danny Do? (The Power to Choose Series)	4,8	8170	13	2019	Adir Levy
What to Expect When You're Expecting	4,4	3341	9	2011	Heidi Murkoff
Wheat Belly: Lose the Wheat, Lose the Weight, and Find Your Path Back to Health	4,4	7497	6	2012	William Davis
When Breath Becomes Air	4,8	13779	14	2016	Paul Kalanithi
Where the Crawdads Sing	4,8	87841	15	2019	Delia Owens
Where the Wild Things Are	4,8	9967	13	2009	Maurice Sendak
Whose Boat Is This Boat?: Comments That Don't Help in the Aftermath of a Hurricane	4,6	6669	12	2018	The Staff of The Late Show withÂ...
Wild: From Lost to Found on the Pacific Crest Trail	4,4	17044	18	2012	Cheryl Strayed
Winter of the World: Book Two of the Century Trilogy	4,5	10760	15	2012	Ken Follett
Women Food and God: An Unexpected Path to Almost Everything	4,2	1302	11	2010	Geneen Roth
Wonder	4,8	21625	9	2013	R. J. Palacio
Wrecking Ball (Diary of a Wimpy Kid Book 14)	4,9	9413	8	2019	Jeff Kinney
You Are a Badass: How to Stop Doubting Your Greatness and Start Living an Awesome Life	4,7	14331	8	2016	Jen Sincero

Затем с параметрами, показанными на рисунке 2, запустим режим 2.3.2.2 системы «Эйдос», представляющий собой автоматизированный программный интерфейс (API) с внешними данными табличного типа. На рисунке 2 приведены реально использованные параметры.

2.3.2.2. Универсальный программный интерфейс импорта данных в систему "ЭЙДОС-X++"

Автоматическая формализация предметной области: генерация классификационных и описательных шкал и градаций, а также обучающей и распознаваемой выборки на основе базы исходных данных: "Inp_data"

Задайте тип файла исходных данных: "Inp_data":

- XLS - MS Excel-2003 Стандарт XLS-файла
- XLSX- MS Excel-2007(2010) Стандарт XLSX-файла
- DBF - DBASE IV (DBF/NTX) Стандарт DBF-файла
- CSV - CSV => DBF конвертер Стандарт CSV-файла

Задайте параметры:

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

Требования к файлу исходных данных

Задайте диапазон столбцов классификационных шкал:

Начальный столбец классификационных шкал:

Конечный столбец классификационных шкал:

Задайте диапазон столбцов описательных шкал:

Начальный столбец описательных шкал:

Конечный столбец описательных шкал:

Задайте режим:

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

Задайте способ выбора размера интервалов:

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

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

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

Параметры интерпретации значений текстовых полей "Inp_data":

Интерпретация TXT-полей классов:

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

Интерпретация TXT-полей признаков:

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

Ok Cancel

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

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [9 x 55]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс.шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис.шкалу
Числовые	3	9	3,00	1	10	10,00
Текстовые	0	0	0,00	1	45	45,00
ВСЕГО:	3	9	3,00	2	55	27,50

Задайте число интервалов (градаций) в шкалах:

В классификационных шкалах:

В описательных шкалах:

Пересчитать шкалы и градации Выйти на создание модели

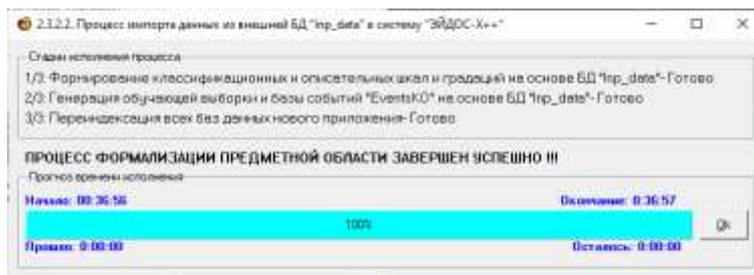


Рисунок 2. Экранные форма программного интерфейса (API) 2.3.2.2 системы «Эйдос» с внешними данными табличного типа

Обратим внимание на то, что заданы адаптивные интервалы, учитывающее неравномерность распределения данных по диапазону значений, что важно при относительно небольшом числе наблюдений. Если бы интервалы были заданы равными по величине, то в них бы учитывалось сильно отличающееся число наблюдений, а в некоторых интервалах их бы могло не оказаться вовсе. И в классификационных, и в описательных шкалах задано 5 числовых интервальных значения.

На рисунке 3 приведен Help данного режима, в котором объясняется принцип организации таблицы исходных данных для данного режима. Здесь же обратим внимание на то, что в таблице 3 как значения рейтинга книги, так и отзывы на нее могут быть представлены как числовыми, так и текстовыми значениями.



Рисунок 3. Экранные формы HELP программного интерфейса (API) 2.3.2.2²

² Все рисунки в статье приведены с достаточно высоким разрешением и при увеличении масштаба просмотра вполне читабельны

В результате работы режима сформировано 3 классификационных шкалы с суммарным количеством градаций (классов) 9 (таблица 4) и 2 описательных шкал с суммарным числом градаций 55. С использованием классификационных и описательных шкал и градаций (таблицы 4 и 5) исходные данные (таблица 3) были закодированы и в результате получена обучающая выборка (таблица 6):

Таблица 4 – Классификационные шкалы и градации
(рейтинг читателей, отзывы и цена на книгу)

KOD_CLSC	NAME_CLSC
1	РЕЙТИНГ ЧИТАТЕЛЕЙ
2	ОТЗЫВОВ
3	ЦЕНА(\$)

Таблица 5 – Описательные шкалы и градации
(параметры тензиограмм динамического поверхностного натяжения
на границе раздела молоко/воздух)

KOD_OPSC	NAME_OPSC
1	ГОД ВЫПУСКА
2	АВТОР

Таблица 6 – Обучающая выборка

NAME_OBJ	N2	N3	N4	N5
10-Day Green Smoothie Cleanse	3	4	7	7
11/22/63: A Novel	2	4	8	2
12 Rules for Life: An Antidote to Chaos	3	4	8	9
1984 (Signet Classics)	3	4	7	8
5,000 Awesome Facts (About Everything!) (National Geographic Kids)	3	4	7	10
A Dance with Dragons (A Song of Ice and Fire)	1	4	7	2
A Game of Thrones / A Clash of Kings / A Storm of Swords / A Feast of Crows / A Dance with Dragons	3	4	9	5
The Very Hungry Caterpillar	3	4	7	10
The Whole30: The 30-Day Guide to Total Health and Food Freedom	2	4	8	6
The Wonderful Things You Will Be	3	4	7	7

The Wright Brothers	3	4	8	6
Things That Matter: Three Decades of Passions, Pastimes and Politics [Deckled Edge]	3	4	8	4
Thinking, Fast and Slow	2	4	8	2
Thirteen Reasons Why	2	4	7	8
Thomas Jefferson: The Art of Power	2	4	8	3
Three Cups of Tea: One Man's Mission to Promote Peace - One School at a Time	1	4	7	1
Thug Kitchen: The Official Cookbook: Eat Like You Give a F*ck (Thug Kitchen Cookbooks)	2	4	8	5
Tina Fey: Bossypants	1	4	7	2
To Kill a Mockingbird	3	4	7	4
Tools of Titans: The Tactics, Routines, and Habits of Billionaires, Icons, and World-Class Performers	2	4	8	8
Towers of Midnight (Wheel of Time, Book Thirteen)	3	4	8	1
True Compass: A Memoir	2	4	8	1
Twilight (The Twilight Saga, Book 1)	3	4	7	1
Ultimate Sticker Book: Frozen: More Than 60 Reusable Full-Color Stickers	2	4	7	5
Unbroken: A World War II Story of Survival, Resilience, and Redemption	3	5	8	1
Under the Dome: A Novel	1	4	8	1
Unfreedom of the Press	3	4	7	10
Unicorn Coloring Book: For Kids Ages 4-8 (US Edition) (Silly Bear Coloring Books)	3	4	7	10
Uninvited: Living Loved When You Feel Less Than, Left Out, and Lonely	3	4	7	7

Watchmen	3	4	9	1
Water for Elephants: A Novel	2	4	7	2
What Happened	2	4	8	8
What If?: Serious Scientific Answers to Absurd Hypothetical Questions	3	4	8	5
What Pet Should I Get? (Classic Seuss)	3	4	7	6
What Should Danny Do? (The Power to Choose Series)	3	4	7	10
What to Expect When You're Expecting	1	4	7	2
Wheat Belly: Lose the Wheat, Lose the Weight, and Find Your Path Back to Health	1	4	7	3
When Breath Becomes Air	3	4	7	7
Where the Crawdads Sing	3	6	8	10
Where the Wild Things Are	3	4	7	1
Whose Boat Is This Boat?: Comments That Don't Help in the Aftermath of a Hurricane	2	4	7	9
Wild: From Lost to Found on the Pacific Crest Trail	1	4	8	3
Winter of the World: Book Two of the Century Trilogy	2	4	8	3
Women Food and God: An Unexpected Path to Almost Everything	1	4	7	1
Wonder	3	4	7	4
Wrecking Ball (Diary of a Wimpy Kid Book 14)	3	4	7	10
You Are a Badass: How to Stop Doubting Your Greatness and Start Living an Awesome Life	3	4	7	7

Обучающая выборка (таблица 6), по сути, представляет собой нормализованные исходные данные, т.е. таблицу исходных данных

(таблица 3), закодированную с помощью классификационных и описательных шкал и градаций (таблицы 4 и 5). Желтым фоном выделены классификационные шкалы. Таким образом, созданы все необходимые и достаточные условия для выполнения следующего этапа АСК-анализа: т.е. для синтеза и верификации моделей.

Задача 3: синтез и верификация статистических и системно-когнитивных моделей и выбор наиболее достоверной из них для решения задач

Синтез и верификация статистических и системно-когнитивных моделей (СК-моделей) осуществляется в режиме 3.5 системы «Эйдос» (рисунок 4).



3.3. Модель "1. ABS - частный критерий: количество встреч сценариев: "Класс:приним" у объекта обучающей"

Год	Наименование сценария	1	2	3	4	5	6	7	8	9	Средн	Средн
год	наименование сценария	ЧЕТАТЕЛИ	ЧЕТАТЕЛИ	ЧЕТАТЕЛИ	ОТЪЕЗД	ОТЪЕЗД	ОТЪЕЗД	ЦЕНА	ЦЕНА	ЦЕНА		
1/9	1/9	4.333333	4.333333	4.333333	1/9	1/9	1/9	1/9	1/9	1/9		
1	ГОД ВЫПСКА-1716-2009-000000-2010-000000	2	1	0	0	1		4	4	1	27	3.0
2	ГОД ВЫПСКА-2716-2010-000000-2011-000000	2	2		6			4	5		18	2.0
3	ГОД ВЫПСКА-3716-2011-000000-2012-000000	2	2		4			1	5		12	1.0
4	ГОД ВЫПСКА-4716-2012-000000-2013-000000			2	2			2	1		8	1.0
5	ГОД ВЫПСКА-5716-2013-000000-2014-000000		2	2	4			1	2	1	12	1.0
6	ГОД ВЫПСКА-6716-2014-000000-2015-000000		1	2	2			1	2		8	1.0
7	ГОД ВЫПСКА-7716-2015-000000-2016-000000			5	5			5			15	1.4
8	ГОД ВЫПСКА-8716-2016-000000-2017-000000		1	1	4			2	2		12	1.3
9	ГОД ВЫПСКА-9716-2017-000000-2018-000000		1	1	2			1	1		8	0.4
10	ГОД ВЫПСКА-10*0-2019-000000-2019-000000		2	6			1	6	1		25	2.0
11	ABTOP Alan Levy		1	1				1			3	0.3
12	ABTOP Alan Moore			1	1					1	3	0.3
13	ABTOP Chako Kuznetsov			1	1				1		3	0.3
14	ABTOP Ched Staged	2			1				1		3	0.3
15	ABTOP David Kahneman		1		1				1		3	0.3
16	ABTOP David McCullough			1	1				1		3	0.3
17	ABTOP Dalia Dweck			1	1		1		1		3	0.3
18	ABTOP DK		1		1			1			3	0.3
19	ABTOP Dr. Seuss			1	1				1		3	0.3
20	ABTOP Edward M. Kennedy		1		1				1		3	0.3
21	ABTOP Emily Winfield Martin			1	1			1			3	0.3
22	ABTOP Eric Carle			1	1			1			3	0.3
23	ABTOP Geena Path	2			1			1			3	0.3

Рисунок 5. Матрица абсолютных частот (фрагмент)

3.3. Модель "4. INF1 - частный критерий: количество людей по А.Карлосу: вероятность из INF1"

Год	Наименование сценария	1	2	3	4	5	6	7	8	9	Средн	Средн
год	наименование сценария	ЧЕТАТЕЛИ	ЧЕТАТЕЛИ	ЧЕТАТЕЛИ	ОТЪЕЗД	ОТЪЕЗД	ОТЪЕЗД	ЦЕНА	ЦЕНА	ЦЕНА		
1/9	1/9	4.333333	4.333333	4.333333	1/9	1/9	1/9	1/9	1/9	1/9		
1	ГОД ВЫПСКА-1716-2009-000000-2010-000000	0.370	-0.512	0.002	-0.042	0.920		-0.144	0.004	0.520	1.233	0.968
2	ГОД ВЫПСКА-2716-2010-000000-2011-000000	0.405	-0.233		-0.024			0.004	-0.079		0.279	0.870
3	ГОД ВЫПСКА-3716-2011-000000-2012-000000	0.405	0.233		-0.024			0.467	-0.079		0.263	0.870
4	ГОД ВЫПСКА-4716-2012-000000-2013-000000			0.221	0.024			0.004	-0.079		0.263	0.870
5	ГОД ВЫПСКА-5716-2013-000000-2014-000000		-0.233	-0.057	0.024			0.467	0.004	0.991	0.970	0.649
6	ГОД ВЫПСКА-6716-2014-000000-2015-000000		-0.105	0.105	0.024			-0.306	0.211		0.240	0.649
7	ГОД ВЫПСКА-7716-2015-000000-2016-000000			0.221	0.024			0.211			0.211	0.211
8	ГОД ВЫПСКА-8716-2016-000000-2017-000000		-0.160	-0.444	-0.024			-0.078	0.004		0.270	0.211
9	ГОД ВЫПСКА-9716-2017-000000-2018-000000		-0.221	-0.057	0.024			-0.078	0.004		0.270	0.211
10	ГОД ВЫПСКА-10*0-2019-000000-2019-000000		0.221	-0.062		1.020		0.225	-0.554		1.013	0.649
11	ABTOP Alan Levy			0.221	0.024			0.211			1.274	2.121
12	ABTOP Alan Moore			0.221	0.024							0.096
13	ABTOP Chako Kuznetsov			0.221	0.024					0.520	1.554	1.294
14	ABTOP Ched Staged	0.405			0.024				0.520		1.294	0.896
15	ABTOP David Kahneman		-0.221		0.024				0.520		1.294	0.896
16	ABTOP David McCullough			0.221	0.024				0.520		1.294	0.896
17	ABTOP Dalia Dweck			0.221	0.024		2.140		0.520		3.025	1.959
18	ABTOP DK		-0.221		0.024			0.211			1.099	0.649
19	ABTOP Dr. Seuss			0.221	0.024			0.211			1.099	0.649
20	ABTOP Edward M. Kennedy		-0.221		0.024				0.520		1.294	0.896
21	ABTOP Emily Winfield Martin			0.221	0.024			0.211			1.099	0.649
22	ABTOP Eric Carle			0.221	0.024			0.211			1.099	0.649
23	ABTOP Geena Path	0.370			0.024			0.211			1.099	0.649

Рисунок 6. Матрица информативностей INF1 (фрагмент)

3.3. Модель "6. INF3 - частный критерий: та же табл. разница между фактическим и ожидаемым абс. частотам"

Год	Наименование сценария	1	2	3	4	5	6	7	8	9	Средн	Средн
год	наименование сценария	ЧЕТАТЕЛИ	ЧЕТАТЕЛИ	ЧЕТАТЕЛИ	ОТЪЕЗД	ОТЪЕЗД	ОТЪЕЗД	ЦЕНА	ЦЕНА	ЦЕНА		
1/9	1/9	4.333333	4.333333	4.333333	1/9	1/9	1/9	1/9	1/9	1/9		
1	ГОД ВЫПСКА-1716-2009-000000-2010-000000	0.949	-1.489	0.021	-0.617	0.639		-0.191	-0.170	0.553	0.417	-0.255
2	ГОД ВЫПСКА-2716-2010-000000-2011-000000	0.979	1.249	-0.319	0.255	-0.128		-0.128	0.553	-0.299	-0.255	-0.178
3	ГОД ВЫПСКА-3716-2011-000000-2012-000000	0.229	0.894	-2.213	0.170	-0.085		-0.085	-1.298	1.468	-0.128	-0.178
4	ГОД ВЫПСКА-4716-2012-000000-2013-000000	-0.511	-0.838	1.248	0.124	-0.064		-0.064	0.277	-0.249	-0.128	-0.178
5	ГОД ВЫПСКА-5716-2013-000000-2014-000000	-0.481	0.894	-0.213	0.178	-0.085		-0.085	-1.298	0.468	0.225	-0.178
6	ГОД ВЫПСКА-6716-2014-000000-2015-000000	-0.511	0.170	0.149	0.124	-0.064		-0.064	-0.723	0.851	-0.128	-0.178
7	ГОД ВЫПСКА-7716-2015-000000-2016-000000	-0.051	-1.202	2.224	0.213	-0.104		-0.104	0.129	-1.015	-0.213	-0.178
8	ГОД ВЫПСКА-8716-2016-000000-2017-000000	-0.681	-1.219	-0.174	-0.174	-0.085		-0.085	-0.298	0.468	-0.178	-0.178
9	ГОД ВЫПСКА-9716-2017-000000-2018-000000	-0.240	-0.447	-0.104	0.081	-0.043		-0.043	-0.149	0.234	-0.001	-0.178
10	ГОД ВЫПСКА-10*0-2019-000000-2019-000000	-0.191	-1.926	2.129	-0.292	-0.149		-0.149	0.851	-1.481	-0.298	-0.178
11	ABTOP Alan Levy	-0.170	-0.277	0.447	0.043	-0.021		-0.021	0.426	-0.203	-0.043	-0.043
12	ABTOP Alan Moore	-0.170	-0.277	0.447	0.043	-0.021		-0.021	-0.574	0.317	-0.043	-0.043
13	ABTOP Chako Kuznetsov	-0.170	-0.277	0.447	0.043	-0.021		-0.021	-0.574	0.317	-0.043	-0.043
14	ABTOP Ched Staged	0.830	-0.277	-0.553	0.043	-0.021		-0.021	-0.574	0.317	-0.043	-0.043
15	ABTOP David Kahneman	-0.170	-0.277	-0.553	0.043	-0.021		-0.021	-0.574	0.317	-0.043	-0.043
16	ABTOP David McCullough	-0.170	-0.277	0.447	0.043	-0.021		-0.021	-0.574	0.317	-0.043	-0.043
17	ABTOP Dalia Dweck	-0.170	-0.277	0.447	-0.957	-0.021		0.939	-0.574	0.317	-0.043	-0.043
18	ABTOP DK	-0.170	-0.221	-0.053	0.043	-0.021		-0.021	0.426	-0.203	-0.043	-0.043
19	ABTOP Dr. Seuss	-0.170	-0.277	0.447	0.043	-0.021		-0.021	0.426	-0.203	-0.043	-0.043
20	ABTOP Edward M. Kennedy	-0.170	-0.221	-0.553	0.043	-0.021		-0.021	-0.574	0.317	-0.043	-0.043
21	ABTOP Emily Winfield Martin	-0.170	-0.277	0.447	0.043	-0.021		-0.021	-0.574	0.317	-0.043	-0.043
22	ABTOP Eric Carle	-0.170	-0.277	0.447	0.043	-0.021		-0.021	0.426	-0.203	-0.043	-0.043
23	ABTOP Geena Path	0.830	-0.277	-0.553	0.043	-0.021		-0.021	-0.574	0.317	-0.043	-0.043

Рисунок 7. Модель INF3 (фрагмент)

Отметим, что в АСК-анализе и СК-моделях степень выраженности различных свойств объектов наблюдения рассматривается с единственной точки зрения: с точки зрения того, какое **количество информации** содержится в них о том, к каким обобщающим категориям (классам) будут принадлежать или не принадлежать эти объекты. Поэтому не играет никакой роли в каких единицах измерения измеряются те или иные свойства объектов наблюдения, а также в каких единицах измеряются результаты влияния этих свойств, натуральных, в процентах или стоимостных. Это и есть решение проблемы сопоставимости в АСК-анализе и системе «Эйдос», отличающее их от других интеллектуальных технологий.

Верификация статистических и системно-когнитивных моделей

Оценка достоверности моделей в системе «Эйдос» осуществляется путем решения задачи классификации объектов обучающей выборки по обобщенным образам классов и подсчета количества истинных положительных и отрицательных, а также ложных положительных и отрицательных решений по F-мере Ван Ризбергера, а также по критериям L1- L2-мерам проф. Е.В.Луценко, которые предложены для того, чтобы смягчить или полностью преодолеть некоторые недостатки F-меры. В режиме 3.4 системы «Эйдос» изучается достоверность каждой частной модели в соответствии с этими мерами достоверности (рисунок 8).

Из рисунка 8 мы видим, что в данном интеллектуальном приложении по F-критерию Ван Ризбергера наиболее достоверной является СК-модель INF3 с интегральным критерием «Сумма знаний» ($F=0,885$ при максимуме 1,000), что неплохо, а по критерию L1 проф.Е.В.Луценко [3] наиболее достоверной также является СК-модель INF3, но с интегральным критерием «Сумма знаний» ($L1=0,935$ при максимуме 1,000), что является очень хорошим результатом.

Имя модели	Интегральный критерий	F-мера	Л1-мера	Л2-мера	Л3-мера	Л4-мера	Л5-мера	Л6-мера	Л7-мера	Л8-мера	Л9-мера	Л10-мера	Л11-мера	Л12-мера	Л13-мера	Л14-мера	Л15-мера	Л16-мера	Л17-мера	Л18-мера	Л19-мера	Л20-мера		
1. INF1 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,906	0,972	37,278	1,506	1,492	8,038	0,912	8,999	0,900	8,289	8												
2. INF2 - частный критерий: количество ложных решений	Сумма ложных решений	1,000	0,974	37,308	1,506	1,492	8,038	0,912	8,999	0,900	8,289	8												
3. INF3 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,906	0,972	37,278	1,506	1,492	8,038	0,912	8,999	0,900	8,289	8												
4. INF4 - частный критерий: количество ложных решений	Сумма ложных решений	1,000	0,974	37,308	1,506	1,492	8,038	0,912	8,999	0,900	8,289	8												
5. INF5 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,906	0,972	37,278	1,506	1,492	8,038	0,912	8,999	0,900	8,289	8												
6. INF6 - частный критерий: количество ложных решений	Сумма ложных решений	1,000	0,974	37,308	1,506	1,492	8,038	0,912	8,999	0,900	8,289	8												
7. INF7 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,750	0,885	17,694	1,857	0,819	8,121	0,945	8,684	0,901	8,183	8												
8. INF8 - частный критерий: количество ложных решений	Сумма ложных решений	0,600	0,887	22,047	0,824	1,054	8,284	0,954	8,907	0,971	8,181	8												
9. INF9 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,750	0,844	17,694	1,857	0,619	8,121	0,945	8,684	0,901	8,183	8												
10. INF10 - частный критерий: количество ложных решений	Сумма ложных решений	0,600	0,887	22,047	0,824	1,054	8,284	0,954	8,907	0,971	8,181	8												
11. INF11 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,823	0,885	24,178	2,243	1,190	4,704	0,967	8,888	0,999	8,226	8												
12. INF12 - частный критерий: количество ложных решений	Сумма ложных решений	0,227	0,885	35,634	0,800	1,287	4,714	0,967	8,905	0,906	8,387	8												
13. INF13 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,787	0,871	19,778	2,285	0,919	4,174	0,984	8,683	0,906	8,176	8												
14. INF14 - частный критерий: количество ложных решений	Сумма ложных решений	0,494	0,910	8,294	0,803	0,242	8,015	0,963	8,994	0,900	0,050	8												
15. INF15 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,787	0,872	14,778	2,285	0,919	4,174	0,984	8,683	0,906	8,176	8												
16. INF16 - частный критерий: количество ложных решений	Сумма ложных решений	0,494	0,910	8,294	0,803	0,242	8,015	0,963	8,994	0,900	0,050	8												
17. INF17 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,730	0,835	24,094	1,417	1,255	4,257	0,955	8,746	0,926	8,259	8												
18. INF18 - частный критерий: количество ложных решений	Сумма ложных решений	0,823	0,892	7,545	0,803	0,749	8,292	0,987	8,963	0,934	0,045	8												
19. INF19 - частный критерий: количество истинных решений	Классификация объектов по обр.	0,730	0,835	24,094	1,417	1,255	4,257	0,955	8,746	0,926	8,259	8												
20. INF20 - частный критерий: количество ложных решений	Сумма ложных решений	0,823	0,892	7,545	0,803	0,749	8,292	0,987	8,963	0,934	0,045	8												

Рисунок 8. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергера и L1- и L2-критериям проф.Е.В.Луценко [10]

Это позволяет сделать обоснованные выводы о том, что:

1. Это подтверждает наличие и адекватное отражение в СК-модели сильной причинно-следственной зависимости между автором книги и отзывами на книгу, ее ценой и оценкой читателей.

2. Эта причинно-следственная зависимость адекватно отражена в системно-когнитивных моделях INF1 и INF3.

На рисунке 9 приведено частотное распределения числа истинных и ложных положительных и отрицательных решений по результатам оценки книг на основе авторов, их написавших, в СК-модели INF3 по данным обучающей выборки:

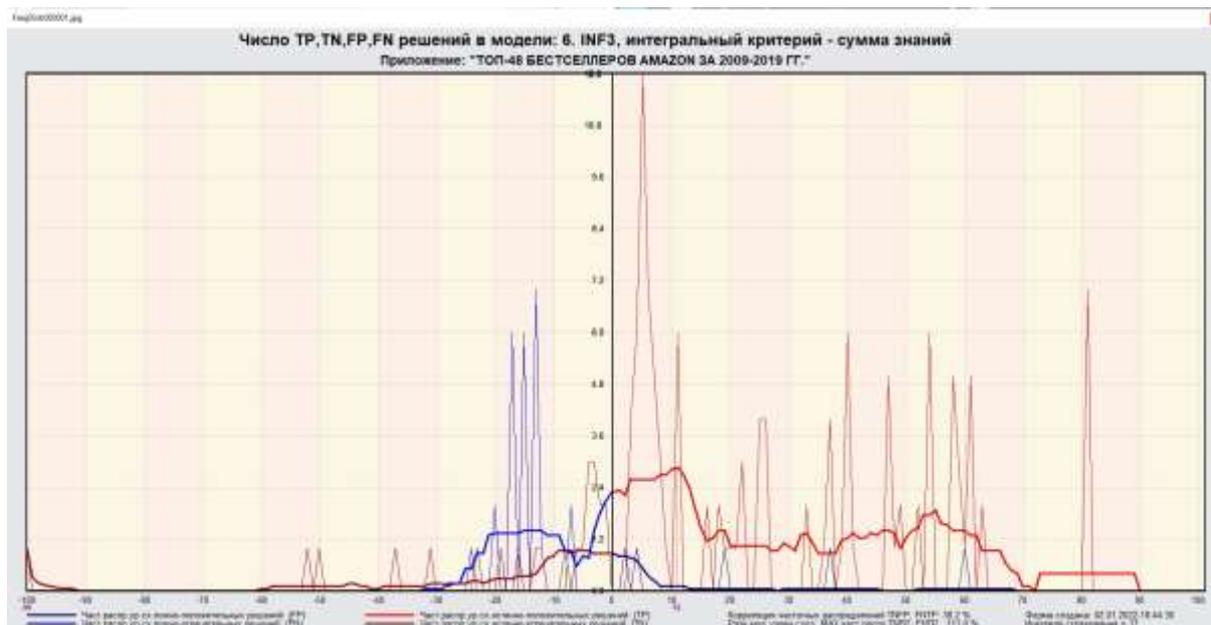


Рисунок 9. Частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности в СК-модели Inf3

Рисунок 9 содержит изображения двух частотных распределений, похожих на нормальные, сдвинутых относительно друг друга по фазе.

Левое распределение, меньшее по амплитуде включает истинно-отрицательные и ложно-положительные решения, а правое, большее по амплитуде, включает ложные-отрицательные и истинно-положительные решения.

Сдвиг этих распределений относительно друг друга и другие различия между ними и позволяют решать задачу идентификации оценки книг на рынке и их отзывы по имени автора и году выпуска книги, а так же многие другие задачи.

Видно, что для отрицательных решений количество истинных решений всегда значительно превосходит количество ложных решений, причем при уровнях различия больше примерно 55% ложные отрицательные решения вообще практически отсутствуют.

Видно также, что для положительных решений картина более сложная и включает 3 диапазона уровней сходства

1) при любых уровнях сходства количество истинных решений всегда больше, чем ложных;

2) от 0% до 35% больше ложных решений, а от 35% и более количество положительных решений превышает ложные

На рисунке 10 приведен Help по режиму 3.4, в котором описаны меры достоверности моделей, применяемые в системе «Эйдос»:

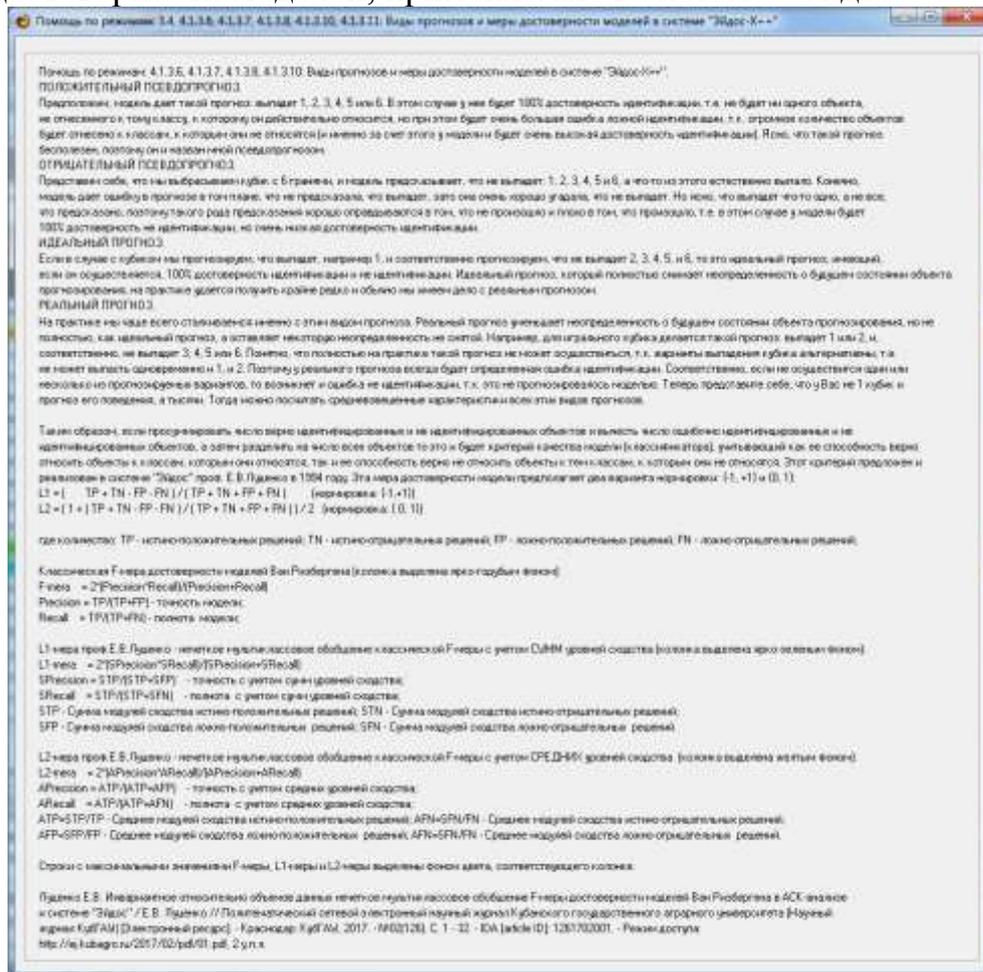


Рисунок 10. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена и L1- и L2-критериям проф.Е.В.Луценко [10]

Выбор наиболее достоверной модели и присвоение ей статуса текущей

В соответствии со схемой обработки данных, информации и знаний в системе «Эйдос» (рисунок 1), присвоим СК-модели INF1 статус текущей модели. Для этого запустим режим 5.6 с параметрами, приведенными на экранной форме (рисунок 11):

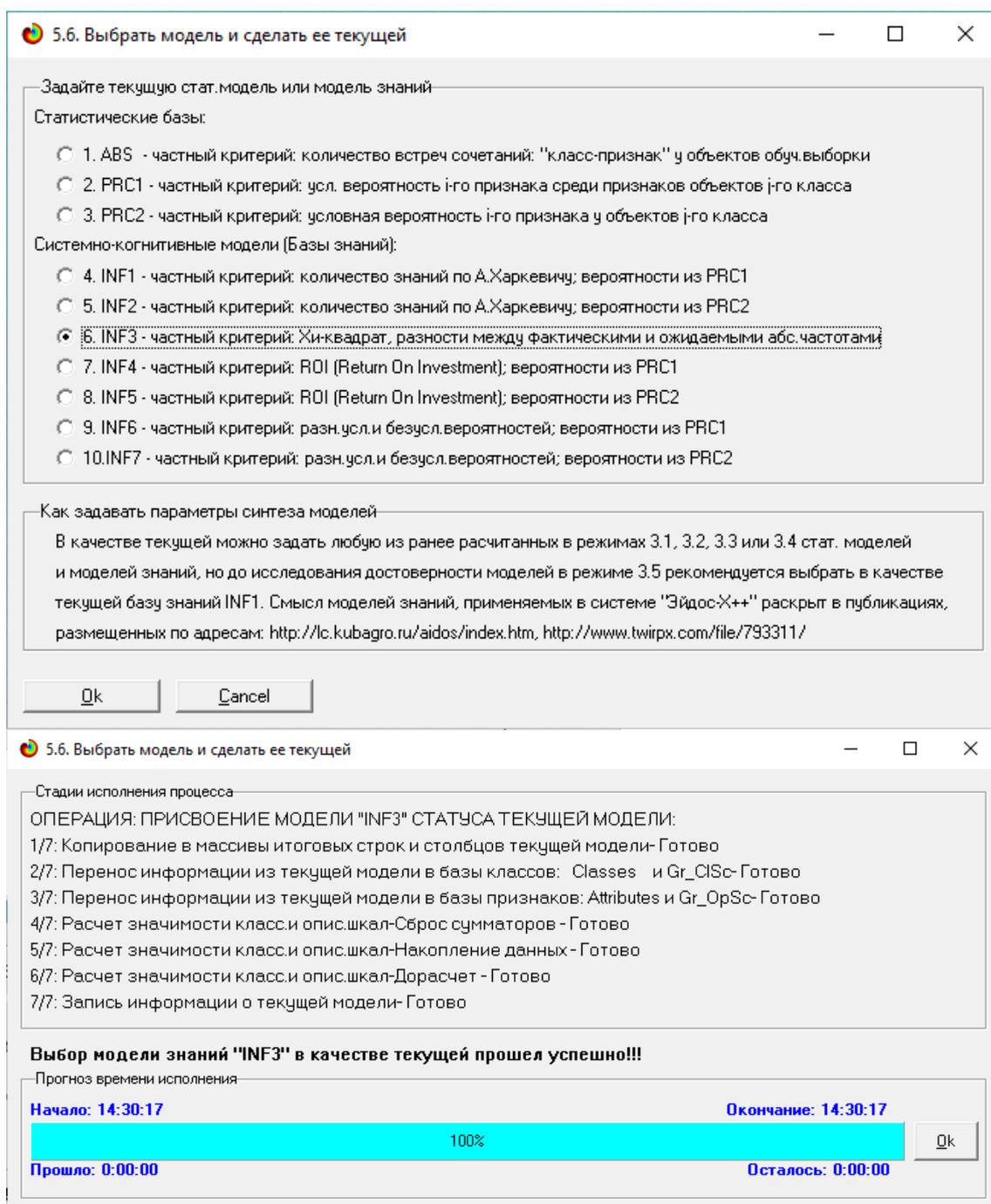
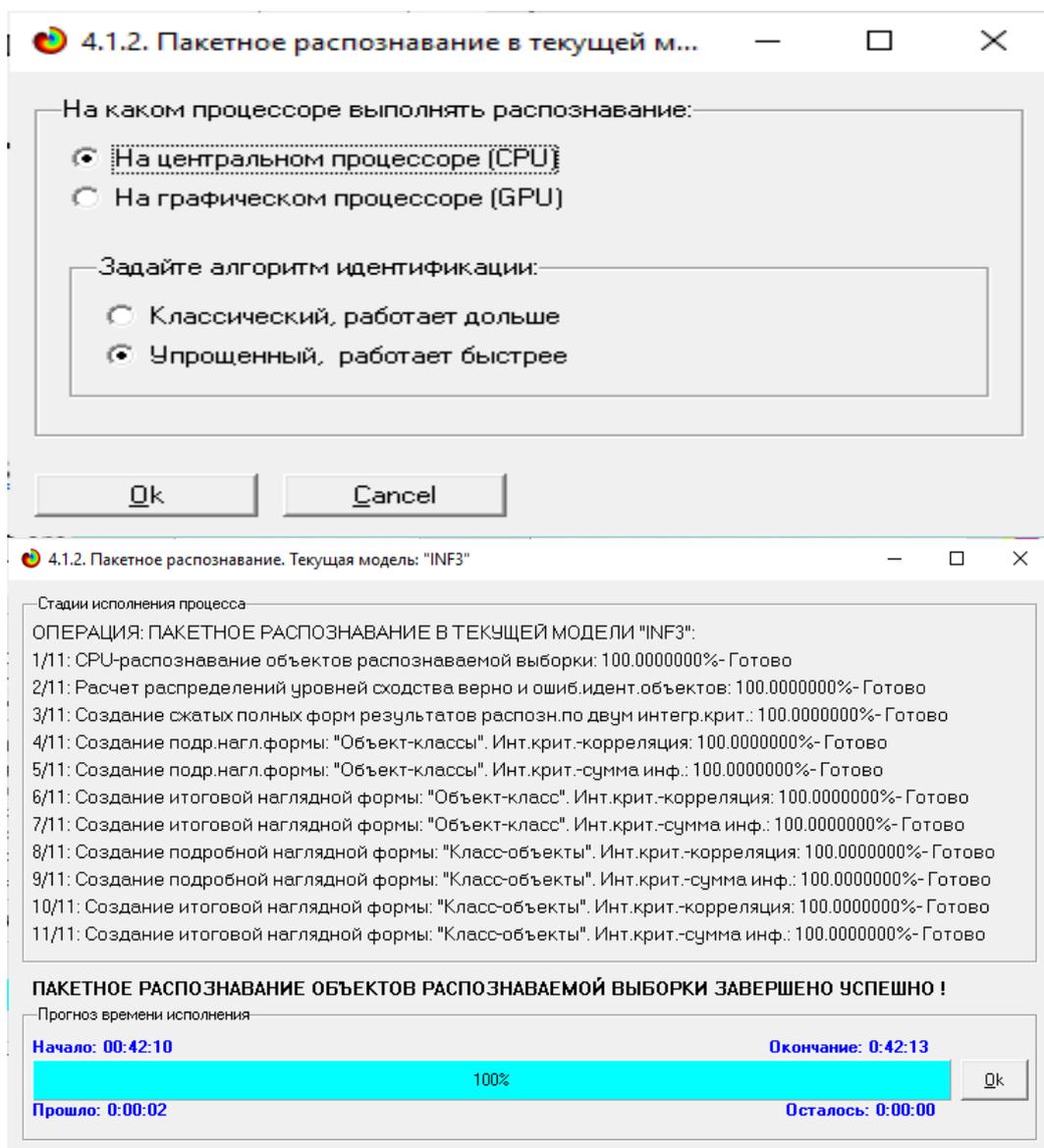


Рисунок 11. Экранные формы придания наиболее достоверной по F-критерию СК-модели Inf3 статуса текущей модели

Задача 4: решение различных задач в наиболее достоверной модели

Подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация)

Решим задачу системной идентификации, т.е. измерения параметров топ-48 бестселлеров amazon за 2009–2019 гг. обучающей выборки в наиболее достоверной СК-модели INF3 на GPU (рисунок 12).



4.1.3. Вывод результатов распознавания	4.1.3.1. Подробно наглядно: "Объект - классы"	
	4.1.3.2. Подробно наглядно: "Класс - объекты"	
	4.1.3.3. Итоги наглядно: "Объект - класс"	
	4.1.3.4. Итоги наглядно: "Класс - объект"	
	4.1.3.5. Подробно сжато: "Объекты - классы"	
	<hr/>	
	4.1.3.6. Обобщ. форма по достов. моделей при разных интегральных крит.	
	4.1.3.7. Обобщ. стат. анализ результатов идент. по моделям и инт. крит.	
	4.1.3.8. Стат. анализ результ. идент. по классам, моделям и инт. крит.	
	4.1.3.9. Достоверность идент. объектов при разных моделях и инт. крит.	
	4.1.3.10. Достоверность идент. классов при разных моделях и инт. крит.	
	4.1.3.11. Распределения уровн. сходства при разных моделях и инт. крит.	
4.1.3.12. Объединение в одной БД строк по самым достоверным моделям		

Рисунок 12. Экранные формы отображения процесса решения задачи системной идентификации в текущей модели

Из рисунка 11 видно, что процесс идентификации занял 0 секунды.

Отметим, что это связано с техникой, на которой производится данная работа. На другом (более старом устройстве) данный процесс мог бы занять большой объем времени и 99,999% этого времени заняла бы не сама идентификация на GPU, а создание 11 выходных форм на основе результатов этого прогнозирования. Эти формы отражают результаты прогнозирования в различных разрезах и обобщениях:

Приведем две из этих 10 форм: 4.1.3.1 и 4.1.3.2 (рисунок 13).

4.1.3.1. Визуализация результатов распознавания в отношении "Объект-классы", Текущая модель: "INF3"

Распознаваемые объекты	
Код	Наименование
1	10 Day Green Smoothie Cleanse
2	11/22/63: A Novel
3	12 Rules for Life: An Antidote t...
4	1984 (Signet Classics)
5	5,000 Awesome Facts (About ...
6	A Dance with Dragons (A Son...
7	A Game of Thrones / A Clash ...
8	The Very Hungry Caterpillar
9	The Whole30: The 30-Day Gu...
10	The Wonderful Things You Wi...
11	The Wright Brothers
12	Things That Matter: Three De...
13	Thinking, Fast and Slow
14	Thirteen Reasons Why
15	Thomas Jefferson: The Art of ...
16	Three Cups of Tea: One Man's...
17	Thug Kitchen: The Official Co...
18	Tina Fey: Bossypants
19	To Kill a Mockingbird
20	Tools of Titans: The Tactics, ...
21	Towers of Midnight (Wheel of ...
22	True Compass: A Memoir
23	Twilight (The Twilight Saga, B...

Интегральный критерий сходства: "Семантический резонанс знаний"				
Код	Наименование класса	Сходство	Ф.	Сходство
7	ЦЕНА(\$)-1/3 (0.0, 14.0)	36.79...	v	██████████
3	РЕЙТИНГ ЧИТАТЕЛЕЙ-3/3 (4.7, 4.9)	29.01...	v	██████████
4	ОТЗЫВОВ-1/3 (438.0, 29572.3)	10.74...	v	██
6	ОТЗЫВОВ-3/3 (58706.7, 87841.0)	-6.972...		██
5	ОТЗЫВОВ-2/3 (29572.3, 58706.7)	-7.142...		██
9	ЦЕНА(\$)-3/3 (28.0, 42.0)	-10.45...		██
1	РЕЙТИНГ ЧИТАТЕЛЕЙ-1/3 (4.2, 4.4)	-17.84...		██
2	РЕЙТИНГ ЧИТАТЕЛЕЙ-2/3 (4.4, 4.7)	-24.35...		██
8	ЦЕНА(\$)-2/3 (14.0, 28.0)	-35.54...		██

Интегральный критерий сходства: "Сумма знаний"				
Код	Наименование класса	Сходство	Ф.	Сходство
3	РЕЙТИНГ ЧИТАТЕЛЕЙ-3/3 (4.7, 4.9)	60.57...	v	██████████
7	ЦЕНА(\$)-1/3 (0.0, 14.0)	57.69...	v	██████████
4	ОТЗЫВОВ-1/3 (438.0, 29572.3)	5.769...	v	██
5	ОТЗЫВОВ-2/3 (29572.3, 58706.7)	-2.684...		██
6	ОТЗЫВОВ-3/3 (58706.7, 87841.0)	-2.884...		██
9	ЦЕНА(\$)-3/3 (28.0, 42.0)	-5.769...		██
1	РЕЙТИНГ ЧИТАТЕЛЕЙ-1/3 (4.2, 4.4)	-23.07...		██
2	РЕЙТИНГ ЧИТАТЕЛЕЙ-2/3 (4.4, 4.7)	-37.49...		██
8	ЦЕНА(\$)-2/3 (14.0, 28.0)	-51.92...		██

Помощь | 9 классов | Классы с MaxMin УрСх | 9 классов с MaxMin УрСх | ВСЕ классы | ВКЛ. фильтр по класс шкале | ВЫКЛ. фильтр по класс шкале | Граф. диаграммы

4.1.3.2. Визуализация результатов распознавания в отношении: "Класс-объекты". Текущая модель: "BIF3"

Классы		Интегральный критерий сходства: "Семантический резонанс значений"				
Код	Наим. класса	Код	Наименование объекта	Сходство	Ф	Сходство
1	РЕЙТИНГ ЧИТАТЕЛЕЙ-1/3-(4.2000000, 4.4...	18	Tina Fey: Bossypants	49,07	v	██████████
2	РЕЙТИНГ ЧИТАТЕЛЕЙ-2/3-(4.4333333, 4.6...	36	What to Expect When You're Expecting	49,07	v	██████████
3	РЕЙТИНГ ЧИТАТЕЛЕЙ-3/3-(4.6666667, 4.9...	2	11/22/63: A Novel	46,09	v	██████████
4	ОТЗЫВОВ-1/3-(4.38.0000000, 2.9572.3333333)	6	A Dance with Dragons (A Song of Ice and...	46,09	v	██████████
5	ОТЗЫВОВ-2/3-(2.9572.3333333, 5.8706.6666)	16	Three Cups of Tea: One Man's Mission to...	40,15	v	██████████
6	ОТЗЫВОВ-3/3-(5.8706.6666667, 8.7841.0000...	44	Women Food and God: An Unexpected P...	40,15	v	██████████
7	ЦЕНА(\$)-1/3-(0.0000000, 14.0000000)	37	Wheat Belly: Lose the Wheat, Lose the W...	37,54	v	██████████
8	ЦЕНА(\$)-2/3-(14.0000000, 28.0000000)	42	Wild: From Lost to Found on the Pacific C...	37,54	v	██████████
9	ЦЕНА(\$)-3/3-(28.0000000, 42.0000000)	26	Under the Dome: A Novel	37,17	v	██████████
		3	12 Rules for Life: An Antidote to Chaos	8,922		██████

		Интегральный критерий сходства: "Суть значений"			
Код	Наименование объекта	Сходство	Ф	Сходство	
18	Tina Fey: Bossypants	63,46	v	██████████	
36	What to Expect When You're Expecting	63,46	v	██████████	
2	11/22/63: A Novel	59,61	v	██████████	
6	A Dance with Dragons (A Song of Ice and...	59,61	v	██████████	
16	Three Cups of Tea: One Man's Mission to...	51,92	v	██████████	
44	Women Food and God: An Unexpected P...	51,92	v	██████████	
37	Wheat Belly: Lose the Wheat, Lose the W...	48,55	v	██████████	
42	Wild: From Lost to Found on the Pacific C...	48,55	v	██████████	
26	Under the Dome: A Novel	48,07	v	██████████	
3	12 Rules for Life: An Antidote to Chaos	-11,53		██████	

Помощь Поиск объекта В начало БД В конец БД Показать/Скрыть Следующая 9 записей Все записки Печать XLS Печать TXT Печать ALL

Рисунок 13. Выходные формы по результатам прогнозирования топ-48 бестселлеров amazon за 2009–2019 гг.

Символ «v» стоит против тех результатов идентификации, которые подтвердились на опыте, т.е. соответствуют факту. Из рисунка 13 видно, что результаты идентификации являются хорошими, естественно при учете информации из рисунка 9 о том, что достоверные прогнозы в данной модели имеют уровень сходства выше 57%, т.е. по сути результаты с более низки уровнем сходства надо просто игнорировать.

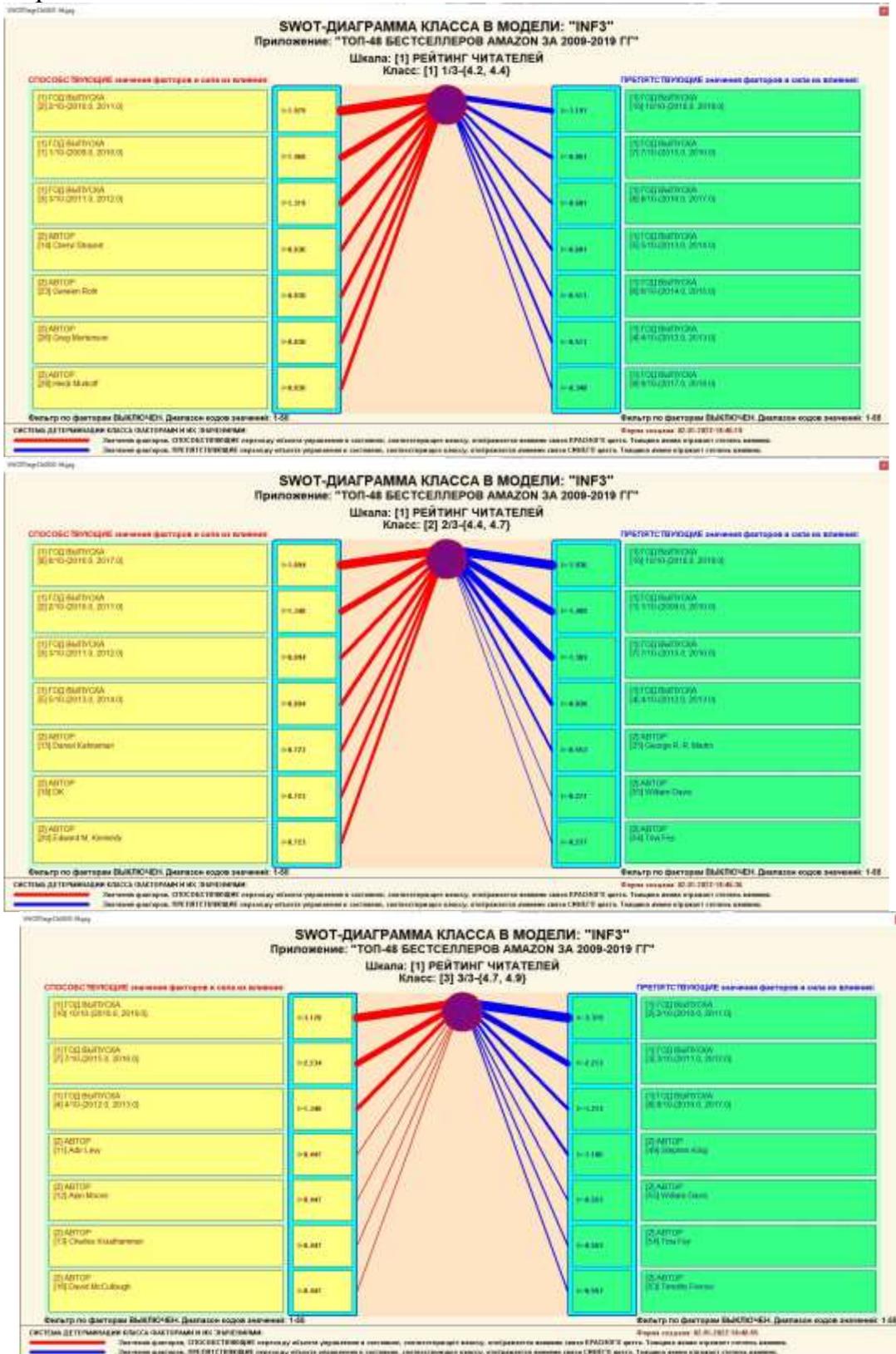
Подзадача 4.2. Поддержка принятия решений (SWOT-анализ)

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 [11].

Применительно к задаче, решаемой в данной работе, SWOT-анализ показывает степень влияния различных авторов на их произведение и их популярность.

В системе «Эйдос» в режиме 4.4.8 поддерживается решение этой задачи. При этом **выявляется система детерминации заданного класса**, т.е. система значений факторов, обуславливающих переход объекта моделирования и управления в состояние, соответствующее данному классу, а также препятствующих этому переходу.

На рисунках 14 приведены SWOT-диаграммы наглядно отражающие силу и направление влияния различных факторов влияющих на популярность и оценки книги.



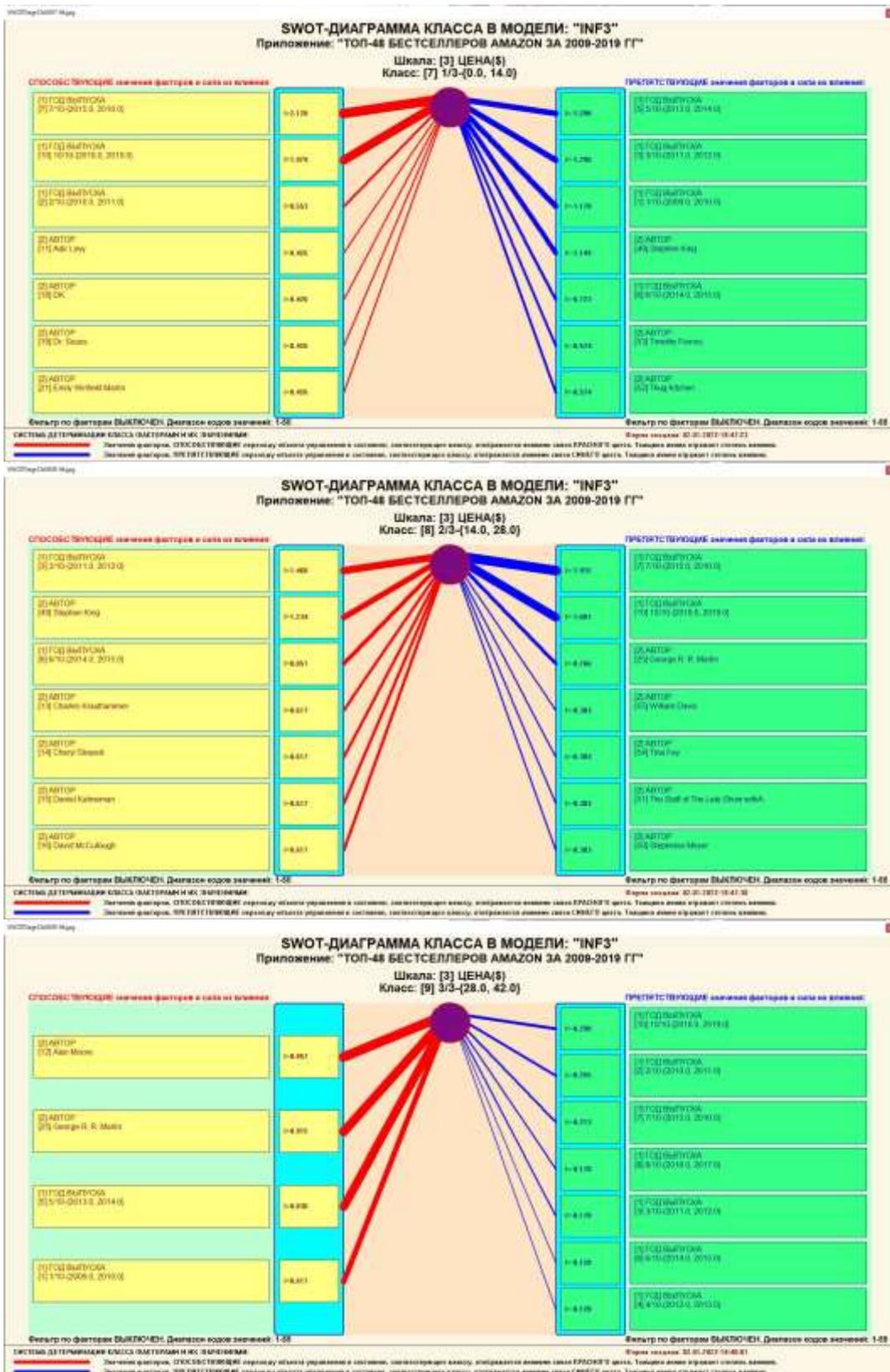


Рисунок 14. SWOT-диаграммы влияния года выпуска книги и ее автора

Эти SWOT-диаграммы наглядно отражают силу и направление влияния различных критериев на людей по всему миру.

Отметим также, что система «Эйдос» обеспечивала решение *всегда*, т.е. даже в самых ранних DOS-версиях и в реализациях системы «Эйдос» на других языках и типах компьютеров. Например, первый акт внедрения системы «Эйдос», где об этом упоминается в явном виде, датируется 1987 годом, а первый подобный расчет относится к 1981 году.

УТВЕРЖДАЮ
Заведующий Краснодарским
сектором ИСИ АН СССР, к.ф.н.
А.А. Хагуров
1987г.

УТВЕРЖДАЮ
Директор Северо-Кавказского филиала
ВНИЦ "АИУС-агроресурсы", к.э.н.
Э.М. Трахов
1987г.

А К Т

Настоящий акт составлен комиссией в составе: Кириченко М.М., Ляшко Г.А., Самсонов Г.А., Коренец В.И., Луценко Е.В. в том, что в соответствии с договором о научно-техническом сотрудничестве между Северо-Кавказским филиалом ВНИЦ "АИУС-агроресурсы" и Краснодарским сектором Института социологических исследований АН СССР Северо-Кавказским филиалом ВНИЦ "АИУС-агроресурсы" выполнены следующие работы:

- осуществлена постановка задачи: "Обработка на ЭВМ социологических анкет Крайагропрома";
- разработаны математическая модель и программное обеспечение подсистемы распознавания образов, позволяющие решать данную задачу в среде персональной технологической системы ВЕГА-М;
- на профессиональной персональной ЭВМ "Искра-226" осуществлены расчёты по задаче в объёме:

Входная информация составила 425 анкет по 9-ти предприятиям.
Выходная информация - 4 вида выходных форм объёмом 90 листов формата А3 и 20 листов формата А4 содержит:

- процентное распределение ответов в разрезе по социальным типам корреспондентов;
- распределение информативностей признаков (в битах) для распознавания социальных типов корреспондентов;
- позитивные и негативные информационные портреты 30-ти социальных типов на языке 212 признаков;
- обобщённая характеристика информативности признаков для выбора такого минимального набора признаков, который содержит максимум информации о распознаваемых объектах (оптимизация анкет).

Работы выполнены на высоком научно-методическом уровне и в срок.

От ИСИ АН СССР:
Мл.научный сотрудник
М.М. Кириченко
19.05. 1987г.

Мл.научный сотрудник
Г.А. Ляшко
19.05. 1987г.

От СКФ ВНИЦ "АИУС-агроресурсы":
Зав.отделом аэрокосмических и
тематических изысканий №4, к.э.н.
Г.А. Самсонов
19.05. 1987г.

Главный конструктор проекта
В.И. Коренец
19.05.87. 1987г.

Главный конструктор проекта
Е.В. Луценко
19.05.87. 1987г.

Но тогда SWOT-диаграммы назывались позитивным и негативным информационными портретами классов.

Информация о системе значений факторов, обуславливающих переход объекта моделирования в различные будущие состояния, соответствующие классам, может быть приведена не только в диаграммах,

показанных на рисунках 14, но и во многих других табличных и графических формах, которые в данной работе не приводятся только из-за ограниченности ее объема. В частности в этих формах может быть выведена значительно более полная информация (в т.ч. вообще вся имеющая в модели). Подобная подробная информация содержится в базах данных, расположенных по пути: c:\Aidos-X\AID_DATA\A0000001\System\SWOTCls#####Infl.DBF, где: «#####» – код класса с ведущими нулями.

На рисунке 15 приведены примеры нескольких инвертированных SWOT-диаграмм:

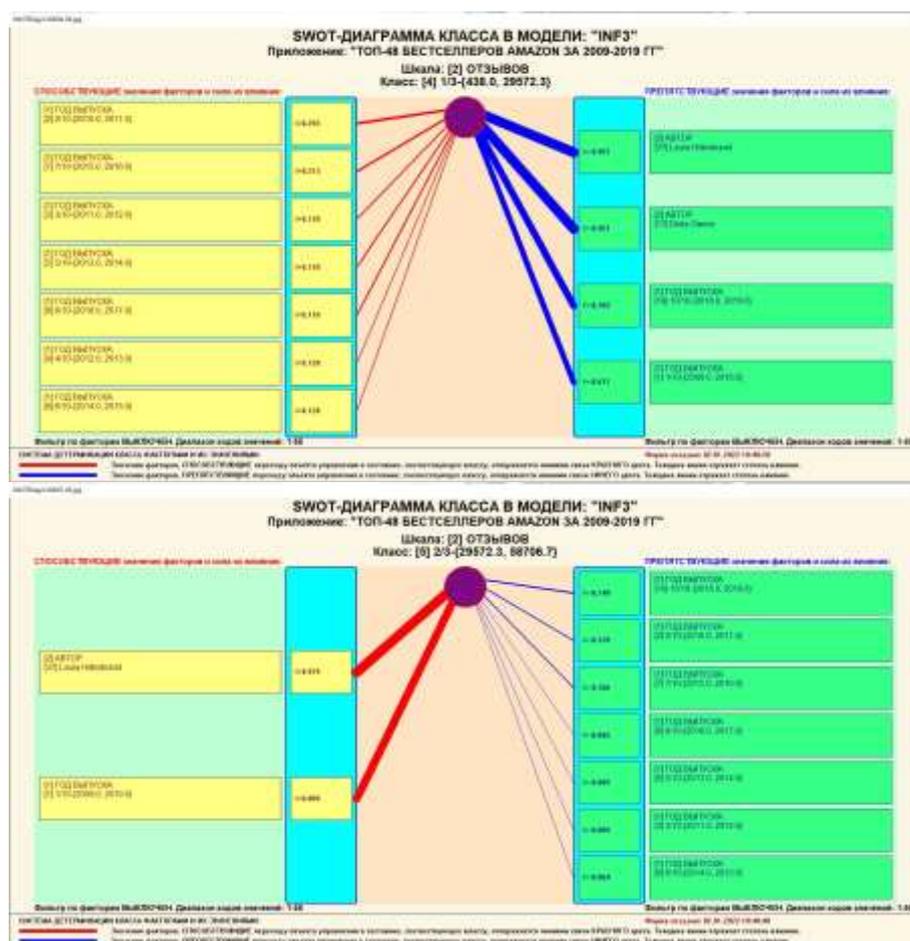


Рисунок 15. Примеры SWOT-диаграмм, отражающих силу и направление влияния конкретного значения параметра на популярность книги

Из рисунка 15 видно, как год выпуска книги влияет на оценки читателей, цену книги и отзывы.

В заключение отметим, что 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 чаще всего является

необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего опыта и профессиональной 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 более 30 лет, но к сожалению она сравнительно малоизвестна – это интеллектуальная система «Эйдос» [11, 18, 19].

Подзадача 4.3. 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 «Эйдос» 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.

4.3.1. Когнитивные диаграммы классов

Эти диаграммы отражают сходство/различие классов. Мы получаем их в режимах 4.2.2.1 и 4.2.2.2 (рисунок 16).

Отметим также, что на когнитивной диаграмме, приведенной на рисунке 16, показаны **количественные** оценки сходства/различия таких важнейших показателей популярности, как рейтинг, отзывы и цена. Важно, что эти результаты сравнения 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.

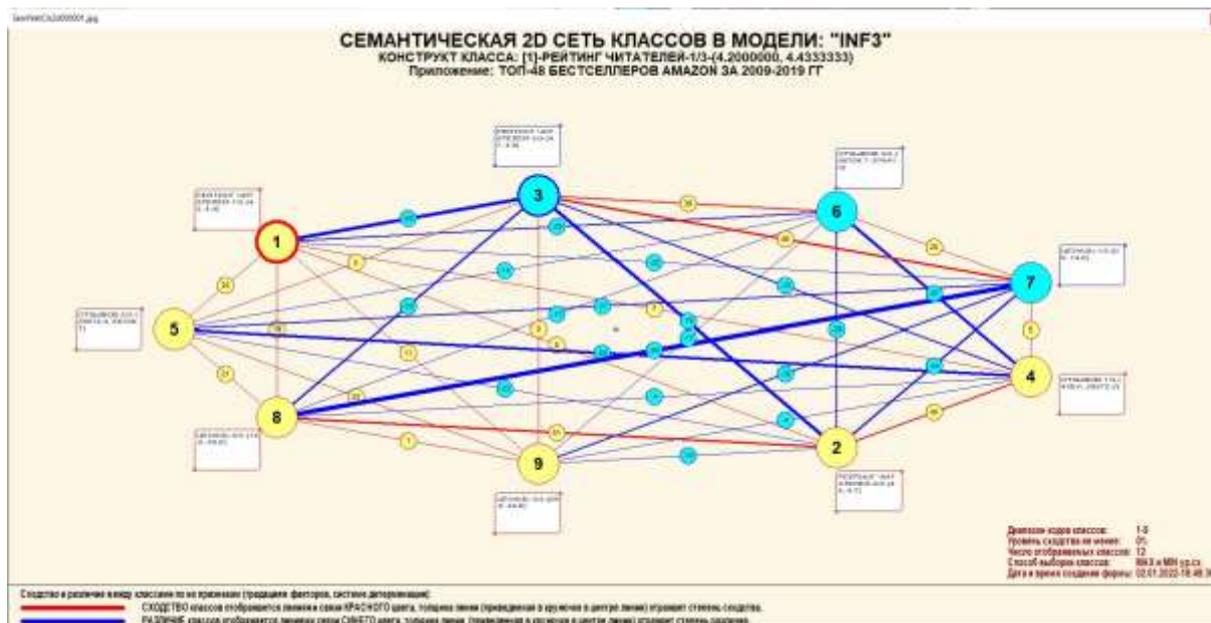


Рисунок 16. Когнитивная диаграмма классов, отражающая сходство/различие популярных книг.

В системе «Эйлос» есть возможность управлять параметрами формирования и вывода изображения, приведенного на рисунке 16. Для этого используется диалоговое окно, приведенное на рисунке 17.

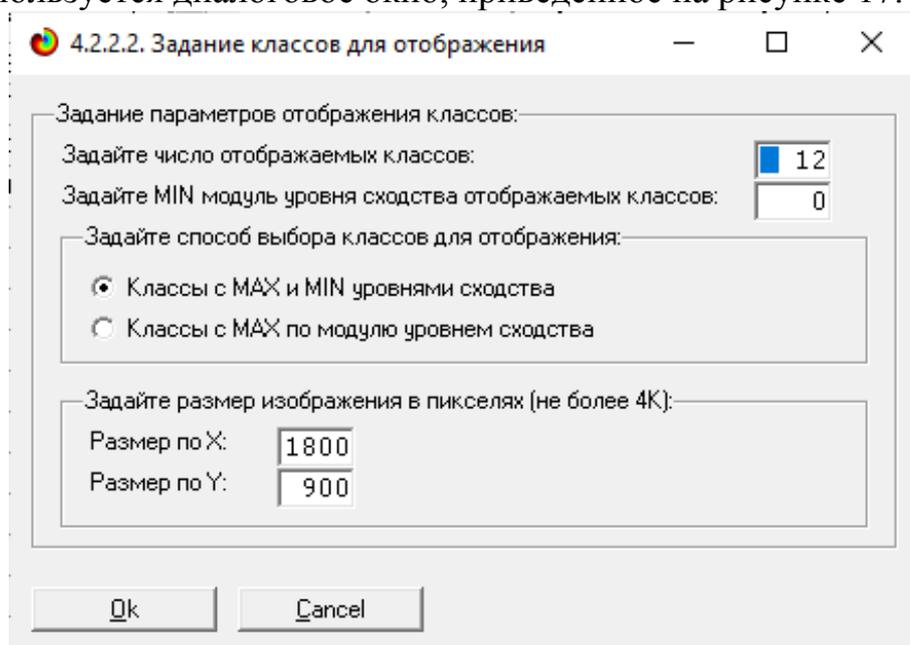


Рисунок 17. Диалоговое окно управления параметрами формирования и вывода изображения когнитивной диаграммы классов

4.3.2. Агломеративная когнитивная кластеризация классов

Информация о сходстве/различии классов, содержащаяся в матрице сходства, может быть визуализирована не только в форме, когнитивных диаграмм, пример которой приведен на рисунке 16, но и в форме

агломеративных дендрограмм, полученных в результате *когнитивной кластеризации* [5] (рисунок 18):

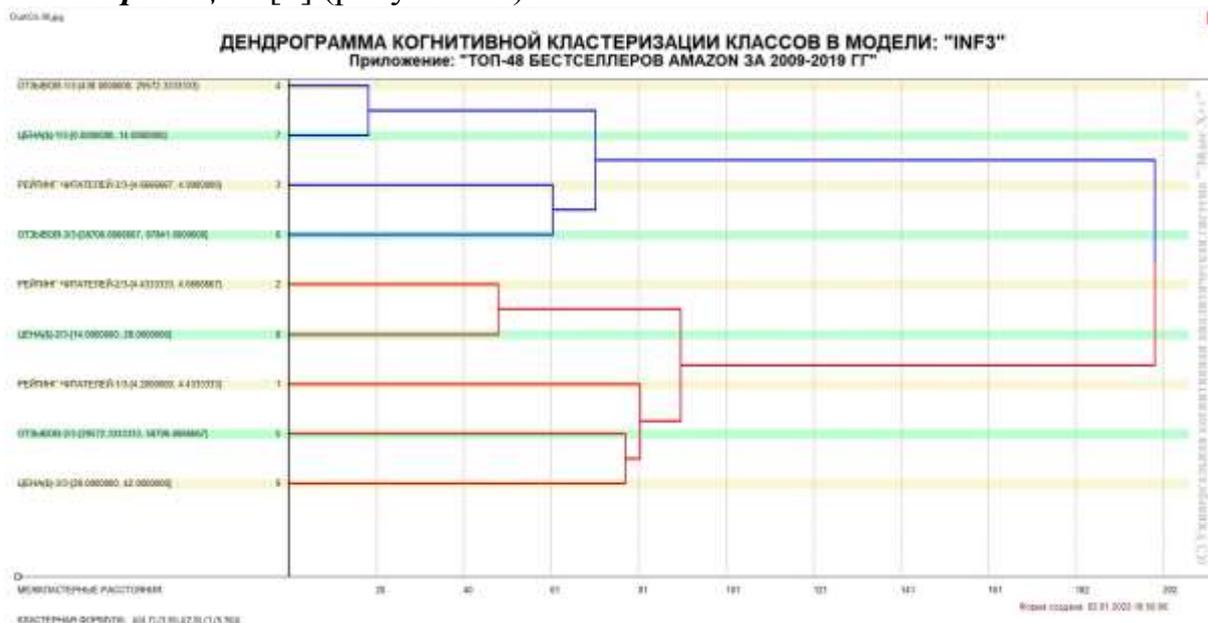


Рисунок 18. Дендрограмма когнитивной агломеративной кластеризации, отражающая сходство/различие популярности разных книг.

Из рисунков 16 и 18 мы видим, что популярность в некоторой степени схожа, и, следовательно, корректно ставить задачу их одновременного достижения, а другие по системе значений этих параметров сильно отличаются, и, следовательно, являются взаимоисключающими, т.е. альтернативными и цель их одновременного достижения является некорректной и недостижимой, т.к. для достижения одного из альтернативных результатов необходимы одни значения параметров авторов, а для достижения другого – совершенно другие, которые не могут наблюдаться одновременно с первыми.

Из дендрограммы когнитивной агломеративной кластеризации классов, приведенной на рисунке 18, мы видим также, что все классы по популярности книг образуют два противоположных кластера, являющихся полюсами конструкта, по системе значений обуславливающих значениям параметров диаграммы.

В верхнем кластере объединены (за редким исключением) результаты с очень высокими, высокими и средними значениями показателей качества, а в нижнем – с очень низкими, низкими и средними.

Из этого можно сделать вывод о том, что высокие и очень высокие показатели популярности по рейтингу бестселлеров amazon за 2009–2019 гг. наблюдаются в одних и тех же пробах и достижимы одновременно, как и очень низкие, низкие и средние показатели качества.

Отметим, что аналогичная картина наблюдается не только в литературе, но и в многих других отраслях.

На рисунке 19 мы видим график изменения межкластерных расстояний:

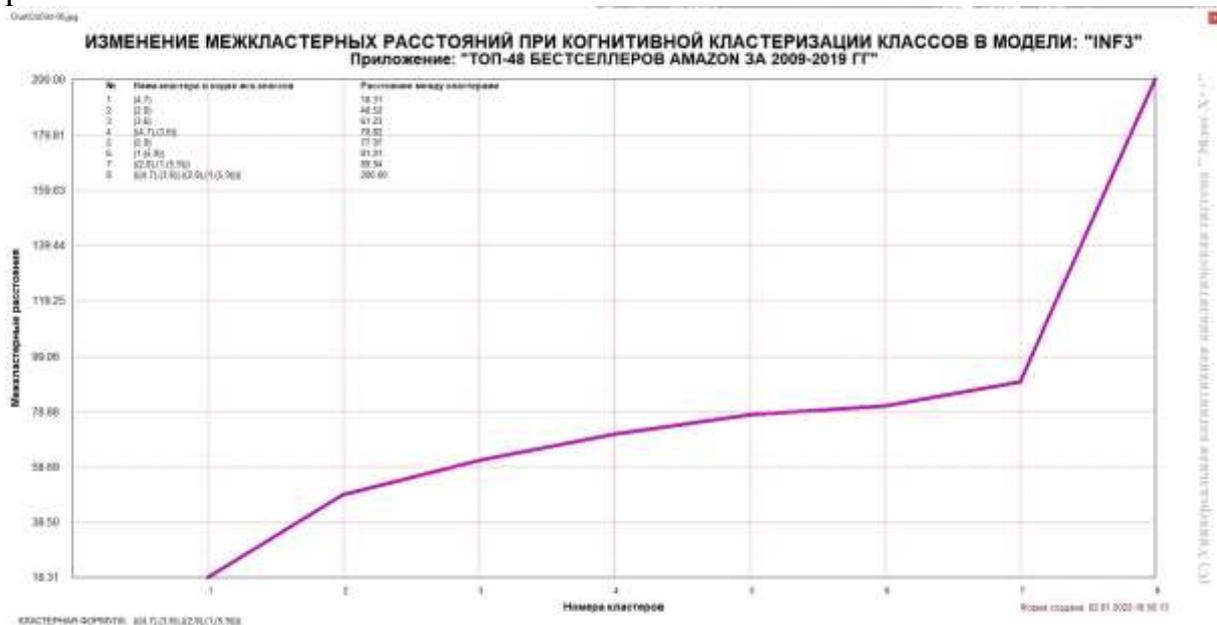


Рисунок 19. График изменения межкластерных расстояний

4.3.3. Когнитивные диаграммы значений факторов

Эти диаграммы отражают сходство/различие значений параметров популярности по их смыслу, т.е. по имени автора и году выхода книги в свет. Эти диаграммы мы получаем в режимах 4.3.2.1 и 4.3.2.2 (рисунок 20).



Рисунок 20. сходство/различие значений параметров популярности книг по их смыслу, т.е. по имени автора и году выхода книги в свет

Из рисунка 20 видно, что все значения факторов образуют два крупных кластера, противоположных по их смыслу. Эти кластеры образуют полюса конструкта.

Отметим, что на когнитивной диаграмме, приведенной на рисунке 20, показаны *количественные* оценки сходства/различия значений факторов, полученные 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.

Диаграмма, приведенная на рисунке 20, получена при параметрах, приведенных на рисунке 21.

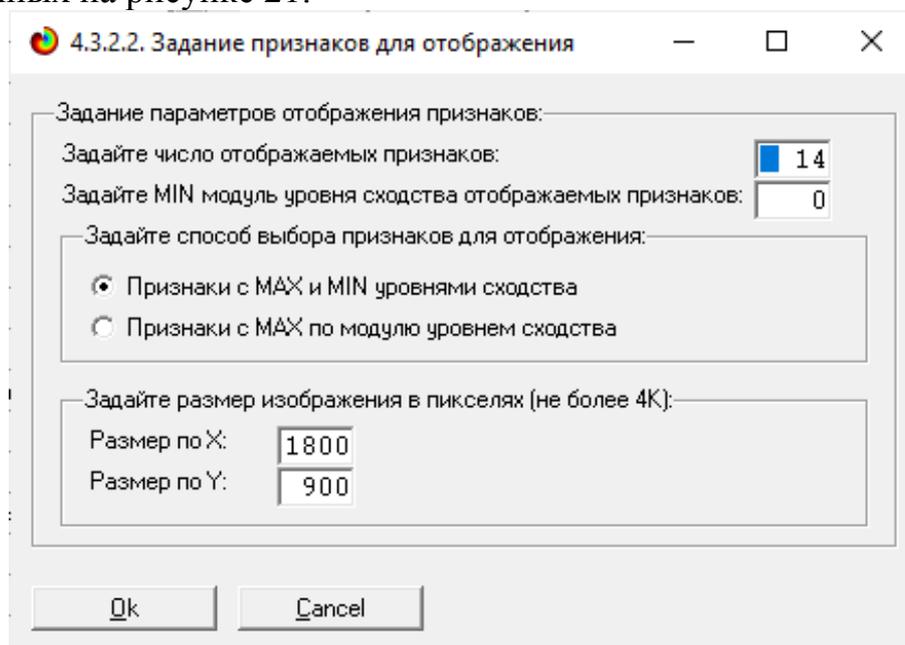


Рисунок 21. Параметры отображения когнитивной диаграммы, приведенной на рисунке 20

4.3.4. Агломеративная когнитивная кластеризация значений факторов

На рисунке 22 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 которой приведен на рисунке 20.

ч

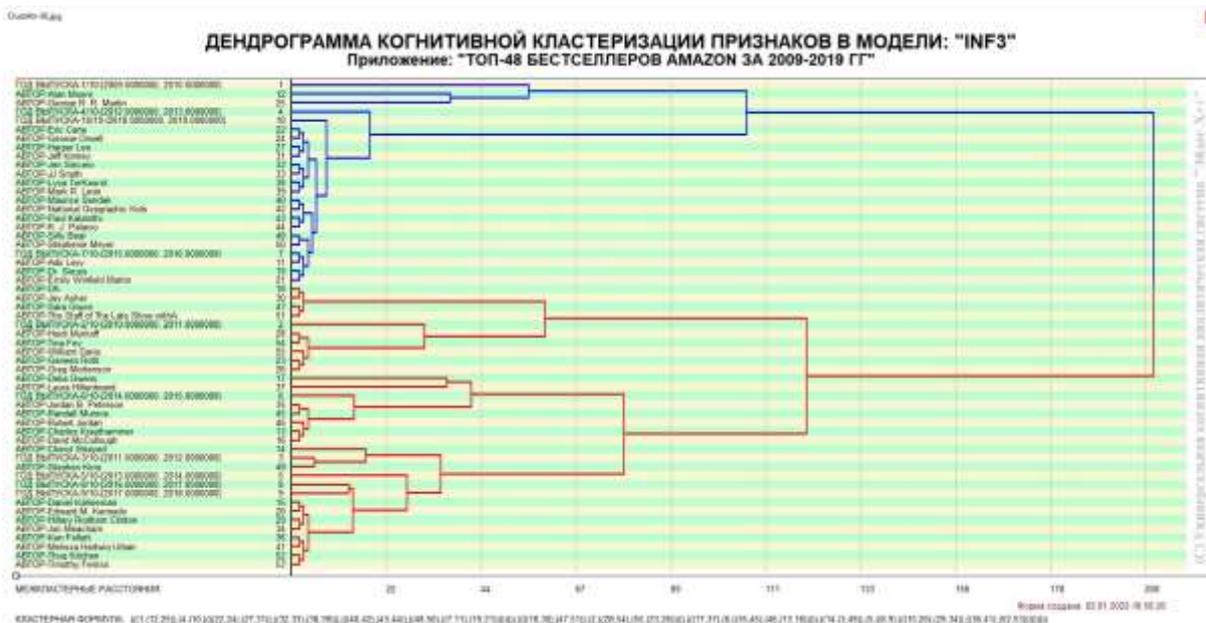


Рисунок 22. Дендрограмма агломеративной когнитивной кластеризации года публикации книг и авторов

Из дендрограммы на рисунке 22 мы видим, что все значения факторов образуют 2 четко выраженных кластера, объединенных в полюса конструкта (показаны синими и красным цветами).

Хорошо видна группировка значений параметров популярности разных книг по их смыслу, т.е. по содержащейся в них информации о годе и авторе. **Значения факторов на полюсах конструкта факторов (рисунок 22) обуславливают переход объекта моделирования в состояния, соответствующие классам, представленным на полюсах конструкта классов (рисунки 18 и 20).**

На рисунке 23 приведен график межкластерных расстояний значений.



Рисунок 23. График изменения межкластерных расстояний при когнитивной кластеризации значений факторов

4.3.5. Нелокальные нейроны и нелокальные нейронные сети

На рисунке 24 приведены пример нелокального нейрона, а на рисунке 25 и фрагмент одного слоя нелокальной нейронной сети:

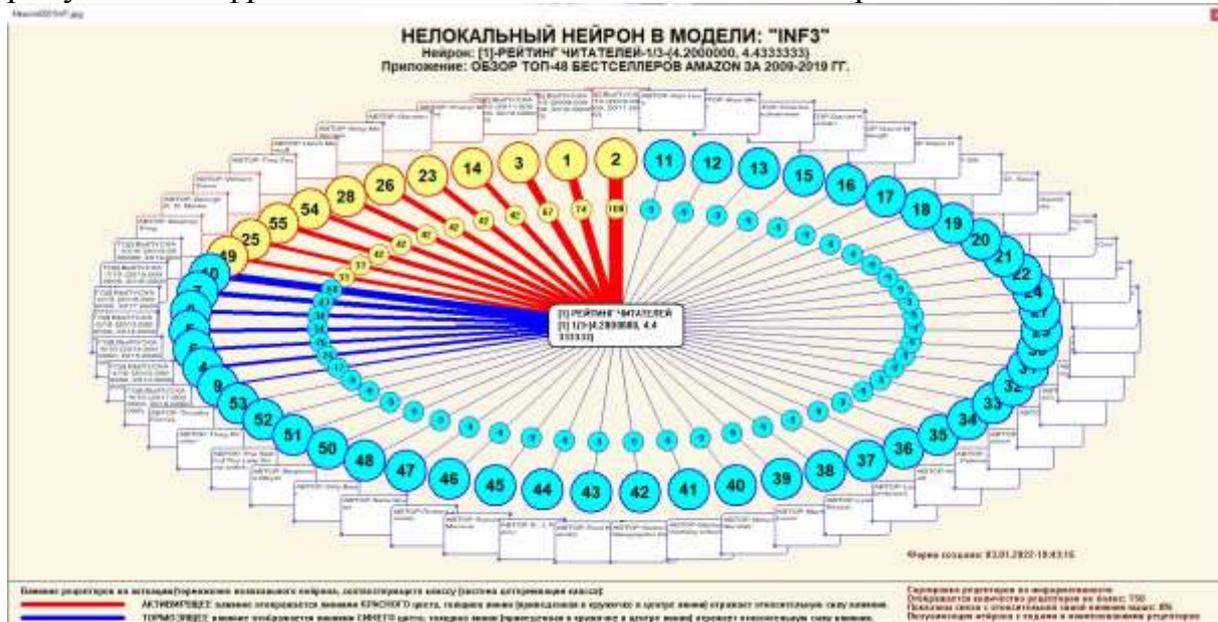


Рисунок 24. Пример нелокального нейрона, отражающего силу и направление влияния показателя рейтинга на показатели авторов и года

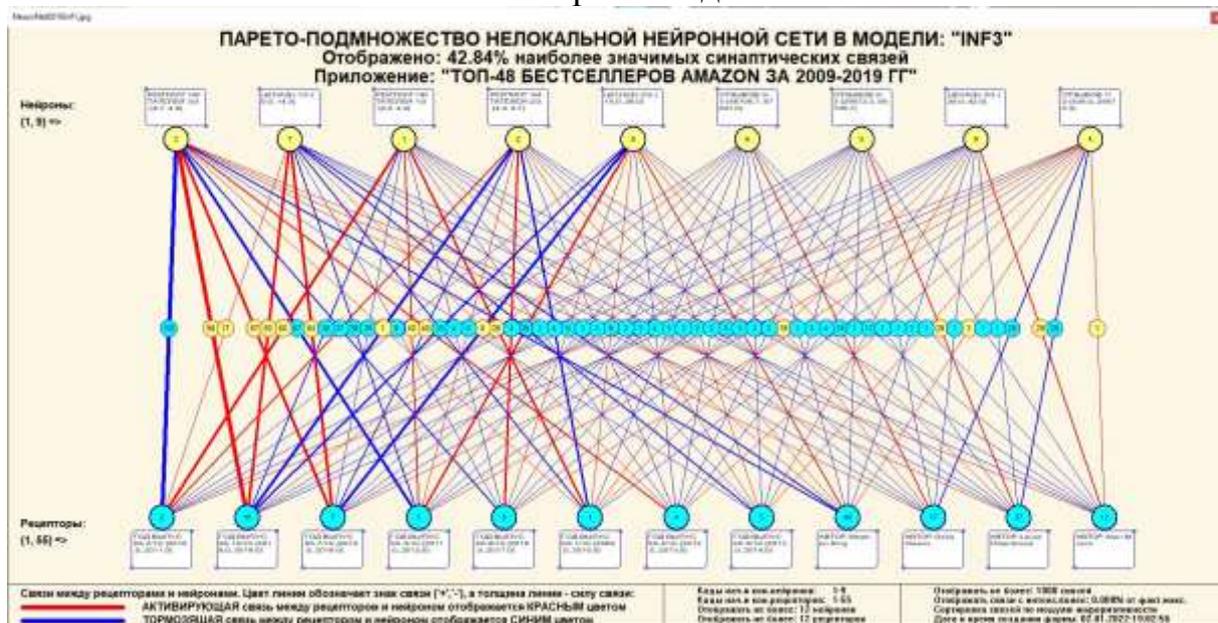


Рисунок 25. Один слой нелокальной нейронной сети, отражающий силу и направление влияния года выхода и авторов книг на рейтинг, отзывы и цены этих книг

В приведенном фрагменте слоя нейронной сети нейроны соответствуют показателям количества популярности, а рецепторы – критериям популярности (отзывы, рейтинги и ...). Нейроны расположены

слева на право в порядке убывания силы детерминации, т.е. слева находятся результаты, наиболее жестко обусловленные обуславливающими их значениями факторами, а с права – менее жестко обусловленные.

Модель знаний системы «Эйдос» относится к *нечетким декларативным* гибридным моделям и объединяет в себе некоторые особенности нейросетевой [13, 20] и фреймовой моделей представления знаний [20]. Классы в этой модели соответствуют нейронам и фреймам, а признаки рецепторам и шпациям (описательные шкалы – слотам).

От фреймовой модели представления знаний модель системы «Эйдос» отличается своей эффективной и простой программной реализацией, полученной за счет того, что разные фреймы отличаются друг от друга не набором слотов и шпаций, а лишь информацией в них. Поэтому в системе «Эйдос» при увеличении числа фреймов само количество баз данных не увеличивается, а увеличивается лишь их размерность.

От нейросетевой модели представления знаний модель системы «Эйдос» отличается тем, что [13, 20]:

1) весовые коэффициенты на рецепторах не подбираются итерационным методом обратного распространения ошибки, а считаются прямым счетом на основе хорошо теоретически обоснованной модели, основанной на теории информации (это напоминает байесовские сети);

2) весовые коэффициенты имеют хорошо теоретически обоснованную содержательную интерпретацию, основанную на теории информации;

3) нейросеть является нелокальной, как сейчас говорят «полносвязной».

4.3.6. 3d-интегральные когнитивные карты

На рисунке 26 приведен фрагмент 3d-интегральной когнитивной карты, отражающая фрагмент около 43% СК-модели Inf3.

3d-интегральная когнитивная карта является отображением на одном рисунке когнитивных диаграмм классов и значений факторов, отображенных соответственно на рисунках 16 и 20, и одного слоя нейронной сети, приведенного на рисунке 25.



Рисунок 26. 3d-интегральная когнитивная карта в СК-модели Inf3

4.3.7. Когнитивные функции

Вместо описания того, что представляют собой когнитивные функции, приведем help соответствующего режима системы «Эйдос» (рисунок 27) и сошлемся на работу, в которой это довольно подробно описано [14].

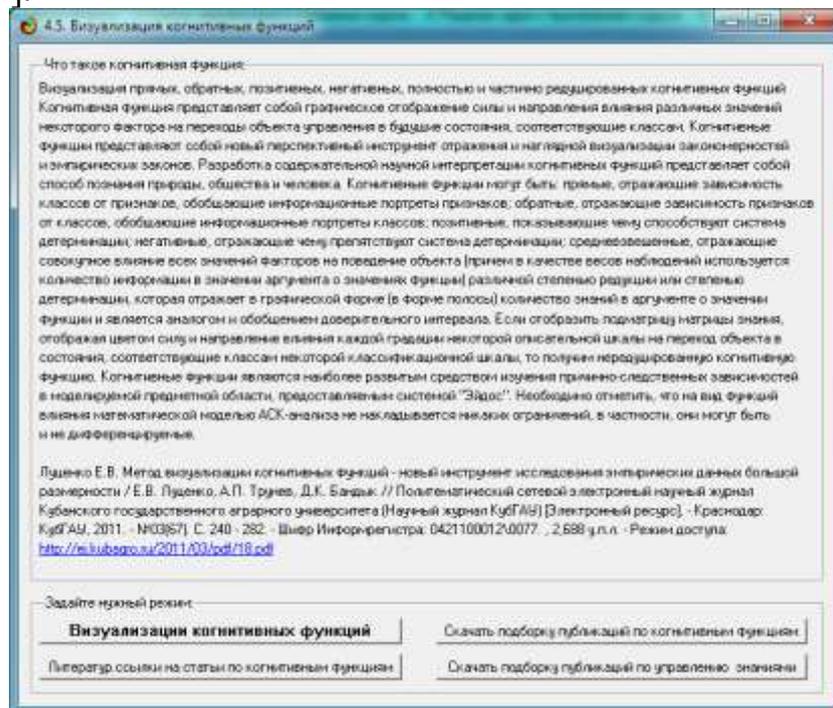


Рисунок 27. Help режима визуализации когнитивных функций

Когнитивная функция представляет собой графическое отображение силы и направления влияния различных значений некоторого фактора (признаков) на переходы объекта управления в будущие состояния,

соответствующие классам. Классы являются градациями классификационных шкал.

Когнитивные функции представляют собой новый перспективный инструмент отражения и наглядной визуализации эмпирических закономерностей и эмпирических законов. Разработка содержательной научной интерпретации когнитивных функций представляет собой способ познания природы, общества и человека.

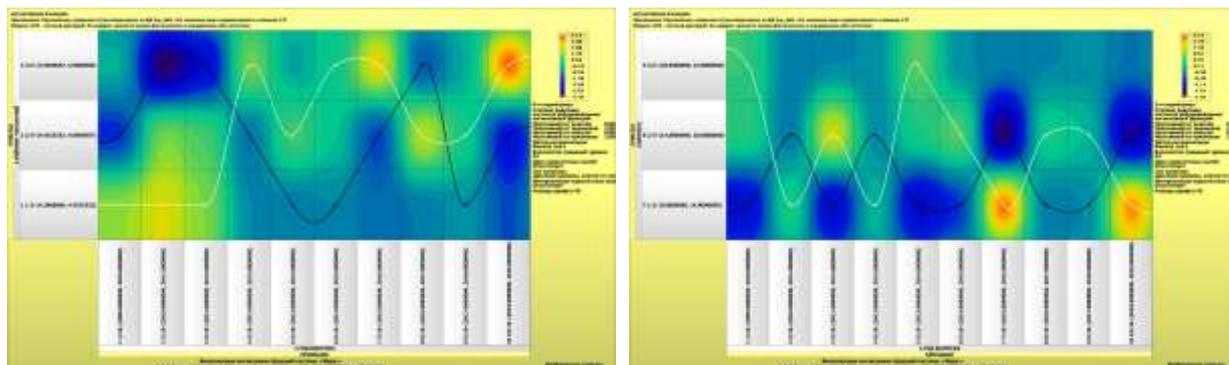
Когнитивные функции могут быть: прямые, отражающие зависимость классов от признаков, обобщающие информационные портреты признаков; обратные, отражающие зависимость признаков от классов, обобщающие информационные портреты классов; позитивные, показывающие чему способствуют система детерминации (обозначены белой линией); негативные, отражающие чему препятствуют система детерминации (обозначены черной линией); 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 разной толщины) 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.

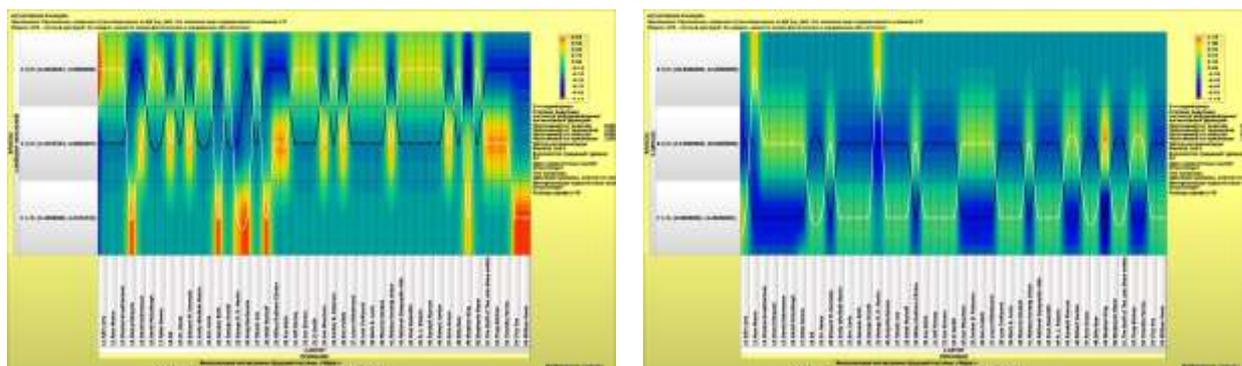
Когнитивные функции являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой "Эйдос".

Необходимо отметить, что *на вид функций влияния математической моделью АСК-анализа не накладывается никаких ограничений*, в частности, они могут быть и не дифференцируемые [14].

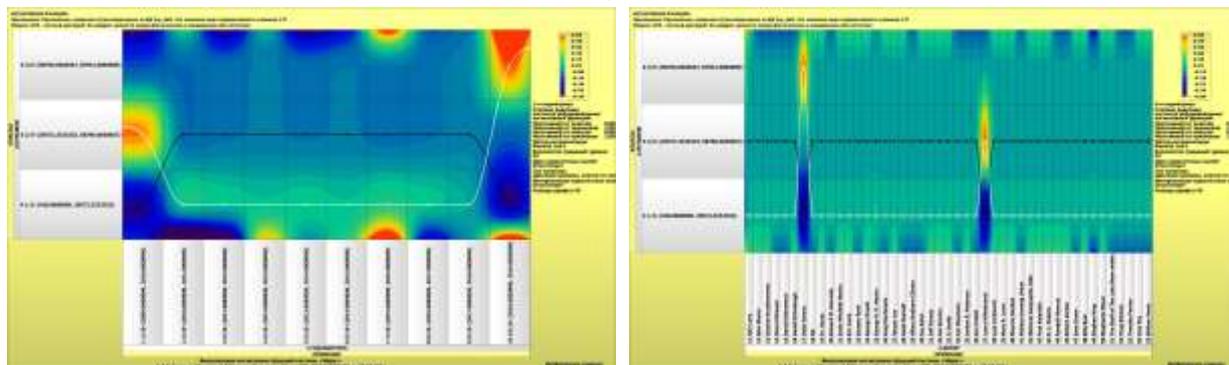
На рисунках 28 приведены примеры некоторых когнитивных функций, наглядно отражающих популярность книг из топ-48 бестселлеров amazon за 2009–2019 гг. (т.е. степени выраженности) различных параметров измерения количества популярности бестселлеров amazon за 2009–2019 гг..



28-1



28-2



28-3

Рисунок 28. Примеры когнитивных функций в СК-модели INF3, отражающие силу и направление влияния значений рейтинга, цены и отзывов на популярность бестселлеров разных авторов на amazon за 2009–2019 гг.

Из когнитивных функций, приведенных на рисунке 28, хорошо видно, что *зависимости между значениями параметров рейтинга, отзывов и показателями отзывов в основном имеют довольно сложный характер.*

Исключением является лишь когнитивные функции, приведенные на рисунках 28-2. На рисунках 28-3 мы видим не так четко выраженную, можно сказать «почти обратно-пропорционально» зависимость. На

рисунке 28-1 отображена «почти прямо-пропорциональная» зависимость авторов и годов выхода книг в период 2009 года - 2019 года.

4.3.8. Сила и направление влияния значений параметров популярности книг и сила влияния самих параметров популярности на авторов и года книг

На рисунках 5, 6, 7 приведены фрагменты некоторых статистических и системно-когнитивных моделей, отражающих моделируемую предметную область.

Строки матриц моделей соответствуют значениям факторов, т.е. значениям параметров популярности (градации описательных шкал).

Колонки матриц моделей соответствуют различным классам, отражающим различное содержание.

Числовые значения в ячейках матриц моделей, находящихся на пересечении строк и колонок, отражают направление (знак) и силу влияния конкретного значения параметра, соответствующего строке, на конкретное значение показателя, соответствующего колонке.

Если какое-то значение параметра автора или года слабо влияет на популярность книг, то в соответствующей строке матрицы модели будут малые по модулю значения разных знаков, если же влияние сильное – то и значения будут большие по модулю разных знаков.

Если значение показателя популярности способствует получению некоторого определенного показателя рейтинга и т.д., то в соответствующей этому результату ячейке матрицы модели будут положительные значения, если же понижает – то и значения будут отрицательные.

Из этого следует, что суммарную силу влияния того или иного значения параметра популярности на показатели рейтинга или отзывов (т.е. ценность данного значения параметра количества для решении задачи квалиметрии и других задач) можно количественно оценивать **степенью вариабельности значений** в строке матрицы модели, соответствующей этому значению.

Существует много мер вариабельности значений: это и среднее модулей отклонения от среднего, и дисперсия, и среднеквадратичное отклонение и другие. В АСК-анализе и системе «Эйдос» для этой цели принято использовать среднеквадратичное отклонение. Численно оно равно стандартному отклонению и вычисляется по той же формуле, но мы предпочитаем не использовать термин «стандартное отклонение», т.к. он предполагает нормальность распределения исследуемых последовательностей чисел, а значит и проверку соответствующих статистических гипотез.

Самая правая колонка в матрицах моделей на рисунках 5, 6, 7 содержит количественную оценку вариабельности значений строки модели (среднеквадратичное отклонение), которая и представляет собой ценность

Из рисунка 32 и таблицы 7 видно, что 1,4% наиболее ценных для решения задачи квалиметрии рейтинга значений популярности обуславливают 74% суммарной ценности, а 5% суммарной ценности обеспечиваются 17% наиболее ценных значений популярности.

Обращает внимание, что наиболее ценным является высокое значение параметра автор - параметр АВТОР-Paul Kalanithi, а наименее ценным – низкое значение параметра ГОД ВЫПУСКА-10/10-{2018.0000000, 2019.0000000}.

4.3.9. Степень детерминированности пола и возраста пострадавших значениями параметров количества

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается *степенью варибельности значений факторов* (градаций описательных шкал) в колонке матрицы модели, соответствующей данному классу (таблица 9).

В данной работе у нас классами являются рейтинг, отзывы и цена книг, а значениями градаций описательных шкал – их авторы и года выходов.

На рисунке 33 мы видим Парето-кривую степени детерминированности классов значениями параметров популярности нарастающим итогом.

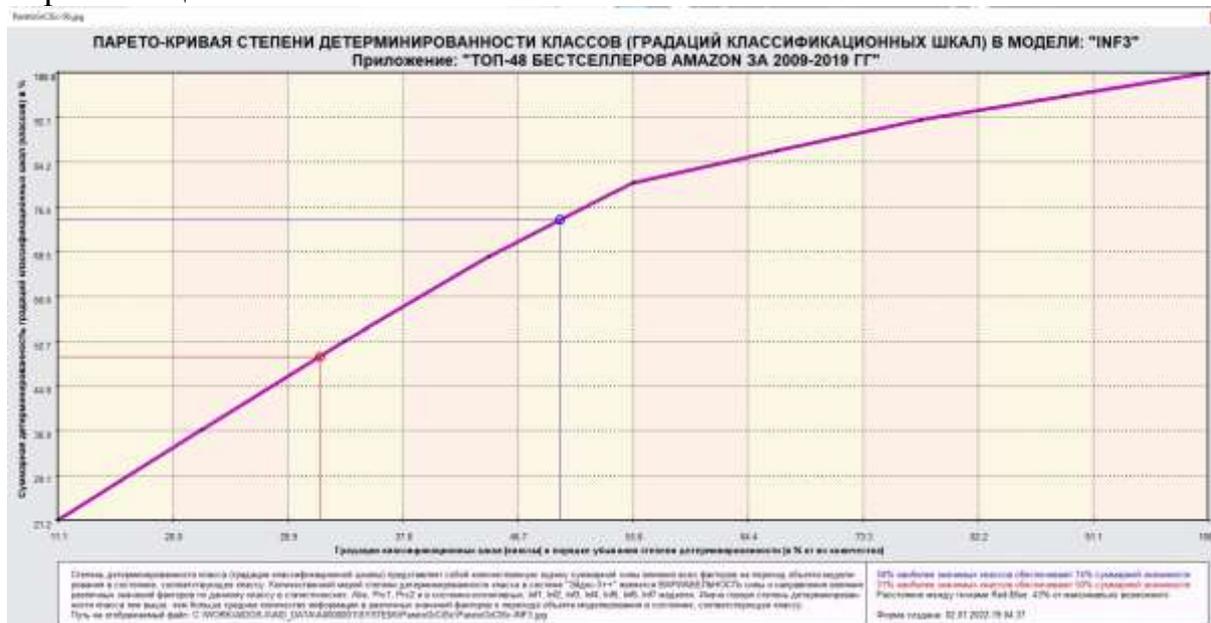


Рисунок 30. Парето-кривая степени детерминированности классов

Степень детерминированности (обусловленности) всей классификационной шкалы является средним от степени детерминированности ее градаций, т.е. классов (таблица 10).

Таблица 8 – Классификационные шкалы, ранжированные по убыванию средней степени детерминированности их градаций в СК-модели INF3

№	Код	Наименование показателя качества	Значимость (%)	Значимость нарастающим итогом (%)
1	1	РЕЙТИНГ ЧИТАТЕЛЕЙ, %	49,928	49,928
2	3	ЦЕНА(\$), %	36,327	86,256
3	2	ОТЗЫВОВ, %	13,744	100,000

Из таблицы 10 видно, что наиболее высокую степень детерминированности обуславливающими их факторами имеет содержание отзывов, а количество почти в 8 раза слабее.

4.3.10. Устойчивость содержания рейтинга и цен от значений параметров авторов книг

Устойчивость зависимостей показателей рейтинга и цен от обуславливающих их значений параметров авторов подразумевает **непрерывность** и **монотонность** этих зависимостей.

Непрерывность зависимостей показателей рейтинга, цен и отзывов от обуславливающих их значений параметров авторов означает, что малые изменения значений параметров популярности детерминируют малые изменения показателей качества рейтинга, а более значительные изменения значения параметров популярности обуславливают и более существенные изменения показателей качества, т.е. *степень изменения значений показателей рейтинга, цен и отзывов соответствует степени изменения обуславливающих их параметров авторов.*

Если непрерывность нарушается, то незначительное изменения значения параметров авторов и года выхода может привести как к малым, так и к значительным изменениям значений показателей качества рейтинга, цен и отзывов, а большие изменения значений параметров авторов и года выхода могут оказать как сильное, так и незначительное влияние на изменение значений показателей качества.

Если в системе управления **нарушается непрерывность управления**, то это воспринимается как ее поломка, неисправность и непригодность для выполнения своей функции.

Например, если нарушается непрерывность зависимости тяги двигателя машины от степени нажатия педали газа, то при плавном увеличении газа машина будет не плавно разгоняться, а начнет дергаться и может вообще заглохнуть, как это бывает у новичков, которые еще не научились правильно трогаться с места.

Монотонность зависимостей показателей пола и возраста пострадавших от значений параметров количества означает, что:

– если параметр количества *способствует* повышению качества пола и возраста пострадавших, то увеличение значения параметра количества приводит к повышению показателя пола и возраста, т.е. между параметром количества и показателем пола и возраста наблюдается прямо-пропорциональная зависимость;

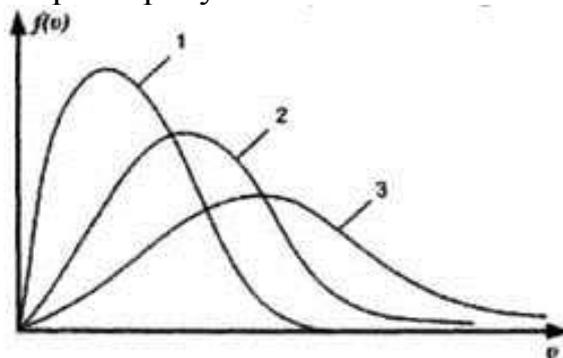
– если же параметр количества *препятствует* повышению пола и возраста, то увеличение значения параметра количества приводит к понижению показателя пола и возраста пострадавших, т.е. между параметром количества и показателем пола и возраста пострадавших наблюдается обратно-пропорциональная зависимость.

Монотонность управления характерна для *линейных* систем управления и нарушается в *нелинейных* системах управления [21]. Система управления является линейной, если для нее выполняется *принцип суперпозиции*, т.е. результат совместного действия на нее совокупности факторов является *суммой* действий каждого из них по отдельности [21].

Если в системе управления *нарушается монотонность управления*, то это может приводить к тому, что при увеличении значения фактора результат может сначала увеличиваться практически пропорционально степени увеличения этого значения, затем *скорость* увеличения результата начинает уменьшаться и затем стабилизируется, а при дальнейшем увеличении значения фактора результат начинает уменьшаться вплоть до нуля или даже отрицательных значений (например, вместо прибыли получены убытки). Могут наблюдаться и другие нелинейные зависимости значения функции от значения аргумента.

По сути, *при нарушении монотонности управления меняется знак первой производной результата управления по значению фактора, нарушается знакоопределенность этой первой производной*³. Понятно, что *немонотонные функции не являются непрерывными*.

Принципиальный вид кривой влияния интенсивности фактора на результат в нелинейной системе при этом получается очень похожий у всех факторов (для примера на рисунке 34 показаны 3 из них):



³ Это вызывает ассоциации с классическим понятием устойчивости управления по Ляпунову.

Рисунок 31. Принципиальный вид кривой влияния интенсивности фактора на нелинейный объект управления⁴.

Например, если по оси X показать интенсивность полива какой-либо конкретной культуры, а по Y урожайность, то график на рисунке 31 можно интерпретировать таким образом, что при полном отсутствии полива урожайность будет минимальной, при его увеличении урожайность будет возрастать сначала быстро, потом все медленнее, затем достигнет максимума, а потом при дальнейшем увеличении полива она начнет уменьшаться пока опять не достигнет минимума (нуля, если культура не рис), когда все поле превратится в озеро.

*Принципиально важно, что один и тот же полив будет действовать по-разному при условии одновременного действия других факторов, причем **при этом смещается точка оптимума**, т.е. при действии других факторов оптимальный полив становится другой, в чем и проявляется **нелинейность** системы и взаимодействие факторов, нарушение для них принципа суперпозиции (кривые 1, 2, 3 на рисунке 34).*

Нарушение монотонности управления может приводить к **различным видам зависимостей** результатов от значений управляющих факторов: это могут быть зависимости, типа показанных на рисунке 34; *периодические* зависимости (ярким примером является таблица Д.И.Менделеева, в которой свойства химических элементов изменяются периодически при линейном увеличении заряда ядра), а также сложные зависимости, в которых трудно найти какую-либо закономерность (напоминающие *случайные*).

Таким образом, у нас есть все основания разделить все значения параметров популярности, действующие на рейтинг, отзывы и цену книг, относящиеся к одной классификационной шкале, на **три основные группы** (рисунок 31):

1. **Способствующие** получению более высоких результатов.
2. **Препятствующие** получению более высоких результатов.
3. **Действующие сложным и неоднозначным образом.**

7. Выводы

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

⁴ Источник рисунка: http://san-of-war2.narod.ru/fiziks/fiziks_image481.jpg На самом деле на рисунке показано распределение Максвелла молекул газа по скоростям при разных температурах. Удивительно, но подобный вид имеет влияние интенсивности различных факторов на различные нелинейные объекты управления

В результате проделанной работы, с помощью системы «Эйдос» были созданы 3 статистические и 7 системно-когнитивных моделей, в которых непосредственно на основе эмпирических данных сформированы обобщенные образы классов по различным значениям показателей популярности посредством именитых авторов и годами выпуска этих самых книг, изучено влияние значений различных параметров рейтинга, отзывов и цен книг на эти показатели, и, на основе этого, решены задачи квалиметрии, классификации и исследования моделируемой предметной области путем исследования ее модели.

Все это, является примером успешного применения когнитивных и информационных технологий для решения задач литературного характера и может быть основанием для предложения нового научного направления.

Литература

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.

2. 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. <http://elibrary.ru/item.asp?id=18632909>

3. 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.

4. 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.

5. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-X++» / 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.

6. 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.

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. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. <http://elibrary.ru/item.asp?id=21358220>

9. Луценко Е.В., Открытая масштабируемая интерактивная интеллектуальная online среда «Эйдос» («Эйдос-online»). Свид. РосПатента РФ на программу для ЭВМ, Заявка № 2017618053 от 07.08.2017, Гос.рег.№ 2017661153, зарегистр. 04.10.2017. – Режим доступа: <http://lc.kubagro.ru/aidos/2017661153.jpg>, 2 у.п.л.

10. Луценко Е.В. Открытая масштабируемая интерактивная интеллектуальная online среда для обучения и научных исследований на базе АСК-анализа и системы «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №06(130). С. 1 – 55. – IDA [article ID]: 1301706001. – Режим доступа: <http://ej.kubagro.ru/2017/06/pdf/01.pdf>, 3,438 у.п.л. http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf

11. Луценко Е.В. Проблемы и перспективы теории и методологии научного познания и автоматизированный системно-когнитивный анализ как автоматизированный метод научного познания, обеспечивающий содержательное феноменологическое моделирование / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №03(127). С. 1 – 60. – IDA [article ID]: 1271703001. – Режим доступа: <http://ej.kubagro.ru/2017/03/pdf/01.pdf>, 3,75 у.п.л.

12. Луценко Е.В. Универсальная автоматизированная система распознавания образов "Эйдос" (версия 4.1).-Краснодар: КЮИ МВД РФ, 1995.- 76с. <http://elibrary.ru/item.asp?id=18630282>

13. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос». Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0. <http://elibrary.ru/item.asp?id=22401787>

14. Луценко Е. В., Лойко В. И., Лаптев В. Н. Системы представления и приобретения знаний : учеб. пособие / Е. В. Луценко, В. И. Лойко, В. Н. Лаптев. – Краснодар : Экоинвест, 2018. – 513 с. ISBN 978-5-94215-415-8. <https://elibrary.ru/item.asp?id=35641755>

15. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-Х++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С.

164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 у.п.л.

16. URL[электронный ресурс] Kaggle - <https://www.kaggle.com/sootersaalu/amazon-top-50-best-selling-books-2009-2019>

Literatura

1. Lucenko E.V. Sintez adaptivny`x intellektual`ny`x izmeritel`ny`x sistem s primeneniem ASK-analiza i sistemy` «E`jdos» i sistemnaya identifikaciya v e`konometrike, biometrii, e`kologii, pedagogike, psixologii i medicine / E.V. Lucenko // Politematicheskij setevoy e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2016. – №02(116). S. 1 – 60. – IDA [article ID]: 1161602001. – Rezhim dostupa: <http://ej.kubagro.ru/2016/02/pdf/01.pdf>, 3,75 u.p.l.

2. Lucenko E.V. Avtomatizirovanny`j sistemno-kognitivny`j analiz v upravlenii aktivny`mi ob`ektami (sistemnaya teoriya informacii i ee primeneniye v issledovanii e`konomicheskix, social`no-psixologicheskix, texnologicheskix i organizacionno-texnicheskix sistem): Monografiya (nauchnoe izdanie). – Krasnodar: KubGAU. 2002. – 605 s. <http://elibrary.ru/item.asp?id=18632909>

3. Lucenko E.V. Metrizaciya izmeritel`ny`x shkal razlichny`x tipov i sovmestnaya sopostavimaya kolichestvennaya obrabotka raznorodny`x faktorov v sistemno-kognitivnom analize i sisteme «E`jdos» / E.V. Lucenko // Politematicheskij setevoy e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2013. – №08(092). S. 859 – 883. – IDA [article ID]: 0921308058. – Rezhim dostupa: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 u.p.l.

4. Lucenko E.V. Invariantnoe otnositel`no ob`emov danny`x nechetkoe mul`tiklassovoe obobshhenie F-mery` dostovernosti modelej Van Rizbergena v ASK-analize i sisteme «E`jdos» / E.V. Lucenko // Politematicheskij setevoy e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2017. – №02(126). S. 1 – 32. – IDA [article ID]: 1261702001. – Rezhim dostupa: <http://ej.kubagro.ru/2017/02/pdf/01.pdf>, 2 u.p.l.

5. Lucenko E.V. Kolichestvenny`j avtomatizirovanny`j SWOT- i PEST-analiz sredstvami ASK-analiza i intellektual`noj sistemy` «E`jdos-X++» / E.V. Lucenko // Politematicheskij setevoy e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2014. – №07(101). S. 1367 – 1409. – IDA [article ID]: 1011407090. – Rezhim dostupa: <http://ej.kubagro.ru/2014/07/pdf/90.pdf>, 2,688 u.p.l.

6. Lucenko E.V. Metod kognitivnoj klasterizacii ili klasterizaciya na osnove znaniy (klasterizaciya v sistemno-kognitivnom analize i intellektual`noj sisteme «E`jdos») / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setevoy e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2011. – №07(071). S. 528 – 576. – Shift Informregistra: 0421100012\0253, IDA [article ID]: 0711107040. – Rezhim dostupa: <http://ej.kubagro.ru/2011/07/pdf/40.pdf>, 3,062 u.p.l.

7. Lucenko E.V. Sistemnaya teoriya informacii i nelokal`ny`e interpretiruemy`e nejronny`e seti pryamogo scheta / E.V. Lucenko // Politematicheskij setevoy e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2003. – №01(001). S. 79 – 91. –

IDA [article ID]: 0010301011. – Rezhim dostupa: <http://ej.kubagro.ru/2003/01/pdf/11.pdf>, 0,812 u.p.l.

8. Orlov A.I., Lucenko E.V. Sistemnaya nechetkaya interval'naya matematika. Monografiya (nauchnoe izdanie). – Krasnodar, KubGAU. 2014. – 600 s. ISBN 978-5-94672-757-0. <http://elibrary.ru/item.asp?id=21358220>

9. Lucenko E.V., Otkry'taya masshtabiruemaya interaktivnaya intellektual'naya on-line sreda «E`jdos» («E`jdos-online»). Svid. RosPatenta RF na programmu dlya E`VM, Zayavka № 2017618053 ot 07.08.2017, Gos.reg.№ 2017661153, zaregistr. 04.10.2017. – Rezhim dostupa: <http://lc.kubagro.ru/aidos/2017661153.jpg>, 2 u.p.l.

10. Lucenko E.V. Otkry'taya masshtabiruemaya interaktivnaya intellektual'naya on-line sreda dlya obucheniya i nauchny`x issledovanij na baze ASK-analiza i sistemy` «E`jdos» / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2017. – №06(130). S. 1 – 55. – IDA [article ID]: 1301706001. – Rezhim dostupa: <http://ej.kubagro.ru/2017/06/pdf/01.pdf>, 3,438 u.p.l. http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf

11. Lucenko E.V. Problemy` i perspektivy` teorii i metodologii nauchnogo poznaniya i avtomatizirovanny`j sistemno-kognitivny`j analiz kak avtomatizirovanny`j metod nauchnogo poznaniya, obespechivayushhij soderzhatel`noe fenomenologicheskoe modelirovanie / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2017. – №03(127). S. 1 – 60. – IDA [article ID]: 1271703001. – Rezhim dostupa: <http://ej.kubagro.ru/2017/03/pdf/01.pdf>, 3,75 u.p.l.

12. Lucenko E.V. Universal'naya avtomatizirovannaya sistema raspoznavaniya obrazov "E`jdos" (versiya 4.1).-Krasnodar: KYuI MVD RF, 1995.- 76s. <http://elibrary.ru/item.asp?id=18630282>

13. Lucenko E.V. Universal'naya kognitivnaya analiticheskaya sistema «E`jdos». Monografiya (nauchnoe izdanie). – Krasnodar, KubGAU. 2014. – 600 s. ISBN 978-5-94672-830-0. <http://elibrary.ru/item.asp?id=22401787>

14. Lucenko E. V., Lojko V. I., Laptev V. N. Sistemy` predstavleniya i priobreteniya znaniy : ucheb. posobie / E. V. Lucenko, V. I. Lojko, V. N. Laptev. – Krasnodar : E`koinvest, 2018. – 513 s. ISBN 978-5-94215-415-8. <https://elibrary.ru/item.asp?id=35641755>

15. Lucenko E.V. Modelirovanie slozhny`x mnogofaktorny`x nelinejny`x ob`ektov upravleniya na osnove fragmentirovanny`x zashumlenny`x e`mpiricheskix danny`x bol'shoj razmernosti v sistemno-kognitivnom analize i intellektual`noj sisteme «E`jdos-X++» / E.V. Lucenko, V.E. Korzhakov // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2013. – №07(091). S. 164 – 188. – IDA [article ID]: 0911307012. – Rezhim dostupa: <http://ej.kubagro.ru/2013/07/pdf/12.pdf>, 1,562 u.p.l.

16. URL[электронный ресурс] Kaggle - <https://www.kaggle.com/sootersaalu/amazon-top-50-best-selling-books-2009-2019>