

УДК 004.8

Автоматизированный системно-когнитивный анализ поведения покупателей

Саркитова Оксана Назыровна
студент факультета ПИ, группы ПИ1822
sarkitovao@mail.ru

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

Целью данной работы является изучение поведение покупателей. Достижение данной цели позволяет получить практические навыки работы с универсальной когнитивной аналитической системой «Эйдос-Х++». Для достижения поставленной цели применяется Автоматизированный системно-когнитивный анализ и использующий его программный инструментарий – интеллектуальная система «Эйдос».

Ключевые слова: ПОВЕДЕНИЕ ПОКУПАТЕЛЕЙ, СТАТИСТИКА, АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, АСК-АНАЛИЗ, СИСТЕМА «ЭЙДОС».

Automated system-cognitive analysis of customer behavior

Sarkitova Oksana Nazyrovna
student of the faculty of PI, group PI1822
sarkitovao@mail.ru

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

The aim of this work is to study the behavior of buyers. Achieving this goal allows you to gain practical skills in working with the Eidos-X ++ universal cognitive analytical system. To achieve this goal, the Automated System-Cognitive Analysis and the software tools that use it - the intelligent system "Eidos", are used.

Key words: CUSTOMER BEHAVIOR , STATISTICS, AUTOMATED SYSTEM-COGNITIVE ANALYSIS, ASC-ANALYSIS, "AIDOS" SYSTEM.

СОДЕРЖАНИЕ

КОГНИТИВНАЯ СТРУКТУРИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ	7
ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ И ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ	8
РЕШЕНИЕ РАЗЛИЧНЫХ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	20
ПРОГНОЗИРОВАНИЕ (ДИАГНОСТИКА, КЛАССИФИКАЦИЯ, РАСПОЗНАВАНИЕ, ИДЕНТИФИКАЦИЯ).....	20
ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ (SWOT-АНАЛИЗ)	22
ИССЛЕДОВАНИЕ МОДЕЛИРУЕМОЙ ПРЕДМЕТНОЙ ОБЛАСТИ ПУТЕМ ИССЛЕДОВАНИЯ ЕЕ МОДЕЛИ	24
<i>Когнитивные диаграммы классов</i>	24
<i>Нелокальные нейроны и нелокальные нейронные сети</i>	25
<i>3d-интегральные когнитивные карты</i>	28
<i>Сила и направление влияния значений исследуемых факторов на факт покупки</i>	29
<i>Степень детерминированности классов</i>	32
ВЫВОД.....	33
СПИСОК ЛИТЕРАТУРЫ	34

Введение

Цель данной работы – исследование данных о поведении покупателей на наличие зависимости между фактом покупки товара и характеристикой потребителя.

Для достижения поставленной цели будут использоваться инструментальные средства программной среды «Aidos-X», утилизирующие метод АСК-анализа, предполагающий выполнение следующих этапов:

1. Когнитивная структуризация предметной области
2. Подготовка исходных данных и формализация предметной области
3. Синтез и верификация статистических и системно-когнитивных моделей, и выбор наиболее достоверной модели
4. Решение различных задач в наиболее достоверной модели:
 - a. Прогнозирование (диагностика, классификация, распознавание, идентификация)
 - b. Поддержка принятия решений
 - c. 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)

АСК-анализ – 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 имеет свой программный инструментарий – универсальную когнитивную аналитическую систему «AIDOS-X++», выгодно отличающуюся от других интеллектуальных систем следующими возможностями:

- 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>), в которых не требуется автоматического, т.е. без непосредственного участия человека в реальном времени решения задач идентификации, прогнозирования, принятия решений и исследования предметной области;
- Находится в полном открытом бесплатном доступе (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. не требует от пользователя специальной подготовки в области технологий искусственного интеллекта: «имеет нулевой порог входа» (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 Эйдос-приложений (в настоящее время их 31 и 293, соответственно) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);
- Поддерживает online среду накопления знаний и широко используется во всем мире (<http://aidos.byethost5.com/map5.php>);
- Обеспечивает мультиязычную поддержку интерфейса на 51 языке. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;

- Наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решение этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA);
- Обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе: 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;
- Вместо того, чтобы предъявлять к исходным данным практически неосуществимые требования (вроде нормальности распределения, абсолютной точности и полных повторений всех сочетаний значений факторов и их полной независимости и аддитивности) автоматизированный системно-когнитивный анализ (АСК-анализ) предлагает без какой-либо предварительной обработки осмыслить эти данные и тем самым преобразовать их в информацию, а затем преобразовать эту информацию в знания путем ее применения для достижения целей (т.е. для управления) и решения задач классификации, поддержки принятия решений и содержательного эмпирического исследования моделируемой предметной области.

В чем сила подхода, реализованного в системе Эйдос? В том, что она реализует подход, эффективность которого не зависит от того, что мы думаем о предметной области и думаем ли вообще. Она формирует модели непосредственно на основе эмпирических данных, а не на основе наших представлений о механизмах реализации закономерностей в этих данных. Именно поэтому Эйдос-модели эффективны даже если наши представления о предметной области ошибочны или вообще отсутствуют.

В этом и слабость этого подхода, реализованного в системе Эйдос. Модели системы Эйдос — это феноменологические модели, отражающие эмпирические закономерности в фактах обучающей выборки, т. е. они не отражают причинно-следственного механизма детерминации, а только сам факт и характер детерминации. Содержательное объяснение этих эмпирических закономерностей формулируется уже на теоретическом уровне познания в теоретических научных законах.

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, алгоритм поиска которого представлен на рисунке 1.

**Последовательность обработки данных, информации и знаний в системе «Эйдос»,
повышение уровня системности данных, информации и знаний,
повышение уровня системности моделей**

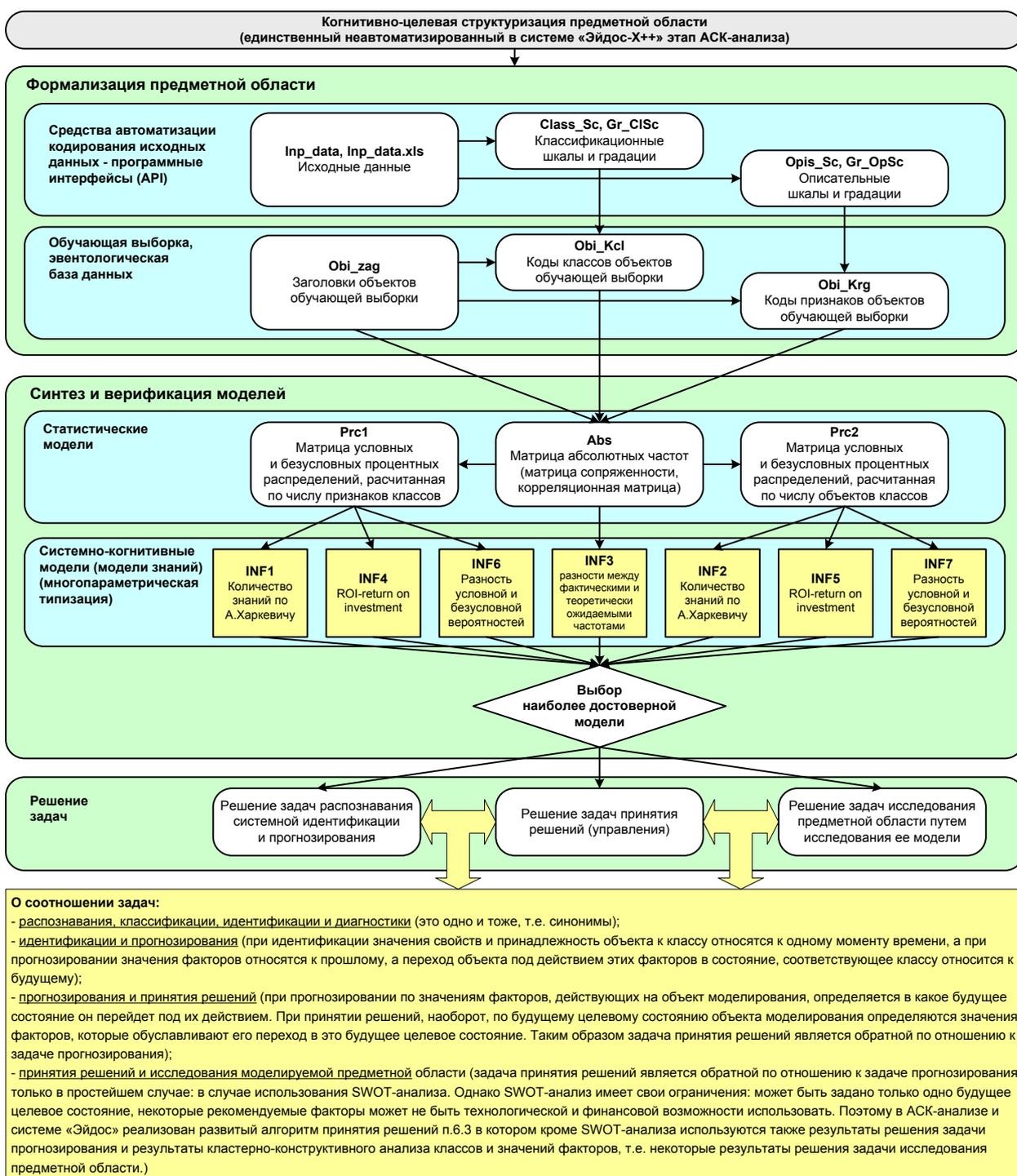


Рисунок 1. Последовательность решения задач в АСК-анализе и системе «Эйдос»

Когнитивная структуризация предметной области

На этапе когнитивно-целевой структуризации предметной области не формализуемым путем было решено, что рассматривать в качестве факторов, действующих на моделируемый объект, а что – в качестве результатов действия этих факторов.

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

1. Содержательная интерпретация СКМ – 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
2. Даже если содержательной интерпретации не разработано, не исключается возможность пользоваться СКМ для достижения заданных результатов и поставленных целей

В ходе проведения исследования для составления классификационной шкалы был выбран параметр «Покупка», а для составления описательных – параметры «Пол», «Возраст» и «Заработная плата»:

Таблица 1. Классификационная шкала

№ п/п	Наименование
1	Покупка

Таблица 2. Описательные шкалы

№ п/п	Наименование
1	Пол
2	Возраст
3	Заработная плата

Подготовка исходных данных и формализация предметной области

Исходные данные были взяты с web-портала Kaggle и представляют из себя данные о 400 клиентах розничной компании с включением информации о совершённых ими покупках.

Таблица 3. Фрагмент исходных данных (2019 год)

ID	Пол	Возраст	Заработная плата	Покупка
15624510	Male	19	19000	0
15810944	Male	35	20000	0
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	0
15728773	Male	27	58000	0
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	0
15727311	Female	35	65000	0
15570769	Female	26	80000	0
15606274	Female	26	52000	0
15746139	Male	20	86000	0
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1
15617482	Male	45	26000	1
15704583	Male	46	28000	1
15621083	Female	48	29000	1
15649487	Male	45	22000	1
15736760	Female	47	49000	1
15714658	Male	48	41000	1
15599081	Female	45	22000	1
15705113	Male	46	23000	1
15631159	Male	47	20000	1
15792818	Male	49	28000	1
15633531	Female	47	30000	1
15744529	Male	29	43000	0

Здесь факт покупки отображается в двоичном формате, то есть 0 – товар не куплен, 1 – товар куплен. Столбец «ID» в исследовании использоваться не будет, так как не является репрезентативным.

Для исследования этих данных с помощью выбранного программного обеспечения был выполнен их импорт.

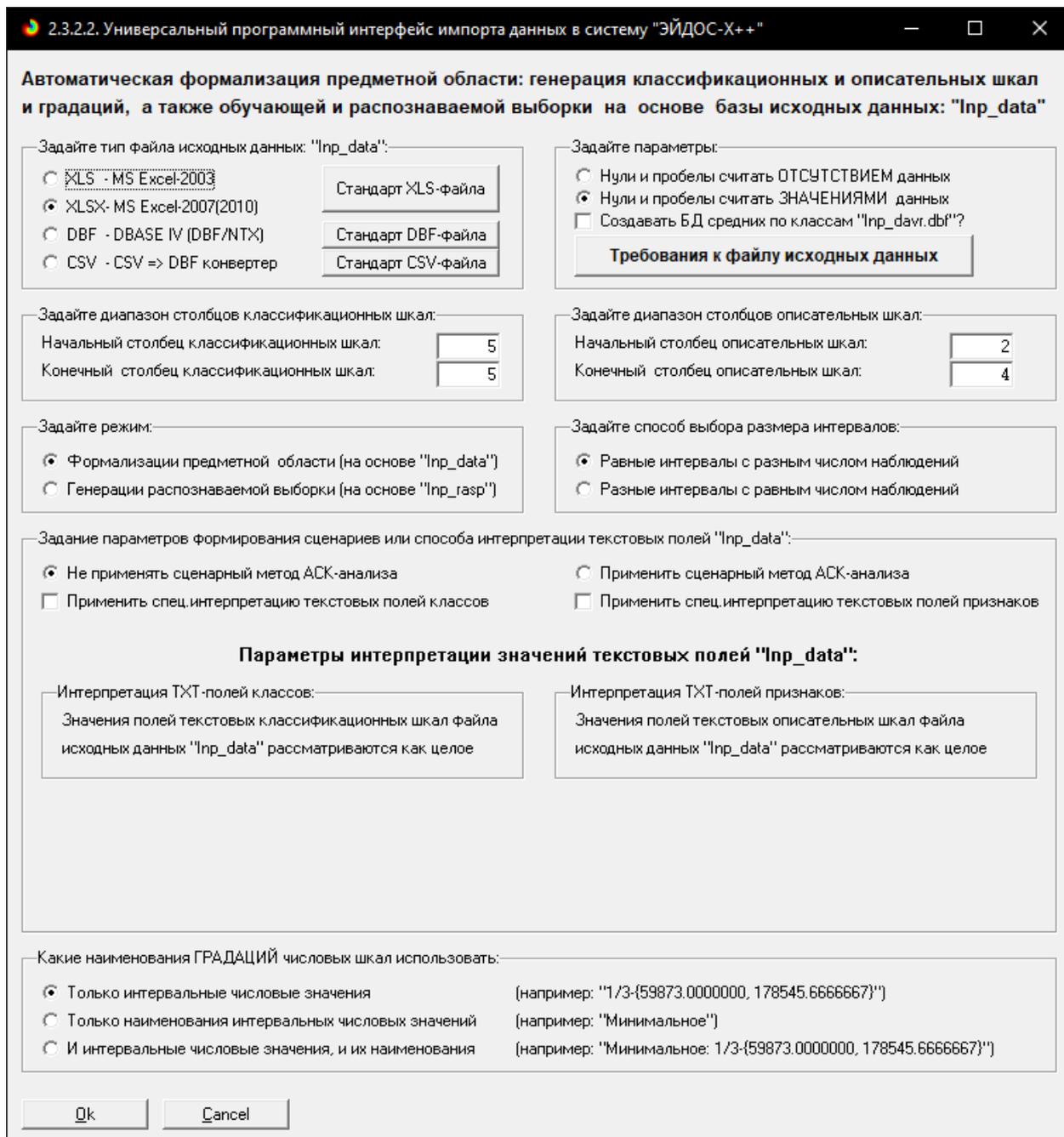


Рисунок 2. Интерфейс импорта данных в систему

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

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [2 x 22]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс. шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис. шкалу
Числовые	1	2	2,00	2	20	10,00
Текстовые	0	0	0,00	1	2	2,00
ВСЕГО:	1	2	2,00	3	22	7,33

Задайте число интервалов (градаций) в шкале:

В классификационных шкалах: В описательных шкалах:

Рисунок 3. Интерфейс настройки размерности модели

Важно, что заданные адаптивные интервалы учитывают 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.

На рисунке 4 приведена справочная форма данного режима, в котором объясняется принцип организации таблицы исходных данных для данного режима.

Помощь по режиму 2.3.2.2 для случая Excel-файлов исходных данных

Режим 2.3.2.2: Универсальный программный интерфейс импорта данных из внешней базы данных "Inp_data.xls" в систему "Эйдос-Х++" и формализации предметной области.

- Данный программный интерфейс обеспечивает формализацию предметной области, т.е. анализ файла исходных данных (Inp_data.xls(x)), формирование классификационных и описательных шкал и градаций, а затем кодирование файла исходных с их использованием.
- Файл исходных данных должен иметь имя: Inp_data.xls(x), а файл распознаваемой выборки имя: Inp_rasp.xls(x). Файлы Inp_data.xls(x) и Inp_rasp.xls(x) должны находиться в папке ../AIDOS-Х/AID_DATA/Inp_data/. Эти файлы имеют совершенно одинаковую структуру.
- 1-я строка этого файла должна содержать наименования колонок на любом языке, в т.ч. и русском. Эти наименования должны быть во всех колонках, при этом переносы по словам разрешены, а объединение ячеек, разрыв строки знак абзаца не допускаются. Эти наименования должны быть короткими, но понятными, т.к. они будут в выходных формах, а к ним еще будут добавляться наименования градаций. В числовых шкалах надо обязательно указывать единицы измерения и число знаков после запятой в колонке должно быть одинаковое.
- 1-я колонка содержит наименование объекта обучающей выборки или наименование наблюдения. Оно может быть длинным: до 255 символов.
- Каждая строка этого файла, начиная со 2-й, содержит данные об одном объекте обучающей выборки или одном наблюдении. В MS Excel-2003 в листе может быть до 65536 строк и до 256 колонок. В листе MS Excel-2010 и более поздних возможно до 1048576 строк и 16384 колонок.
- Столбцы, начиная со 2-го, являются классификационными и описательными шкалами и могут быть текстового (номинального / порядкового) или числового типа (с десятичными знаками после запятой).
- Столбцы присваиваются числовой тип, если все значения его ячеек числового типа. Если хотя бы одно значение является текстовым (не числом, в т.ч. пробелом), то столбцу присваивается текстовый тип. Это означает, что нули должны быть указаны нулями, а не пробелами.
- Столбцы со 2-го по N-й являются классификационными шкалами (выходными параметрами) и содержат данные о классах (будущих состояниях объекта управления), к которым принадлежат объекты обучающей выборки.
- Столбцы с N+1 по последний являются описательными шкалами (свойствами или факторами) и содержат данные о признаках (т.е. значениях свойств или значениях факторов), характеризующих объекты обучающей выборки.
- В результате работы режима формируется файл INP_NAME.TXT стандарта MS DOS (кириллица), в котором наименования классификационных и описательных шкал являются СТРОКАМИ. Система формирует классификационные и описательные шкалы и градации. Для этого в каждом числовом столбце система находит минимальное и максимальное числовые значения и формирует заданное количество числовых интервалов, после чего числовые значения заменяются их интервальными значениями. В текстовых столбцах система находит уникальные текстовые значения. Каждое УНИКАЛЬНОЕ интервальное числовое или текстовое значение считается градацией классификационной или описательной шкалы, характеризующей объект. В каждой шкале ее градации сортируются по алфавиту. С использованием шкал и градаций кодируются исходные данные в результате чего генерируется обучающая выборка, каждый объект которой соответствует одной строке файла исходных данных NP_DATA и содержит коды классов, соответствующие фактам совпадения числовых или уникальных текстовых значений классов с градациями классификационных шкал и коды признаков, соответствующие фактам совпадения числовых или уникальных текстовых значений признаков с градациями описательных шкал
- Распознаваемая выборка формируется на основе файла INP_RASP аналогично, за исключением того, что классификационные и описательные шкалы и градации не создаются, а используются ранее созданные в модели, и базы распознаваемой выборки могут не включать коды классов, если столбцы классов в файле INP_RASP были пустыми. Структура файла INP_RASP должна быть такая же, как INP_DATA, т.е. они должны ПОЛНОСТЬЮ совпадать по наименованиям столбцов, но могут иметь разное количество строк с разными значениями в них.

Принцип организации таблицы исходных данных:

Наименование объекта обучающей выборки	Наименование 1-й классификационной шкалы	Наименование 2-й классификационной шкалы	...	Наименование 1-й описательной шкалы	Наименование 2-й описательной шкалы	...
1-й объект обучающей выборки (1-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
2-й объект обучающей выборки (2-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
...

Определения основных терминов и профилактика типичных ошибок при подготовке Excel-файла исходных данных

Рисунок 4 Экранная форма «HELP» режима 2.3.2.2

В результате работы режима были сформированы:

- 1 классификационная шкала «Покупка» с 2 градациями
- 3 описательных шкалы – «Пол», «Возраст» и «Заработная плата», количество градаций для которых составило 2, 10 и 10 соответственно

Таблица 4. Описательная шкала «Возраст» и её градации

1	1/10- {18; 22,2}
2	1/10- {22,2; 26,4}
3	1/10- {26,4; 30,6}
4	1/10- {30,6; 34,8}
5	1/10- {34,8; 39}
6	1/10- {39; 43,2}
7	1/10- {43,2; 47,4}
8	1/10- {47,4; 51,6}
9	1/10- {51,6; 55,8}
10	1/10- {55,8; 60}

Таблица 5. Описательная шкала «Заработная плата» и её градации

1	1/10- {15000; 28500}
2	1/10- {28500; 42000}
3	1/10- {42000; 55500}
4	1/10- {55500; 69000}
5	1/10- {69000; 82500}
6	1/10- {82500; 96000}
7	1/10- {96000; 109500}
8	1/10- {109500; 123000}
9	1/10- {123000; 136500}
10	1/10- {136500; 150000}

Таким образом, созданы все необходимые и достаточные условия для выполнения следующего этапа АСК-анализа – синтеза и верификации моделей.

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) моделей был осуществлён в режиме 3.5 системы, пользовательский интерфейс которой изображён на рисунке 5.

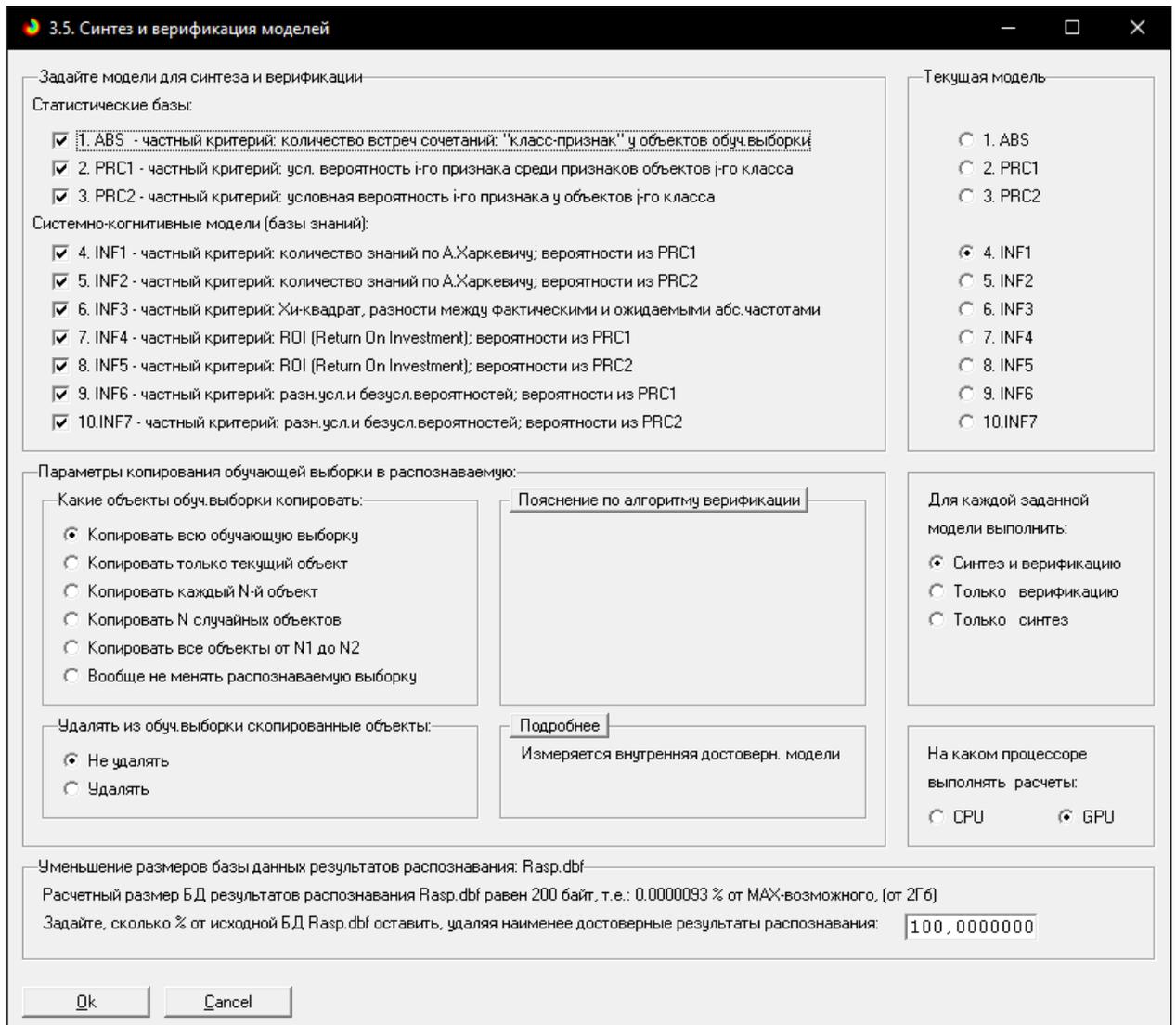


Рисунок 5. Интерфейс выбора моделей для синтеза и верификации

Из-за небольшого размера исходной выборки было решено использовать 100% от исходной выборки данных, то есть без удаления кортежей с наиболее недостоверными результатами распознавания.

Фрагменты самих построенных статистических и системно-когнитивных моделей (СК-модели) приведены на рисунках 6, 7:

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

Код признака	Наименование описательной шкалы и градации	1. PURCHASED 1/2 {0.0000000, 0.5000000}	2. PURCHASED 2/2 {0.5000000, 1.0000000}	Сумма	Среднее	Средн. квадрат. откл.
1	GENDER-Female	127	77	204	102.00	35.36
2	GENDER-Male	130	66	196	98.00	45.25
3	AGE-1/10-{18.0000000, 22.2000000}	28		28	14.00	19.80
4	AGE-2/10-{22.2000000, 26.4000000}	37		37	18.50	26.16
5	AGE-3/10-{26.4000000, 30.6000000}	40	6	46	23.00	24.04
6	AGE-4/10-{30.6000000, 34.8000000}	27	8	35	17.50	13.44
7	AGE-5/10-{34.8000000, 39.0000000}	70	22	92	46.00	33.94
8	AGE-6/10-{39.0000000, 43.2000000}	38	12	50	25.00	18.38
9	AGE-7/10-{43.2000000, 47.4000000}	9	26	35	17.50	12.02
10	AGE-8/10-{47.4000000, 51.6000000}	5	26	31	15.50	14.85
11	AGE-9/10-{51.6000000, 55.8000000}	1	17	18	9.00	11.31
12	AGE-10/10-{55.8000000, 60.0000000}	2	26	28	14.00	16.97
13	ESTIMATEDSALARY-1/10-{15000.0000000, 28500.0000000}	35	16	51	25.50	13.44
14	ESTIMATEDSALARY-2/10-{28500.0000000, 42000.0000000}	21	23	44	22.00	1.41
15	ESTIMATEDSALARY-3/10-{42000.0000000, 55500.0000000}	47	6	53	26.50	28.99
16	ESTIMATEDSALARY-4/10-{55500.0000000, 69000.0000000}	49	2	51	25.50	33.23
17	ESTIMATEDSALARY-5/10-{69000.0000000, 82500.0000000}	67	14	81	40.50	37.48
18	ESTIMATEDSALARY-6/10-{82500.0000000, 96000.0000000}	27	17	44	22.00	7.07
19	ESTIMATEDSALARY-7/10-{96000.0000000, 109500.0000000}	1	17	18	9.00	11.31
20	ESTIMATEDSALARY-8/10-{109500.0000000, 123000.0000000}	7	13	20	10.00	4.24
21	ESTIMATEDSALARY-9/10-{123000.0000000, 136500.0000000}	2	14	16	8.00	8.49
22	ESTIMATEDSALARY-10/10-{136500.0000000, 150000.0000000}	1	21	22	11.00	14.14
	Сумма числа признаков	771	429	1200		
	Среднее	35	20		27.27	

Рисунок 6. Фрагмент матрицы абсолютных частот

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

Код признака	Наименование описательной шкалы и градации	1. PURCHASED 1/2 {0.0000000, 0.5000000}	2. PURCHASED 2/2 {0.5000000, 1.0000000}	Сумма	Среднее	Средн. квадрат. откл.
1	GENDER-Female	-4.070	4.070			5.756
2	GENDER-Male	4.070	-4.070			5.756
3	AGE-1/10-{18.0000000, 22.2000000}	10.010	-10.010			14.156
4	AGE-2/10-{22.2000000, 26.4000000}	13.228	-13.227			18.707
5	AGE-3/10-{26.4000000, 30.6000000}	10.445	-10.445			14.771
6	AGE-4/10-{30.6000000, 34.8000000}	4.512	-4.512			6.382
7	AGE-5/10-{34.8000000, 39.0000000}	10.890	-10.890			15.401
8	AGE-6/10-{39.0000000, 43.2000000}	5.875	-5.875			8.309
9	AGE-7/10-{43.2000000, 47.4000000}	-13.488	13.488			19.074
10	AGE-8/10-{47.4000000, 51.6000000}	-14.917	14.917			21.097
11	AGE-9/10-{51.6000000, 55.8000000}	-10.565	10.565			14.941
12	AGE-10/10-{55.8000000, 60.0000000}	-15.990	15.990			22.613
13	ESTIMATEDSALARY-1/10-{15000.0000000, 28500.0000000}	2.233	-2.233			3.157
14	ESTIMATEDSALARY-2/10-{28500.0000000, 42000.0000000}	-7.270	7.270			10.281
15	ESTIMATEDSALARY-3/10-{42000.0000000, 55500.0000000}	12.948	-12.948			18.311
16	ESTIMATEDSALARY-4/10-{55500.0000000, 69000.0000000}	16.233	-16.233			22.956
17	ESTIMATEDSALARY-5/10-{69000.0000000, 82500.0000000}	14.958	-14.958			21.153
18	ESTIMATEDSALARY-6/10-{82500.0000000, 96000.0000000}	-1.270	1.270			1.796
19	ESTIMATEDSALARY-7/10-{96000.0000000, 109500.0000000}	-10.565	10.565			14.941
20	ESTIMATEDSALARY-8/10-{109500.0000000, 123000.0000000}	-5.850	5.850			8.273
21	ESTIMATEDSALARY-9/10-{123000.0000000, 136500.0000000}	-8.280	8.280			11.710
22	ESTIMATEDSALARY-10/10-{136500.0000000, 150000.0000000}	-13.135	13.135			18.576
	Сумма	0.000	0.000			
	Среднее	0.000				

Рисунок 7. Фрагмент матрицы разности между фактическими и ожидаемыми абсолютными частотами

Стоит отметить, что в АСК-анализе и СКМ 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 Ризбергена, 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 модели в соответствии с этими мерами достоверности.

3.4. Обобщ.форма по достов.моделей при разн.инт.крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Ван Ризбергена	Средняя модаль. уровень истинно-полож. решений (S1P)	Средняя модаль. уровень истинно-отриц. решений (S1N)	Средняя модаль. уровень ложно-полож. решений
1. ABS - частный критерий: количество встреч сочетаний "клас...	Корреляция абс частот с обр...	400	400		400		0.500	1.000	0.667	265.570		168
1. ABS - частный критерий: количество встреч сочетаний "клас...	Сумма абс частот по признак...	400	400		400		0.500	1.000	0.667	266.854		181
2. PR1C1 - частный критерий: усл. вероятность и-го признака сред...	Корреляция усл.отн частот с о...	400	400		400		0.500	1.000	0.667	265.570		168
2. PR1C1 - частный критерий: усл. вероятность и-го признака сред...	Сумма усл.отн частот по призна...	400	400		400		0.500	1.000	0.667	313.835		253
3. PR1C2 - частный критерий: условная вероятность и-го признака...	Корреляция усл.отн частот с о...	400	400		400		0.500	1.000	0.667	265.570		168
3. PR1C2 - частный критерий: условная вероятность и-го признака...	Сумма усл.отн частот по призна...	400	400		400		0.500	1.000	0.667	313.835		253
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	400	364	316	84	36	0.813	0.910	0.858	144.610	118.594	16
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	400	337	351	49	63	0.873	0.843	0.858	44.362	112.768	3
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	400	364	316	84	36	0.813	0.910	0.858	144.610	118.594	16
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	400	337	351	49	63	0.873	0.843	0.858	44.362	112.769	3
6. INF3 - частный критерий: Хинкватер, разности между фактич...	Семантический резонанс зна...	400	339	339	61	61	0.848	0.848	0.848	187.901	187.901	16
6. INF3 - частный критерий: Хинкватер, разности между фактич...	Сумма знаний	400	339	339	61	61	0.848	0.848	0.848	187.901	187.901	16
7. INF4 - частный критерий: ROI (Return On Investment), вероят...	Семантический резонанс зна...	400	341	360	40	59	0.895	0.853	0.873	177.234	193.087	11
7. INF4 - частный критерий: ROI (Return On Investment), вероят...	Сумма знаний	400	353	336	64	47	0.847	0.883	0.864	101.780	86.161	10
8. INF5 - частный критерий: ROI (Return On Investment), вероят...	Семантический резонанс зна...	400	341	360	40	59	0.895	0.853	0.873	177.234	193.087	11
8. INF5 - частный критерий: ROI (Return On Investment), вероят...	Сумма знаний	400	353	336	64	47	0.847	0.883	0.864	101.780	86.161	10
9. INF6 - частный критерий: разн усл и безусловностей; вер...	Семантический резонанс зна...	400	334	337	63	66	0.841	0.835	0.838	165.600	180.728	15
9. INF6 - частный критерий: разн усл и безусловностей; вер...	Сумма знаний	400	339	321	79	61	0.811	0.848	0.829	133.697	150.248	15
10. INF7 - частный критерий: разн усл и безусловностей; вер...	Семантический резонанс зна...	400	334	337	63	66	0.841	0.835	0.838	165.600	180.728	15
10. INF7 - частный критерий: разн усл и безусловностей; вер...	Сумма знаний	400	339	321	79	61	0.811	0.848	0.829	133.697	150.248	15

Рисунок 8. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена

3.4. Обобщ.форма по достов.моделей при разн.инт.крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Средняя модаль. уровень истинно-отриц. решений (S1N)	Средняя модаль. уровень ложно-полож. решений (S1P)	S-T точность модели	S-Полнота модели	L1-мера проф. Е.В.Луценко	Средний модаль. уровень истинно-отриц. решений	Средний модаль. уровень истинно-отриц. решений	Средний модаль. уровень ложно-полож. решений	Средний модаль. уровень ложно-отриц. решений	A-То
1. ABS - частный критерий: количество встреч сочетаний "клас...	Корреляция абс частот с обр...		168.355	0.612	1.000	0.759	0.664		0.421		
1. ABS - частный критерий: количество встреч сочетаний "клас...	Сумма абс частот по признак...		181.431	0.595	1.000	0.746	0.667		0.454		
2. PR1C1 - частный критерий: усл. вероятность и-го признака сред...	Корреляция усл.отн частот с о...		168.355	0.612	1.000	0.759	0.664		0.421		
2. PR1C1 - частный критерий: усл. вероятность и-го признака сред...	Сумма усл.отн частот по призна...		253.749	0.553	1.000	0.712	0.785		0.634		
3. PR1C2 - частный критерий: условная вероятность и-го признака...	Корреляция усл.отн частот с о...		168.355	0.612	1.000	0.759	0.664		0.421		
3. PR1C2 - частный критерий: условная вероятность и-го признака...	Сумма усл.отн частот по призна...		253.749	0.553	1.000	0.712	0.785		0.634		
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	118.594	16.013	7.378	0.900	0.951	0.925	0.397	0.375	0.191	0.205
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	112.768	6.843	9.350	0.920	0.826	0.871	0.132	0.321	0.078	0.148
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	118.594	16.013	7.378	0.900	0.951	0.925	0.397	0.375	0.191	0.205
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	112.769	6.843	9.350	0.920	0.826	0.871	0.132	0.321	0.078	0.148
6. INF3 - частный критерий: Хинкватер, разности между фактич...	Семантический резонанс зна...	187.901	16.850	16.850	0.918	0.918	0.918	0.554	0.554	0.276	0.276
6. INF3 - частный критерий: Хинкватер, разности между фактич...	Сумма знаний	187.901	16.850	16.850	0.918	0.918	0.918	0.554	0.554	0.276	0.276
7. INF4 - частный критерий: ROI (Return On Investment), вероят...	Семантический резонанс зна...	193.087	11.035	16.377	0.941	0.915	0.928	0.520	0.553	0.276	0.278
7. INF4 - частный критерий: ROI (Return On Investment), вероят...	Сумма знаний	86.161	10.339	6.656	0.908	0.939	0.923	0.288	0.256	0.162	0.142
8. INF5 - частный критерий: ROI (Return On Investment), вероят...	Семантический резонанс зна...	193.087	11.035	16.377	0.941	0.915	0.928	0.520	0.553	0.276	0.278
8. INF5 - частный критерий: ROI (Return On Investment), вероят...	Сумма знаний	86.161	10.339	6.656	0.908	0.939	0.923	0.288	0.256	0.162	0.142
9. INF6 - частный критерий: разн усл и безусловностей; вер...	Семантический резонанс зна...	180.728	15.166	19.704	0.916	0.894	0.905	0.496	0.536	0.241	0.299
9. INF6 - частный критерий: разн усл и безусловностей; вер...	Сумма знаний	150.248	15.696	14.317	0.895	0.903	0.899	0.394	0.468	0.199	0.235
10. INF7 - частный критерий: разн усл и безусловностей; вер...	Семантический резонанс зна...	180.728	15.166	19.704	0.916	0.894	0.905	0.496	0.536	0.241	0.299
10. INF7 - частный критерий: разн усл и безусловностей; вер...	Сумма знаний	150.248	15.696	14.317	0.895	0.903	0.899	0.394	0.468	0.199	0.235

Рисунок 9. Экранная форма с информацией о достоверности моделей по L1-критерию проф. Е. В. Луценко

3.4. Обобщ.форма по достов.моделей при разн.инт.крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Средний модаль. уровень истинно-отриц. решений	A-Точность модели AРrecision = AТР/AТР+	A-Полнота модели AРesall = AТР/AТР+	L2-мера проф. Е.В.Луценко	Процент правильной идентификац...	Процент правильной не идентификац...	Процент ошибочной идентификац...	Процент ошибочной не идентификац...	Процент правильных результатов	Дата получения результата	Время получен. резуль
1. ABS - частный критерий: количество встреч сочетаний "клас...	Корреляция абс частот с обр...		0.612	1.000	0.759	100.000		100.000		50.000	04.06.2021	01.00
1. ABS - частный критерий: количество встреч сочетаний "клас...	Сумма абс частот по признак...		0.595	1.000	0.746	100.000		100.000		50.000	04.06.2021	01.00
2. PR1C1 - частный критерий: усл. вероятность и-го признака сред...	Корреляция усл.отн частот с о...		0.612	1.000	0.759	100.000		100.000		50.000	04.06.2021	01.00
2. PR1C1 - частный критерий: усл. вероятность и-го признака сред...	Сумма усл.отн частот по призна...		0.553	1.000	0.712	100.000		100.000		50.000	04.06.2021	01.00
3. PR1C2 - частный критерий: условная вероятность и-го признака...	Корреляция усл.отн частот с о...		0.612	1.000	0.759	100.000		100.000		50.000	04.06.2021	01.00
3. PR1C2 - частный критерий: условная вероятность и-го признака...	Сумма усл.отн частот по призна...		0.553	1.000	0.712	100.000		100.000		50.000	04.06.2021	01.00
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	0.205	0.676	0.660	0.668	91.000	69.393	30.607	9.000	80.197	04.06.2021	01.00
4. INF1 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	0.148	0.627	0.470	0.537	84.250	89.151	10.849	15.750	86.701	04.06.2021	01.00
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Семантический резонанс зна...	0.205	0.676	0.660	0.668	91.000	69.393	30.607	9.000	80.197	04.06.2021	01.00
5. INF2 - частный критерий: количество знаний по АХаркевичу; в...	Сумма знаний	0.148	0.627	0.470	0.537	84.250	89.151	10.849	15.750	86.701	04.06.2021	01.00
6. INF3 - частный критерий: Хинкватер, разности между фактич...	Семантический резонанс зна...	0.276	0.667	0.667	0.667	84.750	80.037	19.963	15.250	82.394	04.06.2021	01.00
6. INF3 - частный критерий: Хинкватер, разности между фактич...	Сумма знаний	0.276	0.667	0.667	0.667	84.750	80.037	19.963	15.250	82.394	04.06.2021	01.00
7. INF4 - частный критерий: ROI (Return On Investment), вероят...	Семантический резонанс зна...	0.278	0.653	0.652	0.653	85.250	84.820	15.180	14.750	85.035	04.06.2021	01.00
7. INF4 - частный критерий: ROI (Return On Investment), вероят...	Сумма знаний	0.142	0.641	0.671	0.655	88.250	87.065	12.935	11.750	87.657	04.06.2021	01.00
8. INF5 - частный критерий: ROI (Return On Investment), вероят...	Семантический резонанс зна...	0.278	0.653	0.652	0.653	85.250	84.820	15.180	14.750	85.035	04.06.2021	01.00
8. INF5 - частный критерий: ROI (Return On Investment), вероят...	Сумма знаний	0.142	0.641	0.671	0.655	88.250	87.065	12.935	11.750	87.657	04.06.2021	01.00
9. INF6 - частный критерий: разн усл и безусловностей; вер...	Семантический резонанс зна...	0.299	0.673	0.624	0.648	83.500	79.759	20.241	16.500	81.630	04.06.2021	01.00
9. INF6 - частный критерий: разн усл и безусловностей; вер...	Сумма знаний	0.235	0.665	0.627	0.645	84.750	77.533	22.467	15.250	81.142	04.06.2021	01.00
10. INF7 - частный критерий: разн усл и безусловностей; вер...	Семантический резонанс зна...	0.299	0.673	0.624	0.648	83.500	79.759	20.241	16.500	81.630	04.06.2021	01.00
10. INF7 - частный критерий: разн усл и безусловностей; вер...	Сумма знаний	0.235	0.665	0.627	0.645	84.750	77.533	22.467	15.250	81.142	04.06.2021	01.00

Рисунок 10. Экранная форма с информацией о достоверности моделей по L2-критерию проф. Е. В. Луценко

Данная форма достоверностей позволяет сделать вывод о высокой достоверности модели 'INF4' с интегральным критерием «Семантический резонанс знаний», F-мерой, равной 0,873, L1-мерой, равной 0,928 и L2-мерой, равной 0,653. Полученные степени достоверности позволяют сделать вывод о наличии в СКМ «INF4» сильной корреляции между выбранными ранее факторами и фактом покупки товара.

На рисунке 11 приведено частотное распределения числа истинных и ложных положительных и отрицательных решений по результатам идентификации классов в СК-модели INF4 по данным обучающей выборки:

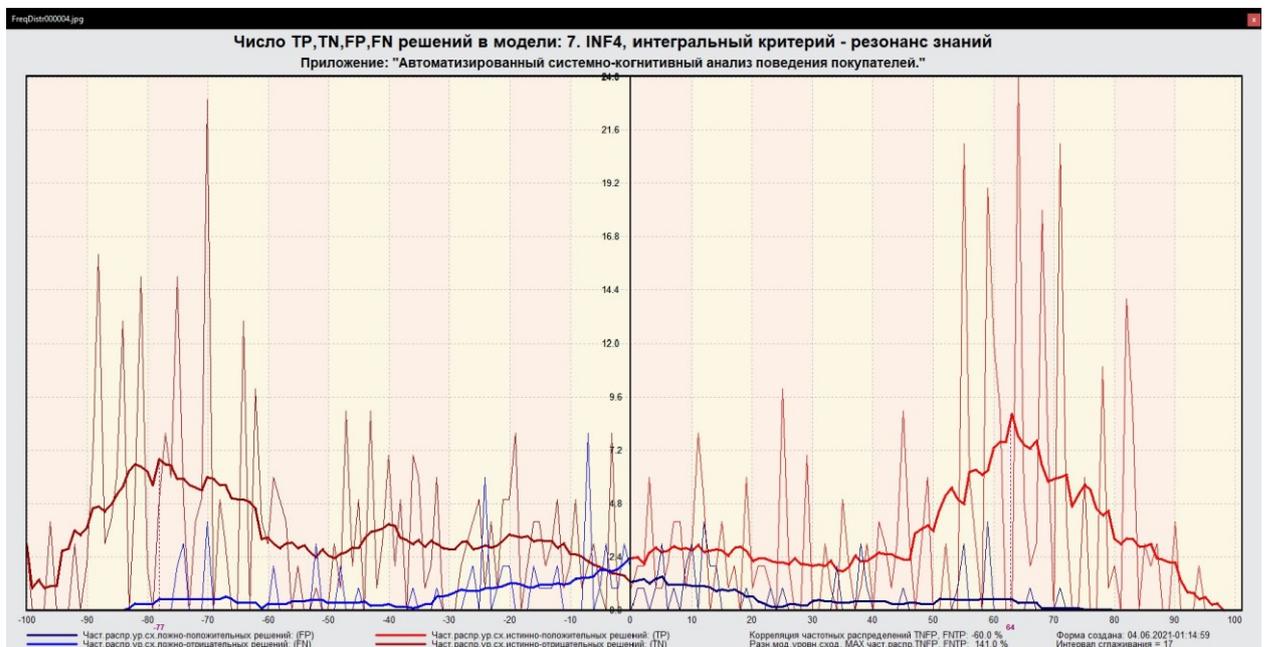


Рисунок 11. Частотные распределения числа истинных и ложных положительных и отрицательных решений в СКМ

Анализ полученного графика распределений частот показал, что решения распределяются симметрично уровню сходства примерно -5% и что при уровне сходства ~68% количество истинно-ложных решений становится незначительным.

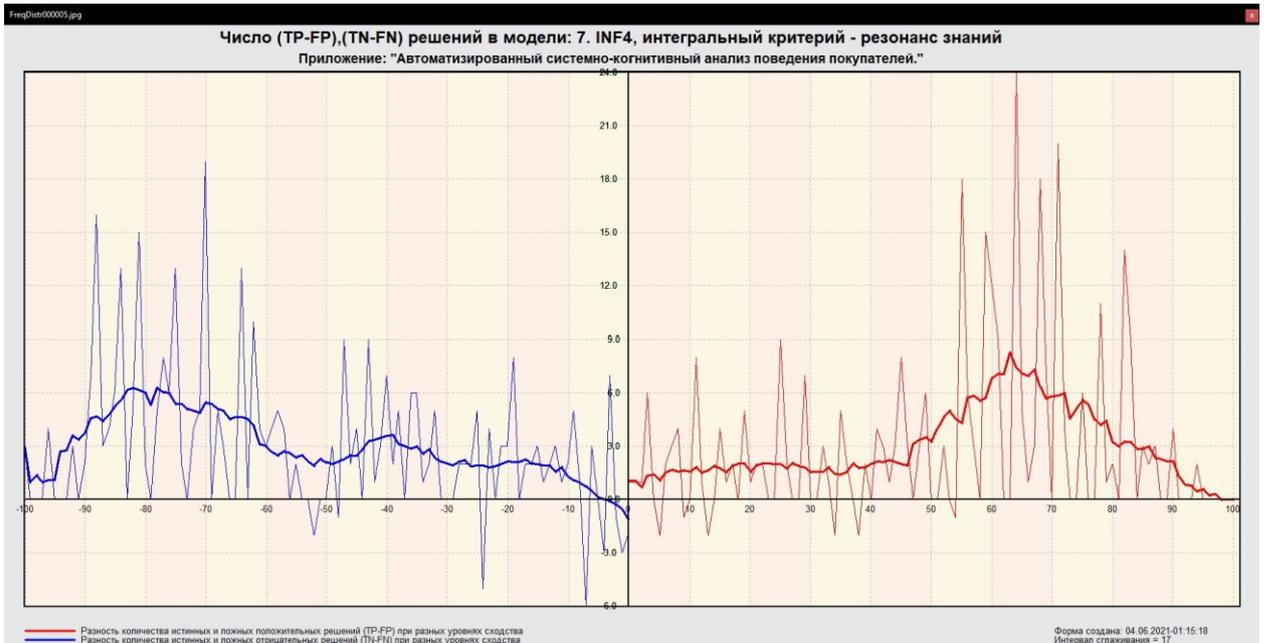


Рисунок 12. Частотное распределение разностей истинных и ложных решений

Анализируя же только истинные решения, можно заметить, что на всём отрезке положительного сходства истинные решения преобладают над количеством ложноположительных, начиная с частоты 1% и достигая максимума в 8% при уровне сходства 64%. Аналогично, количество истинно ложных решений начинает превышать количество ложных такого же типа только при уровне различия -4% и достигает своего максимума при уровне -80%.

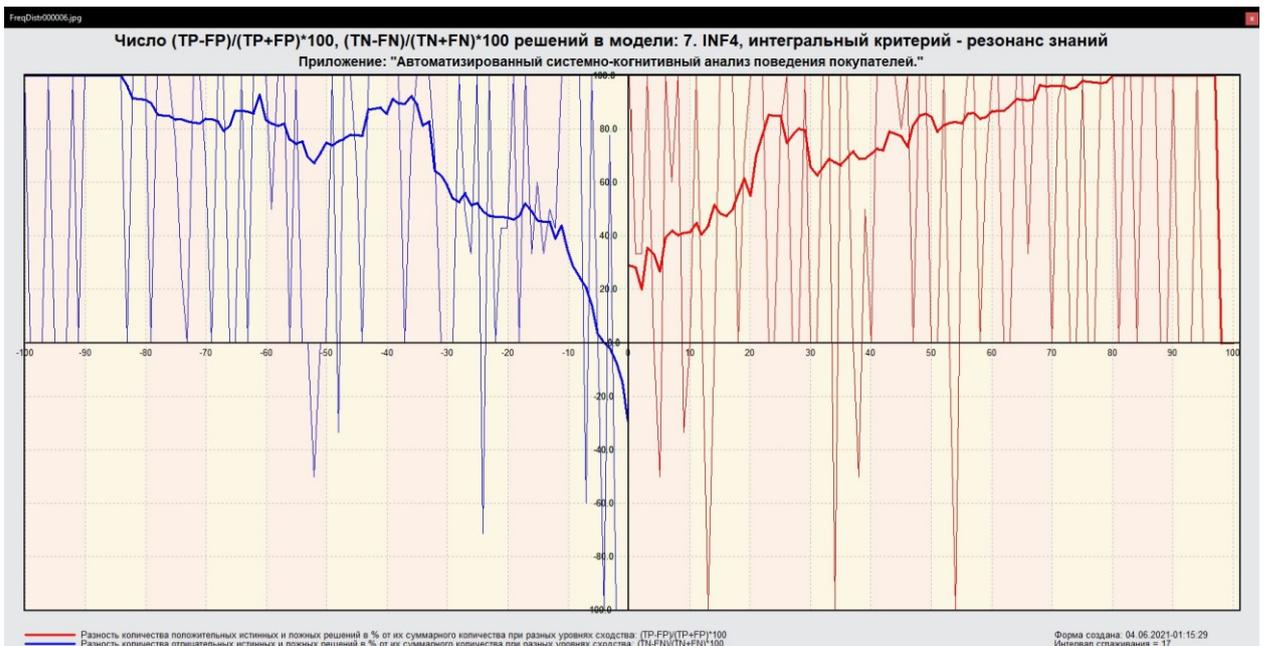


Рисунок 13. Частотное распределение доли разности истинных и ложных решений от общего количества решений

На этом графике видно, что доля истинных среди общего их количества достигает 100% при уровнях сходства, равных -82% для отрицательных и 80% - для положительных решений.

На рисунке 14 приведен вид справочного интерфейса по режиму 3.4, в котором описаны меры достоверности моделей, применяемые в системе «Эйдос»:

Помощь по режимам: 3.4, 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++".

ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Предположим, модель дает такой прогноз, что выпадет все: и 1, и 2, и 3, и 4, и 5, и 6. Понятно, что из всего этого выпадет лишь что-то одно. В этом случае модель не предскажет, что не выпадет, но зато она обязательно предскажет, что выпадет. Однако при этом очень много объектов будет отнесено к классам, к которым они не относятся. Тогда вероятность истинно-положительных решений у модели будет 1/6, а вероятность ложно-положительных решений - 5/6. Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.

ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Представим себе, что мы выбрасываем кубик с 6 гранями, и модель предсказывает, что ничего не выпадет, т.е. не выпадет ни 1, ни 2, ни 3, ни 4, ни 5, ни 6, но что-то из этого, естественно, обязательно выпадет. Конечно, модель не предсказала, что выпадет, зато она очень хорошо предсказала, что не выпадет. Вероятность истинно-отрицательных решений у модели будет 5/6, а вероятность ложно-отрицательных решений - 1/6. Такой прогноз гораздо достовернее, чем положительный псевдопрогноз, но тоже бесполезен.

ИДЕАЛЬНЫЙ ПРОГНОЗ.
Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.

РЕАЛЬНЫЙ ПРОГНОЗ.
На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представьте себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В.Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {-1, +1} и {0, 1}:

$$L_a = \frac{TP + TN - FP - FN}{TP + TN + FP + FN} \quad (\text{нормировка: } \{-1, +1\})$$

$$L_b = \frac{1 + (TP + TN - FP - FN)}{(TP + TN + FP + FN)} / 2 \quad (\text{нормировка: } \{0, 1\})$$

где количество: TP - истинно-положительных решений; TN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;

Классическая F-мера достоверности моделей Ван Ризбергена (колонка выделена ярко-голубым фоном):

$$F\text{-мера} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Recall = TP / (TP + FN) - полнота модели;

Precision = TP / (TP + FP) - точность модели;

L1-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СУММ уровней сходства (колонка выделена ярко-зеленым фоном):

$$L1\text{-мера} = 2 * (SPrecision * SRecall) / (SPrecision + SRecall)$$

SRecall = STP / (STP + SFN) - полнота с учетом сумм уровней сходства;

SFPrecision = STP / (STP + SFP) - точность с учетом сумм уровней сходства;

STP - Сумма модулей сходства истинно-положительных решений; SFN - Сумма модулей сходства ложно-отрицательных решений;

SFP - Сумма модулей сходства ложно-положительных решений; STN - Сумма модулей сходства истинно-отрицательных решений.

L2-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желтым фоном):

$$L2\text{-мера} = 2 * (APrecision * ARecall) / (APrecision + ARecall)$$

APrecision = ATP / (ATP + AFP) - точность с учетом средних уровней сходства;

ARecall = ATP / (ATP + AFN) - полнота с учетом средних уровней сходства;

ATP = STP / TP - Среднее модулей сходства истинно-положительных решений; AFN = SFN / FN - Среднее модулей сходства истинно-отрицательных решений;

AFP = SFP / FP - Среднее модулей сходства ложно-положительных решений; AFN = SFN / FN - Среднее модулей сходства ложно-отрицательных решений.

Строки с максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке.

Из графиков частотных распределений истинно-положительных, истинно-отрицательных, ложно-положительных и ложно-отрицательных решений видно, что чем выше модуль уровня сходства, тем больше доля истинных решений. Это значит, что модуль уровня сходства является адекватной мерой степени истинности решения и степени уверенности системы в этом решении. Поэтому система "Эйдос" имеет адекватный критерий достоверности собственных решений, с помощью которого она может отфильтровать заведомо ложные решения.

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.

Рисунок 14. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена и L1- и L2-критериям проф. Е. В. Луценко

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 INF4 статус текущей модели. Для этого запустим режим 5.6 с параметрами, приведенными на экранной форме (рисунок 11):

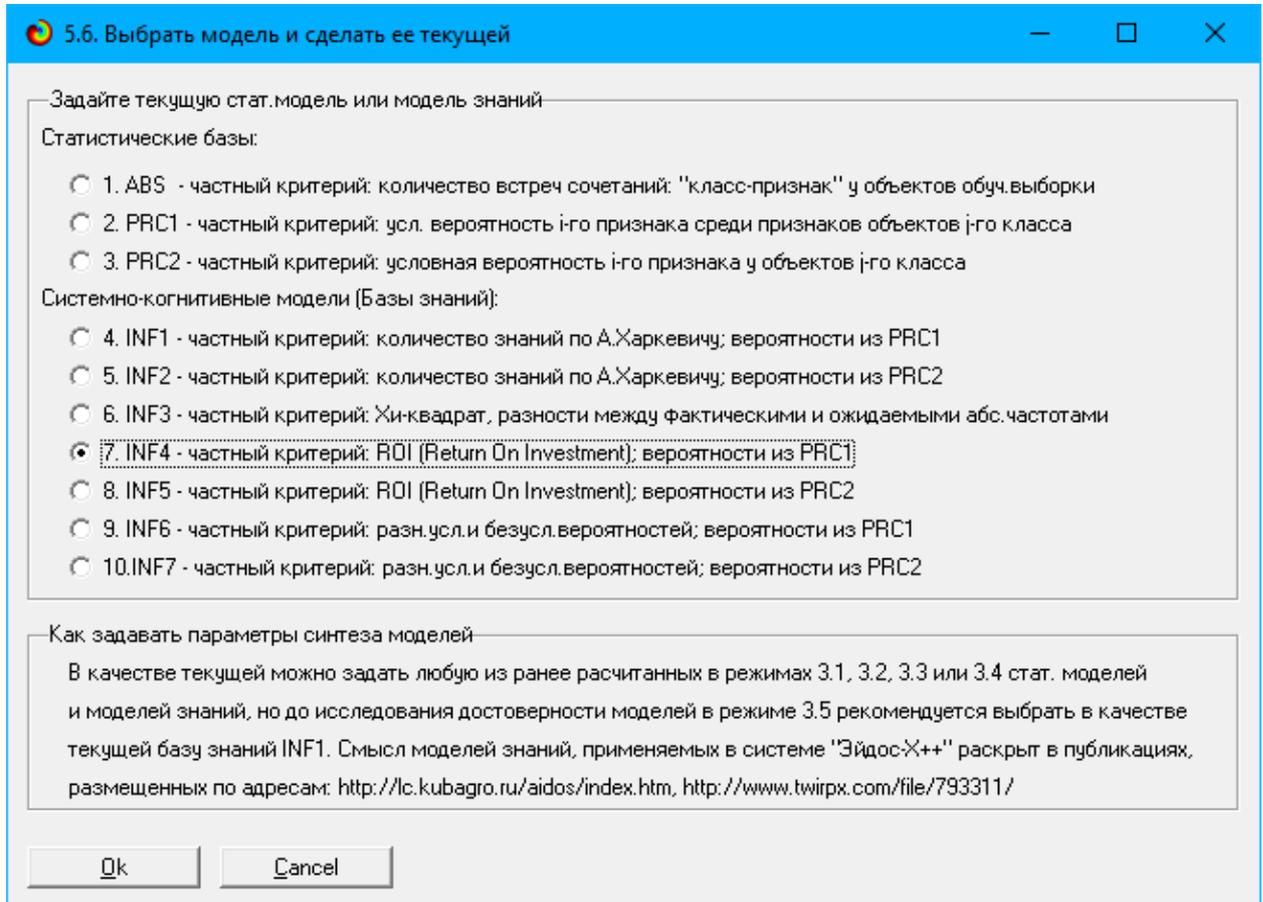


Рисунок 15. Экранная форма присвоения наиболее достоверной СКМ «Inf4» статуса текущей

Решение различных задач в наиболее достоверной модели Прогнозирование (диагностика, классификация, распознавание, идентификация)

Для решения задачи системной идентификации – определения класса на основе обучающей выборки – было проведено пакетное распознавание в наиболее достоверной СК-модели «INF4» на GPU.

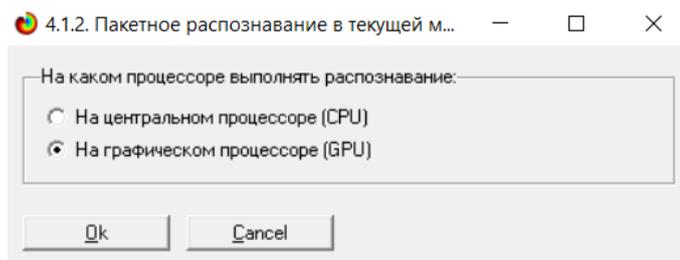


Рисунок 16. Пользовательский интерфейс выбора режима пакетного распознавания

В результате пакетного распознавания получены таблицы критериев сходства имеющихся классов. Формы с этими таблицами отражают результаты прогнозирования в различных разрезах и обобщениях и приведены на рисунках ниже.

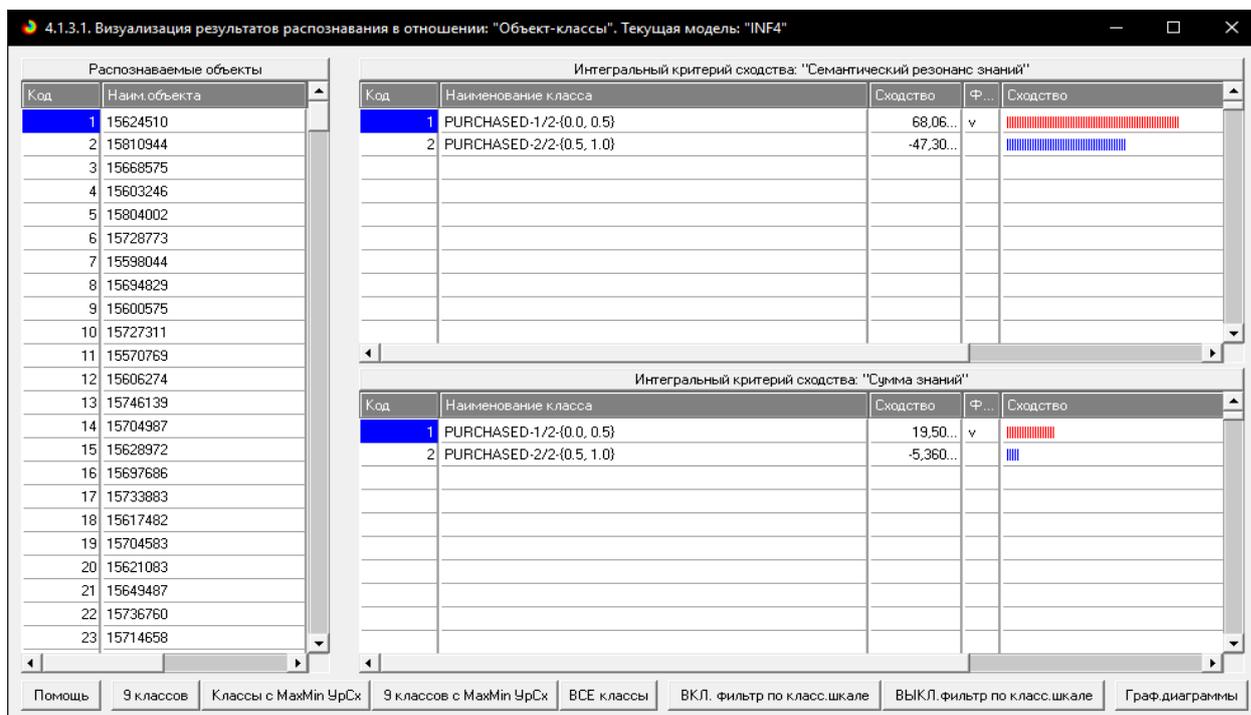


Рисунок 17. Результаты распознавания в отношениях «Объект-классы»

4.1.3.2. Визуализация результатов распознавания в отношении: "Класс-объекты", Текущая модель: "INF4"

Классы		Интегральный критерий сходства: "Семантический резонанс знаний"				
Код	Наим. класса	Код	Наименование объекта	Сходство	Ф...	Сходство
1	PURCHASED-1/2-(0.0000000, 0.5000000)	66	15718071	94,37...	v	████████████████████
2	PURCHASED-2/2-(0.5000000, 1.0000000)	136	15578006	94,37...	v	████████████████████
		63	15595135	90,47...	v	████████████████████
		69	15577178	90,47...	v	████████████████████
		134	15638963	90,47...	v	████████████████████
		142	15666141	90,47...	v	████████████████████
		77	15746737	87,28...	v	████████████████████
		83	15709476	87,28...	v	████████████████████
		87	15694946	87,28...	v	████████████████████
		3	15668575	83,38...	v	████████████████████

Интегральный критерий сходства: "Сумма знаний"				
Код	Наименование объекта	Сходство	Ф...	Сходство
66	15718071	32,19...	v	████████████████████
136	15578006	32,19...	v	████████████████████
63	15595135	30,31...	v	████████████████████
69	15577178	30,31...	v	████████████████████
134	15638963	30,31...	v	████████████████████
142	15666141	30,31...	v	████████████████████
77	15746737	28,77...	v	████████████████████
83	15709476	28,77...	v	████████████████████
87	15694946	28,77...	v	████████████████████
3	15668575	26,89...	v	████████████████████

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

Рисунок 18. Результаты распознавания в отношениях «Класс-объекты»

Символ «√» стоит против тех результатов идентификации, которые соответствуют фактическим значениям. В ходе анализа полученных значений идентификации был сделан вывод, что результаты идентификации с уровнем сходства ниже 27,35% можно игнорировать при дальнейшем анализе.

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.

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 19–20 приведены SWOT-диаграммы для классов, соответствующим совершённой и несовершённой покупке, наглядно отражающие силу и направление влияния различных значений исследуемых факторов.

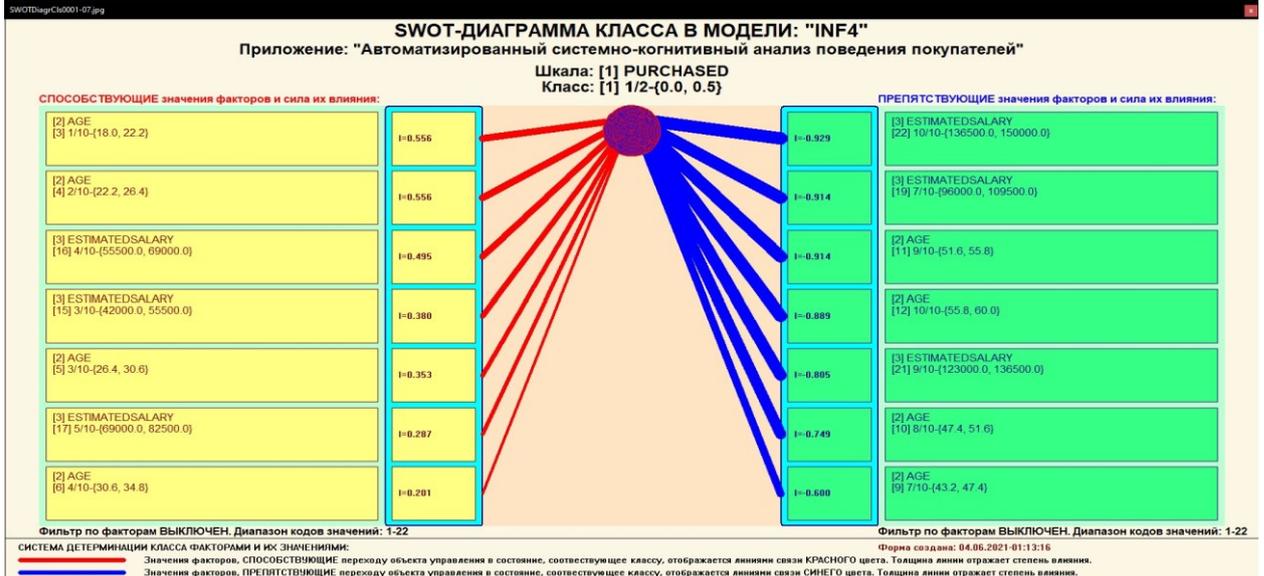


Рисунок 19. SWOT-диаграмма факторов, определяющих отсутствие покупки
 По этой диаграмме видно, что покупки реже всего совершали потребители 18–26 лет с заработной платой в 55000–82500 у. е.

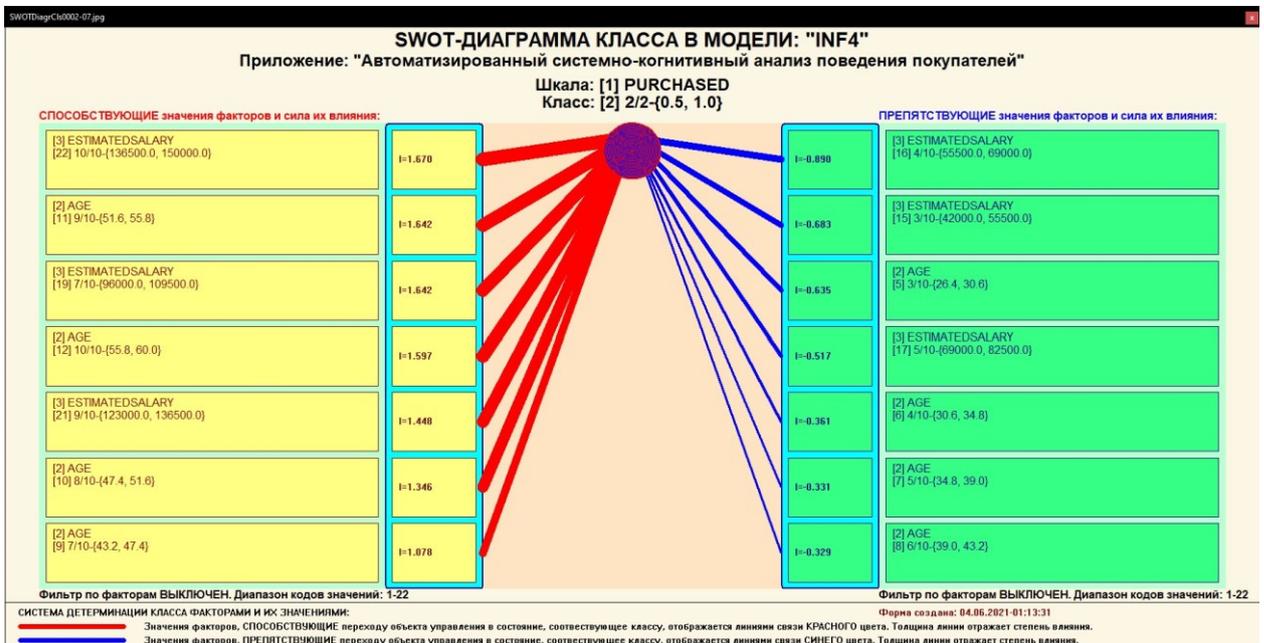


Рисунок 20. SWOT-диаграмма факторов, определяющих наличие покупки

Анализ же этой диаграммы показал, что совершение покупки наиболее характерно для потребителей 47–60 лет с заработной платой 96000–150000 у. е.

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.

Далее будут приведены результаты, полученные в ходе такого исследования – кластерно-конструктивный анализ классов и признаков, нелокальные нейроны, нелокальная нейронная сеть и интегральная когнитивная 3d-карта.

Когнитивные диаграммы классов

На когнитивной диаграмме, приведенной на рисунке 21, наглядно отражаются количественные оценки сходства и различия классов по связанным с ними значениям их характеристик. Важно, что эти результаты сравнения получены с применением системно-когнитивной модели, созданной непосредственно на основе эмпирических данных, а не как принято делать традиционно – на основе экспертных оценок не формализуемым путем на основе интуиции и профессиональной компетенции.

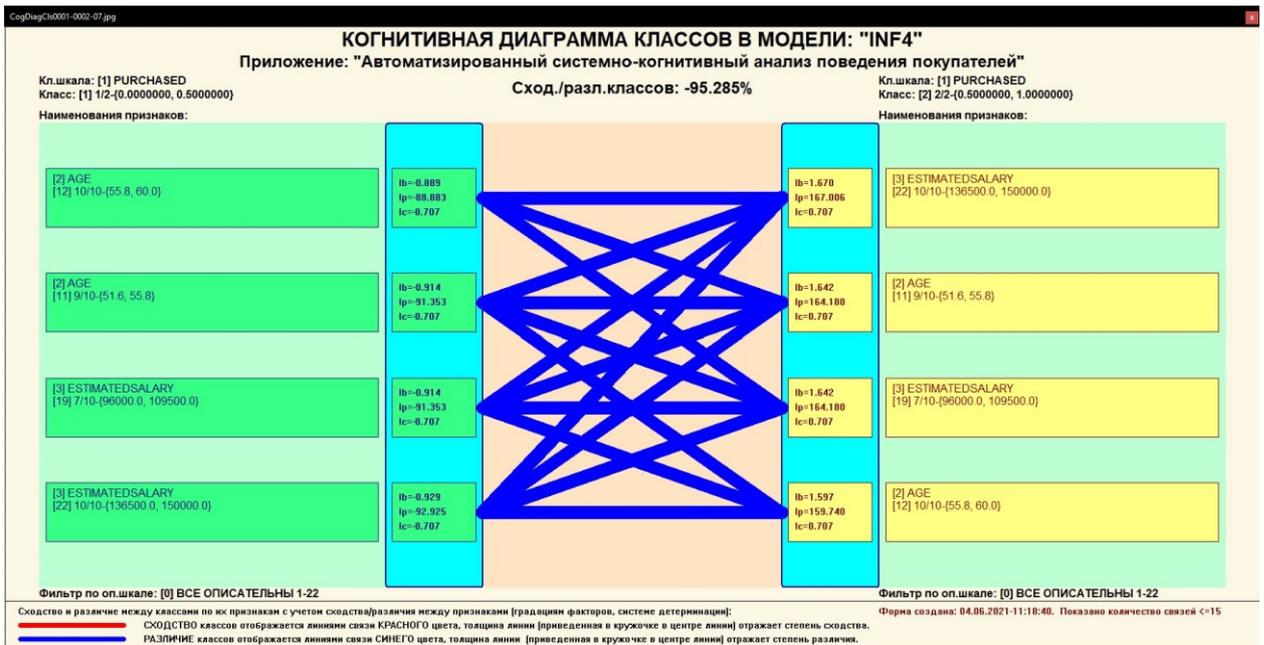


Рисунок 21. Когнитивная диаграмма, сравнивающая 2 исследуемых класса

Анализ полученных диаграмм показывает о разительном различии в характерности изучаемых факторов на принадлежность к тому или иному классу.

Нелокальные нейроны и нелокальные нейронные сети

Для определения способствующих и тормозящих факторов того или иного класса также были использованы нелокальные нейроны. Ниже приведены диаграммы таких структур для обоих классов модели.

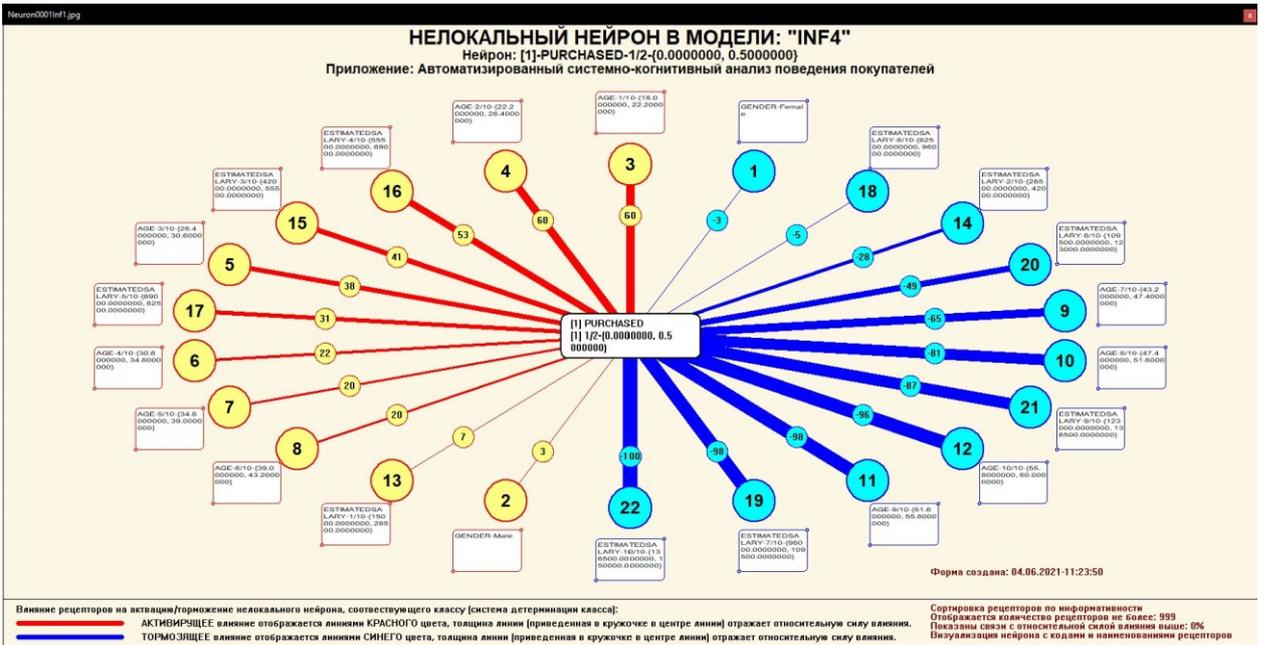


Рисунок 22. Нелокальный нейрон класса, соответствующего отсутствию покупки

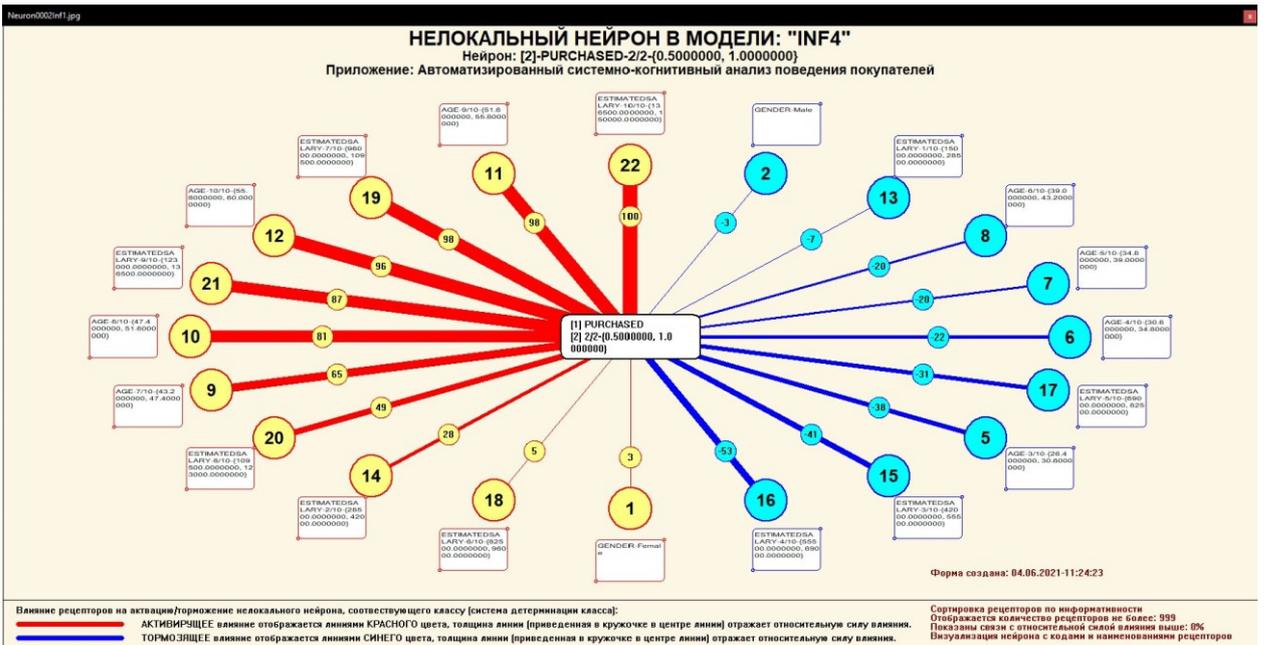


Рисунок 23. Нелокальный нейрон класса, соответствующего наличию покупки

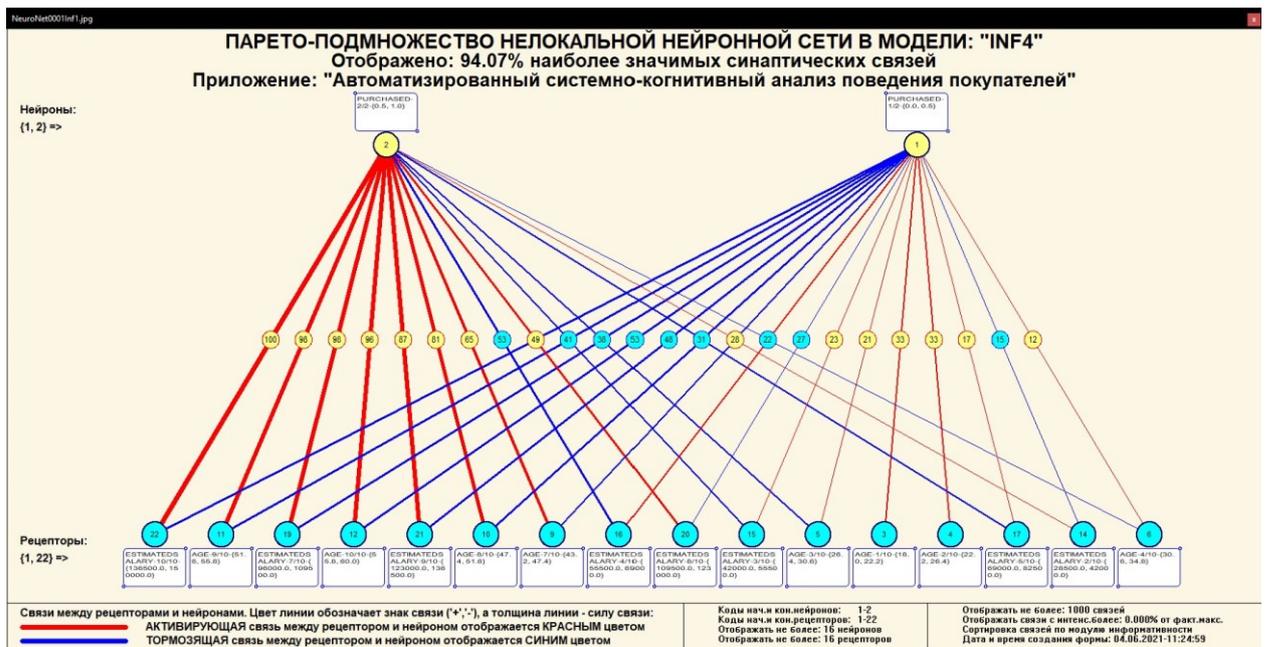


Рисунок 24. Один слой нелокальной нейронной сети, отражающий силу и направление влияния значений исследуемых факторов на факт совершения покупки

В приведенном фрагменте слоя нейронной сети нейроны соответствуют классам, а рецепторы – определяющим их факторам. Нейроны расположены слева направо в порядке убывания степени определенности.

Модель знаний системы «Эйдос» относится к *нечетким декларативным* гибридным моделям и объединяет в себе некоторые особенности нейросетевой и фреймовой моделей представления знаний. 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:

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 – или, иначе, «полносвязной».

3d-интегральные когнитивные карты

На рисунке 25 приведен фрагмент 3d-интегральной когнитивной карты, отражающий фрагмент около 95% СКМ «Inf4».

3d-интегральная когнитивная карта является отображением на одном рисунке когнитивных диаграмм классов и значений факторов.

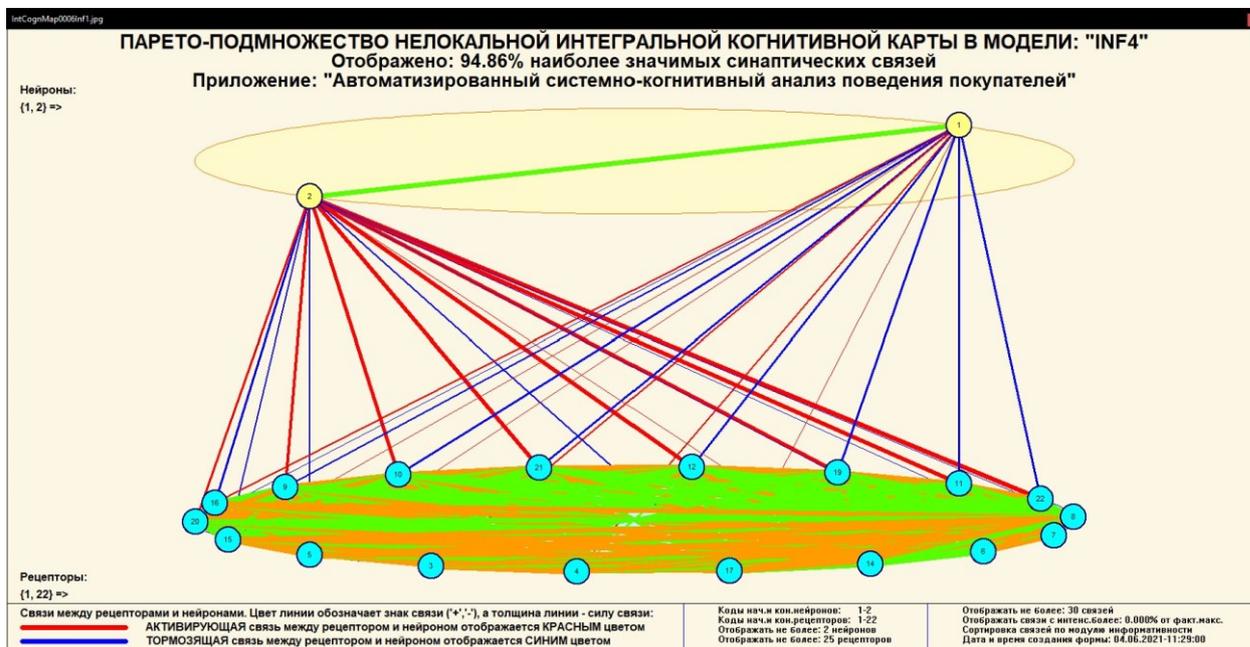


Рисунок 25. Интегральная когнитивная 3d-карта в СКМ «Inf4»

Сила и направление влияния значений исследуемых факторов на факт покупки

Строки матриц моделей соответствуют значениям факторов, то есть градациям описательных шкал.

Колонки матриц моделей соответствуют различным классам, отражающим градации классификационных шкал.

Числовые значения в ячейках матриц моделей, находящихся на пересечении строк и колонок, отражают направление и силу влияния конкретного значения фактора, соответствующего строке, на значение класса.

Из этого следует, что суммарную силу влияния того или иного значения характеристики покупки на его класс (т. е. ценность данного значения характеристики для решения задачи квалиметрии и других задач) можно количественно оценивать степенью вариабельности значений в строке матрицы модели, соответствующей этому значению характеристики класса.

Существует много мер вариативности значений: среднее модулей отклонения от среднего, дисперсия, среднеквадратичное отклонение и другие. В АСК-анализе и, в частности, в системе «Aidos-X» для этой цели используется среднеквадратичное отклонение.

Самая правая колонка в матрицах моделей содержит количественную оценку вариативности значений строки модели, которая и представляет собой ценность значения характеристики покупки, соответствующего строке, для решения задачи квалиметрии и других задач, рассмотренных в работе.

При рассмотрении матрицы модели по правой колонке в порядке убывания, а потом сложении значений в ней в нарастающий итог, была получена логистическая Парето-кривая, отражающая зависимость ценности модели от числа наиболее ценных признаков в ней.

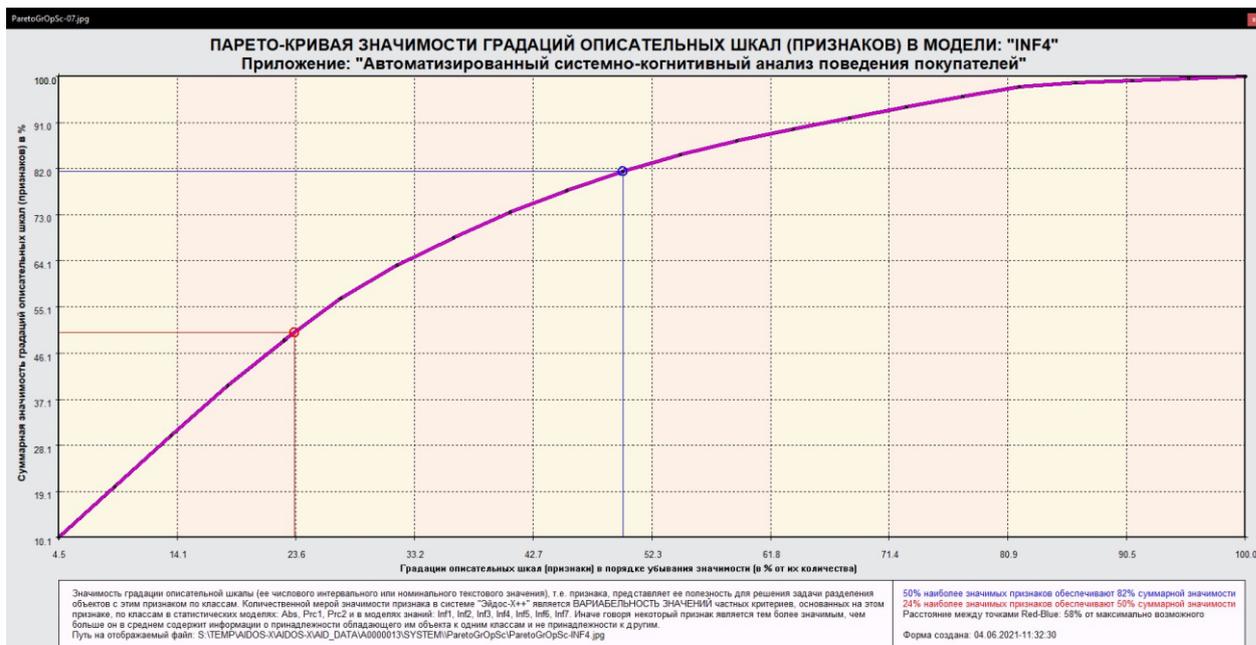


Рисунок 34. Парето-кривая значимости градаций описательных шкал

Таблица 6. Часть Парето-таблицы значимости градаций описательных шкал

№	Код атрибута	Наименование	Значимость, %	Значимость нараст. итогом, %
1	22	ESTIMATEDSALARY-10/10-{136500, 150000}	1,8379903	10,1333823
2	11	AGE-9/10-{51.6, 55.8}	1,8068946	20,0953248
3	19	ESTIMATEDSALARY-7/10-{96000, 109500}	1,8068946	30,0572673
4	12	AGE-10/10-{55.8, 60}	1,7580299	39,7498043
5	21	ESTIMATEDSALARY-9/10-{123000, 136500}	1,5931115	48,5330975
6	10	AGE-8/10-{47.4, 51.6}	1,4813926	56,7004515
7	9	AGE-7/10-{43.2, 47.4}	1,1863129	63,2409442
8	16	ESTIMATEDSALARY-4/10-{55500, 69000}	0,9798307	68,6430397
9	20	ESTIMATEDSALARY-8/10-{109500, 123000}	0,9004543	73,6075096
10	15	ESTIMATEDSALARY-3/10-{42000, 55500}	0,7520486	77,7537752
11	5	AGE-3/10-{26.4, 30.6}	0,6990149	81,6076504
12	17	ESTIMATEDSALARY-5/10-{69000, 82500}	0,5684730	84,7418096
13	14	ESTIMATEDSALARY-2/10-{28500., 42000}	0,5086483	87,5461376
14	6	AGE-4/10-{30.6, 34.8}	0,3969036	89,7343841
15	3	AGE-1/10-{18, 22.2}	0,3934485	91,9035817
16	4	AGE-2/10-{22.2, 26.4}	0,3934485	94,0727792
17	7	AGE-5/10-{34.8, 39}	0,3643979	96,0818122
18	8	AGE-6/10-{39, 43.2}	0,3617210	98,0760867
19	13	ESTIMATEDSALARY-1/10-{15000, 28500}	0,1347588	98,8190517
20	18	ESTIMATEDSALARY-6/10-{82500, 96000}	0,0888560	99,3089410
21	2	GENDER-Male	0,0639255	99,6613812
22	1	GENDER-Female	0,0614186	100,0000000

Из таблицы 6 видно, что факторы пола покупателей не оказывают практически никакого влияния на то, будет ли совершена покупка, в отличие от факторов возраста и заработной платы.

Также заметно, что наиболее сильное влияние оказывают значения характеристик:

1. ESTIMATEDSALARY-10/10- {136500, 150000}
2. AGE-9/10- {51.6, 55.8}
3. ESTIMATEDSALARY-7/10- {96000, 109500}
4. AGE-10/10- {55.8, 60}
5. ESTIMATEDSALARY-9/10- {123000, 136500}

а наиболее низкое:

1. AGE-6/10- {39, 43.2}
2. ESTIMATEDSALARY-1/10- {15000, 28500}
3. ESTIMATEDSALARY-6/10- {82500, 96000}
4. GENDER-Male
5. GENDER-Female

Ценность же всей описательной шкалы или фактора, для решения этих задач можно количественно оценивать как среднее от ценности значений этого параметра.

Таблица 7. Парето-таблица значимости описательных шкал

№	Код шкалы	Наименование	Количество признаков	Значимость, %	Значимость нараст. итогом, %
1	3	ESTIMATEDSALARY	22	49,2027110	49,2027110
2	2	AGE	12	47,4349348	96,6376458
3	1	GENDER	2	3,3623542	100,0000000

Из таблицы 7 видно, что наиболее сильное влияние на факт покупки оказывают в приблизительно равной степени заработная плата и возраст покупателя и незначительно низкой – его пол.

Степень детерминированности классов

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

В данной работе классы определяют наличие или отсутствие покупки тем или иным клиентом, а значениями градаций описательных шкал – возраст, пол и заработная плата покупателя.

На рисунке 27 изображена Парето-кривая степени детерминированности классов значениями характеристик нарастающим ИТОГОМ.

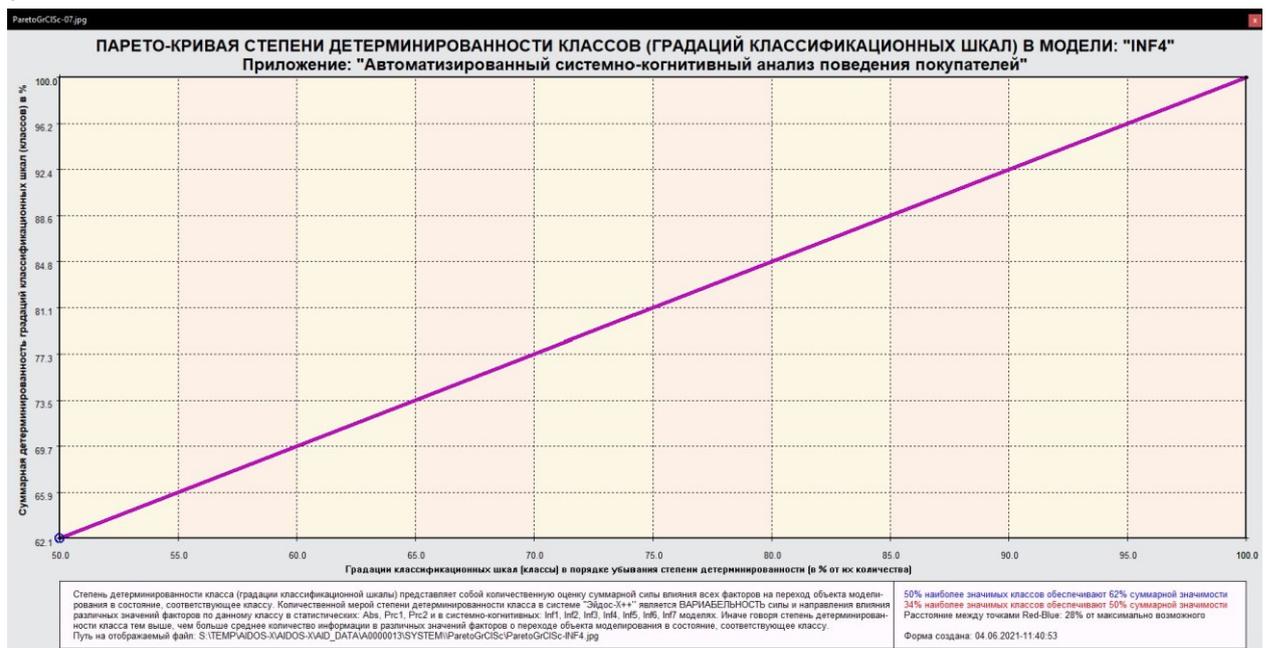


Рисунок 27. Парето-кривая степени детерминированности классов

Таблица 8. Парето-таблица степеней детерминированности классов

№	Код шкалы	Наименование	Значимость, %	Значимость нараст. итогом, %
1	2	PURCHASED-2/2-{0.5000000, 1.0000000}	62,1128419	62,1128419
2	1	PURCHASED-1/2-{0.0000000, 0.5000000}	37,8871581	100,0000000

Данные таблицы 8 позволяют сделать вывод о высокой детерминированности класса, соответствующего совершённой покупке и о заметно более слабой – у класса, соответствующего несовершённой покупке.

При этом степень детерминированности этих классов отличается в почти два раза что довольно существенно.

Вывод

Анализ выбранной предметной области «Поведение покупателей» с помощью инструментальных средств программной среды «Aidos-X» показал высокую зависимость факта покупки товара потребителем от его покупательской способности, определяемой заработной платой покупателя, и его возрастом.

Список литературы

1. Лойко В.И. Подходы к автоматизации процессов управления производством продукции растениеводства / В.И. Лойко, С.А. Курносков, В.В. Ткаченко, Н.А. Ткаченко // Экономико-правовые аспекты реализации стратегии модернизации России: поиск модели эффективного социохозяйственного развития: сб. стат. междунар. науч.-практ. конф., Сочи, 5-9 октября 2016 г. – М.: НИИ ЭИП2016. С. 128–132.
2. 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>
3. Луценко Е. В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с. <http://elibrary.ru/item.asp?id=18632909>
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. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системнокогнитивном 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.
8. 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-X++» / 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.
9. 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

10. 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.

11. Луценко Е. В. Системно-когнитивное моделирование влияния агротехнологий на урожайность и качество пшеницы и решение задач прогнозирования, поддержки принятия решений и исследования предметной области / Е. В. Луценко, Е.К. Печурина // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2019. – №03(147). С. 62–128. – IDA [article ID]: 1471903015. – Режим доступа: <http://ej.kubagro.ru/2019/03/pdf/15.pdf> 4,188 у.п.л.

12. 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.

13. 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>

14. Customer Behaviour. Classification dataset – [Электронный ресурс] URL: <https://www.kaggle.com/denisadutca/customer-behaviour>