

УДК 004.8

Прогноз бинарной классификации для типа рака молочной железы

Шулика Виолетта Валерьевна
студент факультета ПИ, группы ИТ2103
violettashulika@yandex.ru

Щербак Евгения Юрьевна
студент факультета ПИ, группы ИТ2103
Shcherbak.evgesha1@gmail.com

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

Целью данной работы является изучение характеристик бинарной классификации для типа рака молочной железы. Достижение данной цели представляет большой личный интерес. Для нас это позволит получить знания в работе с универсальной когнитивной аналитической системой «Эйдос-X++», а также получить зачет. Для достижения поставленной цели применяется Автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальная система «Эйдос».

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

Binary Classification Prediction for type of Breast Cancer

Shulika Violetta Valerievna
student of the faculty of PI, group IT2103
violettashulika@yandex.ru

Shcherbak Evgenia Yuryevna
student of the faculty of PI, group IT2103
Shcherbak.evgesha1@gmail.com

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

The purpose of this work is to study the characteristics of the binary classification for the type of breast cancer. The achievement of this goal is of great personal interest. For us, this will allow us to gain knowledge in working with the universal analytical system "Eidos-X ++", as well as get a credit. To achieve this goal, the Automated System-Cognitive Analysis (ASK-analysis) and its software tools are used - the intelligent system "Eidos".

Key words: AUTOMATED SYSTEM-COGNITIVE ANALYSIS, ASC-ANALYSIS, "EIDOS" SYSTEM.

СОДЕРЖАНИЕ

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	11
ЗАДАЧА 4: РЕШЕНИЕ РАЗЛИЧНЫХ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	19
Подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация)	19
Подзадача 4.2. Поддержка принятия решений (SWOT-анализ)	21
Подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели	24
4.3.1. Когнитивные диаграммы классов	25
4.3.2. Агломеративная когнитивная кластеризация классов	26
4.3.3. Когнитивные диаграммы значений факторов	27
4.3.4. Агломеративная когнитивная кластеризация значений факторов	28
4.3.5. Нелокальные нейроны и нелокальные нейронные сети	28
4.3.6. 3d-интегральные когнитивные карты	29
4.3.7. Когнитивные функции	29
4.3.8. Сила и направление влияния значений характеристик болезни и влияния этих характеристик	31
4.3.9. Степень детерминированности класса (вида) заболеваемости	32
7. ВЫВОДЫ	33
СПИСОК ЛИТЕРАТУРЫ	33

Введение

Целью данной работы является изучение характеристик бинарной классификации для типа рака молочной железы.

Достижение данной цели представляет большой личный интерес. Для нас это позволит получить знания в работе с универсальной когнитивной аналитической системой «Эйдос-X++», а также получить зачет.

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).

Эти задачи, по сути, представляют собой **этапы** автоматизированного системно-когнитивный анализа (АСК-анализ), который 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>), 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 307, соответственно) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);

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

- обеспечивает мультязычную поддержку интерфейса на 51 языке. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;

- наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решения этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA);

- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе: 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;

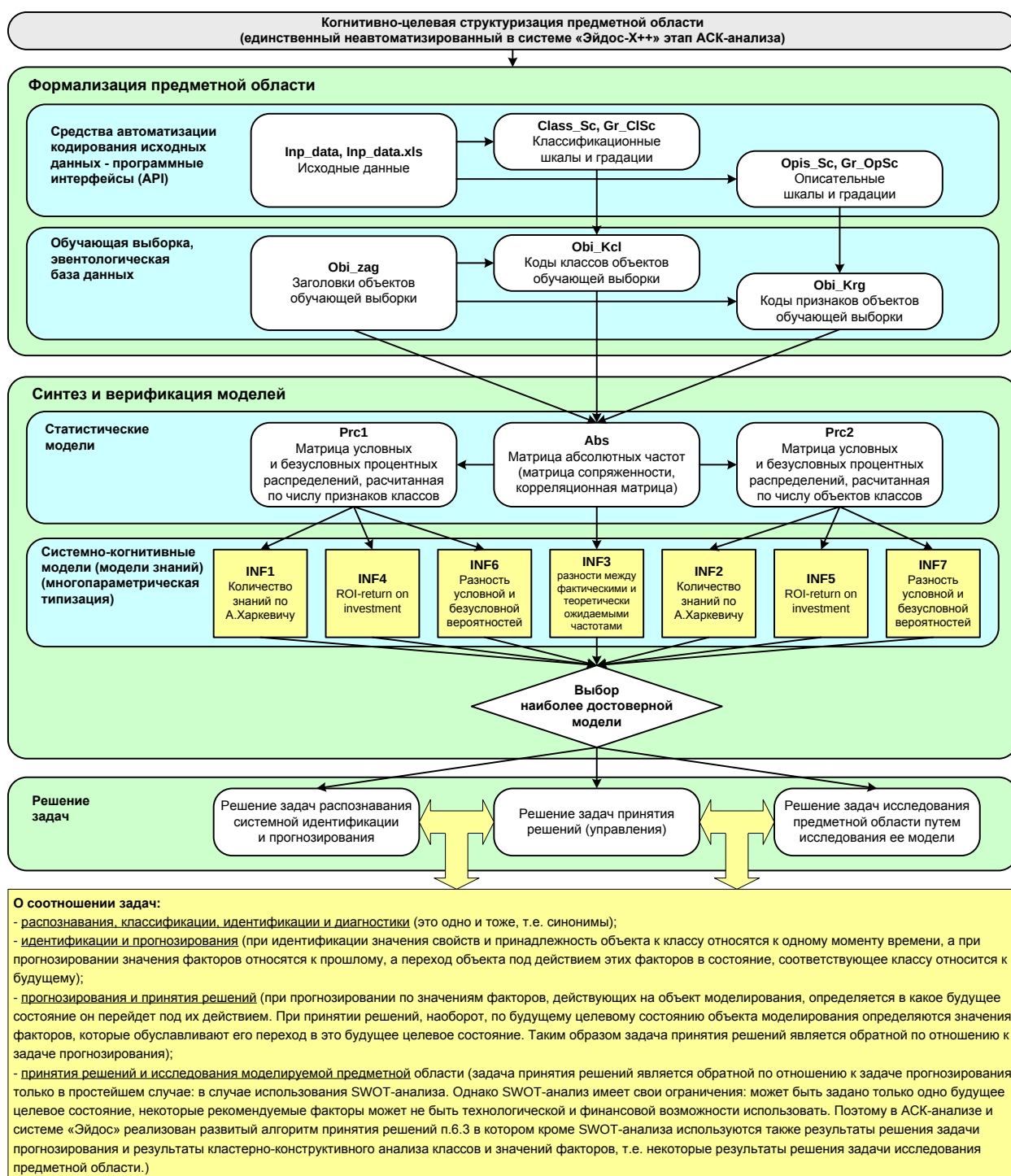
- вместо того, чтобы предъявлять к исходным данным практически неосуществимые требования (вроде нормальности распределения, абсолютной точности и полных повторностей всех сочетаний значений факторов и их полной независимости и аддитивности) автоматизированный системно-когнитивный анализ (АСК-анализ) 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 (рисунок 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»**

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 характеристики признаков рака молочной железы (таблица 1), а в качестве факторов, влияющих на эти результаты – его средние/юго-восточные/худшие текстура, площадь, периметр, гладкость, компактность, вогнутость, вогнутые точки, симметрии, фрактальные размерности, радиусы. (таблица 2):

Таблица 1 – Классификационная шкала

Код	Наименование
1	Уникальный идентификатор

Таблица 2 – Описательные шкалы

Код	Наименование
1	Диагноз
2	Средний/ЮВ/Худший радиус
3	Средний/ЮВ/Худший площадь
4	Средний/ЮВ/Худший периметр
5	Средний/ЮВ/Худший вогнутость
6	Средний/ЮВ/Худший текстура
7	Средний/ЮВ/Худший гладкость
8	Средний/ЮВ/Худший компактность
9	Средний/ЮВ/Худший вогнутые точки

10	Средний/ЮВ/Худший значение симметрии
11	Средний/ЮВ/Худший фрактальной размерности
12	ЮВ радиуса
13	ЮВ текстуры
14	ЮВ периметр

Задача 2: подготовка исходных данных и формализация предметной области

Исходные данные для данной статьи (рисунок 2) получены из других подробных таблиц по болезни, найденных в интернете.

	A	B	C	D	E	F	G	H	I	J	K
1	уникальн	диагноз	средний г	средняя т	средний г	средняя г	среднее ε	средняя κ	средняя ν	средние ε	среднее ε
2	842302	M	17,99	10,38	122,8	1001	0,1184	0,2776	0,3001	0,1471	0,2419
3	842517	M	20,57	17,77	132,9	1326	0,08474	0,07864	0,0869	0,07017	0,1812
4	84300903	M	19,69	21,25	130	1203	0,1096	0,1599	0,1974	0,1279	0,2069
5	84348301	M	11,42	20,38	77,58	386,1	0,1425	0,2839	0,2414	0,1052	0,2597
6	84358402	M	20,29	14,34	135,1	1297	0,1003	0,1328	0,198	0,1043	0,1809
7	843786	M	12,45	15,7	82,57	477,1	0,1278	0,17	0,1578	0,08089	0,2087
8	844359	M	18,25	19,98	119,6	1040	0,09463	0,109	0,1127	0,074	0,1794
9	84458202	M	13,71	20,83	90,2	577,9	0,1189	0,1645	0,09366	0,05985	0,2196
10	844981	M	13	21,82	87,5	519,8	0,1273	0,1932	0,1859	0,09353	0,235
11	84501001	M	12,46	24,04	83,97	475,9	0,1186	0,2396	0,2273	0,08543	0,203
12	845636	M	16,02	23,24	102,7	797,8	0,08206	0,06669	0,03299	0,03323	0,1528
13	84610002	M	15,78	17,89	103,6	781	0,0971	0,1292	0,09954	0,06606	0,1842
14	846226	M	19,17	24,8	132,4	1123	0,0974	0,2458	0,2065	0,1118	0,2397
15	846381	M	15,85	23,95	103,7	782,7	0,08401	0,1002	0,09938	0,05364	0,1847
16	84667401	M	13,73	22,61	93,6	578,3	0,1131	0,2293	0,2128	0,08025	0,2069
17	84799002	M	14,54	27,54	96,73	658,8	0,1139	0,1595	0,1639	0,07364	0,2303
18	848406	M	14,68	20,13	94,74	684,5	0,09867	0,072	0,07395	0,05259	0,1586
19	84862001	M	16,13	20,68	108,1	798,8	0,117	0,2022	0,1722	0,1028	0,2164
20	849014	M	19,81	22,15	130	1260	0,09831	0,1027	0,1479	0,09498	0,1582
21	8510426	B	13,54	14,36	87,46	566,3	0,09779	0,08129	0,06664	0,04781	0,1885
22	8510653	B	13,08	15,71	85,63	520	0,1075	0,127	0,04568	0,0311	0,1967
23	8510824	B	9,504	12,44	60,34	273,9	0,1024	0,06492	0,02956	0,02076	0,1815
24	8511123	B	15,24	14,26	102,5	704,4	0,1073	0,2135	0,2077	0,08756	0,2591

Рисунок 2 – Исходные данные для ввода в систему «Эйдос»

Затем с параметрами, показанными на рисунке 3, запустим режим 2.3.2.2 системы «Эйдос», представляющий собой автоматизированный программный интерфейс (API) с внешними данными табличного типа.

Задайте тип файла исходных данных: "Inp_data":

☐ XLS - MS Excel-2003 Стандарт XLS-файла
☒ XLSX- MS Excel-2007(2010)
☐ 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_rasp")

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

Какие наименования ГРАДАЦИЙ числовых шкал использовать:

☐ Только интервальные числовые значения (например: "1/3-{59873.0000000, 178545.6666667}")
☒ Только наименования интервальных числовых значений (например: "Минимальное")
☐ И интервальные числовые значения, и их наименования (например: "Минимальное: 1/3-{59873.0000000, 178545.6666667}")

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

ЗАДАНИЕ В ДИАЛОГЕ РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [2 x 360]

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	0	0	0,00	30	360	12,00
Текстовые	1	2	2,00	0	0	0,00
ВСЕГО:	1	2	2,00	30	360	12,00

Задайте число интервалов (градаций) в шкале:

В описательных шкалах:

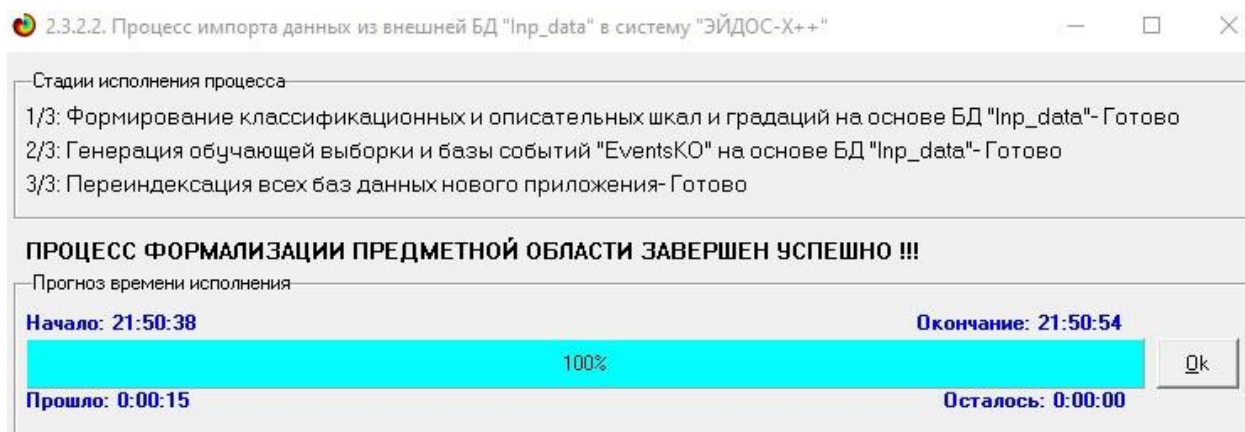


Рисунок 3. Экранная форма программного интерфейса (API) 2.3.2.2 системы «Эйдос» с внешними данными табличного типа

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. В описательных шкалах задано 3 числовых интервальных значения.

На рисунке 4 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 значения параметров могут быть представлены как числовыми, так и текстовыми значениями.



Рисунок 4. Экранная форма HELP программного интерфейса (API) 2.3.2.2

В результате работы режима сформирована 1 классификационная шкала с суммарным количеством градаций (классов) 4 (таблица 3) и 15

описательных шкал с суммарным числом градаций 45 (таблица 4). С использованием классификационных и описательных шкал и градаций исходные данные (рисунок 2) были закодированы и в результате получена обучающая выборка (таблица 5):

Таблица 3 – Классификационные шкалы и градации

Код	Название
1	1/4-Heavy
2	2/4-Pistol
3	3/4-Rifle
4	4/4-SMG

Таблица 4 – Описательные шкалы и градации

Код	Название
1	1/3-{200.0000000, 1050.0000000}
2	2/3-{1050.0000000, 1800.0000000}
3	3/3-{1800.0000000, 5200.0000000}
4	1/3-{100.0000000, 300.0000000}
5	2/3-{300.0000000, 300.0000000}
6	3/3-{300.0000000, 900.0000000}
7	1/3-{0.6666667, 1.7500000}
8	2/3-{1.7500000, 5.6666667}
9	3/3-{5.6666667, 47.5000000}
10	1/3-{150.0000000, 220.0000000}
11	2/3-{220.0000000, 230.0000000}
12	3/3-{230.0000000, 240.0000000}
13	1/3-{41.0000000, 352.0000000}
14	2/3-{352.0000000, 666.0000000}
15	3/3-{666.0000000, 857.0000000}
16	1/3-{26.0000000, 30.0000000}
17	2/3-{30.0000000, 36.0000000}
18	3/3-{36.0000000, 256.0000000}
19	1/3-{70.0000000, 284.0000000}
20	2/3-{284.0000000, 333.0000000}
21	3/3-{333.0000000, 467.0000000}
22	1/3-{47.0000000, 64.0000000}
23	2/3-{64.0000000, 77.6500000}
24	3/3-{77.6500000, 100.0000000}
25	1/3-{0.0000000, 1.0000000}
26	2/3-{1.0000000, 2.0000000}
27	3/3-{2.0000000, 2.5000000}
28	1/3-{5.0000000, 13.0000000}
29	2/3-{13.0000000, 30.0000000}
30	3/3-{30.0000000, 150.0000000}
31	1/3-{8.0000000, 52.0000000}
32	2/3-{52.0000000, 90.0000000}
33	3/3-{90.0000000, 300.0000000}
34	1/3-{2.1000000, 2.5000000}
35	2/3-{2.5000000, 3.1000000}
36	3/3-{3.1000000, 5.7000000}
37	1/3-{16.0000000, 21.0000000}
38	2/3-{21.0000000, 28.0000000}
39	3/3-{28.0000000, 165.0000000}
40	1/3-{2.0000000, 15.0000000}
41	2/3-{15.0000000, 24.0000000}
42	3/3-{24.0000000, 76.0000000}
43	1/3-{0.2000000, 0.6000000}
44	2/3-{0.6000000, 2.0000000}
45	3/3-{2.0000000, 62.0000000}

Таблица 5 – Обучающая выборка

№	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	уникальн	диагноз	средний р	средняя т	средний п	средняя п	среднее з	средняя к	средняя в	средние в	среднее з	среднее з	юго-восто	ЮВ тексту	ЮВ перим	ЮВ площадь
2	842302	M	17,99	10,38	122,8	1001	0,1184	0,2776	0,3001	0,1471	0,2419	0,07871	1,095	0,9053	8,589	153,4
3	842517	M	20,57	17,77	132,9	1326	0,08474	0,07864	0,0869	0,07017	0,1812	0,05667	0,5435	0,7339	3,398	74,08
4	84300903	M	19,69	21,25	130	1203	0,1096	0,1599	0,1974	0,1279	0,2069	0,05999	0,7456	0,7869	4,585	94,03
5	84348301	M	11,42	20,38	77,58	386,1	0,1425	0,2839	0,2414	0,1052	0,2597	0,09744	0,4956	1,156	3,445	27,23
6	84358402	M	20,29	14,34	135,1	1297	0,1003	0,1328	0,198	0,1043	0,1809	0,05883	0,7572	0,7813	5,438	94,44
7	843786	M	12,45	15,7	82,57	477,1	0,1278	0,17	0,1578	0,08089	0,2087	0,07613	0,3345	0,8902	2,217	27,19
8	844359	M	18,25	19,98	119,6	1040	0,09463	0,109	0,1127	0,074	0,1794	0,05742	0,4467	0,7732	3,18	53,91
9	84458202	M	13,71	20,83	90,2	577,9	0,1189	0,1645	0,09366	0,05985	0,2196	0,07451	0,5835	1,377	3,856	50,96
10	844981	M	13	21,82	87,5	519,8	0,1273	0,1859	0,09353	0,235	0,07389	0,3063	1,002	2,406	24,32	
11	84501001	M	12,46	24,04	83,97	475,9	0,1186	0,2396	0,2273	0,08543	0,203	0,08243	0,2976	1,599	2,039	23,94
12	845636	M	16,02	23,24	102,7	797,8	0,08206	0,06669	0,03299	0,03323	0,1528	0,05697	0,3795	1,187	2,466	40,51
13	84610002	M	15,78	17,89	103,6	781	0,0971	0,1292	0,09954	0,06606	0,1842	0,06082	0,5058	0,9849	3,564	54,16
14	846226	M	19,17	24,8	132,4	1123	0,0974	0,2458	0,2065	0,1118	0,2397	0,078	0,9555	3,568	11,07	116,2
15	846381	M	15,85	23,95	103,7	782,7	0,08401	0,1002	0,09938	0,05364	0,1847	0,05338	0,4033	1,078	2,903	36,58
16	84667401	M	13,73	22,61	93,6	578,3	0,1131	0,2293	0,2128	0,08025	0,2069	0,07682	0,2121	1,169	2,061	19,21
17	84799002	M	14,54	27,54	96,73	658,8	0,1139	0,1595	0,1639	0,07364	0,2303	0,07077	0,37	1,033	2,879	32,55
18	848406	M	14,68	20,13	94,74	684,5	0,09867	0,072	0,07395	0,05259	0,1586	0,05922	0,4727	1,24	3,195	45,4
19	84862001	M	16,13	20,68	108,1	798,8	0,117	0,2022	0,1722	0,1028	0,2164	0,07356	0,5692	1,073	3,854	54,18
20	849014	M	19,81	22,15	130	1260	0,09831	0,1027	0,1479	0,09498	0,1582	0,05395	0,7582	1,017	5,865	112,4
21	8510426	B	13,54	14,36	87,46	566,3	0,09779	0,08129	0,06664	0,04781	0,1885	0,05766	0,2699	0,7886	2,058	23,56
22	8510653	B	13,08	15,71	85,63	520	0,1075	0,127	0,04568	0,0311	0,1967	0,06811	0,1852	0,7477	1,383	14,67
23	8510824	B	9,504	12,44	60,34	273,9	0,1024	0,06492	0,02956	0,02076	0,1815	0,06905	0,2773	0,9768	1,909	15,7
24	8511133	M	15,34	14,26	102,5	704,4	0,1073	0,2135	0,2077	0,09756	0,2521	0,07032	0,4388	0,7096	3,384	44,91
25	851509	M	21,16	23,04	137,2	1404	0,09428	0,1022	0,1097	0,08632	0,1769	0,05278	0,6917	1,127	4,303	93,99
26	852552	M	16,65	21,38	110	904,6	0,1121	0,1457	0,1525	0,0917	0,1995	0,0633	0,8068	0,9017	5,455	102,6
27	852631	M	17,14	16,4	116	912,7	0,1186	0,2276	0,2229	0,1401	0,304	0,07413	1,046	0,976	7,276	111,4
28	852763	M	14,58	21,53	97,41	644,8	0,1054	0,1868	0,1425	0,08783	0,2252	0,06924	0,2545	0,9832	2,11	21,05
29	852781	M	18,61	20,25	122,1	1094	0,0944	0,1066	0,149	0,07731	0,1697	0,05699	0,8529	1,849	5,632	93,54
30	852973	M	15,3	25,27	102,4	732,4	0,1082	0,1697	0,1683	0,08751	0,1926	0,0654	0,439	1,012	3,498	43,5
31	853201	M	17,57	15,05	115	955,1	0,09847	0,1157	0,09875	0,07953	0,1739	0,06149	0,6003	0,8225	4,655	61,1
32	853401	M	18,63	25,11	124,8	1088	0,1064	0,1887	0,2319	0,1244	0,2183	0,06197	0,8307	1,466	5,574	105
33	853612	M	11,84	18,7	77,93	440,6	0,1109	0,1516	0,1218	0,05182	0,2301	0,07799	0,4825	1,03	3,475	41
34	85382601	M	17,02	23,98	112,8	899,3	0,1197	0,1496	0,2417	0,1203	0,2248	0,06382	0,6009	1,398	3,999	67,78
35	854002	M	19,27	26,47	127,9	1162	0,09401	0,1719	0,1657	0,07593	0,1853	0,06261	0,5558	0,6062	3,528	68,17
36	854039	M	16,13	17,88	107	807,2	0,104	0,1559	0,1354	0,07752	0,1998	0,06515	0,334	0,6857	2,183	35,03
37	854253	M	16,74	21,59	110,1	869,5	0,0961	0,1336	0,1348	0,06018	0,1896	0,05656	0,4615	0,9197	3,008	45,19

Обучающая выборка (таблица 5), по сути, представляет собой нормализованные исходные данные, т.е. таблицу исходных данных (рисунок 2), закодированную с помощью классификационных и описательных шкал и градаций (таблицы 3 и 4). 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 5).

3.3. Синтез и верификация моделей

Задайте модели для синтеза и верификации

Статистические базы:

- ☒ 1. ABS - частный критерий: количество встреч сочетаний: "класс-признак" у объектов обуч.выборки
- ☒ 2. PRC1 - частный критерий: усл. вероятность i-го признака среди признаков объектов j-го класса
- ☒ 3. PRC2 - частный критерий: условная вероятность i-го признака у объектов j-го класса

Системно-когнитивные модели (базы знаний):

- ☒ 4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1
- ☒ 5. INF2 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC2
- ☒ 6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс. частотами
- ☒ 7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1
- ☒ 8. INF5 - частный критерий: ROI (Return On Investment); вероятности из PRC2
- ☒ 9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вероятности из PRC1
- ☒ 10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; вероятности из PRC2

Текущая модель

- ☐ 1. ABS
- ☐ 2. PRC1
- ☐ 3. PRC2
- ☒ 4. INF1
- ☐ 5. INF2
- ☐ 6. INF3
- ☐ 7. INF4
- ☐ 8. INF5
- ☐ 9. INF6
- ☐ 10. INF7

Параметры копирования обучающей выборки в распознаваемую (бутстрепный подход):

Какие объекты обуч.выборки копировать:

- ☒ Копировать всю обучающую выборку
- ☐ Копировать только текущий объект
- ☐ Копировать каждый N-й объект
- ☐ Копировать N случайных объектов
- ☐ Копировать объекты от N1 до N2 (fastest)
- ☐ Вообще не менять распознаваемую выборку

Пояснение по алгоритму верификации

Удалять из обуч.выборки скопированные объекты:

- ☒ Не удалять
- ☐ Удалять

Подробнее

Измеряется внутренняя достоверн. модели

Для каждой заданной модели выполнять:

- ☒ Синтез и верификацию
- ☐ Только верификацию
- ☐ Только синтез

Задайте процессор

- ☒ CPU
- ☐ GPU

Задайте алгоритм:

- ☐ Классика - дольше
- ☒ Упрощенно-быстрее

Использование только наиболее достоверных результатов распознавания: Rasp.dbf и целесообразность применения бутстрепного подхода

Расчетный размер БД результатов распознавания Rasp.dbf равен 57094 байт, т.е.: 0.0026586 % от MAX-возможного, (от 2Гб)

Задайте, сколько % от исходной БД Rasp.dbf оставить, удаляя наименее достоверные результаты распознавания:

В применении бутстрепного подхода нет необходимости. Синтез и верификация моделей будут выполнены на основе всей выборки.

Ok

Cancel

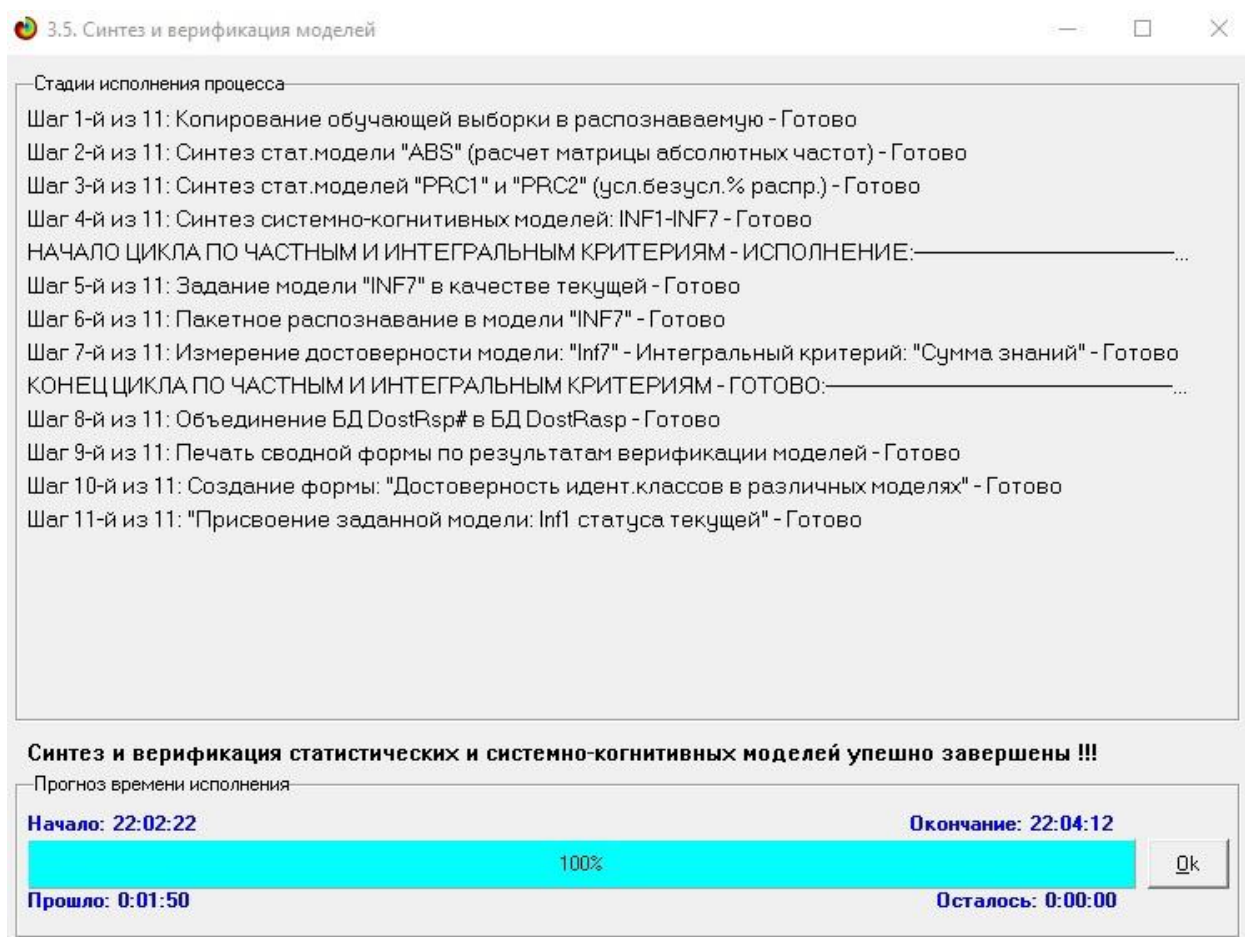


Рисунок 5. Экранные формы режима синтеза и верификации статистических и системно-когнитивных моделей системы «Эйдос»

Обратим внимание на то, что на рисунке 5 в правом нижнем углу окна задана опция: «Расчеты проводить на графическом процессоре (GPU)».

Из рисунка 5 видно, что весь процесс синтеза и верификации моделей занял 14 секунд. Отметим, что при синтезе и верификации моделей использовался графический процессор (GPU) видеокарты. На центральном процессоре (CPU) выполнение этих операций занимает значительно большее время (на некоторых задачах это происходит в десятки, сотни и даже тысячи раз дольше). Таким образом, неграфические вычисления на графических процессорах видеокарты делает возможной обработку больших объемов исходных данных за разумное время. В процесс синтеза и верификации моделей осуществляется также расчет 10 выходных форм, на что уходит более 99% времени исполнения.

Фрагменты самих созданных 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. ДИАГНОЗ В	2. ДИАГНОЗ М	Сумма	Среднее	Средн. квадр. откл.
1	СРЕДНИЙ РАДИУС-1-е из 12	13		13	6.50	9.19
2	СРЕДНИЙ РАДИУС-2-е из 12	52		52	26.00	36.77
3	СРЕДНИЙ РАДИУС-3-е из 12	120	6	126	63.00	80.61
4	СРЕДНИЙ РАДИУС-4-е из 12	121	26	147	73.50	67.18
5	СРЕДНИЙ РАДИУС-5-е из 12	45	44	89	44.50	0.71
6	СРЕДНИЙ РАДИУС-6-е из 12	5	35	40	20.00	21.21
7	СРЕДНИЙ РАДИУС-7-е из 12	1	39	40	20.00	26.87
8	СРЕДНИЙ РАДИУС-8-е из 12		40	40	20.00	28.28
9	СРЕДНИЙ РАДИУС-9-е из 12		10	10	5.00	7.07
10	СРЕДНИЙ РАДИУС-10-е из 12		6	6	3.00	4.24
11	СРЕДНИЙ РАДИУС-11-е из 12		3	3	1.50	2.12
12	СРЕДНИЙ РАДИУС-12-е из 12		3	3	1.50	2.12
13	СРЕДНЯЯ ТЕКСТУРА-1-е из 12	10	2	12	6.00	5.66
14	СРЕДНЯЯ ТЕКСТУРА-2-е из 12	56	2	58	29.00	38.18
15	СРЕДНЯЯ ТЕКСТУРА-3-е из 12	99	16	115	57.50	58.69
16	СРЕДНЯЯ ТЕКСТУРА-4-е из 12	98	38	136	68.00	42.43
17	СРЕДНЯЯ ТЕКСТУРА-5-е из 12	49	69	118	59.00	14.14
18	СРЕДНЯЯ ТЕКСТУРА-6-е из 12	18	44	62	31.00	18.38
19	СРЕДНЯЯ ТЕКСТУРА-7-е из 12	9	25	34	17.00	11.31
20	СРЕДНЯЯ ТЕКСТУРА-8-е из 12	14	10	24	12.00	2.83
21	СРЕДНЯЯ ТЕКСТУРА-9-е из 12	3	3	6	3.00	
22	СРЕДНЯЯ ТЕКСТУРА-10-е из 12	1	2	3	1.50	0.71
23	СРЕДНЯЯ ТЕКСТУРА-11-е из 12					
24	СРЕДНЯЯ ТЕКСТУРА-12-е из 12		1	1	0.50	0.71
25	СРЕДНИЙ ПЕРИМЕТР-1-е из 12	13		13	6.50	9.19
26	СРЕДНИЙ ПЕРИМЕТР-2-е из 12	55		55	27.50	38.89

Рисунок 6. Матрица абсолютных частот (фрагмент)

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

Код признака	Наименование описательной шкалы и градации	1. ДИАГНОЗ В	2. ДИАГНОЗ М	Сумма	Среднее	Средн. квадр. откл.
1	СРЕДНИЙ РАДИУС-1-е из 12	4.865	-4.865			6.880
2	СРЕДНИЙ РАДИУС-2-е из 12	19.459	-19.459			27.520
3	СРЕДНИЙ РАДИУС-3-е из 12	41.152	-41.152			58.197
4	СРЕДНИЙ РАДИУС-4-е из 12	29.010	-29.010			41.027
5	СРЕДНИЙ РАДИУС-5-е из 12	-10.694	10.694			15.124
6	СРЕДНИЙ РАДИУС-6-е из 12	-20.031	20.031			28.328
7	СРЕДНИЙ РАДИУС-7-е из 12	-24.031	24.031			33.985
8	СРЕДНИЙ РАДИУС-8-е из 12	-25.031	25.031			35.399
9	СРЕДНИЙ РАДИУС-9-е из 12	-6.258	6.258			8.850
10	СРЕДНИЙ РАДИУС-10-е из 12	-3.755	3.755			5.310
11	СРЕДНИЙ РАДИУС-11-е из 12	-1.877	1.877			2.655
12	СРЕДНИЙ РАДИУС-12-е из 12	-1.877	1.877			2.655
13	СРЕДНЯЯ ТЕКСТУРА-1-е из 12	2.491	-2.491			3.522
14	СРЕДНЯЯ ТЕКСТУРА-2-е из 12	19.705	-19.705			27.867
15	СРЕДНЯЯ ТЕКСТУРА-3-е из 12	27.035	-27.035			38.234
16	СРЕДНЯЯ ТЕКСТУРА-4-е из 12	12.894	-12.894			18.235
17	СРЕДНЯЯ ТЕКСТУРА-5-е из 12	-24.842	24.842			35.132
18	СРЕДНЯЯ ТЕКСТУРА-6-е из 12	-20.798	20.798			29.413
19	СРЕДНЯЯ ТЕКСТУРА-7-е из 12	-12.277	12.277			17.362
20	СРЕДНЯЯ ТЕКСТУРА-8-е из 12	-1.019	1.019			1.441
21	СРЕДНЯЯ ТЕКСТУРА-9-е из 12	-0.755	0.755			1.067
22	СРЕДНЯЯ ТЕКСТУРА-10-е из 12	-0.877	0.877			1.241
23	СРЕДНЯЯ ТЕКСТУРА-11-е из 12					
24	СРЕДНЯЯ ТЕКСТУРА-12-е из 12	-0.626	0.626			0.885
25	СРЕДНИЙ ПЕРИМЕТР-1-е из 12	4.865	-4.865			6.880
26	СРЕДНИЙ ПЕРИМЕТР-2-е из 12	20.582	-20.582			29.107

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 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 частной модели в соответствии с этими мерами достоверности (рисунок 8).

3.4. Обобщ.форма по доств.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
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс.частот с обр...	569	569	4	53		0.915
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс.частот по признак...	569	569		57		0.909
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн.частот с о...	569	569	4	53		0.915
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн.частот по приз...	569	569		57		0.909
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн.частот с о...	569	569	4	53		0.915
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн.частот по приз...	569	569		57		0.909
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	569	501	53	4	68	0.992
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	569	496	55	2	73	0.996
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	569	501	53	4	68	0.992
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	569	497	55	2	72	0.996
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Семантический резонанс зна...	569	518	44	13	51	0.976
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Сумма знаний	569	518	44	13	51	0.976
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	569	464	53	4	105	0.991
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	569	538	54	3	31	0.994
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	569	464	53	4	105	0.991
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	569	537	54	3	32	0.994
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Семантический резонанс зна...	569	513	44	13	56	0.975
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Сумма знаний	569	518	43	14	51	0.974
10. INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Семантический резонанс зна...	569	513	45	12	56	0.977
10. INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Сумма знаний	569	519	43	14	50	0.974

мощь по мерам достоверно | мощь по частотным распределен | TP, TN, FP, FN | $TP \cdot FP / (TN \cdot FN)$ | $T \cdot F / (T + F) \cdot 100$ | задать интервал сглаживания

Рисунок 8. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена и L1-критерию проф. Е.В.Луценко

Из рисунка 8 мы видим, что в данном интеллектуальном приложении по F-критерию Ван Ризбергена наиболее достоверной является СК-модель INF3 с интегральным критерием «Сумма знаний» ($F=0,810$ при максимуме 1,000), что является довольно хорошим результатом, по критерию L1 проф.

Е.В.Луценко та же модель является наиболее достоверной ($L1=0,960$ при максимуме 1,000), что является отличным результатом.

Это подтверждает наличие и адекватное отражение в СК-модели INF3 сильной причинно-следственной зависимости между видом заболеваемости и ее характеристиками.

На рисунке 9 приведено частотное распределение числа истинных и ложных положительных и отрицательных решений по результатам идентификации характеристик болезни в СК-модели INF3 по данным обучающей выборки:

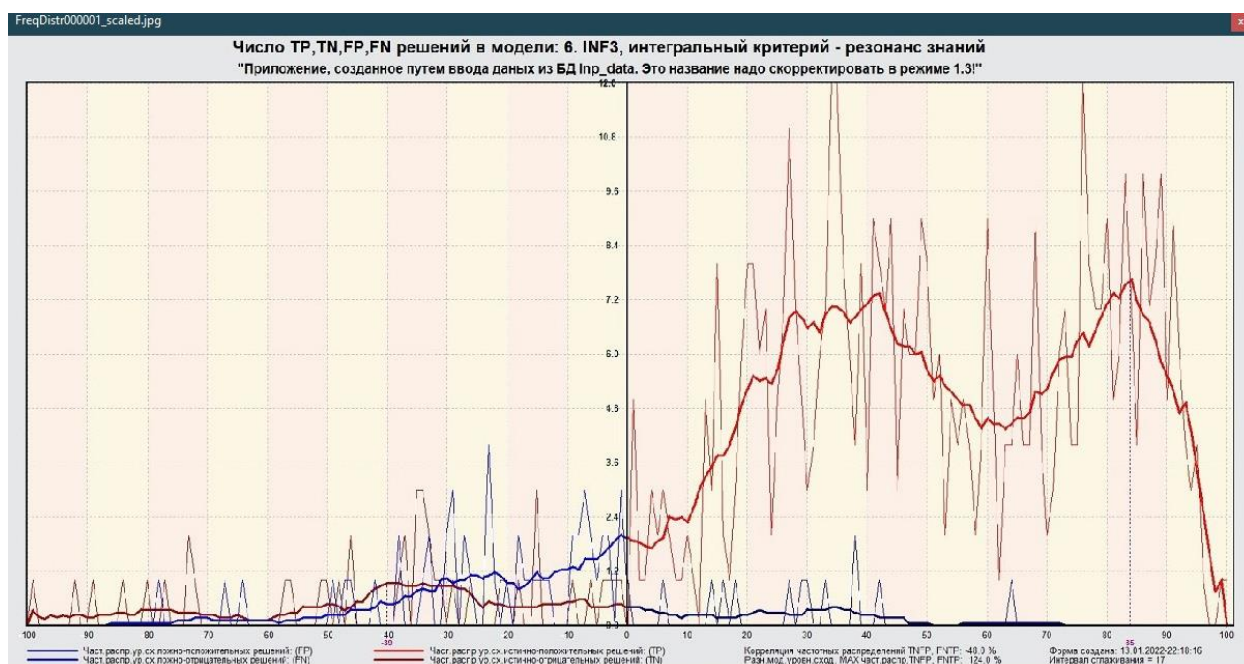


Рисунок 9. Частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности в СК-модели Inf3

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

Левое распределение, большее по амплитуде включает истинно-отрицательные и ложно-положительные решения, а правое, меньшее 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 болезни по характеристикам и другие задачи.

Видно, что для отрицательных решений количество истинных решений всегда значительно превосходит количество ложных решений, причем при уровнях различия больше примерно 5% ложные отрицательные решения вообще практически отсутствуют.

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

1) при уровнях сходства от 0% до 30% количество ложных решений больше числа истинных;

2) при уровнях сходства от 30% до примерно 42% есть и истинные и ложные положительные решения, но число истинных решений больше числа ложных и доля истинных решений возрастает при увеличении уровня сходства;

3) при уровнях сходства выше 42% ложные положительные решения не встречаются.

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

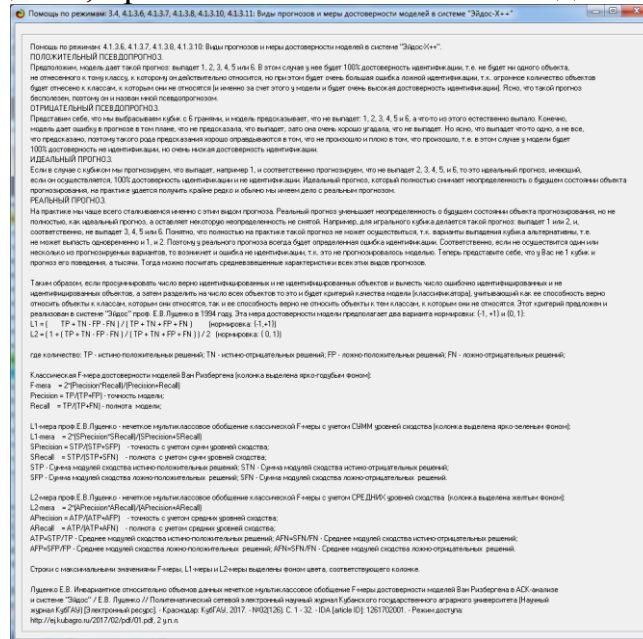


Рисунок 10. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергера и L1- и L2-критериям проф. Е.В.Луценко

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 INF3 статус текущей модели. Для этого запустим режим 5.6 с параметрами, приведенными на экранной форме (рисунок 11):

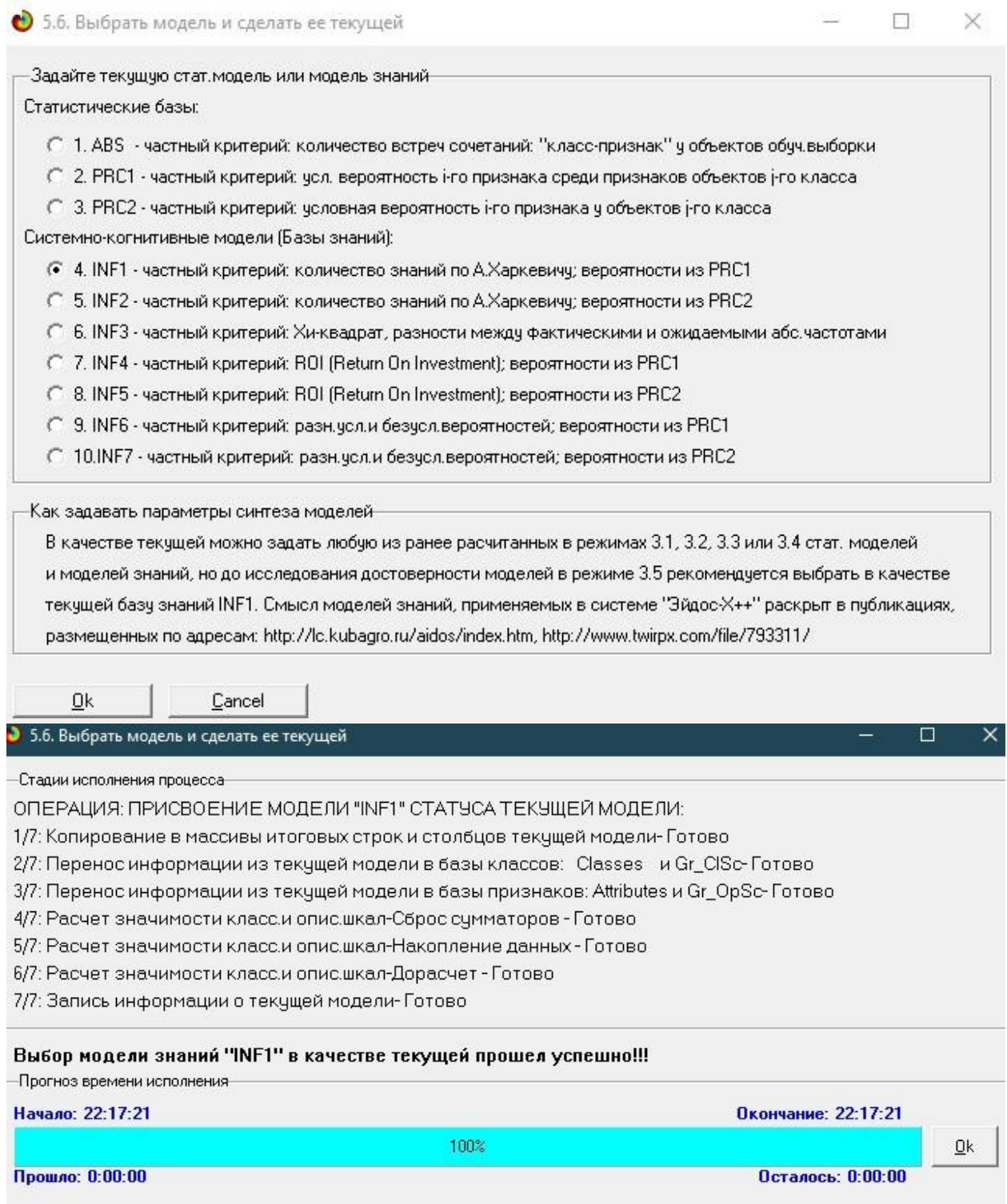


Рисунок 11. Экранные формы придания наиболее достоверной СК-модели Inf3 статуса текущей модели

Задача 4: решение различных задач в наиболее достоверной модели

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

Решим задачу системной идентификации, т.е. определение класса болезни на основе обучающей выборки в наиболее достоверной СК-модели INF3 на GPU (рисунок 12).

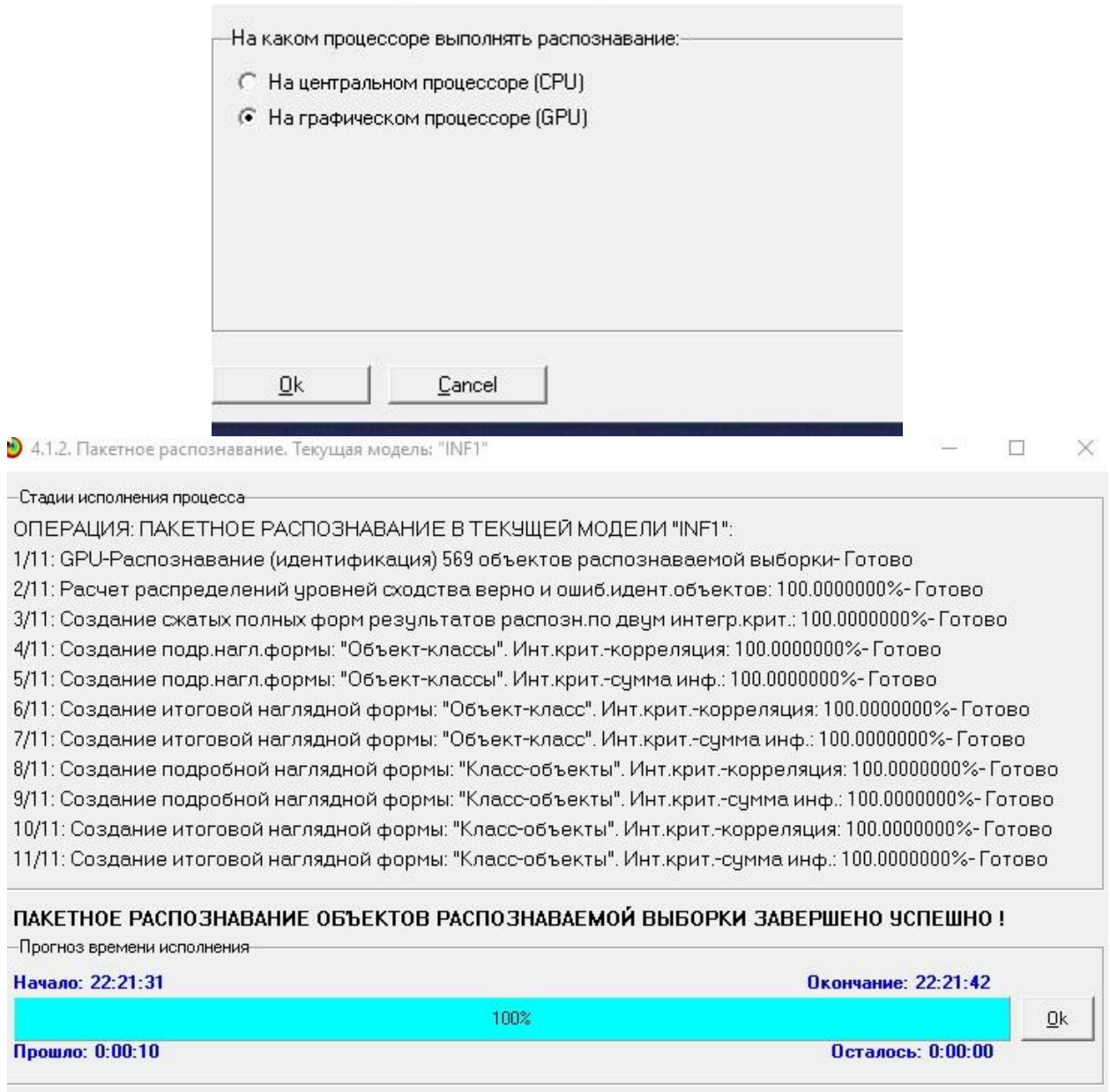


Рисунок 12. Экранные формы, которые отображают процесс решения задачи системной идентификации в текущей модели

Из рисунка 12 видно, что процесс идентификации занял 2 секунды.

Отметим, что 99% этого времени заняла не сама идентификация на GPU, а создание 10 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).

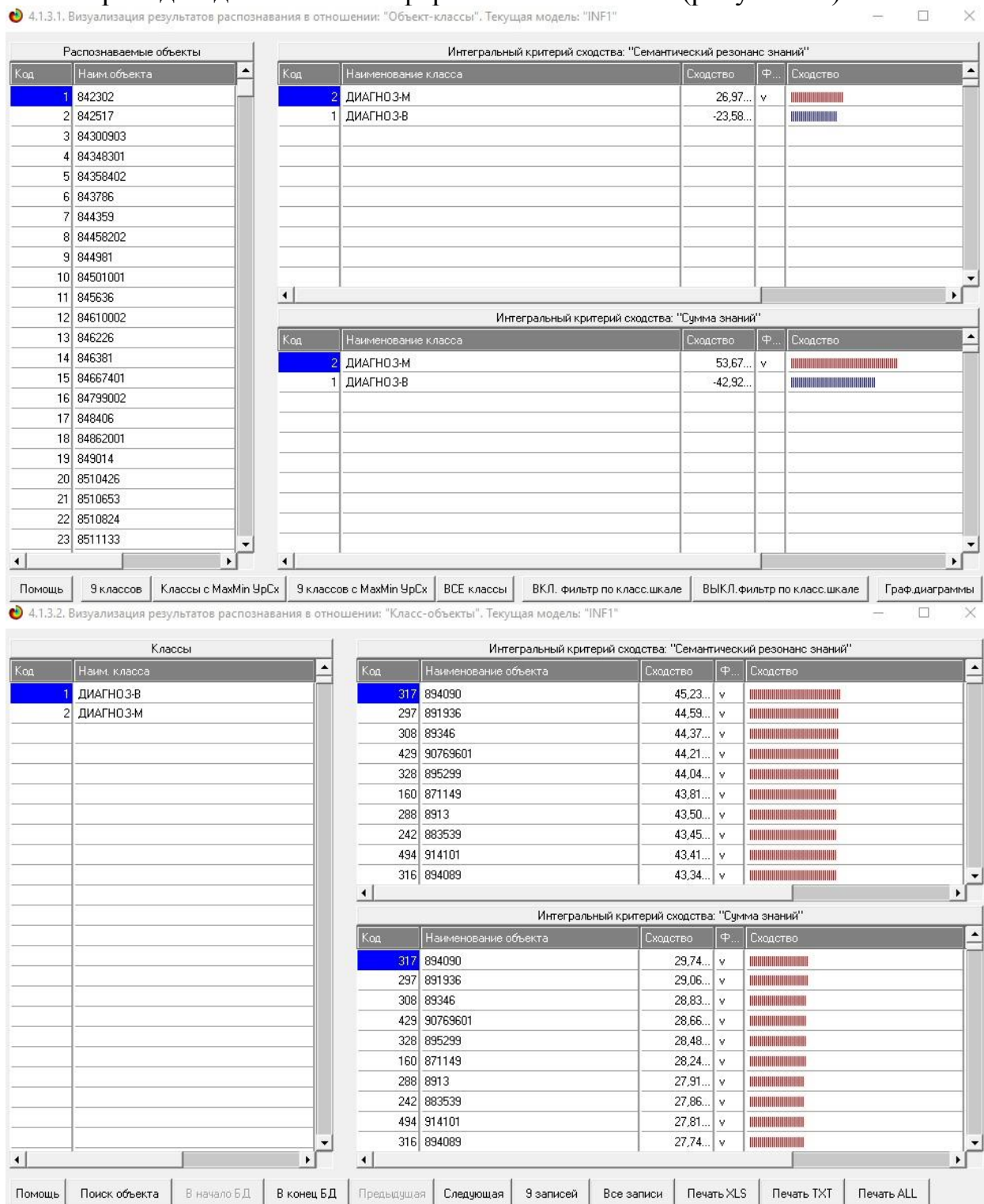


Рисунок 13. Выходные формы по результатам идентификации вида болезни по характеристикам

Символ «✓» стоит против тех результатов идентификации, которые подтвердились на опыте, т.е. соответствуют факту. Из рисунка 13 видно, что результаты идентификации являются отличными, естественно при учете информации из рисунка 9 о том, что достоверные прогнозы в данной модели имеют уровень сходства выше 42%, т.е., по сути, результаты с более низки уровнем сходства надо просто игнорировать.

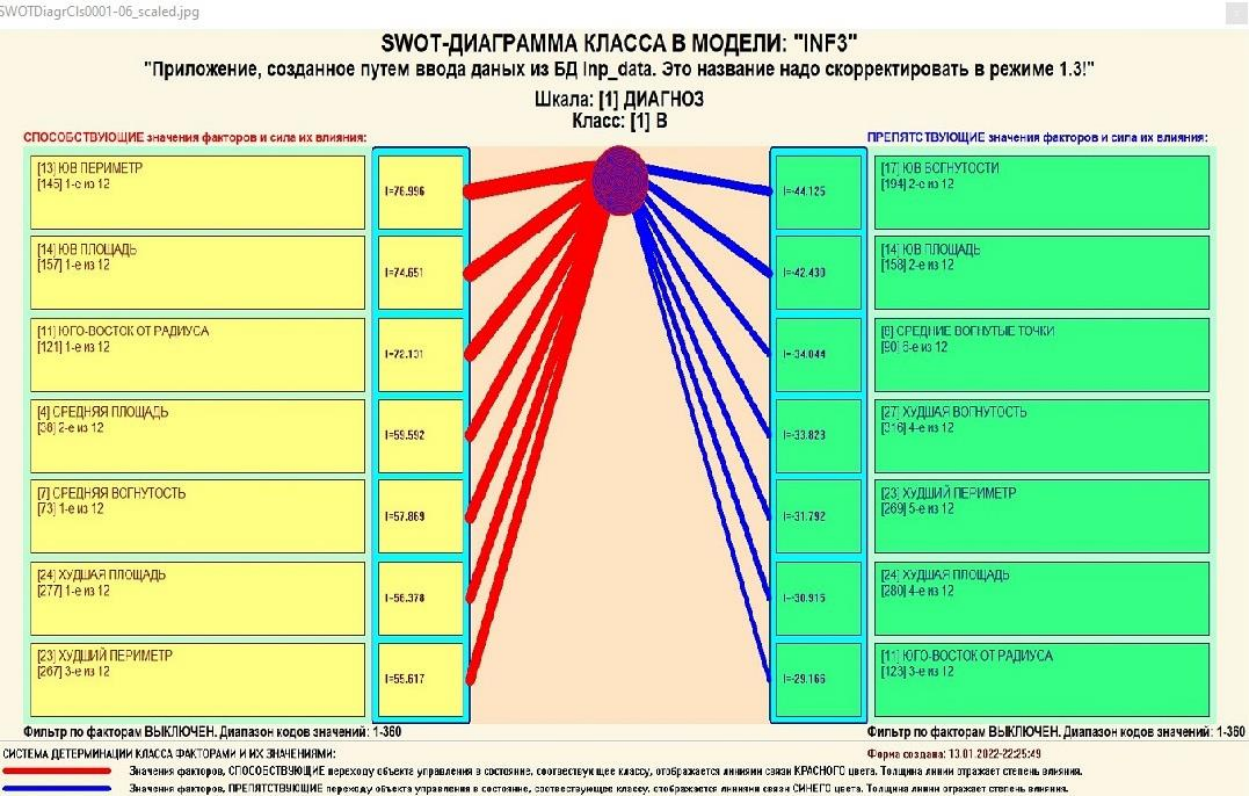
Подзадача 4.2. Поддержка принятия решений (SWOT-анализ)

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.

Применительно к задаче, решаемой в данной работе, SWOT-анализ показывает степень влияния различных значений характеристик заболеваемости на ее вид.

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 значений характеристик болезни на ее вид.



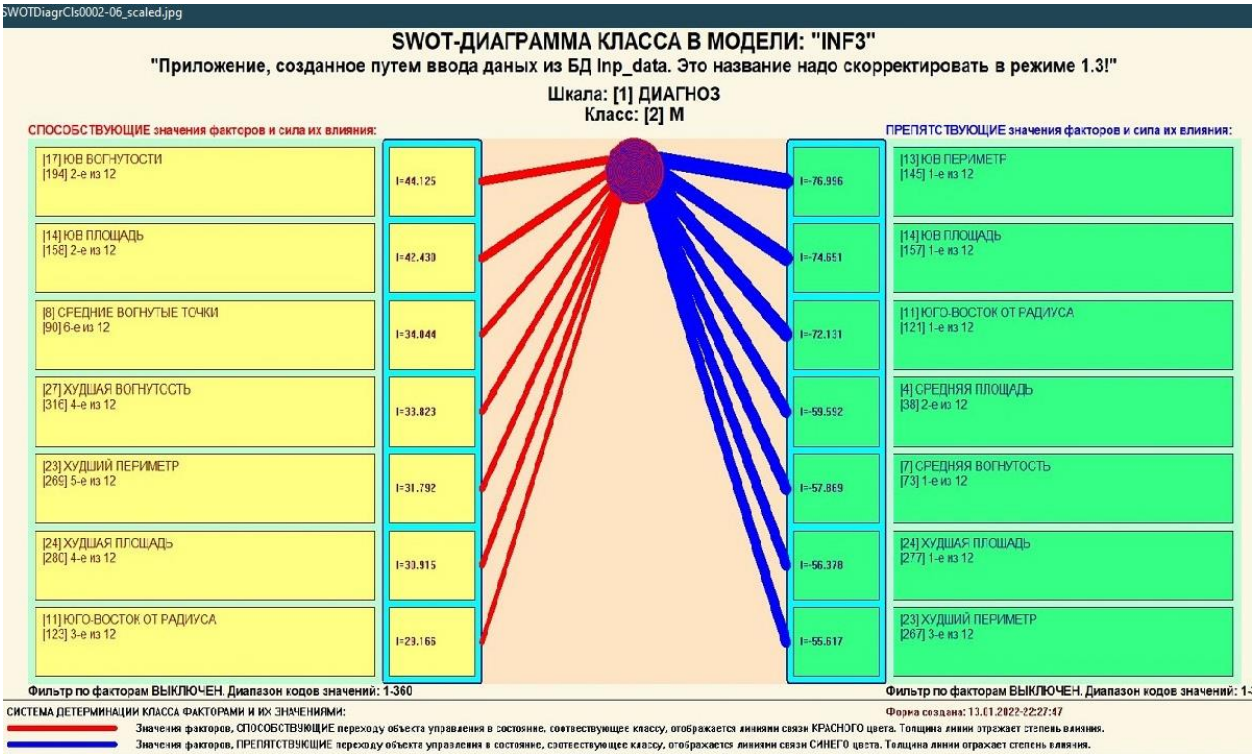


Рисунок 14. SWOT-диаграммы детерминации характеристик болезни

Эти SWOT-диаграммы наглядно отражают силу и направление влияния различных значений характеристик болезни на ее вид.

Отметим также, что система «Эйдос» обеспечивала решение этой задачи *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.

А К Т

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

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

Кириченко М.М. Кириченко
19.05.1987г.

Мл.научный сотрудник

Ляшко Г.А. Ляшко
19.05.1987г.

От СКФ ВНИЦ "АИУС-агроресурсы":

Зав.отделом аэрокосмических и
тематических изысканий №4, к.э.н.
Самсонов Г.А. Самсонов
19.05.1987г.

Главный конструктор проекта
Коренец В.И. Коренец
19.05.1987г.

Главный конструктор проекта
Луценко Е.В. Луценко
19.05.1987г.

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: \Aidos-X\AID_DATA\A0000001\System\SWOTCls####Inf3.DBF, где: «####» - код класса с ведущими нулями. Эти базы открываются в MS Excel.

На рисунке 15 приведены пример инвертированной SWOT-диаграммы, отражающей различные виды болезни на ее характеристики.

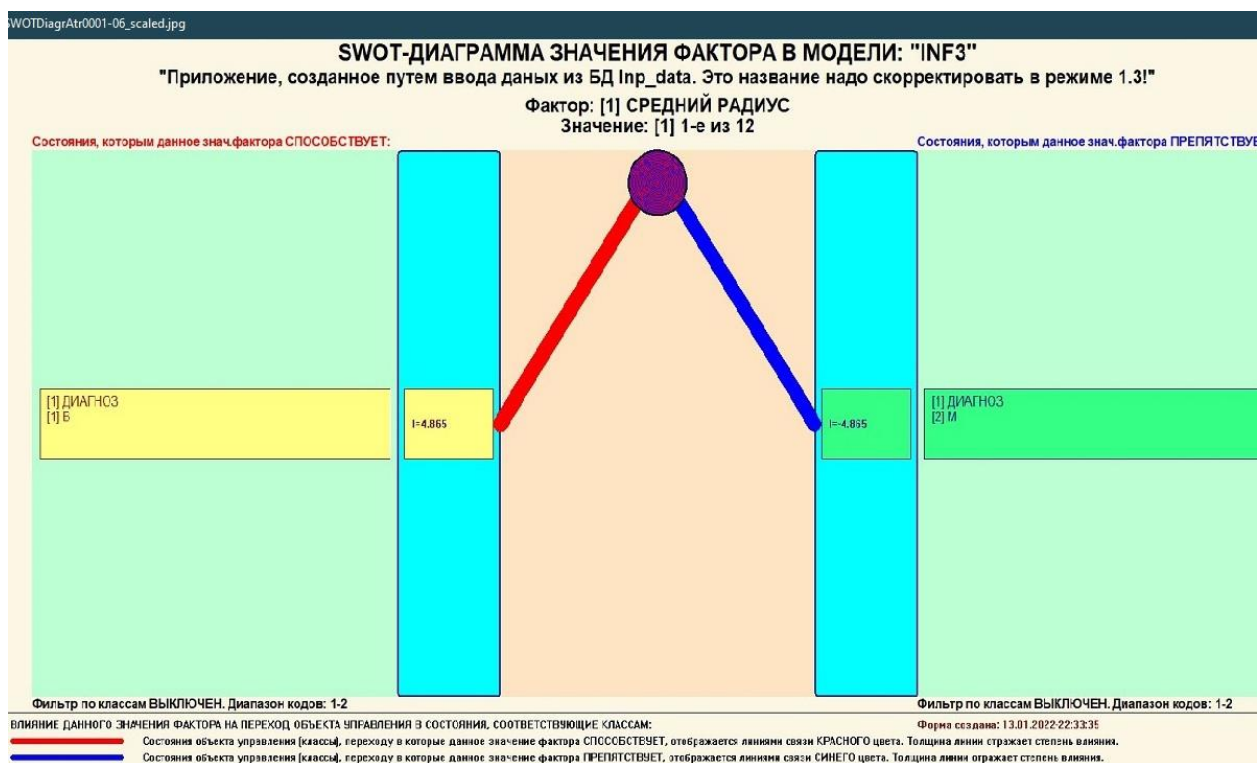


Рисунок 15. Пример SWOT-диаграммы, отражающей влияние болезни на ее характеристики

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».

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 признаков болезни 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 традиционно делается на основе экспертных оценок неформализуемым путем на основе опыта, интуиции и профессиональной компетенции.

SemNetCls2d000003_scaled.jpg

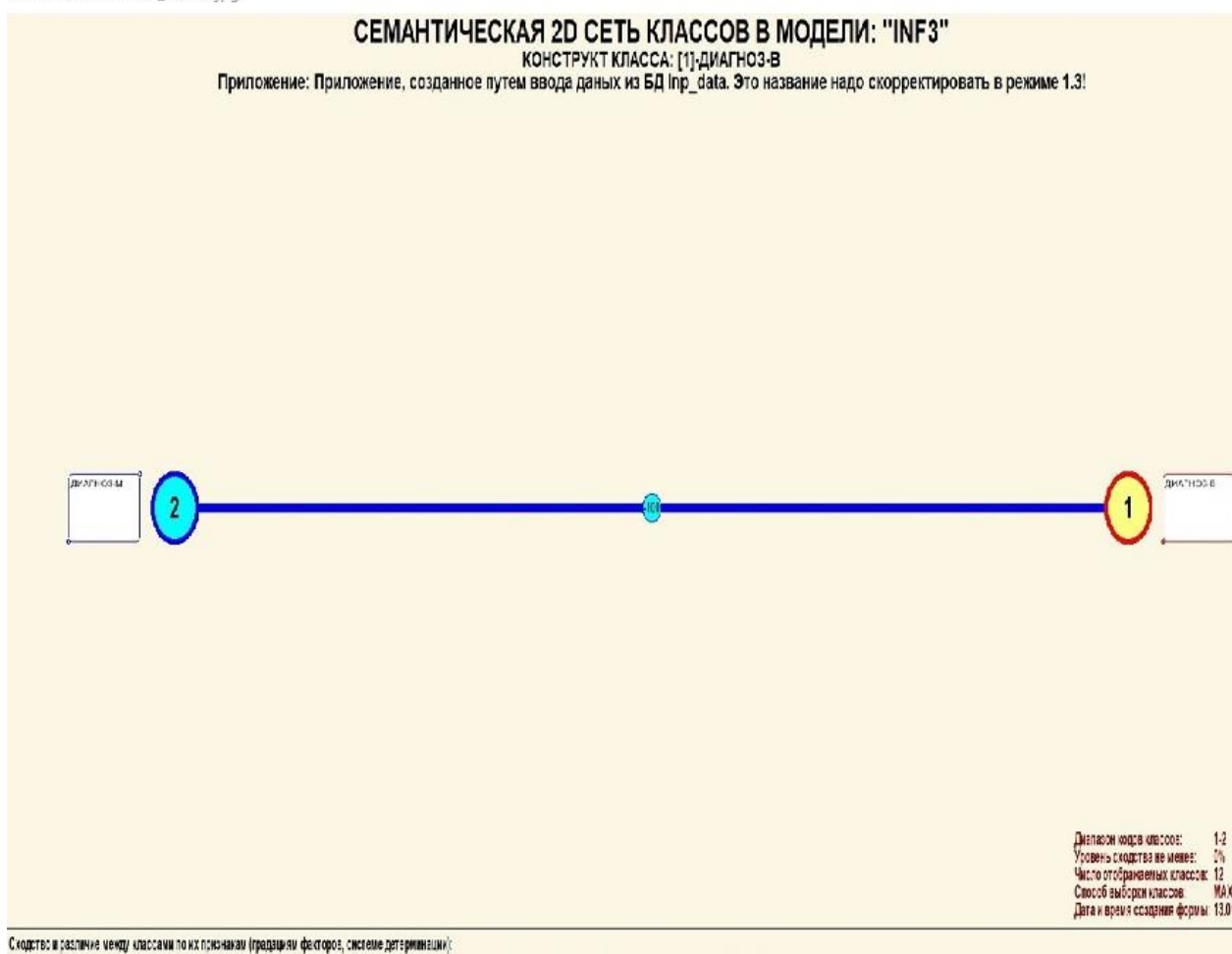


Рисунок 16. Когнитивная диаграмма классов, отражающая сходства/различия признаков болезни по связанным с ней значениями и характеристиками.

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

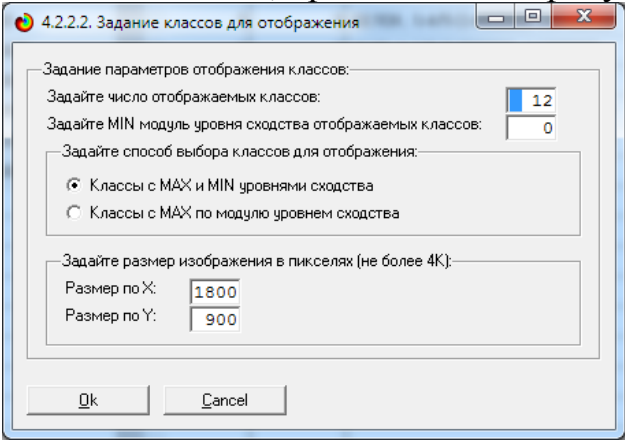


Рисунок 17. Диалоговое окно управления параметрами формирования и вывода изображения когнитивной диаграммы классов

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* (рисунок 18):

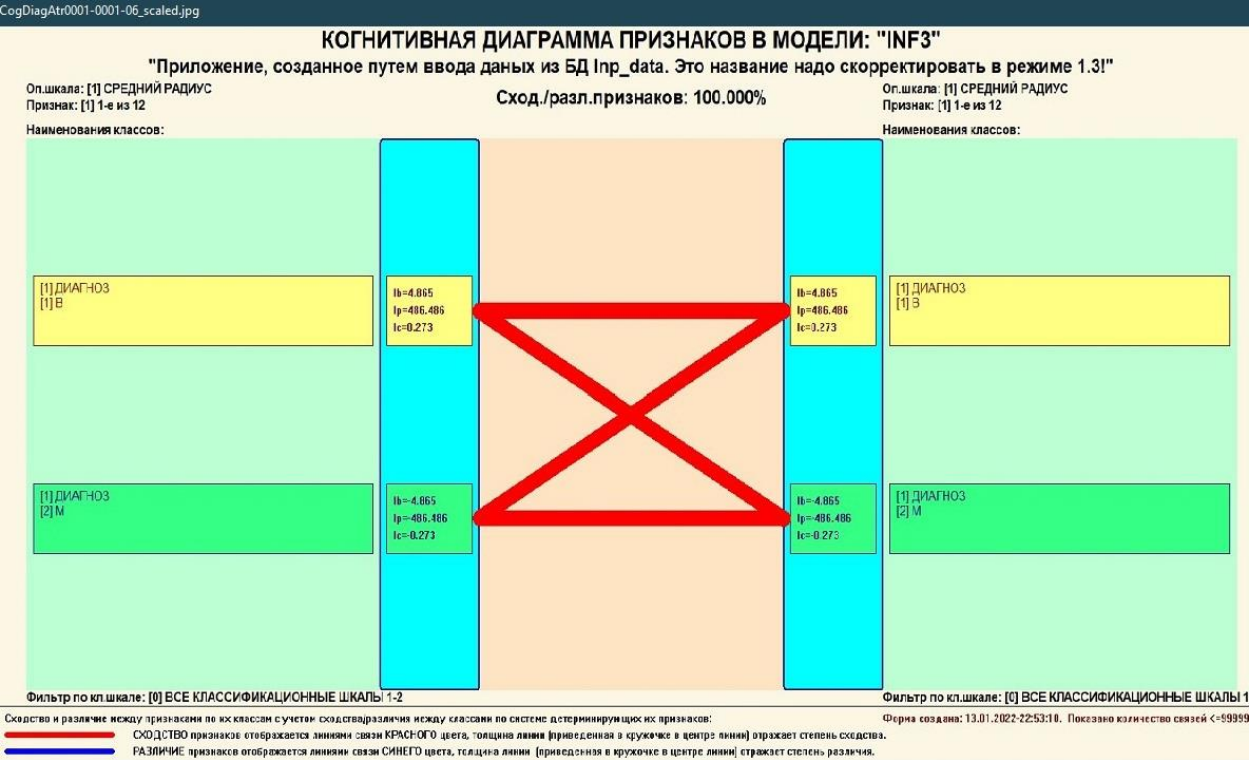


Рисунок 18. Когнитивная дендрограмма признаков в модели болезни

Из рисунков 16 и 18 мы видим, что некоторые признаки болезни сходны по детерминирующей их системе значений характеристик, 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 характеристик, 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 образуют два противоположных кластера, являющихся полюсами конструкта, по системе значений обуславливающих значения параметров их характеристик.

На рисунке 19 мы видим график изменения межкластерных расстояний:

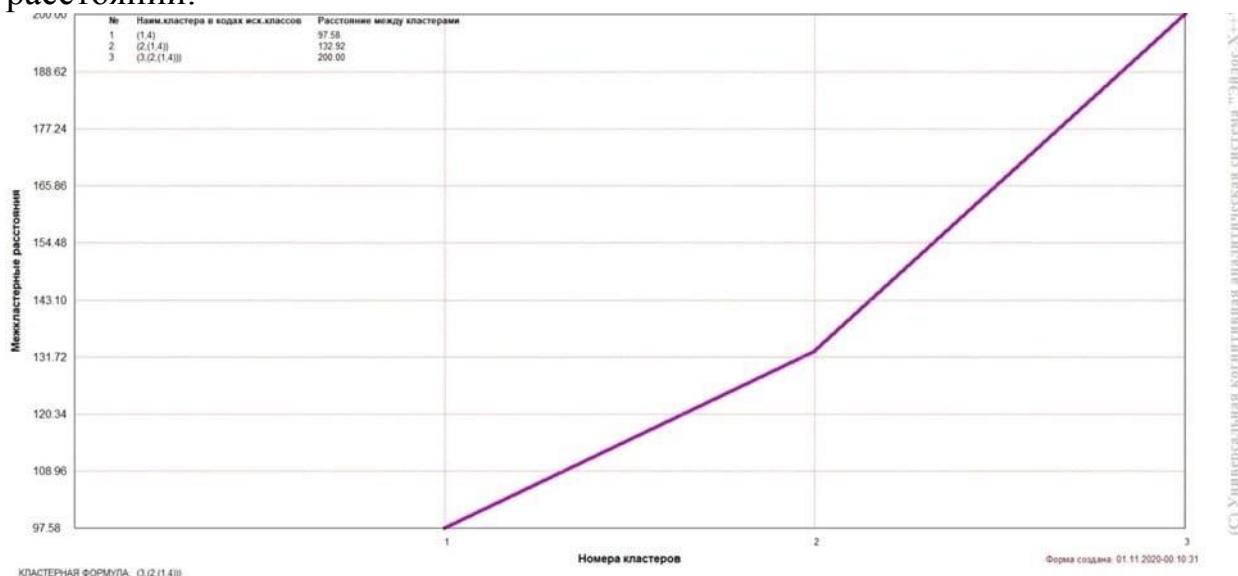


Рисунок 19. График изменения межкластерных расстояний

4.3.3. Когнитивные диаграммы значений факторов

Эти диаграммы отражают сходство/различие значений параметров характеристик по их смыслу, т.е. по содержащейся в них информации о виде заболеваемости. Эти диаграммы мы получаем в режимах 4.3.2.1 и 4.3.2.2

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.

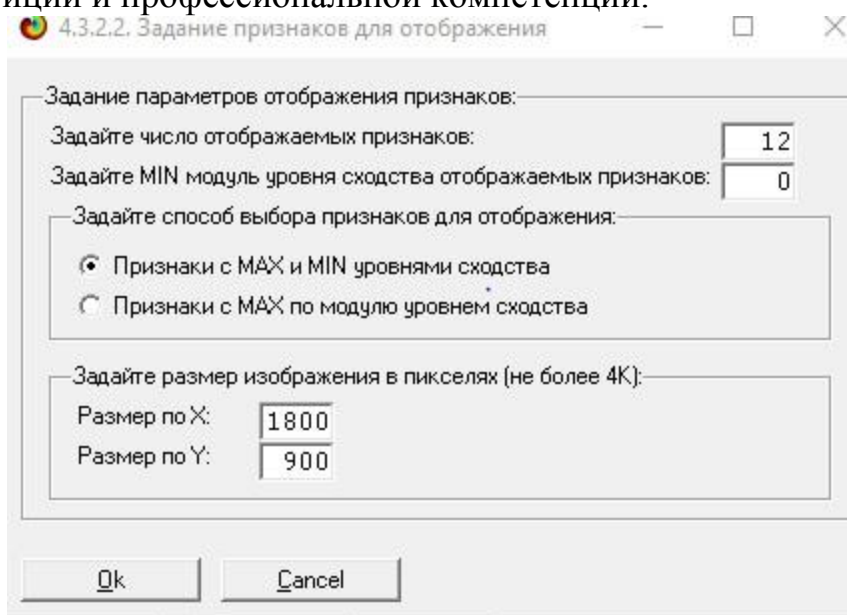


Рисунок 21. Параметры отображения когнитивной диаграммы, приведенной на рисунке 20

4.3.4. Агломеративная когнитивная кластеризация значений факторов

На рисунке 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.

Хорошо видна группировка признаков по их смыслу, т.е. по содержащейся в них информации о классе болезни. *Значения факторов на полюсах конструктора факторов обуславливают переход объекта моделирования в состояния, соответствующие классам, представленным на полюсах конструктора классов (рисунки 18 и 20).*

4.3.5. Нелокальные нейроны и нелокальные нейронные сети

В приведенном фрагменте слоя нейронной сети нейроны соответствуют классу (виду) болезни, а рецепторы – их характеристикам. Нейроны расположены слева направо в порядке убывания силы детерминации, т.е. слева находятся результаты, наиболее жестко обусловленные обуславливающими их значениями факторами, а справа – менее жестко обусловленные.

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, как сейчас говорят «полносвязной».

4.3.6. 3d-интегральные когнитивные карты

На рисунке приведен фрагмент 3d-интегральной когнитивной карты, отражающий фрагмент около 49% 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.

4.3.7. Когнитивные функции

Вместо описания того, что представляют собой когнитивные функции, приведем help соответствующего режима системы «Эйдос» (рисунок 22).

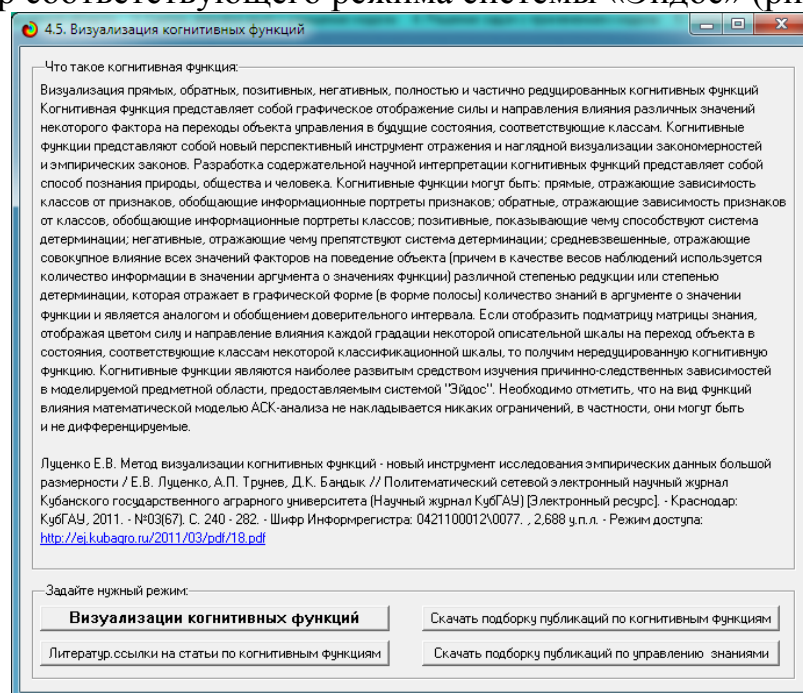


Рисунок 22. Help режима визуализации когнитивных функций

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.

На рисунке 23 приведены когнитивные функции, наглядно отражающие силу и направление влияния значений (т.е. степени выраженности) различных признаков болезни на ее вид (класс).

Мы бы очень хотели сделать эти функции, но нам не позволяет разрешение

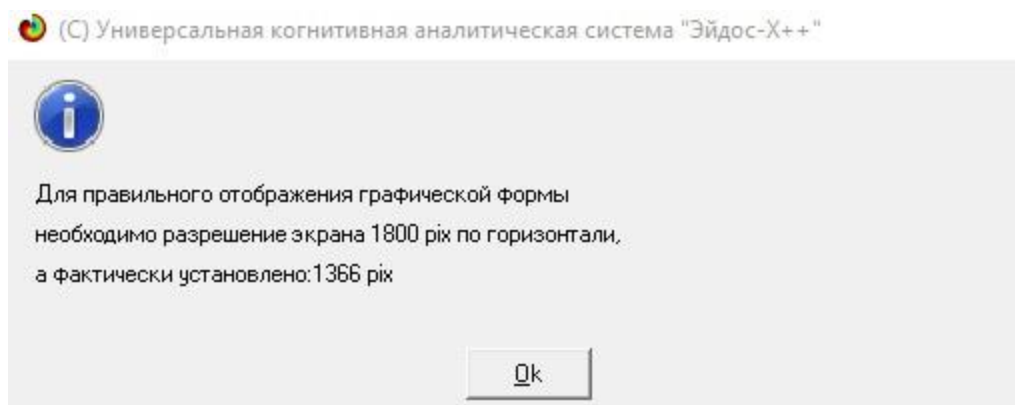


Рисунок 23. Примеры когнитивных функций в СК-модели INF3, отражающих силу и направление влияния значений характеристик болезни на ее вид

Из когнитивных функций, приведенных на рисунке 23, хорошо видно, что *зависимости между значениями характеристиками болезни и ее классом в основном имеют довольно предсказуемый характер.*

4.3.8. Сила и направление влияния значений характеристик болезни и влияния этих характеристик

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

Строки матриц моделей соответствуют значениям факторов, т.е. значениям характеристик болезни (градации описательных шкал).

Колонки матриц моделей соответствуют различным классам, отражающим вид (класс) болезни (градации классификационных шкал).

Числовые значения в ячейках матриц моделей, находящихся на пересечении строк и колонок, отражают направление (знак) и силу влияния конкретного значения характеристики болезни, соответствующего строке, на конкретное значение его класса (вида).

Если какое-то значение характеристики слабо влияет на класс болезни, 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 какой-либо характеристики способствует получению некоторого определенного класса болезни, 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 характеристики болезни 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 болезни.

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 6, 7 содержит количественную оценку вариабельности значений строки модели (среднеквадратичное отклонение), которая и представляет собой ценность значения характеристики болезни, соответствующего строке, для решения задачи квалиметрии и других задач, рассмотренных в работе.

Если рассортировать матрицу модели по этой самой правой колонке в порядке убывания, а потом просуммировать значения в ней нарастающим итогом, то получим логистическую Парето-кривую, отражающую зависимость ценности модели от числа наиболее ценных признаков в ней.

50% наиболее ценных для решения задачи идентификации характеристик болезни обуславливают 65% суммарной ценности, а 50% суммарной ценности обеспечиваются 36% наиболее ценных характеристик болезни.

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

4.3.9. Степень детерминированности класса (вида) заболеваемости

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 у нас классами являются признаки болезни, а значениями градаций описательных шкал – ее характеристики.

Мы видим Парето-кривую степени детерминированности классов значениями характеристик нарастающим итогом.

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.

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 признакам болезни, изучено влияние характеристик болезни, и, на основе этого, решены задачи идентификации, классификации и исследования моделируемой предметной области путем исследования ее модели.

Со всеми моделями, созданными в данной статье, можно ознакомиться установив облачное Эйдос-приложение №308 в режиме 1.3 системы «Эйдос».

Саму систему «Эйдос» можно скачать с сайта автора по ссылке:
http://lc.kubagro.ru/aidos/_Aidos-X.htm.

Список литературы

1. Лойко В.И. Подходы к автоматизации процессов управления производством продукции растениеводства / В.И. Лойко, С.А. Курносов, В.В. Ткаченко, Н.А. Ткаченко // Экономико-правовые аспекты реализации стратегии модернизации России: поиск модели эффективного социоэкономического 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 Ризбергера 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-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: <http://ej.kubagro.ru/2014/07/pdf/90.pdf> 2,688 у.п.л.
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-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.

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>