

УДК 004.8

Автоматизированный системно-когнитивный анализ вероятности появления диабета

Камардина Виктория Романовна
студентка факультета ИСИТ, группы ИТ1821
vikuliakam2000@mail.ru

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

Целью данной работы является реализация возможности диагностического предсказания, а именно, есть ли у пациента диабет или нет, которое делается на основе определенных диагностических измерений. Данная цель является важной с медицинской точки зрения при выявлении не столь очевидных закономерностей при диабете. Читателям данной статья будет полезна для получения знаний в области математики, медицины, изучения математической вероятности и расширения кругозора. Также в данной статье будет показана работа с универсальной когнитивной аналитической системой «Эйдос-Х++». При написании статьи был использован Автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальная система «Эйдос», а также набор данных из Национального института диабета, болезней органов пищеварения, почек.

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

UDC 004.8

Automated system-cognitive analysis of the likelihood of diabetes

Kamardina Viktory Romanovna
student of the faculty of ISIT, group IT1821
vikuliakam2000@mail.ru

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

The aim of this work is to realize the possibility of diagnostic prediction, namely, whether a patient has diabetes or not, which is made on the basis of certain diagnostic measurements. This goal is medically important in identifying less obvious patterns in diabetes. For readers, this article will be useful for gaining knowledge in the field of mathematics, medicine, studying mathematical probability and expanding their horizons. Also in this article will be shown the work with the universal cognitive analytical system "Aidos-X ++". When writing the article, we used the Automated System-Cognitive Analysis (ASK-analysis) and its software toolkit - the intelligent system "Aidos", as well as a set of data from the National Institute of Diabetes, Diseases of the Digestive System, Kidneys.

Key words: AUTOMATED SYSTEM-COGNITIVE ANALYSIS, ASC-ANALYSIS, "AIDOS" SYSTEM, DATA SET

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	2
ЗАДАЧА 1: КОГНИТИВНАЯ СТРУКТУРИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ	6
ЗАДАЧА 2: ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ И ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ	7
ЗАДАЧА 3: СИНТЕЗ И ВЕРИФИКАЦИЯ СТАТИСТИЧЕСКИХ И СИСТЕМНО-КОГНИТИВНЫХ МОДЕЛЕЙ И ВЫБОР НАИБОЛЕЕ ДОСТОВЕРНОЙ ИЗ НИХ ДЛЯ РЕШЕНИЯ ЗАДАЧ	11
ЗАДАЧА 4: РЕШЕНИЕ РАЗЛИЧНЫХ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	17
Подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация)	17
Подзадача 4.2. Поддержка принятия решений (SWOT-анализ)	19
Подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели	21
4.3.1. Когнитивные диаграммы классов.....	22
4.3.2. Когнитивные диаграммы значений факторов	23
4.3.3. Агломеративная когнитивная кластеризация значений факторов	24
4.3.4. Нелокальные нейроны и нелокальные нейронные сети.....	25
4.3.5. 3d-интегральные когнитивные карты	27
4.3.6. Когнитивные функции	27
4.3.7. Сила и направление влияния значений медицинских характеристик и их влияние на результат наличия или отсутствия диабета.....	33
4.3.8. Степень детерминированности результатов	35
7. ВЫВОДЫ	35
СПИСОК ЛИТЕРАТУРЫ	36

Введение

Целью данной работы является реализация возможности автоматизированного диагностического предсказания, а именно, есть ли у пациента диабет или нет, которое делается на основе определенных диагностических измерений.

Достижение данной цели является важной задачей в области медицины для выявления определенных закономерностей. Также достижение данной цели способствует получению знаний в математике и получению навыков работы с универсальной когнитивной аналитической системой «Эйдос-Х++». Данная система позволяет проводить автоматизированный системно-когнитивный анализ (АСК-анализ).

АСК-анализ представляет собой универсальный метод искусственного интеллекта, представляющий собой автоматизированный системный анализ структурированный по базовым когнитивным операциям.

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 предлагается применить для их решения.

АСК-анализ представляет собой метод искусственного интеллекта, разработанный профессором 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-Х++» (система «Эйдос»), которая поддерживается профессором Е.В. Луценко и регулярно обновляется.

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>). Система является автоматизированной, и при решении задач требуется непосредственное участие 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) и лицензией;

- 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 238, 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 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 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.

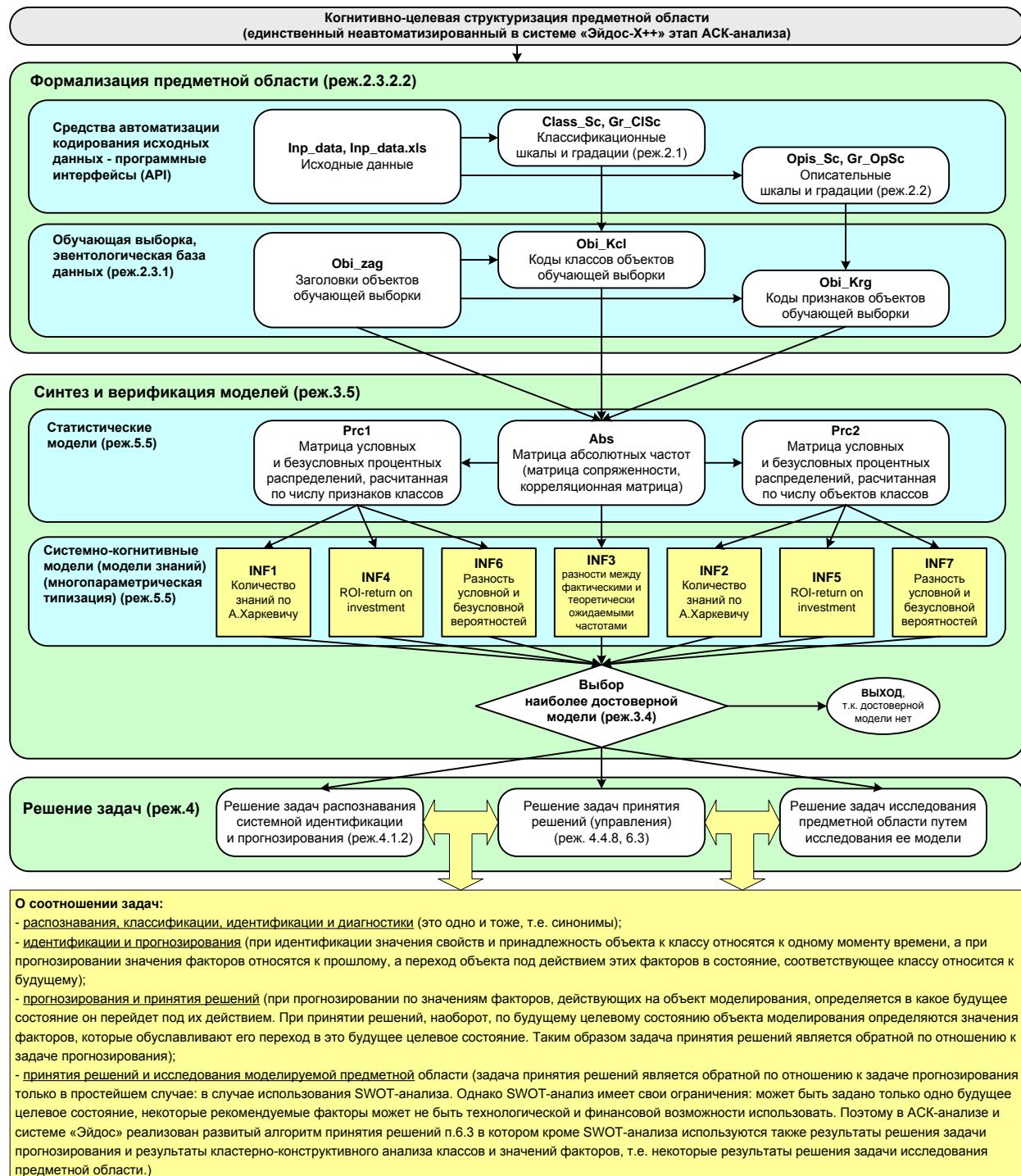
[В чем сила подхода, реализованного в системе Эйдос?](#) В том, что она реализует подход, эффективность которого не зависит от того, что мы думаем о предметной области и думаем ли вообще. Она формирует модели непосредственно на основе эмпирических данных, а не на основе наших представлений о механизмах 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.

Все выше перечисленное является выгодным отличием «Эйдос X++» от других систем искусственного интеллекта. Выбор АСК-анализа и его

программного инструментария – интеллектуальной системы «Эйдос» в качестве метода и инструмента решения поставленной проблемы и достижения цели работы (рисунок 1) Выбор обусловлен доступностью приложения, простотой в использовании, наличием множества инструкций по использованию и понимании полученной информации.

**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 неформализуемым путем решается на качественном уровне, что будет рассматриваться 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 являются причинами, и то, что казалось бы является их последствиями, на самом деле являются 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 (наименований) выберем категорию результат (таблица 1), который указывает на наличие или отсутствие у женщины диабета и заполняется словами «есть» и «нет». А в качестве факторов (описательных шкал), влияющих на эти результаты зададим количество беременностей, концентрацию глюкозы, диастолическое артериальное давление, толщину кожи (кожной складки трицепса), инсулин, ИМТ (индекс массы тела), наличие диабета у родственников, возраст (таблица 2):

Таблица 1 – Классификационная шкала

OUTCOME
есть
нет
есть

Таблица 2 – Описательные шкалы

RENUMBER
PREGNANCIE
GLUCOSE
BLOODPRESS
SKINTHICKN
INSULIN
BMI
DIABETESPE

AGE

Задача 2: подготовка исходных данных и формализация предметной области

Исходные данные для данной статьи (рисунок 2) взяты из Национального института диабета, болезней органов пищеварения и почек. Цель набора данных - диагностически предсказать, есть ли у пациента диабет, на основе определенных диагностических измерений, включенных в набор данных. На выбор этих экземпляров из более крупной базы данных было наложено несколько ограничений. В частности, все пациенты здесь - женщины старше 21 года, принадлежащие к индейцам пима. Наборы данных состоят из нескольких медицинских предикторов и одной целевой переменной Outcome. Переменные-предикторы включают количество беременностей у пациентки, их ИМТ, уровень инсулина, возраст и т. д.

A	B	C	D	E	F	G	H	I	J
RECNUMBER	PREGNANCIE	GLUCOSE	BLOODPRESS	SKINTHICKN	INSULIN	BMI	DIABETESPE	AGE	OUTCOME
1	6	148	72	35	0	33,6	0,627	50	есть
2	1	85	66	29	0	26,6	0,351	31	нет
3	8	183	64	0	0	23,3	0,672	32	есть
4	1	89	66	23	94	28,1	0,167	21	нет
5	0	137	40	35	168	43,1	2,288	33	есть
6	5	116	74	0	0	25,6	0,201	30	нет
7	3	78	50	32	88	31,0	0,248	26	есть
8	10	115	0	0	0	35,3	0,134	29	нет
9	2	197	70	45	543	30,5	0,158	53	есть
10	8	125	96	0	0	0,0	0,232	54	есть
11	4	110	92	0	0	37,6	0,191	30	нет
12	10	168	74	0	0	38,0	0,537	34	есть
13	10	139	80	0	0	27,1	1,441	57	нет
14	1	189	60	23	846	30,1	0,398	59	есть
15	5	166	72	19	175	25,8	0,587	51	есть
16	7	100	0	0	0	30,0	0,484	32	есть
17	0	118	84	47	230	45,8	0,551	31	есть
18	7	107	74	0	0	29,6	0,254	31	есть
19	1	103	30	38	83	43,3	0,183	33	нет
20	1	115	70	30	96	34,6	0,529	32	есть
21	3	126	88	41	235	39,3	0,704	27	нет
22	8	99	84	0	0	35,4	0,388	50	нет
23	7	196	90	0	0	39,8	0,451	41	есть
24	9	119	80	35	0	29,0	0,263	29	есть
25	11	143	94	33	146	36,6	0,254	51	есть
26	10	125	70	26	115	31,1	0,205	41	есть
27	7	147	76	0	0	39,4	0,257	43	есть
28	1	97	66	15	140	23,2	0,487	22	нет
29	13	145	82	19	110	22,2	0,245	57	нет
30	5	117	92	0	0	34,1	0,337	38	нет
31	5	109	75	26	0	36,0	0,546	60	нет
32	3	158	76	36	245	31,6	0,851	28	есть
33	3	88	58	11	54	24,8	0,267	22	нет
34									

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.

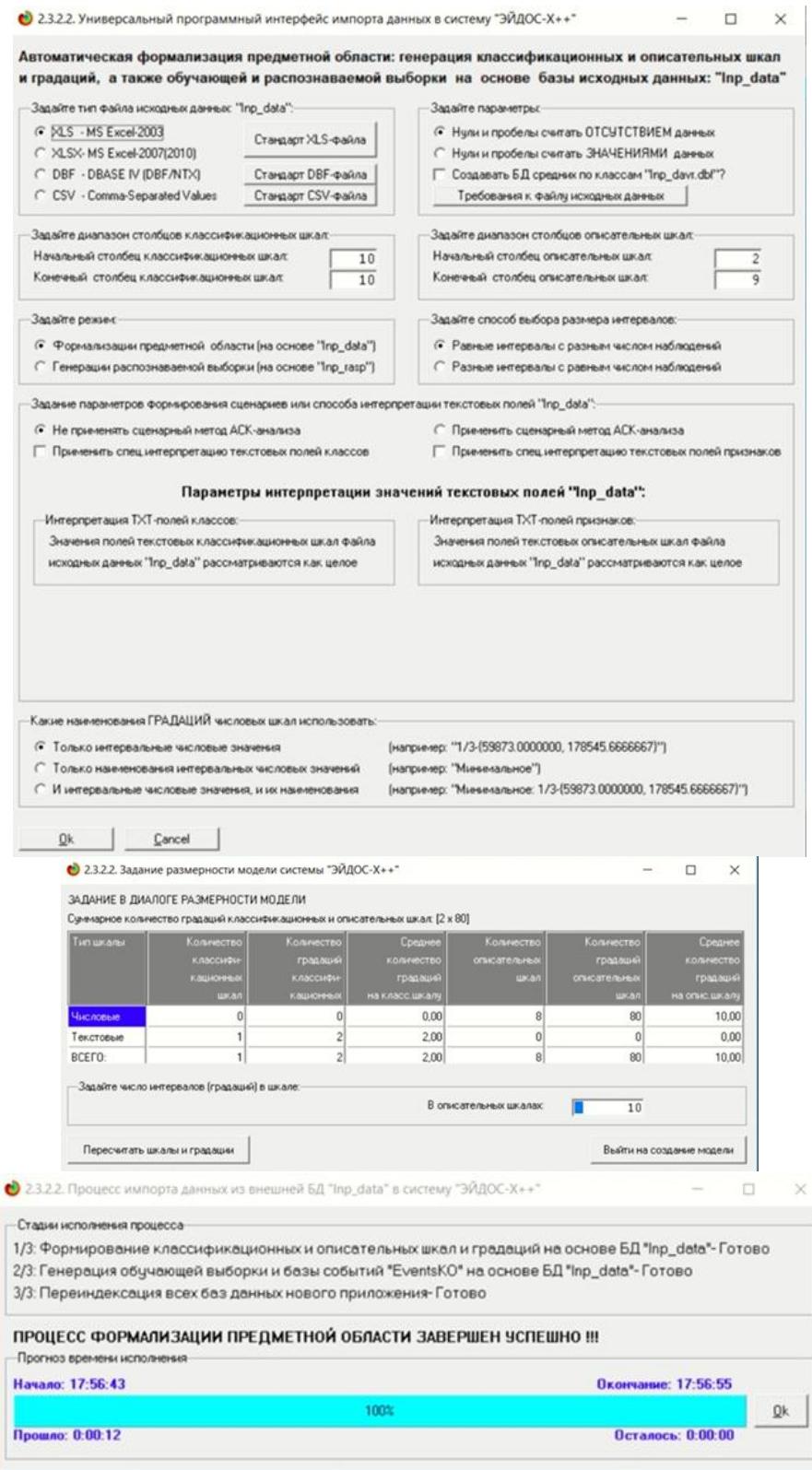


Рисунок 3. Экранные формы программного интерфейса (API) 2.3.2.2 системы «Эйдос» с внешними данными табличного типа

Обратим внимание на то, что заданы равные интервалы с разным числом наблюдений, в них учитывается сильно отличающиеся наблюдения в числовых значениях. В описательных шкалах задано 8 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) 2 (таблица 3) и 9 описательных шкал с большим числом градаций (таблица 4). Результаты работы режима сохраняются в документ INP_NAME.TXT в виде строк с названием описательных и классификационных шкал. С использованием классификационных и описательных шкал и градаций исходные данные (рисунок 2) были закодированы и в результате получена обучающая выборка (рисунок 5). Классами являются будущие состояния объекта управления:

Таблица 3 – Классификационные шкалы и градации
(результат)

Код	OUTCOME
1	Нет
2	Есть

Таблица 4 – Описательные шкалы и градации
(диагностические характеристики)

Код градации	Наименование градации описательной шкалы
1	1/10{1.0000000,2.6000000}
2	2/10-{2.6000000, 4.2000000}
3	3/10-{4.2000000, 5.8000000}
4	4/10-{5.8000000, 7.4000000}
5	5/10-{7.4000000, 9.0000000}
6	6/10-{9.0000000, 10.6000000}
7	7/10-{10.6000000, 12.2000000}
8	8/10-{12.2000000, 3.8000000}

9	9/10-{13.8000000, 15.4000000}
10	10/10-{15.4000000, 17.0000000}

N1	A	B	C	D	E	F	G	H	I	N10
N2		N3	N4	N5	N6	N7	N8	N9		
2	1,0000000	6,0000000	148,0000000	72,0000000	35,0000000	0,0000000	33,0000000	0,0000000	50,0000000	Гбвм
3	2,0000000	1,0000000	85,0000000	66,0000000	29,0000000	0,0000000	26,0000000	0,0000000	31,0000000	Гв
4	3,0000000	8,0000000	183,0000000	64,0000000	0,0000000	0,0000000	23,0000000	0,0000000	32,0000000	Гбвм
5	4,0000000	1,0000000	89,0000000	66,0000000	23,0000000	94,0000000	28,0000000	0,0000000	21,0000000	-Гв
6	5,0000000	0,0000000	137,0000000	40,0000000	35,0000000	168,0000000	43,0000000	2,0000000	33,0000000	Гбвм
7	6,0000000	5,0000000	116,0000000	74,0000000	0,0000000	0,0000000	25,0000000	0,0000000	30,0000000	-Гв
8	7,0000000	3,0000000	78,0000000	50,0000000	32,0000000	88,0000000	31,0000000	0,0000000	26,0000000	Гбвм
9	8,0000000	10,0000000	115,0000000	0,0000000	0,0000000	0,0000000	35,0000000	0,0000000	29,0000000	-Гв
10	9,0000000	2,0000000	197,0000000	70,0000000	45,0000000	543,0000000	30,0000000	0,0000000	53,0000000	Гбвм
11	10,0000000	8,0000000	125,0000000	96,0000000	0,0000000	0,0000000	0,0000000	0,0000000	54,0000000	Гбвм
12	11,0000000	4,0000000	110,0000000	92,0000000	0,0000000	0,0000000	37,0000000	0,0000000	30,0000000	-Гв
13	12,0000000	10,0000000	168,0000000	74,0000000	0,0000000	0,0000000	38,0000000	0,0000000	34,0000000	Гбвм
14	13,0000000	10,0000000	139,0000000	80,0000000	0,0000000	0,0000000	27,0000000	1,0000000	57,0000000	Гв
15	14,0000000	1,0000000	189,0000000	60,0000000	23,0000000	846,0000000	30,0000000	0,0000000	59,0000000	Гбвм
16	15,0000000	5,0000000	166,0000000	72,0000000	19,0000000	175,0000000	25,0000000	0,0000000	51,0000000	Гбвм
17	16,0000000	7,0000000	100,0000000	0,0000000	0,0000000	0,0000000	30,0000000	0,0000000	32,0000000	Гбвм
18	17,0000000	0,0000000	118,0000000	84,0000000	47,0000000	230,0000000	45,0000000	0,0000000	31,0000000	Гбвм
19	18,0000000	7,0000000	107,0000000	74,0000000	0,0000000	0,0000000	29,0000000	0,0000000	31,0000000	Гбвм
20	19,0000000	1,0000000	103,0000000	30,0000000	38,0000000	83,0000000	43,0000000	0,0000000	33,0000000	-Гв
21	20,0000000	1,0000000	115,0000000	70,0000000	30,0000000	96,0000000	34,0000000	0,0000000	32,0000000	Гбвм
22	21,0000000	3,0000000	126,0000000	88,0000000	41,0000000	235,0000000	39,0000000	0,0000000	27,0000000	-Гв
23	22,0000000	8,0000000	99,0000000	84,0000000	0,0000000	0,0000000	35,0000000	0,0000000	50,0000000	-Гв
24	23,0000000	7,0000000	196,0000000	90,0000000	0,0000000	0,0000000	39,0000000	0,0000000	41,0000000	Гбвм
25	24,0000000	9,0000000	119,0000000	80,0000000	35,0000000	0,0000000	29,0000000	0,0000000	29,0000000	Гбвм
26	25,0000000	11,0000000	143,0000000	94,0000000	33,0000000	146,0000000	36,0000000	0,0000000	51,0000000	Гбвм
27	26,0000000	10,0000000	125,0000000	70,0000000	26,0000000	115,0000000	31,0000000	0,0000000	41,0000000	Гбвм
28	27,0000000	7,0000000	147,0000000	76,0000000	0,0000000	0,0000000	39,0000000	0,0000000	43,0000000	Гбвм
29	28,0000000	1,0000000	97,0000000	66,0000000	15,0000000	140,0000000	23,0000000	0,0000000	22,0000000	-Гв
30	29,0000000	13,0000000	145,0000000	82,0000000	19,0000000	110,0000000	22,0000000	0,0000000	57,0000000	-Гв

Рисунок 5 – Обучающая выборка

Обучающая выборка (рисунок 5) 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: для синтеза и верификации статистических и системно-когнитивных 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) моделей осуществляется в режиме 3.5 системы «Эйдос X++» (рисунок 6).

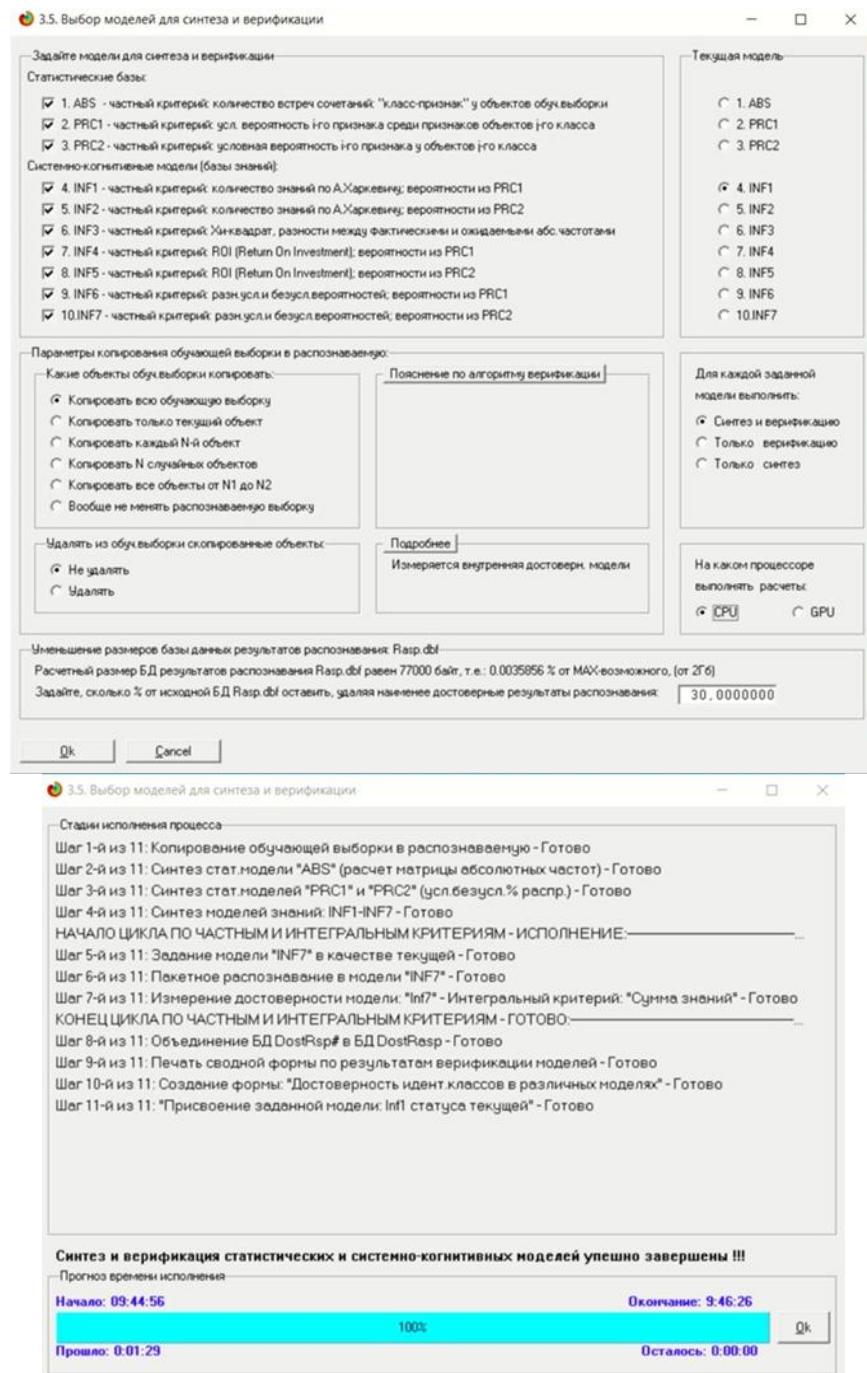


Рисунок 6. Экранные формы режима синтеза и верификации статистических и системно-когнитивных моделей системы «Эйдос»

Обратим внимание на то, что на рисунке 6 в правом нижнем углу окна задана опция: «Расчеты проводить на центральном процессоре (CPU)». Также 30% от исходной модели оставляется, и удаляются наименее достоверные результаты распознания

Из рисунка 6 видно, что весь процесс синтеза и верификации моделей занял 9.46 секунд. Отметим, что при синтезе и верификации моделей использовался центральный процессор (CPU) и выполнение этих операций на нем обычно занимает значительно большее время, чем GPU (на некоторых задачах это происходит в десятки, сотни и даже тысячи раз дольше), но в данной работе представлена небольшая сжатая база данных. Таким образом,

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 7, 8:

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

Код признака	Наименование описательной шкалы и градации	1. OUTCOME ЕСТЬ	2. OUTCOME НЕТ	Сумма	Среднее	Средн. квадр. откл.
1	PREGNANCIE-1/10-(1.0000000, 2.6000000)	48	190	238	119.00	100.41
2	PREGNANCIE-2/10-(2.6000000, 4.2000000)	50	93	143	71.50	30.41
3	PREGNANCIE-3/10-(4.2000000, 5.8000000)	21	36	57	28.50	10.61
4	PREGNANCIE-4/10-(5.8000000, 7.4000000)	41	54	95	47.50	9.19
5	PREGNANCIE-5/10-(7.4000000, 9.0000000)	40	26	66	33.00	9.90
6	PREGNANCIE-6/10-(9.0000000, 10.6000000)	10	14	24	12.00	2.83
7	PREGNANCIE-7/10-(10.6000000, 12.2000000)	11	9	20	10.00	1.41
8	PREGNANCIE-8/10-(12.2000000, 13.8000000)	5	5	10	5.00	
9	PREGNANCIE-9/10-(13.8000000, 15.4000000)	3		3	1.50	2.12
10	PREGNANCIE-10/10-(15.4000000, 17.0000000)	1		1	0.50	0.71
11	GLUCOSE-1/10-(44.0000000, 59.5000000)		4	4	2.00	2.83
12	GLUCOSE-2/10-(59.5000000, 75.0000000)		21	21	10.50	14.85
13	GLUCOSE-3/10-(75.0000000, 90.5000000)	7	78	85	42.50	50.20
14	GLUCOSE-4/10-(90.5000000, 106.0000000)	24	139	163	81.50	81.32
15	GLUCOSE-5/10-(106.0000000, 121.5000000)	42	105	147	73.50	44.55
16	GLUCOSE-6/10-(121.5000000, 137.0000000)	53	80	133	66.50	19.09
17	GLUCOSE-7/10-(137.0000000, 152.5000000)	41	39	80	40.00	1.41
18	GLUCOSE-8/10-(152.5000000, 168.0000000)	38	20	58	29.00	12.73
19	GLUCOSE-9/10-(168.0000000, 183.5000000)	34	6	40	20.00	19.80
20	GLUCOSE-10/10-(183.5000000, 199.0000000)	27	5	32	16.00	15.56
21	BLOODPRESS-1/10-(24.0000000, 33.8000000)	1	2	3	1.50	0.71
22	BLOODPRESS-2/10-(33.8000000, 43.6000000)	1	1	2	1.00	
23	BLOODPRESS-3/10-(43.6000000, 53.4000000)	9	26	35	17.50	12.02
24	BLOODPRESS-4/10-(53.4000000, 63.2000000)	22	96	118	59.00	52.33

Рисунок 7. 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. OUTCOME ЕСТЬ	2. OUTCOME НЕТ	Сумма	Среднее	Средн. квадр. откл.
1	PREGNANCIE-1/10-(1.0000000, 2.6000000)	-34.267	34.267			48.461
2	PREGNANCIE-2/10-(2.6000000, 4.2000000)	0.571	-0.571			0.807
3	PREGNANCIE-3/10-(4.2000000, 5.8000000)	1.297	-1.297			1.835
4	PREGNANCIE-4/10-(5.8000000, 7.4000000)	8.162	-8.162			11.543
5	PREGNANCIE-5/10-(7.4000000, 9.0000000)	17.186	-17.186			24.305
6	PREGNANCIE-6/10-(9.0000000, 10.6000000)	1.704	-1.704			2.410
7	PREGNANCIE-7/10-(10.6000000, 12.2000000)	4.087	-4.087			5.780
8	PREGNANCIE-8/10-(12.2000000, 13.8000000)	1.543	-1.543			2.183
9	PREGNANCIE-9/10-(13.8000000, 15.4000000)	1.963	-1.963			2.776
10	PREGNANCIE-10/10-(15.4000000, 17.0000000)	0.654	-0.654			0.925
11	GLUCOSE-1/10-(44.0000000, 59.5000000)	-1.383	1.383			1.955
12	GLUCOSE-2/10-(59.5000000, 75.0000000)	-7.259	7.259			10.266
13	GLUCOSE-3/10-(75.0000000, 90.5000000)	-22.381	22.381			31.652
14	GLUCOSE-4/10-(90.5000000, 106.0000000)	-32.343	32.343			45.739
15	GLUCOSE-5/10-(106.0000000, 121.5000000)	-8.812	8.812			12.462
16	GLUCOSE-6/10-(121.5000000, 137.0000000)	7.027	-7.027			9.938
17	GLUCOSE-7/10-(137.0000000, 152.5000000)	13.347	-13.347			18.876
18	GLUCOSE-8/10-(152.5000000, 168.0000000)	17.952	-17.952			25.388
19	GLUCOSE-9/10-(168.0000000, 183.5000000)	20.174	-20.174			28.530
20	GLUCOSE-10/10-(183.5000000, 199.0000000)	15.939	-15.939			22.541
21	BLOODPRESS-1/10-(24.0000000, 33.8000000)	-0.037	0.037			0.052
22	BLOODPRESS-2/10-(33.8000000, 43.6000000)	0.309	-0.309			0.437
23	BLOODPRESS-3/10-(43.6000000, 53.4000000)	-3.098	3.098			4.381
24	BLOODPRESS-4/10-(53.4000000, 63.2000000)	-18.788	18.788			26.570

Рисунок 8. 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 образом 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 9).

3.4. Обобщ форма по достоверности моделей при разнокрит. Текущая модель: "INF1"											
Наименование модели и частного критерия	Интегральный критерий	F-мера Ван Ризбергена	Сумма моделей, уровней сход. истинно-полож. решений (STP)	Сумма моделей, уровней сход. истинно-отриц. решений (STN)	Сумма моделей, уровней сход. ложно-полож. решений (SFP)	Сумма моделей, уровней сход. ложно-отриц. решений (SFN)	S-Точность модели	S-Полнота модели	L1-мера проф. Е.В.Луценко	Средний модуль уровней сход. истинно-полож. решений	Средний модуль уровней сход. истинно-отриц. решений
1. ABS - частный критерий: количество встреч соединений "ханс." с диагнозом диабетом	Корреляция abs. частот с общ. суммой	0.869	410.882	0.076	92.260	0.170	0.817	1.000	0.899	0.537	0.038
1. ABS - частный критерий: количество встреч соединений "ханс." с диагнозом диабетом	Сумма abs. частот по признаку	0.870	384.764	76.398	0.834	1.000	0.910	0.501	0.501	0.537	0.038
2. PRC1 - частный критерий: усл. вероятность иго признака сред. от частоты	Корреляция усл. от частот с о. вероятн. иго признака	0.869	410.882	0.076	92.260	0.170	0.817	1.000	0.899	0.537	0.038
2. PRC1 - частный критерий: усл. вероятность иго признака сред. от частоты	Сумма усл. от частот по признаку	0.870	446.565	108.732	0.804	1.000	0.891	0.581	0.581	0.537	0.038
3. PRC2 - частный критерий: условная вероятность иго признака	Корреляция усл. от частот с о. вероятн. иго признака	0.869	410.880	0.076	92.259	0.170	0.817	1.000	0.899	0.537	0.038
3. PRC2 - частный критерий: условная вероятность иго признака	Сумма усл. от частот по признаку	0.870	444.673	107.743	0.805	1.000	0.892	0.579	0.579	0.537	0.038
4. INF1 - частный критерий: количество знаний по АХаркевичу, в. Семантический резонанс зна.	Семантический резонанс зна.	0.812	127.081	58.011	6.130	28.758	0.954	0.815	0.879	0.227	0.328
4. INF1 - частный критерий: количество знаний по АХаркевичу, в. Сумма знаний	Сумма знаний	0.762	64.943	48.973	4.441	24.303	0.936	0.728	0.819	0.129	0.272
5. INF2 - частный критерий: количество знаний по АХаркевичу, в. Семантический резонанс зна.	Семантический резонанс зна.	0.813	127.389	58.013	6.163	28.815	0.954	0.816	0.879	0.227	0.326
5. INF2 - частный критерий: количество знаний по АХаркевичу, в. Сумма знаний	Сумма знаний	0.773	64.005	49.044	4.127	23.754	0.939	0.729	0.821	0.125	0.269
6. INF3 - частный критерий: Хинквадрат, разности между фактами. Семантический резонанс зна.	Семантический резонанс зна.	0.807	188.562	56.979	9.472	30.905	0.952	0.859	0.903	0.334	0.345
6. INF3 - частный критерий: Хинквадрат, разности между фактами. Сумма знаний	Сумма знаний	0.807	178.416	53.727	8.844	28.775	0.953	0.861	0.905	0.316	0.326
7. INF4 - частный критерий: ROI (Return On Investment); вероятн. Семантический резонанс зна.	Семантический резонанс зна.	0.738	161.078	84.641	5.658	66.583	0.966	0.708	0.817	0.343	0.434
7. INF4 - частный критерий: ROI (Return On Investment); вероятн. Сумма знаний	Сумма знаний	0.807	133.509	43.543	10.524	21.589	0.927	0.861	0.893	0.238	0.261
8. INF5 - частный критерий: ROI (Return On Investment); вероятн. Семантический резонанс зна.	Семантический резонанс зна.	0.739	161.479	84.525	5.701	66.561	0.966	0.708	0.817	0.343	0.433
8. INF5 - частный критерий: ROI (Return On Investment); вероятн. Сумма знаний	Сумма знаний	0.811	135.582	44.995	10.094	21.551	0.931	0.863	0.895	0.239	0.271
9. INF6 - частный критерий: разн. усл и без усл. вероятностей; вер. Семантический резонанс зна.	Семантический резонанс зна.	0.802	187.420	57.623	9.337	31.744	0.953	0.855	0.901	0.336	0.349
9. INF6 - частный критерий: разн. усл и без усл. вероятностей; вер. Сумма знаний	Сумма знаний	0.807	119.360	55.167	7.238	26.563	0.943	0.818	0.876	0.212	0.334
10.INF7 - частный критерий: разн. усл и без усл. вероятностей; вер. Семантический резонанс зна.	Семантический резонанс зна.	0.803	188.189	57.695	9.277	31.721	0.953	0.856	0.902	0.336	0.354
10.INF7 - частный критерий: разн. усл и без усл. вероятностей; вер. Сумма знаний	Сумма знаний	0.803	119.761	55.589	6.962	27.204	0.945	0.815	0.875	0.213	0.343

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

Из рисунка 9 видно, что в данном интеллектуальном приложении по F-критерию Ван Ризбергена наиболее достоверной является СК-модель INF3 с интегральным критерием «Сумма знаний» ($F=0,807$ при максимуме 1 000), что является довольно хорошим результатом, по критерию L1 проф. Е.В.Луценко та же модель является наиболее достоверной ($L1=0,892$ при максимуме 1 000). Это является отличным результатом.

Это подтверждает наличие и адекватное отражение в СК-модели INF3 сильной причинно-следственной зависимости между результатом наличия или отсутствия диабета и медицинских характеристик, отраженных в базе данных.

На рисунке 10 приведено частотное распределение числа истинных и ложных положительных и отрицательных решений по результатам идентификации характеристик результата наличия или отсутствия диабета в СК-модели INF3 по данным обучающей выборки:

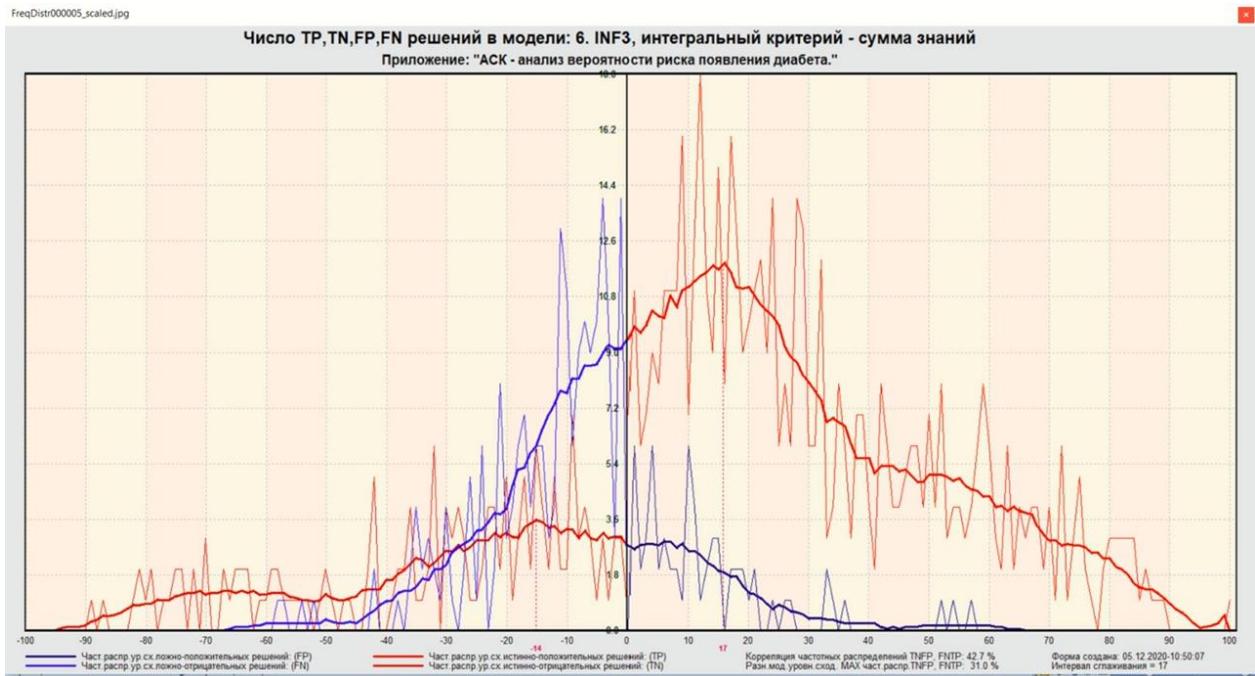


Рисунок 10. Частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности в СК-модели INF3

Рисунок 10 содержит изображения двух частотных распределений похожих на нормальные, которые сдвинуты относительно друг друга по фазе.

Правое распределение, большее по амплитуде включает истинно-отрицательные и ложно-положительные решения, а левое, 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 результата наличия или отсутствия диабета по определенным характеристикам, обозначенным в базе данных, и другие задачи.

Видно, что для отрицательных решений количество истинных решений всегда значительно превосходит количество ложных решений, причем при уровнях различия больше примерно 10% 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 68% 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 68% ложные положительные решения не встречаются.

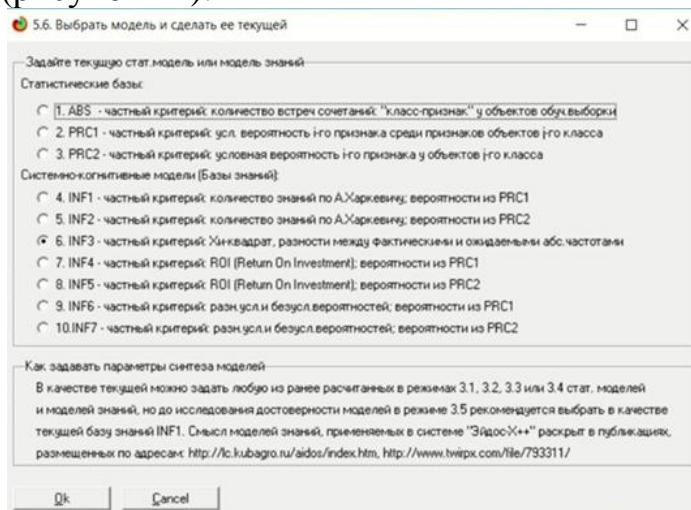
На рисунке 11 приведен Help по режиму 3.4, в котором описаны меры достоверности моделей, применяемые в системе «Эйдос»:



Рисунок 11. 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 INF3 статус текущей модели. Для этого запустим режим 5.6 с параметрами, приведенными на экранной форме (рисунок 12):



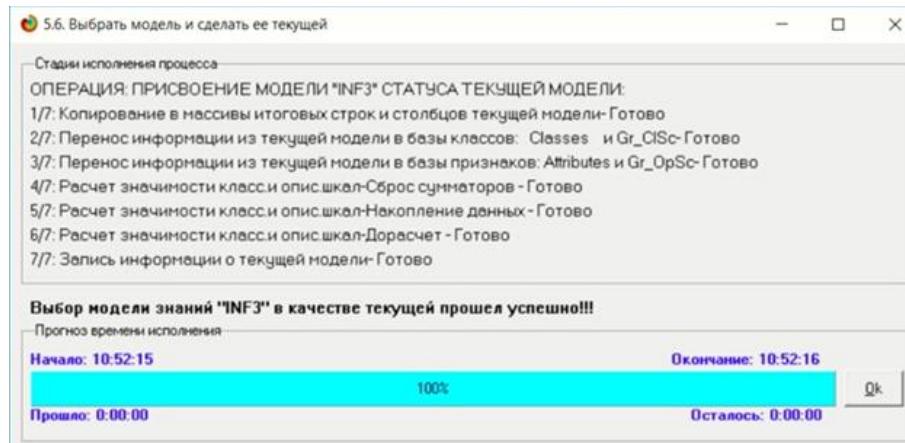


Рисунок 12. Экранные формы приятия наиболее достоверной СК-модели Inf3 статуса текущей модели

Задача 4: решение различных задач в наиболее достоверной модели

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

Решим задачу системной идентификации, т.е. определение результата на основе обучающей выборки в наиболее достоверной СК-модели INF3 на GPU (рисунок 13).

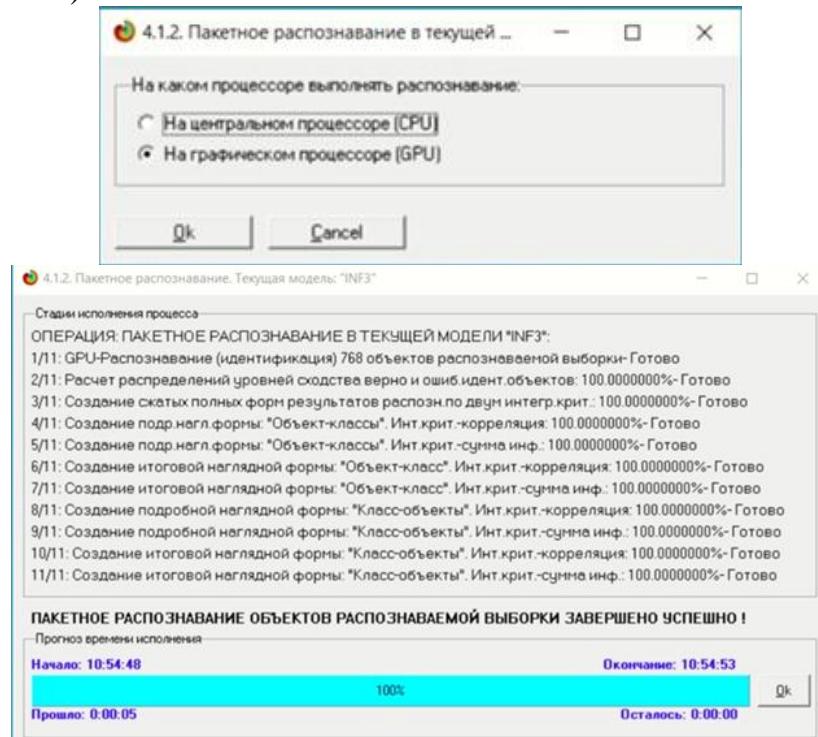


Рисунок 13. Экранные формы, которые отображают процесс решения задачи системной идентификации в текущей модели

Из рисунка 13 видно, что процесс идентификации занял 5 секунд.

Отметим, что 99% этого времени заняла не сама идентификация на CPU, а создание выходных форм на основе результатов этого прогнозирования. Эти формы отражают результаты прогнозирования в различных разрезах и обобщениях:

Приведем две из этих форм: 4.1.3.1 и 4.1.3.2 (рисунок 14).

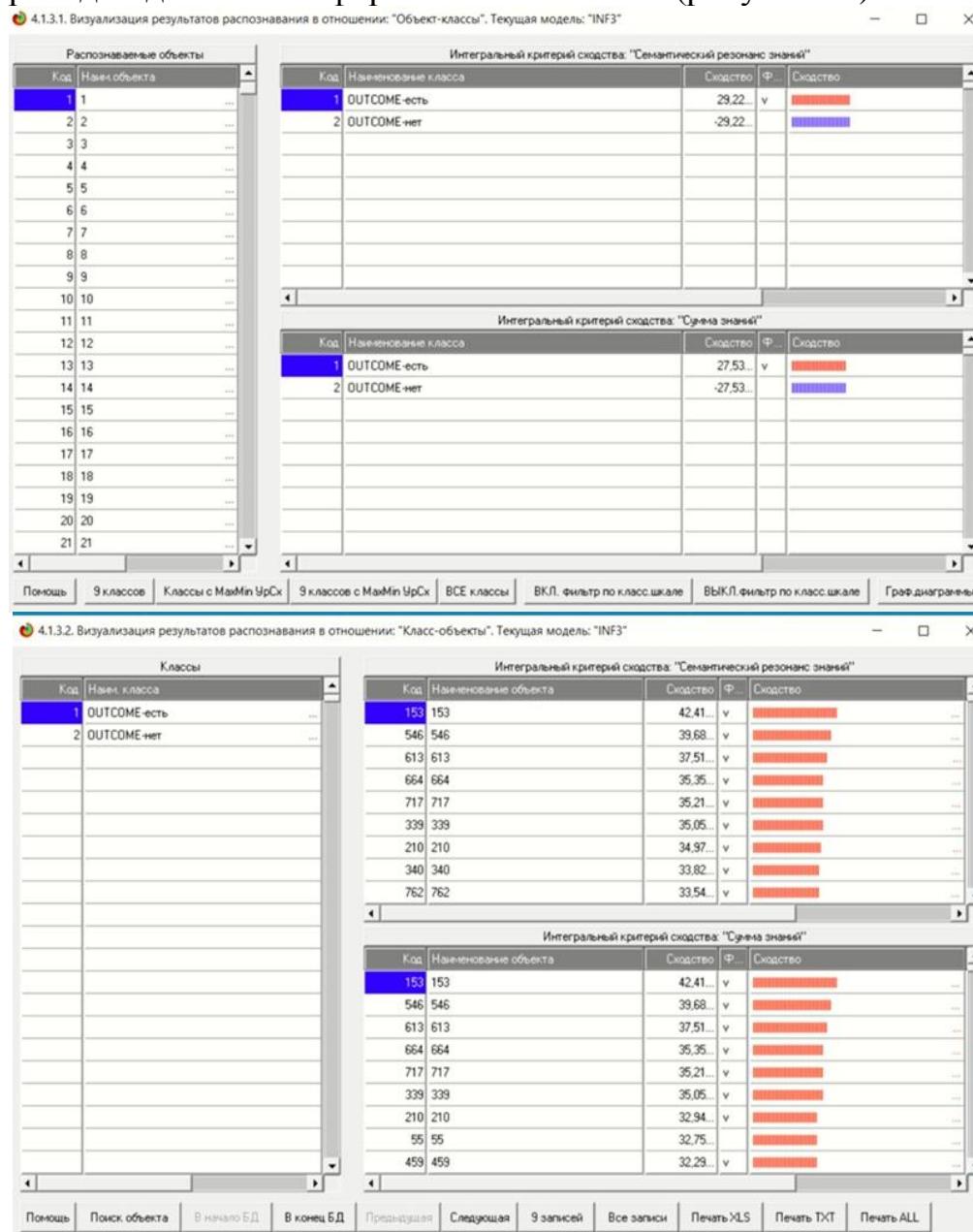


Рисунок 14. Выходные формы по результатам наличия и отсутствия диабета по его характеристикам

Символ « \vee » стоит против тех результатов идентификации, которые подтвердились на опыте, т.е. соответствуют факту. Из рисунка 14 видно, что результаты идентификации являются отличными, естественно при учете информации из рисунка 10 о том, что достоверные прогнозы в данной модели имеют уровень сходства выше 42%, т.е., по сути, результаты с более низким 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 на результат наличия или отсутствия диабета.

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 15 приведены SWOT-диаграммы наглядно отражающие силу и направление влияния различных значений медицинских показателей на вероятность наличия риска диабета.

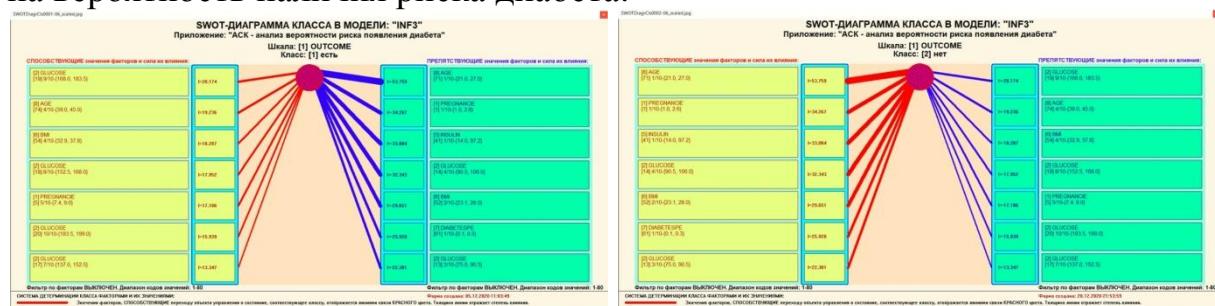


Рисунок 15. SWOT-диаграммы детерминации диагностических характеристик

Эти SWOT-диаграммы наглядно отражают силу и направление влияния различных значений медицинских показателей на вероятность наличия риска диабета

Отметим также, что система «Эйдос» обеспечивала решение этой задачи *всегда*, т.е. даже в самых ранних DOS-версиях при реализации системы «Эйдос» и даже на других языках и различных компьютерах с разной архитектурой о операционной системой. Далее рассмотрен первый акт внедрения системы «Эйдос», где упоминается в явном виде об этом решении, который датируется 1987 годом. Первый же подобный расчет относится к 1981 году.

<p style="text-align: center;">У Т В Е Р Ж Д А Ъ</p> <p>Заведующий Краснодарским сектором ИСИ АН СССР, к.ф.н. <i>А.Хагуров</i> 1987г.</p>	<p style="text-align: center;">У Т В Е Р Ж Д А Ъ</p> <p>Директор Северо-Кавказского филиала ВНИЦ "АИУС-агроресурсы", к.э.н. <i>З.М.Трахов</i> 1987г.</p>
  <p style="font-size: small; margin-top: -10px;">А К Т</p>	
<p>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 ра- боты:</p> <ul style="list-style-type: none"> - осуществлена постановка задачи: "Обработка на ЭВМ социологических анкет Крайагропрома"; - разработаны математическая модель и программное обеспечение под- системы распознавания образов, позволяющие решать данную задачу в среде персональной технологической системы ВЕГА-М; - на профессиональной персональной ЭВМ "Искра-226" осуществлены рас- четы по задаче в объёме: <p>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:</p> <ul style="list-style-type: none"> - процентное распределение ответов в разрезе по социальным типам корреспондентов; - распределение информативностей признаков (в битах) для распозна- вания 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). <p>Работы выполнены на высоком научно-методическом уровне и в срок.</p> <p>От ИСИ АН СССР: Мл. научный сотрудник <i>М.М.Кириченко</i> <i>19.05.1987г.</i></p> <p>От СКФ ВНИЦ "АИУС-агроресурсы": Зав. отделом аэрокосмических и тематических изысканий №4, к.э.н. <i>Г.А.Самсонов</i> <i>19.05.</i> <i>1987г.</i></p> <p>Главный конструктор проекта <i>В.И.Коренец</i> <i>19.05.82.</i> <i>1987г.</i></p> <p>Главный конструктор проекта <i>Е.В.Луценко</i> <i>19.05.87</i> <i>1987г.</i></p>	

Рисунок 16. Акт по обработке анкет

Но тогда SWOT-диаграммы назывались позитивными и негативными 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 15, но и во многих других табличных и графических формах. В этих формах может быть выведена значительно более полная информация (в т. ч. вообще вся имеющаяся 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 17 приведены примеры нескольких инвертированных SWOT-диаграмм, отражающих силу и направление влияния различных результатов на его диагностические характеристики.

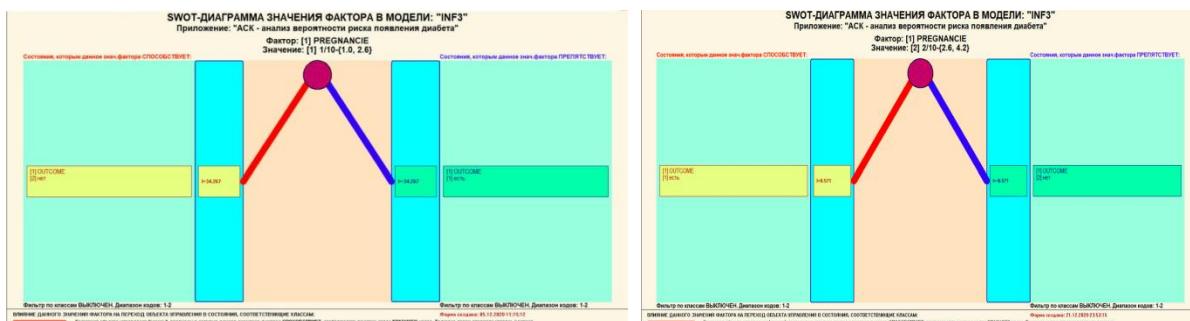


Рисунок 17. Примеры SWOT-диаграмм, отражающих силу и направление влияния показателей на результат наличия или отсутствия диабета

Наибольшее влияние, как оказалось имеет 2-часовой сывороточный инсулин и количество беременностей.

В заключение отметим, что SWOT-анализ является широко известным и общепризнанным методом стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых чаще всего является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего опыта и профессиональной компетенции. Но возможности экспертов имеют свои ограничения и часто по различным причинам они не могут и не хотят это сделать. Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов. Эта проблема может решаться путем автоматизации функций экспертов, т.е. путем измерения силы и направления влияния факторов непосредственно на основе эмпирических данных. Подобная технология разработана более 30 лет назад, но, к сожалению, она сравнительно малоизвестна.

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 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 18, показаны **количественные** оценки сходства/различия результатов 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.

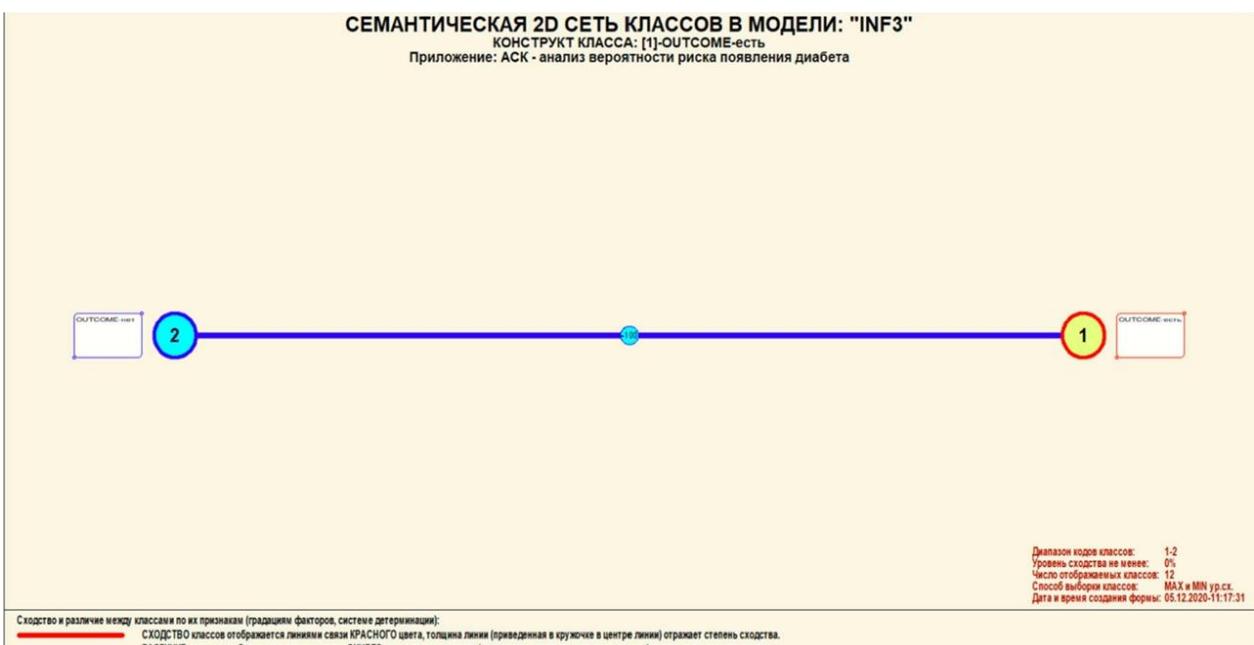


Рисунок 18. Когнитивная диаграмма классов, отражающая сходства/различия результатов по связанным с ними значениям их характеристик.

В системе «Эйдос» есть возможность управлять параметрами формирования и вывода изображения, приведенного на рисунке 18. Для этого используется диалоговое окно, приведенное на рисунке 19.

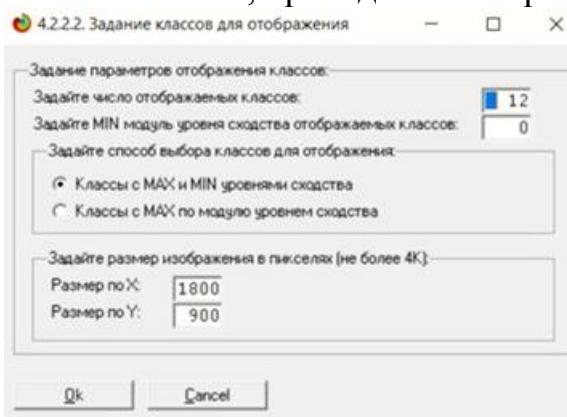


Рисунок 19. Диалоговое окно управления параметрами формирования и вывода изображения когнитивной диаграммы классов

4.3.2. Когнитивные диаграммы значений факторов

Эти диаграммы отражают сходство/различие значений параметров характеристик по их смыслу, т.е. по содержащейся в них информации о результатах. Эти диаграммы мы получаем в режимах 4.3.2.1 и 4.3.2.2 (рисунок 20).

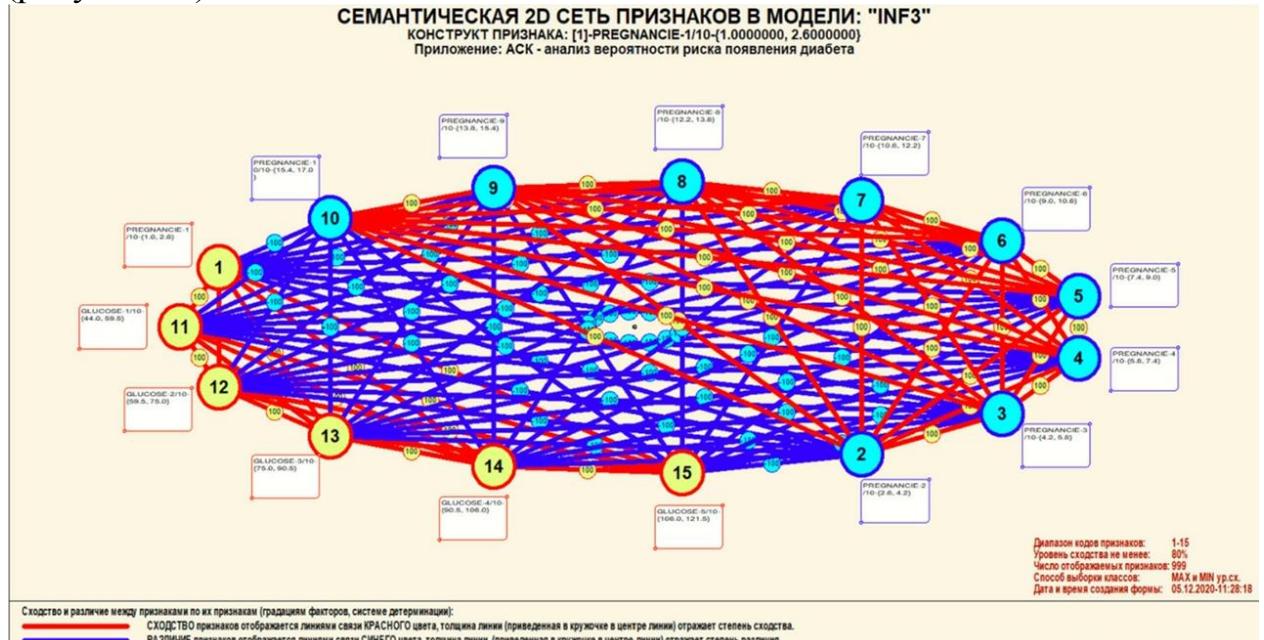


Рисунок 20. сходство/различие результатов в соответствии с их характеристиками

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

Отметим, что на когнитивной диаграмме, приведенной на рисунке 20, показаны **количественные** оценки сходства/различия значений факторов, полученные с применением системно-когнитивной модели, созданной *непосредственно на основе эмпирических данных*, а не как традиционно делается на основе экспертных оценок неформализуемым путем на основе опыта, интуиции и профессиональной компетенции.

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

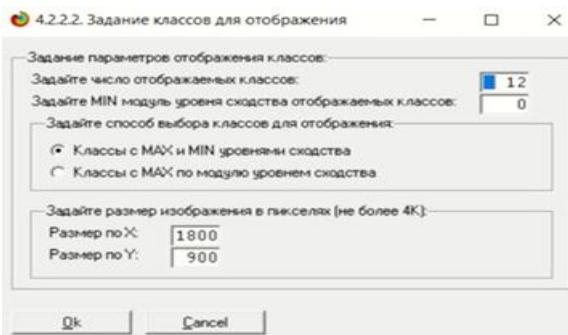
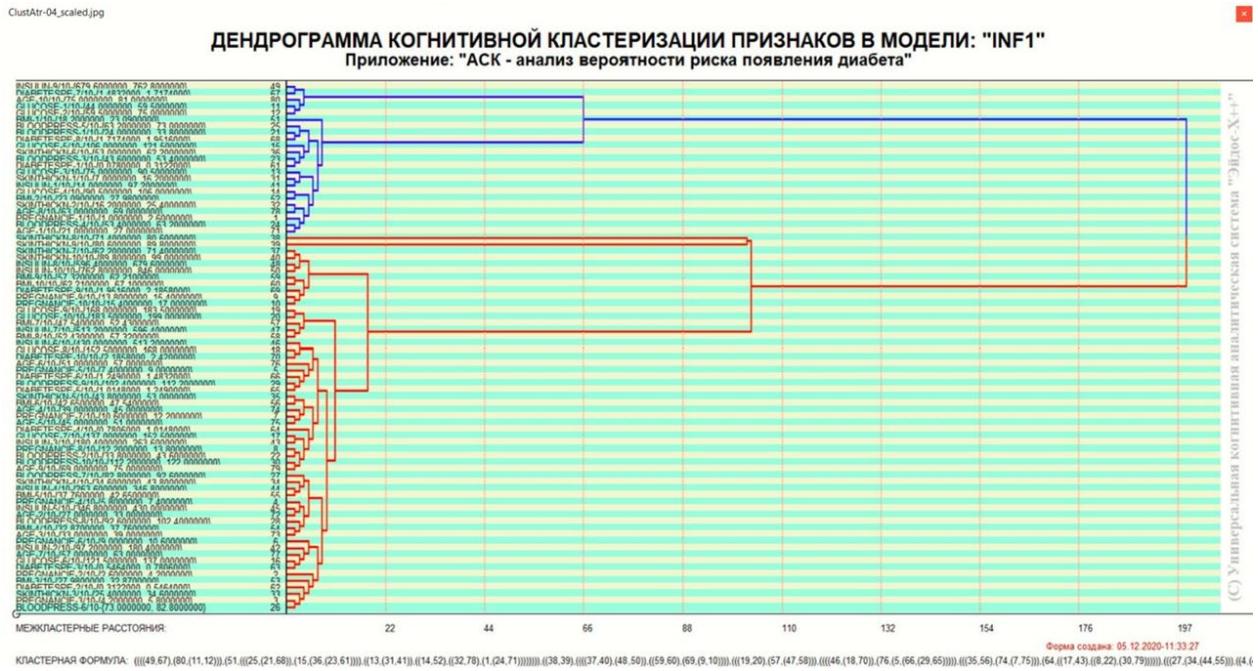


Рисунок 21. Параметры отображения когнитивной диаграммы, приведенной на рисунке 20

4.3.3. Агломеративная когнитивная кластеризация значений факторов

На рисунке 22 приведена агломеративная дендрограмма когнитивной кластеризации значений факторов и график изменения межклusterных расстояний, полученные на основе той же матрицы сходства признаков по их смыслу, что и в когнитивных диаграммах, пример которой приведен на рисунке 20.



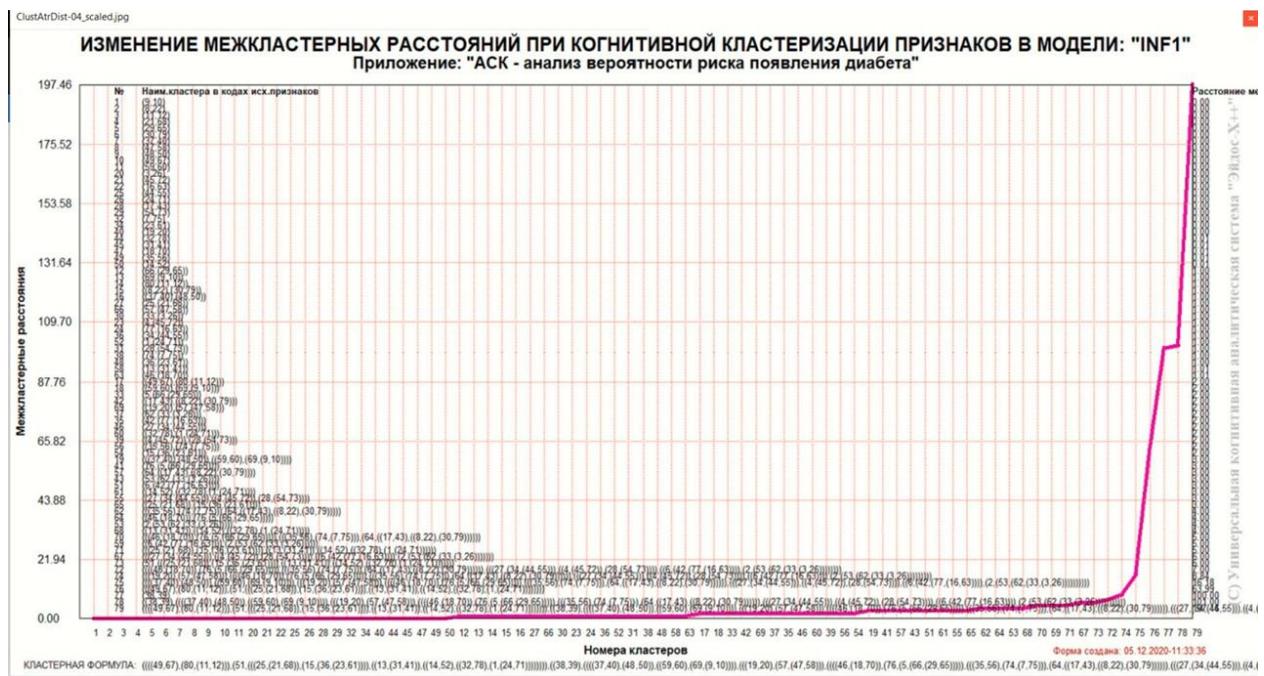


Рисунок 23. График изменения межкластерных расстояний при когнитивной кластеризации значений факторов

4.3.4. Нелокальные нейроны и нелокальные нейронные сети

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

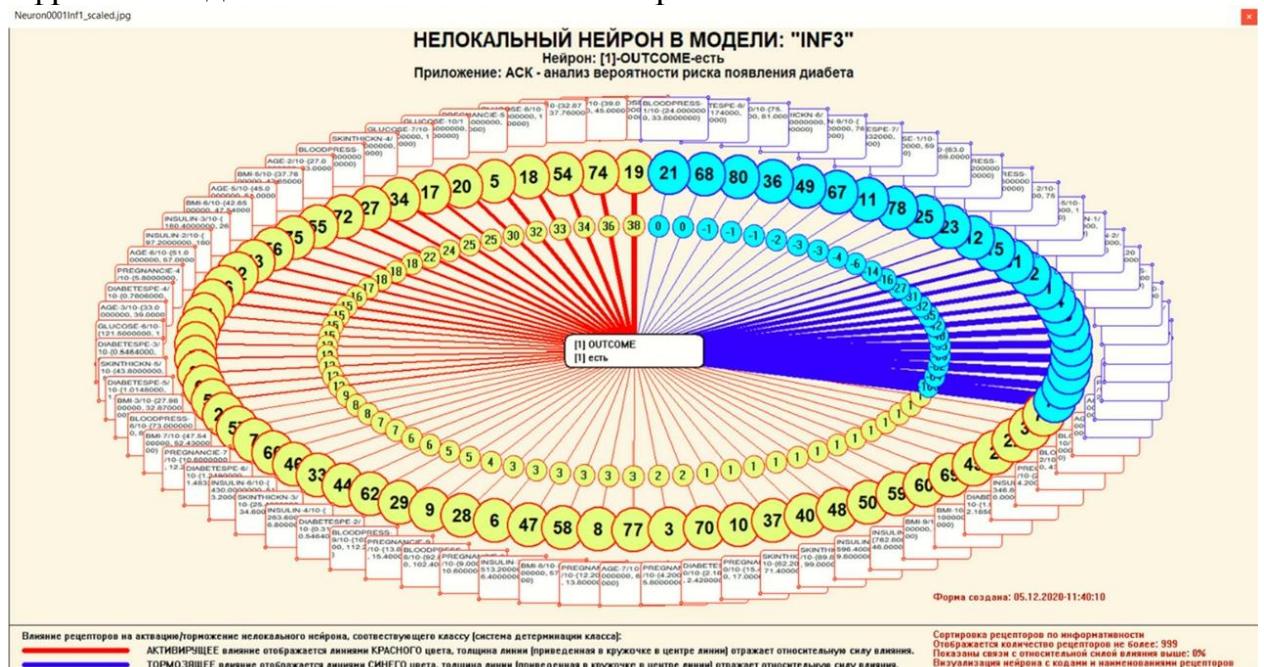


Рисунок 24. Пример нелокального нейрона, отражающего силу и направление влияния различных характеристик на наличие или отсутствие диабета

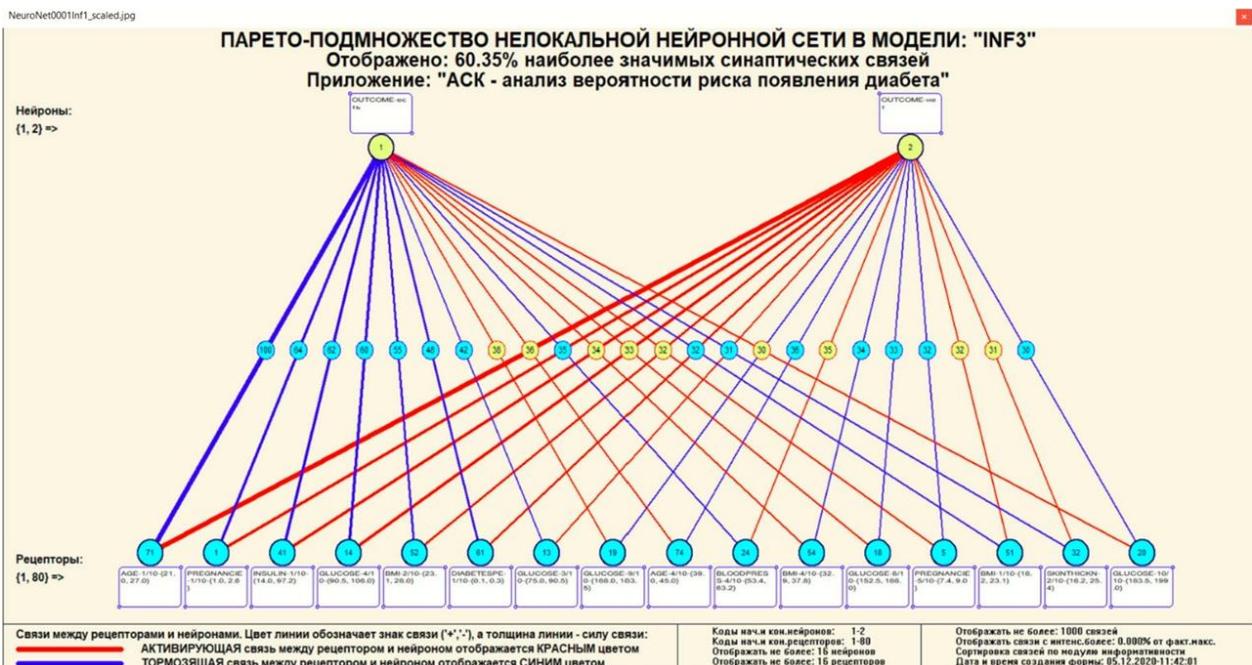


Рисунок 25. Один слой нелокальной нейронной сети, отражающий силу и направление влияния медицинских характеристик на наличие или отсутствие диабета

В приведенном фрагменте слоя нейронной сети нейроны соответствуют результатам наличия или отсутствия диабета, 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:

- 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);

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

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

4.3.5. 3d-интегральные когнитивные карты

На рисунке 26 приведен фрагмент 3d-интегральной когнитивной карты, СК-модели INF3.

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

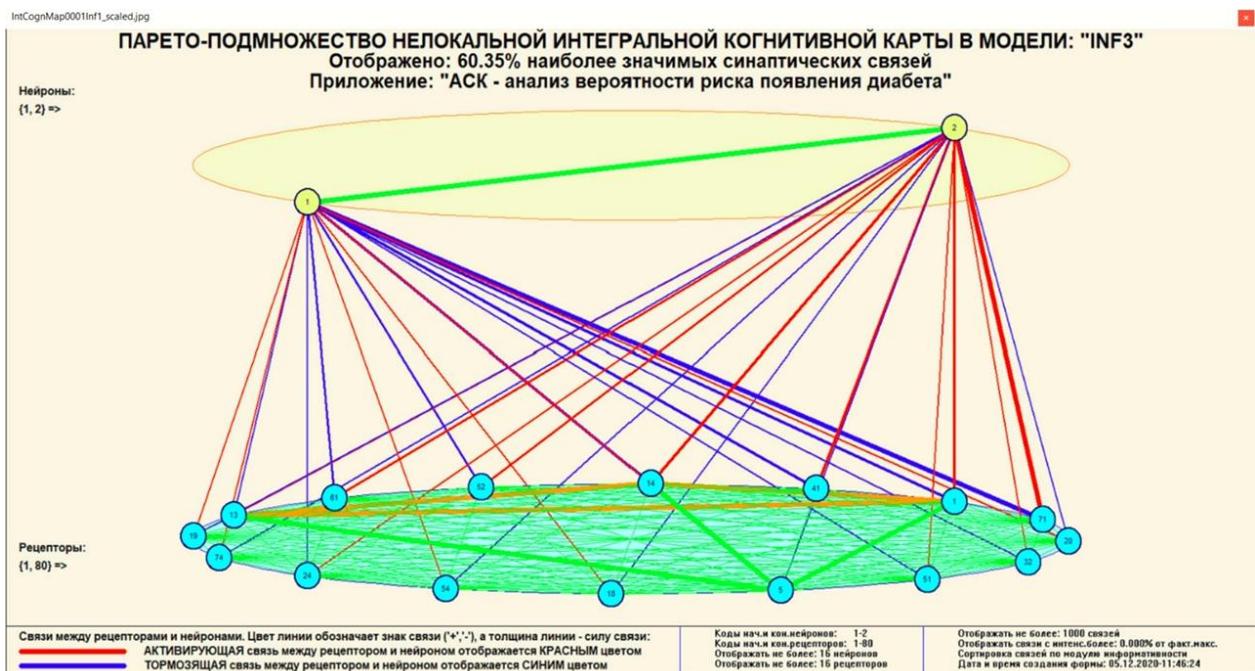


Рисунок 26. 3d-интегральная когнитивная карта в СК-модели INF3

4.3.6. Когнитивные функции

Вместо описания того, что представляют собой когнитивные функции, приведем help соответствующего режима системы «Эйдос» (рисунок 27).

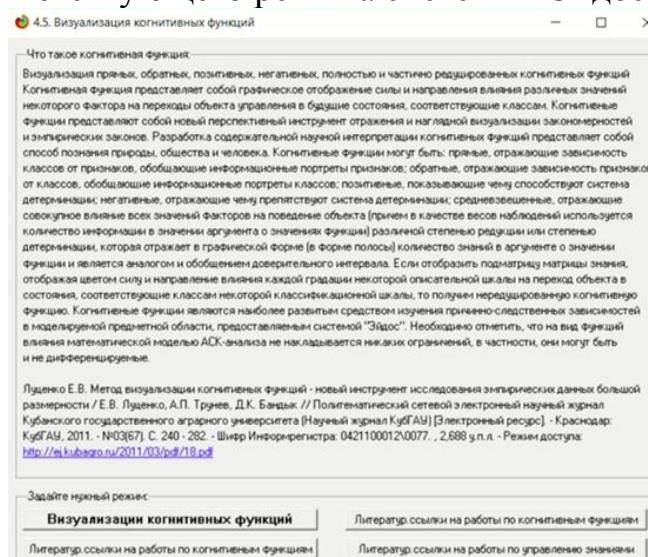


Рисунок 27. Help режима визуализации когнитивных функций

Когнитивная функция представляет собой графическое отображение силы и направления влияния различных значений некоторого фактора (признаков) на переходы объекта управления в будущие состояния, соответствующие классам. Классы являются градациями классификационных шкал.

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

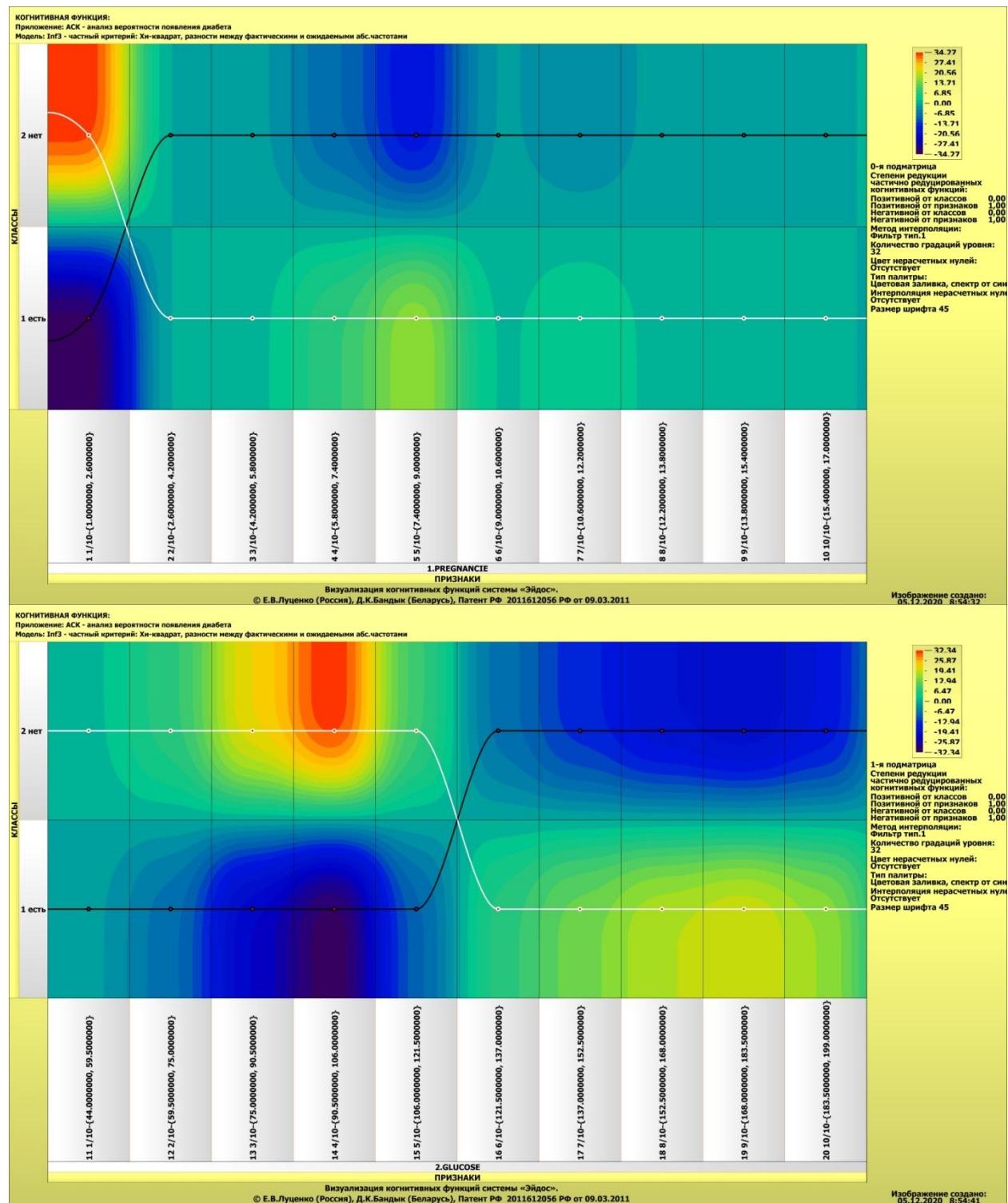
Когнитивные функции могут быть: прямые, отражающие зависимость классов от признаков, обобщающие информационные портреты признаков; обратные, отражающие зависимость признаков от классов, обобщающие информационные портреты классов; позитивные, показывающие чему способствуют системы детерминации (обозначены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

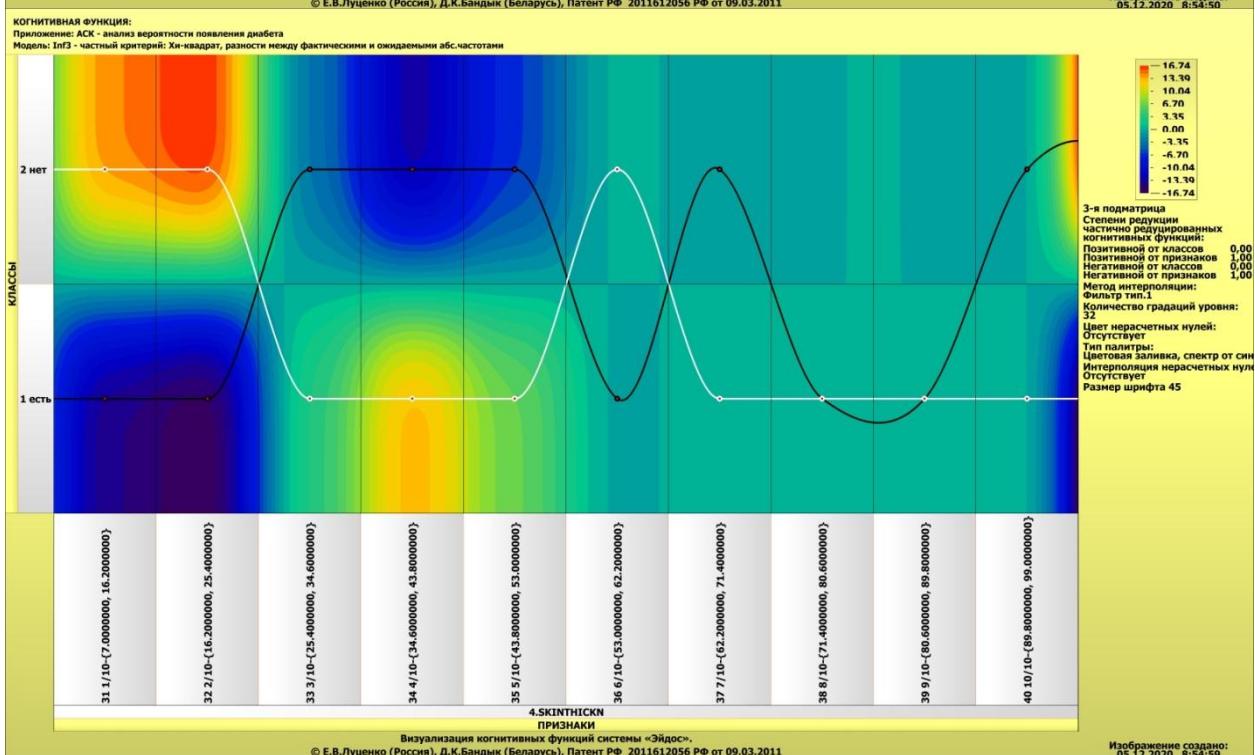
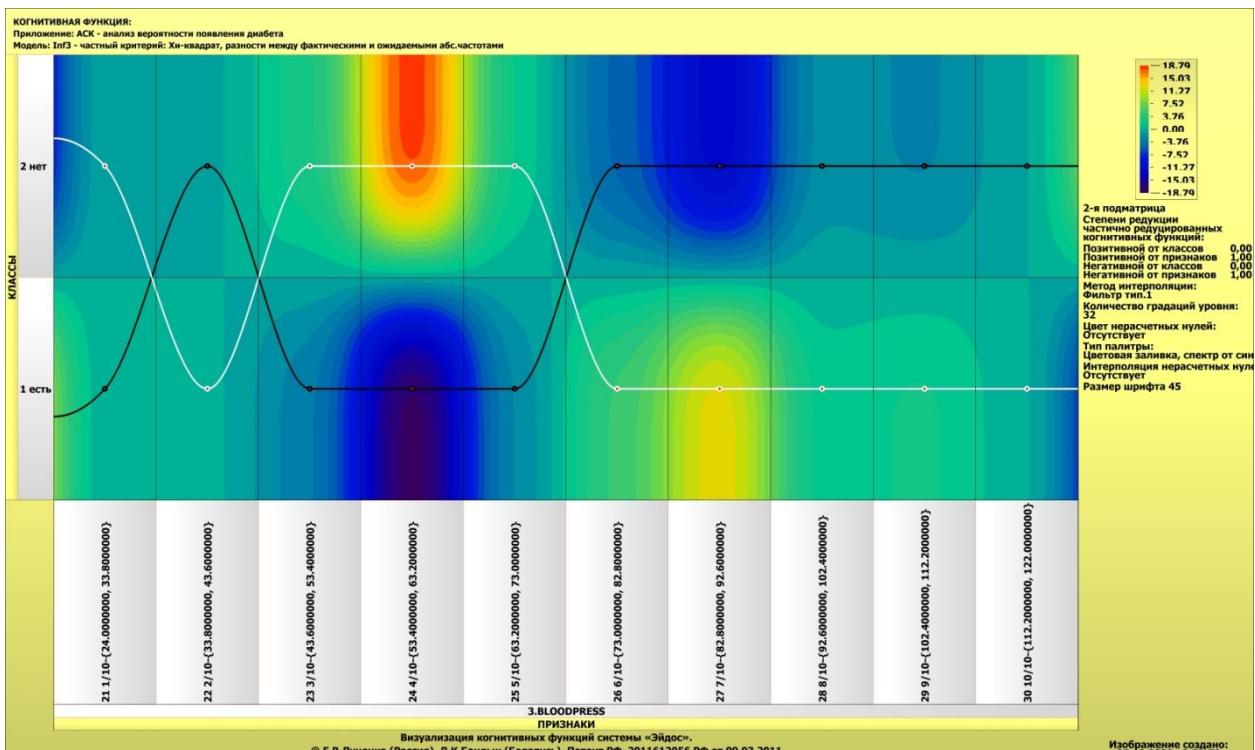
Если отобразить подматрицу матрицы знания, отображая цветом силу и направление влияния каждой градации некоторой описательной шкалы на переход объекта в состояния, соответствующие классам некоторой классификационной шкалы, то получим нередуцированную когнитивную функцию.

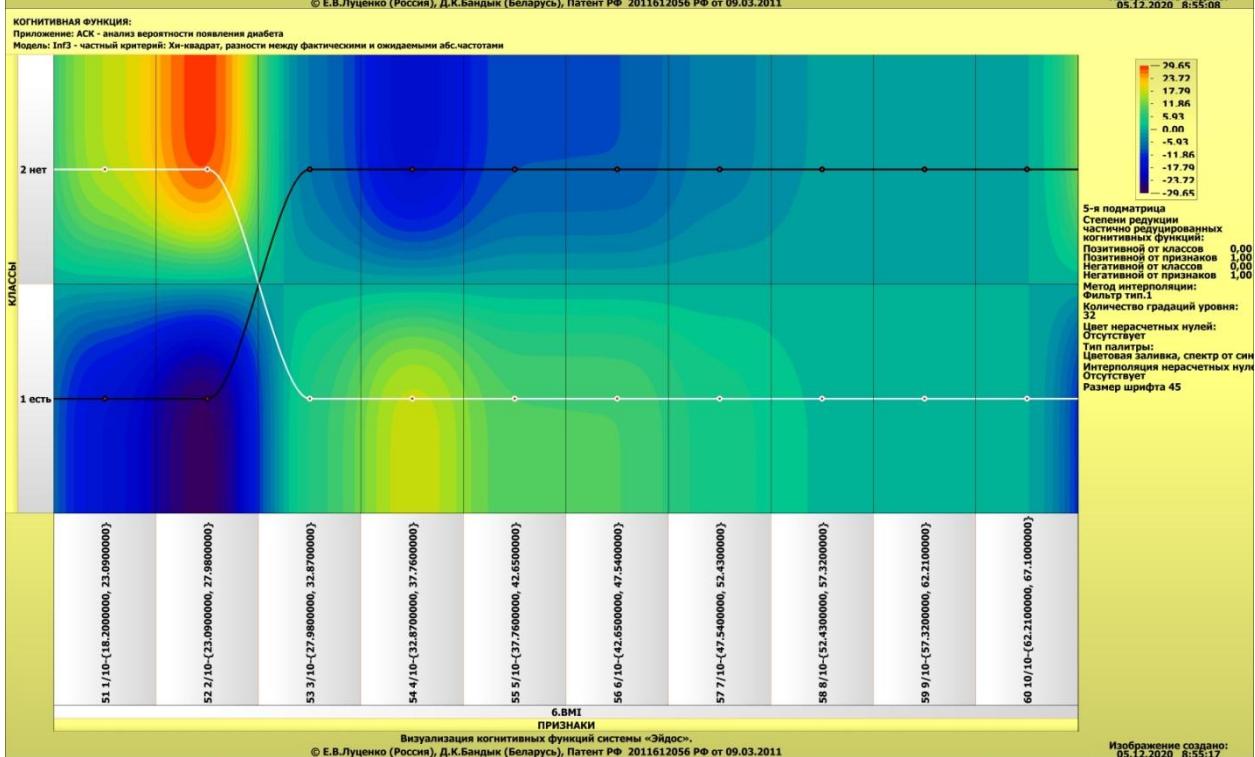
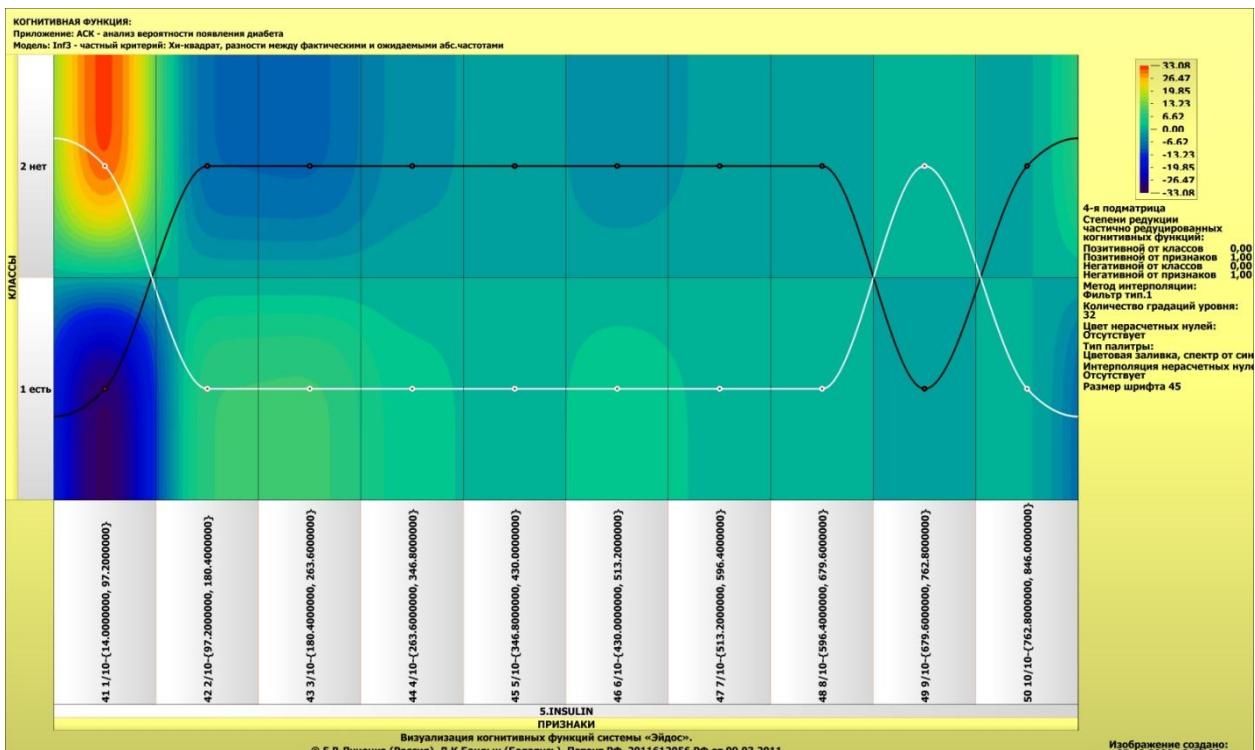
Когнитивные функции являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой "Эйдос".

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

На рисунках 28 приведены когнитивные функции, наглядно отражающие силу и направление влияния значений (т.е. степени выраженности) различных медицинских характеристик (класс).







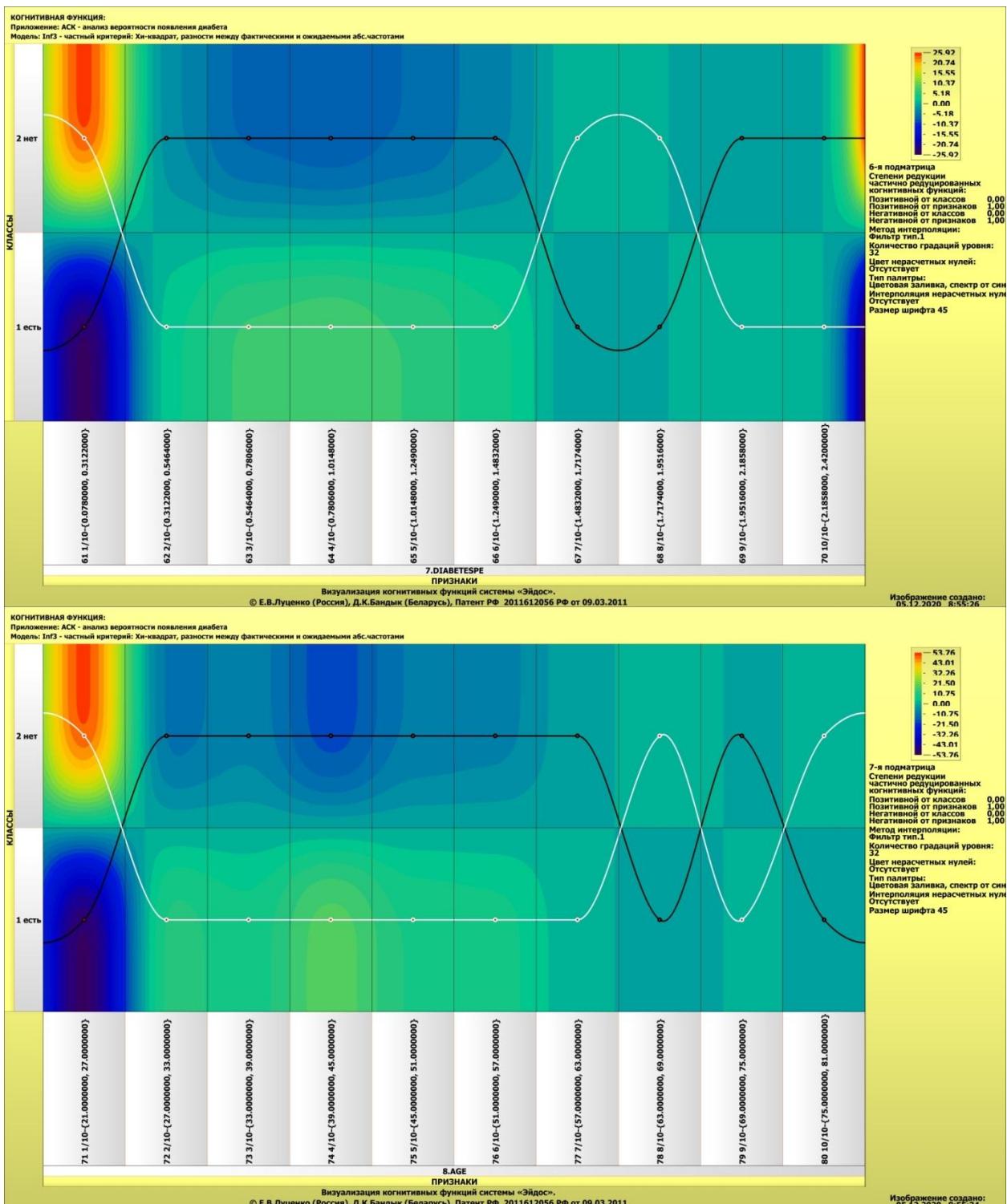


Рисунок 28. Примеры когнитивных функций в СК-модели INF3, отражающих силу и направление влияния значений медицинских характеристик на наличие или отсутствие диабета

Из когнитивных функций, приведенных на рисунке 28, хорошо видно, что **зависимости между значениями медицинских характеристик и влиянием их на наличие или отсутствия диабета в имеют предсказуемый характер.**

4.3.7. Сила и направление влияния значений медицинских характеристик и их влияние на результат наличия или отсутствия диабета

На рисунках 7, 8 приведены фрагменты некоторых статистических и системно-когнитивных моделей, отражающих моделируемую предметную область.

Строки матриц моделей соответствуют значениям факторов, т.е. значениям результатов (градации описательных шкал(есть, нет)).

Колонки матриц моделей соответствуют различным диагностическим характеристикам, отражающим результат наличия или отсутствия диабета (градации классификационных шкал).

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

Если какое-то значение характеристики слабо влияет на класс результат, 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 результата, 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 на его результат (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 результата.

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 7, 8 содержит количественную оценку вариабельности значений строки модели (среднеквадратичное отклонение), которая и представляет собой ценность значения характеристик медицинских, 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 (рисунок 29, рисунок 30).

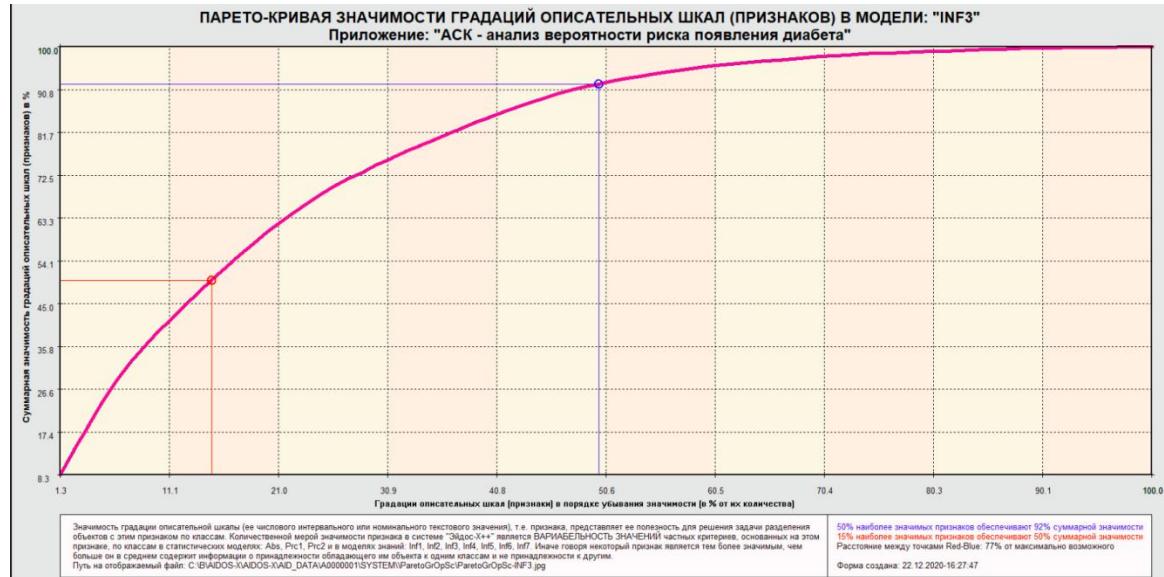


Рисунок 29. Парето-кривая значимости градаций описательных шкал

A	B	C	D	E	F	G	H	I	J	
NUM	NUM_PRC	KOD_ATR	NAME_ATR	KOD_OPSC	ZNACH_ATR	ZN_ATRNIT	ZNACH_PRC	ZN_PRCNIT	DELETE	
1	1,2500000	71	AGE-1/10-[21.000000, 27.000000)	8	76,0271855	76,0271855	8,2619983	8,2619983		
2	2,5000000	1	PREGNANCY-1/10-[1.000000, 2.600000)	1	48,4611907	124,4883762	5,2663567	13,5283550		
3	3,7500000	41	INSULIN-3/10-[14.000000, 97.200000)	5	46,7875776	171,2759538	5,0844824	18,6128375		
4	5,0000000	14	GLUCOSE-4/10-[90.500000, 106.000000)	2	45,7394669	217,0154207	4,9705825	23,5834200		
5	6,2500000	52	BMI-2/10-[23.090000, 27.980000)	6	41,9336388	258,9487845	4,5569671	28,1403871		
6	7,5000000	61	DIABETESPE-1/10-[0.0708000, 0.312200)	7	36,6565417	295,6053262	3,9835262	32,1239133		
7	8,7500000	13	GLUCOSE-3/10-[75.000000, 90.500000)	2	31,6517342	327,2570604	3,4396456	35,5635589		
8	10,0000000	19	GLUCOSE-9/10-[168.000000, 183.500000)	2	28,5297415	355,7868019	3,1003736	38,6639325		
9	11,2500000	74	AGE-4/10-[39.000000, 45.000000)	8	27,2040969	382,9908988	2,9563136	41,6202461		
10	12,5000000	24	BLOODPRESS-4/10-[53.400000, 63.200000)	3	26,5701844	409,5610832	2,8874253	44,5076714		
11	13,7500000	54	BMI-4/10-[32.870000, 37.760000)	6	25,8624206	435,4235038	2,8105115	47,3181829		
12	15,0000000	18	GLUCOSE-8/10-[152.500000, 168.000000)	2	25,3875119	460,8110157	2,7589024	50,0770852		
13	16,2500000	5	PREGNANCIE-5/10-[7.400000, 9.000000)	1	24,3052352	485,1162509	2,6412896	52,7183748		
14	17,5000000	51	BMI-1/10-[20.200000, 23.090000)	6	24,0053619	509,1216128	2,6087019	55,5270767		
15	18,7500000	32	SKINTHICKN-2/10-[16.200000, 25.400000)	4	23,6713227	532,7929355	2,5724013	57,8994779		
16	20,0000000	20	GLUCOSE-10/10-[183.500000, 199.000000)	2	22,5409504	555,3338859	2,4495619	60,3490398		
17	21,2500000	31	SKINTHICKN-1/10-[7.000000, 16.200000)	4	20,3561600	575,6900459	2,2121371	62,5611769		
18	22,5000000	17	GLUCOSE-7/10-[137.000000, 152.500000)	2	18,8757169	594,5657628	2,0512550	64,6124319		
19	23,7500000	34	SKINTHICKN-4/10-[34.600000, 43.800000)	4	18,6846495	613,2504123	2,0304914	66,6429233		
20	25,0000000	27	BLOODPRESS-7/10-[82.800000, 92.600000)	3	17,9684755	631,2188878	1,9526636	68,5955868		
21	26,2500000	72	AGE-2/10-[27.000000, 33.000000)	8	16,8520327	648,0709205	1,8313379	70,4269247		
22	27,5000000	55	BMI-5/10-[37.760000, 42.650000)	6	13,4984989	661,5694194	1,4669039	71,8938286		
23	28,7500000	75	AGE-5/10-[45.000000, 51.000000)	8	13,3576295	674,9270489	1,4515954	73,3454240		
24	30,0000000	56	BMI-6/10-[42.650000, 47.540000)	6	13,3053291	688,2323780	1,4459118	74,7913359		
25	31,2500000	43	INSULIN-3/10-[180.400000, 263.600000)	5	12,7118903	700,942683	1,3814219	76,1727577		
26	32,5000000	15	GLUCOSE-5/10-[106.000000, 121.500000)	2	12,4622150	713,4064833	1,3542892	77,5270470		
27	33,7500000	42	INSULIN-2/10-[97.200000, 180.400000)	5	12,1205540	725,5270373	1,3171604	78,8442074		
28	35,0000000	72	AGE-6/10-[51.000000, 57.000000)	8	11,7160800	737,2431173	1,2732055	80,1174129		
29	36,2500000	4	PREGNANCY-4/10-[5.800000, 7.400000)	1	11,5431470	748,7862643	1,2544126	81,3718255		
30	37,5000000	64	DIABETESPE-4/10-[0.7806600, 1.014800)	7	11,2453762	760,0316405	1,2220534	82,5938789		

Рисунок 30. Парето-таблица значимости градаций описательных шкал, т.е. сила влияния значений медицинских характеристик на результат в СК-модели INF3

Из рисунка 29 и рисунка 30 видно, что 50% наиболее ценных для решения задачи определения вероятности наличия или отсутствия диабета обеспечивают 92% суммарной ценности, а 15 % наиболее ценных признаков обеспечивают 50% суммарной ценности. Обращаем внимание, что наиболее ценным является 2-часовой сывороточный инсулин, возраст и количество

беременностей, а наименее ценным – давление и толщина кожи, причем наименее ценный фактор является бессмысленным.

4.3.8. Степень детерминированности результатов

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается **степенью вариабельности значений факторов** (градаций описательных шкал) в колонке матрицы модели, соответствующей данному классу (рисунок 32).

В данной работе у нас классами являются результаты, а значениями градаций описательных шкал – их характеристики.

На рисунке 31 мы видим Парето-кривую степени детерминированности классов результатов значениями характеристик нарастающим итогом.

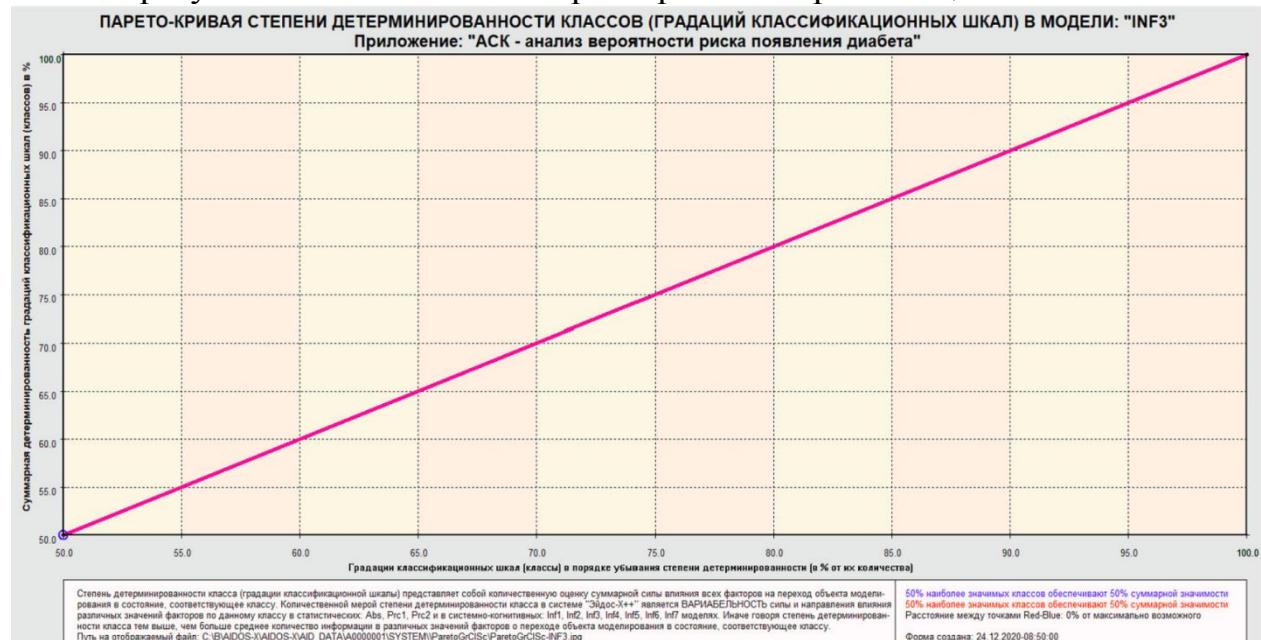


Рисунок 31. Парето-кривая степени детерминированности классов

	A	B	C	D	E	F	G	H	I	J
1	NUM	NUM_PRC	KOD_CLS	NAME_CLS	KOD_CLSC	ZNACH_CLS	ZN_CLSNIT	ZNACH_PRC	ZN_PRCNIT	DELETE
2	1	50,000000		1 OUTCOME-есть		1	12,9420896	12,9420896	50,000000	50,000000
3	2	100,000000		2 OUTCOME-нет		1	12,9420896	25,8841792	50,000000	100,000000
4										

Рисунок 32. Парето-таблица степеней детерминированности (обусловленности) классов результатов в СК-модели INF3

Из рисунка 32 видно, что значения характеристик детерминируют (обуславливают результаты) одинаково.

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. А именно с помощью 2-часового сывороточного инсулина, количества беременностей и возраста у женщин можно определить вероятность риска наличия диабета.

В результате проделанной работы, с помощью системы «Эйдос» были созданы различные модели, в которых непосредственно на основе эмпирических данных сформированы обобщенные образы результатов наличия или отсутствия диабета, изучено влияние диагностических характеристик на эти результаты, 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 № 239 в режиме 1.3 системы «Эйдос».

Список литературы

1. 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>
2. Луценко Е.В. Универсальный информационный вариационный принцип развития систем / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 117 – 193. – Шифр Информрегистра: 0420800012\0091, IDA [article ID]: 0410807010. – Режим доступа: <http://ej.kubagro.ru/2008/07/pdf/10.pdf>, 4,812 у.п.л.
3. Lutsenko E. V. System analysis and decision-making (Automated system-cognitive analysis and solving problems of identification, decision-making and research of the simulated subject area): textbook / E. V. Lutsenko. - Krasnodar: ECSC "Eidos", 2020. - 1031 p. // August 2020, DOI: [10.13140/RG.2.2.27247.05289](https://doi.org/10.13140/RG.2.2.27247.05289), License [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/), [https://www.researchgate.net/publication/343998862 SYSTEM ANALYSIS AND DECISION MAKING Automated system-cognitive analysis and solving problems of identification decision-making and research of the simulated subject area](https://www.researchgate.net/publication/343998862_SYSTEM_ANALYSIS_AND_DECISION_MAKING_Automated_system-cognitive_analysis_and_solving_problems_of_identification_decision-making_and_research_of_the_simulated_subject_area), см. учебный вопрос-2.8.5. Повышение уровня системности объекта управления как цель управления.
4. Lutsenko E.V. On higher forms of consciousness, the prospects of man, technology and society (selected works) // August 2019, DOI: [10.13140/RG.2.2.21336.24320](https://doi.org/10.13140/RG.2.2.21336.24320), License [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/), [https://www.researchgate.net/publication/335057548 On HIGHER FORMS of CONSCIOUSNESS the PROSPECTS of MAN TECHNOLOGY AND SOCIETY selected works](https://www.researchgate.net/publication/335057548_On_HIGHER_FORMS_of_CONSCIOUSNESS_the_PROSPECTS_of_MAN_TECHNOLOGY_AND_SOCIETY_selected_works)
5. Lutsenko E.V. ABOUT THE INTERFACE: "SOUL-COMPUTER» (artificial intelligence: problems and solutions within the system information and functional paradigm of society development) // April 2019, DOI: [10.13140/RG.2.2.23132.85129](https://doi.org/10.13140/RG.2.2.23132.85129), https://www.researchgate.net/publication/332464278_ABOUT_THE_INTERFACE_SOUL-COMPUTER_artificial_intelligence_problems_and_solutions_within_the_system_information_and_functional_paradigm_of_society_development
6. 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>
7. Луценко Е.В. Количественные меры возрастания эмерджентности в процессе эволюции систем (в рамках системной теории информации) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного

аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2006. – №05(021). С. 355 – 374. – Шифр Информрегистра: 0420600012\0089, IDA [article ID]: 0210605031. – Режим доступа: <http://ej.kubagro.ru/2006/05/pdf/31.pdf>, 1,25 у.п.л.

8. Луценко Е.В. Существование, несуществование и изменение как эмерджентные свойства систем // Квантовая Магия, том 5, вып. 1, стр. 1215-1239, 2008. <http://quantmagic.narod.ru/volumes/VOL512008/p1215.html>

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

10. Сайт проф.Е.В.Луценко: <http://lc.kubagro.ru/>

11. Страница Е.В.Луценко: https://www.researchgate.net/profile/Eugene_Lutsenko

12. 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.

13. ПАМЯТИ ПРОФЕССОРА О.Г. КУКОСЯНА // – Режим доступа: <https://www.elibrary.ru/item.asp?id=20266263>

14. Луценко Е.В. Формирование субъективных (виртуальных) моделей физической и социальной реальности сознанием человека и неоправданное приятие им онтологического статуса (гипостазирование) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №09(113). С. 1 – 32. – IDA [article ID]: 1131509001. – Режим доступа: <http://ej.kubagro.ru/2015/09/pdf/01.pdf>, 2 у.п.л.

15. Луценко Е.В. Принципы и перспективы корректной содержательной интерпретации субъективных (виртуальных) моделей физической и социальной реальности, формируемых сознанием человека / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №01(115). С. 22 – 75. – IDA [article ID]: 1151601003. – Режим доступа: <http://ej.kubagro.ru/2016/01/pdf/03.pdf>, 3,375 у.п.л.

16. Луценко Е.В. Сценарный АСК-анализ как метод разработки на основе эмпирических данных базисных функций и весовых коэффициентов для разложения в ряд функции состояния объекта или ситуации по теореме А.Н.Колмогорова (1957) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №07(161). С. 76 – 120. – IDA [article ID]: 1612007009. – Режим доступа: <http://ej.kubagro.ru/2020/07/pdf/09.pdf>, 2,812 у.п.л.

17. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. –

IDA [article ID]: 0921308058. – Режим доступа: <http://ej.kubagro.ru/2013/08/pdf/58.pdf>, 1,562 у.п.л.

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

19. 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 // Политехнический 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.

1.