УДК 004.8
06.01.01 Общее земледелие, растениеводство

АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ ЗАВИСИМОСТИ СУБЪЕКТИВНЫХ СОМЕЛЬЕ-ОЦЕНОК КАЧЕСТВА ВИНА ОТ ЕГО ОБЪЕКТИВНЫХ ФИЗИКО-ХИМИЧЕСКИХ СВОЙСТВ

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор
Scopus Author ID: 57188763047
РИНЦ SPIN-код: 9523-7101
prof.lutsenko@gmail.com http://lc.kubagro.ru

Печурина Елена Каримовна
РИНЦ SPIN-код: 1952-4286
geskov@mail.ru

Сергеев Александр Эдуардович
к.ф.-м.н, доцент
РИНЦ SPIN-код: 7837-9566
Кубанский государственный аграрный университет имени И.Т.Трубилина, Краснодар, Россия

Сомелье оценивает качество вина на основе своих субъективных ощущений. При этом то, что говорит сомелье, когда оценивает вино, непосвященному в это искусство трудно или вообще невозможно понять рационально. Сам процесс оценивания качества вина сомелье не поддается формализации и осуществляется полностью на чувственном уровне. Иногда разные сомелье по-разному оценивают одно и тоже вино, разлитое из одной и той же бочки в бутылки разной престижности с наклейками, отличающимися количеством звездочек. В этой связи возникает по крайней мере два закономерных и естественных вопроса. Первый вопрос о том, связаны ли как-либо субъективные сомелье-оценки качества вина с его объективными физико-химическими свойствами? Второй вопрос возникает в случае положительного ответа на первый: можно ли анализируя объективными методами физико-химические свойства вина предсказать его субъективную оценку различными сомелье или некоторым «обобщенным сомелье», обобщающим много подобных субъективных оценок? Данная статья посвящена получению аргументированных ответов на эти вопросы. Целью данной работы, представляющей большой научный и практический интерес, является создание модели, обеспечивающей автоматизированную оценку качества вина на основе анализа его объективных физико-химические свойства, совпадающую с его сомелье-оценкой. Для достижения этой цели применяется Автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальная

UDC 004.8
General agriculture and crop production

AUTOMATED SYSTEM-COGNITIVE ANALYSIS OF THE DEPENDENCE OF SUBJECTIVE SOMMELIER WINE QUALITY ASSESSMENT ON ITS OBJECTIVE PHYSICAL AND CHEMICAL PROPERTIES

Lutsenko Evgeniy Veniaminovich
Dr.Sci.Econ., Cand.Tech.Sci., professor
Scopus Author ID: 57188763047
RSCI SPIN-code: 9523-7101
prof.lutsenko@gmail.com http://lc.kubagro.ru

Pechurina Elena Karimovna
RSCI SPIN-code: 1952-4286
geskov@mail.ru

Sergeev Aleksandr Eduardovich
Cand.Phys.-Math.Sci., associate Professor
RSCI SPIN-code: 7837-9566
Kuban State Agrarian University, Krasnodar, Russia

Sommelier evaluates the quality of wine on the basis of their subjective feelings. At the same time, what the sommelier says when evaluating wine, it is difficult or impossible to rationally understand for the uninitiated to this art. The process of assessing the quality of wine by sommelier can not be formalized and is carried out entirely at the sensual level. Sometimes, different sommeliers differently evaluate the same wine poured from the same barrel into bottles of different prestige, with stickers differing in the number of stars. This raises at least two legitimate and natural questions. The first question is whether any subjective sommelier evaluations of the quality of wine are connected with its objective physical and chemical properties? The second question arises in the case of a positive answer to the first one: is it possible to analyze the objective methods of physical and chemical properties of wine to predict its subjective assessment by various sommeliers or some "generalized sommelier", generalizing many such subjective assessments? This article is devoted to obtaining reasoned answers to these questions. The purpose of this work, which is of great scientific and practical interest, is to create a model that provides an automated assessment of the quality of wine based on the analysis of its objective physical and chemical properties, coinciding with its sommelier-evaluation. To achieve this goal, we use Automated system-cognitive analysis (ASC-analysis) and its software tools – the intelligent system called "Eidos". A detailed numerical example based on 1599 real-world examples of sommelier evaluation of wine quality with known physical and chemical properties is considered. In addition to the answer to the two questions in the article, there is a study of the created

система «Эйдос». Рассматривается подробный численный пример, основанный на 1599 реальных примерах оценки сомелье качества вин с известными физико-химическими свойствами. Кроме ответа на два поставленных вопроса, в статье приводится и исследование созданной системно-когнитивной модели

Ключевые слова: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, АСК-АНАЛИЗ, СИСТЕМА «ЭЙДОС», СОМЕЛЬЕ, КАЧЕСТВО ВИНА, ФИЗИКО-ХИМИЧЕСКИЕ СВОЙСТВА ВИНА
Doi: 10.21515/1990-4665-149-015

system-cognitive model

Keywords: AUTOMATED SYSTEM-COGNITIVE ANALYSIS, ASC-ANALYSIS, "EIDOS" SYSTEM, SOMMELIER, WINE QUALITY, PHYSICAL AND CHEMICAL PROPERTIES OF WINE

СОДЕРЖАНИЕ

ВВЕДЕНИЕ **2**

ЗАДАЧА 1: КОГНИТИВНАЯ СТРУКТУРИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ **6**

ЗАДАЧА 2: ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ И ФОРМАЛИЗАЦИЯ ПРЕДМЕТНОЙ ОБЛАСТИ **6**

ЗАДАЧА 3: СИНТЕЗ И ВЕРИФИКАЦИЯ СТАТИСТИЧЕСКИХ И СИСТЕМНО-КОГНИТИВНЫХ МОДЕЛЕЙ И ВЫБОР НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ **11**

ЗАДАЧА 4: РЕШЕНИЕ РАЗЛИЧНЫХ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ **18**

ПОДЗАДАЧА 4.1. ПРОГНОЗИРОВАНИЕ (ДИАГНОСТИКА, КЛАССИФИКАЦИЯ, РАСПОЗНАВАНИЕ, ИДЕНТИФИКАЦИЯ) 18
ПОДЗАДАЧА 4.2. ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЙ (SWOT-АНАЛИЗ) 20
ПОДЗАДАЧА 4.3. ИССЛЕДОВАНИЕ МОДЕЛИРУЕМОЙ ПРЕДМЕТНОЙ ОБЛАСТИ ПУТЕМ ИССЛЕДОВАНИЯ ЕЕ МОДЕЛИ 25
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 *29*
4.3.5. Нелокальные нейроны и нелокальные нейронные сети *30*
4.3.6. 3d-интегральные когнитивные карты *32*
4.3.7. Когнитивные функции *32*
4.3.8. Сила влияния значений физико-химических свойств вина и самих этих свойств на сомелье-оценки качества вина *35*
4.3.8. Степень детерминированности (обусловленности) сомелье-оценок качества вина его объективными физико-химическими свойствами *39*

5. ВЫВОДЫ **40**

СПИСОК ЛИТЕРАТУРЫ **41**

Введение

Сомелье оценивает качество вина на основе своих субъективных ощущений.

Однако, то, что говорит сомелье, когда оценивает вино, непосвященному в это искусство трудно или вообще невозможно понять рационально. Кроме того считается, что сам процесс оценивания качества вина сомелье не поддается формализации и осуществляется полностью на чувственном уровне. Более того, иногда разные сомелье по-разному

оценивают одно и тоже вино, разлитое из одной и той же бочки в бутылки разной престижности с наклейками, отличающимися количеством звездочек. Это можно считать последней каплей дегтя, переполнившей бочку с вином, т.е. **проблемой**, которую надо решать.

В этой связи возникает по крайней мере два закономерных и естественных вопроса.

Первый вопрос о том, связаны ли как-либо субъективные сомелье оценки качества вина с его объективными физико-химическими свойствами?

Второй вопрос возникает в случае положительного ответа на первый: можно ли анализируя объективными методами физико-химические свойства вина предсказать его субъективную оценку различными сомелье или даже некоторым «обобщенным сомелье», обобщающим много подобных субъективных оценок?

Данная статья посвящена получению аргументированных ответов на эти вопросы.

Целью данной работы, представляющей большой научный и практический интерес, является создание модели, обеспечивающей автоматизированную оценку качества вина на основе анализа его объективных физико-химические свойства, совпадающую с его сомелье-оценкой.

Для достижения этой цели применяется Автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – интеллектуальная система «Эйдос». Рассматривается подробный численный пример, основанный на 1599 реальных примерах оценки сомелье качества вин с известными физико-химическими свойствами. Кроме ответа на два поставленных вопроса в статье приводится и исследование созданной системно-когнитивной модели.

Для достижения этой 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 и поэтому и предлагается применить для их решения.

АСК-анализ представляет собой метод искусственного интеллекта, разработанный проф. Е.В. Луценко в 2002 году [1] 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»).

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

- разработана в универсальной постановке, не зависящей от предметной области. Поэтому она является универсальной и может быть применена во многих предметных областях (http://lc.kubagro.ru/aidos/index.htm);
- находится в полном открытом бесплатном доступе (http://lc.kubagro.ru/aidos/_Aidos-X.htm), причем с актуальными исходными текстами (http://lc.kubagro.ru/__AIDOS-X.txt);
- является одной из первых отечественных систем искусственного интеллекта персонального уровня, т.е. она не требует от пользователя специальной подготовки в области технологий искусственного интеллекта (есть акт внедрения системы «Эйдос» 1987 года) (http://lc.kubagro.ru/aidos/aidos02/PR-4.htm);
- обеспечивает устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных взаимозависимых (нелинейных) данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения (т.е. не предъявляет жестких требований к данным, которые невозможно выполнить, а обрабатывает те данные, которые есть);
- содержит большое количество локальных (поставляемых с инсталляцией) и облачных учебных и научных приложений (в настоящее время их 31 и 149, соответственно) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);
- обеспечивает мультиязычную поддержку интерфейса на 44 языках. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;
- поддерживает on-line среду накопления знаний и широко используется во всем мире (http://aidos.byethost5.com/map3.php);

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

- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры некоторых форм можно посмотреть в работе: http://lc.kubagro.ru/aidos/aidos18_LLS/aidos18_LLS.pdf).

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

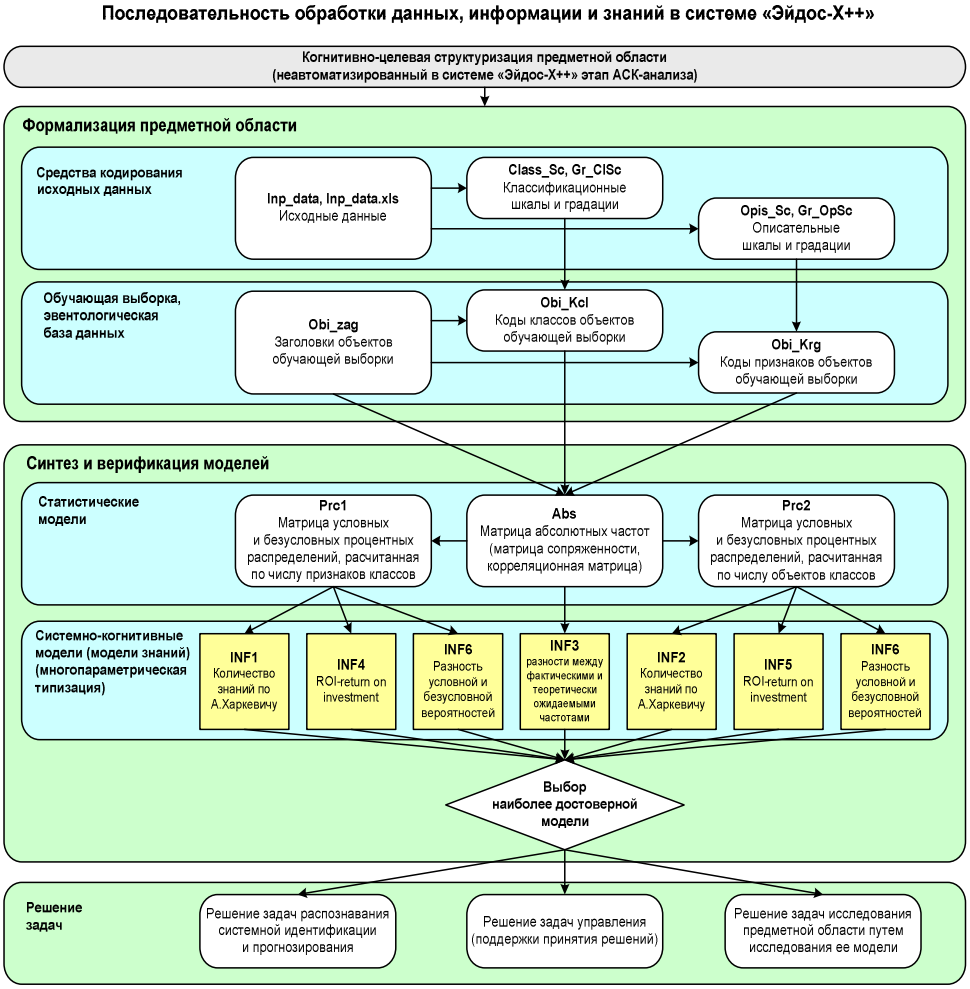


Рисунок 1. Последовательность преобразования данных в информацию, а ее в знания и решения задач в АСК-анализе и системе «Эйдос»

Рассмотрим решение поставленных задач в 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, 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.

В данной работе в качестве классификационной шкалы выбрана субъективная сомелье-оценка качества вина, а в качестве факторов, влияющих на эту оценку – различные устанавливаемые объективными методами физико-химические свойства вина.

Задача 2: подготовка исходных данных и формализация предметной области

Исходные данные для данной статьи (таблица 1) взяты с известных сайтов с данными для машинного обучения Kaggle и UCI [2, 3]. Авторами этих данных являются авторы работы: [4].

Из-за ограничений на объем статьи в таблице 1 приведен фрагмент исходных данных из 40 строк, что составляет около 2.5% всех наблюдений (полная таблица включает 1599 наблюдение).

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

Обратим внимание, что заданы адаптивные интервалы, учитывающее неравномерность распределения данных по значениям. В классификационных шкалах задано 5 числовых интервальных значений (5-бальная шкала), а в описательных – 10.

На рисунке 3 приведен Help данного режима, в котором объясняется принцип организации таблицы исходных для данного режима.

Таблица 1 – Исходные данные для ввода в систему «Эйдос» (фрагмент)[1]

fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
1	7,4	0,70	0,00	1,9	0,076	11	34	0,9978	3,51	0,56	9,4
2	7,8	0,88	0,00	2,6	0,098	25	67	0,9968	3,20	0,68	9,8
3	7,8	0,76	0,04	2,3	0,092	15	54	0,9970	3,26	0,65	9,8
4	11,2	0,28	0,56	1,9	0,075	17	60	0,9980	3,16	0,58	9,8
5	7,4	0,70	0,00	1,9	0,076	11	34	0,9978	3,51	0,56	9,4
6	7,4	0,66	0,00	1,8	0,075	13	40	0,9978	3,51	0,56	9,4
7	7,9	0,60	0,06	1,6	0,069	15	59	0,9964	3,30	0,46	9,4
8	7,3	0,65	0,00	1,2	0,065	15	21	0,9946	3,39	0,47	10,0
9	7,8	0,58	0,02	2,0	0,073	9	18	0,9968	3,36	0,57	9,5
10	7,5	0,50	0,36	6,1	0,071	17	102	0,9978	3,35	0,80	10,5
11	6,7	0,58	0,08	1,8	0,097	15	65	0,9959	3,28	0,54	9,2
12	7,5	0,50	0,36	6,1	0,071	17	102	0,9978	3,35	0,80	10,5
13	5,6	0,62	0,00	1,6	0,089	16	59	0,9943	3,58	0,52	9,9
14	7,8	0,61	0,29	1,6	0,114	9	29	0,9974	3,26	1,56	9,1
15	8,9	0,62	0,18	3,8	0,176	52	145	0,9986	3,16	0,88	9,2
16	8,9	0,62	0,19	3,9	0,170	51	148	0,9986	3,17	0,93	9,2
17	8,5	0,28	0,56	1,8	0,092	35	103	0,9969	3,30	0,75	10,5
18	8,1	0,56	0,28	1,7	0,368	16	56	0,9968	3,11	1,28	9,3
19	7,4	0,59	0,08	4,4	0,086	6	29	0,9974	3,38	0,50	9,0
20	7,9	0,32	0,51	1,8	0,341	17	56	0,9969	3,04	1,08	9,2
21	8,9	0,22	0,48	1,8	0,077	29	60	0,9968	3,39	0,53	9,4
22	7,6	0,39	0,31	2,3	0,082	23	71	0,9982	3,52	0,65	9,7
23	7,9	0,43	0,21	1,6	0,106	10	37	0,9966	3,17	0,91	9,5
24	8,5	0,49	0,11	2,3	0,084	9	67	0,9968	3,17	0,53	9,4
25	6,9	0,40	0,14	2,4	0,085	21	40	0,9968	3,43	0,63	9,7
26	6,3	0,39	0,16	1,4	0,080	11	23	0,9955	3,34	0,56	9,3
27	7,6	0,41	0,24	1,8	0,080	4	11	0,9962	3,28	0,59	9,5
28	7,9	0,43	0,21	1,6	0,106	10	37	0,9966	3,17	0,91	9,5
29	7,1	0,71	0,00	1,9	0,080	14	35	0,9972	3,47	0,55	9,4
30	7,8	0,65	0,00	2,0	0,082	8	16	0,9964	3,38	0,59	9,8
31	6,7	0,68	0,07	2,4	0,089	17	82	0,9958	3,35	0,54	10,1
32	6,9	0,69	0,00	2,5	0,105	22	37	0,9966	3,46	0,57	10,6
33	8,3	0,66	0,12	2,3	0,083	15	113	0,9966	3,17	0,66	9,8
34	6,9	0,61	0,12	10,7	0,073	40	83	0,9993	3,45	0,52	9,4
35	5,2	0,32	0,25	1,8	0,103	13	50	0,9957	3,38	0,55	9,2
36	7,8	0,65	0,00	5,5	0,086	5	18	0,9986	3,40	0,55	9,6
37	7,8	0,60	0,14	2,4	0,086	3	15	0,9975	3,42	0,60	10,8
38	8,1	0,38	0,28	2,1	0,066	13	30	0,9968	3,23	0,73	9,7
39	5,7	1,13	0,09	1,5	0,172	7	19	0,9940	3,50	0,48	9,8
40	7,3	0,45	0,36	5,9	0,074	12	87	0,9978	3,33	0,83	10,5

[1] Полную таблицу исходных данных, представленную в таблице 1, можно скачать с FTP-сервера системы «Эйдос» по ссылке: http://aidos.byethost5.com/Source_data_applications/Applications-000148/Inp_data.xls

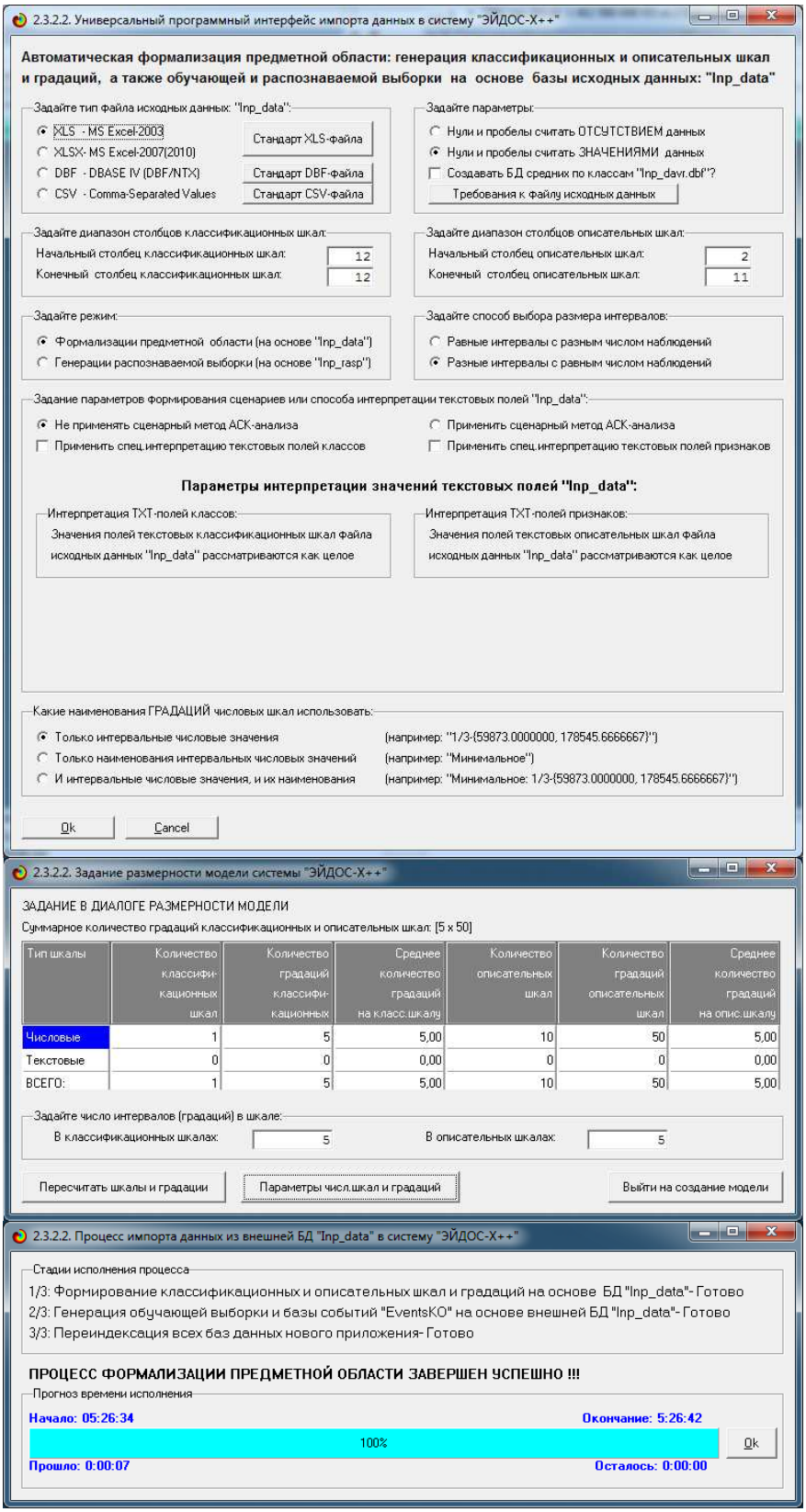


Рисунок 2. Экранные форма программного интерфейса (API) 2.3.2.2 системы «Эйдос» с внешними данными табличного типа[2]

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

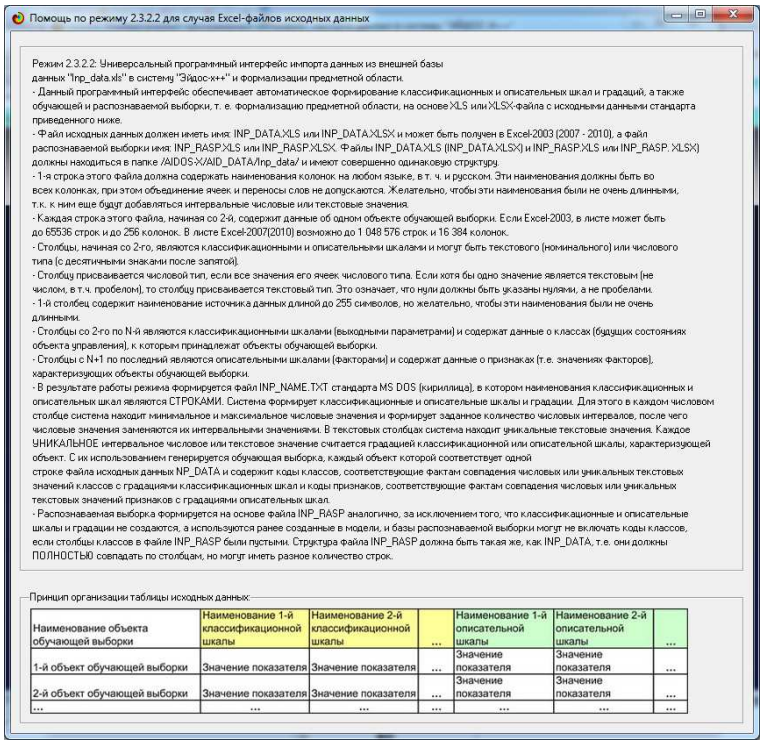


Рисунок 3. Экранные форма HELP программного интерфейса (API) 2.3.2.2

В результате работы режима сформировано классификационная шкалы с суммарным количеством градаций (классов) 5 (таблица 2) и 10 описательных шкал с суммарным числом градаций 50 (таблица 3, рисунок 4).

Таблица 2 – Классификационные шкалы и градации (субъективные сомелье-ценки оценки качества вина)

KOD_CLS	NAME_CLS
1	QUALITY-1/5-{8.4000000, 9.5000000}
2	QUALITY-2/5-{9.5000000, 9.9000000}
3	QUALITY-3/5-{9.9000000, 10.5000000}
4	QUALITY-4/5-{10.5000000, 11.3000000}
5	QUALITY-5/5-{11.3000000, 14.9000000}

Таблица 3 – Описательные шкалы и градации (объективные физико-химические свойства вина)

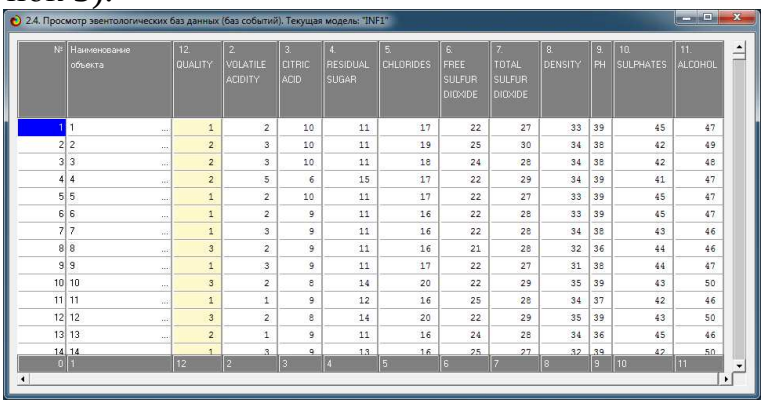
KOD_ATR	NAME_ATR
1	VOLATILE ACIDITY-1/5-{4.6000000, 7.0000000}
2	VOLATILE ACIDITY-2/5-{7.0000000, 7.6000000}
3	VOLATILE ACIDITY-3/5-{7.6000000, 8.3000000}
4	VOLATILE ACIDITY-4/5-{8.3000000, 9.7000000}
5	VOLATILE ACIDITY-5/5-{9.7000000, 15.9000000}
6	CITRIC ACID-1/5-{0.1200000, 0.3700000}
7	CITRIC ACID-2/5-{0.3700000, 0.4700000}
8	CITRIC ACID-3/5-{0.4700000, 0.5700000}
9	CITRIC ACID-4/5-{0.5700000, 0.6600000}
10	CITRIC ACID-5/5-{0.6600000, 1.5800000}
11	RESIDUAL SUGAR-1/5-{0.0000000, 0.0700000}
12	RESIDUAL SUGAR-2/5-{0.0700000, 0.2100000}
13	RESIDUAL SUGAR-3/5-{0.2100000, 0.3200000}
14	RESIDUAL SUGAR-4/5-{0.3200000, 0.4600000}
15	RESIDUAL SUGAR-5/5-{0.4600000, 1.0000000}
16	CHLORIDES-1/5-{0.9000000, 1.8000000}
17	CHLORIDES-2/5-{1.8000000, 2.1000000}
18	CHLORIDES-3/5-{2.1000000, 2.3000000}
19	CHLORIDES-4/5-{2.3000000, 2.7000000}

20	CHLORIDES-5/5-{2.7000000, 15.5000000}
21	FREE SULFUR DIOXIDE-1/5-{0.0120000, 0.0670000}
22	FREE SULFUR DIOXIDE-2/5-{0.0670000, 0.0760000}
23	FREE SULFUR DIOXIDE-3/5-{0.0760000, 0.0820000}
24	FREE SULFUR DIOXIDE-4/5-{0.0820000, 0.0940000}
25	FREE SULFUR DIOXIDE-5/5-{0.0940000, 0.6110000}
26	TOTAL SULFUR DIOXIDE-1/5-{1.0000000, 6.0000000}
27	TOTAL SULFUR DIOXIDE-2/5-{6.0000000, 11.0000000}
28	TOTAL SULFUR DIOXIDE-3/5-{11.0000000, 16.0000000}
29	TOTAL SULFUR DIOXIDE-4/5-{16.0000000, 24.0000000}
30	TOTAL SULFUR DIOXIDE-5/5-{24.0000000, 72.0000000}
31	DENSITY-1/5-{6.0000000, 19.0000000}
32	DENSITY-2/5-{19.0000000, 30.0000000}
33	DENSITY-3/5-{30.0000000, 45.0000000}
34	DENSITY-4/5-{45.0000000, 69.0000000}
35	DENSITY-5/5-{69.0000000, 289.0000000}
36	PH-1/5-{0.9900700, 0.9953400}
37	PH-2/5-{0.9953400, 0.9963000}
38	PH-3/5-{0.9963000, 0.9971400}
39	PH-4/5-{0.9971400, 0.9981700}
40	PH-5/5-{0.9981700, 1.0036900}
41	SULPHATES-1/5-{2.7400000, 3.1800000}
42	SULPHATES-2/5-{3.1800000, 3.2800000}
43	SULPHATES-3/5-{3.2800000, 3.3500000}
44	SULPHATES-4/5-{3.3500000, 3.4200000}
45	SULPHATES-5/5-{3.4200000, 4.0100000}
46	ALCOHOL-1/5-{0.3300000, 0.5400000}
47	ALCOHOL-2/5-{0.5400000, 0.5900000}
48	ALCOHOL-3/5-{0.5900000, 0.6500000}
49	ALCOHOL-4/5-{0.6500000, 0.7600000}
50	ALCOHOL-5/5-{0.7600000, 2.0000000}

```
ПАРАМЕТРЫ ШКАЛ И ГРАДАЦИЙ С АДАПТИВНЫМИ ГРАНИЦАМИ И ПРИМЕРНО РАВНЫМ КОЛИЧЕСТВОМ НАБЛЮДЕНИЙ ПО ГРАДАЦИЯМ
с коррекцией ошибки округления числа наблюдений по интервалу градации при переходе к следующей градации

КЛАССИФИКАЦИОННАЯ ШКАЛА:код: [   1], наим.: "QUALITY", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
             1 Наим.градации: 1/5-{ 8.4000000,  9.5000000}, размер интервала= 1.1000000, расч./факт.число наблюдений на градацию: 319/319
             2 Наим.градации: 2/5-{ 9.5000000,  9.9000000}, размер интервала= 0.4000000, расч./факт.число наблюдений на градацию: 320/320
             3 Наим.градации: 3/5-{ 9.9000000, 10.5000000}, размер интервала= 0.6000000, расч./факт.число наблюдений на градацию: 320/320
             4 Наим.градации: 4/5-{10.5000000, 11.3000000}, размер интервала= 0.8000000, расч./факт.число наблюдений на градацию: 320/320
             5 Наим.градации: 5/5-{11.3000000, 14.9000000}, размер интервала= 3.6000000, расч./факт.число наблюдений на градацию: 320/320

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [   1], наим.: "VOLATILE ACIDITY", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
             1 Наим.градации: 1/5-{ 4.6000000,  7.0000000}, размер интервала= 2.4000000, расч./факт.число наблюдений на градацию: 319/319
             2 Наим.градации: 2/5-{ 7.0000000,  7.6000000}, размер интервала= 0.6000000, расч./факт.число наблюдений на градацию: 320/320
             3 Наим.градации: 3/5-{ 7.6000000,  8.3000000}, размер интервала= 0.7000000, расч./факт.число наблюдений на градацию: 320/320
             4 Наим.градации: 4/5-{ 8.3000000,  9.7000000}, размер интервала= 1.4000000, расч./факт.число наблюдений на градацию: 320/320
             5 Наим.градации: 5/5-{ 9.7000000, 15.9000000}, размер интервала= 6.2000000, расч./факт.число наблюдений на градацию: 320/320

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [   2], наим.: "CITRIC ACID", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
             6 Наим.градации: 1/5-{0.1200000, 0.3700000}, размер интервала=0.2500000, расч./факт.число наблюдений на градацию: 319/319
             7 Наим.градации: 2/5-{0.3700000, 0.4700000}, размер интервала=0.1000000, расч./факт.число наблюдений на градацию: 320/320
             8 Наим.градации: 3/5-{0.4700000, 0.5700000}, размер интервала=0.1000000, расч./факт.число наблюдений на градацию: 320/320
             9 Наим.градации: 4/5-{0.5700000, 0.6600000}, размер интервала=0.0900000, расч./факт.число наблюдений на градацию: 320/320
            10 Наим.градации: 5/5-{0.6600000, 1.5800000}, размер интервала=0.9200000, расч./факт.число наблюдений на градацию: 320/320

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [   3], наим.: "RESIDUAL SUGAR", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
            11 Наим.градации: 1/5-{0.0000000, 0.0700000}, размер интервала=0.0700000, расч./факт.число наблюдений на градацию: 319/319
            12 Наим.градации: 2/5-{0.0700000, 0.2100000}, размер интервала=0.1400000, расч./факт.число наблюдений на градацию: 320/320
            13 Наим.градации: 3/5-{0.2100000, 0.3200000}, размер интервала=0.1100000, расч./факт.число наблюдений на градацию: 320/320
            14 Наим.градации: 4/5-{0.3200000, 0.4600000}, размер интервала=0.1400000, расч./факт.число наблюдений на градацию: 320/320
            15 Наим.градации: 5/5-{0.4600000, 1.0000000}, размер интервала=0.5400000, расч./факт.число наблюдений на градацию: 320/320

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [   4], наим.: "CHLORIDES", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
            16 Наим.градации: 1/5-{ 0.9000000,  1.8000000}, размер интервала= 0.9000000, расч./факт.число наблюдений на градацию: 319/319
            17 Наим.градации: 2/5-{ 1.8000000,  2.1000000}, размер интервала= 0.3000000, расч./факт.число наблюдений на градацию: 320/320
            18 Наим.градации: 3/5-{ 2.1000000,  2.3000000}, размер интервала= 0.2000000, расч./факт.число наблюдений на градацию: 320/320
            19 Наим.градации: 4/5-{ 2.3000000,  2.7000000}, размер интервала= 0.4000000, расч./факт.число наблюдений на градацию: 320/320
            20 Наим.градации: 5/5-{ 2.7000000, 15.5000000}, размер интервала=12.8000000, расч./факт.число наблюдений на градацию: 320/320

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [   5], наим.: "FREE SULFUR DIOXIDE", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
            21 Наим.градации: 1/5-{0.0120000, 0.0670000}, размер интервала=0.0550000, расч./факт.число наблюдений на градацию: 319/319
            22 Наим.градации: 2/5-{0.0670000, 0.0760000}, размер интервала=0.0090000, расч./факт.число наблюдений на градацию: 320/320
            23 Наим.градации: 3/5-{0.0760000, 0.0820000}, размер интервала=0.0060000, расч./факт.число наблюдений на градацию: 320/320
            24 Наим.градации: 4/5-{0.0820000, 0.0940000}, размер интервала=0.0120000, расч./факт.число наблюдений на градацию: 320/320
            25 Наим.градации: 5/5-{0.0940000, 0.6110000}, размер интервала=0.5170000, расч./факт.число наблюдений на градацию: 320/320

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [   6], наим.: "TOTAL SULFUR DIOXIDE", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
            26 Наим.градации: 1/5-{ 1.0000000,  6.0000000}, размер интервала= 5.0000000, расч./факт.число наблюдений на градацию: 319/319
            27 Наим.градации: 2/5-{ 6.0000000, 11.0000000}, размер интервала= 5.0000000, расч./факт.число наблюдений на градацию: 320/320
            28 Наим.градации: 3/5-{11.0000000, 16.0000000}, размер интервала= 5.0000000, расч./факт.число наблюдений на градацию: 320/320
            29 Наим.градации: 4/5-{16.0000000, 24.0000000}, размер интервала= 8.0000000, расч./факт.число наблюдений на градацию: 320/320
            30 Наим.градации: 5/5-{24.0000000, 72.0000000}, размер интервала=48.0000000, расч./факт.число наблюдений на градацию: 320/320

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [   7], наим.: "DENSITY", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
            31 Наим.градации: 1/5-{  6.0000000,  19.0000000}, размер интервала= 13.0000000, расч./факт.число наблюдений на градацию: 319/319
            32 Наим.градации: 2/5-{ 19.0000000,  30.0000000}, размер интервала= 11.0000000, расч./факт.число наблюдений на градацию: 320/320
            33 Наим.градации: 3/5-{ 30.0000000,  45.0000000}, размер интервала= 15.0000000, расч./факт.число наблюдений на градацию: 320/320
            34 Наим.градации: 4/5-{ 45.0000000,  69.0000000}, размер интервала= 24.0000000, расч./факт.число наблюдений на градацию: 320/320
            35 Наим.градации: 5/5-{ 69.0000000, 289.0000000}, размер интервала=220.0000000, расч./факт.число наблюдений на градацию: 320/320

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [   8], наим.: "PH", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
            36 Наим.градации: 1/5-{0.9900700, 0.9953400}, размер интервала=0.0052700, расч./факт.число наблюдений на градацию: 319/319
            37 Наим.градации: 2/5-{0.9953400, 0.9963000}, размер интервала=0.0009600, расч./факт.число наблюдений на градацию: 320/320
            38 Наим.градации: 3/5-{0.9963000, 0.9971400}, размер интервала=0.0008400, расч./факт.число наблюдений на градацию: 320/320
            39 Наим.градации: 4/5-{0.9971400, 0.9981700}, размер интервала=0.0010300, расч./факт.число наблюдений на градацию: 320/320
            40 Наим.градации: 5/5-{0.9981700, 1.0036900}, размер интервала=0.0055200, расч./факт.число наблюдений на градацию: 320/320

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [   9], наим.: "SULPHATES", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
            41 Наим.градации: 1/5-{2.7400000, 3.1800000}, размер интервала=0.4400000, расч./факт.число наблюдений на градацию: 319/319
            42 Наим.градации: 2/5-{3.1800000, 3.2800000}, размер интервала=0.1000000, расч./факт.число наблюдений на градацию: 320/320
            43 Наим.градации: 3/5-{3.2800000, 3.3500000}, размер интервала=0.0700000, расч./факт.число наблюдений на градацию: 320/320
            44 Наим.градации: 4/5-{3.3500000, 3.4200000}, размер интервала=0.0700000, расч./факт.число наблюдений на градацию: 320/320
            45 Наим.градации: 5/5-{3.4200000, 4.0100000}, размер интервала=0.5900000, расч./факт.число наблюдений на градацию: 320/320

ОПИСАТЕЛЬНАЯ ШКАЛА:код: [  10], наим.: "ALCOHOL", набл.на шкалу (всего):1599, тип/число градаций в шкале:"Равное число событий в интервалах"/5
            46 Наим.градации: 1/5-{0.3300000, 0.5400000}, размер интервала=0.2100000, расч./факт.число наблюдений на градацию: 319/319
            47 Наим.градации: 2/5-{0.5400000, 0.5900000}, размер интервала=0.0500000, расч./факт.число наблюдений на градацию: 320/320
            48 Наим.градации: 3/5-{0.5900000, 0.6500000}, размер интервала=0.0600000, расч./факт.число наблюдений на градацию: 320/320
            49 Наим.градации: 4/5-{0.6500000, 0.7600000}, размер интервала=0.1100000, расч./факт.число наблюдений на градацию: 320/320
            50 Наим.градации: 5/5-{0.7600000, 2.0000000}, размер интервала=1.2400000, расч./факт.число наблюдений на градацию: 320/320
```

Рисунок 4. Классификационные и описательные шкалы и градации

С использованием классификационных и описательных шкал и градаций исходные данные были закодированы и получена обучающая выборка (рисунок 5):



2.4. Просмотр эвентологических баз данных (баз событий). Текущая модель: "INF1"

№	Наименование объекта	12. QUALITY	2. VOLATILE ACIDITY	3. CITRIC ACID	4. RESIDUAL SUGAR	5. CHLORIDES	6. FREE SULFUR DIOXIDE	7. TOTAL SULFUR DIOXIDE	8. DENSITY	9. PH	10. SULPHATES	11. ALCOHOL
1	1 ...	1	2	10	11	17	22	27	33	39	45	47
2	2 ...	2	3	10	11	19	25	30	34	38	42	49
3	3 ...	2	3	10	11	18	24	28	34	38	42	48
4	4 ...	2	5	6	15	17	22	29	34	39	41	47
5	5 ...	1	2	10	11	17	22	27	33	39	45	47
6	6 ...	1	2	9	11	16	22	28	33	39	45	47
7	7 ...	1	3	9	11	16	22	28	34	38	43	46
8	8 ...	3	2	9	11	16	21	28	32	36	44	46
9	9 ...	1	3	9	11	17	22	27	31	38	44	47
10	10 ...	3	2	8	14	20	22	29	35	39	43	50
11	11 ...	1	1	9	12	16	25	28	34	37	42	46
12	12 ...	3	2	8	14	20	22	29	35	39	43	50
13	13 ...	2	1	9	11	16	24	28	34	36	45	46
14	14	1	3	9	13	16	25	27	32	39	42	50
0	1	12	2	3	4	5	6	7	8	9	10	11

Рисунок 5. Обучающая выборка (фрагмент)

Обучающая выборка по сути представляет собой нормализованные исходные данные, т.е. таблицу исходных данных (таблица 1), закодированную с помощью классификационных и описательных шкал и градаций (таблицы 2 и 3).

Номера колонок в обучающей выборке (внизу) совпадают с номерами колонок в таблице 1.

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 модели

Синтез и верификация моделей осуществляется в режиме 3.5 системы «Эйдос» (рисунок 6).

Обратим внимание на то, что на рисунке 6 в правом нижнем углу окна задана опция: «Расчеты проводить на графическом процессор (GPU)».

Стадия процесса исполнения синтеза и верификации моделей и прогноз времени его окончания отображается на экранной форме (рисунок 7).

Из рисунка 7 видно, что весь процесс синтеза и верификации моделей занял 1 минуту 42 секунды. 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.

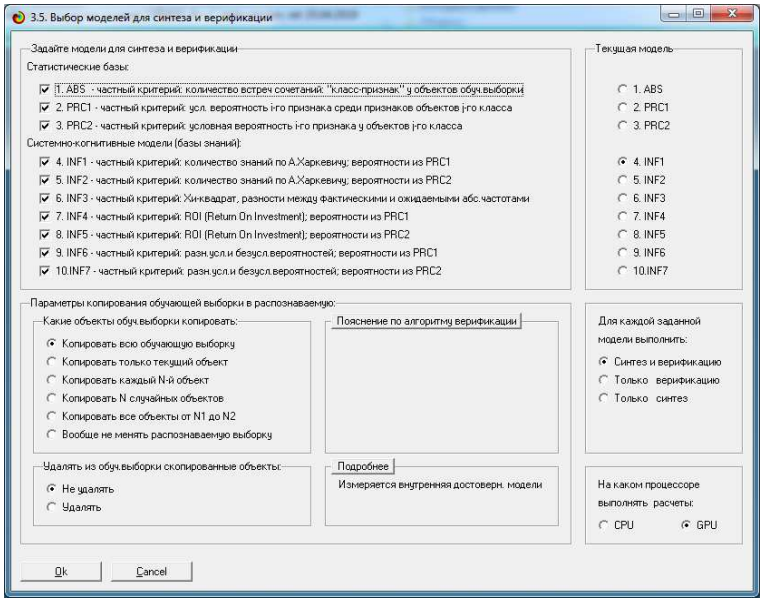


Рисунок 6. Экранная форма режима синтеза и верификации статистических и системно-когнитивных моделей системы «Эйдос»

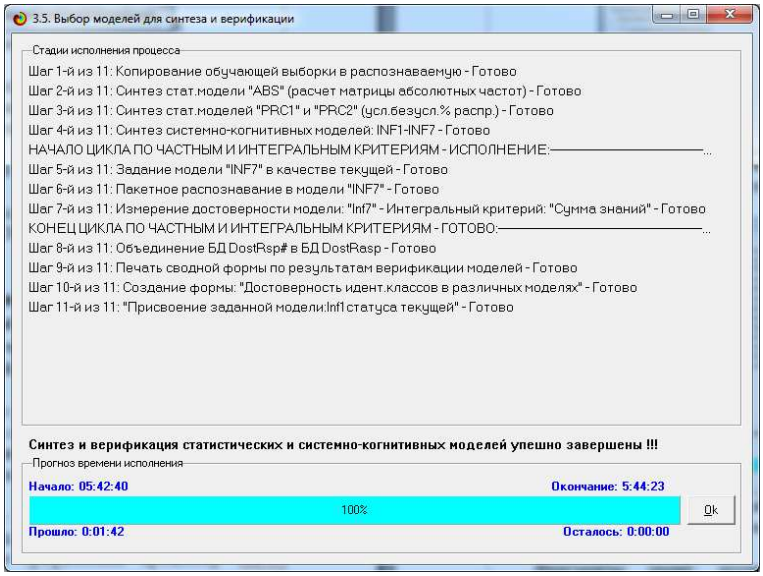


Рисунок 7. Экранная форма с отображением стадии процесса исполнения синтеза и верификации моделей и прогноза времени его окончания

Фрагменты самих созданных статистических и системно-когнитивных моделей (СК-модели) приведены на рисунках 8, 9, 10:

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

Код признака	Наименование описательной шкалы и градации	1. QUALITY 1/5 (8.4000000, 9.5000000)	2. QUALITY 2/5 (9.5000000, 9.9000000)	3. QUALITY 3/5 (9.9000000, 10.5000000)	4. QUALITY 4/5 (10.5000000, 11.3000000)	5. QUALITY 5/5 (11.3000000, 14.9000000)	Сумма	Среднее	Средн. квадр. откл.
1	VOLATILE ACIDITY-1/5-(4.6000000, 7.0000000)	64	41	66	85	106	362	72.40	24.42
2	VOLATILE ACIDITY-2/5-(7.0000000, 7.6000000)	97	60	66	50	37	310	62.00	22.44
3	VOLATILE ACIDITY-3/5-(7.6000000, 8.3000000)	119	50	38	51	39	297	59.40	33.86
4	VOLATILE ACIDITY-4/5-(8.3000000, 9.7000000)	91	51	61	51	59	313	62.60	16.52
5	VOLATILE ACIDITY-5/5-(9.7000000, 15.9000000)	65	41	73	67	71	317	63.40	12.92
6	CITRIC ACID-1/5-(0.1200000, 0.3700000)	51	25	54	95	111	336	67.20	35.03
7	CITRIC ACID-2/5-(0.3700000, 0.4700000)	87	37	67	55	70	316	63.20	18.58
8	CITRIC ACID-3/5-(0.4700000, 0.5700000)	99	43	62	54	54	312	62.40	21.55
9	CITRIC ACID-4/5-(0.5700000, 0.6600000)	106	76	60	40	38	320	64.00	28.18
10	CITRIC ACID-5/5-(0.6600000, 1.5800000)	93	62	61	60	39	315	63.00	19.30
11	RESIDUAL SUGAR-1/5-(0.0000000, 0.0700000)	70	76	56	75	63	340	68.00	8.46
12	RESIDUAL SUGAR-2/5-(0.0700000, 0.2100000)	109	64	70	43	38	324	64.80	28.17
13	RESIDUAL SUGAR-3/5-(0.2100000, 0.3200000)	111	51	73	51	34	320	64.00	29.70
14	RESIDUAL SUGAR-4/5-(0.3200000, 0.4600000)	73	16	45	68	94	296	59.20	29.78
15	RESIDUAL SUGAR-5/5-(0.4600000, 1.0000000)	73	36	60	67	83	319	63.80	17.68
16	CHLORIDES-1/5-(0.9000000, 1.8000000)	113	55	63	52	64	347	69.40	24.91
17	CHLORIDES-2/5-(1.8000000, 2.1000000)	119	74	79	71	60	403	80.60	22.57
18	CHLORIDES-3/5-(2.1000000, 2.3000000)	60	48	39	47	49	243	48.60	7.50
19	CHLORIDES-4/5-(2.3000000, 2.7000000)	65	39	57	70	60	291	58.20	11.82
20	CHLORIDES-5/5-(2.7000000, 15.5000000)	79	27	66	64	79	315	63.00	21.32
21	FREE SULFUR DIOXIDE-1/5-(0.0120000, 0.0670000)	37	21	48	85	133	324	64.80	44.81
22	FREE SULFUR DIOXIDE-2/5-(0.0670000, 0.0760000)	104	49	63	59	68	343	68.60	20.98

Рисунок 8. Матрица абсолютных частот (фрагмент)

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. QUALITY 1/5 (8.4000000, 9.5000000)	2. QUALITY 2/5 (9.5000000, 9.9000000)	3. QUALITY 3/5 (9.9000000, 10.5000000)	4. QUALITY 4/5 (10.5000000, 11.3000000)	5. QUALITY 5/5 (11.3000000, 14.9000000)	Сумма	Среднее	Средн. квадр. откл.
1	VOLATILE ACIDITY-1/5-(4.6000000, 7.0000000)	-0.104	-0.071	-0.010	0.051	0.097	-0.036	-0.007	0.083
2	VOLATILE ACIDITY-2/5-(7.0000000, 7.6000000)	0.033	0.058	0.027	-0.039	-0.118	-0.039	-0.008	0.071
3	VOLATILE ACIDITY-3/5-(7.6000000, 8.3000000)	0.092	0.025	-0.095	-0.024	-0.095	-0.098	-0.020	0.080
4	VOLATILE ACIDITY-4/5-(8.3000000, 9.7000000)	0.015	0.017	0.006	-0.037	-0.008	-0.007	-0.001	0.022
5	VOLATILE ACIDITY-5/5-(9.7000000, 15.9000000) ...	-0.068	-0.039	0.046	0.025	0.033	-0.003	-0.001	0.050
6	CITRIC ACID-1/5-(0.1200000, 0.3700000)	-0.141	-0.171	-0.040	0.095	0.126	-0.131	-0.026	0.134
7	CITRIC ACID-2/5-(0.3700000, 0.4700000)	0.002	-0.063	0.026	-0.021	0.030	-0.025	-0.005	0.038
8	CITRIC ACID-3/5-(0.4700000, 0.5700000)	0.036	-0.023	0.011	-0.023	-0.029	-0.028	-0.006	0.028
9	CITRIC ACID-4/5-(0.5700000, 0.6600000)	0.047	0.107	-0.003	-0.101	-0.119	-0.069	-0.014	0.096
10	CITRIC ACID-5/5-(0.6600000, 1.5800000)	0.019	0.062	0.004	0.000	-0.109	-0.023	-0.005	0.063
11	RESIDUAL SUGAR-1/5-(0.0000000, 0.0700000) ...	-0.067	0.093	-0.034	0.036	-0.012	0.014	0.003	0.063
12	RESIDUAL SUGAR-2/5-(0.0700000, 0.2100000) ...	0.050	0.063	0.031	-0.086	-0.122	-0.064	-0.013	0.085
13	RESIDUAL SUGAR-3/5-(0.2100000, 0.3200000) ...	0.058	0.011	0.044	-0.042	-0.146	-0.075	-0.015	0.083
14	RESIDUAL SUGAR-4/5-(0.3200000, 0.4600000) ...	-0.024	-0.248	-0.054	0.045	0.117	-0.163	-0.033	0.137
15	RESIDUAL SUGAR-5/5-(0.4600000, 1.0000000) ...	-0.042	-0.071	-0.003	0.024	0.069	-0.023	-0.005	0.055
16	CHLORIDES-1/5-(0.9000000, 1.8000000)	0.043	0.010	-0.011	-0.057	-0.014	-0.029	-0.006	0.036
17	CHLORIDES-2/5-(1.8000000, 2.1000000)	0.019	0.045	0.007	-0.018	-0.065	-0.011	-0.002	0.042
18	CHLORIDES-3/5-(2.1000000, 2.3000000)	-0.024	0.063	-0.041	0.004	0.008	0.010	0.002	0.039
19	CHLORIDES-4/5-(2.3000000, 2.7000000)	-0.048	-0.030	0.007	0.056	0.013	-0.001	0.000	0.041
20	CHLORIDES-5/5-(2.7000000, 15.5000000)	-0.020	-0.137	0.023	0.016	0.060	-0.058	-0.012	0.076
21	FREE SULFUR DIOXIDE-1/5-(0.0120000, 0.06700...	-0.209	-0.204	-0.060	0.077	0.178	-0.217	-0.043	0.171
22	FREE SULFUR DIOXIDE-2/5-(0.0670000, 0.07600...	0.025	-0.015	-0.008	-0.024	0.004	-0.018	-0.004	0.019

Рисунок 9. Матрица информативностей INF1 (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. QUALITY 1/5 (8.4000000, 9.5000000)	2. QUALITY 2/5 (9.5000000, 9.9000000)	3. QUALITY 3/5 (9.9000000, 10.5000000)	4. QUALITY 4/5 (10.5000000, 11.3000000)	5. QUALITY 5/5 (11.3000000, 14.9000000)	Сумма	Среднее	Средн. квадр. откл.
1	VOLATILE ACIDITY-1/5-(4.6000000, 7.0000000)	-34.707	-14.013	-2.823	16.177	35.366	0.000	0.000	27.025
2	VOLATILE ACIDITY-2/5-(7.0000000, 7.6000000)	12.472	12.889	7.063	-8.937	-23.488	0.000	0.000	15.836
3	VOLATILE ACIDITY-3/5-(7.6000000, 8.3000000)	38.017	4.865	-18.465	-5.465	-18.951	0.000	0.000	23.446
4	VOLATILE ACIDITY-4/5-(8.3000000, 9.7000000)	5.654	3.433	1.493	-8.507	-2.073	0.000	0.000	5.538
5	VOLATILE ACIDITY-5/5-(9.7000000, 15.9000000) ...	-21.437	-7.174	12.732	6.732	9.146	0.000		14.161
6	CITRIC ACID-1/5-(0.1200000, 0.3700000)	-40.617	-26.062	-9.880	31.120	45.439	0.000	0.000	36.945
7	CITRIC ACID-2/5-(0.3700000, 0.4700000)	0.836	-11.023	6.922	-5.078	8.341	0.000		8.147
8	CITRIC ACID-3/5-(0.4700000, 0.5700000)	13.927	-4.415	2.683	-5.317	-6.878	0.000		8.606
9	CITRIC ACID-4/5-(0.5700000, 0.6600000)	18.745	27.370	-0.838	-20.838	-24.439	0.000	0.000	23.090
10	CITRIC ACID-5/5-(0.6600000, 1.5800000)	7.109	14.129	1.113	0.113	-22.463	0.000	0.000	13.748
11	RESIDUAL SUGAR-1/5-(0.0000000, 0.0700000) ...	-22.708	24.330	-8.640	10.360	-3.341	0.000		18.033
12	RESIDUAL SUGAR-2/5-(0.0700000, 0.2100000) ...	20.655	14.762	8.402	-18.598	-25.220			20.598
13	RESIDUAL SUGAR-3/5-(0.2100000, 0.3200000) ...	23.745	2.370	12.162	-9.838	-28.439			20.143
14	RESIDUAL SUGAR-4/5-(0.3200000, 0.4600000) ...	-7.710	-28.983	-11.275	11.725	36.244	0.000	0.000	24.888
15	RESIDUAL SUGAR-5/5-(0.4600000, 1.0000000) ...	-13.982	-12.478	-0.648	6.352	20.756			14.342
16	CHLORIDES-1/5-(0.9000000, 1.8000000)	18.383	2.266	-2.971	-13.971	-3.707	0.000	0.000	11.841
17	CHLORIDES-2/5-(1.8000000, 2.1000000)	9.114	12.756	2.382	-5.618	-18.634	0.000	0.000	12.552
18	CHLORIDES-3/5-(2.1000000, 2.3000000)	-6.259	11.071	-7.199	0.801	1.585	0.000	0.000	7.361
19	CHLORIDES-4/5-(2.3000000, 2.7000000)	-14.347	-5.223	1.675	14.675	3.220	0.000		10.743
20	CHLORIDES-5/5-(2.7000000, 15.5000000)	-6.891	-20.871	6.113	4.113	17.537	0.000	0.000	14.533
21	FREE SULFUR DIOXIDE-1/5-(0.0120000, 0.06700...	-51.345	-28.238	-13.598	23.402	69.780			47.528
22	FREE SULFUR DIOXIDE-2/5-(0.0670000, 0.07600...	10.474	-3.126	-2.211	-6.211	1.073	0.000	0.000	6.405

Рисунок 10. Модель INF3 (фрагмент)

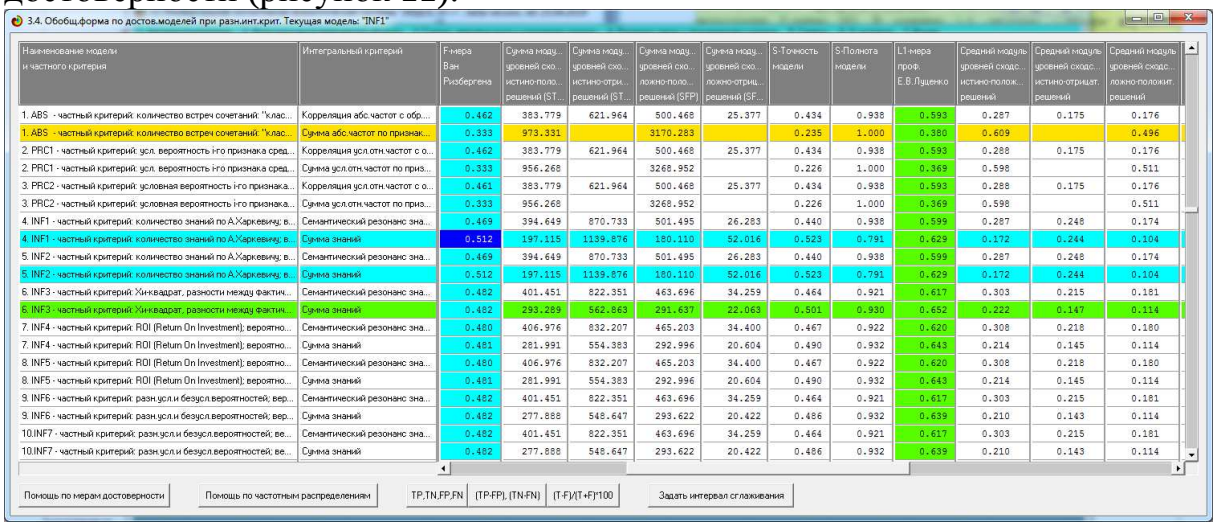
Отметим, что в АСК-анализе и СК-моделях степень выраженности различных физико-химических свойств вина рассматривается с одной единственной точки зрения: какое ***количество информации*** содержится в них о том, какую оценку качества даст этому вину сомелье [5]. Поэтому не играет никакой роли в каких единицах измерения измеряются те или иные

физико-химические свойства вина, а также в каких единицах измерения дает оценку качества вина сомелье [5]. Это решение проблемы сопоставимости в АСК-анализе и системе «Эйдос», отличающее их от других интеллектуальных технологий.

Верификация статистических и системно-когнитивных моделей

Отметим, что для исследования достоверности мы используем СК-модели с 5 адаптивными градациями классификационной шкалы и 50 адаптивными градациями в описательных шкалах, т.к. при таком варианте типа и числа градаций результаты более наглядные, чем при других (из исследованных).

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, смягчающие и преодолевающие недостатки F-меры [6]. В режиме 3.4 системы «Эйдос» изучается достоверность каждой частной модели в соответствии с этими мерами достоверности (рисунок 11).



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

Наименование модели и частного критерия	Интегральный критерий	F-мера Ван Ризбергена	Сумма моду... уровней ско... истинно-поло... решений (ST...	Сумма моду... уровней ско... истинно-отри... решений (ST...	Сумма моду... уровней ско... ложно-поло... решений (SFP)	Сумма моду... уровней ско... ложно-отриц... решений (SF...	S-Точность модели	S-Полнота модели	L1-мера проф. Е.В.Луценко	Средний модуль уровней сходс... истинно-положит... решений	Средний модуль уровней сходс... истинно-отрицат... решений	Средний модуль уровней сходс... ложно-положит... решений
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Корреляция абс.частот с обр...	0.462	383.779	621.964	500.468	25.377	0.434	0.938	0.593	0.287	0.175	0.176
1. ABS - частный критерий: количество встреч сочетаний: "клас...	Сумма абс.частот по признак...	0.333	973.331		3170.283		0.235	1.000	0.380	0.609		0.496
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн.частот с о...	0.462	383.779	621.964	500.468	25.377	0.434	0.938	0.593	0.288	0.175	0.176
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн.частот по приз...	0.333	956.268		3268.952		0.226	1.000	0.369	0.598		0.511
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн.частот с о...	0.461	383.779	621.964	500.468	25.377	0.434	0.938	0.593	0.288	0.175	0.176
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн.частот по приз...	0.333	956.268		3268.952		0.226	1.000	0.369	0.598		0.511
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	0.469	394.649	870.733	501.495	26.283	0.440	0.938	0.599	0.287	0.248	0.174
4. INF1 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	0.512	197.115	1139.876	180.110	52.016	0.523	0.791	0.629	0.172	0.244	0.104
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Семантический резонанс зна...	0.469	394.649	870.733	501.495	26.283	0.440	0.938	0.599	0.287	0.248	0.174
5. INF2 - частный критерий: количество знаний по А.Харкевичу: в...	Сумма знаний	0.512	197.115	1139.876	180.110	52.016	0.523	0.791	0.629	0.172	0.244	0.104
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Семантический резонанс зна...	0.482	401.451	822.351	463.696	34.259	0.464	0.921	0.617	0.303	0.215	0.181
6. INF3 - частный критерий: Хи-квадрат, разности между фактич...	Сумма знаний	0.482	293.289	562.863	291.637	22.063	0.501	0.930	0.652	0.222	0.147	0.114
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.480	406.976	832.207	465.203	34.400	0.467	0.922	0.620	0.308	0.218	0.180
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.481	281.991	554.383	292.996	20.604	0.490	0.932	0.643	0.214	0.145	0.114
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.480	406.976	832.207	465.203	34.400	0.467	0.922	0.620	0.308	0.218	0.180
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.481	281.991	554.383	292.996	20.604	0.490	0.932	0.643	0.214	0.145	0.114
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Семантический резонанс зна...	0.482	401.451	822.351	463.696	34.259	0.464	0.921	0.617	0.303	0.215	0.181
9. INF6 - частный критерий: разн.усл.и безусл.вероятностей; вер...	Сумма знаний	0.482	277.888	548.647	293.622	20.422	0.486	0.932	0.639	0.210	0.143	0.114
10.INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Семантический резонанс зна...	0.482	401.451	822.351	463.696	34.259	0.464	0.921	0.617	0.303	0.215	0.181
10.INF7 - частный критерий: разн.усл.и безусл.вероятностей; ве...	Сумма знаний	0.482	277.888	548.647	293.622	20.422	0.486	0.932	0.639	0.210	0.143	0.114

Помощь по мерам достоверности | Помощь по частотным распределениям | TP,TN,FP,FN | (TP-FP), (TN-FN) | (T-F)/(T+F)*100 | Задать интервал сглаживания

Рисунок 11. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена и L1- и L2-критериям проф.Е.В.Луценко [34]

Из рисунка 10 мы видим, что в данном интеллектуальном приложении по F-критерию Ван Ризбергена наиболее достоверной (F=0,626 при максимуме 1,000) является модель INF1 с интегральным критерием «Сумма знаний» (критерий L1 для этой модели имеет значение 0,771), ***что подтверждает наличие зависимости субъективных сомелье-оценок качества вина от его объективных физико-химических свойств***. Однако сами значения F- и L1 критериев достоверности моделей говорят о том, что эта зависимость эта не очень сильная.

Отметим, что различные классы по качеству вина прогнозируются с разной достоверностью в разных моделях (т.е. с разными частными критериями) и с разными интегральными критериями (таблица 4):

Таблица 4 – Достоверность идентификации качества вина по F-критерию в разных моделях и с разными интегральными критериями

Код класса	1	2	3	4	5	
Наименование класса	QUALITY-1/5-{8.40,9.50}	QUALITY-2/5-{9.50,9.90}	QUALITY-3/5-{9.90,10.50}	QUALITY-4/5-{10.50,11.30}	QUALITY-5/5-{11.30,14.90}	Средневзве-шенное
Максимальная достоверность	0,701	0,517	0,574	0,601	0,721	0,626
Модель с максимальной достоверностью	INF1	INF1	INF1	INF1	INF1	INF1
Интегральный критерий с максимальной достоверностью	Сумма	Сумма	Сумма	Сумма	Сумма	Сумма
ABSK	0,448	0,283	0,332	0,341	0,368	0,356
PRC1K	0,448	0,283	0,332	0,341	0,368	0,356
PRC2K	0,448	0,283	0,332	0,341	0,368	0,356
INF1K	0,619	0,430	0,482	0,540	0,617	0,539
INF2K	0,619	0,430	0,482	0,540	0,617	0,539
INF3K	0,632	0,421	0,492	0,529	0,620	0,540
INF4K	0,643	0,455	0,516	0,541	0,632	0,560
INF5K	0,643	0,455	0,516	0,541	0,632	0,560
INF6K	0,630	0,408	0,502	0,528	0,611	0,537
INF7K	0,630	0,408	0,502	0,528	0,611	0,537
ABSI	0,429	0,264	0,319	0,319	0,327	0,333
PRC1I	0,429	0,264	0,319	0,319	0,327	0,333
PRC2I	0,429	0,264	0,319	0,319	0,327	0,333
INF1I	**0,701**	**0,517**	**0,574**	**0,601**	**0,721**	**0,626**
INF2I	0,701	0,517	0,574	0,601	0,721	0,626
INF3I	0,632	0,421	0,492	0,529	0,620	0,540
INF4I	0,643	0,414	0,498	0,520	0,597	0,536
INF5I	0,643	0,414	0,498	0,520	0,597	0,536
INF6I	0,621	0,380	0,486	0,516	0,600	0,519
INF7I	0,621	0,380	0,486	0,516	0,600	0,519

Из таблицы 4 видно, что:

- по F-критерию наиболее достоверной при идентификации со всеми классами является модель INF1 с интегральным критерием «Сумма знаний»;

- с достоверностью выше средневзвешенной идентифицируются классы с *максимальным* и *минимальным* качеством вина, а ниже средневзвешенной – классы со средним качеством вина, а также ниже и выше среднего (средневзвешенная достоверность является очевидной, естественной и обоснованной базой сравнения).

По сути это означает, что сомелье наиболее четко различают только очень хорошее и откровенно плохое вино, а с винами среднего качества они чаще расходятся (путаются) в оценках (т.е. едва различают вина ниже среднего, среднего и выше среднего качества).

Это позволяет обоснованно предположить, что субъективные сомелье-оценки качества вина менее достоверны, чем оценки, основанные на технологиях искусственного интеллекта и объективных физико-химических свойствах вина. Иначе говоря, учитывая точность сомелье-

оценок, можно предположить, что даже 5 градаций качества вина для них это слишком много, и вполне можно было бы ограничиться 3 или даже 2 градациями. Сейчас же используется шкала, в которой 100 (!) градаций, что, конечно, очень сильно переоценивает возможности сомелье и может рассматриваться не как объективная оценка их возможностей, а скорее как большой им комплемент и аванс, а вероятнее просто маркетинговым ходом. А вот для объективной оценки физико-химических свойств вина использование 50 или даже 100 градаций (при достаточно большом объеме выборки) вполне оправдано и обоснованно.

На рисунке 12 приведены частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности, по результатам прогнозирования субъективных сомелье-оценок качества вина на основе его объективных физико-химических свойств в СК-модели INF1 по данным обучающей выборки, включающей 1599 наблюдений.

А

Б

В

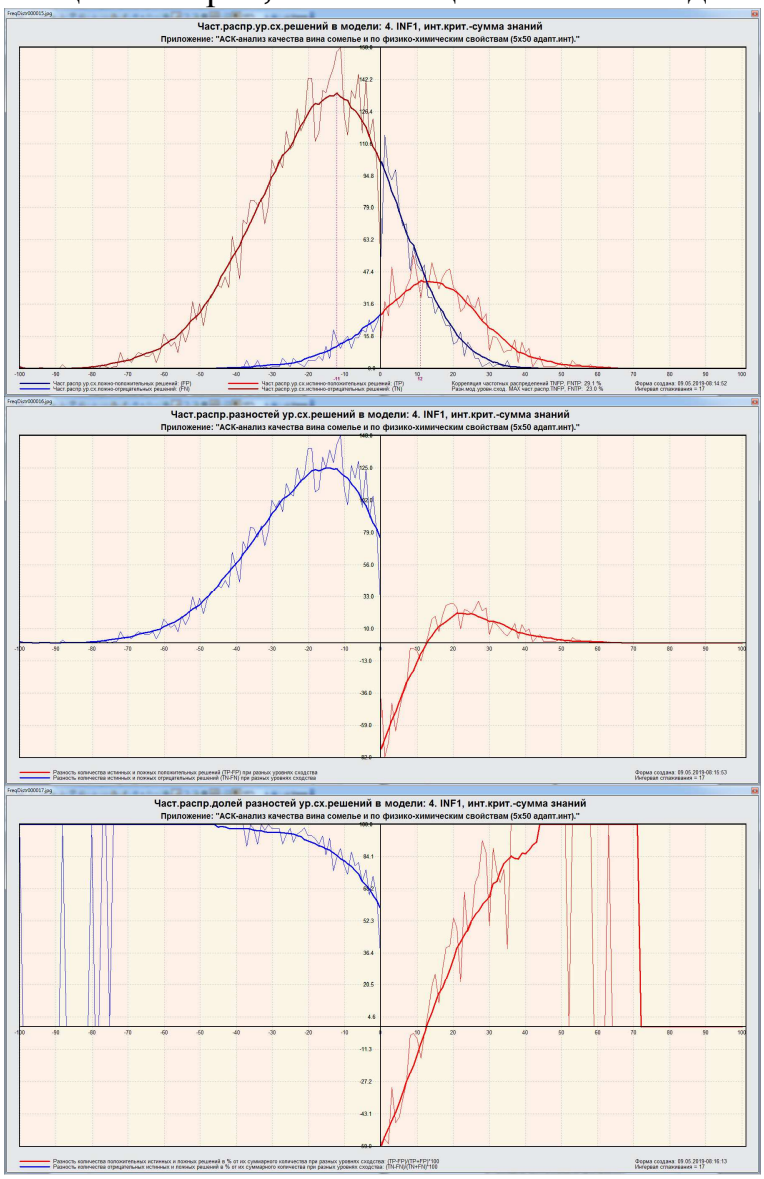
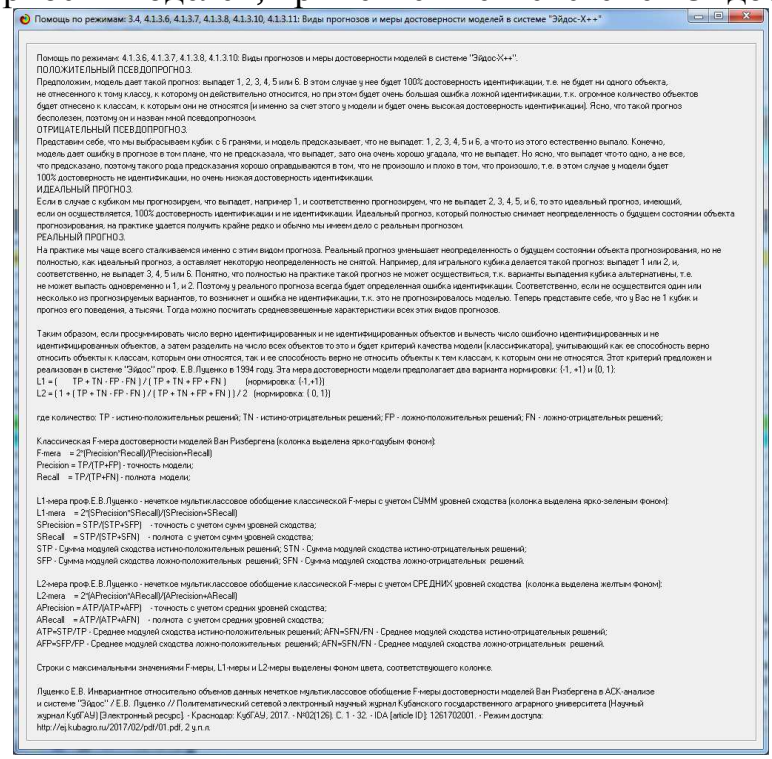


Рисунок 12. Частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности в СК-модели Inf1

Рисунок 12-А содержит изображения двух частотных распределений, очень похожих на нормальные, сдвинутых относительно друг друга по фазе. Левое, 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. Видно, что для отрицательных решений количество истинных решений всегда значительно превосходит количество ложных решений, причем при уровнях различия больше примерно 40% ложные отрицательные решения вообще отсутствуют. Видно также, что для положительных решений картина более сложная и включает 3 диапазона уровней сходства 1) при уровнях сходства от 0% до примерно 14% количество ложных решений больше числа истинных; 2) при уровнях сходства от 14% до примерно 44% есть и истинные и ложные положительные решения, но число истинных больше числа ложных; 3) при уровнях сходства выше 44% встречаются только истинные решения.

Из рисунка 12-В видно, что чем выше уровень сходства, тем больше доля истинных решений; при уровнях сходства выше примерно 14% число истинных решений превосходит число ложных решений (это видно также и из рисунка 12-Б).

На рисунке 13 приведен Help по режиму 3.4, в котором описаны меры достоверности моделей, применяемые в системе «Эйдос»:


Помощь по режимам: 3.4, 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10, 4.1.3.11: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++"

Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++":
ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Предположим, модель дает такой прогноз: выпадет 1, 2, 3, 4, 5 или 6. В этом случае у нее будет 100% достоверность идентификации, т.е. не будет ни одного объекта, не отнесенного к тому классу, к которому он действительно относится, но при этом будет очень большая ошибка ложной идентификации, т.к. огромное количество объектов будет отнесено к классам, к которым они не относятся (и именно за счет этого у модели и будет очень высокая достоверность идентификации). Ясно, что такой прогноз бесполезен, поэтому он и назван ложной псевдопрогнозом.
ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Представим себе, что мы выбрасываем кубик с 6 гранями, и модель предсказывает, что не выпадет: 1, 2, 3, 4, 5 и 6, а что-то из этого естественно выпало. Конечно, модель дает ошибку в прогнозе в том плане, что не предсказала, что выпадет, зато она очень хорошо угадала, что не выпадет. Но ясно, что выпадет что-то одно, а не все, что предсказано, поэтому такого рода предсказания хорошо оправдываются в том, что не произошло и плохо в том, что произошло, т.е. в этом случае у модели будет 100% достоверность не идентификации, но очень низкая достоверность идентификации.
ИДЕАЛЬНЫЙ ПРОГНОЗ.
Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.
РЕАЛЬНЫЙ ПРОГНОЗ.
На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один из несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представьте себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В.Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {-1, +1} и {0, 1}:
L1 = (TP + TN - FP - FN) / (TP + TN + FP + FN) {нормировка: {-1,+1}}
L2 = (1 + [TP + TN - FP - FN] / [TP + TN + FP + FN]) / 2 {нормировка: { 0, 1}}

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

Классическая F-мера достоверности моделей Ван Ризбергена (колонка выделена ярко-голубым фоном):
F-mera = 2*(Precision*Recall)/(Precision+Recall)
Precision = TP/(TP+FP) - точность модели;
Recall = TP/(TP+FN) - полнота модели;

L1-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СУММ уровней сходства (колонка выделена ярко-зеленым фоном):
L1-mera = 2*(SPrecision*SRecall)/(SPrecision+SRecall)
SPrecision = STP/(STP+SFP) - точность с учетом сумм уровней сходства;
SRecall = STP/(STP+SFN) - полнота с учетом сумм уровней сходства;
STP - Сумма модулей сходства истинно-положительных решений; STN - Сумма модулей сходства истинно-отрицательных решений;
SFP - Сумма модулей сходства ложно-положительных решений; SFN - Сумма модулей сходства ложно-отрицательных решений.

L2-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желтым фоном):
L2-mera = 2*(APrecision*ARecall)/(APrecision+ARecall)
APrecision = ATP/(ATP+AFP) - точность с учетом средних уровней сходства;
ARecall = ATP/(ATP+AFN) - полнота с учетом средних уровней сходства;
ATP=STP/TP - Среднее модулей сходства истинно-положительных решений; AFN=SFN/FN - Среднее модулей сходства истинно-отрицательных решений;
AFP=SFP/FP - Среднее модулей сходства ложно-положительных решений; AFN=SFN/FN - Среднее модулей сходства ложно-отрицательных решений.

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

Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе "Эйдос" / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №02(126). С. 1 - 32. - IDA [article ID]: 1261702001. - Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf, 2 у.п.л.

Рисунок 13. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена и L1- и L2-критериям проф.Е.В.Луценко [6]

Выбор наиболее достоверной модели и присвоение ей статуса текущей

В соответствии со схемой обработки данных, информации и знаний в системе «Эйдос» (рисунок 1), присвоим СК-модели INF1 статус текущей модели. Для это запустим режим 5.6 с параметрами, приведенными на экранной форме (рисунок 14):

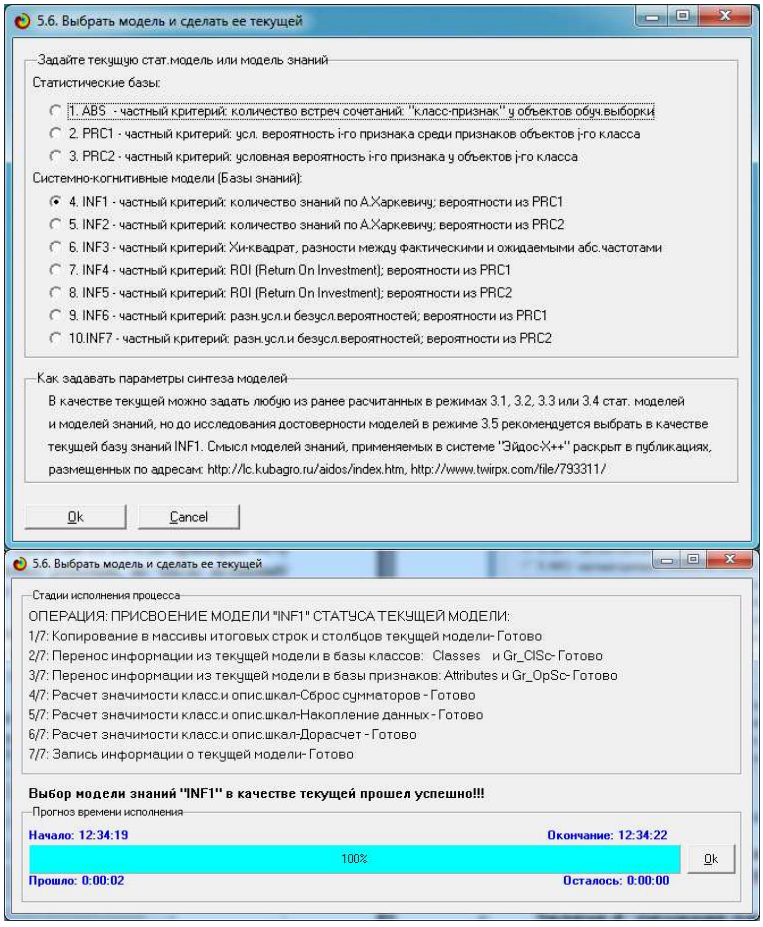


Рисунок 14. Экранные формы придания наиболее достоверной по L2-критерию СК-модели Inf4 статуса текущей модели

Задача 4: решение различных задач в наиболее достоверной модели

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

Решим задачу прогнозирования обучающей выборки в наиболее достоверной СК-модели INF1 на GPU. Для этого запустим режим 4.1.2 (рисунок 15).

Из рисунка 15 видно, что идентификация 1599 наблюдений заняло 8 секунд.

Отметим, что 99,999% этого времени заняло не само прогнозирование на GPU, а создание 10 выходных форм на основе результатов этого прогнозирования. Эти формы отражают результаты прогнозирования в различных разрезах и обобщениях:

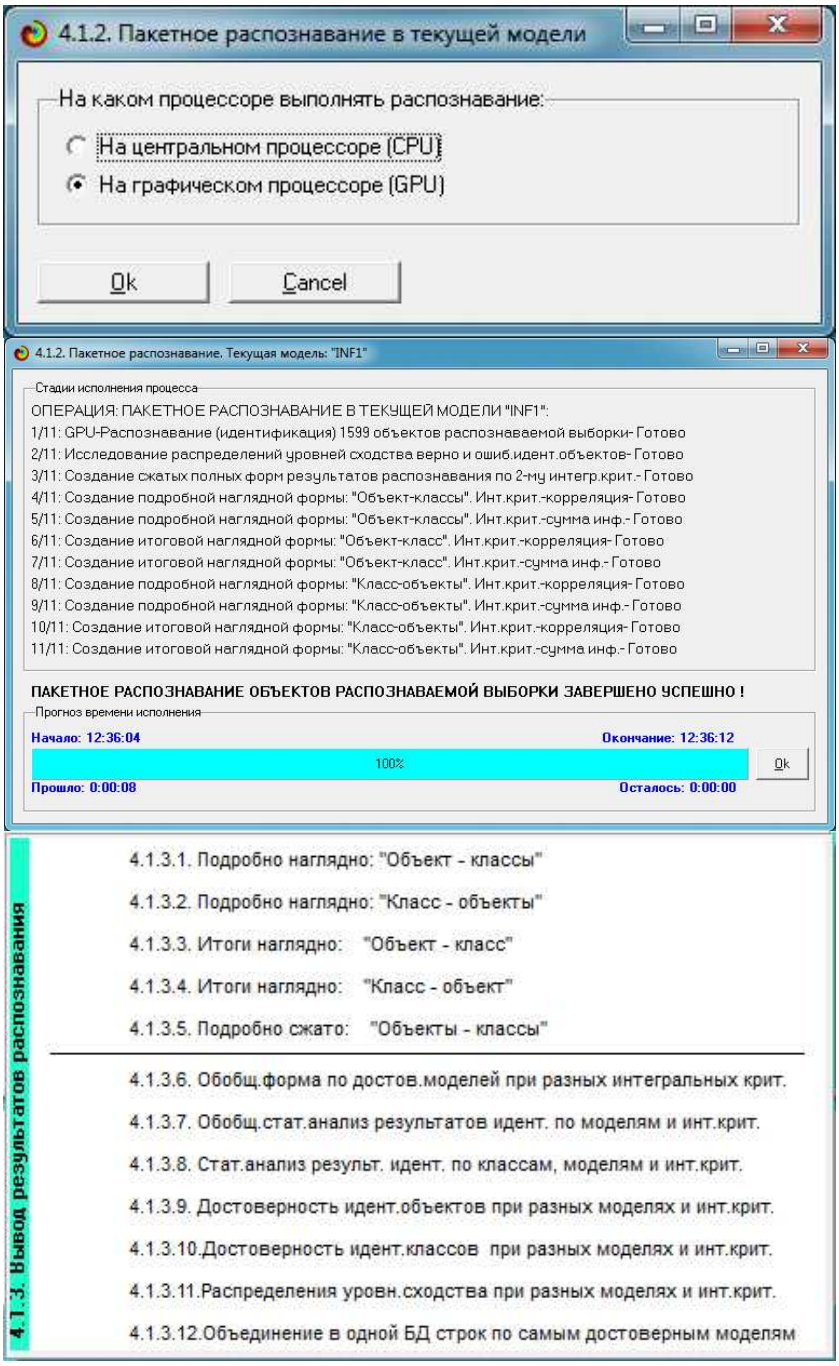


Рисунок 15. Экранные формы отображения процесса решения задачи прогнозирования в текущей модели

Приведем две из этих 10 форм: 4.1.3.1 и 4.1.3.2 (рисунок 16).

Символ «√» стоит против тех результатов прогнозирования, которые подтвердились на опыте, т.е. соответствуют факту. Из рисунка 15 видно, что результаты прогнозирования являются очень хорошими, естественно при учете информации из рисунка 12-Б о том, что достоверные прогнозы в данной модели имеют уровень сходства выше 14%, т.е. по сути прогнозы с более низки уровнем сходства надо просто игнорировать.

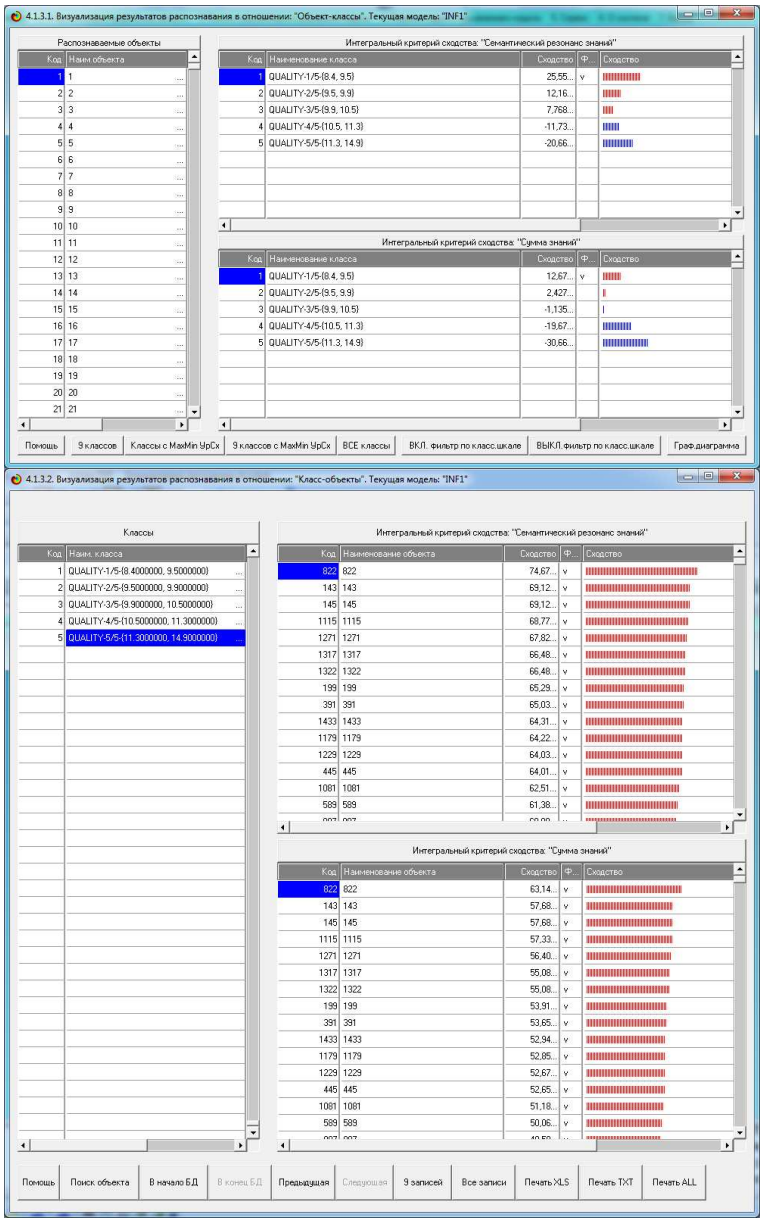


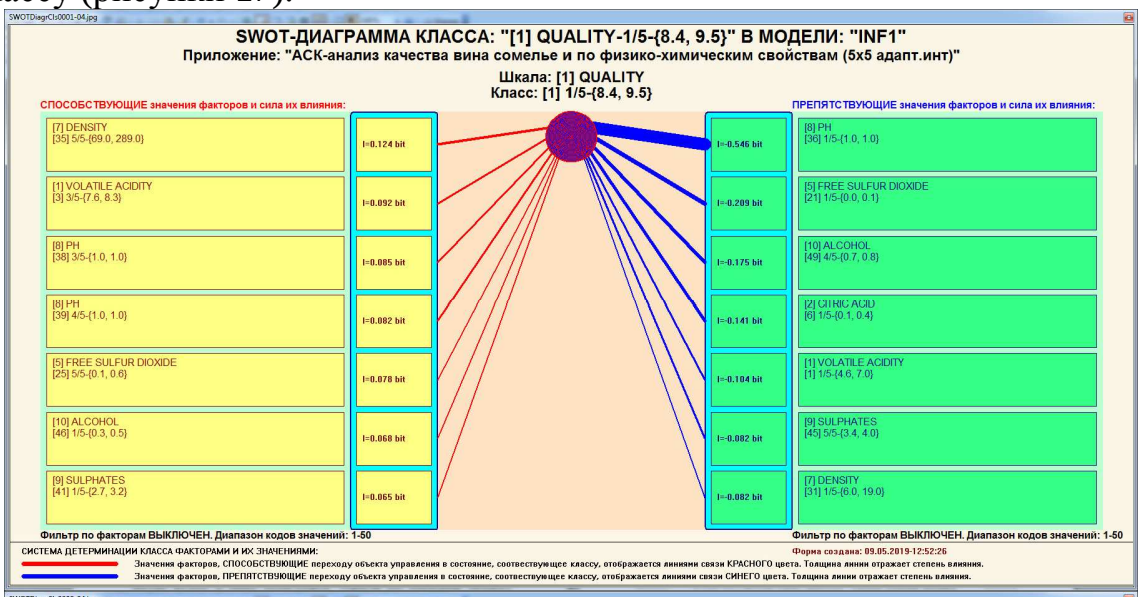
Рисунок 16. Выходные формы по результатам идентификации качества вина на основе его физико-химических свойств

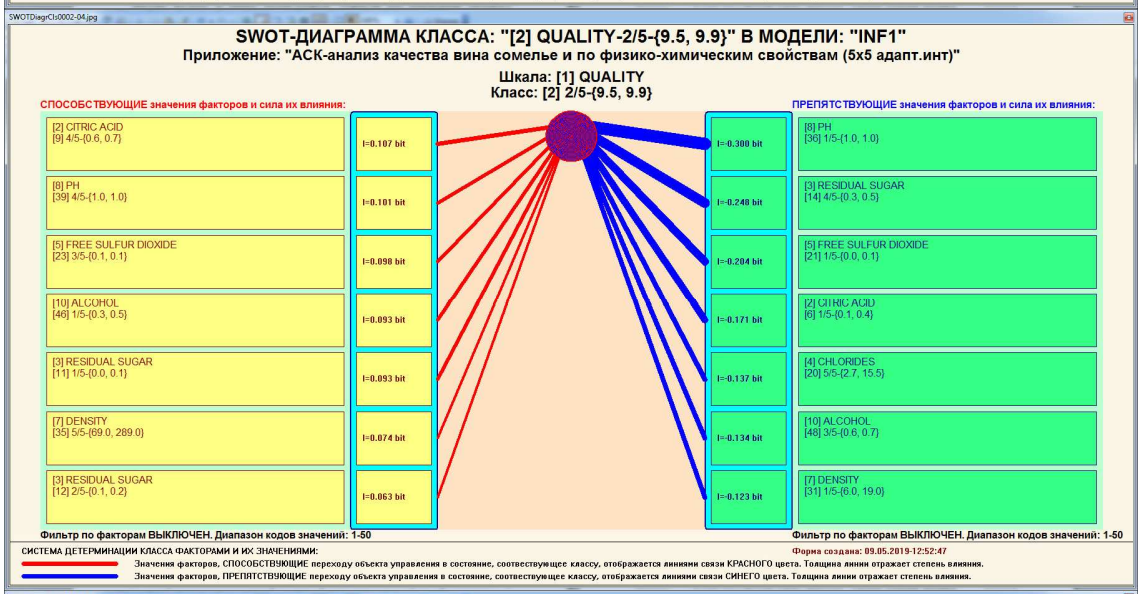
Подзадача 4.2. Поддержка принятия решений (SWOT-анализ)

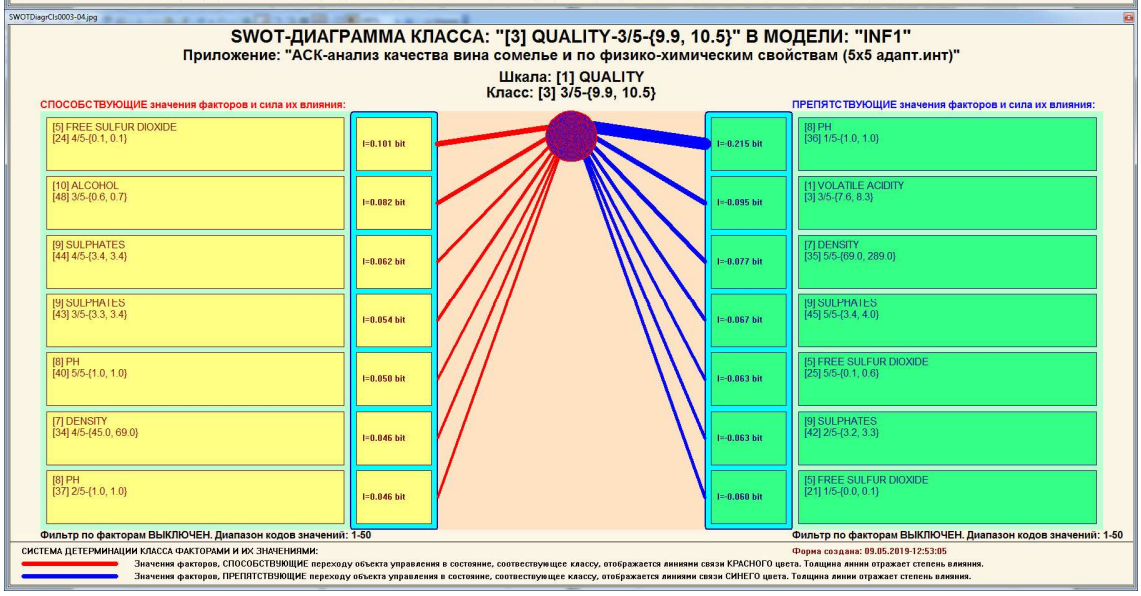
При принятии решений определяется сила и направление влияния факторов на принадлежность состояний объекта моделирования к тем или иным классам, соответствующим различным будущим состояниям. По сути это решение задачи SWOT-анализа [7]. Применительно к задаче, решаемой в данной работе, SWOT-анализ показывает степень влияния различных значений физико-химических свойств вина на его принадлежность к различным классам по качеству с точки зрения сомелье.

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 (рисунки 17).







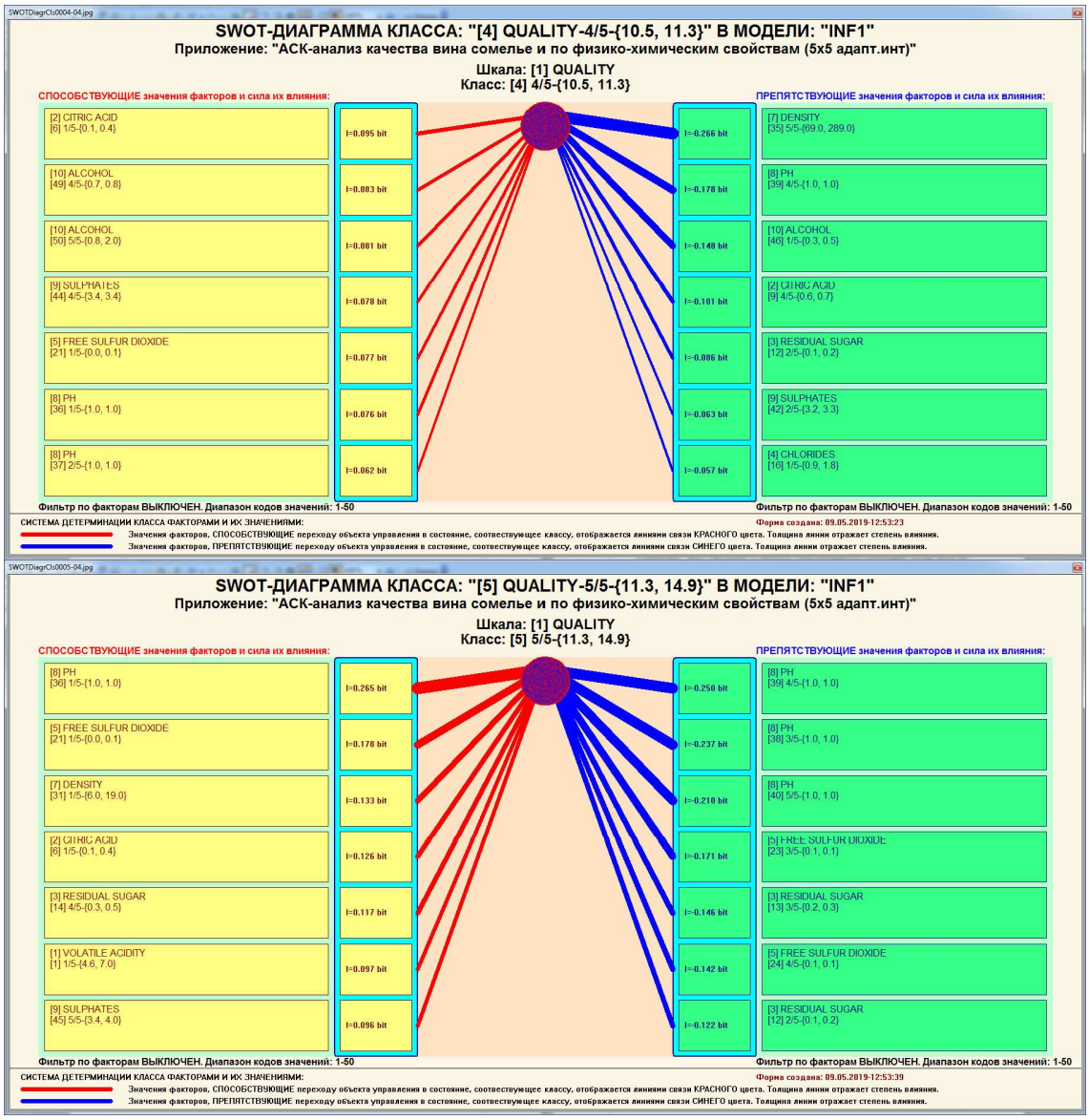
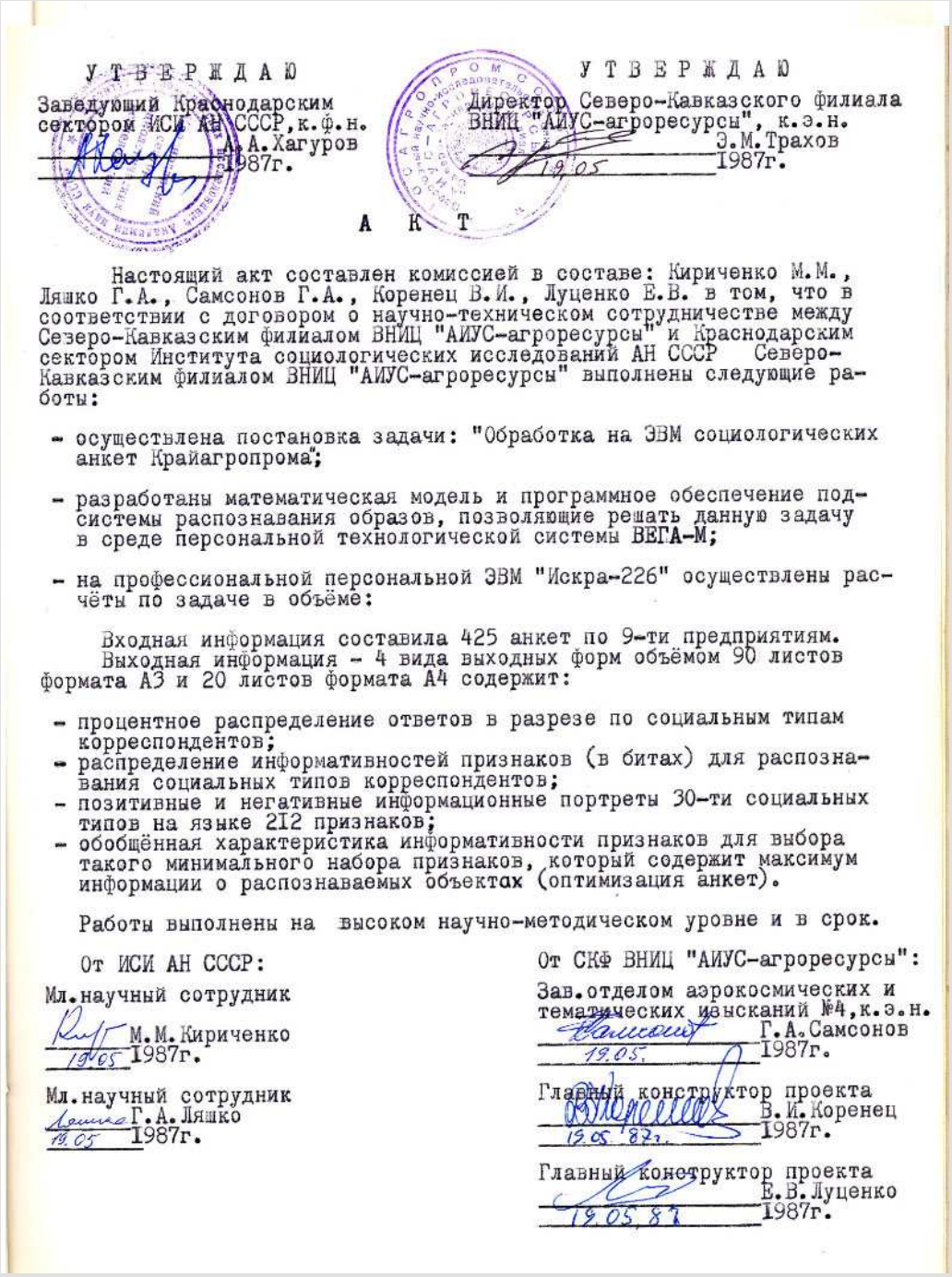


Рисунок 17. SWOT-диаграммы, отражающие степень влияния различных значений физико-химических свойств вина на его принадлежность к различным классам по качеству

Выходные формы, приведенные на рисунках 17, как говорят «интуитивно понятны», т.е. не требуют особых комментариев. Отметим лишь, что информация быть приведена не только в приведенных, но и во многих других табличных и графических формах, которые в данной работе не приводятся только из-за ограниченности ее размера. В частности в этих формах может быть выведена значительно более полная информация (в т.ч. вообще вся имеющая в модели). Подобная подробная информация содержится в базах данных, расположенных по пути: c:\Aidos-X\AID_DATA\A0000003\System\SWOTCls####Inf1.DBF, где: «####» – код класса с ведущими нулями. Эти базы открываются в MS Excel.

Отметим также, что система «Эйдос» обеспечивала решение этой ***всегда***, т.е. даже в самых ранних DOS-версиях и в реализациях системы

«Эйдос» на других языках и типах компьютеров. Например, первый акт внедрения системы «Эйдос», где об этом упоминается в явном виде, датируется 1987 годом, а первый подобный расчет относится к 1981 году.



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

УТВЕРЖДАЮ
Директор Северо-Кавказского филиала ВНИЦ "АИУС-агроресурсы", к.э.н.
Э.М.Трахов
19.05 1987г.

А К Т

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

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

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

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

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

От ИСИ АН СССР:

Мл.научный сотрудник
М.М.Кириченко
19.05 1987г.

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

От СКФ ВНИЦ "АИУС-агроресурсы":

Зав.отделом аэрокосмических и тематических изысканий №4, к.э.н.
Г.А.Самсонов
19.05. 1987г.

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

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

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

На рисунке 18 приведены примеры инвертированных SWOT-диаграмм, отражающих влияние заданных значений физико-химических свойств вина на сомелье-оценку его качества:

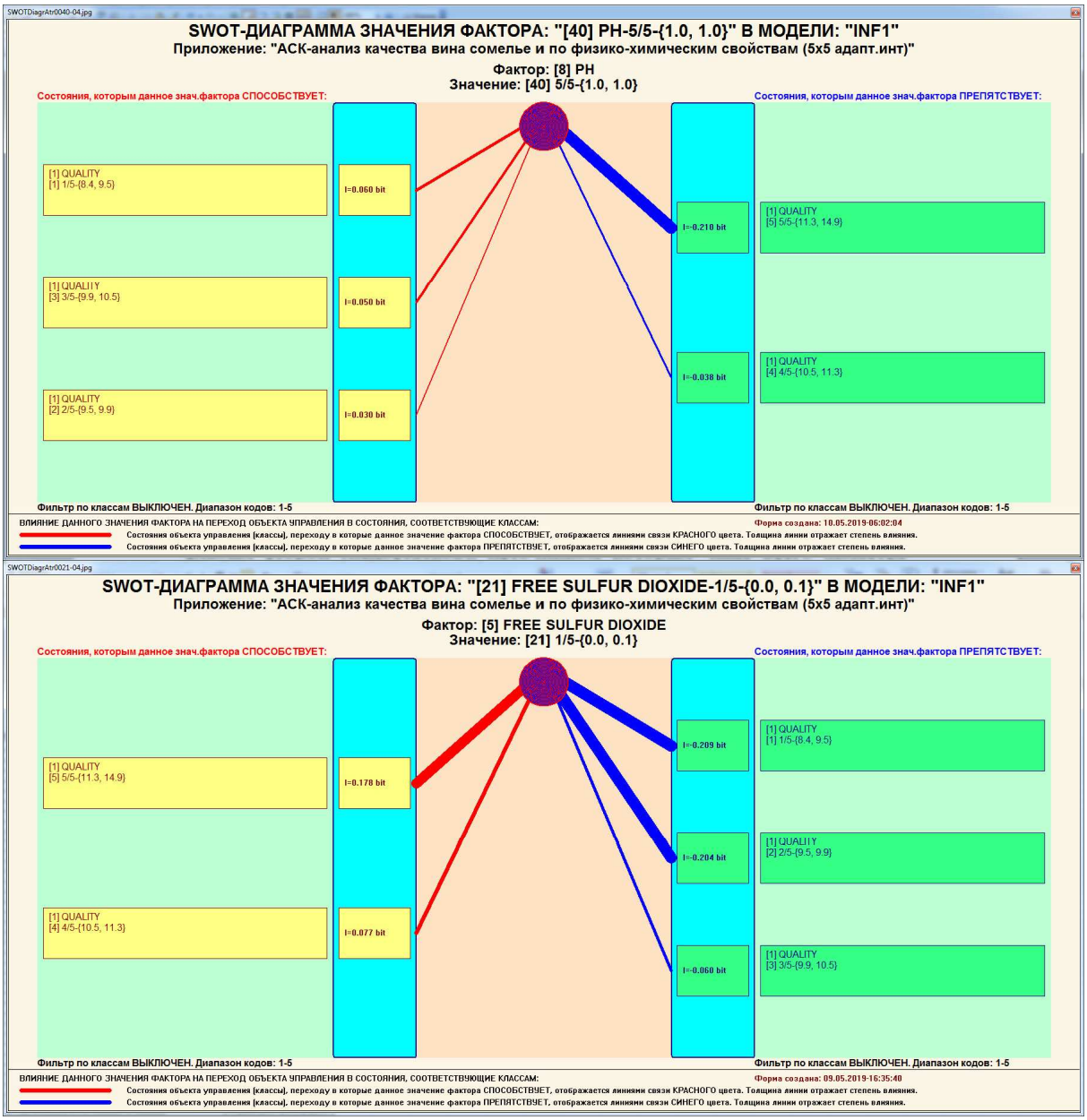


Рисунок 18. Примеры инвертированных SWOT-диаграмм, отражающих влияние заданных значений физико-химических свойств вина на сомелье-оценку его качества

Из рисунка 18 видно, что кислые вина сомелье оценивают как вина низкого и среднего качества, а вина с низким содержанием свободного диоксида серы, как вина высокого качества.

В заключение отметим, что SWOT-анализ является широко известным и общепризнанным метод 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 она малоизвестна – это интеллектуальная система «Эйдос» [6].

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 19):

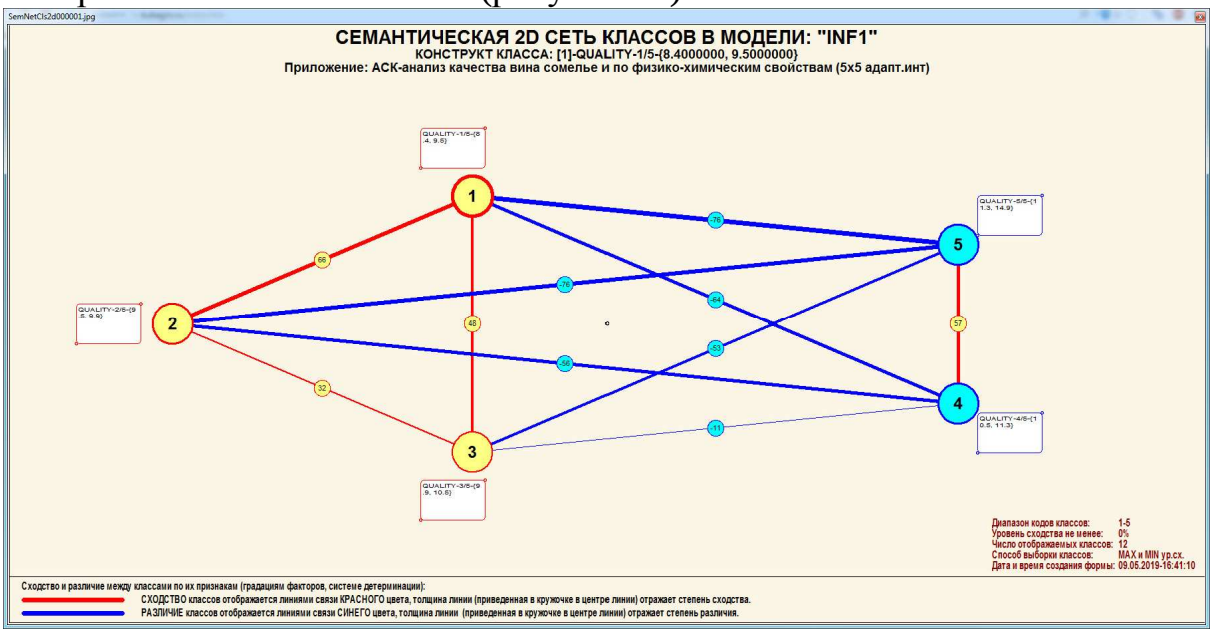


Рисунок 19. Когнитивная диаграмма классов

Отметим также, что на когнитивной диаграмме, приведенной на рисунке 19, показаны ***количественные*** оценки сходства/различия

различных опасных природно-климатических явлений, 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.

В системе «Эйлос» есть возможность управлять параметрами формирования и вывода изображения, приведенного на рисунке 19. Для этого используется диалоговое окно, приведенное на рисунке 20.

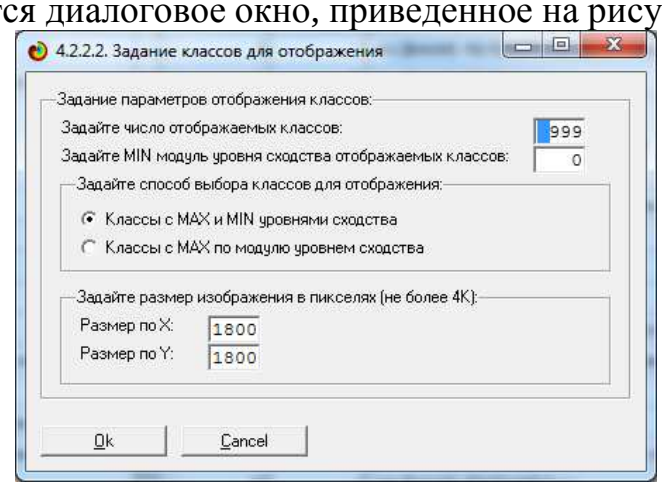


Рисунок 20. Диалоговое окно управления параметрами формирования и вывода изображения когнитивной диаграммы классов

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 18, но и в форме агломеративных дендрограмм, полученных в результате ***когнитивной кластеризации*** [8] (рисунок 21):

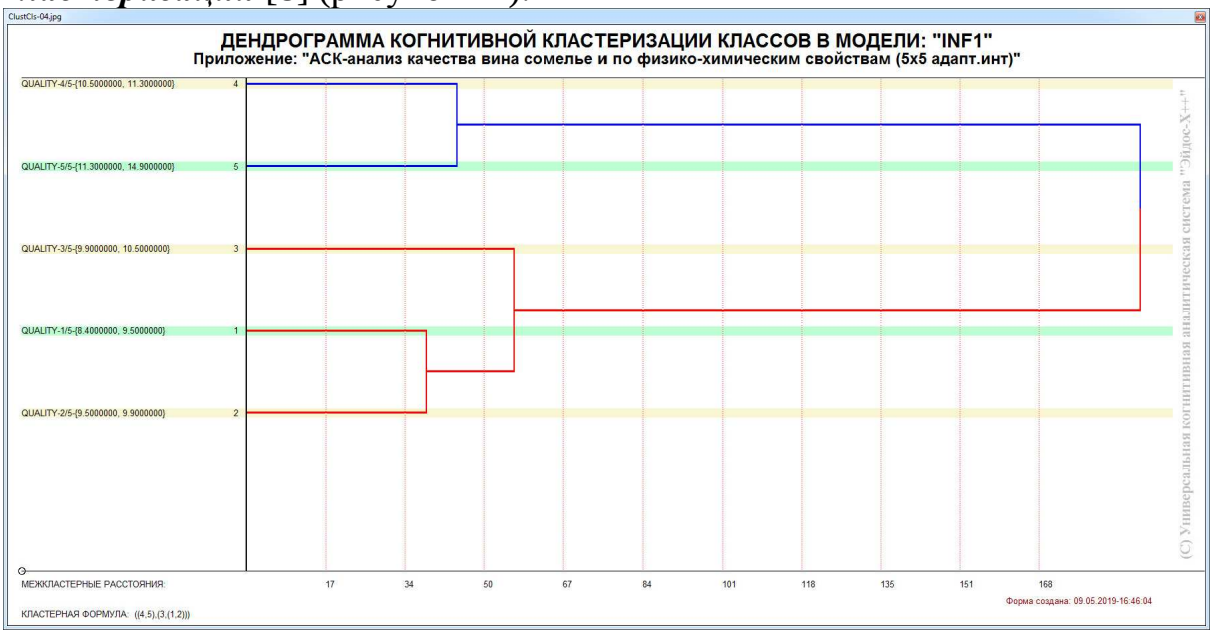


Рисунок 21. Дендрограмма когнитивной агломеративной кластеризации классов по качеству вина

Из рисунка 21 мы видим, что некоторые классы качества вина сходны по системе значений физико-химических свойств, и, следовательно, могут наблюдаться одновременно, а другие по этой системе свойств сильно отличаются, и, следовательно, являются взаимоисключающими, т.е. альтернативными. Из дендрограммы когнитивной агломеративной кластеризации классов, приведенной на рисунке 21, мы видим также, что все классы качества вина образуют два противоположных по системе значений физико-химических свойств кластера, являющихся полюсами конструкта: в верхнем кластере собраны классы с высоким и очень высоким качеством вина, а в нижнем – с очень низким, низким и средним качеством вина.

На рисунке 22 мы видим график изменения межкластерных расстояний:

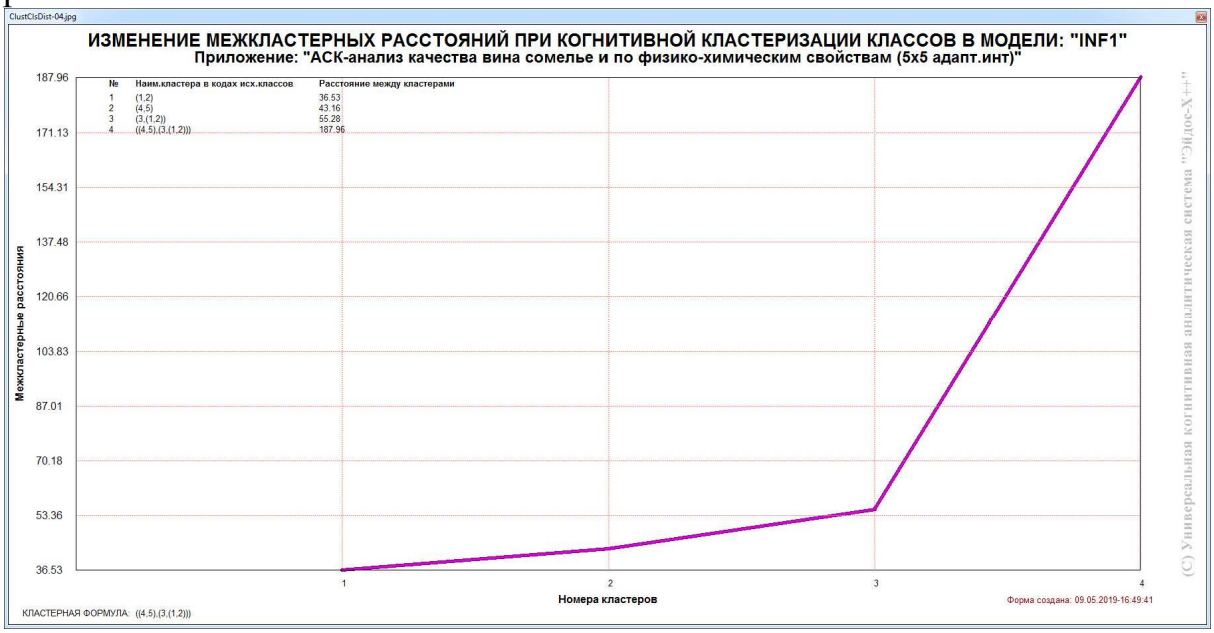


Рисунок 22. График изменения межкластерных расстояний

4.3.3. Когнитивные диаграммы значений факторов

Эти диаграммы отражают сходство/различие значений физико-химических свойств по их смыслу, т.е. по содержащейся в них информации о качестве вина с этими свойствами. Эти диаграммы мы получаем в режимах 4.3.2.1 и 4.3.2.2 (рисунок 23).

Из рисунка 23 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 23, 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.

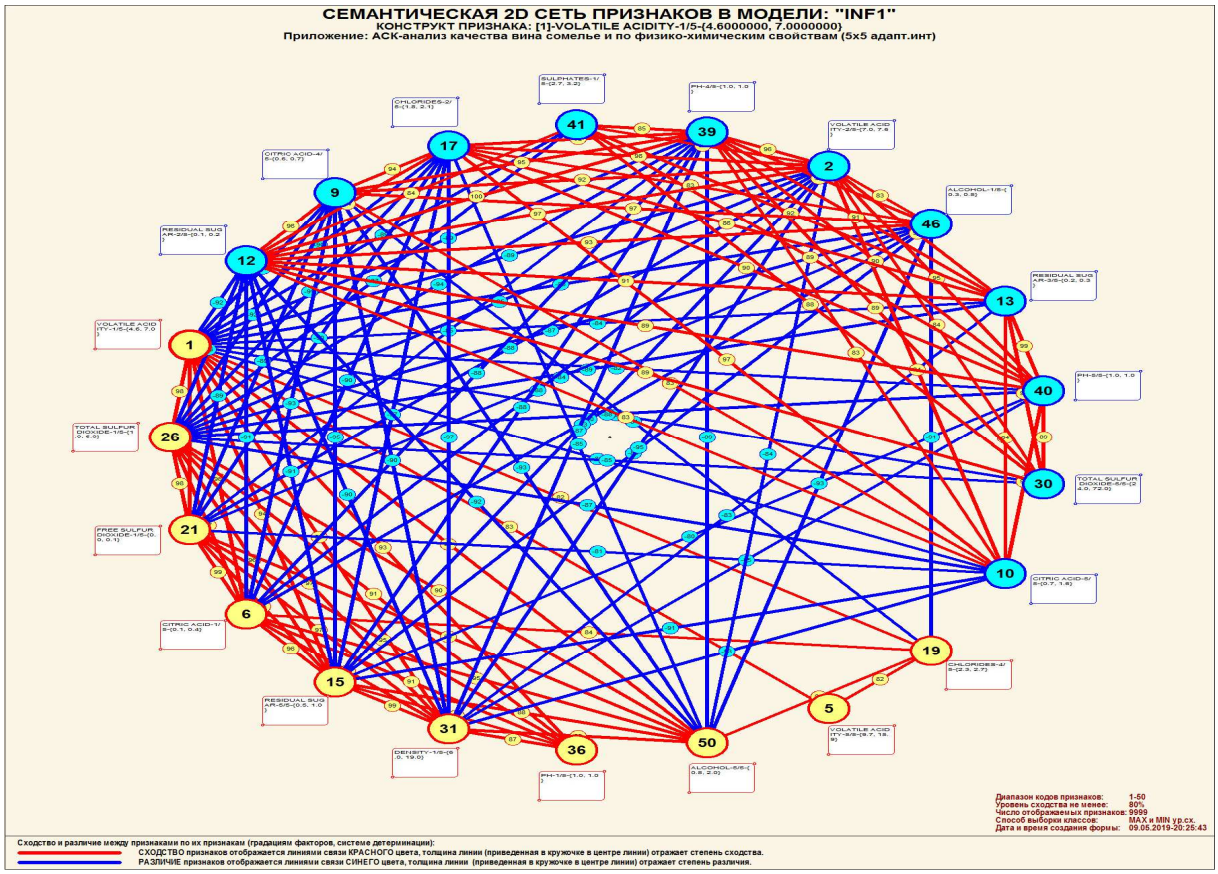


Рисунок 23. Когнитивная диаграмма и конструкт сходства/различия значений физико-химических свойств вина по их смыслу, т.е. по содержащейся в них информации о качестве вина с этими свойствами

Диаграмма, приведенная на рисунке 23, получена при параметрах, приведенных на рисунке 24.

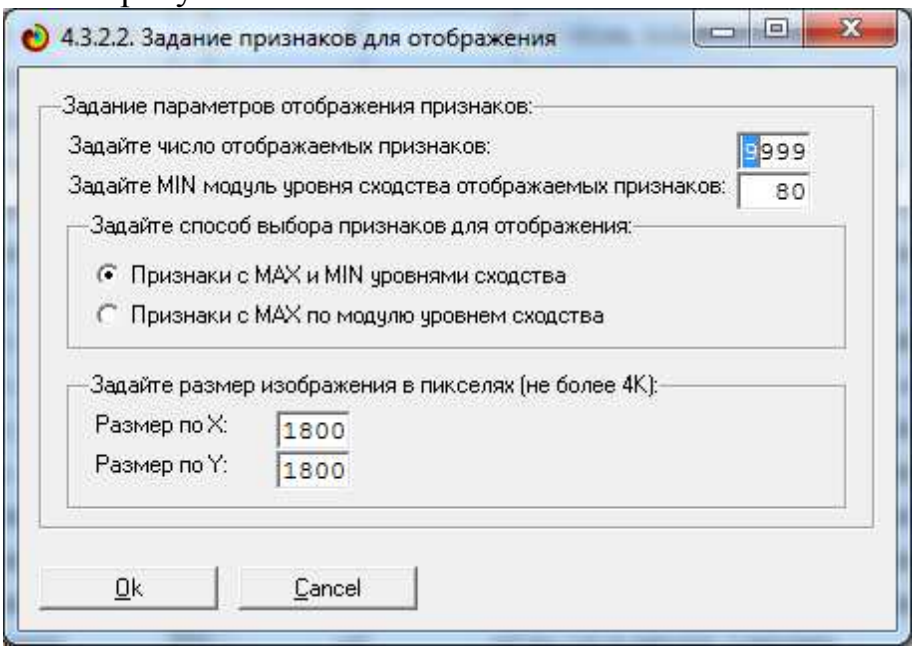


Рисунок 24. Параметры отображения когнитивной диаграммы, приведенной на рисунке 23

4.3.4. Агломеративная когнитивная кластеризация значений факторов

На рисунке 25 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 23.

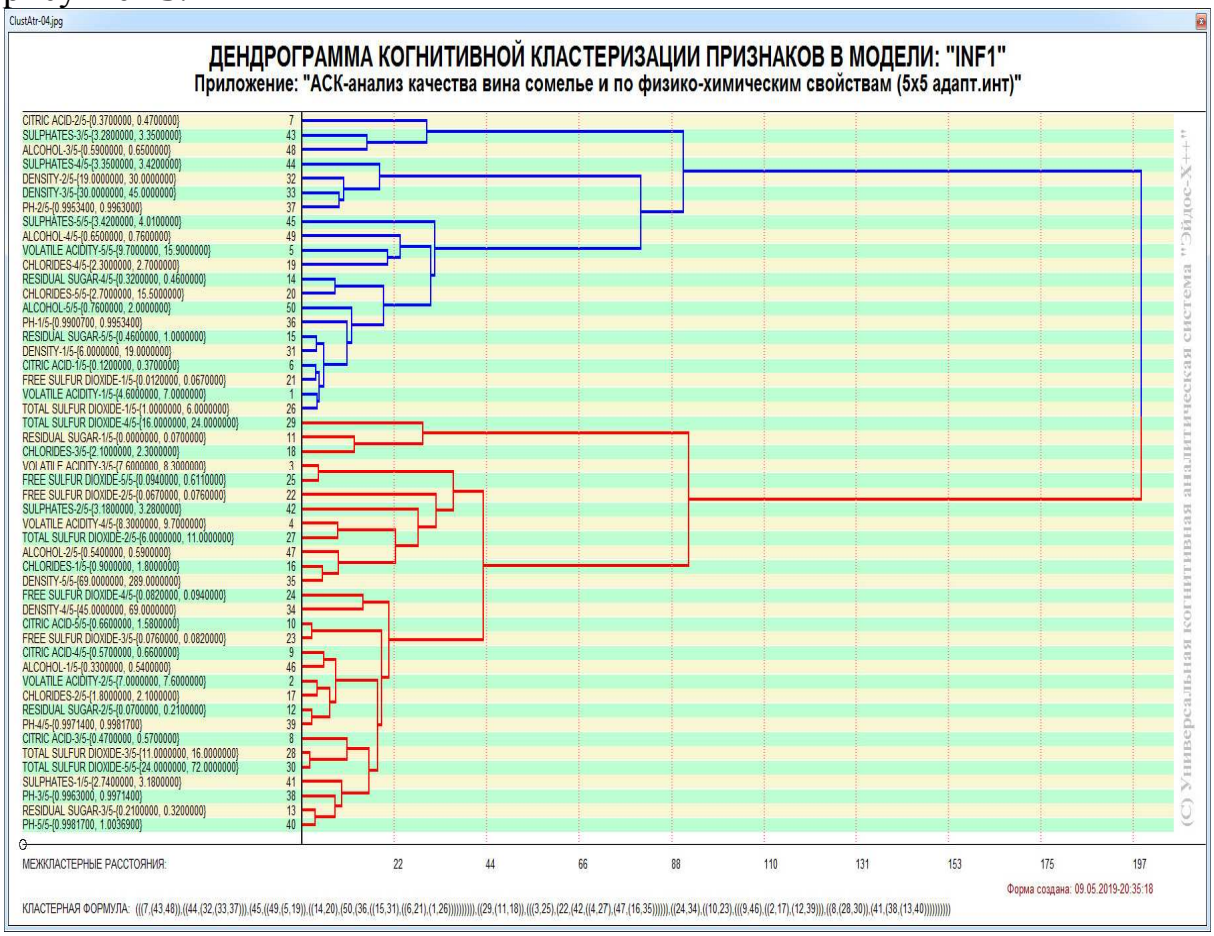


Рисунок 25. Дендрограмма агломеративной когнитивной кластеризации значений физико-химических свойств вина по их смыслу, т.е. по содержащейся в них информации о качестве вина с этими свойствами

Из дендрограммы на рисунке 25 мы видим, что все значения факторов образуют 2 четко выраженных кластера, объединенных в полюса конструкта (показаны синими и красным цветами).

Хорошо видна группировка значений физико-химических свойств вина по наблюдающимся при них качествам вина. ***Значения факторов на полюсах конструкта факторов (рисунок 25) обуславливают переход объекта моделирования в состояния, соответствующие классам, представленным на полюсах конструкта классов (рисунки 19 и 21).***

На рисунке 26 приведен график межкластерных расстояний значений физико-химических свойств вина.

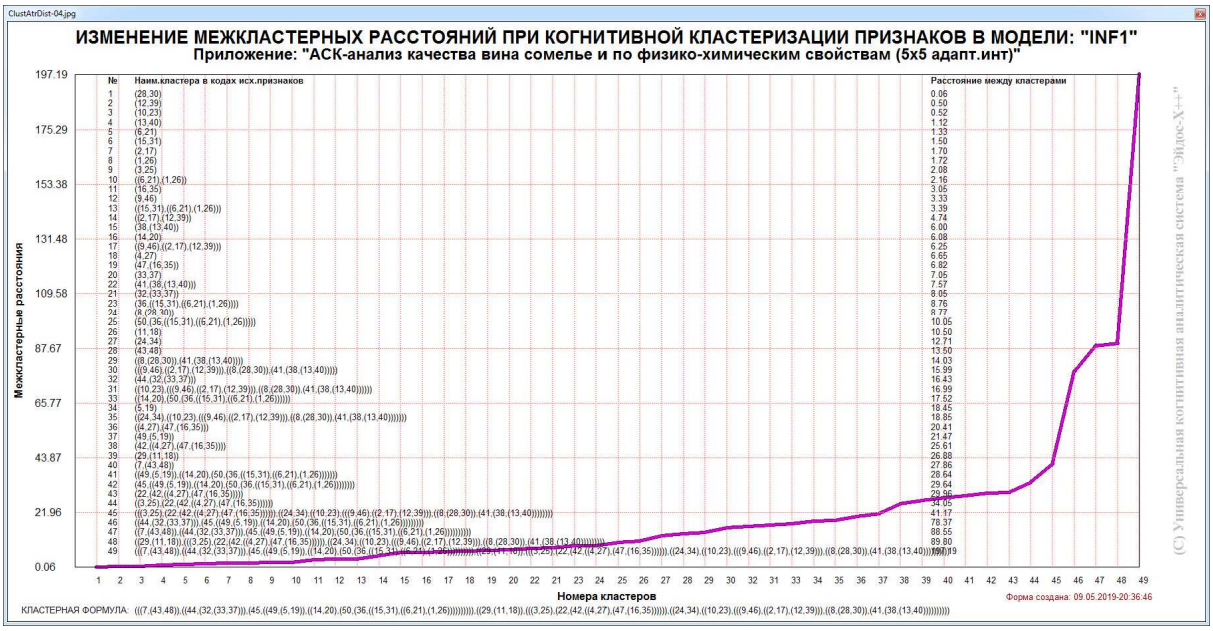


Рисунок 26. График изменения межкластерных расстояний при когнитивной кластеризации значений факторов

4.3.5. Нелокальные нейроны и нелокальные нейронные сети

На рисунке 27 приведены пример нелокального нейрона, а на рисунке 28 и фрагмент одного слоя нелокальной нейронной сети:

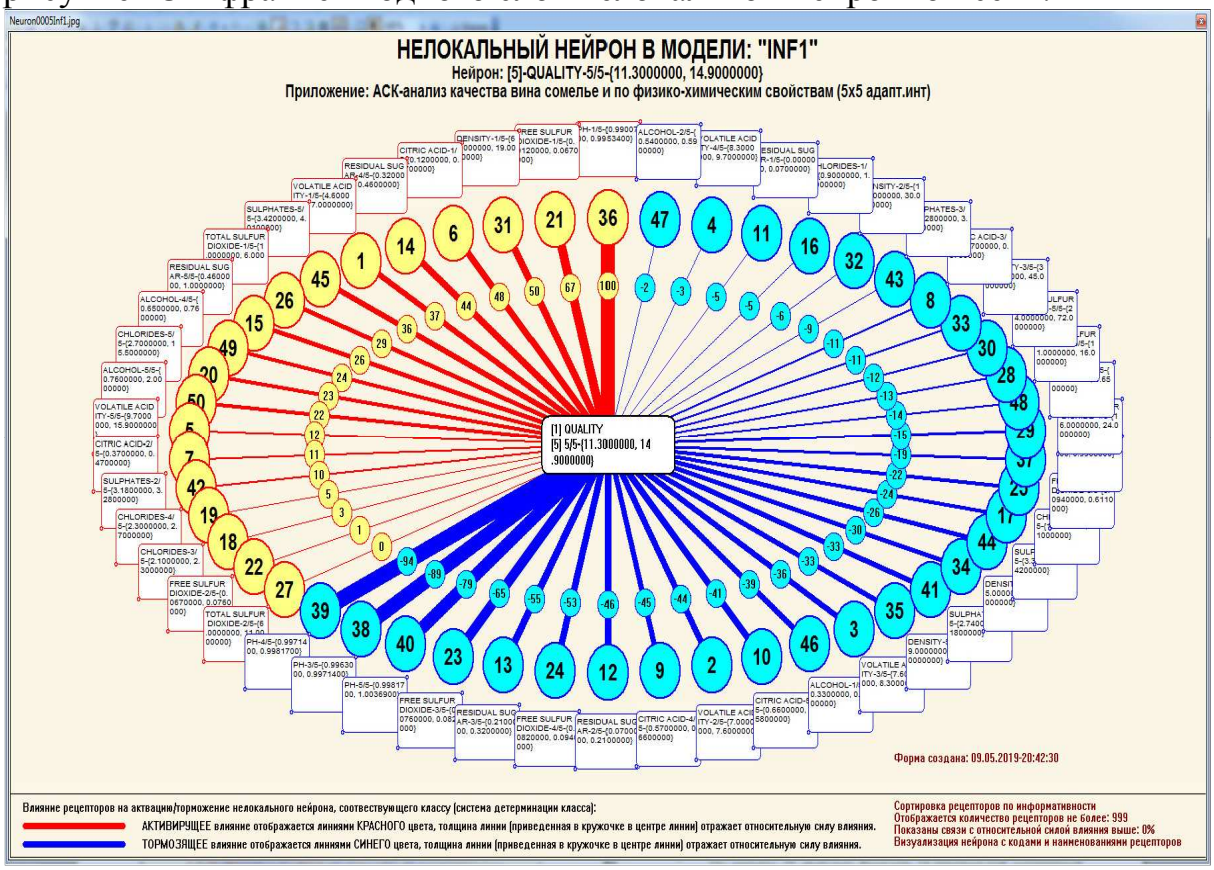


Рисунок 27. Пример нелокального нейрона

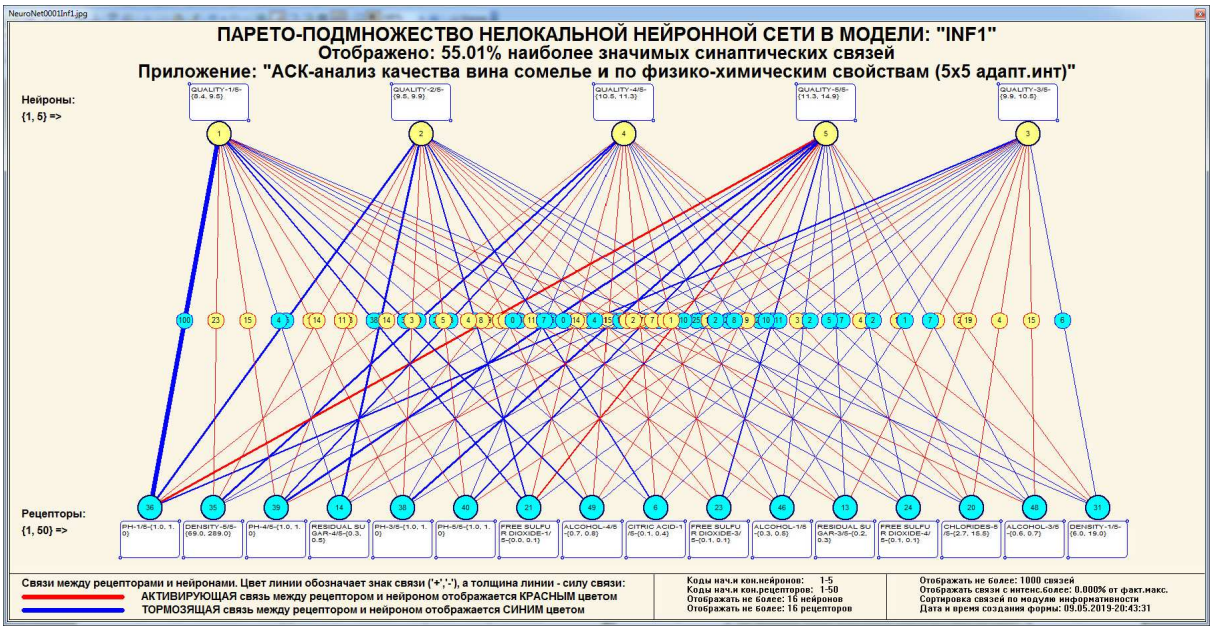


Рисунок 28. Один слой нелокальной нейронной сети (фрагмент)

В приведенном слое нейронной сети нейроны соответствуют классам качества вина, а рецепторы различным обуславливающим их значениям физико-химических свойств. Нейроны расположены слева на право в порядке убывания силы детерминации, т.е. слева находятся наиболее жестко обусловленные явления, а с права – менее жестко обусловленные.

Модель знаний системы «Эйдос» относится к ***нечетким декларативным*** гибридным моделям и объединяет в себе некоторые особенности нейросетевой [9] 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 [9], как сейчас говорят «полносвязной».

4.3.6. 3d-интегральные когнитивные карты

На рисунке 29 приведен фрагмент 3d-интегральной когнитивной карты, отражающая СК-модель Inf3. 3d-интегральная когнитивная карта является отображением на одном рисунке когнитивных диаграмм классов и значений факторов, отображенных соответственно на рисунках 19 и 23, и одного слоя нейронной сети, приведенного на рисунке 28.

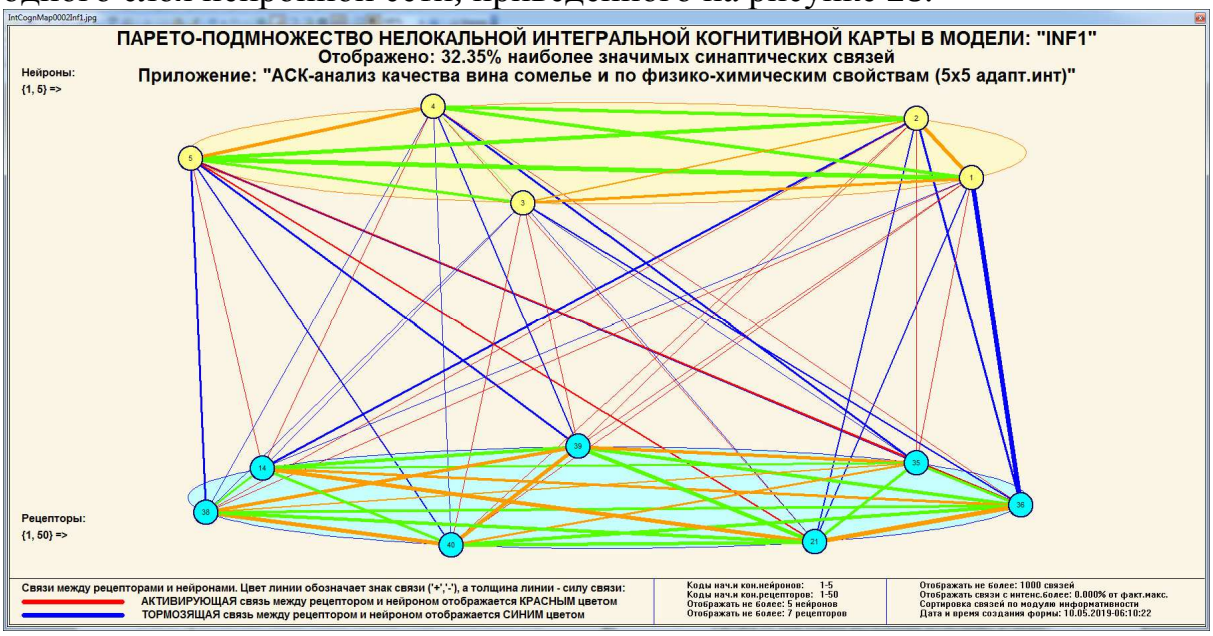


Рисунок 29. 3d-интегральная когнитивная карта в СК-модели Inf3

4.3.7. Когнитивные функции

Вместо описания того, что представляют собой когнитивные функции, приведем help соответствующего режима системы «Эйдос» (рисунок 30) и сошлемся на работу, в которой это описано [10].

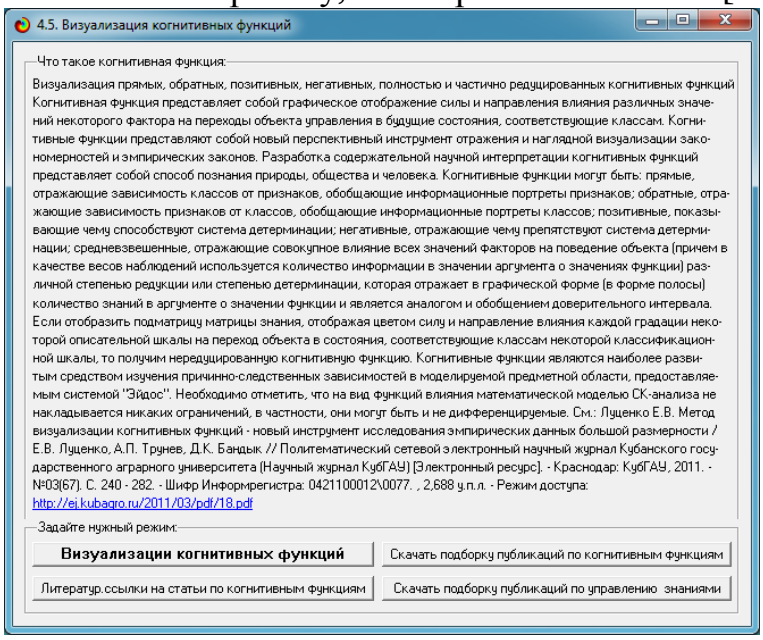
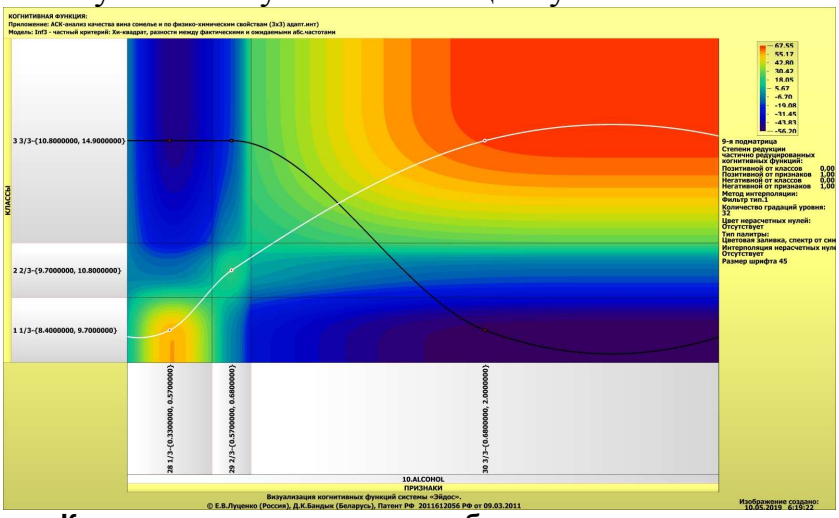
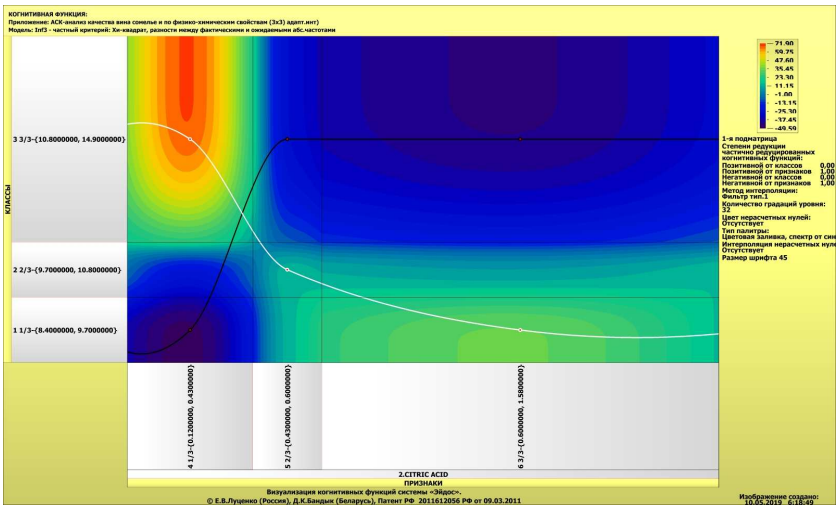


Рисунок 30. Help режима визуализации когнитивных функций

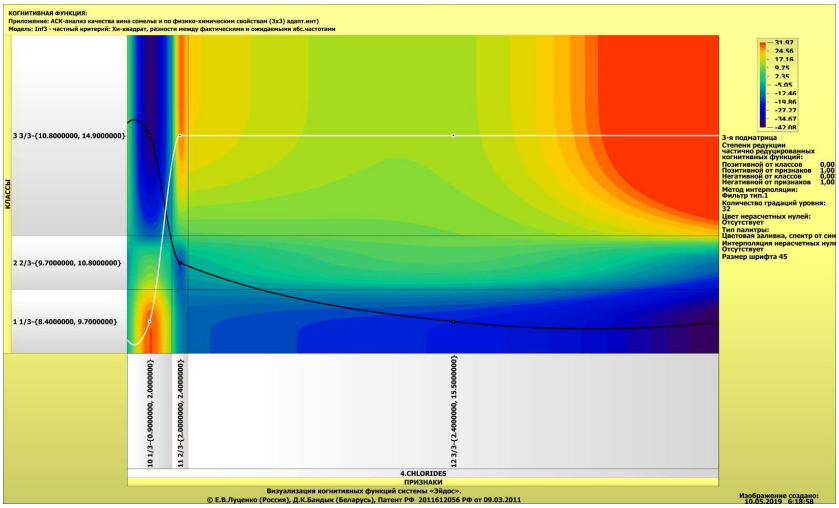
На рисунках 31 приведены некоторые когнитивные функции, наглядно отражающие силу и направление влияния значений (т.е. степени выраженности) различных объективно-установленных физико-химических свойств вина на субъективную сомелье-оценку качества этого вина.



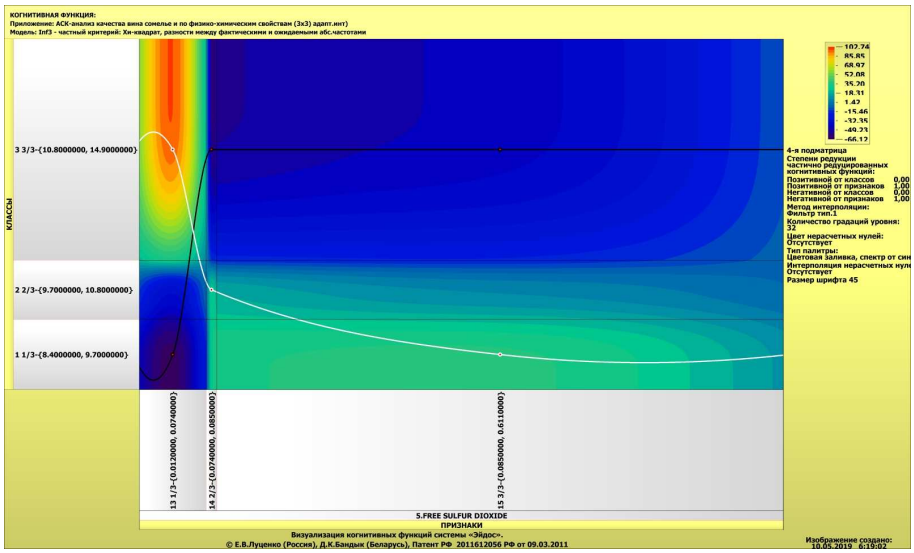
Качество вина тем выше, чем больше в нем алкоголя



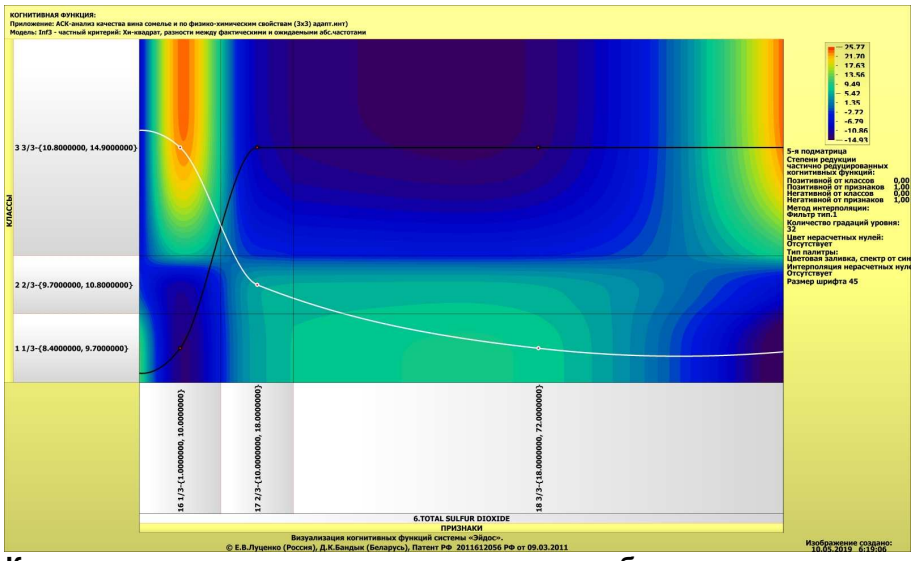
Качество вина тем выше, чем меньше в нем лимонной кислоты



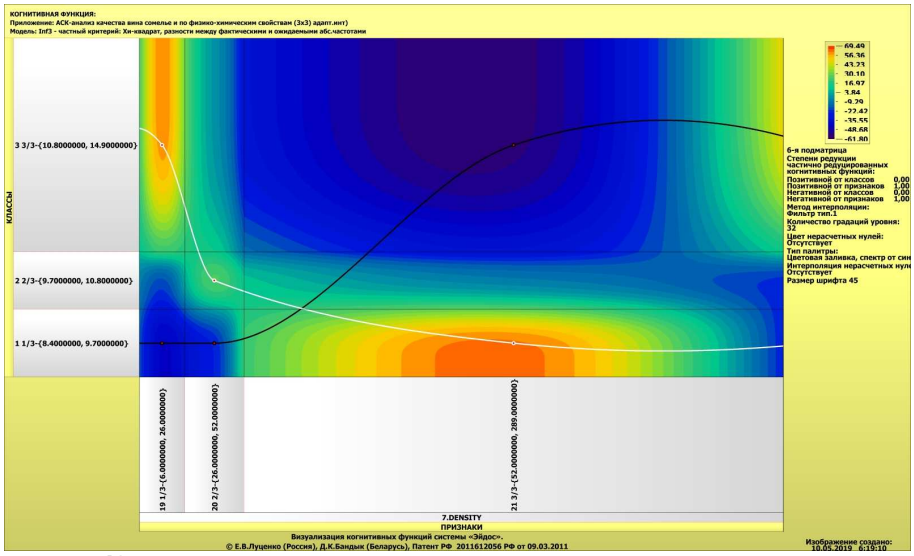
Качество вина тем выше, чем больше в нем хлорида



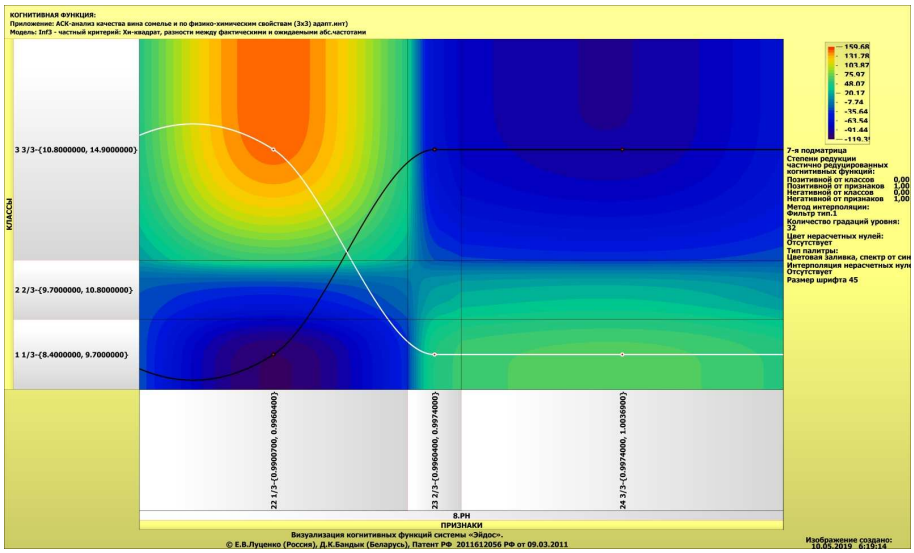
Качество вина тем выше, чем меньше в нем свободного диоксида серы



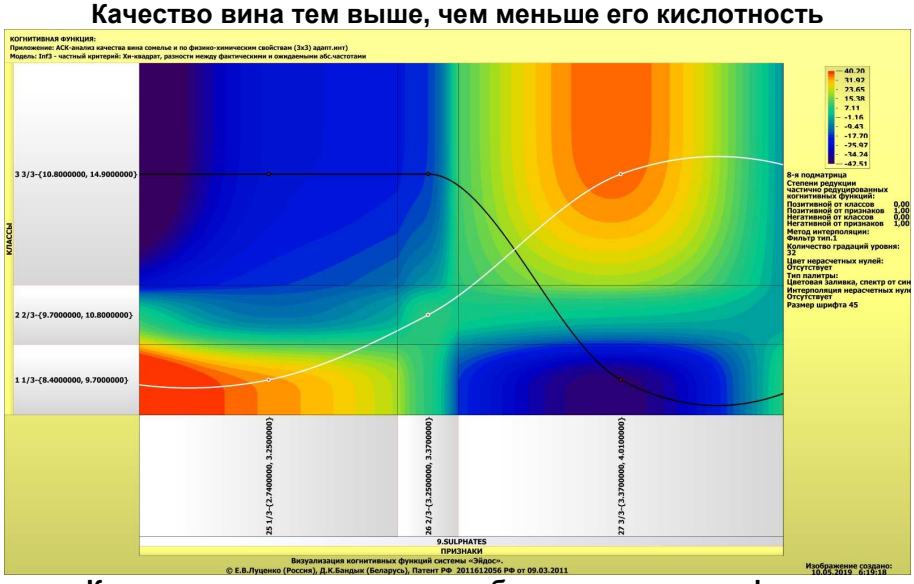
Качество вина тем выше, чем меньше в нем общего диоксида серы



Качество вина тем выше, чем меньше его плотность



Качество вина тем выше, чем меньше его кислотность



Качество вина тем выше, чем больше в нем сульфатов

Рисунок 31. Примеры когнитивных функций

Знание того, какие объективные физико-химические свойства вина обуславливают высокую оценку его качества сомелье, можно приготовить некую смесь различных веществ, которая будет оцениваться не только простыми потребителями вина, но даже и экспертами, как вино достаточно высокого качества. Пример этого мы видим в нашумевшем видео: https://www.youtube.com/watch?v=hGje6LFGQY0.

4.3.8. Сила влияния значений физико-химических свойств вина и самих этих свойств на сомелье-оценки качества вина

На рисунках 8, 9, 10 приведены фрагменты некоторых статистических и системно-когнитивных моделей, отражающих моделируемую предметную область.

Строки матриц моделей соответствуют значениям факторов, т.е. степени выраженности различных физико-химических свойств вина

(градации описательных шкал), устанавливаемых объективными методами в результате лабораторных исследований.

Колонки матриц моделей соответствуют различным классам по субъективными сомелье-оценкам качества вина (градации классификационных шкал).

Числовые значения в ячейках матриц моделей, находящихся на пересечении строк и колонок, отражают направление (знак) и силу влияния значения физико-химического свойства, соответствующего сроке, на принадлежность вина к классу качества, соответствующего колонке.

Если какое-то определенное значение физико-химического свойства слабо влияет на сомелье-оценку качества вина, то в соответствующей строке матрицы модели будут малые по модулю значения, если же влияние сильное – то и значения будут большие по модулю.

Если значение физико-химического свойства повышает сомелье-оценку качества вина, то в соответствующей строке матрицы модели будут положительные значения, если же понижает – то и значения будут отрицательные.

Из этого понятно, что суммарную силу влияния того или иного значения физико-химического свойства вина на сомелье-оценку его качества (т.е. ценность данного значения физико-химического свойства для решении задачи идентификации и других задач) можно количественно оценивать ***степенью вариабельности значений*** в строке матрицы модели, соответствующей этому значению свойства.

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 8, 9, 10 содержит количественную оценку вариабельности значений строки модели (среднеквадратичное отклонение), которая и представляет собой ценность значения физико-химического свойства, соответствующего строке, для решения задач идентификации качества вина и решения 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 32, таблица 5).

Ценность же физико-химического свойства (всей описательной шкалы или фактора), для решения этих задач можно количественно оценивать как среднее от ценности значений этого свойства (таблица 6).

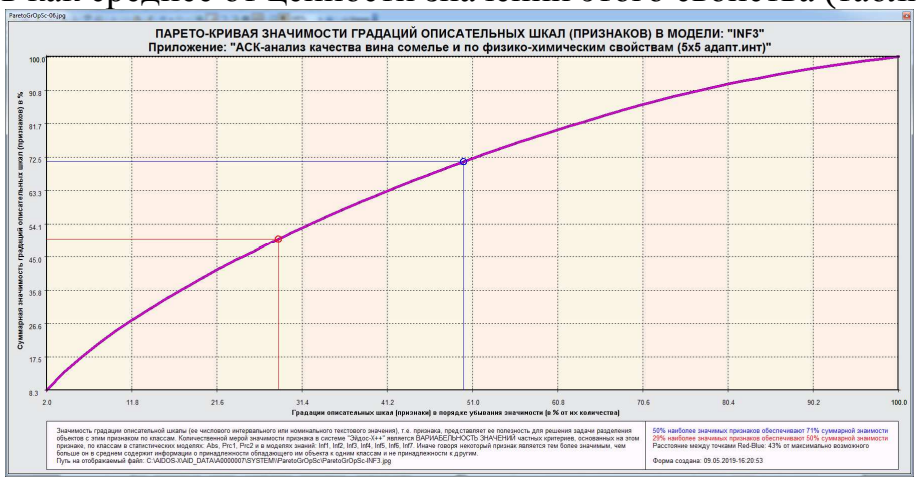


Рисунок 32. Парето-кривая значимости градаций описательных шкал

Таблица 5 – Парето-таблица значимости градаций описательных шкал

№	Код значения физ.-хим. свойства вина	Наименование значения физико-химического свойства вина	Код физ.-хим. свойства вина	Ценность значения физ.-хим. свойства вина	Ценность значения физ.-хим. свойства вина нарастающим итогом	Ценность значения физ.-хим. свойства вина в %	Ценность значения физ.-хим. свойства вина в % нарастающим итогом
1	36	PH-1/5-{0.9900700, 0.9953400}	8	0,32	0,32	8,43	8,43
2	21	FREE SULFUR DIOXIDE-1/5-{0.0120000, 0.0670000}	5	0,17	0,49	4,53	12,96
3	39	PH-4/5-{0.9971400, 0.9981700}	8	0,16	0,65	4,27	17,23
4	35	DENSITY-5/5-{69.0000000, 289.0000000}	7	0,15	0,81	4,06	21,29
5	14	RESIDUAL SUGAR-4/5-{0.3200000, 0.4600000}	3	0,14	0,94	3,63	24,92
6	6	CITRIC ACID-1/5-{0.1200000, 0.3700000}	2	0,13	1,08	3,55	28,47
7	38	PH-3/5-{0.9963000, 0.9971400}	8	0,12	1,20	3,23	31,70
8	40	PH-5/5-{0.9981700, 1.0036900}	8	0,11	1,31	2,96	34,66
9	46	ALCOHOL-1/5-{0.3300000, 0.5400000}	10	0,10	1,42	2,77	37,43
10	49	ALCOHOL-4/5-{0.6500000, 0.7600000}	10	0,10	1,52	2,71	40,14
11	31	DENSITY-1/5-{6.0000000, 19.0000000}	7	0,10	1,62	2,63	42,77
12	23	FREE SULFUR DIOXIDE-3/5-{0.0760000, 0.0820000}	5	0,10	1,72	2,61	45,38
13	9	CITRIC ACID-4/5-{0.5700000, 0.6600000}	2	0,10	1,81	2,54	47,93
14	24	FREE SULFUR DIOXIDE-4/5-{0.0820000, 0.0940000}	5	0,09	1,91	2,45	50,37
15	12	RESIDUAL SUGAR-2/5-{0.0700000, 0.2100000}	3	0,09	1,99	2,25	52,62
16	1	VOLATILE ACIDITY-1/5-{4.6000000, 7.0000000}	1	0,08	2,07	2,20	54,82
17	13	RESIDUAL SUGAR-3/5-{0.2100000, 0.3200000}	3	0,08	2,16	2,18	57,00
18	48	ALCOHOL-3/5-{0.5900000, 0.6500000}	10	0,08	2,24	2,13	59,13
19	3	VOLATILE ACIDITY-3/5-{7.6000000, 8.3000000}	1	0,08	2,32	2,13	61,25
20	20	CHLORIDES-5/5-{2.7000000, 15.5000000}	4	0,08	2,39	2,00	63,26
21	50	ALCOHOL-5/5-{0.7600000, 2.0000000}	10	0,07	2,47	1,95	65,20
22	45	SULPHATES-5/5-{3.4200000, 4.0100000}	9	0,07	2,54	1,92	67,12
23	2	VOLATILE ACIDITY-2/5-{7.0000000, 7.6000000}	1	0,07	2,61	1,88	69,01
24	44	SULPHATES-4/5-{3.3500000, 3.4200000}	9	0,07	2,68	1,86	70,86
25	10	CITRIC ACID-5/5-{0.6600000, 1.5800000}	2	0,06	2,74	1,67	72,54
26	11	RESIDUAL SUGAR-1/5-{0.0000000, 0.0700000}	3	0,06	2,81	1,65	74,19
27	26	TOTAL SULFUR DIOXIDE-1/5-{1.0000000, 6.0000000}	6	0,06	2,87	1,56	75,75
28	25	FREE SULFUR DIOXIDE-5/5-{0.0940000, 0.6110000}	5	0,06	2,92	1,52	77,27
29	41	SULPHATES-1/5-{2.7400000, 3.1800000}	9	0,06	2,98	1,46	78,73
30	15	RESIDUAL SUGAR-5/5-{0.4600000, 1.0000000}	3	0,05	3,03	1,45	80,18
31	37	PH-2/5-{0.9953400, 0.9963000}	8	0,05	3,09	1,43	81,61
32	42	SULPHATES-2/5-{3.1800000, 3.2800000}	9	0,05	3,14	1,42	83,03
33	32	DENSITY-2/5-{19.0000000, 30.0000000}	7	0,05	3,19	1,34	84,37
34	5	VOLATILE ACIDITY-5/5-{9.7000000, 15.9000000}	1	0,05	3,24	1,32	85,70

35	34	DENSITY-4/5-{45.0000000, 69.0000000}	7	0,05	3,29	1,22	86,92
36	47	ALCOHOL-2/5-{0.5400000, 0.5900000}	10	0,04	3,33	1,14	88,06
37	17	CHLORIDES-2/5-{1.8000000, 2.1000000}	4	0,04	3,37	1,11	89,17
38	19	CHLORIDES-4/5-{2.3000000, 2.7000000}	4	0,04	3,41	1,07	90,24
39	29	TOTAL SULFUR DIOXIDE-4/5-{16.0000000, 24.0000000}	6	0,04	3,45	1,06	91,30
40	18	CHLORIDES-3/5-{2.1000000, 2.3000000}	4	0,04	3,49	1,04	92,34
41	7	CITRIC ACID-2/5-{0.3700000, 0.4700000}	2	0,04	3,53	1,01	93,35
42	33	DENSITY-3/5-{30.0000000, 45.0000000}	7	0,04	3,57	0,98	94,33
43	16	CHLORIDES-1/5-{0.9000000, 1.8000000}	4	0,04	3,61	0,96	95,29
44	43	SULPHATES-3/5-{3.2800000, 3.3500000}	9	0,03	3,64	0,91	96,20
45	8	CITRIC ACID-3/5-{0.4700000, 0.5700000}	2	0,03	3,67	0,74	96,94
46	27	TOTAL SULFUR DIOXIDE-2/5-{6.0000000, 11.0000000}	6	0,03	3,70	0,71	97,65
47	28	TOTAL SULFUR DIOXIDE-3/5-{11.0000000, 16.0000000}	6	0,02	3,72	0,63	98,28
48	30	TOTAL SULFUR DIOXIDE-5/5-{24.0000000, 72.0000000}	6	0,02	3,74	0,62	98,91
49	4	VOLATILE ACIDITY-4/5-{8.3000000, 9.7000000}	1	0,02	3,76	0,59	99,49
50	22	FREE SULFUR DIOXIDE-2/5-{0.0670000, 0.0760000}	5	0,02	3,78	0,51	100,00

Таблица 6 – Парето-таблица значимости описательных шкал

№	Код физ.-хим. свойства вина	Наименование физ.-хим. свойства вина	Число градаций физ.-хим. свойства вина	Код минимальной градации физ.-хим. свойства вина	Код максимальной градации физ.-хим. свойства вина	Значимость физ.-хим. свойства вина	Значимость физ.-хим. свойства вина нарастающим итогом	Значимость физ.-хим. свойства вина в %	Значимость физ.-хим. свойства вина нарастающим итогом в %
1	8	PH	5	36	40	**0,15**	0,15	**20,32**	20,32
2	5	FREE SULFUR DIOXIDE	5	21	25	0,09	0,24	11,61	31,93
3	3	RESIDUAL SUGAR	5	11	15	0,08	0,33	11,17	43,10
4	10	ALCOHOL	5	46	50	0,08	0,41	10,70	53,79
5	7	DENSITY	5	31	35	0,08	0,48	10,24	64,03
6	2	CITRIC ACID	5	6	10	0,07	0,56	9,52	73,55
7	1	VOLATILE ACIDITY	5	1	5	0,06	0,62	8,12	81,67
8	9	SULPHATES	5	41	45	0,06	0,68	7,56	89,23
9	4	CHLORIDES	5	16	20	0,05	0,72	6,19	95,42
10	6	TOTAL SULFUR DIOXIDE	5	26	30	**0,03**	0,76	**4,58**	100,00

Из таблицы 6 видно, что наибольшую роль в оценке качества вина сомелье играет такое физико-химическое свойство, как кислотность (pH) вина, и далее в порядке убывания значимости идут:

- Ph — Водородный показатель
- Free sulfur dioxide — Свободный диоксид серы
- Residual sugar — Остаточный сахар
- Alcohol — Алкоголь
- Density — Плотность
- Citric acid — Лимонная кислота
- Volatile acidity — Летучая кислотность
- Sulphates — Сульфаты
- Chlorides — Хлорид
- Total sulfur dioxide — Общий диоксид серы

причем значимость последнего физико-химического свойства в этом рейтинге в 5 раз ниже значимости первого, т.е. разница в значимости весьма существенная.

4.3.8. Степень детерминированности (обусловленности) сомелье-оценок качества вина его объективными физико-химическими свойствами

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается ***степенью вариабельности значений*** описательных шкал в колонке матрицы модели, соответствующей данному классу (таблица 7). На рисунке 33 мы видим Парето-кривую степени детерминированности классов нарастающим итогом.

Рисунок 33. Парето-кривая степени детерминированности классов

Таблица 7 – Парето-таблица степеней детерминированности (обусловленности) классов, т.е. сомелье-оценок качества вина в СК-модели INF1

№	Код	Наименование класса	Степень детерминированности (бит)	Сумма степени детерминированности (бит)	Степень детерминированности (%)	Сумма степени детерминированности (%)
1	5	QUALITY-5/5-{11.3000000, 14.9000000}	**0,103**	0,103	**24,881**	24,881
2	1	QUALITY-1/5-{8.4000000, 9.5000000}	0,102	0,205	24,759	49,640
3	2	QUALITY-2/5-{9.5000000, 9.9000000}	0,088	0,293	21,240	70,880
4	4	QUALITY-4/5-{10.5000000, 11.3000000}	0,069	0,362	16,564	87,443
5	3	QUALITY-3/5-{9.9000000, 10.5000000}	**0,052**	0,414	**12,557**	100,000

Из таблицы 7 мы видим, что объективные физико-химические свойства вина наиболее жестко детерминируют (обуславливают) сомелье-оценку качества вина как максимальную и минимальную, а наиболее слабо – среднюю. При этом степень детерминированности наиболее и наименее детерминированных классов отличается примерно в два раза, что весьма существенно. Оценки выше и ниже среднего по степени

детерминированности занимают промежуточное положение между максимальными и минимальными оценками качестве вина с одной стороны, и средней его оценкой с другой. Это значит, что для сомелье проще всего идентифицировать вина максимального и минимального качества, а сложнее всего среднего качества. Вина качества выше и ниже среднего идентифицируются сомелье лучше, чем среднего, но хуже, чем максимального и минимального. Все это весьма логично и хорошо вписывается в предыдущие результаты анализа.

Однако, из данного исследования можно сделать обоснованный вывод о том, что шкала сомелье-оценок качества вина со 100 градациями – это большой аванс и комплемент сомелье, а фактически вполне достаточно 5, или даже 3 градаций, т.к. именно при таком количестве градаций субъективные сомелье-оценки хорошо совпадают с оценками качества вина на основе его объективных физико-химических свойств.

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

Поскольку в СК-модели, рассматриваемой в данной работе, только одна классификационная шкала, то мы не можем ранжировать классификационные шкалы в порядке убывания средней степени детерминированности их градаций.

5. 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, 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 классам качества вина, изучено влияние значений различных объективных физико-химических свойств вина на субъективную сомелье-оценку качества вина, и, на основе этого, решены задачи идентификации, принятия решений и исследования моделируемой предметной области путем исследования ее модели.

Со всеми моделями, созданными в данной статье, можно ознакомиться установив облачное Эйдос-приложение №148 в режиме 1.3 системы «Эйдос».

Автор благодарен доктору биологических наук профессору Андрею Георгиевичу Кощаеву https://kubsau.ru/university/rectorate/ за предоставленную возможность опубликования данной статьи.

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. Исходные данные: https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009

3. Исходные данные: https://archive.ics.uci.edu/ml/datasets/wine+quality

4. P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009. https://doi.org/10.1016/j.dss.2009.05.016

5. 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.

6. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №02(126). С. 1 – 32. – IDA [article ID]: 1261702001. – Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf, 2 у.п.л.

7. 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.

8. 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.

9. 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.

10. 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

Spisok literatury`

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

2. Isxodny`e danny`e: https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009

3. Isxodny`e danny`e: https://archive.ics.uci.edu/ml/datasets/wine+quality

4. P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009. https://doi.org/10.1016/j.dss.2009.05.016

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

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

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

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

9. Lucenko E.V. Sistemnaya teoriya informacii i nelokal`ny`e interpretiruemy`e nejronny`e seti pryamogo scheta / E.V. Lucenko // Politematicheskij setevoj e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchny`j zhurnal KubGAU) [E`lektronny`j resurs]. – Krasnodar: KubGAU, 2003. – №01(001). S. 79 – 91. – IDA [article ID]: 0010301011. – Rezhim dostupa: http://ej.kubagro.ru/2003/01/pdf/11.pdf, 0,812 u.p.l.

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