

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ И.Т.ТРУБИЛИНА»

Факультет _____

РЕЦЕНЗИЯ

на курсовой проект (работу)

Студента (ки): Якимова Александра Юрьевича

курса 3 очной (заочной) формы обучения

направления подготовки 09.03.02

направленность (профиль) Информационные системы и технологии.

Наименование темы «Разработка системно-когнитивной модели классификации грибов на основе данных репозитория UCI».

Рецензент:

(Ф.И.О., ученое звание и степень, должность)

Оценка качества выполнения курсового проекта (работы)

№ п/п	Показатель	Оценка соответствия заданию (по 5-и балльной шкале)
1.	Актуальность тематики работы	5
2.	Степень полноты обзора состояния проблемы и корректность постановки цели и задач исследования	4
3.	Уровень и корректность использования в работе различных методов исследований	5
4.	Степень комплексности работы, применения в ней знаний общепрофессиональных и специальных дисциплин	5
5.	Ясность, четкость, последовательность и обоснованность изложения	5
6.	Применение современных технологий обработки информации	5
7.	Качество оформления работы (общий уровень грамотности, стиль изложения, качество иллюстраций, соответствие требованиям по оформлению)	4
8.	Ответы на вопросы при защите	5

Достоинства работы

Применение современных математических моделей и программного
инструмента

Недостатки работы

Недостаточно качественное оформление

Итоговая оценка при защите _____

Рецензент _____ (____ Е.В. Луценко ____)

«____» _____ 201__ г.

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ
ФЕДЕРАЦИИ

ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ
УНИВЕРСИТЕТ
ИМЕНИ И.Т. ТРУБИЛИНА»

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

к курсовой работе

по дисциплине: ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ
СИСТЕМЫ

на тему:

**Разработка системно-когнитивной модели классификации грибов на
основе данных репозитория UCI**

выполнил студент группы ИТ1401 **Якимов Александр Юрьевич**

Допущена к защите _____

Руководитель проекта _____ Луценко Евгений Вениаминович,
д.э.н., к.т.н., профессор
(подпись, расшифровка подписи)

Нормоконтролер _____ Николаева Ирина Валентиновна, к.т.н.,
доцент
(подпись, расшифровка подписи)

Защищена _____ Оценка _____

(дата)

Члены комиссии _____ В.И. Лойко
_____ Е.В. Луценко
_____ И. В. Николаева
(подпись, дата, расшифровка подписи)

Краснодар
2017 г.

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМЕНИ И.Т. ТРУБИЛИНА»

Факультет прикладной информатики
Кафедра компьютерных технологий и систем

УТВЕРЖДАЮ:

Зав. кафедрой КТС _____ В. И. Лойко

ЗАДАНИЕ
на курсовую работу

Студентке: ИТ1401 группы 3 курса

Факультета прикладной информатики

Специальности: 09.03.02 Информационные системы и технологии
(шифр)

Якимов Александр Юрьевич
(Ф.И.О.)

Тема проекта: **Разработка системно-когнитивной модели классификации грибов на
основе данных репозитория UCI**

Содержание задания: Проанализировать методы формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами, принятия решений и исследования моделируемой предметной области путем исследования ее модели

Объем работы:

а) пояснительная записка к работе _____ листа формата А4

б) графическая часть _____ лист формата А4

Рекомендуемая литература: _ Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для бакалавриата. 7-е изд., перераб. и доп.- Краснодар: КубГАУ – 2016, – 615 с., в электронном виде на сайте автора: <http://lc.kubagro.ru/aidos/p14.htm>

Срок выполнения проекта: с “ ” _____ по “ ” _____ 2017 г.

Срок защиты: “ ” _____ 2017 г.

Дата выдачи задания: “ ” _____ 2017 г.

Дата сдачи проекта на кафедру: “ ” _____ 2017 г.

Руководитель проекта: _____ Луценко Евгений Вениаминович,
д.э.н., к.т.н., профессор

(подпись, Ф.И.О., звание, степень)

Задание принял студент _____ “ ” _____ 2017 г.

(подпись, дата)

Краснодар
2017 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	6
1.СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	7
1.1. Описание решения.....	7
1.2. Преобразование исходных данных из CSV-формата в файл исходных данных MS Excel	8
1.3. Синтез и верификация статистических и интеллектуальных моделей	14
1.4. Виды моделей системы «Эйдос»	16
1.5. Результаты верификации моделей.....	17
2.РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	24
2.1. Решение задачи	24
2.2. Когнитивные функции	27
2.3. SWOT и PEST матрицы и диаграммы.....	29
2.4. Нелокальные Нейронные Сети И Нейроны.....	31
2.5 Кластерный и Конструктивный Анализ.....	33
ЗАКЛЮЧЕНИЕ	34
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	35

ВВЕДЕНИЕ

Создание систем искусственного интеллекта является одним из важных и перспективных направлений развития современных информационных технологий. Так как существует множество альтернатив систем искусственного интеллекта, то возникает необходимость оценки качества математических моделей этих систем. В данной работе рассмотрено решение задачи по поиску зависимости состава продуктов быстрого питания от выбранной категории.

Для достижения поставленной цели необходимы свободный доступ к тестовым исходным данным и методика, которая поможет преобразовать эти данные в форму, которая необходима для работы в системе искусственного интеллекта. Удачным выбором является сборник баз данных Kaggle.

В данной курсовой работе использована база данных «menu» из банка исходных данных по задачам искусственного интеллекта – репозитория Kaggle.

Для решения задачи используем стандартные возможности Microsoft Office Word и Excel, блокнот, а также систему искусственного интеллекта "Эйдос- X++".

1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

1.1. Описание решения

В соответствии с методологией АСК-анализа решение поставленной задачи проведем в четыре этапа:

1. Преобразование исходных данных из csv-формата в промежуточные файлы MS Excel.
2. Преобразование исходных данных из промежуточных файлов MS Excel в базы данных системы "Эйдос".
3. Синтез и верификация моделей предметной области.
4. Применение моделей для решения задач идентификации, прогнозирования и исследования предметной области.

1.2. Преобразование исходных данных из HTML-формата в файл исходных данных MS Excel

Из электронного ресурса баз данных Kaggle возьмем базу данных классификации грибов – «mushrooms.csv», которую оставим без изменений по абсолютной гиперссылке: [база данных](#).

Общее описание задачи:

1. Код
2. Класс
3. Форма шляпки
4. Поверхность шляпки
5. Цвет шляпки
6. Запах
7. Крепление гимениальной пластинки
8. Размер гимениальной пластинки
9. Интервал между гимениальной пластинкой
10. Вмятины
11. Цвет гимениальной пластинки
12. Форма стебля
13. Форма черенка

14. Поверхность стебля сверху
15. Поверхность стебля снизу
16. Цвет поверхности стебля сверху
17. Цвет поверхности стебля снизу
18. Тип вуали
19. Цвет вуали
20. Количество колец
21. Тип колец
22. Печать спор
23. Популяция
24. Обитание

Столбцы 4-24 описательные шкалы.

Столбцы 2-3 являются классификационной шкалой. Этот столбец показывает класс гриба.

Описание столбцов:

classes: edible=e, poisonous=p

cap-shape: bell=b, conical=c, convex=x, flat=f, knobbed=k, sunken=s

cap-surface: fibrous=f, grooves=g, scaly=y, smooth=s

cap-color:

brown=n, buff=b, cinnamon=c, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y

bruises: bruises=t, no=f

odor:

almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s

gill-attachment: attached=a, descending=d, free=f, notched=n

gill-spacing: close=c, crowded=w, distant=d

gill-size: broad=b, narrow=n

gill-color: black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y

stalk-shape: enlarging=e, tapering=t

stalk-root: bulbous=b, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r, missing=?

stalk-surface-above-ring: fibrous=f, scaly=y, silky=k, smooth=s

stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s

stalk-color-above-ring:

brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y

stalk-color-below-ring:

brown=n,buff=b,cinnamon=c,gray=g,orange=o,pink=p,red=e,white=w,yellow=y

veil-type: partial=p,universal=u

veil-color: brown=n,orange=o,white=w,yellow=y

ring-number: none=n,one=o,two=t

ring-type:

cobwebby=c,evanescent=e,flaring=f,large=l,none=n,pendant=p,sheathing=s,zone=z

spore-print-color:

black=k,brown=n,buff=b,chocolate=h,green=r,orange=o,purple=u,white=w,yellow=y

population: abundant=a,clustered=c,numerous=n,scattered=s,several=v,solitary=y

habitat: grasses=g,leaves=l,meadows=m,paths=p,urban=u,waste=w,woods=d

Обучающая выборка:

Таблица 1 – Mushrooms.xls

id	class	cap-shape	cap-surf	cap-color	bruises	odor	gill-attach	gill-spacing	gill-size	gill-color	stalk-shape	stalk-root	stalk-surf	stalk-surf	stalk-color	stalk-color	veil-type	veil-color	ring-num	ring-type	spore-print	population	habitat
1	p	x	s	n	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	k	s	u
2	e	x	s	y	t	a	f	c	b	k	e	c	s	s	w	w	p	w	o	p	n	n	g
3	e	b	s	w	t	l	f	c	b	n	e	c	s	s	w	w	p	w	o	p	n	n	m
4	p	x	y	w	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	k	s	u
5	e	x	s	g	f	n	f	w	b	k	t	e	s	s	w	w	p	w	o	e	n	a	g
6	e	x	y	y	t	a	f	c	b	n	e	c	s	s	w	w	p	w	o	p	k	n	g
7	e	b	s	w	t	a	f	c	b	g	e	c	s	s	w	w	p	w	o	p	k	n	m
8	e	b	y	w	t	l	f	c	b	n	e	c	s	s	w	w	p	w	o	p	n	s	m
9	p	x	y	w	t	p	f	c	n	p	e	e	s	s	w	w	p	w	o	p	k	v	g
10	e	b	s	y	t	a	f	c	b	g	e	c	s	s	w	w	p	w	o	p	k	s	m
11	e	x	y	y	t	l	f	c	b	g	e	c	s	s	w	w	p	w	o	p	n	n	g
12	e	x	y	y	t	a	f	c	b	n	e	c	s	s	w	w	p	w	o	p	k	s	m
13	e	b	s	y	t	a	f	c	b	w	e	c	s	s	w	w	p	w	o	p	n	s	g
14	p	x	y	w	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	v	u
15	e	x	f	n	f	n	f	w	b	n	t	e	s	f	w	w	p	w	o	e	k	a	g
16	e	s	f	g	f	n	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	y	u
17	e	f	f	w	f	n	f	w	b	k	t	e	s	s	w	w	p	w	o	e	n	a	g
18	p	x	s	n	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	k	s	g
19	p	x	y	w	t	p	f	c	n	n	e	e	s	s	w	w	p	w	o	p	n	s	u
20	p	x	s	n	t	p	f	c	n	k	e	e	s	s	w	w	p	w	o	p	n	s	u

Поскольку ввод исходных данных в систему «Эйдос» планируется осуществить с помощью ее универсального программного интерфейса импорта данных из внешних баз данных, который работает с файлами MS Excel, то преобразуем данные из csv-файла в xls-файл, для удобства дальнейшего редактирования и выполним следующие операции.

Конвертируем исходный файл из csv формата в xls и запишем его с именем: Inp_data.xls в папку: c:\Aidos-X\AID_DATA\Inp_data\. В результате получим таблицу исходных данных, полностью подготовленную для обработки в системе «Эйдос» и записанную в нужную папку в виде файла нужного типа с нужным именем.

Автоматизированная формализация предметной области путем импорта исходных данных из внешних баз данных в систему "Эйдос".

Для загрузки базы исходных данных в систему «Эйдос» необходимо воспользоваться универсальным программным интерфейсом для ввода данных из внешних баз данных табличного вида, т.е. режимом 2.3.2.2.

Задайте тип файла исходных данных: "Inp_data":

☒ XLS - MS Excel-2003 Стандарт XLS-файла
☐ XLSX- MS Excel-2007(2010)
☐ DBF - DBASE IV (DBF/NTX) Стандарт DBF-файла
☐ CSV - Comma-Separated Values Стандарт CSV-файла

Задайте параметры:

☒ Нули и пробелы считать ОТСУТСТВИЕМ данных
☐ Нули и пробелы считать ЗНАЧЕНИЯМИ данных
☒ Создавать БД средних по классам "Inp_davr.dbf"?
 Требования к файлу исходных данных

Задайте диапазон столбцов классификационных шкал:

Начальный столбец классификационных шкал:
 Конечный столбец классификационных шкал:

Задайте диапазон столбцов описательных шкал:

Начальный столбец описательных шкал:
 Конечный столбец описательных шкал:

Задайте режим:

☒ Формализации предметной области (на основе "Inp_data")
☐ Генерации распознаваемой выборки (на основе "Inp_rasp")

Задайте способ выбора размера интервалов:

☐ Равные интервалы с разным числом наблюдений
☒ Разные интервалы с равным числом наблюдений

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

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

Параметры интерпретации значений текстовых полей "Inp_data":

Интерпретация TXT-полей классов:

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

Интерпретация TXT-полей признаков:

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

Какие наименования ГРАДАЦИЙ числовых шкал использовать:

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

Ok Cancel

Рисунок 1. Экранная форма Универсального программного интерфейса импорта данных в систему "Эйдос" (режим 2.3.2.2.)

В экранной форме, приведенной на рисунке 2, задать настройки, показанные на рисунке:

- "Задайте тип файла исходных данных Inp_data": "XLS – MS Excel-2003";
- "Задайте диапазон столбцов классификационных шкал": "Начальный столбец классификационных шкал" – 2, "Конечный столбец классификационных шкал" – 3(последний столбец в таблице);

– "Задайте диапазон столбцов описательных шкал": "Начальный столбец описательных шкал" – 4, "Конечный столбец описательных шкал" – 24;

– "Задание параметров формирования сценариев или способа интерпретации текстовых полей": "Не применять сценарный метод АСК-анализа".

После нажать кнопку "ОК". Далее открывается окно, где размещена информация о размерности модели (рисунок 2). В этом окне необходимо нажать кнопку "Выйти на создание модели".

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

ИНФОРМАЦИЯ О РАЗМЕРНОСТИ МОДЕЛИ

Суммарное количество градаций классификационных и описательных шкал: [6 x 55]

Тип шкалы	Количество классификационных шкал	Количество градаций классификационных	Среднее количество градаций на класс.шкалу	Количество описательных шкал	Количество градаций описательных шкал	Среднее количество градаций на опис.шкалу
Числовые	0	0	0,00	0	0	0,00
Текстовые	2	6	3,00	21	55	2,62
ВСЕГО:	2	6	3,00	21	55	2,62

Задайте число интервалов (градаций) в шкале: _____

Пересчитать шкалы и градации Выйти на создание модели

Рисунок 2. Задание размерности модели системы "Эйдос"

Далее открывается окно, отображающее стадию процесса импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос" (рисунок 3), а также прогноз времени завершения этого процесса. В том окне необходимо дождаться завершения формализации предметной области и нажать кнопку "ОК".

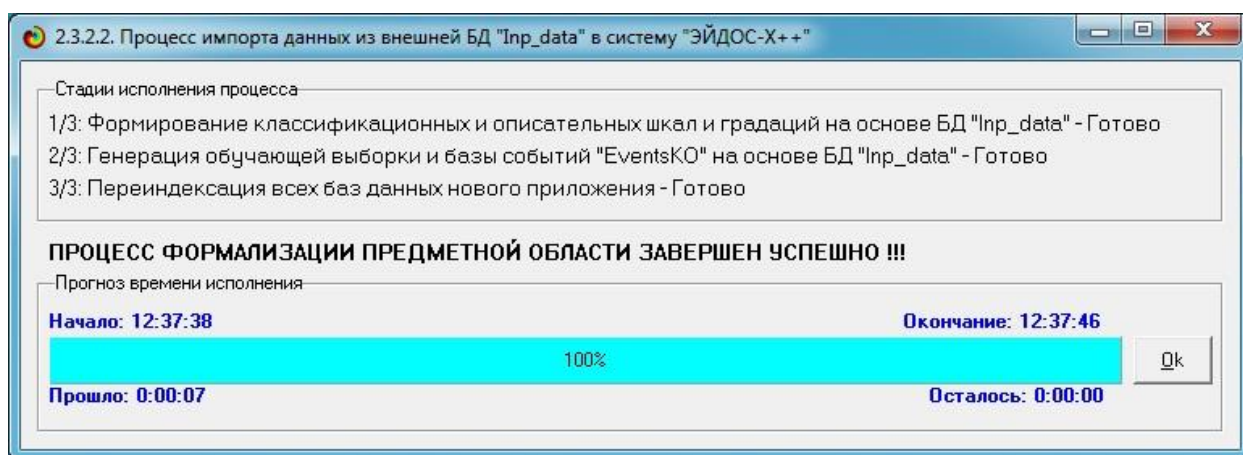


Рисунок 3. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

В процессе формализации предметной области обнаружилось 6 шкал, без вариабельности градации, поэтому они были проигнорированы.

Шкалы: 8(gill-attachment), 16(stalk-color-above-ring), 17(stalk-color-below-ring), 18(veil-type), 19(veil-color), 20(ring-number).

В результате формируются классификационные и описательные шкалы и градации, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Этим самым полностью автоматизировано выполняется 2-й этап АСК-анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1 (рисунок 4).

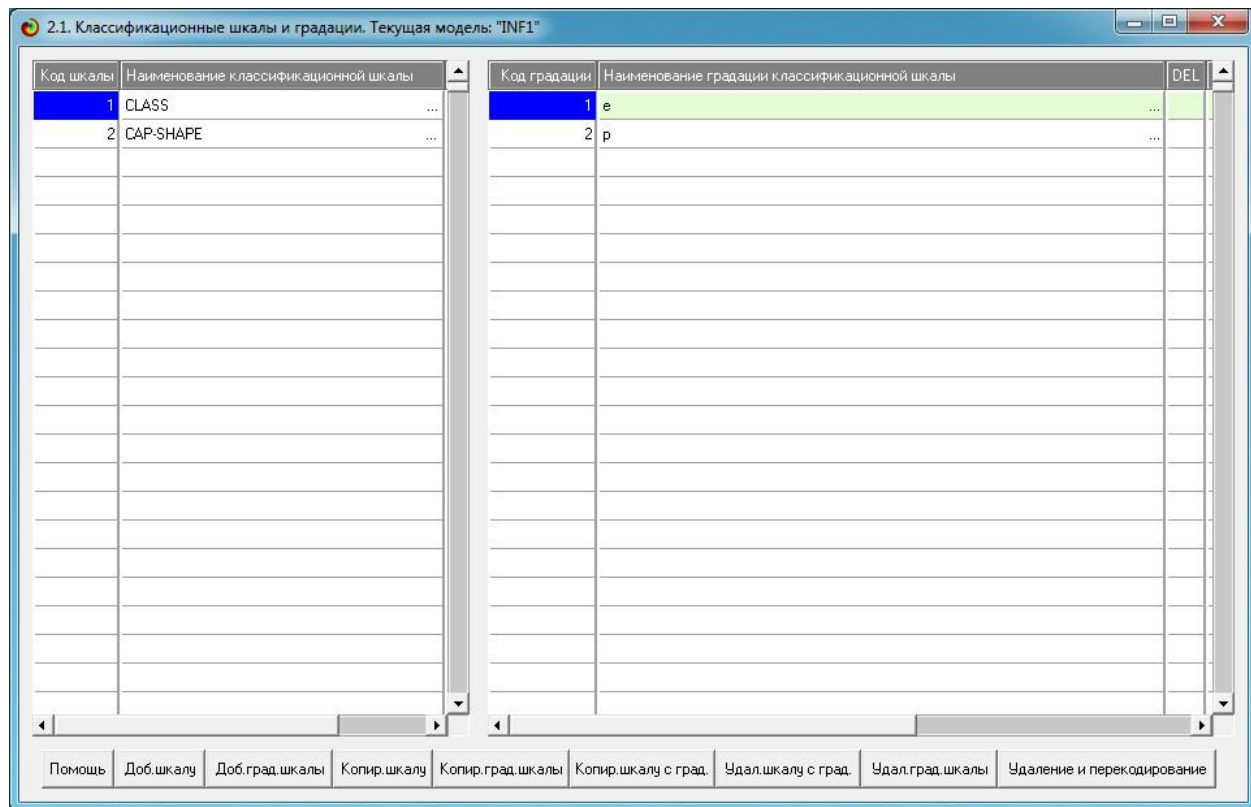


Рисунок 4. Классификационные шкалы и градации (фрагмент)

Для просмотра описательных шкал и градаций необходимо запустить режим 2.2 (рисунок 5):

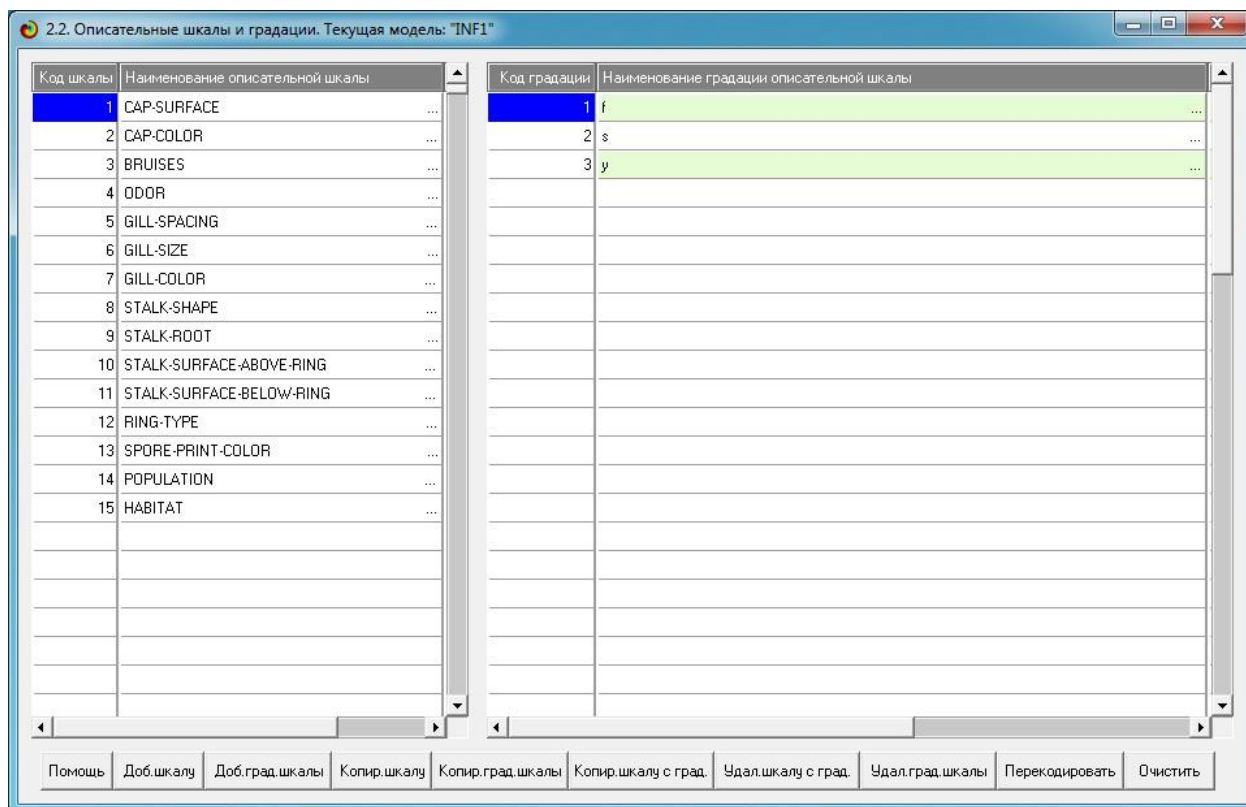


Рисунок 5. Описательные шкалы и градации (фрагмент)

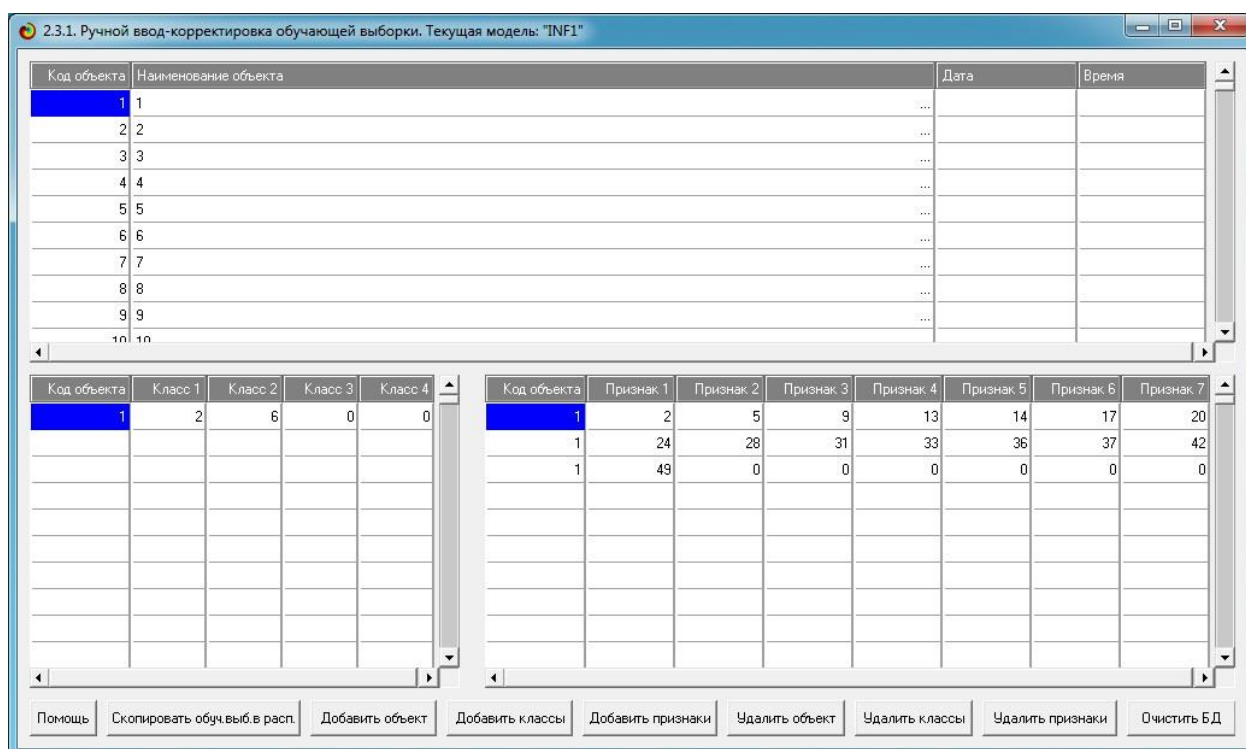


Рисунок 6. Обучающая выборка (фрагмент)

Тем самым создаются все необходимые и достаточные предпосылки для выявления силы и направления причинно-следственных связей между значениями факторов и результатами их совместного системного воздействия (с учетом нелинейности системы [11]).

1.3. Синтез и верификация статистических и интеллектуальных моделей

Далее запускаем режим 3.5, в котором задаются модели для синтеза и верификации, а также задается модель, которой по окончании режима присваивается статус текущей (рисунок 7).

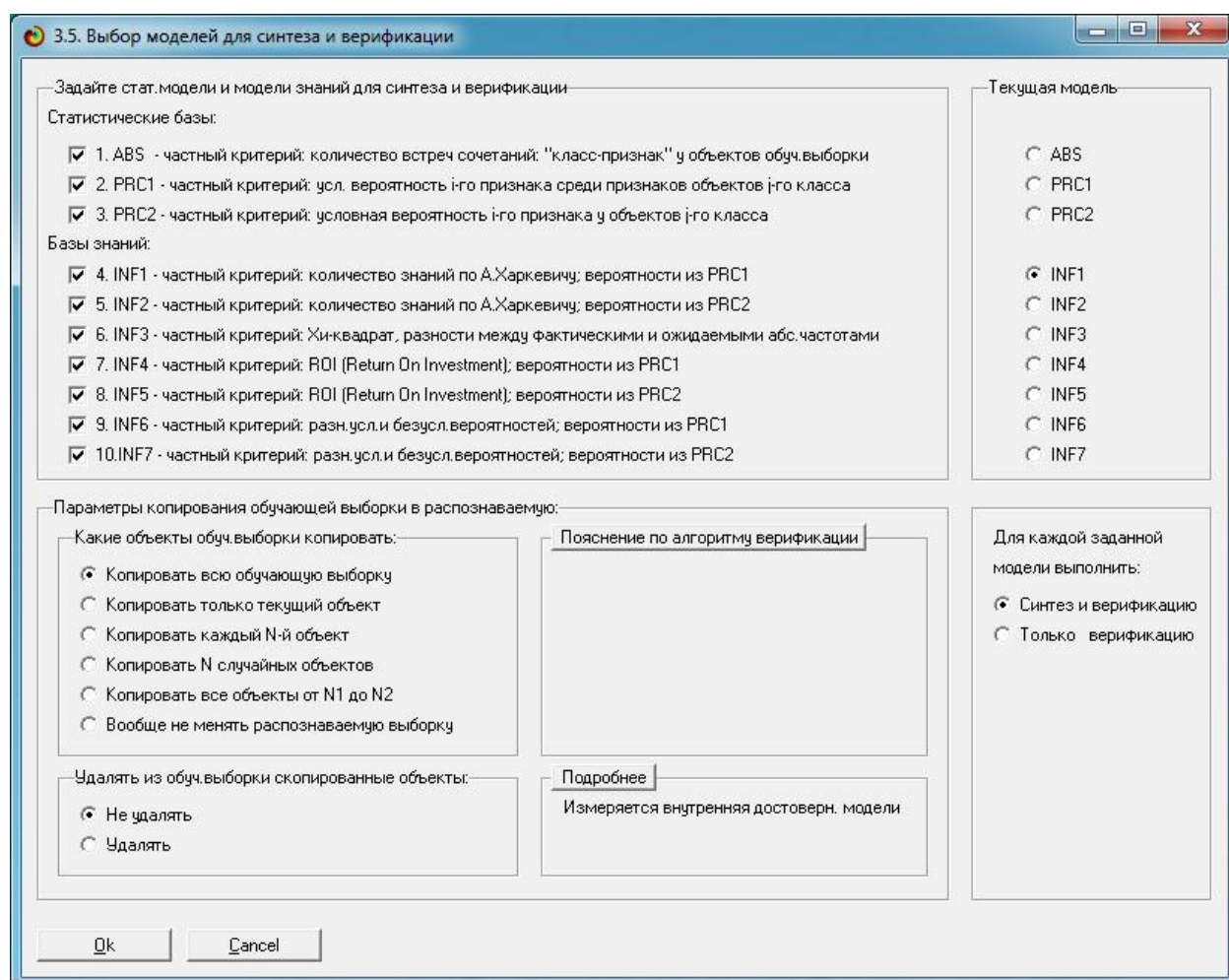


Рисунок 7. Выбор моделей для синтеза и верификации, а также текущей модели

В данном режиме имеется много различных методов верификации моделей, в том числе и поддерживающие бутстрепный метод. Но мы используем параметры по умолчанию, приведенные на рисунке 8. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 9.

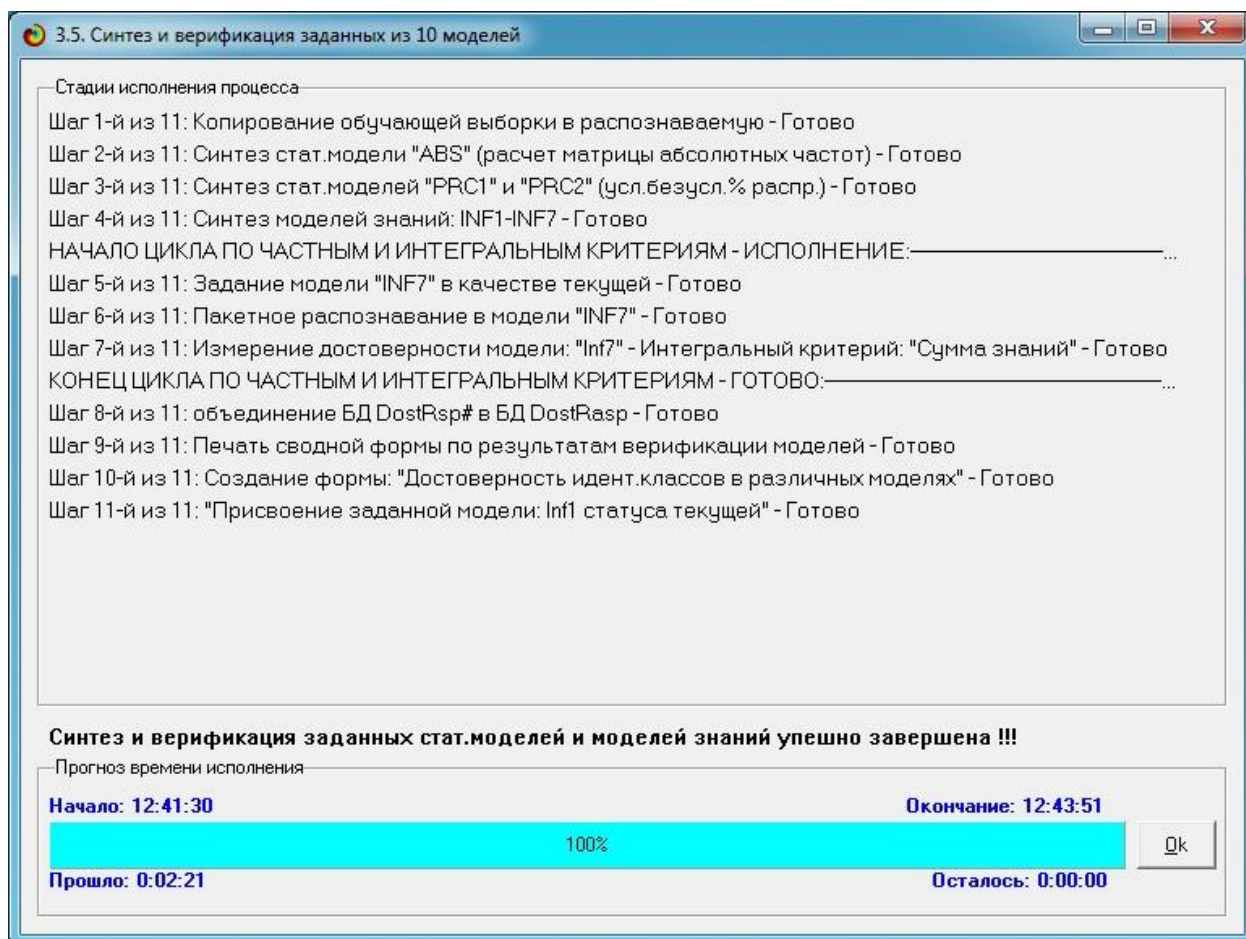


Рисунок 8. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 141 секунд. При этом верификация (оценка достоверности моделей) проводилась на всех 299 примерах наблюдения из обучающей выборки. В результате выполнения режима 3.5 созданы все модели, со всеми частными критериями, перечисленные на рисунке 9, но ниже мы приведем лишь некоторые из них (таблицы 2, 3, 4).

1.4. Виды моделей системы «Эйдос»

Рассмотрим решение задачи идентификации на примере модели INF1, в которой рассчитано количество информации по А.Харкевичу, которое мы получаем о принадлежности идентифицируемого объекта к каждому из классов, если знаем, что у этого объекта есть некоторый признак.

По сути, частные критерии представляют собой просто формулы для преобразования матрицы абсолютных частот (таблица 2) в матрицы условных и безусловных процентных распределений, и матрицы знаний (таблицы 3 и 4) (проф. В.И.Лойко).

Таблица 2 – Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений (фрагменты)

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

Код признака	Наименование описательной шкалы и градации	1. CLASS E	2. CLASS P	3. CAP SHAPE B	4. CAP SHAPE F	5. CAP SHAPE S	6. CAP SHAPE X	Сумма	Среднее	Средн. квадрат
1	CAP-SURFACE-I	54			20	13	21	108	18.000	19.910
2	CAP-SURFACE-s	97	12	41	11		57	218	36.333	36.560
3	CAP-SURFACE-y	115	21	38	24		74	272	45.333	42.027
4	CAP-COLOR-g	24			9	6	9	48	8.000	8.832
5	CAP-COLOR-n	44	13		20	7	30	114	19.000	16.050
6	CAP-COLOR-w	95	20	44	10		61	230	38.333	35.680
7	CAP-COLOR-y	103		35	16		52	206	34.333	39.297
8	BRUISES-I	52			17	13	22	104	17.333	19.201
9	BRUISES-I	214	33	79	38		130	494	82.333	78.482
10	ODOR-a	106		36	17		53	212	35.333	40.347
11	ODOR-I	108		43	17		48	216	36.000	40.817
12	ODOR-n	52			17	13	22	104	17.333	19.201
13	ODOR-p		33		4		29	66	11.000	15.620
14	GILL-SPACING-c	220	33	79	35	13	126	506	84.333	77.881
15	GILL-SPACING-w	46			20		26	92	15.333	18.875
16	GILL-SIZE-b	212		79	30		103	424	70.667	80.909
17	GILL-SIZE-n	54	33		25	13	49	174	29.000	20.756
18	GILL-COLOR-g	51		23	3	3	22	102	17.000	19.483
19	GILL-COLOR-h	3			2		1	6	1.000	1.265
20	GILL-COLOR-k	50	12	17	7	6	32	124	20.667	17.201
21	GILL-COLOR-n	68	7	21	10	1	43	150	25.000	25.745
22	GILL-COLOR-p	30	7		15	3	19	74	12.333	11.237
23	GILL-COLOR-w	64	7	18	18		35	142	23.667	23.054
24	STALK-SHAPE-e	220	33	79	35	13	126	506	84.333	77.881
25	STALK-SHAPE-j	46			20		26	92	15.333	18.875

Таблица 3 – Матрица информативности (модель INF1) в битах (фрагмент)

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

Код признака	Наименование описательной шкалы и градации	1. CLASS E	2. CLASS P	3. CAP SHAPE B	4. CAP SHAPE F	5. CAP SHAPE S	6. CAP SHAPE X	Сумма	Среднее	Средн. квадрат
1	CAP-SURFACE-f	0.033			0.199	0.486	-0.076	0.642	0.107	0.207
2	CAP-SURFACE-s	0.000	-0.001	0.100	-0.171		0.008	-0.063	-0.010	0.088
3	CAP-SURFACE-y	-0.014	0.095	0.016	-0.012		0.019	0.104	0.017	0.041
4	CAP-COLOR-g	0.033			0.202	0.497	-0.086	0.646	0.108	0.213
5	CAP-COLOR-n	-0.040	0.206		0.183	0.295	0.010	0.654	0.109	0.137
6	CAP-COLOR-w	-0.021	0.129	0.105	-0.213		0.012	0.013	0.002	0.121
7	CAP-COLOR-y	0.033		0.071	-0.048		-0.002	0.055	0.009	0.040
8	BRUISES-f	0.033			0.163	0.497	-0.052	0.641	0.107	0.204
9	BRUISES-f	-0.008	0.054	0.054	-0.051		0.010	0.060	0.010	0.040
10	ODOR-s	0.033		0.071	-0.039		-0.005	0.061	0.010	0.038
11	ODOR-f	0.033		0.116	-0.044		-0.038	0.067	0.011	0.059
12	ODOR-n	0.033			0.163	0.497	-0.052	0.641	0.107	0.204
13	ODOR-p		0.626		-0.118		0.155	0.663	0.110	0.267
14	GILL-SPACING-c	-0.006	0.047	0.047	-0.081	0.047	-0.006	0.049	0.008	0.051
15	GILL-SPACING-w	0.033			0.244		0.030	0.308	0.051	0.096
16	GILL-SIZE-b	0.033		0.098	-0.074		-0.013	0.044	0.007	0.057
17	GILL-SIZE-n	-0.102	0.351		0.127	0.351	0.029	0.755	0.126	0.189
18	GILL-COLOR-g	0.033		0.152	-0.324	0.086	-0.047	-0.100	-0.017	0.165
19	GILL-COLOR-h	0.033			0.366		-0.120	0.279	0.047	0.165
20	GILL-COLOR-k	-0.028	0.160	0.011	-0.139	0.227	0.004	0.235	0.039	0.133
21	GILL-COLOR-n	0.005	-0.048	0.016	-0.091	-0.336	0.034	-0.419	-0.070	0.138
22	GILL-COLOR-p	-0.026	0.153		0.224	0.177	0.003	0.531	0.089	0.108
23	GILL-COLOR-w	0.004	-0.032	-0.012	0.091		-0.009	0.042	0.007	0.043
24	STALK-SHAPE-e	-0.006	0.047	0.047	-0.081	0.047	-0.006	0.049	0.008	0.051
25	STALK-SHAPE-f	0.033			0.244		0.030	0.308	0.051	0.096

Таблица 4 – Матрица знаний (модель INF3) (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. CLASS E	2. CLASS P	3. CAP SHAPE B	4. CAP SHAPE F	5. CAP SHAPE S	6. CAP SHAPE X	Сумма	Среднее	Средн. квадрат
1	CAP-SURFACE-f	5.960	-5.960	-14.268	10.067	10.652	-6.452			10.306
2	CAP-SURFACE-s	0.030	-0.030	12.201	-9.050	-4.739	1.589			7.152
3	CAP-SURFACE-y	-5.990	5.990	2.067	-1.017	-5.913	4.863			5.209
4	CAP-COLOR-g	2.649	-2.649	-6.341	4.585	4.957	-3.201			4.692
5	CAP-COLOR-n	-6.709	6.709	-15.060	9.515	4.522	1.023			9.261
6	CAP-COLOR-w	-7.308	7.308	13.615	-11.154	-5.000	2.538			9.466
7	CAP-COLOR-y	11.368	-11.368	7.786	-2.946	-4.478	-0.361			8.342
8	BRUISES-f	5.739	-5.739	-13.739	7.435	10.739	-4.435			9.433
9	BRUISES-f	-5.739	5.739	13.739	-7.435	-10.739	4.435			9.433
10	ODOR-s	11.699	-11.699	7.993	-2.498	-4.609	-0.886			8.554
11	ODOR-f	11.920	-11.920	14.465	-2.866	-4.696	-6.903			10.689
12	ODOR-n	5.739	-5.739	-13.739	7.435	10.739	-4.435			9.433
13	ODOR-p	-29.358	29.358	-8.719	-2.070	-1.435	12.224			19.777
14	GILL-SPACING-c	-5.077	5.077	12.154	-11.538	2.000	-2.615			8.285
15	GILL-SPACING-w	5.077	-5.077	-12.154	11.538	-2.000	2.615			8.285
16	GILL-SIZE-b	23.398	-23.398	22.987	-8.997	-9.217	-4.773			19.037
17	GILL-SIZE-n	-23.398	23.398	-22.987	8.997	9.217	4.773			19.037
18	GILL-COLOR-g	5.629	-5.629	9.525	-6.381	0.783	-3.926			6.494
19	GILL-COLOR-h	0.331	-0.331	-0.793	1.448	-0.130	-0.525			0.805
20	GILL-COLOR-k	-5.157	5.157	0.619	-4.405	3.304	0.482			4.102
21	GILL-COLOR-n	1.278	-1.278	1.184	-3.796	-2.261	4.873			3.096
22	GILL-COLOR-p	-2.916	2.916	-9.776	8.194	1.391	0.191			6.028
23	GILL-COLOR-w	0.836	-0.836	-0.759	4.940	-3.087	-1.094			2.724
24	STALK-SHAPE-e	-5.077	5.077	12.154	-11.538	2.000	-2.615			8.285
25	STALK-SHAPE-f	5.077	-5.077	-12.154	11.538	-2.000	2.615			8.285

1.5. Результаты верификации моделей

Результаты верификации (оценки достоверности) моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями приведены на рисунке 9.

4.1.3.6. Обобщенная форма достов.моделей при разн.крит. Текущая модель: "INF1"												
Наименование модели и частного критерия	Интегральный критерий	Полнота модели	F-мера Ван Ризбергена	Средняя модуль сходства истинно-полож. решений (ST)	Средняя модуль сходства истинно-отриц. решений (ST)	Средняя модуль сходства ложно-полож. решений (SFP)	Средняя модуль сходства ложно-отриц. решений (SF)	S-Точность модели	S-Полнота модели	L-Точность проф. Е.В. Луценко	Средний модуль сходства истинно-полож. решений	Средний модуль сходства ложно-отриц. решений
1. ABS - частный критерий: количество встреч соитаний "клас...	Корреляция абс частот с обр...	0.977	0.499	389.169	3.007	476.276	1.082	0.450	0.997	0.628	0.666	0.666
1. ABS - частный критерий: количество встреч соитаний "клас...	Средняя абс частот по призна...	1.000	0.500	340.082		225.328		0.601	1.000	0.751	0.569	0.569
2. PR1 - частный критерий: усл. вероятность иго признака сред...	Корреляция усл.отн частот с о...	0.977	0.499	389.169	3.007	476.276	1.082	0.450	0.997	0.620	0.666	0.666
2. PR1 - частный критерий: усл. вероятность иго признака сред...	Средняя усл.отн частот по призна...	1.000	0.500	384.037		647.999		0.372	1.000	0.542	0.642	0.642
3. PR2 - частный критерий: условная вероятность иго признака...	Корреляция усл.отн частот с о...	0.977	0.499	389.166	3.007	476.273	1.082	0.450	0.997	0.620	0.666	0.666
3. PR2 - частный критерий: условная вероятность иго признака...	Средняя усл.отн частот по призна...	1.000	0.500	384.037		647.999		0.372	1.000	0.542	0.642	0.642
4. INF1 - частный критерий: количество знаний по А.Харкевич; в...	Семантический резонанс зна...	0.540	0.564	102.831	327.955	79.079	46.210	0.565	0.690	0.621	0.318	0.318
4. INF1 - частный критерий: количество знаний по А.Харкевич; в...	Средняя знания	0.766	0.489	71.102	44.553	139.910	4.050	0.337	0.946	0.497	0.155	0.155
5. INF2 - частный критерий: количество знаний по А.Харкевич; в...	Семантический резонанс зна...	0.540	0.564	102.831	327.956	79.079	46.210	0.565	0.690	0.621	0.318	0.318
5. INF2 - частный критерий: количество знаний по А.Харкевич; в...	Средняя знания	0.766	0.489	71.102	44.553	139.910	4.050	0.337	0.946	0.497	0.155	0.155
6. INF3 - частный критерий: Хингвард: разности между фактис...	Семантический резонанс зна...	0.699	0.622	160.600	348.786	122.978	43.713	0.578	0.794	0.669	0.403	0.403
6. INF3 - частный критерий: Хингвард: разности между фактис...	Средняя знания	0.699	0.622	133.718	213.142	108.817	21.394	0.578	0.862	0.686	0.320	0.320
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.462	0.520	86.108	340.006	65.341	56.997	0.569	0.602	0.585	0.312	0.312
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Средняя знания	0.794	0.492	54.597	14.446	103.438	1.313	0.345	0.977	0.510	0.115	0.115
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.462	0.520	86.108	340.006	65.341	56.997	0.569	0.602	0.585	0.312	0.312
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Средняя знания	0.794	0.492	54.597	14.446	103.438	1.313	0.345	0.977	0.510	0.115	0.115
9. INF6 - частный критерий: разн усл и безуслов вероятностей; вер...	Семантический резонанс зна...	0.671	0.614	150.666	311.076	109.409	54.490	0.579	0.734	0.648	0.376	0.376
9. INF6 - частный критерий: разн усл и безуслов вероятностей; вер...	Средняя знания	0.699	0.451	66.493	41.129	164.707	4.050	0.288	0.943	0.441	0.159	0.159
10. INF7 - частный критерий: разн усл и безуслов вероятностей; ве...	Семантический резонанс зна...	0.671	0.614	150.666	311.076	109.409	54.490	0.579	0.734	0.648	0.376	0.376
10. INF7 - частный критерий: разн усл и безуслов вероятностей; ве...	Средняя знания	0.699	0.451	66.493	41.129	164.707	4.050	0.288	0.943	0.441	0.159	0.159

Рисунок 9. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели ABS, INF3 при интегральном критерии «Семантический резонанс знаний». При этом точность модели составляет 0,560. Таким образом, уровень достоверности прогнозирования с применением модели приблизительно равен показателю экспертных оценок, достоверность которых считается равной примерно 69%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием (рисунок 11).

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++"

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 мной 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 или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представьте себе, что у вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, к 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}:

$$L1 = \frac{TP + TN - FP - FN}{TP + TN + FP + FN}$$
 (нормировка: {-1, +1})

$$L2 = \frac{(TP + TN - FP - FN)}{(TP + TN + FP + FN)} / 2$$
 (нормировка: {0, 1})
где: TP - истинно-положительное решение; TN - истинно-отрицательное решение; FP - ложно-положительное решение; FN - ложно-отрицательное решение;

Классическая F-мера достоверности моделей Ван Ризбергена:
Precision = TP / (TP + FP) - точность модели;
Recall = TP / (TP + FN) - полнота модели;
F-мера = 2 * (Precision * Recall) / (Precision + Recall).

В АСК-анализе и системе "Эйдос" предлагается L-мера, представляющая собой нечеткое мультиклассовое обобщение классической F-меры достоверности моделей Ван Ризбергена:
SPrecision = STP / (STP + SFP) - нечеткая мультиклассовая точность модели;
SRecall = STP / (STP + SFN) - нечеткая мультиклассовая полнота модели;
L-мера = 2 * (SPrecision * SRecall) / (SPrecision + SRecall) - нечеткая мультиклассовая достоверность модели.

Где:
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.

Рисунок 10 Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

Также обращает на себя внимание, что статистические модели, как правило, дают более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, и практически никогда – более высокую. Этим и оправдано применение моделей знаний и интеллектуальных технологий. На рисунке 10 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и неидентифицированных ситуаций в наиболее достоверной модели INF5.

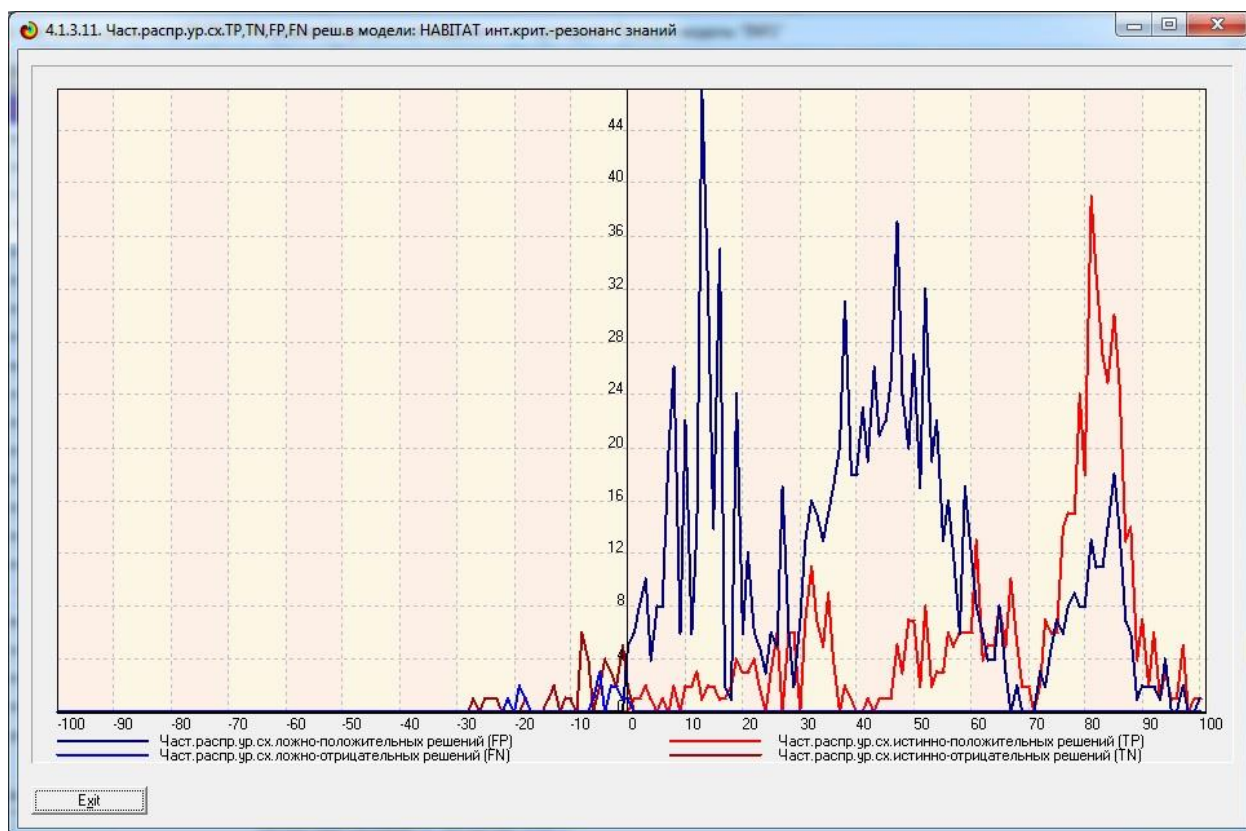


Рисунок 11. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF3

Из рисунка 11 видно, что:

– наиболее достоверная модель INF3 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 9);

– модуль уровня сходства-различия в наиболее достоверной модели INF3 для верно идентифицированных и верно неидентифицированных объектов значительно выше, чем для ошибочно идентифицированных и ошибочно неидентифицированных. Это верно практически для всего диапазона уровней сходства-различия, кроме небольших по модулю значений в диапазоне от 0 до 60% уровня сходства. Для очень больших значений уровней сходства-различия (более 70%) также различие между верно и ошибочно идентифицированными и неидентифицированными ситуациями практически отсутствует.

Любые данные о наблюдениях можно считать суммой истинного значения и шума, причем ни первое, ни второе неизвестны. Поэтому имеет смысл сравнить созданные модели с чисто случайными моделями, совпадающими по основным характеристикам. В системе «Эйдос» есть лабораторная работа № 2.01: «Исследование RND-модели при различных объемах выборки». Если данная работа устанавливается при отсутствии текущего приложения, то все параметры создаваемых моделей задаются вручную, если же текущая модель существует, как в нашем случае, то все основные ее параметры определяются автоматически.

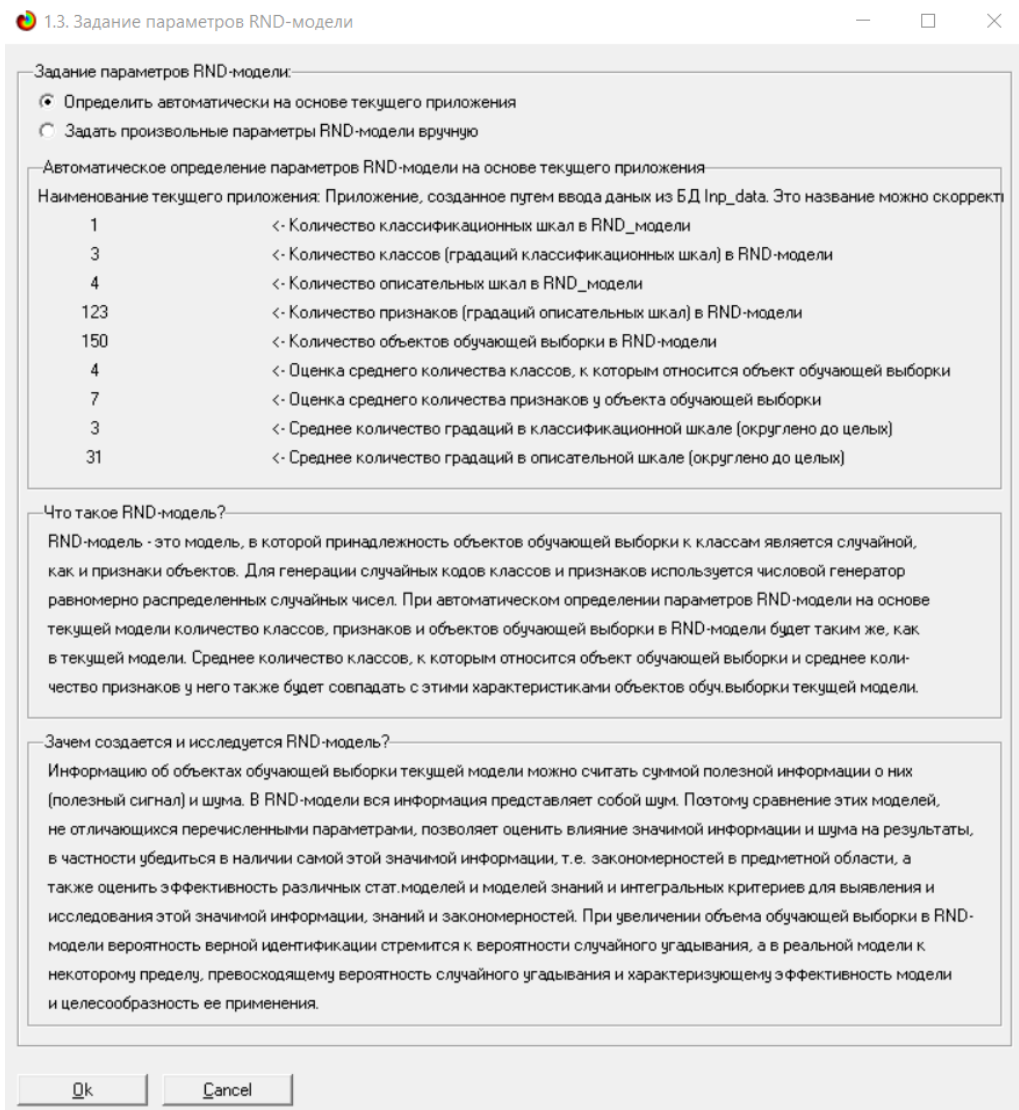


Рисунок 12. Экранная форма управления созданием случайных моделей, совпадающих с текущей по размерностям основных баз данных

На рисунке 12 показано частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF3.

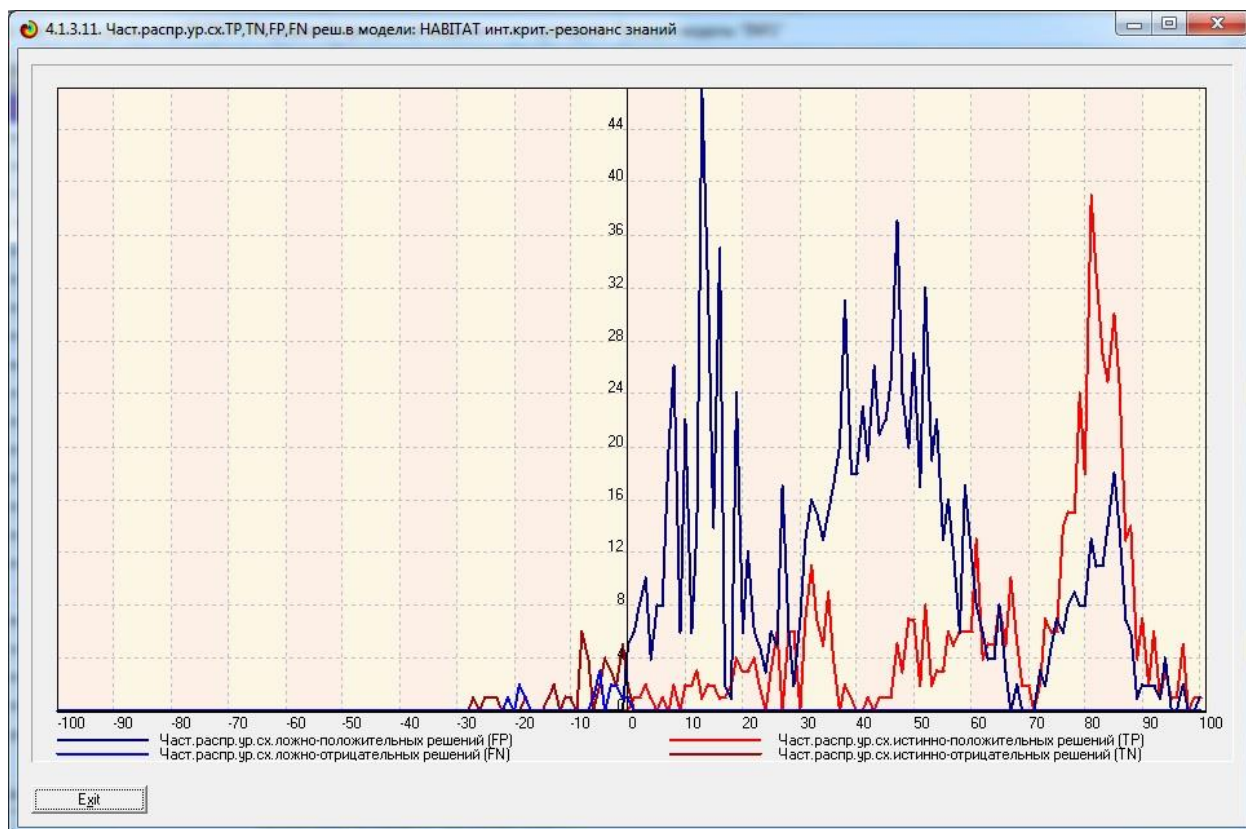


Рисунок 13. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний в случайной модели INF3

Совершенно очевидное различие частотных распределений уровней сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования и случайной модели (рисунки 11 и 13) объясняется тем, что в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны. Если же такой информации в модели нет, то и распределение получается типа, приведенного на рисунке 13.

На рисунке 14 приведены данные по достоверности статистических и когнитивных моделей, созданных на основе случайной выборки.

413.6. Обобщенная форма по достов. моделям при разн. крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	чистота	5-Полнота модели	L1-мера проп. Е.В. Личенко	Средней модель: уровень спад. истинно-полож. решений	Средней модель: уровень спад. истинно-отрицат. решений	Средней модель: уровень спад. ложно-полож. решений	Средней модель: уровень спад. ложно-отрицат. решений	A-Точность модели AРесонанс = АТР/АТР	A-Полнота модели AРесонанс = АТР/АТР	L2-мера проп. Е.В. Личенко
1. ABS - частный критерий: количество встреч соитаний: "х.лас"	Корреляция абс. частот с обр...	0.450	0.997	0.620	0.666	0.077	0.411	0.079	0.618	0.896	0.732
1. ABS - частный критерий: количество встреч соитаний: "х.лас"	Средняя абс. частот по призна...	0.601	1.000	0.751	0.569		0.188		0.751	1.000	0.858
2. PRC1 - частный критерий: усл. вероятность н-го признака сред.	Корреляция усл. отн. частот с о...	0.450	0.997	0.620	0.666	0.077	0.411	0.079	0.618	0.896	0.732
2. PRC1 - частный критерий: усл. вероятность н-го признака сред.	Средняя усл. отн. частот по приз...	0.372	1.000	0.542	0.642		0.542		0.542	1.000	0.703
3. PRC2 - частный критерий: условная вероятность н-го признака...	Корреляция усл. отн. частот с о...	0.450	0.997	0.620	0.666	0.077	0.411	0.079	0.618	0.896	0.732
3. PRC2 - частный критерий: условная вероятность н-го признака...	Средняя усл. отн. частот по приз...	0.372	1.000	0.542	0.642		0.542		0.542	1.000	0.703
4. INF1 - частный критерий: количество знаний по А.Харкевич: в...	Семантический резонанс зна...	0.565	0.690	0.621	0.318	0.168	0.353	0.337	0.474	0.655	0.550
4. INF1 - частный критерий: количество знаний по А.Харкевич: в...	Средняя знания	0.337	0.946	0.497	0.155	0.029	0.171	0.117	0.475	0.843	0.608
5. INF2 - частный критерий: количество знаний по А.Харкевич: в...	Семантический резонанс зна...	0.565	0.690	0.621	0.318	0.168	0.353	0.337	0.474	0.655	0.550
5. INF2 - частный критерий: количество знаний по А.Харкевич: в...	Средняя знания	0.337	0.946	0.497	0.155	0.029	0.171	0.117	0.475	0.843	0.608
6. INF3 - частный критерий: Хинтвадт: разности между факти...	Семантический резонанс зна...	0.578	0.794	0.669	0.403	0.243	0.375	0.402	0.518	0.624	0.566
6. INF3 - частный критерий: Хинтвадт: разности между факти...	Средняя знания	0.570	0.862	0.695	0.320	0.119	0.307	0.246	0.510	0.729	0.600
7. INF4 - частный критерий: ROI (Return On Investment): вероятно...	Семантический резонанс зна...	0.569	0.602	0.595	0.312	0.177	0.349	0.337	0.472	0.638	0.542
7. INF4 - частный критерий: ROI (Return On Investment): вероятно...	Средняя знания	0.345	0.977	0.510	0.115	0.011	0.121	0.043	0.488	0.915	0.637
8. INF5 - частный критерий: ROI (Return On Investment): вероятно...	Семантический резонанс зна...	0.569	0.602	0.595	0.312	0.177	0.349	0.337	0.472	0.638	0.542
8. INF5 - частный критерий: ROI (Return On Investment): вероятно...	Средняя знания	0.345	0.977	0.510	0.115	0.011	0.121	0.043	0.488	0.915	0.637
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Семантический резонанс зна...	0.579	0.734	0.648	0.376	0.277	0.355	0.350	0.514	0.576	0.543
9. INF6 - частный критерий: разн. усл. и безуслов. вероятностей; вер...	Средняя знания	0.288	0.943	0.441	0.159	0.023	0.196	0.115	0.448	0.876	0.592
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Семантический резонанс зна...	0.579	0.734	0.648	0.376	0.277	0.355	0.350	0.514	0.576	0.543
10. INF7 - частный критерий: разн. усл. и безуслов. вероятностей; ве...	Средняя знания	0.288	0.943	0.441	0.159	0.023	0.196	0.115	0.448	0.876	0.592

Понимать

Рисунок 14. Достоверность статистических и когнитивных моделей, созданных на основе случайной выборки

На основе его сравнения с рисунком 9 можно сделать следующие выводы:

- достоверность лучшей модели INF3, отражающей реальный объект моделирования, примерно на 20% выше, чем аналогичной случайной модели;
- различие между достоверностью статистических моделей и моделей знаний, созданных на основе случайной выборки, значительно меньше, чем у моделей, отражающих реальный объект моделирования;
- в реальных моделях кроме шума есть также и информация об истинных причинно-следственных взаимосвязях факторов и их значений с одной стороны, и состояниями объекта моделирования, которые ими обуславливаются, с другой стороны, причем примерно 1/3 достоверности обусловлена отражением в реальных моделях закономерностей предметной области, а 2/3 достоверности обусловлено наличием шума в исходных данных. На основании этого можно предположить, что в исходных данных уровень сигнала о реальных причинно-следственных связях в моделируемой предметной области примерно в два раза ниже уровня шума.

2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи

В соответствии с технологией АСК-анализа зададим текущей модель INF3 (режим 5.6) (рисунок 15) и проведем пакетное распознавание в режиме 4.2.1.

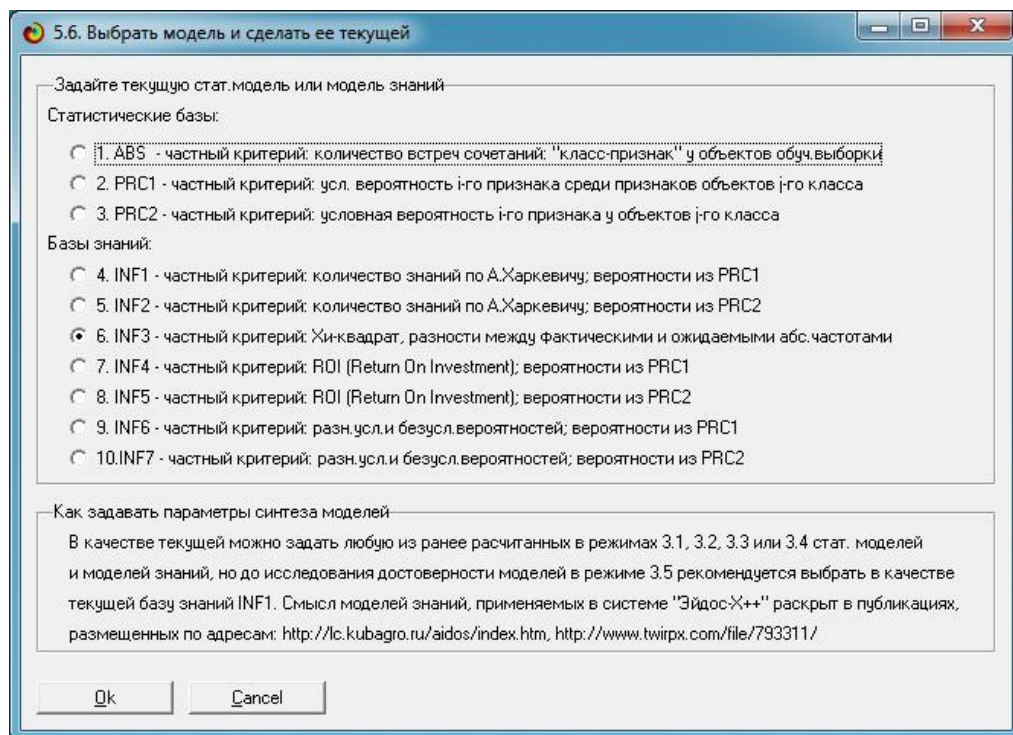
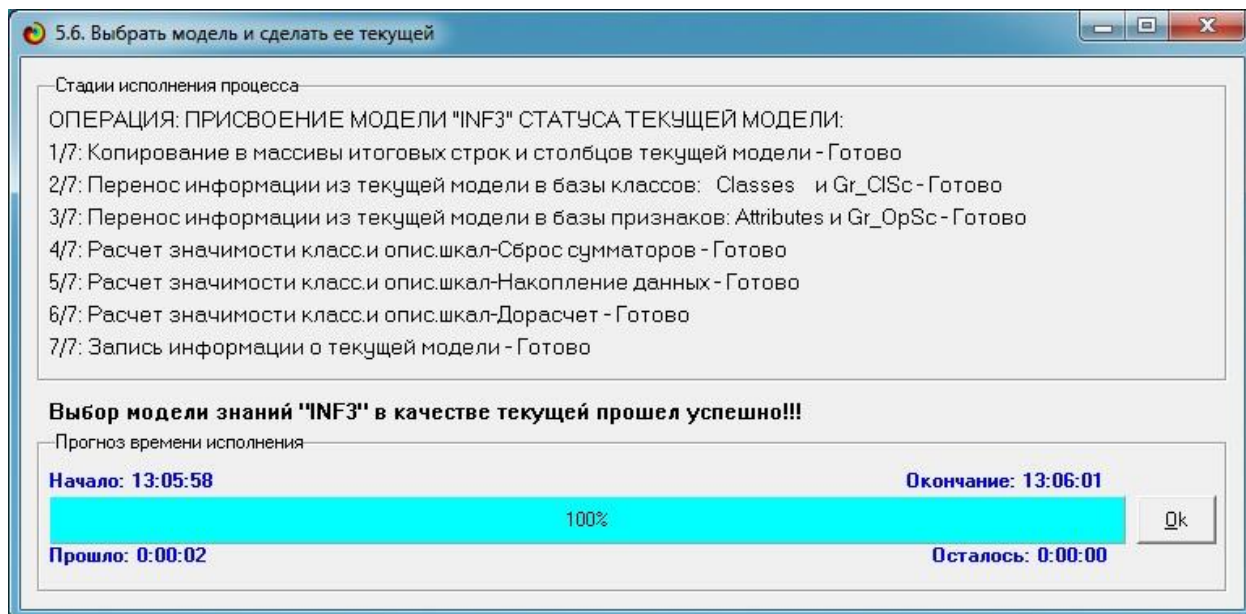


Рисунок 15. Экранные формы режима задания модели в качестве текущей

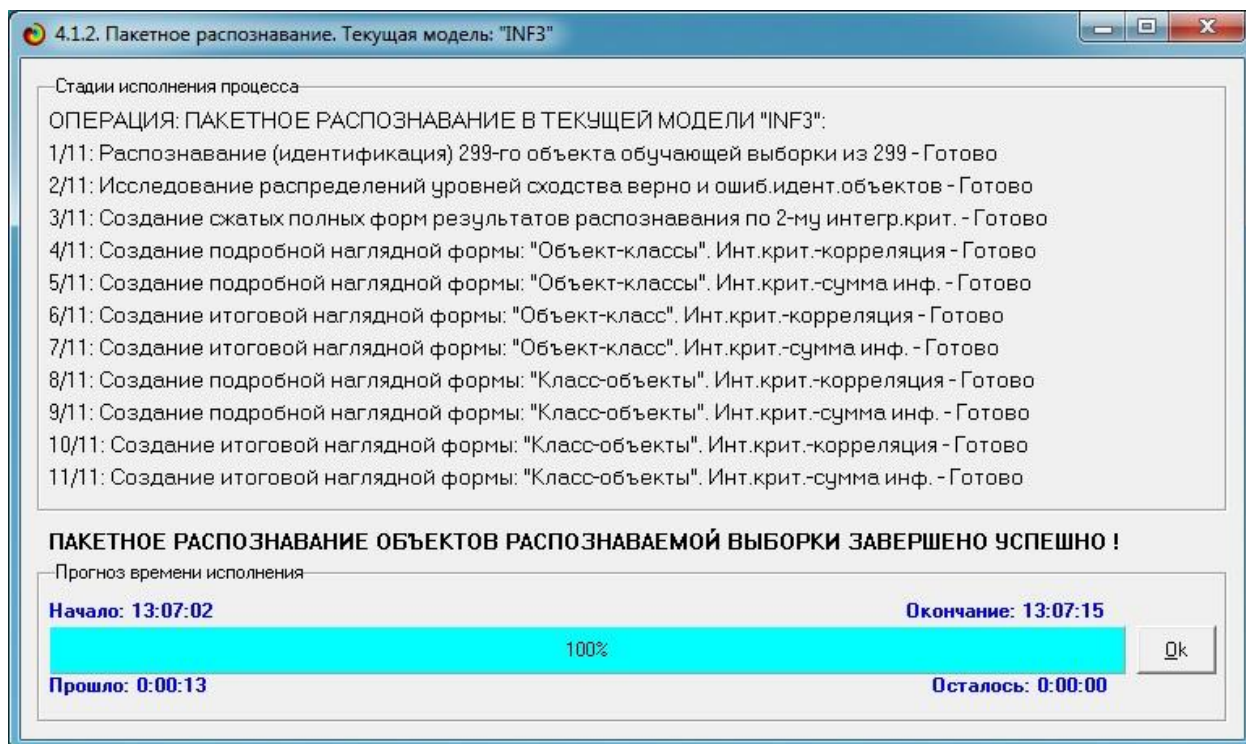


Рисунок 16. Экранная форма режима пакетного распознавания в текущей модели

В результате пакетного распознавания в текущей модели создается ряд баз данных, которые визуализируются в выходных экранных формах, отражающих результаты решения задачи идентификации и прогнозирования.

Режим 4.1.3 системы «Эйдос» обеспечивает отображение результатов идентификации и прогнозирования в различных формах:

1. Подробно наглядно: "Объект – классы".
2. Подробно наглядно: "Класс – объекты".
3. Итоги наглядно: "Объект – классы".
4. Итоги наглядно: "Класс – объекты".
5. Подробно сжато: "Объект – классы".
6. Обобщенная форма по достоверности моделей при разных интегральных критериях.

7. Обобщенный статистический анализ результатов идентификации по моделям и интегральным критериям.

8. Статистический анализ результатов идентификации по классам, моделям и интегральным критериям.

9. Распознавание уровня сходства при разных моделях и интегральных критериях.

10. Достоверность идентификации классов при разных моделях и интегральных критериях.

Ниже кратко рассмотрим некоторые из них.

На рисунках 17 и 18 приведены примеры прогнозов высокой и низкой достоверности частоты и классов ирисов в наиболее достоверной модели INF3 на основе наблюдения предыстории их развития:

Классы	
Код	Наим. класса
1	CLASS-e
2	CLASS-p
3	CAP-SHAPE-b
4	CAP-SHAPE-f
5	CAP-SHAPE-s
6	CAP-SHAPE-x

Интегральный критерий сходства: "Семантический резонанс знаний"				
Код	Наименование объекта	Сходство	Ф...	Сходство
31	31	39,20...	v	██████████
296	296	39,06...	v	██████████
64	64	35,36...	v	██████████
179	179	35,36...	v	██████████
114	114	34,24...	v	██████████
63	63	34,14...	v	██████████
125	125	34,14...	v	██████████
180	180	34,00...	v	██████████
184	184	34,00...	v	██████████

Интегральный критерий сходства: "Сумма знаний"				
Код	Наименование объекта	Сходство	Ф...	Сходство
31	31	29,45...	v	██████████
296	296	29,34...	v	██████████
64	64	26,56...	v	██████████
179	179	26,56...	v	██████████
114	114	25,72...	v	██████████
63	63	25,64...	v	██████████
125	125	25,64...	v	██████████
180	180	25,54...	v	██████████
184	184	25,54...	v	██████████

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

Рисунок 17. Пример идентификации классов в модели INF3

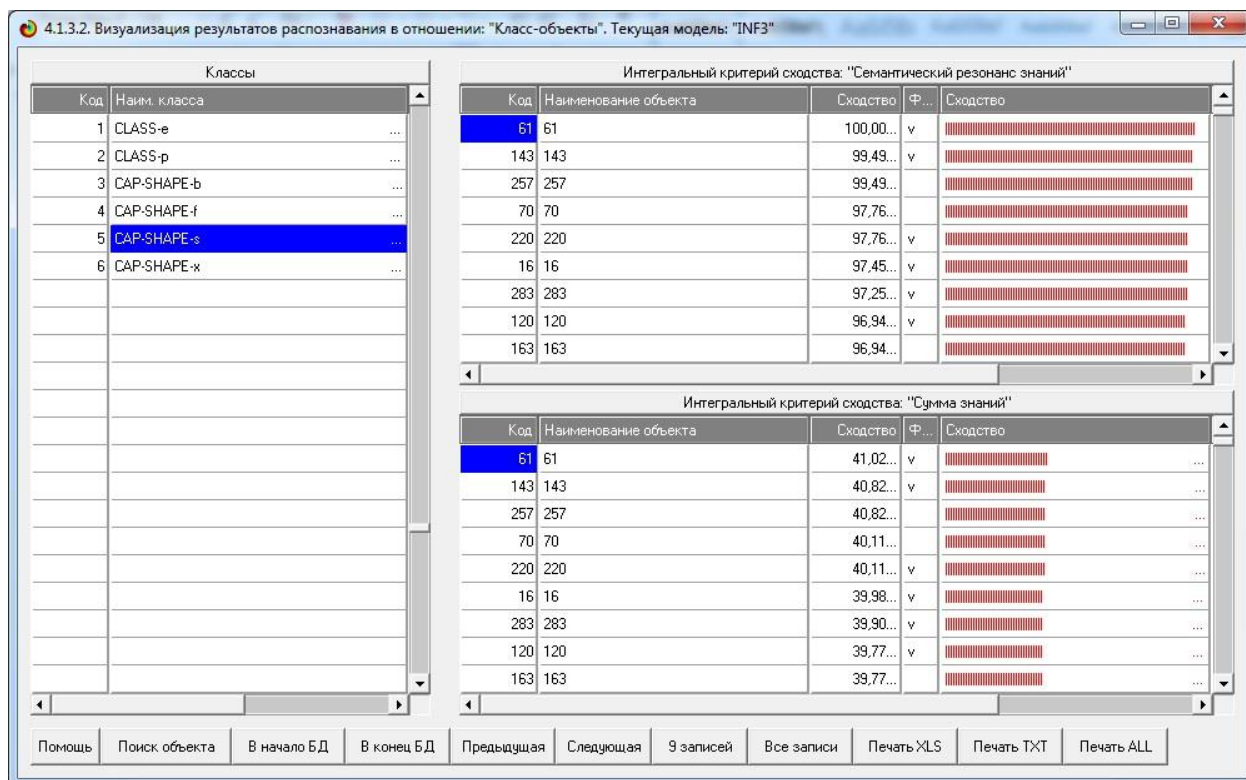


Рисунок 18. Пример идентификации классов в модели INF3

2.2. Когнитивные функции

Рассмотрим режим 4.5, в котором реализована возможность визуализации когнитивных функций для любых моделей и любых сочетаний классификационных и описательных шкал (рисунок 19)

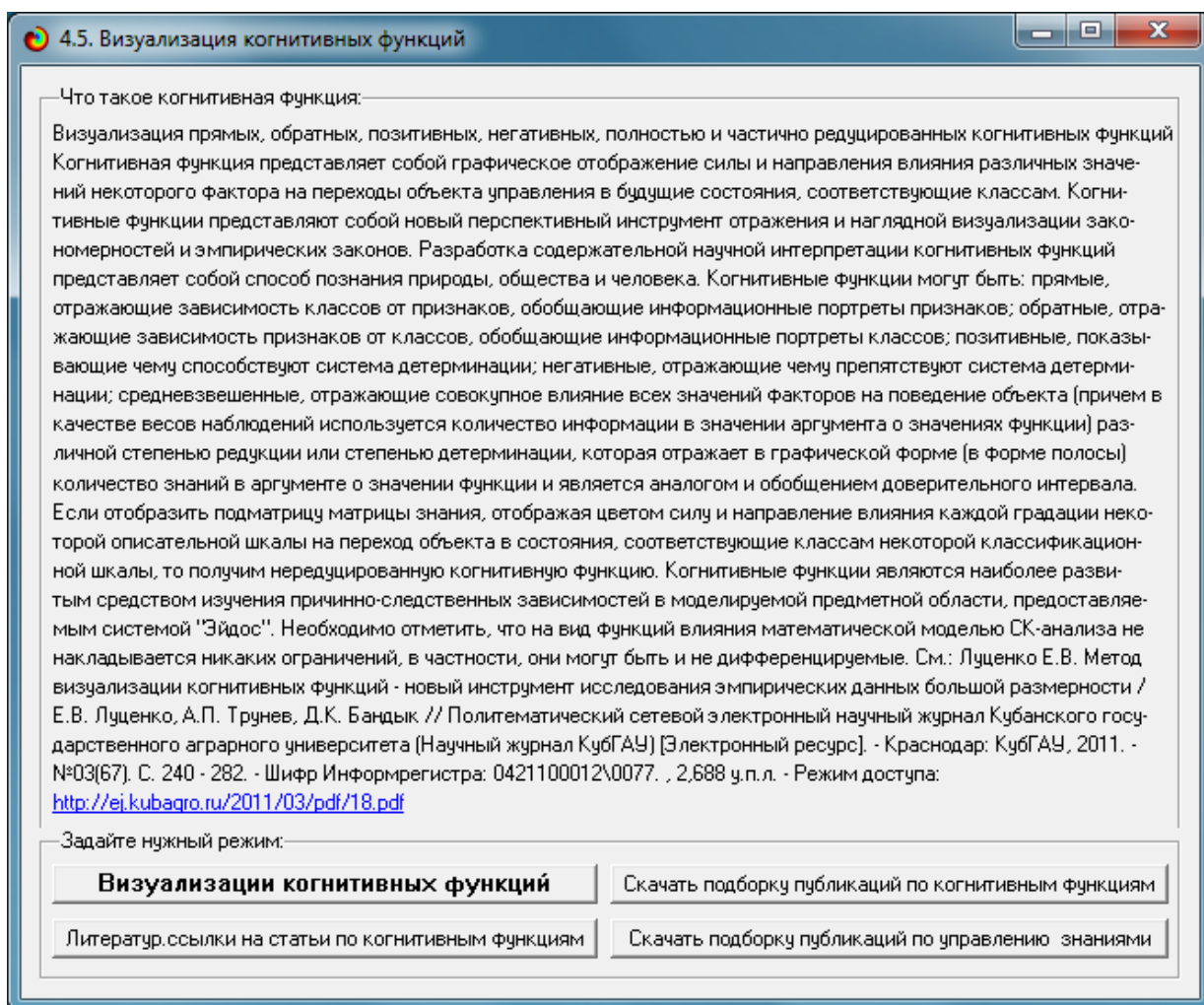


Рисунок 19. Экранная форма режима 4.5 системы «Эйдос-X++» «Визуализация когнитивных функций»

Применительно к задаче, рассматриваемой в данной работе, когнитивная функция показывает, какое количество информации содержится в различных значениях факторов о том, что объект моделирования перейдет в те или иные будущие состояния. Когнитивным функциям посвящено много работ автора 9, но наиболее новой и обобщающей из них является работа. Поэтому здесь не будем останавливаться на описании того, что представляют

собой когнитивные функции в АСК-анализе. На рисунках 20, 21. приведены визуализации некоторых когнитивных функций данного приложения для модели INF3.

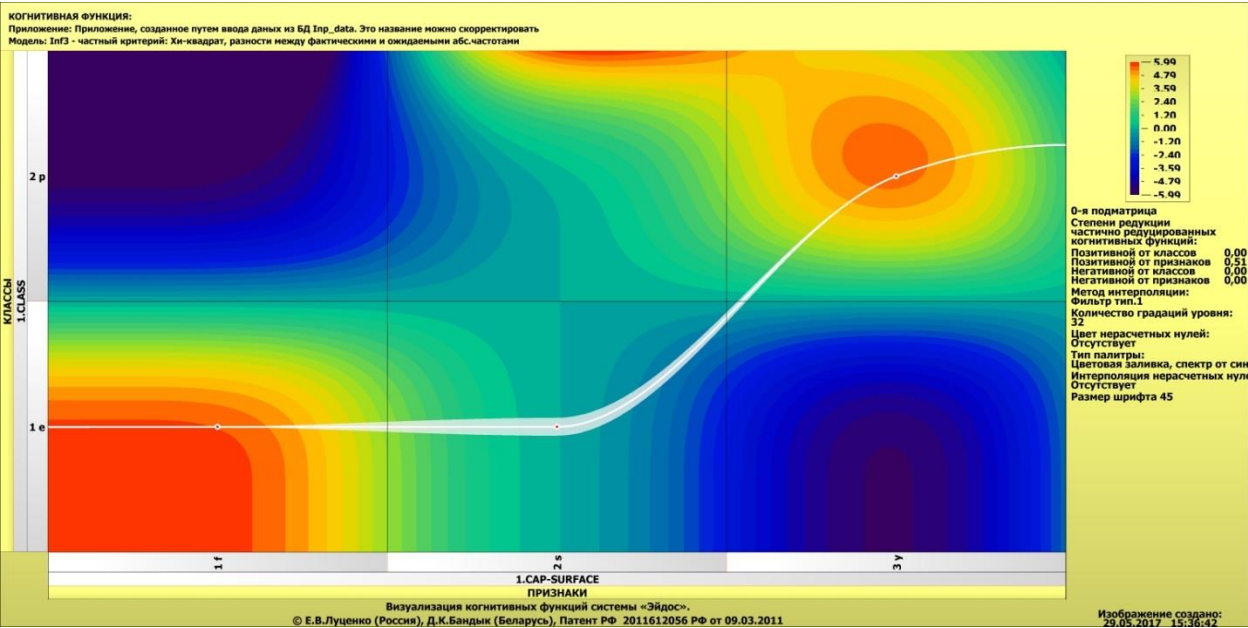


Рисунок 20. Визуализация когнитивных функций зависимости класса гриба (ядовитый, не ядовитый) от типа поверхности шляпки (волокнистый, чешуйчатый, гладкий) INF3

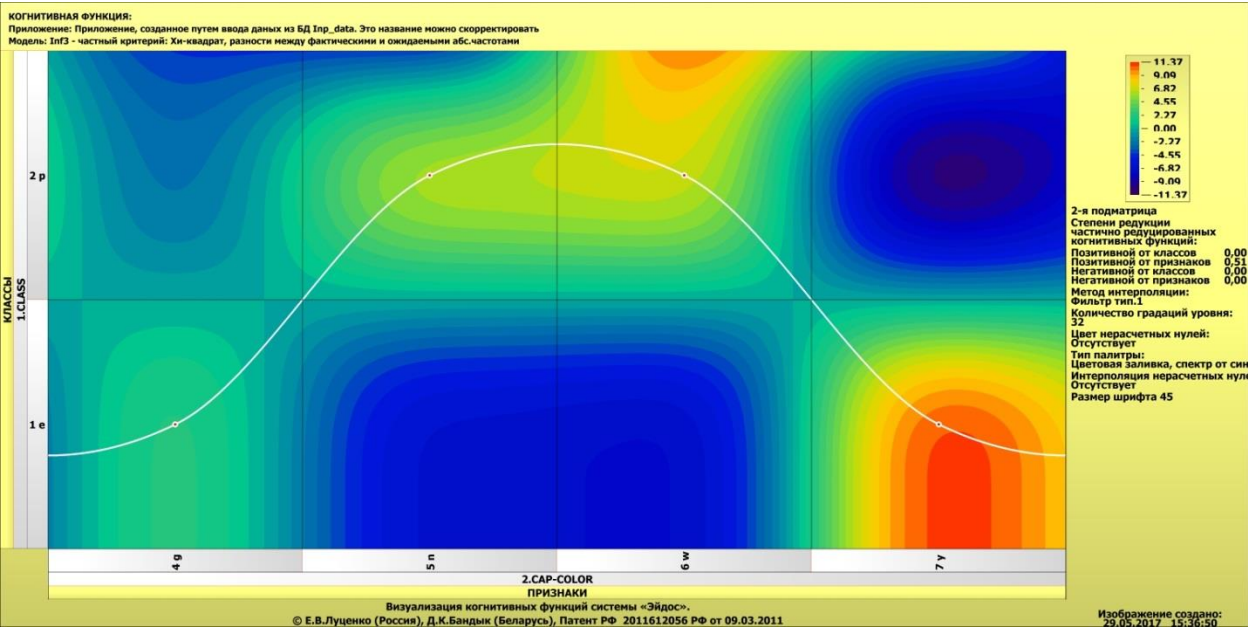


Рисунок 21. Визуализация когнитивных функций зависимости класса гриба (ядовитый, не ядовитый) от цвета шляпки (коричневый, серый, белый, желтый) INF3

В работе приведено 2 визуализации когнитивных функций из десятков, которые не приводятся в связи с ограниченным объемом работы.

2.3. SWOT и PEST матрицы и диаграммы

SWOT-анализ является широко известным и общепризнанным методом стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают не формализуемым путем (интуитивно), на основе своего профессионального опыта и 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 около 30 лет, но она малоизвестна – это интеллектуальная система «Эйдос». Данная система всегда обеспечивала возможность проведения количественного автоматизированного SWOT-анализа без использования экспертных оценок непосредственно на основе эмпирических данных. Результаты SWOT-анализа выводились в форме информационных портретов. В версии системы под MS Windows: «Эйдос-X++» предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (рисунок 22).

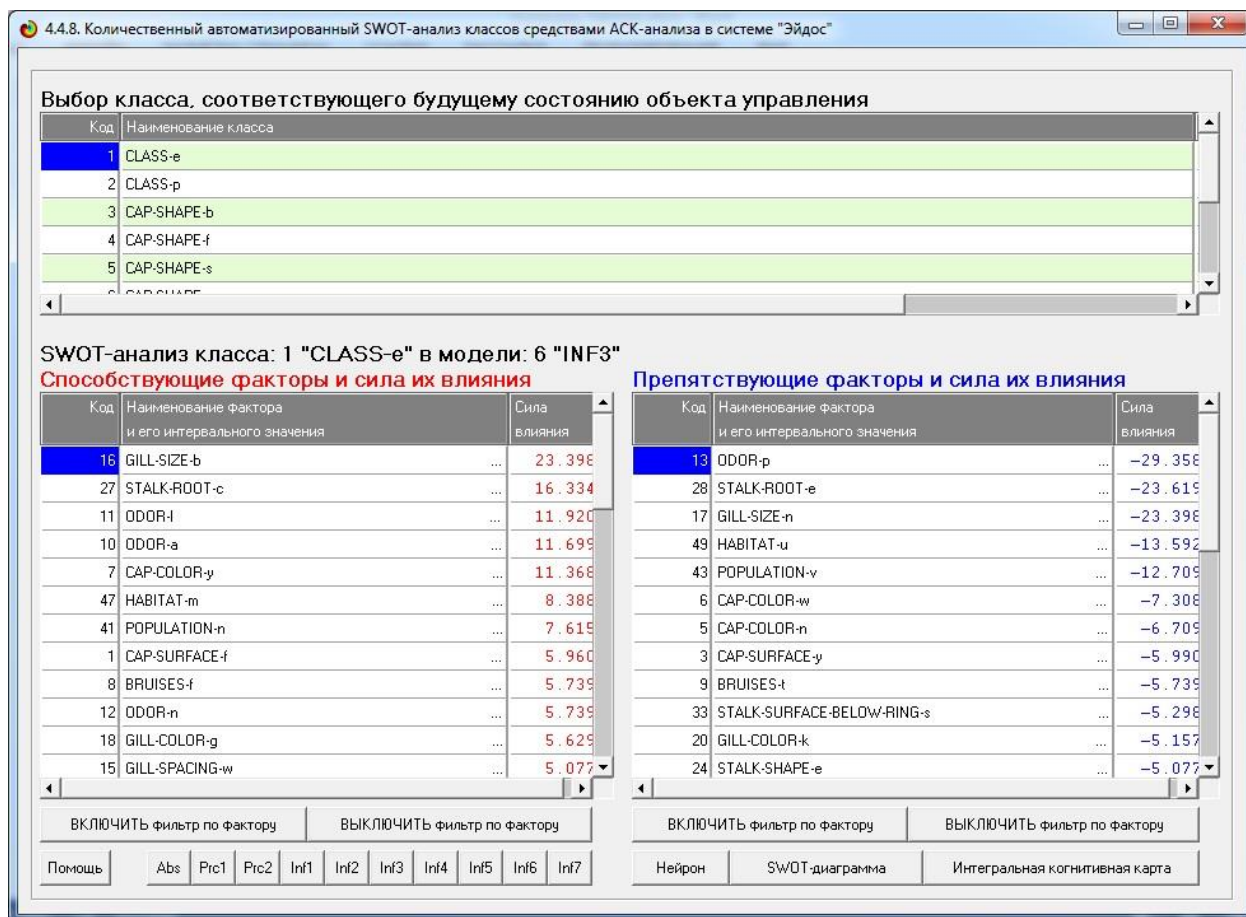


Рисунок 22. Пример SWOT-матрицы в модели INF3

На рисунке 23 приведены примеры инвертированной SWOT- матрицы и инвертированной SWOT-диаграммы в модели INF3.

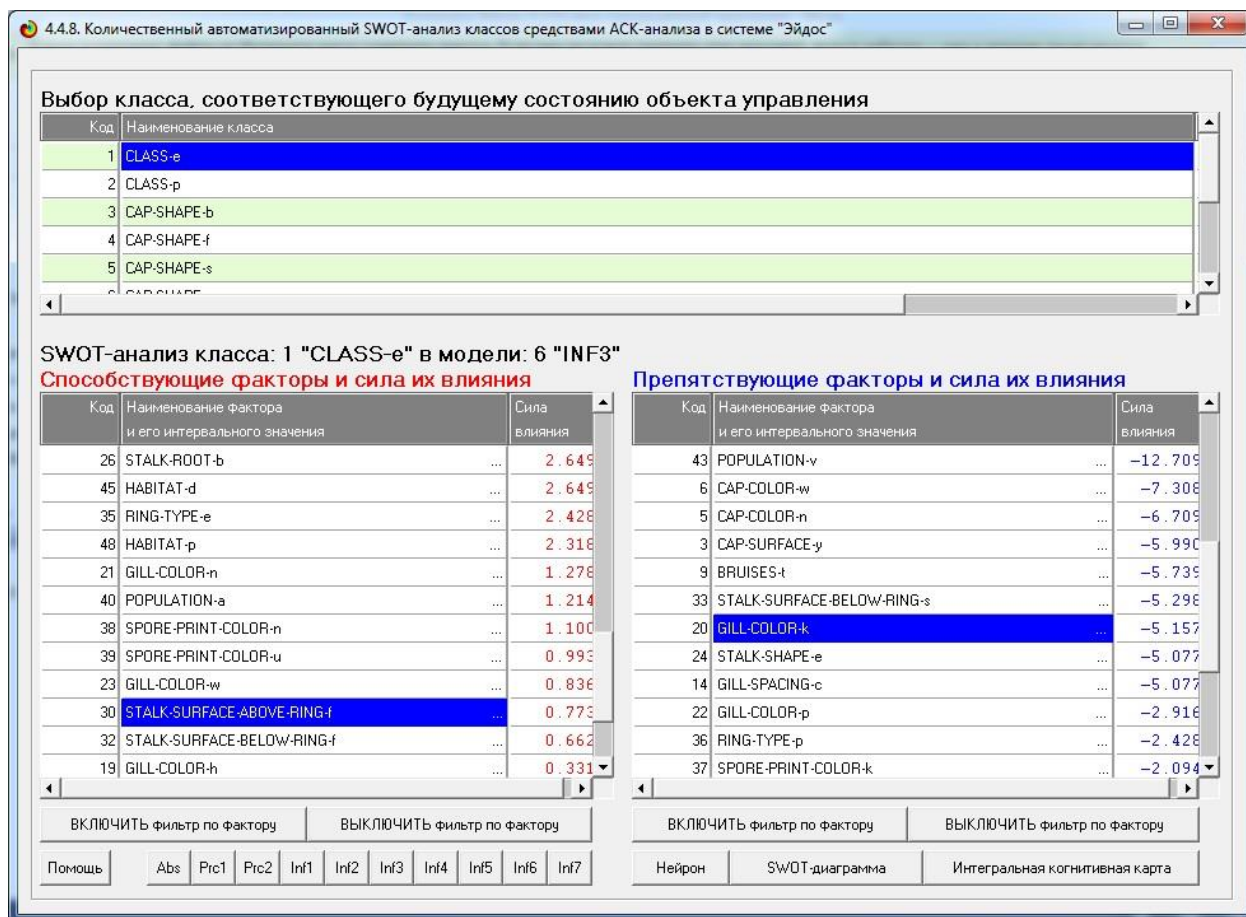


Рисунок 23. Пример SWOT-матрицы в модели INF3

2.4. НЕЛОКАЛЬНЫЕ НЕЙРОННЫЕ СЕТИ И НЕЙРОНЫ.

Каждому классу системно-когнитивной модели соответствует не локальный нейрон совокупность, которых образует не локальную нейронную сеть [9] В качестве примеров приведены рисунок 24-26.

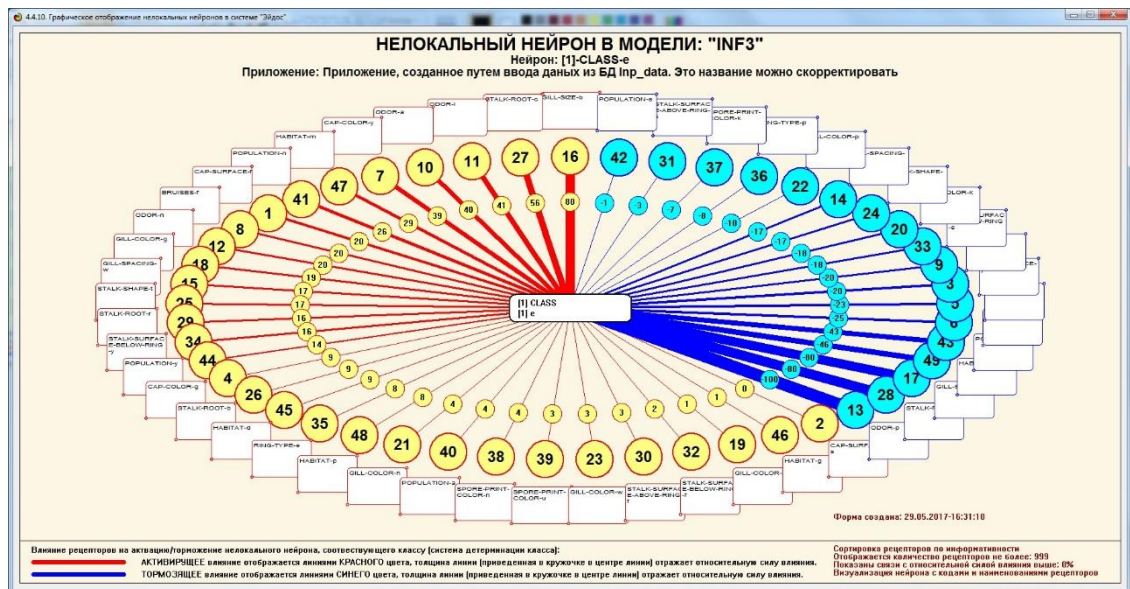


Рисунок 24. Пример нелокальные нейронные сети и нейроны модели INF3

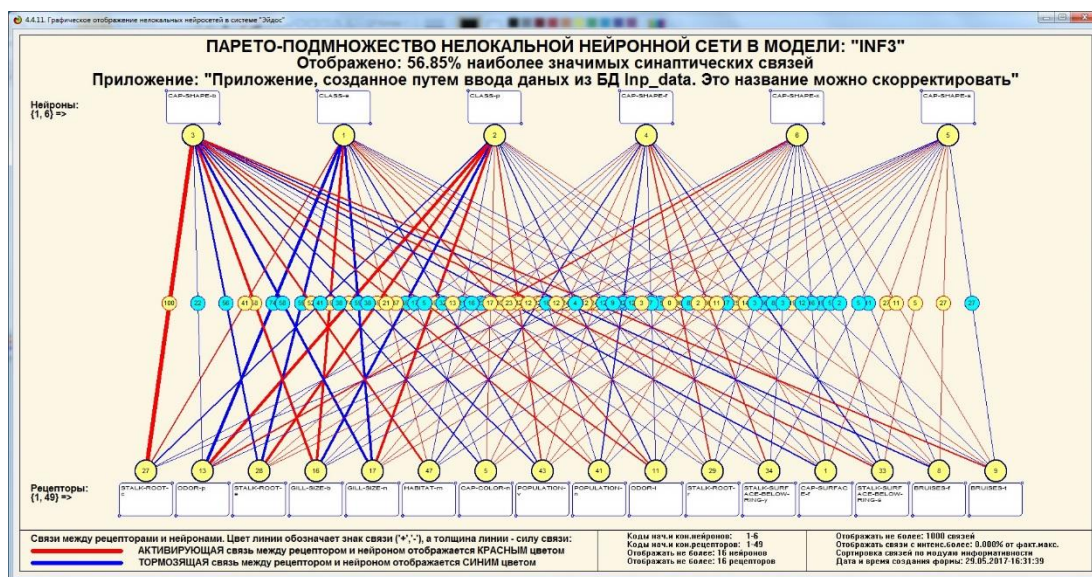


Рисунок 25. Пример нелокальные нейронные сети и нейроны модели INF3

2.5 КЛАСТЕРНЫЙ И КОНСТРУКТИВНЫЙ АНАЛИЗ

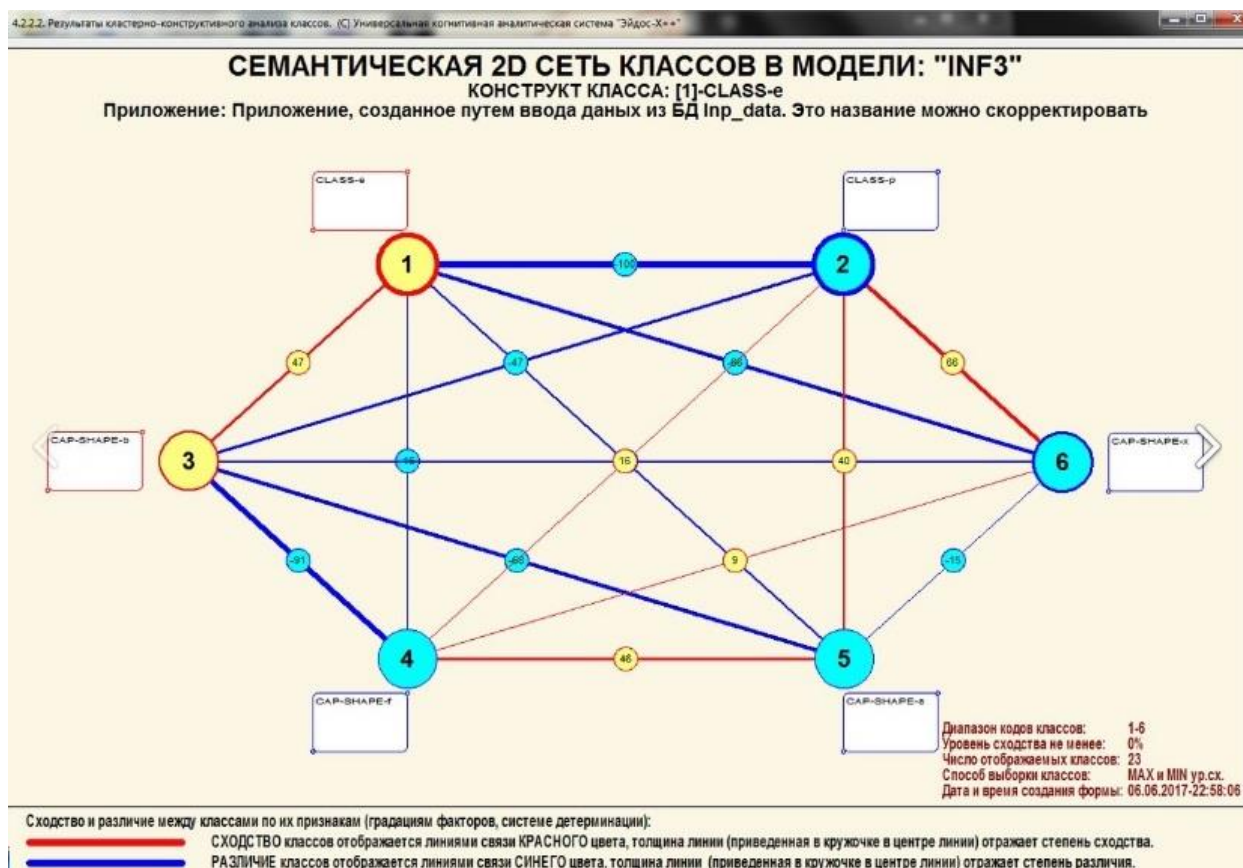


Рисунок 26. Пример кластерный и конструктивный анализ модели INF3

ЗАКЛЮЧЕНИЕ

Так как существует множество систем искусственного интеллекта, то возникает необходимость сопоставимой оценки качества их математических моделей. Одним из вариантов решения этой задачи является тестирование различных системы на общей базе исходных данных, для чего очень удобно использовать общедоступную базу Kaggle. В данной курсовой работе приводится развернутый пример использования базы данных Kaggle для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». При этом наиболее достоверной в данном приложении оказались модели INF5, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,687, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 69%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется метрика, сходная с F-критерием. Также обращает на себя внимание, что статистические модели в данном приложении дают примерно на 21% более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, что, как правило, наблюдается и в других приложениях. Этим и оправдано применение моделей знаний.

На основе базы данных Kaggle, рассмотренной в данной курсовой работе, построить модели прогнозирования не с помощью АСК-анализа и реализующей его системы «Эйдос», а с применением других математических методов и реализующих их программных систем, то можно сопоставимо сравнить их качество.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Луценко Е.В. Методика использования репозитория UCI для оценки качества математических моделей систем искусственного интеллекта / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: http://lc.kubagro.ru/My_training_schedule.doc

КубГАУ, 2003. – №02(002). С. 120 – 145. – IDA [article ID]: 0020302012. – Режим доступа: <http://ej.kubagro.ru/2003/02/pdf/12.pdf>, 1,625 у.п.л.

2. Луценко Е.В. АСК-анализ, моделирование и идентификация живых существ на основе их фенотипических признаков / Е.В. Луценко, Ю.Н. Пенкина // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2014. – №06(100). С. 1346 – 1395. – IDA [article ID]: 1001406090. – Режим доступа: <http://ej.kubagro.ru/2014/06/pdf/90.pdf>, 3,125 у.п.л.

3. Луценко Е.В. Теоретические основы, технология и инструментарий автоматизированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: Куб- ГАУ, 2013. – №04(088). С. 340 – 359. – IDA [article ID]: 0881304022. – Режим доступа: <http://ej.kubagro.ru/2013/04/pdf/22.pdf>, 1,25 у.п.л.

4. 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.

5. [Электронный ресурс]. Статья "baza dannix sotrudnikov": <http://allexcel.ru/gotovye-tablitsy-excel-besplatno>, свободный. - Загл. с экрана. Яз. англ.

6. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: <http://lc.kubagro.ru/>, свободный. - Загл. с экрана. Яз. рус.

7. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее время / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(054). С. 48 – 77. – Шифр Информрегистра: 0420900012\0110, IDA [article ID]: 0540910004. – Режим доступа: <http://ej.kubagro.ru/2009/10/pdf/04.pdf>, 1,875 у.п.л.

8. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос- Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2012. – №09(083). С. 328 – 356. – IDA [article ID]: 0831209025. – Режим доступа: <http://ej.kubagro.ru/2012/09/pdf/25.pdf>, 1,812 у.п.л.

9. Луценко Е.В., Боровко А.Ю. ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА И КЛАССОВ СОЛНЕЧНЫХ ВСПЫШЕК НА ОСНОВЕ ИХ ПРЕДЫСТОРИИ ПО ДАННЫМ РЕПОЗИТОРИЯ UCI С ПРИМЕНЕНИЕМ АСК- АНАЛИЗА И ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ «ЭЙДОС» // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – IDA [article ID]: 0831209025. – Режим доступа: <http://ej.kubagro.ru/2014/10/pdf/99.pdf>