

Федеральное государственное бюджетное образовательное
учреждение высшего образования
**КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
ИМ. И.Т. ТРУБИЛИНА**

Факультет заочного обучения

КУРСОВАЯ РАБОТА

по дисциплине: Системы представления и приобретения знаний
на тему: «Анализ детской смертности от заболевания корью
по регионам мира с 2000 по 2017 года»

выполнил студент группы ИТз 1841 Алиев Марат Мурадович

Руководитель проекта: Луценко Евгений Вениаминович

Защищен _____
(дата)

Оценка _____

Краснодар
2019 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ.....	4
1.1. Описание решения.....	4
1.2. Преобразование исходных данных из CSV - формата в файл исходных данных MS Excel	4
1.3. Синтез и верификация статистических и интеллектуальных моделей ...	9
1.4. Виды моделей системы «Эйдос»	11
1.5. Результаты верификации моделей.....	13
2. РЕШЕНИЕ ЗАДАЧ ИДЕНТИФИКАЦИИ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	17
2.1. Решение задачи	17
2.2. Когнитивные функции	20
ЗАКЛЮЧЕНИЕ.....	27
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	28

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 проведен анализ детской смертности от заболевания корью по регионам мира с 2000 по 2017 года.

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 база данных ВОЗ для систем искусственного интеллекта репозитория ВОЗ.

В данной работе использована база данных «Number of child deaths by region Measles» из банка исходных данных Global Health Observatory data repository.

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 CSV - формата в файл исходных данных MS Excel

Из электронного ресурса баз данных Global Health Observatory data repository возьмем базу данных – «Number of child deaths by region Measles», которую оставим без изменений.

Общее описание задачи:

1. Год
2. Возраст
3. Количество умерших
4. Регион
5. Заболевание
6. Тип ситуации

Столбцы 2-6 описательные шкалы.

Столбец 7 является классификационной шкалой.

Обучающая выборка:

Таблица 1 – Number of child deaths by region Measles.xls

Номер	Год	Возраст	Количество умерших	Регион	Заболевание	Тип ситуации
1	2017	0-4 years	43382	Африка	Корь	Чрезвычайная
2	2016	0-4 years	91225	Африка	Корь	Чрезвычайная
3	2015	0-4 years	71849	Африка	Корь	Чрезвычайная
4	2014	0-4 years	58900	Африка	Корь	Чрезвычайная
5	2013	0-4 years	97993	Африка	Корь	Чрезвычайная
6	2012	0-4 years	62861	Африка	Корь	Чрезвычайная
7	2011	0-4 years	92407	Африка	Корь	Чрезвычайная
8	2010	0-4 years	59510	Африка	Корь	Чрезвычайная
9	2009	0-4 years	63308	Африка	Корь	Чрезвычайная
10	2008	0-4 years	73558	Африка	Корь	Чрезвычайная
11	2007	0-4 years	45833	Африка	Корь	Чрезвычайная
12	2006	0-4 years	74368	Африка	Корь	Чрезвычайная
13	2005	0-4 years	224419	Африка	Корь	Чрезвычайная
14	2004	0-4 years	131746	Африка	Корь	Чрезвычайная
15	2003	0-4 years	310264	Африка	Корь	Чрезвычайная
16	2002	0-4 years	280527	Африка	Корь	Чрезвычайная
17	2001	0-4 years	290839	Африка	Корь	Чрезвычайная
18	2000	0-4 years	320126	Африка	Корь	Чрезвычайная
19	2017	0-4 years	0	Южная и Северная Америка	Корь	Благоприятная
20	2016	0-4 years	0	Южная и Северная Америка	Корь	Благоприятная
21	2015	0-4 years	0	Южная и Северная Америка	Корь	Благоприятная
22	2014	0-4 years	0	Южная и Северная Америка	Корь	Благоприятная

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-файл в xls-файл, для чего выполним следующие операции.

Скопируем получившуюся таблицу из MS Word в MS Excel и запишем ее с именем: Inp_data.xls в папку: c:\Aidos-X\AID_DATA\Inp_data\. В файле Inp_data.xls добавим пустую колонку на позиции «А» и автоматически пронумеруем все строки. 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
- XLSX- MS Excel-2007(2010)
- DBF - DBASE IV (DBF/NTX)
- CSV - Comma-Separated Values

Стандарт XLS-файла

Стандарт DBF-файла

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 1, 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" – 7, "Конечный столбец классификационных шкал" – 7;
- "Задайте диапазон столбцов описательных шкал": "Начальный столбец описательных шкал" – 2, "Конечный столбец описательных шкал" – 6;
- "Задание параметров формирования сценариев или способа интерпретации текстовых полей": "Не применять сценарный метод АСК- анализа и спец.интерпретацию ТХТ-полей".

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".

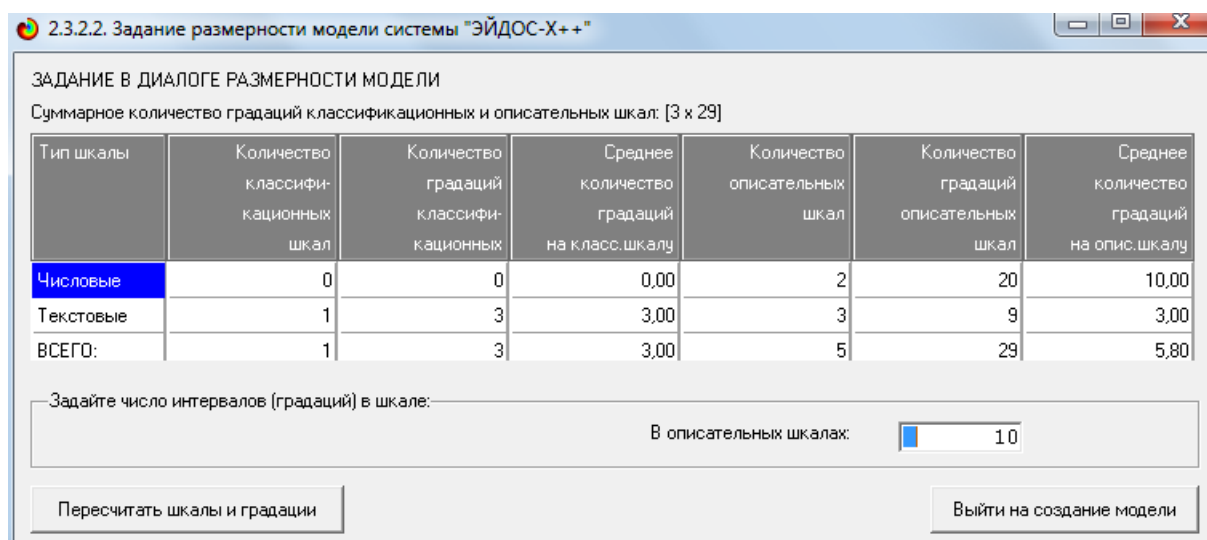


Рисунок 2. Задание размерности модели системы "Эйдос"

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".

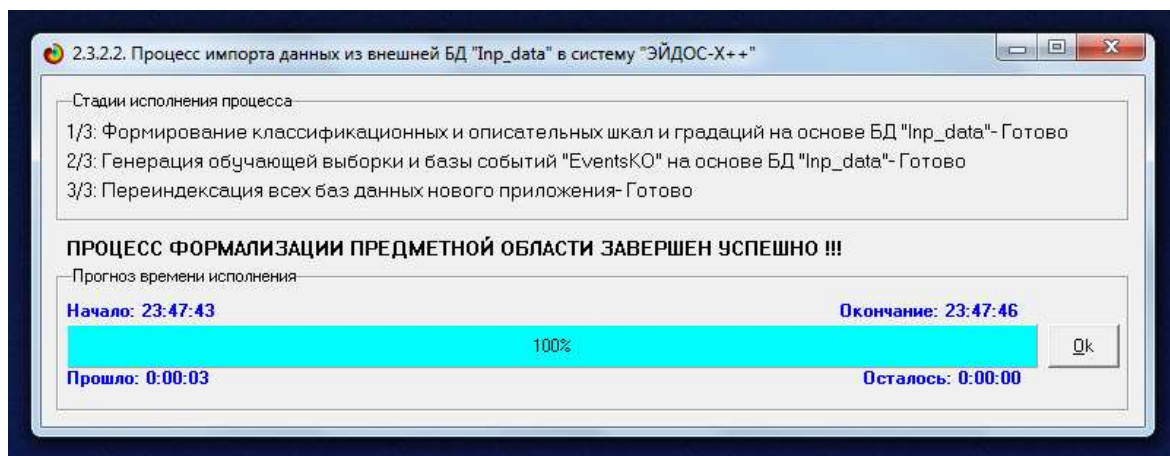


Рисунок 3. Процесс импорта данных из внешней БД "Inp_data.xls" в систему "Эйдос"

В 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).

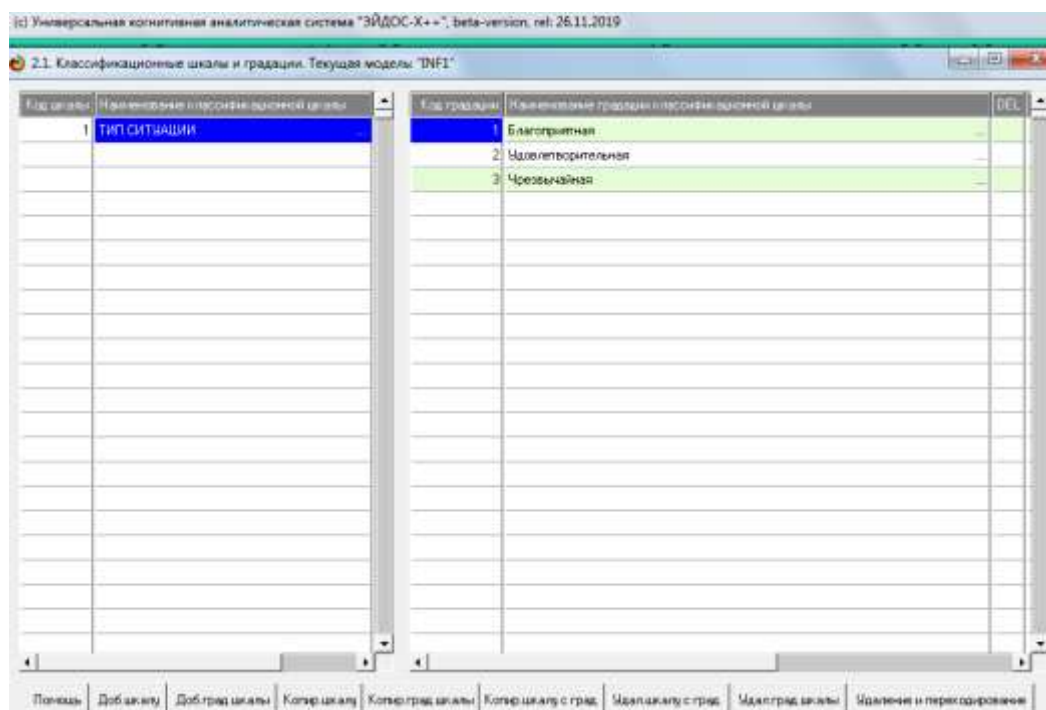


Рисунок 4. Классификационные шкалы и градации

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

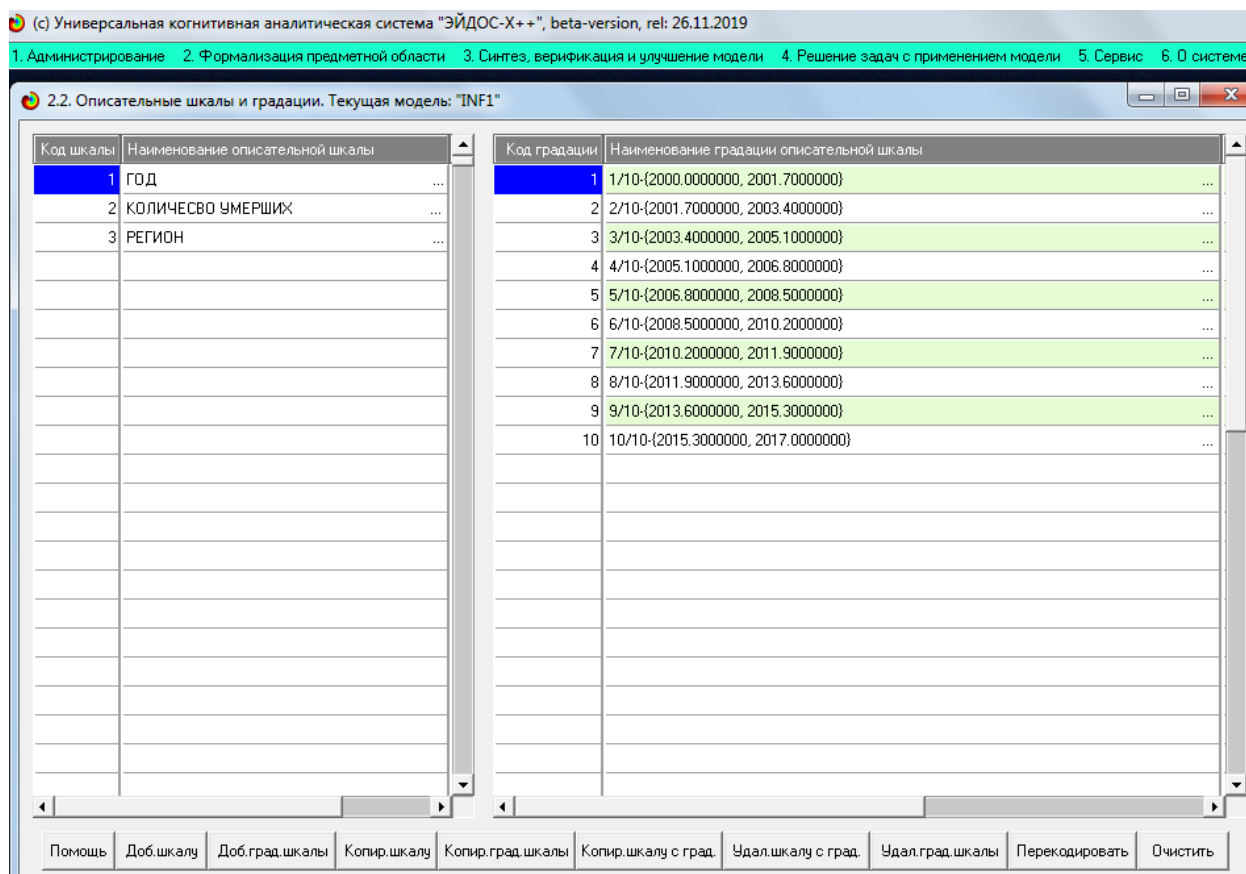


Рисунок 5. Описательные шкалы и градации

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

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 6).

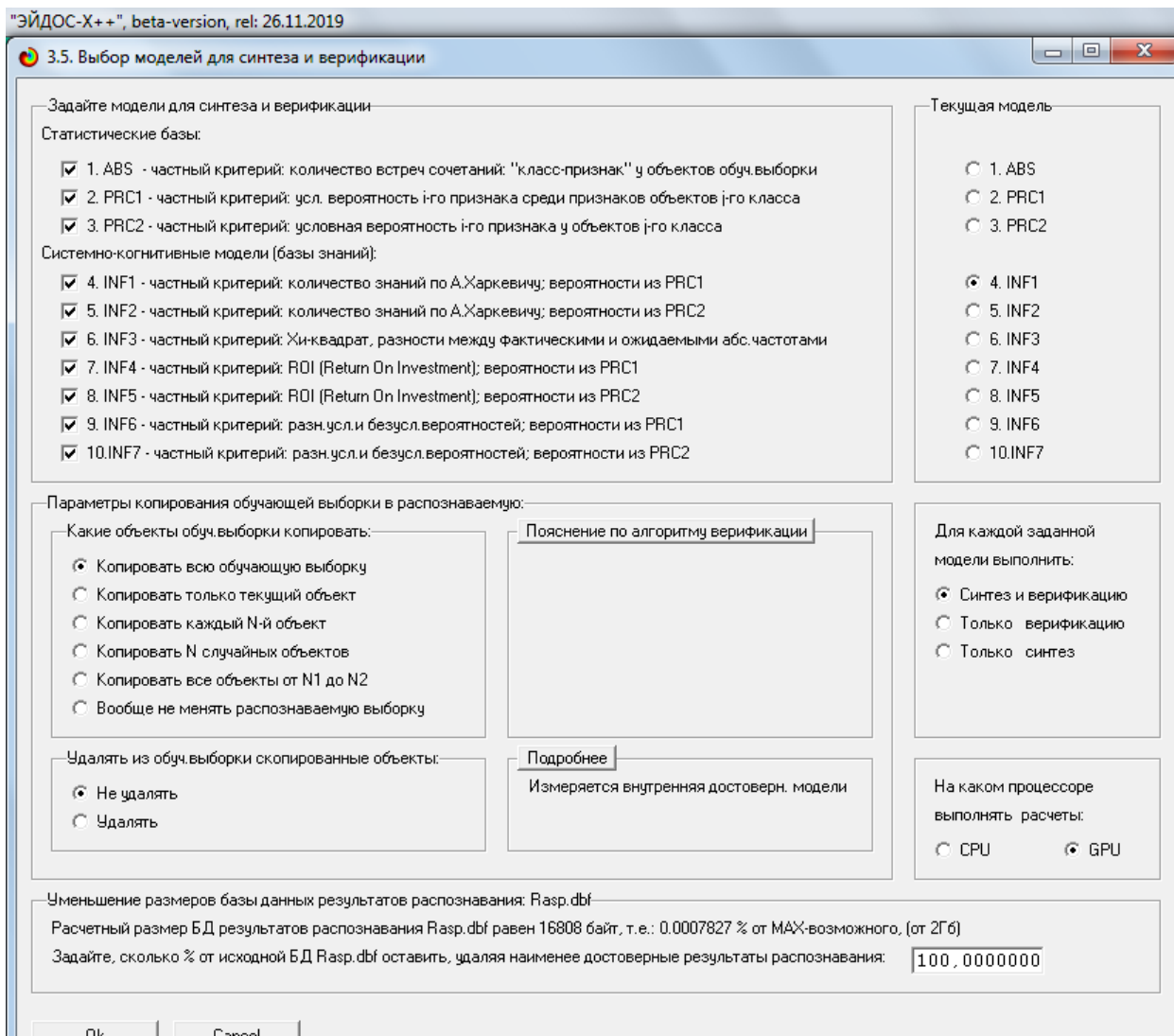


Рисунок 6. Выбор моделей для синтеза и верификации, а также текущей модели

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 6. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 7.

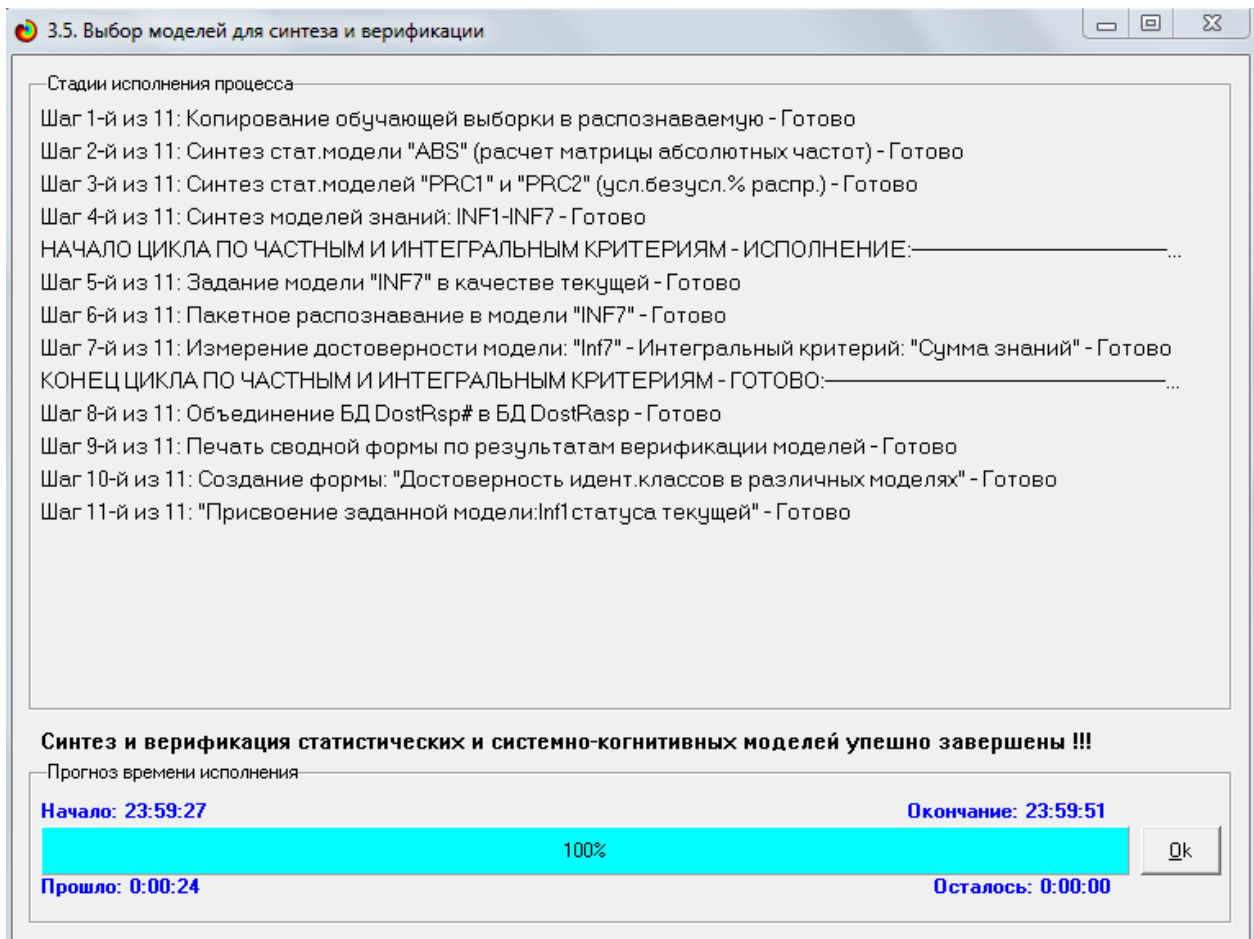


Рисунок 7. Синтез и верификация статистических моделей и моделей знаний

Интересно заметить (см. рисунок 7), что синтез и верификация всех 10 моделей на данной задаче заняли 24 секунды, несмотря на то, что обрабатывалось 127 строк. В результате выполнения режима 3.5 созданы все модели, со всеми частными критериями, перечисленные на рисунке 6, 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).

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

Код признака	Наименование описательной шкалы и градации	1. ТИП СИТУАЦИИ БЛАГОПРИЯТ...	2. ТИП СИТУАЦИИ УДОВОЛЕТВОР...	3. ТИП СИТУАЦИИ ЧРЕЗВЫЧАЙН...	Сумма	Среднее	Средн. квадрат. откл.
1	ГОД-1/10-(2000.0000000, 2001.7000000)	2	6	6	14	4.67	2.31
2	ГОД-2/10-(2001.7000000, 2003.4000000)	2	6	6	14	4.67	2.31
3	ГОД-3/10-(2003.4000000, 2005.1000000)	2	6	6	14	4.67	2.31
4	ГОД-4/10-(2005.1000000, 2006.8000000)	1	3	3	7	2.33	1.15
5	ГОД-5/10-(2006.8000000, 2008.5000000)	2	6	6	14	4.67	2.31
6	ГОД-6/10-(2008.5000000, 2010.2000000)	2	6	6	14	4.67	2.31
7	ГОД-7/10-(2010.2000000, 2011.9000000)	1	3	3	7	2.33	1.15
8	ГОД-8/10-(2011.9000000, 2013.6000000)	2	6	6	14	4.67	2.31
9	ГОД-9/10-(2013.6000000, 2015.3000000)	2	6	6	14	4.67	2.31
10	ГОД-10/10-(2015.3000000, 2017.0000000)	2	6	6	14	4.67	2.31
11	КОЛИЧЕСТВО УМЕРШИХ-1/10-(1.0000000, 48294.70...	5	36	26	67	22.33	15.82
12	КОЛИЧЕСТВО УМЕРШИХ-2/10-(48294.7000000, 9658...		1	14	15	5.00	7.81
13	КОЛИЧЕСТВО УМЕРШИХ-3/10-(96588.4000000, 1448...		6	9	15	5.00	4.58
14	КОЛИЧЕСТВО УМЕРШИХ-4/10-(144882.1000000, 193...		4		4	1.33	2.31
15	КОЛИЧЕСТВО УМЕРШИХ-5/10-(193175.8000000, 241...		1	1	2	0.67	0.58
16	КОЛИЧЕСТВО УМЕРШИХ-6/10-(241469.5000000, 289...			1	1	0.33	0.58
17	КОЛИЧЕСТВО УМЕРШИХ-7/10-(289763.2000000, 338...		1	3	4	1.33	1.53
18	КОЛИЧЕСТВО УМЕРШИХ-8/10-(338056.9000000, 386...		1		1	0.33	0.58
19	КОЛИЧЕСТВО УМЕРШИХ-9/10-(386350.6000000, 434...						
20	КОЛИЧЕСТВО УМЕРШИХ-10/10-(434644.3000000, 48...		4		4	1.33	2.31
21	РЕГИОН-Австралия и океания		18		18	6.00	10.39
22	РЕГИОН-Африка			18	18	6.00	10.39
23	РЕГИОН-Ближний восток			18	18	6.00	10.39
24	РЕГИОН-Во всем мире		18		18	6.00	10.39

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

Код признака	Наименование описательной шкалы и градации	1. ТИП СИТУАЦИИ БЛАГОПРИЯТ...	2. ТИП СИТУАЦИИ УДОВОЛЕТВОР...	3. ТИП СИТУАЦИИ ЧРЕЗВЫЧАЙН...	Сумма	Среднее	Средн. квадрат. откл.
1	ГОД-1/10-(2000.0000000, 2001.7000000)	0.055	-0.009	-0.009	0.046	0.015	0.043
2	ГОД-2/10-(2001.7000000, 2003.4000000)	0.065	-0.009	-0.009	0.046	0.015	0.043
3	ГОД-3/10-(2003.4000000, 2005.1000000)	0.065	-0.009	-0.009	0.046	0.015	0.043
4	ГОД-4/10-(2005.1000000, 2006.8000000)	0.055	-0.009	-0.009	0.046	0.015	0.043
5	ГОД-5/10-(2006.8000000, 2008.5000000)	0.065	-0.009	-0.009	0.046	0.015	0.043
6	ГОД-6/10-(2008.5000000, 2010.2000000)	0.065	-0.009	-0.009	0.046	0.015	0.043
7	ГОД-7/10-(2010.2000000, 2011.9000000)	0.055	-0.009	-0.009	0.046	0.015	0.043
8	ГОД-8/10-(2011.9000000, 2013.6000000)	0.055	-0.009	-0.009	0.046	0.015	0.043
9	ГОД-9/10-(2013.6000000, 2015.3000000)	0.065	-0.009	-0.009	0.046	0.015	0.043
10	ГОД-10/10-(2015.3000000, 2017.0000000)	0.065	-0.009	-0.009	0.046	0.015	0.043
11	КОЛИЧЕСТВО УМЕРШИХ-1/10-(1.0000000, 4829...	-0.110	0.051	-0.036	-0.095	-0.032	0.091
12	КОЛИЧЕСТВО УМЕРШИХ-2/10-(48294.7000000, 9658...		-0.509	0.280	-0.310	-0.103	0.366
13	КОЛИЧЕСТВО УМЕРШИХ-3/10-(96588.4000000, 1448...		-0.028	0.081	0.053	0.018	0.057
14	КОЛИЧЕСТВО УМЕРШИХ-4/10-(144882.1000000, 193...		0.218		0.218	0.073	0.126
15	КОЛИЧЕСТВО УМЕРШИХ-5/10-(193175.8000000, 241...		0.032	0.022	0.054	0.021	0.018
16	КОЛИЧЕСТВО УМЕРШИХ-6/10-(241469.5000000, 289...			0.218	0.218	0.073	0.126
17	КОЛИЧЕСТВО УМЕРШИХ-7/10-(289763.2000000, 338...		-0.154	0.141	-0.013	-0.004	0.148
18	КОЛИЧЕСТВО УМЕРШИХ-8/10-(338056.9000000, 386...		0.218		0.218	0.073	0.126
19	КОЛИЧЕСТВО УМЕРШИХ-9/10-(386350.6000000, 434...						
20	КОЛИЧЕСТВО УМЕРШИХ-10/10-(434644.3000000, 48...		0.218		0.218	0.073	0.126
21	РЕГИОН-Австралия и океания		0.218		0.218	0.073	0.126
22	РЕГИОН-Африка			0.218	0.218	0.073	0.126
23	РЕГИОН-Ближний восток			0.218	0.218	0.073	0.126
24	РЕГИОН-Во всем мире		0.218		0.218	0.073	0.126

Таблица 4 – Матрица знаний (фрагмент)

5.5. Модель: "7. INF4 - частный критерий: ROI (Return On Investment); вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. ТИП СИТУАЦИИ БЛАГОПРИЯТ...	2. ТИП СИТУАЦИИ УДОВЛЕТВОРИ...	3. ТИП СИТУАЦИИ ЧРЕЗВЫЧАЙНАЯ	Сумма	Среднее	Средн. квадр. откл.
1	ГОД-1/10-(2000.000000, 2001.7000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
2	ГОД-2/10-(2001.7000000, 2003.4000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
3	ГОД-3/10-(2003.4000000, 2005.1000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
4	ГОД-4/10-(2005.1000000, 2006.8000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
5	ГОД-5/10-(2006.8000000, 2008.5000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
6	ГОД-6/10-(2008.5000000, 2010.2000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
7	ГОД-7/10-(2010.2000000, 2011.9000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
8	ГОД-8/10-(2011.9000000, 2013.6000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
9	ГОД-9/10-(2013.6000000, 2015.3000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
10	ГОД-10/10-(2015.3000000, 2017.0000000) ...	0.272	-0.034	-0.034	0.203	0.068	0.177
11	КОЛИЧЕСТВО УМЕРШИХ-1/10-(1.0000000, 4829...	-0.336	0.211	-0.126	-0.251	-0.084	0.276
12	КОЛИЧЕСТВО УМЕРШИХ-2/10-(4829.7000000, ...		-0.850	1.103	0.253	0.084	0.979
13	КОЛИЧЕСТВО УМЕРШИХ-3/10-(96588.4000000, ...		-0.099	0.352	0.253	0.084	0.237
14	КОЛИЧЕСТВО УМЕРШИХ-4/10-(144882.1000000, ...		1.253		1.253	0.418	0.723
15	КОЛИЧЕСТВО УМЕРШИХ-5/10-(193175.8000000, ...		0.127	0.127	0.253	0.084	0.073
16	КОЛИЧЕСТВО УМЕРШИХ-6/10-(241469.5000000, ...			1.253	1.253	0.418	0.723
17	КОЛИЧЕСТВО УМЕРШИХ-7/10-(289763.2000000, ...		-0.437	0.690	0.253	0.084	0.568
18	КОЛИЧЕСТВО УМЕРШИХ-8/10-(338056.9000000, ...		1.253		1.253	0.418	0.723
19	КОЛИЧЕСТВО УМЕРШИХ-9/10-(386350.6000000, ...						
20	КОЛИЧЕСТВО УМЕРШИХ-10/10-(434644.3000000, ...		1.253		1.253	0.418	0.723
21	РЕГИОН-Австралия и океания		1.253		1.253	0.418	0.723
22	РЕГИОН-Африка			1.253	1.253	0.418	0.723
23	РЕГИОН-Ближний восток			1.253	1.253	0.418	0.723
24	РЕГИОН-Во всем мире		1.253		1.253	0.418	0.723

1.5. Результаты верификации моделей

Результаты верификации (оценки достоверности) моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями приведены на рисунке 8.

4.1.3.6. Обобщенная форма по достов. моделей при различии крит. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Число признаков объектов	Число признаков по критерию (TP)	Число признаков отрицательных (TN)	Число признаков по критерию (FP)	Число признаков отрицательных (FN)	Точность модели	Плотность признаков	Функция Вил Руберова	Среднее значение качества модели
1. INF1 - частный критерий: количество водных объектов, % из...	Корреляция объектов с обб...	126	126	117	135		0.483	1.000	0.651	84
2. INF2 - частный критерий: количество водных объектов, % из...	Средн. аби. частот по призна...	126	126	252			0.333	1.000	0.580	87
3. PRC1 - частный критерий: усл. вероятность по признаку сред...	Корреляция усл. частот с о...	126	126	117	135		0.483	1.000	0.651	84
4. PRC2 - частный критерий: усл. вероятность по признаку сред...	Средн. усл. частот по призна...	126	126	252			0.333	1.000	0.580	89
5. INF1 - частный критерий: количество знаний по А/Харковск. в...	Семантический резонанс: зна...	126	124	251	1	2	0.992	0.994	0.988	42
6. INF2 - частный критерий: количество знаний по А/Харковск. в...	Средн. знаний	126	125	165	87	1	0.590	0.992	0.748	61
7. INF3 - частный критерий: количество знаний по А/Харковск. в...	Семантический резонанс: зна...	126	124	251	1	2	0.992	0.994	0.988	44
8. INF4 - частный критерий: количество знаний по А/Харковск. в...	Средн. знаний	126	125	211	41	1	0.753	0.992	0.856	77
9. INF5 - частный критерий: количество знаний по А/Харковск. в...	Семантический резонанс: зна...	126	126	252			1.000	1.000	1.000	62
10. INF6 - частный критерий: количество знаний по А/Харковск. в...	Средн. знаний	126	126	252			1.000	1.000	1.000	93
7. INF4 - частный критерий: ROI (Return On Investment), вероятн...	Семантический резонанс: зна...	126	125	251	1	1	0.992	0.992	0.992	45
8. INF5 - частный критерий: ROI (Return On Investment), вероятн...	Средн. знаний	126	126	165	87	1	0.592	1.000	0.743	38
9. INF6 - частный критерий: ROI (Return On Investment), вероятн...	Семантический резонанс: зна...	126	125	251	1	1	0.992	0.992	0.992	46
10. INF7 - частный критерий: ROI (Return On Investment), вероятн...	Средн. знаний	126	126	211	41		0.754	1.000	0.868	48
11. INF8 - частный критерий: разн. усл. и без усл. вероятностей, вер...	Семантический резонанс: зна...	126	126	228	32		0.797	1.000	0.887	58
12. INF9 - частный критерий: разн. усл. и без усл. вероятностей, вер...	Средн. знаний	126	126	164	86		0.594	1.000	0.746	37
13. INF10 - частный критерий: разн. усл. и без усл. вероятностей, ве...	Семантический резонанс: зна...	126	126	228	32		0.797	1.000	0.887	63
14. INF11 - частный критерий: разн. усл. и без усл. вероятностей, ве...	Средн. знаний	126	126	211	41		0.754	1.000	0.868	47

Рисунок 8. Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели INF3 при интегральном критерии «Семантический резонанс знаний». При этом точность модели (F-мера Ван Ризбергена) составляет 1, а точность модели (L1-мера профессора Луценко 1) Таким образом, уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 100%. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется F-мера Ван Ризбергена и L-мера, представляющая собой ее нечеткое мультиклассовое обобщение, предложенное профессором Е.В.Луценко (рисунок 9).



Рисунок 9. 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 не идентифицированных ситуаций в наиболее достоверной модели INF3.

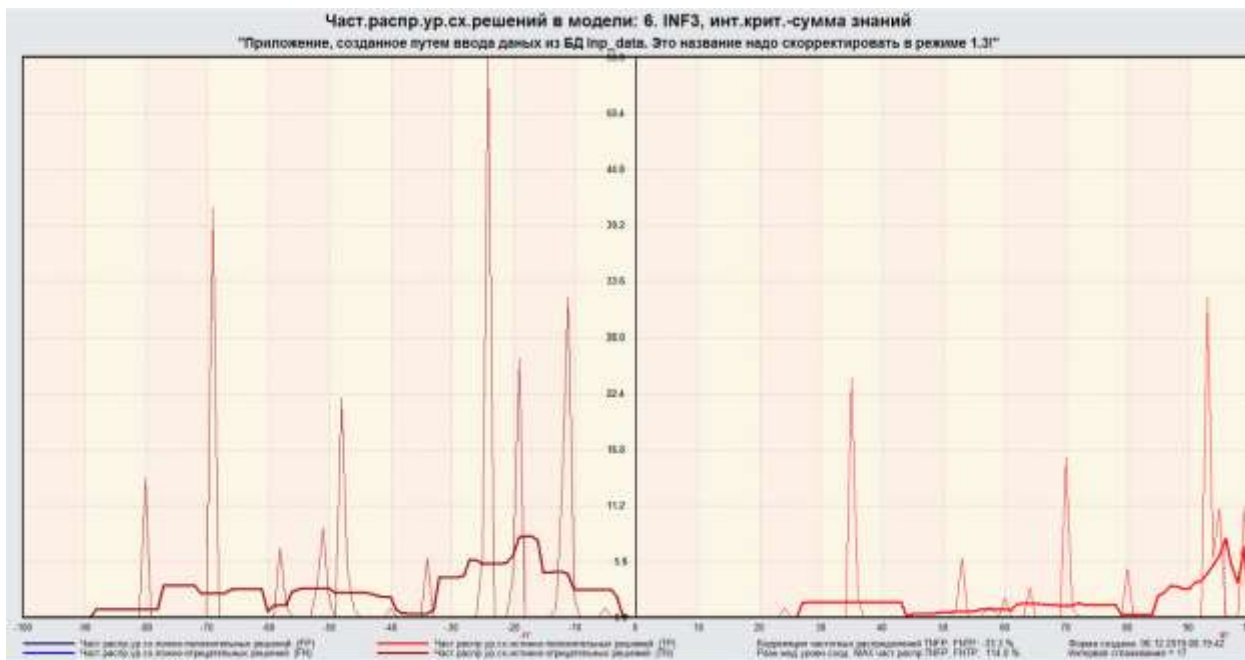


Рисунок 10. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF3

Из рисунка 10 видно, что:

- наиболее достоверная модель INF3 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 8);
- 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 15% уровня сходства. Для очень больших значений уровней сходства-различия (более 90%) также различие между верно и ошибочно идентифицированными и не идентифицированными ситуациями практически отсутствует.

Любые данные о наблюдениях можно считать суммой истинного значения и шума, причем ни первое, ни второе неизвестны.

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 11) и проведем пакетное распознавание в режиме 4.1.2. (рисунок 12).

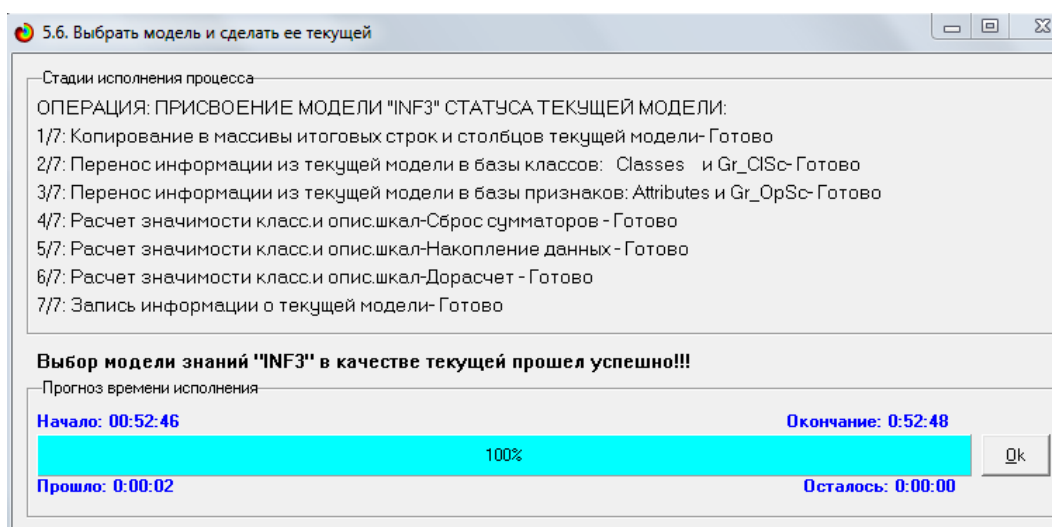


Рисунок 11. Экранные формы режима задания модели в качестве текущей

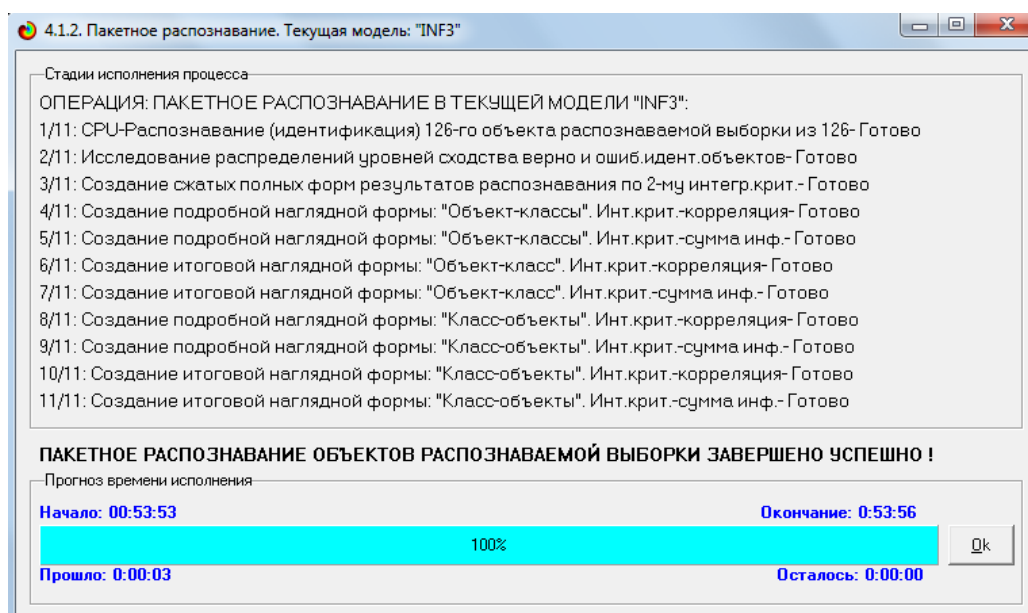


Рисунок 12. Экранная форма режима пакетного распознавания в текущей модели

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 13 и 14 приведены примеры прогнозов высокой и низкой достоверности частоты и классов ирисов в наиболее достоверной модели INF3 на основе наблюдения предыстории их развития:

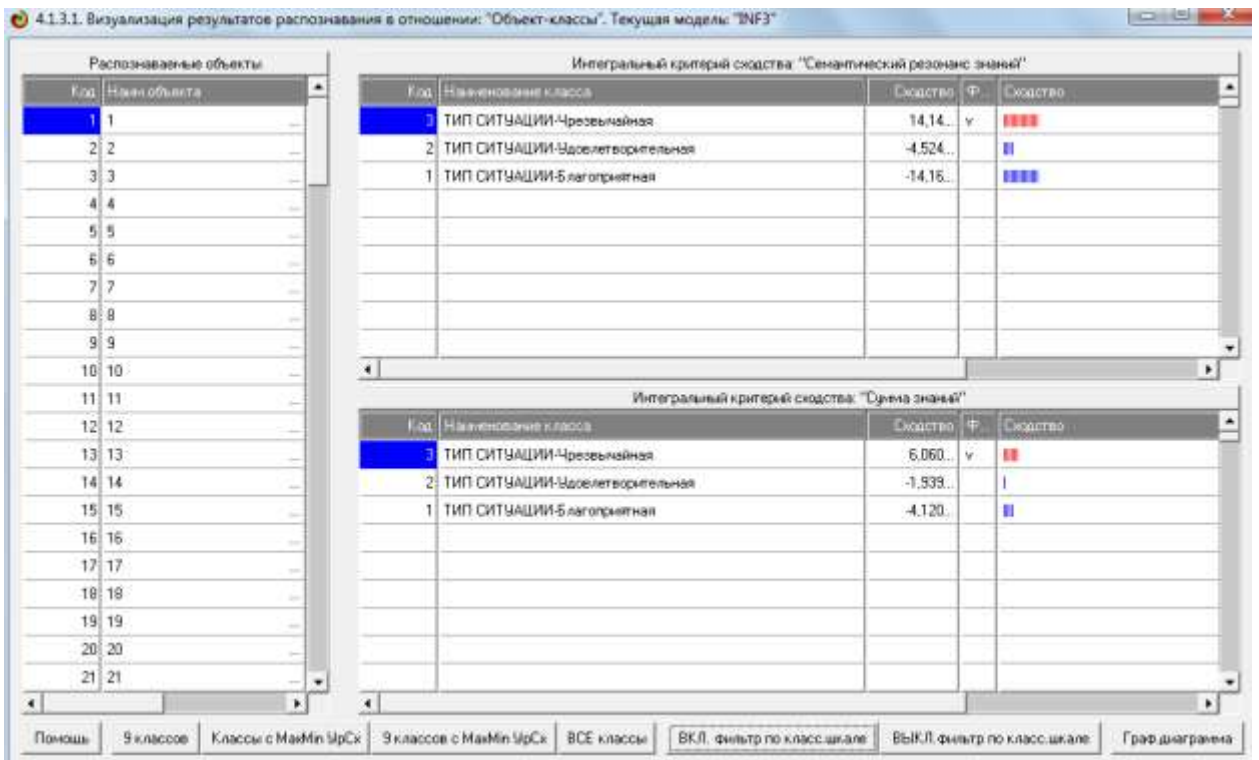


Рисунок 13. Пример идентификации классов в модели INF3

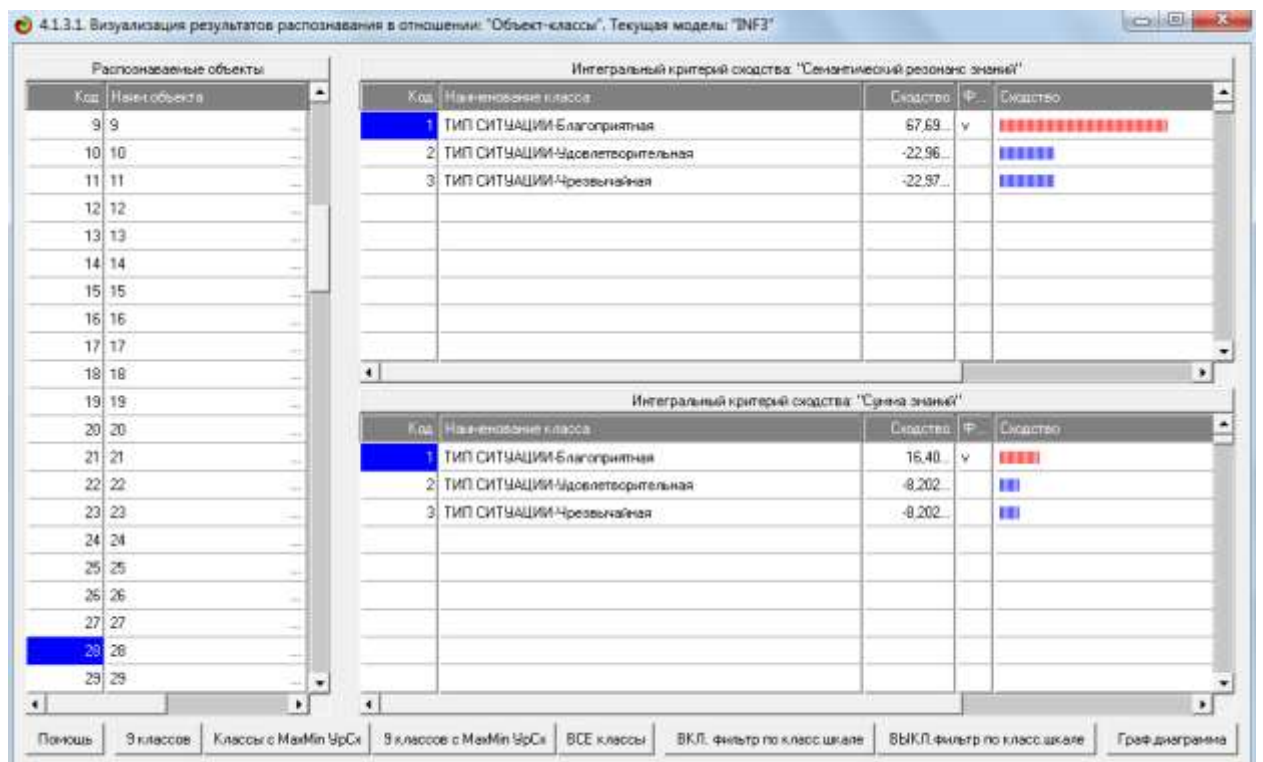


Рисунок 14. Пример идентификации классов в модели INF3

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 15)

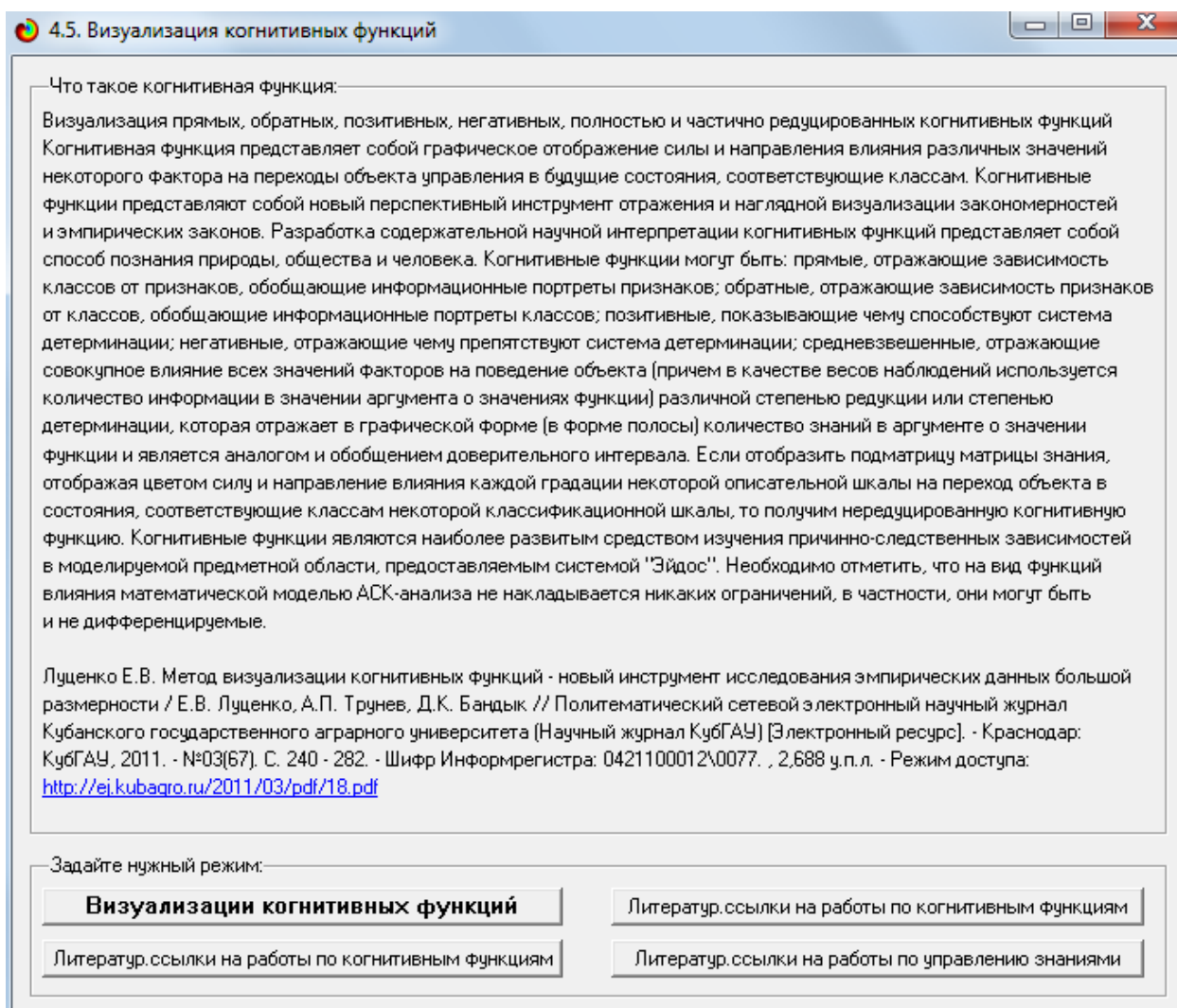


Рисунок 15. Экранная форма режима визуализации когнитивных функций

В когнитивных функциях количество информации в значениях аргумента о значениях функции отображается цветом (красным максимальное, синим минимальное), линией соединены значения функции, о которых в значении аргумента содержится максимальное количество информации, ширина линии (аналог доверительного интервала) отражает степень неопределенности значения функции, которое тем ниже, чем больше информации о нем в значении функции (рис. 16–17):

На рисунке 16 представлена когнитивная функция, показывающая, что в меньшей широте и долготе преобладают регионы по чрезвычайные ситуации, связанных с местом проживания детей. В средней широте и долготе, связанные по удовлетворительной ситуации, а в большей долготе регионы, связанные с благоприятной ситуацией.

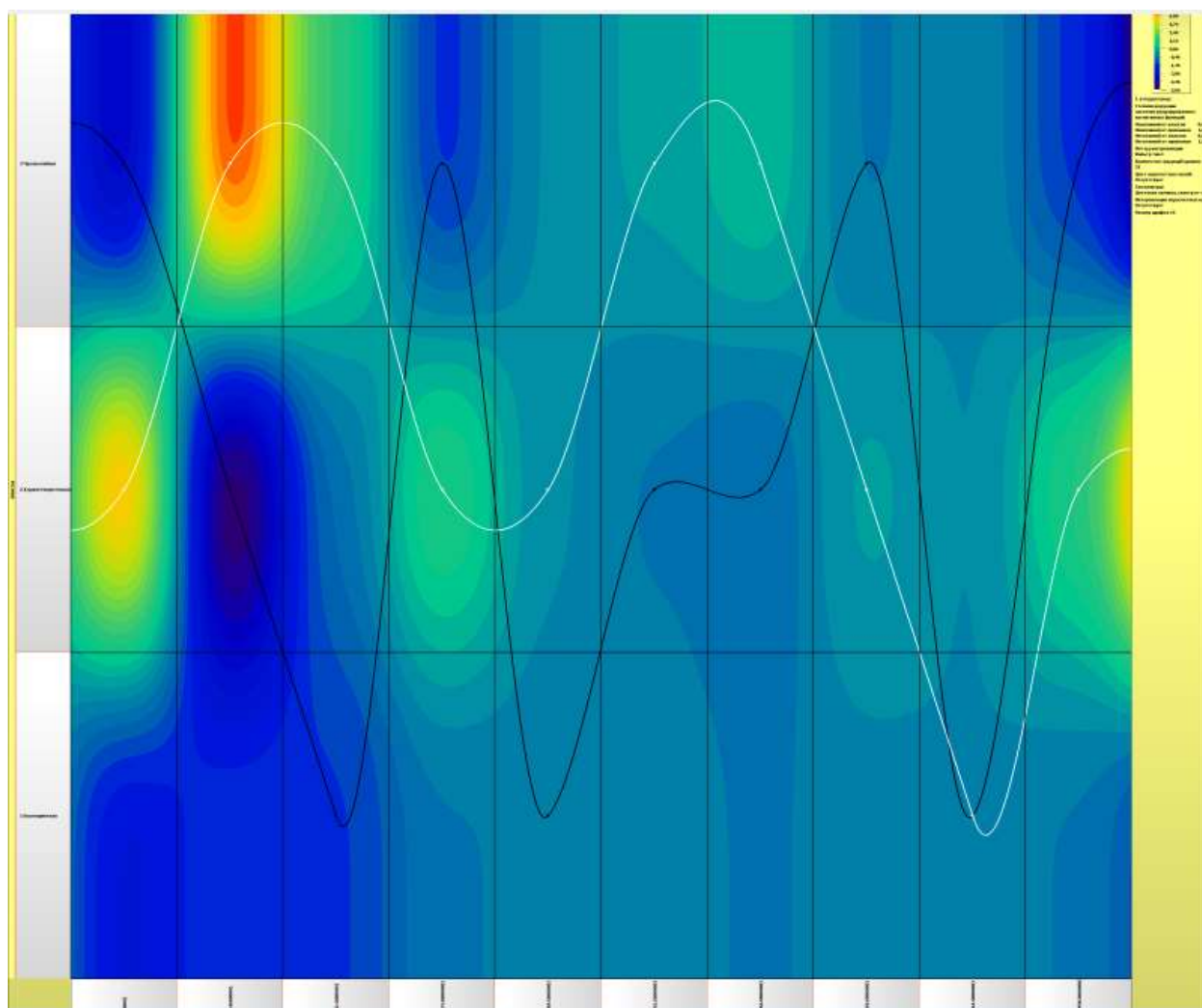


Рисунок 16. Когнитивная функция, отражающая взаимосвязь широты и долготы с типом ситуации

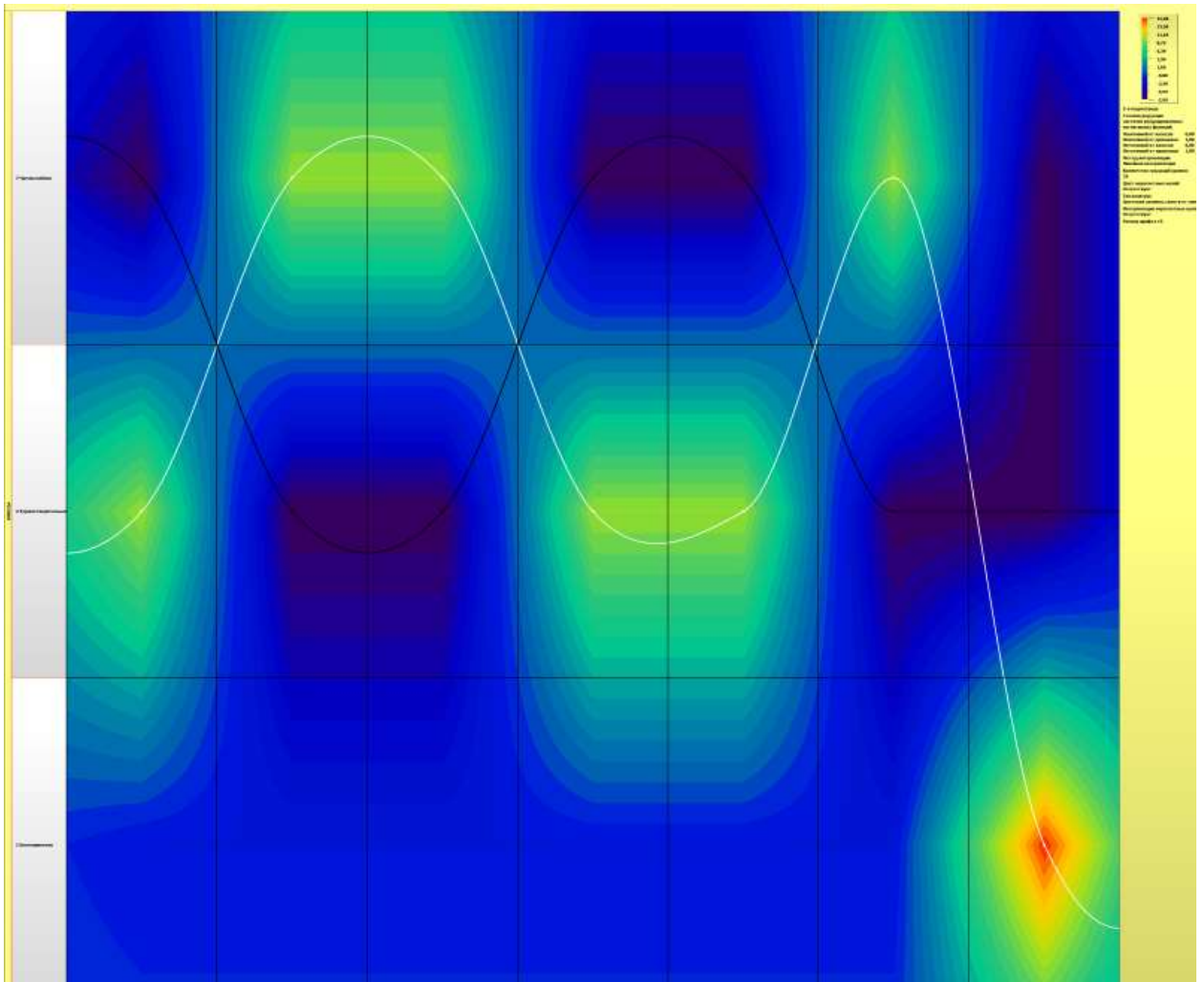


Рисунок 17. Когнитивная функция, отражающая взаимосвязь региона смертности, заболевания и количества умерших

Сходство-различие обобщенных образов различных результатов научной деятельности по характерным для них системам значений показателей. Результаты сравнения классов по системе характерных для типов чрезвычайных ситуаций приведены на рис. 18, 19:

4.2.2.2. Результаты кластерно-конструктивного анализа классов

Конструкт класса:1 "ТИП СИТУАЦИИ-Благоприятная" в модели:6 "INF3"

Код	Наименование класса	№	Код класса	Наименование класса	Сходство
1	ТИП СИТУАЦИИ-Благоприятная	1	1	ТИП СИТУАЦИИ-Благоприятная	100.000
2	ТИП СИТУАЦИИ-Удовлетворительная	2	3	ТИП СИТУАЦИИ-Чрезвычайная	-33.833
3	ТИП СИТУАЦИИ-Чрезвычайная	3	2	ТИП СИТУАЦИИ-Удовлетворительная	-34.036

Помощь Abs Prc1 Prc2 Inf1 Inf2 Inf3 Inf4 Inf5 Inf6 Inf7 График ВКЛ. фильтр по кл.шкале ВЫКЛ. фильтр по кл.шкале Параметры Показать ВСЕ

Рисунок 18. Информационные портреты классов

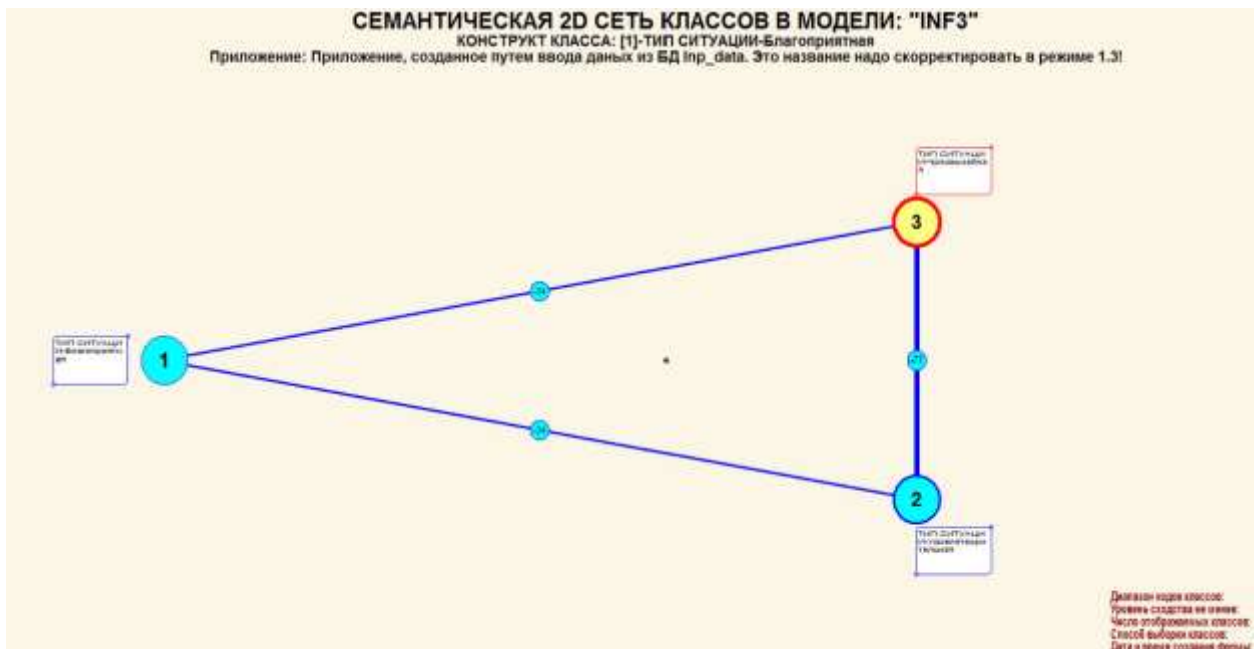


Рисунок 19. Результаты сравнения классов по системе характерных для значений смертности от кори по регионам

Из когнитивной диаграммы, приведенной на рисунке 19, мы видим, что, наиболее благоприятная ситуация заболеваемости и смертности корью находится в

странах американского контингента, а наиболее плохая в африканском регионе. Отметим также, что приведенная когнитивная диаграмма формируется системой «Эйдос» автоматически на основе созданных моделей.

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 20).

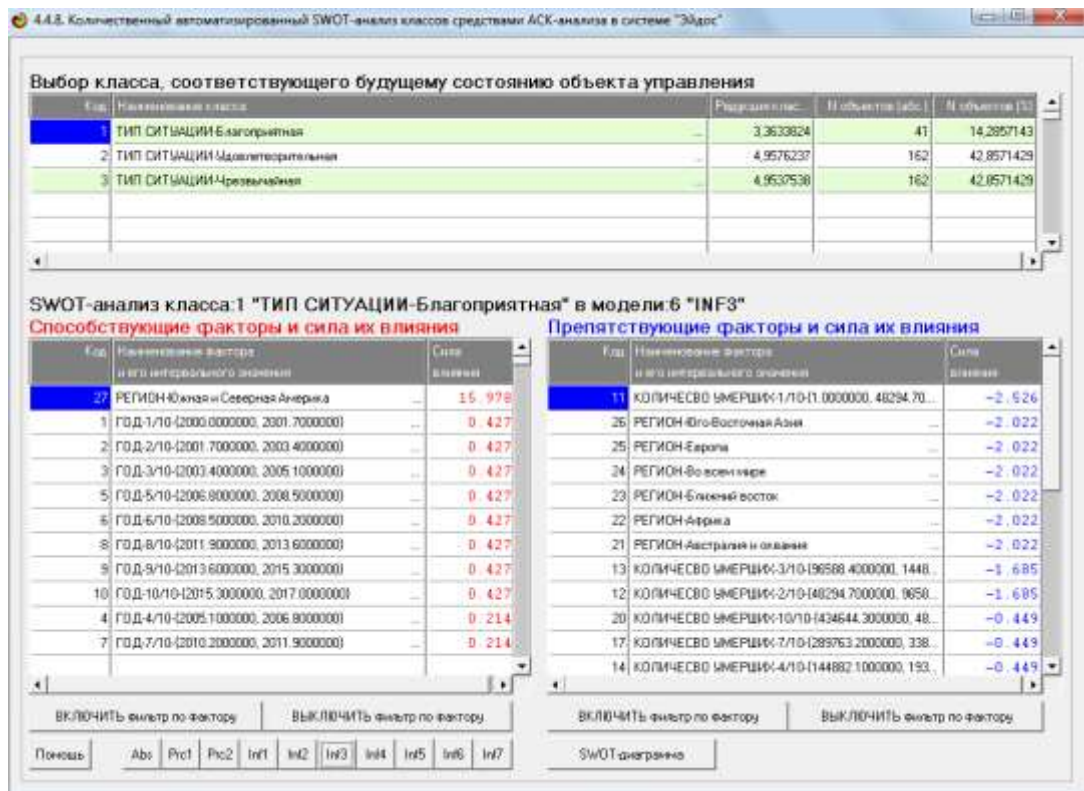


Рисунок 20. Пример SWOT-матрицы в модели INF3

На рисунке 20 приведены примеры инвертированной SWOT- матрицы и инвертированной SWOT-диаграммы в модели INF3.

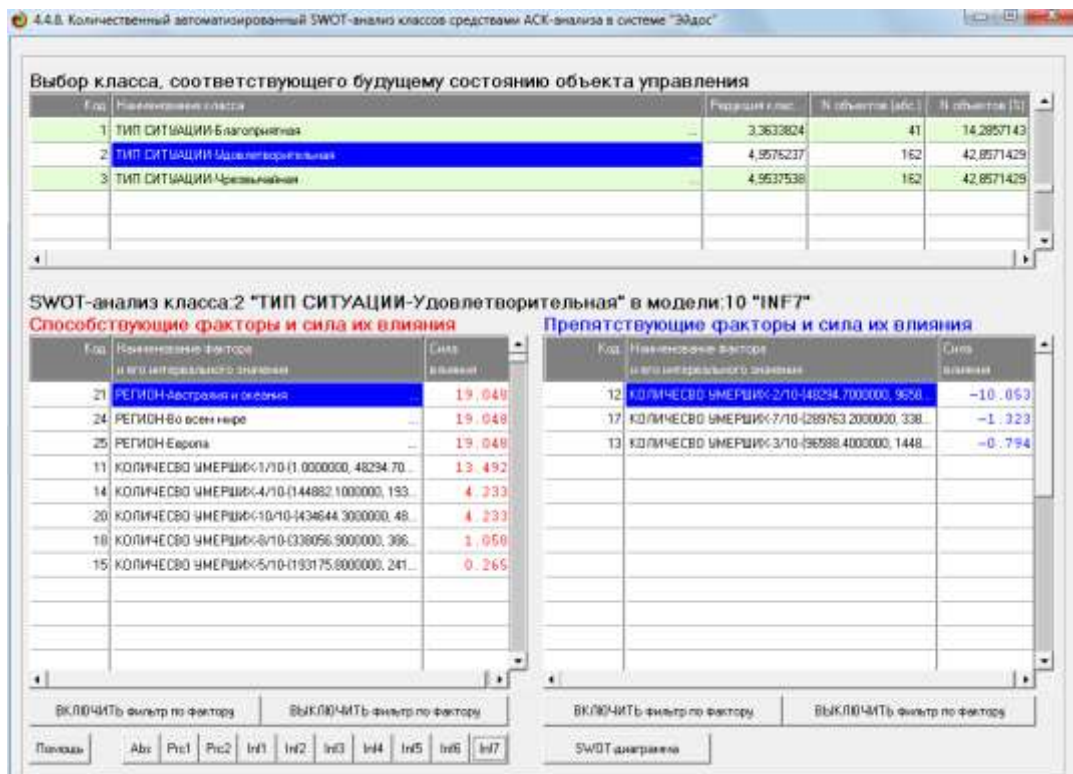


Рисунок 21. Табличная выходная форма количественного автоматизированного SWOT и PEST-анализа средствами системы «Эйдос»

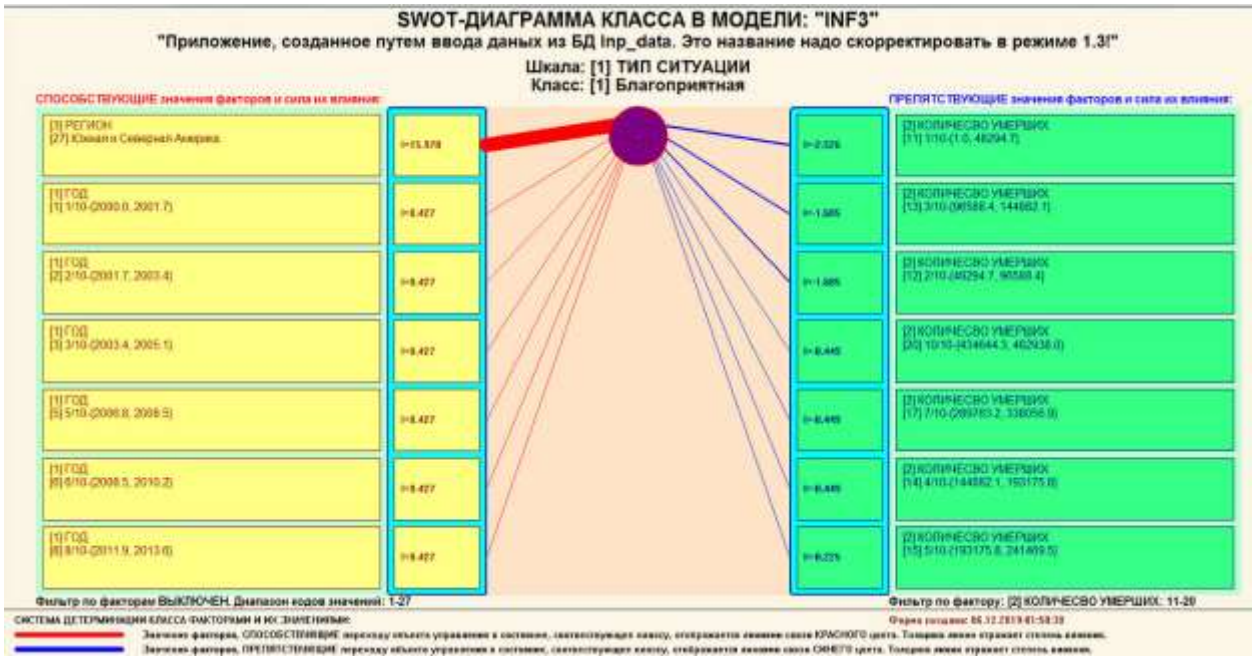


Рисунок 22. Графическая выходная форма количественного автоматизированного SWOT- и PEST-анализа средствами системы «Эйдос»

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 репозитория ВОЗ. В данной работе приводится развернутый пример использования базы данных репозитория ВОЗ для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». при интегральном критерии «Семантический резонанс знаний». При этом точность модели (F-мера Ван Ризбергена) составляет 1, а точность модели (L1-мера профессора Луценко 1) Таким образом, уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 100%, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 75%. 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 ВОЗ, рассмотренной в данной 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. 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.
2. Луценко Е.В. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос» и системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №02(116). С. 1 – 60. – IDA [article ID]: 1161602001. – Режим доступа: <http://ej.kubagro.ru/2016/02/pdf/01.pdf>, 3,75 у.п.л. 3.
3. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с.
4. [Электронный ресурс]. Статья "Emergency – 911 Calls": <https://www.kaggle.com/mchirico/montcoalert>, свободный. - Загл. с экрана. Яз. англ.
5. Сайт профессора Е.В.Луценко [Электронный ресурс]. Режим доступа: <http://lc.kubagro.ru/>, свободный. - Загл. с экрана. Яз. рус.
6. Луценко Е.В. Количественная оценка степени манипулирования индексом Хирша и его модификация, устойчивая к манипулированию / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №07(121). С. 202 – 234. –

IDA [article ID]: 1211607005. – Режим доступа: <http://ej.kubagro.ru/2016/07/pdf/05.pdf>, 2,062 у.п.л.. <http://dx.doi.org/10.21515/1990-4665-121-0057>.

7 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.

8. Луценко Е.В., Боровко А.Ю. ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА И КЛАССОВ СОЛНЕЧНЫХ ВСПЫШЕК НА ОСНОВЕ ИХ ПРЕДЫСТОРИИ ПО ДАННЫМ РЕПОЗИТОРИЯ UC1 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>