

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:

Автоматизированный системно-когнитивный анализ матчей
европейского футбола на основе данных репозитория UC1

выполнил студент группы ИТ1401 Германий Эдуард Валерьевич
Допущен 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**

Студенту: ИТ1421 группы 3 курса
Факультета прикладной информатики
Специальности: 09.03.02 Информационные системы и технологии
(шифр)

Гермонию Эдуарду Валерьевичу
(Ф.И.О.)

Тема проекта: **Автоматизированный системно-когнитивный анализ матчей
европейского футбола 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.

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

Факультет прикладной информатики

РЕЦЕНЗИЯ

на курсовой проект (работу)

Студента (ки) Германия Эдуарда Валерьевича
курса 3 очной (заочной) формы обучения
направления подготовки Информационные системы и технологии
направленность(профиль) Интеллектуальные информационные системы и технологии.
Наименование темы «Автоматизированный системно-когнитивный анализ матчей европейского футбола на основе данных репозитория UCI».

Рецензент: Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор
(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	5
3.	Уровень и корректность использования в работе различных методов исследований	4
4.	Степень комплексности работы, применения в ней знаний общепрофессиональных и специальных дисциплин	4
5.	Ясность, четкость, последовательность и обоснованность изложения	5
6.	Применение современных технологий обработки информации	4
7.	Качество оформления работы (общий уровень грамотности, стиль изложения, качество иллюстраций, соответствие требованиям по оформлению)	5
8.	Ответы на вопросы при защите	5

Достоинства работы

Актуальность темы применения современных математических моделей и программного инструментария для проведения АСК-анализа на основе базы данных

Недостатки работы

Отдельные погрешности в оформлении

Итоговая оценка при защите _____ 5(отл.) _____

Рецензент _____ Е.В. Луценко

« _____ » _____ 201 _____ г.

РЕФЕРАТ

Курсовая работа содержит: 27 страниц, 19 рисунков, 4 таблицы, 9 литературных источников.

Ключевые слова: КОГНИТИВНЫЙ АНАЛИЗ, МОДЕЛИ, ШКАЛЫ, КЛАССЫ.

В данной курсовой работе были проанализированы методы формирования обобщенных образов классов и решения задач идентификации конкретных объектов с классами, принятия решений и исследования моделируемой предметной области путем исследования ее модели.

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	7
1.3. Ввод выборки в систему Aidos-X.....	9
1.4. Синтез и верификация статистических и интеллектуальных моделей.....	12
1.5. Виды моделей системы «Aidos».....	13
1.5. Результаты верификации моделей	15
2. РЕШЕНИЕ ЗАДАЧ НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ	17
2.1. Решение задачи идентификации	17
2.2. Выбор модели и пакетное распознавание	17
2.3 Визуализация когнитивных функций	20
2.4 SWOT и PEST матрицы и диаграммы	20
2.5 Нелокальные нейронные сети и нейтроны.....	23
2.6 Кластерный и конструктивный анализ	25
Заключение	26
Список используемых источников.....	27

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 определению зависимостей в футбольных матчах самых популярных лиг Европы сезона 2015/2016 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 сборник баз данных Kaggle.

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

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

С электронного ресурса [kaggle.com](https://www.kaggle.com/secareanualin/football-events) возьмем набор данных: events.csv.
<https://www.kaggle.com/secareanualin/football-events>

Данный файл содержит:

- № события.
- Голы хозяев.
- Голы гостей.
- Разница счета.
- Дата.
- Страна.
- Лига.
- Хозяева.
- Гости.

В качестве разделителей используются запяты. Из всех вышеперечисленных характеристик нас интересуют:

- Голы хозяев.
- Голы гостей.
- Разница счета.

Это нам позволит понять, какие команды чаще всего забивают дома или в гостях, а также команды, которые чаще всего пропускают в своем чемпионате.

В данном наборе данных исследование будет проводиться только внутри определенной лиги, так как не рассматриваются другие турниры. В итоге получается следующая таблица для работы:

Таблица 1 – Фрагмент обучающей выборки

№	Голы хозяев	Голы гостей	Разница счета	Дата	Лига	Страна	Хозяева	Гости
1	0	1	-1	07.08.15	F1	france	Lille	Paris Saint-Germain
2	0	1	-1	08.08.15	E0	england	Bournemouth	Aston Villa
3	2	2	0	08.08.15	E0	england	Everton	Watford
4	0	2	-2	08.08.15	F1	france	Montpellier	Angers
5	2	2	0	08.08.15	E0	england	Chelsea	Swansea
6	2	1	1	08.08.15	F1	france	Bastia	Stade Rennes
7	0	0	0	08.08.15	F1	france	Troyes	GFC Ajaccio
8	4	2	2	08.08.15	E0	england	Leicester City	Sunderland
9	1	2	-1	08.08.15	F1	france	Nice	AS Monaco
10	1	0	1	08.08.15	F1	france	Nantes	Guingamp
11	1	3	-2	08.08.15	E0	england	Norwich City	Crystal Palace
12	1	0	1	08.08.15	E0	england	Manchester Utd	Tottenham
13	0	1	-1	08.08.15	F1	france	Marseille	Caen
14	1	2	-1	09.08.15	F1	france	Bordeaux	Stade de Reims
15	2	2	0	09.08.15	E0	england	Newcastle	Southampton
16	0	0	0	09.08.15	F1	france	Lyon	Lorient
17	0	1	-1	09.08.15	E0	england	Stoke City	Liverpool
18	2	1	1	09.08.15	F1	france	Toulouse	St Etienne

Таблица отсортирована по дате проведения матча для более удобного восприятия имеющихся данных.

Классификационные шкалы следующие: голы хозяев, голы гостей, разница счета. Остальные – описательные.

1.3. Ввод выборки в систему Aidos-X

Теперь, когда мы имеем обучающую выборку в формате *.xlsx, можно импортировать ее в систему Aidos-X.

Скопируем выборку в папку d:\Aidos-X\AID_DATA\Inp_data\ и переименуем ее в Inp_data. Затем, запустив систему, воспользуемся универсальным программным интерфейсом для ввода данных из внешних баз табличного вида (режим 2.3.2.2). Настройка ввода будет такая, как на рисунке 1 – ввод обучающей выборки.

The image shows a configuration dialog box for importing data into the Aidos-X system. The dialog is organized into several sections:

- 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"?
- Задайте диапазон столбцов классификационных шкал:**
 - Начальный столбец классификационных шкал: 2
 - Конечный столбец классификационных шкал: 4
- Задайте диапазон столбцов описательных шкал:**
 - Начальный столбец описательных шкал: 5
 - Конечный столбец описательных шкал: 9
- Задайте режим:**
 - Формализации предметной области (на основе "Inp_data")
 - Генерации распознаваемой выборки (на основе "Inp_gasp")
- 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)")

Buttons: Ok, Cancel

Рисунок 1 – Ввод обучающей выборки

Стоит отметить следующие настройки:

- Тип файла – xlsx.
- Классификационные шкалы – 2-4.
- Описательные шкалы – 5-9.
- Не применяется спец. интерпретация текстовых полей классов.

Так как предварительно числовые шкалы были разбиты на интервалы так, как это необходимо для исследования, перерасчет шкал после ввода выборки производить не надо.

После импорта автоматически формируются классификационные и описательные шкалы, с применением которых исходные данные кодируются и представляются в форме эвентологических баз данных. Этим самым полностью автоматизировано выполняется этап «Формализация предметной области».

Классификационные шкалы можно просмотреть в режиме 2.1 (рисунок 2).

Код шкалы	Наименование классификационной шкалы
1	ГОЛЫ ХОЗЯЕВ
2	ГОЛЫ ГОСТЕЙ
3	РАЗНИЦА СЧЕТА

Код градации	Наименование градации классификационной шкалы	DEL
1	Очень малое: 1/5-{1.0000000, 2.8000000}	
2	Малое: 2/5-{2.8000000, 4.6000000}	
3	Среднее: 3/5-{4.6000000, 6.4000000}	
4	Большое: 4/5-{6.4000000, 8.2000000}	
5	Очень большое: 5/5-{8.2000000, 10.0000000}	

Рисунок 2 – Классификационные шкалы

Описательные шкалы – в режиме 2.2 (рисунок 3).

Для ручного ввода-корректировки выборки существует режим 2.3.1, его можно увидеть на рисунке 4. Установка значений описательных и классификационных шкал объектов осуществляется по их номерам.

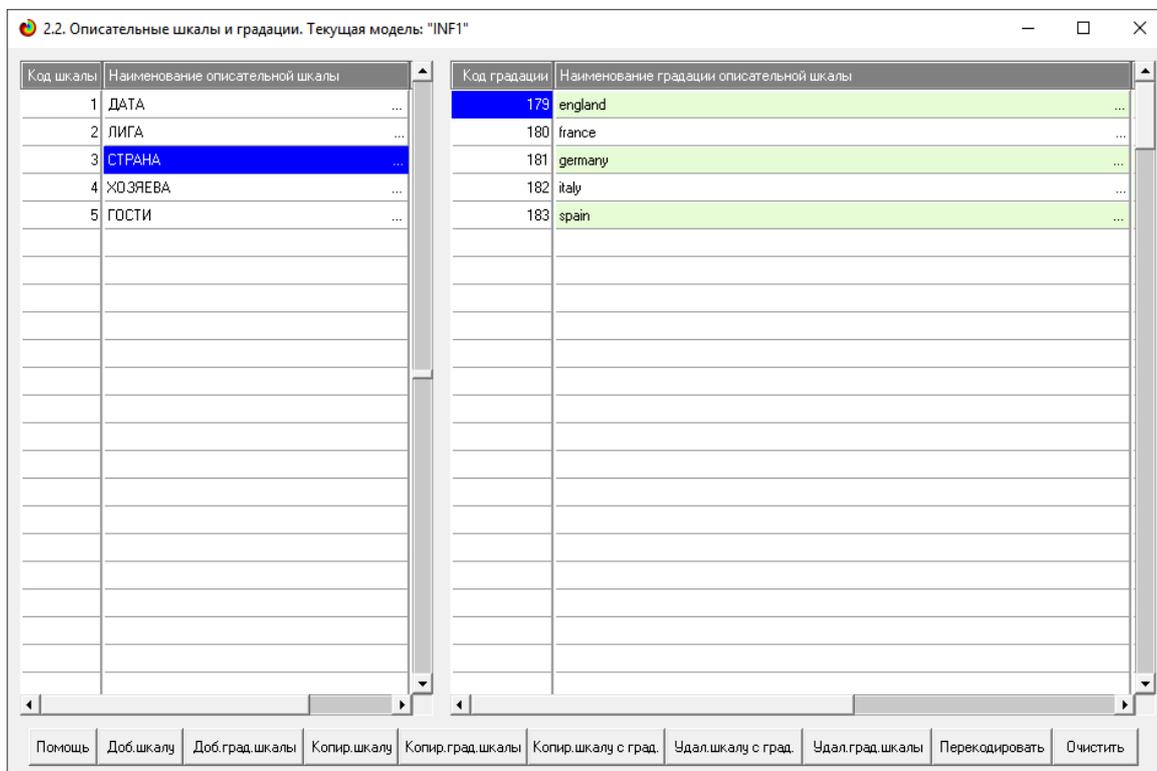


Рисунок 3 – Описательные шкалы

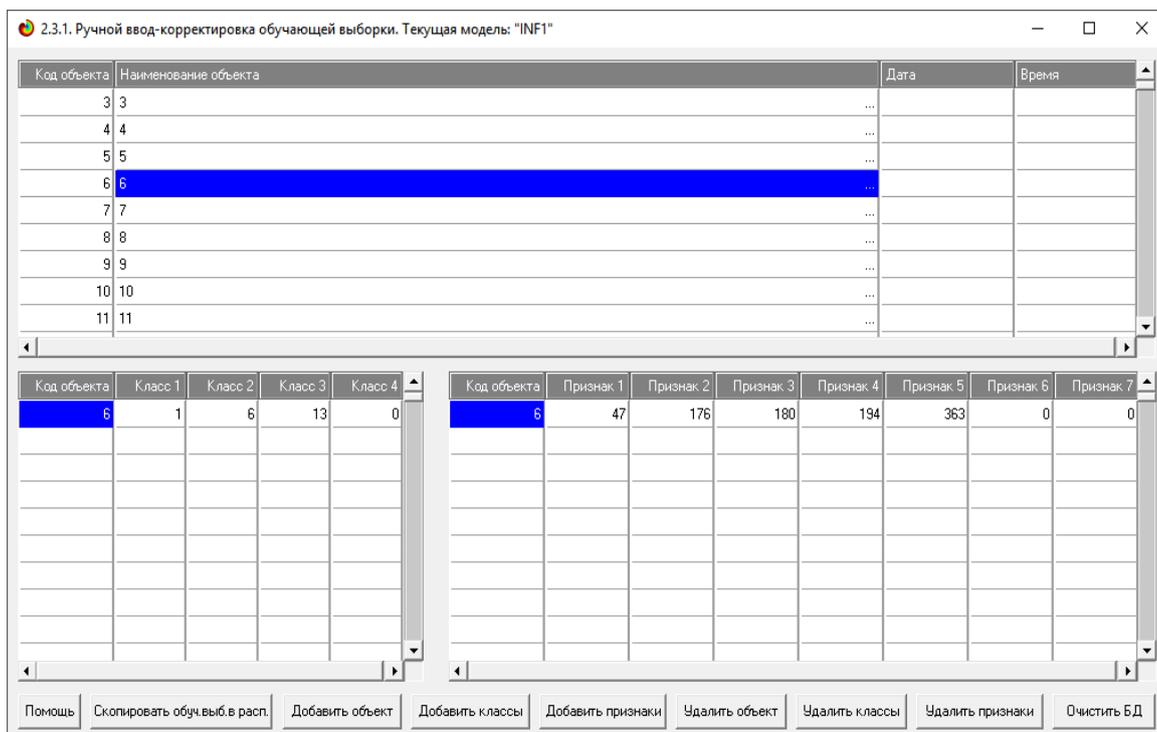


Рисунок 4 – Ручной ввод-корректировка выборки

Тем самым создаются все необходимые и достаточные предпосылки для выявления связей между значениями факторов и результатами их совместного системного воздействия.

1.4. Синтез и верификация статистических и интеллектуальных моделей

Далее запускаем режим 3.5 – Синтез и верификация моделей. Настраиваем так, как показано на рисунке 5.

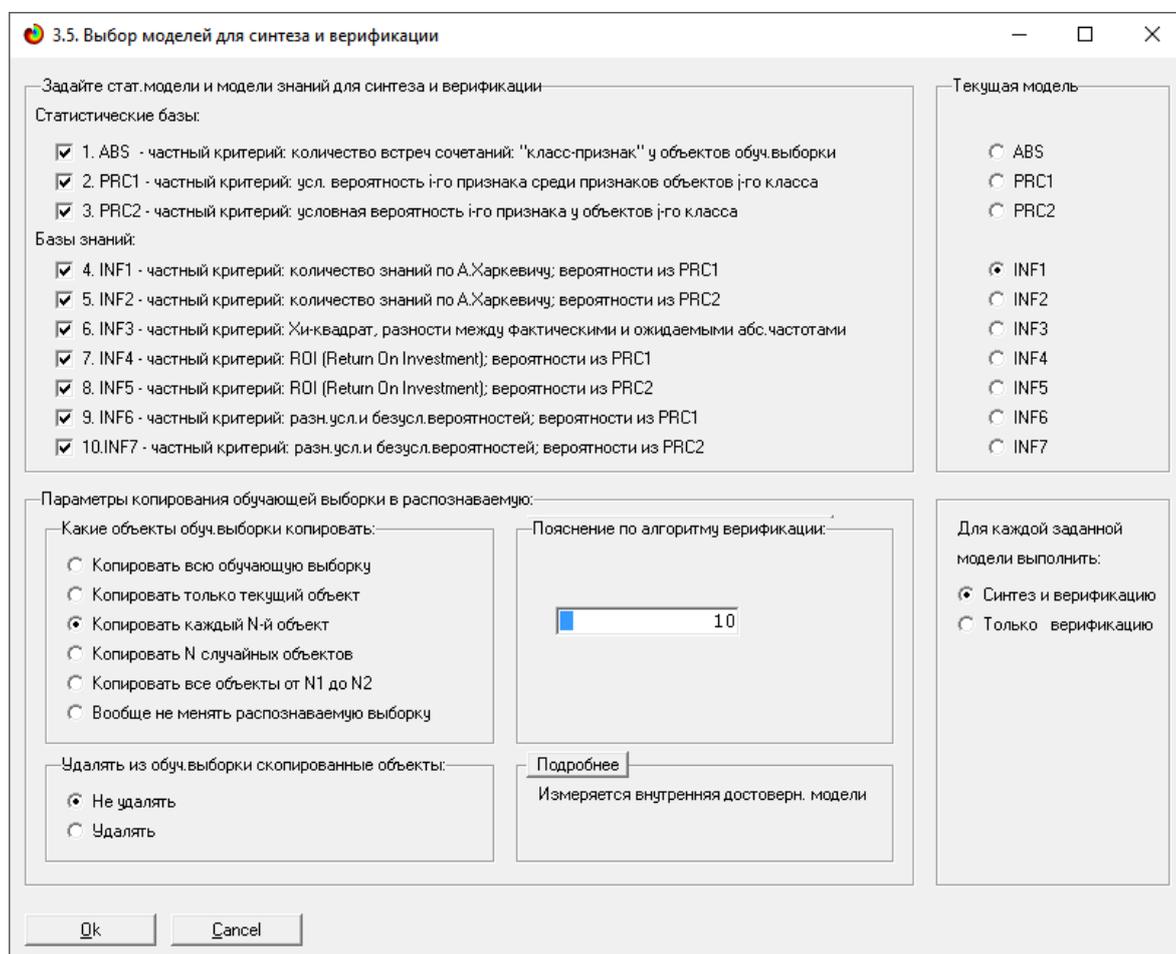


Рисунок 5 – Настройка синтеза и верификации моделей

Так как в выборке около 2000 значений, для распознавания будем использовать каждый 10й объект. Из-за сортировки таблицы по дате это будет надежнее, чем использование 200 рандомных записей.

После построения моделей нажимаем «ОК» на форме, которая показана на рисунке 6.

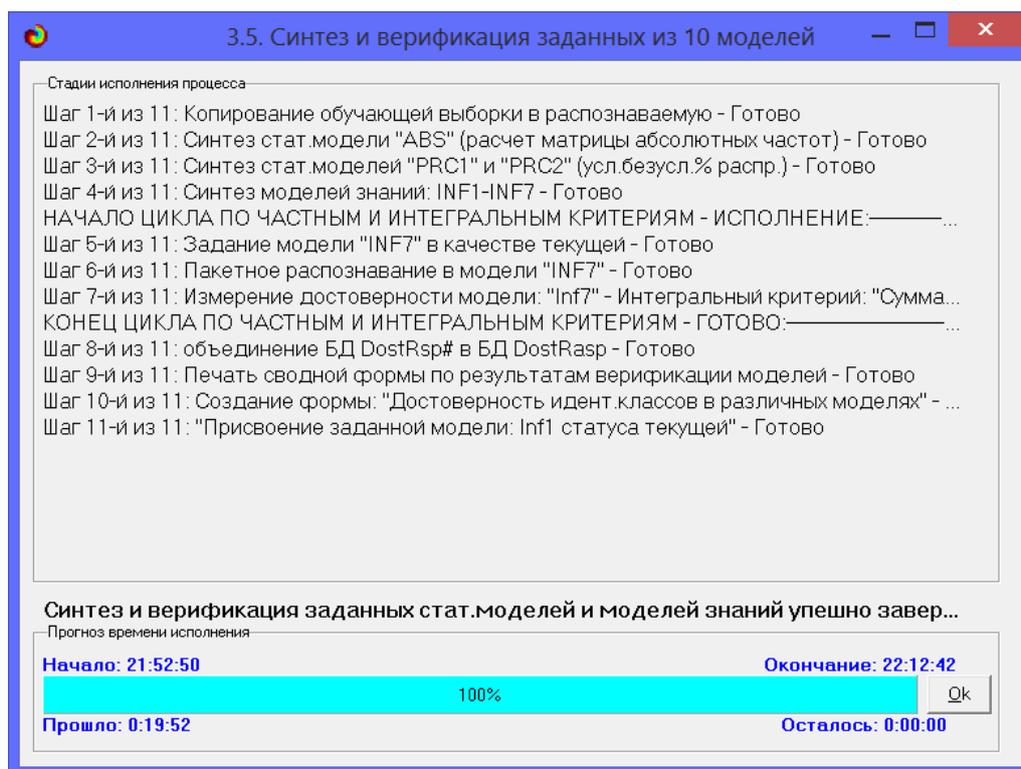


Рисунок 6 – Завершение процесса синтеза и верификации моделей

После этого можно перейти непосредственно к выбору наиболее достоверной модели.

1.5. Виды моделей системы «Aidos»

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 обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. ГОЛЫ ХОЗЯЕВ ОЧЕНЬ МАЛОЕ 1/5 (1.000000, 2.800000)	2. ГОЛЫ ХОЗЯЕВ МАЛОЕ: 2/5 (2.800000, 4.600000)	3. ГОЛЫ ХОЗЯЕВ СРЕДНЕЕ: 3/5 (4.600000, 6.400000)	4. ГОЛЫ ХОЗЯЕВ БОЛЬШОЕ: 4/5 (6.400000, 8.200000)	5. ГОЛЫ ХОЗЯЕВ ОЧЕНЬ БОЛЬШОЕ: 5/5 (8.200000, 10.000000)	6. ГОЛЫ ГОСТЕЙ ОЧЕНЬ МАЛОЕ: 1/5 (1.000000, 2.600000)	7. ГОЛЫ ГОСТЕЙ МАЛОЕ: 2/5 (2.600000, 4.200000)	8. ГОЛЫ ГОСТЕЙ СРЕДНЕЕ: 3/5 (4.200000, 5.800000)	9. ГОЛЫ ГОСТЕЙ БОЛЬШОЕ: 4/5 (5.800000, 7.400000)	10. ГОЛЫ ГОСТЕЙ ОЧЕНЬ БОЛЬШОЕ: 5/5 (7.400000, 9.000000)	11. РАЗНИЦА СЧЕТА ОЧЕНЬ МАЛОЕ: 1/5 (5.600000, 2.200000)	12. РАЗНИЦА СЧЕТА МАЛОЕ: 2/5 (5.600000, 2.200000)
1	ДАТА-01.02.16	1					1						
2	ДАТА-01.03.16	6	2				4	2					1
3	ДАТА-01.04.16	2	1				1						
4	ДАТА-01.05.16	7	3				7	1					
5	ДАТА-01.11.15	9	2	1			5	2					
6	ДАТА-01.12.15												
7	ДАТА-02.01.16	8	1				4						
8	ДАТА-02.02.16	5	1				5	1					1
9	ДАТА-02.03.16	11	4				7	2					1
10	ДАТА-02.04.16	12	7	1			16	4					3
11	ДАТА-02.05.16	3	1	1			4	1					
12	ДАТА-02.10.15	2						1					
13	ДАТА-02.11.15	1	1				3						
14	ДАТА-02.12.15	4					3	1					1
15	ДАТА-03.01.16	6					6	2					2
16	ДАТА-03.02.16	10	5				9						
17	ДАТА-03.03.16	2							1	1			2
18	ДАТА-03.04.16	13	3	1			11	3					2
19	ДАТА-03.10.15	17	2	1			15	1					

Таблица 3 – Матрица информативностей (модель ABS) в битах

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRС1"

Код признака	Наименование описательной шкалы и градации	1. ГОЛЫ ХОЗЯЕВ ОЧЕНЬ МАЛОЕ: 1/5 (1.000000, 2.800000)	2. ГОЛЫ ХОЗЯЕВ МАЛОЕ: 2/5 (2.800000, 4.600000)	3. ГОЛЫ ХОЗЯЕВ СРЕДНЕЕ: 3/5 (4.600000, 6.400000)	4. ГОЛЫ ХОЗЯЕВ БОЛЬШОЕ: 4/5 (6.400000, 8.200000)	5. ГОЛЫ ХОЗЯЕВ ОЧЕНЬ БОЛЬШОЕ: 5/5 (8.200000, 10.000000)	6. ГОЛЫ ГОСТЕЙ ОЧЕНЬ МАЛОЕ: 1/5 (1.000000, 2.600000)	7. ГОЛЫ ГОСТЕЙ МАЛОЕ: 2/5 (2.600000, 4.200000)	8. ГОЛЫ ГОСТЕЙ СРЕДНЕЕ: 3/5 (4.200000, 5.800000)	9. ГОЛЫ ГОСТЕЙ БОЛЬШОЕ: 4/5 (5.800000, 7.400000)	10. ГОЛЫ ГОСТЕЙ ОЧЕНЬ БОЛЬШОЕ: 5/5 (7.400000, 9.000000)	11. РАЗНИЦА СЧЕТА ОЧЕНЬ МАЛОЕ: 1/5 (5.600000, 2.200000)	12. РАЗНИЦА СЧЕТА МАЛОЕ: 2/5 (5.600000, 2.200000)	13. РАЗНИЦА СЧЕТА СРЕДНЕЕ: 3/5 (5.600000, 2.200000)
1	ДАТА-01.02.16	0.140					0.154							0.051
2	ДАТА-01.03.16	0.045	0.096				-0.100	0.236					0.321	-0.114
3	ДАТА-01.04.16	0.080	0.290				-0.177							-0.280
4	ДАТА-01.05.16	-0.024	0.126				-0.010	-0.163						-0.015
5	ДАТА-01.11.15	0.028	-0.079	0.323			-0.188	0.061						0.051
6	ДАТА-01.12.15													0.480
7	ДАТА-02.01.16	0.140	-0.191				-0.116							0.051
8	ДАТА-02.02.16	-0.027	-0.174				-0.013	-0.034					0.321	-0.045
9	ДАТА-02.03.16	0.056	0.141				-0.107	0.011					0.095	0.000
10	ДАТА-02.04.16	-0.096	0.174	0.086			0.030	0.095					0.338	-0.071
11	ДАТА-02.05.16	-0.059	-0.007	0.664			0.067	0.132						-0.146
12	ДАТА-02.10.15	0.140						0.490						0.051
13	ДАТА-02.11.15	-0.243	0.238				0.200							-0.061
14	ДАТА-02.12.15	0.080					-0.018	0.159					0.514	0.076
15	ДАТА-03.01.16	0.062					0.076	0.254					0.608	-0.027
16	ДАТА-03.02.16	0.009	0.219				-0.018							0.022
17	ДАТА-03.03.16	0.080						0.430	1.583				1.055	
18	ДАТА-03.04.16	0.051	-0.041	0.202			0.000	0.099					0.296	-0.061
19	ДАТА-03.10.15	0.083	-0.271	0.130			0.048	-0.402						-0.006

Таблица 4 – Матрица знаний (модель ABS)

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

Код признака	Наименование описательной шкалы и градации	1. ГОЛЫ ХОЗЯЕВ ОЧЕНЬ МАЛОЕ: 1/5 (1.000000, 2.800000)	2. ГОЛЫ ХОЗЯЕВ МАЛОЕ: 2/5 (2.800000, 4.600000)	3. ГОЛЫ ХОЗЯЕВ СРЕДНЕЕ: 3/5 (4.600000, 6.400000)	4. ГОЛЫ ХОЗЯЕВ БОЛЬШОЕ: 4/5 (6.400000, 8.200000)	5. ГОЛЫ ХОЗЯЕВ ОЧЕНЬ БОЛЬШОЕ: 5/5 (8.200000, 10.000000)	6. ГОЛЫ ГОСТЕЙ ОЧЕНЬ МАЛОЕ: 1/5 (1.000000, 2.600000)	7. ГОЛЫ ГОСТЕЙ МАЛОЕ: 2/5 (2.600000, 4.200000)	8. ГОЛЫ ГОСТЕЙ СРЕДНЕЕ: 3/5 (4.200000, 5.800000)	9. ГОЛЫ ГОСТЕЙ БОЛЬШОЕ: 4/5 (5.800000, 7.400000)	10. ГОЛЫ ГОСТЕЙ ОЧЕНЬ БОЛЬШОЕ: 5/5 (7.400000, 9.000000)	11. РАЗНИЦА СЧЕТА ОЧЕНЬ МАЛОЕ: 1/5 (9.000000, 5.600000)	12. РАЗНИЦА СЧЕТА МАЛОЕ: 2/5 (5.600000, 2.200000)	13. РАЗНИЦА СЧЕТА СРЕДНЕЕ: 3/5 (5.600000, 2.200000)
1	ДАТА-01.02.16	0.302	-0.204	-0.036	-0.001	-0.001	0.326	-0.142	-0.007	-0.001	-0.001	-0.003	-0.057	0.122
2	ДАТА-01.03.16	0.649	0.437	-0.279	-0.005	-0.005	-1.165	0.908	-0.057	-0.010	-0.010	-0.021	0.560	-1.728
3	ДАТА-01.04.16	0.371	0.524	-0.085	-0.002	-0.002	-0.572	-0.332	-0.017	-0.003	-0.003	-0.006	-0.134	-1.048
4	ДАТА-01.05.16	-0.446	0.825	-0.389	-0.007	-0.007	-0.186	-0.519	-0.079	-0.014	-0.014	-0.029	-0.612	-0.361
5	ДАТА-01.11.15	0.624	-0.446	0.563	-0.008	-0.008	-3.085	0.291	-0.089	-0.016	-0.016	-0.032	-0.689	1.469
6	ДАТА-01.12.15	-0.698	-0.204	-0.036	-0.001	-0.001	-0.674	-0.142	-0.007	-0.001	-0.001	-0.003	-0.057	2.122
7	ДАТА-02.01.16	2.416	-0.631	-0.292	-0.005	-0.005	-1.390	-1.140	-0.059	-0.011	-0.011	-0.022	-0.459	0.979
8	ДАТА-02.02.16	-0.351	-0.563	-0.279	-0.005	-0.005	-0.165	-0.092	-0.057	-0.010	-0.010	-0.021	0.560	-0.728
9	ДАТА-02.03.16	1.460	1.214	-0.498	-0.009	-0.009	-2.207	0.053	-0.101	-0.018	-0.018	-0.037	0.216	0.006
10	ДАТА-02.04.16	-3.356	2.515	0.198	-0.015	-0.015	1.178	0.866	-0.163	-0.030	-0.030	-0.059	1.738	-3.307
11	ДАТА-02.05.16	-0.490	-0.019	0.818	-0.003	-0.003	0.631	0.288	-0.037	-0.007	-0.007	-0.014	-0.287	-1.388
12	ДАТА-02.10.15	0.604	-0.408	-0.073	-0.001	-0.001	-1.347	0.715	-0.015	-0.003	-0.003	-0.005	-0.115	0.245
13	ДАТА-02.11.15	-0.861	0.456	-0.097	-0.002	-0.002	1.203	-0.380	-0.020	-0.004	-0.004	-0.007	-0.153	-0.340
14	ДАТА-02.12.15	0.743	-0.951	-0.170	-0.003	-0.003	-0.144	0.335	-0.035	-0.006	-0.006	-0.013	0.732	0.905
15	ДАТА-03.01.16	0.881	-1.495	-0.267	-0.005	-0.005	1.059	0.955	-0.054	-0.010	-0.010	-0.020	1.579	-0.436
16	ДАТА-03.02.16	0.228	2.146	-0.510	-0.009	-0.009	-0.432	-1.994	-0.104	-0.019	-0.019	-0.038	-0.803	0.714
17	ДАТА-03.03.16	0.371	-0.476	-0.085	-0.002	-0.002	-1.572	0.668	0.983	-0.003	-0.003	-0.006	1.866	-2.048
18	ДАТА-03.04.16	1.599	-0.330	0.405	-0.011	-0.011	-0.004	0.673	-0.121	-0.022	-0.022	-0.044	1.063	-2.334
19	ДАТА-03.10.15	3.272	-2.009	0.283	-0.013	-0.013	1.750	-1.801	-0.146	-0.027	-0.027	-0.053	-1.128	-0.259

1.5. Результаты верификации моделей

Результаты верификации моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями, приведены на рисунке 6.

4.1.3.6. Обобщ.форма по достов.моделей при разн.инт.критг.. Текущая модель: "INF1"

Наименование модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительн. решений (TP)	Число истинно-отрицательных решений (TN)	Число ложн. положительных решений (FP)	Число ложн. отрицательных решений (FN)	Точность модели	Полнота модели	Ф-мера Ван Ризбергена	Сумма мод. уровней схо. истинно-поло. решений (ST)	Сумма мод. уровней схо. истинно-отриц. решений (ST)
1. ABS - частный критерий: количество встреч сочтаний "i, клас...	Корреляция абс частот с обр...	4444	4444	5602	17374		0.204	1.000	0.338	1231.618	93.716
1. ABS - частный критерий: количество встреч сочтаний "i, клас...	Сумма абс частот по признак...	4444	4444	5602	17374		0.204	1.000	0.338	2894.570	
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл.отн частот с о...	4444	4444	5602	17374		0.204	1.000	0.338	1231.618	93.716
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма усл.отн частот по приз...	4444	4444	5602	17374		0.204	1.000	0.338	397.651	
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл.отн частот с о...	4444	4444	5602	17374		0.204	1.000	0.338	1231.544	93.715
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл.отн частот по приз...	4444	4444	5602	17374		0.204	1.000	0.338	397.651	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	4444	2463	15990	6986	1981	0.261	0.554	0.355	93.783	614.455
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	4444	2952	12474	10502	1492	0.219	0.664	0.330	125.822	348.368
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	4444	2463	15990	6986	1981	0.261	0.554	0.355	93.783	614.455
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	4444	2952	12474	10502	1492	0.219	0.664	0.330	125.822	348.368
6. INF3 - частный критерий: Хинквадрат, разности между фактич...	Семантический резонанс зна...	4444	2513	13766	9210	1931	0.214	0.565	0.311	364.014	1760.744
6. INF3 - частный критерий: Хинквадрат, разности между фактич...	Сумма знаний	4444	2513	13766	9210	1931	0.214	0.565	0.311	741.908	1511.656
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	4444	1690	19567	3409	2754	0.331	0.380	0.354	67.213	546.558
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	4444	3168	11677	11299	1276	0.219	0.713	0.335	19.063	12.655
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	4444	1690	19567	3409	2754	0.331	0.380	0.354	67.213	546.558
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	4444	3168	11677	11299	1276	0.219	0.713	0.335	19.063	12.655
9. INF6 - частный критерий: разн усл и безуслов вероятностей, вер...	Семантический резонанс зна...	4444	2446	13941	9035	1998	0.213	0.550	0.307	348.347	1265.442
9. INF6 - частный критерий: разн усл и безуслов вероятностей, вер...	Сумма знаний	4444	2513	13506	9470	1931	0.210	0.565	0.306	19.524	122.066
10. INF7 - частный критерий: разн усл и безуслов вероятностей, ве...	Семантический резонанс зна...	4444	2446	13941	9035	1998	0.213	0.550	0.307	348.347	1265.442
10. INF7 - частный критерий: разн усл и безуслов вероятностей, ве...	Сумма знаний	4444	2513	13506	9470	1931	0.210	0.565	0.306	19.524	122.066

Помощь

Рисунок 6 – Результаты верификации моделей

Наиболее достоверной в данном приложении оказалась модель INF5 при интегральном критерии «Семантический резонанс знаний», если учитывать L_меру проф. Е.В. Луценко. Точность модели составляет 0,889, то есть уровень достоверности прогнозирования с применением модели выше, чем экспертных оценок, достоверность которых считается равной примерно 49,4%.

Также статистические модели обычно дают более низкую средневзвешенную достоверность идентификации и неидентификации, чем модели знаний, и практически никогда – более высокую. Соответственно, это оправдывает применение интеллектуальных технологий и моделей знаний. Посмотрим, каковы частные распределения уровней сходства и различия для верно и ошибочно неидентифицированных и идентифицированных ситуаций

в наиболее достоверной модели ABS (рисунок 7). Для этого зайдём в режим 4.1.3.11, выберем модель ABS и запросим график по сумме знаний.

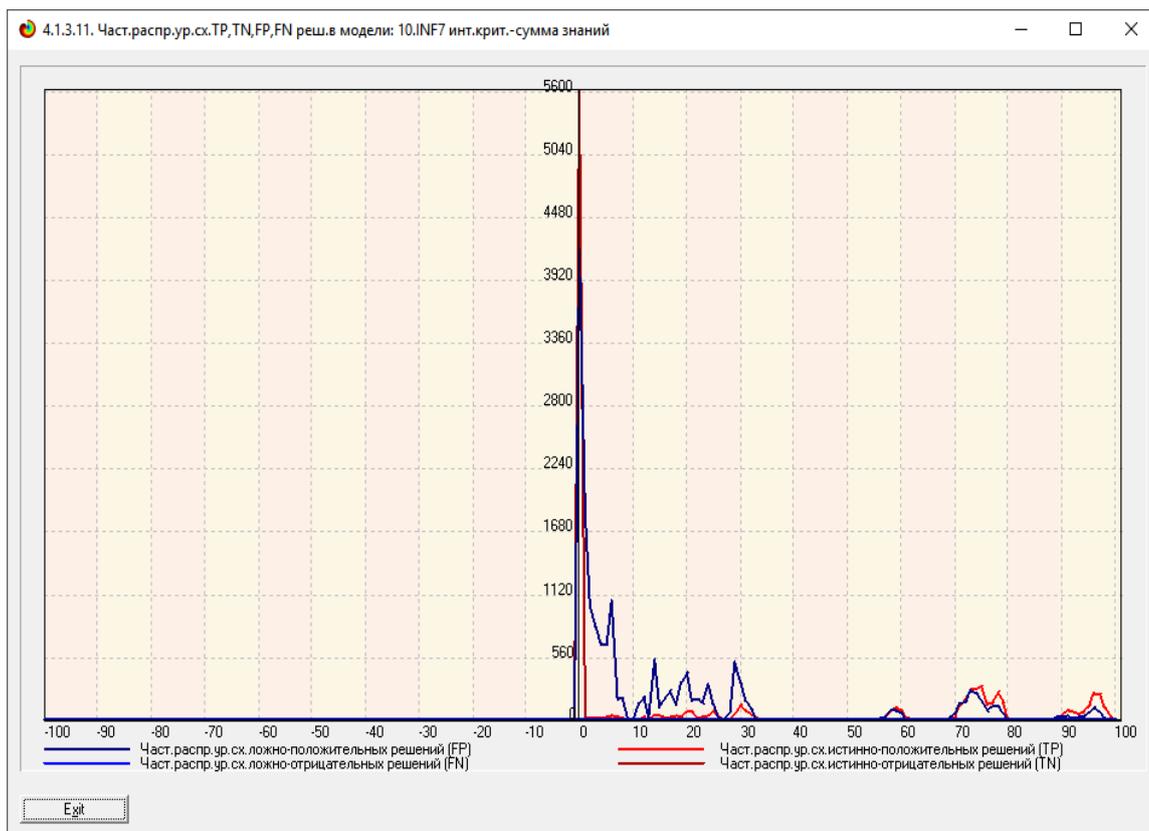


Рисунок 7 – Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели ABS

Как можно видеть, при модели ABS непринадлежность объекта к классу лучше определяется, чем принадлежность. При этом существует вероятность ложного определения непринадлежности или принадлежности.

Выберем модель ABS в качестве основной и перейдем к решению задач идентификации.

2. РЕШЕНИЕ ЗАДАЧ НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи идентификации

Если смотреть с позиции букмекера, то какие задачи при анализе баз данных пользователей можно выделить?

Во-первых, зависимость того – какая из команд играет в гостях, а какая проводит домашний матч. Ведь зная какие команды лучше играют домашние матчи, а какие команды лучше играют гостевые матчи, можно делать ставки на победу одной из команд.

Во-вторых, ставки можно ставить и на количество забитых или же пропущенных голов. Отсюда следует вывод, что можно увидеть закономерности забивания количества голов в матче определенных команд. Но тут стоит учитывать и то, какая команда принимает матч.

2.2. Выбор модели и пакетное распознавание

Зададим модель INF5 в качестве текущей в режиме 5.6 (рисунок 8), после чего проведем пакетное распознавание в режиме 4.1.2 (рисунок 9).

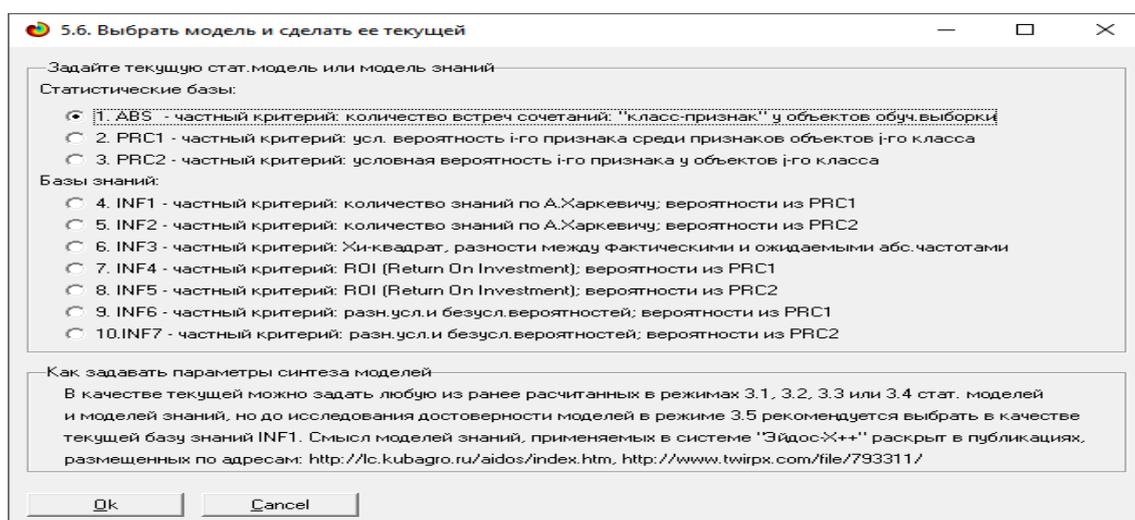


Рисунок 8 – Выбор текущей модели

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

Ниже кратко рассмотрим некоторые из них.

На рисунках 10 и 11 приведены примеры прогнозов высокой и низкой достоверности классов, а также итоги распознавания классов объектов в модели ABS:

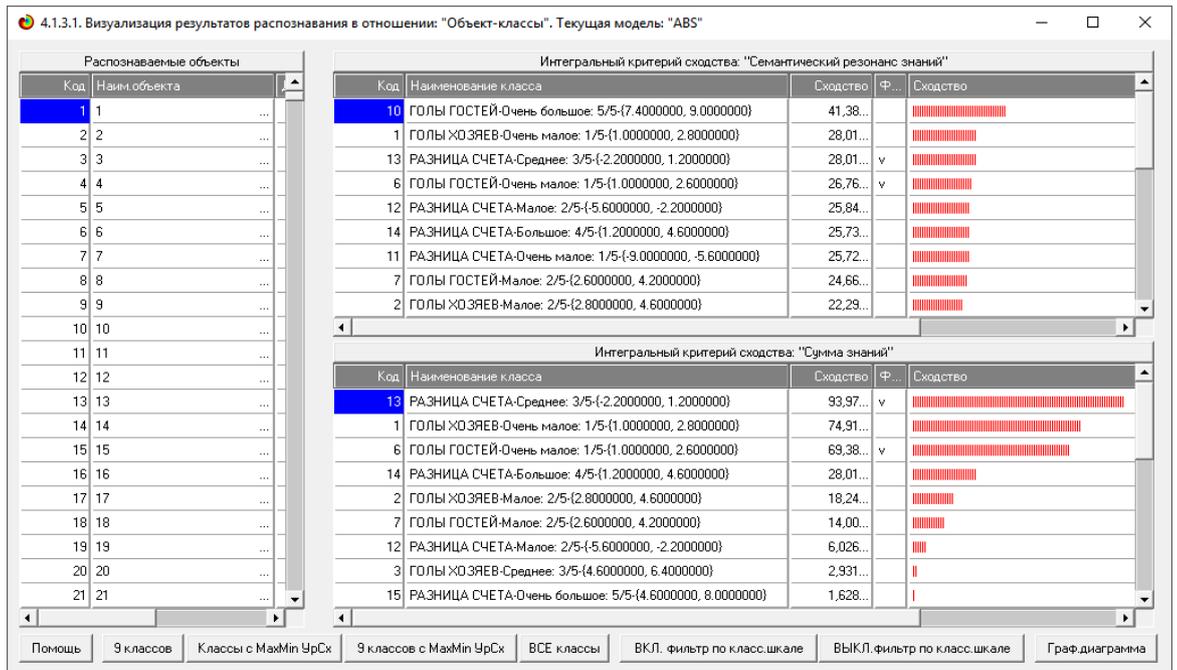


Рисунок 10 – Прогнозы достоверности классов объектов

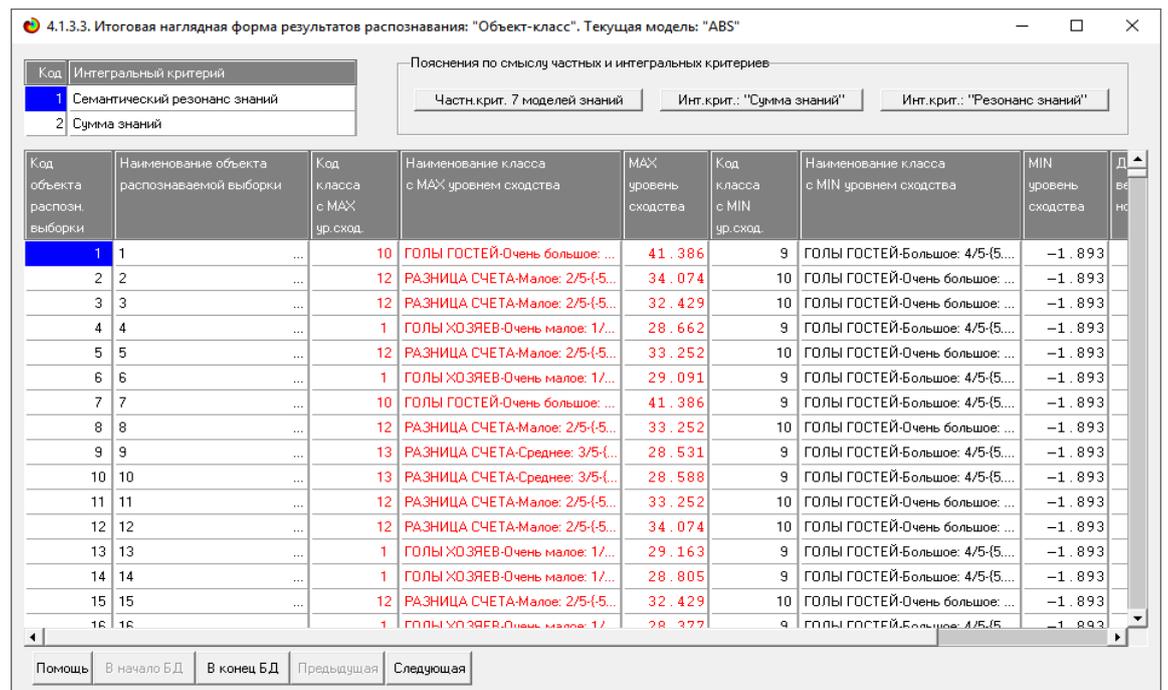


Рисунок 11 – Итоги распознавания классов объектов

2.3 Визуализация когнитивных функций

Применительно к задачам, рассматриваемым 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.

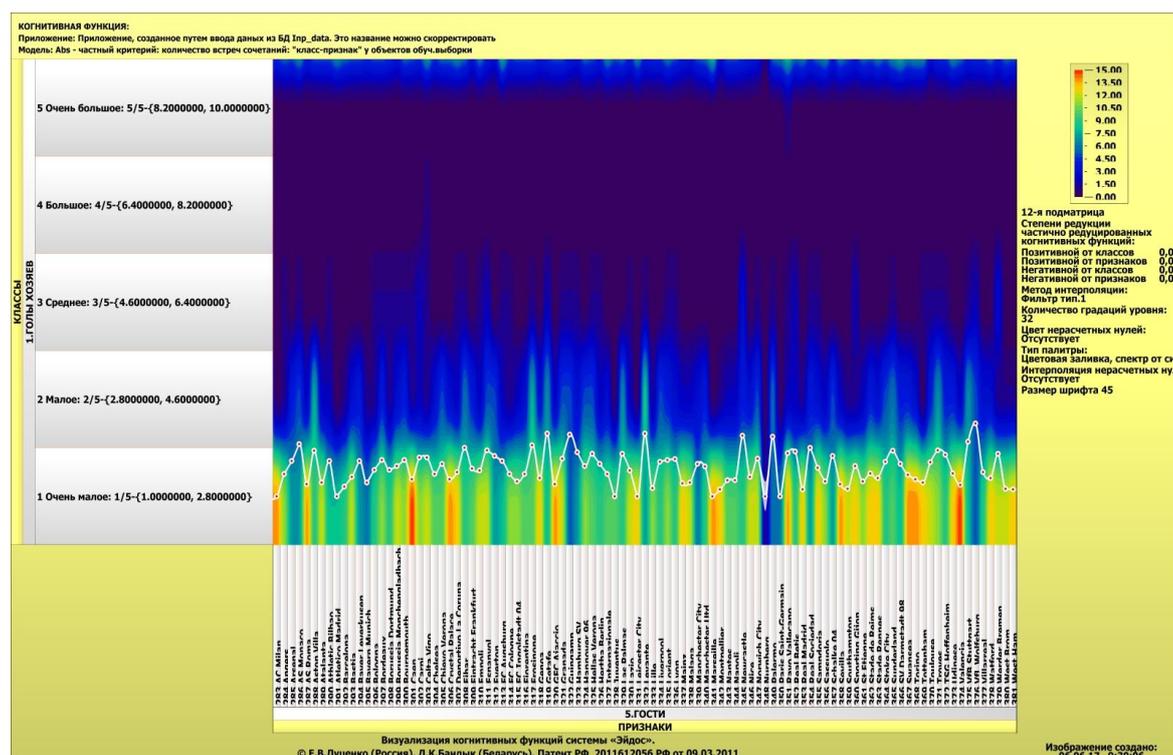


Рисунок 12 – Визуализация зависимости количества голов хозяев от команды гостей

Рассмотрим визуализацию когнитивных функций в рамках определенных задач.

Зависимость количества голов хозяев от команды гостей показывает нам насколько много команда гостей пропускает в матчах своей лиги.

2.4 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 неформализуемым 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.

В версии системы под MS Windows: «Aidos-X++» предложено автоматизированное количественное решение прямой и обратной задач SWOT-анализа с построением традиционных SWOT-матриц и диаграмм (рисунок 13).

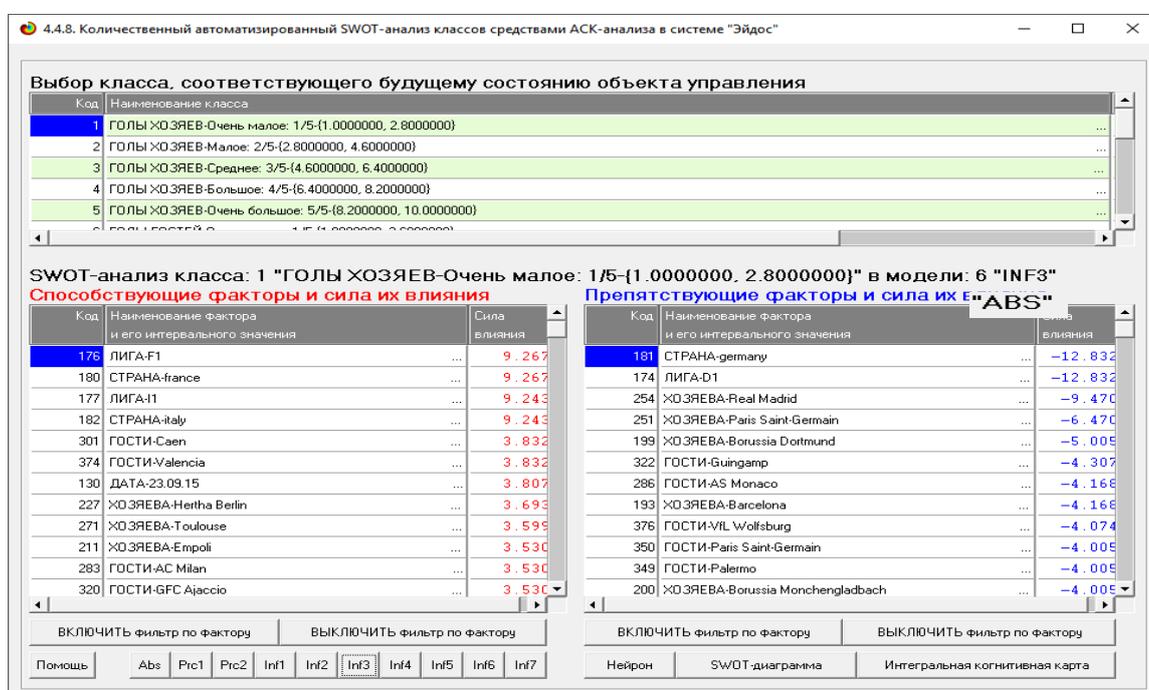


Рисунок 13 – Пример SWOT-Матрицы в модели ABS

На рисунке 14 приведен пример инвертированной SWOT-матрицы.

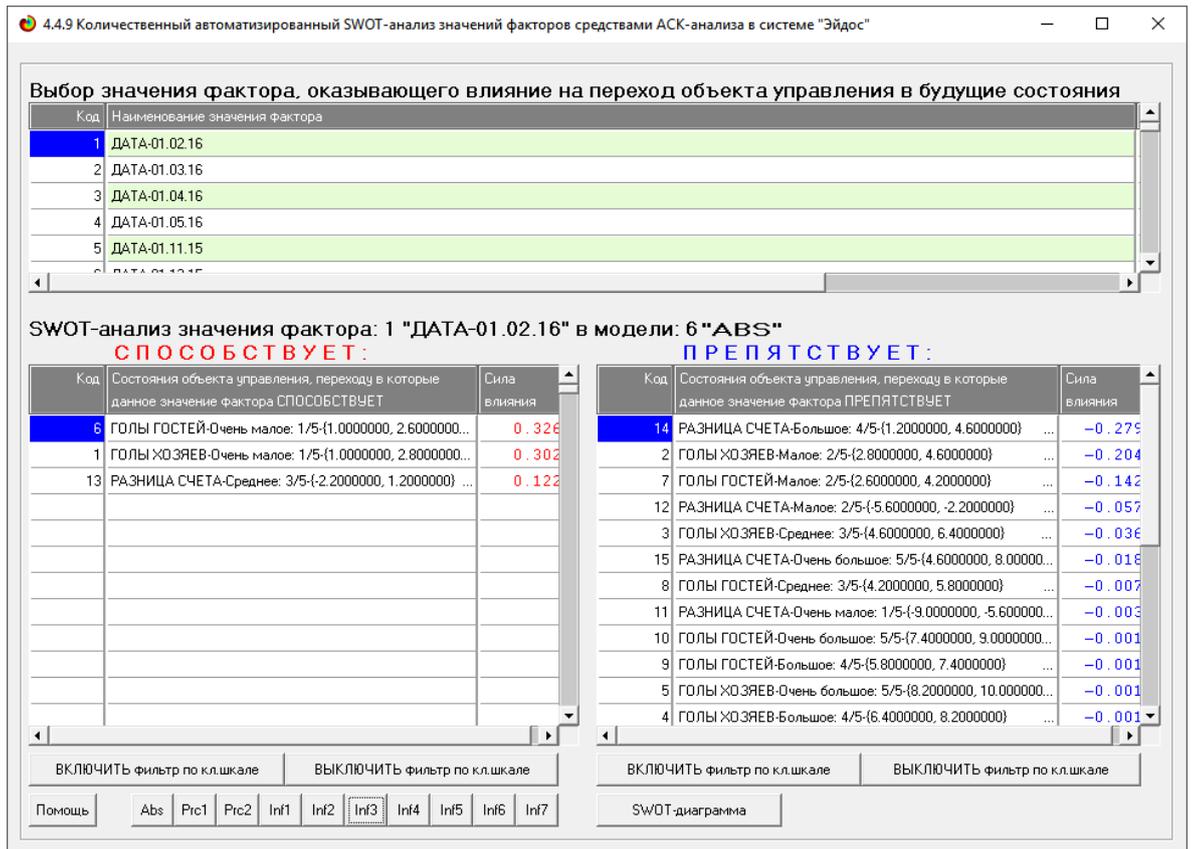


Рисунок 14 – Пример инвертированной SWOT-Матрицы в модели ABS
 А теперь рассмотрим даты, в которые чаще всего забивали хозяева (рисунок 15):

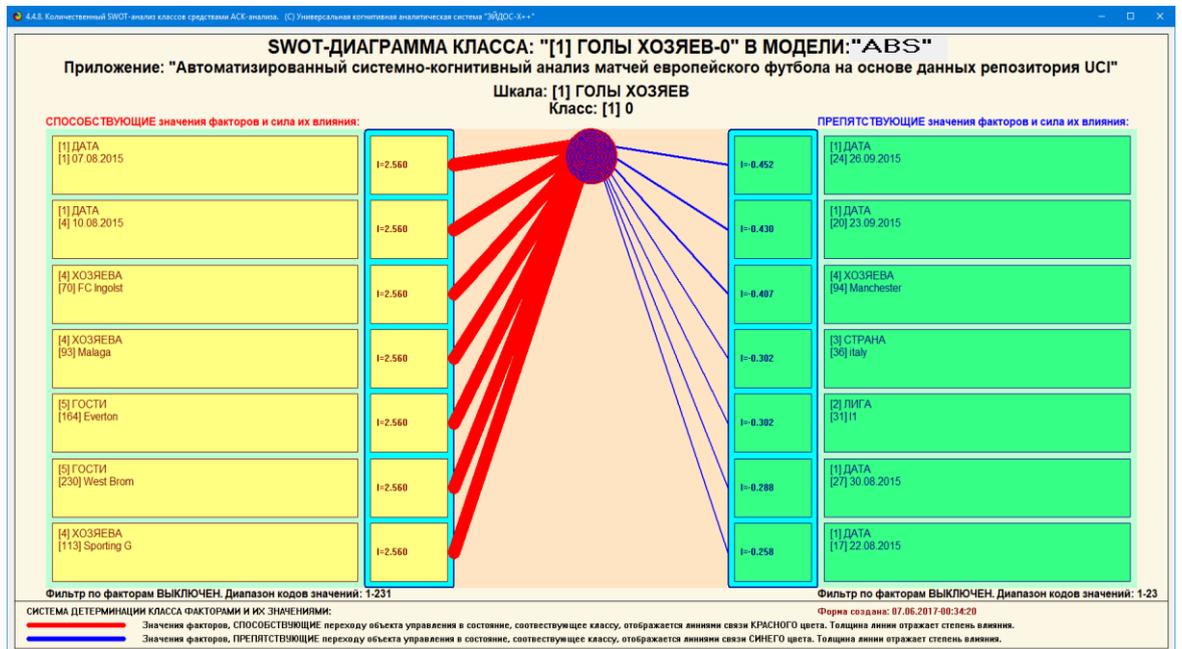


Рисунок 15 – SWOT-матрица

Как можно видеть, большая часть голов была забита хозяевами поля в определенные даты.

А теперь узнаем, в какие даты были самые популярные разницы счета (рисунок 20).

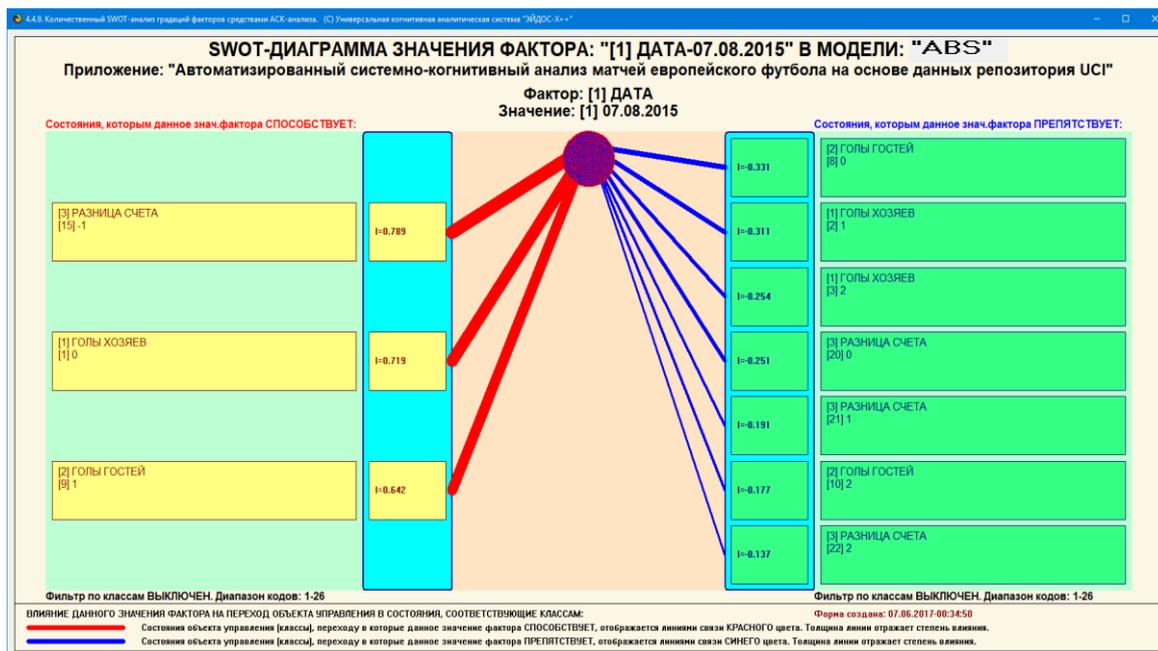


Рисунок 16 – SWOT-диаграмма разницы счета

2.5 Нелокальные нейронные сети и нейтроны

Каждому классу системно-когнитивной модели соответствует нелокальный нейрон, совокупность которых образует не локальную нейронную сеть.

Рассмотрим пару примеров, возвращаясь к нашим задачам.

На рисунке 17 изображен нелокальный нейрон для 30-39 уровней.

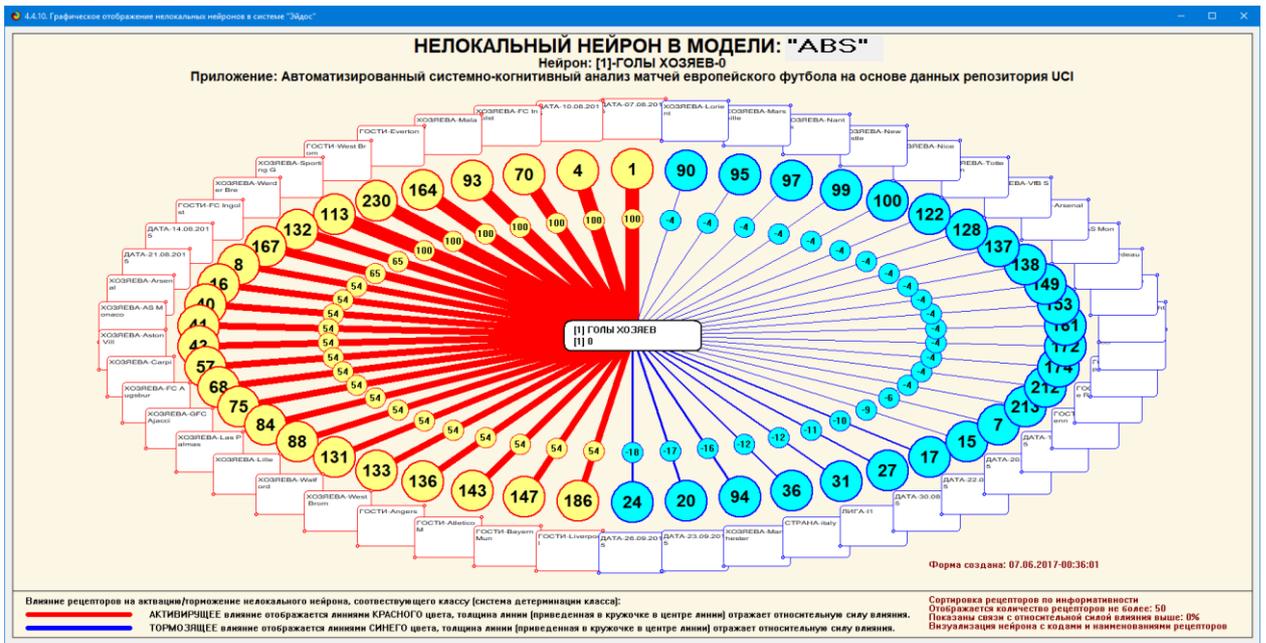


Рисунок 17 – Нейрон для уровней 30-39

На его основании можно определить, что Западный край пользуется большей популярностью, чем Нагорье Арати. Аналогично с Пустошами и Тысячью Игл. В действительности, некоторые из этих локаций были подвергнуты значительным изменениям в дополнении «Cataclysm», что может подтвердить достоверность используемой модели.

На основании нелокальных нейронов строится нейронная сеть, пример которой представлен на рисунке 18.

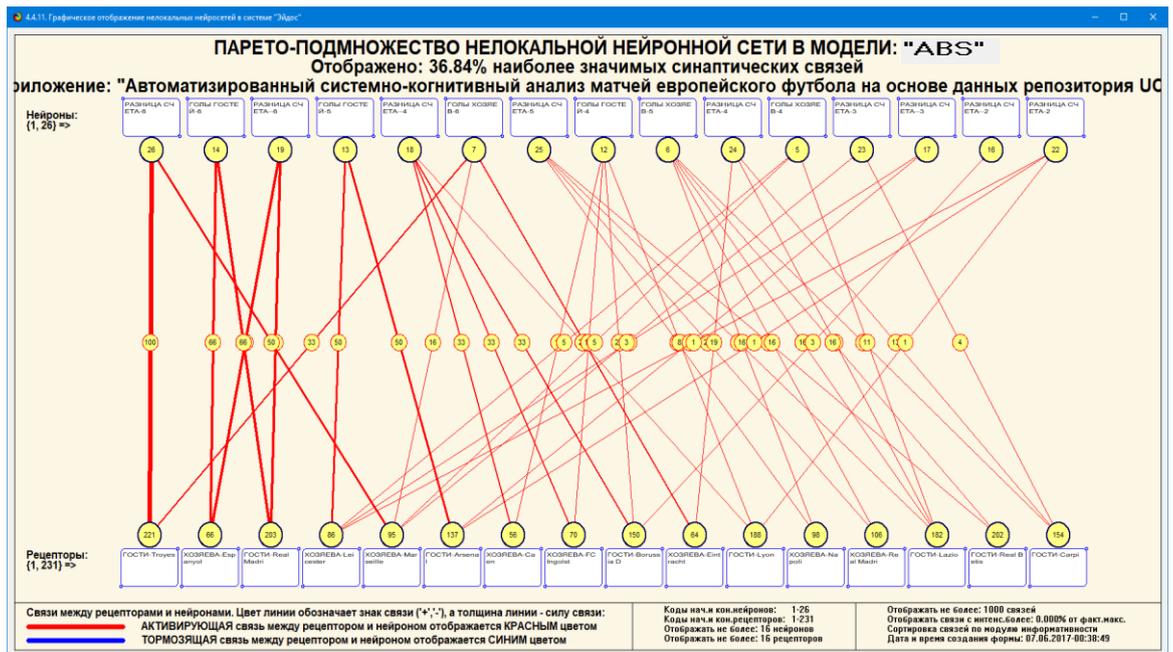


Рисунок 18 – Пример нелокальной нейронной сети в модели INF5

По ней можно, в основном, определить зависимость выбора локации от расы и уровня.

2.6 Кластерный и конструктивный анализ

В режиме 4.2.2.2 после расчета матриц сходства, кластеров и конструкторов (4.2.2.1) строим 2D сеть классов в выбранной модели знаний (рисунок 24).

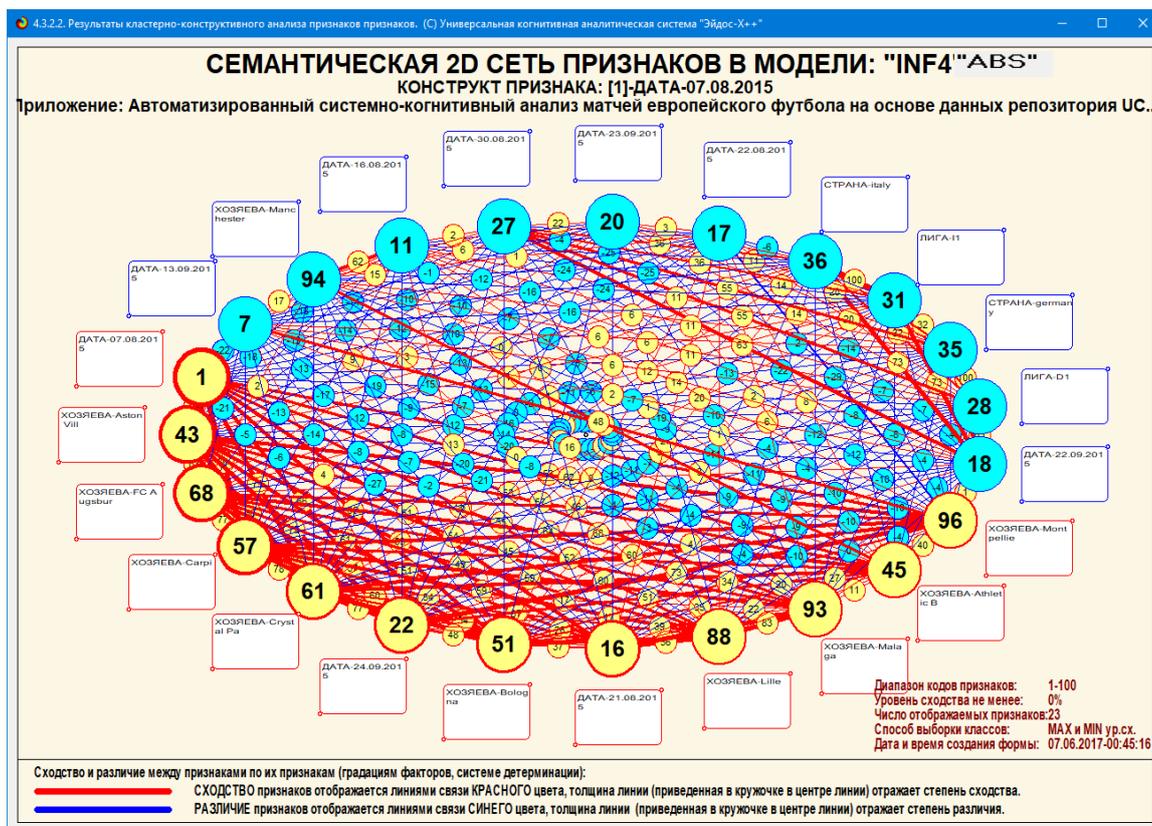


Рисунок 19 – Семантическая 2D сеть классов

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 был показан пример использования общедоступной базы данных «Футбольные события» для построения модели поведения и предпочтения игроков. При этом наиболее достоверной в данном приложении оказалась модель ABS, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Семантический резонанс». Точность модели составляет 0,889, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 49,4%.

При рассмотрении модели в перспективе ретроанализа можно убедиться в достоверности моделей: история изменений контента игры подтверждает выводы, получаемые в процессе выполнения задач курсовой работы.

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

СПИСОК ИСПОЛЬЗУЕМЫХ 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. 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. 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.

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.

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

6. Официальный сайт World of Warcraft [Электронный ресурс]. Режим доступа: <http://eu.battle.net/wow/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- X++» / 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>