

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ
ФЕДЕРАЦИИ

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

Факультет заочного обучения
Кафедра компьютерных технологий и систем

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
к курсовой работе

по дисциплине: СИСТЕМЫ ПРЕДСТАВЛЕНИЯ И ПРИОБРЕТЕНИЯ
ЗНАНИЙ
на тему:

«Автоматизированный системно-когнитивный анализ данных по
здравоохранению больших городов на основе данных портала Kaggle»
выполнил студент группы ИТ31841 Нехай Вячеслав Вячеславович

Допущен к защите _____

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

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

Защищена _____ Оценка_____

(дата)

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

Краснодар

2019 г.

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

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

Факультет заочного обучения
Кафедра компьютерных технологий и систем

УТВЕРЖДАЮ:

Зав. кафедрой КТС _____ В. И. Лойко

**ЗАДАНИЕ
на курсовую работу**

Студенту: ИТ31841 группы 2 курса

Факультета заочного

Специальности: 09.04.02 Информационные системы и технологии
(шифр)

Нехай Вячеславу Вячеславовичу
(Ф.И.О.)

Тема проекта: Автоматизированный системно-когнитивный анализ данных по здравоохранению больших городов на основе данных портала Kaggle

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 “” _____ 2019 г.

Срок защиты: “” _____ 2019 г.

Дата выдачи задания:

Дата сдачи проекта на кафедру: “” _____ 2019 г.

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

(подпись, Ф.И.О., звание, степень)

Задание принял студент _____ “” _____ 2019 г.

(подпись, дата)

Краснодар

2019 г.

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

Факультет заочного обучения

РЕЦЕНЗИЯ
на курсовой проект (работу)

Студента (ки) Нехай Вячеслав Вячеславович
курса 2 заочной формы обучения
направления подготовки Информационные системы и технологии
направленность(профиль) Информационные системы и технологии в науке и управлении

Наименование темы «Автоматизированный системно-когнитивный анализ данных по здравоохранению больших городов на основе данных портала Kaggle».

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
«_____» 2019 г.

РЕФЕРАТ

Курсовая работа содержит: 29 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	13
1.5. Виды моделей системы «Aidos».....	14
1.5. Результаты верификации моделей	16
2. РЕШЕНИЕ ЗАДАЧ НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ.....	18
2.1. Решение задачи идентификации.....	18
2.2. Выбор модели и пакетное распознавание	18
2.3 Визуализация когнитивных функций	21
2.4 SWOT и PEST матрицы и диаграммы	22
2.5 Нелокальные нейронные сети и нейтроны.....	25
2.6 Кластерный и конструктивный анализ	26
Заключение	27
Список используемых 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 рассмотрено решение задачи по определению зависимости заболеваемости жителей больших городов от различных факторов.

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 «Big Cities Health Data» 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** возьмем набор данных «Big Cities Health Data» <https://www.kaggle.com/khoongweihao/big-cities-health-data>.
Данный файл содержит:

Indicator Category - Категория Индикаторов

Indicator - Указатель

Year - Год

Gender - Пол

Race/ Ethnicity - Раса / Этническая Принадлежность

Value - Значение

Place - Место

BCHC Requested Methodology - Bchc методология

Source - Источник

Methods - Методы

Notes - Примечания

В качестве разделителей используются запятые.

С помощью онлайн-конвертера <https://onlineconvertfree.com/ru/convert-format/csv-to-xls> трансформируем формат csv в формат xls.

Добавим в таблицу столбец A и назовем его Number.

Преобразуем таблицу и оставим в ней поля: Indicator Category, Place, Indicator, Year, Gender, Value. Остальные поля удалим как несущественные для анализа.

В итоге получается следующая таблица для работы:

Таблица 1 – Фрагмент обучающей выборки

Big_Cities_Health_Data_Inventory.xls

Number	Indicator Category	Place	Indicator	Year	Gender	Value
1	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2011	Both	60,9
2	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2013	Male	90,1
3	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2013	Female	36,6
4	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2013	Both	62,3
5	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2012	Male	96,4
6	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2012	Female	43,7
7	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2012	Both	68,7
8	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2011	Male	84,8
9	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2011	Both	62,1
10	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2013	Both	77,2
11	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2012	Both	80,3
12	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2011	Both	280,5
13	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2013	Both	166,5
14	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2012	Both	251,1
15	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2011	Both	62
16	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2013	Both	55,1
17	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are difficult as definitions can vary.	2012	Both	64,7
18	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2013	Both	14,2
19	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2013	Male	18,6
20	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2013	Female	10,4
21	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2012	Male	29,2
22	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2012	Female	13,6
23	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2012	Both	21,3

24	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2011	Male	26,2
25	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2011	Female	12
26	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2011	Both	19,7
27	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2013	Both	9,6
28	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2012	Both	18,2
29	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2011	Both	16,3
30	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2013	Both	19,2
31	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2012	Both	30,1
32	Behavioral Health/Substance Abuse	Atlanta (Fulton County), GA	Percent of Adults Who Binge Drank	2011	Both	20,4
33	Behavioral Health/Substance Abuse	Baltimore, MD	Percent of Adults Who Binge Drank	2010	Both	14,5
34	Behavioral Health/Substance Abuse	Baltimore, MD	Percent of Adults Who Binge Drank	2010	Female	9,7
35	Behavioral Health/Substance Abuse	Baltimore, MD	Percent of Adults Who Binge Drank	2010	Male	20,3
36	Behavioral Health/Substance Abuse	Baltimore, MD	Percent of Adults Who Binge Drank	2011	Both	21,3
37	Behavioral Health/Substance Abuse	Baltimore, MD	Percent of Adults Who Binge Drank	2011	Female	22,2
38	Behavioral Health/Substance Abuse	Baltimore, MD	Percent of Adults Who Binge Drank	2011	Male	20,3
39	Behavioral Health/Substance Abuse	Baltimore, MD	Percent of Adults Who Binge Drank	2012	Both	18,8

Классификационные шкалы следующие: Категория Индикаторов, Место. Указатель, Год, Пол, Значение – описательные.

1.3. Ввод выборки в систему Aidos-X

Теперь, когда мы имеем обучающую выборку в формате *.xls, 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.

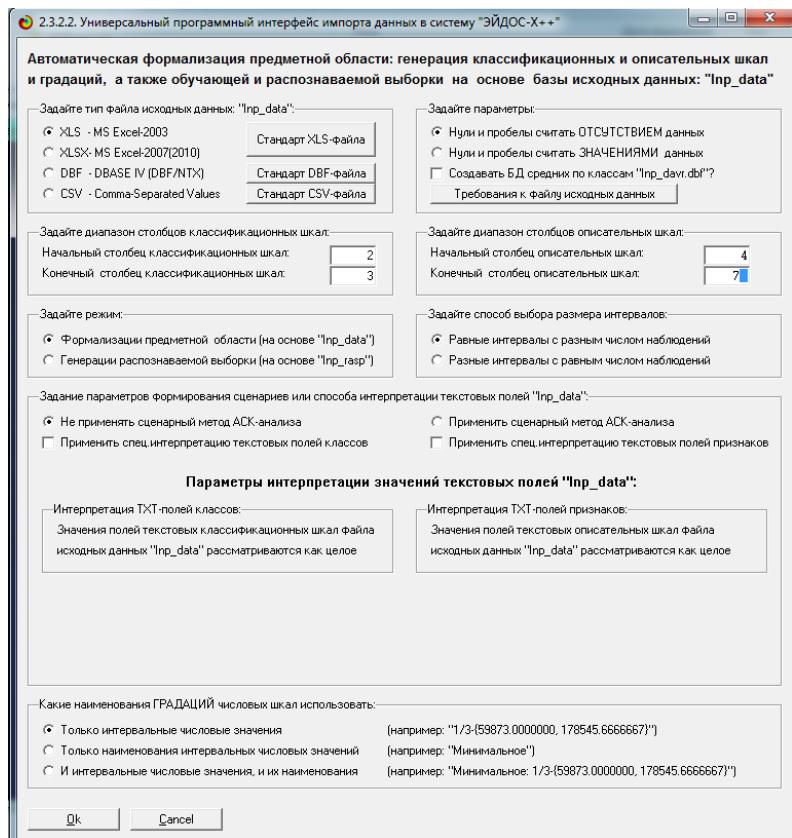


Рисунок 1 – Ввод обучающей выборки

Стоит отметить следующие настройки:

- Тип файла – xls.
- Классификационные шкалы – 2-3.
- Описательные шкалы – 4-7.
- Не применяется спец. интерпретация текстовых полей классов.

Так как предварительно числовые шкалы были разбиты на интервалы так, как это необходимо для исследования, пересчет 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).

2.1. Классификационные шкалы и градации. Текущая модель: "INF1"	
Код шкалы	Наименование классификационной шкалы
Код градации	Наименование градации классификационной шкалы
1	INDICATOR CATEGORY
2	PLACE
1	Behavioral Health/Substance Abuse
2	Cancer
3	Demographics
4	Food Safety
5	HIV/AIDS
6	Infectious Disease
7	Injury and Violence
8	Life Expectancy and Death Rate (Overall)
9	Maternal and Child Health
10	Nutrition, Physical Activity, & Obesity
11	Tobacco

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.

Код шкалы	Наименование описательной шкалы	Код градации	Наименование градации описательной шкалы
1	INDICATOR	1	AIDS Diagnoses Rate (Per 100,000 people)
2	YEAR	2	All-Cause Mortality Rate (Age-Adjusted; Per 100,000 people)
3	GENDER	3	All Types of Cancer Mortality Rate (Age-Adjusted; Per 100,000 people)
4	VALUE	4	Diabetes Mortality Rate (Age-Adjusted; Per 100,000 people)
		5	Drug Abuse-Related Hospitalization Rate (Per 100,000 people) *Comparisons of these data are ...
		6	Female Breast Cancer Mortality Rate (Age-Adjusted; Per 100,000 people)
		7	Firearm Related Mortality Rate (Age-Adjusted; Per 100,000 people)
		8	Heart Disease Mortality Rate (Age-Adjusted; Per 100,000 people)
		9	HIV-Related Mortality Rate (Age-Adjusted; Per 100,000 people)
		10	HIV Diagnoses Rate (Per 100,000 people)
		11	Homicide Rate (Age-Adjusted; Per 100,000 people)
		12	Infant Mortality Rate (Per 1,000 live births)
		13	Life Expectancy at Birth (Years)
		14	Lung Cancer Mortality Rate (Age-Adjusted; Per 100,000 people)
		15	Median Household Income (Dollars)
		16	Motor Vehicle Mortality Rate (Age-Adjusted; Per 100,000 people)
		17	Opioid-Related Mortality Rate (Age-Adjusted; Per 100,000 people) *These data should not be co... 18 Opioid-Related Mortality Rate (Age-Adjusted; Per 100,000 people) *These data should not be co...
		19	Opioid-Related Mortality Rate (Crude Rate; Per 100,000 people) *These data should not be com...
		20	Percent 18+ High School Graduates
		21	Percent Below 200% Poverty Level
		22	Percent Foreign Born

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.

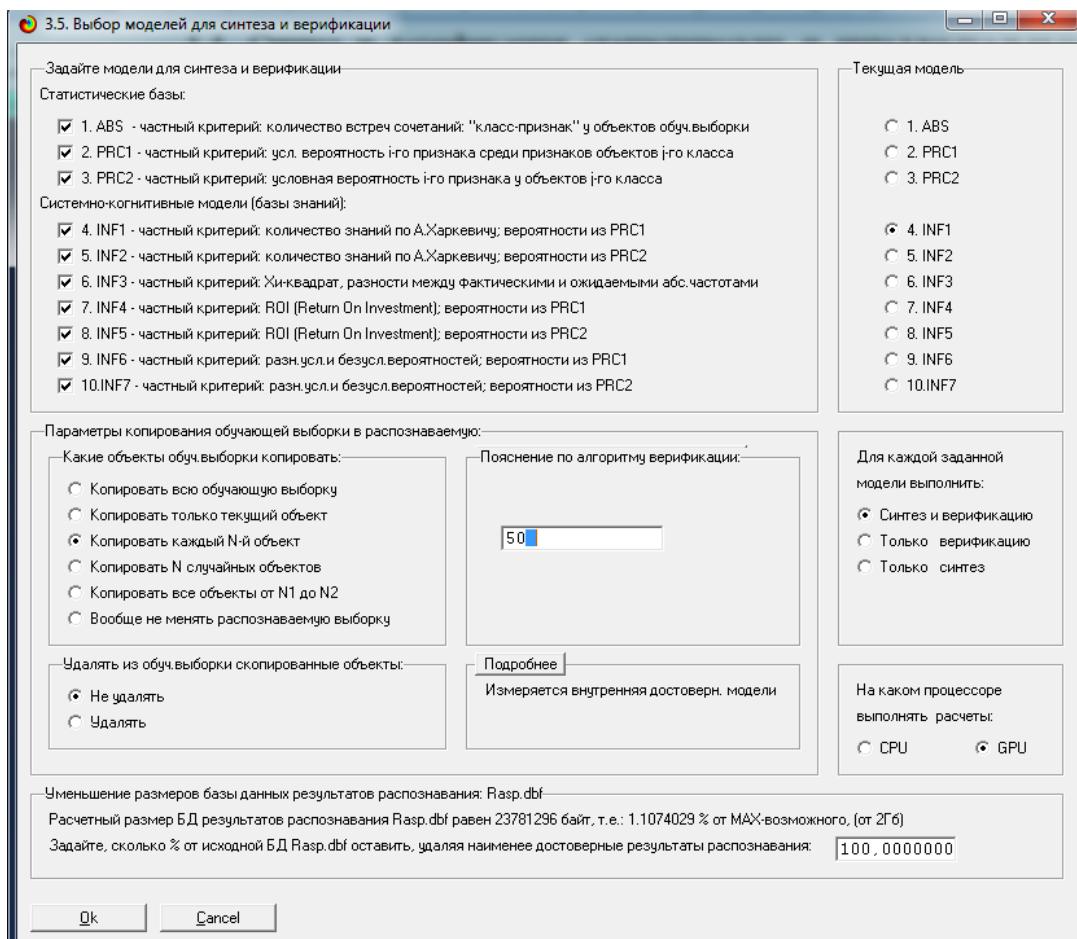


Рисунок 5 – Настройка синтеза и верификации моделей

Так как в выборке около 13500 значений, для распознавания будем использовать каждый 50-й объект. Из-за сортировки таблицы по названию болезни это будет надежнее, чем использование рандомных записей.

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

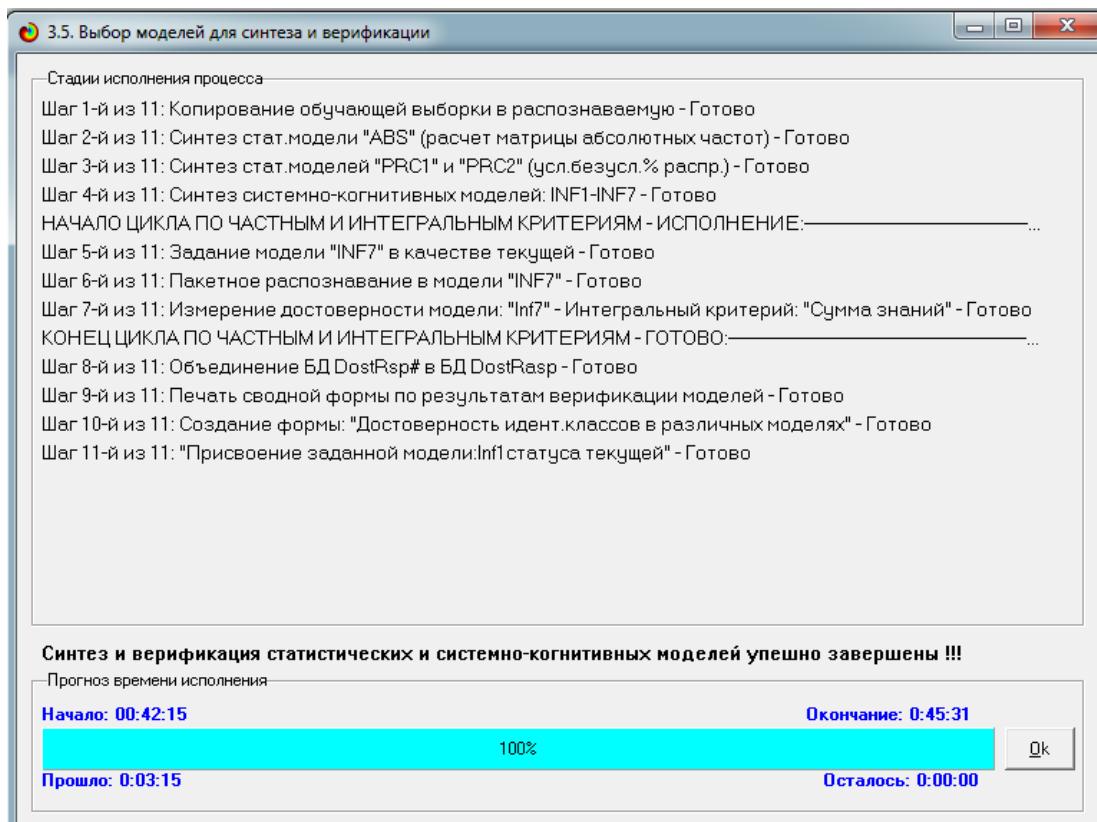


Рисунок 6 – Завершение процесса синтеза и верификации моделей

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 и условных и безусловных процентных распределений (фрагменты)

Код признака	Наименование описательной шкалы и градации	1. INDICATOR CATEGORY BEHAVIORAL HEALTH/SUBSTANCE ABUSE	2. INDICATOR CATEGORY CANCER	3. INDICATOR CATEGORY DEMOGRAPHICS	4. INDICATOR CATEGORY FOOD SAFETY	5. INDICATOR CATEGORY HIV/AIDS	6. INDICATOR CATEGORY INFECTIOUS DISEASE	7. INDICATOR CATEGORY INJURY AND VIOLENCE	8. INDICATOR CATEGORY LIFE EXPECTANCY AND DEATH RATE (OVERALL)	9. INDICATOR CATEGORY MATERNAL AND CHILD HEALTH	10. INDICATOR CATEGORY NUTRITION, PHYSICAL ACTIVITY, & OBESITY	11. INDICATOR CATEGORY TOBACCO
1	INDICATOR-AIDS Diagnoses Rate (Per 100,000 people...)							510				
2	INDICATOR-All-Cause Mortality Rate (Age-Adjusted; Per...								389			
3	INDICATOR-All Types of Cancer Mortality Rate (Age-Adj...											538
4	INDICATOR-Diabetes Mortality Rate (Age-Adjusted; Per ...)											
5	INDICATOR-Drug Abuse-Related Hospitalization Rate (...)	272										
6	INDICATOR-Female Breast Cancer Mortality Rate (Age-Ad...			336								
7	INDICATOR-Firearm Related Mortality Rate (Age-Adjusted;...								440			
8	INDICATOR-Heart Disease Mortality Rate (Age-Adjusted; ...)											554
9	INDICATOR-HIV-Related Mortality Rate (Age-Adjusted; ...)							444				
10	INDICATOR-HIV Diagnoses Rate (Per 100,000 people) ...							600				
11	INDICATOR-Homicide Rate (Age-Adjusted; Per 100,000 ...)								467			
12	INDICATOR-Infant Mortality Rate (Per 1,000 live births) ...											385
13	INDICATOR-Life Expectancy at Birth (Years) ...									155		
14	INDICATOR-Lung Cancer Mortality Rate (Age-Adjusted; ...)			524								
15	INDICATOR-Median Household Income (Dollars) ...							52				
16	INDICATOR-Motor Vehicle Mortality Rate (Age-Adjusted; ...)											
17	INDICATOR-Opioid-Related Mortality Rate (Age-Adj...)	180										
18	INDICATOR-Opioid-Related Mortality Rate (Age-Adj...)		23									
19	INDICATOR-Opioid-Related Mortality Rate (Code-Booked)	0										

Таблица 3 – Матрица информативностей в битах

Код признака	Наименование описательной шкалы и градации	1. INDICATOR CATEGORY BEHAVIORAL HEALTH/SUBSTANCE ABUSE	2. INDICATOR CATEGORY CANCER	3. INDICATOR CATEGORY DEMOGRAPHICS	4. INDICATOR CATEGORY FOOD SAFETY	5. INDICATOR CATEGORY HIV/AIDS	6. INDICATOR CATEGORY INFECTIOUS DISEASE	7. INDICATOR CATEGORY INJURY AND VIOLENCE	8. INDICATOR CATEGORY LIFE EXPECTANCY AND DEATH RATE (OVERALL)	9. INDICATOR CATEGORY MATERNAL AND CHILD HEALTH	10. INDICATOR CATEGORY NUTRITION, PHYSICAL ACTIVITY, & OBESITY	11. INDICATOR CATEGORY TOBACCO
1	INDICATOR-AIDS Diagnoses Rate (Per 100,000 people...)							0.892				
2	INDICATOR-All-Cause Mortality Rate (Age-Adjusted; Per...									1.562		
3	INDICATOR-All Types of Cancer Mortality Rate (Age-Adj...			1.096								0.973
4	INDICATOR-Diabetes Mortality Rate (Age-Adjusted; Per ...)											
5	INDICATOR-Drug Abuse-Related Hospitalization Rate (...)	1.280										
6	INDICATOR-Female Breast Cancer Mortality Rate (...)		1.096									
7	INDICATOR-Firearm Related Mortality Rate (Age-Adj...)								0.954			
8	INDICATOR-Heart Disease Mortality Rate (Age-Adj...)											0.973
9	INDICATOR-HIV-Related Mortality Rate (Age-Adjusted; ...)							0.892				
10	INDICATOR-HIV Diagnoses Rate (Per 100,000 people) ...							0.892				
11	INDICATOR-Homicide Rate (Age-Adjusted; Per 100,000 ...)								0.954			
12	INDICATOR-Infant Mortality Rate (Per 1,000 live bir...)											1.135
13	INDICATOR-Life Expectancy at Birth (Years) ...									1.562		
14	INDICATOR-Lung Cancer Mortality Rate (Age-Adj...)		1.096									
15	INDICATOR-Median Household Income (Dollars) ...				1.606							
16	INDICATOR-Motor Vehicle Mortality Rate (Age-Adjusted; ...)								0.954			
17	INDICATOR-Opioid-Related Mortality Rate (Age-Adj...)	1.280										
18	INDICATOR-Opioid-Related Mortality Rate (Age-Adj...)	1.280										
19	INDICATOR-Opioid-Related Mortality Rate (Code-Booked)	1.700										

Таблица 4 – Матрица знаний

Код признака	Наименование описательной шкалы и градации	1. INDICATOR CATEGORY BEHAVIORAL HEALTH/SUBSTANCE ABUSE	2. INDICATOR CATEGORY CANCER	3. INDICATOR CATEGORY DEMOGRAPHICS	4. INDICATOR CATEGORY FOOD SAFETY	5. INDICATOR CATEGORY HIV/AIDS	6. INDICATOR CATEGORY INFECTIOUS DISEASE	7. INDICATOR CATEGORY INJURY AND VIOLENCE	8. INDICATOR CATEGORY LIFE EXPECTANCY AND DEATH RATE (OVERALL)	9. INDICATOR CATEGORY MATERNAL AND CHILD HEALTH	10. INDICATOR CATEGORY NUTRITION, PHYSICAL ACTIVITY, & OBESITY	11. INDICATOR CATEGORY TOBACCO
1	INDICATOR-AIDS Diagnoses Rate (Per 100,000 people...)	-37.074	-54.008	-19.008	-32.963	427.895	-56.044	-72.262	-20.800	-49.897	-69.546	
2	INDICATOR-All-Cause Mortality Rate (Age-Adjusted; Per...	-28.278	-41.194	-14.499	-25.142	-62.625	-42.748	-55.117	373.135	-38.059	-53.046	
3	INDICATOR-All Types of Cancer Mortality Rate (Age-Adj...	-41.581	511.427	-21.319	-37.970	-92.087	-62.858	-81.047	-23.328	-55.963	-78.001	
4	INDICATOR-Diabetes Mortality Rate (Age-Adjusted; Per ...)	-39.109	-56.973	-20.052	-34.773	-86.613	-59.121	-76.229	-21.942	-52.636	464.635	
5	INDICATOR-Drug Abuse-Related Hospitalization Rate (...)	252.227	-28.804	-10.138	-17.580	-43.790	-29.890	-38.540	-11.093	-26.612	-37.091	
6	INDICATOR-Female Breast Cancer Mortality Rate (...)	-24.425	300.418	-12.523	-21.717	-54.093	-36.923	-47.608	-13.703	-32.873	-45.819	
7	INDICATOR-Firearm Related Mortality Rate (Age-Adj...)	-31.985	-46.595	-16.399	-28.439	-70.836	-48.352	377.656	-17.945	-43.048	-60.001	
8	INDICATOR-Heart Disease Mortality Rate (Age-Adj...)	-40.272	-58.667	-20.648	-35.807	-89.189	-60.880	-78.496	-22.594	-54.202	478.454	
9	INDICATOR-HIV-Related Mortality Rate (Age-Adjusted; ...)	-32.276	-47.019	-16.548	-28.697	372.520	-48.792	-62.910	-18.108	-43.440	-60.546	
10	INDICATOR-HIV Diagnoses Rate (Per 100,000 people) ...	-43.616	-63.539	-22.363	-30.780	503.405	-65.935	-85.014	-24.470	-58.702	-81.819	
11	INDICATOR-Homicide Rate (Age-Adjusted; Per 100,000 ...)	-33.948	-49.454	-17.406	-30.184	-75.183	-51.319	400.831	-19.046	-45.690	-63.683	
12	INDICATOR-Infant Mortality Rate (Per 1,000 live bir...)	-27.987	-40.771	-14.349	-28.884	-61.982	-42.308	-54.551	-15.702	347.333	-52.501	
13	INDICATOR-Life Expectancy at Birth (Years) ...	-11.268	-16.414	-5.777	-10.018	-24.954	-17.033	-21.962	148.678	-15.165	-21.137	
14	INDICATOR-Lung Cancer Mortality Rate (Age-Adj...)	-38.091	468.510	-19.530	-33.868	-84.359	-57.583	-74.245	-21.371	-51.267	-71.455	
15	INDICATOR-Median Household Income (Dollars) ...	-3.780	-5.507	50.062	-3.361	-8.372	-5.714	-7.368	-2.121	-5.088	-7.091	
16	INDICATOR-Motor Vehicle Mortality Rate (Age-Adjusted; ...)	-35.911	-52.313	-18.412	-31.929	-79.530	-54.286	424.005	-20.147	-48.331	-67.364	
17	INDICATOR-Opioid-Related Mortality Rate (Age-Adj...)	166.915	-19.062	-6.709	-11.634	-28.978	-19.780	-25.504	-7.341	-17.611	-24.546	
18	INDICATOR-Opioid-Related Mortality Rate (Age-Adj...)	21.328	-2.436	-0.857	-1.487	-3.703	-2.527	-3.259	-0.938	-2.250	-3.136	
19	INDICATOR-Opioid-Related Mortality Rate (Code-Booked)	99.021	10.144	2.570	4.205	15.455	10.550	12.602	2.915	9.292	12.601	

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.крит. Текущая модель: "INF1"													
Наименование модели и частного критерия	Интегральный критерий	- як	Точность	Полнота	F-мера	Сумма модул. уровней сход. истинно-полоз. решений (STP)	Сумма модул. уровней сход. истинно-отриц. решений (STN)	Сумма модул. уровней сход. ложно-полоз. решений (SFP)	S-Точность	S-Полнота	L1-мера проф. Е.В.Луценко	Средн. уровни истин.	
1. ABS - частный критерий: количество встреч сочтений "класс...".	Корреляция abs. частот с обр...	0.050	1.000	0.096	346.253	0.074	5473.802		0.059	1.000	0.112		
1. ABS - частный критерий: количество встреч сочтений "класс...".	Сумма abs. частот по признакам	0.050	1.000	0.096	192.035			1733.669		0.100	1.000	0.181	
2. PRC1 - частный критерий: усл. вероятность иго признака сред...	Корреляция усл.отн частот с о...	0.050	1.000	0.096	346.235	0.074	5473.496		0.059	1.000	0.112		
2. PRC1 - частный критерий: усл. вероятность иго признака сред...	Сумма усл.отн частот по приз...	0.050	1.000	0.096	279.317			4169.400		0.063	1.000	0.118	
3. PRC2 - частный критерий: условная вероятность иго признака...	Корреляция усл.отн частот с о...	0.050	1.000	0.096	346.277	0.074	5474.170		0.059	1.000	0.112		
3. PRC2 - частный критерий: условная вероятность иго признака...	Сумма усл.отн частот по приз...	0.050	1.000	0.096	275.763			4116.381		0.063	1.000	0.118	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	0.077	0.826	0.141	199.472	448.200	501.509	9.636	0.285	0.954	0.438		
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	0.072	0.828	0.133	176.687	343.891	370.220	7.100	0.323	0.961	0.484		
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	0.075	0.828	0.137	199.453	448.811	500.998	9.645	0.285	0.954	0.439		
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	0.072	0.828	0.133	176.647	343.923	370.553	7.093	0.323	0.961	0.493		
6. INF3 - частный критерий: Универсал, различия между фактами...	Семантический резонанс зна...		0.096	0.798	0.171	169.314	1015.727	823.404	22.301	0.171	0.884	0.286	
6. INF3 - частный критерий: Универсал, различия между фактами...	Сумма знаний		0.096	0.798	0.171	161.678	261.799	111.718	1.597	0.576	0.950	0.728	
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.085	0.822	0.154	163.792	151.295	335.921	3.465	0.328	0.979	0.491		
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.072	0.828	0.132	93.662	29.202	70.304	0.654	0.571	0.993	0.725		
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	0.085	0.824	0.154	163.752	152.036	335.068	3.485	0.328	0.979	0.492		
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	0.072	0.828	0.133	93.798	29.238	70.404	0.653	0.571	0.993	0.725		
9. INF6 - частный критерий: разн.усл.бездл.вероятностей; вер...	Семантический резонанс зна...	0.073	0.798	0.133	160.316	783.991	926.646	22.032	0.147	0.879	0.253		
9. INF6 - частный критерий: разн.усл.бездл.вероятностей; вер...	Сумма знаний	0.073	0.798	0.133	119.945	185.679	281.838	3.073	0.299	0.975	0.457		
10. INF7 - частный критерий: разн.усл.бездл.вероятностей; вер...	Семантический резонанс зна...	0.073	0.794	0.133	160.102	780.561	924.304	21.887	0.148	0.880	0.253		
10. INF7 - частный критерий: разн.усл.бездл.вероятностей; вер...	Сумма знаний	0.073	0.794	0.133	118.120	182.828	278.472	3.012	0.298	0.975	0.456		

Рисунок 6 – Результаты верификации моделей

Наиболее достоверной в данном приложении оказалась модель INF3 при интегральном критерии «Сумма знаний», если учитывать L-1 меру проф. Е.В. Луценко. Точность модели составляет 0,728, 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 INF3 (рисунок 7).

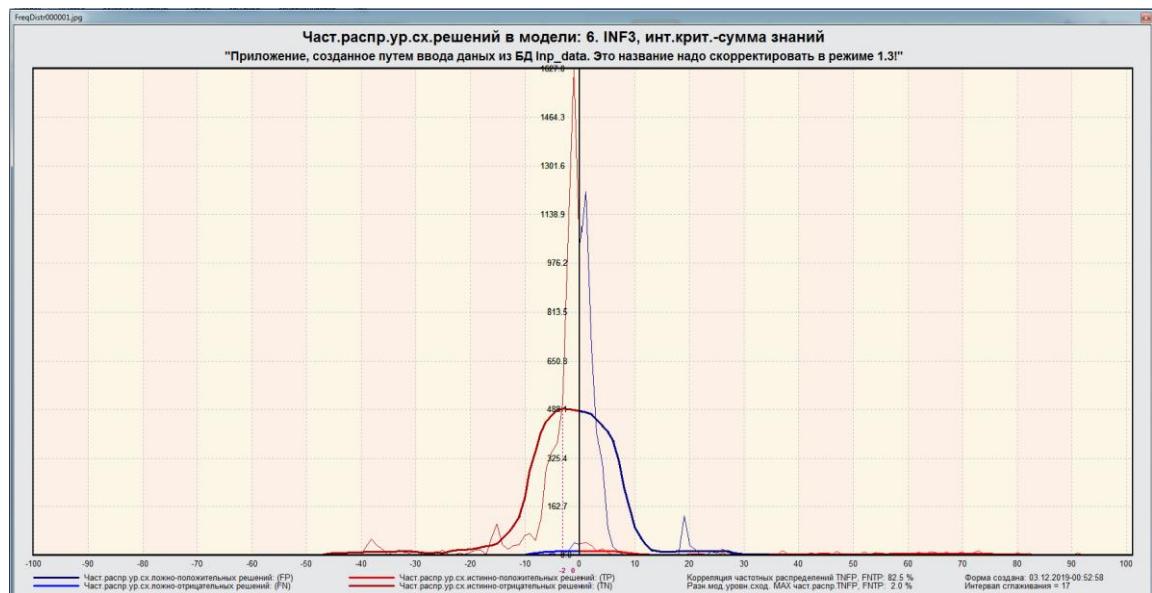


Рисунок 7 – Частотное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF3

Как можно видеть, при модели INF3 непринадлежность объекта к классу лучше определяется, чем принадлежность. При этом существует вероятность ложного определения непренаадлежности или принадлежности.

Выберем модель INF3 в качестве основной и перейдем к решению задач идентификации.

2.РЕШЕНИЕ ЗАДАЧ НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи идентификации

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

Во-первых, зависимость заболеваний от места проживания. Анализ данного параметра позволит наиболее оптимально распределять финансовые и профессиональные ресурсы среди медицинских учреждений.

Во-вторых, можно прогнозировать продолжительность жизни и смертность в зависимости от территории нахождения пациента.

2.2. Выбор модели и пакетное распознавание

Зададим модель INF3 в качестве текущей в режиме 5.6 (рисунок 8), после чего проведем пакетное распознавание в режиме 4.1.2 (рисунок 9).

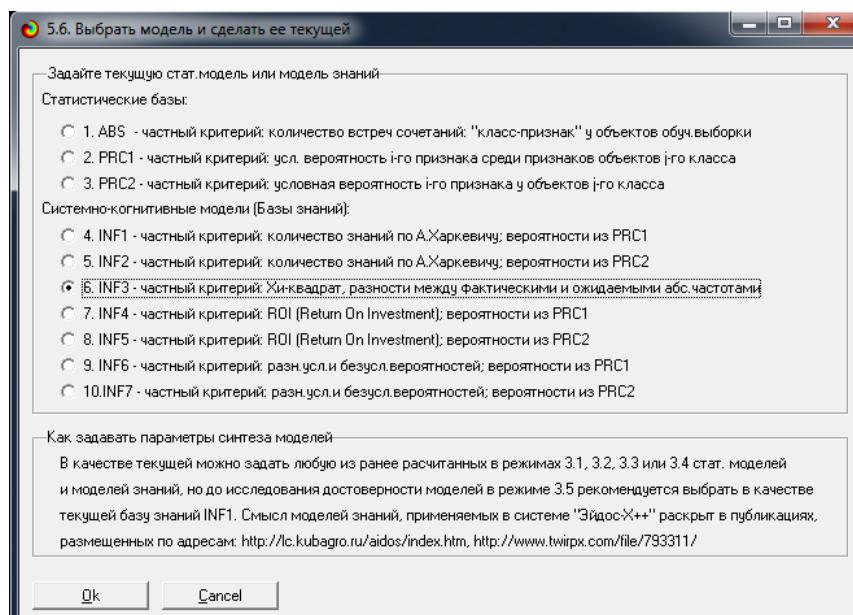


Рисунок 8 – Выбор текущей модели

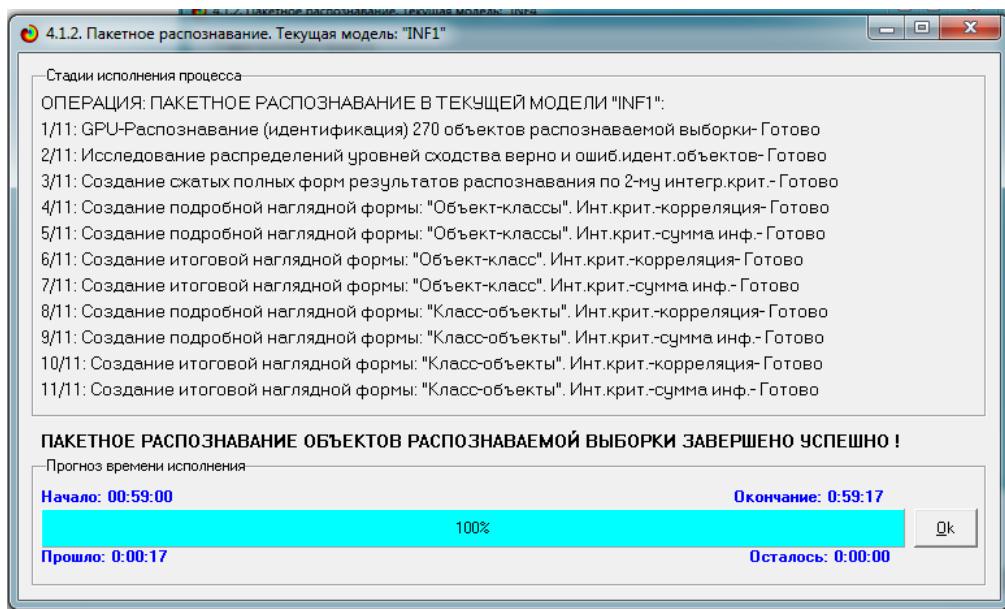


Рисунок 9 – Пакетное распознавание в новой модели

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

Режим 4.1.3 системы «Aidos-X» 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 10 и 11 приведены примеры прогнозов высокой и низкой достоверности классов, а также итоги распознавания классов объектов в модели INF3:

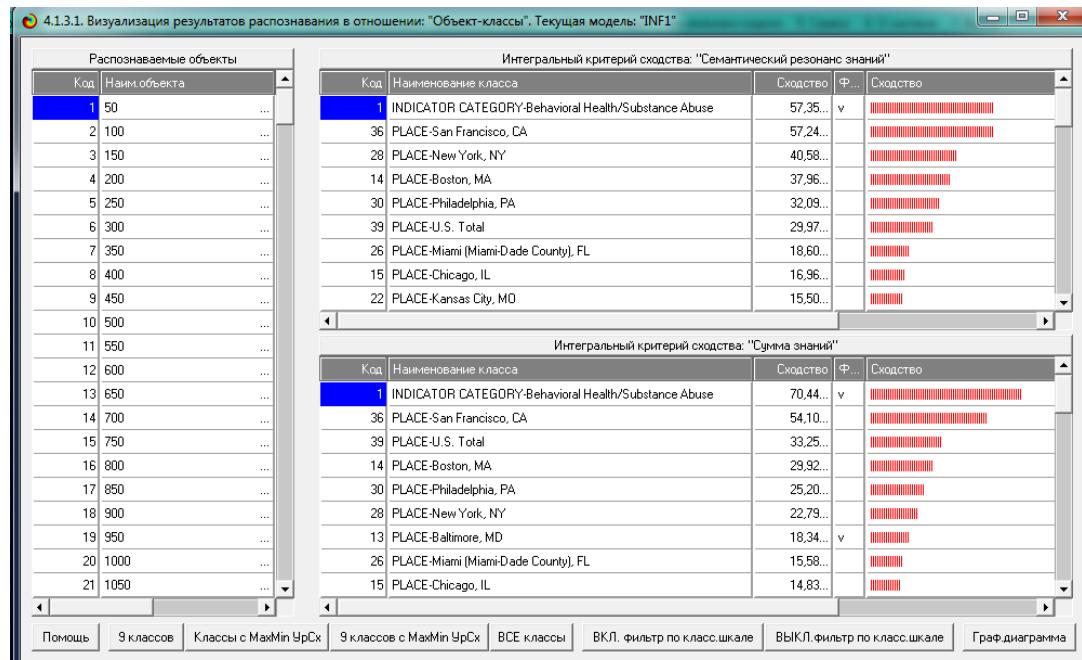


Рисунок 10 – Прогнозы достоверности классов объектов

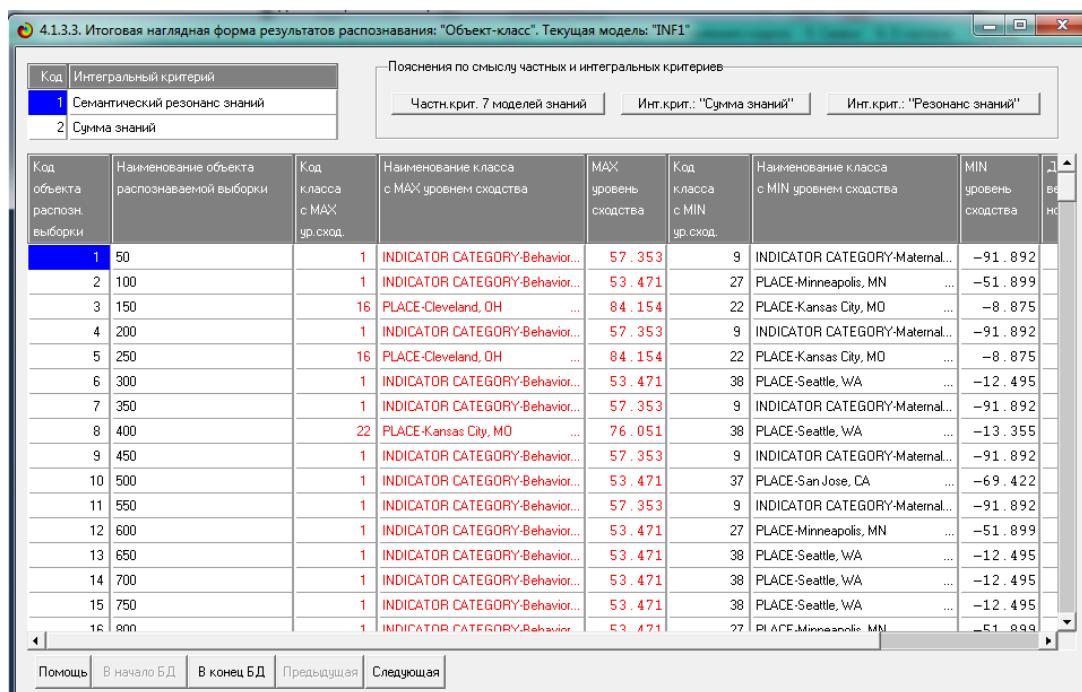


Рисунок 11 – Итоги распознавания классов объектов

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.

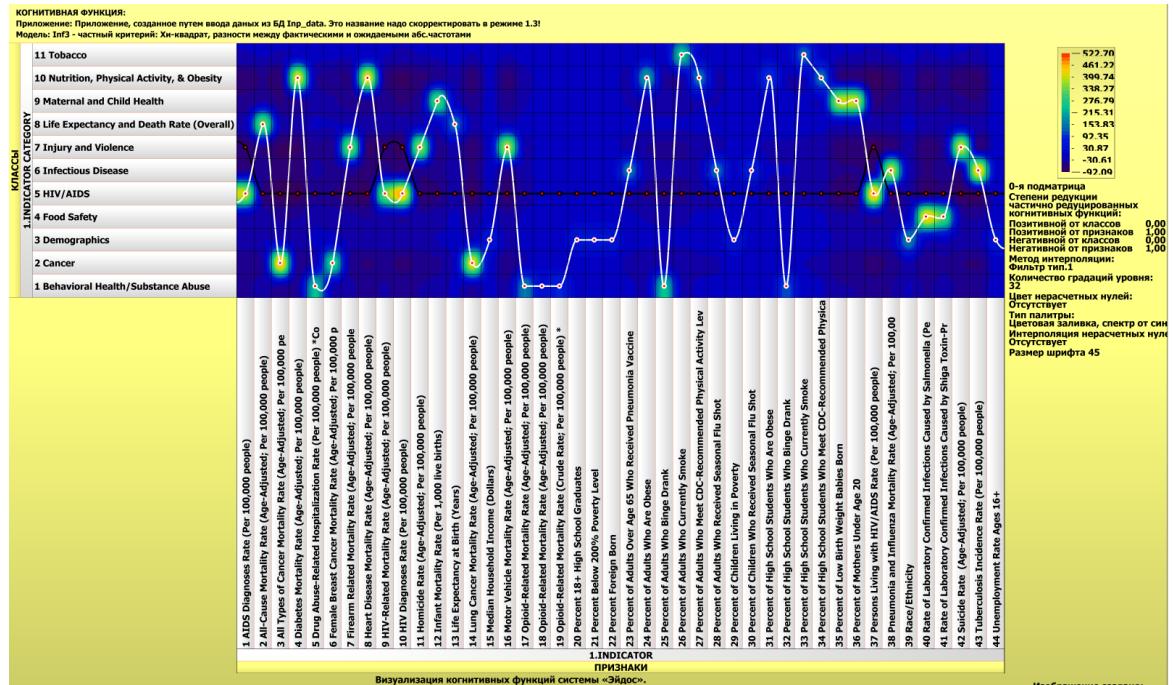


Рисунок 12 – Визуализация зависимости заболеваний от категории

зарегистрированных

Рассмотрим визуализацию когнитивных функций в рамках определенных задач.

Визуализация показывает нам, что наибольшее количество заболеваний связано с табакокурением, причем зарегистрированы обращения как взрослых людей, так и студентов высших учебных заведений. Так же очень высок процент заболеваний связанных с питанием, физической активностью и ожирением.

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).

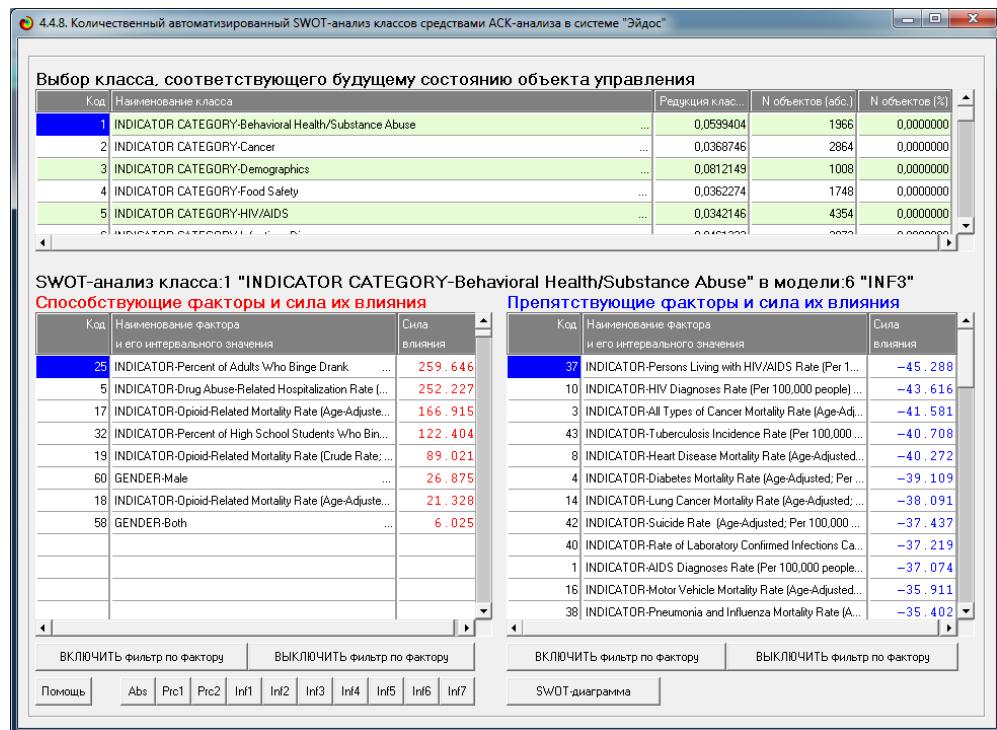


Рисунок 13 – Пример SWOT-Матрицы в модели INF3

На рисунке 14 приведен пример инвертированной SWOT-матрицы.

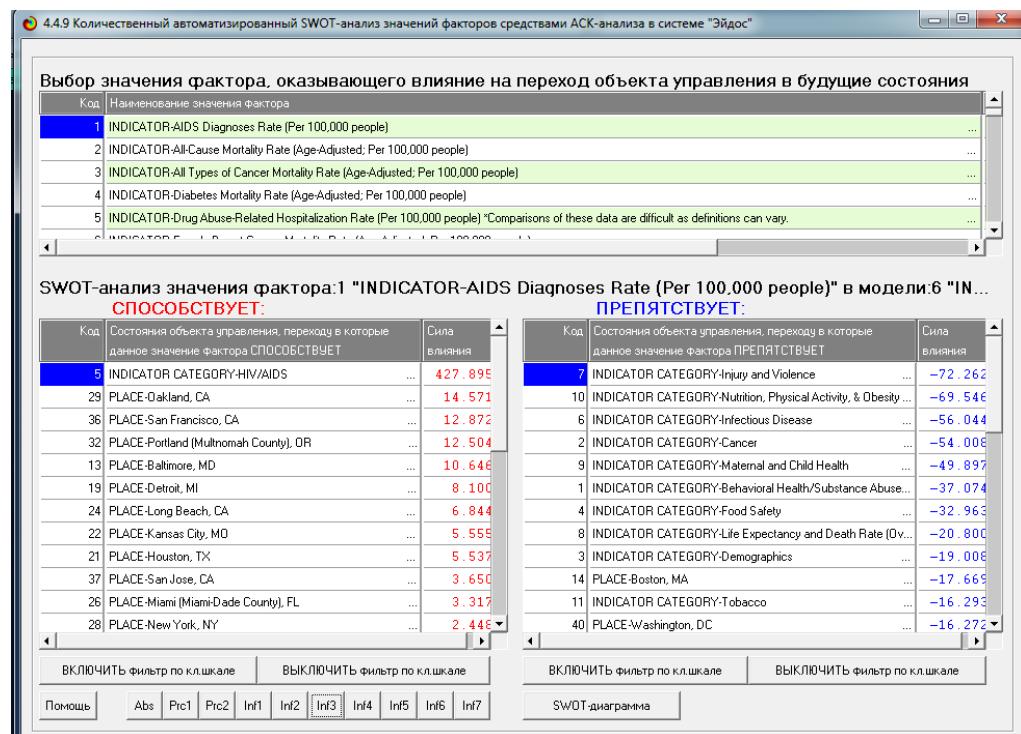


Рисунок 14 – Пример инвертированной SWOT-Матрицы в модели INF3

А теперь рассмотрим у каких категорий людей чаще всего встречается заболевания (рисунок 15):

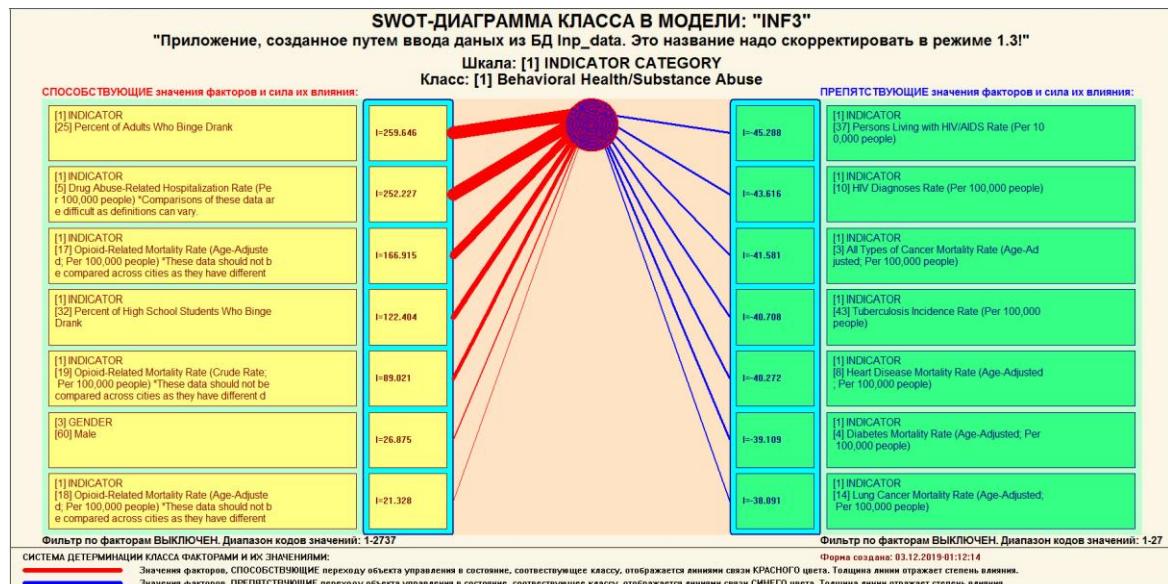


Рисунок 15 –SWOT-матрица

Как можно видеть, большая часть проблем наблюдается у людей с лишним весом и у людей, которые употребляют наркотики.

А теперь узнаем, в каких городах наибольший коэффициент смертности (рисунок 20).

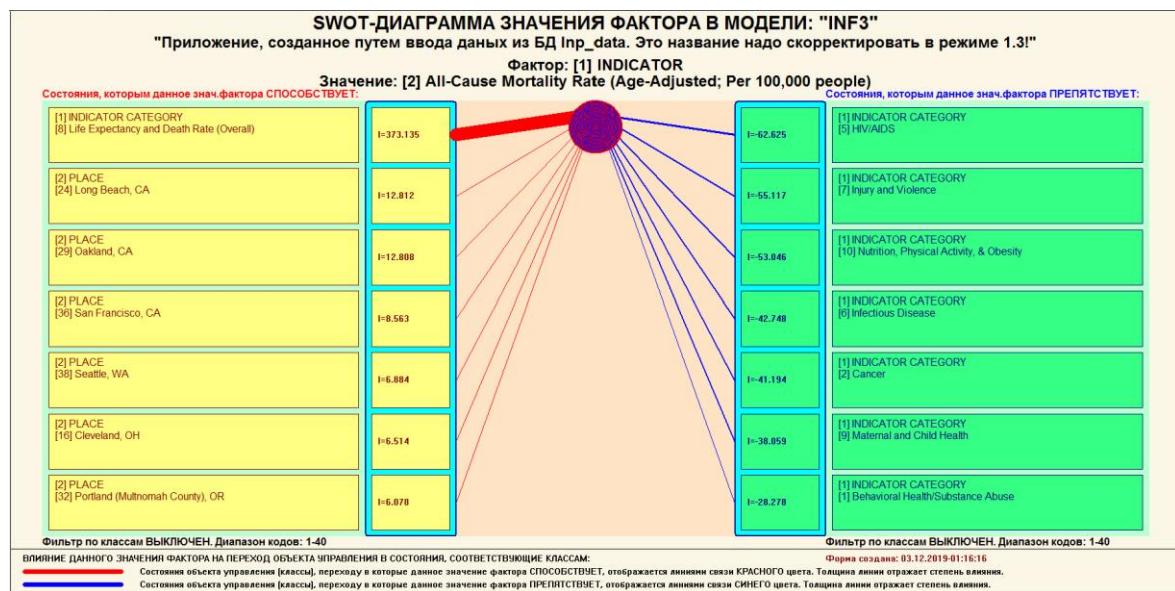


Рисунок 16 – SWOT-диаграмма коэффициента смертности в зависимости от места проживания

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 модели INF3.

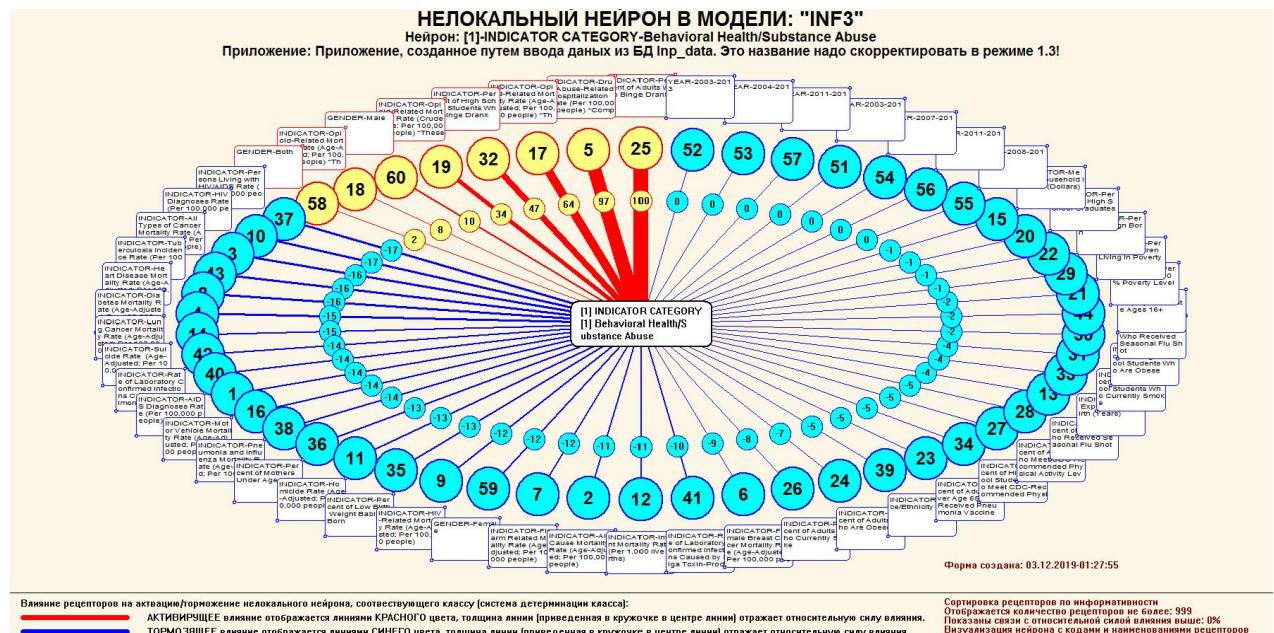


Рисунок 17 – Нейрон для модели INF3

На его основании можно определить правильность всех сделанных ранее выводов.

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

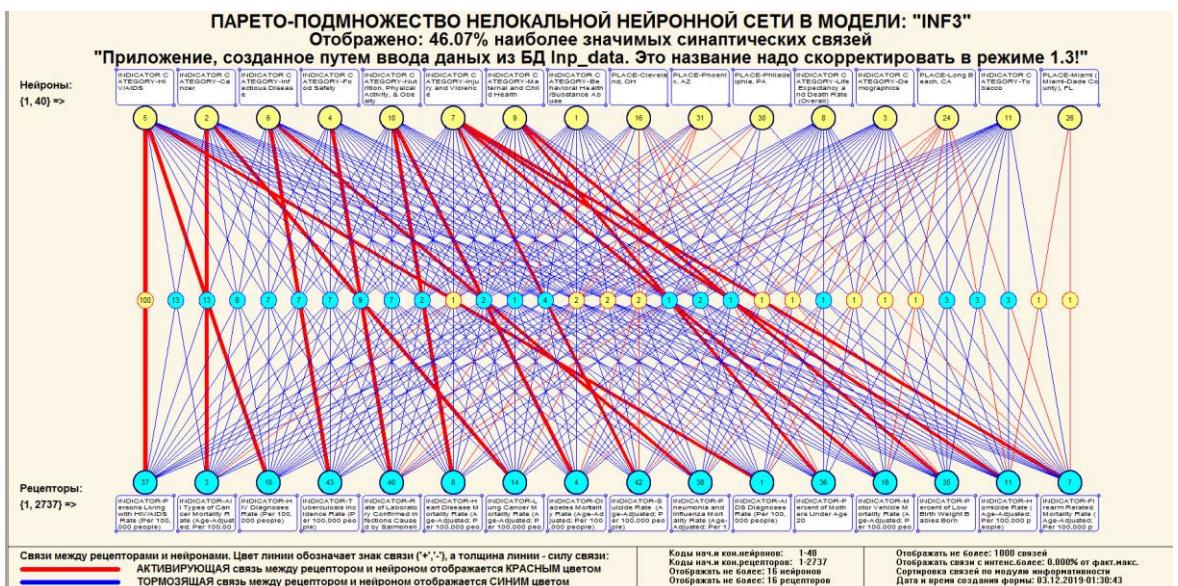


Рисунок 18 – Пример нелокальной нейронной сети в модели INF3

По ней можно, в основном, определить зависимость заболеваемости от места проживания и возрастных факторов.

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).

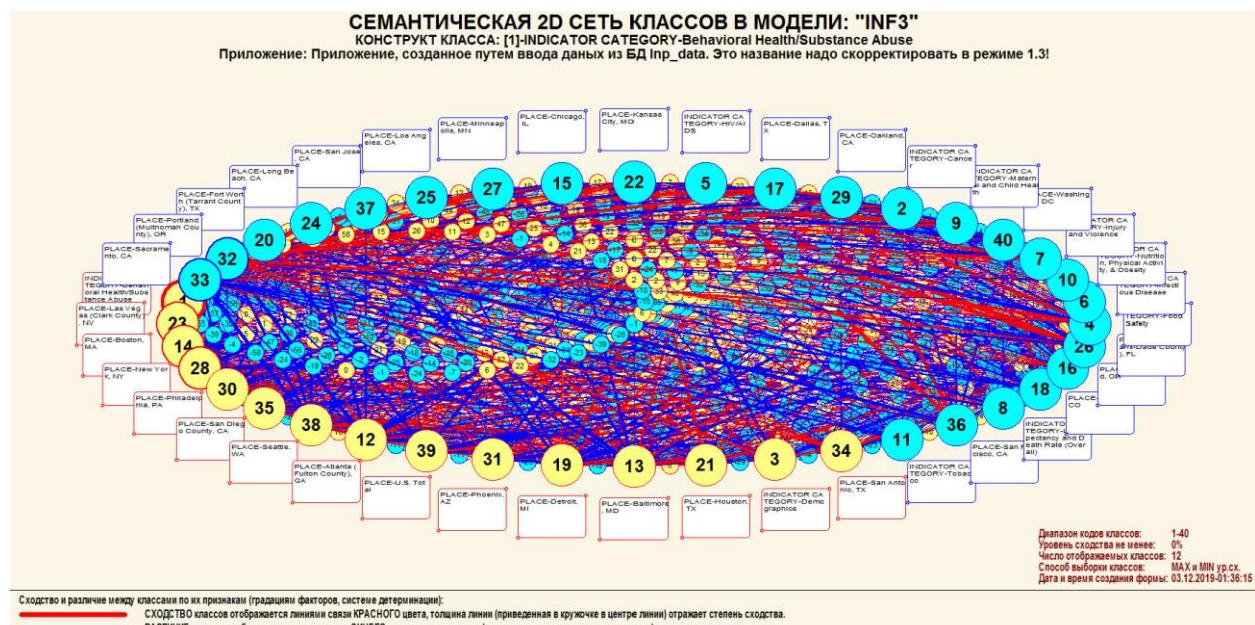


Рисунок 19 – Семантическая 2D сеть классов

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 «Данные по здравоохранению больших городов» для построения модели зависимости заболеваемости в больших городах от различных факторов. При этом наиболее достоверной в данном приложении оказалась модель INF3, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сумма знаний». Точность модели составляет 0,728, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 49,4%.

При рассмотрении модели в перспективе ретроанализа можно убедиться в достоверности моделей: анализ заболеваемости жителей больших городов в зависимости от различных факторов подтверждает выводы, получаемые в процессе выполнения задач курсовой работы.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Луценко Е.В. Методика использования репозитория UCI для оценки качества математических моделей систем искусственного интеллекта / Е.В. Луценко // Полitemатический 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 [articleID]: 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 // Полitemатический 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 автомати-зированного системно-когнитивного анализа и возможности его применения для сопоставимой оценки эффективности вузов / Е.В. Луценко, В.Е. Коржаков // Полitemатический сетевой электронный научный журнал Кубанского государственного аграрно- го 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 "bazadannixsotrudnikov": <http://allexcel.ru/gotovye-tablitsy-excel-besplatno>, свободный. - Загл. с экрана. Яз.анг.

6. Официальный сайт WorldofWarcraft [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 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>