МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ имени И.Т. ТРУБИЛИНА»

Факультет прикладной информатики Кафедра компьютерных технологий и систем

КУРСОВАЯ РАБОТА

по дисциплин	е: <u>Интелло</u>	ектуальные си	стемы и тех	нологии	[
на тему: <u>АС</u>	К-анализ	грамотности	населения	Индии	на	основе	данных
портала Kaggl	l <u>e</u>						
Выполнил сту	дент груп	пы: <u>ИТ1941 М</u>	Іоскаленко д	Алексей	Cep	геевич	
Допущен к заг	щите:						
Руководитель	проекта: Д	Ц.Э.Н., К. Т. Н., I	профессор Л	Гуценко	Е.В.	.(<i>\(\(\)</i>	
				(подпи	ись, ра	асшифровья	подписи)
						·	
Защищен	_16/02/202	21					
Опенка	ОТПИЧНО						

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 курсовую работу

Студента Москаленко Алексея Сергеевича курса 2 очной формы обучения группы ИТ1941 Направление подготовки 09.04.02 «Информационные системы и технологии» Наименование темы «АСК-анализ грамотности населения Индии на основе

Рецензент: <u>Луценко Евгений Вениаминович, д.э.н., к.т.н., профессор</u> (Ф.И.О., ученое звание и степень, должность)

данных портала Kaggle»

Оценка качества выполнения курсовой работы

№ п/п	Показатель	Оценка соответствия («неудовлетворительно», «удовлетворительно», «хорошо», «отлично»)
1	Степень полноты обзора состояния проблемы и корректность постановки цели и задач исследования	5
2	Уровень и корректность использования в работе различных методов исследований	5
3	Степень комплексности работы, применения в ней знаний общепрофессиональных и специальных дисциплин	5
4	Ясность, четкость, последовательность и обоснованность изложения	5
5	Применение современных технологий обработки информации	5
6	Качество оформления работы (общий уровень грамотности, стиль изложения, качество иллюстраций, соответствие требованиям по оформлению)	5
7	Уровень освоения компетенций в результате выполнения курсовой работы (проекта)	5
8	Ответы на вопросы при защите	5

Цостоинства работы
Недостатки работы
Итоговая оценка при защите отлично
Рецензент (Е.В. Луценко)
(16 » февраля 2021 г.

РЕФЕРАТ

Курсовая работа содержит: 36 страниц, 29 рисунков, 17 литературных источников.

Ключевые слова: СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, AIDOS-X, КОГНИТИВНЫЙ АНАЛИЗ, МОДЕЛИ, ШКАЛЫ, КЛАССЫ.

Целью работы провести АСК-анализ грамотности населения Индии по на основе данных портала Kaggle. В данной курсовой работе необходимо 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.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ	6
1.1. Описание решения	6
1.2. ПРЕОБРАЗОВАНИЕ ИСХОДНЫХ ДАННЫХ ИЗ CSV-ФОРМАТ	А В ФАЙЛ
ИСХОДНЫХ ДАННЫХ MS EXCEL	8
1.3. Ввод выборки в систему Aidos-X	9
1.4. Синтез и верификация статистических и интеллек	ГУАЛЬНЫХ
МОДЕЛЕЙ	13
1.5. Виды моделей системы Aidos-X	15
1.6. Результаты верификации моделей	17
2. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ	
модели	20
2.1. Решение задачи идентификации	20
2.2. Кластерно-конструктивный анализ	24
2.3. НЕЛОКАЛЬНЫЕ НЕЙРОННЫЕ СЕТИ И НЕЙТРОНЫ	26
2.4. SWOT и PEST матрицы и диаграммы	29
ЗАКЛЮЧЕНИЕ	33
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ	35

ВВЕДЕНИЕ

Одним ИЗ важных перспективных направлений развития И современных информационных технологий является создание систем искусственного интеллекта. Существует множество альтернатив систем искусственного интеллекта, но возникает необходимость оценки качества математических моделей этих систем. В данной работе рассмотрено решение задачи АСК-анализ грамотности населения Индии по на основе данных портала Kaggle. 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 «Govt Of India Literacy Rate» из банка исходных данных по задачам искусственного интеллекта – Kaggle.

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

1. СИНТЕЗ И ВЕРИФИКАЦИЯ МОДЕЛЕЙ

1.1. Описание решения

В качестве метода исследования, решения проблемы и достижения цели было решено использовать новый метод искусственного интеллекта: Автоматизированный системно-когнитивный анализ или АСК-анализ.

Основной причиной выбора АСК-анализа является то, что он включает теорию и метод количественного выявления в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал и единицах измерения.

Очень важным является также то, что АСК-анализ имеет свой развитый и доступный программный инструментарий, в качестве которого в настоящее время выступает Универсальная когнитивная аналитическая система Aidos-X.

Преимущества данной системы:

- универсальность;
- доступность, то есть данная система находится в полном открытом бесплатном доступе, причем с актуальными исходными текстам;
- одна из первых отечественных систем искусственного интеллекта персонального уровня, т.е. она не требует от пользователя специальной подготовки в области технологий искусственного интеллекта (акт внедрения системы Aidos-X 1987 года);
- обеспечивает устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных взаимозависимых (нелинейных) данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и числовых) и в различных единицах измерения;

- содержит большое количество локальных и облачных учебных и научных приложений (в настоящее время их 31 и 251, соответственно);
 - мультиязычная поддержка интерфейса (больше 50 языков);
- поддерживает online среду накопления знаний и широко используется во всем мире;
- наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решение этих задач в несколько тысяч раз;
- обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение задач классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе;
- хорошо имитирует человеческий стиль мышления: дает результаты анализа, понятные экспертам на основе их опыта, интуиции и профессиональной компетенции.

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-X.
 - 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 pecypca kaggle.com возьмем набор данных «Govt Of India Literacy Rate»

https://www.kaggle.com/doncorleone92/govt-of-india-literacy-rate

Csv файл содержит следующие данные:

Country/ States/ Union Territories Name Наименование территории.

Category Категория региона Индии.

Literacy Rate (Persons) - Total – 2001 Общий коэффициент грамотности за 2001 год.

Literacy Rate (Persons) - Total – 2011 Общий коэффициент грамотности за 2011 год.

Literacy Rate (Persons) - Rural — 2001 Сельский коэффициент грамотности за 2001 год.

Literacy Rate (Persons) - Rural — 2011 Сельский коэффициент грамотности за 2011 год.

Literacy Rate (Persons) - Urban — 2001 Городской коэффициент грамотности за 2001 год.

Literacy Rate (Persons) - Urban — 2011 Городской коэффициент грамотности за 2011 год.

Для конвертации csv-файла в xls был использован онлайн конвертер: https://document.online-convert.com/ru/convert/csv-to-excel

После конвертации необходимо добавить еще один столбец, в данном случае было решено дублировать столбец, содержащий имена участников. Также данный столбец был переименован в «Categoty» и размещен в таблице между столбцами «Country/ States/ Union Territories Name» и «Literacy Rate (Persons) - Total - 2001» (Рисунок 1).

4	Α	В	С	D	Е	F	G	Н	1	J	K	L
1	Category	Country/ States/ Uni	Category	Literacy Rat	Literacy R	Literacy Ra	Literacy R	Literacy R	Literacy Ra	te (Perso	ns) - Urban	- 2011
2	Country	INDIA	Country	64,8	73	58,7	67,8		84,1			
3	State	Andhra Pradesh	State	60,5	67	54,5	60,4	76,1	80,1			
4	State	Arunachal Pradesh	State	54,3	65,4	47,8	59,9	78,3	82,9			
5	State	Assam	State	63,3	72,2	59,7	69,3	85,3	88,5			
6	State	Bihar	State	47	61,8	43,9	59,8	71,9	76,9			
7	State	Chhattisgarh	State	64,7	70,3	60,5	66	80,6	84			
8	State	Goa	State	82	88,7	79,7	86,6	84,4	90			
9	State	Gujarat	State	69,1	78	61,3	71,7	81,8	86,3			
10	State	Haryana	State	67,9	75,6	63,2	71,4	79,2	83,1			
11	State	Himachal Pradesh	State	76,5	82,8	75,1	81,9	88,9	91,1			
12	State	Jammu & Kashmir	State	55,5	67,2	49,8	63,2	71,9	77,1			
13	State	Jharkhand	State	53,6	66,4	45,7	61,1	79,1	82,3			
14	State	Karnataka	State	66,6	75,4	59,3	68,7	80,6	85,8			
15	State	Kerala	State	90,9	94	90	93	93,2	95,1			
16	State	Madhya Pradesh	State	63,7	69,3	57,8	63,9	79,4	82,8			
17	State	Maharashtra	State	76,9	82,3	70,4	77	85,5	88,7			
18	State	Manipur	State	70,5	76,9	67,3	73,4	79,3	85,4			
19	State	Meghalaya	State	62,6	74,4	56,3	69,9	86,3	90,8			
20	State	Mizoram	State	88,8	91,3	81,3	84,1	96,1	97,6			
21	State	Nagaland	State	66,6	79,6	62,8	75,3	84,7	89,6			
22	State	Odisha	State	63,1	72,9	59,8	70,2	80,8	85,7			
23	State	Punjab	State	69,7	75,8	64,7	71,4	79,1	83,2			
24	State	Rajasthan	State	60,4	66,1	55,3	61,4	76,2	79,7			
25	State	Sikkim	State	68,8	81,4	66,8	78,9	83,9	88,7			
26	State	Tamil Nadu	State	73,5	80,1	66,2	73,5	82,5	87			
27	State	Tripura	State	73,2	87,2	69,7	84,9	89,2	93,5			
28	State	Uttar Pradesh	State	56,3	67,7	52,5	65,5	69,8	75,1			
29	State	Uttarakhand	State	71,6	78,8	68,1	76,3	81,4	84,5			
30	State	West Bengal	State	68,6	76,3	63,4	72,1	81,2	84,8			
31		A & N Islands	Union Ter	81,3	86,6	78,7	84,5	86,6	90,1			
32	Union Ter	Chandigarh	Union Ter	81,9	86	75,6	80,7	82,6	86,2			
33	Union Ter	D & N Haveli	Union Ter	57,6	76,2	49,3	64,1	84,4	89,8			
34	Union Ter	Daman & Diu	Union Ter		87,1	75,8	81,4	82,3	89			
35		Lakshadweep	Union Ter	86,7	91,8	85	91,6	88,6	91,9			
36	Union Ter	NCT of Delhi	Union Ter	,	86,2	78,1	81,9	81,9	86,3			
37	Union Ter	Puducherry	Union Ter	81,2	85,8	74	80,1	84,8	88,5			
20	I											

Рисунок 1 - Фрагмент обучающей выборки

Таким образом, в качестве классификационной шкалы решено использовать столбец 2. Описательные шкалы столбцы 3-9.

1.3. Ввод выборки в систему Aidos-X

Теперь, когда имеется обучающая выборка в формате *.xls, можно импортировать ее в систему Aidos-X.

Скопируем данную выборку в папку Inp_data и переименуем ее в Inp_data. Далее был использован универсальный программный интерфейс импорта данных в систему Aidos-X (режим 2.3.2.2), результат заполнения данной формы представлен на рисунке 2.

🕙 2.3.2.2. Универсальный программный интерфейс импорта д	данных в систему "ЭЙДОС-Х++" — 🖂 🗙						
Автоматическая формализация предметной области	и: генерация классификационных и описательных шкал						
и градаций, а также обучающей и распознаваемой г	выборки на основе базы исходных данных: "Inp_data"						
—Задайте тип файла исходных данных: "Inp_data":———————————————————————————————————	— Задайте параметры:						
© XIS . MS EveeL2003	С Нули и пробелы считать ОТСУТСТВИЕМ данных						
С XLSX-MS Excel-2007(2010)	 Нули и пробелы считать ЗНАЧЕНИЯМИ данных 						
C DBF - DBASE IV (DBF/NTX) Стандарт DBF-файла	☐ Создавать БД средних по классам "Inp_davr.dbf"?						
C CSV - CSV => DBF конвертер Стандарт CSV-файла	Требования к файлу исходных данных						
Задайте диапазон столбцов классификационных шкал:	Задайте диапазон столбцов описательных шкал:						
Начальный столбец классификационных шкал: 2	Начальный столбец описательных шкал:						
Конечный столбец классификационных шкал:	Конечный столбец описательных шкал: 9						
Задайте режим:	Задайте способ выбора размера интервалов:						
 Формализации предметной области (на основе "Inp_data") 	 Равные интервалы с разным числом наблюдений; 						
○ Генерации распознаваемой выборки (на основе "Inp_rasp")	С Разные интервалы с равным числом наблюдений						
Применить спец.интерпретацию текстовых полей классов Параметры интерпретации зн	Применить спец.интерпретацию текстовых полей признаков начений текстовых полей "Inp_data":						
—Интерпретация ТХТ-полей классов:-	Интерпретация ТХТ-полей признаков:						
Значения полей текстовых классификационных шкал файла	Значения полей текстовых описательных шкал файла						
исходных данных "Inp_data" рассматриваются как целое	исходных данных "Inp_data" рассматриваются как целое						
—Какие наименования ГРАДАЦИЙ числовых шкал использовать:—							
	(например: "1/3-{59873.0000000, 178545.6666667}")						
·	(например: "Минимальное")						
·	(например: "Минимальное: 1/3-{59873.0000000, 178545.6666667}")						
Ok Cancel							

Рисунок 2 - Ввод обучающей выборки

Стоит отметить следующие настройки:

- Тип файла xls;
- Классификационные шкалы 2;
- Описательные шкалы 3-9;

В форме задания размерности модели системы оставляем всё без изменений (Рисунок 3).



Рисунок 3 - Задание размерности модели системы Aidos-X

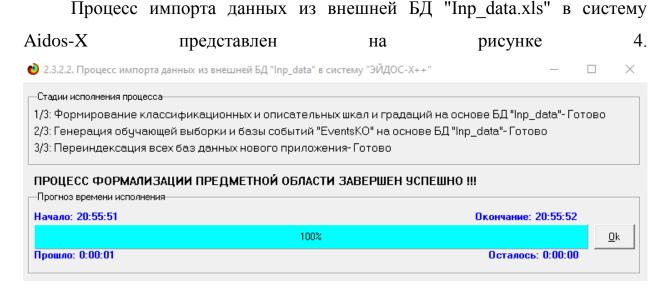


Рисунок 4 – Процесс импорта данных

Так как предварительно числовые шкалы были разбиты на интервалы, то перерасчет шкал после ввода выборки производить не надо. После импорта автоматически формируются классификационные и описательные применением которых исходные шкалы, данные кодируются представляются в форме эвентологических баз данных. Тем самым этап формализации предметной области выполняется полностью автоматизировано.

Классификационная шкала представлена на рисунке 5, её можно просмотреть в режиме 2.1. Описательные шкалы – в режиме 2.2 (рисунок 6).

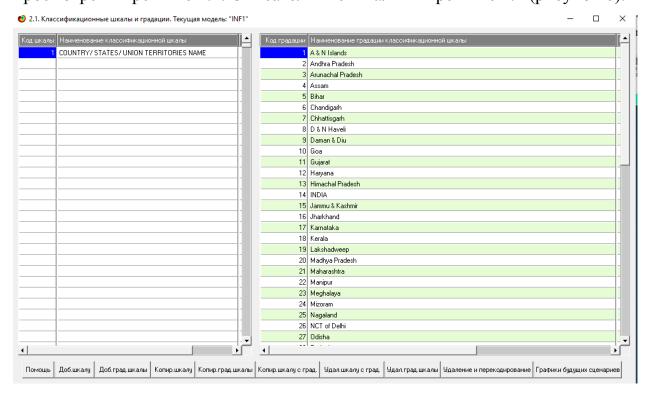


Рисунок 5 - Классификационные шкалы и градации

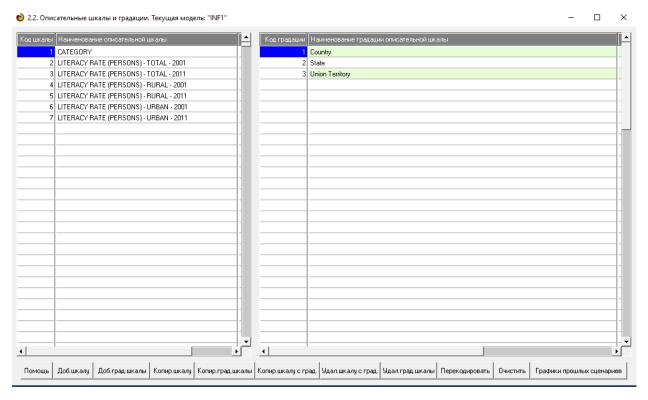


Рисунок 6 - Описательные шкалы и градации (фрагмент)

Для ручного ввода-корректировки обучающей выборки существует режим 2.3.1, он представлен на рисунке 7. Установка значений описательных и классификационных шкал объектов осуществляется по их номерам.

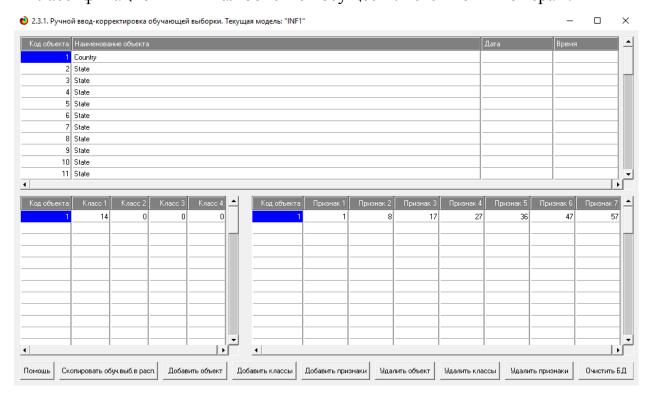


Рисунок 7 - Обучающая выборка (фрагмент)

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

1.4. 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 8).

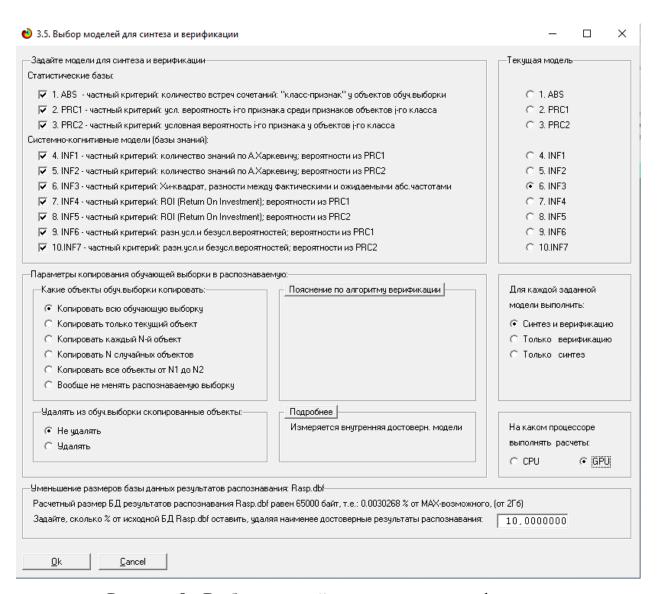


Рисунок 8 - Выбор моделей для синтеза и верификации

В данном режиме имеется много различных методов верификации моделей. Но мы используем параметры, приведенные на рисунке 8. Стадия процесса исполнения режима 3.5 и прогноз времени его окончания отображаются на экранной форме, приведенной на рисунке 8.

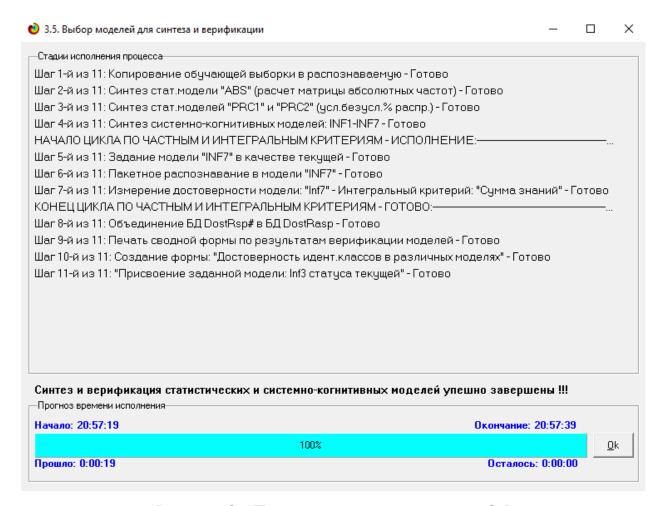


Рисунок 8 - Процесс исполнения режима 3.5

Интересно заметить (см. рисунок 8), что синтез и верификация всех 10 моделей на данной задаче заняли 19 секунд. При этом оценка достоверности моделей проводилась на 36 примерах наблюдения из обучающей выборки. После этого можно перейти непосредственно к выбору наиболее достоверной модели.

1.5. Виды моделей системы Aidos-X

Рассмотрим решение задачи идентификации на примере нескольких моделей, в которой рассчитано количество информации по А. Харкевичу, которое было получено по принадлежности идентифицируемого объекта к каждому из классов, если известно, что у этого объекта есть некоторый признак.

То есть частные критерии представляют собой просто формулы для преобразования матрицы абсолютных частот (рисунок 9) в матрицы условных и безусловных процентных распределений, и матрицы знаний (рисунок 10 и 11).

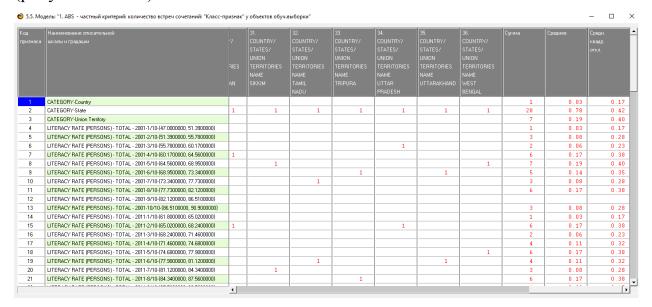


Рисунок 9 - Матрица абсолютных частот (модель ABS) и условных и безусловных процентных распределений

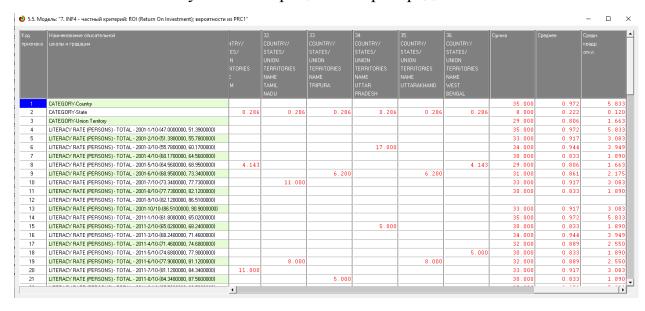


Рисунок 10 - Модель INF4 (фрагмент)

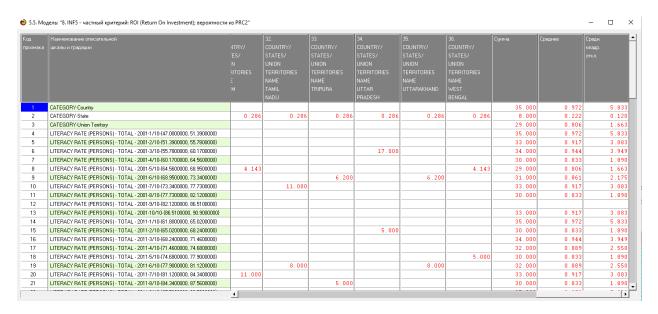


Рисунок 11 - Модель INF5 (фрагмент)

1.6. Результаты верификации моделей

Результаты верификации моделей, отличающихся частными критериями с двумя приведенными выше интегральными критериями представлены на рисунке 12.

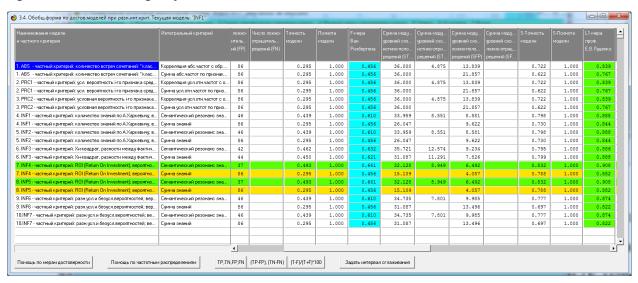


Рисунок 12 - Оценки достоверности моделей

Наиболее достоверной в данном приложении оказались модели INF5 и INF 4 при интегральном критерии «Семантический резонанс знаний», но далее будет рассматриваться только INF5. При этом точность модели (F-мера Ван Ризбергена) составляет 0,661 а точность модели (L1-мера профессора Луценко) - 0,908. L1-мера профессора Луценко является более достоверной,

по сравнению с F-мерой Ван Ризбергена. 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 Aidos-X используется F-мера Ван Ризбергена и L-мера, представляющая собой ее нечеткое мультиклассовое обобщение, предложенное профессором Е.В.Луценко (рисунок 13).

ѐ Помощь по режимам: 3.4, 4.1.3.#: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++" Помощь по режимам: 3.4, 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-Х++". положительный псевлопрогноз. Предположим, модель дает такой прогноз: выпадет 1, 2, 3, 4, 5 или 6. В этом случае у нее будет 100% достоверность идентификации, т.е. не будет ни одного объекта, не отнесенного к тому классу, к которому он действительно относится, но при этом будет очень большая ошибка ложной идентификации, т.к. огромное количество объектов будет отнесено к классам, к которым они не относятся (и именно за счет этого у модели и будет очень высокая достоверность идентификации). Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом. ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ. Представим себе, что мы выбрасываем кубик с 6 гранями, и модель предсказывает, что не выпадет: 1, 2, 3, 4, 5 и 6, а что-то из этого естественно выпало. Конечно, модель дает ошибку в прогнозе в том плане, что не предсказала, что выпадет, зато она очень хорошо угадала, что не выпадет. Но ясно, что выпадет что-то одно, а не все, что предсказано, поэтому такого рода предсказания хорошо оправдываются в том, что не произошло и плохо в том, что произошло, т.е. в этом случае у модели будет 100% достоверность не идентификации, но очень низкая достоверность идентификации. Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий. если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом. На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представите себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов. Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и Lb = (1 + (TP + TN - FP - FN))/(TP + TN + FP + FN))/2 (нормировка: $\{0,1\}$) где количество: ТР - истинно-положительных решений; ТN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений; Классическая F-мера достоверности моделей Ван Ризбергена (колонка выделена ярко-голубым фоном): F-mera = 2"(Precision*Recall)/(Precision+Recall) - достоверность модели Precision = TP/(TP+FP) - точность модели; Recall = TP/(TP+FN) - полнота модели; L1-мера проф. Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СУММ уровней сходства (колонка выделена ярко-зеленым фоном): L1-mera = 2"(SPrecision"SRecall)/(SPrecision+SRecall) SPrecision = STP/(STP+SFP) - точность с учетом сумм уровней сходства;

SRecall = STP/(STP+SFN) - полнота с учетом сумм уровней сходства;

STP - Сумма модулей сходства истинно-положительных решений; STN - Сумма модулей сходства истинно-отрицательных решений; SFP - Сумма модулей сходства ложно-положительных решений; SFN - Сумма модулей сходства ложно-отрицательных решений. L2-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желтым фоном); L2-мера проф. Е.В. тущенко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желт L2-мера = 2"(APrecision*ARecall)/(APrecision+ARecall) APrecision = ATP/(ATP+AFP) - точность с учетом средних уровней сходства; ARecall = ATP/(ATP+AFN) - полнота с учетом средних уровней сходства; ATP=STP/TP - Среднее модулей сходства истинно-положительных решений; AFN=SFN/FN - Среднее модулей сходства ложно-отрицательных решений; AFN=SFN/FN - Среднее модулей сходства ложно-отрицательных решений. Строки с максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке. Из графиков частотных распределений истинно-положительных, истинно-отрицательных и ложно-положительных и ложно-отрицательных решений видно, что чем выше модуль чровня сходства, тем больше доля истинных решений. Значит модуль уровня сходства является адекватной мерой степени истинности решения и уверенности системы в этом решении Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе "Эйдос" / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №02(126). С. 1 - 32. - IDA [article ID]: 1261702001. - Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf, 2 у.п.л.

Рисунок 13- Виды прогнозов и принцип определения достоверности моделей по авторскому варианту метрики, сходной с F-критерием

Также необходимо обратить внимание на то, 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 21 интеллектуальных технологий. На рисунке 14 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и не идентифицированных ситуаций в наиболее достоверной модели INF5.

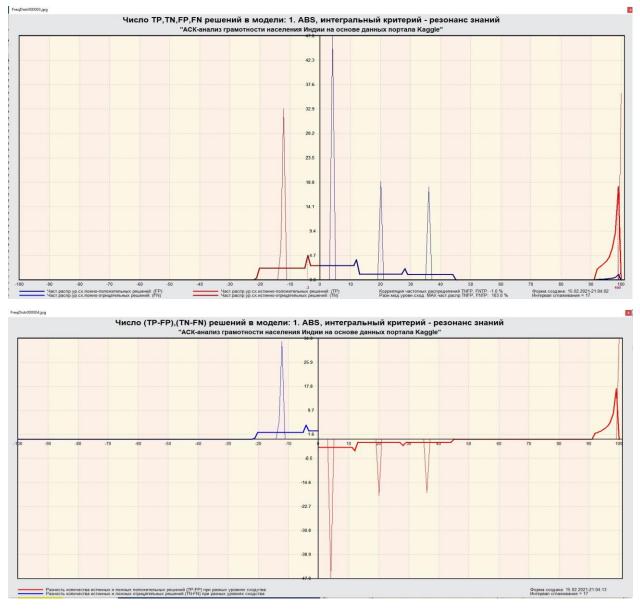


Рисунок 14 - Частотные распределения числа верно и ошибочно идентифицированных и не идентифицированных состояний объекта моделирования в зависимости от уровня сходства в модели INF5

Из рисунка 14 видно, что:

- наиболее точная модель INF5 лучше определяет непринадлежность объекта к классу, чем принадлежность;
- модуль уровня сходства-различия в модели INF5 для верно идентифицированных и верно не идентифицированных объектов значительно выше, чем для ошибочно идентифицированных и ошибочно не идентифицированных. Это верно практически для всего диапазона уровней сходства-различия, кроме небольших по модулю значений в диапазоне от 0 до 20% уровня сходства. Для уровней сходства-различия более 25% ошибочно идентифицированные и не идентифицированными ситуации практически отсутствует.

2. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1. Решение задачи идентификации

В соответствии с технологией АСК-анализа зададим текущей модель INF5 (режим 5.6) (рисунок 15) и проведем пакетное распознавание в режиме 4.2.1.

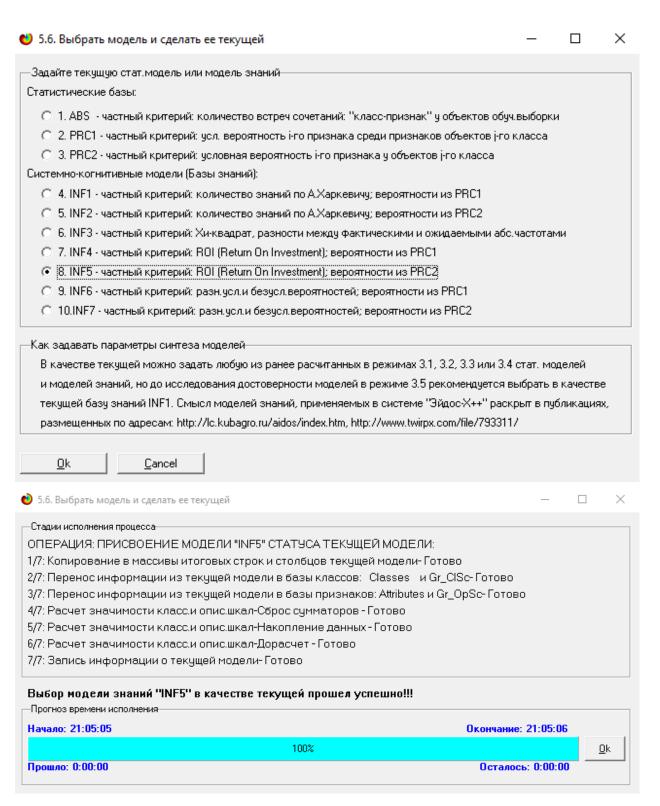


Рисунок 15 - Экранные формы режима задания модели в качестве текущей

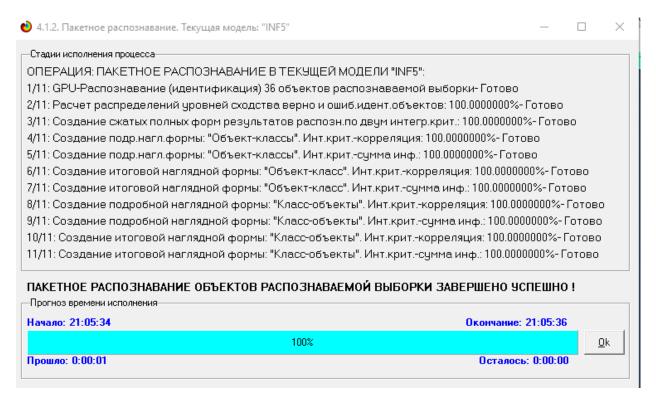


Рисунок 16 - Экранная форма режима пакетного распознавания в текущей модели

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

Режим 4.1.3 системы 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 17 и 18 приведены примеры прогнозов высокой и низкой достоверности частоты объектов и классов в модели INF5 на основе наблюдения предыстории их развития:

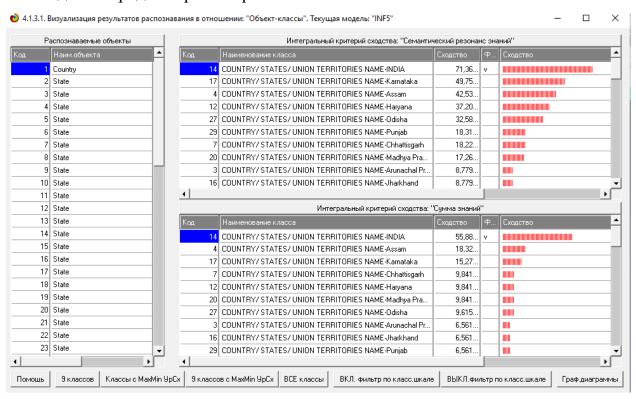


Рисунок 17. Пример идентификации объектов в модели INF5

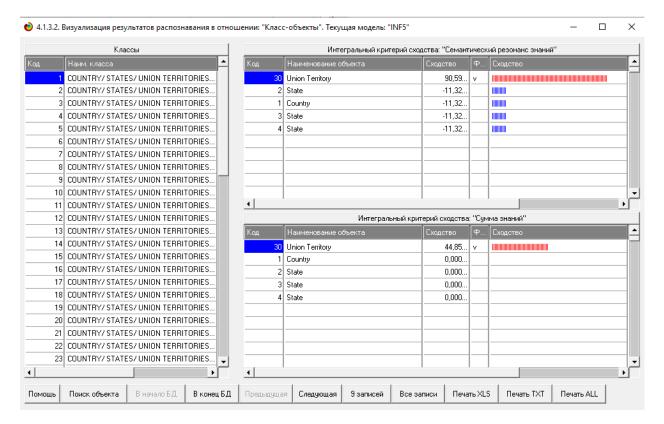


Рисунок 18. Пример идентификации классов в модели INF5

2.2. Кластерно-конструктивный анализ

Сходство-различие обобщенных образов различных результатов научной деятельности по характерным для них системам значений показателей. Результаты сравнения классов по системе характерных приведены на рисунке 19:

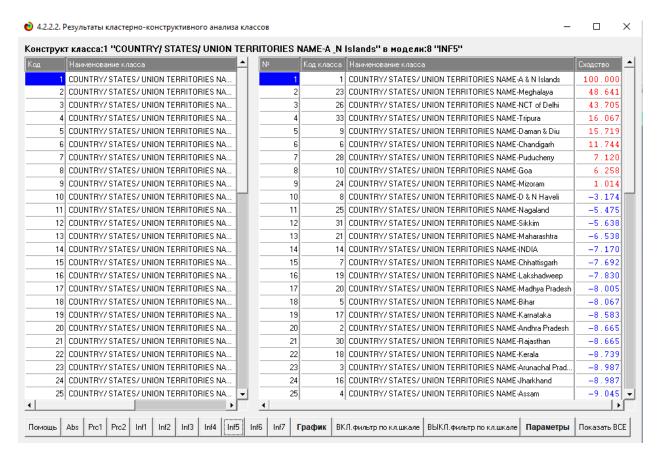


Рисунок 19 - Результаты кластерно-конструктивного анализа классов

На рисунке 20 представлена семантическая сеть классов в модели «INF5».

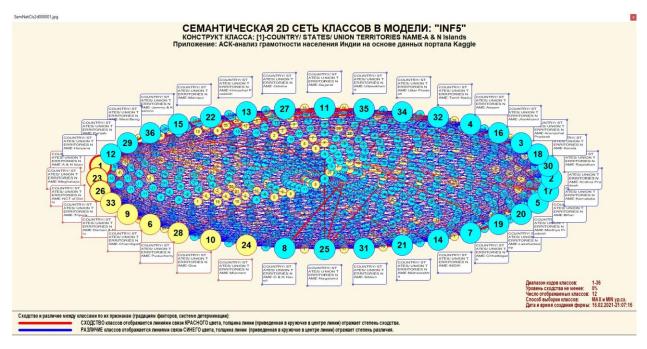


Рисунок 20 - Семантическая 2D сеть классов



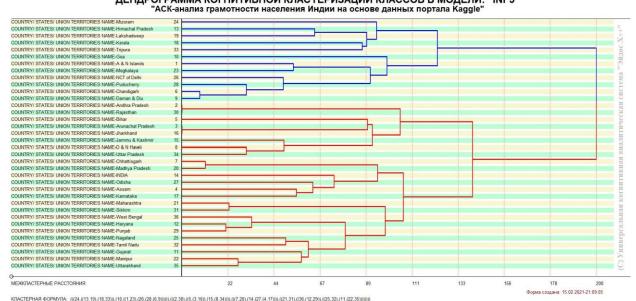


Рисунок 21 – Агломеративная дендрограмма классов



Рисунок 22 – График межкластерных расстояний классов

2.3. Нелокальные нейронные сети и нейтроны

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

Рассмотрим пару примеров, возвращаясь к нашим задачам.

На рисунке 23 изображено графическое отображение нелокальных нейронов в системе Aidos-X

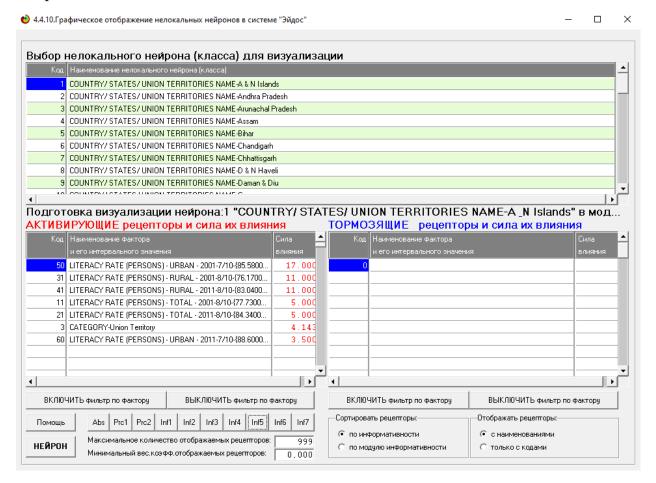


Рисунок 24 - Графическое отображение нелокальных нейронов в системе Aidos-X

Для каждого технологического фактора в соответствии с предложенной моделью определяется величина и направление его влиянии на осуществление всех желаемых и не желаемых хозяйственных ситуаций. Для каждой ситуации эта информация отображается в различных текстовых и графических формах, в частности в форме нелокального нейрона (рисунок 25). На данной диаграмме цвет линии означает знак связи (красный – положительная, синий – отрицательная), а толщина –ее модуль.

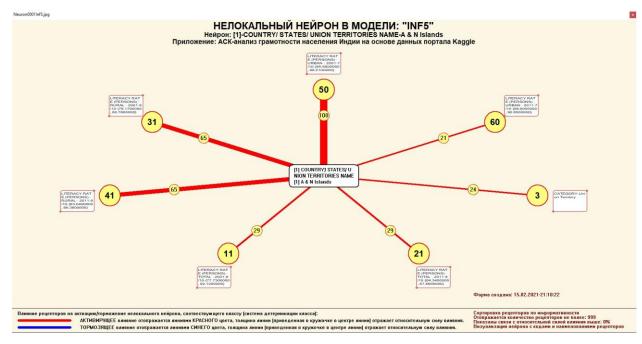


Рисунок 25 - Нелокальный нейрон в модели "INF5"

Дополнение модели нейрона связями факторов позволяет построить классическую когнитивную карту ситуации (будущего состояния АОУ). Детальная внутренняя структура любой связи отображается в форме инвертированной когнитивной диаграммы (рисунок 26). Необходимо отметить, что все указанные графические формы генерируются системой Aidos-X автоматически в соответствии с созданной моделью[9].

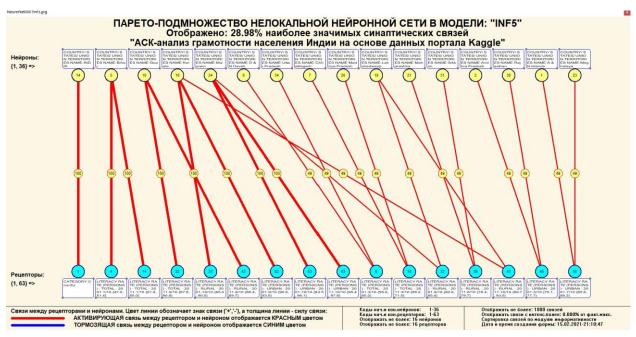


Рисунок 26 - Паретто – подмножество нелокальной нейронной сети

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 не формализуемым путем (интуитивно), основе делают на 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 привлечения экспертов. Эта проблема может без решаться автоматизации функций экспертов, т.е. путем измерения силы и направления факторов непосредственно на основе эмпирических данных. Подобная технология разработана давно, ей уже около 30 лет, но она малоизвестна – это интеллектуальная система Aidos-X. Данная система всегда обеспечивала возможность проведения количественного автоматизированного SWOT-анализа без использования экспертных оценок непосредственно на основе эмпирических данных. Результаты SWOTанализа выводились в форме информационных портретов. В версии системы под MS Windows: Aidos-X++ предложено автоматизированное количественное SWOT-анализа решение прямой обратной задач построением традиционных SWOT-матриц и диаграмм (рисунок 27).

На рисунке 28 приведен пример табличной выходной формы количественного автоматизированного SWOT- и PEST- анализа средствами системы Aidos.

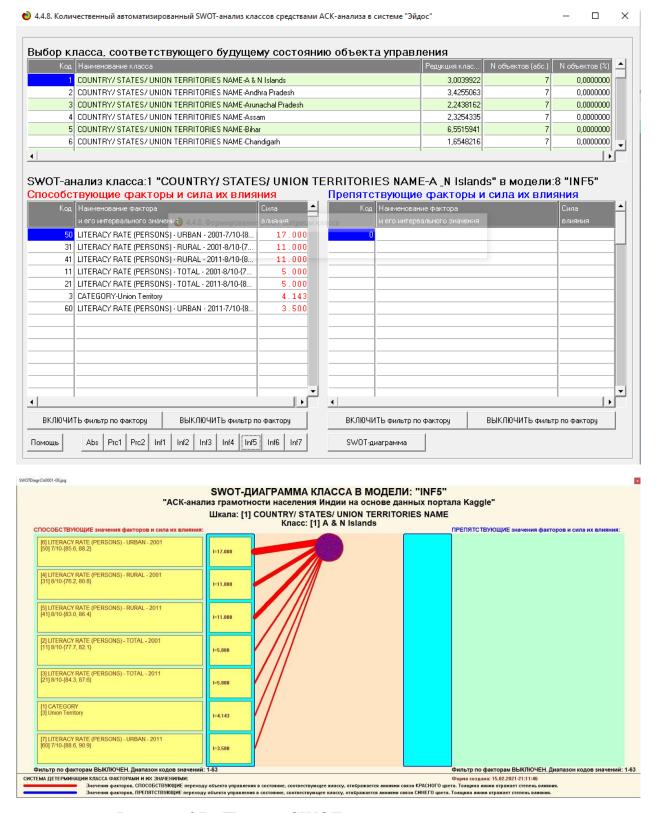


Рисунок 27 - Пример SWOT-матрицы в модели INF5

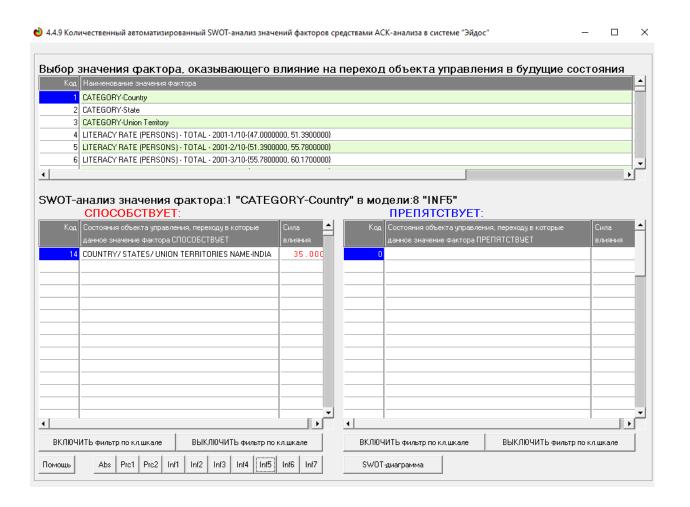


Рисунок 28 - Табличная выходная форма количественного автоматизированного SWOT- и PEST- анализа средствами системы Aidos

На рисунке 29 приведен пример графической выходной формы количественного автоматизированного SWOT- и PEST- анализа средствами системы Aidos.

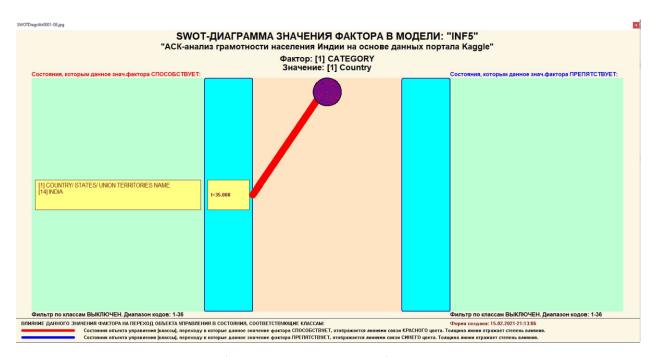


Рисунок 29 - Графическая выходная форма количественного автоматизированного SWOT- и PEST-анализа средствами системы Aidos

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 электронную базу Kaggle.

Сверхзадачей искусственного интеллекта является построение компьютерной интеллектуальной системы, которая обладала бы уровнем эффективности решений неформализованных задач, сравнимым с человеческим или превосходящим его.

Самую существенную часть систем искусственного интеллекта составляют экспертные системы. Экспертная система обычно определяется как программа ЭВМ, моделирующая действия эксперта человека при решении задач в узкой предметной области: составление базы знаний и накопления их.

В данной курсовой работе было показано построение модели на основе статистических данных грамотности населения Индии системой искусственного интеллекта "Aidos-X++". При этом наиболее достоверной в данном приложении оказалась модель INF5, основанная на семантической мере целесообразности информации А.Харкевича при интегральном критерии «Сематический резонанс знаний». Точность модели составляет 0,661, что заметно выше, чем достоверность экспертных оценок, которая считается равной около 49,4%.

АСК-анализ текстов позволяет:

- формировать обобщенные лингвистические образы классов (семантические ядра) на основе фрагментов или примеров относящихся к ним текстов на любом языке;

- количественно сравнивать лингвистический образ конкретного человека, или описание объекта, процесса с обобщенными лингвистическими образами групп (классов);
- сравнивать обобщенные лингвистические образы классов друг с другом и создавать их кластеры и конструкты;
- исследовать моделируемую предметную область путем исследования ее лингвистической системно-когнитивной модели;
- проводить интеллектуальную атрибуцию текстов, т.е. определять вероятное авторство анонимных и псевдонимных текстов, датировку, жанр и смысловую направленность содержания текстов;
- все это можно делать для любого естественного или искусственного языка или системы кодирования.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

- 1. 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.
- 2. Луценко Е.В. АСК-анализ, моделирование и идентификация живых существ на основе их фенотипических признаков / Е.В. Луценко, Ю.Н. Пенкина // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. Краснодар: Куб- ГАУ, 2014. №06(100). С. 1346 1395. IDA [article ID]: 1001406090. Режим доступа: http://ej.kubagro.ru/2014/06/pdf/90.pdf, 3,125 у.п.л.
- 3. Луценко Е.В. 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. сэкрана. Яз.анг.
- 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- 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://ei.kubagro.ru/2012/09/pdf/25.pdf, 1,812 у.п.л.
- 8. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. Краснодар: КубГАУ, 2003. №01(001). С. 79 91. IDA [article ID]: 0010301011. Режим доступа: http://ej.kubagro.ru/2003/01/pdf/11.pdf, 0,812 у.п.л.
- 9. Луценко Е.В. Синтез семантических ядер научных специальностей ВАК РФ и автоматическая классификации статей по научным специальностям с применением АСК-анализа и интеллектуальной системы «Эйдос» (на примере Научного журнала КубГАУ и его научных специальностей: механизации, агрономии и ветеринарии) / Е.В. Луценко, Н.В. Андрафанова, Н.В. Потапова // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. Краснодар: КубГАУ, 2019. №01(145). С. 31 102. IDA [article ID]: 1451901033. Режим доступа: http://ej.kubagro.ru/2019/01/pdf/33.pdf, 4,5 у.п.л.

- 10. Луценко Е.В. Формирование семантического ядра ветеринарии путем Автоматизированного системно-когнитивного анализа паспортов научных специальностей ВАК РФ и автоматическая классификация текстов по направлениям науки / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный Краснодар: КубГАУ, 2018. №10(144). C. 44 – 102. [article ID]: 1441810033. Режим доступа: http://ej.kubagro.ru/2018/10/pdf/33.pdf, 3,688 у.п.л.
- 11. Луценко Е.В. Интеллектуальная привязка некорректных ссылок к литературным источникам в библиографических базах данных с применением АСК-анализа и системы «Эйдос» (на примере Российского индекса научного цитирования РИНЦ) / Е.В. Луценко, В.А. Глухов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №01(125). С. 1 – 65. – IDA [article ID]: 1251701001. Режим доступа: http://ej.kubagro.ru/2017/01/pdf/01.pdf, 4,062 у.п.л.
- 12. Луценко Е.В. Применение АСК-анализа и интеллектуальной системы "Эйдос" для решения в общем виде задачи идентификации литературных источников и авторов по стандартным, нестандартным и некорректным библиографическим описаниям / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. Краснодар: КубГАУ, 2014. №09(103). С. 498 544. IDA [article ID]: 1031409032. Режим доступа: http://ej.kubagro.ru/2014/09/pdf/32.pdf, 2,938 у.п.л.
- 13. Луценко Е.В. АСК-анализ проблематики статей Научного журнала КубГАУ в динамике / Е.В. Луценко, В.И. Лойко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. Краснодар: КубГАУ, 2014. №06(100). С. 109 145. IDA [article ID]: 1001406007. Режим доступа: http://ej.kubagro.ru/2014/06/pdf/07.pdf, 2,312 у.п.л.
- 14. Луценко Е.В. Атрибуция анонимных и псевдонимных текстов в системно-когнитивном анализе / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. Краснодар: КубГАУ, 2004. №03(005). С. 44 64. IDA [article ID]: 0050403003. Режим доступа: http://ej.kubagro.ru/2004/03/pdf/03.pdf, 1,312 у.п.л.
- 15. Луценко Е.В. Атрибуция текстов, как обобщенная задача идентификации и прогнозирования / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. Краснодар: КубГАУ, 2003. №02(002). С. 146 164. IDA [article ID]: 0020302013. Режим доступа: http://ej.kubagro.ru/2003/02/pdf/13.pdf, 1,188 у.п.л.
- 16. Луценко Д.С., Луценко Е.В. Интеллектуальная датировка текста, определение авторства и жанра на примере русской литературы XIX и XX веков, 2020 // Статья в открытом архиве. 38 с. DOI: 10.13140/RG.2.2.28824.01281, https://www.elibrary.ru/item.asp?id=43796415
- 17. Lutsenko D.S., Lutsenko E.V. Intellectual attribution of literary texts (finding the dates of the text, determining authorship and genre on the example of russian literature of the XIX and XX centuries), 2020 // Статья в открытом архиве. 9 р. DOI: 10.13140/RG.2.2.15349.81122, https://www.elibrary.ru/item.asp?id=43794562