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

Лабораторная работа

АСК-анализ изменения спроса на товары аптечного ассортимента.

Выполнил студент 3 курса,
группы ПИ1501
Игнатенко С.С.

Руководитель
профессор, д.э.н., к.т.н., Луценко Е.В.

Краснодар 2018

Оглавление

Введение 3

1.Синтез и верификация моделей 4

1.1. Описание решения 4

1.2. Формирование базы исходных данных 4

1.3. Синтез и верификация моделей предметной области 8

1.4. Результаты верификации моделей 9

2.РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ 13

2.1.Решение задачи 13

2.2. SWOT и PEST матрицы и диаграммы 15

2.3 Нелокальные нейроны 15

2.4 Кластерно-конструктивный анализ признаков 18

2.5 Кластерно-конструктивный анализ классов 22

2.6 Когнитивные функции 25

2.7 Геокогнитивная система 27

ЗАКЛЮЧЕНИЕ 29

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ 30

Введение

Сейчас искусственный интеллект рассматривают как прикладную область исследований, связанных с имитацией отдельных функций интеллекта человека. Распознавание образов, машинный перевод, интеллектуальные агенты, робототехника — это лишь некоторые из направлений, по которым развиваются системы искусственного интеллекта. В данной работе рассмотрено изменения спроса на товары аптечного ассортимента.

Для решения данной задачи необходимы исходные статистические данные. В качестве источника данных была взята единая информационно-статистическая система (Росстат), содержащая официальную статистическую информацию, формируемую субъектами официального статистического учета в рамках Федерального плана статистических работ.

В данной работе использована база данных, содержащая среднее количество совершенных преступлений разными категориями лиц (данные МВД) с 1992 по 2015 года.

Для решения задачи будут использованы программы Microsoft Office Word и Excel, а также система искусственного интеллекта «Aidos-X++».

1.Синтез и верификация моделей

1.1. Описание решения

Порядок решения задачи в соответствии с методологией АСК-анализа:

1. Преобразование базы исходных данных в необходимый для системы формат файла MS Excel.
2. Преобразование исходных данных из файла MS Excel в базы данных системы «Эйдос».
3. Синтез и верификация моделей предметной области.
4. Применение моделей для решения задач идентификации, прогнозирования и исследования предметной области.

1.2.Формирование базы исходных данных

Из Росстат «http://cbsd.gks.ru/ » базы данных - «Анализ изменения спроса на товары аптечного ассортимента**.** «. (из ряда таблиц были выбраны данные и собраны в одну таблицу и представлены на рис 1)

Столбцы 2-3 – классификационные шкалы, столбец 4-11 – описательные шкалы.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Объект	Вид/месяц	Антидепрессанты	Витамины	Средства для похудения	Препараты против аллергии	Успокоительные	Растворы для линз	Препараты для потенции	Противозачаточные средства	Препараты от ОРВИ, гриппа	Косметические средства	
2	1	Январь	23143	1241	43212	21124	12876	65564	23143	24162	15689	89514	
3	2	Февраль	34176	15464	43467	31546	15644	61035	22601	22149	70645	105563	
4	3	Март	46514	16431	54418	32436	16584	54736	19845	1165431	220014	120315	
5	4	Апрель	33216	14683	37514	27521	16463	55646	19955	221325	195312	112031	
6	5	Май	45800	167828	36217	33677	18534	52169	18364	208865	226872	105463	
7	6	Июнь	50457	15887	29132	35685	21304	39143	17746	20463	3334	118087	
8	7	Июль	51512	13946	26048	31693	2073	36117	15428	196062	9796	121923	
9	8	Август	59828	12005	22964	35165	22643	43912	14110	189660	6258	125760	
10	9	Сентябрь	64504	16436	21880	32637	23912	400652	13792	18358	292721	129596	
11	10	Октябрь	51316	18067	70796	37109	25232	37032	11475	176857	295613	133433	
12	11	Ноябрь	45651	17528	19712	28581	26515	34031	19157	170455	298505	137270	
13	12	Декабрь	3465	16929	18628	64053	27792	39871	19639	164053	301397	141106	
14													

Рисунок 1 – «Продажи товаров аптечного ассортимента»

Ввод исходных данных в систему «Эйдос» в формате XLS производится при помощи универсального программного интерфейса импорта данных в режиме 2.3.2.2 (рис. 2).

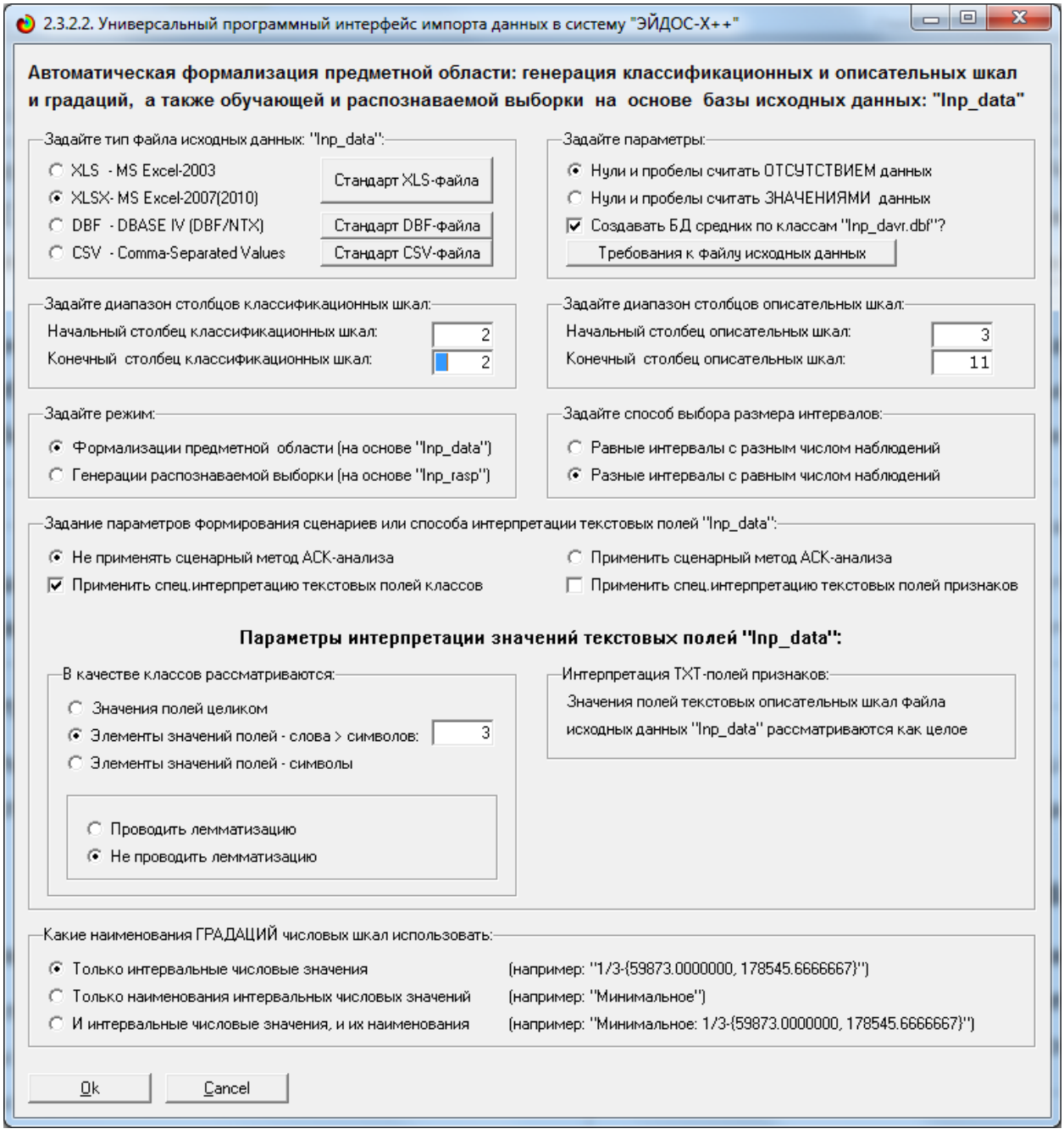


Рисунок 2 - Экранная форма «Универсального программного интерфейса импорта данных» в систему «Эйдос» (режим 2.3.2.2)

После ввода первичных настроек будет предложено задать число интервалов классификационных и описательных шкал (рис. 3).

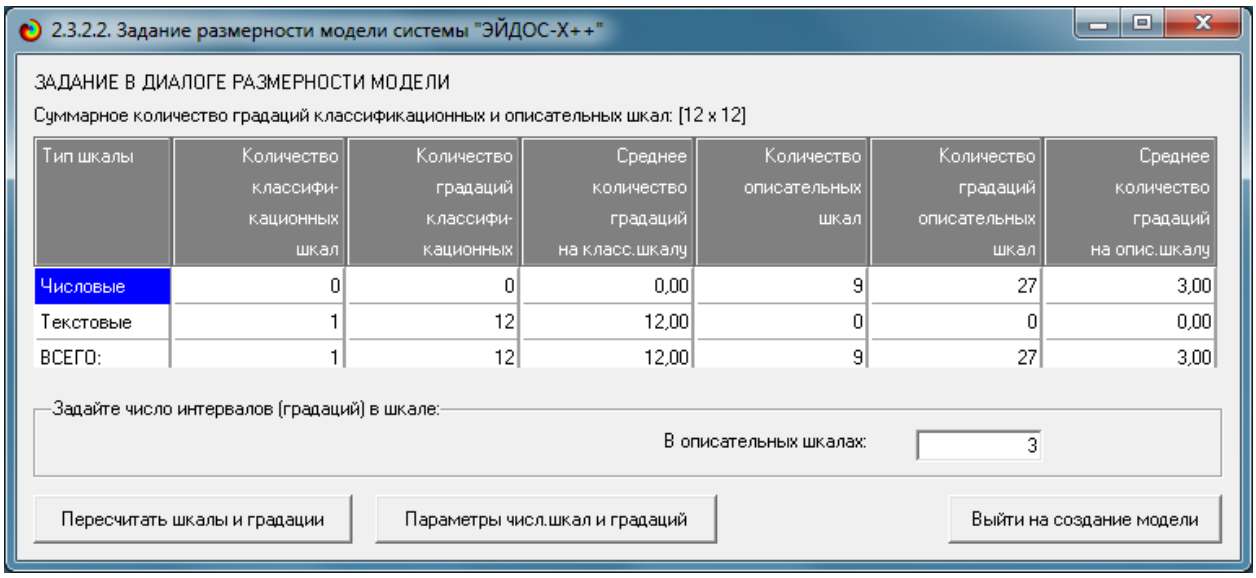


Рисунок 3 - Задание размерности модели системы «Эйдос»

Далее открывается окно с прогрессом импорта данных (рис 4).

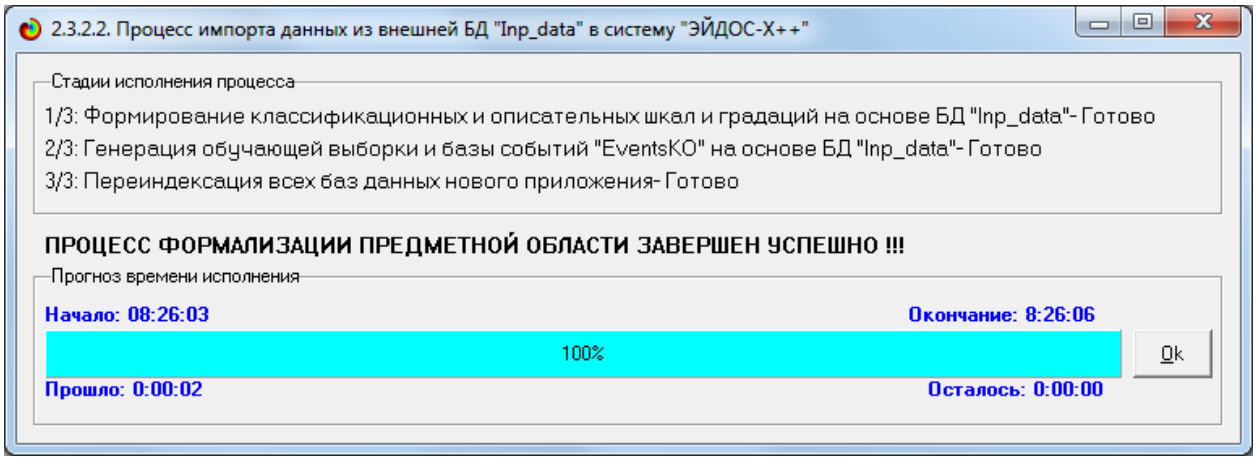


Рисунок 4 - Прогресс процесса импорта данных из внешней БД «input_data.xls» в систему «Эйдос»

В результате формируются классификационные и описательные шкалы и градации, с применением которых исходные данные кодируются и представляются в форме баз данных. Этим самым полностью автоматизировано выполняется 2-й этап АСК-анализа «Формализация предметной области». Для просмотра классификационных шкал и градаций необходимо запустить режим 2.1 (рисунок 5).

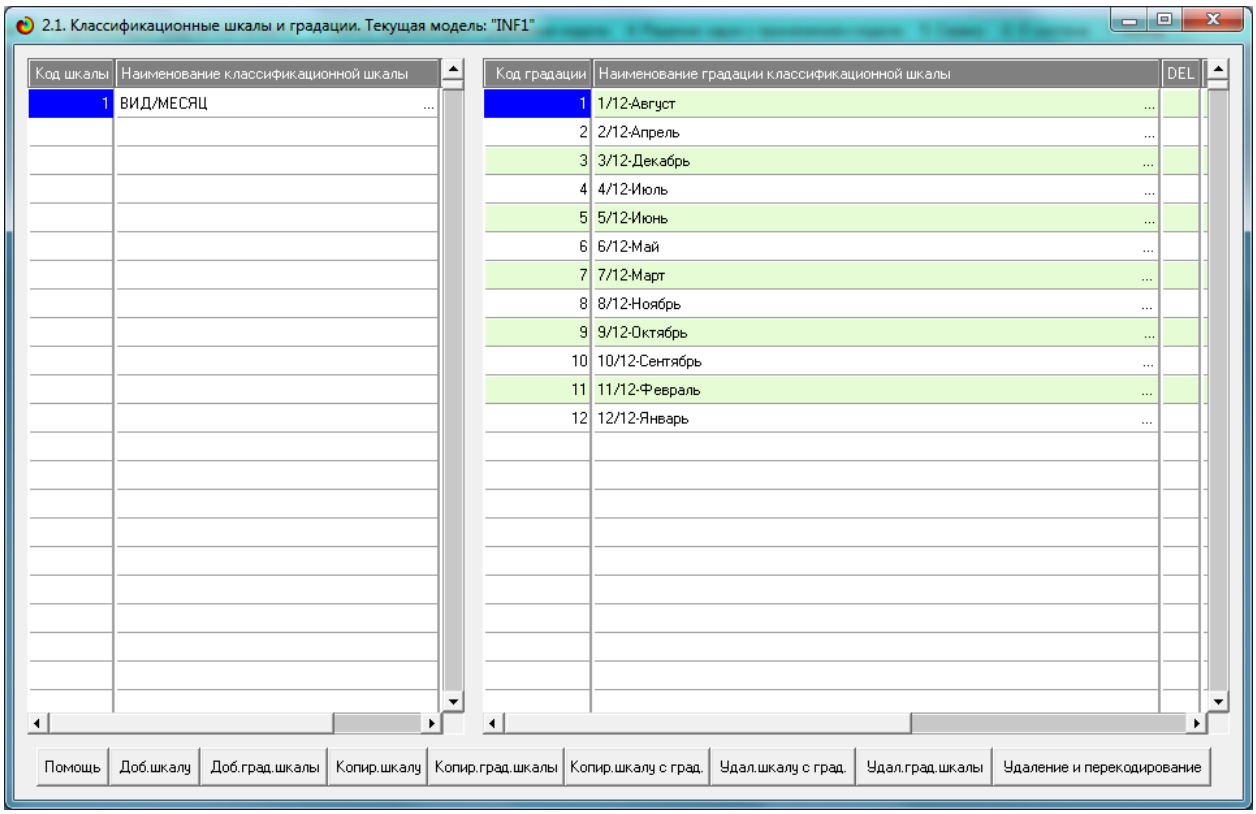


Рисунок 5 - Классификационные шкалы и градации

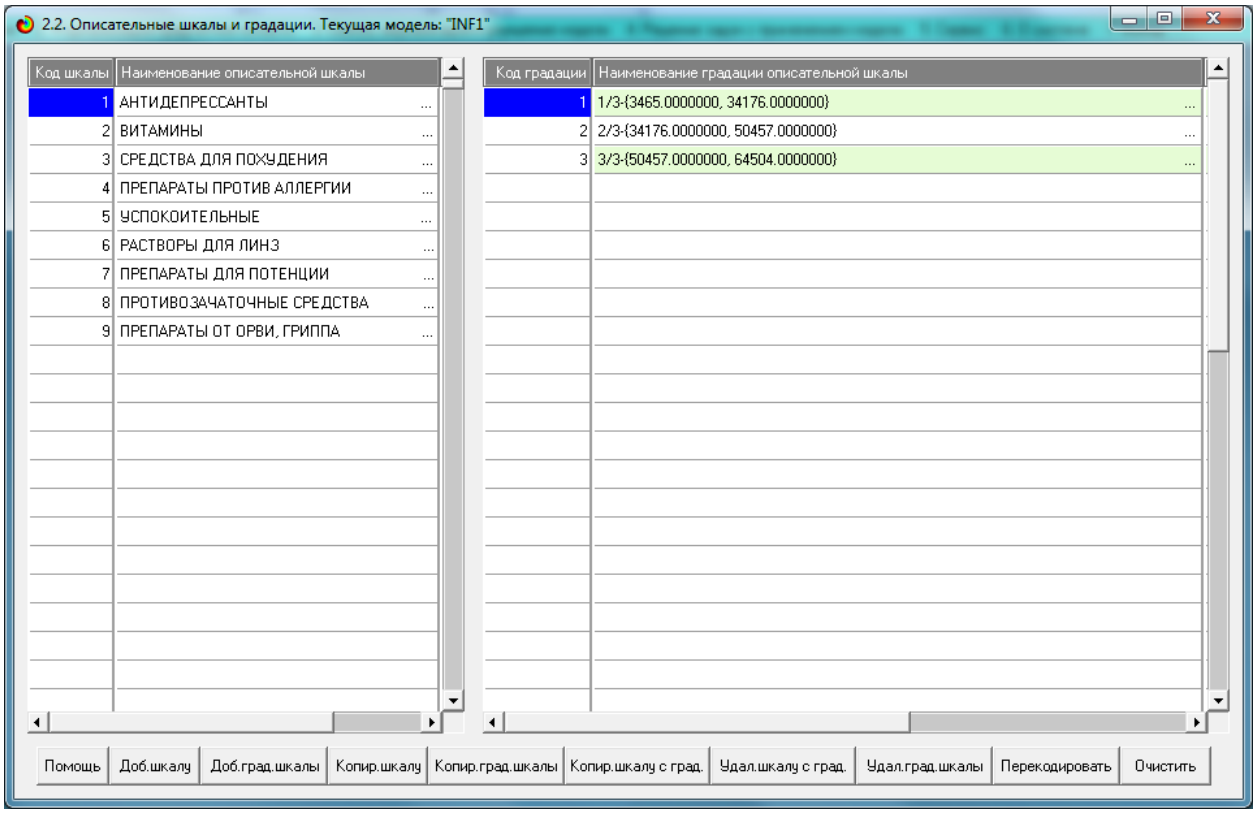


Рисунок 6 - Описательные шкалы и градации

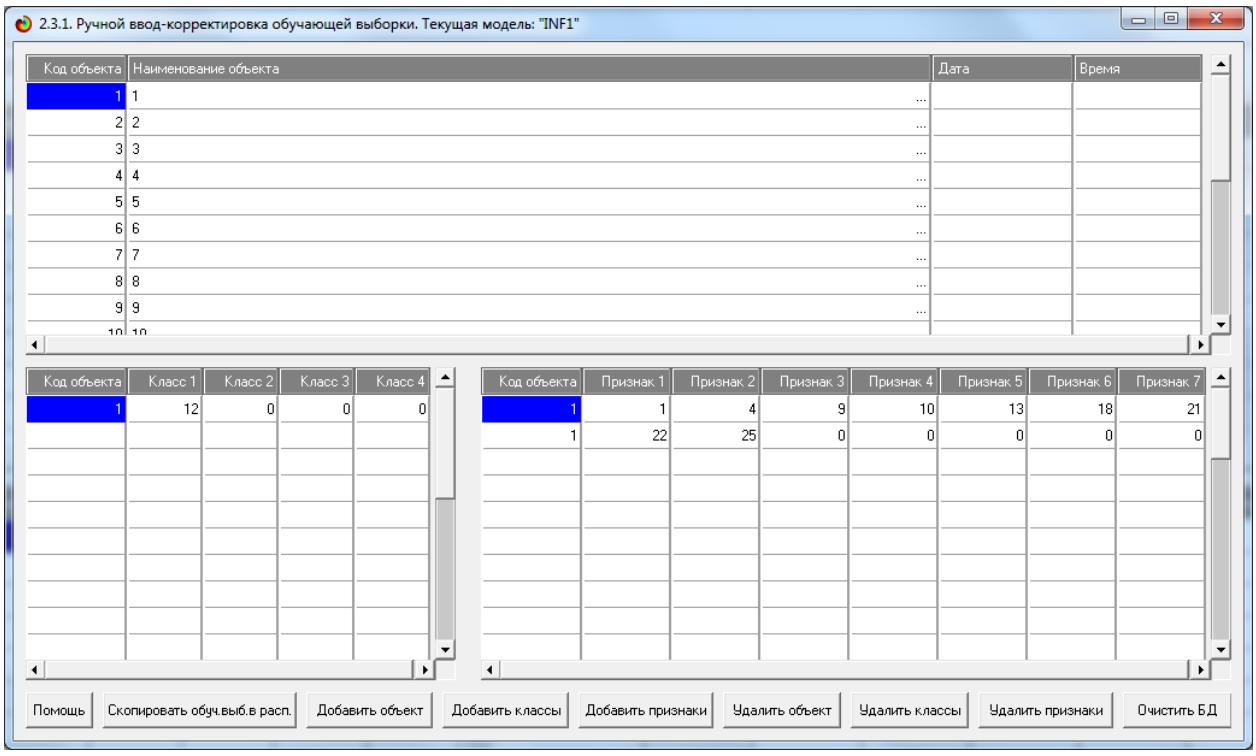


Рисунок 7 - Обучающая выборка (фрагмент)

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.

1.3.Синтез и верификация моделей предметной области

Далее запускается режим 3.5 (рис. 4), в котором задаются модели для синтеза и верификации, а также задается модель, которой после выполнения режима задается статус текущей.

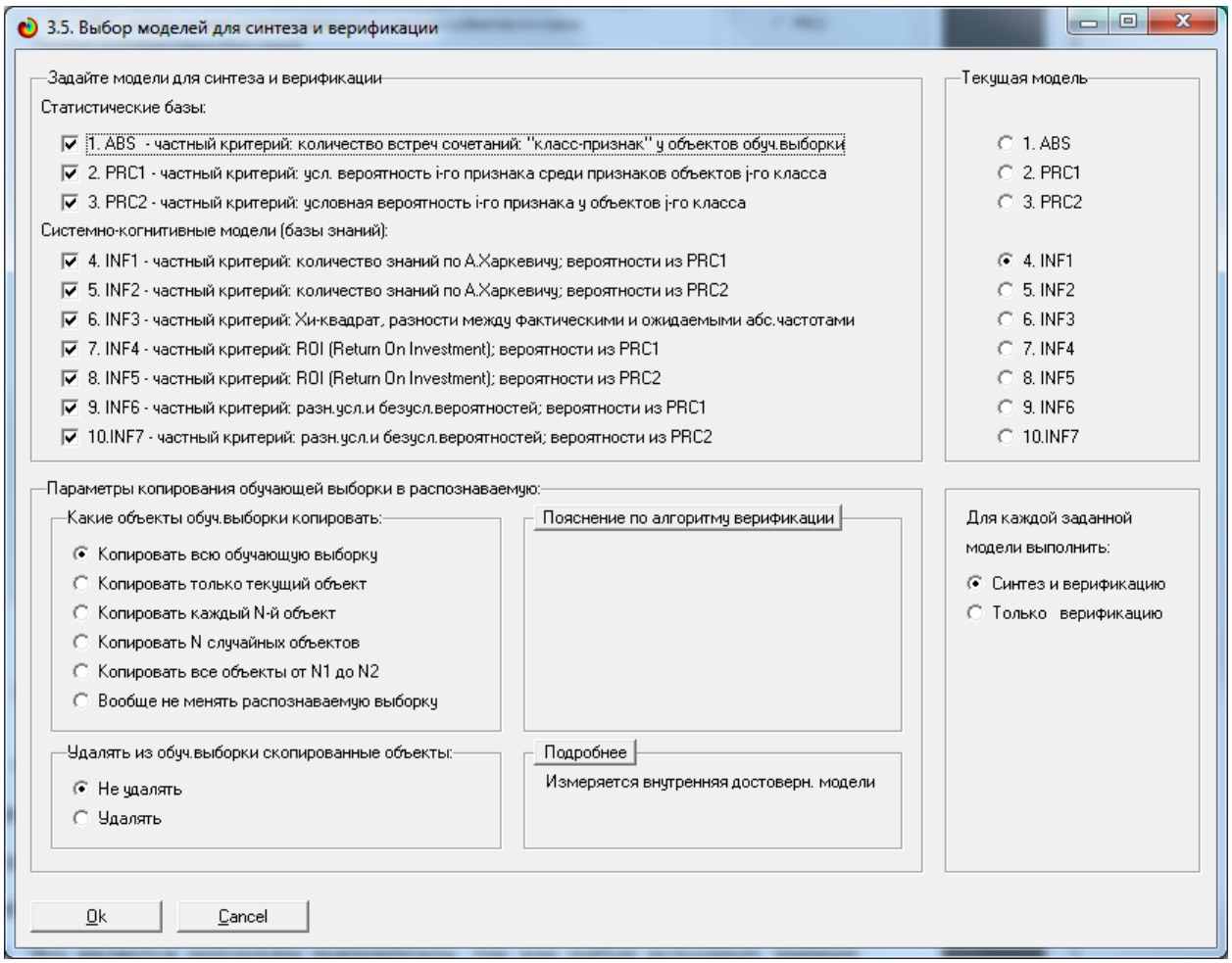


Рисунок 8. Выбор моделей для синтеза и верификации

1.4.Результаты верификации моделей

Фрагменты таблиц результатов верификации моделей представлен на рисунках 9, 10. Наиболее достоверными в данном приложении оказалась модель INF3 (рис. 11) при интегральных критериях: семантический резонанс знаний и сумма знаний, при этом точность модели составляет 0.577, а полнота 0.580. Это является неплохим показателем, так как набор исходных данных имеет сильный статистический разброс.

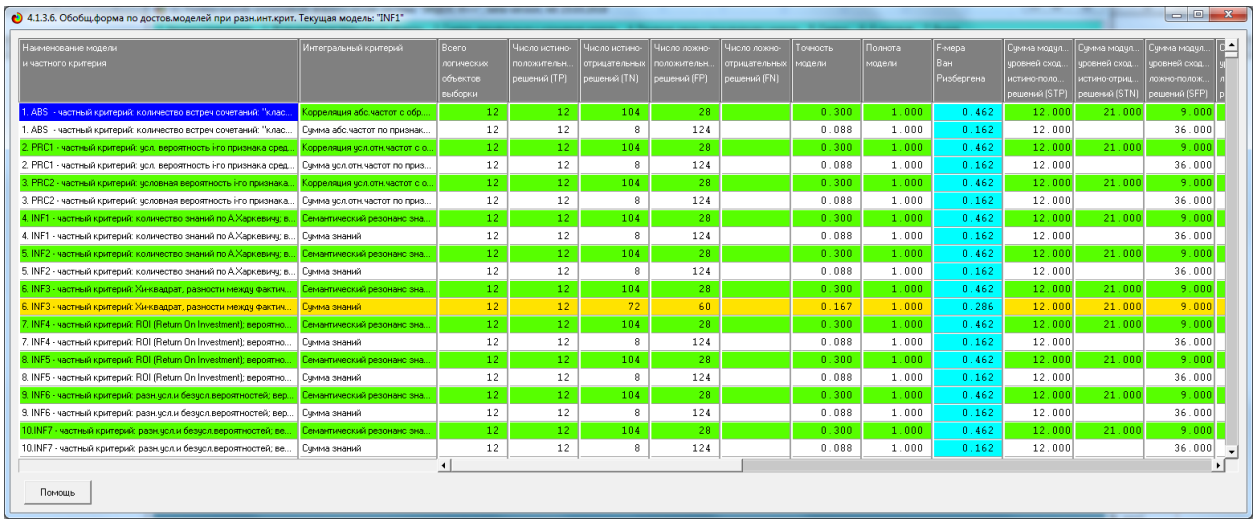


Рисунок 9. Форма достоверности моделей (фрагмент 1)

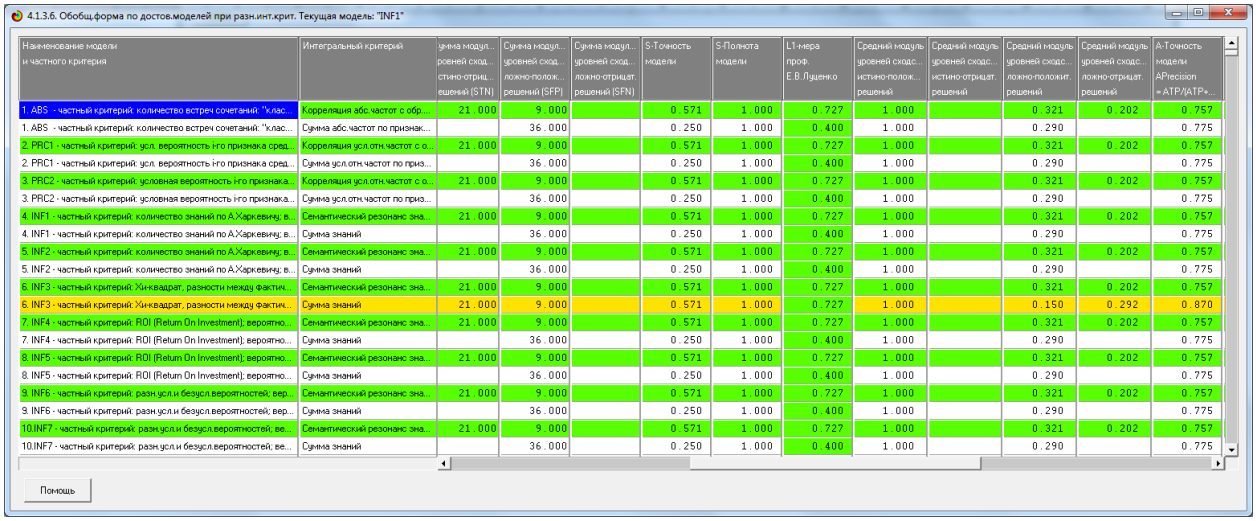


Рисунок 10. Форма достоверности моделей (фрагмент 1)

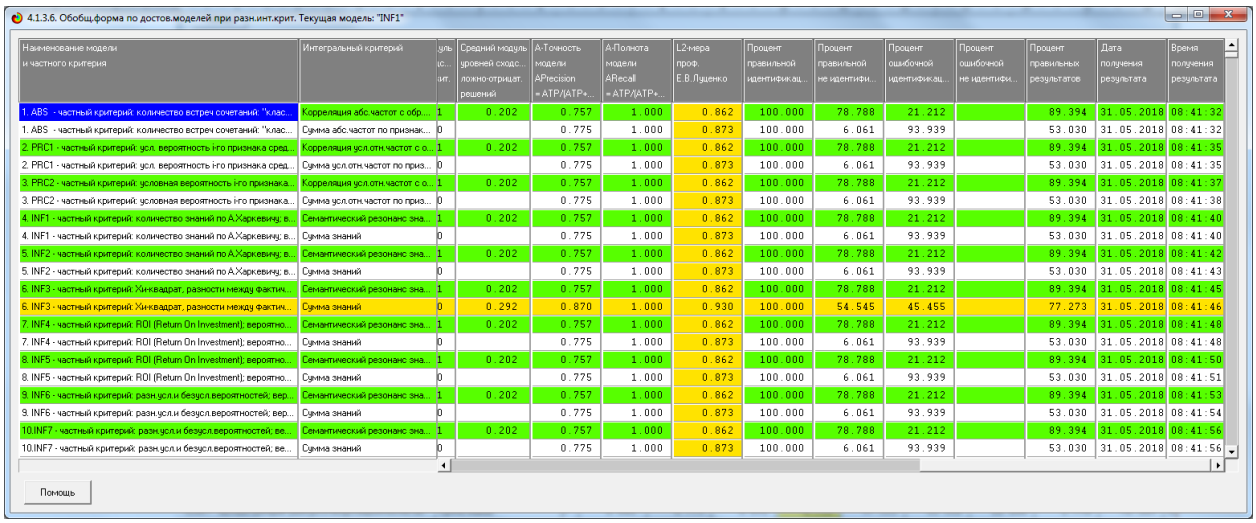
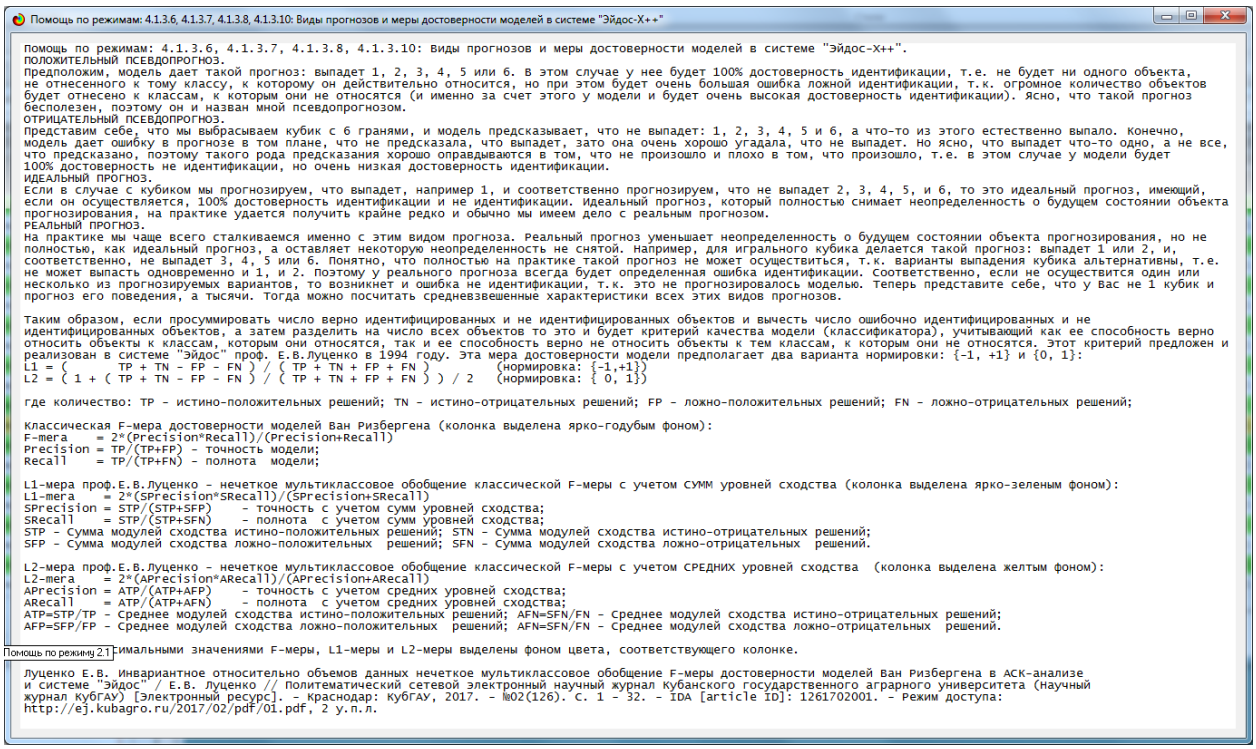


Рисунок 11. Форма достоверности моделей (фрагмент 1)

Вместо описания критериев достоверности модели приведём форму помощи по данному режиму, где они описываются(Рисунок 12).



Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-X++"

Помощь по режимам: 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-X++".
ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Предположим, модель дает такой прогноз: выпадет 1, 2, 3, 4, 5 или 6. В этом случае у нее будет 100% достоверность идентификации, т.е. не будет ни одного объекта, не отнесенного к тому классу, к которому он действительно относится, но при этом будет очень большая ошибка ложной идентификации, т.к. огромное количество объектов будет отнесено к классам, к которым они не относятся (и именно за счет этого у модели и будет очень высокая достоверность идентификации). Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.
ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Представим себе, что мы выбрасываем кубик с 6 гранями, и модель предсказывает, что не выпадет: 1, 2, 3, 4, 5 и 6, а что-то из этого естественно выпало. Конечно, модель дает ошибку в прогнозе в том плане, что не предсказала, что выпадет, зато она очень хорошо угадала, что не выпадет. Но ясно, что выпадет что-то одно, а не все, что предсказано, поэтому такого рода предсказания хорошо оправдываются в том, что не произошло и плохо в том, что произошло, т.е. в этом случае у модели будет 100% достоверность не идентификации, но очень низкая достоверность идентификации.
ИДЕАЛЬНЫЙ ПРОГНОЗ.
Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.
РЕАЛЬНЫЙ ПРОГНОЗ.
На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь 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 реализован в системе "Эйдос" проф. Е.В.Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {-1, +1} и {0, 1}:
L1 = (TP + TN - FP - FN) / (TP + TN + FP + FN)   (нормировка: {-1,+1})
L2 = (1 + (TP + TN - FP - FN) / (TP + TN + FP + FN)) / 2   (нормировка: { 0, 1})

где количество: TP - истино-положительных решений; TN - истино-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;

Классическая F-мера достоверности моделей Ван Ризбергена (колонка выделена ярко-годубым фоном):
F-mera   = 2*(Precision*Recall)/(Precision+Recall)
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 истино-положительных решений; STN - Сумма модулей сходства истино-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-mera   = 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 истино-положительных решений; AFN=SFN/FN - Среднее модулей сходства истино-отрицательных решений;
AFP=SFP/FP - Среднее модулей сходства ложно-положительных решений; AFN=SFN/FN - Среднее модулей сходства ложно-отрицательных решений.

Максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке.

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 12-Описание режима формы достоверности модели

На рисунке 13 приведены частные распределения уровней сходства и различия для верно и ошибочно идентифицированных и неидентифицированных ситуаций в наиболее достоверной модели INF3. видно, что наиболее достоверная модель INF3 лучше определяет непринадлежность объекта к классу, чем принадлежность (что видно также из рисунка 11).

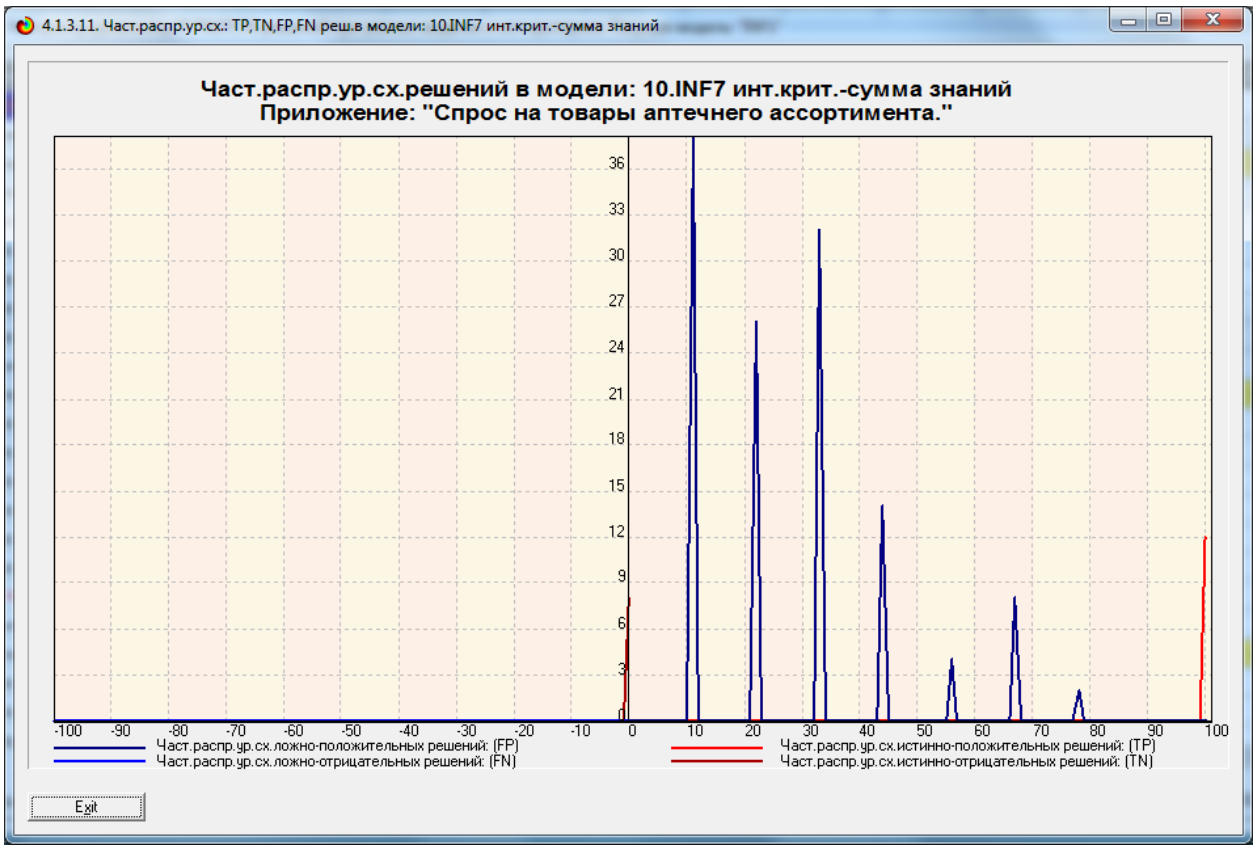


Рисунок 13. Частное распределение сходства-различия верно и ошибочно идентифицированных и неидентифицированных состояний объекта моделирования в модели INF3

Любые данные о наблюдениях можно считать суммой истинного значения и шума, причем ни первое, ни второе неизвестны. Поэтому имеет смысл сравнить созданные модели с чисто случайными моделями, совпадающими по основным характеристикам. В системе «Эйдос» есть лабораторная работа № 2.01: «Исследование RND-модели при различных объемах выборки». Если данная работа устанавливается при отсутствии текущего приложения, то все параметры создаваемых моделей задаются вручную, если же текущая модель существует, как в нашем случае, то все основные ее параметры определяются автоматически (рис. 12)

2. РЕШЕНИЕ ЗАДАЧ В НАИБОЛЕЕ ДОСТОВЕРНОЙ МОДЕЛИ

2.1.Решение задачи

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

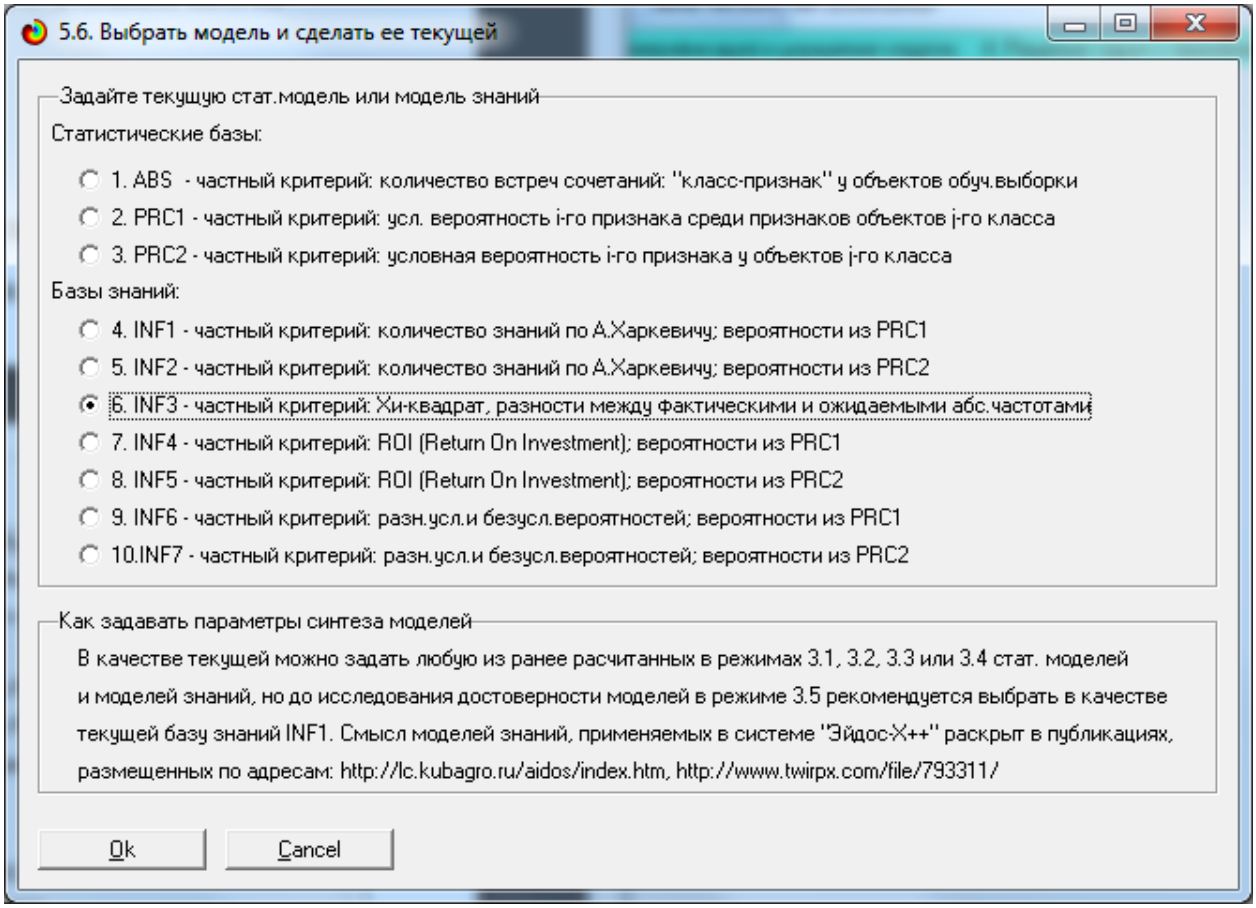


Рисунок 14. Экранные формы режима задания модели в качестве текущей

На рисунках 15 и 16 приведены примеры прогнозов в наиболее достоверной модели INF3:

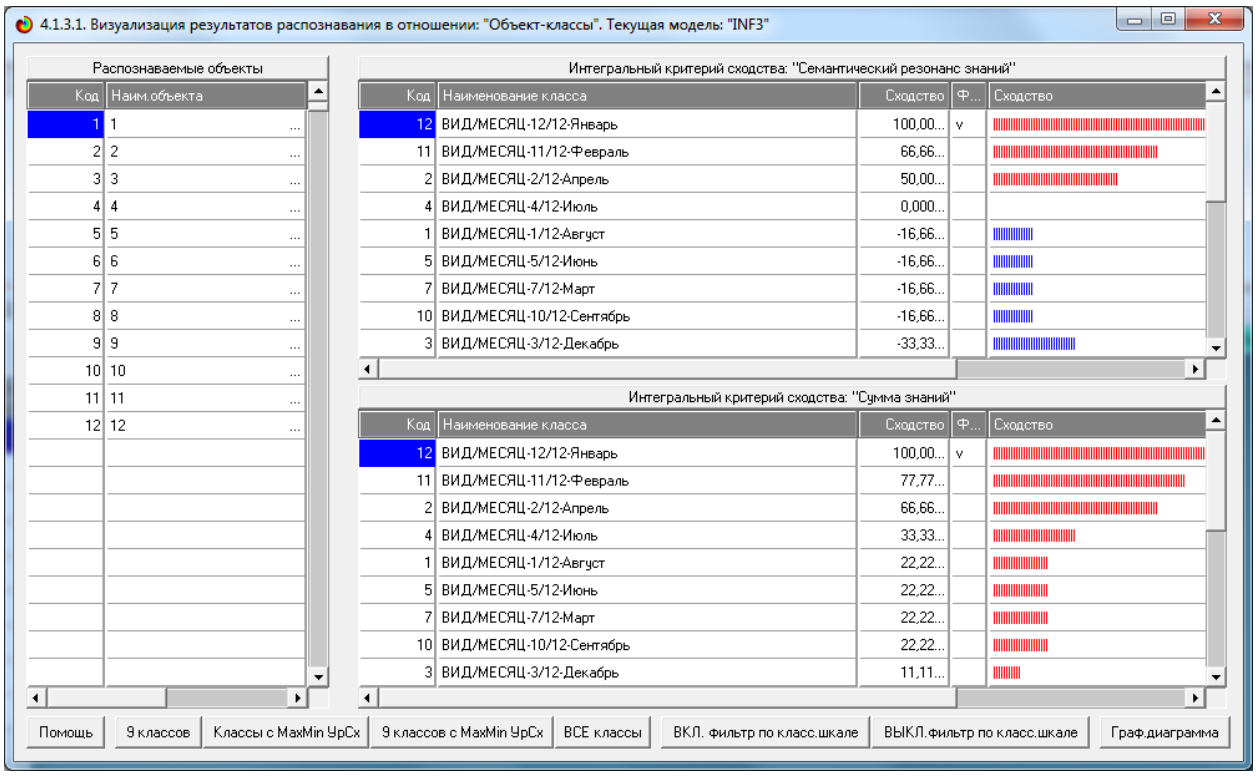


Рисунок 15. Экранные формы режима задания модели в качестве текущей

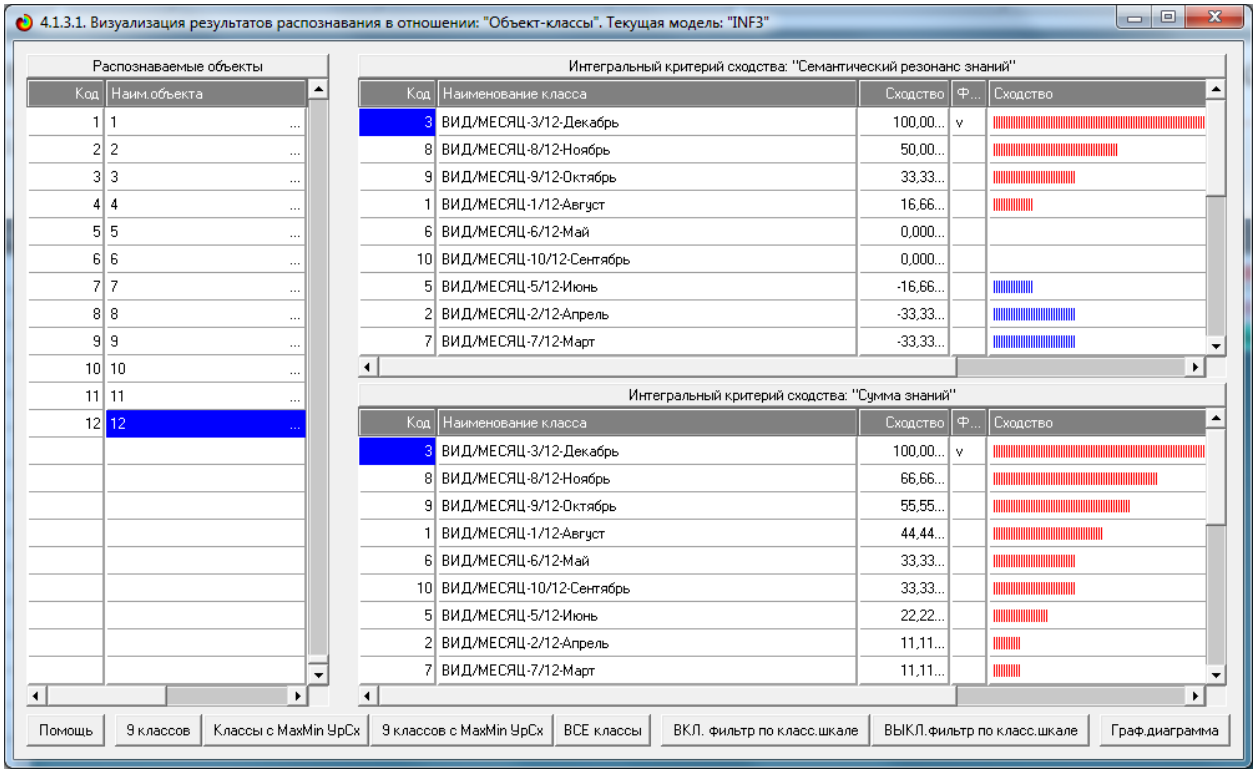


Рисунок 16. Экранные формы режима задания модели в качестве текущей

2.2. SWOT и PEST матрицы и диаграммы

Одним из самым широко известным и распространенным методом стратегического планирования является SWOT – анализ. Однако данный метод очень часто подвергается критике, но это не без причины. В результате использования SWOT – анализа в нем было выявлено множество недостатков. Главной причиной этих недостатков является необходимость привлечения специалиста для оценки и силы и направления влияния факторов. Эту проблему можно решить только с помощью автоматизации функций экспертов. Данную функцию выполняется система «Эйдос». Эта система проводит SWOT – анализ без использования экспертных оценок непосредственно на основе эмпирических данных. В данной работе представлено решение прямой задачи SWOT- анализа с построением традиционных SWOT – матриц и диаграмм (рисунок 16).

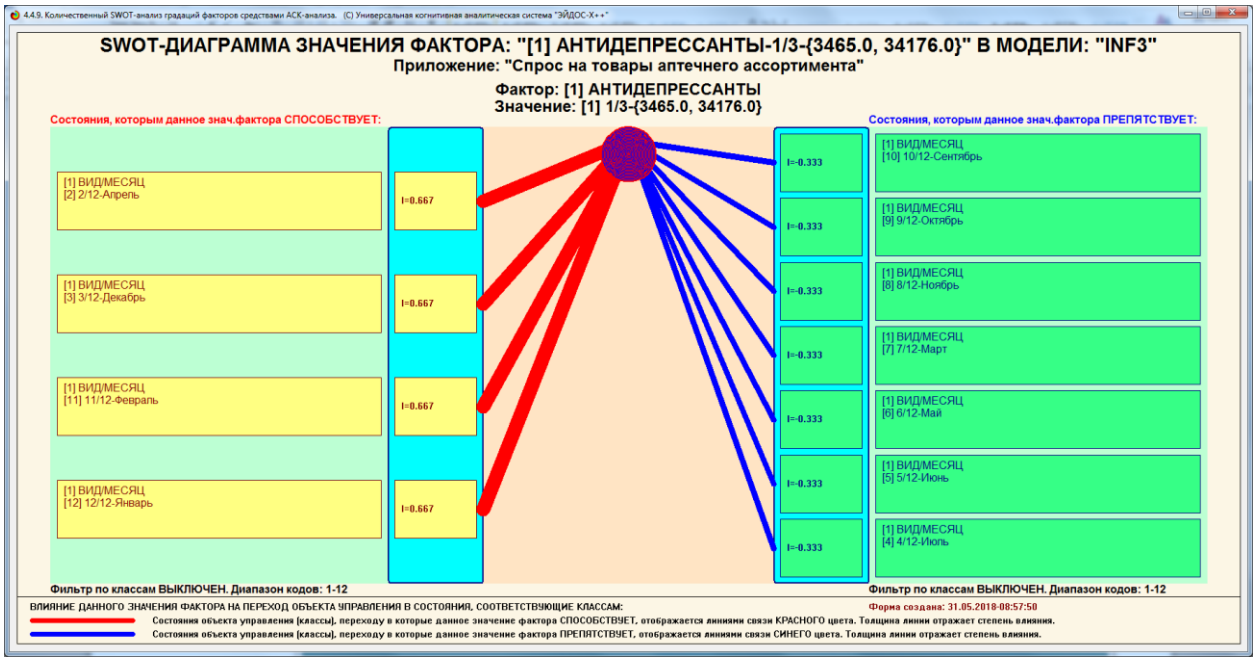


Рисунок 17 – Пример SWOT – матрицы в модели INF3

2.3 Нелокальные нейроны

АСК анализ обеспечивает построение нелокальных нейронов с указанием силы и направления влияния активирующих и тормозящих рецепторов непосредственно на основе эмпирических данных. Пример нелокального нейрона приведен на рисунке 18. Для наглядности на рисунке

19 изображен нелокальный нейрон «Элитное жилье на вторичном рынке» с 3 рецепторами. Красным выделены активизирующее рецепторы, а синим – тормозящие. По рисунку 16 видно, что «Элит. Втор. среднее» (наибольший) – активизирующий рецептор, а «Элит. Втор. малое» (наименьший) – тормозящий [1].

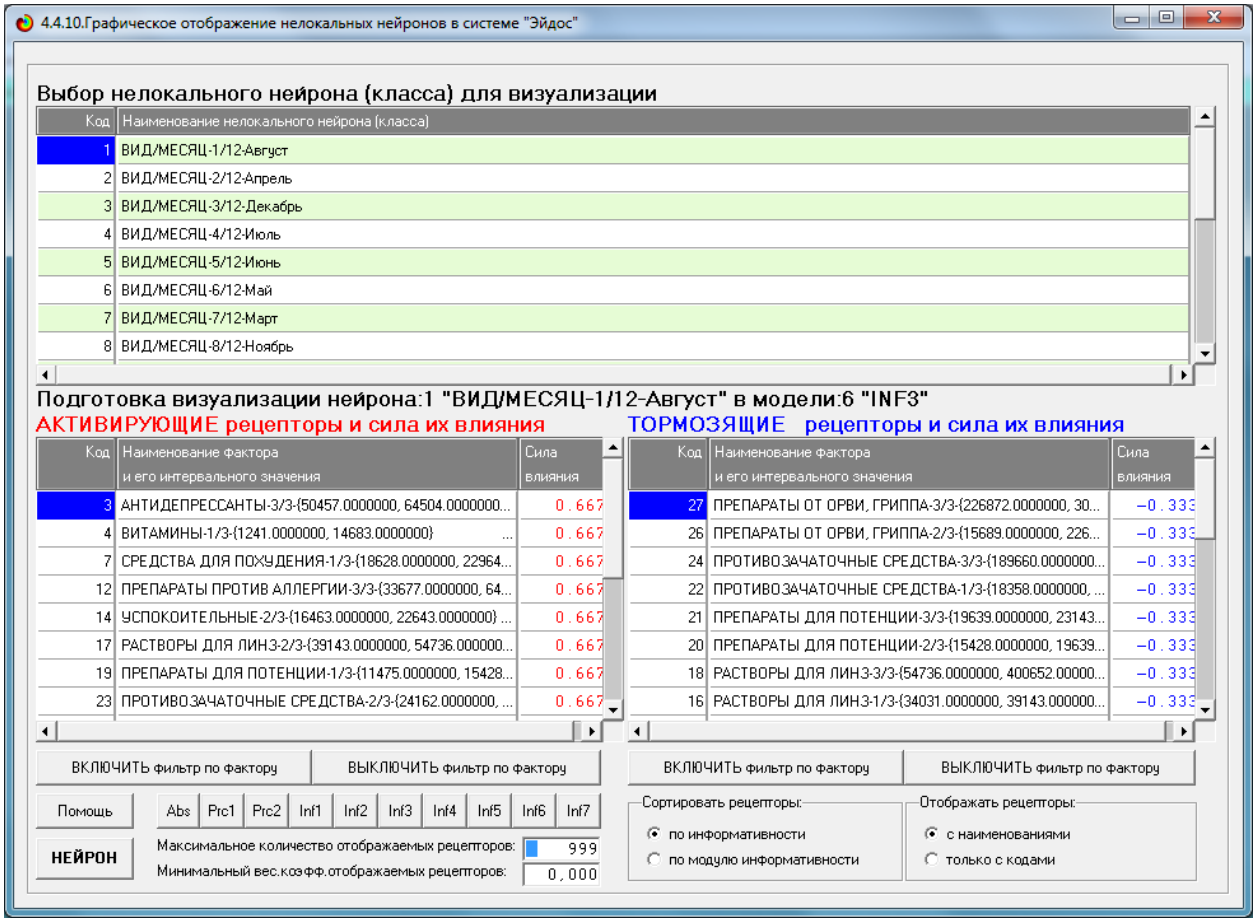


Рисунок 18 − Пример нейрона в модели INF3 (режим 4.4.9)

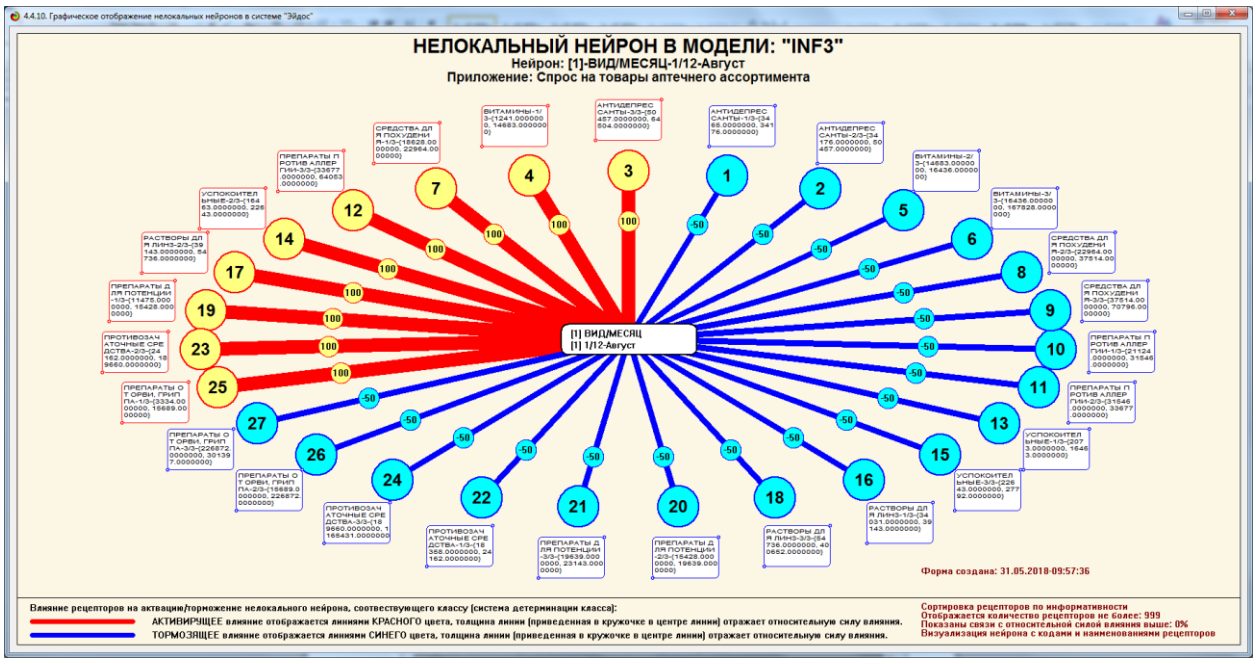


Рисунок 19 – Пример нейрона в модели INF3

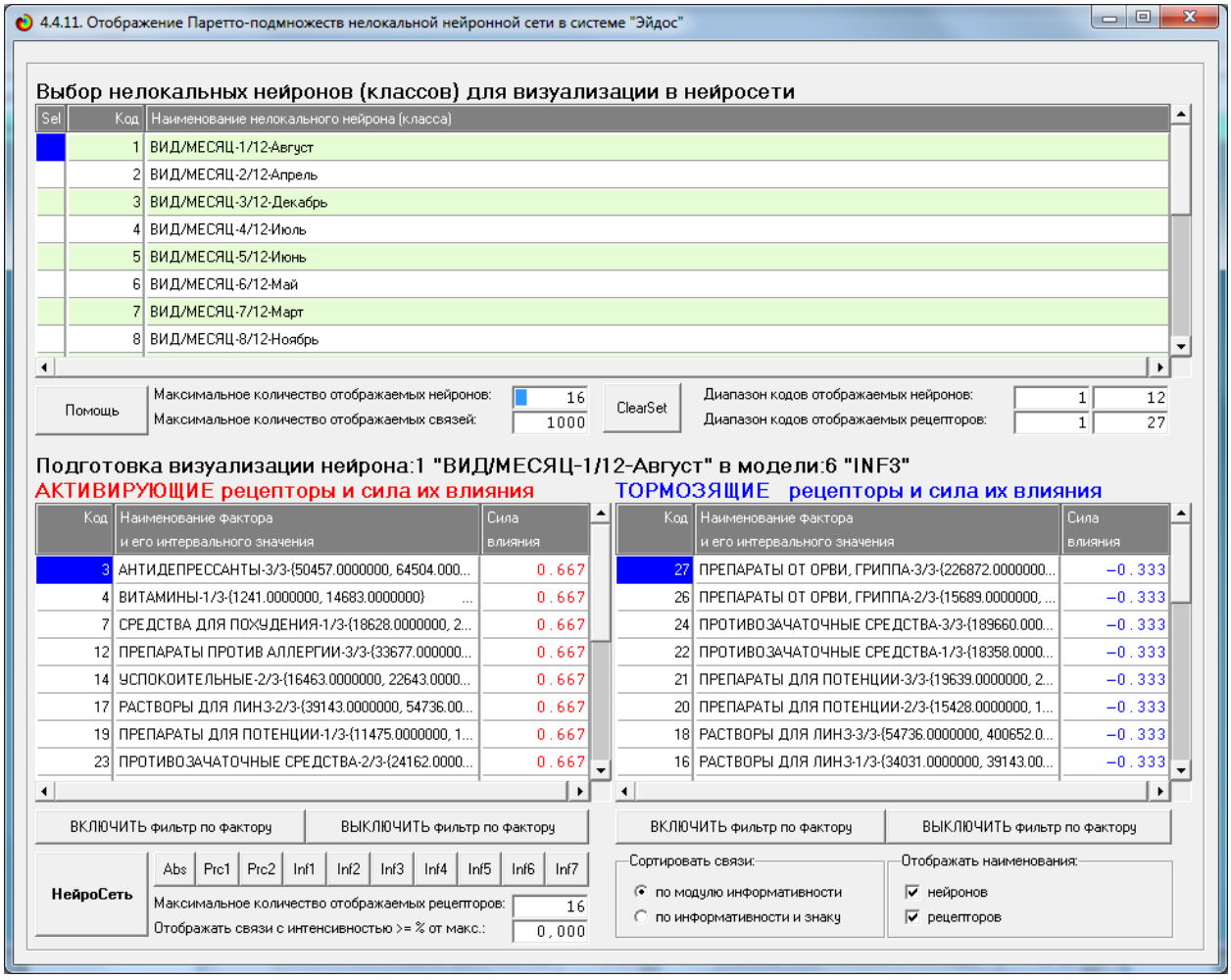


Рисунок 20 - Пример нейронной сети в модели INF3

Однослойная нейронная сеть отображена на рисунке 23.

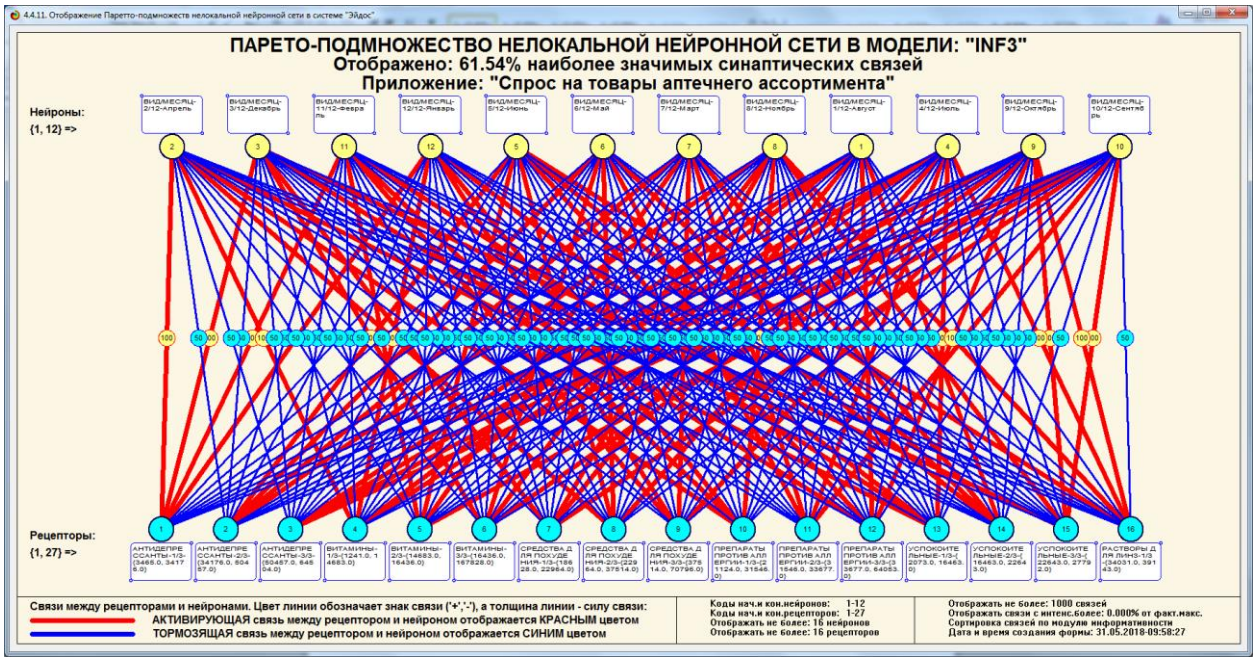


Рисунок 21 - Нейронная сеть

2.4 Кластерно-конструктивный анализ признаков

На рисунке 21 приведены результаты кластерно-конструктивного анализа признаков:

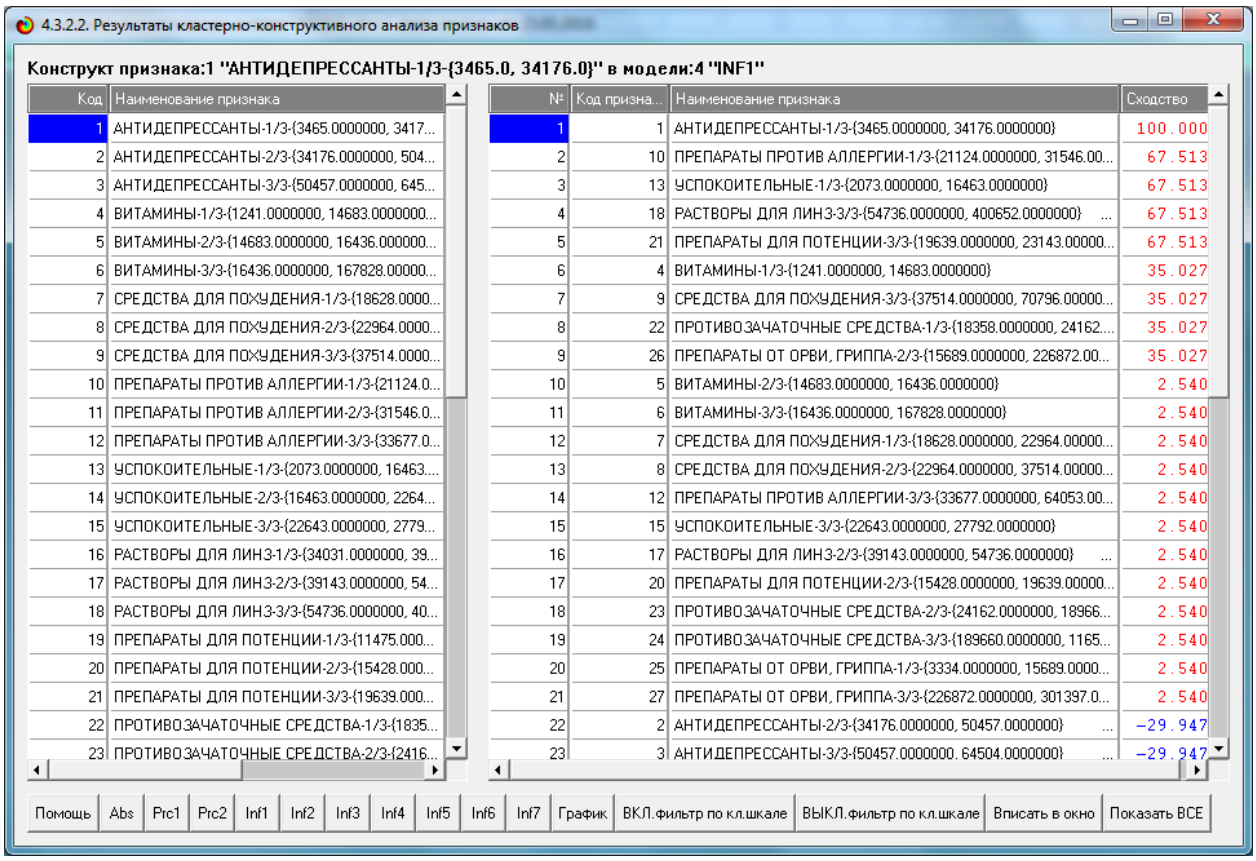


Рисунок 22 – Кластерно-конструктивный анализ признаков (режим 4.3.2.2)

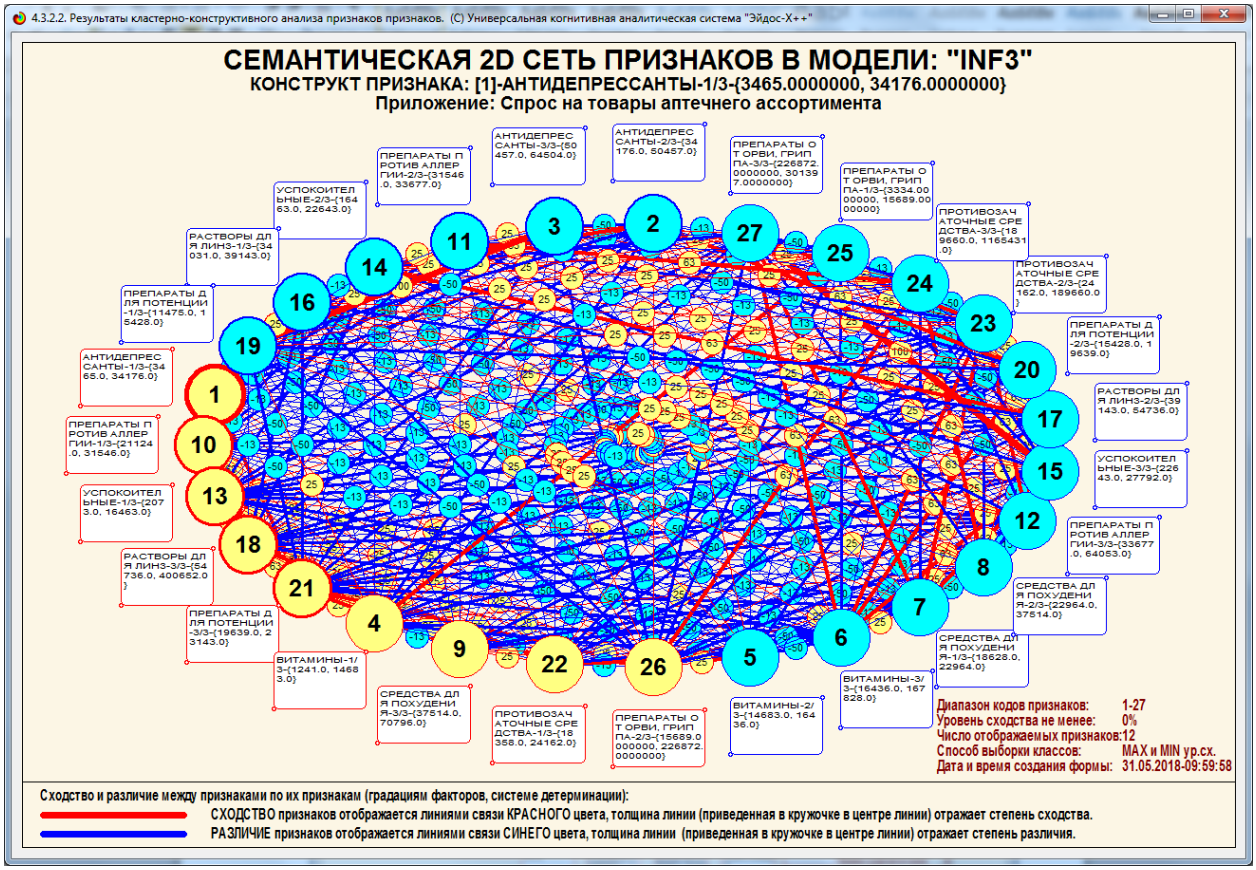


Рисунок 23 - Результаты кластерно-конструктивного анализа признаков

На приведенном графике, сложно различить данные. Поэтому использована возможность ограничить число признаков на этой диаграмме.

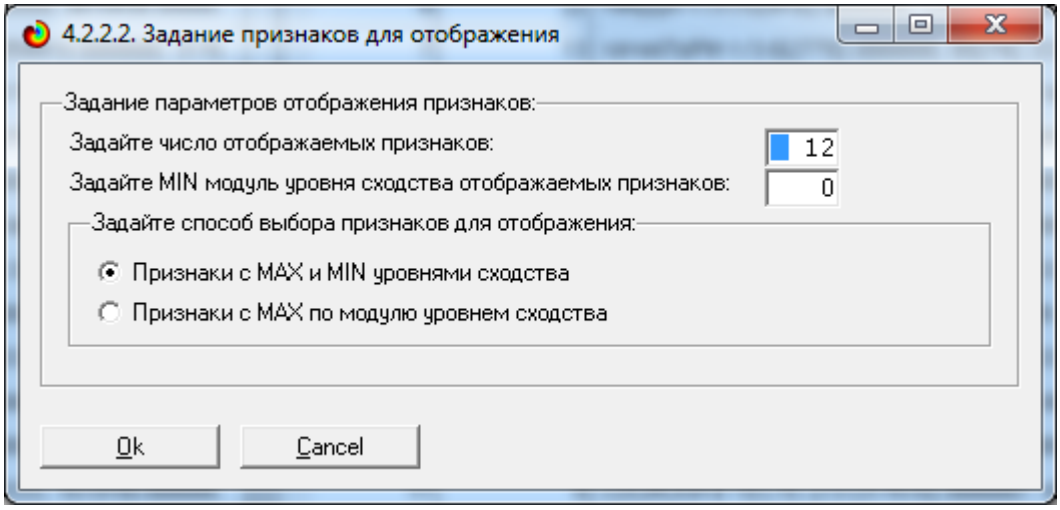


Рисунок 24 – Экранная форма, позволяющая ограничить число признаков на диаграмме

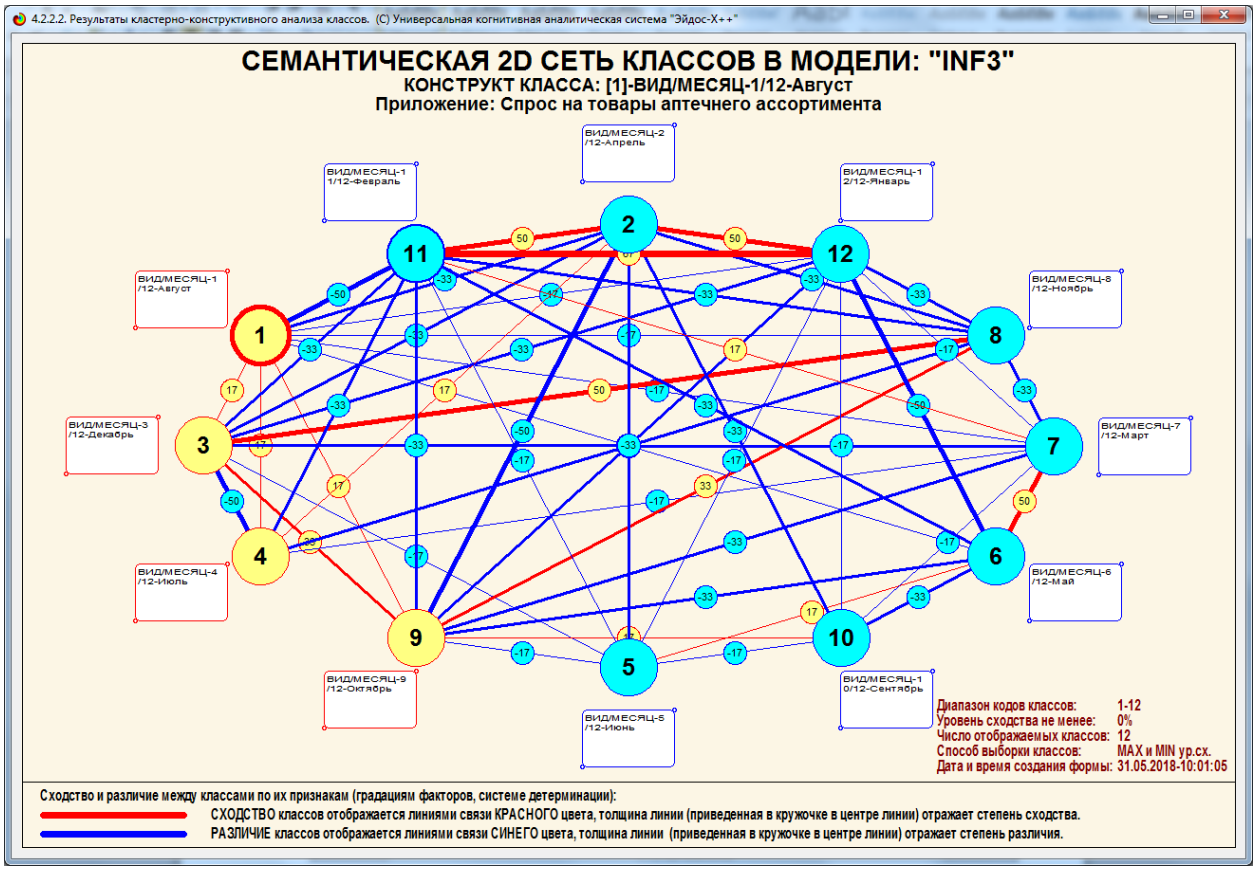


Рисунок 25 – Упрощенная когнитивная диаграмма признаков

Ниже приведены дендрограмма агломеративной кластеризации признаков, график изменения межкластерных расстояний и help режима 4.3.2.3, обеспечивающего их получение.

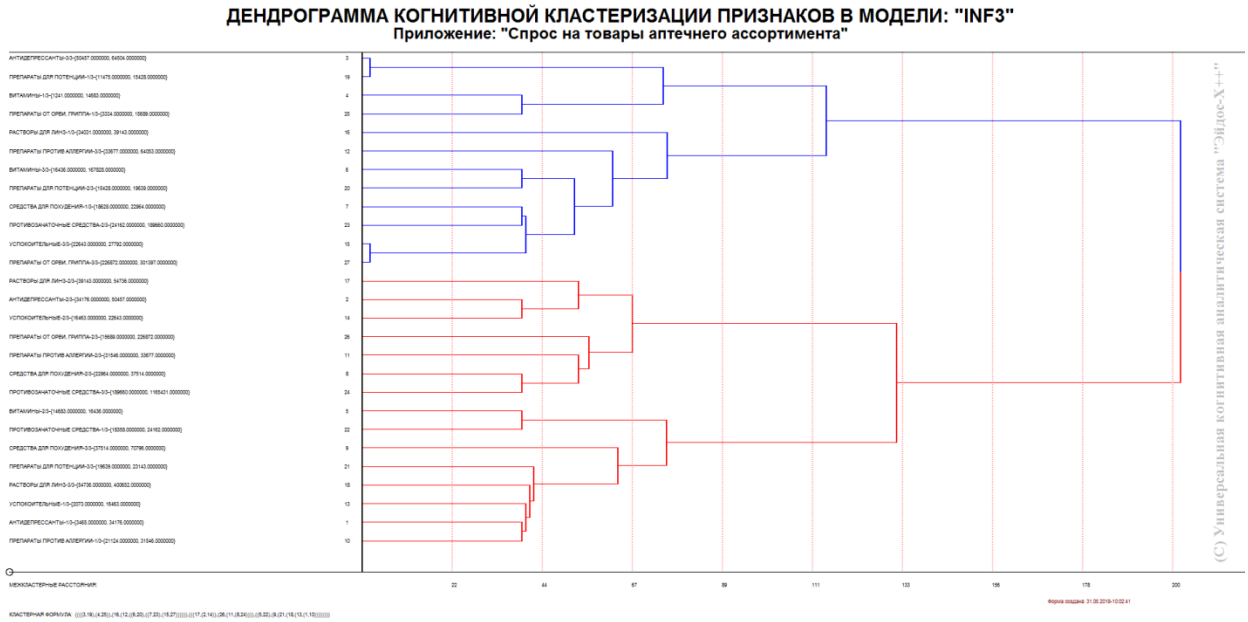


Рисунок 26 – Дендрограмма когнитивной кластеризации признаков

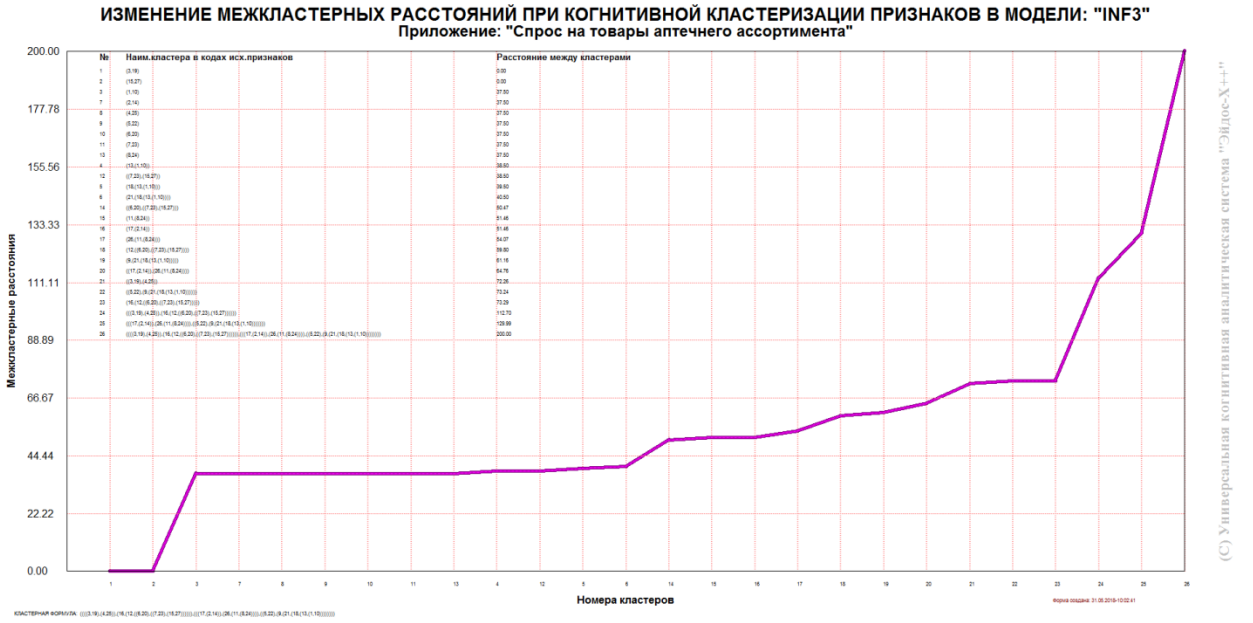


Рисунок 27 – Изменение межкластерных расстояний при когнитивной кластеризации признаков

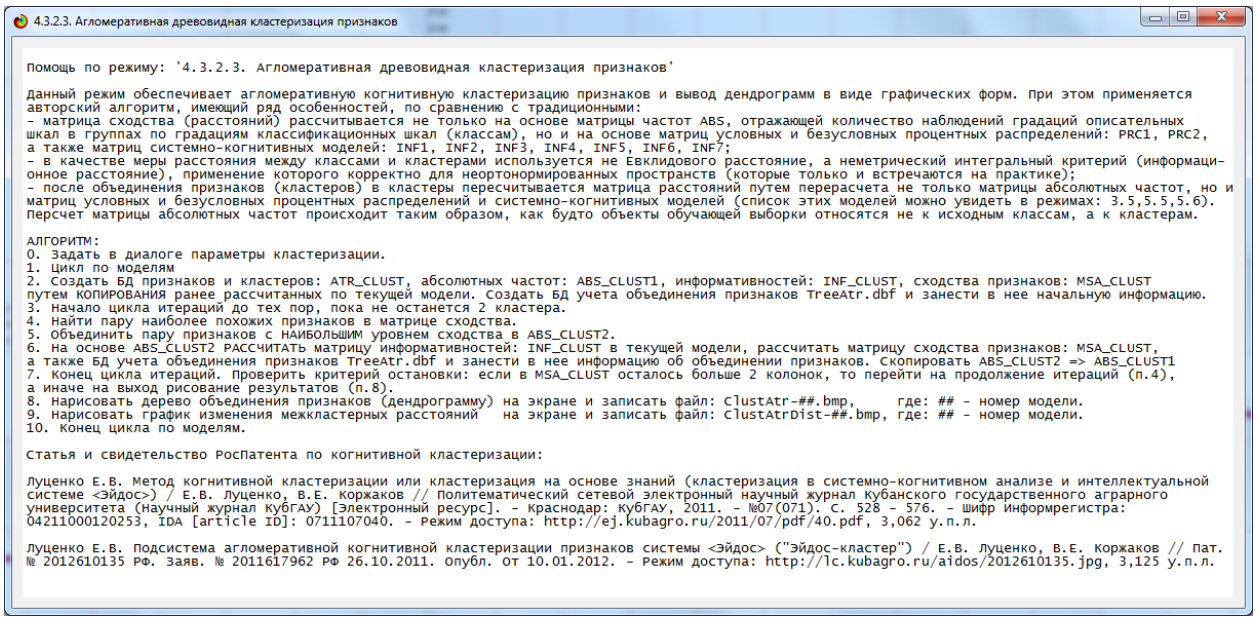


Рисунок 28 – Экран помощи по когнитивной кластеризации признаков

2.5 Кластерно-конструктивный анализ классов

На рисунке 29 приведены результаты кластерно-конструктивного анализа классов:

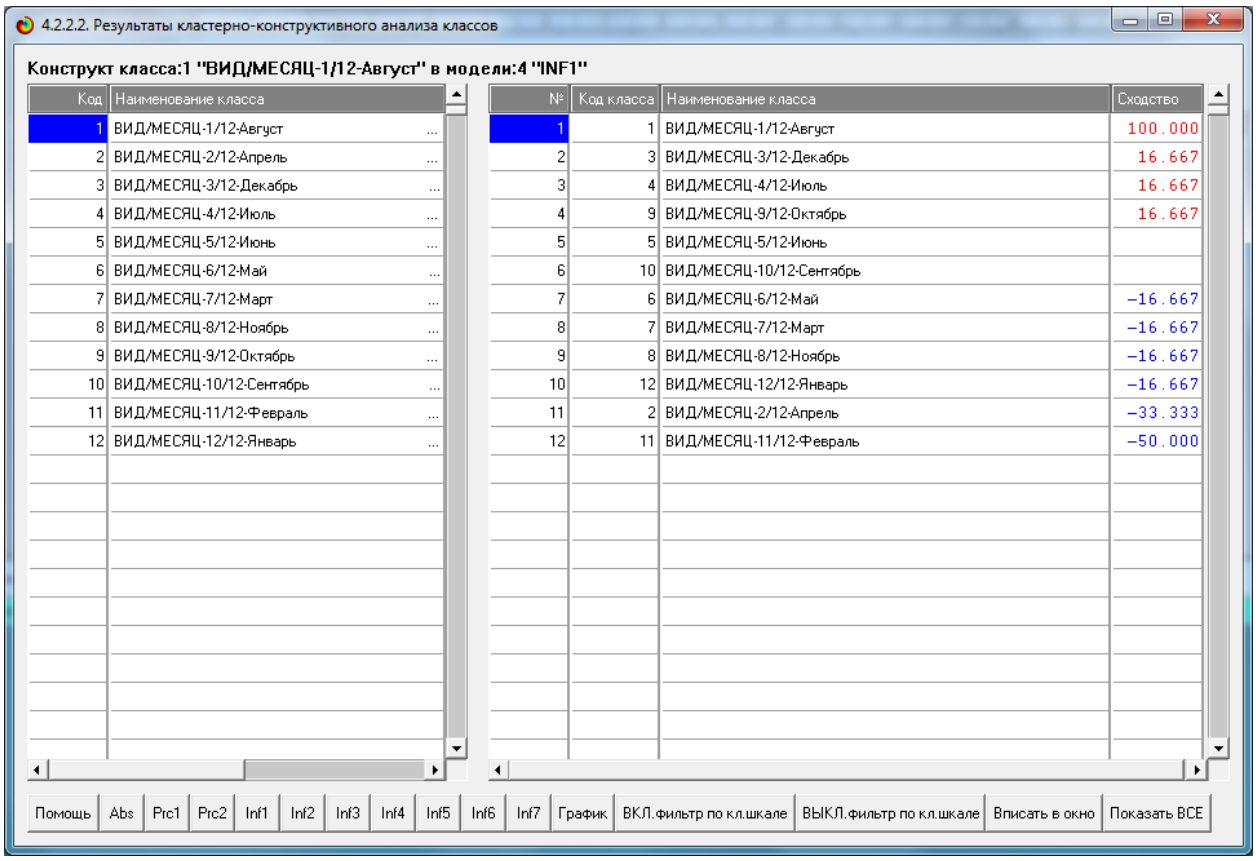


Рисунок 29 - Экранная форма, позволяющая ограничить число классов на диаграмме

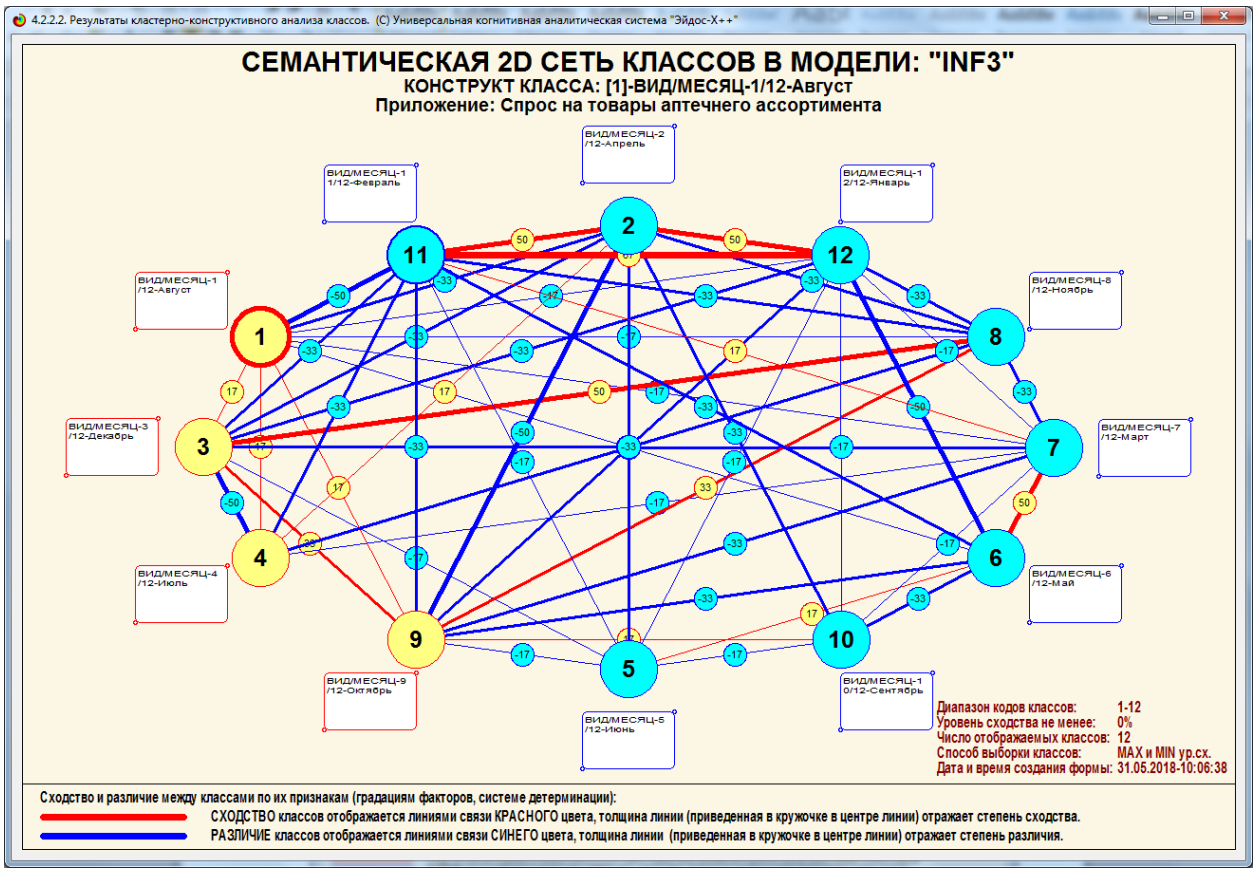


Рисунок 30 - Результаты кластерно-конструктивного анализа признаков

Ниже приведены дендрограмма агломеративной кластеризации классов, график изменения межкластерных расстояний и help режима 4.2.2.3, обеспечивающего их получение.

Рисунок 31 – Дендрограмма когнитивной кластеризации классов

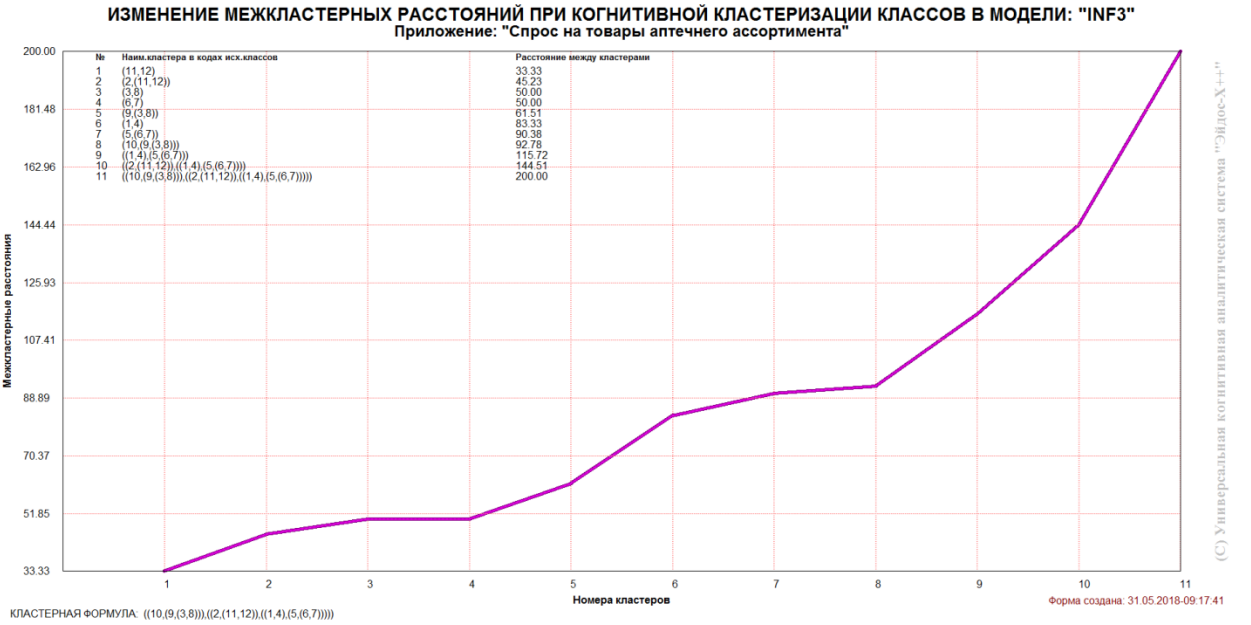


Рисунок 32 – Изменение межкластерных расстояний при когнитивной кластеризации классов

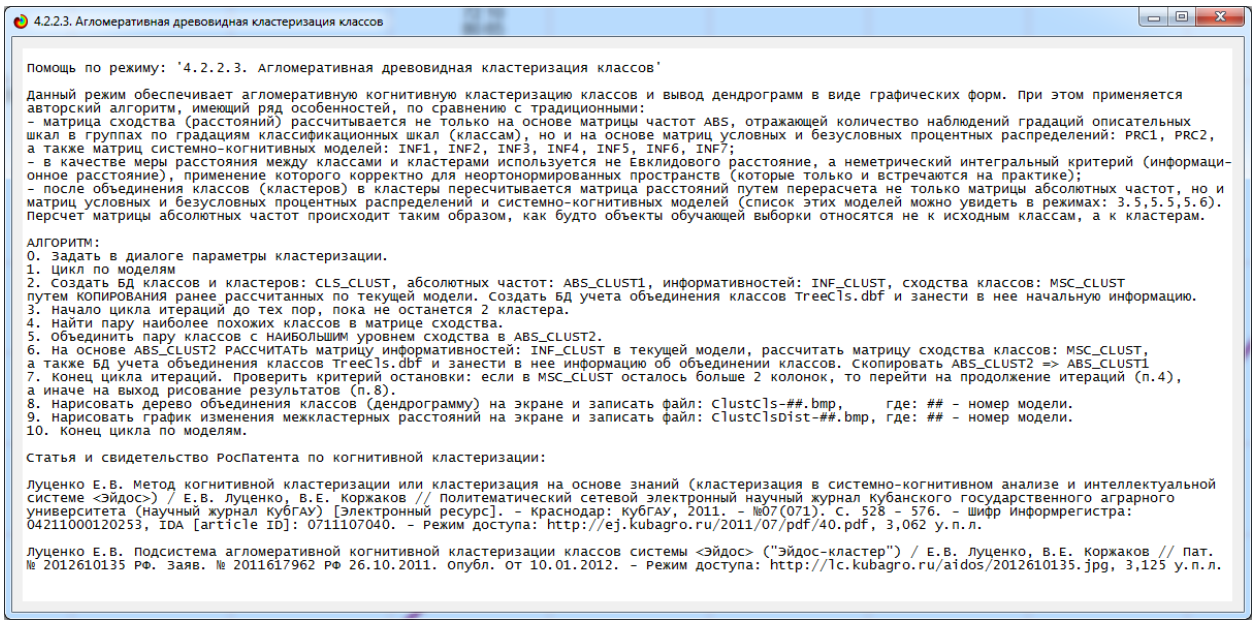


Рисунок 33 – Экран помощи по когнитивной кластеризации классов

2.6 Когнитивные функции

Обеспечивает отображение влияния факторов на результат. Описание режима 4.5

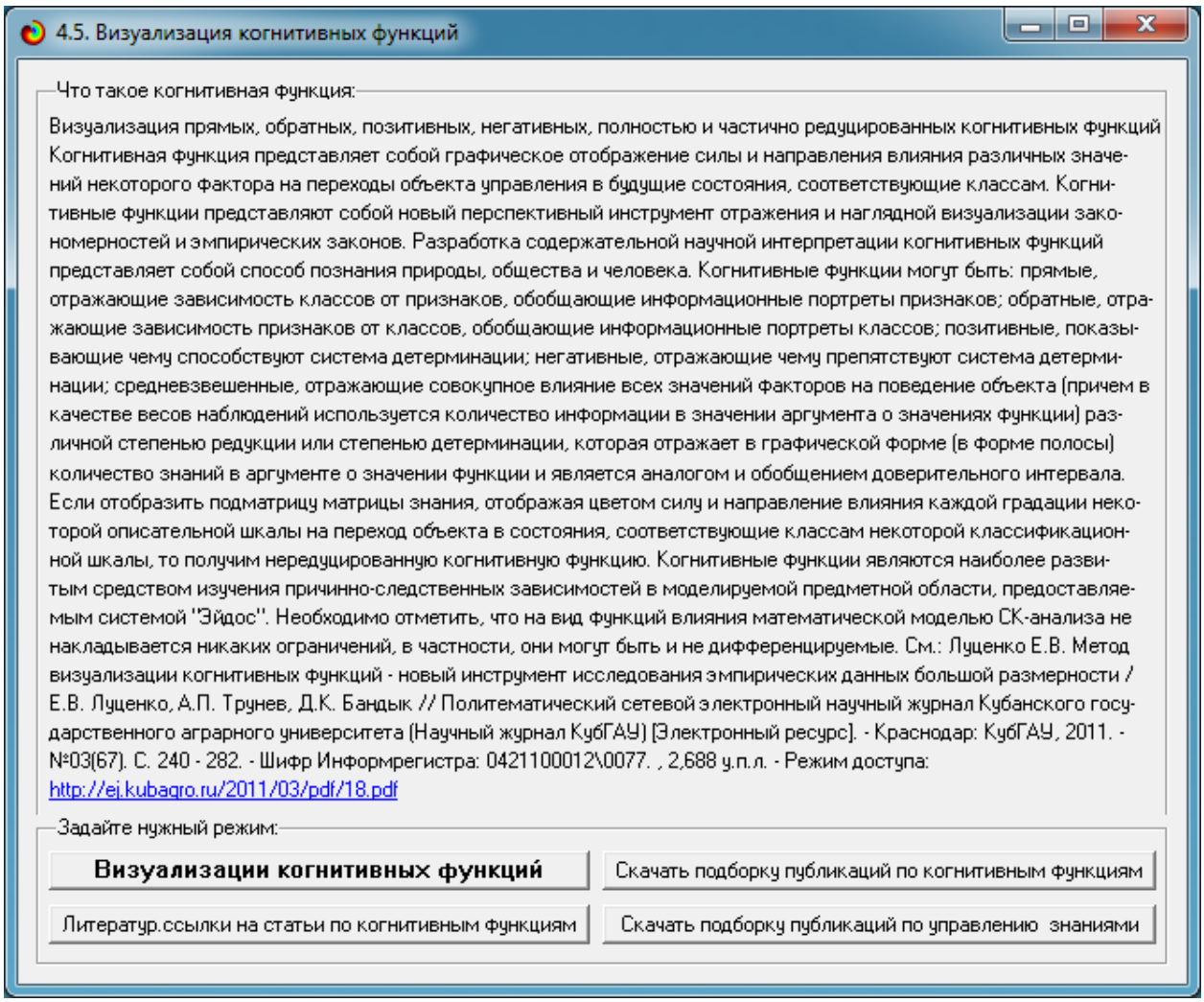


Рисунок 34 – Экранная форма режима 4.5

Ниже приведены когнитивные функции модели.

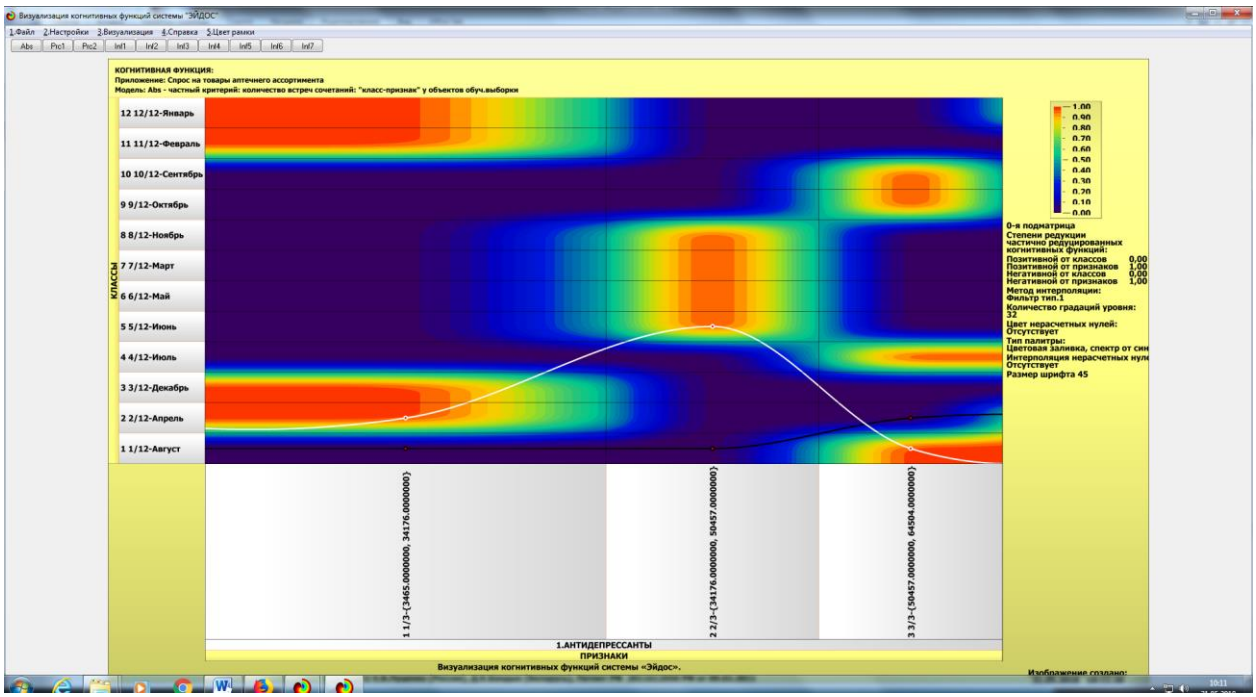


Рисунок 35 – Когнитивная функция

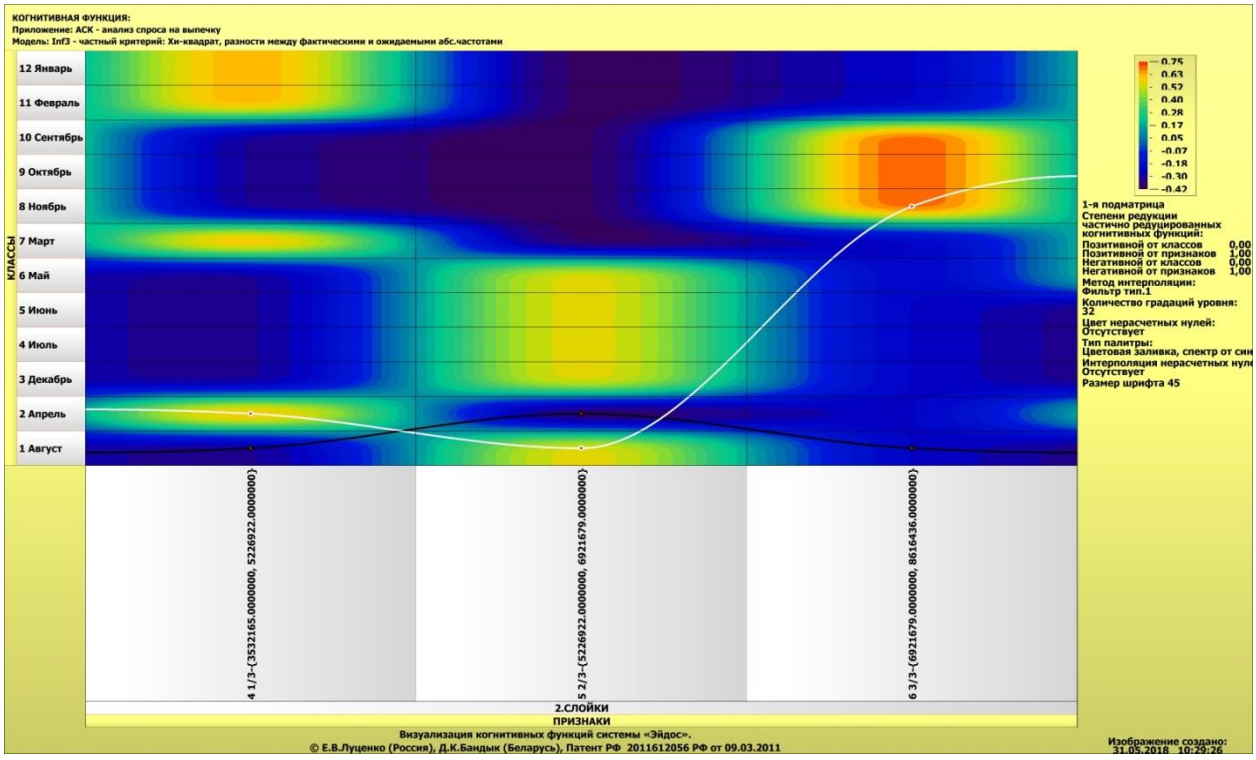


Рисунок 36 - Когнитивная функция

2.7 Геокогнитивная система

В данном режиме можно сформировать облако точек и провести его триангуляцию Делоне с выводом результатов.

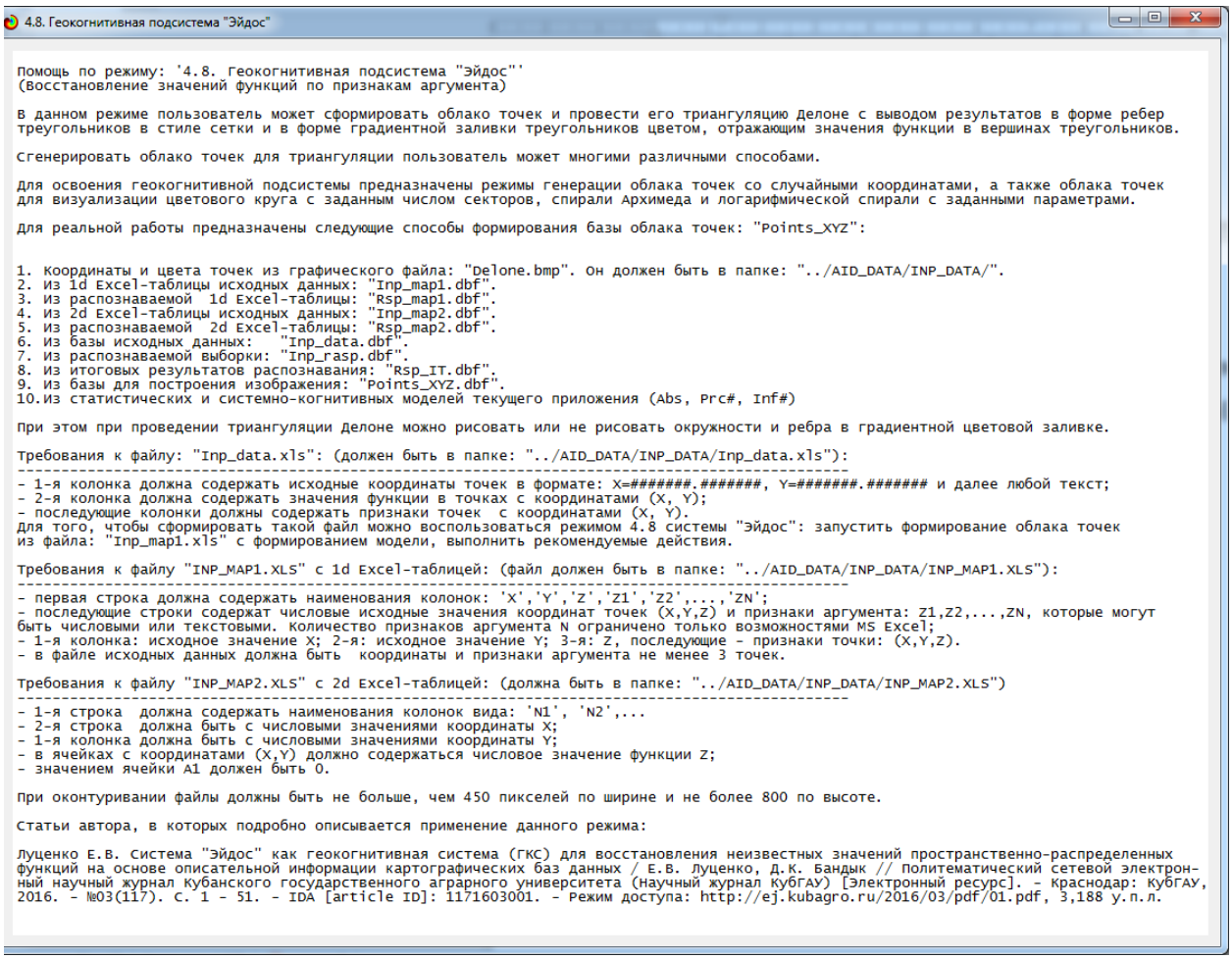


Рисунок 37 – Описание режима 4.8

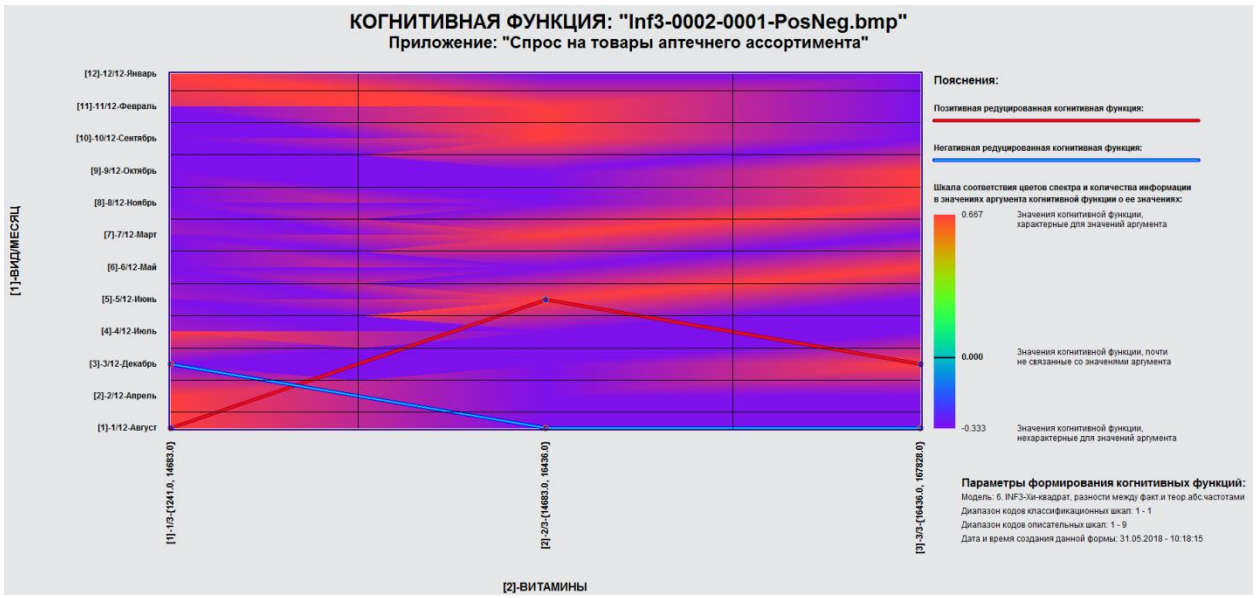


Рисунок 38 - Геокогнитивная модель лаваша

ЗАКЛЮЧЕНИЕ

Человеку трудно оценить достоверность систем искусственного интеллекта, поэтому необходимо сопоставить оценки качества их математических моделей. Одним из вариантов решения этой задачи является тестирование различных системы на общей базе исходных данных.

В данной работе приводится развернутый пример использования базы данных статистики РОССТАТ для оценки качества математических моделей, применяемых в АСК-анализе и его программном инструментарии системе искусственного интеллекта «Эйдос». При этом наиболее достоверной в данном приложении оказались модели INF3 при интегральном критерии «Сумма знаний». Точность модели составляет 0,580, что объясняется сравнительно небольшой выборкой данных. Для оценки достоверности моделей в АСК-анализе и системе «Эйдос» используется F-критерий Ван Ризбергена и его нечеткое мультиклассовое обобщение, предложенное проф.Е.В.Луценко. Также обращает на себя внимание, что статистические модели в данном приложении дают примерно на 20% более низкую средневзвешенную достоверность идентификации и не идентификации, чем модели знаний, что, как правило, наблюдается и в других приложениях. Этим и оправдано применение моделей знаний.

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

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. 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.

2. Луценко Е.В. Открытая масштабируемая интерактивная интеллектуальная on-line среда для обучения и научных исследований на базе АСК-анализа и системы «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №06(130). С. 1 – 55. – IDA [article ID]: 1301706001. – Режим доступа: http://ej.kubagro.ru/2017/06/pdf/01.pdf, 3,438 у.п.л.

3. ЕМИСС - https://fedstat.ru/indicator/31452

4. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с. http://elibrary.ru/item.asp?id=21745340

5. Симанков В.С., Луценко Е.В. Адаптивное управление сложными системами на основе теории распознавания образов. Монография (научное издание). – Краснодар: ТУ КубГТУ, 1999. - 318с. http://elibrary.ru/item.asp?id=18828433

6. Симанков В.С., Луценко Е.В., Лаптев В.Н. Системный анализ в адаптивном управлении: Монография (научное издание). /Под науч. ред. В.С.Симанкова. – Краснодар: ИСТЭК КубГТУ, 2001. – 258с.

http://elibrary.ru/item.asp?id=21747625
7. 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. http://elibrary.ru/item.asp?id=18632909
8. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности 351400 "Прикладная информатика (по отраслям)". – Краснодар: КубГАУ. 2004. – 633 с. http://elibrary.ru/item.asp?id=18632737
9. Луценко Е.В., Лойко В.И., Семантические информационные модели управления агропромышленным комплексом. Монография (научное издание).
– Краснодар: КубГАУ. 2005. – 480 с. http://elibrary.ru/item.asp?id=21720635
10.Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп.– Краснодар: КубГАУ, 2006. – 615 с. http://elibrary.ru/item.asp?id=18632602
11.Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд.,перераб. и доп. – Краснодар: КубГАУ, 2006. – 318с. http://elibrary.ru/item.asp?id=21683721
12.Наприев И.Л., Луценко Е.В., Чистилин А.Н. Образ-Я и стилевые особенности деятельности сотрудников органов внутренних дел в экстремальных условиях. Монография (научное издание). – Краснодар: КубГАУ. 2008. – 262 с. http://elibrary.ru/item.asp?id=21683724