

<p>УДК 004.8 05.13.18 - Математическое моделирование, численные методы и комплексы программ (технические науки)</p> <p>Выявление нетипичных объектов и артефактов в исходных данных, назначение на новые классы нетипичных объектов и удаление артефактов в математических моделях автоматизированного системно-когнитивного анализа</p> <p>Луценко Евгений Вениаминович д.э.н., к.т.н., профессор Web of Science ResearcherID S-8667-2018 Scopus Author ID: 57188763047 РИНЦ SPIN-код: 9523-7101 prof.lutsenko@gmail.com http://lc.kubagro.ru https://www.researchgate.net/profile/Eugene_Lutsenko</p> <p><i>Кубанский Государственный Аграрный университет имени И.Т. Трубилина, Краснодар, Россия</i></p> <p>В любой обучающей (тренировочной) выборке (исходных данных) всегда есть не только истинная информация о моделируемой предметной области, но и дезинформация, да и просто шум. Понятно, что шум и дезинформация в исходных данных приводят к понижению достоверности моделей, созданных на основе этих исходных данных. Поэтому необходимо иметь критерии, позволяющие отличить шум от дезинформации и от истинной информации, а также основанные на этих критериях математические и программные инструменты для выявления и подавления шума в исходных данных, а также для выявления в исходных данных дезинформации и восстановления истинной информации путем анализа дезинформации. Артефактами будем называть объекты обучающей выборки, у которых и/или признаки случайны, и/или классы случайны, и/или связь признаков этих объектов с принадлежностью этих объектов классам также случайна. В этом и состоит шум в исходных данных. Нетипичными будем называть объекты обучающей выборки, у которых признаки, и классы не случайны, и связь признаков (этых объектов) с принадлежностью (этих объектов) к классам, указанная в обучающей выборке, также не случайна, а вполне закономерна, но не та, которая указана в обучающей выборке, а другая, т.е. в обучающей выборке неверно указана принадлежность объектов к классам. В этом и заключается дезинформация в исходных данных. Нетипичные объекты в действительности могут относиться как к тем классам, которые указаны в обучающей выборке, так и к новым классам, которых там нет. Основным критерием, позволяющим отличить шум и дезинформацию от истинной информации является уровень сходства объекта с классами, к которым он относится по данным обучающей выборки. При увеличении уровня сходства закономерно расчет доля истинных решений среди всех решений. Низкий уровень сходства и, особенно, ложно-отрицательные решения при решении задачи идентификации, являются признаками шума и дезинформации в описании объектов обучающей выборки. Отличить шум и дезинформацию друг от друга можно по следующему критерию: если гипотеза о том, что объект нетипичный не подтверждается при итерационном назначении его на новые специально создаваемые для этого классы, т.е. этот процесс «зацикливается» (приводит к повторению ситуации) без повышения достоверности моделей, то описание этого объекта не содержит закономерностей и является шумом, т.е. он является артефактом. Такие объекты надо просто удалять из обучающей выборки. Но если вместе с артефактами удалить и нетипичные объекты, то это приведет к уменьшению</p>	<p>УДК 004.8 05.13.18 - Mathematical modeling, numerical methods and software packages (technical sciences)</p> <p>Identification of atypical objects and artifacts in the source data, assignment to new classes of atypical objects and removal of artifacts in mathematical models of automated system-cognitive analysis</p> <p>Lutsenko Evgeniy Veniaminovich Doctor of Economics, Ph.D., professor Web of Science ResearcherID S-8667-2018 Scopus Author ID: 57188763047 RSCI SPIN code: 9523-7101 prof.lutsenko@gmail.com http://lc.kubagro.ru https://www.researchgate.net/profile/Eugene_Lutsenko</p> <p><i>Kuban State Agrarian University named after I.T. Trubilin, Krasnodar, Russia</i></p> <p>In any training (training) sample (source data) there is always not only true information about the simulated subject area, but also misinformation, and just noise. It is clear that noise and misinformation in the source data lead to a decrease in the reliability of models created on the basis of these source data. Therefore, it is necessary to have criteria to distinguish noise from disinformation and from true information, as well as mathematical and software tools based on these criteria for detecting and suppressing noise in the source data, as well as for detecting disinformation in the source data and restoring true information by analyzing disinformation. Artifacts will be called objects of the training sample, in which and /or features are random, and /or classes are random, and /or the relationship of the features of these objects with the belonging of these objects to classes is also random. This is the noise in the source data. We will call the objects of the training sample atypical, in which both the signs and classes are not random, and the relationship of the signs (of these objects) with the belonging (of these objects) to the classes indicated in the training sample is also not random, but quite natural, but not the one indicated in the training sample, but the other, i.e.e. the class membership of objects is incorrectly indicated in the training sample. This is the misinformation in the source data. Atypical objects can actually belong both to those classes that are specified in the training sample, and to new classes that are not there. The main criterion for distinguishing noise and misinformation from true information is the level of similarity of the object with the classes to which it belongs according to the training sample. With an increase in the level of similarity, the proportion of true solutions among all solutions is naturally calculated. A low level of similarity and, especially, false-negative solutions when solving the identification problem are signs of noise and misinformation in the description of the objects of the training sample. It is possible to distinguish noise and disinformation from each other by the following criterion: if the hypothesis that an object is atypical is not confirmed when iteratively assigning it to new classes specially created for this purpose, i.e. this process "loops" (leads to a repetition of the situation) without increasing the reliability of models, the description of this object does not contain patterns and is noise, i.e. it is an artifact. Such objects should simply be removed from the training sample. But if atypical objects are removed along with artifacts, this will lead to a decrease in the amount of information in the model, its impoverishment (which is called: "pour the baby out of the bath along with the water"). Therefore, in the automated system-cognitive analysis (ASK-analysis) and its software tools, the intelligent system "Eidos" implements both mechanisms based on the two criteria given, allowing first to divide classes into typical and atypical parts with the creation of new classes for</p>
---	--

<p>количества информации в модели, ее обеднению (что называется: «вылить из ванны вместе с водой и ребенком»). Поэтому в автоматизированном системно-когнитивном анализе (ACK-анализ) и его программном инструментарии интеллектуальной системе «Эйдос» реализованы оба механизма на основе двух приведенных критериев, позволяющие сначала разделять классы на типичную и нетипичную части с созданием новых классов для нетипичных объектов, т.е. восстанавливать истинную информацию путем анализа дезинформации, а затем эффективно подавлять шум в исходных данных. В работе приводятся подробные численные примеры, демонстрирующие эти подходы на модельных исходных данных.</p> <p>Ключевые слова: ACK-АНАЛИЗ, АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, КОГНИТИВНАЯ, ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА «ЭЙДОС», ШУМ, ДЕЗИНФОРМАЦИЯ, ОБУЧАЮЩАЯ ВЫБОРКА, ТИПИЧНЫЕ И НЕТИПИЧНЫЕ ОБЪЕКТЫ, АРТЕФАКТЫ, ДИВИЗИОННАЯ КЛАСТЕРИЗАЦИЯ http://dx.doi.org/10.21515/1990-4665-184-012</p>	<p>atypical objects, i.e. recover true information by analyzing misinformation, and then effectively suppress noise in the source data. The paper provides detailed numerical examples demonstrating these approaches on model source data.</p> <p>Keywords: ASK ANALYSIS, AUTOMATED SYSTEM-COGNITIVE ANALYSIS, COGNITIVE, INTELLIGENT SYSTEM "EIDOS", NOISE, DISINFORMATION, TRAINING SAMPLE, TYPICAL AND ATYPICAL OBJECTS, ARTIFACTS, DIVISIONAL CLUSTERING</p>
--	--

СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ	3
2. МАТЕРИАЛЫ (ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ).....	5
3. МЕТОД	8
4. РЕЗУЛЬТАТЫ	9
4.1. Этапы ACK-анализа и их реализация в системе «Эйдос»	9
4.2. Алгоритм выявления нетипичных объектов и артефактов в исходных данных, назначения на новые классы нетипичных объектов и удаления артефактов	10
4.3. Создание исходной модели для итераций	12
Шаг 1. Формализация предметной области	12
Шаг 2. Синтез и верификация статистических и системно-когнитивных моделей	16
Шаг 3. Оценка достоверности статистических и системно-когнитивных моделей.....	19
Шаг 4. Идентификация в наиболее достоверной модели	19
4.4. Шаг 5. Начало 1-й итерации: выявление нетипичных объектов и их назначение на новые классы (разделение классов на типичную и нетипичную части).....	25
4.5. Шаг 6. Проверка условий завершения процесса итераций	27
4.6. Создание модели 1-й итерации	28
Шаг 1. Формализация предметной области	28
Шаг 2. Синтез и верификация статистических и системно-когнитивных моделей	31
Шаг 3. Оценка достоверности статистических и системно-когнитивных моделей.....	31
Шаг 4. Идентификация в наиболее достоверной модели	32
4.7. Шаг 5. Начало 2-й итерации: выявление нетипичных объектов и их назначение на новые классы (разделение классов на типичную и нетипичную части).....	38
4.8. Шаг 6. Проверка условий завершения процесса итераций	41
4.9. Шаг 7. Выявление и удаление артефактов	41
5. ОБСУЖДЕНИЕ	44
6. ВЫВОДЫ	45
ЛИТЕРАТУРА	46
ПРИЛОЖЕНИЯ:.....	48

1. Введение

В любой обучающей (тренировочной) выборке (исходных данных) всегда есть не только истинная информация о моделируемой предметной области, но и дезинформация, да и просто шум. Понятно, что шум и дезинформация в исходных данных приводят к понижению достоверности моделей, созданных на основе этих исходных данных. Поэтому необходимо иметь критерии, позволяющие отличить шум от дезинформации и от истинной информации, а также основанные на этих критериях математические и программные инструменты для выявления и подавления шума в исходных данных, а также для выявления в исходных данных дезинформации и восстановления истинной информации путем анализа дезинформации.

Артефактами будем называть объекты обучающей выборки, у которых и/или признаки случайны, и/или классы случайны, и/или связь признаков этих объектов с принадлежностью этих объектов к классам также случайна.

В этом и состоит **шум** в исходных данных.

Нетипичными будем называть объекты обучающей выборки, у которых и признаки, и классы не случайны, и связь признаков (этих объектов) с принадлежностью (этих объектов) к классам, указанная в обучающей выборке, также не случайна, а вполне закономерна, но не та, которая указана в обучающей выборке, а другая, т.е. в обучающей выборке неверно указана принадлежность объектов к классам.

В этом и заключается **дезинформация** в исходных данных.

Нетипичные объекты в действительности могут относиться как к тем классам, которые указаны в обучающей выборке, так и к новым классам, которых там нет. Основным **критерием**, позволяющим **отличить шум и дезинформацию от истинной информации** является уровень сходства объекта с классами, к которым он относится по данным обучающей выборки. *При увеличении уровня сходства закономерно расчет доли истинных решений среди всех решений.* Низкий уровень сходства и, особенно, ложно-отрицательные решения при решении задачи идентификации, являются признаками шума и дезинформации в описании объектов обучающей выборки.

Отличить шум и дезинформацию друг от друга можно по следующему **критерию**: если гипотеза о том, что объект нетипичный не подтверждается при итерационном назначении его на новые специально создаваемые для этого классы, т.е. этот процесс «зацикливается» (приводит к повторению ситуации) без повышения достоверности моделей, то описание этого объекта не содержит закономерностей и является шумом, т.е. он является артефактом.

Такие объекты надо просто удалять из обучающей выборки. Но если вместе с артефактами удалить и нетипичные объекты, то это приведет к

уменьшению количества информации в модели, ее обеднению (что называется: «вылить из ванны вместе с водой и ребенком»).

Поэтому в автоматизированном системно-когнитивном анализе (АСК-анализ) и его программном инструментарии интеллектуальной системе «Эйдос» реализованы оба механизма на основе двух приведенных критериев, позволяющие **сначала** разделять классы на типичную и нетипичную части с созданием новых классов для нетипичных объектов, т.е. восстанавливать истинную информацию путем анализа дезинформации, а **затем** эффективно подавлять шум в исходных данных.

В работе приводятся подробные численные примеры, демонстрирующие эти подходы на модельных исходных данных.

Традиционно артефакты и нетипичные объекты **не различают** и называют просто артефактами. Артефакты, особенно не задумываясь, просто удаляют из выборки. Правда где-то на краю сознания при этом остается смутное ощущение, что поступая так возможно мы теряем что-то важное, неизвестное об объекте моделирования, о чем мы уже никогда не узнаем. Когда артефакты являются шумом, просто искажающими картину, то это нормально. Но в ряде случаев подобный традиционный подход неприемлем просто потому, что нетипичные объекты обучающей выборки, также как и типичные, действительно относятся к тем классам, для формирования обобщенных образов которых предполагалось их использовать и точно не являются шумом и артефактами, а являются просто дезинфекцией. Поэтому удаление нетипичных объектов из обучающей выборки хотя и позволяет получить формально более корректную модель предметной области, однако это достигается неприемлемой ценой: ценой того, что модель перестает полностью отражать эту предметную область и начинает отражать лишь ее часть.

Решение этой проблемы дается методом дивизионной (или дивизивной) кластеризации, которая отличается от агломеративной (объединительной) кластеризации тем, что кластеры образуются не путем объединения классов, а путем их разделения на типичную и нетипичную части. *При этом нетипичные объекты удаляются не из всей обучающей выборки, а только из тех классов, к которым они относятся и являются нетипичными.* В этом случае *для нетипичных объектов обучающей выборки могут создаваться специальные классы, с теми же наименованиями, что и у классов, из которых эти объекты удалены*, но с добавлением к наименованию класса информации о том, что он создан на определенной итерации для нетипичных объектов (например, может добавляться номер итерации).

Такой вариант дивизивной кластеризации реализован еще в DOS-версии системы «Эйдос» в 2008 году. Однако при разработке новой версии системы «Эйдос» под MS Windows в 2012 году этот режим полностью не вошел в новую систему и был в ней существенно переработан. В частности

он стал работать непосредственно с внешней базой исходных данных для API-2.3.2.2 системы «Эйдос» и из него были убраны итерации, которые в новых версиях также возможны, но теперь осуществляются вручную [1-8].

В данной работе делается следующий шаг по совершенствованию режима разделения классов на типичную и нетипичные части и удаления артефактов. Конечно, здесь возникает вопрос о том, что считать нетипичными объектами обучающей выборки или конкретнее: каковы критерии, позволяющие отличить нетипичные объекты обучающей выборки от типичных.

Это сделано путем уточнения критериев, позволяющих на практике отличить с одной стороны нетипичные объекты и артефакты от истинных данных, а с другой – нетипичные объекты от артефактов. Если ранее нетипичными объектами считались те, которые приводили к ложно-отрицательным решениям, то в новой реализации данного режима, описанном в данной работе, пользователь сам в диалоге может задать минимальный уровень сходства объектов с классом больше нуля, при котором объекты еще считаются типичными, а ниже являются уже нетипичными. При этом объекты, порождающие ложно-отрицательные решения, т.е. имеющие отрицательный уровень сходства с классом, также всегда считаются нетипичными. Удалять нетипичные объекты из обучающей выборки данный режим также позволяет вместе с артефактами.

Но лучше сначала разделить классы на типичные и нетипичные части, а уже после этого удалить артефакты, т.к. тогда нетипичные объекты останутся в обучающей выборке, но уже будут относиться к новым, специально созданным для них классам, т.е. уже станут типичными для этих классов, будут типизированы

В результате работы данного режима повышается степень адекватности и степень адаптивности модели, т.к. динамика предметной области часто приводит к тому, что появляются новые нетипичные для предыдущих периодов объекты, относящиеся как к уже имеющимся в модели, так и к новым классам.

Таким образом, данный режим является обобщением предыдущей реализации.

2. Материалы (подготовка исходных данных)

Для рассмотрения предложенного подхода на численном примере сформируем стандартными средствами MS Excel модельную обучающую выборку таким образом, чтобы для формирования каждого класса использовались объекты обучающей выборки с различной, известной и изменяющейся степенью истинности, типичности и случайности (таблица 1):

Таблица 1 – Исходные данные (полностью)

Разработка автора.

Путь на файл: c:\Aidos-X\AID DATA\Inp_data\Inp_data.xlsx

В таблице 9 приложения приведена таблица MS Excel с показанными формулами расчета таблицы 1. Для получения таблицы 1 из таблицы 5 в ней сделан лист: «Inp_data», в который все значения с расчетного листа 1 скопированы без формул с сохранением исходного форматирования.

В обучающей выборке приведено 180 примеров объектов, каждый из которых описан 10 признаками с кодами от 01 до 10.

Эти объекты используются для формирования обобщенных образов 10 классов.

На каждый класс приходится 10 примеров объектов (группа наблюдений).

Объекты каждой группы наблюдений описаны признаками, совпадающим по наименованию с наименованием класса и с добавлением шума различного вида и уровня. Таким образом, *различные объекты группы наблюдений, соответствующие каждому из классов, обладают различной степенью типичности и нетипичности для него, причем эта степень известна.*

В каждой группе наблюдений представлены:

– 12 объектов с *истинными* описаниями признаков и принадлежности к классам (без фона);

– 2 объекта со *случайными* признаками и истинной принадлежностью к классам (артефактами, розовый фон);

– 4 объекта с *истинными закономерными* признаками, другими, чем у истинно описанных объектов, но с ошибочно указанными классами (нетипичные объекты, *дезинформация*) (светло-зеленый фон).

Для того чтобы система могла отличить истинные описания объектов от шума и дезинформации и шум и дезинформацию друг от друга, необходимо чтобы в обучающей выборке по каждому классу было приведено *несколько примеров* объектов всех этих категорий. Если шум, дезинформация и истинное описание объекта обучающей выборки представлено по одному разу, то отличить их друг от друга нет никакой возможности.

Для того чтобы система могла отличить дезинформацию от шума, т.е. нетипичные объекты от артефактов, нетипичных объектов должно быть больше, чем артефактов по каждому классу. В наших модельных исходных данных (таблица 1) в группах наблюдений по каждому классу нетипичных объектов в два раза больше, чем артефактов: 4 и 2 соответственно.

Для того чтобы система могла отличить истинные описания объектов с одной стороны от шума и дезинформации, т.е. от артефактов и нетипичных объектов с другой стороны, истинных описаний должно быть больше, чем артефактов и нетипичных объектов вместе взятых по каждому классу. В наших модельных исходных данных (таблица 1) в группах наблюдений по каждому классу истинных описаний объектов в два раза больше, чем нетипичных объектов и артефактов вместе взятых: 12 и 4 + 2 соответственно.

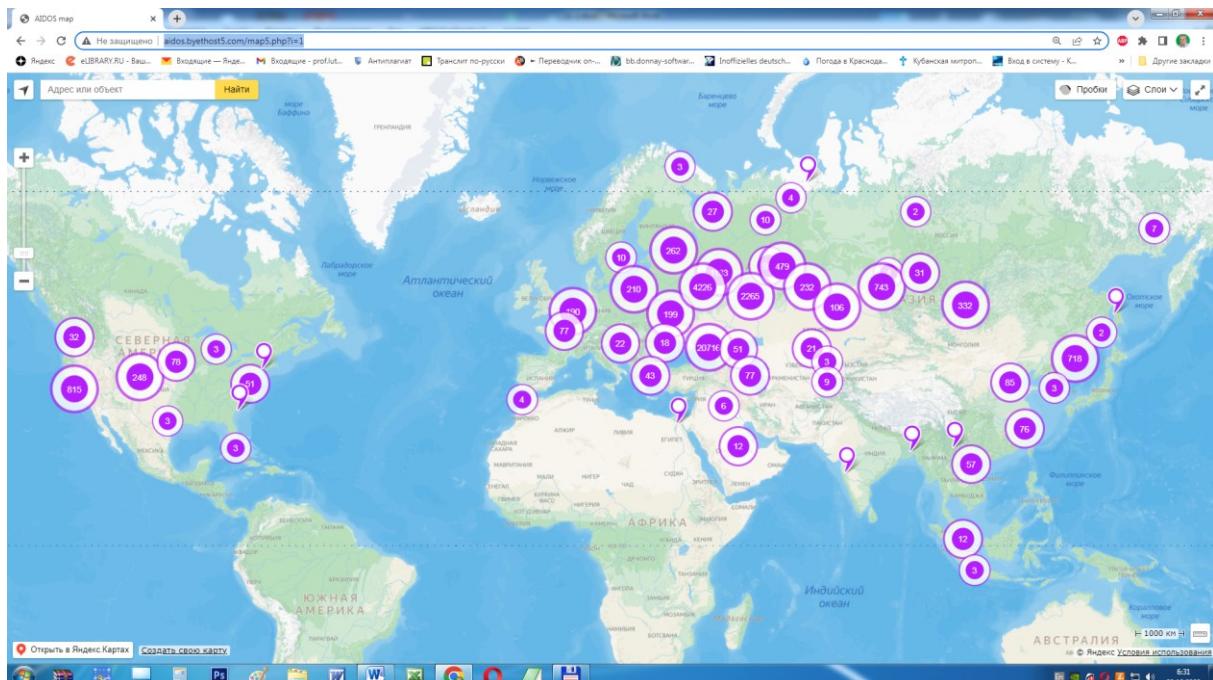
В качестве шума выступают псевдослучайные числа от 0 до 99, генерируемые стандартными средствами MS Excel. Надо отметить, что в качестве шума могут быть сгенерированы и признаки, которые и действительно соответствуют классу, к которому относится объект.

Реальные объекты в обучающих выборках можно рассматривать как взвешенные суперпозиции (линейные комбинации) объектов приведенных

типов. Поэтому подходы к выявлению артефактов и разделению классов на типичную и нетипичную части в реальных обучающих выборках можно изучать на примере исследуемой в данной работе модельной обучающей выборки, представленной в работе 1.

3. Метод

В качестве метода исследования предложенного подхода использован Автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий, в качестве которого в настоящее время выступает интеллектуальная система «Эйдос». Данный метод широко и успешно применяется во всем мире (рисунок 1), широко освещен в большом количестве публикаций (более 700) автора и соавторов на русском и английском языках [1, 2] и в данной работе приводить это описание еще раз целесообразно.



**Рисунок 1. Запуски системы «Эйдос» в мире
(по состоянию с конца 2016 года по конец 2022 года)**
[\(http://aidos.byethost5.com/map5.php\)](http://aidos.byethost5.com/map5.php)

Метод разделения классов на типичную и нетипичную части разработан автором в процессе проведения исследования [3] и описан в ряде работ [4-6]. На данный метод есть свидетельство РосПатента РФ [7].

4. Результаты

4.1. Этапы АСК-анализа и их реализация в системе «Эйдос»

Титульная видеограмма системы «Эйдос» приведена на рисунке 2:

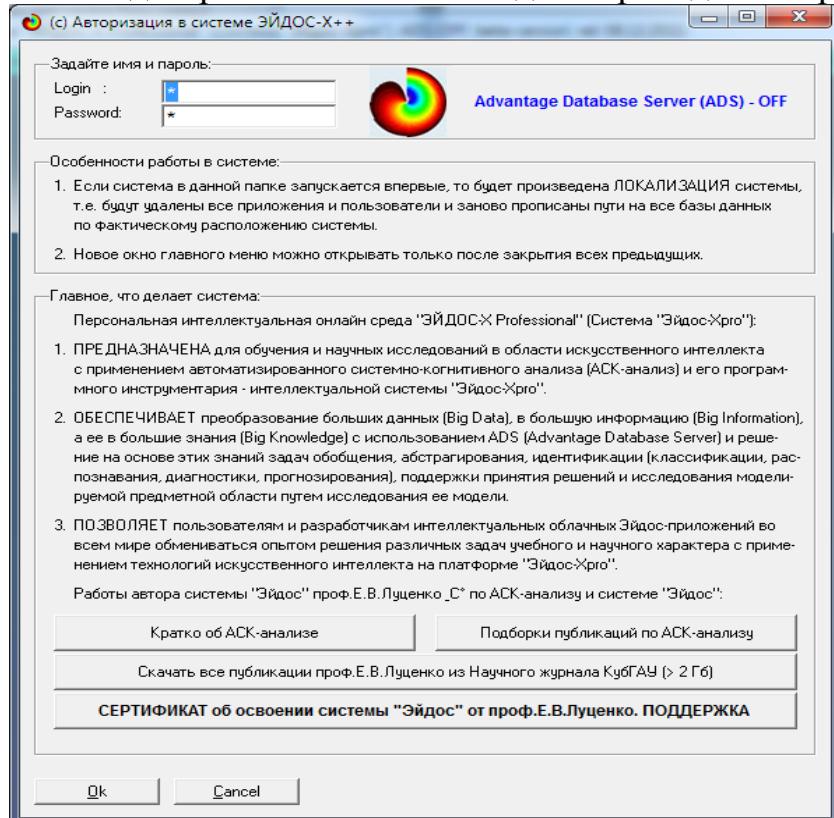


Рисунок 2. Титульная видеограмма системы «Эйдос» (один из вариантов)

АСК-анализ включает следующие этапы (рисунок 3):

1. Когнитивно-целевая структуризация моделируемой предметной области (единственный не автоматизированный в системе «Эйдос» этап АСК-анализа).
2. Формализация предметной области:
 - разработка классификационных и описательных шкал и градаций;
 - кодирование исходных данных с помощью классификационных и описательных шкал и градаций и формирование обучающей выборки.
3. Синтез статистических и системно-когнитивных моделей.
4. Верификация статистических и системно-когнитивных моделей.
5. Выбор наиболее достоверной модели.
6. Решение задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области в наиболее достоверной модели.

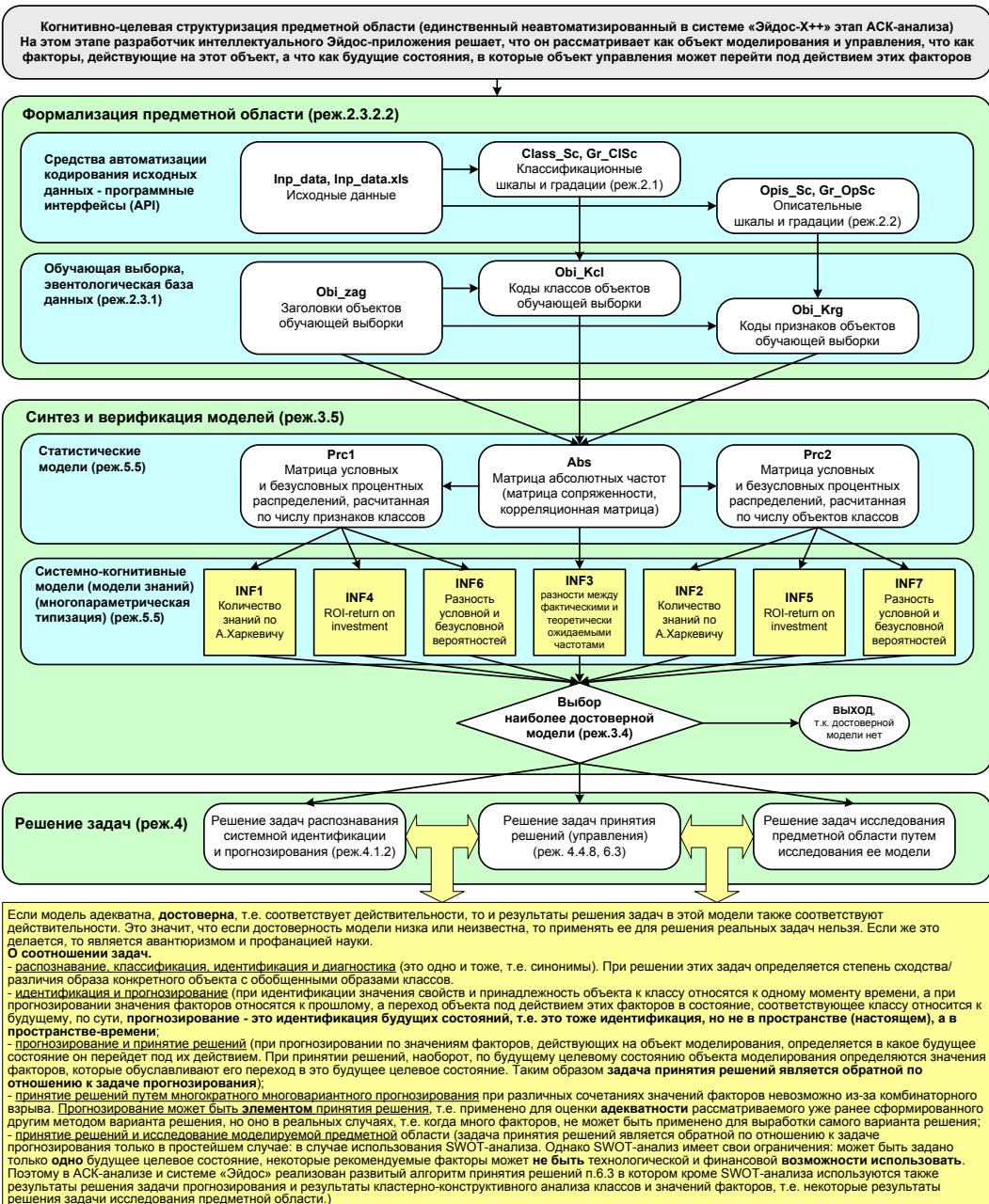


Рисунок 3. Этапы автоматизированного системно-когнитивного анализа

4.2. Алгоритм выявления нетипичных объектов и артефактов в исходных данных, назначения на новые классы нетипичных объектов и удаления артефактов

В предлагаемом алгоритме, приведенные выше этапы АСК-анализа повторяются **итерационно** для выявления нетипичных объектов и назначения их на новые классы, а удаление артефактов осуществляется уже **после** выхода из процесса итераций, когда в их продолжении исчезает смысл.

**АЛГОРИТМ
выявления нетипичных объектов и артефактов в исходных данных,
назначения на новые классы нетипичных объектов и удаления
артефактов в математических моделях АСК-анализа**

Шаг 1. Формализация предметной области в API-2.3.2.2 системы «Эйдос».

Шаг 2. Синтез и верификация статистических и системно-когнитивных моделей в режиме 3.5.

Шаг 3. Оценка достоверности статистических и системно-когнитивных моделей в режиме 3.4.

Шаг 4. Идентификации в наиболее достоверной модели в режимах 5.6 и 4.1.2.

Шаг 5. Начало очередной итерации. Выявление нетипичных объектов и их назначение на новые классы (разделение классов на типичную и нетипичную части) в режиме 3.6.

Шаг 6. Проверка условий завершения процесса итераций и его **завершение** и переход на шаг 7 при выполнении одного или нескольких из следующих условий:

- назначено на новые классы 0 (или другое не изменяющееся в итерациях количество) объектов обучающей выборки;
- достоверность модели достигает приемлемого уровня;
- достоверность модели не меняется в итерациях;
- в итерациях одни и те же объекты назначаются на новые классы (процесс «зацикливается»).

А иначе **продолжение** процесса итераций, т.е. переход на шаг 1

Шаг 7. Выявление и удаление артефактов в режиме 3.6.

Ниже подробно рассмотрим реализацию данного алгоритма на модельном численном примере:

- 2 итерации выявления нетипичных объектов и их назначения на новые классы (разделение классов на типичную и нетипичную части);
- Выявление и удаление артефактов.

4.3. Создание исходной модели для итераций

Шаг 1. Формализация предметной области

На рисунке 4 приведена экранная форма запуска автоматизированного программного интерфейса 2.3.2.2 (API-2.3.2.2) системы «Эйдос»:

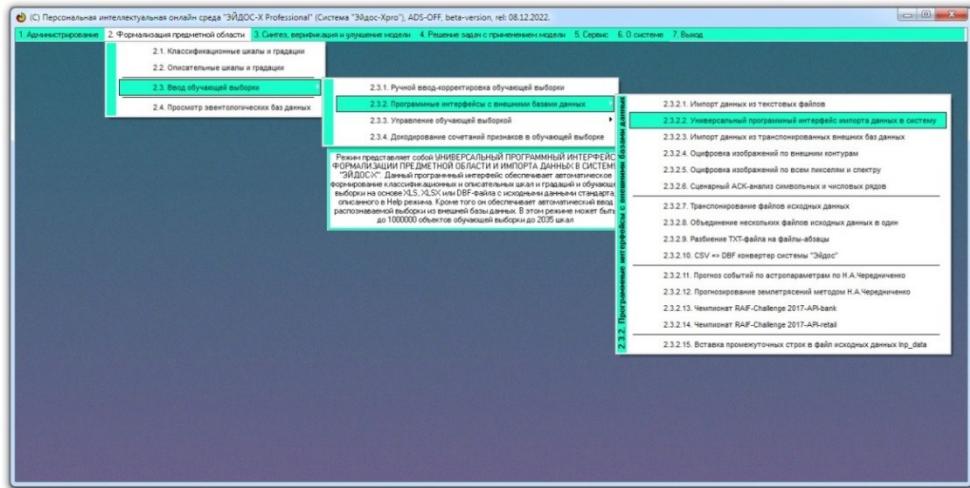


Рисунок 4. Запуск API-2.3.2.2 системы «Эйдос»

На рисунке 5 приведена экранная форма управления API-2.3.2.2 с параметрами, использованными в данной работе. На рисунке 6 приведены экранные формы развернутого help данного режима.

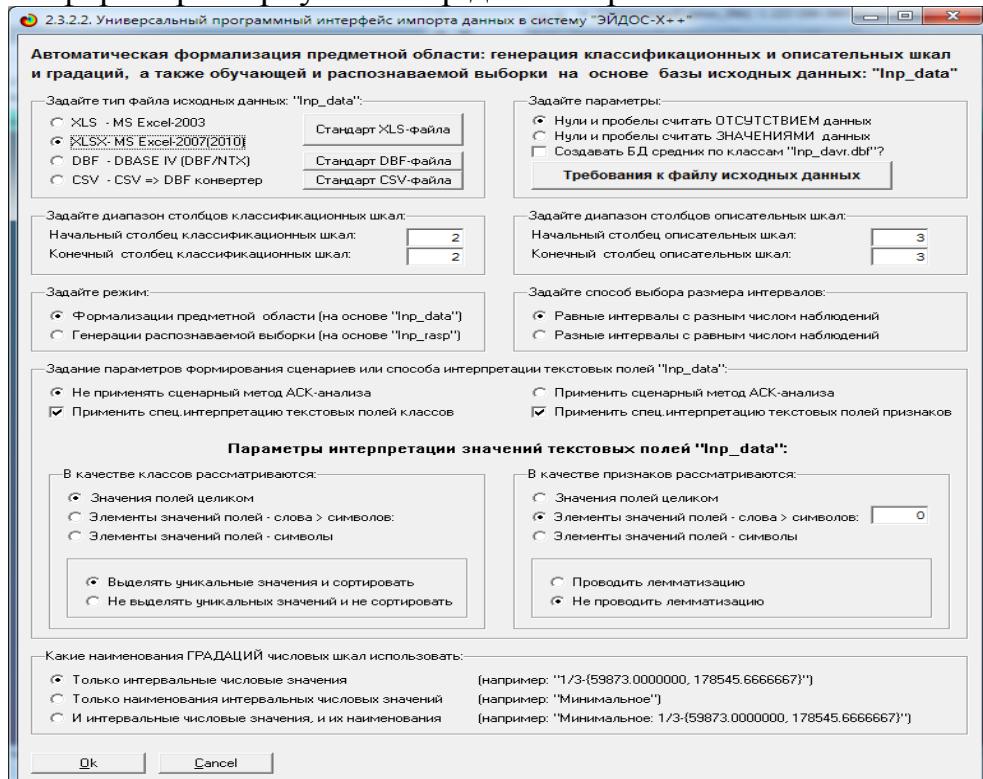


Рисунок 5. Экранная форма управления API-2.3.2.2 системы «Эйдос» с актуальными параметрами для данной работы

Помощь по режиму 2.3.2.2 для случая Excel-файлов исходных данных

Режим 2.3.2.2: Универсальный программный интерфейс импорта данных из внешней базы данных "Inp_data.xls[x]" в систему "ЭйдосX++" и формализации предметной области.

- Данный программный интерфейс обеспечивает формализацию предметной области, т.е. анализ файла исходных данных Inp_data.xls[x], формирование классификационных и описательных шкал и градаций, а затем кодирование файла исходных с их использованием.
- Файл исходных данных должен иметь имя: Inp_data.xls[x], а файл распознаваемой выборки имя: Inp_rasp.xls[x]. Файлы Inp_data.xls[x] и Inp_rasp.xls[x] должны находиться в папке ..\AIDOS_X\AID_DATA\Inp_data/. Эти файлы имеют совершенную одноковую структуру.
- 1-я строка этого файла должна содержать наименования колонок, на любом языке, в т.ч. и русском. Эти наименования должны быть во всех колонках, при этом переносы по словам разрешены, а объединение ячеек, разрывы строки знаки абзаца не допускаются. Эти наименования должны быть короткими, понятными, т.к. они будут в выходных формах, а к ним еще будут добавляться наименования градаций. В числовых шкалах надо ОБЯЗАТЕЛЬНО указывать единицы измерения и число знаков после запятой в колонке должно быть 0.ОДИНАКОВОЕ.
- 1-я колонка содержит наименование объекта обучающей выборки или наименование наблюдения. Оно может быть длинным: до 255 символов.
- Каждая строка этого файла, начиная со 2-го, содержит данные об одном объекте обучающей выборки или одном наблюдении. В MS Excel 2003 в листе может быть до 65536 строк и до 256 колонок. В листе MS Excel 2010 и более поздних возможно до 1048576 строк и 16384 колонок.
- Столбцы, начиная со 2-го по N-й являются классификационными шкалами (выходными параметрами) и содержат данные о классах (будущих состояниях объекта управления), к которым принадлежат объекты обучающей выборки.
- Столбцы присваиваются числового типа, если все значения его ячеек числового типа. Если хотя бы одно значение является текстовым [не число, в т.ч. пробелы], то столбец присваивается текстового типа. Это означает, что нули должны быть указаны нулями, а не пробелами.
- Столбцы со 2-го по N-й являются классификационными шкалами (выходными параметрами) и содержат данные о классах (будущих состояниях объекта управления), к которым принадлежат объекты обучающей выборки.
- В результате работы формируется файл INP_NAME.TXT стандарта MS DOS (кириллица), в котором наименования классификационных и описательных шкал являются СТРОКАМИ. Система формирует классификационные и описательные шкалы и градации. Для этого в каждом числовом столбце система находит минимальное и максимальное числовые значения и формирует заданное количество числовых интервалов, после чего числовые значения заменяются их интервальными значениями. В текстовых столбцах система находит уникальные текстовые значения. Каждое ЧИСЛОВОЕ интервальное числовое или текстовое значение считается градацией классификационной или описательной шкалы, характеризующей объект. В каждой шкале ее градации сортируются по алфавиту. С использованием шкал и градаций кодируются исходные данные в результате чего генерируется обучающая выборка, каждый объект которой соответствует одной строке файла исходных данных NP_DATA и содержит коды классов, соответствующие фактическим совпадениям числовых или уникальных текстовых значений классов с градациями классификационных шкал и коды признаков, соответствующие фактам совпадения числовых или уникальных текстовых значений признаков с градациями описательных шкал.
- Распознаваемая выборка формируется на основе файла INP_RASP аналогично, за исключением того, что классификационные и описательные шкалы и градации не создаются, а используются ранее созданные в модели, и базы распознаваемой выборки могут не включать коды классов, если столбцы классов в файле INP_RASP были пустыми. Структура файла INP_RASP должна быть та же, как INP_DATA, т.е. они должны ПОЛНОСТЬЮ совпадать по наименованиям столбцов, но могут иметь разное количество строк с разными значениями в них.

Принцип организации таблицы исходных данных:

Наименование объекта обучающей выборки	Наименование 1-й классификационной шкалы	Наименование 2-й классификационной шкалы	...	Наименование 1-й описательной шкалы	Наименование 2-й описательной шкалы	...
1-й объект обучающей выборки (1-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
2-й объект обучающей выборки (2-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
...

Определения основных терминов и профилактика типичных ошибок при подготовке Excel-файла исходных данных

Помощь по режиму 2.3.2.2 для случая Excel-файлов исходных данных

Режим 2.3.2.2: Универсальный программный интерфейс импорта данных из внешней базы данных "Inp_data.xls[x]" в систему "ЭйдосX++".

ТЕРМИНЫ АСК-АНАЛИЗА И СИСТЕМЫ "ЭЙДОС":

Шкала представляет собой способ формализации предметной области. Используется числовые и текстовые шкалы, при этом текстовые могут быть номинальными и порядковыми. На номинальных шкалах есть только отношения эквивалентности и неэквивалентности, на порядковых кроме того еще отношения "больше", "меньше", а на числовых - кроме того могут выполняться все арифметические операции. Каждый объект выборки (наблюдение) описан с одной стороны своими признаками, а с другой - принадлежностью к некоторым обобщющим категориям (классам). Такая структура описания называется онтологией или фреймом экспертизы и является базовой для всех моделей представления знаний.

В АСК-анализе и системе "Эйдос" используется три интерпретации шкал и градаций: универсальная, статическая и динамическая:

- в универсальной интерпретации: признаки - это градации описательных шкал;
- в статической интерпретации: описательная шкала - это свойство, а градация [признак] - это степень выраженности этого свойства;
- в динамической интерпретации: описательная шкала - это фактор, а градация [признак] - это значение фактора;
- в универсальной интерпретации: классы - это градации классификационных шкал;
- в статической интерпретации: классификационная шкала - способ классификации обобщающих категорий [классов], к которым в настоящем времени по отношению к признакам относятся состояния объекта прогнозирования или управления;
- в динамической интерпретации: классификационная шкала - способ классификации обобщающих категорий [классов], к которым в будущем времени по отношению к признакам относятся состояния объекта прогнозирования или управления;

ПРОФИЛАКТИКА ОШИБОК В ФАЙЛЕ ИСХОДНЫХ ДАННЫХ:

- 1-я строка файла "Inp_data.xls[x]" должна содержать наименования колонок. Эти наименования должны быть во всех колонках, при этом переносы по словам разрешены, а объединение ячеек, разрывы строки знаки абзаца и неалфавитные символы не допускаются. Эти наименования должны быть короткими, понятными, т.к. они будут в выходных формах, а к ним еще будут добавляться наименования градаций. В числовых шкалах надо обязательно указывать единицы измерения. Число знаков после запятой в числовой колонке должно быть одинаковым.
- 1-я колонка содержит наименование объекта обучающей выборки или наименование наблюдения. Оно может быть длинным: до 255 символов.
- Столбцы, начиная со 2-го, являются классификационными и описательными шкалами и могут быть текстового (номинального / порядкового) или числового типа (со знаками после запятой). Чтобы текстовая шкала была порядковой, нужно чтобы при сортировке по алфавиту градации этой шкалы образовывали последовательность от минимального значения до максимального. Например, текстовая шкала "Размер" с градациями: "очень малое", "малое", "среднее", "большое", "очень большое", будет номинальной шкалой, т.к. при сортировке по алфавиту они расположатся в порядке: "большое", "малое", "среднее", "очень большое", "очень малое", "среднее". Чтобы шкала "Размер" стала порядковой нужно в этом градации присвоить следующие значения: "1/5-очень малое", "2/5-малое", "3/5-среднее", "4/5-большое", "5/5-очень большое".
- Столбцы присваиваются числового типа, если все значения его ячеек числового типа. Если хотя бы одно значение является текстовым [не число, в т.ч. пробелы], то столбец присваивается текстового типа. Это означает, что нули должны быть указаны нулями, а не пробелами.
- Если в системе "Эйдос" в режимах 2.1, 2.2 посмотреть на градации классификационных и описательных шкал, которые должны быть числовыми, то сразу будет видно, в какой форме представлены числа: числовыми диапазонами или просто числами. Если числовыми диапазонами, значит в файле исходных данных в этом отношении все правильно, если же числа, то возможно в Excel-файле нужно заменить десятичные точки на запятые, а также найти и исправить нечисловые данные в числовых по смыслу колонках. Быстро найти их можно перейдя на последнюю строку файла исходных данных и задача расчет суммы колонки. В формуле будет видно с какой строкой идет расчет суммы. Если со 2-й, то значит все верно, иначе будет указана строка, в которой находится нечисловое значение.
- Система "Эйдос" работает с областью данных файла исходных данных, которую можно выделить блоком, поставив курсор в ячейку A1, нажав Ctrl+Home, а затем зажав клавиши Shift+Ctrl нажать End. Если этот блок выходит за пределы области таблицы, фактически занятой данными надо скопировать эту фактическую область данных в буфер обмена, создать новый лист и скопировать в него, а исходный лист удалить.
- Иногда бывает полезно сбросить все форматирование Excel-таблицы исходных данных. Это можно сделать в MS Excel. А можно скопировать таблицу в MS Word, а потом обратно в MS Excel.

Принцип организации таблицы исходных данных:

Наименование объекта обучающей выборки	Наименование 1-й классификационной шкалы	Наименование 2-й классификационной шкалы	...	Наименование 1-й описательной шкалы	Наименование 2-й описательной шкалы	...
1-й объект обучающей выборки (1-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
2-й объект обучающей выборки (2-е наблюдение)	Значение шкалы	Значение шкалы	...	Значение шкалы	Значение шкалы	...
...

Рисунок 6. Help API-2.3.2.2 системы «Эйдос»

На рисунке 7 приведена экранная форма внутреннего калькулятора API-2.3.2.2 системы «Эйдос»:

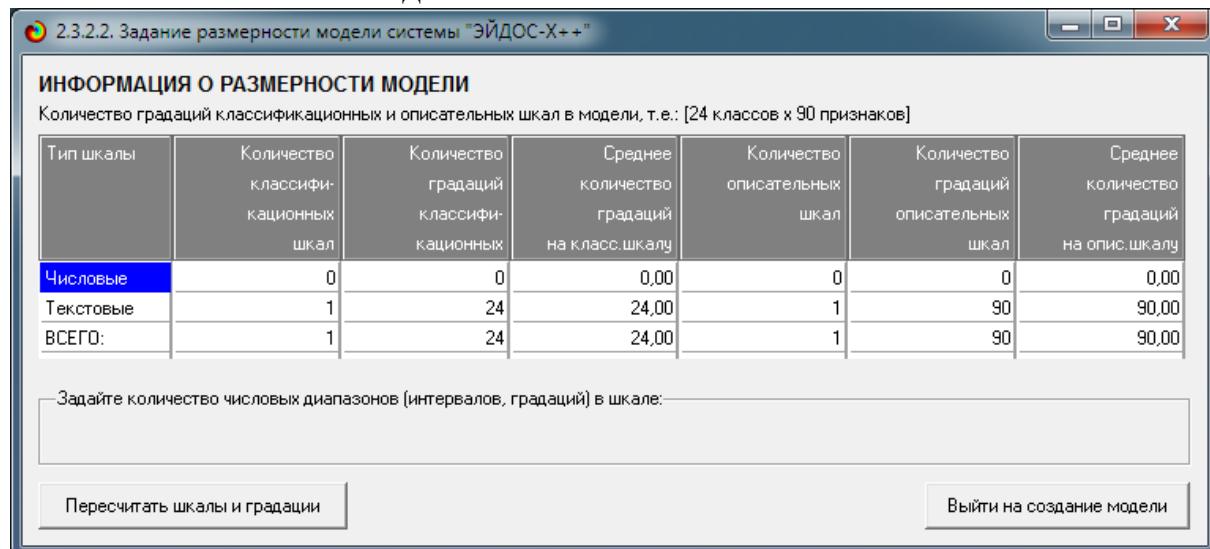


Рисунок 7. Внутренний калькулятор API-2.3.2.2 системы «Эйдос»

В данном случае в модели нет числовых шкал, поэтому калькулятор выполняет чисто информационную функцию.

В результате работы API-2.3.2.2 созданы классификационные и описательные шкалы и градации (рисунок 8, таблицы 2 и 3), а затем с их использованием закодированы исходные данные (таблица 1), в результате чего получена обучающая выборка (таблица 4):

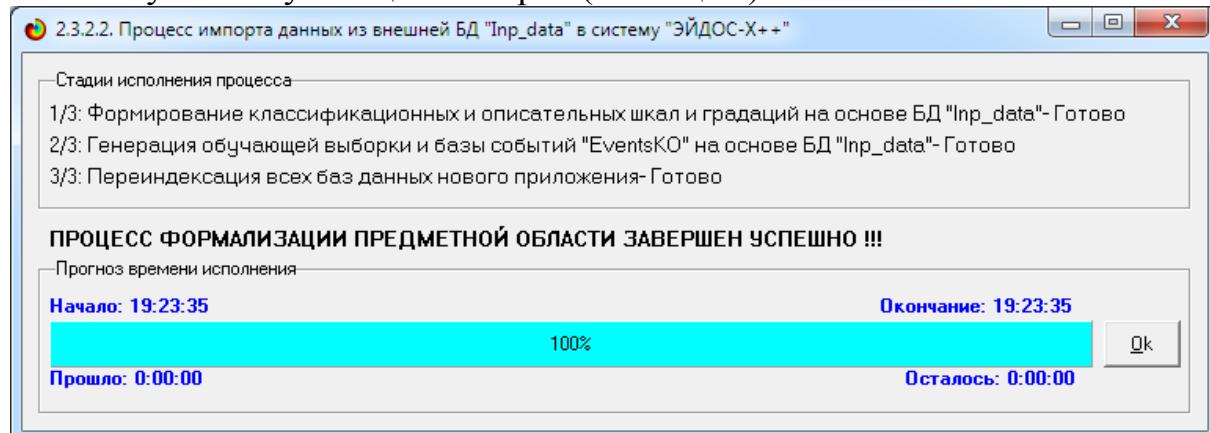


Рисунок 8. Экранная форма отображения процесса и окончания работы API-2.3.2.2 системы «Эйдос»

Таблица 2 – Классификационные шкалы и градации (полностью)

KOD_CLS	NAME_CLS
1	КЛАСС-01
2	КЛАСС-02
3	КЛАСС-03
4	КЛАСС-04
5	КЛАСС-05
6	КЛАСС-06
7	КЛАСС-07
8	КЛАСС-08
9	КЛАСС-09
10	КЛАСС-10

Таблица 3 – Описательные шкалы и градации (полностью)

KOD_ATR	NAME_ATR	KOD_ATR	NAME_ATR	KOD_ATR	NAME_ATR
1	ПРИЗНАКИ-00	25	ПРИЗНАКИ-26	49	ПРИЗНАКИ-59
2	ПРИЗНАКИ-01	26	ПРИЗНАКИ-27	49	ПРИЗНАКИ-59
3	ПРИЗНАКИ-02	27	ПРИЗНАКИ-28	50	ПРИЗНАКИ-60
4	ПРИЗНАКИ-03	28	ПРИЗНАКИ-29	51	ПРИЗНАКИ-61
5	ПРИЗНАКИ-04	29	ПРИЗНАКИ-32	52	ПРИЗНАКИ-62
6	ПРИЗНАКИ-05	30	ПРИЗНАКИ-33	53	ПРИЗНАКИ-63
7	ПРИЗНАКИ-06	31	ПРИЗНАКИ-34	54	ПРИЗНАКИ-64
8	ПРИЗНАКИ-07	32	ПРИЗНАКИ-35	55	ПРИЗНАКИ-65
9	ПРИЗНАКИ-08	33	ПРИЗНАКИ-36	56	ПРИЗНАКИ-67
10	ПРИЗНАКИ-09	34	ПРИЗНАКИ-38	57	ПРИЗНАКИ-69
11	ПРИЗНАКИ-10	35	ПРИЗНАКИ-39	58	ПРИЗНАКИ-71
12	ПРИЗНАКИ-11	36	ПРИЗНАКИ-40	59	ПРИЗНАКИ-73
13	ПРИЗНАКИ-12	37	ПРИЗНАКИ-41	60	ПРИЗНАКИ-76
14	ПРИЗНАКИ-13	38	ПРИЗНАКИ-44	61	ПРИЗНАКИ-80
15	ПРИЗНАКИ-14	39	ПРИЗНАКИ-45	62	ПРИЗНАКИ-81
16	ПРИЗНАКИ-15	40	ПРИЗНАКИ-48	63	ПРИЗНАКИ-83
17	ПРИЗНАКИ-16	41	ПРИЗНАКИ-50	64	ПРИЗНАКИ-85
18	ПРИЗНАКИ-17	42	ПРИЗНАКИ-51	65	ПРИЗНАКИ-86
19	ПРИЗНАКИ-18	43	ПРИЗНАКИ-52	66	ПРИЗНАКИ-88
20	ПРИЗНАКИ-19	44	ПРИЗНАКИ-54	67	ПРИЗНАКИ-89
21	ПРИЗНАКИ-20	45	ПРИЗНАКИ-55	68	ПРИЗНАКИ-91
22	ПРИЗНАКИ-21	46	ПРИЗНАКИ-56	69	ПРИЗНАКИ-92
23	ПРИЗНАКИ-23	47	ПРИЗНАКИ-57	70	ПРИЗНАКИ-97
24	ПРИЗНАКИ-25	48	ПРИЗНАКИ-58	71	ПРИЗНАКИ-98

Таблица 4 – Обучающая выборка (полностью)

NAME_OBJ	N2	N3									
01_01i	1	2	03_10i	3	4	06_01i	6	7	08_10i	8	9
01_02i	1	2	03_11i	3	4	06_02i	6	7	08_11i	8	9
01_03i	1	2	03_12i	3	4	06_03i	6	7	08_12i	8	9
01_04i	1	2	03_13a	3	54	06_04i	6	7	08_13a	8	50
01_05i	1	2	03_14a	3	5	06_05i	6	7	08_14a	8	3
01_06i	1	2	03_15n	3	14	06_06i	6	7	08_15n	8	19
01_07i	1	2	03_16n	3	14	06_07i	6	7	08_16n	8	19
01_08i	1	2	03_17n	3	14	06_08i	6	7	08_17n	8	19
01_09i	1	2	03_18n	3	14	06_09i	6	7	08_18n	8	19
01_10i	1	2	04_01i	4	5	06_10i	6	7	09_01i	9	10
01_11i	1	2	04_02i	4	5	06_11i	6	7	09_02i	9	10
01_12i	1	2	04_03i	4	5	06_12i	6	7	09_03i	9	10
01_13a	1	65	04_04i	4	5	06_13a	6	43	09_04i	9	10
01_14a	1	6	04_05i	4	5	06_14a	6	8	09_05i	9	10
01_15n	1	12	04_06i	4	5	06_15n	6	17	09_06i	9	10
01_16n	1	12	04_07i	4	5	06_16n	6	17	09_07i	9	10
01_17n	1	2	04_08i	4	5	06_17n	6	17	09_08i	9	10
01_18n	1	12	04_09i	4	5	06_18n	6	17	09_09i	9	10
02_01i	2	3	04_10i	4	5	07_01i	7	8	09_10i	9	10
02_02i	2	3	04_11i	4	5	07_02i	7	8	09_11i	9	10
02_03i	2	3	04_12i	4	5	07_03i	7	8	09_12i	9	10
02_04i	2	3	04_13a	4	68	07_04i	7	8	09_13a	9	9
02_05i	2	3	04_14a	4	8	07_05i	7	8	09_14a	9	2
02_06i	2	3	04_15n	4	15	07_06i	7	8	09_15n	9	20
02_07i	2	3	04_16n	4	15	07_07i	7	8	09_16n	9	20
02_08i	2	3	04_17n	4	15	07_08i	7	8	09_17n	9	20
02_09i	2	3	04_18n	4	15	07_09i	7	8	09_18n	9	20
02_10i	2	3	05_01i	5	6	07_10i	7	8	10_01i	10	11
02_11i	2	3	05_02i	5	6	07_11i	7	8	10_02i	10	11
02_12i	2	3	05_03i	5	6	07_12i	7	8	10_03i	10	11
02_13a	2	32	05_04i	5	6	07_13a	7	7	10_04i	10	11
02_14a	2	3	05_05i	5	6	07_14a	7	9	10_05i	10	11
02_15n	2	13	05_06i	5	6	07_15n	7	18	10_06i	10	11
02_16n	2	13	05_07i	5	6	07_16n	7	18	10_07i	10	11
02_17n	2	13	05_08i	5	6	07_17n	7	18	10_08i	10	11
02_18n	2	13	05_09i	5	6	07_18n	7	18	10_09i	10	11
03_01i	3	4	05_10i	5	6	08_01i	8	9	10_10i	10	11
03_02i	3	4	05_11i	5	6	08_02i	8	9	10_11i	10	11
03_03i	3	4	05_12i	5	6	08_03i	8	9	10_12i	10	11
03_04i	3	4	05_13a	5	55	08_04i	8	9	10_13a	10	41
03_05i	3	4	05_14a	5	5	08_05i	8	9	10_14a	10	11
03_06i	3	4	05_15n	5	16	08_06i	8	9	10_15n	10	21
03_07i	3	4	05_16n	5	16	08_07i	8	9	10_16n	10	21
03_08i	3	4	05_17n	5	16	08_08i	8	9	10_17n	10	21
03_09i	3	4	05_18n	5	16	08_09i	8	9	10_18n	10	21

Шаг 2. Синтез и верификация статистических и системно-когнитивных моделей

Таким образом, работа API-2.3.2.2 полностью подготовила следующий этап АСК-анализа: синтез и верификацию статистических и системно-когнитивных моделей. Этот этап выполняется в режиме 3.5 системы «Эйдос» (рисунки 9, 10, 11):

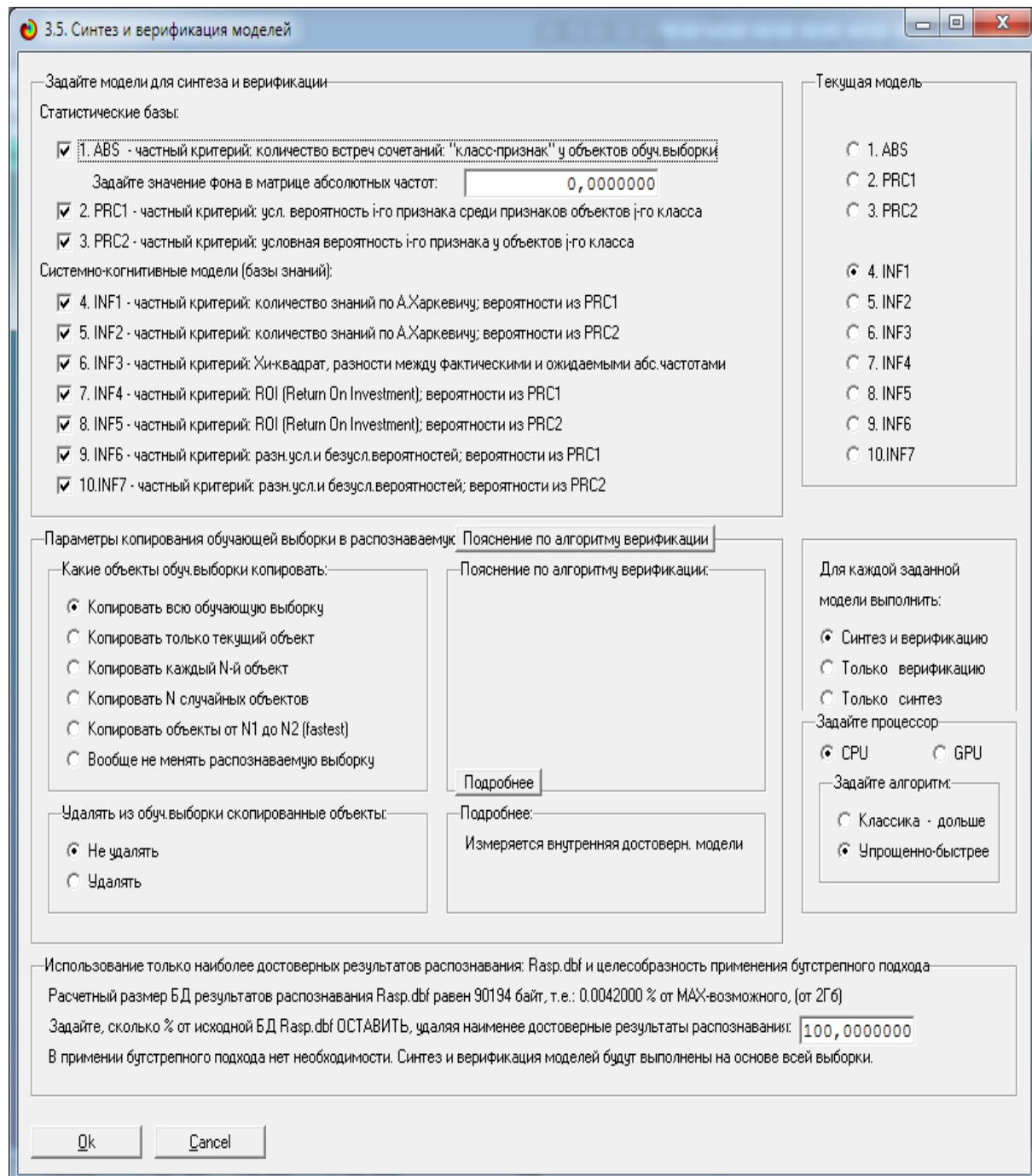


Рисунок 9. Экранная форма управления режимом синтеза и верификации статистических и системно-когнитивных моделей системы «Эйдос»

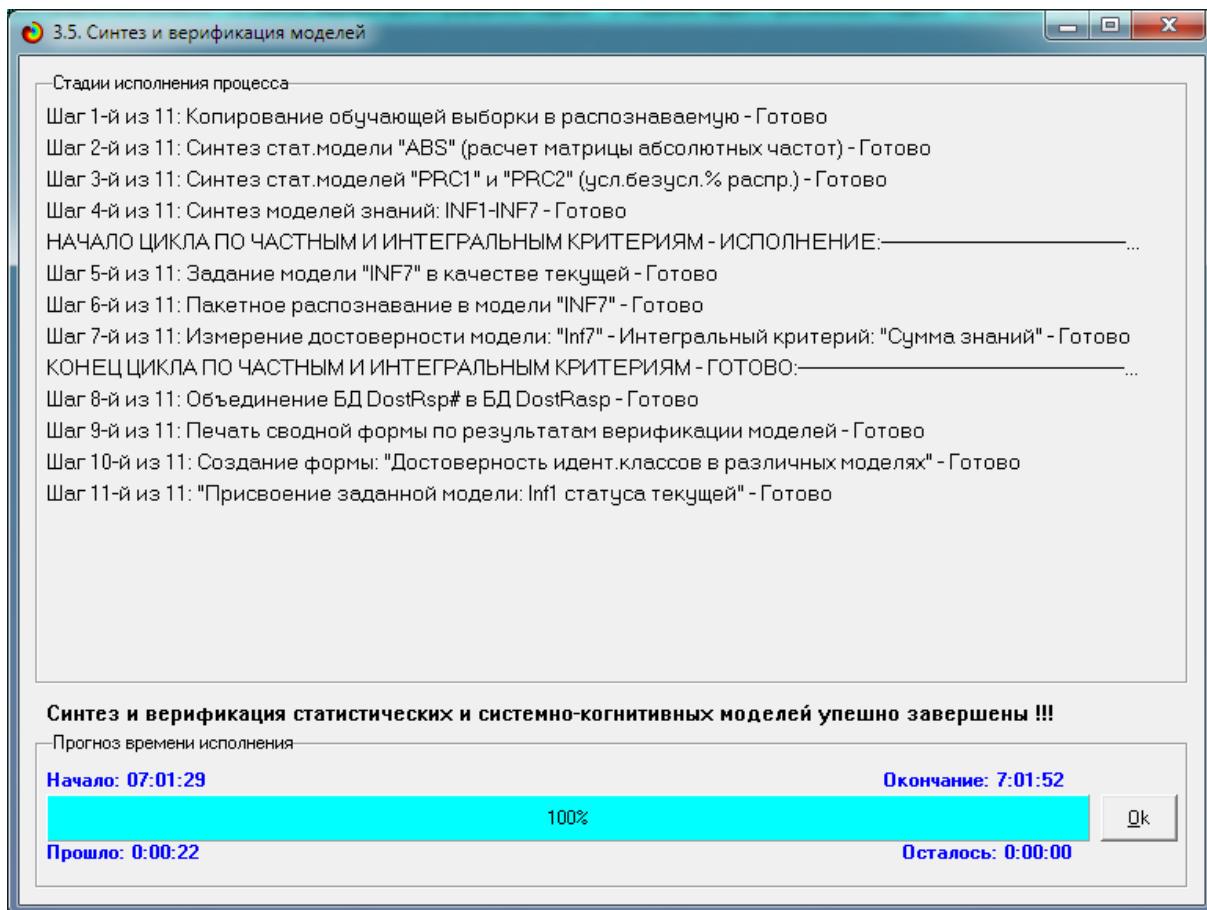


Рисунок 10. Экранная форма отображения процесса и окончания работы режима синтеза и верификации статистических и системно-когнитивных моделей системы «Эйдос»

В результате работы режима синтеза и верификации моделей были созданы 3 статистические и 7 системно-когнитивные модели (рисунки 11)

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 01	2. КЛАСС 02	3. КЛАСС 03	4. КЛАСС 04	5. КЛАСС 05	6. КЛАСС 06	7. КЛАСС 07	8. КЛАСС 08	9. КЛАСС 09	10. КЛАСС 10	Сумма	Среднее	Средн квадр откл
1.0	ПРИЗНАКИ-00					1.0	1.0		1.0		1.0	4.0	0.40	0.52
2.0	ПРИЗНАКИ-01	120.0					1.0		1.0	1.0	2.0	125.0	12.50	37.76
3.0	ПРИЗНАКИ-02	1.0	121.0	1.0			2.0		1.0	1.0		127.0	12.70	38.06
4.0	ПРИЗНАКИ-03		1.0	122.0	1.0		1.0	3.0	1.0		2.0	131.0	13.10	38.28
5.0	ПРИЗНАКИ-04	1.0	2.0	1.0	120.0	1.0						125.0	12.50	37.78
6.0	ПРИЗНАКИ-05	1.0		5.0	2.0	122.0	1.0	4.0	1.0	2.0	2.0	140.0	14.00	37.96
7.0	ПРИЗНАКИ-06		2.0		3.0	4.0	120.0	1.0	2.0	1.0	1.0	134.0	13.40	37.48
8.0	ПРИЗНАКИ-07		2.0		3.0		2.0	122.0	2.0	3.0	2.0	136.0	13.60	38.11
9.0	ПРИЗНАКИ-08	2.0		1.0	1.0	1.0	1.0	1.0		1.0		130.0	13.00	37.95
10.0	ПРИЗНАКИ-09	3.0	1.0	1.0				1.0			122.0	128.0	12.80	38.38
11.0	ПРИЗНАКИ-10	2.0	2.0		1.0	2.0	1.0				122.0	130.0	13.00	38.31
12.0	ПРИЗНАКИ-11	40.0								1.0	41.0	4.10	12.62	
13.0	ПРИЗНАКИ-12		40.0	1.0							41.0	4.10	12.62	
14.0	ПРИЗНАКИ-13			40.0							40.0	4.00	12.65	
15.0	ПРИЗНАКИ-14				40.0						40.0	4.00	12.65	
16.0	ПРИЗНАКИ-15	1.0				41.0					42.0	4.20	12.95	
17.0	ПРИЗНАКИ-16						40.0	1.0			41.0	4.10	12.62	
18.0	ПРИЗНАКИ-17							40.0			40.0	4.00	12.65	
19.0	ПРИЗНАКИ-18								40.0		40.0	4.00	12.65	
20.0	ПРИЗНАКИ-19	1.0						1.0	1.0	40.0	43.0	4.30	12.55	
21.0	ПРИЗНАКИ-20									1.0	40.0	41.0	4.10	12.62
22.0	ПРИЗНАКИ-21	1.0									1.0	0.10	0.32	
23.0	ПРИЗНАКИ-22			1.0							1.0	0.10	0.32	
24.0	ПРИЗНАКИ-25		1.0								1.0	0.10	0.32	
25.0	ПРИЗНАКИ-26					1.0			1.0		1.0	0.30	0.48	
26.0	ПРИЗНАКИ-27		1.0								1.0	0.10	0.32	

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 01_	2. КЛАСС 02_	3. КЛАСС 03_	4. КЛАСС 04_	5. КЛАСС 05_	6. КЛАСС 06_	7. КЛАСС 07_	8. КЛАСС 08_	9. КЛАСС 09_	10. КЛАСС 10_	Безузл. вероятн.	Среднее	Средн квадр откл.	
1.0	ПРИЗНАКИ-00					0.556	0.556	0.556	0.556	0.556	0.556	0.222	0.222	0.287	
2.0	ПРИЗНАКИ-01	66.667					0.556	0.556	0.556	0.556	0.556	6.944	6.944	20.988	
3.0	ПРИЗНАКИ-02	0.556	67.222	0.556			1.111		0.556	0.556		7.056	7.056	21.144	
4.0	ПРИЗНАКИ-03		0.556	67.778	0.556		0.556	1.667	0.556		1.111	7.278	7.278	21.264	
5.0	ПРИЗНАКИ-04	0.556	1.111	0.556	66.667	0.556					1.111	6.944	6.944	20.988	
6.0	ПРИЗНАКИ-05	0.556		2.778	1.111	67.778	0.556	2.222	0.556	1.111	1.111	7.778	7.778	21.098	
7.0	ПРИЗНАКИ-06			1.111		1.667	2.222	66.667	0.556	1.111	0.556	7.444	7.444	20.820	
8.0	ПРИЗНАКИ-07			1.111		1.667		1.111	67.778	1.111	1.667	1.111	7.556	7.556	21.170
9.0	ПРИЗНАКИ-08	1.111			0.556	0.556	0.556	0.556	1.111	67.222	0.556	7.222	7.222	21.085	
10.0	ПРИЗНАКИ-09	1.667	0.556	0.556				0.556				7.111	7.111	21.323	
11.0	ПРИЗНАКИ-10	1.111	1.111		0.556	1.111	0.556					67.778	7.222	7.222	21.283
12.0	ПРИЗНАКИ-11	22.222									0.556		2.278	2.278	7.010
13.0	ПРИЗНАКИ-12		22.222	0.556									2.278	2.278	7.010
14.0	ПРИЗНАКИ-13			22.222									2.222	2.222	7.027
15.0	ПРИЗНАКИ-14				22.222								2.222	2.222	7.027
16.0	ПРИЗНАКИ-15	0.556				22.778							2.333	2.333	7.186
17.0	ПРИЗНАКИ-16						22.222	0.556					2.278	2.278	7.010
18.0	ПРИЗНАКИ-17							22.222					2.222	2.222	7.027
19.0	ПРИЗНАКИ-18								22.222				2.222	2.222	7.027
20.0	ПРИЗНАКИ-19	0.556					0.556	0.556	22.222			2.389	2.389	6.974	
21.0	ПРИЗНАКИ-20									0.556	22.222		2.278	2.278	7.010
22.0	ПРИЗНАКИ-21	0.556										0.056	0.056	0.176	
23.0	ПРИЗНАКИ-23				0.556							0.056	0.056	0.176	
24.0	ПРИЗНАКИ-25			0.556								0.056	0.056	0.176	
25.0	ПРИЗНАКИ-26					0.556			0.556		0.556	0.167	0.167	0.268	
26.0	ПРИЗНАКИ-27	0.556										0.056	0.056	0.176	

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 01_	2. КЛАСС 02_	3. КЛАСС 03_	4. КЛАСС 04_	5. КЛАСС 05_	6. КЛАСС 06_	7. КЛАСС 07_	8. КЛАСС 08_	9. КЛАСС 09_	10. КЛАСС 10_	Сумма	Среднее	Среднеквадр. откл.
1.0	ПРИЗНАКИ-00					0.406	0.406		0.406		0.406	1.624	1.62	0.162
2.0	ПРИЗНАКИ-01		1.002				-1.119		-1.119	-1.119	-0.812	-3.168	-0.317	
3.0	ПРИЗНАКИ-02	-1.126	0.999	-1.126			-0.819		-1.126	-1.126		-4.326	-0.433	
4.0	ПРИЗНАКИ-03	-1.140	0.899	-1.140	-1.140		-1.140	-0.653	-1.140		-0.833	-5.058	-0.506	
5.0	ПРИЗНАКИ-04	-1.119	-0.812	-1.119	1.002	-1.119						-3.168	-0.317	
6.0	ПРИЗНАКИ-05	-1.170		-0.456	-0.862	0.959	-1.170	-0.555	-1.170	-0.862	-0.862	-6.148	-0.615	
7.0	ПРИЗНАКИ-06		-0.843	-0.663	-0.536	0.972	-1.150	-0.843	-1.150	-1.150		-5.364	-0.536	
8.0	ПРИЗНАКИ-07		-0.850	-0.670		-0.850	0.972	-0.850	-0.670	-0.850		-3.766	-0.377	
9.0	ПРИЗНАКИ-08	-0.830		-1.137	-1.137	-1.137	-1.137	-0.830	0.989	-1.137		-6.354	-0.635	
10.0	ПРИЗНАКИ-09	-0.643	-1.130	-1.130			-1.130				0.999	-3.033	-0.303	
11.0	ПРИЗНАКИ-10	-0.830	-0.830		-1.137	-0.830	-1.137				0.992	-3.770	-0.377	
12.0	ПРИЗНАКИ-11	1.010									-0.625	0.384	0.038	
13.0	ПРИЗНАКИ-12		1.010	-0.625								0.384	0.038	
14.0	ПРИЗНАКИ-13			1.020								1.020	0.102	
15.0	ПРИЗНАКИ-14				1.020							1.020	0.102	
16.0	ПРИЗНАКИ-15	-0.636				1.010						0.374	0.037	
17.0	ПРИЗНАКИ-16						1.010	-0.625				0.384	0.038	
18.0	ПРИЗНАКИ-17							1.020				1.020	0.102	
19.0	ПРИЗНАКИ-18								1.020			1.020	0.102	
20.0	ПРИЗНАКИ-19	-0.646					-0.646	-0.646	0.988		-0.951	-0.951	-0.095	
21.0	ПРИЗНАКИ-20									-0.625	1.010	0.384	0.038	
22.0	ПРИЗНАКИ-21	1.020										1.020	0.102	
23.0	ПРИЗНАКИ-23			1.020								1.020	0.102	
24.0	ПРИЗНАКИ-25			1.020								1.020	0.102	
25.0	ПРИЗНАКИ-26					0.534			0.534		0.534	1.601	0.160	
26.0	ПРИЗНАКИ-27		1.020									1.020	0.102	

Рисунок 11. Статистические и системно-когнитивные модели системы «Эйдос» до начала итераций разделения классов на типичную и нетипичную части

Способ расчета этих моделей кратко, но полно описан в работе [8], а также в help режима 5.5 системы «Эйдос».

Шаг 3. Оценка достоверности статистических и системно-когнитивных моделей

Оценка достоверности статистических и системно-когнитивных моделей осуществляется в режиме 3.4 системы «Эйдос».

Достоверность исходных моделей оказалась довольно высокой: F-мера Ван Ризбергена для модели INF3 равна $F=0,854$ при максимуме 1,000 (рисунок 12).

Наконечное название модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	F-мера Van Rizbergena	Сумма модулей уровней схожести истинно-отрицательных решений (SITR)	Сумма модулей уровней схожести истинно-положительных решений (SITP)	Сумма модулей уровней схожести ложноположительных решений (SFN)
1. ABS - частный критерий: количество встреч соединений: "класс... 1. ABS - частный критерий: количество встреч соединений: "класс... 2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция abs. частот с обр...	180	169	1456	164	11	0.508	0.939	0.659	135.161	26.887	20.718
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Сумма abs. частот по признак...	180	180	744	876	0	0.170	1.000	0.291	133.841	20.009	20.718
2. PRC1 - частный критерий: усл. вероятность i-го признака сред...	Корреляция усл. отн. частот с о...	180	169	1456	164	11	0.508	0.939	0.659	135.161	26.887	20.718
3. PRC2 - частный критерий: условная вероятность i-го признака...	Сумма усл. отн. частот с о...	180	169	1456	164	11	0.508	0.939	0.659	135.161	26.887	20.718
3. PRC2 - частный критерий: условная вероятность i-го признака...	Корреляция усл. отн. частот с о...	180	180	744	876	0	0.170	1.000	0.291	133.841	20.009	20.718
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	180	170	1530	90	10	0.654	0.944	0.773	72.048	344.679	3.551
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	180	170	1568	52	10	0.766	0.944	0.846	140.712	608.039	3.751
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	180	170	1530	90	10	0.654	0.944	0.773	72.048	344.679	3.551
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	180	170	1568	52	10	0.766	0.944	0.846	140.712	608.039	3.751
6. INF3 - частный критерий: Числоводят: разности между факты...	Семантический резонанс зна...	180	172	1575	45	8	0.793	0.956	0.866	134.493	144.505	10.853
6. INF3 - частный критерий: Числоводят: разности между фактами...	Сумма знаний	180	172	1569	51	8	0.771	0.956	0.854	132.697	136.826	4.589
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	180	173	1562	39	7	0.816	0.964	0.863	67.760	127.987	4.420
7. INF4 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	180	177	1504	116	3	0.604	0.983	0.748	155.189	70.375	8.949
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Семантический резонанс зна...	180	173	1581	39	7	0.816	0.961	0.863	87.760	127.987	4.420
8. INF5 - частный критерий: ROI (Return On Investment); вероятно...	Сумма знаний	180	177	1504	116	3	0.604	0.983	0.748	155.189	70.375	8.949
9. INF6 - частный критерий: разн.усл.и безразл.вероятностей; вер...	Семантический резонанс зна...	180	167	1558	62	13	0.729	0.928	0.817	133.093	98.280	13.961
9. INF6 - частный критерий: разн.усл.и безразл.вероятностей; вер...	Сумма знаний	180	172	1530	90	8	0.656	0.956	0.778	132.697	74.590	6.765
10.INF7 - частный критерий: разн.усл.и безразл.вероятностей; вер...	Семантический резонанс зна...	180	167	1558	62	13	0.729	0.928	0.817	133.093	98.280	13.961
10.INF7 - частный критерий: разн.усл.и безразл.вероятностей; вер...	Сумма знаний	180	172	1530	90	8	0.656	0.956	0.778	132.697	74.590	6.765

Рисунок 12. Достоверность системно-когнитивной модели INF3 системы «Эйдос» до начала итераций разделения классов на типичную и нетипичную части

Шаг 4. Идентификация в наиболее достоверной модели

В качестве текущей зададим наиболее достоверную по F-критерию Ван Ризбергена модель INF3 (рисунок 13).

Для задания модели в качестве текущей используется режим 5.6 системы «Эйдос» (рисунки 14), а для решения задачи идентификации объектов обучающей выборки в текущей модели режим 4.1.2 (рисунки 15).

Распознаваемая выборка уже содержит обучающую выборку после выполнения режима 3.5 (синтез и верификация модели), поэтому специально создавать распознаваемую выборку нет необходимости и можно сразу запускать распознавание (идентификацию).

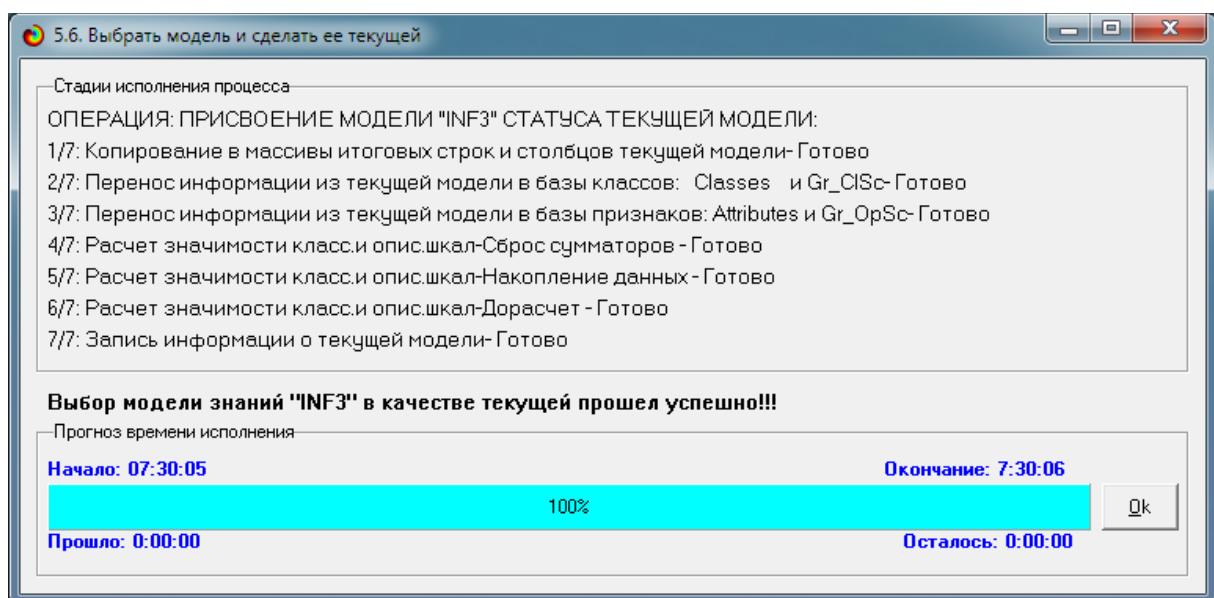
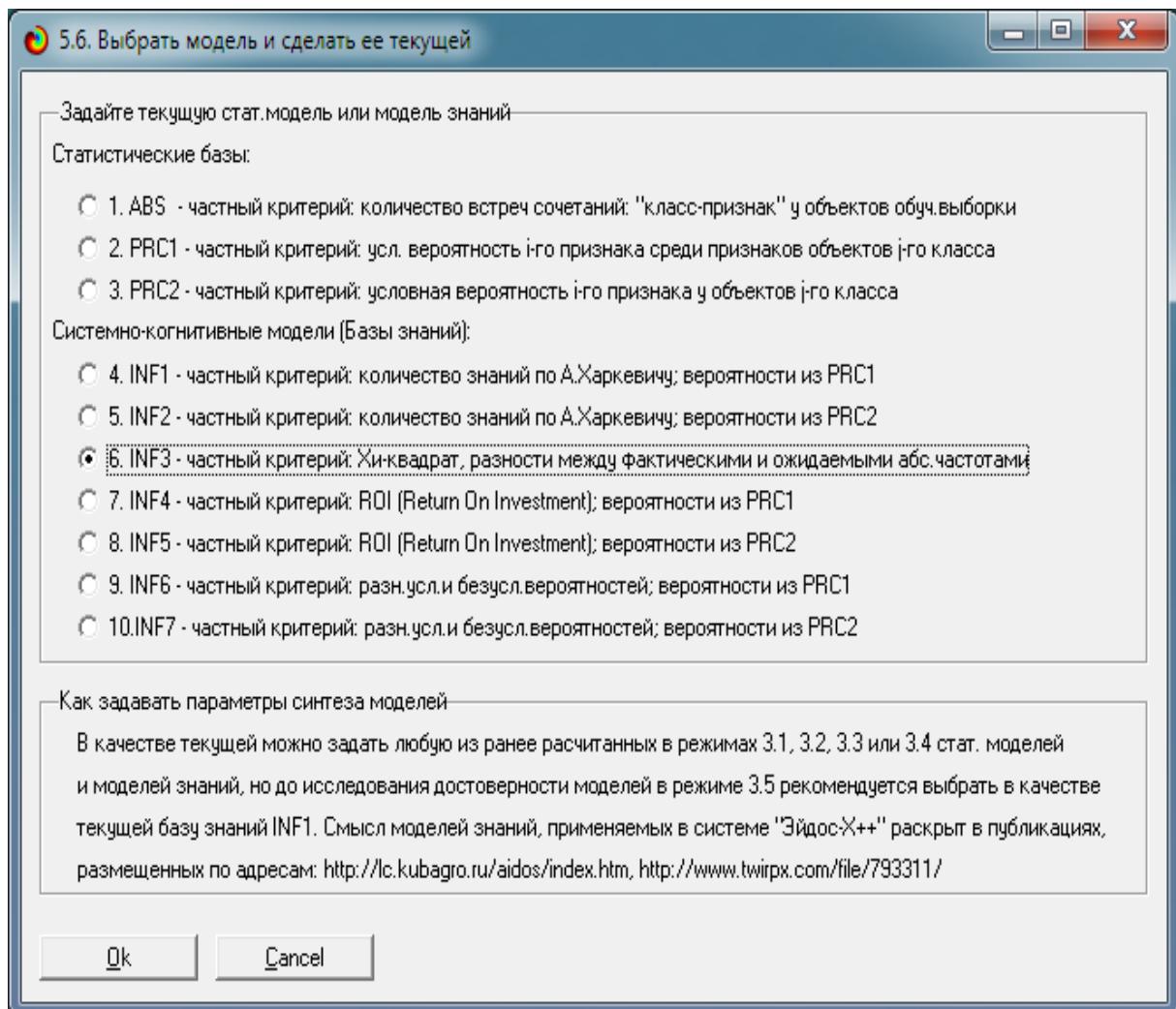


Рисунок 13. Экранные формы режима 5.6 придания модели статуса текущей

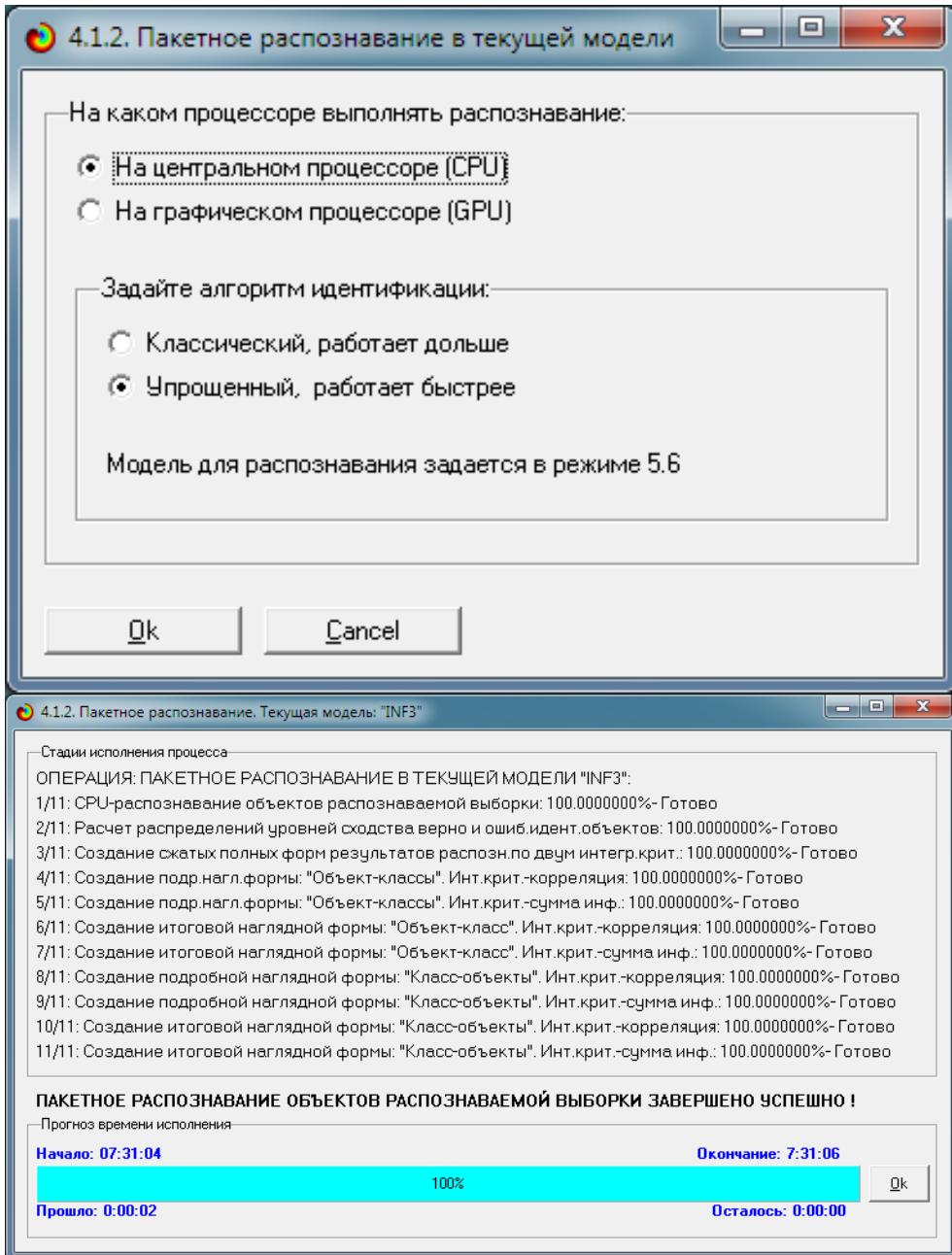


Рисунок 14. Экранные формы режима 4.1.2 пакетного распознавания

Подробные результаты идентификации объектов обучающей выборки с каждым из классов приведены в экранных формах на рисунках 15.

Из этих экранных форм видно, что уровень сходства различных объектов с классами, к которым они относятся, является различным, и есть довольно много объектов с уровнем сходства с классами, к которым они относятся, ниже 50%.

Эти объекты могут являться и **нетипичными для данных классов** объектами, и объектами **нетипичными для всех классов**, т.е. артефактами.

4.1.32. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель "INP".

Код	Название класса
1	KLASS_01
2	KLASS_02
3	KLASS_03
4	KLASS_04
5	KLASS_05
6	KLASS_06
7	KLASS_07
8	KLASS_08
9	KLASS_09
10	KLASS_10

Интегральный критерий класса: "Семантический уровень знаний"

Код	Название объекта	Скоростр.	Ф.	Скорость
1	01_01	99.67	v	
2	01_02	99.67	v	
3	01_03	99.67	v	
4	01_04	99.67	v	
5	01_05	99.67	v	
6	01_06	99.67	v	
7	01_07	99.67	v	
8	01_08	99.67	v	
9	01_09	99.67	v	
10	01_10	99.67	v	
11	01_11	99.67	v	
12	01_12	99.67	v	
15	01_15	33.26	v	
16	01_16	33.26	v	
17	01_17	33.26	v	
18	01_18	33.26	v	
19	01_19	33.26	v	
176	10_14	33.26	v	
177	11_14	33.26	v	
178	11_15	4.77%	v	

Интегральный критерий класса: "Сумма знаний"

Код	Название объекта	Скоростр.	Ф.	Скорость
1	01_01	99.44	v	
2	01_02	99.44	v	
3	01_03	99.44	v	
4	01_04	99.44	v	
5	01_05	99.44	v	
6	01_06	99.44	v	
7	01_07	99.44	v	
8	01_08	99.44	v	
9	01_09	99.44	v	
10	01_10	99.44	v	
11	01_11	99.44	v	
12	01_12	99.44	v	
15	01_15	32.87	v	
16	01_16	32.87	v	
17	01_17	32.87	v	
18	01_18	32.87	v	
19	01_19	32.87	v	
176	10_14	11.34	v	
177	11_14	11.34	v	
178	11_15	1.77%	v	

Помощь Панк объекта В начало БД В конец БД Правдивая Следующая Записей Все записи Печать XLS Печать TXT Печать ALL

4.1.32. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель "INP".

Код	Название класса
1	KLASS_01
2	KLASS_02
3	KLASS_03
4	KLASS_04
5	KLASS_05
6	KLASS_06
7	KLASS_07
8	KLASS_08
9	KLASS_09
10	KLASS_10

Интегральный критерий класса: "Семантический уровень знаний"

Код	Название объекта	Скоростр.	Ф.	Скорость
37	03_01	99.92	v	
38	03_02	99.92	v	
39	03_03	99.92	v	
40	03_04	99.92	v	
41	03_05	99.92	v	
42	03_06	99.92	v	
43	03_07	99.92	v	
44	03_08	99.92	v	
45	03_09	99.92	v	
46	03_10	99.92	v	
47	03_11	99.92	v	
48	03_12	99.92	v	
51	03_15	33.00	v	
52	03_16	33.00	v	
53	03_17	33.00	v	
54	03_18	33.00	v	
59	03_14	24.77	v	
170	07_14	93.93	v	

Интегральный критерий класса: "Сумма знаний"

Код	Название объекта	Скоростр.	Ф.	Скорость
37	03_01	99.72	v	
38	03_02	99.72	v	
39	03_03	99.72	v	
40	03_04	99.72	v	
41	03_05	99.72	v	
42	03_06	99.72	v	
43	03_07	99.72	v	
44	03_08	99.72	v	
45	03_09	99.72	v	
46	03_10	99.72	v	
47	03_11	99.72	v	
48	03_12	99.72	v	
51	03_15	33.00	v	
52	03_16	33.00	v	
53	03_17	33.00	v	
54	03_18	33.00	v	
59	03_14	24.74	v	
170	07_14	11.91	v	

Помощь Панк объекта В начало БД В конец БД Правдивая Следующая Записей Все записи Печать XLS Печать TXT Печать ALL

4.1.32. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель "INP".

Код	Название класса
1	KLASS_01
2	KLASS_02
3	KLASS_03
4	KLASS_04
5	KLASS_05
6	KLASS_06
7	KLASS_07
8	KLASS_08
9	KLASS_09
10	KLASS_10

Интегральный критерий класса: "Семантический уровень знаний"

Код	Название объекта	Скоростр.	Ф.	Скорость
37	03_01	99.92	v	
38	03_02	99.92	v	
39	03_03	99.92	v	
40	03_04	99.92	v	
41	03_05	99.92	v	
42	03_06	99.92	v	
43	03_07	99.92	v	
44	03_08	99.92	v	
45	03_09	99.92	v	
46	03_10	99.92	v	
47	03_11	99.92	v	
48	03_12	99.92	v	
51	03_15	32.96	v	
52	03_16	32.96	v	
53	03_17	32.96	v	
54	03_18	32.96	v	
59	03_14	12.34	v	
170	07_14	11.91	v	

Интегральный критерий класса: "Сумма знаний"

Код	Название объекта	Скоростр.	Ф.	Скорость
37	03_01	99.72	v	
38	03_02	99.72	v	
39	03_03	99.72	v	
40	03_04	99.72	v	
41	03_05	99.72	v	
42	03_06	99.72	v	
43	03_07	99.72	v	
44	03_08	99.72	v	
45	03_09	99.72	v	
46	03_10	99.72	v	
47	03_11	99.72	v	
48	03_12	99.72	v	
51	03_15	32.96	v	
52	03_16	32.96	v	
53	03_17	32.96	v	
54	03_18	32.96	v	
59	03_14	12.34	v	
170	07_14	11.91	v	

Помощь Панк объекта В начало БД В конец БД Правдивая Следующая Записей Все записи Печать XLS Печать TXT Печать ALL

4.1.32. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель "INP".

Код	Название класса
1	KLASS_01
2	KLASS_02
3	KLASS_03
4	KLASS_04
5	KLASS_05
6	KLASS_06
7	KLASS_07
8	KLASS_08
9	KLASS_09
10	KLASS_10

Интегральный критерий класса: "Семантический уровень знаний"

Код	Название объекта	Скоростр.	Ф.	Скорость
73	05_01	99.90	v	
74	05_02	99.90	v	
75	05_03	99.90	v	
76	05_04	99.90	v	
77	05_05	99.90	v	
78	05_06	99.90	v	
79	05_07	99.90	v	
80	05_08	99.90	v	
81	05_09	99.90	v	
82	05_10	99.90	v	
83	05_11	99.90	v	
84	05_12	99.90	v	
85	05_13	99.90	v	
86	05_14	99.90	v	
122	07_14	62.01	v	
95	05_13	41.04	v	
87	05_15	33.98	v	
88	05_16	33.98	v	
89	05_17	33.98	v	
90	05_18	33.98	v	
91	05_19	32.67	v	
170	07_14	9.47	v	

Интегральный критерий класса: "Сумма знаний"

Код	Название объекта	Скоростр.	Ф.	Скорость
73	05_01	99.90	v	
74	05_02	99.90	v	
75	05_03	99.90	v	
76	05_04	99.90	v	
77	05_05	99.90	v	
78	05_06	99.90	v	
79	05_07	99.90	v	
80	05_08	99.90	v	
81	05_09	99.90	v	
82	05_10	99.90	v	
83	05_11	99.90	v	
84	05_12	99.90	v	
85	05_13	99.90	v	
86	05_14	99.90	v	
122	07_14	62.01	v	
95	05_13	41.04	v	
87	05_15	33.98	v	
88	05_16	33.98	v	
89	05_17	33.98	v	
90	05_18	33.98	v	
91	05_19	32.67	v	
170	07_14	9.47	v	

Помощь Панк объекта В начало БД В конец БД Правдивая Следующая Записей Все записи Печать XLS Печать TXT Печать ALL

4.1.32. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель "INP".

Код	Название класса
9	KLASS_01
10	KLASS_02
11	KLASS_03
12	KLASS_04
13	KLASS_05
14	KLASS_06
15	KLASS_07
16	KLASS_08
17	KLASS_09
18	KLASS_10

Интегральный критерий класса: "Семантический уровень знаний"

Код	Название объекта	Скоростр.	Ф.	Скорость
9	05_01	97.61	v	
10	05_02	97.61	v	
11	05_03	97.61	v	
12	05_04	97.61	v	
13	05_05	97.61	v	
14	05_06	97.61	v	
15	05_07	97.61	v	
16	05_08	97.61	v	
17	05_09	97.61	v	
18	05_10	97.61	v	
19	05_11	97.61	v	
20	05_12	97.61	v	
21	05_13	97.61	v	
22	05_14	97.61	v	
23	05_15	97.61	v	
24	05_16	97.61	v	
25	05_17	97.61	v	
26	05_18	97.61	v	
27	05_19	97.61	v	
28	05_10	97.61	v	
29	05_11	97.61	v	
30	05_12	97.61	v	
31	05_13	97.61	v	
32	05_14	97.61	v	
33	05_15	97.61	v	
34	05_16	97.61	v	
35	05_17	97.61	v	
36	05_18	97.61	v	
37	05_19	97.61	v	
38	05_10	97.61	v	
39	05_11	97.61	v	
40	05_12	97.61	v	
41	05_13	97.61	v	
42	05_14	97.61	v	
43	05_15	97.61	v	
44	05_16	97.61	v	
45	05_17	97.61	v	
46				

The screenshot displays four windows of the INF3 software interface, each showing classification results for a specific class. The windows are arranged in a 2x2 grid.

- Top Left Window:** Shows results for "КЛАСС-01_". The table has columns: Код (Code), Название объекта (Object Name), Сходство (Similarity), and Свойство (Property). The similarity values range from 99.90% to 99.95%.
- Top Right Window:** Shows results for "КЛАСС-01_". The table has columns: Код (Code), Название объекта (Object Name), Сходство (Similarity), and Свойство (Property). The similarity values range from 99.55% to 99.95%.
- Bottom Left Window:** Shows results for "Суж-но знан". The table has columns: Код (Code), Название объекта (Object Name), Сходство (Similarity), and Свойство (Property). The similarity values range from 99.26% to 100.00%.
- Bottom Right Window:** Shows results for "Суж-но знан". The table has columns: Код (Code), Название объекта (Object Name), Сходство (Similarity), and Свойство (Property). The similarity values range from 99.55% to 100.00%.

Each window includes a toolbar at the bottom with buttons for Поиска (Search), Показать объекта (Show object), В начало БД (Start of DB), В конец БД (End of DB), Правильная (Correct), Следующая (Next), Запись (Record), Все записи (All records), Печать XLS (Print XLS), Печать TXT (Print TXT), and Печать ALL (Print All).

Рисунок 15. Подробные результаты идентификации объектов обучающей выборки с каждым из классов, к которым они фактически относятся (в наиболее достоверной модели INF3 до начала процесса итераций)¹

Какие выводы можно сделать из экранных форм, приведенных на рисунке 15?

Прежде всего, мы видим, что все нетипичные объекты имеют сходство с классами, к которым они относятся по данным обучающей выборки, менее 35%. Этот уровень сходства обусловлен тем, что для формирования каждого обобщенного образа класса использовалось 12 объектов с **истинными** значениями признаков, 4 объекта с **закономерными ложными** значениями признаков (нетипичные объекты) и 2 объекта со **случайными** значениями признаков (артефакты).

¹ Приведенные экранные формы вполне читабельны при увеличении масштаба изображения

Из рисунков 15 видно, что уровни сходства артефактов с обобщенными образами классов могут принимать значения в широком диапазоне, т.е. могут быть и довольно похожими на класс, но обычно они имеют весьма низкий уровень сходства с классами: менее 15%.

Количество информации о принадлежности объектов к классам, которое содержится в их признаках, наглядно видно из когнитивных функций (см. рисунки 16 и 17) [10]:

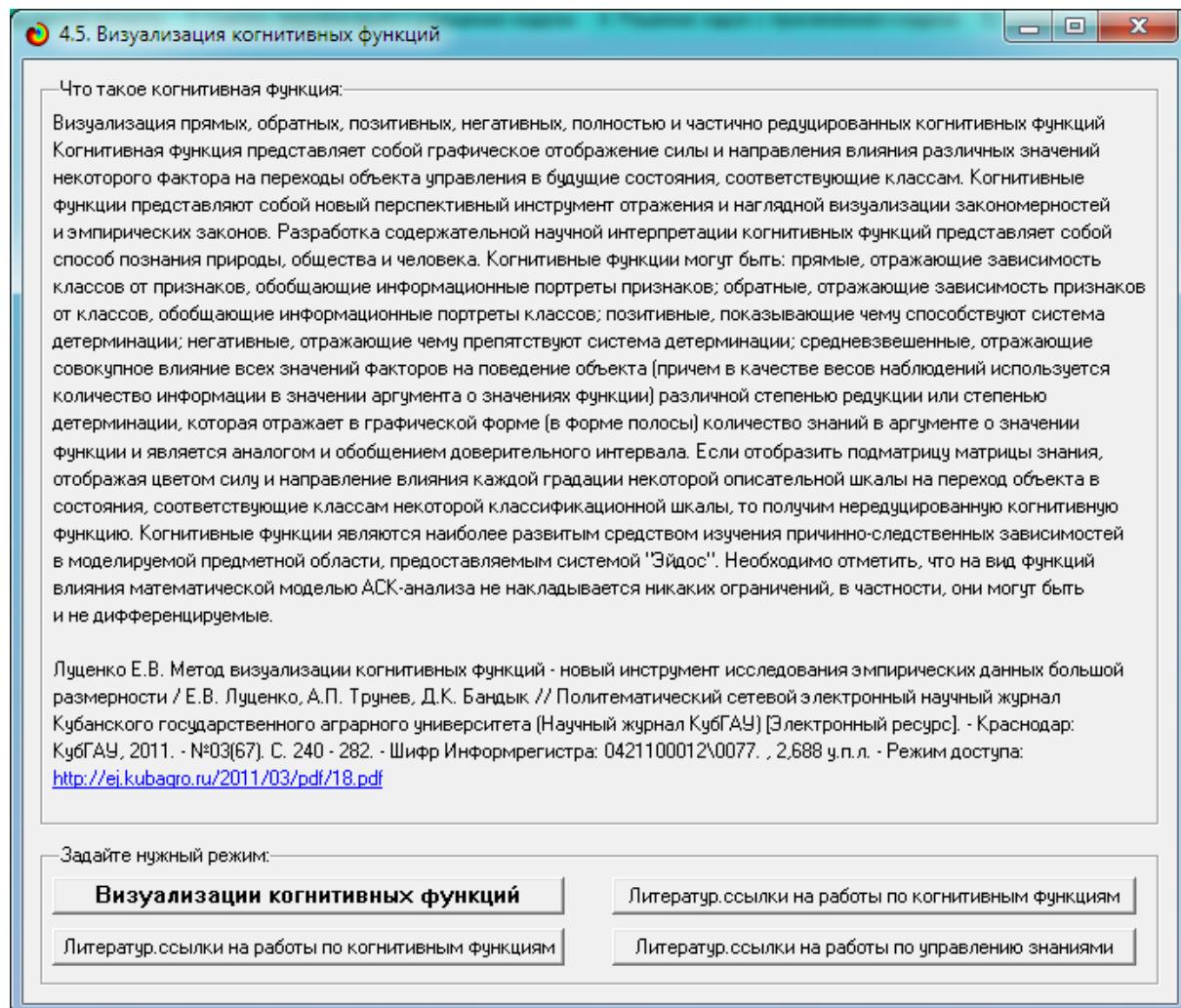


Рисунок 16. Комментарий к режиму формирования когнитивных функций

Из рисунка 17 видно, что:

– признаки в диапазоне от 01 до 10 содержат большое количество информации о принадлежности к одним классам (обозначены цветами теплой гаммы) и непринадлежности к другим классам (холодные цвета). Эти признаки характерны для объектов с истинными описаниями;

– признаки в диапазоне от 11 до 20 содержат большое количество информации о принадлежности к одним классам (обозначены цветами теплой гаммы), но малое количество информации о непринадлежности к другим классам (холодные цвета). Эти признаки характерны для

нетипичных объектов классов, т.е. объектов с закономерными, но ложными описаниями;

– признаки в диапазоне от 21 до 98 описывают объекты со случайными описаниями, т.е. артефакты.

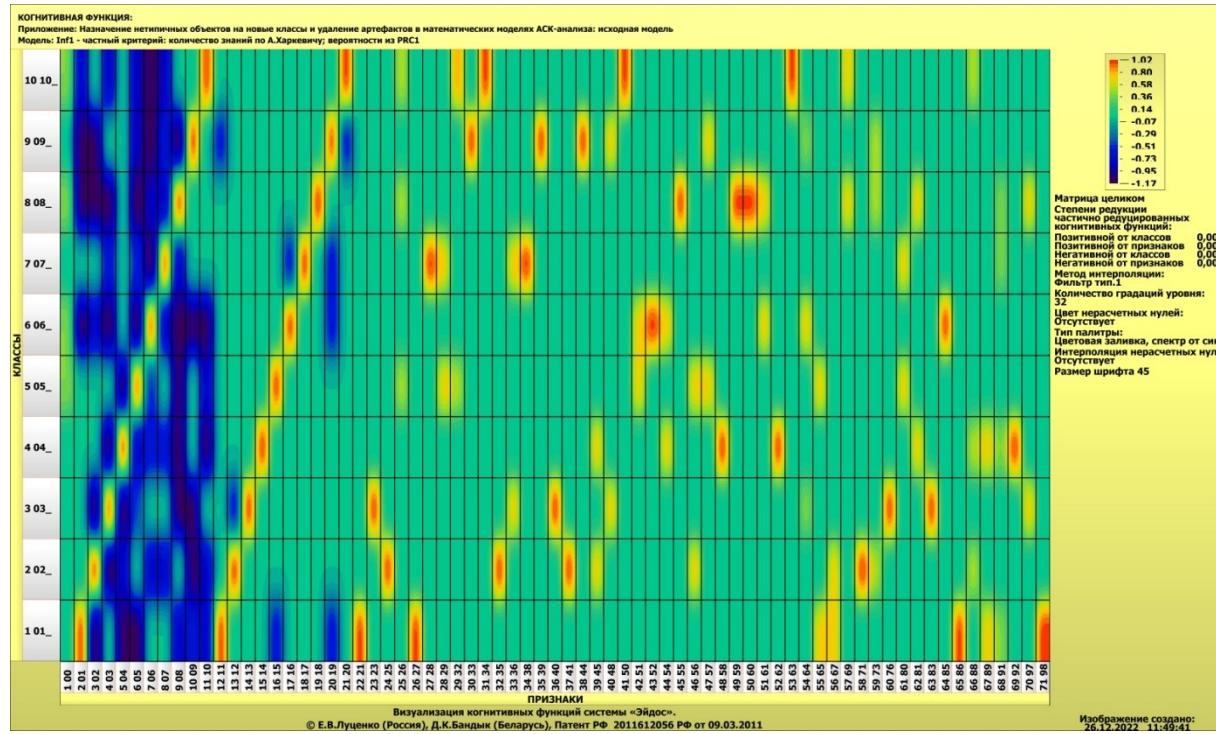


Рисунок 17. Когнитивная функция, построенная в модели INF1 на нулевой итерации (исходная модель)

Рассмотрим теперь как в системе «Эйдос» осуществляется выявление нетипичных объектов и их назначение на новые классы (разделение классов на типичную и нетипичную части).

4.4. Шаг 5. Начало 1-й итерации: выявление нетипичных объектов и их назначение на новые классы (разделение классов на типичную и нетипичную части)

Эти процедуры осуществляются в режиме 3.6 системы «Эйдос» (рисунки 18, 19, 20). Запустим режим 3.6 при параметрах, приведенных на рисунке 18.

В результате работы режима получим экранную форму, приведенную на рисунке 19.

Отметим, что 60 объектов, для которых созданы новые классы, включают 40 нетипичных объекта по 4 на каждый из 10 классов и 20 артефактов по 2 на каждый класс.

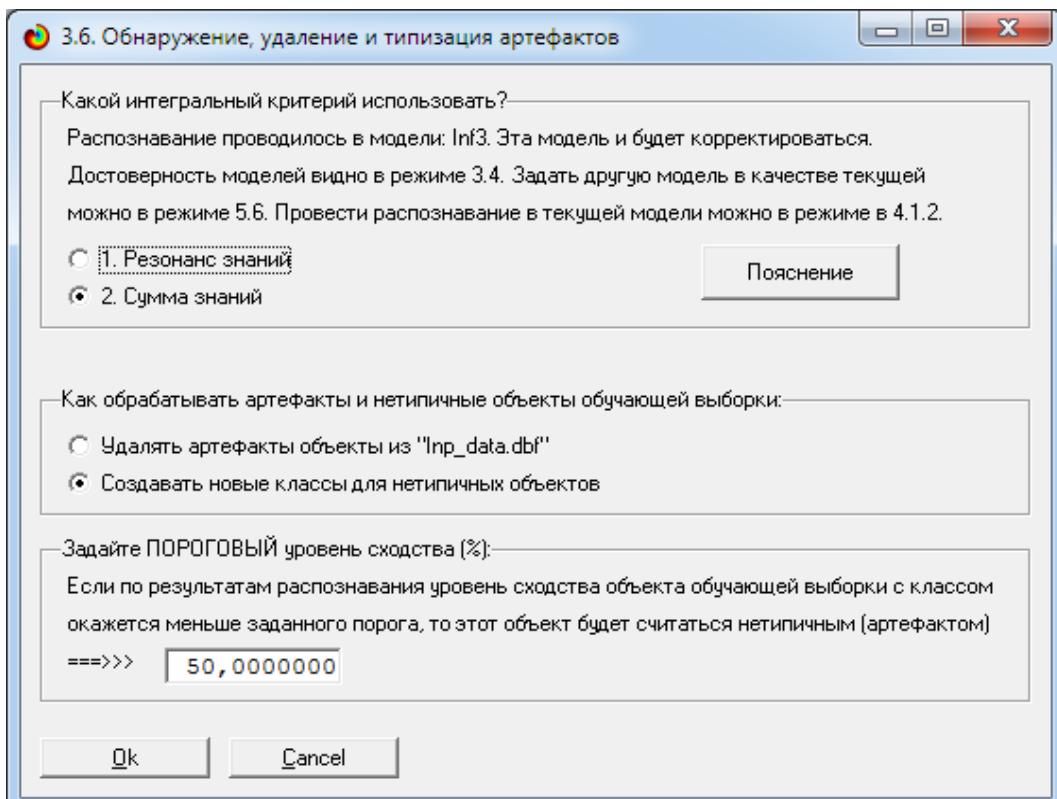


Рисунок 18. Экранная форма управления режимом 3.6 системы «Эйдос»

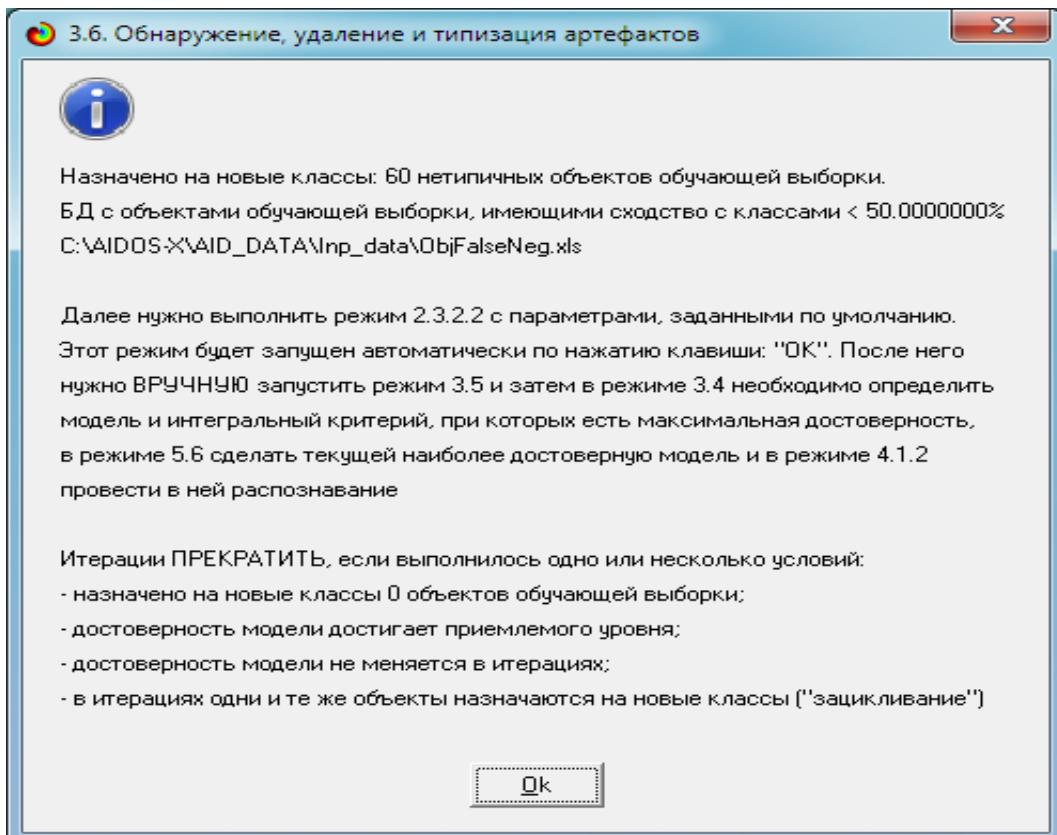


Рисунок 19. Экранная форма по результатам работы режима 3.6

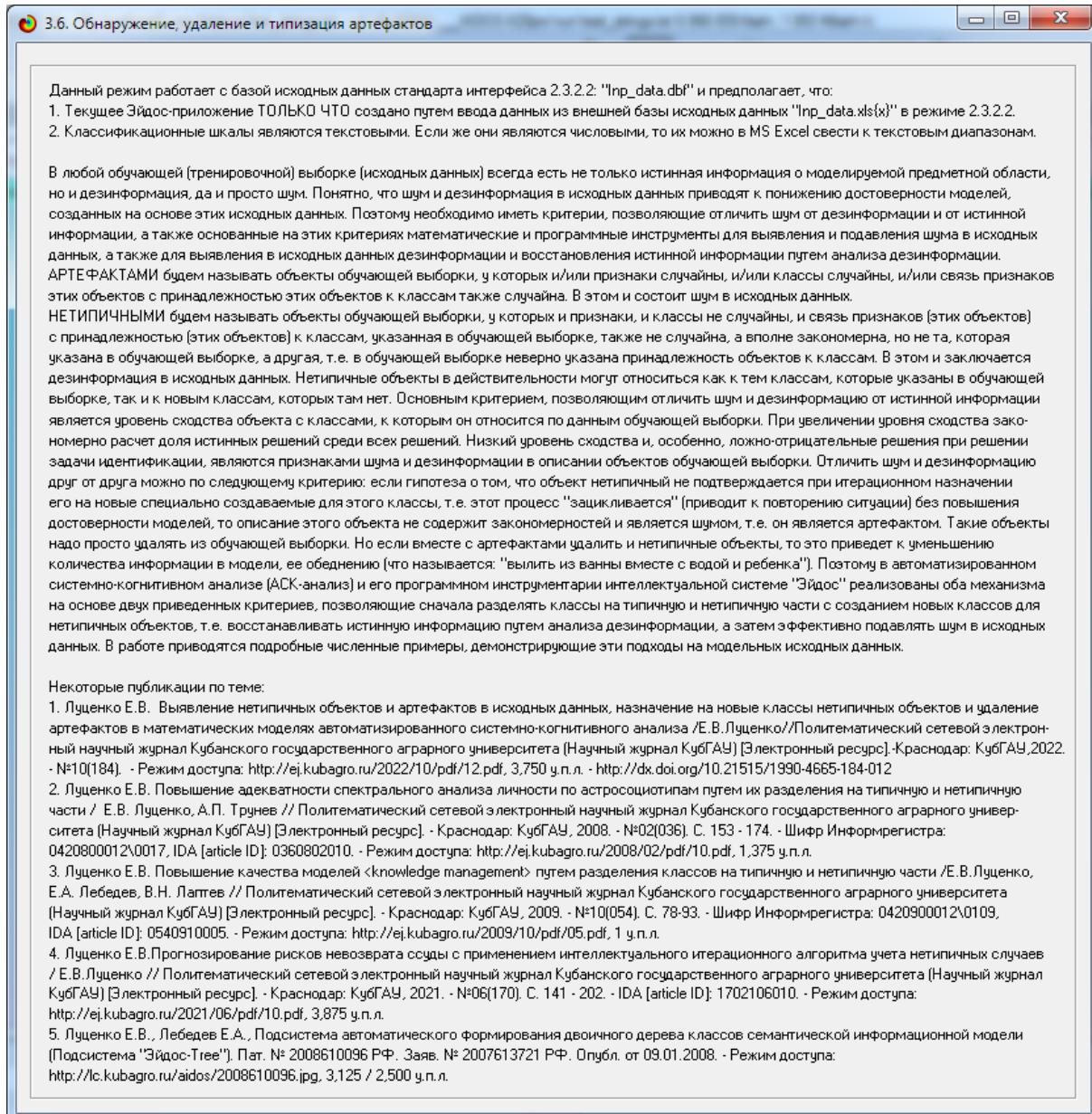


Рисунок 20. Экранная форма help режима 3.6 системы «Эйдос»

4.5. Шаг 6. Проверка условий завершения процесса итераций

Завершение процесса итераций, т.е. переход на шаг 7, осуществляется при выполнении одного или нескольких из следующих условий:

1. Назначено на новые классы 0 (или другое не изменяющееся в итерациях количество) объектов обучающей выборки.
2. Достоверность модели достигает приемлемого уровня.
3. Достоверность модели не меняется в итерациях.
4. В итерациях одни и те же объекты назначаются на новые классы (процесс «зацикливается»).

Иначе **продолжение** процесса итераций, т.е. переход на шаг 1.

Где получить информацию по пунктам условий окончания итераций.

По п.1 информация содержится в форме, приведенной на рисунке 19.

По п.2 и п.3 информация содержится в форме, приведенной на рисунке 12.

По п.4 информация содержится в форме, приведенной в таблице 5.

Поскольку, по сути, в начале 1-й итерации сравнивать исходную ситуацию не с чем, т.к. еще нет никакой *динамики* изменения ситуации, то переходим на последующие этапы 1-й итерации: создание ее модели.

4.6. Создание модели 1-й итерации

Шаг 1. Формализация предметной области

После нажатия «Ok» на экранной форме, приведенной на рисунке 19, *автоматически* запускается API-2.3.2.2. На рисунке 21 приведена экранная форма управления API-2.3.2.2 с параметрами по умолчанию после запуска режима 3.6.

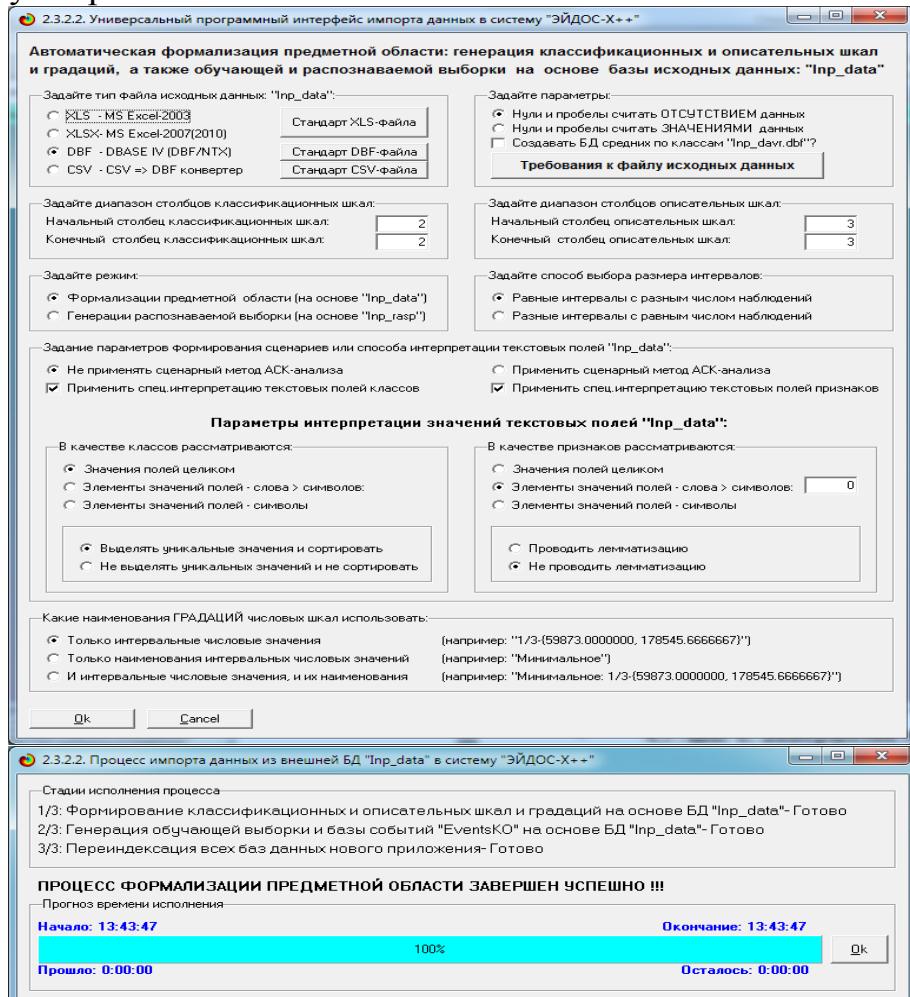


Рисунок 21. Экранные формы управления процесса исполнения API-2.3.2.2 с параметрами по умолчанию после запуска режима 3.6.

Как видно из рисунка 21 в качестве источника исходных данных для создания моделей будет использован файл «Inp_data.dbf», созданный режимом 3.6 на предыдущем шаге. Этот файл полностью аналогичен исходным данным, приведенным в таблице 1, за исключением того, что нетипичные объекты и артефакты **назначены** на новые специально созданные для этого классы (таблица 5).

Таблица 5 – Исходные данные после назначения нетипичных объектов и артефактов на новые классы в 1-й итерации

N1	N2	N3	N1	N2	N3	N1	N2	N3
01_01i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_07i	04	04 04 04 04 04 04 04 04 04 04 04 04 04 04	07_13a	07_1	29 80 38 91 19 28 36 16 03 06
01_02i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_08i	04	04 04 04 04 04 04 04 04 04 04 04 04 04 04	07_14a	07_1	05 07 03 07 05 03 05 08 05 08
01_03i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_09i	04	04 04 04 04 04 04 04 04 04 04 04 04 04 04	07_15n	07_1	17 17 17 17 17 17 17 17 17 17 17 17 17
01_04i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_10i	04	04 04 04 04 04 04 04 04 04 04 04 04 04 04	07_16n	07_1	17 17 17 17 17 17 17 17 17 17 17 17 17
01_05i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_11i	04	04 04 04 04 04 04 04 04 04 04 04 04 04 04	07_17n	07_1	17 17 17 17 17 17 17 17 17 17 17 17 17
01_06i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_12i	04	04 04 04 04 04 04 04 04 04 04 04 04 04 04	07_18n	07_1	17 17 17 17 17 17 17 17 17 17 17 17 17
01_07i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_13a	04_1	54 89 45 92 82 81 58 07 88 91	08_01i	08	08 08 08 08 08 08 08 08 08 08 08 08 08
01_08i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_14a	04_1	05 10 06 07 03 08 05 06 06 07	08_02i	08	08 08 08 08 08 08 08 08 08 08 08 08 08
01_09i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_15n	04_1	14 14 14 14 14 14 14 14 14 14 14 14	08_03i	08	08 08 08 08 08 08 08 08 08 08 08 08 08
01_10i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_16n	04_1	14 14 14 14 14 14 14 14 14 14 14 14	08_04i	08	08 08 08 08 08 08 08 08 08 08 08 08 08
01_11i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_17n	04_1	14 14 14 14 14 14 14 14 14 14 14 14	08_05i	08	08 08 08 08 08 08 08 08 08 08 08 08 08
01_12i	01	01 01 01 01 01 01 01 01 01 01 01 01 01 01	04_18n	04_1	14 14 14 14 14 14 14 14 14 14 14 14	08_06i	08	08 08 08 08 08 08 08 08 08 08 08 08 08
01_13a	01_1	98 27 91 65 89 15 21 19 67 86	05_01i	05	05 05 05 05 05 05 05 05 05 05 05 05 05	08_07i	08	08 08 08 08 08 08 08 08 08 08 08 08 08
01_14a	01_1	10 02 09 08 09 04 10 08 09 05	05_02i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_08i	08	08 08 08 08 08 08 08 08 08 08 08 08
01_15n	01_1	11 11 11 11 11 11 11 11 11 11 11 11 11 11	05_03i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_09i	08	08 08 08 08 08 08 08 08 08 08 08 08
01_16n	01_1	11 11 11 11 11 11 11 11 11 11 11 11 11 11	05_04i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_10i	08	08 08 08 08 08 08 08 08 08 08 08 08
01_17n	01_1	11 11 11 11 11 11 11 11 11 11 11 11 11 11	05_05i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_11i	08	08 08 08 08 08 08 08 08 08 08 08 08
01_18n	01_1	11 11 11 11 11 11 11 11 11 11 11 11 11 11	05_06i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_12i	08	08 08 08 08 08 08 08 08 08 08 08 08
02_01i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_07i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_13a	08_1	55 69 61 81 97 59 26 91 73 60
02_02i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_08i	05_	05 05 05 05 05 05 05 05 05 05 05 05	08_14a	08_1	08 03 05 06 06 00 01 07 07 02
02_03i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_09i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_15n	08_1	18 18 18 18 18 18 18 18 18 18 18 18
02_04i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_10i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_16n	08_1	18 18 18 18 18 18 18 18 18 18 18 18
02_05i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_11i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_17n	08_1	18 18 18 18 18 18 18 18 18 18 18 18
02_06i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_12i	05	05 05 05 05 05 05 05 05 05 05 05 05	08_18n	08_1	18 18 18 18 18 18 18 18 18 18 18 18
02_07i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_13a	05_1	15 57 05 29 56 26 51 32 80 65	09_01i	09	09 09 09 09 09 09 09 09 09 09 09 09
02_08i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_14a	05_1	00 06 06 10 06 05 08 06 10 04	09_02i	09	09 09 09 09 09 09 09 09 09 09 09 09
02_09i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_15n	05_1	15 15 15 15 15 15 15 15 15 15 15 15	09_03i	09	09 09 09 09 09 09 09 09 09 09 09 09
02_10i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_16n	05_1	15 15 15 15 15 15 15 15 15 15 15 15	09_04i	09	09 09 09 09 09 09 09 09 09 09 09 09
02_11i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_17n	05_1	15 15 15 15 15 15 15 15 15 15 15 15	09_05i	09	09 09 09 09 09 09 09 09 09 09 09 09
02_12i	02	02 02 02 02 02 02 02 02 02 02 02 02	05_18n	05_1	15 15 15 15 15 15 15 15 15 15 15 15	09_06i	09	09 09 09 09 09 09 09 09 09 09 09 09
02_13a	02_1	25 41 73 56 45 71 67 04 88 35	06_01i	06	06 06 06 06 06 06 06 06 06 06	09_07i	09	09 09 09 09 09 09 09 09 09 09 09 09
02_14a	02_1	07 10 06 10 09 03 07 06 04 02	06_02i	06	06 06 06 06 06 06 06 06 06 06	09_08i	09_	09 09 09 09 09 09 09 09 09 09 09 09
02_15n	02_1	12 12 12 12 12 12 12 12 12 12 12	06_03i	06	06 06 06 06 06 06 06 06 06 06	09_09i	09	09 09 09 09 09 09 09 09 09 09 09 09
02_16n	02_1	12 12 12 12 12 12 12 12 12 12 12	06_04i	06	06 06 06 06 06 06 06 06 06 06	09_10i	09	09 09 09 09 09 09 09 09 09 09 09 09
02_17n	02_1	12 12 12 12 12 12 12 12 12 12 12	06_05i	06	06 06 06 06 06 06 06 06 06 06	09_11i	09_	09 09 09 09 09 09 09 09 09 09 09 09
02_18n	02_1	12 12 12 12 12 12 12 12 12 12 12	06_06i	06	06 06 06 06 06 06 06 06 06 06	09_12i	09	09 09 09 09 09 09 09 09 09 09 09 09
03_01i	03	03 03 03 03 03 03 03 03 03 03 03	06_07i	06	06 06 06 06 06 06 06 06 06 06	09_13a	09_1	48 44 73 33 57 11 64 20 39 08
03_02i	03	03 03 03 03 03 03 03 03 03 03	06_08i	06	06 06 06 06 06 06 06 06 06 06	09_14a	09_1	05 06 07 05 07 09 09 07 02 01
03_03i	03_	03 03 03 03 03 03 03 03 03 03	06_09i	06	06 06 06 06 06 06 06 06 06 06	09_15n	09_1	19 19 19 19 19 19 19 19 19 19
03_04i	03	03 03 03 03 03 03 03 03 03 03	06_10i	06	06 06 06 06 06 06 06 06 06 06	09_16n	09_1	19 19 19 19 19 19 19 19 19 19
03_05i	03	03 03 03 03 03 03 03 03 03 03	06_11i	06	06 06 06 06 06 06 06 06 06 06	09_17n	09_1	19 19 19 19 19 19 19 19 19 19
03_06i	03	03 03 03 03 03 03 03 03 03 03	06_12i	06	06 06 06 06 06 06 06 06 06 06	09_18n	09_1	19 19 19 19 19 19 19 19 19 19
03_07i	03	03 03 03 03 03 03 03 03 03 03	06_13a	06_1	64 51 64 00 85 54 19 61 85 52	10_01i	10	10 10 10 10 10 10 10 10 10 10
03_08i	03	03 03 03 03 03 03 03 03 03 03	06_14a	06_1	09 01 08 07 02 10 05 03 02 07	10_02i	10	10 10 10 10 10 10 10 10 10 10
03_09i	03	03 03 03 03 03 03 03 03 03 03	06_15n	06_1	16 16 16 16 16 16 16 16 16 16	10_03i	10	10 10 10 10 10 10 10 10 10 10
03_10i	03	03 03 03 03 03 03 03 03 03 03	06_16n	06_1	16 16 16 16 16 16 16 16 16 16	10_04i	10	10 10 10 10 10 10 10 10 10 10
03_11i	03_	03 03 03 03 03 03 03 03 03 03	06_17n	06_1	16 16 16 16 16 16 16 16 16 16	10_05i	10	10 10 10 10 10 10 10 10 10 10
03_12i	03	03 03 03 03 03 03 03 03 03 03	06_18n	06_1	16 16 16 16 16 16 16 16 16 16	10_06i	10	10 10 10 10 10 10 10 10 10 10
03_13a	03_1	48 36 40 83 12 05 23 97 76 64	07_01i	07	07 07 07 07 07 07 07 07 07 07	10_07i	10	10 10 10 10 10 10 10 10 10 10
03_14a	03_1	05 03 03 05 02 08 09 05 04 04	07_02i	07	07 07 07 07 07 07 07 07 07 07	10_08i	10_	10 10 10 10 10 10 10 10 10 10
03_15n	03_1	13 13 13 13 13 13 13 13 13 13	07_03i	07	07 07 07 07 07 07 07 07 07 07	10_09i	10	10 10 10 10 10 10 10 10 10 10
03_16n	03_1	13 13 13 13 13 13 13 13 13 13	07_04i	07	07 07 07 07 07 07 07 07 07 07	10_10i	10	10 10 10 10 10 10 10 10 10 10
03_17n	03_1	13 13 13 13 13 13 13 13 13 13	07_05i	07	07 07 07 07 07 07 07 07 07 07	10_11i	10	10 10 10 10 10 10 10 10 10 10
03_18n	03_1	13 13 13 13 13 13 13 13 13 13	07_06i	07	07 07 07 07 07 07 07 07 07 07	10_12i	10	10 10 10 10 10 10 10 10 10 10
04_01i	04	04 04 04 04 04 04 04 04 04 04	07_07i	07	07 07 07 07 07 07 07 07 07 07	10_13a	10_1	26 03 63 10 69 32 88 34 32 50
04_02i	04	04 04 04 04 04 04 04 04 04 04	07_08i	07	07 07 07 07 07 07 07 07 07 07	10_14a	10_1	01 07 05 01 07 03 00 06 05 10
04_03i	04	04 04 04 04 04 04 04 04 04 04	07_09i	07	07 07 07 07 07 07 07 07 07 07	10_15n	10_1	20 20 20 20 20 20 20 20 20 20
04_04i	04	04 04 04 04 04 04 04 04 04 04	07_10i	07	07 07 07 07 07 07 07 07 07 07	10_16n	10_1	20 20 20 20 20 20 20 20 20 20
04_05i	04_	04 04 04 04 04 04 04 04 04 04	07_11i	07_	07 07 07 07 07 07 07 07 07 07	10_17n	10_1	20 20 20 20 20 20 20 20 20 20
04_06i	04	04 04 04 04 04 04 04 04 04 04	07_12i	07_	07 07 07 07 07 07 07 07 07 07	10_18n	10_1	20 20 20 20 20 20 20 20 20 20

Разработка автора.

Источник данных: c:\Aidos-X\AID_DATA\Inp_data\Inp_data.dbf

В таблице 5 для отметки артефактов

После выполнения API-2.3.2.2 пользователю надо выполнить рекомендации, приведенные на экранной форме (рисунок 22). Этим самым **создаются условия для использования результатов текущей итерации.**

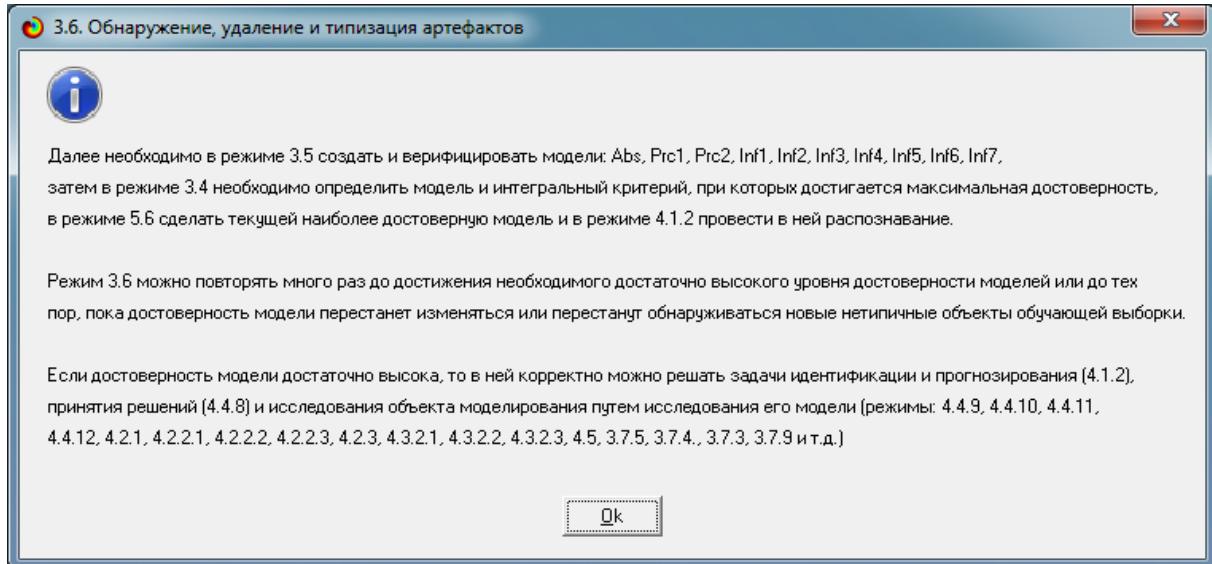


Рисунок 22. Рекомендациями пользователю по окончании работы API-2.3.2.2

В результате работы API-2.3.2.2 созданы классификационные и описательные шкалы и градации (таблицы 6 и 7), а затем с их использованием закодированы исходные данные (таблица 5), в результате чего получена обучающая выборка (таблица 8):

Таблица 6 – Классификационные шкалы и градации (полностью)

KOD_CLS	NAME_CLS
1	КЛАСС-01
2	КЛАСС-01_1
3	КЛАСС-02
4	КЛАСС-02_1
5	КЛАСС-03
6	КЛАСС-03_1
7	КЛАСС-04
8	КЛАСС-04_1
9	КЛАСС-05
10	КЛАСС-05_1

KOD_CLS	NAME_CLS
11	КЛАСС-06
12	КЛАСС-06_1
13	КЛАСС-07
14	КЛАСС-07_1
15	КЛАСС-08
16	КЛАСС-08_1
17	КЛАСС-09
18	КЛАСС-09_1
19	КЛАСС-10
20	КЛАСС-10_1

Таблица 7 – Описательные шкалы и градации (полностью)

KOD_ATR	NAME_ATR
1	ПРИЗНАКИ-00
2	ПРИЗНАКИ-01
3	ПРИЗНАКИ-02
4	ПРИЗНАКИ-03
5	ПРИЗНАКИ-04
6	ПРИЗНАКИ-05
7	ПРИЗНАКИ-06
8	ПРИЗНАКИ-07
9	ПРИЗНАКИ-08
10	ПРИЗНАКИ-09
11	ПРИЗНАКИ-10
12	ПРИЗНАКИ-11
13	ПРИЗНАКИ-12
14	ПРИЗНАКИ-13
15	ПРИЗНАКИ-14
16	ПРИЗНАКИ-15
17	ПРИЗНАКИ-16
18	ПРИЗНАКИ-17
19	ПРИЗНАКИ-18
20	ПРИЗНАКИ-19
21	ПРИЗНАКИ-20
22	ПРИЗНАКИ-21
23	ПРИЗНАКИ-23
24	ПРИЗНАКИ-25
25	ПРИЗНАКИ-26
26	ПРИЗНАКИ-27
27	ПРИЗНАКИ-28
28	ПРИЗНАКИ-29
29	ПРИЗНАКИ-32
30	ПРИЗНАКИ-33
31	ПРИЗНАКИ-34
32	ПРИЗНАКИ-35
33	ПРИЗНАКИ-36
34	ПРИЗНАКИ-38
35	ПРИЗНАКИ-39
36	ПРИЗНАКИ-40
37	ПРИЗНАКИ-41
38	ПРИЗНАКИ-44
39	ПРИЗНАКИ-45
40	ПРИЗНАКИ-48
41	ПРИЗНАКИ-50
42	ПРИЗНАКИ-51
43	ПРИЗНАКИ-52
44	ПРИЗНАКИ-54
45	ПРИЗНАКИ-55
46	ПРИЗНАКИ-56
47	ПРИЗНАКИ-57
48	ПРИЗНАКИ-58

KOD_ATR	NAME_ATR
49	ПРИЗНАКИ-59
49	ПРИЗНАКИ-59
50	ПРИЗНАКИ-60
51	ПРИЗНАКИ-61
52	ПРИЗНАКИ-62
53	ПРИЗНАКИ-63
54	ПРИЗНАКИ-64
55	ПРИЗНАКИ-65
56	ПРИЗНАКИ-67
57	ПРИЗНАКИ-69
58	ПРИЗНАКИ-71
59	ПРИЗНАКИ-73
60	ПРИЗНАКИ-76
61	ПРИЗНАКИ-80
62	ПРИЗНАКИ-81
63	ПРИЗНАКИ-83
64	ПРИЗНАКИ-85
65	ПРИЗНАКИ-86
66	ПРИЗНАКИ-88
67	ПРИЗНАКИ-89
68	ПРИЗНАКИ-91
69	ПРИЗНАКИ-92
70	ПРИЗНАКИ-97
71	ПРИЗНАКИ-98

Таблица 8 – Обучающая выборка (полностью)

NAME_OBJ	N2	N3
01_01i	1	2
01_02i	1	2
01_03i	1	2
01_04i	1	2
01_05i	1	2
01_06i	1	2
01_07i	1	2
01_08i	1	2
01_09i	1	2
01_10i	1	2
01_11i	1	2
01_12i	1	2
01_13a	2	65
01_14a	2	6
01_15n	2	12
01_16n	2	12
01_17n	2	12
01_18n	2	12
02_01i	3	3
02_02i	3	3
02_03i	3	3
02_04i	3	3
02_05i	3	3
02_06i	3	3
02_07i	3	3
02_08i	3	3
02_09i	3	3
02_10i	3	3
02_11i	3	3
02_12i	3	3
02_13a	4	32
02_14a	4	3
02_15n	4	13
02_16n	4	13
02_17n	4	13
02_18n	4	13
03_01i	5	4
03_02i	5	4
03_03i	5	4
03_04i	5	4
03_05i	5	4
03_06i	5	4
03_07i	5	4
03_08i	5	4
03_09i	5	4
03_10i	5	4
03_11i	5	4
03_12i	5	4
03_13a	6	54
03_14a	6	5
03_15n	6	14
03_16n	6	14
03_17n	6	14
03_18n	6	14
04_01i	7	5
04_02i	7	5
04_03i	7	5
04_04i	7	5
04_05i	7	5
04_06i	7	5
04_07i	7	5
04_08i	7	5
04_09i	7	5
04_10i	7	5
04_11i	7	5
04_12i	7	5
04_13a	8	68
04_14a	8	8
04_15n	8	15
04_16n	8	15
04_17n	8	15
04_18n	8	15
05_01i	9	6
05_02i	9	6
05_03i	9	6
05_04i	9	6
05_05i	9	6
05_06i	9	6
05_07i	9	6
05_08i	9	6
05_09i	9	6
05_10i	9	6
05_11i	9	6
05_12i	9	6
05_13a	10	55
05_14a	10	5
05_15n	10	16
05_16n	10	16
05_17n	10	16
05_18n	10	16
06_01i	11	7
06_02i	11	7
06_03i	11	7
06_04i	11	7
06_05i	11	7
06_06i	11	7
06_07i	11	7
06_08i	11	7
06_09i	11	7
06_10i	11	7
06_11i	11	7
06_12i	11	7
06_13a	12	43
06_14a	12	8
06_15n	12	17
06_16n	12	17
06_17n	12	17
06_18n	12	17
07_01i	13	8
07_02i	13	8
07_03i	13	8
07_04i	13	8
07_05i	13	8
07_06i	13	8
07_07i	13	8
07_08i	13	8
07_09i	13	8
07_10i	13	8
07_11i	13	8
07_12i	13	8
07_13a	14	7
07_14a	14	9
07_15n	14	18
07_16n	14	18
07_17n	14	18
07_18n	14	18
08_01i	15	9
08_02i	15	9
08_03i	15	9
08_04i	15	9
08_05i	15	9
08_06i	15	9
08_07i	15	9
08_08i	15	9
08_09i	15	9

Шаг 2. Синтез и верификация статистических и системно-когнитивных моделей

Таким образом, работа API-2.3.2.2 полностью подготовила следующий этап АСК-анализа: синтез и верификацию статистических и системно-когнитивных моделей. Этот этап выполняется в режиме 3.5 системы «Эйдос». В результате работы режима синтеза и верификации моделей были созданы 3 статистические и 7 системно-когнитивные модели. Способ расчета этих моделей кратко, но полно описан в работе [8], а также в help режима 5.5 системы «Эйдос».

Шаг 3. Оценка достоверности статистических и системно-когнитивных моделей

Оценка достоверности статистических и системно-когнитивных моделей осуществляется в режиме 3.4 системы «Эйдос».

Достоверность исходных моделей на 2-й итерации оказалась довольно высокой: F-мера Ван Ризбергена для модели INF3 равна F=0,754 при максимуме 1,000 (рисунок 23):

Наименование модели и частичного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно положительных решений (TP)	Число истинно отрицательных решений (TN)	Число ложно положительных решений (FP)	Число ложно отрицательных решений (FN)	Точность модели	Положка модели	Гибера Ван Рибергена	Сумма между истинно-ложн. соо. и ложн.-истин. соо.	Сумма между ложн. соо. и истинно-ложн. соо.	Сумма между ложн. соо. и ложн.-истин. соо.
1. ABS - частичный критерий количества встреч сочетаний "класс	Корреляция abs частот с обр.	180	180	2482	938	0	0.161	1.000	0.277	160,665	46,565	41,619
1. ABS - частичный критерий количество встреч сочетаний "хоз	Сумма abs частот по признаку	180	180	2390	1030	0	0.149	1.000	0.259	133,472	22,738	22,738
2. PRC1 - частичный критерий л. вероятности ит. признака в сред.	Корреляция ул.отч частот с обр.	180	180	2482	938	0	0.161	1.000	0.277	160,665	46,565	41,619
2. PRC1 - частичный критерий л. вероятности ит. признака в сред.	Сумма ул.отч частот по при.	180	180	2390	1030	0	0.149	1.000	0.259	147,343	34,877	34,877
3. PRC2 - частичный критерий л.вероятность ит. признака	Корреляция ул.отч частот с обр.	180	180	2482	938	0	0.161	1.000	0.277	160,665	46,565	41,619
3. PRC2 - частичный критерий л.вероятность ит. признака	Сумма ул.отч частот по при.	180	180	2390	1030	0	0.149	1.000	0.259	147,343	34,877	34,877
4. INF1 - частичный критерий количества знань об Ахереном в среде	Семантический резонанс энз...	180	170	3323	97	10	0.637	0.944	0.743	140,413	185,859	23,181
4. INF1 - частичный критерий количества знань об Ахереном в среде	Сумма знаний	180	170	3276	144	10	0.541	0.944	0.698	139,046	205,006	13,380
5. INF2 - частичный критерий количества знань об Ахереном в среде	Семантический резонанс энз...	180	170	3323	97	10	0.637	0.944	0.743	140,413	185,859	23,181
5. INF2 - частичный критерий количества знань об Ахереном в среде	Сумма знаний	180	170	3276	144	10	0.541	0.944	0.698	139,046	205,006	13,380
6. INF3 - частичный критерий "известад" различия между фактами	Семантический резонанс энз...	180	169	3321	99	11	0.631	0.939	0.754	160,099	233,542	17,063
6. INF3 - частичный критерий "известад" различия между фактами	Сумма знаний	180	169	3321	99	11	0.631	0.939	0.754	133,564	140,229	6,762
7. INF4 - частичный критерий ROI (Return On Investment); вероятн.	Семантический резонанс энз...	180	170	3337	83	10	0.672	0.944	0.785	144,210	112,150	22,444
7. INF4 - частичный критерий ROI (Return On Investment); вероятн.	Сумма знаний	180	173	3265	155	7	0.527	0.961	0.683	97,403	17,572	5,194
8. INF5 - частичный критерий ROI (Return On Investment); вероятн.	Семантический резонанс энз...	180	170	3337	83	10	0.672	0.944	0.785	144,210	112,150	22,444
8. INF5 - частичный критерий ROI (Return On Investment); вероятн.	Сумма знаний	180	173	3265	155	7	0.527	0.961	0.683	97,403	17,572	5,194
9. INF6 - частичный критерий разн. усл и безсл.вероятностей еф.	Семантический резонанс энз...	180	168	3342	78	12	0.663	0.933	0.789	159,734	105,761	23,905
9. INF6 - частичный критерий разн. усл и безсл.вероятностей еф.	Сумма знаний	180	169	3298	122	11	0.581	0.939	0.718	147,530	39,793	11,226
10. INF7 - частичный критерий разн. усл и безсл.вероятностей еф.	Семантический резонанс энз...	180	168	3342	78	12	0.663	0.933	0.789	159,734	105,761	23,905
10. INF7 - частичный критерий разн. усл и безсл.вероятностей еф.	Сумма знаний	180	169	3298	122	11	0.581	0.939	0.718	147,530	39,793	11,226

Рисунок 23. Достоверность системно-когнитивной модели INF3 системы «Эйдос» на 1-й итерации разделения классов на типичную и нетипичную части

Отметим, что в исходной модели $F=0,854$ (рисунок 13), т.е. достоверность моделей немного понизилась [1-8].

Шаг 4. Идентификация в наиболее достоверной модели

В качестве текущей зададим наиболее достоверную по F-критерию Ван Ризбергена и критерию L1 модель INF3. Для задания модели в качестве текущей используется режим 5.6 системы «Эйдос», а для решения задачи идентификации объектов обучающей выборки в текущей модели режим 4.1.2.

Подобные результаты идентификации объектов обучающей выборки с каждым из классов приведены в экранных формах на рисунках 24.

Интегральный критерий сходимости "Семантический распознавание знаний"						
Код	Название объекта	Статус	Ф	Соцети		
1	KMCC_01	99.98	v	██████████		
2	KMCC_01_1	99.98	v	██████████		
3	KMCC_02	99.98	v	██████████		
4	KMCC_03	99.98	v	██████████		
5	KMCC_03_1	99.98	v	██████████		
6	KMCC_03_1_1	99.98	v	██████████		
7	KMCC_04	99.98	v	██████████		
8	KMCC_04_1	99.98	v	██████████		
9	KMCC_05	99.98	v	██████████		
10	KMCC_05_1	99.98	v	██████████		
11	KMCC_06	99.98	v	██████████		
12	KMCC_06_1	99.98	v	██████████		
13	KMCC_07	35.53		██████		
14	KMCC_07_1	81.32		███		
15	KMCC_08	6504		██		
16	KMCC_08_1	4459		██		
17	KMCC_09	0.370				
18	KMCC_09_1	0.370				
19	KMCC_10	0.974				
20	KMCC_10_1					

Интегральный критерий сходимости "Судья знаний"						
Код	Название объекта	Статус	Ф	Соцети		
1	01_01	100.00	v	██████████		
2	01_02	100.00	v	██████████		
3	01_03	100.00	v	██████████		
4	01_04	100.00	v	██████████		
5	01_05	100.00	v	██████████		
6	01_06	100.00	v	██████████		
7	01_07	100.00	v	██████████		
8	01_08	100.00	v	██████████		
9	01_09	100.00	v	██████████		
10	01_10	100.00	v	██████████		
11	01_11	100.00	v	██████████		
12	01_12	100.00	v	██████████		
13	01_13	100.00	v	██████████		
14	01_14	14.32		███		
15	08_14a	3.600		██		
16	09_14a	2.925		██		
17	09_14a	2.800		██		
18	09_13a	0.125		██		
19	08_13a	0.384		██		
20	08_13a	-0.384		██		

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объект". Текущая модель "INP".

Классы		Интегральный критерий сущности "Семантический уровень знаний"		
Код	Название класса	Наименование объекта	Соцсети	Ф. Соцсети
1	KMACC_01_	19_02_01	99.98.	v ██████████
2	KMACC_01_1	20_02_03	99.98.	v ██████████
3	KMACC_02_	21_02_03	99.98.	v ██████████
4	KMACC_02_1	22_02_04	99.98.	v ██████████
5	KMACC_03_	23_02_05	99.98.	v ██████████
6	KMACC_03_1	24_02_06	99.98.	v ██████████
7	KMACC_04_	25_02_07	99.98.	v ██████████
8	KMACC_04_1	26_02_08	99.98.	v ██████████
9	KMACC_05_	27_02_08	99.98.	v ██████████
10	KMACC_05_1	28_02_10	99.98.	v ██████████
11	KMACC_06_	29_02_11	99.98.	v ██████████
12	KMACC_06_1	30_02_12	99.98.	v ██████████
13	KMACC_07_	104_08_14	97.04.	██████████
14	KMACC_07_1	140_08_14	8.142.	█
15	KMACC_08_	141_08_14	7.28.	█
16	KMACC_08_1	14_01_14	5.038.	█
17	KMACC_09_	159_08_14	4.575.	█
18	KMACC_09_1	50_03_14a	4.189.	█
19	KMACC_10_			
20	KMACC_10_1			

Интегральный критерий сущности "Сумма знаний"

Код	Наименование объекта	Соцсети	Ф. Соцсети
19_02_01	99.98.	v ██████████	
20_02_02	99.98.	v ██████████	
21_02_03	99.98.	v ██████████	
22_02_04	99.98.	v ██████████	
23_02_05	99.98.	v ██████████	
24_02_06	99.98.	v ██████████	
25_02_07	99.98.	v ██████████	
26_02_08	99.98.	v ██████████	
27_02_09	99.98.	v ██████████	
28_02_10	99.98.	v ██████████	
29_02_11	99.98.	v ██████████	
30_02_12	99.98.	v ██████████	
104_08_14	97.04.	██████████	
140_08_14	8.142.	█	
141_08_14	7.28.	█	
14_01_14	5.038.	█	
159_08_14	4.575.	█	
50_03_14a	4.189.	█	

Помощь Помощь Поиск объекта В начало БД В конец БД Правдивость Следующая Запись Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объект". Текущая модель "INP".

Классы		Интегральный критерий сущности "Семантический уровень знаний"		
Код	Название класса	Наименование объекта	Соцсети	Ф. Соцсети
1	KMACC_01_	02_15n	98.95.	v ██████████
2	KMACC_01_1	03_15n	98.95.	v ██████████
3	KMACC_02_	02_17n	98.95.	v ██████████
4	KMACC_02_1	02_18n	98.95.	v ██████████
5	KMACC_03_	03_13a	25.22.	██████████
6	KMACC_03_1	31_02_13a	4.744.	v █
7	KMACC_04_	139_08_13a	0.225.	
8	KMACC_04_1	67_04_13a	0.953.	
9	KMACC_05_	103_06_13a	3.388.	█
10	KMACC_05_1	11_01_13a	-1.732.	█
11	KMACC_06_	51_03_15n	-3.401.	█
12	KMACC_06_1	52_03_16n	-3.401.	█
13	KMACC_07_	53_03_17n	-3.401.	█
14	KMACC_07_1	54_03_18n	-3.401.	█
15	KMACC_08_	68_04_13a	-3.401.	█
16	KMACC_08_1	70_04_15n	-3.401.	█
17	KMACC_09_	71_04_17n	-3.401.	█
18	KMACC_09_1	72_04_18n	-3.401.	█
19	KMACC_10_			
20	KMACC_10_1			

Интегральный критерий сущности "Сумма знаний"

Код	Наименование объекта	Соцсети	Ф. Соцсети
02_15n	34.59.	v ██████████	
03_15n	34.59.	v ██████████	
02_17n	34.59.	v ██████████	
02_18n	34.59.	v ██████████	
03_13a	3.000.	█	
31_02_13a	0.564.	v	
139_08_13a	0.026.		
67_04_13a	-0.101.		
103_06_13a	-0.197.		
11_01_13a	-0.286.		
51_03_15n	-0.471.		
52_03_16n	-0.507.		
53_03_17n	-0.595.		
121_07_13a	-0.808.		
51_03_15n	-1.194.	█	
52_03_16n	-1.194.	█	
53_03_17n	-1.194.	█	
54_03_18n	-1.194.	█	

Помощь Помощь Поиск объекта В начало БД В конец БД Правдивость Следующая Запись Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объект". Текущая модель "INP".

Классы		Интегральный критерий сущности "Семантический уровень знаний"		
Код	Название класса	Наименование объекта	Соцсети	Ф. Соцсети
1	KMACC_01_	03_01	99.98.	v ██████████
2	KMACC_01_1	30_02_03	99.98.	v ██████████
3	KMACC_02_	39_02_03	99.98.	v ██████████
4	KMACC_02_1	40_02_04	99.98.	v ██████████
5	KMACC_03_	41_03_05	99.98.	v ██████████
6	KMACC_03_1	42_03_06	99.98.	v ██████████
7	KMACC_04_	43_03_07	99.98.	v ██████████
8	KMACC_04_1	44_03_08	99.98.	v ██████████
9	KMACC_05_	45_03_09	99.98.	v ██████████
10	KMACC_05_1	46_03_10	99.98.	v ██████████
11	KMACC_06_	47_03_11	99.98.	v ██████████
12	KMACC_06_1	48_03_12	99.98.	v ██████████
13	KMACC_07_	50_03_14a	36.96.	██████████
14	KMACC_07_1	121_07_13a	25.44.	██████████
15	KMACC_08_	179_10_13a	13.40.	██████████
16	KMACC_08_1	122_07_14a	24.19.	██████████
17	KMACC_09_	140_08_14	8.162.	█
18	KMACC_09_1	176_10_14a	7.236.	█
19	KMACC_10_			
20	KMACC_10_1			

Интегральный критерий сущности "Сумма знаний"

Код	Наименование объекта	Соцсети	Ф. Соцсети
03_01	99.64.	v ██████████	
30_02_03	99.64.	v ██████████	
39_02_03	99.64.	v ██████████	
40_02_04	99.64.	v ██████████	
41_03_05	99.64.	v ██████████	
42_03_06	99.64.	v ██████████	
43_03_07	99.64.	v ██████████	
44_03_08	99.64.	v ██████████	
45_03_09	99.64.	v ██████████	
46_03_10	99.64.	v ██████████	
47_03_11	99.64.	v ██████████	
48_03_12	99.64.	v ██████████	
50_03_14a	13.54.	██████████	
121_07_13a	24.19.	██████████	
179_10_13a	8.086.	█	
122_07_14a	13.40.	██████████	
140_08_14	0.591.	█	
176_10_14a	3.600.	█	

Помощь Помощь Поиск объекта В начало БД В конец БД Правдивость Следующая Запись Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объект". Текущая модель "INP".

Классы		Интегральный критерий сущности "Семантический уровень знаний"		
Код	Название класса	Наименование объекта	Соцсети	Ф. Соцсети
1	KMACC_01_	03_15n	99.26.	v ██████████
2	KMACC_01_1	50_03_15n	99.26.	v ██████████
3	KMACC_02_	51_03_17n	98.26.	v ██████████
4	KMACC_02_1	54_03_18n	98.28.	v ██████████
5	KMACC_03_	49_03_13a	5.627.	v █
6	KMACC_03_1	71_05_01	0.847.	
7	KMACC_04_	74_05_03	0.847.	
8	KMACC_04_1	75_05_03	0.847.	
9	KMACC_05_	76_05_04	0.847.	
10	KMACC_05_1	77_05_05	0.847.	
11	KMACC_06_	79_05_01	0.847.	
12	KMACC_06_1	80_05_08	0.847.	
13	KMACC_07_	81_05_09	0.847.	
14	KMACC_07_1	82_05_10	0.847.	
15	KMACC_08_	83_05_11	0.847.	
16	KMACC_08_1	84_05_12	0.847.	
17	KMACC_09_	130_06_13a	0.025.	
18	KMACC_09_1			
19	KMACC_10_			
20	KMACC_10_1			

Интегральный критерий сущности "Сумма знаний"

Код	Наименование объекта	Соцсети	Ф. Соцсети
03_15n	34.62.	v ██████████	
50_03_15n	34.62.	v ██████████	
51_03_16n	34.62.	v ██████████	
52_03_17n	34.62.	v ██████████	
54_03_18n	34.62.	v ██████████	
49_03_13a	0.071.	v	
71_05_01	0.298.		
74_05_03	0.298.		
75_05_03	0.298.		
76_05_04	0.298.		
77_05_05	0.298.		
79_05_01	0.298.		
80_05_08	0.298.		
82_05_10	0.298.		
83_05_11	0.298.		
84_05_12	0.298.		
130_06_13a	0.026.		

Помощь Помощь Поиск объекта В начало БД В конец БД Правдивость Следующая Запись Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объект". Текущая модель "INP".

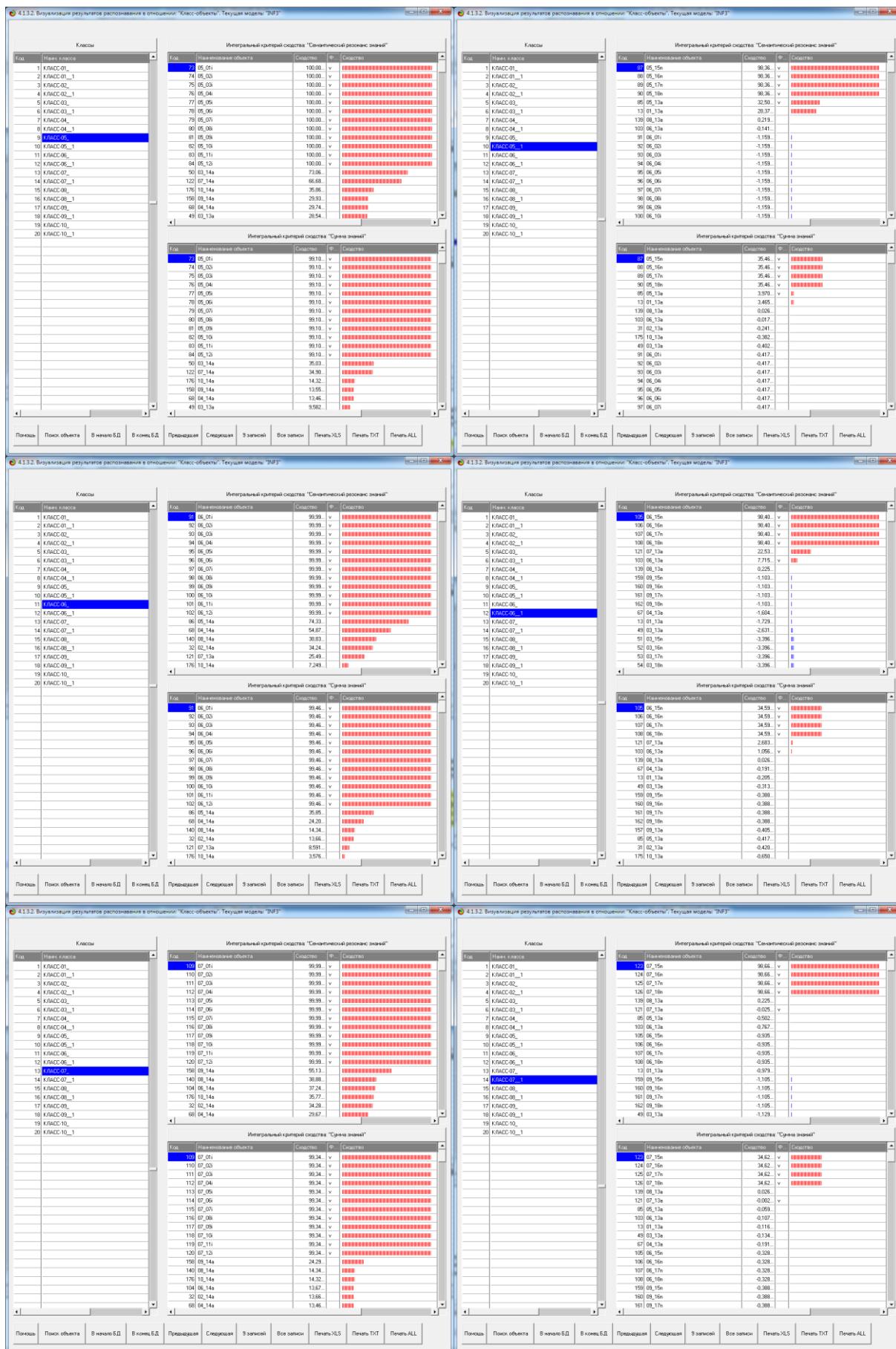
Классы		Интегральный критерий сущности "Семантический уровень знаний"		
Код	Название класса	Наименование объекта	Соцсети	Ф. Соцсети
1	KMACC_01_	04_15n	98.47.	v ██████████
2	KMACC_01_1	70_04_16n	98.47.	v ██████████
3	KMACC_02_	71_04_17n	98.47.	v ██████████
4	KMACC_02_1	72_04_18n	98.47.	v ██████████
5	KMACC_03_	67_04_13a	5.162.	v
6	KMACC_03_1	139_08_13a	0.977.	
7	KMACC_04_	103_06_13a	-0.765.	
8	KMACC_04_1	31_02_13a	-2.029.	█
9	KMACC_05_	48_04_13a	-3.326.	█
10	KMACC_05_1	51_03_15n	-3.295.	█
11	KMACC_06_	52_03_16n	-3.295.	█
12	KMACC_06_1	53_03_17n	-3.295.	█
13	KMACC_07_	54_03_18n	-3.295.	█
14	KMACC_07_1	123_07_15n	-3.295.	█
15	KMACC_08_	124_07_16n	-3.295.	█
16	KMACC_08_1	125_07_17n	-3.295.	█
17	KMACC_09_	126_07_18n	-3.295.	█
18	KMACC_09_1			
19	KMACC_10_			
20	KMACC_10_1			

Интегральный критерий сущности "Сумма знаний"

Код	Наименование объекта	Соцсети	Ф. Соцсети
04_15n	34.62.	v ██████████	
70_04_16n	34.62.	v ██████████	
71_04_17n	34.62.	v ██████████	
72_04_18n	34.62.	v ██████████	
67_04_13a	0.071.	v	
139_08_13a	0.116.		
103_06_13a	-0.107.		
31_02_13a	-0.116.		
48_04_13a	-0.241.		
51_03_15n	-0.402.		
52_03_16n	-0.417.		
53_03_17n	-0.417.		
123_07_15n	-0.581.		
124_07_16n	-0.585.		
125_07_17n	-0.585.		
126_07_18n	-0.585.		

Помощь Помощь Поиск объекта В начало БД В конец БД Правдивость Следующая Запись Все записи Печать XLS Печать TXT Печать ALL

<http://ej.kubagro.ru/2022/10/pdf/12.pdf>



4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель: "INF3".

Код	Название класса
1	KMACC_01
2	KMACC_01_1
3	KMACC_02
4	KMACC_02_1
5	KMACC_03
6	KMACC_03_1
7	KMACC_04
8	KMACC_04_1
9	KMACC_05
10	KMACC_05_1
11	KMACC_06
12	KMACC_06_1
13	KMACC_07
14	KMACC_07_1
15	KMACC_08
16	KMACC_08_1
17	KMACC_09
18	KMACC_09_1
19	KMACC_10
20	KMACC_10_1

Интегральный критерий сходимости "Семантический распознавание знаний"

Код	Название объекта	Соцсети	Ф. Соседство
137	08_01	99.99..	v
126	08_02	99.99..	v
126	08_03	99.99..	v
130	08_04	99.99..	v
131	08_05	99.99..	v
132	08_06	99.99..	v
123	08_07	99.99..	v
134	08_08	99.99..	v
135	08_09	99.99..	v
100	08_10	99.99..	v
137	08_11	99.99..	v
136	08_12	99.99..	v
14	01_14	30.21..	v
197	05_15a	27.99..	v
122	05_15b	24.16..	v
140	05_14a	0.157..	v
104	05_14a	6.524..	v
96	05_14b	5.641..	v
140	05_14b	2.929..	v
104	05_14b	0.157..	v

Показы Поиск объекта В начало БД В конец БД Предыдущая Следующая З записей Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель: "INF3".

Код	Название класса
1	KCLASS_01..
2	KCLASS_01_1
3	KCLASS_02..
4	KCLASS_02_1
5	KCLASS_03..
6	KCLASS_03_1
7	KCLASS_04..
8	KCLASS_04_1
9	KCLASS_05..
10	KCLASS_05_1
11	KCLASS_06..
12	KCLASS_06_1
13	KCLASS_07..
14	KCLASS_07_1
15	KCLASS_08..
16	KCLASS_08_1
17	KCLASS_09..
18	KCLASS_09_1
19	KCLASS_10..
20	KCLASS_10_1

Интегральный критерий сходимости "Сумма знаний"

Код	Название объекта	Соцсети	Ф. Соседство
145	08_01	99.99..	v
146	08_02	99.99..	v
147	08_03	99.99..	v
148	08_04	99.99..	v
149	08_05	99.99..	v
150	08_06	99.99..	v
151	08_07	99.99..	v
152	08_08	99.99..	v
153	08_09	99.99..	v
154	08_10	99.99..	v
155	08_11	99.99..	v
156	08_12	99.99..	v
14	01_14	29.72..	v
157	05_15a	29.72..	v
104	05_14a	6.516..	v
32	02_14a	6.716..	v
92	03_14a	4.111..	v
139	08_13a	0.370..	v

Показы Поиск объекта В начало БД В конец БД Предыдущая Следующая З записей Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель: "INF3".

Код	Название класса
1	KCLASS_01..
2	KCLASS_01_1
3	KCLASS_02..
4	KCLASS_02_1
5	KCLASS_03..
6	KCLASS_03_1
7	KCLASS_04..
8	KCLASS_04_1
9	KCLASS_05..
10	KCLASS_05_1
11	KCLASS_06..
12	KCLASS_06_1
13	KCLASS_07..
14	KCLASS_07_1
15	KCLASS_08..
16	KCLASS_08_1
17	KCLASS_09..
18	KCLASS_09_1
19	KCLASS_10..
20	KCLASS_10_1

Интегральный критерий сходимости "Сумма знаний"

Код	Название объекта	Соцсети	Ф. Соседство
145	08_01	99.92..	v
146	08_02	99.92..	v
147	08_03	99.92..	v
148	08_04	99.92..	v
149	08_05	99.92..	v
150	08_06	99.92..	v
151	08_07	99.92..	v
152	08_08	99.92..	v
153	08_09	99.92..	v
154	08_10	99.92..	v
155	08_11	99.92..	v
156	08_12	99.92..	v
14	01_14	24.50..	v
157	05_15a	1.99..	v
104	05_14a	2.925..	v
32	02_14a	2.919..	v
92	03_14a	2.794..	v
139	08_13a	0.125..	v

Показы Поиск объекта В начало БД В конец БД Предыдущая Следующая З записей Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель: "INF3".

Код	Название класса
1	KCLASS_01..
2	KCLASS_01_1
3	KCLASS_02..
4	KCLASS_02_1
5	KCLASS_03..
6	KCLASS_03_1
7	KCLASS_04..
8	KCLASS_04_1
9	KCLASS_05..
10	KCLASS_05_1
11	KCLASS_06..
12	KCLASS_06_1
13	KCLASS_07..
14	KCLASS_07_1
15	KCLASS_08..
16	KCLASS_08_1
17	KCLASS_09..
18	KCLASS_09_1
19	KCLASS_10..
20	KCLASS_10_1

Интегральный критерий сходимости "Сумма знаний"

Код	Название объекта	Соцсети	Ф. Соседство
145	08_01	99.99..	v
146	08_02	99.99..	v
147	08_03	99.99..	v
148	08_04	99.99..	v
149	08_05	99.99..	v
150	08_06	99.99..	v
151	08_07	99.99..	v
152	08_08	99.99..	v
153	08_09	99.99..	v
154	08_10	99.99..	v
155	08_11	99.99..	v
156	08_12	99.99..	v
14	01_14	30.21..	v
157	05_15a	29.72..	v
104	05_14a	6.516..	v
32	02_14a	6.716..	v
92	03_14a	4.111..	v
139	08_13a	0.370..	v

Показы Поиск объекта В начало БД В конец БД Предыдущая Следующая З записей Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель: "INF3".

Код	Название класса
1	KCLASS_01..
2	KCLASS_01_1
3	KCLASS_02..
4	KCLASS_02_1
5	KCLASS_03..
6	KCLASS_03_1
7	KCLASS_04..
8	KCLASS_04_1
9	KCLASS_05..
10	KCLASS_05_1
11	KCLASS_06..
12	KCLASS_06_1
13	KCLASS_07..
14	KCLASS_07_1
15	KCLASS_08..
16	KCLASS_08_1
17	KCLASS_09..
18	KCLASS_09_1
19	KCLASS_10..
20	KCLASS_10_1

Интегральный критерий сходимости "Сумма знаний"

Код	Название объекта	Соцсети	Ф. Соседство
145	08_01	99.99..	v
146	08_02	99.99..	v
147	08_03	99.99..	v
148	08_04	99.99..	v
149	08_05	99.99..	v
150	08_06	99.99..	v
151	08_07	99.99..	v
152	08_08	99.99..	v
153	08_09	99.99..	v
154	08_10	99.99..	v
155	08_11	99.99..	v
156	08_12	99.99..	v
14	01_14	30.21..	v
157	05_15a	29.72..	v
104	05_14a	6.516..	v
32	02_14a	6.716..	v
92	03_14a	4.111..	v
139	08_13a	0.370..	v

Показы Поиск объекта В начало БД В конец БД Предыдущая Следующая З записей Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель: "INF3".

Код	Название класса
1	KCLASS_01..
2	KCLASS_01_1
3	KCLASS_02..
4	KCLASS_02_1
5	KCLASS_03..
6	KCLASS_03_1
7	KCLASS_04..
8	KCLASS_04_1
9	KCLASS_05..
10	KCLASS_05_1
11	KCLASS_06..
12	KCLASS_06_1
13	KCLASS_07..
14	KCLASS_07_1
15	KCLASS_08..
16	KCLASS_08_1
17	KCLASS_09..
18	KCLASS_09_1
19	KCLASS_10..
20	KCLASS_10_1

Интегральный критерий сходимости "Сумма знаний"

Код	Название объекта	Соцсети	Ф. Соседство
145	08_01	99.99..	v
146	08_02	99.99..	v
147	08_03	99.99..	v
148	08_04	99.99..	v
149	08_05	99.99..	v
150	08_06	99.99..	v
151	08_07	99.99..	v
152	08_08	99.99..	v
153	08_09	99.99..	v
154	08_10	99.99..	v
155	08_11	99.99..	v
156	08_12	99.99..	v
14	01_14	30.21..	v
157	05_15a	29.72..	v
104	05_14a	6.516..	v
32	02_14a	6.716..	v
92	03_14a	4.111..	v
139	08_13a	0.370..	v

Показы Поиск объекта В начало БД В конец БД Предыдущая Следующая З записей Все записи Печать XLS Печать TXT Печать ALL

4.1.3.2. Визуализация результатов распознавания в отношении "Класс-объекты". Текущая модель: "INF3".

Код	Название класса
1	KCLASS_01..
2	KCLASS_01_1
3	KCLASS_02..
4	KCLASS_02_1
5	KCLASS_03..
6	KCLASS_03_1
7	KCLASS_04..
8	KCLASS_04_1
9	KCLASS_05..
10	KCLASS_05_1
11	KCLASS_06..
12	KCLASS_06_1
13	KCLASS_07..
14	KCLASS_07_1
15	KCLASS_08..
16	KCLASS_08_1
17	KCLASS_09..
18	KCLASS_09_1
19	KCLASS_10..
20	KCLASS_10_1

Интегральный критерий сходимости "Сумма знаний"

Код	Название объекта	Соцсети	Ф. Соседство
145	08_01	99.99..	v
146	08_02	99.99..	v
147	08_03	99.99..	v
148	08_04	99.99..	v
149	08_05	99.99..	v
150	08_06	99.99..	v
151	08_07	99.99..	v
152	08_08	99.99..	v
153	08_09	99.99..	v
154	08_10	99.99..	v
155	08_11	99.99..	v
156	08_12	99.99..	v
14	01_14	30.21..	v
157	05_15a	29.72..	v
104	05_14a	6.516..	v
32	02_14a	6.716..	v
92	03_14a	4.111..	v
139	08_13a	0.370..	v

Показы Поиск объекта В начало БД В конец БД Предыдущая Следующая З записей Все записи Печать XLS Печать TXT Печать ALL

Рисунок 24. Подробные результаты идентификации объектов обучающей выборки с каждым из классов, к которым они фактически относятся (исходной наиболее достоверной модели INF3 на 1-й итерации)²

² Приведенные экранные формы вполне читабельны при увеличении масштаба изображения

В модели 1-й итерации у нас представлены объекты с истинными описаниями признаков и принадлежности к классам (имена таких объектов оканчиваются буквой “и”) и артефакты со случайными признаками и ошибочно указанной принадлежностью к классам (имена таких объектов оканчиваются буквой “а”).

В исходной модели, у нас были еще и нетипичные объекты, с истинными описаниями признаков и ошибочным отнесением к классам. Теперь эти объекты правильно отнесены к новым специально для них созданным классам, имена которых образованы из имен классов, к которым они были ошибочно отнесены по данным обучающей выборки, и номера итерации. Таким образом, в результате первой итерации нетипичные объекты (имена таких объектов оканчиваются буквой “н”) переведены в категорию объектов с истинными описаниями.

Из экранных форм, приведенных на рисунке 24, видно, что:

– уровень сходства объектов с истинными описаниями, и тех, которые и ранее были истинными (и), **и тех, которые ранее были нетипичными (н), а теперь стали истинными**, по интегральному критерию «резонанс знаний» с классами близок к 100%, а по интегральному критерию «сумма знаний» составляет около 35%;

– уровень сходства артефактов (а) с классами, к которым они отнесены, остается очень низким и иногда может быть даже отрицательным, что порождает ложно отрицательные решения (пример чего приведен на рисунке 25).

4.1.3.2. Визуализация результатов распознавания в отношении: "Класс-объекты". Текущая модель: "INF3"				
Классы		Интегральный критерий сходства: "Семантический резонанс знаний"		
Код	Нам. класса	Код	Наименование объекта	Сходство
1	КЛАСС-01_	116	07_08i	-11,54...
2	КЛАСС-01_1	117	07_09i	-11,54...
3	КЛАСС-02_	118	07_10i	-11,54...
4	КЛАСС-02_1	119	07_11i	-11,54...
5	КЛАСС-03_	120	07_12i	-11,54...
6	КЛАСС-03_1	14	01_14a	-12,45... v
7	КЛАСС-04_	86	05_14a	-15,95...
8	КЛАСС-04_1	50	03_14a	-16,65...
9	КЛАСС-05_	122	07_14a	-17,42...
10	КЛАСС-05_1	158	09_14a	-19,22...
11	КЛАСС-06_			
12	КЛАСС-06_1			
13	КЛАСС-07_			
14	КЛАСС-07_1			
15	КЛАСС-08_			
16	КЛАСС-08_1			
17	КЛАСС-09_			
18	КЛАСС-09_1			
19	КЛАСС-10_			
20	КЛАСС-10_1			

Интегральный критерий сходства: "Сумма знаний"				
Код	Наименование объекта	Сходство	Ф...	Сходство
108	06_18n	-1,223...		
177	10_15n	-1,223...		
178	10_16n	-1,223...		
179	10_17n	-1,223...		
180	10_18n	-1,223...		
14	01_14a	-2,077... v		
127	08_01i	-2,089...		
128	08_02i	-2,089...		
129	08_03i	-2,089...		
130	08_04i	-2,089...		

Помощь	Поиск объекта	В начало БД	В конец БД	Предыдущая	Следующая	9 записей	Все записи	Печать XLS	Печать TXT	Печать ALL
--------	---------------	-------------	------------	------------	-----------	-----------	------------	------------	------------	------------

Рисунок 25. Пример ложно-отрицательного решения, порожденного идентификацией артефакта

Иначе говоря, артефакты и на 1-й итерации остаются объектами *нетипичными для тех классов, к которым они отнесены, хотя ясно, что они являются объектами нетипичными для всех классов*. Но так как признаки артефактов являются случайными, то нельзя исключить и того, что случайно у артефактов могут оказаться и комбинации признаков, характерные для большинства объектов тех классов, к которым эти артефакты отнесены, т.е. случайно артефакты могут оказаться и похожими на обобщенный образ класса, которому отнесены, но это маловероятное совпадение.

На каждой итерации классы могут быть разделены только на две части, поэтому если разнообразие объектов с их признаками, отнесенных к одному классу, очень велико, то может потребоваться и несколько итераций для создания классов для каждой группы таких объектов, сходных друг с другом.

На 1-й итерации мы видим именно такую ситуацию: *к новым классам отнесены как нетипичные объекты, так и артефакты*, которые существенно отличаются друг от друга. Причем, если нетипичные объекты образуют группу сходных друг с другом объектов, то артефакты могут объединиться в группы только случайно и это маловероятно.

Ясно, что если в этих условиях продолжить процесс разделения классов на типичную и нетипичную части с параметрами 1-й итерации, т.е. если считать нетипичными объекты, имеющие сходство с классами, к которым они относятся по данным обучающей выборки, менее 50%, то и на следующей итерации все нетипичные объекты и артефакты опять окажутся в этой категории и для них будут созданы новые классы. В результате все ранее созданные классы просто поменяют названия и больше ничего не изменится, т.е. процесс «**зациклится**».

Из этой ситуации есть два разумных выхода:

1. Изменить критерий, позволяющий отличить нетипичные объекты от артефактов с 50%, например на 20% (при интегральном критерии «сумма знаний»).

2. Выйти из процесса итераций на процесс удаления артефактов.

Оба эти варианта являются вполне обоснованными, и выбор одного из них остается за пользователем. В последующих разделах данной работы рассмотрим (кратко) оба этих варианта.

Количество информации о принадлежности объектов к классам, которое содержится в их признаках, наглядно видно из когнитивных функций (рисунок 26) [10]:

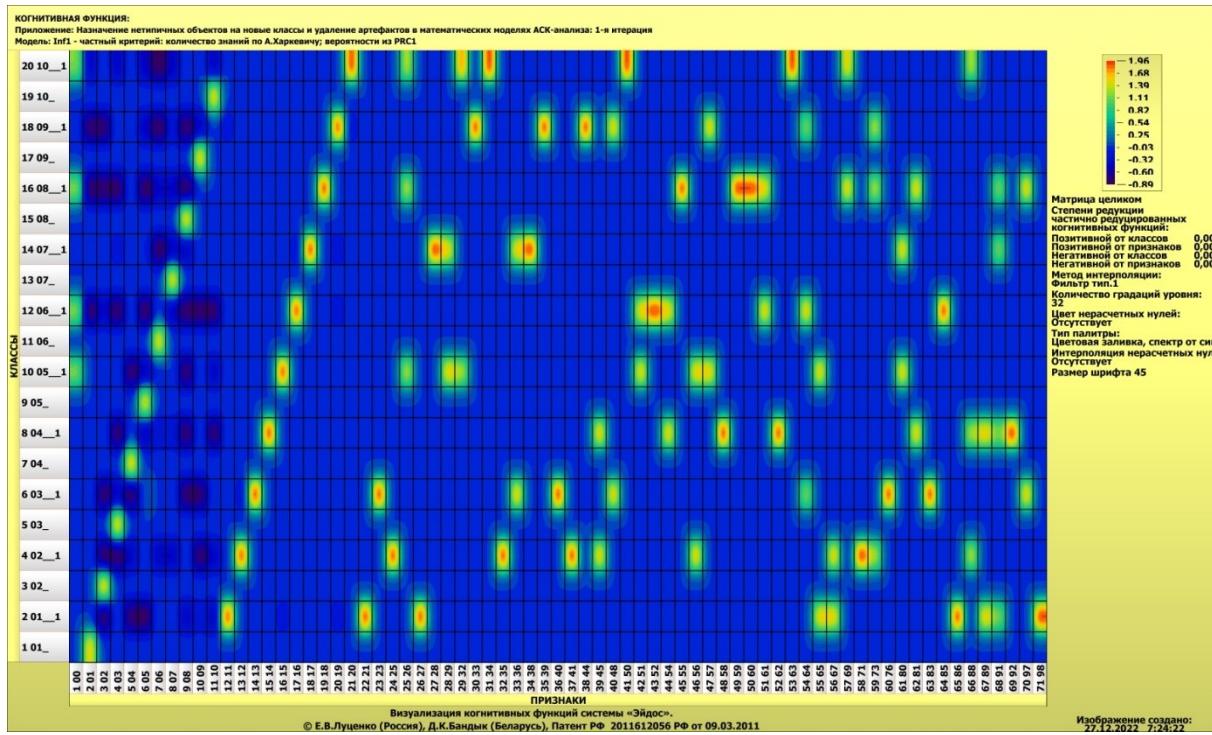


Рисунок 26. Когнитивная функция, построенная в модели INF1 на 1-й итерации

Из когнитивной функции, приведенной на рисунке 26 видно, что признаки, описывающие объекты, четко разделены по классам:

- признаки от 01 до 10 описывают объекты истинно описанных классов;
- признаки от 11 до 98 описывают нетипичные объекты с закономерной дезинформацией в признаках и артефакты со случайными признаками.

Таким образом, исходные классы в модели 1-й итерации уже разделены на классы с истинно-описанными объектами и классы с нетипичными объектами и артефактами.

4.7. Шаг 5. Начало 2-й итерации: выявление нетипичных объектов и их назначение на новые классы (разделение классов на типичную и нетипичную части)

Эти процедуры осуществляются в режиме 3.6 системы «Эйдос». Запустим режим 3.6 при параметрах, приведенных на рисунке 27.

В результате работы режима получим экранную форму, приведенную на рисунке 28.

Отметим, что 20 объектов, для которых созданы новые классы, включают 20 артефактов по 2 на каждый из 10 классов.

Затем в режиме 2.3.2.2 производится формализация предметной области и в режиме 3.5 создаются статистические и системно-когнитивные модели.

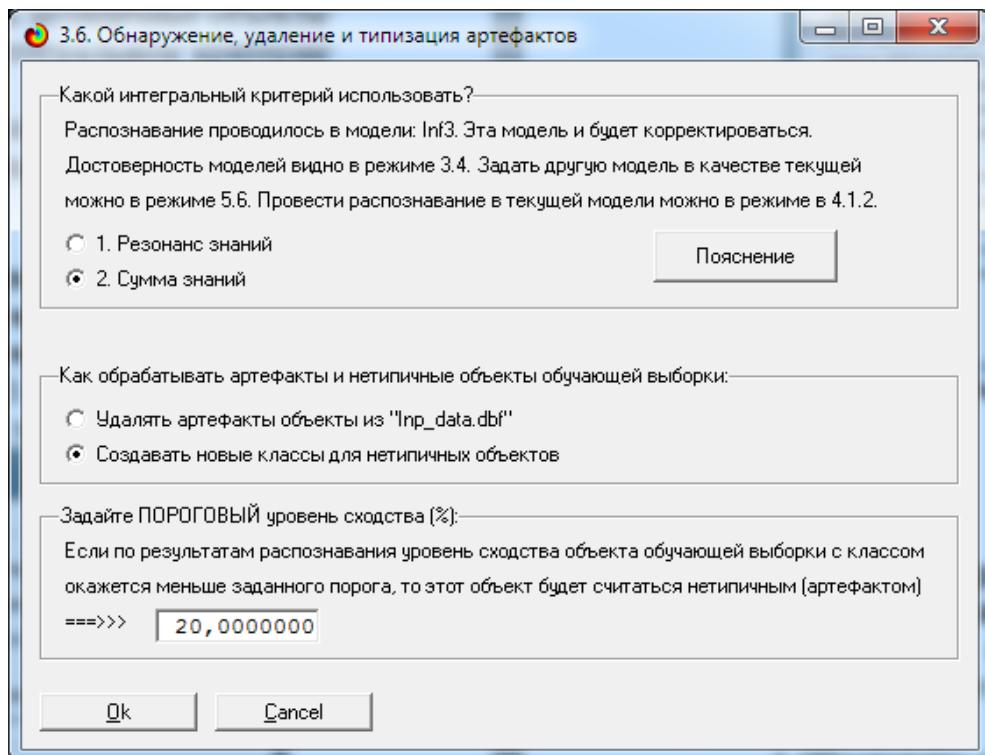


Рисунок 27. Экранная форма управления режимом 3.6 системы «Эйдос» для 2-й итерации

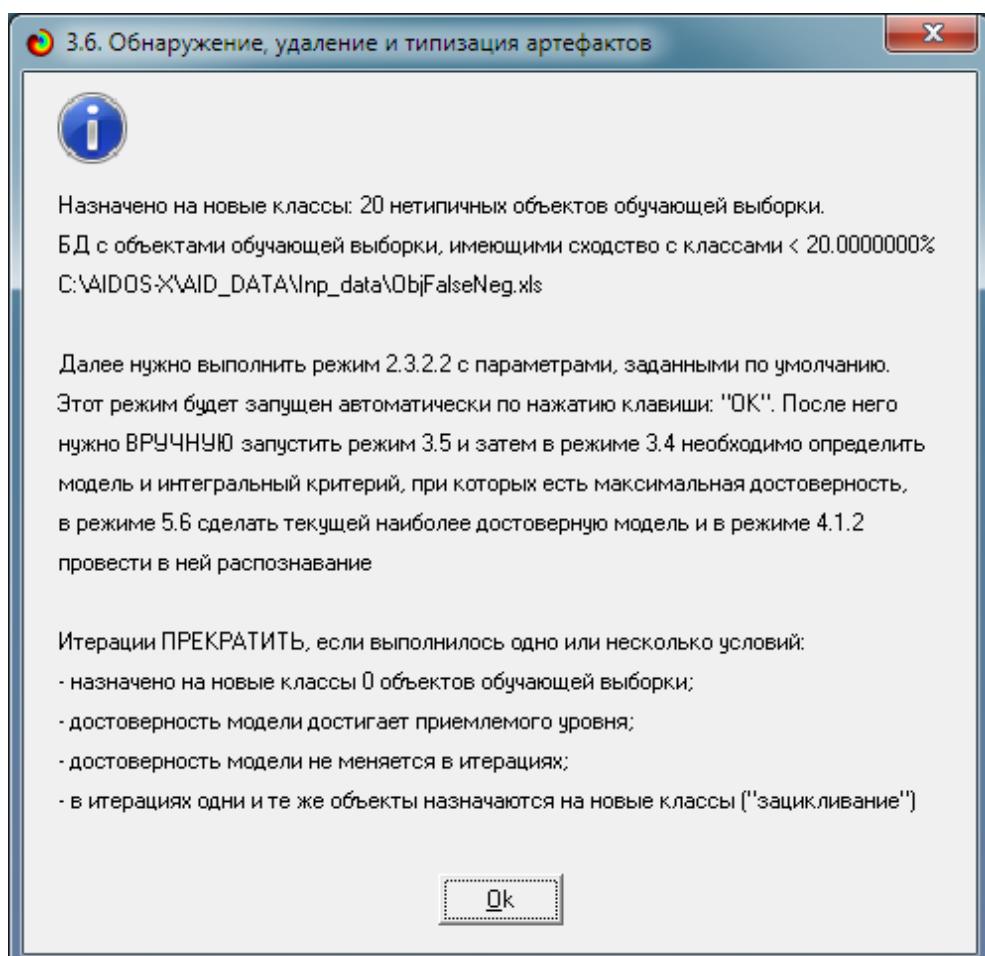


Рисунок 28. Экранная форма по результатам работы режима 3.6

В результате созданы модели чрезвычайно высокого уровня достоверности, особенно по L1 и L2 критериям проф. Е.В.Луценко [9] (рисунок 29):

The screenshot displays two tables of data from a software application. Both tables have identical columns and rows, representing different models and their performance metrics.

Название модели и частного критерия	Интегральный критерий	L1-нера prof. Е.В.Луценко	Сумма модуля уровней сходс... истинно-полож... решений (ST... р.)	Сумма модуля уровней сходс... истинно-отрицат... решений (SFP... р.)	S-Точность модели	S-Полнота модели	L1-нера prof. Е.В.Луценко	Средний модуль уровней сходс... истинно-полож... решений	Средний модуль уровней сходс... истинно-отрицат... решений	Средний модуль уровней сходс... ложно-положит... решений		
1. ABS - частный критерий количество встреч сочетаний: "классы"	Корреляция abs.частот с обр...	0,267	172,574	105,190	270,315		0,390	1,000	0,561	0,959	0,025	0,274
1. ABS - частный критерий количество встреч сочетаний: "классы"	Сумма abs.частот по признакам	0,257	133,605	22,805			0,854	1,000	0,921	0,742		0,022
2. PRC1 - частный критерий усл. вероятность Iго признака сред...	Корреляция усл.частот с обр...	0,267	172,574	105,190	270,315		0,390	1,000	0,561	0,959	0,025	0,274
2. PRC1 - частный критерий усл. вероятность Iго признака сред...	Сумма усл.частот по признакам	0,257	161,630	82,930			0,661	1,000	0,796	0,898		0,080
3. PRC2 - частный критерий условная вероятность Iго признака	Корреляция усл.частот с обр...	0,267	172,574	105,189	270,311		0,390	1,000	0,561	0,959	0,025	0,274
3. PRC2 - частный критерий условная вероятность Iго признака	Сумма усл.частот по признакам	0,257	161,630	82,930			0,661	1,000	0,796	0,898		0,080
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	0,988	167,928	137,213	30,276	0,397	0,847	0,998	0,918	0,971	0,028	0,129
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	0,366	130,177	37,271	67,305	0,025	0,659	1,000	0,795	0,727	0,008	0,109
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	0,988	167,928	137,213	30,276	0,397	0,847	0,998	0,918	0,971	0,028	0,129
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	0,366	130,177	37,271	67,305	0,025	0,659	1,000	0,795	0,727	0,008	0,109
6. INF3 - частный критерий: Харкевич, разности между фактами	Семантический резонанс зна...	0,995	168,872	444,797	97,415	0,021	0,634	1,000	0,776	0,943	0,095	0,178
6. INF3 - частный критерий: Харкевич, разности между фактами	Сумма знаний	0,395	133,749	143,890	10,142	0,000	0,930	1,000	0,961	0,747	0,031	0,019
7. INF4 - частный критерий: ROI (Return On Investment), вероятно...	Семантический резонанс зна...	0,778	167,116	130,155	24,587	0,711	0,872	0,996	0,838	0,983	0,025	0,283
7. INF4 - частный критерий: ROI (Return On Investment), вероятно...	Сумма знаний	0,361	130,177	2,192	13,392	0,001	0,836	1,000	0,913	0,382	0,000	0,021
8. INF5 - частный критерий: ROI (Return On Investment), вероятно...	Семантический резонанс зна...	0,778	167,116	130,155	24,587	0,711	0,872	0,996	0,930	0,983	0,025	0,283
8. INF5 - частный критерий: ROI (Return On Investment), вероятно...	Сумма знаний	0,361	68,314	2,192	13,392	0,001	0,836	1,000	0,913	0,382	0,000	0,021
9. INF6 - частный критерий: разн.усл.и безузл.вероятностей; вер...	Семантический резонанс зна...	0,388	170,081	163,526	118,112	0,138	0,590	0,999	0,742	0,956	0,035	0,211
9. INF6 - частный критерий: разн.усл.и безузл.вероятностей; вер...	Сумма знаний	0,362	154,642	8,871	30,729	0,003	0,834	1,000	0,910	0,864	0,002	0,049
10. INF7 - частный критерий: разн.усл.и безузл.вероятностей; вер...	Семантический резонанс зна...	0,388	170,081	163,526	118,112	0,138	0,590	0,999	0,742	0,956	0,035	0,211
10. INF7 - частный критерий: разн.усл.и безузл.вероятностей; вер...	Сумма знаний	0,362	154,642	8,871	30,729	0,003	0,834	1,000	0,910	0,864	0,002	0,049

Название модели и частного критерия	Интегральный критерий	L1-нера prof. Е.В.Луценко	Средний модуль уровней сходс... истинно-полож... решений	Средний модуль уровней сходс... истинно-отрицат... решений	Средний модуль уровней сходс... ложно-положит... решений	A-Точность модели АРесоф = АТР/АТР...	A-Полнота модели АРесоф = АТР/АТР...	L2-нера prof. Е.В.Луценко	Прогноз правильной идентификации...	Прогноз ошибкой идентификации...		
1. ABS - частный критерий количество встреч сочетаний: "классы"	Корреляция abs.частот с обр...	0,961	0,959	0,025	0,274	0,778	1,000	0,875	100,000	91,434	8,566	
1. ABS - частный критерий количество встреч сочетаний: "классы"	Сумма abs.частот по признакам	0,921	0,742		0,022	0,971	1,000	0,985	100,000	91,110	8,890	
2. PRC1 - частный критерий усл. вероятность Iго признака сред...	Корреляция усл.частот с обр...	0,981	0,959	0,025	0,274	0,778	1,000	0,875	100,000	91,434	8,566	
2. PRC1 - частный критерий усл. вероятность Iго признака сред...	Сумма усл.частот по признакам	0,798	0,998		0,080	0,918	1,000	0,957	100,000	91,110	8,690	
3. PRC2 - частный критерий условная вероятность Iго признака	Корреляция усл.частот с обр...	0,981	0,959	0,025	0,274	0,778	1,000	0,875	100,000	91,434	8,566	
3. PRC2 - частный критерий условная вероятность Iго признака	Сумма усл.частот по признакам	0,796	0,898		0,080	0,918	1,000	0,957	100,000	91,110	8,890	
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	0,916	0,971	0,028	0,129	0,057	0,883	0,945	0,913	96,111	96,135	3,865
4. INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	0,795	0,727	0,008	0,109	0,025	0,870	0,966	0,916	99,444	93,732	6,268
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	0,916	0,971	0,028	0,129	0,057	0,883	0,945	0,913	96,111	96,135	3,865
5. INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	0,796	0,727	0,008	0,109	0,025	0,870	0,966	0,916	99,444	93,732	6,268
6. INF3 - частный критерий: Харкевич, разности между фактами	Семантический резонанс зна...	0,797	0,943	0,095	0,178	0,021	0,841	0,978	0,908	99,444	94,181	5,819
6. INF3 - частный критерий: Харкевич, разности между фактами	Сумма знаний	0,963	0,747	0,031	0,019	0,000	0,976	0,999	0,987	99,444	94,181	5,819
7. INF4 - частный критерий: ROI (Return On Investment), вероятно...	Семантический резонанс зна...	0,930	0,983	0,025	0,283	0,071	0,777	0,933	0,848	94,444	97,059	2,941
7. INF4 - частный критерий: ROI (Return On Investment), вероятно...	Сумма знаний	0,931	0,382	0,000	0,021	0,001	0,947	0,998	0,972	99,444	93,657	6,343
8. INF5 - частный критерий: ROI (Return On Investment), вероятно...	Семантический резонанс зна...	0,930	0,983	0,025	0,283	0,071	0,777	0,933	0,848	94,444	97,059	2,941
8. INF5 - частный критерий: ROI (Return On Investment), вероятно...	Сумма знаний	0,931	0,382	0,000	0,021	0,001	0,947	0,998	0,972	99,444	93,657	6,343
9. INF6 - частный критерий: разн.усл.и безузл.вероятностей; вер...	Семантический резонанс зна...	0,742	0,956	0,035	0,211	0,069	0,819	0,933	0,872	98,889	94,106	5,894
9. INF6 - частный критерий: разн.усл.и безузл.вероятностей; вер...	Сумма знаний	0,910	0,864	0,002	0,049	0,003	0,946	0,997	0,971	99,444	93,675	6,325
10. INF7 - частный критерий: разн.усл.и безузл.вероятностей; вер...	Семантический резонанс зна...	0,742	0,956	0,035	0,211	0,069	0,819	0,933	0,872	98,889	94,106	5,894
10. INF7 - частный критерий: разн.усл.и безузл.вероятностей; вер...	Сумма знаний	0,910	0,864	0,002	0,049	0,003	0,946	0,997	0,971	99,444	93,675	6,325

Рисунок 29. Экранные формы режима 3.4 оценки достоверности моделей

Как и ожидалось, на 2-й итерации созданы отдельные классы для артефактов.

Таким образом, на 1-й итерации исходные классы разделены на классы для истинно-описанных объектов с одной стороны, и нетипичных объектов (с истинными описаниями признаков и дезинформацией о принадлежности к классам) и артефактов (со случайными признаками и случайной принадлежностью к классам) с другой стороны.

На 2-й итерации классы с нетипичными объектами и артефактами в свою очередь разделены на классы, содержащие только нетипичные объекты, и только артефакты. Причем если в классах для нетипичных объектов попало по 4 нетипичных объекта все с высоким уровнем сходства с классом, то в классы по артефактам попали по 2 объекта, причем один с довольно высоким уровнем сходства с классом, а

другой с довольно низким. Это значит, что при продолжении итераций будут созданы классы, содержащие по 1 артефакту.

Из когнитивной функции, приведенной на рисунке 30 и построенной в модели INF1 на 2-й итерации видно, что классы, полученные на 1-й итерации и содержащие нетипичные объекты и артефакты, разделены на 2-й итерации на классы, содержащие только нетипичные объекты и только артефакты:

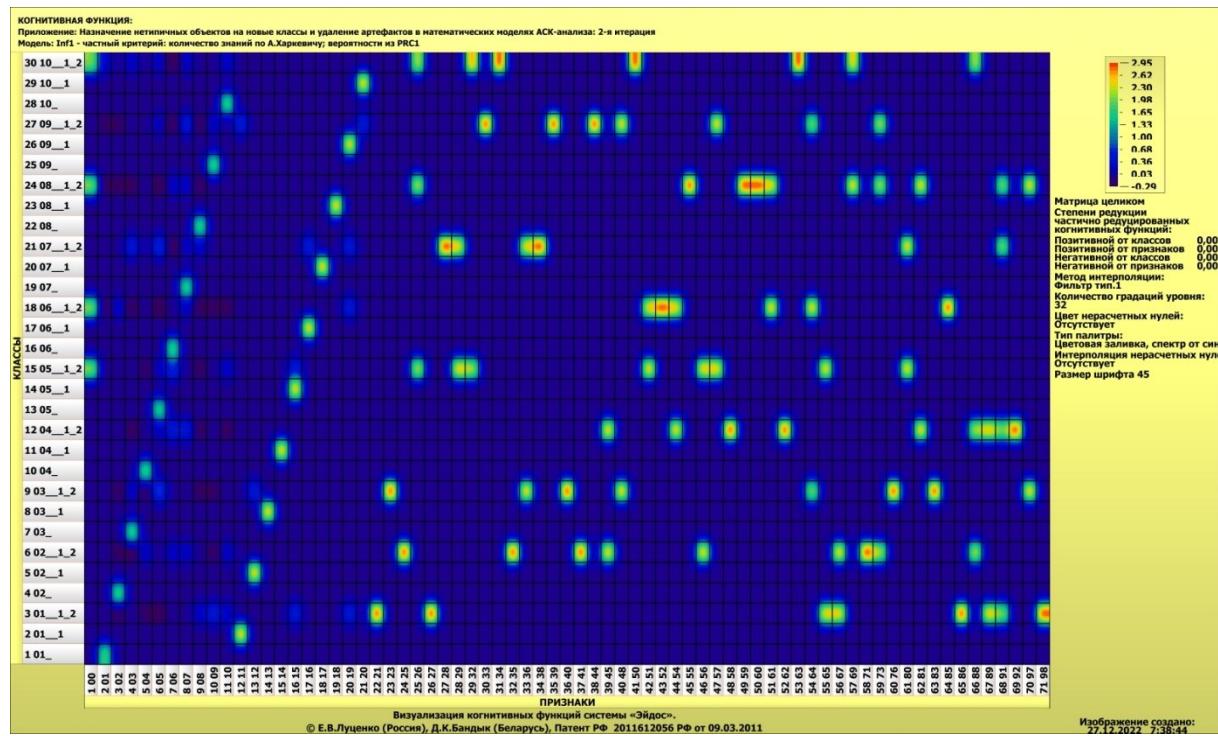


Рисунок 30. Когнитивная функция модели INF1 2-й итерации

4.8. Шаг 6. Проверка условий завершения процесса итераций

Поскольку, по сути, в конце 2-й итерации созданы классы, четко разделяющие все виды объектов обучающей выборки, то на этом процесс итераций разделения классов на типичную и нетипичную части целесообразно прекратить и перейти к удалению артефактов.

4.9. Шаг 7. Выявление и удаление артефактов

Для выявления и удаления артефактов запустим режим 3.6 с параметрами, приведенными на рисунке 31:

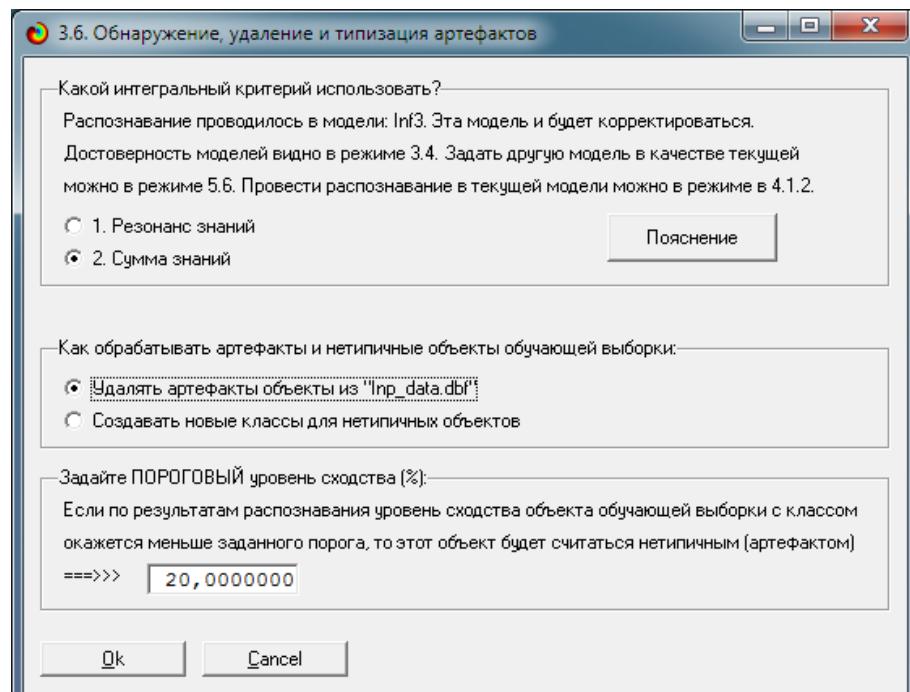


Рисунок 31. Экранная форма управления режимом 3.6 выявления и удаления артефактов

В результате выполнения данного режима получим отчет в экранной форме на рисунках 32 и 33:

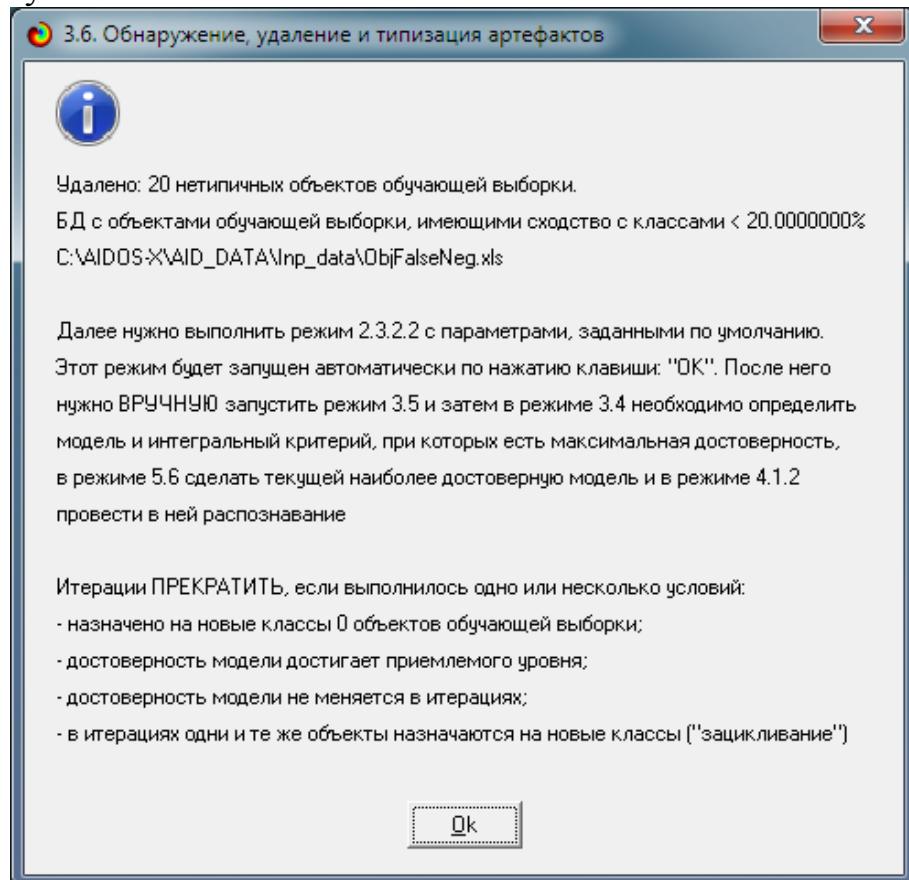


Рисунок 32. Экранная форма-отчет по результатам удаления артефактов

Файл: c:\Aidos-X\AID_DATA\Inp_data\ObjFalseNeg.xls

	A	B	C
1	N1	N2	N3
2	01_13a		98 27 91 65 89 15 21 19 67 86
3	01_14a		10 02 09 08 09 04 10 08 09 05
4	02_13a		25 41 73 56 45 71 67 04 88 35
5	02_14a		07 10 06 10 09 03 07 06 04 02
6	03_14a		05 03 03 05 02 08 09 05 05 04
7	03_13a		48 36 40 83 12 05 23 97 76 64
8	04_13a		54 89 45 92 62 81 58 07 88 91
9	04_14a		05 10 06 07 03 08 05 06 06 07
10	05_14a		00 06 06 10 06 05 08 06 10 04
11	05_13a		15 57 05 29 56 26 51 32 80 65
12	06_13a		64 51 64 00 85 54 19 61 85 52
13	06_14a		09 01 08 07 02 10 05 03 02 07
14	07_14a		05 07 03 07 05 03 05 08 05 08
15	07_13a		29 80 38 91 19 28 36 16 03 06
16	08_13a		55 69 61 81 97 59 26 91 73 60
17	08_14a		08 03 05 06 06 00 01 07 07 02
18	09_13a		48 44 73 33 57 11 64 20 39 08
19	09_14a		05 06 07 05 07 09 09 07 02 01
20	10_13a		26 03 63 10 69 32 88 34 32 50
21	10_14a		01 07 05 01 07 03 00 06 05 10

Разработка автораИсточник: c:\Aidos-X\AID_DATA\Inp_data\ObjFalseNeg.xls

Рисунок 33. Отчет по результатам удаления артефактов из исходных данных

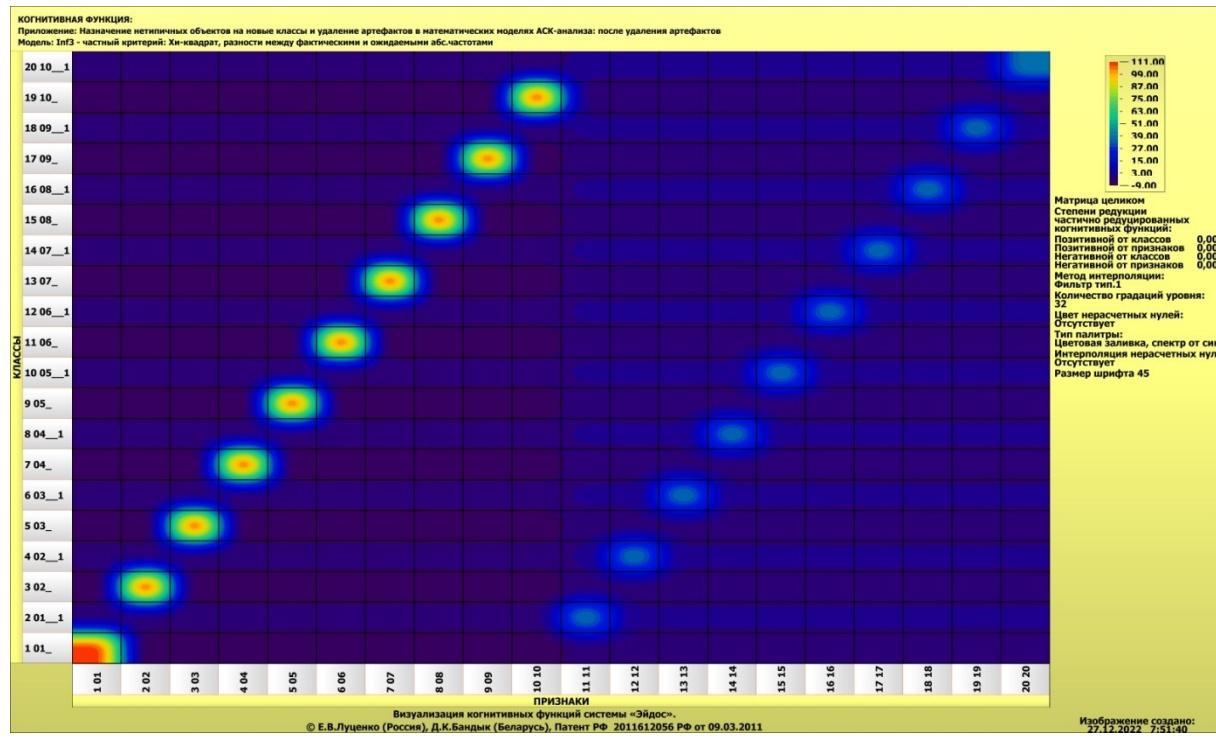
В результате удаления артефактов все модели стали **абсолютно безошибочными**, т.е. не дают ни одного ложного решения, ни ложно-положительного, ни ложно-отрицательного (рисунок 34).

3.4. Обобщ.форма по достоверн.моделей при разн.нкт.крит. Текущая модель: "INF1"

Название модели и частного критерия	Интегральный критерий	Всего логических объектов выборки	Число истинно-положительных решений (TP)	Число истинно-отрицательных решений (TN)	Число ложно-положительных решений (FP)	Число ложно-отрицательных решений (FN)	Точность модели	Полнота модели	Фамилия Ван Рибергена	Сумма наход. уровней схемы истинностного решения (ST...)	Сумма наход. уровней схемы истинностного решения (SF)	Сумма наход. уровней схемы ложноположительных решений
1_INF1 - частный критерий: количество патронов со стечением % из...	Корреляция abs.частот с обр...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
1_INF2 - частный критерий: количество патронов со стечением % из...	Сумма abs.частот по признак...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
2_PR1 - частный критерий: усл. вероятность итого признака сред...	Корреляция усл.отн частот с о...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
2_PR2 - частный критерий: усл. вероятность итого признака сред...	Сумма усл.отн частот по приз...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
3_PR2 - частный критерий: усл.вероятность итого признака	Корреляция усл.отн частот с о...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
3_PR2 - частный критерий: усл.вероятность итого признака	Сумма усл.отн частот по приз...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
4_INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
4_INF1 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	160	160	3040			1.000	1.000	1.000	124.262	124.262	
5_INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Семантический резонанс зна...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
5_INF2 - частный критерий: количество знаний по А.Харкевичу, в...	Сумма знаний	160	160	3040			1.000	1.000	1.000	124.262	124.262	
6_INF3 - частный критерий: Хи-квадрат, разности между фактами	Семантический резонанс зна...	160	160	3040			1.000	1.000	1.000	159.922	199.853	
6_INF3 - частный критерий: Хи-квадрат, разности между фактами	Сумма знаний	160	160	3040			1.000	1.000	1.000	134.054	134.054	
7_INF4 - частный критерий: ROI (Return On Investment), вероятно...	Семантический резонанс зна...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
7_INF4 - частный критерий: ROI (Return On Investment), вероятно...	Сумма знаний	160	160	3040			1.000	1.000	1.000	77.949	77.949	
8_INF5 - частный критерий: ROI (Return On Investment), вероятно...	Семантический резонанс зна...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
8_INF5 - частный критерий: ROI (Return On Investment), вероятно...	Сумма знаний	160	160	3040			1.000	1.000	1.000	77.949	77.949	
9_INF6 - частный критерий: разница и безразл.вероятностей, вер...	Семантический резонанс зна...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
9_INF6 - частный критерий: разница и безразл.вероятностей, вер...	Сумма знаний	160	160	3040			1.000	1.000	1.000	153.846	153.846	
10_INF7 - частный критерий: разница и безразл.вероятностей, вер...	Семантический резонанс зна...	160	160	3040			1.000	1.000	1.000	160.000	160.000	
10_INF7 - частный критерий: разница и безразл.вероятностей, вер...	Сумма знаний	160	160	3040			1.000	1.000	1.000	153.846	153.846	

Рисунок 34. Экранная форма по результату оценки достоверности моделей, созданных после 2-х итераций и удаления артефактов

Из когнитивной функции, приведенной на рисунке 36 и построенной в модели INF3 после 2-х итераций и удаления артефактов видно, что из исходных данных удалены все артефакты, а из модели удалены все классы, полученные на 2-й итерации и содержащие исключительно артефакты и все признаки, использовавшиеся для описания исключительно артефактов.



**Рисунок 35. Когнитивная функция, созданная в модели INF3
после 2-х итераций и удаления артефактов**

На основе полученных результатов можно сделать обоснованный вывод о том, что *все ложно-положительные и ложно-отрицательные решения были обусловлены только идентификацией артефактов, которые, таким образом, являются нетипичными для всех классов, за исключением классов, созданных на примере единственного артефакта.*

Артефактами являются объекты обучающей выборки, в которых и признаки, и принадлежность к классам являются совершенно случайными, т.е. шумом.

5. Обсуждение

Таким образом, в данной работе на модельных исходных данных продемонстрированы эффективные теоретические, математические, алгоритмические и программные подходы выявления дезинформации и шума в исходных данных, выявления истинной информации путем анализа дезинформации и разделения классов на типичную и нетипичную части, а также подавления шума путем выявления и удаления артефактов из исходных данных.

6. Выводы

В любой обучающей (тренировочной) выборке (исходных данных) всегда есть не только истинная информация о моделируемой предметной области, но и дезинформация, да и просто шум. Понятно, что шум и дезинформация в исходных данных приводят к понижению достоверности моделей, созданных на основе этих исходных данных. Поэтому необходимо иметь критерии, позволяющие отличить шум от дезинформации и от истинной информации, а также основанные на этих критериях математические и программные инструменты для выявления и подавления шума в исходных данных, а также для выявления в исходных данных дезинформации и восстановления истинной информации путем анализа дезинформации. Артефактами будем называть объекты обучающей выборки, у которых и/или признаки случайны, и/или классы случайны, и/или связь признаков этих объектов с принадлежностью этих объектов к классам также случайна. В этом и состоит шум в исходных данных. Нетипичными будем называть объекты обучающей выборки, у которых и признаки, и классы не случайны, и связь признаков (этих объектов) с принадлежностью (этих объектов) к классам, указанная в обучающей выборке, также не случайна, а вполне закономерна, но не та, которая указана в обучающей выборке, а другая, т.е. в обучающей выборке неверно указана принадлежность объектов к классам. В этом и заключается дезинформация в исходных данных. Нетипичные объекты в действительности могут относиться как к тем классам, которые указаны в обучающей выборке, так и к новым классам, которых там нет. Основным критерием, позволяющим отличить шум и дезинформацию от истинной информации является уровень сходства объекта с классами, к которым он относится по данным обучающей выборки. При увеличении уровня сходства закономерно расчет доля истинных решений среди всех решений. Низкий уровень сходства и, особенно, ложно-отрицательные решения при решении задачи идентификации, являются признаками шума и дезинформации в описании объектов обучающей выборки. Отличить шум и дезинформацию друг от друга можно по следующему критерию: если гипотеза о том, что объект нетипичный не подтверждается при итерационном назначении его на новые специально создаваемые для этого классы, т.е. этот процесс «зацикливается» (приводит к повторению ситуации) без повышения достоверности моделей, то описание этого объекта не содержит закономерностей и является шумом, т.е. он является артефактом. Такие объекты надо просто удалять из обучающей выборки. Но если вместе с артефактами удалить и нетипичные объекты, то это приведет к уменьшению количества информации в модели, ее обеднению (что называется: «вылить из ванны вместе с водой и ребенком»). Поэтому в автоматизированном системно-когнитивном анализе (АСК-анализ) и его программном инструментарии интеллектуальной системе «Эйдос»

реализованы оба механизма на основе двух приведенных критериев, позволяющие сначала разделять классы на типичную и нетипичную части с созданием новых классов для нетипичных объектов, т.е. восстанавливать истинную информацию путем анализа дезинформации, а затем эффективно подавлять шум в исходных данных. В работе приводятся подробные численные примеры, демонстрирующие эти подходы на модельных исходных данных.

Для тех, кто заинтересовался данной работой и хотел бы лично ознакомиться с ней в самой системе «Эйдос» такая возможность предоставляется, т.к. данной интеллектуальное облачное приложение размещено в Эйдос-облаке под номером **344**.

Все желающие могут совершенно **бесплатно** скачать систему на сайте ее автора и разработчика проф.Е.В.Луценко [1] на странице по ссылке: http://lc.kubagro.ru/aidos/_Aidos-X.htm, а затем в диспетчере приложений системы «Эйдос» (режим 1.3) скачать и установить данное приложение в его начальном состоянии.

Литература

1. Сайт проф.Е.В.Луценко: <http://lc.kubagro.ru/>
2. Блог проф.Е.В.Луценко: <https://www.researchgate.net/profile/Eugene-Lutsenko>
3. Наприев, И. Л. Образ-Я и стилевые особенности деятельности сотрудников органов внутренних дел в экстремальных условиях / И. Л. Наприев, Е. В. Луценко, А. Н. Чистилин. – Краснодар : Кубанский государственный аграрный университет имени И.Т. Трубилина, 2008. – 263 с. – ISBN 978-5-9266-0288-0. – EDN SGNDDV.
4. Луценко, Е. В. Повышение адекватности спектрального анализа личности по астросоциотипам путем их разделения на типичную и нетипичную части / Е. В. Луценко, А. П. Трунев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2008. – № 36. – С. 39-60. – EDN JWXXWIP.
5. Луценко, Е. В. Повышение качества моделей "knowledge management" путем разделения классов на типичную и нетипичную части / Е. В. Луценко, Е. А. Лебедев, В. Н. Лаптев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2009. – № 54. – С. 9-24. – EDN KYBNTH.
6. Луценко, Е. В. Прогнозирование рисков невозврата ссуды с применением интеллектуального итерационного алгоритма учета нетипичных случаев / Е. В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2021. – № 170. – С. 141-202. – DOI 10.21515/1990-4665-170-009. – EDN HPDDEU.
7. Луценко Е.В., Лебедев Е.А., Подсистема автоматического формирования двоичного дерева классов семантической информационной модели (Подсистема "Эйдос-Tree"). Пат. № 2008610096 РФ. Заяв. № 2007613721 РФ. Опубл. от 09.01.2008. – Режим доступа: <http://lc.kubagro.ru/aidos/2008610096.jpg>, 3,125 / 2,500 у.п.л.
8. Lutsenko E.V. Mathematical relationship of Karl Pearson's χ -square measure with the return on investment (ROI) coefficient and with the semantic measure of expediency of Alexander Harkevich's information // October 2022, DOI: [10.13140/RG.2.2.31687.78240](https://doi.org/10.13140/RG.2.2.31687.78240), License [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/), <https://www.researchgate.net/publication/364949578>

9. 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. – 2017. – № 126. – С. 1-32. – DOI 10.21515/1990-4665-126-001. – EDN XXXBDV.

10. Работы проф.Е.В.Луценко & C° по когнитивным функциям:
http://lc.kubagro.ru/aidos/Works_on_cognitive_functions.htm

Literatura

1. Sajt prof.E.V.Lucenko: <http://lc.kubagro.ru/>
2. Blog prof.E.V.Lucenko: <https://www.researchgate.net/profile/Eugene-Lutsenko>
3. Napriev, I. L. Obraz-Ya i stilevye osobennosti deyatel'nosti sotrudnikov organov vnutrennih del v ekstremal'nyh usloviyah / I. L. Napriev, E. V. Lucenko, A. N. Chistilin. – Krasnodar : Kubanskij gosudarstvennyj agrarnyj universitet imeni I.T. Trubilina, 2008. – 263 s. – ISBN 978-5-9266-0288-0. – EDN SGNDDV.
4. Lucenko, E. V. Povyshenie adekvatnosti spektral'nogo analiza lichnosti po astrosociotipam putem ih razdeleniya na tipichnuyu i netipichnuyu chasti / E. V. Lucenko, A. P. Truney // Politematiceskij setevoj elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. – 2008. – № 36. – S. 39-60. – EDN JWYWIP.
5. Lucenko, E. V. Povyshenie kachestva modelej \"knowledge management\" putem razdeleniya klassov na tipichnuyu i netipichnuyu chasti / E. V. Lucenko, E. A. Lebedev, V. N. Laptev // Politematiceskij setevoj elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. – 2009. – № 54. – S. 9-24. – EDN KYBNTH.
6. Lucenko, E. V. Prognozirovaniye riskov nevozvrata ssudy s primeneniem intellektual'nogo iteracionnogo algoritma ucheta netipichnyh sluchaev / E. V. Lucenko // Politematiceskij setevoj elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. – 2021. – № 170. – S. 141-202. – DOI 10.21515/1990-4665-170-009. – EDN HPDDEU.
7. Lucenko E.V., Lebedev E.A., Podistema avtomaticheskogo formirovaniya dvoichnogo dereva klassov semanticeskoy informacionnoj modeli (Podistema \"Ejdos-Tree\"). Pat. № 2008610096 RF. Zayav. № 2007613721 RF. Opubl. ot 09.01.2008. – Rezhim dostupa: <http://lc.kubagro.ru/aidos/2008610096.jpg>, 3,125 / 2,500 u.p.l.
8. Lutsenko E.V. Mathematical relationship of Karl Pearson's χ^2 -square measure with the return on investment (ROI) coefficient and with the semantic measure of expediency of Alexander Harkevich's information // October 2022, DOI: 10.13140/RG.2.2.31687.78240, License CC BY 4.0, <https://www.researchgate.net/publication/364949578>
9. Lucenko, E. V. Invariantnoe otnositel'no ob'emov dannyh nechetkoe mul'tiklassovoe obobshchenie f-mery dostovernosti modelej Van Rizbergena v ASK-analize i sisteme \"Ejdos\" / E. V. Lucenko // Politematiceskij setevoj elektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. – 2017. – № 126. – S. 1-32. – DOI 10.21515/1990-4665-126-001. – EDN XXXBDV.
10. Raboty prof.E.V.Lucenko & C° po kognitivnym funkciyam:
http://lc.kubagro.ru/aidos/Works_on_cognitive_functions.htm

Приложения:

Таблица 9 – Способ расчета исходных данных в MS Excel (полностью)

³ При увеличении масштаба просмотра таблица вполне читабельна

Разработка автора

Путь на источник: "c:\Aidos-X\AID DATA\Inp data\Inp data с расчетом-артефакты и нетипичные с форм 19 02 2022.xlsx"