УДК 303.732.4

**МЕТОД КОГНИТИВНОЙ КЛАСТЕРИЗАЦИИ ИЛИ КЛАСТЕРИЗАЦИЯ НА ОСНОВЕ ЗНАНИЙ (Кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос»)**

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор
*Кубанский государственный аграрный университет, Россия, 350044, Краснодар, Калинина, 13,* prof.lutsenko@gmail.com
Коржаков Валерий Евгеньевич
к. т. н., доцент
*Адыгейский государственный университет Адыгея, Россия,* korve@yandex.ru

В статье на небольшом численном примере рассматриваются новые математическая модель, алгоритм и результаты агломеративной кластеризации, основные отличия которых от ранее известных стоят в том, что: а) в них параметры обобщенного образа кластера не вычисляются как средние от исходных объектов (классов) или центры тяжести, а определяются с помощью той же самой базовой когнитивной операции АСК-анализа, которая применяется и для формирования обобщенных образов классов на основе примеров объектов и которая действительно обеспечивает обобщение; б) в качестве критерия сходства используется не евклидово расстояние или его варианты, а интегральный критерий неметрической природы: «суммарное количество информации», применение которого теоретически корректно и дает хорошие результаты в неортонормированных пространствах, которые обычно и встречаются на практике; в) кластерный анализ проводится не на основе исходных переменных или матрицы сопряженности, зависящих от единиц измерения по осям, а в когнитивном пространстве, в котором по всем осям (описательным шкалам) используется одна единица измерения: количество информации, и поэтому результаты кластеризации не зависят от исходных единиц измерения признаков объектов. Имеется и ряд других менее существенных отличий. Все это позволяет получить результаты кластеризации, понятные специалистам и поддающиеся содержательной интерпретации, хорошо согласующиеся с оценками экспертов, их опытом и интуитивными ожиданиями, что часто представляет собой проблему для классических методов кластеризации. Описанные методы теоретически обоснованы в системно-когнитивном анализе (СК-анализ) и реализованы в его программном инструментарии – интеллектуальной системе «Эйдос»

Ключевые слова: АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ, ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА «ЭЙДОС», КОГНИТИВНОЕ ПРОСТРАНСТВО, АГЛОМЕРАТИВНАЯ КЛАСТЕРИЗАЦИЯ

UDC 303.732.4

**METHOD OF COGNITIVE CLUSTERIZATION OR CLUSTERIZATION ON THE BASIS OF KNOWLEDGE (Clusterization in system-cognitive analysis and intellectual system "Eidos")**

Lutsenko Evgeny Veniaminovich
Dr.Sci.Econ., Cand.Tech.Sci., professor
*Kuban State Agrarian University, Krasnodar, Russia*

Korzhakov Valery Evgenievich
Cand.Tech.Sci., assistant professor
*Adygh State University, Adygheya, Russia*

In this article, on a small and evident numerical example, methodological aspects of a process engineering of detection of knowledge from the trial-and-error data explicitly are considered, representation of knowledge and its usage for problem solving of forecasting, decision making and data domain examination in system-cognitive analysis (SC-analysis) and its programmatic toolkit - intellectual "Eidos" system are shown

Keywords: COMPUTERIZED SYSTEM-COGNITIVE ANALYSIS, INTELLECTUAL SYSTEM "EIDOS", COGNITIVE SPACE, AGGLOMERATIVE CLUSTERIZATION

***«Мышление – это обобщение, абстрагирование, сравнение, и классификация»***
***Патанджали[1], II в. до н. э.***

***"Истинное знание – это знание причин"***
**Френсис Бэкон (1561–1626 гг.)**

**Кластерный анализ[2]** (англ. *Data clustering*) – это задача разбиения заданной выборки *объектов* (ситуаций) на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались. Кластерный анализ очень широко применяется как в науке, так и в различных направлениях практической деятельности. Значение кластерного анализа невозможно переоценить, оно широко известно[3] и нет необходимости его специально обосновывать.

Существует *большое количество* различных методов кластерного анализа, хорошо описанных в многочисленной специальной литературе [1, 3] и прекрасных обзорных статьях [2-5]. Поэтому в данной статье мы не ставим перед собой задачу дать еще одно подобное описание, а обратим основное внимание на **проблемы**, существующие в кластерном анализе и вариант их решения, предлагаемый в автоматизированном системно-когнитивном анализе (АСК-анализ). Эти проблемы, *в основном*, хорошо известны специалистам, и поэтому наш краткий обзор будет практически полностью основан на уже упомянутых работах [2-5]. Необходимо специально отметить, что специалисты небезуспешно работают над решением этих проблем, предлагая все новые и новые варианты, которые и являются различными вариантами кластерного анализа. Мы в данной статье также предложим еще один ранее не описанный в специальной литературе (т.е. новый, авторский) теоретически обоснованный и программно-реализованный вариант решения некоторых из этих проблем, а также проиллюстрируем его на простом численном примере.

Почему же разработано так много различных методов кластерного анализа, почему это было необходимо? Кажутся почти очевидными мысли о том, что различные методы кластерного анализа дают результаты *различного качества*, т.е. одни методы *в определенном смысле* «лучше», а другие «хуже», и это действительно так [6], и, следовательно, по-видимому, *должен* существовать только один-единственный метод кластеризации, *всегда* (т.е. на любых данных) дающий «правильные» результаты, тогда как все остальные методы являются «неправильными». Однако если задать аналогичный вопрос по поводу, например, автомобиля или одежды, то становится ясным, что нет просто наилучшего автомобиля, а есть лучшие по определенным критериям-требованиям или лучшие для определен-

[1] http://ru.wikipedia.org/wiki/Патанджали
[2] http://ru.wikipedia.org/wiki/Кластерный%20анализ
[3] http://yandex.ru/yandsearch?text=кластерный%20анализ

ных *целей*. При этом сами критерии также должны быть обоснованы и не просто могут быть различными, но и должны быть различными при различных целях, чтобы отражать цель и соответствовать ей. Так автомобиль, лучший для семейного отдыха не являются лучшим для гонок Формулы-1 или для представительских целей. Аналогично можно обоснованно утверждать, что одни методы кластерного анализа являются более подходящими для кластеризации данных определенной структуры, а другие – другой, т.е. не существует одного наилучшего во всех случаях *универсального метода кластеризации*, но существуют методы более универсальные и методы менее универсальные. Но все же многообразие разработанных методов кластерного анализа на наш взгляд указывает не только на это, но и на то, что *их можно рассматривать как различные более или менее успешные варианты решения или попытки решения тех или иных проблем, существующих в области кластерного анализа*.

Для структурирования дальнейшего изложения сформулируем требования к исходным данным в кластерном анализе и фундаментальные вопросы, которые решают разработчики различных методов кластерного анализа.

Считается[4], что кластерный анализ предъявляет следующие *требования к исходным данным*:

1. Показатели не должны коррелировать между собой.
2. Показатели должны быть безразмерными.
3. Распределение показателей должно быть близко к нормальному.
4. Показатели должны отвечать требованию «устойчивости», под которой понимается отсутствие влияния на их значения случайных факторов.
5. Выборка должна быть однородна, не содержать «выбросов».

Даже поверхностный анализ сформулированных требований к исходным данным сразу позволяет утверждать, что *на практике они в полной мере никогда не выполняются*, а приведение исходных данных к виду, удовлетворяющему этим требованиям, или очень сложно, т.е. представляет собой **проблему**, и не одну, или *даже теоретически невозможно* в полной мере. В любом случае пытаться это делать можно *различными способами*, хотя *чаще всего на практике этого не делается вообще* или потому, что необходимость этого плохо осознается исследователем, или чаще потому, что в его распоряжении нет соответствующих инструментов, реализующих необходимые методы[5]. Конечно, в последнем случае не приходится удивляться тому, что результаты кластерного анализа получаются мягко сказать «несколько странными», а если они соответствуют здравому смыслу и точке зрения экспертов, то можно сказать, что это получилось случайно или потому, что «просто повезло».

[4] http://ru.wikipedia.org/wiki/Кластерный%20анализ

[5] Справедливости ради отметим, что подобных инструментов вообще *мало* и они практически недоступны исследователям

Остановимся подробнее на анализе перечисленных требований к исходным данным, а также проблем, возникающих при попытке их выполнения и решения.

*Первое требование* связано с использованием в большинстве методов кластеризации *евклидова расстояния* или различных его вариантов в качестве меры близости объектов и кластеров. Другими словами это требование означает, что описательные шкалы, рассматриваемые как оси семантического пространства, должны быть *ортонормированны*, т.к. *в противном случае применение евклидова расстояния и большинства других метрик (таблица 1) (кроме расстояния Махалонобиса) теоретически необоснованно и некорректно*.

**Таблица 1 – ОСНОВНЫЕ ТИПЫ МЕТРИК ПРИ КЛАСТЕР-АНАЛИЗЕ[6]**

| № | Наименование метрики | Тип признаков | Формула для оценки меры близости (метрики) |
|---|---|---|---|
| 1 | Эвклидово расстояние | Количественные | $d_{ik} = \left( \sum_{j=1}^{N} (x_{ij} - x_{kj})^2 \right)^{\frac{1}{2}}$ |
| 2 | Мера сходства Хэмминга | Номинальные (качественные) | $\mu_{ij}^{H} = \frac{n_{ik}}{N}$ где $n_{ik}$ число совпадающих признаков у образцов — $X_i$ и $X_k$ |
| 3 | Мера сходства Роджерса–Танимото | Номинальные шкалы | $\mu_{ij}^{R-T} = n''_{ik}(n'_i + n'_k - n''_{ik})$ где $n''_{ik}$ число совпадающих единичных признаков у образцов — $X_i$ и $X_k$; $n'_i$, $n'_k$ общее число — единичных признаков у образцов $X_i$ и $X_k$ соответственно; |
| 4 | Манхэттенская метрика | Количественные | $d_{ik}^{(1)} = \sum_{j=1}^{N} \lvert x_{ij} - x_{kj} \rvert$ |
| 5 | Расстояние Махаланобиса | Количественные | $d_{ik}^{M} = (x_{ij} - x_{kj})^T W^{-1} (x_{ij} - x_{kj})$, где **W** ковариационная матрица выборки; — $\mathbb{X} = (X_1, X_2, \ldots, X_n)$; |
| 6 | Расстояние Журавлева | Смешанные | $d_{ik} = \sum_{j=1}^{N} I_{ik}^{j}$, $I_{ik}^{j} = \begin{cases} 1, \text{если } \lvert x_{ij} - x_{kj} \rvert < \varepsilon \\ 0, \text{иначе} \end{cases}$ где |

[6] Источник: проф. Зайченко Ю.П. http://www.masters.donntu.edu.ua/2005/kita/kapustina/library/cluster.htm

Существуют и другие метрики, в частности: квадрат евклидова расстояния, расстояние городских кварталов (манхэттенское расстояние), расстояние Чебышева, степенное расстояние, процент несогласия, метрики Рао, Хемминга, Роджерса-Танимото, Жаккара, Гауэра, Воронина, Миркина, Брея-Кертиса, Канберровская и многие другие [2, 4]. Когда *корреляции между переменными равны нулю*, расстояние Махаланобиса эквивалентно квадрату евклидового расстояния [2]. Это означает, что метрику Махаланобиса можно считать обобщением евклидовой метрики для неортонормированных пространств[7].

Но на практике это требование *никогда* в полной мере не выполняется, а для его выполнения необходимо выполнить операцию ортонормирования семантического пространства, при которой из модели тем или иным методом[8] (реализованным в программной системе, в которой проводится кластерный анализ) *исключаются* все шкалы, коррелирующие между собой.

Таким образом, первое требование к исходным данным порождает две проблемы:

*Проблема 1.1 выбора метрики, корректной для неортонормированных пространств.*

*Проблема 1.2 ортонормирования пространства.*

*Второе требование* (безразмерности показателей) вытекает из того, что *выбор единиц измерения по осям существенно влияет на результаты кластеризации.* Казалось бы, одного этого должно быть достаточно для того, чтобы не делать этого, т.к. выбор единиц измерения, по сути, произволен (определяется исследователем), вследствие чего и результаты кластеризации, вместо того чтобы объективно отражать структуру данных и описываемой ими объективной реальности, также становятся произвольными и зависящими не только от самой исследуемой реальности, но и от произвола исследователя (причем неизвестно от чего больше: от реальности или исследователя). По сути, *автоматизированная система кластеризации превращается в этих условиях из инструмента исследования структуры объективной реальности в автоматизированный инструмент рисования таких дендрограмм, какие больше нравятся пользователю*. Непонятно также, какой содержательный смысл могут иметь, например корни квадратные из сумм квадратов разностей координат объектов, классов или кластеров, *измеряемых в различных единицах измерения. Разве корректно складывать величины даже одного рода, измеряемые в различных единицах измерения, а тем более разного рода?* Даже если сложить величины одного рода, но измеренные в разных единицах измерения, например *расстояния* от школы до подъезда дома 1.2 (километра), и от подъезда дома

[7] http://matlab.exponenta.ru/fuzzylogic/book1/12_1_3.php http://d3lpirt.narod.ru/dm/dm.htm

[8] Например, для ортонормирования семантического пространства может быть применен метод главных компонент: http://ru.wikipedia.org/wiki/Метод%20главных%20компонент

до квартиры 25 (метров), то получится 26,2 *непонятно чего*. Если же сложить разнородные по смыслу величины, т.е. *величины различной природы*, такие, например, как квадрат разности веса студентов с квадратом разности их роста, возраста, успеваемости и т.д., а потом еще извлечь из этой суммы квадратный корень, то получится просто *бессмысленная величина*, которая в традиционном кластерном анализе почему-то называется «Евклидово расстояние». В школе на уроке физики в 8-м классе за подобные действия сразу бы поставили «Неуд»[9]. Однако, как это ни удивительно, то, что «не прошло бы» на уроке физики в средней школе является вполне устоявшейся практикой в статистике и ее научных применениях.

В подтверждение тому, что подобная практика действительно существует, авторы не могут удержаться от искушения и не привести пространную цитату из работы [4]: «Заметим, что *евклидово расстояние* (и его квадрат) вычисляется по исходным, а не по стандартизованным данным. *Это обычный способ его вычисления*, который имеет определенные преимущества (например, расстояние между двумя объектами не изменяется при введении в анализ нового объекта, который может оказаться выбросом). Тем не менее, *на расстояния могут сильно влиять различия между осями, по координатам которых вычисляются эти расстояния. К примеру, если одна из осей измерена в сантиметрах, а вы потом переведете ее в миллиметры (умножая значения на 10), то окончательное евклидово расстояние (или квадрат евклидова расстояния), вычисляемое по координатам, сильно изменится, и, как следствие, результаты кластерного анализа могут сильно отличаться от предыдущих*.» (выделено нами, авт.)[10]. В работе [4] просто констатируется факт этой ситуации, но ему не дается никакой *оценки*. Наша же оценка этой практике по перечисленным выше причинам *отрицательная*. Приведем еще цитату из той же работы [4]: «**Степенное расстояние.** Иногда желают (!!!?)[11] прогрессивно увеличить или уменьшить вес, относящийся к размерности, для которой соответствующие объекты сильно отличаются. Это может быть достигнуто с использованием *степенного расстояния*. Степенное расстояние вычисляется по формуле:

$$\text{расстояние}(x,y) = \left(\sum\nolimits_i |x_i - y_i|^p\right)^{1/r}$$

где *r* и *p* - параметры, определяемые пользователем. Несколько примеров вычислений могут показать, как "работает" эта мера. Параметр *p* ответственен за постепенное взвешивание разностей по отдельным координатам, параметр *r* ответственен за прогрессивное взвешивание больших расстояний между объектами. Если оба параметра - *r* и *p*, равны двум, то это расстояние совпадает с расстоянием Евклида». Мы считаем, что еще какие-то комментарии здесь излишни.

Таким образом, второе требование к исходным данным порождает следующую проблему 2.1:

*<u>Проблема 2.1</u> сопоставимой обработки описаний объектов, описанных признаками различной природы, измеряемыми в различных единицах измерения (проблема размерностей).*

---

[9] Конечно, есть случаи, когда производят определенные математические операции над величинами различной природы, измеряемыми в различных единицах измерения, и это вполне корректно, правда это не операция сложения. Например, в физике так производятся вычисления *по формулам*. Но эти формулы теоретически обоснованы в соответствующих физических теориях. Если математические операции производятся так, что это не соответствует обоснованным формулам, то в результате получаются бессмысленные величины неизвестных науке размерностей. В этом случае говорят о проверке размерностей и нарушении размерностей. Такое впечатление, что в статистике подобные нарушения размерностей просто стали нормой.

[10] Пространные цитаты здесь и далее для удобства читателей приведены мелким шрифтом.

[11] Пометка (!!!?) наша, авт.

Отметим также, что объекты чаще всего описаны не только признаками, измеряемыми в различных единицах измерения, но как количественными, так и качественными признаками, которые соответственно являются градациями как числовых шкал, так и номинальных (текстовых) шкал. Существует метрика для номинальных шкал: это «Процент несогласия» [4], однако для количественных шкал применяются другие метрики. *Каким образом и с помощью какой комбинации классических метрик вычислять расстояния между объектами, описанными как количественными, так и качественными признаками, а также между кластерами, в которые они входят, вообще не понятно.* Это порождает проблему 2.2.:

*Проблема 2.2 формализации описаний объектов, имеющих как количественные, так и качественные признаки.*

*Третье требование* (нормальности распределения показателей) вытекает из того, что статистическое обоснование корректности вышеперечисленных метрик существенным образом основано на этом предположении, т.е. эти метрики являются параметрическими. На практике это означает, что перед применением кластерного анализа с этими метриками необходимо доказать гипотезу о нормальности исходных данных либо применить процедуру их нормализации. И первое, и второе, весьма **проблематично** и на практике не делается, более того, даже вопрос об этом чаще всего не ставится. Процедура нормализации (или взвешивания, ремонта) исходных данных обычно предполагает удаление из исходной выборки тех данных, которые нарушают их нормальность. Ясно, что это непредсказуемым образом может повлиять на результаты кластеризации, которые, скорее всего, существенно изменяться и их уже нельзя будет признать результатами кластеризации исходных данных. Отметим, что на практике исходные данные, не подчиняющиеся нормальному распределению, встречаются достаточно часто, что и делает актуальными методы непараметрической статистики.

Таким образом, 3-е требование к исходным данным порождает проблемы 3.1., 3.2. и 3.3.:

*Проблема 3.1 доказательства гипотезы о нормальности исходных данных.*

*Проблема 3.2 нормализации исходных данных.*

*Проблема 3.3 применения непараметрических методов кластеризации, корректно работающих с ненормализованными данными.*

Что можно сказать о четвертом и пятом требованиях?[12] Эти требования взаимосвязаны, т.к. случайные факторы и порождают «выбросы». На практике, строго говоря, эти требования никогда не выполняются и вообще звучат *несколько наивно*, если учесть, что как случайные часто рассматри-

[12] 4. Показатели должны отвечать требованию «устойчивости», под которой понимается отсутствие влияния на их значения случайных факторов. 5. Выборка должна быть однородна, не содержать «выбросов».

ваются неизвестные факторы, а их влияние даже теоретически, т.е. в принципе, исключить невозможно. С другой стороны эти требования «удобны» тем, что неудачные, неадекватные или не интерпретируемые результаты кластеризации, полученные тем или иным методом кластерного анализа, всегда можно «списать» на эти неизвестные «случайные» факторы или скрытые параметры и порожденные ими выбросы. А поскольку ответственность за обеспечение отсутствия шума и выбросов в исходных данных возложена *этими требованиями* на самого исследователя, то получается, что если что-то получилось не так, то это связано уж не столько с методом кластеризации, сколько с каким-то недоработками самого исследователя. По этим причинам более логично и главное, более *продуктивно* было бы предъявить эти требования не к исходным данным и обеспечивающему их исследователю, а к самому методу кластерного анализа, *который, по мнению авторов, должен корректно работать в случае наличия шума и выбросов в исходных данных*.

Таким образом, четвертое и пятое требования приводят к двум проблемам:

*Проблема 4 разработки такого метода кластерного анализа, математическая модель и алгоритм и которого органично включали бы фильтр, подавляющий шум в исходных данных, в результате чего данный метод кластеризации корректно работал бы при наличии шума в исходных данных.*

*Проблема 5 разработки метода кластерного анализа, математическая модель и алгоритм и которого обеспечивали бы выявление «выбросов» (артефактов) в исходных данных и позволяли либо вообще не показывать их в дендрограммах, либо показывать, но так, чтобы было наглядно видно, что это артефакты.*

**Далее рассмотрим, как решаются (или не решаются) сформулированные выше проблемы в классических методах кластерного анализа.** Для удобства дальнейшего изложения повторим формулировки этих проблем.

*Проблема 1.1 выбора метрики, корректной для неортонормированных пространств.*

*Проблема 1.2 ортонормирования пространства.*

*Проблема 2.1 сопоставимой обработки описаний объектов, описанных признаками различной природы, измеряемыми в различных единицах измерения (проблема размерностей).*

*Проблема 2.2 формализации описаний объектов, имеющих как количественные, так и качественные признаки.*

*Проблема 3.1 доказательства гипотезы о нормальности исходных данных.*

*Проблема 3.2 нормализации исходных данных.*

*Проблема 3.3 применения непараметрических методов кластеризации, корректно работающих с ненормализованными данными.*

*Проблема 4 разработки такого метода кластерного анализа, математическая модель и алгоритм и которого органично включали бы фильтр, подавляющий шум в исходных данных, в результате чего данный метод кластеризации корректно работал бы при наличии шума в исходных данных.*

*Проблема 5 разработки метода кластерного анализа, математическая модель и алгоритм и которого обеспечивали бы выявление «выбросов» (артефактов) в исходных данных и позволяли либо вообще не показывать их в дендрограммах, либо показывать, но так, чтобы было наглядно видно, что это артефакты.*

Сделать это *удобнее всего*, рассматривая какие ответы предлагают классические методы кластерного анализа на сформулированные в в работе [2] вопросы:

– как вычислять координаты кластера из двух более объектов;

– как вычислять расстояние до таких "полиобъектных" кластеров от "монокластеров" и между "полиобъектными" кластерами.

Дело в том, что эти вопросы имеют фундаментальное значение для кластерного анализа, т.к. разнообразные комбинации используемых метрик и методов вычисления координат и взаимных расстояний кластеров и порождают все многообразие методов кластерного анализа [2]. Мы бы несколько переформулировали эти вопросы, а также добавили бы еще один:

1. Каким методом вычислять *координаты* кластера, состоящего из одного и более объектов, т.е. каким образом *объединять объекты* в кластеры.

2. Каким методом *сравнивать* кластеры, т.е. как вычислять *расстояния* между кластерами, состоящими из различного количества объектов (одного и более).

3. Каким методом *объединять кластеры*, т.е. формировать *обобщенные* («многообъектные») кластеры.

*Вопрос 1-й.* Чаше всего ни в теории и математических моделях кластерного анализа, ни на практике между кластером, состоящим из одного объекта («монооообъектным» кластером) и самим объектом не делается *никакого различия*, т.е. считается, что это одно и тоже. «В агломеративно-иерархических методах (aggomerative hierarhical algorithms) … первоначально все объекты (наблюдения) рассматриваются как отдельные, самостоятельные кластеры состоящие всего лишь из одного элемента» [2]. В работе [4] также говорится, что древовидная «Диаграмма начинается с каждого объекта в классе (в левой части диаграммы)». Это решение сразу же *порождает* многие из вышеперечисленных проблем (1.1., 1.2., 2.1, 2.2), т.к. объекты могут быть описаны как количественными, так и качественными признаками различной природы, измеряемыми в различных едини-

цах измерения, причем эти признаки *взаимосвязаны* (коррелируют) между собой.

Казалось бы, *проблему размерностей* (2.1) решает кластеризация не исходных переменных, а матриц сопряженности, содержащих *абсолютные частоты* наблюдения признаков по объектам или классам. Однако при таком подходе, например при сравнении моделей автомобилей, *четыре и два* цилиндра у этих моделей, а также *четыре и два болта*, которыми у них прикручен номер, будут давать одинаковый вклад в сходство-различие этих моделей, что едва ли разумно и приемлемо [8]. Тем ни менее матрица сопряженности анализируется в социологических и социометрических исследованиях, а в статистических системах, в разделах справки, посвященных кластерному анализу, приводятся примеры подобного рода.

Другое предложение по решению проблемы размерностей (2.1) основано на четком пожимании того, что изменение единиц измерения переменной меняет среднее ее значений и их разброс от этого среднего. Например, переход от сантиметров к миллиметрам увеличивает среднее и среднее отклонение от среднего в 10 раз. Речь идет о методе нормализации или стандартизации исходных данных, когда значения переменных заменяются их стандартизированными значения или z-вкладами [15]. Z-вклад показывает, сколько стандартных отклонений отделяет данное наблюдение от среднего значения:

$$Z_i = \frac{x_i - \bar{x}}{s},$$

где $x_i$ – значение данного наблюдения, $\bar{x}$ – среднее, $s$ – стандартное отклонение. Однако этот метод имеет серьезный недостаток, описанный в литературе [2, 4, 15]. Дело в том, что нормализация значений переменных приводит к тому, что независимо от значений их среднего и вариабельности до нормализации (т.е. значимости, измеряемой стандартным отклонением), после нормализации среднее становится равным нулю, а стандартное отклонение 1. Это значит, что *нормализация выравнивает средние и отклонения по всем переменным, снижая, таким образом, вес значимых переменных, оказывающих большое влияние на объект, и завышая роль малозначимых переменных, оказывающих меньшее влияни*е и искажая, таким образом, картину. На взгляд авторов это на вряд ли приемлемо. Другой важный недостаток, который в отличие от первого не отмечается в специальной литературе, состоит в том, что стандартизированные значения сложно как-то содержательно интерпретировать, т.е. устранение влияния единиц измерения достигается ценой потери смысла переменных, который как раз и содержался в единицах их измерения. В результате нормализации все переменные становятся как бы «на одно лицо». Это также недопустимо. Таким образом, можно обоснованно сделать вывод о том, *нормализация и стандартизация исходных данных – это весьма радикальное*

*решение проблемы 2.1 «в лоб и в корне», но решение неприемлемо дорогой ценой.*

В классических методах кластерного анализа предлагается два основных варианта *ответов* на 1-й вопрос:

1. Вообще не формировать обобщенных классов или кластеров из объектов, а на всех этапах кластеризации рассматривать только сами первичные объекты.

2. Формировать обобщенные кластеры путем вычисления неких статистических характеристик кластера на основе характеристик входящих в него объектов.

О 1-м варианте ответа в работе [4] говорится: «Диаграмма начинается с каждого объекта в классе (в левой части диаграммы). Теперь представим себе, что постепенно (очень малыми шагами) вы "ослабляете" ваш критерий о том, какие объекты являются уникальными, а какие нет. Другими словами, вы понижаете порог, относящийся к решению об объединении двух или более объектов в один кластер. В результате, вы *связываете* вместе всё большее и большее число объектов и агрегируете (*объединяете*) все больше и больше кластеров, состоящих из все сильнее различающихся элементов». Этот подход, когда кластеры реально не формируются, т.к. им не соответствуют какие-либо конструкции математической модели, представляется авторам сомнительным, т.к., во-первых, как было показано выше, это порождает проблемы 1.1., 1.2., 2.1, 2.2, а во-вторых, никак не решает проблемы 3.1, 3.2, 3.3, 4 и 5. *Между тем сам способ формирования кластеров из объектов, по мнению авторов, призван стать средством решения всех этих проблем.*

2-й вариант ответа представляется более обоснованным, однако он сам в свою очередь порождает вопросы о степени корректности и научной обоснованности того или иного метода вычисления обобщенных характеристик кластера и главное о том, *в какой степени этот метод позволяет решить сформулированные выше проблемы*. Описание кластера на основе входящих в него объектов традиционно включает *центр кластера*, в качестве которого обычно используется *среднее* или *центр тяжести* от характеристик входящих в него объектов [2], а также какую-либо количественную оценку степени рассеяния объектов кластера от его центра (как правило, это дисперсия). Ответ на 2-й вопрос является продолжением ответа на 1-й вопрос.

*Вопрос 2-й*. В работах [2, 3, 4] и других по кластерному анализу описывается большое количество различных мер и методов, которые можно применить как для измерения расстояний между кластерами, так и расстояний от объекта до кластеров. Например, в *невзвешенном центроидном методе* при определении расстояния от объекта до кластера, по сути, определяется расстояние до его центра [4]. В методе *невзвешенного попарного среднего* расстояние между двумя кластерами вычисляется как среднее расстояние между всеми парами объектов в них [4]. При этом, как правило, не решаются перечисленные выше проблемы, т.к. *не устраняются их причины*: а именно средние вычисляются на основе мер расстояния, коррект-

ных только для ортонормированных пространств и при этом часто используются размерные или нормализованные формы представления признаков объектов, не формализуется описание объектов, обладающих как количественными, так и качественными признаками. Ответ на 3-й вопрос является продолжением ответа на 2-й вопрос.

*Вопрос 3-й.* При объединении кластеров характеристики вновь образованного обобщенного кластера обычно пересчитываются тем же методом, каким они рассчитывались для исходных кластеров. Это сохраняет нерешенными и все проблемы, которые были при определении характеристик исходных кластеров и расстояний между этими кластерами.

**Далее рассмотрим вариант решения некоторых из сформулированных выше проблем кластерного анализа, предлагаемый в АСК-анализе и реализованный в интеллектуальной системе «Эйдос».**

Обратимся к эпиграфам к данной статье: «Мышление – это обобщение, абстрагирование, сравнение, и классификация» (Патанджали, II в. до н. э.), «Истинное знание – это знание причин» (Френсис Бэкон, 1561–1626 гг.). Итак, мышление, как процесс это [в том числе] классификация, результатом же мышления является знание, причем истинное знание есть знание причин. Истинное мышление есть мышление, дающее истинное знание. Соответственно ложное мышление – это мышление, приводящее к заблуждениям. Поэтому *истинное мышление – это [в том числе] истинная (правильная, адекватная) классификация объектов по причинам их поведения, т.е. по системе их детерминации*. Правильной классификацией будем считать ту, которая совпадает с классификацией экспертов, основанной на их высоком уровне компетенции, профессиональной интуиции и большом практическом опыте.

Если, как это принято в АСК-анализе [14], факторы формализовать в виде шкал различного типа (номинальных, порядковых и числовых), признаки рассматривать как значения факторов, т.е. их интервальные значения, более или менее жестко детерминирующих поведение объекта, а классы как будущие состояния, в которые объект переходит под влиянием различных значений этих факторов, то можно сказать, что *признаки формализуют причины переходов объекта в состояния, соответствующие классам или кластерам*. Если учесть, что классификация – это кластерный анализ, то можно сделать обоснованные выводы о том, что ***кластерный анализ это и есть мышление (но мышление не сводится только к кластерному анализу), а результаты кластерного анализа представляют собой знания.*** *Степень истинности этих знаний*, полученных в результате кластерного анализа, т.е. их адекватность или соответствие действительности, *полностью определяются степенью истинности метода кластерного анализа, с помощью которого они получены*. Поэтому столь важно решить сформулированные выше проблемы кластерного анализа.

В свою очередь *классификация (в т.ч. кластерный анализ) как процесс основана на обобщении и сравнении*. В монографии 2002 года [9] предлагается пирамида иерархической структуры процесса познания, входящая в базовую когнитивную концепцию (рисунок 1):

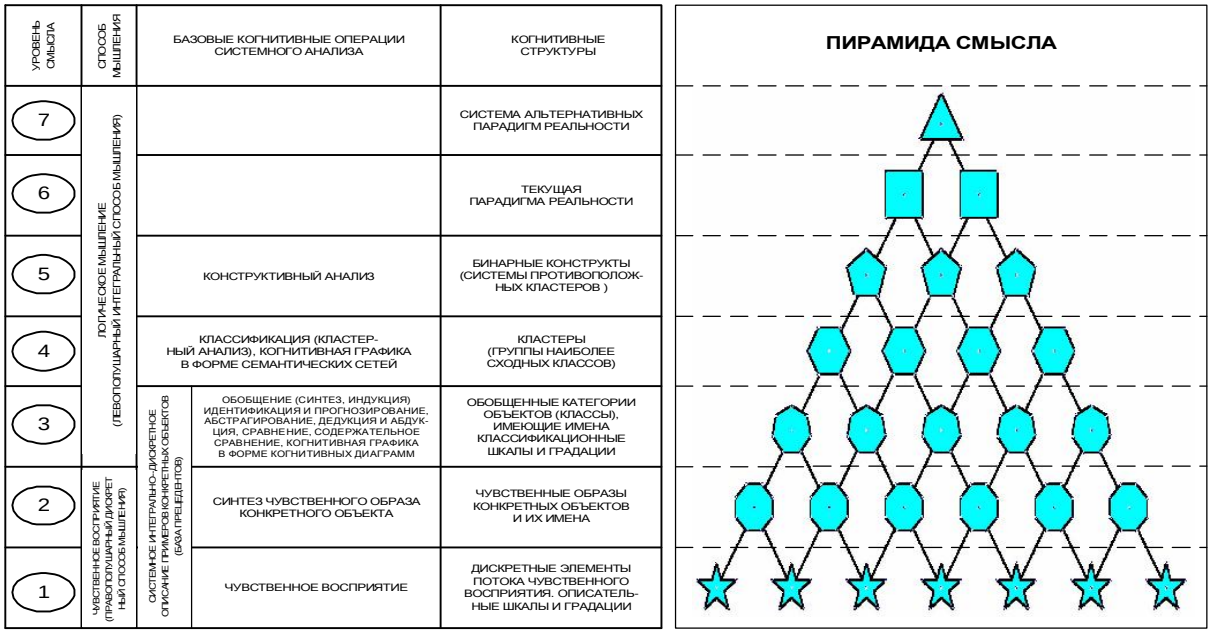



**Рисунок 1. Обобщенная схема иерархической структуры процесса познания согласно базовой формализуемой когнитивной концепции[13]**

В этой же монографии [9] предлагается математическая модель, основанная на семантической теории информации, обеспечивающая высокую степень формализацию данной когнитивной концепции, достаточную для разработки алгоритмов[14], структур данных и программной реализации в виде интеллектуальной программной системы. Такая система была создана автором и постоянно развивается, это система «Эйдос» [9, 12, 13].

***Суть* предлагаемых в АСК-анализе решений сформулированных выше проблем кластерного анализа состоит в следующем[15].**

Основная ***идея*** решения проблем кластерного анализа, состоит в том, что для решения задачи кластеризации предлагается использовать математическое представление объектов не виде переменных со значениями, измеряемыми в различных единицах измерения и в шкалах разного типа, и не матрицу сопряженности с абсолютными частотами встреч признаков по классам или нормализованными Z-вкладами, а *базы знаний*, рассчитанные на основе матрицы сопряженности (матрицы абсолютных частот) с использованием различных аналитических выражений для частных критериев. При этом для всех значений всех переменных используется *одна и та же размерность – это размерность* ***количества информации*** (бит, байт и

[13] http://lc.kubagro.ru/aidos/aidos02/2.3.htm
[14] http://lc.kubagro.ru/aidos/aidos02/4.2.htm http://lc.kubagro.ru/aidos/aidos02/4.3.htm
[15] Данные предложения приведены в том же порядке, что и переформулированные нами фундаментальные вопросы кластерного анализа согласно работе [2]

т.д.), что обеспечивает расчет на основе исходных данных силы и направления влияния на объект всех факторов и их значений и сопоставимую обработку значений переменных, изначальное (в исходных данных) представленных в разных единицах измерения и в шкалах разного типа (количественных – числовых, и качественных – текстовых).

1. Расстояния между объектом и кластером, а также между кластерами предлагается определять с использованием неметрических интегральных критериев, корректных для неортормированных пространств, *одним и тем же методом*: *по суммарному количеству информации*, которое содержится (соответственно) в системе признаков объекта о принадлежности к классу или кластеру, или которое содержится в обобщенных образах двух классов или кластеров об их принадлежности друг к другу.

2. *Координаты* кластера, возникающего как при включении в него *одного* единственного объекта, так и при *объединении многих объектов* в кластеры вычисляются *тем же самым методом*, что и координаты кластера, возникающего при объединении нескольких кластеров, а именно путем применения базовой когнитивной операции (БКОСА): «Обобщение», «Синтез», «Индукция» (БКОСА-3) АСК-анализа.

3. *Объединять* кластеры, т.е. формировать *обобщенные* («многообъектные») кластеры при объединении кластеров предлагается *тем же самым методом*, что и обобщенные образы классов при объединении конкретных образов объектов, т.е. путем применения базовой когнитивной операции (БКОСА): «Обобщение», «Синтез», «Индукция» (БКОСА-3) АСК-анализа.

Основная идея сводится к тому, чтобы кластеризовать не размерные переменные, абсолютные или относительные частоты или Z-вклады, а ***знания***. Предложения 1-3 являются непосредственными ответами на сформулированные выше фундаментальные вопросы кластерного анализа.

**Остановимся подробнее на математическом и алгоритмическом описании этих предложений и затем проиллюстрируем их на простом и наглядном численном примере.**

*Основная идея*. Вспомним приведенный выше пример кластеризации моделей автомобилей, в котором четыре или два цилиндра в двигателе давали такой вклад в сходство-различие моделей, как четыре или два болта, которыми прикручивается регистрационный номер. Из этого примера ясно, что при сравнении объектов и кластеров основную роль должно играть не само *количество* разных деталей или элементов конструкции, а, например, *их влияние на стоимость модели*, выраженное в долларах или на *степень ее пригодности (полезности) для поставленной цели*, тоже выраженное *в одних и тех же* для всех переменных и их значений *единицах измерения*. В АСК-анализе предлагается более радикальное решение: измерять степень и направление влияния всех переменных и их значений на поведение объекта или принадлежность его к тому или иному классу или класте-

ру *в одних и тех же универсальных единицах измерения, а именно единицах измерения количества информации. Ведь по сути, когда мы узнаем о том, что некий объект обладает определенным признаком, то мы получаем из этого факта некое количество информации о том, что принадлежит к определенной категории (классу, кластеру). А уж сами эти категории могут иметь совершенно различный смысл, в частности классифицировать текущие или будущие состояния объектов, или степень их полезности для достижения тех или иных целей. И что очень важно, при этом не играет абсолютно никакой роли в каких единицах измерения в какой шкале, количественной или качественной, изначально измерялся этот признак: килограммах, долларах, Омах, джоулях, или еще каких-то других.*

*Предложение 1-е*. В этом смысле в АСК-анализе исчезает существенное различие между классом и кластером и эти термины можно использовать как *синонимы*. Классы в АСК-анализе могут быть различаться степенью обобщенности: чем больше объектов в классе и чем выше вариабельность этих объектов по их признакам, тем шире представляемая ими генеральная совокупность, по отношению к которой они представляют собой репрезентативную выборку, тем выше степень обобщения в объединяющем их классе. Классы включают один или насколько объектов. Наименьшей степенью обобщения обладают классы, включающие лишь один объект, но и они совершенно не тождественны объекту исходной выборки, т.к. в математической мидели АСК-анализа у них совершенно различные математические формы представления. Кластеры обычно являются классами более высокой степени обобщения, т.к. включают один или несколько классов.

Как реализуется базовая когнитивная операция АСК-анализа «Обобщение», «Синтез», «Индукция» (БКОСА-3) будет рассмотрено ниже при кратком изложении математической модели АСК-анализа.

*Предложения 2-е* и *3-е* необходимо рассматривать в комплексе, т.к. их смысл в том, что объект при когнитивной кластеризации имеет другую математическую форму, чем объект в исходных данных, а именно такую же форму, как класс и как кластер, т.е. в АСК-анализе возможны классы и кластеры, включающие как один, так и много объектов. При этом для формирования класса состоящего из одного объекта, т.е. при добавлении в пустой кластер первого объекта, используется та же самая математическая процедура, что и при добавлении в него второго и вообще любого нового объекта (в АСК-анализе она называется БКОСА-3), и эта же самая процедура БКОСА-3 используется и при объединении классов или кластеров. При этом *само объединение классов (кластеров) осуществляется путем создания «с нуля» нового класса (кластера) из всех объектов, входящих в объединяемые классы (кластеры), а затем удаления исходных классов (кластеров). Новый объединенный класс (кластер) создается «с нуля» тем же самым методом (*БКОСА-3*), каким впервые создается любой новый*

*класс (*кластер). Теперь рассмотрим, как же это реализовано математически и алгоритмически.

**Математическая модель АСК-анализа.**

Математическая модель, которая стала основой модели АСК-анализа [16], была разработана автором в 1979 году [12], впервые опубликована в 1993 году [17], а затем и в последующих статьях и монографиях [9, 18, 19, 20], основной из которых является [9], а также в учебных пособиях [10, 11]. Поскольку эта модель описана во многих статьях и монографиях, в данной статье мы лишь кратко изложим ее суть.

В качестве формальной модели классов и признаков используются соответственно классификационные и описательные шкалы и градации.

Класс формализуется в виде градации классификационной шкалы. Если шкала числовая, то градации шкал представляют собой интервальные значения (числовые интервалы или диапазоны), если же признак качественный, то градация шкалы представляет собой просто уникальное текстовое наименование. Числовым интервалам также присваиваются текстовые наименования.

Признак формализуются в виде шкалы, а значения признака в виде градаций шкалы. Если признак количественный (числовой), то градации шкал представляют собой интервальные значения (числовые интервалы или диапазоны), если же признак качественный, то градация шкалы представляет собой просто уникальное текстовое наименование. Числовым интервалам также присваиваются текстовые наименования.

Математически и классификационные, и описательные шкалы представляются в форме векторов, а градации – в форме значений координат этих векторов, которые могут принимать значения n, где n={0, 1, 2, 3…}, т.е. 0 и натуральные числа.

Описание объекта исходной выборки формализуется в виде вектора, координаты которого имеют значение n, если соответствующий признак встречается n раз, в т.ч. 0, если признак *отсутствует* у объекта.

Например, признак: буква «м» присутствует в объекте: слово «молоко» 1 раз, поэтому значение соответствующего ему элемента вектора этого объекта будет равно 1, признак: буква «о» присутствует в объекте: слово «молоко» 3 раза, поэтому значение соответствующего ему элемента вектора этого объекта будет равно 3, а признак буква «ы» *отсутствует* у этого объекта, поэтому значение соответствующего этому признаку элемента вектора будет равно 0. При программной реализации классификационные и описательные шкалы и градации представляют собой справочники классов и признаков.

С использованием формального описания всех объектов исходной выборки рассчитывается *таблица сопряженности классов и признаков*, которая в АСК-анализе называется «матрица абсолютных частот» [21] (таблица 2).

**Таблица 2 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ**

| | | Классы | | | | | Сумма |
|---|---|---|---|---|---|---|---|
| | | 1 | ... | j | ... | W | |
| Значения факторов | 1 | $N_{11}$ | | $N_{1j}$ | | $N_{1W}$ | |
| | ... | | | | | | |
| | i | $N_{i1}$ | | $N_{ij}$ | | $N_{iW}$ | $N_i = \sum_{j=1}^{W} N_{ij}$ |
| | ... | | | | | | |
| | M | $N_{M1}$ | | $N_{Mj}$ | | $N_{MW}$ | |
| Суммарное количество признаков | | | | $N_j = \sum_{i=1}^{M} N_{ij}$ | | | $N = \sum_{i=1}^{W}\sum_{j=1}^{M} N_{ij}$ |
| Суммарное количество объектов обучающей выборки | | | | $N_j$ | | | $N$ |

***Алгоритм формирования матрицы абсолютных частот.***

Объекты обучающей выборки описываются векторами (массивами) $\vec{L} = \{L_i\}$ имеющихся у них признаков:

$$\vec{L} = \{L_i\} = \begin{cases} n, \text{ если у объекта } i\text{-й признак встречается } n \text{ раз}; \\ 0, \text{ если у объекта нет } i\text{-го признака}. \end{cases}$$

Первоначально в матрице абсолютных частот все значения равны нулю. Затем организуется цикл по объектам обучающей выборки. Если предъявленного объекта относящегося к *j-му* классу есть *i-й* признак, то:

$$N_{ij} = N_{ij} + 1;\ N_i = N_i + 1;\ N_j = N_j + 1;\ N = N + 1$$

Отметим, что уже при расчете матрицы абсолютных частот закладываются основы для *решения проблем 4 и 5*. Способ формирования матрицы абсолютных частот можно рассматривать как ***многоканальную*** *систему выделения полезного сигнала из шума*. Представим себе, что все объекты, предъявляемые для формирования обобщенного образа некоторого класса в действительности являются различными реализациями одного объекта – "Эйдоса" (в смысле Платона), ***по-разному*** *зашумленного различными случайными обстоятельствами* (по-разному, т.к. это шум). И наша задача состоит в том, чтобы подавить этот шум и выделить из него то общее и существенное, что отличает объекты данного класса от объектов других классов. Учитывая, что шум чаще всего является "белым" и имеет свойство при суммировании с самим собой стремиться к нулю, а сигнал при этом

наоборот возрастает пропорционально количеству ***слагаемых***, то увеличение объема обучающей выборки (в случае если сигнал эргодичный, т.е. закономерности в предметной области не меняются) приводит ко все лучшему отношению сигнал/шум в матрице абсолютных частот, т.е. к выделению полезной информации из шума. Примерно так мы начинаем постепенно понимать смысл фразы, которую мы сразу не расслышали по телефону и несколько раз переспрашивали. При этом в повторах шум не позволяет понять то одну, то другую часть фразы, но в конце-концов за счет использования памяти и интеллектуальной обработки информации мы понимаем ее всю. Так и *объекты, описанные признаками, можно рассматривать как зашумленные фразы, несущие нам информацию об обобщенных образах классов: "Эйдосах" [22], к которым они относятся.* ***И эту информацию мы выделяем из шума при синтезе модели.***

**Различные аналитические формы частных критериев в матрицах знаний и неметрических интегральных критериев при определении информационных расстояний.**

*Непосредственно на основе матрицы абсолютных частот* (таблиц 2) рассчитывается матрица знаний (таблица 4). При этом используются различные выражения для количества знаний (таблица 3), которое в последующем, при решении задач идентификации, прогнозирования, принятия решений и исследования предметной области используются как частные критерии в неметрических интегральных критериях,

**Таблица 3 – РАЗЛИЧНЫЕ АНАЛИТИЧЕСКИЕ ФОРМЫ ЧАСТНЫХ КРИТЕРИЕВ В МАТРИЦАХ ЗНАНИЙ И НЕМЕТРИЧЕСКИХ ИНТЕГРАЛЬНЫХ КРИТЕРИЯХ ПРИ ОПРЕДЕЛЕНИИ ИНФОРМАЦИОННЫХ РАССТОЯНИЙ**

| Наименование модели знаний и частный критерий | Выражение для частного критерия | |
|---|---|---|
| | через относительные частоты | через абсолютные частоты |
| СИМ-1, частный критерий: количество знаний по А.Харкевичу-Е.Луценко, 1-й вариант расчета вероятностей: *Nj* – суммарное количество признаков по *j*-му классу (предпоследняя строка таблицы 2) | $I_{ij} = \Psi \times Log_2 \frac{P_{ij}}{P_i}$ | $I_{ij} = \Psi \times Log_2 \frac{N_{ij} N}{N_i N_j}$ |
| СИМ-2, частный критерий: количество знаний по А.Харкевичу-Е.Луценко, 2-й вариант расчета вероятностей: *Nj* – суммарное количество объектов по *j*-му классу (последняя строка таблицы 2) | $I_{ij} = \Psi \times Log_2 \frac{P_{ij}}{P_i}$ | $I_{ij} = \Psi \times Log_2 \frac{N_{ij} N}{N_i N_j}$ |
| СИМ-3, частный критерий: разности между фактическими и теоретически ожидаемыми по критерию хи-квадрат абсолютными частотами | --- | $I_{ij} = N_{ij} - \frac{N_i N_j}{N}$ |
| СИМ-4, частный критерий: ROI - Return On Investment | $I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$ | $I_{ij} = \frac{N_{ij} N}{N_i N_j} - 1$ |
| СИМ-5, частный критерий: разность условной и безусловной вероятностей | $I_{ij} = P_{ij} - P_i$ | $I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$ |

где:

**Александр Александрович Харкевич**
**(21.1(3.2).1904 – 30.3.1965)**

$\Psi = \frac{Log_2 W}{Log_2 N}$ – упрощенное выражение для нормировочного коэффициента, переводящего количество информации в биты [21], предложенного в работе [17], обоснованного в [9] и названного автором коэффициентом эмерджентности А.А.Харкевича[16] в честь этого выдающегося советского ученого, внесшего огромный в клад в создание семантической теории информации и *фактически предложившего количественную меру знаний*, директора Института проблем передачи информации АН СССР академика АН СССР.

$s_I = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}\left(I_{ij} - \bar{I}_j\right)}$ Среднеквадратичное отклонение количества знаний во всех значениях факторов о переходе объекта в *j*-е состояние от среднего количества знаний об этом в этих значениях

$\bar{I}_j = \frac{1}{M}\sum_{i=1}^{M} I_{ij}$ Среднее количество знаний во всех значениях факторов о переходе объекта в *j*-е состояние

$s_L = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}\left(L_i - \bar{L}\right)}$ Среднеквадратичное отклонение значений вектора описания объекта от среднего этих значений

$\bar{L} = \frac{1}{M}\sum_{i=1}^{M} L_i$ Среднее значений вектора описания объекта

$$P_i = \frac{N_i}{N};\ P_j = \frac{N_j}{N};$$

$$N_i = \sum_{j=1}^{W} N_{ij};\ N_j = \sum_{i=1}^{M} N_{ij};$$

$$N = \sum_{i=1}^{M} N_i = \sum_{j=1}^{W} N_j = \sum_{i=1}^{W}\sum_{j=1}^{M} N_{ij}$$

В таблице 3 использованы обозначения:

$N_{ij}$ – суммарное количество наблюдений в исследуемой выборке *факта*: "действовало *i-е* значение фактора и объект перешел в *j-е* состояние";

$N_j$ – суммарное количество встреч различных значений факторов у объектов, перешедших в *j-е* состояние;

$N_i$ – суммарное количество встреч *i-го* значения фактора у всех объектов исследуемой выборки;

[16] Источник информации: http://www.help-rus-student.ru/text/86/120.htm

$N$ – суммарное количество встреч различных значений факторов у всех объектов исследуемой выборки.

$P_{ij}$ – условная вероятность перехода объекта в *j-е* состояние *при условии* действия на него *i-го* значения фактора;

$P_j$ – безусловная вероятность перехода объекта в *j-е* состояние (вероятность самопроизвольного перехода или вероятность перехода, посчитанная по всей выборке, т.е. при действии *любого* значения фактора).

$P_i$ – безусловная вероятность встречи *i-го* значения фактора или вероятность его встречи по всей выборке.

**Таблица 4 – МАТРИЦА ЗНАНИЙ**

| | | Классы | | | | | Значимость фактора |
|---|---|---|---|---|---|---|---|
| | | **1** | **...** | **j** | **...** | **W** | |
| Значения факторов | **1** | $I_{11}$ | | $I_{1j}$ | | $I_{1W}$ | $s_1 = \sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{1j}-\bar{I}_1\right)^2}$ |
| | **...** | | | | | | |
| | **i** | $I_{i1}$ | | $I_{ij}$ | | $I_{iW}$ | $s_i = \sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{ij}-\bar{I}_i\right)^2}$ |
| | **...** | | | | | | |
| | **M** | $I_{M1}$ | | $I_{Mj}$ | | $I_{MW}$ | $s_M = \sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{Mj}-\bar{I}_M\right)^2}$ |
| **Степень редукции класса** | | $s_1$ | | $s_j$ | | $s_W$ | $H = \sqrt[2]{\frac{1}{(W\cdot M-1)}\sum_{j=1}^{W}\sum_{i=1}^{M}\left(I_{ij}-\bar{I}\right)^2}$ |

Здесь – $\bar{I}_i$ это *среднее* количество знаний в *i-м* значении фактора:

$$\bar{I}_i = \frac{1}{W}\sum_{j=1}^{W} I_{ij}$$

***Количественные значения коэффициентов $I_{ij}$ таблицы 4 являются знаниями о том, что "объект перейдет в j-е состояние" если "на объект действует i-е значение фактора".***

Утверждение о том, что это именно ***знания***, а не данные или информация (или что-либо еще), требует специального серьезного обоснования, которое дано автором в работах [8, 9] и ряде других работ, начиная с [17] и здесь не приводится в связи с доступностью этих работ в Internet и достаточно большого объема этого обоснования.

Принципиально важно, что эти весовые коэффициенты не определяются экспертами на основе опыта интуитивным неформализуемым способом, а *рассчитываются непосредственно на основе эмпирических данных на основе теоретически обоснованных моделей, хорошо зарекомендовавших себя на практике при решении широкого круга задач в различных предметных областях.*

Когда количество информации $I_{ij}>0$ – ***i–й*** фактор способствует переходу объекта управления в ***j-е*** состояние, когда $I_{ij}<0$ – препятствует этому переходу, когда же $I_{ij}=0$ – никак не влияет на это. В векторе ***i-го*** фактора (строка матрицы информативностей) отображается, какое количество информации о переходе объекта управления в каждое из будущих состояний содержится в том факте, что данный фактор действует. В векторе ***j-го*** состояния класса (столбец матрицы информативностей) отображается, какое количество информации о переходе объекта управления в соответствующее состояние содержится в каждом из факторов.

*Таким образом, матрица информативностей (таблица 4) является обобщенной таблицей решений, в которой входы (факторы) и выходы (будущие состояния объекта управления) связаны друг с другом не с помощью классических (Аристотелевских) импликаций, принимающих только значения: "Истина" и "Ложь", а* ***различными значениями истинности, выраженными в битах*** *и принимающими значения от положительного теоретически-максимально-возможного ("Максимальная степень истинности"), до теоретически неограниченного отрицательного ("Степень ложности").*

Фактически предложенная модель позволяет осуществить *синтез обобщенных таблиц решений* для различных предметных областей непосредственно на основе эмпирических исходных данных и *продуцировать на их основе прямые и обратные правдоподобные (нечеткие) логические рассуждения по неклассическим схемам* ***с различными расчетными значениями истинности***, являющимся обобщением классических импликаций.

Таким образом, данная *модель позволяет рассчитать какое* ***количество знаний*** *содержится в любом факте о наступлении любого события в любой предметной области, причем для этого не требуется повторности этих фактов и событий.* Если же эти повторности осуществляются и при этом наблюдается некоторая вариабельность значений факторов, обуславливающих наступление тех или иных событий, то модель обеспечивает ***многопараметрическую типизацию***, т.е. синтез обобщенных образов классов или категорий наступающих событий с количественной оценкой степени и знака влияния на их наступление различных значений факторов. Причем эти значения факторов могут быть как количественными, так и качественными и измеряться в любых единицах измерения, в любом случае в модели оценивается количество знаний которое в них содержится о насту-

плении событий, переходе объекта управления в определенные состояния или просто о его принадлежности к тем или иным классам.

Все эти модели (представленные в таблице 3) можно считать различными *вариациями* одной *базовой модели знаний*, в которой мерой связи между признаком и классом является отношение условной вероятности наблюдения признака у объектов класса к безусловной вероятности его наблюдения по всей выборке, так как *отличаются они только способами нормировки частных критериев **к нулю** при отсутствии причинно-следственной связи между значением фактора и поведением объекта управления или его принадлежностью к тому или иному классу*. Все эти модели (кроме СИМ-5) поддерживаются новой версией системой "Эйдос".

Нормировать частные критерии к нулю при отсутствии связи (когда условная вероятность наблюдения признака у объектов класса равно безусловной вероятности его наблюдения по всей выборке: $P_{ij}=P_i$) необходимо, чтобы их было удобно использовать в *аддитивном* интегральном критерии. Это можно сделать разными способами. Например, в ROI (СИМ-4) из отношения условной вероятности к безусловной просто вычитается 1.

Критерий А.Харкевича тесно связан с критерием хи-квадрат. Рассмотрим выражение для частного критерия через абсолютные частоты в первой семантической информационной модели (СИМ-1):

$$I_{ij} = \Psi \times Log_2 \frac{N_{ij} N}{N_i N_j}$$

Преобразуем это выражение, учитывая, что логарифм отношения равен разности логарифмов:

$$I_{ij} = \Psi \times \left( Log_2 N_{ij} - Log_2 \frac{N_i N_j}{N} \right)$$

Сравнивая полученное выражение с выражением для частного критерия на основе хи-квадрат в СИМ-3 из таблицы 3

$$I_{ij} = N_{ij} - \frac{N_i N_j}{N}$$

видим, что они отличаются только шкалой измерения (логарифмическая шкала или нет) и постоянным множителем, т.е. по сути, единицами измерения. Величина $N_{ij}$ представляет собой фактическое количество наблюдений *i*-го признака у объектов *j*-го класса, а $t_{ij} = \frac{N_i N_j}{N}$ – теоретически ожидаемое в соответствии с критерием хи-квадрат число таких наблюдений.

Также и логарифм отношения условной и безусловной вероятности СИМ-1 и СИМ-2 является разностью их логарифмов:

$$I_{ij} = \Psi \times Log_2 \frac{P_{ij}}{P_i} = \Psi \times \left( LogP_{ij} - LogP_i \right)$$

и отличается от разности этих вероятностей (СИМ-5) только постоянным для каждой модели множителем и применением логарифмической шкалы вместо линейной.

В настоящее время для каждой модели в АСК-анализе используется *два аддитивных интегральных критерия*: это свертка (сумма частных критериев по тем признакам, которые встречаются у объекта) и нормированная свертка или корреляция:

$$I_j = \sum_{i=1}^{M} I_{ij} L_i, \qquad I_j = \frac{1}{s_I s_L} \sum_{i=1}^{M} \left( I_{ij} - \bar{I}_j \right) \left( L_i - \bar{L} \right),$$

Отметим, что при расчете интегрального критерия на основе матрицы знаний закладываются основы для *решения проблем 4 и 5*. Дело в том, что по своей математической форме интегральный критерий является сверткой (скалярным произведением вектора объекта и вектора класса) или нормированной сверткой, т.е. корреляцией. Это означает, что если эти вектора являются суммой двух сигналов: полезного и белого шума, то *при расчете неметрического интегрального критерия белый шум будет подавляться*, т.к. корреляция белого шума с самим собой (автокорреляция) стремится к нулю по самому определению белого шума. Поэтому интегральный критерий сходства объекта со случным набором признаков с любыми образами классов, или реального объекта с образами классов, сформированными случайным образом, будет близок нулю. Это означает, что ***выбранный интегральный критерий сходства является высокоэффективным средством подавления белого шума и выделения знаний из шума,*** который неизбежно присутствует в эмпирических данных.

Важно также отметить ***неметрическую природу*** предложенного в АСК-анализе интегрального критерия сходства, благодаря чему *его применение является* ***корректным*** *и при* ***неортонормированном*** *семантическом информационном пространстве, каким оно в подавляющем количестве случае и является, т.е. в общем случае*. *В этом состоит предлагаемое в АСК-анализе решение проблем 1.1 и 1.2.*

Метод кластеризации, реализованный в АСК-анализе, в котором сравниваются и объединяются когнитивные модели объектов и классов (кластеров), т.е. их модели, основанные на знаниях и представленные в матрице знаний, будем называть «*Метод когнитивной кластеризации*» или *кластеризацией, основанной на знаниях*. Ясно, что кластеризация, основанная на знаниях, может быть реализована уже не в статистических системах, а только *методами искусственного интеллекта*, т.е. в интеллектуальных системах, работающих с базами знаний. При этом, конечно, может быть разработано много методов кластеризации, основанных на зна-

ниях (не меньше чем уже существующих), отличающихся способами вычисления и представления знаний в различных интеллектуальных системах, а также способами использования знаний для формирования кластеров. В любом случае *кластеризация на основе знаний* – это новое перспективное направление исследований и разработок, в котором уже есть достижения[17].

**Рассмотрим *предлагаемый* алгоритм когнитивной кластеризации в графической и текстовой форме (рисунок 3):**

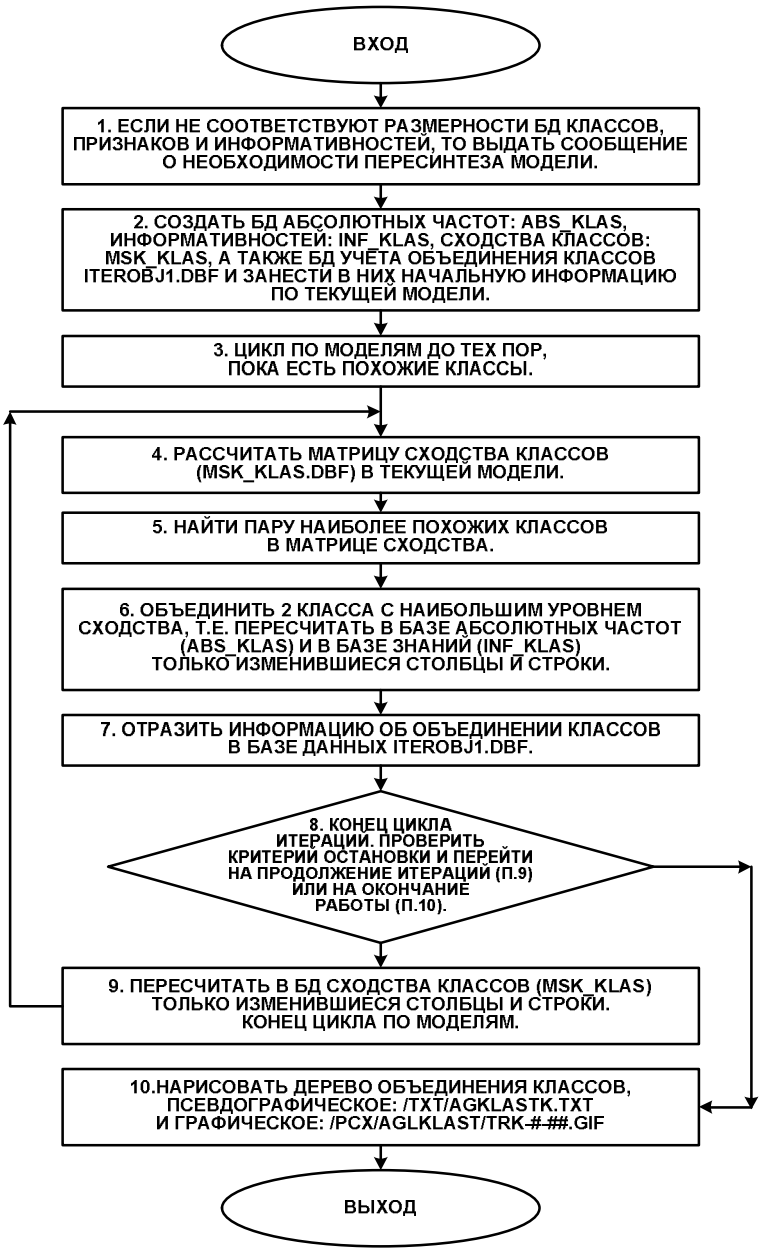



**Рисунок 2. Алгоритм когнитивной кластеризации или кластеризации, основанной на знаниях**

[17] Еще в 1982 Кохоненом представлены модели нейронной сети, обучающейся без учителя, решающей задачи кластеризации, визуализации данных и другие задачи предварительного анализа данных.

Дадим необходимые пояснения к приведенному алгоритму.

*1. Если не соответствуют размерности баз данных (БД) классов, признаков и информативностей, то выдать сообщение о необходимости пересинтеза модели.*

*2. Создать БД абсолютных частот: ABS_KLAS, информативностей: INF_KLAS, сходства классов: MSK_KLAS, а также БД учета объединения классов IterObj1.dbf и занести в них начальную информацию по текущей модели.*

Данный режим реализован в модуле _5126 системы «Эйдос» и обеспечивает работу с любой из четырех моделей или со всеми этими моделями по очереди, поддерживаемых системой и приведенных в таблице 3. При этом в базах данных этих моделей ничего не изменяется.

*3. Цикл по моделям до тех пор, пока есть похожие классы.*

*4. Рассчитать матрицу сходства классов MSK_KLAS в текущей модели.*

Эта матрица рассчитываемся на основе матрицы ***знаний*** модели, заданной при запуске режима (СИМ-1 – СИМ-4), путем расчета *корреляции* обобщенных образов классов (т.е. векторов или профилей классов).

*5. Найти пару наиболее похожих классов в матрице сходства.*

Здесь определяются два класса, у которых на предыдущем шаге было обнаружено наивысшее сходство. При этом при запуске режима задается параметр: «Исключать ли артефакты (выбросы)». Если задано исключать, то рассматриваются только положительные уровни сходства, если нет – то и отрицательные, т.е. в этом случае могут быть объединены и *непохожие* классы, но наименее непохожие из всех, если других нет. Считается, что непохожие кассы являются исключениями или «выбросами».

*6. Объединить 2 класса с наибольшим уровнем сходства.*

Данный пункт алгоритма требует наиболее детальных пояснений. Как же объединяются классы в методе когнитивной кластеризации? Сначала *суммируются* абсолютные частоты этих классов в таблице 2, причем сумма рассчитывается в столбце класса с меньшим кодом, а затем частоты класса с большим кодом обнуляются. После этого в базе знаний (таблица 4) с использованием частного критерия соответствующей модели (таблица 3) пересчитываются *только изменившиеся* столбцы и строки, т.е. пересчитывается столбец класса с меньшим кодом, а столбец класса с большим кодом обнуляется.

*7. Отразить информацию об объединении классов в БД IterObj1.dbf.*

*8. Конец цикла итераций. Проверить критерий остановки и перейти на продолжение итераций (п.9) или на окончание работы (п.10).*

*9. Пересчитать в базе данных сходства классов (MSK_KLAS) только изменившиеся столбцы и строки. Конец цикла по моделям.*

*10. Нарисовать дерево объединения классов, псевдографическое: /TXT/AgKlastK.txt и графическое: /PCX/AGLKLAST/TrK-#-##.GIF.*

**Далее рассмотрим работу предлагаемой математической модели и реализующего ее алгоритма когнитивной кластеризации на простом численном примере.**

***Численный пример*** основан на варианте той же задачи из работы [7], которая использовалась в статье [8]. В книге Д.Мичи и Р.Джонстона "Компьютер – творец" [7] эта задача приводится (на страницах: 205-208) в качестве примера задачи, решаемой методами искусственного интеллекта. Авторами этой задачи являются Рышард Михальски и Джеймс Ларсон. Суть этой задачи сводится к тому, чтобы выработать правила, обеспечивающие идентификацию и классификацию железнодорожных составов на основе их формализованных описаний (рисунок 2).

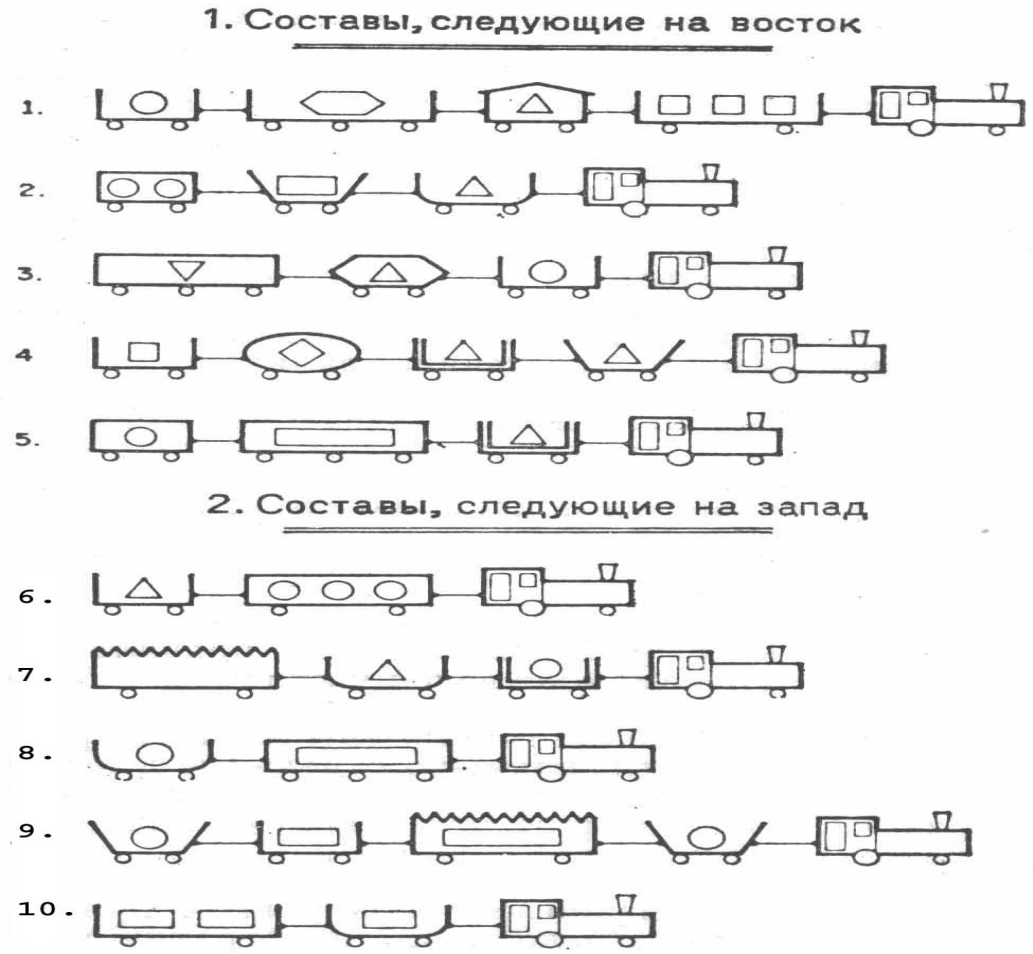



Рисунок 3. Исходные данные по численному примеру в графическом виде

Важно отметить, что в данной задаче речь идет о классификации *изображений* объектов. Признаки этих объектов в данном случае выявляются человеком, однако в принципе могут быть разработаны программы, выявляющие подобные признаки непосредственно из графических файлов с изображениями объектов. Выбор данной задачи не накладывает ограничений на выводы, полученные в результате ее исследования [8].

**Этапы АСК-анализа предметной области**

В работе [9] предложены следующие этапы АСК-анализа предметной области:

1. Когнитивная структуризация предметной области, при которой определяется, что мы хотим прогнозировать и на основе чего (конструирование классификационных и описательных *шкал*).

2. Формализация предметной области:

– разработка *градаций* классификационных и описательных шкал (номинального, порядкового и числового типа);

– использование разработанных на предыдущих этапах классификационных и описательных шкал и градаций для формального описания (кодирования) *исходных данных* (исследуемой выборки).

3. Синтез и верификация (оценка степени адекватности) модели.

4. *Если модель адекватна*, то ее использование для решения задач идентификации, прогнозирования и принятия решений, а также для исследования моделируемой предметной области.

Этап формализации предметной области при небольших размерностях модели (т.е. когда мало классов и признаков) и небольших объемах исходных данных может быть выполнен вручную. Однако даже в этом случае какие-либо изменения на первых этапах создания модели могут быть весьма трудоемкими. А такие изменения могут быть необходимыми и могут осуществляться *многократно*, так как качество модели определяется только на следующем этапе после ее синтеза в процессе верификации. Если же размерность модели и объем исходных данных велики, то вручную выполнить этап формализации предметной области весьма проблематично. Поэтому в системе «Эйдос» реализовано много программных интерфейсов с внешними базами исходных данных различной структуры, которые позволяют автоматизировать этап формализации предметной области, т.е. выполнить его автоматически с учетом параметров формализации, заданных исследователем в диалоге. В численном примере, приведенном в данной статье, авторы воспользовались программным интерфейсом формализации предметной области системы «Эйдос», реализованным в программном модуле (режиме) _152, который мы и рассмотрим.

**Программный интерфейс ввода исходных в систему «Эйдос».**

Скриншот Help режима _152 приведен на рисунке 4, а скриншот экранной формы с диалогом пользователя по заданию параметров формализации предметной области – на рисунках 5 и 6. ***Форма представления исходных данных*** для данного режима, заполненная реальными данными по рассматриваемому численному примеру, приведена в таблице 5. Исходные данные представляются в форме денормализованной таблицы MS Excel, которая затем записывается из самого Excel в виде файла базы данных стандарта DBF 4 (dBASE IV) (*.dbf), непосредственно воспринимаемого системой «Эйдос».

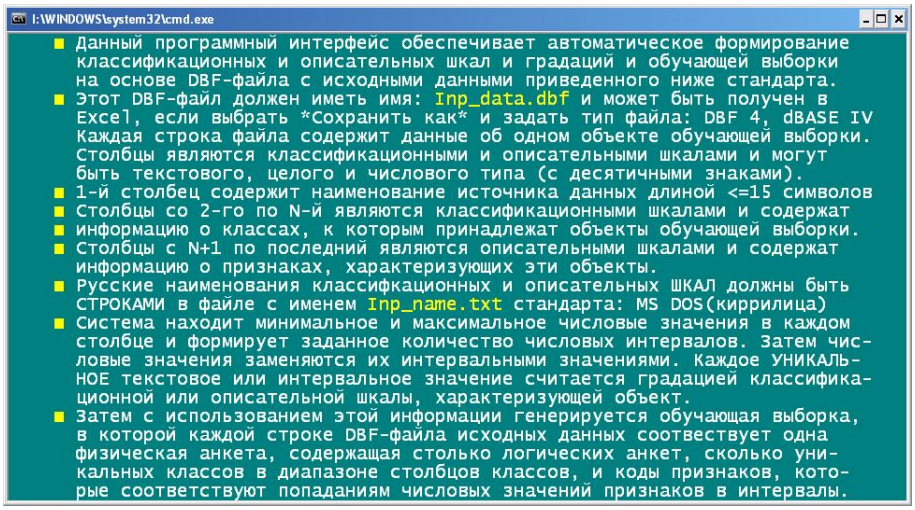



**Рисунок 4. Скриншот Help программного интерфейса формализации предметной области системы «Эйдос» (режим _152)**

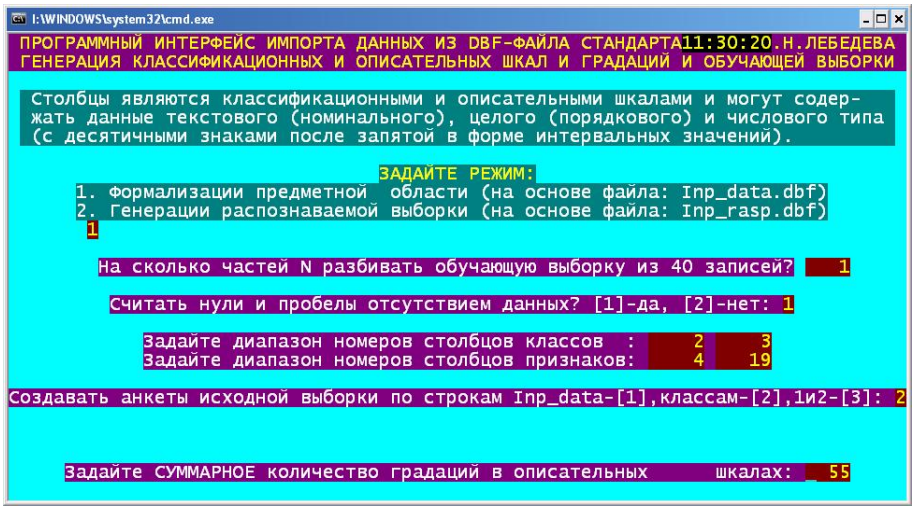



**Рисунок 5. Скриншот экранной формы диалога пользователя по заданию параметров формализации предметной области системы «Эйдос» (режим _152)**


| | Классификационные | | | Описательные | | |
|---|---|---|---|---|---|---|
| | Шкалы | Градации | Гр/шк | Шкалы | Градации | Гр/шк |
| Числовые | 0 | 0 | 0 | 2 | 10 | 5 |
| Текстовые | 2 | 12 | 6 | 14 | 45 | 3 |
| ВСЕГО: | 2 | 12 | 6 | 16 | 55 | 3 |

**Рисунок 6. Скриншот экранной формы диалога пользователя по заданию параметров формализации предметной области системы «Эйдос» (режим _152)**

**Таблица 5 – MS EXEL-ФОРМА ДЛЯ ПРЕДСТАВЛЕНИЯ ИСХОДНЫХ ДАННЫХ ПО ЧИСЛЕННОМУ ПРИМЕРУ**

| Источник информации | Классификационные шкалы | | Описательные шкалы | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Состав следует на | Название состава | Количество вагонов | Суммарный вес груза | Форма вагона | Длина вагона | Количество осей вагона | Грузоподъемность вагона | Вид стенок вагона | Вид крыши вагона | Груз-отсутствует | Груз-треугольник_прямой | Груз-треугольник_перевернутый | Груз-ромб | Груз-овал | Груз-квадрат | Груз-прямоугольник_короткий | Груз-прямоугольник_длинный |
| Состав-01 в целом | ВОСТОК | Состав-01 | **4** | **61,000** | | | | | | | | | | | | | | |
| Состав-02 в целом | ВОСТОК | Состав-02 | **3** | **21,000** | | | | | | | | | | | | | | |
| Состав-03 в целом | ВОСТОК | Состав-03 | **3** | **24,000** | | | | | | | | | | | | | | |
| Состав-04 в целом | ВОСТОК | Состав-04 | **3** | **45,000** | | | | | | | | | | | | | | |
| Состав-05 в целом | ВОСТОК | Состав-05 | **3** | **22,000** | | | | | | | | | | | | | | |
| Состав-06 в целом | ЗАПАД | Состав-06 | **2** | **20,000** | | | | | | | | | | | | | | |
| Состав-07 в целом | ЗАПАД | Состав-07 | **3** | **14,000** | | | | | | | | | | | | | | |
| Состав-08 в целом | ЗАПАД | Состав-08 | **2** | **11,000** | | | | | | | | | | | | | | |
| Состав-09 в целом | ЗАПАД | Состав-09 | **3** | **18,000** | | | | | | | | | | | | | | |
| Состав-10 в целом | ЗАПАД | Состав-10 | 2 | **12,000** | | | | | | | | | | | | | | |
| Сост-01,ваг-1 | ВОСТОК | Состав-01 | | | Прямоугольная | Короткий | 2 | 40,0 | Одинарные | Отсутствует | | | | | 1 | | | |
| Сост-01,ваг-2 | ВОСТОК | Состав-01 | | | Прямоугольная | Длинный | 3 | 80,0 | Одинарные | Отсутствует | | | | 1 | | | | |
| Сост-01,ваг-3 | ВОСТОК | Состав-01 | | | Прямоугольная | Короткий | 2 | 40,0 | Одинарные | Треугольная | | 1 | | | | | | |
| Сост-01,ваг-4 | ВОСТОК | Состав-01 | | | Прямоугольная | Длинный | 2 | 60,0 | Одинарные | Отсутствует | | | | | | 3 | | |
| Сост-02,ваг-1 | ВОСТОК | Состав-02 | | | Прямоугольная | Короткий | 2 | 40,0 | Одинарные | Прямая | | | | | 2 | | | |
| Сост-02,ваг-2 | ВОСТОК | Состав-02 | | | V-образная | Короткий | 2 | 30,0 | Одинарные | Отсутствует | | | | | | | 1 | |
| Сост-02,ваг-3 | ВОСТОК | Состав-02 | | | U-образная | Короткий | 2 | 30,0 | Одинарные | Отсутствует | | 1 | | | | | | |
| Сост-03,ваг-1 | ВОСТОК | Состав-03 | | | Прямоугольная | Длинный | 3 | 80,0 | Одинарные | Прямая | | | 1 | | | | | |
| Сост-03,ваг-2 | ВОСТОК | Состав-03 | | | Ромбовидная | Короткий | 2 | 40,0 | Одинарные | Прямая | | 1 | | | | | | |
| Сост-03,ваг-3 | ВОСТОК | Состав-03 | | | Прямоугольная | Короткий | 2 | 40,0 | Одинарные | Отсутствует | | | | | 1 | | | |
| Сост-04,ваг-1 | ВОСТОК | Состав-04 | | | Прямоугольная | Короткий | 2 | 40,0 | Одинарные | Отсутствует | | | | | | 1 | | |
| Сост-04,ваг-2 | ВОСТОК | Состав-04 | | | Овальная | Короткий | 2 | 40,0 | Одинарные | Овальная | | | | 1 | | | | |
| Сост-04,ваг-3 | ВОСТОК | Состав-04 | | | Прямоугольная | Короткий | 2 | 40,0 | Двойные | Отсутствует | | 1 | | | | | | |
| Сост-04,ваг-4 | ВОСТОК | Состав-04 | | | U-образная | Короткий | 2 | 30,0 | Одинарные | Отсутствует | | 1 | | | | | | |
| Сост-05,ваг-1 | ВОСТОК | Состав-05 | | | Прямоугольная | Короткий | 2 | 40,0 | Одинарные | Прямая | | | | | 1 | | | |
| Сост-05,ваг-2 | ВОСТОК | Состав-05 | | | Прямоугольная | Длинный | 3 | 80,0 | Одинарные | Прямая | | | | | | | | 1 |
| Сост-05,ваг-3 | ВОСТОК | Состав-05 | | | Прямоугольная | Короткий | 2 | 40,0 | Двойные | Отсутствует | | 1 | | | | | | |
| Сост-06,ваг-1 | ЗАПАД | Состав-06 | | | Прямоугольная | Короткий | 2 | 40,0 | Одинарные | Отсутствует | | 1 | | | | | | |
| Сост-06,ваг-2 | ЗАПАД | Состав-06 | | | Прямоугольная | Длинный | 2 | 60,0 | Одинарные | Прямая | | | | | 3 | | | |
| Сост-07,ваг-1 | ЗАПАД | Состав-07 | | | Прямоугольная | Длинный | 2 | 60,0 | Одинарные | Гофрированная | 1 | | | | | | | |
| Сост-07,ваг-2 | ЗАПАД | Состав-07 | | | U-образная | Короткий | 2 | 30,0 | Одинарные | Отсутствует | | 1 | | | | | | |
| Сост-07,ваг-3 | ЗАПАД | Состав-07 | | | Прямоугольная | Короткий | 2 | 40,0 | Двойные | Отсутствует | | | | | 1 | | | |
| Сост-08,ваг-1 | ЗАПАД | Состав-08 | | | U-образная | Короткий | 2 | 30,0 | Одинарные | Отсутствует | | | | | 1 | | | |
| Сост-08,ваг-2 | ЗАПАД | Состав-08 | | | Прямоугольная | Длинный | 3 | 80,0 | Одинарные | Прямая | | | | | | | | 1 |
| Сост-09,ваг-1 | ЗАПАД | Состав-09 | | | V-образная | Короткий | 2 | 30,0 | Одинарные | Отсутствует | | | | | 1 | | | |
| Сост-09,ваг-2 | ЗАПАД | Состав-09 | | | Прямоугольная | Короткий | 2 | 40,0 | Одинарные | Отсутствует | | | | | | | 1 | |
| Сост-09,ваг-3 | ЗАПАД | Состав-09 | | | Прямоугольная | Длинный | 2 | 60,0 | Одинарные | Гофрированная | | | | | | | | 1 |
| Сост-09,ваг-4 | ЗАПАД | Состав-09 | | | V-образная | Короткий | 2 | 30,0 | Одинарные | Отсутствует | | | | | 1 | | | |
| Сост-10,ваг-1 | ЗАПАД | Состав-10 | | | Прямоугольная | Длинный | 2 | 60,0 | Одинарные | Отсутствует | | | | | | | 2 | |
| Сост-10,ваг-2 | ЗАПАД | Состав-10 | | | U-образная | Короткий | 2 | 30,0 | Одинарные | Отсутствует | | | | | | | 1 | |

**Результаты формализации предметной области.**

В результате работы программного интерфейса _152 автоматически формируются классификационные и описательные шкалы и градации и с их использованием кодируются исходные данные, в результат чего формируется обучающая выборка (таблицы 6 – 9):

**Таблица 6 – СПРАВОЧНИК КЛАССОВ (КЛАССИФИКАЦИОННЫХ ШКАЛ И ГРАДАЦИЙ)**

| KOD | NAME |
|---|---|
| 1 | Состав следует на ВОСТОК |
| 2 | Состав следует на ЗАПАД |
| 3 | Состав-01 |
| 4 | Состав-02 |
| 5 | Состав-03 |
| 6 | Состав-04 |
| 7 | Состав-05 |
| 8 | Состав-06 |
| 9 | Состав-07 |
| 10 | Состав-08 |
| 11 | Состав-09 |
| 12 | Состав-10 |

В таблице 1, по сути, приведены *исходные кластеры*, первые два из которых являются составными или «полиобъектными» (решение о принадлежности объектов к тому или иному составному классу принималось экспертом – учителем), а последующие 10 – «монообъетными». Первый полиобъектный класс состоит из объектов с кодами 3-7, а второй – 8-12, монообъектные классы состоят из объектов с кодами от 3 до 12.

**Таблица 7 – СПРАВОЧНИК ПРИЗНАКОВ (ОПИСАТЕЛЬНЫХ ШКАЛ И ГРАДАЦИЙ)**

| Код | Наименование | Ед.изм. | Тип шкалы |
|---|---|---|---|
| 1 | КОЛИЧЕСТВО ВАГОНОВ-2 | Шт. | Порядковая (целочисленная) |
| 2 | КОЛИЧЕСТВО ВАГОНОВ-3 | | |
| 3 | КОЛИЧЕСТВО ВАГОНОВ-4 | | |
| 4 | СУММАРНЫЙ ВЕС ГРУЗА: 1/5-{11.00, 21.00} | Тонны | Числовая (в интервальных значениях) |
| 5 | СУММАРНЫЙ ВЕС ГРУЗА: 2/5-{21.00, 31.00} | | |
| 6 | СУММАРНЫЙ ВЕС ГРУЗА: 3/5-{31.00, 41.00} | | |
| 7 | СУММАРНЫЙ ВЕС ГРУЗА: 4/5-{41.00, 51.00} | | |
| 8 | СУММАРНЫЙ ВЕС ГРУЗА: 5/5-{51.00, 61.00} | | |
| 9 | ФОРМА ВАГОНА-U-образная | | Текстовая (номинальная) |
| 10 | ФОРМА ВАГОНА-V-образная | | |
| 11 | ФОРМА ВАГОНА-Овальная | | |
| 12 | ФОРМА ВАГОНА-Прямоугольная | | |
| 13 | ФОРМА ВАГОНА-Ромбовидная | | |
| 14 | ДЛИНА ВАГОНА-Длинный | | Порядковая (целочисленная) |
| 15 | ДЛИНА ВАГОНА-Короткий | | |
| 16 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-2 | Шт. | Порядковая (целочисленная) |
| 17 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-3 | | |
| 18 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00} | Тонны | Числовая (в интервальных значениях) |
| 19 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00} | | |
| 20 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00} | | |
| 21 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00} | | |
| 22 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00} | | |
| 23 | ВИД СТЕНОК ВАГОНА-Двойные | Шт. | Порядковая (целочисленная) |
| 24 | ВИД СТЕНОК ВАГОНА-Одинарные | | |
| 25 | ВИД КРЫШИ ВАГОНА-Гофрированная | | Текстовая (номинальная) |
| 26 | ВИД КРЫШИ ВАГОНА-Овальная | | |
| 27 | ВИД КРЫШИ ВАГОНА-Отсутствует | | |
| 28 | ВИД КРЫШИ ВАГОНА-Прямая | | |
| 29 | ВИД КРЫШИ ВАГОНА-Треугольная | | |
| 30 | ГРУЗ-ОТСУТСТВУЕТ-0001 | | Текстовая (номинальная) |
| 31 | ГРУЗ-ТРЕУГОЛЬНИК_ПРЯМОЙ-0001 | | |

| | | | |
|---|---|---|---|
| 32 | ГРУЗ-ТРЕУГОЛЬНИК_ПЕРЕВЕРНУТЫЙ-0001 | | |
| 33 | ГРУЗ-РОМБ-0001 | | |
| 34 | ГРУЗ-ОВАЛ-0001 | | |
| 35 | ГРУЗ-ОВАЛ-0002 | | |
| 36 | ГРУЗ-ОВАЛ-0003 | | |
| 37 | ГРУЗ-КВАДРАТ-0001 | | |
| 38 | ГРУЗ-КВАДРАТ-0003 | | |
| 39 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0001 | | |
| 40 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0002 | | |
| 41 | ГРУЗ-ПРЯМОУГОЛЬНИК_ДЛИННЫЙ-0001 | | |

Отметим, что эти признаки объектов формализуются в текстовых (номинальных), порядковых (целочисленных) и числовых (со знаками после запятой) шкалах и измеряются в разных единицах измерения, которые можно ввести только для числовых и порядковых шкал.

Соответственно и *исходные данные (исследуемая выборка)* представлены в форме переменных с количественными и качественными значениями, измеряемыми в различных единицах измерения и формализуемыми в шкалах различного типа. Но в исходных данных есть информация не только о признаках объектов, но и об их принадлежности к тем или иным классам (полиобъектным или монообъектным). Вся эта информация представлена в обучающей выборке, стоящей из двух баз данных, базы заголовков и базы признаков, связанных отношением «один ко многим» по полю «Код объекта» (таблицы 8 и 9):

**Таблица 8 – ОБУЧАЮЩАЯ ВЫБОРКА (база заголовков)**

| Код объекта | Наименование объекта | Коды классов | |
|---|---|---|---|
| 41 | ВОСТОК | 1 | |
| 42 | ЗАПАД | 2 | |
| 43 | Состав-01 | 3 | 1 |
| 44 | Состав-02 | 4 | 1 |
| 45 | Состав-03 | 5 | 1 |
| 46 | Состав-04 | 6 | 1 |
| 47 | Состав-05 | 7 | 1 |
| 48 | Состав-06 | 8 | 2 |
| 49 | Состав-07 | 9 | 2 |
| 50 | Состав-08 | 10 | 2 |
| 51 | Состав-09 | 11 | 2 |
| 52 | Состав-10 | 12 | 2 |

**Таблица 9 – ОБУЧАЮЩАЯ ВЫБОРКА (база признаков)**

| Код объекта | Коды признаков | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 41 | 3 | 8 | 2 | 4 | 5 | 2 | 5 | 2 | 7 | 2 | 5 |
| 41 | 15 | 16 | 18 | 19 | 24 | 27 | 34 | 12 | 14 | 17 | 22 |
| 41 | 27 | 33 | 12 | 15 | 16 | 18 | 19 | 24 | 29 | 31 | 12 |
| 41 | 16 | 20 | 21 | 24 | 27 | 38 | 12 | 15 | 16 | 18 | 19 |
| 41 | 28 | 35 | 10 | 15 | 16 | 18 | 24 | 27 | 39 | 9 | 15 |
| 41 | 18 | 24 | 27 | 31 | 12 | 14 | 17 | 22 | 24 | 28 | 32 |
| 41 | 15 | 16 | 18 | 19 | 24 | 28 | 31 | 12 | 15 | 16 | 18 |
| 41 | 24 | 27 | 34 | 12 | 15 | 16 | 18 | 19 | 24 | 27 | 37 |
| 41 | 15 | 16 | 18 | 19 | 24 | 26 | 33 | 12 | 15 | 16 | 18 |
| 41 | 23 | 27 | 31 | 9 | 15 | 16 | 18 | 24 | 27 | 31 | 12 |
| 41 | 16 | 18 | 19 | 24 | 28 | 34 | 12 | 14 | 17 | 22 | 24 |
| 41 | 41 | 12 | 15 | 16 | 18 | 19 | 23 | 27 | 31 | | |
| 42 | 1 | 4 | 2 | 4 | 1 | 4 | 2 | 4 | 1 | 4 | 12 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 42 | 16 | 18 | 19 | 24 | 27 | 31 | 12 | 14 | 16 | 20 | 21 |
| 42 | 28 | 36 | 12 | 14 | 16 | 20 | 21 | 24 | 25 | 30 | 9 |
| 42 | 16 | 18 | 24 | 27 | 31 | 12 | 15 | 16 | 18 | 19 | 23 |
| 42 | 34 | 9 | 15 | 16 | 18 | 24 | 27 | 34 | 12 | 14 | 17 |
| 42 | 24 | 28 | 41 | 10 | 15 | 16 | 18 | 24 | 27 | 34 | 12 |
| 42 | 16 | 18 | 19 | 24 | 27 | 39 | 12 | 14 | 16 | 20 | 21 |
| 42 | 25 | 41 | 10 | 15 | 16 | 18 | 24 | 27 | 34 | 12 | 14 |
| 42 | 20 | 21 | 24 | 27 | 40 | 9 | 15 | 16 | 18 | 24 | 27 |
| 42 | 39 | | | | | | | | | | |
| 43 | 3 | 8 | 12 | 15 | 16 | 18 | 19 | 24 | 27 | 34 | 12 |
| 43 | 17 | 22 | 24 | 27 | 33 | 12 | 15 | 16 | 18 | 19 | 24 |
| 43 | 31 | 12 | 14 | 16 | 20 | 21 | 24 | 27 | 38 | | |
| 44 | 2 | 4 | 5 | 12 | 15 | 16 | 18 | 19 | 24 | 28 | 35 |
| 44 | 15 | 16 | 18 | 24 | 27 | 39 | 9 | 15 | 16 | 18 | 24 |
| 44 | 31 | | | | | | | | | | |
| 45 | 2 | 5 | 12 | 14 | 17 | 22 | 24 | 28 | 32 | 13 | 15 |
| 45 | 18 | 19 | 24 | 28 | 31 | 12 | 15 | 16 | 18 | 19 | 24 |
| 45 | 34 | | | | | | | | | | |
| 46 | 2 | 7 | 12 | 15 | 16 | 18 | 19 | 24 | 27 | 37 | 11 |
| 46 | 16 | 18 | 19 | 24 | 26 | 33 | 12 | 15 | 16 | 18 | 19 |
| 46 | 27 | 31 | 9 | 15 | 16 | 18 | 24 | 27 | 31 | | |
| 47 | 2 | 5 | 12 | 15 | 16 | 18 | 19 | 24 | 28 | 34 | 12 |
| 47 | 17 | 22 | 24 | 28 | 41 | 12 | 15 | 16 | 18 | 19 | 23 |
| 47 | 31 | | | | | | | | | | |
| 48 | 1 | 4 | 12 | 15 | 16 | 18 | 19 | 24 | 27 | 31 | 12 |
| 48 | 16 | 20 | 21 | 24 | 28 | 36 | | | | | |
| 49 | 2 | 4 | 12 | 14 | 16 | 20 | 21 | 24 | 25 | 30 | 9 |
| 49 | 16 | 18 | 24 | 27 | 31 | 12 | 15 | 16 | 18 | 19 | 23 |
| 49 | 34 | | | | | | | | | | |
| 50 | 1 | 4 | 9 | 15 | 16 | 18 | 24 | 27 | 34 | 12 | 14 |
| 50 | 22 | 24 | 28 | 41 | | | | | | | |
| 51 | 2 | 4 | 10 | 15 | 16 | 18 | 24 | 27 | 34 | 12 | 15 |
| 51 | 18 | 19 | 24 | 27 | 39 | 12 | 14 | 16 | 20 | 21 | 24 |
| 51 | 41 | 10 | 15 | 16 | 18 | 24 | 27 | 34 | | | |
| 52 | 1 | 4 | 12 | 14 | 16 | 20 | 21 | 24 | 27 | 40 | 9 |
| 52 | 16 | 18 | 24 | 27 | 39 | | | | | | |

В системе «Эйдос» есть режим _25, экранная форма которого приведена на рисунке 7, обеспечивающий как расчет всех четырех типов моделей (СИМ-1 – СИМ-4), отличающихся видом частных критериев (таблица 3), так и измерение их достоверности с двумя видами интегральных критериев: сверткой и корреляцией.

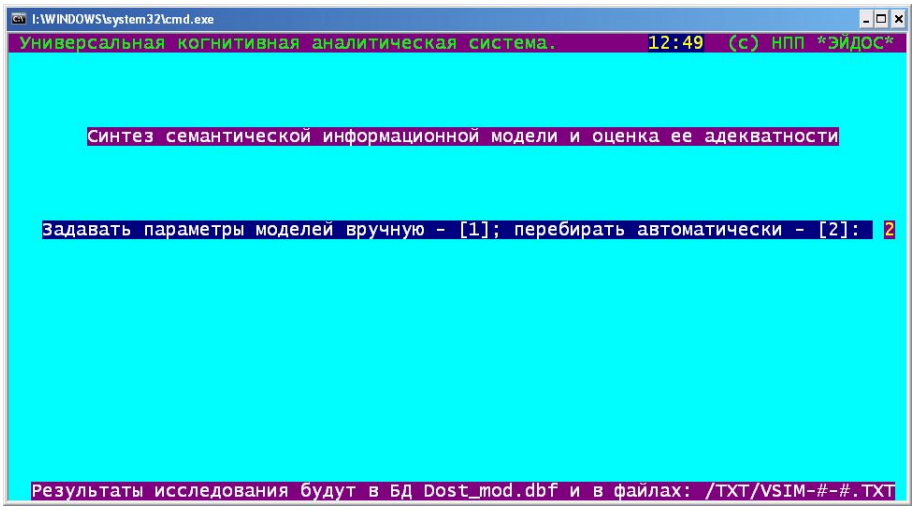



**Рисунок 7. Экранная форма режима _25 системы «Эйдос»**

В результате работы режима _25 формируется матрица абсолютных частот (таблица 10) и матрицы знаний четырех моделей (таблицы 11 – 14):

## Таблица 10 – МАТРИЦА АБСОЛЮТНЫХ ЧАСТОТ

| Код | Наименование | Классы | | | | | | | | | | | Сумма |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| 1 | КОЛИЧЕСТВО ВАГОНОВ-2 | | 6 | | | | | | 1 | | 1 | | 1 | 9 |
| 2 | КОЛИЧЕСТВО ВАГОНОВ-3 | 8 | 4 | | 1 | 1 | 1 | 1 | | 1 | | 1 | | 18 |
| 3 | КОЛИЧЕСТВО ВАГОНОВ-4 | 2 | | 1 | | | | | | | | | | 3 |
| 4 | СУММАРНЫЙ ВЕС ГРУЗА: 1/5-{11.00, 21.00} | 2 | 10 | | 1 | | | | 1 | 1 | 1 | 1 | 1 | 18 |
| 5 | СУММАРНЫЙ ВЕС ГРУЗА: 2/5-{21.00, 31.00} | 6 | | | 1 | 1 | | 1 | | | | | | 9 |
| 6 | СУММАРНЫЙ ВЕС ГРУЗА: 3/5-{31.00, 41.00} | | | | | | | | | | | | | |
| 7 | СУММАРНЫЙ ВЕС ГРУЗА: 4/5-{41.00, 51.00} | 2 | | | | | 1 | | | | | | | 3 |
| 8 | СУММАРНЫЙ ВЕС ГРУЗА: 5/5-{51.00, 61.00} | 2 | | 1 | | | | | | | | | | 3 |
| 9 | ФОРМА ВАГОНА-U-образная | 4 | 6 | | 1 | | 1 | | | 1 | 1 | | 1 | 15 |
| 10 | ФОРМА ВАГОНА-V-образная | 1 | 4 | | | | | | | | | 2 | | 7 |
| 11 | ФОРМА ВАГОНА-Овальная | 1 | | | | | 1 | | | | | | | 2 |
| 12 | ФОРМА ВАГОНА-Прямоугольная | 23 | 16 | 4 | 1 | 2 | 2 | 3 | 2 | 2 | 1 | 2 | 1 | 59 |
| 13 | ФОРМА ВАГОНА-Ромбовидная | 1 | | | | 1 | | | | | | | | 2 |
| 14 | ДЛИНА ВАГОНА-Длинный | 5 | 9 | 1 | | 1 | | | | 1 | 1 | 1 | 1 | 20 |
| 15 | ДЛИНА ВАГОНА-Короткий | 24 | 11 | 2 | 3 | 2 | 3 | 2 | 1 | 1 | 1 | 3 | | 53 |
| 16 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-2 | 26 | 22 | 3 | 3 | 1 | 4 | 2 | 2 | 3 | 1 | 3 | 2 | 72 |
| 17 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-3 | 6 | 1 | 1 | | 1 | | 1 | | | | | | 10 |
| 18 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00} | 26 | 16 | 2 | 3 | 2 | 4 | 2 | 1 | 2 | 1 | 3 | 1 | 63 |
| 19 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00} | 18 | 6 | 2 | 1 | 2 | 3 | 2 | 1 | 1 | | 1 | | 37 |
| 20 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00} | 2 | 8 | 1 | | | | | 1 | 1 | | 1 | 1 | 15 |
| 21 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00} | 2 | 8 | 1 | | | | | 1 | 1 | | 1 | 1 | 15 |
| 22 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00} | 6 | 1 | 1 | | 1 | | 1 | | | 1 | | | 11 |
| 23 | ВИД СТЕНОК ВАГОНА-Двойные | 3 | 2 | | | | | 1 | | 1 | | | | 7 |
| 24 | ВИД СТЕНОК ВАГОНА-Одинарные | 28 | 22 | 4 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 4 | 2 | 77 |
| 25 | ВИД КРЫШИ ВАГОНА-Гофрированная | | 3 | | | | | | | 1 | | | | 4 |
| 26 | ВИД КРЫШИ ВАГОНА-Овальная | 2 | | | | | 1 | | | | | | | 3 |
| 27 | ВИД КРЫШИ ВАГОНА-Отсутствует | 17 | 16 | 3 | 1 | | 3 | | 1 | 1 | 1 | 3 | 2 | 48 |
| 28 | ВИД КРЫШИ ВАГОНА-Прямая | 9 | 4 | | 1 | 2 | | 2 | 1 | | 1 | | | 20 |
| 29 | ВИД КРЫШИ ВАГОНА-Треугольная | 1 | | | | | | | | | | | | 1 |
| 30 | ГРУЗ-ОТСУТСТВУЕТ-0001 | | 2 | | | | | | | 1 | | | | 3 |
| 31 | ГРУЗ-ТРЕУГОЛЬНИК_ПРЯМОЙ-0001 | 12 | 4 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | | | | 24 |
| 32 | ГРУЗ-ТРЕУГОЛЬНИК_ПЕРЕВЕРНУТЫЙ-0001 | 2 | | | | 1 | | | | | | | | 3 |
| 33 | ГРУЗ-РОМБ-0001 | 4 | | 1 | | | 1 | | | | | | | 6 |
| 34 | ГРУЗ-ОВАЛ-0001 | 6 | 8 | 1 | | 1 | | 1 | | 1 | 1 | 2 | | 21 |
| 35 | ГРУЗ-ОВАЛ-0002 | 2 | | | 1 | | | | | | | | | 3 |
| 36 | ГРУЗ-ОВАЛ-0003 | | 2 | | | | | | 1 | | | | | 3 |
| 37 | ГРУЗ-КВАДРАТ-0001 | 2 | | | | | 1 | | | | | | | 3 |
| 38 | ГРУЗ-КВАДРАТ-0003 | 2 | | 1 | | | | | | | | | | 3 |
| 39 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0001 | 2 | 4 | | 1 | | | | | | | 1 | 1 | 9 |
| 40 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0002 | | 2 | | | | | | | | | | 1 | 3 |
| 41 | ГРУЗ-ПРЯМОУГОЛЬНИК_ДЛИННЫЙ-0001 | 2 | 4 | | | | | 1 | | | 1 | 1 | | 9 |
| | Суммарное количество признаков | 261 | 201 | 31 | 23 | 23 | 31 | 23 | 17 | 23 | 15 | 30 | 16 | 694 |

## Таблица 11 – МАТРИЦА ЗНАНИЙ СИМ-1 (В САНТИБИТАХ)

| Код | Наименование | Классы | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | КОЛИЧЕСТВО ВАГОНОВ-2 | | 46 | | | | | | 83 | | 90 | | 86 |
| 2 | КОЛИЧЕСТВО ВАГОНОВ-3 | 9 | -15 | | 28 | 28 | 12 | 28 | | 28 | | 14 | |
| 3 | КОЛИЧЕСТВО ВАГОНОВ-4 | 31 | | 110 | | | | | | | | | |
| 4 | СУММАРНЫЙ ВЕС ГРУЗА: 1/5-{11.00, 21.00} | -67 | 36 | | 28 | | | | 45 | 28 | 52 | 14 | 48 |
| 5 | СУММАРНЫЙ ВЕС ГРУЗА: 2/5-{21.00, 31.00} | 31 | | | 66 | 66 | | 66 | | | | | |
| 6 | СУММАРНЫЙ ВЕС ГРУЗА: 3/5-{31.00, 41.00} | | | | | | | | | | | | |
| 7 | СУММАРНЫЙ ВЕС ГРУЗА: 4/5-{41.00, 51.00} | 31 | | | | | 110 | | | | | | |
| 8 | СУММАРНЫЙ ВЕС ГРУЗА: 5/5-{51.00, 61.00} | 31 | | 110 | | | | | | | | | |
| 9 | ФОРМА ВАГОНА-U-образная | -19 | 18 | | 38 | | 22 | | | 38 | 62 | | 58 |
| 10 | ФОРМА ВАГОНА-V-образная | -53 | 37 | | | | | | | | | 103 | |
| 11 | ФОРМА ВАГОНА-Овальная | 16 | | | | | 132 | | | | | | |
| 12 | ФОРМА ВАГОНА-Прямоугольная | 2 | -4 | 23 | -37 | 1 | -15 | 23 | 18 | 1 | -13 | -13 | -17 |
| 13 | ФОРМА ВАГОНА-Ромбовидная | 16 | | | | 149 | | | | | | | |
| 14 | ДЛИНА ВАГОНА-Длинный | -22 | 24 | 6 | | 23 | | | | 23 | 46 | 8 | 42 |
| 15 | ДЛИНА ВАГОНА-Короткий | 10 | -18 | -9 | 29 | 7 | 13 | 7 | -14 | -31 | -7 | 15 | |
| 16 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-2 | -2 | 3 | -4 | 13 | -48 | 12 | -10 | 7 | 13 | -24 | -2 | 10 |
| 17 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-3 | 26 | -58 | 44 | | 61 | | 61 | | | | | |
| 18 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00} | 5 | -7 | -19 | 20 | -2 | 19 | -2 | -24 | -2 | -17 | 5 | -20 |
| 19 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00} | 14 | -32 | 10 | -11 | 27 | 33 | 27 | 5 | -11 | | -26 | |
| 20 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00} | -57 | 33 | 22 | | | | | 55 | 38 | | 24 | 58 |
| 21 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00} | -57 | 33 | 22 | | | | | 55 | 38 | | 24 | 58 |
| 22 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00} | 20 | -63 | 39 | | 55 | | 55 | | | 79 | | |
| 23 | ВИД СТЕНОК ВАГОНА-Двойные | 7 | -1 | | | | | 80 | | 80 | | | |
| 24 | ВИД СТЕНОК ВАГОНА-Одинарные | -2 | -1 | 8 | 9 | 9 | -7 | -13 | 3 | -13 | 10 | 10 | 7 |
| 25 | ВИД КРЫШИ ВАГОНА-Гофрированная | | 52 | | | | | | | 111 | | | |
| 26 | ВИД КРЫШИ ВАГОНА-Овальная | 31 | | | | | 110 | | | | | | |
| 27 | ВИД КРЫШИ ВАГОНА-Отсутствует | -3 | 8 | 18 | -25 | | 18 | | -9 | -25 | -2 | 20 | 32 |
| 28 | ВИД КРЫШИ ВАГОНА-Прямая | 10 | -20 | | 23 | 61 | | 61 | 39 | | 46 | | |
| 29 | ВИД КРЫШИ ВАГОНА-Треугольная | 54 | | | | | | | | | | | |
| 30 | ГРУЗ-ОТСУТСТВУЕТ-0001 | | 46 | | | | | | | 126 | | | |
| 31 | ГРУЗ-ТРЕУГОЛЬНИК_ПРЯМОЙ-0001 | 16 | -30 | -4 | 13 | 13 | 34 | 13 | 29 | 13 | | | |
| 32 | ГРУЗ-ТРЕУГОЛЬНИК_ПЕРЕВЕРНУТЫЙ-0001 | 31 | | | | 126 | | | | | | | |
| 33 | ГРУЗ-РОМБ-0001 | 31 | | 72 | | | 72 | | | | | | |
| 34 | ГРУЗ-ОВАЛ-0001 | -15 | 15 | 4 | | 20 | | 20 | | 20 | 43 | 43 | |
| 35 | ГРУЗ-ОВАЛ-0002 | 31 | | | 126 | | | | | | | | |
| 36 | ГРУЗ-ОВАЛ-0003 | | 46 | | | | | | 143 | | | | |
| 37 | ГРУЗ-КВАДРАТ-0001 | 31 | | | | | 110 | | | | | | |
| 38 | ГРУЗ-КВАДРАТ-0003 | 31 | | 110 | | | | | | | | | |
| 39 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0001 | -29 | 23 | | 66 | | | | | | | 52 | 86 |
| 40 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0002 | | 46 | | | | | | | | | | 146 |
| 41 | ГРУЗ-ПРЯМОУГОЛЬНИК_ДЛИННЫЙ-0001 | -29 | 23 | | | | | 66 | | | 90 | 52 | |

**Таблица 12 – МАТРИЦА ЗНАНИЙ СИМ-2 (в 0.01 исходных ед.изм.)**

| Код | Наименование | Классы | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | КОЛИЧЕСТВО ВАГОНОВ-2 | | 129 | | | | | | 129 | | 129 | | 129 |
| 2 | КОЛИЧЕСТВО ВАГОНОВ-3 | 70 | -30 | | 29 | 29 | 29 | 29 | | 29 | | 29 | |
| 3 | КОЛИЧЕСТВО ВАГОНОВ-4 | 129 | | 287 | | | | | | | | | |
| 4 | СУММАРНЫЙ ВЕС ГРУЗА: 1/5-{11.00, 21.00} | -130 | 103 | | 29 | | | | 29 | 29 | 29 | 29 | 29 |
| 5 | СУММАРНЫЙ ВЕС ГРУЗА: 2/5-{21.00, 31.00} | 129 | | | 129 | 129 | | 129 | | | | | |
| 6 | СУММАРНЫЙ ВЕС ГРУЗА: 3/5-{31.00, 41.00} | | | | | | | | | | | | |
| 7 | СУММАРНЫЙ ВЕС ГРУЗА: 4/5-{41.00, 51.00} | 129 | | | | | 287 | | | | | | |
| 8 | СУММАРНЫЙ ВЕС ГРУЗА: 5/5-{51.00, 61.00} | 129 | | 287 | | | | | | | | | |
| 9 | ФОРМА ВАГОНА-U-образная | -3 | 55 | | 55 | | 55 | | | 55 | 55 | | 55 |
| 10 | ФОРМА ВАГОНА-V-образная | -93 | 107 | | | | | | | | | 265 | |
| 11 | ФОРМА ВАГОНА-Овальная | 87 | | | | | 346 | | | | | | |
| 12 | ФОРМА ВАГОНА-Прямоугольная | 52 | -1 | 58 | -142 | -42 | -42 | 16 | -42 | -42 | -142 | -42 | -142 |
| 13 | ФОРМА ВАГОНА-Ромбовидная | 87 | | | | 346 | | | | | | | |
| 14 | ДЛИНА ВАГОНА-Длинный | -13 | 72 | 14 | | 14 | | | | 14 | 14 | 14 | 14 |
| 15 | ДЛИНА ВАГОНА-Короткий | 73 | -39 | -27 | 32 | -27 | 32 | -27 | -127 | -127 | -127 | 32 | |
| 16 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-2 | 40 | 16 | -13 | -13 | -171 | 29 | -71 | -71 | -13 | -171 | -13 | -71 |
| 17 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-3 | 114 | -145 | 114 | | 114 | | 114 | | | | | |
| 18 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00} | 60 | -10 | -52 | 7 | -52 | 48 | -52 | -152 | -52 | -152 | 7 | -152 |
| 19 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00} | 83 | -75 | 25 | -75 | 25 | 83 | 25 | -75 | -75 | | -75 | |
| 20 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00} | -103 | 97 | 55 | | | | | 55 | 55 | | 55 | 55 |
| 21 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00} | -103 | 97 | 55 | | | | | 55 | 55 | | 55 | 55 |
| 22 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00} | 100 | -158 | 100 | | 100 | | 100 | | | 100 | | |
| 23 | ВИД СТЕНОК ВАГОНА-Двойные | 65 | 7 | | | | | 165 | | 165 | | | |
| 24 | ВИД СТЕНОК ВАГОНА-Одинарные | 42 | 7 | 19 | -22 | -22 | -22 | -81 | -81 | -81 | -81 | 19 | -81 |
| 25 | ВИД КРЫШИ ВАГОНА-Гофрированная | | 146 | | | | | | | 246 | | | |
| 26 | ВИД КРЫШИ ВАГОНА-Овальная | 129 | | | | | 287 | | | | | | |
| 27 | ВИД КРЫШИ ВАГОНА-Отсутствует | 38 | 29 | 46 | -113 | | 46 | | -113 | -113 | -113 | 46 | -13 |
| 28 | ВИД КРЫШИ ВАГОНА-Прямая | 72 | -45 | | 14 | 114 | | 114 | 14 | | 14 | | |
| 29 | ВИД КРЫШИ ВАГОНА-Треугольная | 187 | | | | | | | | | | | |
| 30 | ГРУЗ-ОТСУТСТВУЕТ-0001 | | 129 | | | | | | | 287 | | | |
| 31 | ГРУЗ-ТРЕУГОЛЬНИК_ПРЯМОЙ-0001 | 87 | -71 | -13 | -13 | -13 | 87 | -13 | -13 | -13 | | | |
| 32 | ГРУЗ-ТРЕУГОЛЬНИК_ПЕРЕВЕРНУТЫЙ-0001 | 129 | | | | 287 | | | | | | | |
| 33 | ГРУЗ-РОМБ-0001 | 129 | | 187 | | | 187 | | | | | | |
| 34 | ГРУЗ-ОВАЛ-0001 | 7 | 48 | 7 | | 7 | | 7 | | 7 | 7 | 107 | |
| 35 | ГРУЗ-ОВАЛ-0002 | 129 | | | 287 | | | | | | | | |
| 36 | ГРУЗ-ОВАЛ-0003 | | 129 | | | | | | 287 | | | | |
| 37 | ГРУЗ-КВАДРАТ-0001 | 129 | | | | | 287 | | | | | | |
| 38 | ГРУЗ-КВАДРАТ-0003 | 129 | | 287 | | | | | | | | | |
| 39 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0001 | -30 | 70 | | 129 | | | | | | | 129 | 129 |
| 40 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0002 | | 129 | | | | | | | | | | 287 |
| 41 | ГРУЗ-ПРЯМОУГОЛЬНИК_ДЛИННЫЙ-0001 | -30 | 70 | | | | | 129 | | | 129 | 129 | |

**Таблица 13 – МАТРИЦА ЗНАНИЙ СИМ-3 (в 0.01 исходных ед.изм.)**

| Код | Наименование | Классы | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | КОЛИЧЕСТВО ВАГОНОВ-2 | -338 | 339 | -40 | -30 | -30 | -40 | -30 | 78 | -30 | 81 | -39 | 79 |
| 2 | КОЛИЧЕСТВО ВАГОНОВ-3 | 123 | -121 | -80 | 40 | 40 | 20 | 40 | -44 | 40 | -39 | 22 | -41 |
| 3 | КОЛИЧЕСТВО ВАГОНОВ-4 | 87 | -87 | 87 | -10 | -10 | -13 | -10 | -7 | -10 | -6 | -13 | -7 |
| 4 | СУММАРНЫЙ ВЕС ГРУЗА: 1/5-{11.00, 21.00} | -477 | 479 | -80 | 40 | -60 | -80 | -60 | 56 | 40 | 61 | 22 | 59 |
| 5 | СУММАРНЫЙ ВЕС ГРУЗА: 2/5-{21.00, 31.00} | 262 | -261 | -40 | 70 | 70 | -40 | 70 | -22 | -30 | -19 | -39 | -21 |
| 6 | СУММАРНЫЙ ВЕС ГРУЗА: 3/5-{31.00, 41.00} | | | | | | | | | | | | |
| 7 | СУММАРНЫЙ ВЕС ГРУЗА: 4/5-{41.00, 51.00} | 87 | -87 | -13 | -10 | -10 | 87 | -10 | -7 | -10 | -6 | -13 | -7 |
| 8 | СУММАРНЫЙ ВЕС ГРУЗА: 5/5-{51.00, 61.00} | 87 | -87 | 87 | -10 | -10 | -13 | -10 | -7 | -10 | -6 | -13 | -7 |
| 9 | ФОРМА ВАГОНА-U-образная | -164 | 166 | -67 | 50 | -50 | 33 | -50 | -37 | 50 | 68 | -65 | 65 |
| 10 | ФОРМА ВАГОНА-V-образная | -163 | 197 | -31 | -23 | -23 | -31 | -23 | -17 | -23 | -15 | 170 | -16 |
| 11 | ФОРМА ВАГОНА-Овальная | 25 | -58 | -9 | -7 | -7 | 91 | -7 | -5 | -7 | -4 | -9 | -5 |
| 12 | ФОРМА ВАГОНА-Прямоугольная | 81 | -109 | 136 | -96 | 4 | -64 | 104 | 55 | 4 | -28 | -55 | -36 |
| 13 | ФОРМА ВАГОНА-Ромбовидная | 25 | -58 | -9 | -7 | 93 | -9 | -7 | -5 | -7 | -4 | -9 | -5 |
| 14 | ДЛИНА ВАГОНА-Длинный | -252 | 321 | 11 | -66 | 34 | -89 | -66 | -49 | 34 | 57 | 14 | 54 |
| 15 | ДЛИНА ВАГОНА-Короткий | 407 | -435 | -37 | 124 | 24 | 63 | 24 | -30 | -76 | -15 | 71 | -122 |
| 16 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-2 | -108 | 115 | -22 | 61 | -139 | 78 | -39 | 24 | 61 | -56 | -11 | 34 |
| 17 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-3 | 224 | -190 | 55 | -33 | 67 | -45 | 67 | -24 | -33 | -22 | -43 | -23 |
| 18 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00} | 231 | -225 | -81 | 91 | -9 | 119 | -9 | -54 | -9 | -36 | 28 | -45 |
| 19 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00} | 409 | -472 | 35 | -23 | 77 | 135 | 77 | 9 | -23 | -80 | -60 | -85 |
| 20 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00} | -364 | 366 | 33 | -50 | -50 | -67 | -50 | 63 | 50 | -32 | 35 | 65 |
| 21 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00} | -364 | 366 | 33 | -50 | -50 | -67 | -50 | 63 | 50 | -32 | 35 | 65 |
| 22 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00} | 186 | -219 | 51 | -36 | 64 | -49 | 64 | -27 | -36 | 76 | -48 | -25 |
| 23 | ВИД СТЕНОК ВАГОНА-Двойные | 37 | -3 | -31 | -23 | -23 | -31 | 77 | -17 | 77 | -15 | -30 | -16 |
| 24 | ВИД СТЕНОК ВАГОНА-Одинарные | -96 | -30 | 56 | 45 | 45 | -44 | -55 | 11 | -55 | 34 | 67 | 22 |
| 25 | ВИД КРЫШИ ВАГОНА-Гофрированная | -150 | 184 | -18 | -13 | -13 | -18 | -13 | -10 | 87 | -9 | -17 | -9 |
| 26 | ВИД КРЫШИ ВАГОНА-Овальная | 87 | -87 | -13 | -10 | -10 | 87 | -10 | -7 | -10 | -6 | -13 | -7 |
| 27 | ВИД КРЫШИ ВАГОНА-Отсутствует | -105 | 210 | 86 | -59 | -159 | 86 | -159 | -18 | -59 | -4 | 93 | 89 |
| 28 | ВИД КРЫШИ ВАГОНА-Прямая | 148 | -179 | -89 | 34 | 134 | -89 | 134 | 51 | -66 | 57 | -86 | -46 |
| 29 | ВИД КРЫШИ ВАГОНА-Треугольная | 62 | -29 | -4 | -3 | -3 | -4 | -3 | -2 | -3 | -2 | -4 | -2 |
| 30 | ГРУЗ-ОТСУТСТВУЕТ-0001 | -113 | 113 | -13 | -10 | -10 | -13 | -10 | -7 | 90 | -6 | -13 | -7 |
| 31 | ГРУЗ-ТРЕУГОЛЬНИК_ПРЯМОЙ-0001 | 297 | -295 | -7 | 20 | 20 | 93 | 20 | 41 | 20 | -52 | -104 | -55 |
| 32 | ГРУЗ-ТРЕУГОЛЬНИК_ПЕРЕВЕРНУТЫЙ-0001 | 87 | -87 | -13 | -10 | 90 | -13 | -10 | -7 | -10 | -6 | -13 | -7 |
| 33 | ГРУЗ-РОМБ-0001 | 174 | -174 | 73 | -20 | -20 | 73 | -20 | -15 | -20 | -13 | -26 | -14 |
| 34 | ГРУЗ-ОВАЛ-0001 | -190 | 192 | 6 | -70 | 30 | -94 | 30 | -51 | 30 | 55 | 109 | -48 |
| 35 | ГРУЗ-ОВАЛ-0002 | 87 | -87 | -13 | 90 | -10 | -13 | -10 | -7 | -10 | -6 | -13 | -7 |
| 36 | ГРУЗ-ОВАЛ-0003 | -113 | 113 | -13 | -10 | -10 | -13 | -10 | 93 | -10 | -6 | -13 | -7 |
| 37 | ГРУЗ-КВАДРАТ-0001 | 87 | -87 | -13 | -10 | -10 | 87 | -10 | -7 | -10 | -6 | -13 | -7 |
| 38 | ГРУЗ-КВАДРАТ-0003 | 87 | -87 | 87 | -10 | -10 | -13 | -10 | -7 | -10 | -6 | -13 | -7 |
| 39 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0001 | -138 | 139 | -40 | 70 | -30 | -40 | -30 | -22 | -30 | -19 | 61 | 79 |
| 40 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0002 | -113 | 113 | -13 | -10 | -10 | -13 | -10 | -7 | -10 | -6 | -13 | 93 |
| 41 | ГРУЗ-ПРЯМОУГОЛЬНИК_ДЛИННЫЙ-0001 | -138 | 139 | -40 | -30 | -30 | -40 | 70 | -22 | -30 | 81 | 61 | -21 |

**Таблица 14 – МАТРИЦА ЗНАНИЙ СИМ-4 (в 0.01 исходных ед.изм.)**

| Код | Наименование | Классы | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | КОЛИЧЕСТВО ВАГОНОВ-2 | -100 | 130 | -100 | -100 | -100 | -100 | -100 | 354 | -100 | 414 | -100 | 382 |
| 2 | КОЛИЧЕСТВО ВАГОНОВ-3 | 18 | -23 | -100 | 68 | 68 | 24 | 68 | -100 | 68 | -100 | 29 | -100 |
| 3 | КОЛИЧЕСТВО ВАГОНОВ-4 | 77 | -100 | 646 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 |
| 4 | СУММАРНЫЙ ВЕС ГРУЗА: 1/5-{11.00, 21.00} | -70 | 92 | -100 | 68 | -100 | -100 | -100 | 127 | 68 | 157 | 29 | 141 |
| 5 | СУММАРНЫЙ ВЕС ГРУЗА: 2/5-{21.00, 31.00} | 77 | -100 | -100 | 235 | 235 | -100 | 235 | -100 | -100 | -100 | -100 | -100 |
| 6 | СУММАРНЫЙ ВЕС ГРУЗА: 3/5-{31.00, 41.00} | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | СУММАРНЫЙ ВЕС ГРУЗА: 4/5-{41.00, 51.00} | 77 | -100 | -100 | -100 | -100 | 646 | -100 | -100 | -100 | -100 | -100 | -100 |
| 8 | СУММАРНЫЙ ВЕС ГРУЗА: 5/5-{51.00, 61.00} | 77 | -100 | 646 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 |
| 9 | ФОРМА ВАГОНА-U-образная | -29 | 38 | -100 | 101 | -100 | 49 | -100 | -100 | 101 | 208 | -100 | 189 |
| 10 | ФОРМА ВАГОНА-V-образная | -62 | 97 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | 561 | -100 |
| 11 | ФОРМА ВАГОНА-Овальная | 33 | -100 | -100 | -100 | -100 | 1019 | -100 | -100 | -100 | -100 | -100 | -100 |
| 12 | ФОРМА ВАГОНА-Прямоугольная | 4 | -6 | 52 | -49 | 2 | -24 | 53 | 38 | 2 | -22 | -22 | -26 |
| 13 | ФОРМА ВАГОНА-Ромбовидная | 33 | -100 | -100 | -100 | 1409 | -100 | -100 | -100 | -100 | -100 | -100 | -100 |
| 14 | ДЛИНА ВАГОНА-Длинный | -34 | 55 | 12 | -100 | 51 | -100 | -100 | -100 | 51 | 131 | 16 | 117 |
| 15 | ДЛИНА ВАГОНА-Короткий | 20 | -28 | -16 | 71 | 14 | 27 | 14 | -23 | -43 | -13 | 31 | -100 |
| 16 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-2 | -4 | 6 | -7 | 26 | -58 | 24 | -16 | 13 | 26 | -36 | -4 | 20 |
| 17 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-3 | 60 | -65 | 124 | -100 | 202 | -100 | 202 | -100 | -100 | -100 | -100 | -100 |
| 18 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00} | 10 | -12 | -29 | 44 | -4 | 42 | -4 | -35 | -4 | -27 | 10 | -31 |
| 19 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00} | 29 | -44 | 21 | -18 | 63 | 82 | 63 | 10 | -18 | -100 | -37 | -100 |
| 20 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00} | -65 | 84 | 49 | -100 | -100 | -100 | -100 | 172 | 101 | -100 | 54 | 189 |
| 21 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00} | -65 | 84 | 49 | -100 | -100 | -100 | -100 | 172 | 101 | -100 | 54 | 189 |
| 22 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00} | 45 | -69 | 104 | -100 | 174 | -100 | 174 | -100 | -100 | 321 | -100 | -100 |
| 23 | ВИД СТЕНОК ВАГОНА-Двойные | 14 | -1 | -100 | -100 | -100 | -100 | 331 | -100 | 331 | -100 | -100 | -100 |
| 24 | ВИД СТЕНОК ВАГОНА-Одинарные | -3 | -1 | 16 | 18 | 18 | -13 | -22 | 6 | -22 | 20 | 20 | 13 |
| 25 | ВИД КРЫШИ ВАГОНА-Гофрированная | -100 | 159 | -100 | -100 | -100 | -100 | -100 | -100 | 654 | -100 | -100 | -100 |
| 26 | ВИД КРЫШИ ВАГОНА-Овальная | 77 | -100 | -100 | -100 | -100 | 646 | -100 | -100 | -100 | -100 | -100 | -100 |
| 27 | ВИД КРЫШИ ВАГОНА-Отсутствует | -6 | 15 | 40 | -37 | -100 | 40 | -100 | -15 | -37 | -4 | 45 | 81 |
| 28 | ВИД КРЫШИ ВАГОНА-Прямая | 20 | -31 | -100 | 51 | 202 | -100 | 202 | 104 | -100 | 131 | -100 | -100 |
| 29 | ВИД КРЫШИ ВАГОНА-Треугольная | 166 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 |
| 30 | ГРУЗ-ОТСУТСТВУЕТ-0001 | -100 | 130 | -100 | -100 | -100 | -100 | -100 | -100 | 906 | -100 | -100 | -100 |
| 31 | ГРУЗ-ТРЕУГОЛЬНИК_ПРЯМОЙ-0001 | 33 | -42 | -7 | 26 | 26 | 87 | 26 | 70 | 26 | -100 | -100 | -100 |
| 32 | ГРУЗ-ТРЕУГОЛЬНИК_ПЕРЕВЕРНУТЫЙ-0001 | 77 | -100 | -100 | -100 | 906 | -100 | -100 | -100 | -100 | -100 | -100 | -100 |
| 33 | ГРУЗ-РОМБ-0001 | 77 | -100 | 273 | -100 | -100 | 273 | -100 | -100 | -100 | -100 | -100 | -100 |
| 34 | ГРУЗ-ОВАЛ-0001 | -24 | 32 | 7 | -100 | 44 | -100 | 44 | -100 | 44 | 120 | 120 | -100 |
| 35 | ГРУЗ-ОВАЛ-0002 | 77 | -100 | -100 | 906 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 |
| 36 | ГРУЗ-ОВАЛ-0003 | -100 | 130 | -100 | -100 | -100 | -100 | -100 | 1261 | -100 | -100 | -100 | -100 |
| 37 | ГРУЗ-КВАДРАТ-0001 | 77 | -100 | -100 | -100 | -100 | 646 | -100 | -100 | -100 | -100 | -100 | -100 |
| 38 | ГРУЗ-КВАДРАТ-0003 | 77 | -100 | 646 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 |
| 39 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0001 | -41 | 53 | -100 | 235 | -100 | -100 | -100 | -100 | -100 | -100 | 157 | 382 |
| 40 | ГРУЗ-ПРЯМОУГОЛЬНИК_КОРОТКИЙ-0002 | -100 | 130 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | -100 | 1346 |
| 41 | ГРУЗ-ПРЯМОУГОЛЬНИК_ДЛИННЫЙ-0001 | -41 | 53 | -100 | -100 | -100 | -100 | 235 | -100 | -100 | 414 | 157 | -100 |

**Рассмотрим формальные модели объектов в исходных данных, матрице абсолютных частот (сопряженности), а также в базах знаний.**

*1. В исходной выборке* объект (мы выбрали «Состав-1») представляется в виде:

```
          ОПИСАНИЕ ОБЪЕКТА обучающей выборки N°43
12-09-11  06:29:53                                              г.Краснодар
=============================================================================
|  Код |            Наименования классов  распознавания                    |
=============================================================================
|    1 | СОСТАВ СЛЕДУЕТ НА-ВОСТОК
|    3 | НАЗВАНИЕ СОСТАВА-Состав-01
=============================================================================
|  Код |            С о д е р ж а н и е   в о п р о с а                    |
=============================================================================
|    3 | КОЛИЧЕСТВО ВАГОНОВ-4
|    8 | СУММАРНЫЙ ВЕС ГРУЗА: 5/5-{51.00, 61.00}
|   12 | ФОРМА ВАГОНА-Прямоугольная
|   15 | ДЛИНА ВАГОНА-Короткий
|   16 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-2
|   18 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00}
|   19 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00}
|   24 | ВИД СТЕНОК ВАГОНА-Одинарные
|   27 | ВИД КРЫШИ ВАГОНА-Отсутствует
|   34 | ГРУЗ-ОВАЛ-0001
|   12 | ФОРМА ВАГОНА-Прямоугольная
|   17 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-3
|   22 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 5/5-{70.00, 80.00}
|   24 | ВИД СТЕНОК ВАГОНА-Одинарные
|   27 | ВИД КРЫШИ ВАГОНА-Отсутствует
|   33 | ГРУЗ-РОМБ-0001
|   12 | ФОРМА ВАГОНА-Прямоугольная
|   15 | ДЛИНА ВАГОНА-Короткий
|   16 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-2
|   18 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 1/5-{30.00, 40.00}
|   19 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 2/5-{40.00, 50.00}
|   24 | ВИД СТЕНОК ВАГОНА-Одинарные
|   31 | ГРУЗ-ТРЕУГОЛЬНИК_ПРЯМОЙ-0001
|   12 | ФОРМА ВАГОНА-Прямоугольная
|   14 | ДЛИНА ВАГОНА-Длинный
|   16 | КОЛИЧЕСТВО ОСЕЙ ВАГОНА-2
|   20 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 3/5-{50.00, 60.00}
|   21 | ГРУЗОПОДЪЕМНОСТЬ ВАГОНА: 4/5-{60.00, 70.00}
|   24 | ВИД СТЕНОК ВАГОНА-Одинарные
|   27 | ВИД КРЫШИ ВАГОНА-Отсутствует
|   38 | ГРУЗ-КВАДРАТ-0003
=============================================================================
Универсальная когнитивная аналитическая система                 НПП *ЭЙДОС*
```

Как видно, это описание включает признаки различной природы, измеряемые в различных единицах измерения и в шкалах различного типа: номинальных (текстовых), порядковых и числовых, причем последние системой «Эйдос» представлены в виде шкал интервальных числовых значений.

*2. В матрице абсолютных частот* столбцы соответствуют классам. Но мы можем приводить пример с данным объектом (для удобства он выделен на светло-желтом фоне), т.к. ему соответствует монообъектный класс. В матрице абсолютных частот (таблица 10) этот же объект представлен в виде вектора с элементами, соответствующими признакам (градациям описательных шкал), имеющими значения:

0 – если, признак не встречается у объекта;

n – если, признак встречается n раз.

*3. В матрицах знаний* (таблицы 11-14) этот же объект представлен в виде вектора с элементами, соответствующими признакам (градациям описательных шкал), имеющими значения количества знаний о принадлежности или непринадлежности объекта к классу, если у него есть определенный признак. В матрице знаний СИМ-1 это количество знаний представлено в сантибитах (сотых долях бита), если оно больше 0, то это знания о принадлежности, если меньше – то о непринадлежности.

Мы видим, насколько *существенно* отличаются модели объекта в классических методах кластерного анализа, которые оперируют 1-й и 2-й формами представления, и в АСК-анализе и системе «Эйдос», оперирующей 3-й формой представления, основанной на базах знаний с различными частными критериями для расчета количества знаний. Соответственно различаются и результаты кластерного анализа в классических методах и методе когнитивной кластеризации. Кроме того, эти формы представления объектов порождает или позволяет решить ряд проблем кластерного анализа.

Модели СИМ-1, СИМ-2, СИМ-3 и СИМ-4, отличающиеся частными критериями, имеют различную ***достоверность*** при использовании различных интегральных критериев, которая рассчитывается, в частности, в режиме _25 системы «Эйдос» (таблица 15):

**Таблица 15 – ДОСТОВЕРНОСТЬ МОДЕЛЕЙ СИМ-1, СИМ-2, СИМ-3 И СИМ-4 С РАЗЛИЧНЫМИ ИНТЕГРАЛЬНЫМИ КРИТЕРИЯМИ**

| Тип модели | Интегральный критерий | Дата и время расчета | | Достоверность идентификации | Достоверность неидентификации | Средняя достоверность |
|---|---|---|---|---|---|---|
| СИМ-1 | Корреляция | 11-09-11 | 14:32:20 | 70,727 | 93,400 | 82,064 |
| СИМ-1 | Свертка | 11-09-11 | 14:32:23 | 100,000 | 37,906 | 68,953 |
| СИМ-2 | Корреляция | 11-09-11 | 14:32:28 | 81,701 | 98,344 | 90,023 |
| СИМ-2 | Свертка | 11-09-11 | 14:32:31 | 97,558 | 56,038 | 76,798 |
| СИМ-3 | Корреляция | 11-09-11 | 14:32:37 | 100,000 | 72,525 | 86,263 |
| СИМ-3 | Свертка | 11-09-11 | 14:32:39 | 100,000 | 77,469 | 88,734 |
| СИМ-4 | Корреляция | 11-09-11 | 14:32:45 | 100,000 | 71,981 | 85,991 |
| СИМ-4 | Свертка | 11-09-11 | 14:32:47 | 100,000 | 84,613 | 92,306 |

На разных исходных данных преимущества по достоверности имеют различные модели.

***Неортонормированность*** когнитивного пространства признаков в рассматриваемом численном примере *подтверждается* тем, что корреляционная матрица признаков *не является диагональной матрицей*, т.к. в *ней есть не нулевые корреляции между разными признаками*, а не только каждого признака с самим собой (на диагонали) (таблица 16). Когнитивное пространство классов также неортнормированно:

**Таблица 16 – МАТРИЦА СХОДСТВА КЛАССОВ**

| KOD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 100,00 | -54,69 | 27,08 | 2,95 | 24,24 | 31,94 | 10,48 | -31,49 | -26,22 | -28,30 | -60,81 | -46,20 |
| 2 | -54,69 | 100,00 | -19,88 | -2,23 | -34,99 | -15,79 | -44,68 | 37,65 | 42,73 | 8,70 | 34,55 | 50,12 |
| 3 | 27,08 | -19,88 | 100,00 | -21,97 | -5,99 | -9,80 | -3,49 | -7,59 | -12,96 | -7,12 | -14,14 | -11,41 |
| 4 | 2,95 | -2,23 | -21,97 | 99,99 | 1,09 | -10,87 | 3,68 | -9,10 | -4,32 | 1,24 | 8,10 | 9,25 |
| 5 | 24,24 | -34,99 | -5,99 | 1,09 | 100,00 | -17,31 | 30,65 | -9,14 | -14,39 | 9,38 | -13,25 | -17,43 |
| 6 | 31,94 | -15,79 | -9,80 | -10,87 | -17,31 | 100,00 | -19,16 | -16,57 | -16,24 | -16,72 | -18,00 | -16,76 |
| 7 | 10,48 | -44,68 | -3,49 | 3,68 | 30,65 | -19,16 | 100,00 | -7,04 | 5,96 | 31,48 | -1,81 | -23,80 |
| 8 | -31,49 | 37,65 | -7,59 | -9,10 | -9,14 | -16,57 | -7,04 | 100,00 | 4,23 | 21,11 | -5,14 | 24,11 |
| 9 | -26,22 | 42,73 | -12,96 | -4,32 | -14,39 | -16,24 | 5,96 | 4,23 | 100,00 | 1,01 | -4,53 | 3,31 |
| 10 | -28,30 | 8,70 | -7,12 | 1,24 | 9,38 | -16,72 | 31,48 | 21,11 | 1,01 | 100,00 | 16,36 | 25,77 |
| 11 | -60,81 | 34,55 | -14,14 | 8,10 | -13,25 | -18,00 | -1,81 | -5,14 | -4,53 | 16,36 | 99,99 | 14,37 |
| 12 | -46,20 | 50,12 | -11,41 | 9,25 | -17,43 | -16,76 | -23,80 | 24,11 | 3,31 | 25,77 | 14,37 | 100,00 |

В системе «Эйдос» реализовано несколько итерационных алгоритмов ортонормирования как когнитивного пространства признаков, так и когнитивного пространства классов.

При ортнормировании когнитивного пространства признаков матрица сходства признаков приводятся к диагональному виду, т.к. из модели *удаляются* признаки, сходные друг с другом, а информативность оставшихся соответственно увеличивается. Например, если ввести в модель еще одну описательную шкалу, точно совпадающую с одной из уже в ней имеющихся, то это приведет к тому, что количество информации из градаций ранее имевшейся шкалы распределится поровну между ней и градациями новой шкалы. Это означает, что предложенные модели вычисления количества знаний, представленные в таблице 2, дают различное количество знаний в признаке о принадлежности объекта к классам в зависимости от того, присутствуют ли в модели признаки, сходные с ним по смыслу или нет, т.е. по сути, *эти меры учитывают степень неортонормированности когнитивного пространства*. Поэтому они и корректно работают в пространствах различной степени ортонормированности, т.е. в неортонормированных пространствах.

При ортнормировании когнитивного пространства классов корреляционная матрица классов приводятся *к диагональному виду*, т.к. из модели удаляются классы, сходные друг с другом, а информативность оставшихся соответственно увеличивается.

*При ортонормировании осуществляется максимальное уменьшение размерности пространства при минимальной потере информации в модели*. Художник изображает трехмерную сцену на двумерном холсте, т.е. понижает размерность пространства, сохраняя при этом наиболее существенную информацию. Чем выше талант художника, тем лучше ему это удается и тем легче ценителям искусства по двумерному изображению восстановить соответствующий трехмерный образ. Однако *необратимая* потеря информации при ортонормировании все же *неизбежна*. Например, мы уже *никогда* не узнаем, где прячется четвертый медвежонок на картине И.И.Шишкина «Утро в сосновом лесу» (рисунок 8):

**Рисунок 8. Картина И.И.Шишкина «Утро в сосновом лесу»**

и по этому поводу нам лишь остается строить по этому поводу различные гипотезы, между тем художнику это было точно известно.

Известен закон необходимого разнообразия, предложенный Уильямом Россом Эшби и играющий фундаментальную роль в кибернетике[18]. Смысл этого закона в том, что *для того, чтобы адекватное управление было возможным, необходимо чтобы управляющая система была сложнее объекта управления.* Можно предположить, что этот принцип выполняется потому, что *более сложная система отображает более простую, адекватно, без потери информации, однако более простая система отображает более простую неадекватно, с необратимой потерей информации*.

Поэтому в реальных исследованиях ортонормирование осуществлять не всегда целесообразно и его нет смысла проводить без необходимости.

[18] http://ru.wikipedia.org/wiki/Эшби,%20Уильям

**Таблица 17 – МАТРИЦА СХОДСТВА ПРИЗНАКОВ**

| KOD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 100 | -65 | -26 | 68 | -46 | 0 | -26 | -26 | 53 | -3 | -23 | -7 | -23 | 60 | -38 | 3 | -43 | -71 | -22 | 50 | 50 | -3 | -32 | 32 | -12 | -26 | 19 | 8 | -21 | -14 | -19 | -25 | -37 | 10 | -25 | 51 | -26 | -26 | 24 | 53 | 27 |
| 2 | -65 | 100 | -25 | -26 | 70 | 0 | 1 | -25 | -10 | -10 | 1 | -9 | 37 | -26 | 39 | -17 | 57 | 56 | 34 | -30 | -30 | 36 | 55 | -30 | 10 | 1 | -50 | 47 | -4 | 17 | 50 | 36 | -18 | 9 | 36 | -41 | 1 | -25 | -4 | -40 | -5 |
| 3 | -26 | -25 | 100 | -37 | -16 | 0 | -6 | 100 | -34 | -21 | -9 | 46 | -10 | -26 | -12 | -2 | 35 | -27 | 16 | -13 | -13 | 21 | -16 | 20 | -16 | -6 | 24 | -24 | 19 | -15 | -18 | -7 | 68 | -29 | -7 | -15 | -6 | 100 | -27 | -15 | -27 |
| 4 | 68 | -26 | -37 | 100 | -35 | 0 | -37 | -37 | 72 | 38 | -24 | -26 | -23 | 74 | -37 | 15 | -49 | -43 | -46 | 80 | 80 | -18 | -7 | 25 | 20 | -37 | -2 | 0 | -79 | 19 | -26 | -34 | -47 | 41 | -7 | 34 | -37 | -37 | 53 | 37 | 40 |
| 5 | -46 | 70 | -16 | -35 | 100 | 0 | -16 | -16 | -26 | -28 | -19 | -3 | 51 | -32 | 58 | -41 | 58 | 42 | 38 | -50 | -50 | 36 | 23 | -5 | -28 | -16 | -41 | 64 | 13 | -27 | 21 | 53 | -26 | -12 | 53 | -26 | -16 | -16 | -3 | -26 | -3 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | -26 | 1 | -6 | -37 | -16 | 0 | 100 | -6 | -8 | -21 | 99 | -20 | -10 | -35 | 29 | 26 | -7 | 54 | 49 | -35 | -35 | -12 | -16 | -35 | -16 | 100 | 24 | -24 | 19 | -15 | 53 | -7 | 68 | -35 | -7 | -15 | 100 | -6 | -27 | -15 | -27 |
| 8 | -26 | -25 | 100 | -37 | -16 | 0 | -6 | 100 | -34 | -21 | -9 | 46 | -10 | -26 | -12 | -2 | 35 | -27 | 16 | -13 | -13 | 21 | -16 | 20 | -16 | -6 | 24 | -24 | 19 | -15 | -18 | -7 | 68 | -29 | -7 | -15 | -6 | 100 | -27 | -15 | -27 |
| 9 | 53 | -10 | -34 | 72 | -26 | 0 | -8 | -34 | 100 | 1 | -1 | -60 | -27 | 77 | -16 | 20 | -42 | -13 | -28 | 38 | 38 | -0 | -1 | 14 | 23 | -8 | -3 | -10 | -45 | 24 | -18 | -33 | -27 | 25 | 13 | -22 | -8 | -34 | 51 | 48 | 29 |
| 10 | -3 | -10 | -21 | 38 | -28 | 0 | -21 | -21 | 1 | 100 | -13 | -20 | -12 | 23 | 5 | 4 | -35 | 6 | -60 | 44 | 44 | -32 | -13 | 28 | 6 | -21 | 33 | -27 | -53 | 3 | -40 | -19 | -26 | 67 | -19 | 2 | -21 | -21 | 47 | 2 | 46 |
| 11 | -23 | 1 | -9 | -24 | -19 | 0 | 99 | -9 | -1 | -13 | 100 | -21 | -10 | -27 | 27 | 26 | -9 | 52 | 48 | -23 | -23 | -13 | -15 | -34 | -14 | 99 | 27 | -22 | 3 | -13 | 51 | -9 | 66 | -28 | -9 | -13 | 99 | -9 | -20 | -13 | -20 |
| 12 | -7 | -9 | 46 | -26 | -3 | 0 | -20 | 46 | -60 | -20 | -21 | 100 | 7 | -25 | -49 | -23 | 48 | -52 | 36 | 7 | 7 | 26 | 39 | -35 | 5 | -20 | 4 | 26 | 8 | 6 | 5 | 8 | 19 | -7 | -58 | 34 | -20 | 46 | -63 | -25 | 1 |
| 13 | -23 | 37 | -10 | -23 | 51 | 0 | -10 | -10 | -27 | -12 | -10 | 7 | 100 | 10 | 15 | -80 | 49 | 5 | 38 | -22 | -22 | 34 | -15 | 25 | -14 | -10 | -6 | 49 | 1 | -13 | 10 | 99 | -14 | 8 | -10 | -13 | -10 | -10 | -20 | -13 | -20 |
| 14 | 60 | -26 | -26 | 74 | -32 | 0 | -35 | -26 | 77 | 23 | -27 | -25 | 10 | 100 | -41 | -26 | -28 | -49 | -30 | 53 | 53 | 10 | -7 | 29 | 23 | -35 | 22 | 1 | -56 | 22 | -48 | 2 | -40 | 54 | -33 | -14 | -35 | -26 | 38 | 53 | 43 |
| 15 | -38 | 39 | -12 | -37 | 58 | 0 | 29 | -12 | -16 | 5 | 27 | -49 | 15 | -41 | 100 | -8 | 29 | 66 | 26 | -53 | -53 | 17 | -32 | 30 | -70 | 29 | 11 | 21 | 19 | -69 | 30 | 18 | 11 | -15 | 59 | -37 | 29 | -12 | 32 | -11 | -3 |
| 16 | 3 | -17 | -2 | 15 | -41 | 0 | 26 | -2 | 20 | 4 | 26 | -23 | -80 | -26 | -8 | 100 | -54 | 18 | -35 | 33 | 33 | -67 | 11 | -34 | 30 | 26 | -3 | -66 | 1 | 30 | 11 | -78 | 18 | -44 | 27 | 20 | 26 | -2 | 34 | 25 | -33 |
| 17 | -43 | 57 | 35 | -49 | 58 | 0 | -7 | 35 | -42 | -35 | -9 | 48 | 49 | -28 | 29 | -54 | 100 | -1 | 74 | -39 | -39 | 78 | 29 | -4 | -39 | -7 | -1 | 65 | 14 | -34 | 41 | 50 | 20 | -5 | -7 | -31 | -7 | 35 | -35 | -31 | -3 |
| 18 | -71 | 56 | -27 | -43 | 42 | 0 | 54 | -27 | -13 | 6 | 52 | -52 | 5 | -49 | 66 | 18 | -1 | 100 | 9 | -55 | -55 | -18 | 6 | -24 | -1 | 54 | -24 | -12 | 19 | 0 | 32 | 8 | 19 | -12 | 55 | -45 | 54 | -27 | 5 | -38 | -17 |
| 19 | -22 | 34 | 16 | -46 | 38 | 0 | 49 | 16 | -28 | -60 | 48 | 36 | 38 | -30 | 26 | -35 | 74 | 9 | 100 | -41 | -41 | 60 | 12 | -26 | -44 | 49 | 10 | 54 | 17 | -40 | 67 | 40 | 47 | -34 | -17 | -13 | 49 | 16 | -50 | -21 | -17 |
| 20 | 50 | -30 | -13 | 80 | -50 | 0 | -35 | -13 | 38 | 44 | -23 | 7 | -22 | 53 | -53 | 33 | -39 | -55 | -41 | 100 | 100 | -40 | 2 | 9 | 31 | -35 | 21 | -24 | -72 | 30 | -22 | -32 | -28 | 17 | -32 | 46 | -35 | -13 | 50 | 49 | 6 |
| 21 | 50 | -30 | -13 | 80 | -50 | 0 | -35 | -13 | 38 | 44 | -23 | 7 | -22 | 53 | -53 | 33 | -39 | -55 | -41 | 100 | 100 | -40 | 2 | 9 | 31 | -35 | 21 | -24 | -72 | 30 | -22 | -32 | -28 | 17 | -32 | 46 | -35 | -13 | 50 | 49 | 6 |
| 22 | -3 | 36 | 21 | -18 | 36 | 0 | -12 | 21 | -0 | -32 | -13 | 26 | 34 | 10 | 17 | -67 | 78 | -18 | 60 | -40 | -40 | 100 | 16 | 16 | -42 | -12 | -7 | 76 | 4 | -36 | 29 | 34 | 6 | 29 | -12 | -33 | -12 | 21 | -37 | -33 | 39 |
| 23 | -32 | 55 | -16 | -7 | 23 | 0 | -16 | -16 | -1 | -13 | -15 | 39 | -15 | -7 | -32 | 11 | 29 | 6 | 12 | 2 | 2 | 16 | 100 | -80 | 57 | -16 | -41 | 21 | -7 | 60 | 14 | -16 | -23 | 16 | -16 | -18 | -16 | -16 | -26 | -18 | 19 |
| 24 | 32 | -30 | 20 | 25 | -5 | 0 | -35 | 20 | 14 | 28 | -34 | -35 | 25 | 29 | 30 | -34 | -4 | -24 | -26 | 9 | 9 | 16 | -80 | 100 | -53 | -35 | 25 | 6 | -12 | -54 | -28 | 23 | -10 | 19 | 23 | 3 | -35 | 20 | 42 | 15 | 7 |
| 25 | -12 | 10 | -16 | 20 | -28 | 0 | -16 | -16 | 23 | 6 | -14 | 5 | -14 | 23 | -70 | 30 | -39 | -1 | -44 | 31 | 31 | -42 | 57 | -53 | 100 | -16 | -43 | -40 | -13 | 100 | -24 | -16 | -22 | 14 | -16 | -1 | -16 | -16 | -12 | -1 | -12 |
| 26 | -26 | 1 | -6 | -37 | -16 | 0 | 100 | -6 | -8 | -21 | 99 | -20 | -10 | -35 | 29 | 26 | -7 | 54 | 49 | -35 | -35 | -12 | -16 | -35 | -16 | 100 | 24 | -24 | 19 | -15 | 53 | -7 | 68 | -35 | -7 | -15 | 100 | -6 | -27 | -15 | -27 |
| 27 | 19 | -50 | 24 | -2 | -41 | 0 | 24 | 24 | -3 | 33 | 27 | 4 | -6 | 22 | 11 | -3 | -1 | -24 | 10 | 21 | 21 | -7 | -41 | 25 | -43 | 24 | 100 | -34 | -11 | -45 | -27 | -7 | 37 | 3 | -52 | -17 | 24 | 24 | 29 | 54 | 9 |
| 28 | 8 | 47 | -24 | 0 | 64 | 0 | -24 | -24 | -10 | -27 | -22 | 26 | 49 | 1 | 21 | -66 | 65 | -12 | 54 | -24 | -24 | 76 | 21 | 6 | -40 | -24 | -34 | 100 | -10 | -36 | 40 | 47 | -34 | 23 | 3 | 10 | -24 | -24 | -29 | -35 | 36 |
| 29 | -21 | -4 | 19 | -79 | 13 | 0 | 19 | 19 | -45 | -53 | 3 | 8 | 1 | -56 | 19 | 1 | 14 | 19 | 17 | -72 | -72 | 4 | -7 | -12 | -13 | 19 | -11 | -10 | 100 | -12 | 15 | 16 | 19 | -48 | 16 | -12 | 19 | 19 | -42 | -12 | -42 |
| 30 | -14 | 17 | -15 | 19 | -27 | 0 | -15 | -15 | 24 | 3 | -13 | 6 | -13 | 22 | -69 | 30 | -34 | 0 | -40 | 30 | 30 | -36 | 60 | -54 | 100 | -15 | -45 | -36 | -12 | 100 | -17 | -15 | -21 | 14 | -15 | -4 | -15 | -15 | -13 | -4 | -13 |
| 31 | -19 | 50 | -18 | -26 | 21 | 0 | 53 | -18 | -18 | -40 | 51 | 5 | 10 | -48 | 30 | 11 | 41 | 32 | 67 | -22 | -22 | 29 | 14 | -28 | -24 | 53 | -27 | 40 | 15 | -17 | 100 | 12 | 25 | -35 | 12 | 17 | 53 | -18 | -31 | -37 | -30 |
| 32 | -25 | 36 | -7 | -34 | 53 | 0 | -7 | -7 | -33 | -19 | -9 | 8 | 99 | 2 | 18 | -78 | 50 | 8 | 40 | -32 | -32 | 34 | -16 | 23 | -16 | -7 | -7 | 47 | 16 | -15 | 12 | 100 | -12 | 1 | -7 | -15 | -7 | -7 | -26 | -14 | -26 |
| 33 | -37 | -18 | 68 | -47 | -26 | 0 | 68 | 68 | -27 | -26 | 66 | 19 | -14 | -40 | 11 | 18 | 20 | 19 | 47 | -28 | -28 | 6 | -23 | -10 | -22 | 68 | 37 | -34 | 19 | -21 | 25 | -12 | 100 | -43 | -12 | -21 | 68 | 68 | -36 | -21 | -36 |
| 34 | 10 | 9 | -29 | 41 | -12 | 0 | -35 | -29 | 25 | 67 | -28 | -7 | 8 | 54 | -15 | -44 | -5 | -12 | -34 | 17 | 17 | 29 | 16 | 19 | 14 | -35 | 3 | 23 | -48 | 14 | -35 | 1 | -43 | 100 | -34 | -20 | -35 | -29 | 8 | -20 | 84 |
| 35 | -25 | 36 | -7 | -7 | 53 | 0 | -7 | -7 | 13 | -19 | -9 | -58 | -10 | -33 | 59 | 27 | -7 | 55 | -17 | -32 | -32 | -12 | -16 | 23 | -16 | -7 | -52 | 3 | 16 | -15 | 12 | -7 | -12 | -34 | 100 | -15 | -7 | -7 | 36 | -14 | -26 |
| 36 | 51 | -41 | -15 | 34 | -26 | 0 | -15 | -15 | -22 | 2 | -13 | 34 | -13 | -14 | -37 | 20 | -31 | -45 | -13 | 46 | 46 | -33 | -18 | 3 | -1 | -15 | -17 | 10 | -12 | -4 | 17 | -15 | -21 | -20 | -15 | 100 | -15 | -15 | -13 | -5 | -13 |
| 37 | -26 | 1 | -6 | -37 | -16 | 0 | 100 | -6 | -8 | -21 | 99 | -20 | -10 | -35 | 29 | 26 | -7 | 54 | 49 | -35 | -35 | -12 | -16 | -35 | -16 | 100 | 24 | -24 | 19 | -15 | 53 | -7 | 68 | -35 | -7 | -15 | 100 | -6 | -27 | -15 | -27 |
| 38 | -26 | -25 | 100 | -37 | -16 | 0 | -6 | 100 | -34 | -21 | -9 | 46 | -10 | -26 | -12 | -2 | 35 | -27 | 16 | -13 | -13 | 21 | -16 | 20 | -16 | -6 | 24 | -24 | 19 | -15 | -18 | -7 | 68 | -29 | -7 | -15 | -6 | 100 | -27 | -15 | -27 |
| 39 | 24 | -4 | -27 | 53 | -3 | 0 | -27 | -27 | 51 | 47 | -20 | -63 | -20 | 38 | 32 | 34 | -35 | 5 | -50 | 50 | 50 | -37 | -26 | 42 | -12 | -27 | 29 | -29 | -42 | -13 | -31 | -26 | -36 | 8 | 36 | -13 | -27 | -27 | 100 | 65 | 5 |
| 40 | 53 | -40 | -15 | 37 | -26 | 0 | -15 | -15 | 48 | 2 | -13 | -25 | -13 | 53 | -11 | 25 | -31 | -38 | -21 | 49 | 49 | -33 | -18 | 15 | -1 | -15 | 54 | -35 | -12 | -4 | -37 | -14 | -21 | -20 | -14 | -5 | -15 | -15 | 65 | 100 | -13 |
| 41 | 27 | -5 | -27 | 40 | -3 | 0 | -27 | -27 | 29 | 46 | -20 | 1 | -20 | 43 | -3 | -33 | -3 | -17 | -17 | 6 | 6 | 39 | 19 | 7 | -12 | -27 | 9 | 36 | -42 | -13 | -30 | -26 | -36 | 84 | -26 | -13 | -27 | -27 | 5 | -13 | 100 |

На рисунке 9 приведен скриншот экранной формы режима _5126, обеспечивающей задание в диалоге параметров когнитивной кластеризации:

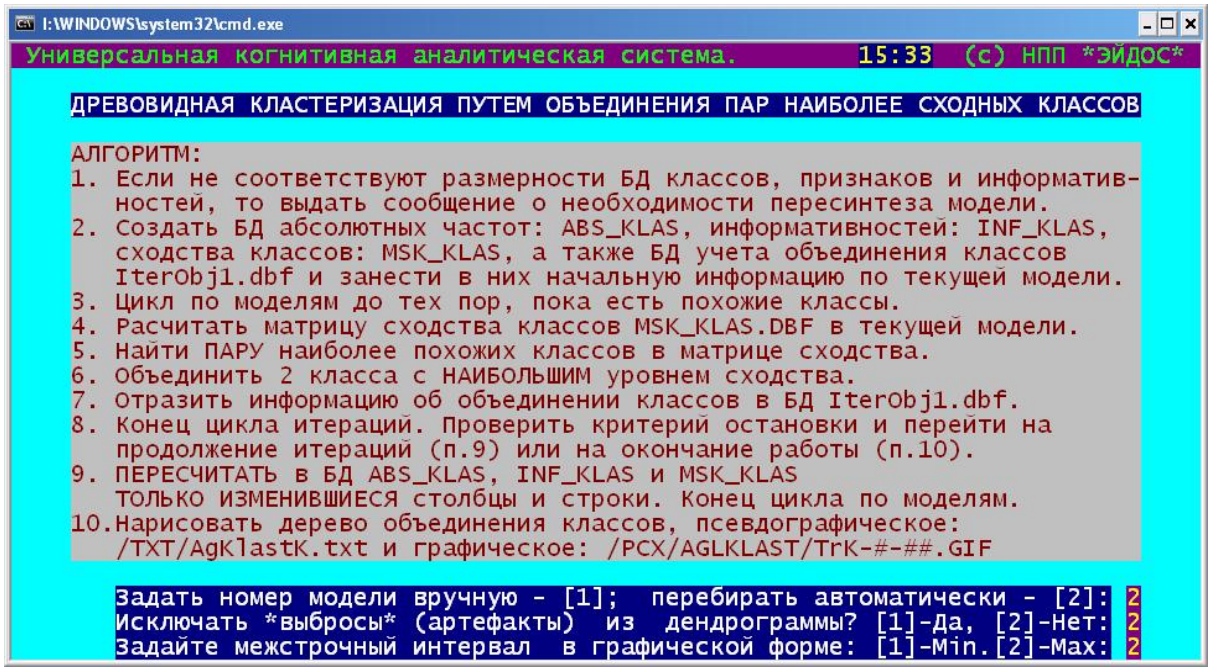
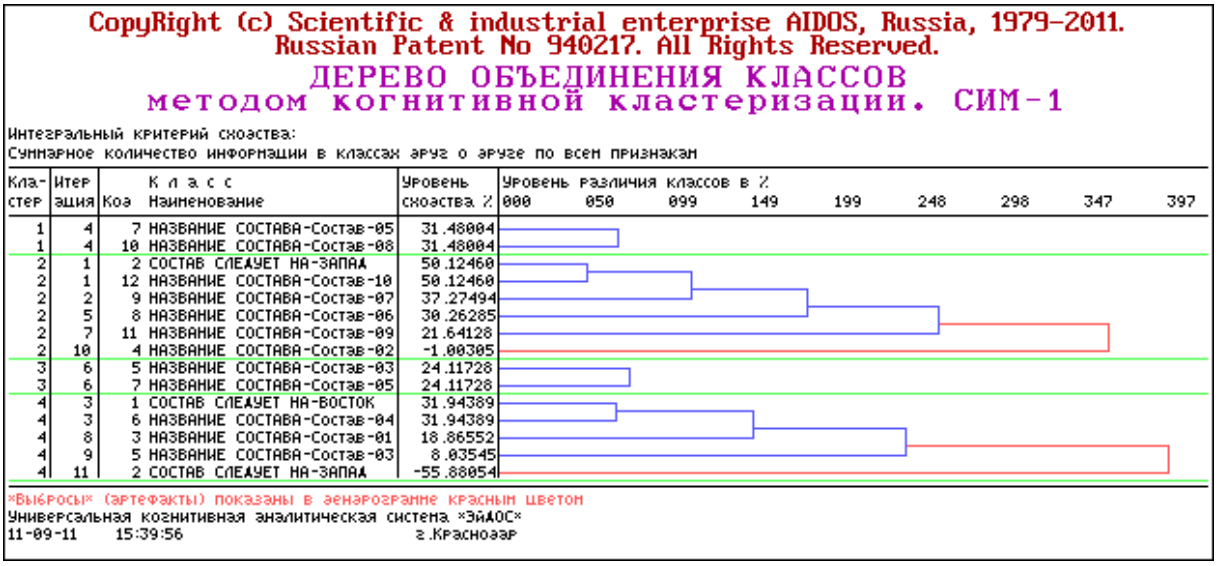
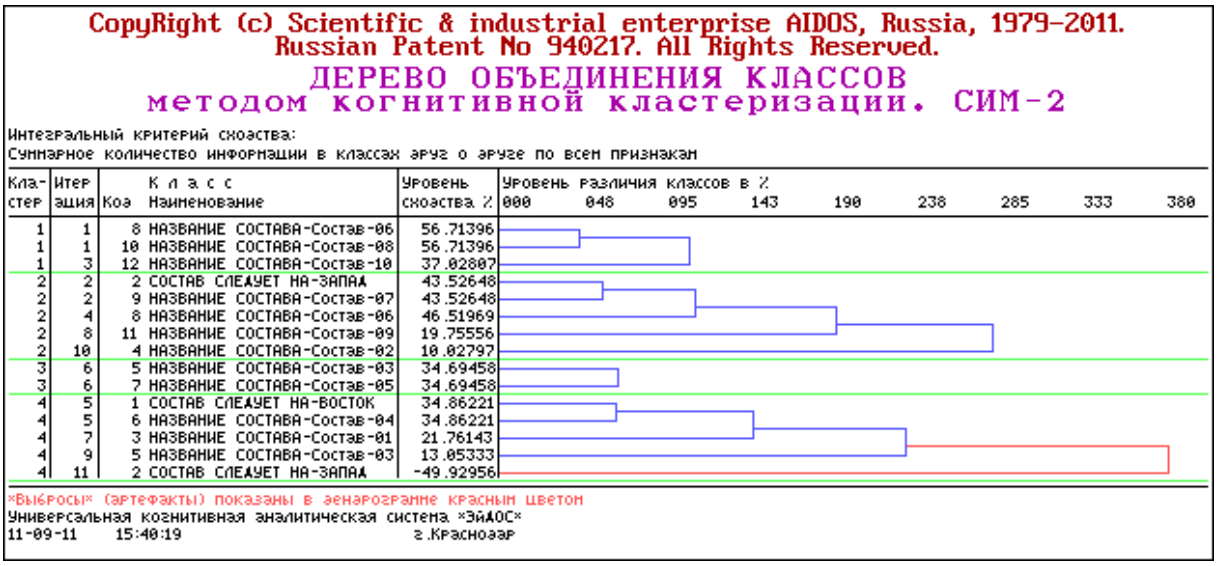



**Рисунок 9. Скриншот экранной формы режима _5126 системы «Эйдос», обеспечивающей задание в диалоге параметров когнитивной кластеризации**

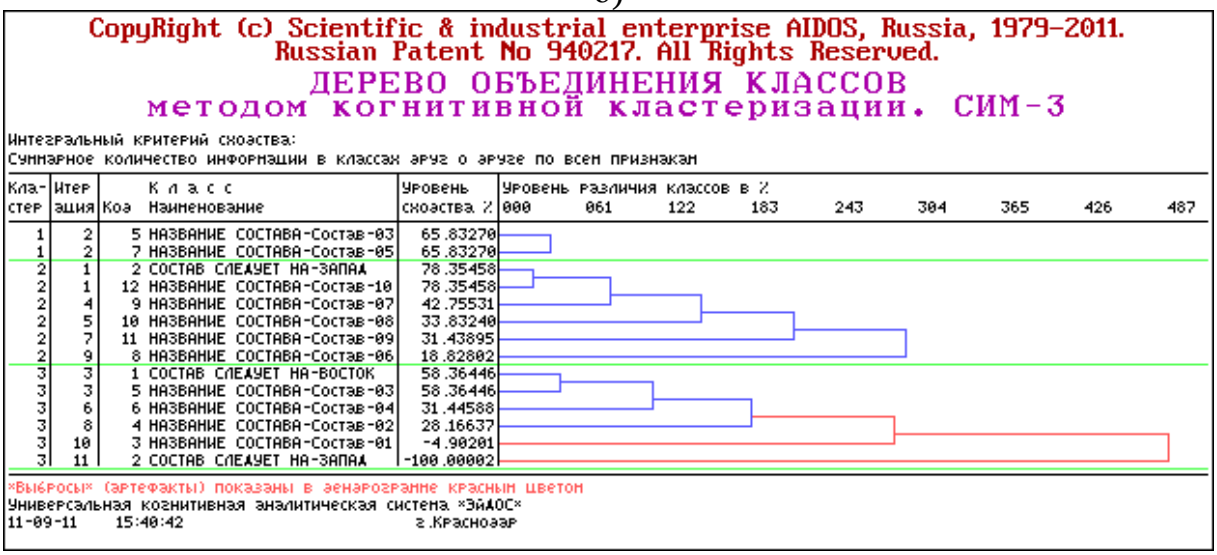
В результате работы данного режима формируются дендрограммы результатов когнитивной кластеризации и графики пошагового изменения межкластерного расстояния, приведенные на рисунках 10 а), б), в) и г):



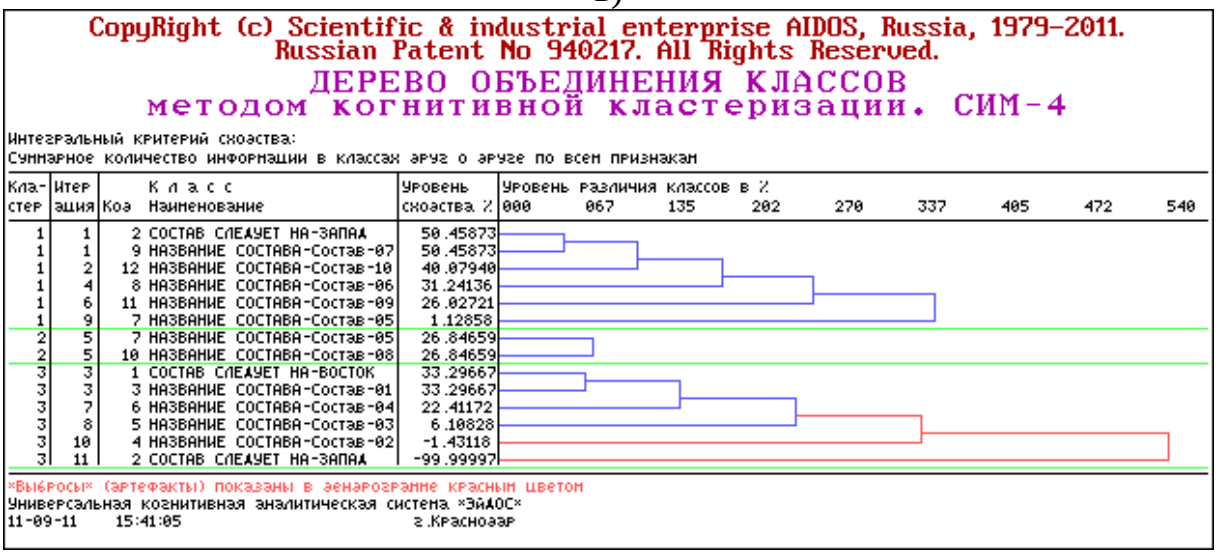

а)




б)




в)




г)

**Рисунок 10. Дендрограммы когнитивной кластеризации, полученные в режиме _5126 системы «Эйдос» на рассматриваемом численном примере**

Из рисунков 10 мы видим, что когнитивная кластеризация может начинаться как с монообъектных, так и с полиобъектных классов. Во втором случае классы создаются путем объединения объектов на основе априорной информации, источником которой является учитель (эксперт). Поэтому когнитивная кластеризация представляет собой сочетание обучения с учителем (экспертом) и без учителя, т.е. самообучения, причем учитель принимает участие лишь в формировании исходной модели для последующей кластеризации без учителя. На рисунке 11 приведены Графики пошагового изменения межкластерного расстояния при когнитивной кластеризации в моделях с разными частными критериями знаний, представленными в таблице 3:

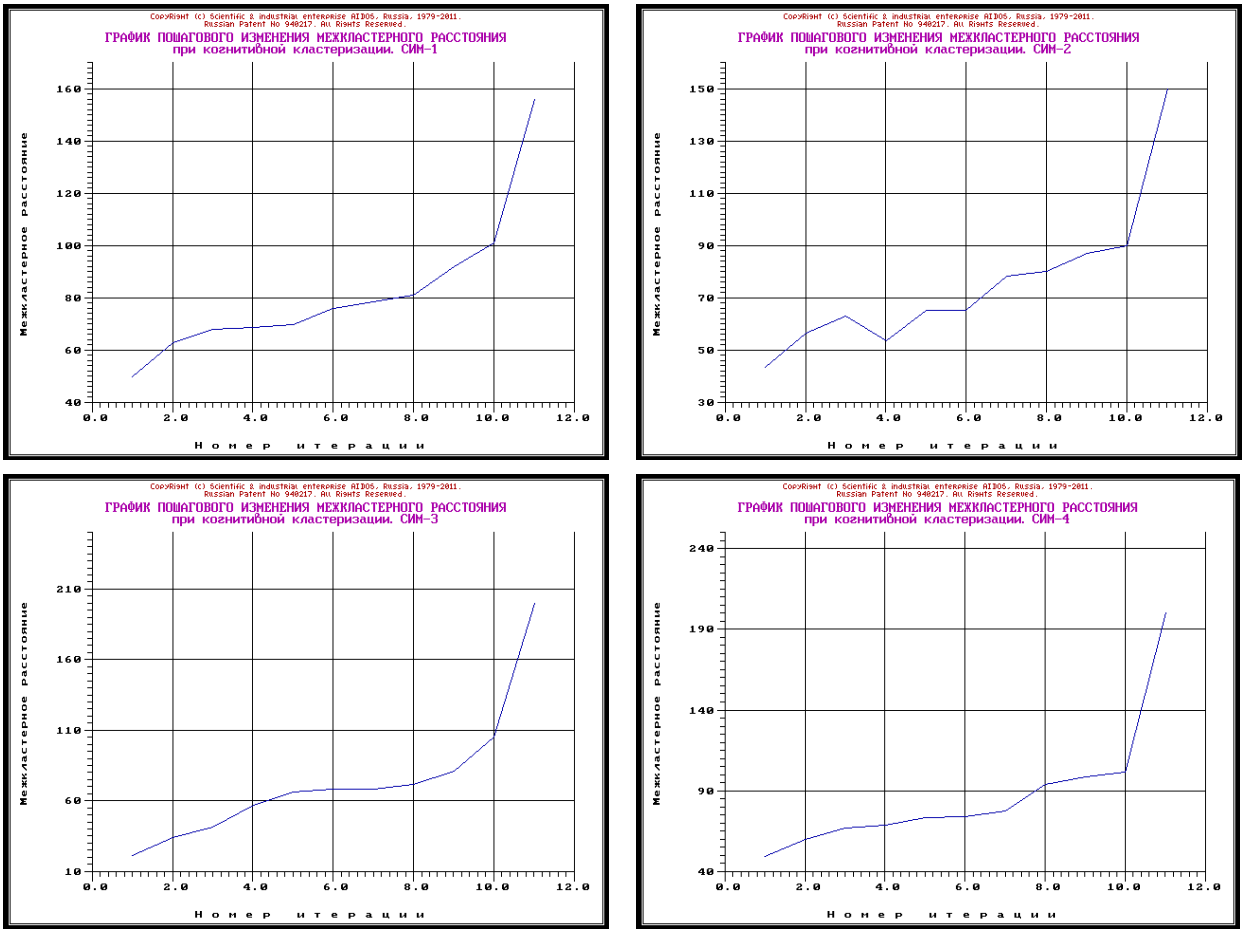



**Рисунок 11. Графики пошагового изменения межкластерного расстояния при когнитивной кластеризации, полученные в режиме _5126 системы «Эйдос» на рассматриваемом численном примере**

Таким образом, у нас есть возможность подвести итоги и кратко сформулировать, как решаются проблемы кластерного анализа в методе когнитивной кластеризации, основанном на АСК-анализе и реализованном в интеллектуальной системе «Эйдос» (таблица 18):

**Таблица 18 – ПРОБЛЕМЫ КЛАСТЕРИЗАЦИИ И ИХ РЕШЕНИЯ ПРЕДЛАГАЕМЫЕ В АСК-АНАЛИЗЕ И СИСТЕМЕ «ЭЙДОС»**

| № | Формулировка проблемы кластерного анализа | Предлагаемое в АСК-анализе и реализованное в системе «Эйдос» решение |
|---|---|---|
| 1. | *Проблема 1.1 выбора метрики, корректной для неортонормированных пространств.* | Предлагается применить неметрический интегральный критерий, представляющий собой суммарное количество информации в системе признаков о принадлежности объекта к классу («информационное расстояние»), никак не основанный на предположении об ортонормированности пространства и корректно работающий в неортонормированных пространствах |
| 2. | *Проблема 1.2 ортонормирования пространства.* | Поддерживается |
| 3. | *Проблема 2.1 сопоставимой обработки описаний объектов, описанных признаками различной природы, измеряемыми в различных единицах измерения (проблема размерностей).* | Объекты формально описываются в виде векторов, затем рассчитывается матрица абсолютных частот и на ее основе – матрица знаний, с использованием которой все признаки измеряются в одних единицах измерения: единицах измерения количества данных, информации и знаний – битах, байтах, и т.д. |
| | *Проблема 2.2 формализации описаний объектов, имеющих как количественные, так и качественные признаки.* | Числовые шкалы преобразуются в интервальные значения, после чего градации всех типов шкал обрабатываются единообразно (см.п.3) |
| 4. | *Проблема 3.1 доказательства гипотезы о нормальности исходных данных.* | Нет необходимости, т.к. предлагаемые частные и интегральные критерии не предполагают нормальности исходных данных |
| 5. | *Проблема 3.2 нормализации исходных данных.* | Реализованы режимы ремонта или взвешивания исходных данных. |
| 6. | *Проблема 3.3 применения непараметрических методов кластеризации, корректно работающих с ненормализованными данными.* | Предлагаемые методы являются непараметрическими и корректно работают с ненормализованными данными |
| 7. | *Проблема 4 разработки такого метода кластерного анализа, математическая модель и алгоритм и которого органично включали бы фильтр, подавляющий шум в исходных данных, в результате чего данный метод кластеризации корректно работал бы при наличии шума в исходных данных.* | Предлагаемый метод включает фильтр подавления шума на уровне формирования матрицы абсолютных частот и самой математической форме интегрального критерия. Кроме того, реализованы режимы удаления или корректной обработки артефактов, выбросов (нетипичных объектов) и малопредставленных данных, по которым нет достаточной статистики в исходных данных |
| 8. | *Проблема 5 разработки метода кластерного анализа, математическая модель и алгоритм и которого обеспечивали бы выявление «выбросов» (артефактов) в исходных данных и позволяли либо вообще не показывать их в дендрограммах, либо показывать, но так, чтобы было наглядно видно, что это артефакты.* | Поддерживается исключение выбросов и артефактов из дендрограмм, либо их отображение специальным для них образом. |

Отметим, что в АСК-анализе и системе «Эйдос» реализованы и другие методы кластеризации, *также основанные на знаниях*:

– дивизивная кластеризация (см., например: [23, 24]);

– кластерно-конструктивный анализ классов и признаков [9].

Дивизивная (разделительная, в отличие от агломартивной, т.е. объединяющей) кластеризация используется в системе «Эйдос» для того разделять классы на типичную и нетипичную части. Предполагается, что если объекты не были отнесены к классу, к которому они на самом деле отно-

сятся, то они являются нетипичными для него (исключениями), и это является достаточным основанием для того, чтобы создать для них новый класс с тем же наименованием и добавлением номера итерации. Такой подход приводит к резкому уменьшению ошибок неидентификации при примерно том же уровне ошибок ложной идентификации, что приводит к существенному улучшению достоверности модели (рисунок 12):

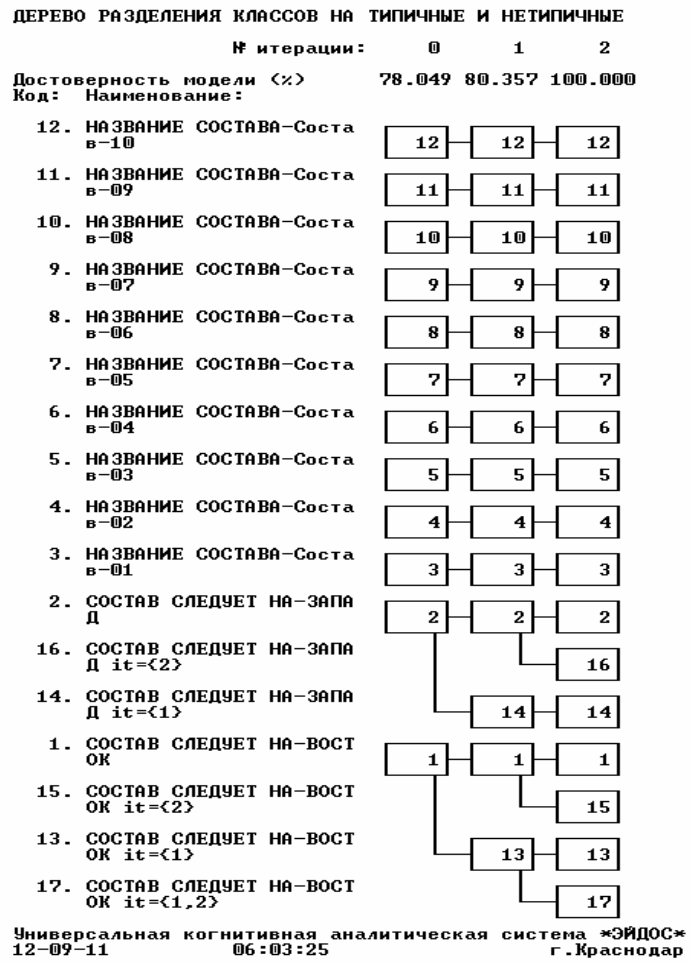



**Рисунок 12. Дендрограмма дивизивной кластеризации, полученная в режиме _34 системы «Эйдос» на рассматриваемом численном примере**

Конструкты представляют собой понятия, имеющие противоположные смысловые полюса, в качестве которых у нас выступают наиболее непохожие кластеры, а также спектр промежуточных по смыслу классов. Конструкты принадлежат к наивысшему иерархическому уровню процесса познания, выше которого только парадигма реальности (рисунок 1) и их можно рассматривать как оси координат нашего когнитивного пространства [9]. Система «Эйдос» формирует конструкты на основе исследования модели предметной области. Роль конструктов невозможно переоценить, т.к. когда мы познаем мы применяем уже имеющиеся у нас конструкты, уточняем или расширяем область их применения и создаем новые конструкты (таблица 19).

**Таблица 19 – КОНСТРУКТ: «ЗАПАД-ВОСТОК»**

| Код класса | Наименование класса | Уровень сходства |
|---|---|---|
| 2 | СОСТАВ СЛЕДУЕТ НА-ЗАПАД | 100,00 |
| 12 | НАЗВАНИЕ СОСТАВА-Состав-10 | 50,12 |
| 9 | НАЗВАНИЕ СОСТАВА-Состав-07 | 42,73 |
| 8 | НАЗВАНИЕ СОСТАВА-Состав-06 | 37,65 |
| 11 | НАЗВАНИЕ СОСТАВА-Состав-09 | 34,55 |
| 10 | НАЗВАНИЕ СОСТАВА-Состав-08 | 8,70 |
| 4 | НАЗВАНИЕ СОСТАВА-Состав-02 | -2,23 |
| 6 | НАЗВАНИЕ СОСТАВА-Состав-04 | -15,79 |
| 3 | НАЗВАНИЕ СОСТАВА-Состав-01 | -19,88 |
| 5 | НАЗВАНИЕ СОСТАВА-Состав-03 | -34,99 |
| 7 | НАЗВАНИЕ СОСТАВА-Состав-05 | -44,68 |
| 1 | СОСТАВ СЛЕДУЕТ НА-ВОСТОК | -54,69 |

Таким образом, в данной статье на небольшом численном примере рассматриваются новые алгоритмы и результаты агломеративной кластеризации, основные отличия которых от ранее известных стоят в том, что:

а) в них параметры обобщенного образа кластера не вычисляются как средние от исходных объектов (классов) или центры тяжести, а определяются с помощью той же самой базовой когнитивной операции АСК-анализа, которая применяется и для формирования обобщенных образов классов на основе примеров объектов и которая действительно обеспечивает обобщение;

б) в качестве критерия сходства используется не евклидово расстояние или его варианты, а интегральный критерий неметрической природы: «суммарное количество информации», применение которого теоретически корректно и дает хорошие результаты в неортонормированных пространствах, которые обычно и встречаются на практике;

в) кластерный анализ проводится не на основе исходных переменных или матрицы сопряженности, зависящих от единиц измерения по осям, а в когнитивном пространстве, в котором по всем осям (описательным шкалам) используется одна единица измерения: количество информации, и поэтому результаты кластеризации не зависят от исходных единиц измерения признаков объектов.

Имеется и ряд других менее существенных отличий. Все это позволяет получить результаты кластеризации, понятные специалистам и поддающиеся содержательной интерпретации, хорошо согласующиеся с оценками экспертов, их опытом и интуитивными ожиданиями, что часто представляет собой проблему для классических методов кластеризации. Описанные методы теоретически обоснованы в системно-когнитивном анализе (СК-анализ) и реализованы в его программном инструментарии – интеллектуальной системе «Эйдос»,

Основной **вывод**, который, по мнению авторов можно обоснованно сделать по материалам данной статьи, состоит в том, что, не смотря на существование огромного количества различных методов кластеризации, в этой области существует ряд нерешенных проблем, ждущих своего решения. Анализ этих проблем позволяет высказать **гипотезу**, что для их реше-

ния необходимо выйти за пределы понятийного поля чисто математических рассуждений и привлечь представления из области искусственного интеллекта, в частности основываться на четкой дефиниции содержания таких основополагающих понятий, как данные, информация и знания [8]. Данная статья и содержит описание авторского варианта реализации этой идеи. Здесь же хотелось бы отметить, что кластеризация классическим методом матрицы знаний, полученной вне статистической системы, реализующий кластерный анализ, не дает желаемых результатов, т.к. только 1-я итерация получается соответствующей предлагаемому подходу, а последующие дают ошибочные результаты, т.к. в статистических системах не реализовано операции обобщения и добавление объекта к кластеру или объединение классов в кластер осуществляется иначе, чем формирование самих классов в исходной матрице знаний.

Предлагаемый метод когнитивной кластеризации не лишен и некоторых ***недостатков и ограничений***, преодоление которых является одним из ***перспективных*** направлений развития этого метода.

Из *недостатков* следует прежде всего указать большие затраты вычислительных ресурсов и машинного времени на решение задачи кластеризации, чем у классических методов, обусловленные значительным объемом и более высокой сложностью вычислений. Другим недостатком является нежесткое ограничения текущей версии системы «Эйдос» на размерности модели, которые планируется преодолеть и которые постепенно преодолеваются. Версия системы «Эйдос» весны 2011 года обеспечивала объем обучающей выборки не более 100000 объектов, в текущей версии это ограничение снято и теперь система может работать с миллионами и даже десятками объектов. Но осталось ограничение на размерность баз знаний: не более 4000 классов и 4000 градаций факторов. Это ограничение также в перспективе планируется снять.

В качестве *перспективы* авторы рассматривают разработку *режимов, обеспечивающих*:

– когнитивную кластеризацию признаков;

– двухвходовую кластеризацию (одновременно и классов, и признаков), что оправдано тем, что при кластеризации классов изменяется и смысл признаков;

– моделей, основанных на новых частных критериях знаний (в частности, СИМ-5).

Материалы данной статьи могут быть использованы при разработке интеллектуальных систем, а также при проведении лабораторных работ по дисциплинам: «Интеллектуальные информационные системы» для специальности: 080801.65 – Прикладная информатика (по областям) и «Представление знаний в информационных системах» для специальности: 230201.65 – Информационные системы и технологии.

***Библиографический список***[19]


1. Мандель И.Д. Кластерный анализ. - М.: Финансы и статистика. 1988. – 176с.
2. Леонов В.П. Краткий обзор методов кластерного анализа. Сайт: http://www.biometrica.tomsk.ru/cluster_2.htm http://www.biometrica.tomsk.ru/cluster_3.htm
3. Леонов В.П. Литература и сайты по кластерному анализу. Сайт: http://www.biometrica.tomsk.ru/cluster_4.htm
4. Сайт Института Космических Исследований РАН: http://www.iki.rssi.ru/magbase/REFMAN/STATTEXT/modules/stcluan.html#general
5. Сайт Internet-сообщества закупщиков: http://zakup.vl.ru/132-metodi_klastern.html
6. Баран О.И., Григорьев Ю.А., Жилина Н.М. Алгоритмы и критерии качества кластеризации // Общественное здоровье и здравоохранение: материалы XLV науч.-практ. конф. с международным участием «Гигиена, организация здравоохранения и профпатология» и семинара «Актуальные вопросы современной профпатологии», Новокузнецк, 17-18 ноября 2010 / под ред. В.В.Захаренкова. Кемерово: Примула, 2010. – С. 21-26.
7. Мичи Д., Джонстон Р. Компьютер – творец. – М.: Мир, 1987. –251 с.
8. Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(70). С. 233 – 280. – Режим доступа: http://ej.kubagro.ru/2011/06/pdf/18.pdf, 3 у.п.л.
9. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605с. – Режим доступа: http://lc.kubagro.ru/aidos/aidos02/index.htm
10. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп.– Краснодар: КубГАУ, 2006. – 615 с. – Режим доступа: http://lc.kubagro.ru/aidos/aidos06_lec/index.htm
11. Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп. – Краснодар: КубГАУ, 2006. – 318с. – Режим доступа: http://lc.kubagro.ru/aidos/aidos06_lab/index.htm
12. Луценко Е.В. 30 лет системе «Эйдос» – одной из старейших отечественных универсальных систем искусственного интеллекта, широко применяемых и развивающихся и в настоящее время / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(54). С. 48 – 77. – Шифр Информрегистра: 0420900012\0110. – Режим доступа: http://ej.kubagro.ru/2009/10/pdf/04.pdf, 1,875 у.п.л.
13. Луценко Е.В. Универсальная когнитивная аналитическая система "ЭЙДОС". Пат. № 2003610986 РФ. Заяв. № 2003610510 РФ. Опубл. от 22.04.2003.
14. Луценко Е.В. Типовая методика и инструментарий когнитивной структуризации и формализации задач в СК-анализе / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №01(3). С. 388 – 414. – Режим доступа: http://ej.kubagro.ru/2004/01/pdf/16.pdf, 1,688 у.п.л.
15. Близоруков М. Г. Статистические методы анализа рынка: Учебно-метод. пособие / Близоруков М. Г. – Екатеринбург: Ин-т управления и предпринимательства Урал. гос. ун-та, 2008. – 75 с. – Режим доступа: http://elar.usu.ru/bitstream/1234.56789/1671/6/1334937_schoolbook.pdf
16. Луценко Е.В. Семантическая информационная модель СК-анализа / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграр-


[19] Для удобства читателей ряд работ из списка литературы приведен на сайте автора: http://lc.kubagro.ru/

ного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №02(36). С. 193 – 211. – Шифр Информрегистра: 0420800012\0015. – Режим доступа: http://ej.kubagro.ru/2008/02/pdf/12.pdf, 1,188 у.п.л.

17. Луценко Е.В. Автоматизированная система распознавания образов, математическая модель и опыт применения. В сб.: "В.И.Вернадский и современность (к 130-летию со дня рождения)". Тезисы научно-практической конференции. – Краснодар: КНА, НПП «Эйдос», 1993. – С. 37-42.
18. Луценко Е.В. Универсальная автоматизированная система распознавания образов "Эйдос" (версия 4.1).-Краснодар: КЮИ МВД РФ, 1995.- 76с
19. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с.
20. Симанков В.С., Луценко Е.В. Адаптивное управление сложными системами на основе теории распознавания образов. Монография (научное издание). – Краснодар: ТУ КубГТУ, 1999. - 318с.
21. Луценко Е.В. Математическая сущность системной теории информации (СТИ) (Системное обобщение формулы Больцмана-Найквиста-Хартли, синтез семантической теории информации Харкевича и теории информации Шеннона) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №08(42). С. 76 – 103. – Шифр Информрегистра: 0420800012\0114. – Режим доступа: http://ej.kubagro.ru/2008/08/pdf/04.pdf, 1,75 у.п.л.
22. Луценко Е.В. СК-анализ и система "Эйдос" в свете философии Платона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №01(45). С. 91 – 100. – Шифр Информрегистра: 0420900012\0010. – Режим доступа: http://ej.kubagro.ru/2009/01/pdf/08.pdf, 0,625 у.п.л.
23. 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 (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №02(36). С. 153 – 174. – Шифр Информрегистра: 0420800012\0017. – Режим доступа: http://ej.kubagro.ru/2008/02/pdf/10.pdf, 1,375 у.п.л.
24. 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 (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №10(54). С. 78 – 93. – Шифр Информрегистра: 0420900012\0109. – Режим доступа: http://ej.kubagro.ru/2009/10/pdf/05.pdf, 1 у.п.л.