

<sup>1</sup> Российский университет дружбы народов, г. Москва, Россия<sup>2</sup> Федеральный исследовательский центр «Информатика и управление» РАН, г. Москва, Россия

## МЕТОДЫ ВЫЯВЛЕНИЯ КЛАСТЕРОВ ЯЧЕЕК В РАЗРЕЖЕННЫХ КУБАХ ДАННЫХ МНОГОМЕРНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

### Аннотация

*Моделью данных в информационных системах, построенных на базе многомерного подхода, является многомерный куб данных. В системах с многоаспектным описанием предметной области кубы данных характеризуются большой разреженностью, что затрудняет организацию хранения данных и порождает сложности в процессе анализа информации. Значимые ячейки куба могут быть представлены как допустимые сочетания значений измерений. Анализируя семантически связанные значения разных измерений можно выявить кластеры ячеек – наборы ячеек, которые обладают сходными свойствами. При этом для измерений, которые являются размерностями многомерного куба, выявляются группы их значений, которые семантически связаны с группами значений других измерений. Для построения кластеров ячеек могут быть использованы логические методы интеллектуального анализа. В рамках логического подхода сочетание значений измерений представляется в виде конъюнкции пар «Измерение» – «Значение измерения». К сочетаниям значений измерений может быть применено преобразование, сходное с преобразованием приведения к совершенной дизъюнктивной нормальной форме. Оно позволяет выявить группы значений измерений и кластеры сочетаний значений измерений. Эти кластеры могут быть использованы как элементы модели данных информационной системы.*

### Ключевые слова

*Многомерная модель данных; разреженный куб; множество допустимых сочетаний; кластер сочетаний значений измерений, интеллектуальный анализ данных, индуктивный вывод.*

Fomin M.B.<sup>1</sup>, Smirnov I.V.<sup>2,1</sup><sup>1</sup> Peoples' Friendship University of Russia, Moscow, Russia<sup>2</sup> Federal Research Center Computer Science and Control of the Russian Academy of Sciences, Moscow, Russia

## METHODS FOR IDENTIFYING CLUSTERS OF CELLS IN SPARSE DATA CUBES OF MULTIDIMENSIONAL INFORMATION SYSTEMS

### Abstract

*The data model in information system based on the multi-dimensional approach is a multi-dimensional data cube. Systems with a multi-aspect description of the subject area are characterized by large data cubes with sparseness. It complicates the data storage organization and creates difficulties in the process of data analysis. Possible cube cells can be represented as possible member combinations. The analysis of semantically related members belonging to different dimensions allows identifying clusters – sets of cells that have similar properties. Clusters are constructed from groups of values of dimensions, which are semantically related to groups of values of other dimensions. Logical methods of intellectual analysis can be used to construct clusters of cells. In the framework of a logical approach, a member combination is represented as a conjunction of the pairs “Measurement” – “Measurement value”. The identified clusters can be used as the elements in the data model of the information system.*

### Keywords

*Multidimensional data model; sparse cube; set of possible member combinations; cluster of member combinations, data mining, inductive inference.*

## Введение

В информационной системе, в которой показатели, характеризующие наблюдаемое явление, представлены в многомерной форме, размерностями куба данных являются измерения. Каждое измерение соответствует некоторому аспекту анализа этого наблюдаемого явления. Если система содержит большой объем семантически разнородных данных, многомерный куб характеризуется высокой разреженностью и неравномерностью заполнения [1]. Каждая значимая ячейка куба соответствует некоторому факту. Данные, которые описывают факты многомерного куба, появляются в результате сбора первичной информации, возможно, из разных источников, или в результате преобразования данных, имеющихся в информационной системе. После заполнения информационной системы данными формируется структура значимых ячеек многомерного куба. В процессе анализа наблюдаемого явления, для описания которого создана информационная система, важно иметь описание этой структуры. Для её выявления могут быть использованы методы интеллектуального анализа данных, основанные на индуктивном обобщении небольших выборок данных.

## Описание фактов в многомерном кубе данных

Факты, относящиеся к наблюдаемому явлению, являющемуся объектом анализа, описываются значениями показателей, заданными в ячейках многомерного куба  $H$ . Каждому аспекту анализа наблюдаемого явления соответствует одно из измерений многомерного куба [2]. Полный набор измерений образует множество  $D(H) = \{D^1, D^2, \dots, D^n\}$ , где  $D^i$  –  $i$ -е измерение,  $n = \dim(H)$  – размерность многомерного куба. Измерение задается множеством значений измерения:  $D^i = \{d_{k_1}^i, d_{k_2}^i, \dots, d_{k_i}^i\}$ , где  $k_i$  – число значений  $i$ -го измерения. Значения измерения  $D^i$  выбираются из множества позиций классификатора, который соответствует тому аспекту наблюдаемого явления, который связан с измерением  $D^i$ .

Многомерный куб данных является структурированным набором ячеек. Каждой ячейке  $c$  многомерного куба может быть сопоставлено сочетание значений измерений  $c = (d_{i_1}^1, d_{i_2}^2, \dots, d_{i_n}^n)$ , по одному значению для каждого из измерений [3]. В случае разреженного куба не все возможные сочетания значений измерений соответствуют значимым ячейкам куба.

Если многомерный куб содержит семантически разнородные данные, возможна ситуация, когда значения некоторых измерений не могут быть заданы в сочетании с имеющимся набором значений других измерений. В такой ситуации при описании значимой ячейки многомерного куба значения некоторых измерений не могут быть определены. Для задания значений этих семантически неопределенных измерений может применяться специальное значение «Не используется» [3]. Структуру многомерного куба данных информационной системы в этом случае можно описать как множество допустимых сочетаний значений измерений. В сочетаниях этого множества могут использоваться значения, взятые из классификаторов, соответствующих измерениям, и специальное значение «Не используется». Для обозначения множества допустимых сочетаний значений измерений будем использовать аббревиатуру «МДС».

Введение в модель данных многомерного куба специального значения «Не используется» приводит к необходимости ввести дополнительные требования для значимых ячеек. Если в МДС есть сочетание, в котором значение «Не используется» задано для одного или нескольких измерений (набор 1) в сочетании с некоторым набором значений других измерений (набор 2), то в МДС не должно существовать еще одного сочетания с такими же значениями измерений из набора 2 (иными словами, измерение либо используется, либо не используется в сочетании с некоторым набором значений остальных измерений). В МДС не может существовать сочетания, в котором для всех измерений задано значение «Не используется» – соответствующая ячейка не имеет смысла.

Наблюдаемое явление характеризуется значениями показателей, заданными в значимых ячейках многомерного куба. Полный набор показателей образует множество  $V(H) = \{v_1, v_2, \dots, v_p\}$ , где  $v_j$  –  $j$ -й показатель,  $p$  – число показателей в гиперкубе. В значимой ячейке могут быть заданы не все показатели из  $V(H)$ . Такая ситуация возникает в случае семантического несоответствия между значениями измерений, задающими ячейку, и некоторыми показателями. При описании МДС для каждой значимой ячейки  $c$  требуется задать своё множество  $V(c) = \{v_1, v_2, \dots, v_{p_c}\}$ , состоящее из определенных в этой ячейке показателей,  $1 \leq p_c \leq p$ . Для описания в ячейке  $c$  показателей, не входящих во множество  $V(c)$ , будем применять специальное значение «Не используется». Должно выполняться правило: множество показателей  $V(c)$ , заданных в значимой ячейке  $c$ , не может быть пустым. Описание показателей в незначимых ячейках многомерного куба, соответствующих сочетаниям значений измерений, не входящим в МДС, не имеет смысла.

## Интеллектуальный анализ данных и обработка многомерных кубов данных

Решаемая нами задача состоит в выявлении структуры разреженного многомерного куба данных, сформированного фактами, которые появились в информационной системе в результате сбора данных или в результате преобразования уже имеющихся данных. Для решения этой задачи предлагается использовать подходы интеллектуального анализа данных.

Технология интеллектуального анализа данных (Data Mining) предназначена для поиска в больших объемах данных неочевидных закономерностей [4]. Инструменты Data Mining могут самостоятельно находить такие закономерности и самостоятельно строить гипотезы о взаимосвязях между объектами информационной системы. В основе технологии лежит, в том числе, концепция шаблонов, описывающих закономерности, свойственные подвыборкам данных, которые могут быть выражены в форме, понятной человеку. Традиционные методы анализа данных, прежде всего статистические, ориентированы на проверку заранее сформулированных гипотез, в то время как одно из основных свойств Data Mining – поиск неочевидных закономерностей. В аспекте исследуемой задачи интересны предпринимаемые в последнее время попытки использовать системы Data Mining как часть технологии хранилищ данных.

К области Data Mining также относят задачи классификации, кластеризации, регрессионного анализа (поиск зависимости выходных данных от входных) и задачу выявления шаблонов (поиск наиболее типичных или нетипичных шаблонов в последовательности данных), которые не связаны с потребностями, возникающими при анализе сочетаемости значений измерений, однако могут быть использованы при анализе значений показателей в зависимости от значений измерений.

Методы интеллектуального анализа данных можно классифицировать с разных точек зрения. По способам обработки данных их можно разделить на следующие классы: статистические, логические, эволюционные, нейросетевые. По типу предоставляемых данных, методы анализа данных разделяются на методы обучения «с учителем» (supervised), когда обучающие примеры помечены, «без учителя» (unsupervised), когда обучающие примеры не помечены, и другие виды обучения.

Статистические методы анализа данных обычно опираются на вероятностные подходы к анализу информации и часто оперируют данными, представленными в числовом виде. Результатом выполнения статистических методов являются распределения, наборы чисел, весов, которые являются или статистиками, или расстояниями, или коэффициентами в уравнениях, описывающих разделяющие числовые данные гиперплоскости. Типичными представителями таких методов являются Байесовский классификатор или метод опорных векторов SVM [5]. Статистические методы эффективны при больших выборках обучающих примеров, а результаты их работы трудно интерпретируемы человеком, что является недостатком данных методов. В аспекте настоящего исследования можно сделать следующий вывод по поводу применимости статистических методов анализа: они не будут эффективны по причине недостаточного объема данных и невозможности одновременного использования данных о сочетаемости значений измерений при описании фактов в разных кубах данных. Ведь разным кубам соответствуют разные наборы измерений, и сочетаемость их значений при этом имеет разную семантику.

Логические методы анализа данных используют обобщение или специализацию данных, которые представлены в виде логических структур. Составляющие таких структур обычно записываются символами или цепочками символов, поэтому логические методы называют иногда символьными методами. Типичными представителями таких методов являются методы AQ [6] и ДСМ [7], которые применяются в том числе и для выявления причинно-следственных отношений в данных [8]. Замечательным свойством логических методов является то, что результаты их работы представляют собой легко интерпретируемые человеком правила. Интерпретируемость результатов в некоторых задачах является критически важным, поэтому применение таких методов часто становится безальтернативным. Не менее важное значение имеет также способность логических методов работать эффективно на обучающих данных небольшого объема (от нескольких обучающих примеров). Учитывая вышесказанное, для решения поставленной задачи выявления структуры разреженного многомерного куба данных наиболее целесообразно применять логические методы интеллектуального анализа данных.

### Анализ структуры многомерного куба данных методами data mining

Продемонстрируем возможность использования логических методов Data Mining на примере анализа списка сочетаний значений измерений многомерного куба, описывающего факты наблюдаемого явления «Выдача кредитов в северных областях РФ».

Пусть факты этого наблюдаемого явления определяется следующими измерениями:

D <sup>1</sup>	Вид кредита,	значения:	потребительский; ипотечный; на пополнение оборотных средств; на покупку оборудования; межбанковский;
----------------	--------------	-----------	--

D <sup>2</sup>	Заемщик,	значения:	юридическое лицо; физическое лицо;
D <sup>3</sup>	Пол,	значения:	мужской; женский;
D <sup>4</sup>	Вид деятельности,	значения:	розничная торговля; производство одежды; банковская деятельность;
D <sup>5</sup>	Регион,	значения:	Ярославская область; Тверская область.

МДС для соответствующего многомерного куба имеет вид, представленный в Таблице 1.

Таблица 1. Описание МДС для наблюдаемого явления «Выдача кредитов в северных областях РФ»

Вид кредита	Заемщик	Пол	Вид деятельн.	Регион
потребительский	физ. лицо	мужской	не используется	Яросл. область
потребительский	физ. лицо	женский	не используется	Яросл. область
ипотечный	физ. лицо	мужской	не используется	Яросл. область
ипотечный	физ. лицо	женский	не используется	Яросл. область
на пополнение оборот. средств	юр. лицо	не используется	розничн. торг.	Яросл. область
на покупку оборудования	юр. лицо	не используется	розничн. торг.	Яросл. область
на пополнение оборот. средств	юр. лицо	не используется	произв. одежды	Яросл. область
на покупку оборудования	юр. лицо	не используется	произв. одежды	Яросл. область
межбанковский	юр. лицо	не используется	банковская деят.	Яросл. область
потребительский	физ. лицо	мужской	не используется	Тверская область
потребительский	физ. лицо	женский	не используется	Тверская область
ипотечный	физ. лицо	мужской	не используется	Тверская область
ипотечный	физ. лицо	женский	не используется	Тверская область
на пополнение оборот. средств	юр. лицо	не используется	розничн. торг.	Тверская область
на покупку оборудования	юр. лицо	не используется	розничн. торг.	Тверская область
на пополнение оборот. средств	юр. лицо	не используется	произв. одежды	Тверская область
на покупку оборудования	юр. лицо	не используется	произв. одежды	Тверская область
межбанковский	юр. лицо	не используется	банковская деят.	Тверская область

Представленная в таблице 1 структура МДС определяется тем, что для юридических лиц не определен пол, при описании физических лиц не используется вид деятельности, заемщики разных типов могут брать кредиты разных видов. Такая сложная структура наблюдаемого явления «Выдача кредитов» является причиной разреженности многомерного куба данных, описывающего факты этого явления.

В терминологии методов интеллектуального анализа задачу распознавания структуры многомерного куба данных на основе анализа МДС, представленного в таблице 1, можно поставить следующим образом:

Имеется некоторый класс объектов, который описывается набором сочетаний значений измерений (эти значения берутся из соответствующих измерениям справочников). Требуется сформировать описание этих объектов в компактной форме. Задача разбивается на две подзадачи:

1. Вычленение групп значений измерений из информации о сочетаемости значений этих измерений в МДС;
2. Построение подмножеств сочетаний значений измерений, в которых эти группы значений играют одинаковую роль.

Таким образом, задача распознавания структуры многомерного куба данных является описательной.

Исходные данные для задачи представлены в виде таблицы сочетаний. Такое представление с точки зрения методов анализа является множеством объектов, где описание каждого объекта задаётся признаками – столбцами таблицы, а каждый объект задаётся строкой в таблице. Тогда задача состоит в получении описания множества объектов через возможные сочетания допустимых значений измерений для рассматриваемых объектов.

Предварительный анализ исходных данных показывает, что они обладают следующими особенностями:

- Значения измерений образуют дискретный набор;
- Количество имеющихся сочетаний ограничено;
- Наличие одинаковых сочетаний исключено.

Эти особенности позволяют классифицировать задачу как задачу интеллектуального анализа логического типа. Важным аргументом в пользу выбора логических методов решения является необходимость интерпретировать результаты решения. Вывод: могут быть использованы логические методы (сочетания значений измерений могут быть интерпретированы как логические структуры), решаемая задача является задачей кластеризации.

Может быть предложен следующий алгоритм, состоящий из четырех шагов, на входе которого – список сочетаний значений измерений, на выходе – описание МДС в логической форме с использованием кластеров сочетаний, построенных из групп значений измерений, которые выявлены в процессе выполнения алгоритма.

Из анализа структуры МДС, представленного в виде списка сочетаний, видно, что включение в МДС значений измерения «Регион» имеет особенности: набор сочетаний, со значением «Ярославская область» повторяет набор сочетаний со значением «Тверская область». Это означает, что сочетаемость значений остальных измерений не зависит от значений измерения «Регион». Полный набор измерений может быть разбит на два слоя: первый – состоящий из единственного измерения «Регион», второй – состоящий из измерений {«Вид кредита», «Заемщик», «Пол», «Вид деятельности»}. Сочетаемость значений измерений в одном слое не зависит от значений измерений другого слоя. Сочетания в слоях можно анализировать независимо друг от друга, а полный набор сочетаний МДС представлять в виде декартова произведения сочетаний значений измерений всех слоев.

Следует заметить, что методы анализа данных работают при отсутствии какой-либо априорной информации о наблюдаемом явлении, и поэтому связи между двумя измерениями должны восприниматься просто как информация о совместной встречаемости значений этих измерений из соответствующих групп значений.

Действия по выявлению этих групп значений можно производить путем выполнения несколько шагов.

Шаг 1. Разбиение полного набора измерений на слои измерений.

Разбиение должно быть произведено таким образом, чтобы сочетаемость значений измерений из одного слоя в МДС не зависела бы от значений измерений из другого слоя.

В рассматриваемом иллюстративном примере имеется два слоя. Итоговое полное описание МДС может быть получено декартовым произведением описания для сочетаний из слоя 2 и группы значений {«Ярославская область», «Тверская область»} из слоя 1. Слой 1 содержит одно измерение и имеет вырожденное описание. Дальнейшее изложение алгоритма будем производить для слоя 2.

Шаг 2. Получение начального множества описаний.

На втором шаге для слоя 2 каждый объект (строка) представляется в виде конъюнкций пар ИмяИзмерения.Значение. При этом получаем следующие логические выражения, описывающие объекты:

ВидКредита.Потребительский & Заемщик.ФизЛицо & Пол.Мужской &

ВидДеятельности.НеИспользуется;

ВидКредита.Потребительский & Заемщик.ФизЛицо & Пол.Женский &

ВидДеятельности.НеИспользуется;

ВидКредита.Ипотечный & Заемщик.ФизЛицо & Пол.Мужской & ВидДеятельности.НеИспользуется;

ВидКредита.Ипотечный & Заемщик.ФизЛицо & Пол.Женский & ВидДеятельности.НеИспользуется;

ВидКредита.НаПополнение & Заемщик.ЮрЛицо & Пол.НеИспользуется &

ВидДеятельности.РознТорговля;

ВидКредита.НаПополнениеОборудования & Заемщик.ЮрЛицо & Пол.НеИспользуется &

ВидДеятельности.РознТорговля;

ВидКредита.НаПополнениеОС & Заемщик.ЮрЛицо & Пол.НеИспользуется &

ВидДеятельности.ПроизводствоОдежды;

ВидКредита.НаПополнениеОборудования & Заемщик.ЮрЛицо & Пол.НеИспользуется &

ВидДеятельности.ПроизводствоОдежды;

ВидКредита.Межбанковский & Заемщик.ЮрЛицо & Пол.НеИспользуется &

ВидДеятельности.БанковскаяДеятельность.

Шаг 3. Индуктивное обобщение начального множества сочетаний.

Соединим последовательно дизъюнкциями выражения, отличающиеся только одним конъюнктом (этот процесс похож на приведение к совершенной дизъюнктивной нормальной форме). В результате получим следующие выражения:

ВидКредита.{Потребительский | Ипотечный} & Заемщик.ФизЛицо & Пол.{Мужской | Женский} &

ВидДеятельности.НеИспользуется;

ВидКредита.{НаПополнениеОС | НаПополнениеОборудования} & Заемщик.ЮрЛицо &

Пол.НеИспользуется & ВидДеятельности.{РознТорговля | ПроизводствоОдежды};  
ВидКредита.Межбанковский & Заемщик.ЮрЛицо & Пол.НеИспользуется &  
ВидДеятельности.БанковскаяДеятельность.

Полученные выражения описывают множества связанных значений измерений в слое 2 в смысле совместной встречаемости этих значений.

Шаг 4. Формирование групп значений измерений.

В результате преобразования формируются группы значений измерений с одинаковой сочетаемостью в МДС. Каждое выражение в полученном описании задает отдельный кластер сочетаний значений измерений слоя 2. Кластер задается группами значений измерений, по одной группе для каждого измерения из слоя. Сочетания кластера могут быть получены декартовым произведением этих групп значений.

В рассматриваемом иллюстративном примере выявлены группы значений, которые приведены ниже (если в выражение включены все значения некоторого измерения, соответствующую группу мы будем обозначать «Все»):

Группы для измерения «Вид кредита»: КрФЛ={Потребительский + Ипотечный},  
КрПредпр={НаПополнениеОС + НаПокупкуОборудования}, {Межбанковский}.

Группы для измерения «Заемщик»: {ФизЛицо}, {ЮрЛицо}.

Группы для измерения «Пол»: {Мужской + Женский = Все}, {НеИспользуется}.

Группы для измерения «Вид деятельности»: ВДПредпр={РознТорговля + ПроизводствоОдежды},  
{БанковскаяДеятельность}, {НеИспользуется}.

Шаг 5. Получение логического описания МДС по слоям.

После того, как группы получили наименование, выражения, сформированные на шаге 2, приводятся к итоговому выражению для сочетаний слоя 2 в виде, содержащем описание кластеров сочетаний:

$MDC(Слой_2) = (ВидКредита.КрФЛ \& Заемщик.ФизЛицо \& Пол.Все \& Вид \text{ деятельности } НеИспользуется) | (ВидКредита.КрПредпр \& Заемщик.ЮрЛицо \& Пол.НеИспользуется \& ВидДеятельности.ВДПредпр) | (Вид \text{ кредита.Межбанковский} \& Заемщик.ЮрЛицо \& Пол.НеИспользуется \& ВидДеятельности.Банковская).$

Первый слой содержит единственное измерение «Регион». МДС для этого слоя описывается единственной группой значений измерений:  $MDC(Слой_1) = \{«Ярославская область» + «Тверская область»\}$ .

Шаг 6. Получение логического описания множества СДЗ.

Результирующее МДС может быть получено соединением описаний сочетаний по слоям:  
 $MDC = MDC(Слой_1) | MDC(Слой_2).$

## Полученные результаты

Предложен алгоритм анализа структуры разреженного многомерного куба данных.

Выявлены элементы структуры разреженного многомерного куба данных (слой измерений, группа значений измерений, кластер сочетаний значений измерений), которые могут быть использованы при описании метаданных многомерной информационной системы.

Рассмотренный иллюстративный пример показывает, что логический вариант метода Data Mining может быть использован при построении описания МДС по имеющемуся списку сочетаний. Однако для разработки алгоритма в случае куба большой размерности со сложным вариантом сочетаемости значений измерений при описании структуры многомерного куба данных путем выявления кластеров сочетаний может потребоваться дополнительное разбиение МДС на компоненты. Такая ситуация может возникнуть, если в описании наблюдаемого явления можно выделить несколько семантических компонент и список сочетаний значений измерений разбивается на подмножества, в каждом из которых существует свой вариант разделения измерений на слои. В этой ситуации полное множество сочетаний должно быть разбито на подмножества, соответствующие компонентам наблюдаемого явления, и для каждого подмножества применен описанный выше алгоритм.

Анализ списка сочетаний значений измерений с использованием логических методов интеллектуального анализа имеет еще одну важную сторону – методологическую. Даже исследование простого иллюстративного примера позволило выявить набор конструкций, при помощи которых описывается структура разреженного многомерного куба данных. В качестве таковых можно указать:

- группы значений измерений, объединенные одинаковым поведением по отношению к выявленным в сочетаниях значений измерений связям;
- кластеры сочетаний значений измерений, объединяющие сочетания по критерию схожего поведения по отношению к связям между измерениями;
- слои измерений, разделяющие все имеющиеся в кубе измерения на непересекающиеся наборы. Сочетаемость значений в каждом из слоев определяется только связями между измерениями внутри этого слоя и не зависит от того, какие значения в сочетаниях принимают измерения, лежащие вне слоя.

Приведенные особенности задачи перехода от задания структуры многомерного куба данных списком

сочетаний к кластерному заданию, позволяет классифицировать ее как задачу логического интеллектуального анализа. Предложенный алгоритм представляет принципы решения задачи. Такая задача не может быть алгоритмизирована однозначно. Она может модифицироваться для случая каждого конкретного многомерного куба данных с учетом особенностей его структуры.

### Заключение

Данные многомерной информационной системы, использующей большой набор характеристик наблюдаемого явления, имеют структуру разреженного и неравномерно заполненного многомерного куба. Анализ структуры многомерного куба, описывающего факты, полученные в результате сбора данных или преобразования информации других фактов, является сложной задачей. Для её решения можно использовать логические методы интеллектуального анализа. Предлагаемый алгоритм анализа структуры разреженного многомерного куба позволяет выявить элементы структуры куба данных (слои измерений, группы значений измерений) и представить совокупность ячеек куба в виде набора кластеров ячеек, обладающих сходными свойствами. Каждый кластер задается набором сочетаний значений измерений, которые могут быть получены декартовым произведением групп значений измерений. Выявленные элементы структуры многомерного куба можно использовать при описании метаданных многомерной информационной системы. Такой способ описания метаданных хорошо отражает семантику наблюдаемого явления, для описания которого разработана информационная система.

### Благодарности

Работа выполнена при финансовой поддержке Министерства образования и науки РФ по Программе повышения конкурентоспособности РУДН «5-100» среди ведущих мировых научно-образовательных центров на 2016-2020 гг.

### Литература

1. Thomsen E. OLAP Solution: Building Multidimensional Information System. — NY: Willey Computer Publishing, 2002. — 688p.
2. Висков А.В., Фомин М.Б. Моделирование аналитических измерений в многомерных базах данных // Вестник Иркутского государственного технического университета. — 2012. — Т. 63, № 4. — С. 15-19.
3. Висков А.В., Фомин М.Б. Методы описания допустимых сочетаний реквизитов-признаков при использовании многомерных моделей в инфокоммуникационных системах // Т-Comm. – Телекоммуникации и Транспорт. — 2012. — № 7. — С. 45-47.
4. Fayyad U., Piatetsky-Shapiro G., Smyth P. From data mining to knowledge discovery: an overview // Advances in knowledge discovery and data mining. — Menlo Park: AAAI press, 1996. — P. 1-34.
5. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979. — 448 с.
6. Michalski R.S. Attributional calculus: a logic and representation language for natural induction // Reports of the Machine Learning and Inference Laboratory, MLI 04-2, George Mason University, Fairfax, 2004 [электронный ресурс] // URL: <http://mars.gmu.edu/handle/1920/1487/> (датаобращения 04.09.2017)
7. Автоматическое порождение гипотез в интеллектуальных системах. Сост. Панкратова Е.С., Финн В.К. — М.: Либроком, 2009. — 528 с.
8. Панов А.И. Выявление причинно-следственных связей в данных психологического тестирования логическими методами // Искусственный интеллект и принятие решений. — 2013. — № 1. — С. 24-32.

### References

1. Thomsen E. OLAP Solution: Building Multidimensional Information System. — NY: Willey Computer Publishing, 2002. — 688p.
2. Viskov A.V., Fomin M.B. Analytical dimensions modeling in multidimensional databases // Vestnik IrGTU [Proceedings of Irkutsk State Technical University]. — 2012. — Т. 63, № 4. S. 15-19.
3. Viskov A.V., Fomin M.B. Methods of a description of the permissible combination of detail-attributes when using multidimensional models in infocommunication systems // T-Comm: Telecommunications and Transport. — 2012. — № 7. — S. 45-47.
4. Fayyad U., Piatetsky-Shapiro G., Smyth P. From data mining to knowledge discovery: an overview // Advances in knowledge discovery and data mining. — Menlo Park: AAAI press, 1996. — P. 1-34.
5. Vapnik V. Vosstanovlenie zavisimostei ot empiricheskikh dannyh // M.: Nauka, 1979. — 448 s.
6. Michalski R.S. Attributional calculus: a logic and representation language for natural induction // Reports of the Machine Learning and Inference Laboratory, MLI 04-2, George Mason University, Fairfax, 2004 [электронный ресурс] // URL: <http://mars.gmu.edu/handle/1920/1487/> (датаобращения 04.09.2017)
7. Avtomaticheskoe porozhdenie gipotez v intellektuaknykh sistemakh. Editors Pankratova E.S., Finn V.K. — M.: Librocom, 2009. — 528 s.
8. Panov A.I. Extraction of cause-effect relationships from psychological test data using logical methods // Scientific and Technical Information Processing— 2014. — Т. 41, № 5. — S. 275-282.

### Об авторах:

**Фомин Максим Борисович**, кандидат физико-математических наук, доцент, доцент кафедры информационных технологий, Российский университет дружбы народов, [fomin\\_mb@rudn.university](mailto:fomin_mb@rudn.university)

**Смирнов Иван Валентинович**, кандидат физико-математических наук, доцент, заведующий

лабораторией «Компьютерная лингвистика и интеллектуальный анализ информации» Института системного анализа, Федеральный исследовательский центр «Информатика и управление» Российской академии наук; доцент кафедры информационных технологий, Российский университет дружбы народов, [ivs@isa.ru](mailto:ivs@isa.ru)

**Note on the authors:**

**Fomin Maxim B.**, Candidate of Physico-Mathematical Sciences, Associate Professor, Associate Professor of the Department of Information Technology, Peoples' Friendship University of Russia, [fomin\\_mb@rudn.university](mailto:fomin_mb@rudn.university)

**Smirnov Ivan V.**, Candidate of Physico-Mathematical Sciences, Associate Professor, head of the laboratory for Computational Linguistics and Intelligent Data Analysis, The Institute for Systems Analysis, Federal Research Center Computer Science and Control of the Russian Academy of Sciences; Associate Professor of the Department of Information Technologies, Peoples' Friendship University of Russia, [ivs@isa.ru](mailto:ivs@isa.ru)