

# Информатика и её применения

Том 12 Выпуск 1 Год 2018

## СОДЕРЖАНИЕ

Онтология цифровой семантической библиотеки LibMeta <b>О. М. Атаева, В. А. Серебряков</b>	<b>2</b>
Диаграммы уязвимости потоковых сетевых систем <b>Ю. Е. Малашенко, И. А. Назарова, Н. М. Новикова</b>	<b>11</b>
Автоматизированная система весового проектирования самолетов <b>Л. Л. Вышинский, Ю. А. Флеров, Н. И. Широков</b>	<b>18</b>
Континуальный критерий VaR на сценарных рынках <b>Г. А. Агасандян</b>	<b>31</b>
О формальной постановке задач поиска сгущений в разреженных булевых матрицах <b>И. С. Алешин</b>	<b>40</b>
Учет предварительных оценок скорости порождения сходств спаривающей цепью Маркова <b>Д. В. Виноградов</b>	<b>49</b>
Метод интерполяционного аналитического моделирования одномерных распределений в стохастических системах <b>И. Н. Сеницын</b>	<b>55</b>
О некоторых возможностях управления ресурсами при организации проактивного противодействия компьютерным атакам <b>А. А. Грушо, М. И. Забейло, А. А. Зацаринный, Е. Е. Тимонина</b>	<b>62</b>
Реконструкция осей главных компонент <b>М. П. Кривенко</b>	<b>71</b>
Система ситуационного управления как мультисервисная технология в облачной среде <b>А. А. Зацаринный, А. П. Сучков</b>	<b>78</b>
Эксплоративный патентный поиск <b>И. В. Соченков, Д. В. Зубарев, И. А. Тихомиров</b>	<b>89</b>
Теория категорий как математическая прагматика модельно-ориентированной системной инженерии <b>С. П. Ковалёв</b>	<b>95</b>
Байесовские модели тестирования больших групп обслуживающих приборов <b>А. А. Кудрявцев, О. В. Шестаков</b>	<b>105</b>
Разработка и предварительное исследование стохастической полумарковской модели управления запасом непрерывного продукта при постоянно происходящем потреблении <b>П. В. Шнурков, А. Ю. Егоров</b>	<b>109</b>
Финансирование фундаментальных исследований: концептуальный облик системы поддержки принятия решений с использованием методов наукометрии и анализа данных <b>О. В. Сюнтюренко</b>	<b>118</b>
Об авторах	<b>128</b>
Правила подготовки рукописей	<b>130</b>
Requirements for manuscripts	<b>133</b>

# ОНТОЛОГИЯ ЦИФРОВОЙ СЕМАНТИЧЕСКОЙ БИБЛИОТЕКИ LibMeta

О. М. Атаева<sup>1</sup>, В. А. Серебряков<sup>2</sup>

**Аннотация:** При разработке цифровых библиотек особое внимание уделяют модели данных содержимого библиотеки. При этом контент цифровых библиотек может быть описан различными форматами и представлен различными способами. Библиотека, определяемая с помощью системы LibMeta, рассматривается как хранилище структурированных разнообразных данных с возможностью их интеграции с другими источниками данных и предполагает возможность специфицирования своего контента за счет описания предметной области. В качестве средства формализации выступает онтология контента семантической библиотеки. Также вводятся основные понятия для описания задачи интеграции данных из источников Linked Open Data (LOD), понятия для определения произвольного тезауруса. Онтология построена таким образом, чтобы иметь возможность определения семантической библиотеки в произвольной предметной области.

**Ключевые слова:** семантические библиотеки; модель данных; онтологии; источники данных; поиск в LOD

**DOI:** 10.14357/19922264180101

## 1 Введение

В различных предметных областях модель данных содержимого цифровых семантических библиотек может существенно отличаться как по типам ресурсов, так и по их структуре. При разработке таких библиотек особое внимание уделяют модели данных содержимого библиотеки.

Говоря о библиотеках, авторы прежде всего имеют в виду разработанную информационную систему для создания семантических библиотек LibMeta [1–3], с помощью которой создается и описывается семантическая библиотека некоторой предметной области.

LibMeta представляет собой информационную систему, которая реализует функциональность, необходимую для работы с контентом семантической библиотеки. LibMeta не является традиционной системой управления электронными библиотеками (СУЭБ).

Развитие современных технологий подталкивает к переопределению как понятия библиотеки, так и контента библиотеки, в качестве которых не обязательно могут выступать традиционные описания печатных изданий, но и любые другие типы цифровых объектов. При этом контент цифровых библиотек может быть описан различными форматами и представлен различными способами.

Библиотека, реализуемая с помощью LibMeta, рассматривается как хранилище структурированных разнообразных данных с возможностью их интеграции с другими источниками данных и предполагает возможность специфицирования своего контента путем описания предметной области.

Определение предметной области задается тезаурусом [4], который содержит основные термины этой предметной области, связанные иерархическими и горизонтальными связями между собой. Содержимое библиотеки задается типами ресурсов, описание которых задает, в свою очередь, множество допустимых объектов, возможно объединенных в разнообразные коллекции, составляющие вместе с тезаурусом ее контент.

Статья посвящена исследованию средств представления знаний о контенте семантической библиотеки. Эти средства необходимы для автоматизации описания ресурсов библиотеки конкретной предметной области и возможности их автоматизированной интеграции с данными внешних открытых источников. Необходимым условием для этого является структуризация и формализация знаний в области описания контента семантической библиотеки.

При реализации LibMeta авторы руководствовались набором основных задач, которые должна решать разрабатываемая система:

<sup>1</sup>Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, oli@ultimeta.ru

<sup>2</sup>Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, serebr@ultimeta.ru

- (1) библиотека должна поддерживать возможность использования медийных объектов или ссылки на них при описании своих объектов, включая текст, аудио- и видеофайлы или любую их комбинацию. Это требование отражается в названии словом «цифровая»;
- (2) типы используемых ресурсов и связи между ними должны быть описаны средствами системы в рамках определенных в предыдущей работе понятий, составляющих семантическое описание ресурсов контента библиотеки. При этом, согласно принципам LOD, при описании ресурсов поддерживается использование классов и свойств ранее используемых онтологий в сообществе, поддерживающем LOD. Эта поддержка выражается либо в непосредственном использовании готовых онтологий при описании ресурсов и связей между ними, либо в возможности ссылок на их элементы, используя связи на уровне описания ресурсов. Это требование отражается в названии словом «семантическая»;
- (3) библиотека должна служить интеграционным узлом, предоставляя возможность связывания своих данных с данными из разных источников, которые включены в облако LOD. Должна также обеспечиваться возможность извлекать данные этой библиотеки в машиночитаемом формате. Это требование отражается в названии словом «открытая»;
- (4) пользователи библиотеки должны иметь возможность организовывать свои коллекции по интересующему их научному направлению, добавляя новые термины в предметный тезаурус, уточняя таким образом область своих интересов. Пользователи должны также иметь возможность осуществлять поиск не только среди объектов в рамках системы, но и по источникам данных, без необходимости использования специализированного языка для поисковых запросов. Это требование отражается в названии словом «персональная».

Основные требования, предъявляемые при этом к контенту системы, — *универсальность, структурированность, адаптируемость* — не противоречат этим свойствам и обеспечивают поддержку настраиваемого хранилища метаданных для объектов и расширяемый набор информационных ресурсов. *Универсальность* обеспечивает описания типов ее ресурсов и объектов независимо от предметной области и области интересов пользователей. *Структурированность* описания обеспечивает

поддержку связей между различными типами ресурсов как внутри системы, так и вне нее, исходя из определений LOD. *Адаптируемость* описания ресурсов обеспечивает возможность добавления новых свойств и связей в процессе развития системы и обеспечивает настройку пользовательских интерфейсов под эти изменения.

В качестве средства формализации выступает онтология [5] контента семантической библиотеки. На основе этого описания можно выделить основные понятия описания задачи интеграции данных из открытых источников.

В качестве открытых источников рассматриваются источники данных, включенные в LOD [6] и соответствующие основным требованиям, предъявляемым к таким источникам данных.

В качестве основных разделов освещаемой задачи рассмотрим определение тезауруса и основные стандарты, выделим понятия, необходимые для описания контента семантической библиотеки в произвольной предметной области, определим основные понятия, необходимые для описания задачи интеграции данных из открытых источников и выделим основные типы связей между этими понятиями.

## 2 Тезаурус и стандарты

Для описания какой-либо предметной области всегда используется определенный набор терминов, каждый из которых обозначает или описывает какую-либо концепцию из этой предметной области. Совокупность терминов, описывающих предметную область с указанием семантических отношений (связей) между ними, является тезаурусом. Такие отношения в тезаурусе всегда указывают на наличие смысловой (семантической) связи между терминами.

При этом модель тезауруса не должна быть ориентирована ни на одну из конкретных предметных областей и быть достаточно гибкой для того, чтобы позволить всегда сохранять актуальность словаря и удобство его использования для определения любой предметной области.

Тезаурус с наличием связей различных типов позволяет реализовать гибкий настраиваемый поиск, результатом которого будет список объектов предметной области, соответствующий выбранным терминам.

Рассматриваемая в статье модель тезауруса соответствует стандарту ISO 2788-1986. Этот стандарт определяет тезаурус как набор терминов, связанных между собой соответствующими связями (отношениями).

Термины могут иметь следующие атрибуты:

- SN — Scope Note. Комментарий к термину. Например, представляет вербальное пояснение термина или правила его использования;
- TT — Top Term. Признак, выделяющий термины на самом верхнем уровне иерархии (термины наиболее общих понятий в иерархии понятий).

Связи между терминами могут быть следующими:

- USE — связывает термин с наиболее предпочтительным термином для понятия. А USE  $B$  означает, что термин  $B$  является наиболее предпочтительным для понятия, обозначаемого термином  $A$ ;
- UF — Used For. Обращение связи USE. Связывает наиболее подходящий термин с синонимами и квазисинонимами (менее подходящими терминами);
- BT — Broader Term. Связь термина с термином более общего понятия. А BT  $B$  означает, что термин  $B$  обозначает более общее понятие по сравнению с понятием, обозначаемым термином  $A$ ;
- BTG — Broader Term Generic. Вариант связи BT в случае, когда термин характеризует разновидность понятия, определяемого более общим термином. Например, «попугай» и «птицы». Наличие связи BTG подразумевает наличие связи BT;
- BTP — Broader Term Partitive. Вариант связи BT в случае, когда термин характеризует часть понятия, определяемого более общим термином. Например, «математика» и «теория чисел». Наличие связи BTP подразумевает наличие связи BT;
- NT, NTG и NTP — Narrower Term, Narrower Term Generic и Narrower Term Partitive — обращение связей BT, BTG и BTP соответственно;
- RT — Related Term. Ассоциативная связь. Используется для семантически связанных между собою терминов, не находящихся при этом в одной иерархии и не являющихся синонимами или квазисинонимами. Эта связь проставляется в тех случаях, когда пользователю тезауруса может быть полезно осуществлять поиск или индексацию не только по данному термину, но и по связанному с ним.

### 3 ОНТОЛОГИЯ

Исходя из вышесказанного, тезаурус — это полный систематизированный набор терминов о какой-либо области знаний и больше относится к лексике, используемой в конкретной области, тогда как онтология описывает ресурсы предметной области и их взаимосвязи. Для каждой предметной области набор ресурсов может отличаться как по формату, так и по набору самих ресурсов. Поэтому, задавая определение самой библиотеки, предлагается использовать для описания ресурсов, составляющих контент конкретной предметной области, понятия, общие для любой из них, т. е. набор понятий, формирующих описание контента библиотеки, должен быть настолько универсальным, чтобы мог адаптироваться под нужды конкретной области.

Так как одной из основных задач, решаемых в рамках библиотеки, как было сказано выше, является интеграция данных из различных источников, такой подход позволяет реализовать средства интеграции данных в рамках библиотеки, адаптируемые под условия любой предметной области без оглядки на ее специфику.

Понятия, составляющие онтологию библиотеки LibMeta, условно делятся на предназначенные для:

- описания контента предметной области;
- формирования тезауруса любой предметной области;
- описания тематических коллекций;
- описания задачи интеграции контента библиотеки с данными источников из LOD.

Между этими группами понятий определены семантически значимые связи.

Рассмотрим далее основные формальные определения, необходимые для описания онтологии.

**Определение 1.** *Контент библиотеки*  $C = \langle IR, A, IO \rangle$  определяется типами ее информационных ресурсов, описанных связанными с ними наборами атрибутов  $A$  и набором входных данных, определяющих информационные объекты  $IO$ , которые являются непосредственно объектами, хранящимися в библиотеке.

**Определение 2.** *Тезаурус библиотеки*  $TH = \langle T, R \rangle$  определяется терминами  $T$  и связями  $R$  между ними. Набор терминов  $T$ , составляющих описание предметной области, строго задан.

**Определение 3.** *Семантические метки*  $M = \{m_i\}$  информационного объекта — это термины, которые не попали в тезаурус, но являются необходимыми для специфицирования тематики информационного объекта. Семантические метки не связаны,

в отличие от терминов тезауруса, связями между собой или с терминами тезауруса, но дают возможность дополнительного тематического разделения информационных объектов в рамках предметной области.

**Определение 4.** *Задача интеграции данных библиотеки*  $IT = \langle DS, R, A, M, D, D_S \rangle$  с внешними источниками  $DS$  определяется типами ресурсов библиотеки и набором их атрибутов  $A$ , отображением  $M$  ресурсов  $R$  на схему источника данных  $S$  и набором связей  $D_S$  с данными из источника.

**Определение 5.** *Коллекция информационных объектов*  $C = \langle IO, T, M, DS \rangle$  представляет собой набор объектов, объединенных на основе совокупности признаков:

- (1) по их термину тезауруса предметной области;
- (2) по семантическим меткам;
- (3) по источнику данных, из которого поступили объекты.

В коллекцию могут входить объекты различных типов ресурсов, заданных при описании контента библиотеки. При этом коллекции по каждому признаку могут формироваться автоматически и будем называть их автоматическими коллекциями. В случае, когда признаки определяет пользователь, будем называть такие коллекции просто *коллекциями*.

**Определение 6.** *Семантически значимыми связями библиотеки*  $P = \{P_i\}$  назовем связи, определенные между контентом библиотеки, ее предметной областью (тезаурусом), семантическими метками и объектами источника данных. Выделим следующие основные связи:

- $P_1(t, io)$  — термин тезауруса — информационный объект;
- $P_2(io, t)$  — информационный объект — термин тезауруса;
- $P_3(r, s)$  — информационный ресурс — класс объектов источника, где информационный ресурс — это общее определение для информационных объектов, хранящихся в системе; таким образом, фактически информационные объекты являются экземплярами информационных ресурсов;
- $P_4(a, s_a)$  — атрибут информационного ресурса — свойство класса источника;
- $P_5(io, o_s)$  — информационный объект — экземпляр класса из источника данных;

- $P_6(m, io)$  — семантическая метка — информационный объект;
- $P_7(io, m)$  — информационный объект — семантическая метка.

На основе введенных явных связей можно определить связи, которые назовем *неявными значимыми связями* (т. е. заданными по некоторым определенным заранее правилам) между семантическими метками и терминами тезауруса и объектами как самой библиотеки, так и экземплярами связанных данных из источников:

- $P_8(m, t) \leftarrow P_6(m, io) \wedge P_2(io, t)$  семантическая метка — информационный объект — термин тезауруса;
- $P_9(t, m) \leftarrow P_1(t, io) \wedge P_7(io, m)$  термин тезауруса — информационный объект — семантическая метка;
- $P_{10}(m, o_s) \leftarrow P_6(m, io) \wedge P_5(io, o_s)$  семантическая метка — информационный объект — экземпляр класса из источника данных;
- $P_{11}(t, o_s) \leftarrow P_1(t, io) \wedge P_5(io, o_s)$  термин тезауруса — информационный объект — экземпляр класса из источника данных.

Для представления онтологии LibMeta был выбран язык описания онтологий OWL (Web Ontology Language)<sup>1</sup>. Такая онтология состоит из классов, свойств классов и индивидов. В терминах OWL  $P_1$  инверсивно  $P_2$ ,  $P_6$  инверсивно  $P_7$ ,  $P_8$  инверсивно  $P_9$  и  $P_{10}$  инверсивно  $P_{11}$ . При этом правила для неявных связей задаются с помощью правил SWRL (Semantic Web Rule Language)<sup>2</sup>. Правила SWRL как расширение OWL помогают описать абстрактный механизм оперирования объектами предметной области и ее закономерности. Правила SWRL дают возможность выводить новые факты из существующих утверждений, что повышает эффективность описания предметной области.

В соответствии с определениями были введены основные классы онтологии. Исходя из определения 1, вводятся классы:

1. IResource (информационный ресурс библиотеки), который содержит общую информацию о типе ресурса, название, *URI* (Universal Resource Identifier)<sup>3</sup> и информацию об используемом наборе атрибутов для описания структуры ресурса.

<sup>1</sup><https://www.w3.org/TR/owl-ref>.

<sup>2</sup><https://www.w3.org/Submission/SWRL>.

<sup>3</sup><https://tools.ietf.org/html/rfc3986>.



2. IObject (информационный объект библиотеки), который фактически представляет собой экземпляр некоторого ресурса и по составу атрибутов соответствует набору атрибутов связанного с ним ресурса. Для описания соответствующих значений для информационного объекта имеется многозначное свойство value, значениями которого являются экземпляры вспомогательного класса AttributeValue, содержащие информацию о конкретном значении объекта и соответствующем атрибуте.
3. Attribute (атрибут, элемент описания информационного ресурса), который имеет следующие свойства:
  - (а) name — название;
  - (б) type — содержит информацию о типе значений этого атрибута и может включать такие значения, как *строка*, *число*, *дата*, *тип ресурса* (т. е. значениями являются объекты некоторого выбранного типа ресурса);
  - (в) view — указывает на область применения атрибута в рамках системы. Может иметь значения *поисковый* (участвует в формировании поисковых форм), *идентифицирующий* (является обязательным) и *описательный* (содержит дополнительную информацию об описываемом объекте).
4. AttributeSet (набор атрибутов, группирующий атрибуты, соответствующие одному представлению ресурса).

Исходя из определения 2, согласно описанному ранее стандарту ISO 2788-1986 для тезаурусов, вводятся классы:

5. Thesaurus (тезаурус предметной области) — содержит в себе общую информацию о тезаурусе: название и авторов (организации и персоны). Наличие этой сущности позволяет загружать готовые тезаурусы, не смешивая их с теми, что уже, быть может, есть в системе.
6. Concept — сущность, содержащая информацию о понятиях тезауруса. Содержит следующие атрибуты:
  - (а) Name — название понятия. В случае, если понятие не может иметь названия, представленного в виде текста, используется какой-либо идентификатор;
  - (б) RepresentationType — тип представления понятия. Понятие не всегда можно описать словами, иногда для этого гораздо больше подходит формула или изображение, поэтому необходимо иметь возможность добавления понятия в любом виде;

- (в) Image — изображение;
- (г) Note — примечание.

7. ConceptGroup — тематическое разделение понятий тезауруса.
8. HierarchicalRel — иерархические связи, определяющие древовидную структуру словаря. Содержит атрибуты, определяющие связи в соответствии со стандартом (BT, BTG, BTP).
9. FamilyRel — горизонтальные связи. Они задают родственные отношения между понятиями и позволяют находить публикации по похожим тематикам. Содержит также атрибуты, определяющие связи в соответствии со стандартом (NT, NTG, NTP).
10. PreferredTerm — дескрипторы понятия. Каждому понятию соответствует единственный дескриптор на каждом языке.
11. NonPreferredTerm — сюда включаются синонимы. Один дескриптор может иметь множество синонимов. В этот класс объектов добавлен атрибут Visibility — свойство, отвечающее за видимость термина. Имеет два значения — global и private, глобальная и приватная области видимости соответственно. Этот атрибут введен для решения проблемы множественных терминологий — разные люди могут называть одни и те же объекты по-разному (пусть даже эти названия будут похожи). Для того чтобы каждому пользователю было комфортно работать в системе, ему дается возможность создавать свои термины, если таковых нет в глобальной части тезауруса. Эти термины он может связывать с другими терминами из глобальной части и размечать ими свои публикации. Таким образом, если два пользователя создали в своих локальных репозиториях удобные для них ключевые слова, разместили ими свои публикации и связали эти ключевые слова с одним и тем же термином из глобального тезауруса, то они смогут находить и получать публикации друг друга, пользуясь при этом своими терминологиями.
12. Term — общий класс, объединяющий дескрипторы и синонимы. Содержит набор свойств, который при необходимости позволяет произвольно расширять текстовые описания терминов и определять связи с информационными объектами системы.

Исходя из определений 3 и 5, вводятся классы:

13. SemanticTag — класс семантических меток, который обладает следующими свойствами:

- (a) title — краткое название семантической метки;
  - (б) description — расширенное описание семантической метки.
14. ICollection — класс коллекций, определенных человеком, который обладает следующими свойствами:
- (a) name — название коллекции;
  - (б) definition — описание коллекции;
  - (в) resources — типы ресурсов, включаемых в эту коллекцию.
- Исходя из определения 4 вводятся классы:
15. DataSource (источники данных LOD) — класс, который имеет следующие свойства:
- (a) name — название источника;
  - (б) description — описание источника;
  - (в) url — точка входа для извлечения данных;
  - (г) resourceMapping — содержит информацию о типах ресурсов, отображаемых на этот источник, и соответствующие классы источника. Значениями являются экземпляры класса ResourceMapping.
16. ResourceMapping — класс, содержащий информацию об отображаемых на источник данных информационных ресурсах библиотеки:
- (a) resource — тип ресурсов, отображаемых на этот источник;
  - (б) class — ссылка на соответствующий класс источника данных;
  - (в) attributeMappings — содержит экземпляры класса AttributeMapping, содержащих информацию об отображении соответствующих ресурсу атрибутов.
17. AttributeMapping — класс, содержащий информацию об отображаемых на источник данных атрибутах из набора атрибутов, соответствующего информационному ресурсу библиотеки:
- (a) attribute — атрибут, отображаемый на этот источник;
  - (б) property — ссылка на соответствующее свойство класса источника данных.
- На рис. 1 и 2 приведены примеры описания конкретного информационного ресурса и информационного объекта в терминах этой онтологии согласно определению 1.

```

- <rdf:RDF>
- <lbm:InformationResource rdf:about="http://libmeta.ru/resource/person">
  <lbm:title>Person</lbm:title>
  <lbm:label>Персона</lbm:label>
  <lbm:description>Ресурс соответствующий персонам</lbm:description>
- <lbm:properties>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#activity"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#additionEmployer"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#additionPosition"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#address"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#bio"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#dateOfBirth"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#dateOfDeath"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#employer"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#first"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#keywords"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#last"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#middle"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#placeOfBirth"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#position"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#seeAlso"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#source"/>
  <lbm:property rdf:resource="http://libmeta.ru/attribute#email"/>
  </lbm:properties>
  <lbm:dateCreated> 06-09-2016 13:32 </lbm:dateCreated>
  <lbm:dateUpdated> 06-09-2016 13:32 </lbm:dateUpdated>
  </lbm:InformationResource>
- <lbm:InformationResource rdf:about="http://libmeta.ru/resource/publication">

```

Рис. 1 Пример описания информационного ресурса в терминах онтологии LibMeta

```

- <rdf:RDF>
- <lbm:InformationObject rdf:about="http://libmeta.ru/resource/publication#vmj#2758">
  <lbm:type rdf:resource="http://libmeta.ru/resource/publication"/>
  <lbm:description/>
  <lbm:dateCreated> 08-09-2016 01:17 </lbm:dateCreated>
  <lbm:dateUpdated> 08-09-2016 01:17 </lbm:dateUpdated>
- <lbm:properties>
- <lbm:property>
  <lbm:type rdf:resource="http://libmeta.ru/attribute#annt"/>
  <lbm:value>
    Устанавливается, что композиция квазидифференцируемых отображений квазидифференцируема и выводится явная формула д
    технику дезинтегрирования, установлено, что в специальных случаях выполняется аналог классического "цепного правила" &l
    квазидифференциалов. Получены следствия для вычисления квазидифференциалов супремума, инфимума и интегрального опе
  </lbm:value>
</lbm:property>
- <lbm:property>
  <lbm:type rdf:resource="http://libmeta.ru/attribute#auth"/>
  <lbm:value>Басаева Елена Казбековна</lbm:value>
</lbm:property>
- <lbm:property>
  <lbm:type rdf:resource="http://libmeta.ru/attribute#auth"/>
  <lbm:value>Кусраев Анатолий Георгиевич</lbm:value>
</lbm:property>
- <lbm:property>
  <lbm:type rdf:resource="http://libmeta.ru/attribute#issueDate"/>
  <lbm:value>2003</lbm:value>
</lbm:property>
- <lbm:property>
  <lbm:type rdf:resource="http://libmeta.ru/attribute#magazine"/>
  <lbm:value>ВМЖ</lbm:value>
</lbm:property>

```

Рис. 2 Пример описания информационного объекта в терминах онтологии LibMeta

#### 4 Использование онтологии контента библиотеки и тезауруса предметной области при конструировании семантической библиотеки в LibMeta

Для применения тезауруса конкретной предметной области и онтологии контента библиотеки необходимо придерживаться следующей последовательности их использования при конструировании семантической библиотеки в рамках LibMeta:

- (1) на основе введенной модели задается набор информационных ресурсов, используемых в библиотеке. Для этого необходимо представить описания содержимого будущей библиотеки в терминах предложенной модели. На базе классов, заданных для описания контента библиотеки, реализован модуль, в котором задаются базовые свойства, атрибуты для ресурсов и связи между ними;
- (2) осуществляется окончательная настройка структуры тезауруса. На базе определенных классов согласно определению тезауруса ре-

ализован модуль для его построения, в котором задаются используемые связи между терминами, расширяется при необходимости описание термина, определяются связи с ресурсами системы;

- (3) для выбора семантических меток можно использовать дополнительные словари по предметной области или оставить возможность их определения (доопределения) позднее;
- (4) на основе заданных классов согласно определению задачи интеграции реализован модуль, в рамках которого осуществляется подключение внешних источников данных. Это действие можно выполнить на любом этапе жизнедеятельности системы;
- (5) на основе заданных классов согласно определению коллекций реализован модуль, в рамках которого осуществляются создание коллекций и их наполнение; это можно выполнить также на любом этапе.

На основе выполненных действий происходит автоматическая адаптация пользовательских интерфейсов системы под заданные описания ресурсов, составляющих содержимое библиотеки. Пользовательский интерфейс делится условно на следующие категории:



- интерфейсы поиска;
- интерфейсы просмотра;
- интерфейсы редактирования;
- интерфейсы загрузки данных.

## Пример

На основе предложенной модели была сконструирована библиотека для предметной области обыкновенных дифференциальных уравнений (ОДУ). В качестве тезауруса использован тезаурус ОДУ, разработанный коллективом специалистов в этой области [7].

Объектами библиотеки рассматривались журнальные математические статьи. В качестве примеров типов ресурсов, соответственно, рассматривались *Авторы* и *Публикации*. Был определен набор атрибутов для каждого типа ресурсов в рамках минимального набора свойств на основе Dublin Core<sup>1</sup> для публикаций и FOAF (Friend of a Friend)<sup>2</sup> для описания авторов. В качестве примера источника данных рассматривались данные о персонах из системы MathNet<sup>3</sup>, которые были смоделированы в виде источника, интегрированного в LOD. Были определены отображения атрибутов *Авторов* на свойства персон из этого источника и выявлены связи у почти 50% авторов рассматриваемых публикаций, при этом авторов было около 700. Средствами системы для каждой публикации на основе ее названия, аннотации и ключевых слов были выявлены связи с тезаурусом ОДУ. В качестве семантических меток были использованы термины математической энциклопедии<sup>4</sup> [8], что позволило дополнительно выявить смежные предметные области и произвести дополнительное тематическое разбиение публикаций в рамках предметной области. Такое связывание позволило выявить с некоторой долей вероятности статьи, относящиеся к предметной области ОДУ, и организовать их в коллекции на основе тезауруса и выявленных семантических меток. Было использовано описание около 2000 публикаций, из них около 30% были отнесены к области ОДУ и имели связи со смежными предметными областями, выявленными согласно семантическим меткам.

## 5 Дальнейшее направление работ

Работа с полными текстами предоставленных статей пока находится в активной стадии. Предпо-

лагается создание информационного образа статей для выделения микротезауруса на основе семантических меток и терминов предметной области по каждой статье с дальнейшим определением возможностей расширения используемых тезаурусов или для создания облака ключевых понятий отдельных областей знания.

Отдельной задачей является семантическая обработка формул из полных текстов и определение их ключевых слов с возможностью дальнейшего поиска по формулам, а также выделение отдельных направлений и математических школ. При этом формулы рассматриваются как отдельный тип ресурсов системы.

## Литература

1. *Серебряков В. А., Атаева О. М.* Персональная цифровая библиотека LibMeta как среда интеграции связанных открытых данных // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Тр. XVI Всеросс. науч. конф. RCDL'2014. — Дубна: ОИЯИ, 2014. С. 66–71.
2. *Серебряков В. А., Атаева О. М.* Основные понятия формальной модели семантических библиотек и формализация процессов интеграции в ней // Программные продукты и системы, 2015. № 4. С. 180–187.
3. *Серебряков В. А., Атаева О. М.* Информационная модель открытой персональной семантической библиотеки LibMeta // Научный сервис в сети Интернет: Тр. XVIII Всеросс. науч. конф. — М.: ИПМ им. М. В. Келдыша, 2016. С. 304–313.
4. *Нуеен М. Х., Аджиев А. С.* Описание и использование тезаурусов в информационных системах, подходы и реализация // Электронные библиотеки, 2004. Т. 7. № 1. С. 16–45.
5. *Gruber T. R.* A translation approach to portable ontologies // Knowl. Acquis., 1993. Vol. 5. No. 2. P. 199–220.
6. *Bizer C., Heath T., Berners-Lee T.* Linked data — the story so far // Int. J. Semantic Web Inf., 2009. Vol. 5. No. 3. P. 1–22.
7. *Моисеев Е. И., Муромский А. А., Тучкова Н. П.* Тезаурус информационно-поисковый по предметной области «обыкновенные дифференциальные уравнения». — М.: МАКС Пресс, 2005. 116 с.
8. Математическая энциклопедия: В 5 т. / Гл. ред. И. М. Виноградов. — М.: Советская энциклопедия, 1977.

Поступила в редакцию 03.05.17

<sup>1</sup><http://dublincore.org>.

<sup>2</sup><http://xmlns.com/foaf/spec>.

<sup>3</sup><http://www.mathnet.ru>.

<sup>4</sup><https://www.encyclopediaofmath.org>.

## ONTOLOGY OF THE DIGITAL SEMANTIC LIBRARY LibMeta

V. A. Serebryakov and O. M. Ataeva

A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** During development of digital libraries, particular attention is paid to the library content data model. In this case, the content of digital libraries can be described in various formats and presented in various ways. The library defined by the LibMeta system is considered as a storehouse of structured diverse data with the possibility of their integration with other data sources and assumes the possibility of specifying its content by describing the subject area. The ontology of the semantic library content serves as a means of formalization. It also introduces the basic concepts for describing the task of data integration from sources of Linked Open Data (LOD), concepts for defining an arbitrary thesaurus. The ontology is constructed in such a way that it is possible to determine the semantic library in an arbitrary domain.

**Keywords:** semantic library; data model; ontology; data source; search in LOD

**DOI:** 10.14357/19922264180101

### References

1. Serebryakov, V. A., and O. M. Ataeva. 2014. Personal'naya tsifrovaya biblioteka LibMeta kak sreda integratsii svyazannykh otkrytykh dannykh [Personal Digital Library Libmeta as an integration environment of linked data]. *Tr. XVI Vseross. nauch. konf. RCDL'2014* [16th All-Russia Scientific Conference RCDL'2014 Proceedings]. Dubna: OIYI. 66–71.
2. Serebryakov, V. A., and O. M. Ataeva. 2015. Osnovnye ponyatiya formal'noy modeli semanticheskikh bibliotek i formalizatsiya protsessov integratsii v ney [The basic concepts of a formal model of semantic libraries and formalization of the integration processes in it]. *Programmye produkty i sistemy* [Software Systems] 4:180–187.
3. Serebryakov, V. A., and O. M. Ataeva. 2016. Informatsionnaya model' otkrytoy personal'noy semanticheskoy biblioteki LibMeta [Information model of the open personal semantic library LibMeta]. *Nauchnyy servis v seti Internet: Tr. XVIII Vseross. nauch. konf.* [Scientific service in the Internet: 18th All-Russia Scientific Conference Proceedings]. Moscow: IPM. 304–313.
4. Nguen, M. H., and A. S. Adzhiev. 2004. Opisanie i ispol'zovanie tezaurusov v informatsionnykh sistemakh, podkhody i realizatsiya [Description and use of thesauri in information systems, approaches and implementation]. *Elektronnye biblioteki* [Digital Library] 7(1):16–45.
5. Gruber, T. R. 1993. A translation approach to portable ontologies. *Knowl. Acquis.* 5(2):199–220.
6. Bizer, C., T. Heath, and T. Berners-Lee. 2009. Linked data — the story so far. *Int. J. Semantic Web Inf.* 5(3):1–22.
7. Moiseev, E. I., A. A. Muromskiy, and N. P. Tuchkova. 2005. *Tezaurus informatsionno-poiskovyy po predmetnoy oblasti "obyknovennye differentsial'nye uravneniya"* [Information search with thesaurus in application area of ordinary differential equations]. Moscow: MAKS Press. 116 p.
8. Vinogradov, I. M., ed. 1977. *Matematicheskaya enciklopediya: V 5 t.* [Mathematical encyclopedia: In 5 vols.]. Moscow: Sovetskaya Entsiklopediya.

Received May 3, 2017

### Contributors

**Ataeva Olga M.** (b. 1978) — junior scientist, A. A. Dorodnicyn Computing Centre, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; oli@ultimeta.ru

**Serebryakov Vladimir A.** (b. 1946) — Doctor of Science in physics and mathematics, professor, Head of Department, A. A. Dorodnicyn Computing Centre, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; serebr@ultimeta.ru

# ДИАГРАММЫ УЯЗВИМОСТИ ПОТОКОВЫХ СЕТЕВЫХ СИСТЕМ

Ю. Е. Малашенко<sup>1</sup>, И. А. Назарова<sup>2</sup>, Н. М. Новикова<sup>3</sup>

**Аннотация:** Рассматривается метод анализа изменений функциональных возможностей сетевой системы после разрушающих воздействий. Для описания процессов передачи потоков взаимозаменяемых продуктов различным равноправным пользователям используется модель однопродуктовой сети. Для каждого возможного случая выхода из строя физических или логических элементов системы вычисляются оценки ущерба для всех пользователей. Под ущербом от повреждений понимается величина невыполненных требований потребителей на объемы поставок. В качестве оценок используются оптимальные решения последовательности задач минимизации суммарных относительных ущербов. На основании полученных результатов для различных конфигураций разрушений строятся диаграммы уязвимости исходной сетевой системы. Диаграммы позволяют проводить априорный анализ как небольших, так и критически опасных повреждений, вследствие которых потоки к некоторым пользователям оказываются равными нулю. Предложенный метод может быть использован при исследовании структурно-функциональной уязвимости потоковых сетевых систем.

**Ключевые слова:** однопродуктовая потоковая сеть; функциональная уязвимость; оценки ущерба

**DOI:** 10.14357/19922264180102

## 1 Введение

В литературе в качестве меры уязвимости обычно используется оценка снижения работоспособности системы вследствие полного или частичного разрушения ее элементов [1]. В [2, 3] для перераспределения потоков после повреждения, оценки ущерба пользователей и поиска узких мест сети решалась последовательность задач дискретной оптимизации. В [4] на примере однопродуктовой потоковой модели изучались возможности анализа изменений функциональных характеристик при понижении пропускной способности дуг сетевой системы.

В настоящей работе рассматривается многостокковая потоковая система, структура которой задается графом. Функциональные возможности многополюсной сети определяются величинами потоков, которые могут передаваться одновременно по всем стоковым дугам, направленным в единый фиктивный узел-сток. Считается, что точно известен вектор требований на передачу потока каждому из пользователей сети. Повреждение задается подмножеством разрушенных дуг, пропускная способность которых полагается равной нулю.

В [5] был предложен метод оценки изменения функциональных возможностей сети. Для построения гарантированных оценок — апостериорно

ри — последовательно определяется лексикографический векторный максимум величин потоков, передаваемых в сети после повреждения. В настоящей работе предлагается — априори — находить оценки ущерба для всех пользователей сети при различных повреждающих воздействиях и на их основе строить диаграммы уязвимости исходной системы. Перейдем к описанию модели.

## 2 Модель сети

Рассмотрим проблему оценки ущерба пользователей потоковой системы после крупномасштабных повреждений и/или целенаправленных разрушающих воздействий. Следуя [6], сеть передачи единственного вида продукта, или однопродуктовую многополюсную сетевую систему, будем описывать ориентированным графом  $\bar{G} = (\bar{V}, \bar{L})$  без петель, который определяется множеством вершин (узлов)  $\bar{V} = \{v_1, v_2, \dots, v_N\}$ , где  $|\bar{V}| = N$ , и множеством направленных дуг  $\bar{L} = \{l_{ij} \mid i \in \mathcal{N}, j \in \mathcal{N}, i \neq j\}$ , соединяющих вершины, где  $\mathcal{N}$  — множество индексов вершин. Здесь  $l_{ij} = (v_i, v_j)$  — дуга, ведущая из вершины  $v_i$  в вершину  $v_j$ ,  $|\bar{L}| = L$ .

Обозначим через  $\mathcal{V}_S$  и  $\mathcal{V}_R$  множества вершин графа  $\bar{G}$ , являющихся соответственно источниками и стоками для потока, который передается по многополюсной сети;  $\mathcal{N}_S$  и  $\mathcal{N}_R$  — множества

<sup>1</sup>Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, malash09@ccas.ru

<sup>2</sup>Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, irina-nazar@yandex.ru

<sup>3</sup>Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, n\_novikova@umail.ru

индексов вершин-источников и вершин-стоков:  $\mathcal{V}_\ominus = \{v_i \mid i \in \mathcal{N}_\ominus\}$ ,  $|\mathcal{V}_\ominus| = S$ ,  $\mathcal{V}_\ominus \subset \mathcal{V}$ ,  $\mathcal{N}_\ominus \subset \mathcal{N}$ ,  $S \geq 1$ ,  $\mathcal{V}_\otimes = \{v_i \mid i \in \mathcal{N}_\otimes\}$ ,  $|\mathcal{V}_\otimes| = K$ ,  $\mathcal{V}_\otimes \subset \mathcal{V}$ ,  $\mathcal{N}_\otimes \subset \mathcal{N}$ ,  $K \geq 1$ ,  $\mathcal{V}_\ominus \cap \mathcal{V}_\otimes = \emptyset$ . Считается, что на дугах графа  $\overline{\mathcal{G}}$  заданы веса — значения  $d_{ij}$  пропускной способности дуг  $l_{ij}$ . Вектор  $d$  определяет максимально достижимую величину потока по дугам,  $d = \{d_{ij} \mid d_{ij} \geq 0, l_{ij} \in \overline{\mathcal{L}}\}$ . К графу  $\overline{\mathcal{G}}$  добавим:

$v_0$  — единственный источник потока бесконечной мощности и дуги  $(v_0, v_j)$ ,  $j \in \mathcal{N}_\ominus$ , соединяющие  $v_0$  с каждым узлом-источником. Для каждой дуги  $l_{0j}$  определим верхнее ограничение  $d_{0j}$ , которое соответствует величине максимального потока из источника  $v_j$  в систему. Будем считать, что дуги  $(v_0, v_j)$ ,  $j \in \mathcal{N}_\ominus$ , являются дугами-источниками, и обозначим их множество через  $\hat{\mathcal{L}} = \{l_{0j} \mid j \in \mathcal{N}_\ominus\}$ ,  $|\hat{\mathcal{L}}| = S$ ;

$v_{N+1}$  — единственный узел-сток бесконечного объема и дуги  $(v_i, v_{N+1})$ ,  $i \in \mathcal{N}_\otimes$ , соединяющие каждый узел-сток с  $v_{N+1}$ . Для каждой дуги  $l_{i(N+1)}$  определим верхнее ограничение  $d_{i(N+1)}$ , которое соответствует величине максимального возможного потока из системы в  $v_{N+1}$ . Значение  $d_{i(N+1)}$  задает верхний предел для величины потока, который покидает систему по дуге из узла  $v_i$ . Назовем дуги  $(v_i, v_{N+1})$ ,  $i \in \mathcal{N}_\otimes$ , дугами-стоками, или стоковыми дугами, и обозначим их множество через  $\tilde{\mathcal{L}} = \{l_{i(N+1)} \mid i \in \mathcal{N}_\otimes\}$ ,  $|\tilde{\mathcal{L}}| = K$ . Ориентированный граф, который определяется множествами вершин  $\mathcal{V} = \overline{\mathcal{V}} \cup \{v_0, v_{N+1}\}$  и дуг  $\mathcal{L} = \tilde{\mathcal{L}} \cup \overline{\mathcal{L}} \cup \hat{\mathcal{L}}$ , обозначим  $\mathcal{G} = \langle \mathcal{V}, \mathcal{L} \rangle$ .

Для графа  $\mathcal{G}$  введем обозначения:  $x_{ij}$  — поток по дуге  $l_{ij}$ ,  $l_{ij} \in \mathcal{L}$ , протекающий в соответствии с ее направлением;  $\mathcal{N}_j^-$  — множество индексов узлов-предшественников  $j$ -го (узлов, из которых исходят дуги, ведущие в  $j$ -й узел),  $\mathcal{N}_j^- \subset \mathcal{N} \cup \{0\}$ ;  $\mathcal{N}_j^+$  — множество индексов узлов-последователей  $j$ -го (узлов, в которые ведут дуги, исходящие из  $j$ -го узла),  $\mathcal{N}_j^+ \subset \mathcal{N} \cup \{N+1\}$ .

Поток  $x = \langle x_{0j}, \dots, x_{ij}, \dots, x_{i(N+1)} \rangle$ , где

$$i \in \mathcal{N} \cup \{0\}, j \in \mathcal{N} \cup \{N+1\}, i \neq j, l_{ij} \in \mathcal{L},$$

и если  $i = 0$ , то  $j \neq N+1$ , (1)

проходящий по дугам  $l_{ij} \in \mathcal{L}$ , должен удовлетворять:

- условию сохранения потока в транзитных узлах, т. е.

$$\sum_{i \in \mathcal{N}_j^-} x_{ij} = \sum_{i \in \mathcal{N}_j^+} x_{ji}, \quad j \in \mathcal{N}, \quad (2)$$

- ограничению на пропускную способность соответствующих дуг, т. е.

$$0 \leq x_{ij} \leq d_{ij}, \quad l_{ij} \in \mathcal{L}, \quad \text{для } i, j \text{ выполняется (1)}. \quad (3)$$

Обозначим через  $\mathcal{X}$  множество всех допустимых потоков в сети,

$$\mathcal{X} = \{x \mid \text{выполняется (1)–(3)}\}.$$

При анализе функциональных возможностей системы будем рассматривать потоки по стоковым дугам. Последние перенумеруем по некоторому правилу натуральными числами от 1 до  $K$ , т. е. установим взаимно однозначное соответствие  $l_k = l_{j(N+1)}$ ,  $k = \overline{1, K}$ ,  $j \in \mathcal{N}_\otimes$ .

Пусть  $d_k$  — пропускная способность стоковой дуги  $l_k$ ,  $d_k = d_{j(N+1)}$ ,  $j \in \mathcal{N}_\otimes$ ,  $\bar{x}_k$  — величина потока по стоковой дуге  $l_k$ ,  $\bar{x}_k = x_{j(N+1)}$ ,  $j \in \mathcal{N}_\otimes$ . Таким образом, вектор  $\bar{x} = \langle \bar{x}_1, \dots, \bar{x}_k, \dots, \bar{x}_K \rangle$  покомпонентно определяет величины потоков, которые передаются по каждой стоковой дуге сети в соответствии с некоторым допустимым потоком  $x \in \mathcal{X}$ . Обозначим множество всех допустимых векторов величин потоков по стоковым дугам  $\bar{x}$  через

$$\overline{\mathcal{X}} = \{\bar{x} \mid \bar{x}_k = x_{j(N+1)}, k = \overline{1, K}, j \in \mathcal{N}_\otimes, x \in \mathcal{X}\}.$$

Предположим, что за каждой стоковой дугой стоит некий абстрактный пользователь (потребитель), имеющий определенное требование (запрос)  $f_k$ ,  $k = \overline{1, K}$ , на обеспечение продуктом, который передается по сети. Вектор  $f = \langle f_1, \dots, f_k, \dots, f_K \rangle$ ,  $0 \leq f_k \leq d_k$ ,  $k = \overline{1, K}$ , описывает требования всех потребителей сети. Считается, что недопоставка продукта одному пользователю не может быть восполнена за счет поставки дополнительного объема продукта другому пользователю. Указанное предположение делает пользователей сети в определенном смысле равноправными, а их требования и потоки — не взаимозаменяемыми.

### 3 Оценка ущерба поврежденной сети

Пусть до момента  $t_0$  однопродуктовая сеть работала в некотором стационарном режиме и по дугам сети передавался поток  $x^0 = \langle x_{0j}^0, \dots, x_{ij}^0, \dots, x_{i(N+1)}^0 \rangle$ ,  $x^0 \in \mathcal{X}$ , в том числе по стоковым дугам — вектор  $\bar{x}^0 = \langle \bar{x}_1^0, \dots, \bar{x}_k^0, \dots, \bar{x}_K^0 \rangle$ ,  $\bar{x}^0 \in \overline{\mathcal{X}}$ ,  $\bar{x}_k^0 > 0$ ,  $k = \overline{1, K}$ . Поток  $x^0$  будем называть исходным, а величины компонент  $x_{ij}^0$  — исходными значениями  $x^0$  до повреждающего воздействия.

Пусть стационарно работающая однопродуктовая сеть подверглась в момент  $t_0$  воздействию  $W$ , в результате которого были разрушены несколько дуг  $l_{ij}$ . В рамках рассматриваемой постановки



предполагается, что могут быть повреждены любые дуги, кроме стоковых. Однако, какие дуги сети будут разбиты, заранее не известно, а множество вершин остается неизменным. Пропускная способность разрушенных дуг полагается равной нулю. Обозначим через  $\mathcal{L}(W)$  множество поврежденных дуг. Тогда  $\mathcal{L}(W) \subset [\overline{\mathcal{L}} \cup \hat{\mathcal{L}}]$ ,  $\mathcal{L}(W) \cap \hat{\mathcal{L}} = \emptyset$ ;  $\mathcal{G}(W)$  — граф сети;  $\mathcal{S}(W) = \langle \mathcal{G}(W); f \rangle$  — сеть после разрушающего воздействия  $W$ . Таким образом, воздействие полностью определяется множеством дуг  $\mathcal{L}(W)$ . Пусть  $d(W)$  — вектор пропускной способности дуг поврежденной сети  $\mathcal{S}(W)$ , для каждой компоненты  $d_{ij}(W)$  которого выполняется:

$$d_{ij}(W) = \begin{cases} 0, & \text{если } l_{ij} \in \mathcal{L}(W), \mathcal{L}(W) \subset [\overline{\mathcal{L}} \cup \hat{\mathcal{L}}]; \\ d_{ij}, & \text{если } l_{ij} \in \mathcal{L} \setminus \mathcal{L}(W). \end{cases} \quad (4)$$

Для допустимого потока  $x(W)$  после разрушения должно выполняться (1). Распределение потоков по дугам сети  $\mathcal{S}(W)$  описывается компонентами вектора  $x(W) = \langle x_{0j}(W), \dots, x_{ij}(W), \dots, x_{i(N+1)}(W) \rangle$ .

Для любых допустимых потоков  $x(W)$  в поврежденной сети  $\mathcal{S}(W)$  должны выполняться стандартные ограничения

$$0 \leq x_{ij}(W) \leq d_{ij}(W), \quad (5)$$

где  $d_{ij}(W)$  определяется (4),  $l_{ij} \in \mathcal{L}$ , и закон сохранения потока в каждом узле

$$\sum_{i \in \mathcal{N}_j^-} x_{ij}(W) = \sum_{i \in \mathcal{N}_j^+} x_{ji}(W), \quad j \in \mathcal{N}. \quad (6)$$

Множество  $\mathcal{X}(W)$  допустимых потоков в поврежденной сети  $\mathcal{S}(W)$  определяется условиями (1), (5) и (6):

$$\mathcal{X}(W) = \{x(W) \mid \text{выполняется (1), (5), (6)}\}.$$

Пусть  $\overline{x}_k(W)$  — величина потока по  $k$ -й стоковой дуге после разрушающего воздействия. Тогда  $\overline{x}(W)$  — вектор величин потоков по всем стоковым дугам,  $\overline{x}(W) = \langle \overline{x}_1(W), \dots, \overline{x}_k(W), \dots, \overline{x}_K(W) \rangle$ . Множество

$$\overline{\mathcal{X}}(W) = \{\overline{x}(W) \mid \overline{x}_k(W) = x_{j(N+1)}(W), \\ k = \overline{1, K}, j \in \mathcal{N}_{\overline{R}}, x(W) \in \mathcal{X}(W)\}$$

описывает достижимые величины потоков по стоковым дугам после разрушения.

В дальнейших рассуждениях значения компонент вектора требований потоков по стоковым дугам сети положим равными соответствующим исходным величинам до повреждения, т. е.  $f_k = \overline{x}_k^0$ ,

$k = \overline{1, K}$ . Кроме того, поскольку считается, что недопоставка продукта одному пользователю не может быть восполнена за счет поставки дополнительного объема продукта другому, то выполняется  $\overline{x}_k(W) \leq \overline{x}_k^0$ ,  $k = \overline{1, K}$ . Введем величину  $u_k(W)$ , численно равную разности между исходным значением, равным требованию  $k$ -го пользователя, и значением потока, протекающего по  $k$ -й стоковой дуге после разрушающего воздействия:

$$u_k(W) = f_k - \overline{x}_k(W) = \overline{x}_k^0 - \overline{x}_k(W), \\ \overline{x}_k(W) \leq \overline{x}_k^0, \quad k = \overline{1, K},$$

которую назовем недопоставкой потока  $k$ -му потребителю, или  $k$ -й недопоставкой. Вектор  $u(W) = \langle u_1(W), \dots, u_k(W), \dots, u_K(W) \rangle$  описывает возможные недопоставки всем пользователям сети после повреждения.

Назовем ущербом  $k$ -го потребителя отношение величины недопоставки  $u_k(W)$  к соответствующему запрошенному количеству потока  $f_k = \overline{x}_k^0$ ,  $k = \overline{1, K}$ , и обозначим эту величину через

$$\omega_k(W) = \frac{u_k(W)}{f_k} = \frac{f_k - \overline{x}_k(W)}{f_k}, \\ 0 \leq \omega_k \leq 1, \quad k = \overline{1, K}.$$

Для оценки суммарного ущерба пользователей решим следующую задачу оптимизации.

**Задача А.** Найти

$$\Omega^* = \min_{x(W) \in \mathcal{X}(W)} \sum_{k=\overline{1, K}} \left( \frac{f_k - \overline{x}_k(W)}{f_k} \right)^2$$

при условии  $f_k = \overline{x}_k^0 \neq 0$ ,  $k = \overline{1, K}$ .

Оптимальное решение задачи А обозначим  $\Omega^*$ . Если  $\Omega^* = 0$ , то исходный вектор  $\overline{x}^0$  можно передать в поврежденной сети. В противном случае значение  $\Omega^*$  равно минимальной сумме квадратов ущерба пользователей от повреждения  $W$ . Поток, обеспечивающий решение задачи — достижение минимального значения  $\Omega^*$ , обозначим  $x^*(W)$ ,  $x^*(W) \in \mathcal{X}(W)$ . Вектору  $x^*(W)$  соответствует вектор  $\overline{x}^*(W)$ , принадлежащий множеству  $\overline{\mathcal{X}}(W)$ , каждая компонента которого отвечает величине потока по стоковой дуге. В случае, когда  $\Omega^* \neq 0$ ,  $\overline{x}^*(W)$  является точкой, ближайшей к вектору требований  $f$ .

Для каждой  $k$ -й стоковой дуги ( $k = \overline{1, K}$ ) вычислим величину  $\omega_k^*(W)$  — значение ущерба  $k$ -го пользователя сети при разрушении  $W$ :

$$\omega_k^*(W) = \frac{f_k - \overline{x}_k^*(W)}{f_k}$$

при условии  $f_k = \overline{x}_k^0 \neq 0$ ,  $k = \overline{1, K}$ .



Следуя [7], отношение величины потока, протекающего по  $k$ -й стоковой дуге после разрушения, к исходному обозначим через  $\theta_k^*(W)$ :

$$\theta_k^*(W) = \frac{\bar{x}_k^*(W)}{f_k} = 1 - \omega_k^*(W)$$

при условии  $f_k = \bar{x}_k^0 \neq 0$ ,  $k = \overline{1, K}$ . Величину  $\theta_k^*(W)$  будем называть мерой обеспеченности требований  $k$ -го потребителя после разрушающего воздействия  $W$ .

Для фиксированного повреждения  $W$  определим величины максимального  $\bar{\omega}^*(W)$  и минимального  $\underline{\omega}^*(W)$  ущерба пользователей сети. По определению

$$\bar{\omega}^*(W) = \max_{k=\overline{1, K}} \omega_k^*(W); \quad \underline{\omega}^*(W) = \min_{k=\overline{1, K}} \omega_k^*(W).$$

Тогда

$$\begin{aligned} \bar{\theta}^*(W) &= \max_{k=\overline{1, K}} \theta_k^*(W) = \max_{k=\overline{1, K}} (1 - \omega_k^*(W)) = \\ &= 1 - \min_{k=\overline{1, K}} \omega_k^*(W) = 1 - \underline{\omega}^*(W); \\ \underline{\theta}^*(W) &= \min_{k=\overline{1, K}} \theta_k^*(W) = \min_{k=\overline{1, K}} (1 - \omega_k^*(W)) = \\ &= 1 - \max_{k=\overline{1, K}} \omega_k^*(W) = 1 - \bar{\omega}^*(W). \end{aligned}$$

Заметим, что пользователей, получивших максимальный (минимальный) ущерб от повреждения  $W$  и, соответственно, имеющих минимальную (максимальную) обеспеченность требований, в зависимости от конфигурации сети может быть любое число от 1 до  $K$ .

Еще одним важным показателем для фиксированного повреждения  $W$  является медиана ущерба  $\tilde{\omega}(W)$ . Для построения медианы  $\tilde{\omega}(W)$  для каждого повреждения  $W$  определим такое значение  $\tilde{\omega}(W)$ , что последнее делит множество значений  $\omega_k^*(W)$  для выбранного  $W$  на две равные части, т.е.

- в половине случаев значения  $\omega_k^*(W) \leq \tilde{\omega}(W)$ ;
- в другой половине случаев  $\tilde{\omega}(W) \leq \omega_k^*(W)$ .

При нечетном  $k$  в качестве оценки медианы выберем среднее по порядку значение, а при четном — полусумму двух средних по порядку значений. Медиану обеспеченности требований  $\tilde{\theta}(W)$  определим аналогично  $\tilde{\omega}(W)$ .

## 4 Диаграммы структурно-функциональной уязвимости

В настоящей работе под мощностью повреждения  $W$  будем понимать число разрушенных дуг.

Напомним, что разбиты могут быть любые дуги, кроме стоковых. Рассмотрим все повреждения, имеющие мощность 3. Пусть  $R(3)$  — общее число таких повреждений:

$$R(3) = \frac{(L+S)(L+S-1)(L+S-2)}{6}.$$

Перенумеруем по некоторому правилу все повреждения, имеющие мощность 3, и обозначим через  $\overline{\mathcal{W}}(3)$  множество всех таких повреждений:

$$\overline{\mathcal{W}}(3) = \{\overline{W}_1, \dots, \overline{W}_w, \dots, \overline{W}_{R(3)}\}.$$

Каждое из повреждений  $\overline{W}_w \in \overline{\mathcal{W}}(3)$  полностью определяется тройкой разрушенных дуг сети  $(l_{ab}, l_{ij}, l_{km})$ . Таким образом, устанавливается взаимно однозначное соответствие между тройками разрушенных дуг и номером  $w$  разрушающего воздействия  $\overline{W}_w$ . Для каждого повреждения  $\overline{W}_w \in \overline{\mathcal{W}}(3)$  решим задачу  $A$  и вычислим вектор величин потоков  $\bar{x}^*(\overline{W}_w)$ , далее для  $k = \overline{1, K}$  найдем значения переменных  $\omega_k^*(\overline{W}_w)$ ,  $\theta_k^*(\overline{W}_w)$  и соответствующие  $\bar{\omega}^*(\overline{W}_w)$ ,  $\underline{\omega}^*(\overline{W}_w)$ ,  $\tilde{\omega}^*(\overline{W}_w)$ ,  $\bar{\theta}^*(\overline{W}_w)$ ,  $\underline{\theta}^*(\overline{W}_w)$  и  $\tilde{\theta}^*(\overline{W}_w)$ .

Исходя из полученных значений  $\underline{\theta}^*(\overline{W}_w)$  и  $\bar{\omega}^*(\overline{W}_w)$ ,  $w = \overline{1, R(3)}$ , переставим и переобозначим элементы множества  $\overline{\mathcal{W}}(3)$  так, чтобы значения  $\underline{\theta}^*(\overline{W}_w)$  были упорядочены по неубыванию. Тогда первым элементом множества  $\mathcal{W}(3) = \{W_1, \dots, W_w, \dots, W_{R(3)}\}$  будем считать такое разрушение мощности 3 — тройку  $(l_{ab}, l_{ij}, l_{km})$ , которое обеспечивает минимальное значение  $\underline{\theta}^*(\overline{W}_w)$  и, соответственно, максимальное значение  $\bar{\omega}^*(\overline{W}_w)$  среди всех  $\overline{W}_w \in \overline{\mathcal{W}}(3)$ :

$$W_1 = \overline{W}_j : \underline{\theta}^*(\overline{W}_j) = \min_{\overline{W}_w \in \overline{\mathcal{W}}(3)} \underline{\theta}^*(\overline{W}_w).$$

При этом будем считать, что для всех  $k = \overline{1, K}$  выполнено  $\omega_k^*(W_1) = \omega_k^*(\overline{W}_j)$ ,  $\theta_k^*(W_1) = \theta_k^*(\overline{W}_j)$  и  $\bar{\omega}^*(W_1) = \bar{\omega}^*(\overline{W}_j)$ ,  $\underline{\omega}^*(W_1) = \underline{\omega}^*(\overline{W}_j)$ ,  $\tilde{\omega}^*(W_1) = \tilde{\omega}^*(\overline{W}_j)$ ,  $\bar{\theta}^*(W_1) = \bar{\theta}^*(\overline{W}_j)$ ,  $\underline{\theta}^*(W_1) = \underline{\theta}^*(\overline{W}_j)$  и  $\tilde{\theta}^*(W_1) = \tilde{\theta}^*(\overline{W}_j)$ .

Аналогично

$$W_2 = \overline{W}_g : \underline{\theta}^*(\overline{W}_g) = \min_{\overline{W}_w \in [\overline{\mathcal{W}}(3) \setminus \{\overline{W}_j\}] } \underline{\theta}^*(\overline{W}_w).$$

И так далее.

Данные о повреждениях — множестве  $\mathcal{W}(3)$  — представлены на рис. 1 в виде диаграммы, построенной по следующему правилу. На рис. 1 проведена горизонтальная жирная линия, проходящая через точку  $(0, 1)$ . По оси абсцисс отложены числа

$$\frac{1}{R(3)}, \frac{2}{R(3)}, \frac{3}{R(3)}, \dots, \frac{w}{R(3)}, \dots, \frac{R(3)}{R(3)},$$

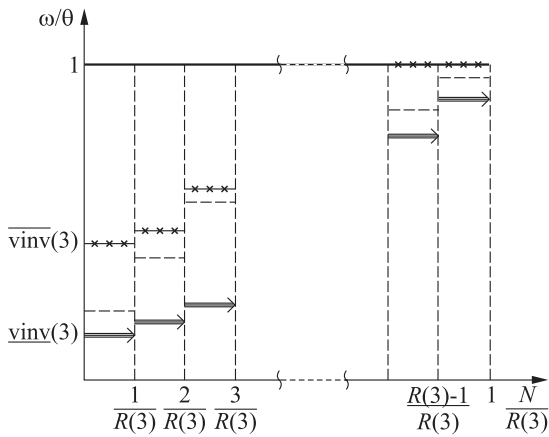


Рис. 1 Диаграмма ущерба

с помощью которых определяется ширина столбцов, соответствующих номерам повреждений, а именно: повреждению  $W_1$  соответствует отрезок  $(0, 1/R(3))$ , повреждению  $W_2$  — отрезок  $(1/R(3), 2/R(3))$ , ..., повреждению  $W_{R(3)}$  — отрезок  $((R(3) - 1)/R(3), R(3)/R(3))$ . Высота столбца по оси ординат численно равна величине  $\underline{\theta}^*(W_w)$ ,  $w = \overline{1, R(3)}$ , и отмечена линией со стрелкой. Высота столбца от линии со стрелкой до горизонтальной жирной линии численно равна  $\overline{\omega}^*(W_w) = 1 - \underline{\theta}^*(W_w)$ ,  $w = \overline{1, R(3)}$ . Штриховые линии соответствуют значениям  $\tilde{\theta}^*(W_w)$ , а линии со звездочками — значениям  $\overline{\theta}^*(W_w)$ ,  $w = \overline{1, R(3)}$ . На рис. 1 величина максимального ущерба пользователей после повреждения любых трех дуг сети слева направо не увеличивается. При этом  $\underline{\theta}^*(W_w)$ ,  $w = \overline{1, R(3)}$ , численно равна доле требований, гарантированно выполненных для любого пользователя при повреждении  $W_w \in \mathcal{W}(3)$ , и эти значения слева направо не убывают.

В литературе ущерб пользователей при повреждениях трактуется как мера уязвимости: чем больше ущерб, тем более уязвимой считается сеть. И наоборот — чем больше поток, который гарантированно можно передать по сети после повреждения, тем сеть менее уязвима. Диаграмму на рис. 1 будем называть VINV(3)-диаграммой (по числу поврежденных дуг и от английских слов vulnerable — уязвимый, invulnerable — неуязвимый). Обозначим через  $\underline{vinv}(3)$  кусочно-постоянную ступенчатую линию, отвечающую значениям  $\underline{\theta}^*(W_w)$ ,  $w = \overline{1, R(3)}$ . Выше линии  $\underline{vinv}(3)$  лежит область значений ущерба, который может быть нанесен пользователям сети, а ниже — область, отвечающая доле требований, которые гарантированно могут быть доставлены пользователям при повреждениях, имеющих мощность три.

Обозначим через  $\overline{vinv}(3)$  кусочно-постоянную ступенчатую линию со звездочками, отвечающую значениям  $\overline{\theta}^*(W_w)$ ,  $w = \overline{1, R(3)}$ . Выше линии  $\overline{vinv}(3)$  лежит область, отвечающая пользователям с «минимальным» ущербом, или тем, кто получил максимальную долю от требований при имеющихся повреждениях сети. Между кривыми  $\underline{vinv}(3)$  и  $\overline{vinv}(3)$  лежит область, позволяющая анализировать ущерб пользователей, опираясь на ступенчатую пунктирную линию значений медиан  $\tilde{\theta}^*(W_w)$ ,  $\tilde{\omega}^*(W_w)$ . Таким образом, с помощью кривых  $\underline{vinv}(3)$  и  $\overline{vinv}(3)$  можно анализировать способность/неспособность системы обеспечивать передачу определенной доли потока в разрушенной сети, т. е. судить об уязвимости/неуязвимости исследуемой сети в терминах сохраненный поток / понесенный ущерб, а также выявлять критически опасные повреждения, вследствие которых потоки к некоторым пользователям оказываются равными нулю. Заметим, что процедуру, описанную выше для построения VINV(3)-диаграммы, можно использовать для изучения функциональных возможностей сети после повреждения любого числа дуг.

На рис. 2 представлена диаграмма, позволяющая сравнивать последствия разрушений по двум расчетным параметрам: максимальный ущерб и величина медианы ущерба. По оси абсцисс отложены значения медиан, по оси ординат — соответствующие значения максимального ущерба. Каждое разрушение отмечено крестиком, крестики в кружочках соответствуют угловым точкам, лежащим на выпуклой оболочке множества всех повреждений мощности три. Повреждения, принадлежащие выпуклой оболочке, будем называть эффективными, поскольку их показатели не доминируются сразу по всем параметрам другими разрушающими воздействиями.

К построению VINV(n)-диаграммы для  $n \geq 1$  можно подойти иначе. Формирование VINV(n)-

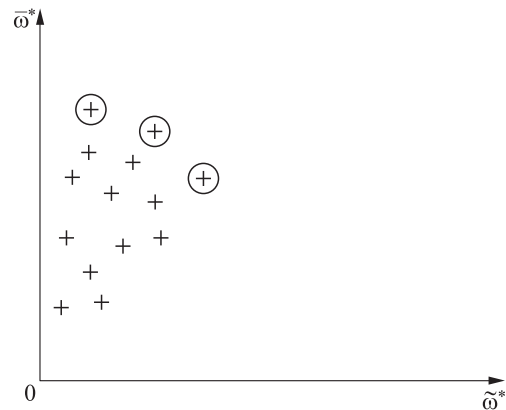


Рис. 2 Диаграмма эффективных повреждений

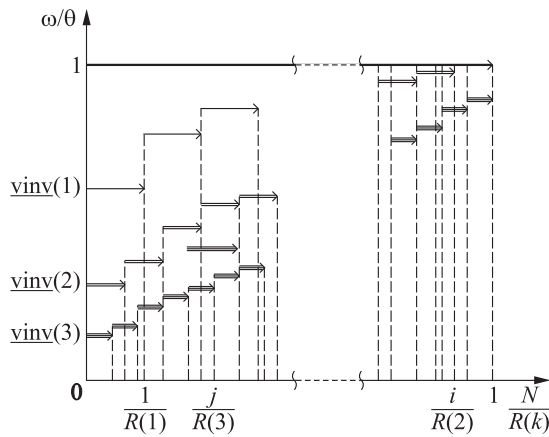


Рис. 3 Диаграмма ущерба для различной мощности повреждения

диаграммы можно начинать с анализа разрушения одной, двух, . . . ,  $n$  дуг, или до получения критических повреждений, при этом рассматривая каждую предыдущую диаграмму как детализацию следующей. Остановимся на этой вычислительной схеме последовательного построения подробнее. Обозначим через  $\zeta_{ij}$  отношение исходной величины потока  $x_{ij}^0$  по выбранной дуге  $l_{ij}$  к ее пропускной способности  $d_{ij}$  и назовем указанную величину загруженностью дуги  $l_{ij}$ :

$$\zeta_{ij} = \frac{x_{ij}^0}{d_{ij}}, \quad (7)$$

где  $l_{ij} \in [\bar{\mathcal{L}} \cup \hat{\mathcal{L}}]$ ,  $x^0 \in \mathcal{X}$ .

Вычислим загруженность  $\zeta_{ij}$  всех дуг сети, кроме стоковых. Затем, переупорядочив полученные  $\zeta_{ij}$  по невозрастанию, построим VINV( $n$ )-диаграммы в первую очередь для дуг с большими значениями  $\zeta_{ij}$ .

Далее исследование уязвимости сети можно проводить по двум сценариям. В первом случае — выбирать пары, тройки и т. д. разрушенных дуг исходя из вычисленных значений загруженности  $\zeta_{ij}$ ,  $l_{ij} \in [\bar{\mathcal{L}} \cup \hat{\mathcal{L}}]$ . Во втором — корректировать значения загруженности дуг исходя из повреждений сети, а именно: для вычисления загруженности неповрежденной дуги  $l_{ij}$  после фиксированного разрушения  $W$  в формуле (7) вместо исходного значения  $x_{ij}^0$  компоненты вектора потока  $x^0$  использовать соответствующую компоненту  $x_{ij}^*(W)$  вектора  $x^*(W)$  — решения задачи  $A$ ,  $x^*(W) \in \mathcal{X}(W)$ :

$$\zeta_{ij}(W) = \frac{x_{ij}^*(W)}{d_{ij}}, \quad (8)$$

где  $l_{ij} \in [\bar{\mathcal{L}} \cup \hat{\mathcal{L}}] \setminus \mathcal{L}(W)$ ,  $x^*(W) \in \mathcal{X}(W)$ .

Далее выбор очередной дуги для анализа разрушений производить исходя из полученных  $\zeta_{ij}(W)$ ,

выбирая, как и выше, дуги с большими значениями. Таким образом, для поиска каждой следующей дуги множества  $\mathcal{L}(W)$  предлагается использовать формулу (8), а дальнейший анализ уязвимости сети проводить исходя из решения задачи  $A$ . Пример возможного расположения кривых  $\text{vinv}(1)$ ,  $\text{vinv}(2)$ ,  $\text{vinv}(3)$  приведен на рис. 3.

## 5 Заключение

Предложенный метод позволяет проводить априорный анализ уязвимости сетевой системы для произвольного фиксированного вектора требований и любых разрушений. Для вычисления ущерба пользователей сети при удалении произвольных наборов дуг с помощью эффективных алгоритмов [8] решается задача  $A$ , причем столько раз, сколько потребуется для получения значимого описания.

На основании полученных результатов после анализа и преобразований строятся подробные диаграммы оценок возможного ущерба для различных значений мощности повреждений. Диаграммы на качественном уровне демонстрируют зависимость функциональных характеристик от структурных повреждений. Особенно важно, что диаграммы точно отражают критические повреждения, при которых потоки по некоторым стоковым дугам становятся равными нулю. Появление нулевых значений в векторе-решении задачи  $A$  указывает на утрату возможности передачи потока соответствующему пользователю. Анализ списка дуг, приводящих к эффективным и критическим повреждениям, позволяет выявить слабые места и структурные особенности сетевой системы. Более детальное исследование структурно-функциональной уязвимости и получение гарантированных оценок ущерба можно проводить с помощью ресурсоемких методов [5].

## Литература

1. Murray A. T. An overview of network vulnerability modeling approaches // GeoJ., 2013. Vol. 78. P. 209–221.
2. Назарова И. А. Лексикографическая задача анализа уязвимости многопродуктовой сети // Изв. РАН. ТиСУ, 2003. № 5. С. 123–134.
3. Назарова И. А. Модели и методы решения задачи анализа уязвимости сетей // Изв. РАН. ТиСУ, 2006. № 4. С. 61–72.
4. Козлов М. В., Малащенко Ю. Е., Назарова И. А., Новикова Н. М. Управление топливно-энергетической системой при крупномасштабных повреждениях. I. Сетевая модель и программная реализация // Изв. РАН. ТиСУ, 2017. № 6. С. 50–73.

5. Малащенко Ю. Е., Назарова И. А., Новикова Н. М. Метод анализа функциональной уязвимости потоковых сетевых систем // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 47–54.
6. Форд Л., Фалкерсон Д. Потоки в сетях / Пер. с англ. — М.: Мир, 1966. 277 с. (Ford L. R., Fulkerson D. R. Flows in networks. — Princeton, NJ, USA: Princeton University Press, 1962. 332 p.)
7. Малащенко Ю. Е., Новикова Н. М. Модели неопределенности в многопользовательских сетях. — М: Эдиториал УРСС, 1999. 160 с.
8. Йенсен П., Барнес Д. Потокое программирование / Пер. с англ. — М.: Радио и связь, 1984. 392 с. (Jensen P. A., Barnes J. W. Network flow programming. — New York, NY, USA: Wiley, 1980. 408 p.)

Поступила в редакцию 04.12.17

## DIAGRAMS OF THE FUNCTIONAL VULNERABILITY OF FLOW NETWORK SYSTEMS

Yu. E. Malashenko, I. A. Nazarova, and N. M. Novikova

A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The method of analysis of changes in functional capabilities of the flow network system after the damaging effects is considered. To describe the processes of streams of different interchangeable products to various peer users, the model of single-product network is used. The damage is defined as the total amount of unsatisfied demands. Damage estimates are calculated for all users and all possible cases of failure of physical or logical elements of the system. Optimal solutions of the sequence of problems of minimization of total relative damages are used as damage estimates. Based on the obtained results for various destruction configurations, the vulnerability diagrams of the initial network system are constructed. Diagrams allow an *a priori* analysis of both small and critical damages, due to which the flows to some users vanish. The proposed method can be used to study the structural and functional vulnerability of flow network systems.

**Keywords:** single-product flow network; functional vulnerability; damage assessment

**DOI:** 10.14357/19922264180102

### References

1. Murray, A. T. 2013. An overview of network vulnerability modeling approaches. *GeoJ.* 78:209–221.
2. Nazarova, I. A. 2003. Lexicographical problem of the analysis of the vulnerability of a multicommodity network. *J. Comput. Syst. Sci. Int.* 42(5):778–789.
3. Nazarova, I. A. 2006. Models and methods for solving the problem of network vulnerability. *J. Comput. Syst. Sci. Int.* 45(4):567–578.
4. Kozlov, M. V., Yu. E. Malashenko, I. A. Nazarova, and N. M. Novikova. 2017. Fuel and energy system control at large-scale damages. I. Network model and software implementation. *J. Comput. Syst. Sci. Int.* 56(6):945–968.
5. Malashenko, Yu. E., I. A. Nazarova, and N. M. Novikova. 2017. Metod analiza funktsional'noy uязvimosti potokovykh setevykh system [Method of the analysis of the functional vulnerability of flow network systems]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):50–73.
6. Ford, L. R., and D. R. Fulkerson. 1962. Flows in networks. Princeton, NJ: Princeton University Press. 332 p.
7. Malashenko, Ju. E., and N. M. Novikova. 1999. Modeli neopredelennosti v mnogopol'zovatel'skikh setyakh [Indeterminacy models in the multiuser networks]. Moscow: Editorial URSS Publ. 160 p.
8. Jensen, P. A., and J. W. Barnes. 1980. Network flow programming. New York, NY: Wiley. 408 p.

Received December 4, 2017

### Contributors

**Malashenko Yuri E.** (b. 1946) — Doctor of Science in physics and mathematics, Head of Laboratory, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; malash09@ccas.ru

**Nazarova Irina A.** (b. 1966) — Candidate of Science (PhD) in physics and mathematics, scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; irina-nazar@yandex.ru

**Novikova Natalya M.** (b. 1953) — Doctor of Science in physics and mathematics, professor; leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; n\_novikova@umail.ru



## АВТОМАТИЗИРОВАННАЯ СИСТЕМА ВЕСОВОГО ПРОЕКТИРОВАНИЯ САМОЛЕТОВ

Л. Л. Вышинский<sup>1</sup>, Ю. А. Флеров<sup>2</sup>, Н. И. Широков<sup>1</sup>

**Аннотация:** Статья посвящена вопросам автоматизации задач весового проектирования самолетов. Весовые и массово-инерционные параметры являются одними из основных величин, влияющих на эксплуатационные характеристики самолетов. Информационной основой системы служит весовая модель самолета. Описывается структура весовой модели и даны характеристики отдельным ее компонентам. Показана программная реализация системы, которая выполнена в рамках архитектуры клиент–сервер. Автоматизированная система весового проектирования (АСВП) реализована с использованием программно-инструментального комплекса «Генератор проектов» (технология ГП), который был разработан в Вычислительном центре Российской академии наук. Создание информационно-вычислительных систем в рамках технологии ГП базируется на так называемом «проектном подходе», когда по формальному описанию системы автоматически генерируются тексты программного кода для клиентских и серверных компонент системы.

**Ключевые слова:** математическое моделирование; автоматизация проектирования; самолет; весовое проектирование; весовая модель; дерево конструкции; генератор проектов; генерация программного кода; архитектура клиент–сервер

**DOI:** 10.14357/19922264180103

### 1 Введение

Развитие и повсеместное использование информационных технологий за последние несколько десятилетий существенно изменили традиционный процесс проектирования и создания различных инженерных систем, сооружений, машин. Во многих проектных организациях давно отказались от когда-то привычных инструментов конструктора — кульмана и логарифмической линейки. Сейчас первые эскизы новых проектов появляются чаще не на бумаге, как было всегда, а на экране монитора. Этому способствует широкий спектр имеющихся систем автоматизированного проектирования. В российских авиационных конструкторских бюро, например, уже давно применяются такие CAD (computer aided design) системы, как NX (Unigraphics), CATIA и др. Эти развитые системы геометрического трехмерного (3D) моделирования позволяют создавать сложные поверхности, конструировать любые детали, осуществлять сборку узлов, агрегатов и сложнейших изделий. Однако построение геометрических моделей изделий является финальной стадией проектирования, за которой следует их реализация «в металле». Построению электронных

геометрических макетов предшествует и сопутствует решение множества расчетных задач, а также задач анализа и оптимизации в разных областях инженерных знаний. В авиастроении это аэродинамика, динамика полета, прочность, системы управления, двигателестроение и пр. Все эти задачи требуют разработки разноплановых математических моделей и специальных вычислительных программ.

Одной из важнейших технических характеристик самолета является его вес. При решении подавляющего большинства проектных и конструкторских задач весовые параметры в том или ином виде участвуют в расчетах. Необходимость проведения весовых расчетов возникает на самых ранних этапах проектирования и сопровождает все дальнейшие стадии разработки и эксплуатации.

В процессе создания и эксплуатации самолетов постоянно контролируются вес и другие массово-инерционные характеристики (МИХ) всех размещаемых на борту систем, агрегатов, узлов и деталей. Количество агрегатов, узлов и деталей современных самолетов исчисляется десятками тысяч, поэтому в авиастроении весовые расчеты, весовой анализ, весовой контроль выливаются в сложную инженер-

<sup>1</sup>Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, Wysh@ccas.ru

<sup>2</sup>Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, fler@ccas.ru



ную проблему и выделяются в целое направление инженерной деятельности, которое принято называть весовым проектированием [1].

Данная статья посвящена вопросам автоматизации задач весового проектирования самолетов. В разные годы Вычислительным центром РАН был разработан и внедрен в эксплуатацию ряд программ, решающих отдельные задачи весовых расчетов летательных аппаратов (ЛА) [2–4]. В настоящей статье представлено описание интегрированной АСВП, предназначенной для использования на всех этапах жизненного цикла изделий. Она разработана как интерактивная многопользовательская информационная система клиент-серверной архитектуры с централизованной базой данных. Информационным ядром и основой АСВП является единая структурно-параметрическая весовая модель самолета, описание которой дает довольно полное представление о задачах, решаемых с помощью АСВП.

## 2 Структурно-параметрическая весовая модель самолета

Самолет является сложным техническим объектом, состоящим из множества различных компонентов, функционально и конструктивно связанных между собой. Под структурно-параметрической весовой моделью самолета здесь понимается база данных, которая содержит всю необходимую информацию для проведения комплекса расчетов МИХ и массово-центровочных данных (МЦД) самолета. Весовая модель состоит из нескольких структур, ориентированных на определенные группы параметров и задач весового проектирования. Ниже перечислены основные структуры весовой модели, реализованные в системе АСВП:

- дерево конструкции самолета;
- иерархия систем координат, связанных с самолетом и его агрегатами;
- геометрические структуры весовой модели самолета;
- каталог целевой нагрузки, размещаемой во внутренних отсеках и на подвесках;
- реестр допустимых вариантов загрузки самолета;
- таблицы тарифовочных характеристик топливных баков;
- таблицы характеристик выработки топлива.

### 2.1 Дерево конструкции самолета

Дерево конструкции самолета является центральной структурой весовой модели, которая отражает членение изделия на его составные части — системы, агрегаты, узлы, детали. В базе данных весовой модели эта структура представлена в виде многоуровневого корневого дерева  $W = (U, V)$ , где вершинам  $U = \{U_i\}$  соответствуют различные элементы конструкции. Ориентированные дуги дерева, идущие из  $U_i$  в  $U_j$ , означают вхождение конструкции  $U_j$  в конструкцию  $U_i$  в качестве ее составной части. Терминальными или висячими вершинами дерева конструкции будем называть вершины, у которых нет ни одной подчиненной конструкции.

Многолетний опыт самолетостроения выработал устоявшиеся конструктивные схемы самолетов различного назначения. Существуют отраслевые стандарты и классификаторы, которые вводят определения основных элементов конструкции самолетов. На рис. 1 показан пример представления в АСВП нескольких верхних уровней дерева конструкции самолета.

Существующие классификаторы отражают лишь самые общие принципы построения конструкции самолетов. Разумеется, каждый новый проект самолета имеет свои конструктивные особенности, которые отражаются на структуре весовой модели. Дерево конструкции строится постепенно, сверху вниз, в течение всего процесса проектирования самолета.

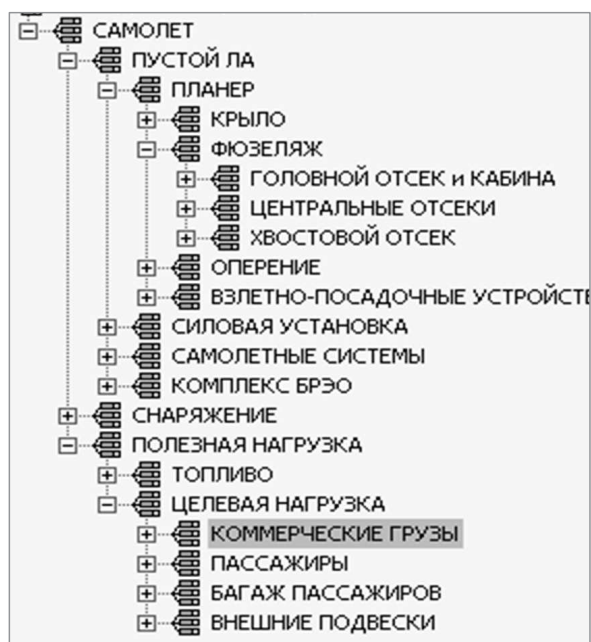


Рис. 1 Дерево конструкции самолета

Понятие «конструкции» в данном контексте используется и для обозначения любой вершины графа, и для всего поддерева подчиненных конструкций этой вершине. Каждая конструкция дерева имеет уникальное в рамках весовой модели обозначение, которое может быть шифром, кодом, идентификатором или чертежным номером конструкции. Разумеется, для более полного и наглядного вербального представления конструкции в структурно-параметрической модели можно задать ее текстовое описание.

**Масса конструкции.** Основную содержательную и необходимую информацию весовой модели содержит список значений МИХ, соответствующих каждой вершине дерева конструкций. Центральным параметром является масса.

На разных стадиях создания самолета, когда неизвестно точное значение массы, прибегают к различным оценкам. В структурно-параметрической весовой модели фиксируются перечисленные ниже оценки массы, которые соответствуют разным этапам проектирования:

$M_{\text{теор}}$  — теоретическая масса — оценка массы, вычисленная на основании некоторой математической модели конструкции;

$M_{\text{лим}}$  — лимитная масса конструкции, устанавливаемая на основании теоретических оценок и используемая для весового контроля в процессе детальной разработки конструкции;

$M_{\text{черт}}$  — чертежная масса конструкции, рассчитанная по чертежу или по электронной геометрической модели конструкции;

$M_{\text{креп}}$  — масса крепежа конструкции — дополнение к чертежной массе, которое учитывает мелкие детали конструкции, предназначенные для соединения подчиненных деталей (защелки, болты, гайки, шайбы и т.п.). Введение такой дополнительной массы позволяет избавиться от десятков и сотен тысяч вершин, которые несут относительно небольшую нагрузку в весовых характеристиках, но тем не менее их учет необходим при контроле веса. Масса крепежа распределяется по подчиненным конструкциям;

$M_{\text{факт}}$  — фактическая масса изготовленной и взвешенной конструкции. Фактическая масса может задаваться не только для изготавливаемых конструкций ЛА, но и для готовых поставляемых изделий при их установке на борту.

Порядок задания оценок массы диктуется логикой развития проекта. В дереве конструкции все оценки массы, кроме  $M_{\text{лим}}$  и  $M_{\text{креп}}$ , суммируются по подчиненным вершинам снизу вверх. Однако

если для некоторых терминальных значений одна или несколько оценок не определены, например некоторые детали конструкции не изготовлены и, стало быть,  $M_{\text{факт}}$  не определена, то и для всех вышестоящих конструкций эти оценки не определены. При задании  $M_{\text{лим}}$  это правило может не соблюдаться.

На основании оценок массы определяется то расчетное значение массы конструкции, которое используется во всех расчетах на текущей стадии проекта:  $M$  — текущая масса конструкции. Значение текущей массы для *нетерминальных* конструкций определяется суммированием по подчиненным конструкциям. Для *терминальных* вершин дерева конструкций применяется процедура определения текущей массы по первому известному значению из следующего списка в указанном порядке:  $M_{\text{факт}}$ ,  $M_{\text{черт}} + M_{\text{креп}}$ ,  $M_{\text{теор}}$ ,  $M_{\text{лим}}$ .

**Геометрия масс конструкции.** Кроме собственно массы в весовой модели задаются или вычисляются значения характеристик, которые принято называть характеристиками геометрии масс:

$X$ ,  $Y$  и  $Z$  — положение центра масс конструкции;  
 $L_x$ ,  $L_y$  и  $L_z$  — габаритные параметры конструкции;

$I_x$ ,  $I_y$  и  $I_z$  — полные плоскостные моменты инерции;

$I_{xy}$ ,  $I_{xz}$  и  $I_{yz}$  — полные центробежные моменты инерции;

$I_x^c$ ,  $I_y^c$  и  $I_z^c$  — собственные плоскостные моменты инерции:

$$I_x^c = I_x - MX^2;$$

$$I_y^c = I_y - MY^2;$$

$$I_z^c = I_z - MZ^2;$$

$I_{xy}^c$ ,  $I_{xz}^c$  и  $I_{yz}^c$  — собственные центробежные моменты инерции:

$$I_{xy}^c = I_{xy} - MXY;$$

$$I_{xz}^c = I_{xz} - MXZ;$$

$$I_{yz}^c = I_{yz} - MYZ;$$

$J_x$ ,  $J_y$  и  $J_z$  — собственные осевые моменты инерции конструкции:

$$J_x = I_y^c + I_z^c;$$

$$J_y = I_x^c + I_z^c;$$

$$J_z = I_y^c + I_x^c;$$

СК — система координат конструкции, в которой задаются характеристики геометрии масс.

Вышестоящие конструкции																
Конструкция	S	M лим	M теор	M черт	M факт	Крепеж	X	Y	Z	LX	LY	LZ	IXY	IYZ	IXZ	Приз
ПУСТОЙ САМОЛЕТ			9588.5490			0.0000										
ПЛАНЕР			4963.9440			0.0000										
ФЮЗЕЛЯЖ			1646.5700			0.0000										

Текущая конструкция													
Конструкция	S	M лим	M теор	M черт	M факт	Крепеж	X	Y	Z	LX	LY	LZ	IX
M1.0200.2.000 ГОЛОВНАЯ ЧАСТЬ ФЮЗ			289.9060			0.0000							

Подчиненные конструкции													
Конструкция	S	M лим	M теор	M черт	M факт	Крепеж	X	Y					
M1.0200.2.020 ПАНЕЛЬ ВЕРХНЯЯ			5.4600			0.0000							
M1.0200.2.030.000 ПАН.Б.ШП1-4			6.0760			0.0000							
M1.0200.2.025 ПАНЕЛЬ ВЕРХН.6-10Ш			21.6600			0.0000							
M1.0200.2.035 ПАНЕЛЬ БОКОВАЯ			15.2760			0.0000							
M1.0230.2.000 НИША ПЕР.Н.ШАССИ			30.2400			0.0000							
M1.0270.2.000 СТВОРКИ ПЕР.Н.Ш.			18.4000			0.0000							
M1.0235.2.000 ПЕРЕГЛЕТ ФОНАРЯ			41.8680			7.2300							
M1.0200.2.050 ПАНЕЛЬ НИЖНЯЯ			35.0530			0.1000							
M1.0201.2.000 ШПАНГОУТ 1			2.5950			0.0000	-0.060000	-0.720000					
M1.0202.2.000 ШПАНГОУТ 2			3.1700			0.0000	0.200000	-0.670000					
M1.0203.2.000 ШПАНГОУТ 3			3.7000			0.0000	0.625000	-0.540000					
M1.0204.2.000 ШПАНГОУТ 4			7.6600			0.0000	0.900000	-0.550000					
M1.0247.2.000 КОНУС			11.6270			0.0000							
M1.0240.2.000 ПОЛ КАБИНЫ			18.9760			0.0000							
M1.0200.2.030.000 ПАН.Б.ШП1-4	S		5.9760			0.0000							

ONLINE/IDLE 127.0.0.1:43110

Рис. 2 Основные параметры конструкций весовой модели самолета

Каждая конструкция привязывается к одной из систем координат, которые описаны в весовой модели. В весовой модели изделия для удобства описания различных агрегатов может быть описано несколько систем координат. Все описанные системы координат организованы в иерархическую структуру. Считается предопределенной глобальная самолетная система координат, в которой могут быть заданы или вычислены координаты всех объектов весовой модели. Каждая система координат в весовой модели задается уникальным именем, положением начала координат относительно вышестоящей системы координат и тремя углами поворота относительно вышестоящей.

Параметр, обозначенный как СК, — это имя одной из систем координат весовой модели. Если СК не задано, то считается, что характеристики геометрии масс заданы в глобальной системе координат. Каждая система координат весовой модели содержит матрицу преобразования координат из самолетной (глобальной) системы координат в данную и обратно. Эта матрица для каждой системы координат есть произведение локальных матриц преобразований в соответствии с положением данной системы в иерархии систем координат. Любое изменение параметров систем координат требует перевычисления матриц преобразования как изме-

ненной системы, так и всех подчиненных. На рис. 2 показана панель параметрического представления конструкций весовой модели.

Так же как и масса, центры тяжести и моменты инерции вычисляются снизу вверх от терминальных конструкций к вышестоящим. При этом осуществляется пересчет характеристик по заданной иерархии систем координат от нижестоящих к вышестоящим и к самолетной системе координат. Расчет МИХ терминальных конструкций осуществляется на основании геометрических моделей. Геометрические модели на этапе рабочего проекта строятся в системах геометрического моделирования. В процессе их построения автоматически вычисляются объемы, массы, положение центра тяжести и другие характеристики геометрии масс. Рассчитанная в системах геометрического моделирования масса с помощью соответствующих интерфейсных средств может быть загружена как  $M_{\text{черт}}$  в весовую модель. (Раньше документация была представлена в виде чертежей на бумажных носителях и  $M_{\text{черт}}$  вручную вычислялась по этим чертежам.) Однако на более ранних этапах проектирования, когда еще не проработана геометрия многих элементов конструкции, весовые расчеты проводятся на основании эскизов и наборов геометрических и конструктивных параметров агре-

готов изделия. Для этого в весовой модели должны быть предусмотрены средства параметрического представления геометрии конструкций. Геометрическое представление конструкций в автоматизированной системе весового проектирования выполняет и немаловажную функцию визуализации конструкций, их компоновки, размещения нагрузки и т.д. В АСВП реализовано несколько форм представления геометрической информации, ориентированных не только на расчет МИХ, но и на визуализацию выполняемых расчетов. Это чертежи геометрических проекций изделия, это табличное задание типовых геометрических конструкций, это каркасное представление трехмерных геометрических моделей и, наконец, задание объемных конструкций триангуляционной (фасеточной) поверхностью. Последний вид представления является наиболее перспективным для точного вычисления МИХ. В АСВП он применяется для расчета тарифных характеристик топливных баков, о чем будет сказано ниже.

**Классификационные признаки конструкции.** В весовой модели кроме числовых параметров определен ряд классификационных признаков конструкций, по которым проводится весовой анализ. Такими маркерами могут быть подразделения, ответственные за разработку конструкции, поставщики или изготовители готовых изделий, принадлежность конструкции к определенным функциональным системам, конструкционные материалы и пр.

**Функциональные подсистемы изделия.** Конструктивное членение самолета может не совпадать с его функциональной структурой. Отдельные элементы функциональных подсистем самолета удобнее описывать в составе конструкции какого-нибудь агрегата планера. Например, некоторая деталь может конструктивно входить в состав крыла, а принадлежать к функциональной подсистеме гидравлики или электрооборудования. Для того чтобы иметь возможность выполнять весовые расчеты, учитывая разные подходы к классификации конструкции самолета, в АСВП отдельно от дерева конструкции ведется реестр подсистем, для которых может быть проведен специальный расчет весовых параметров. В этом реестре ведется полный перечень конструкций весовой модели, входящих в подсистемы реестра, независимо от того, в какой ветви дерева конструкции они находятся. Любая конструкция может быть включена только в одну из подсистем реестра. В зависимости от режима расчетов МИХ конструкций, входящих в подсистему, будут учтены либо в составе вышестоящих агрегатов дерева конструкции, либо отдельно в подсистеме.

**Распределенные характеристики изделия.** Задача вычисления распределенных характеристик изделия является родственной задачей вычисления характеристик геометрии масс. Основное отличие состоит в том, что в данной задаче рассчитываются не интегральные характеристики распределения материала, а сама функция распределения массы по объему конструкции. Такие функции рассчитываются по заданному геометрическому разбиению конструкции на пространственные отсеки. Функции распределения массы по объему конструкции в процессе проектирования используются при построении динамически подобных моделей для проведения некоторых видов испытаний и продувок, а также для выполнения прочностных расчетов.

Каждый отсек разбиения для расчета распределенных характеристик представляет собой выпуклый многогранник, ограниченный конечным набором плоскостей. Задача построения распределенных весовых характеристик состоит в вычислении для каждого отсека массы и положения центра тяжести той части конструкции самолета, которая геометрически расположена внутри этого отсека. Эта задача решается путем нахождения геометрического пересечения каждой терминальной конструкции с каждым отсеком разбиения, и если такое пересечение не пусто, то вычисление массы и центра тяжести той части конструкции, которая попадает в отсек. Некоторые конструкции могут быть объявлены сосредоточенными массами. Использование сосредоточенных масс позволяет исключить все подчиненные конструкции из распределения по отсекам и рассматривать их отдельно для задания сосредоточенных нагрузок. Список сосредоточенных масс с уникальными именами представляет собой отдельную структуру весовой модели. Каждая сосредоточенная масса содержит список ссылок на конструкции весовой модели. Любая конструкция может быть включена не более чем в одну сосредоточенную массу.

**Весовые сводки.** Одной из основных задач АСВП является построение так называемых весовых сводок. Весовые сводки являются документами, сопровождающими построение весовой модели самолета в процессе его создания. В АСВП реализовано несколько форм весовых сводок, которые с разных сторон отражают дерево конструкции самолета или отдельных ветвей этого дерева. Назначение этих сводок и форма представления зависят от решаемых задач. Весовые данные в сводках могут быть представлены либо в табличном виде, либо в виде иерархии конструкций. Могут содержать информацию в детализированном или в укрупненном виде. Отдельные виды весовых сводок представляют распределенные характеристики по отсекам. Весовые



сводки предназначены для решения задач весового контроля и весового анализа.

Весовой контроль при проектировании самолетов представляет собой организационно-техническую систему, нацеленную на создание конструкции минимального веса. Для эффективного весового контроля необходима оперативная информация о текущей массе изделия и любой его части. Весовая информация для весового контроля в АСВП представляется в виде оперативных весовых сводок по отдельным подразделениям предприятия. В таких весовых сводках содержится информация о текущей, теоретической, лимитной, чертежной и фактической массах конструкций, разрабатываемых данным подразделением. Могут также выпускаться оперативные сводки по группе подразделений или по всему проекту. Сводки весового контроля предназначены для использования руководителями проекта.

Весовой анализ также связан с выпуском определенного вида весовых сводок. Для решения задач весового анализа в АСВП осуществляется сортировка и выборки конструкций по определенному классификационному признаку. Например, могут быть рассчитаны массы силового и несилового набора конструкции, массы продольного и поперечного набора, массы конструкций определенного материала, массы готовых изделий или изделий конкретного поставщика и т. д. Весовой контроль и анализ позволяют выявить резервы конструкции, узкие места, тренды в изменении веса конструкции.

## 2.2 Постоянные и переменные структуры весовой модели самолета

Дерево конструкции весовой модели готового изделия не является статической структурой. Конфигурация самолета зависит от конкретных условий его применения. Массово-инерционные характеристики при взлете и посадке отличаются от тех же характеристик в полете, когда убраны стойки шасси. Конфигурация меняется и в полете у самолетов с изменяемым углом стреловидности или с изменяемым вектором тяги. Текущая конфигурация является одним из параметров весовой модели и параметров весовых расчетов. По самому смыслу создания самолета как транспортного средства предполагается, что кроме собственно конструкции, которая обеспечивает выполнение основных задач, на его МИХ существенным образом влияет перевозимая нагрузка. Перевозимая нагрузка есть переменная часть структуры дерева конструкции. Самолетные весовые классификаторы выде-

ляют постоянную часть массы изделия и переменную, состоящую из снаряжения, топлива и целевой нагрузки:

$$M = M_{\text{пустого}} + M_{\text{снаряжения}} + \\ + M_{\text{топлива}} + M_{\text{целевой_нагрузки}} .$$

Все переменные и постоянные компоненты самолета составляют единое целое, и расчет массово-инерционных и центровочных характеристик допустимых конфигураций является одной из главных задач проектирования самолетов любого назначения. Переменные структуры в весовой модели могут задаваться альтернативными конструкциями, когда у некоторой вершины дерева есть несколько вариантов поддеревьев и когда любой из вариантов, но только один из них, может быть активирован в конкретный момент времени. Существует своя специфика задания переменных структур весовой модели для разных содержательных задач.

**Пустой самолет** — это постоянная часть конструкции самолета, которая не меняется в процессе эксплуатации готового изделия. Компонентами пустого самолета являются конструкция планера самолета, силовая установка и ее системы, другие самолетные системы, обеспечивающие управление самолетом, а также специальные системы бортового оборудования, предназначенные для решения целевых задач самолета. В процессе проектирования и при эксплуатации самолетов рассматриваются различные варианты отдельных конструкций планера, а чаще — различные варианты поставляемых готовых изделий. В связи с этим в весовой модели АСВП рассматриваются возможные комбинации вариантов пустого самолета, вариантов снаряжения и полезной нагрузки.

**Снаряжение самолета** устанавливается на борту в процессе предполетной подготовки. Снаряжение самолета принято разделять на основное и дополнительное. Основное снаряжение включает несколько позиций. Это экипаж и системы жизнеобеспечения экипажа, системы жизнеобеспечения пассажиров, заправляемые компоненты и расходные материалы, несгораемый остаток топлива и другие возможные компоненты. Использование различных вариантов экипажа и другого снаряжения самолета связано с различным характером выполняемых задач. Как правило, существует несколько типовых вариантов комплектации экипажа и элементов снаряжения. Весовая модель должна содержать перечень альтернативных вариантов снаряжения и их характеристик. Естественно, что этот перечень может модифицироваться. К дополнительному снаряжению относят временное оборудование и средства, связанные с установкой на



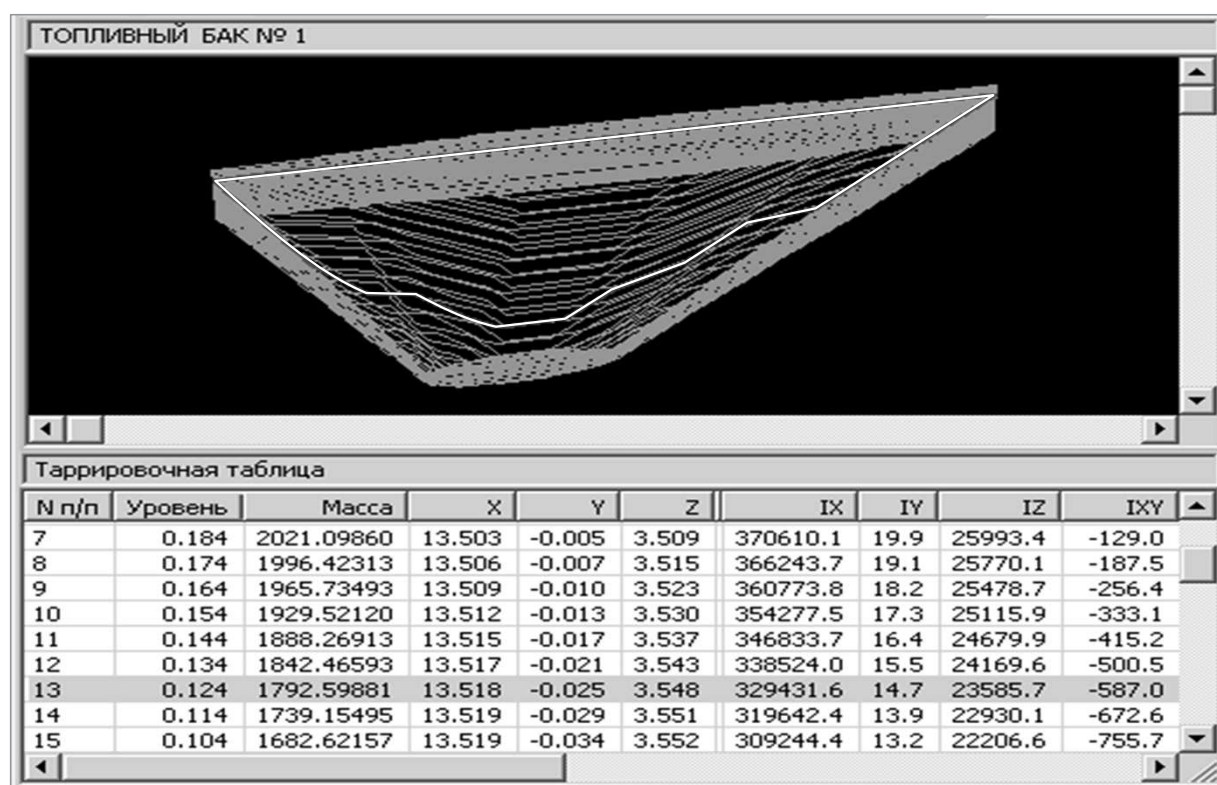


Рис. 3 Тарифовочная таблица топливного бака

борту и закреплением на подвесках перевозимых грузов. Временно устанавливаемое оборудование, как правило, связано со спецификой полетных заданий. Это может быть специальная измерительная аппаратура или оборудование, которое необходимо проверить в условиях реального полета. Перечень такого оборудования и его характеристики в весовой модели должны быть представлены в специальном реестре, или в каталоге. Для установки оборудования, размещения любой коммерческой нагрузки и вооружения в конструкции самолета должны быть предусмотрены специальные места размещения и узлы крепления. Точки размещения оборудования и любых элементов целевой нагрузки задаются своими координатами и установочными углами крепления.

**Топливо** — величина переменная и на земле, при подготовке самолета к вылету, и в воздухе, при выработке топлива, и, если это предусмотрено, при дозаправке в воздухе. Одной из самых сложных и важных задач построения весовой модели является отражение изменяющихся в полете МИХ топлива, находящегося в топливных баках. Топливные баки современных ЛА могут иметь довольно сложные геометрические формы. В процессе выработки топлива все характеристики располагаемого запаса топлива меняются. Необходимо отслеживать эти

изменения в процессе произвольных допустимых эволюций траектории полета. Функции изменения МИХ в зависимости от объема оставшегося топлива задаются тарифовочными характеристиками баков. Для расчета тарифовочных характеристик топливных баков при произвольных углах атаки, углах тангажа и крена в весовой модели наиболее удобно триангуляционное задание баков. В тарифовочной таблице вычисляется масса оставшегося топлива в зависимости от уровня поверхности жидкости в топливном баке. На рис. 3 приведен пример расчета тарифовочной таблицы крыльевого топливного бака.

Если МИХ топлива в конкретном баке по мере его выработки определяются тарифовочной характеристикой, то зависимость МИХ оставшегося топлива определяется последовательностью, в которой осуществляется выработка из разных баков. Топливная система самолета состоит из нескольких баков — как внутренних, так и размещенных на подвесках, а также из системы трубопроводов, перекачивающих насосов и управляющей автоматики. Основой управления расходом топлива является программа, определяющая порядок расходования топлива из разных баков. Переключение перекачки топлива между разными баками осуществляется для обеспечения центровки самолета в заданных

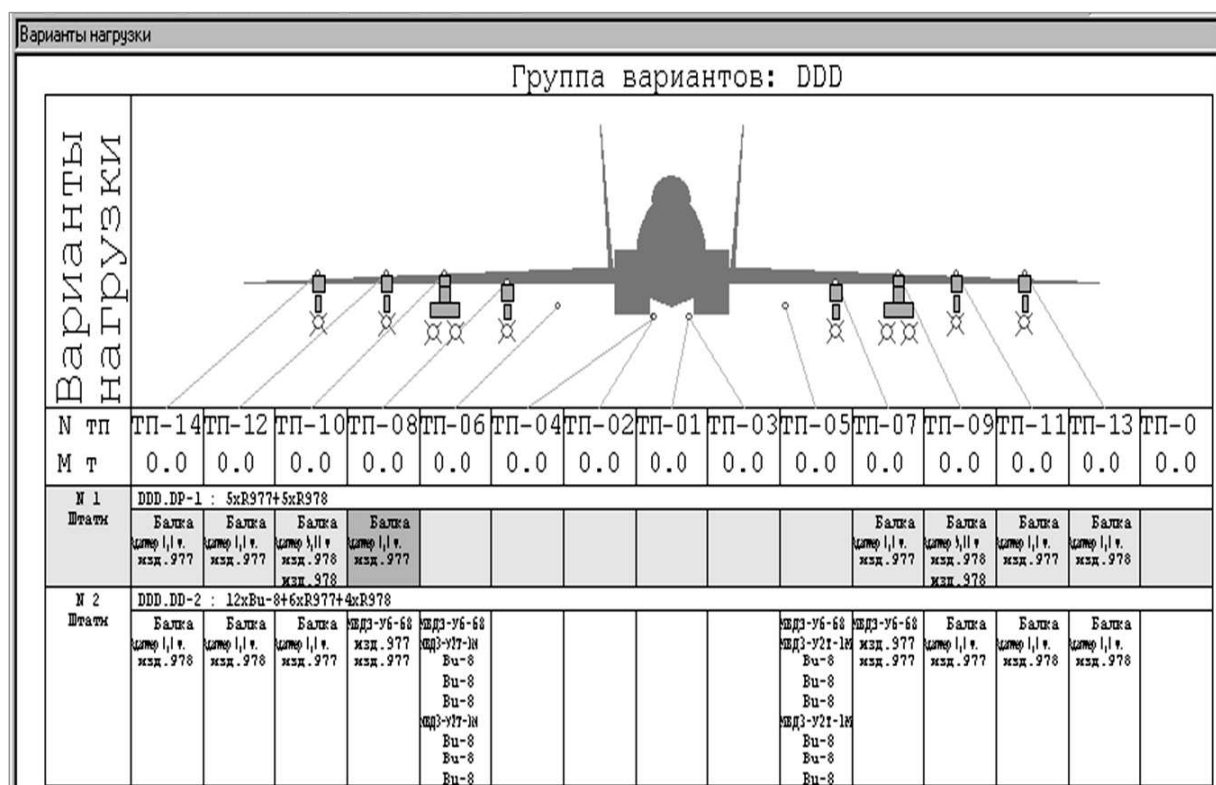


Рис. 4 Варианты размещения целевой нагрузки самолета на подвесках

границах. Одним из критериев при разработке алгоритмов перекачки является число переключений и обеспечение бесперебойной подачи топлива при любых допустимых параметрах траектории полета. Массово-инерционные характеристики топлива в процессе тарировки баков задаются их разбиением плоскопараллельными сечениями на тонкие слои. Для каждого слоя указывается масса, координаты центра тяжести и плоскостные моменты инерции. Программа выработки топлива представляет собой последовательность выработки слоев из разных баков в соответствии с диаграммой переключений. В весовой модели может быть задано несколько вариантов программ расходования топлива. Разумеется, в процессе выполнения полетного задания программа расходования топлива фиксирована. Предварительный расчет характеристик для разных вариантов порядка выработки топлива необходим для выбора наилучшего, удовлетворяющего всем ограничениям.

**Целевая нагрузка** зависит от назначения самолета и от конкретного полетного задания. Для пассажирских самолетов целевая нагрузка — это пассажиры с багажом, для транспортных самолетов — это коммерческие грузы, для военных — подвесное или размещаемое в специальных отсеках вооружение. В полете возможен сброс и десантиро-

вание целевой нагрузки. Комплектация и установка целевой нагрузки представляет собой довольно сложный процесс. Выбор состава грузов и их размещение могут проходить в несколько этапов. Сложность выбора обусловлена большим количеством типов перевозимой нагрузки, наличием большого числа вспомогательных специальных устройств закрепления грузов как во внутренних отсеках самолета, так и на внешних подвесках. На рис. 4 приведена панель формирования различных расчетных вариантов целевой нагрузки самолета. Визуализация этого процесса существенно облегчает решение различных задач анализа допустимой нагрузки как на этапе проектирования самолета, так и при эксплуатации во время подготовки полетных заданий.

Для удобства выбора и проведения расчетов множества вариантов загрузки самолета в рамках весовой модели реализованы каталоги нагрузки — специального оборудования, коммерческой нагрузки, вооружения. В этих каталогах ведутся классификаторы, позволяющие в громадных перечнях находить нужные позиции и их характеристики. Кроме МИХ размещаемой нагрузки в каталогах даются ссылки на их геометрические модели, задаются габариты, другие геометрические параметры. Эти данные нужны для визуализации

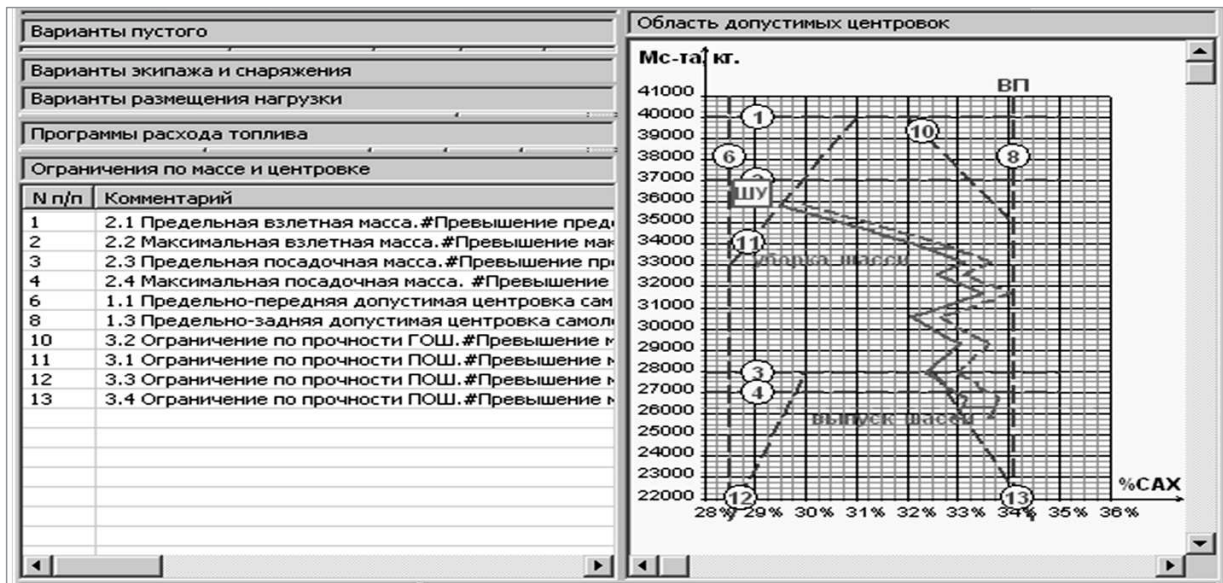


Рис. 5 Область допустимых центровок

ции размещения и компоновки, для вычисления их МИХ. Как правило, существуют довольно жесткие ограничения на размещение нагрузки на борту, а также на внешних узлах крепления. Эти ограничения должны указываться в каталоге и учитываться в процессе формирования вариантов загрузки самолета.

Ограничения, которые проверяются при анализе различных вариантов снаряжения самолета, программы выработки топлива и допустимых вариантов целевой нагрузки, задают область допустимых центровок самолета.

**Область допустимых центровок** является неотъемлемой частью весовой модели и служит одной из основных весовых характеристик самолета, особенно важной и контролируемой в процессе его эксплуатации. На рис. 5 проиллюстрированы ограничения, образующие область допустимых центровок, и приведен график изменения центровки самолета при выработке топлива.

По оси абсцисс на этом графике откладывается центровка самолета, которая определяется как положение центра тяжести самолета на средней аэродинамической хорде в процентах от ее длины. По оси ординат откладывается текущая масса самолета с учетом массы снаряжения, массы целевой нагрузки и текущего запаса топлива. Точки излома на графиках центровки соответствуют моментам переключения подачи топлива с одного бака на другой, которые определяются программой выработки топлива или моментами сброса целевой нагрузки. Двойной график изменения центровки соответствует двум полетным конфигурациям —

с убранными и выпущенными стойками шасси. Ограничения, которые обеспечивают устойчивость и управляемость полета, задаются предельными значениями центровки. Предельно передняя и предельно задняя центровки на графике показаны вертикальными штриховыми линиями. Горизонтальные линии задают ограничения на взлетную и посадочную массы. Ограничения максимальной взлетной и посадочной массы при определенных условиях могут нарушаться, но эти нарушения допускаются в исключительных условиях и сказываются на ресурсных характеристиках самолета. Превышение **предельных** значений взлетной и посадочной массы не допускается. Наклонные штриховые линии на графике задают ограничения, связанные с максимально допустимыми нагрузками на переднюю и главную опоры шасси.

### 3 Программная реализация автоматизированной системы весового проектирования

Представленная здесь структурно-параметрическая весовая модель самолета позволяет решать широкий круг задач весового проектирования. Весовая модель составляет информационную основу, на базе которой могут быть построены различные вычислительные программы и пользовательские модули. Рассматриваемая в данной работе АСВП построена по клиент-серверной архитектуре, где весовая модель служит единым хранилищем

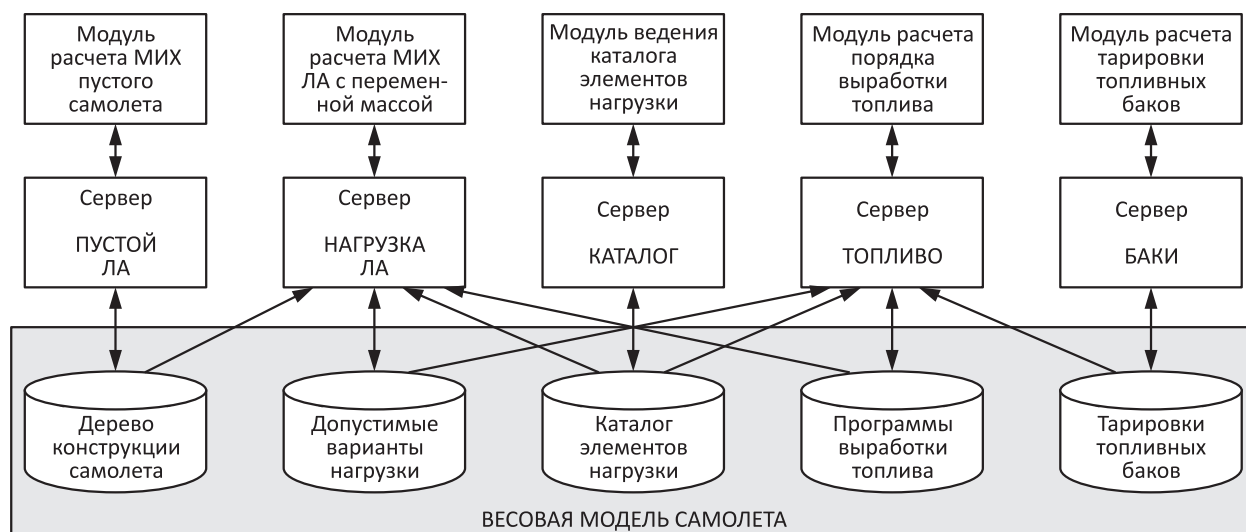


Рис. 6 Архитектура программной реализации исполнительных модулей АСВП

информации, базой данных системы. Непосредственно с информацией, хранящейся в этой базе данных, взаимодействуют различные вычислительные, расчетные программы — серверы, которые кроме расчетных функций обеспечивают информационную связь клиентских модулей с весовой моделью самолета. Непосредственными пользователями клиентских модулей являются конструкторы и проектировщики, решающие различные задачи весового проектирования. Построена АСВП как многопользовательская интерактивная система. На рис. 6 представлена архитектура АСВП, ее основные программные и информационные компоненты.

Ниже перечислены основные функции программных модулей АСВП:

Сервер ПУСТОЙ ЛА + Модуль расчета МИХ пустого самолета:

- создание и модификация дерева конструкции пустого самолета;
- расчет МИХ пустого изделия, всех его систем, узлов, агрегатов и деталей на любых уровнях дерева конструкции;
- весовой анализ и контроль текущего состояния проекта, выполнения лимитных ограничений по весу, осуществление выборок весовой информации по различным признакам — системам, агрегатам, типу конструкции (силовая/несиловая), материалу конструкции, подразделениям и т. д.;
- расчет распределения массы самолета по различным разбиениям на отсеки; эта информация используется для построения ди-

намически подобных моделей и при прочностных расчетах;

- расчет МИХ при различных вариантах полетной конфигурации при убранных и выпущенных стойках шасси, при отклонениях консолей крыла для самолетов с изменяемой геометрией, при отклонении органов управления.

Сервер НАГРУЗКА ЛА + Модуль расчета МИХ самолета с переменной массой:

- создание и модификация реестра допустимых вариантов нагрузки самолета;
- расчеты МИХ снаряженного и загруженного самолета для разных вариантов компоновки и размещения на борту полезной нагрузки;
- расчет изменения МИХ самолета в полете при выработке топлива, дозаправке в воздухе, сбросе нагрузки;
- расчет МИХ самолета в виде табличных зависимостей для различных вариантов снаряжения и размещения нагрузки;
- расчет МИХ самолета в виде графических зависимостей от массы самолета и/или от массы топлива;
- проверка выполнения установленных эксплуатационных ограничений по центровке, взлетной и посадочной массе, нагрузке на опоры шасси для различных вариантов снаряжения и размещения нагрузки; сигнализация в случае нарушения ограничений, а также для различных вариантов программ выработки топлива.



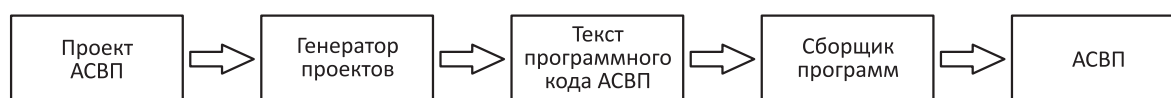


Рис. 7 Проектный подход — технология ГП

Сервер КАТАЛОГ + Модуль ведения каталога элементов нагрузки:

- создание и модификация каталога элементов целевой нагрузки самолета;
- создание и модификация базы данных вариантов размещения и закрепления элементов нагрузки каталога на борту самолета или на подвесках;
- создание и модификация базы данных вспомогательных элементов конструкции установки элементов нагрузки.

Сервер ТОПЛИВО + Модуль расчета порядка выработки топлива:

- создание и модификация базы данных различных вариантов программы выработки топлива;
- расчет МИХ и МЦД для различных вариантов переключения выработки топлива из внутренних, закладных и подвесных баков;
- расчет МИХ и МЦД при различных программах заливки и дозаправки топлива во внутренние, закладные и подвесные баки.

Сервер БАКИ + Модуль расчета тарировки топливных баков:

- создание и модификация базы данных геометрии топливных баков;
- расчет тарировочных характеристик топливных баков при различных углах тангажа и крена.

Программная реализация АСВП велась с использованием инструментального комплекса «Генератор проектов» (технология ГП) [5]. Технология ГП обеспечивает возможность разработки прикладных систем многоуровневой клиент-серверной архитектуры с использованием реляционных и сетевых баз данных со сложным пользовательским и межпрограммным интерфейсом. Создание информационно-вычислительных систем в рамках технологии ГП базируется на так называемом «проектном подходе». Под проектом здесь понимается

пакет документов (файлов), содержащий описание структуры проекта, описание логической структуры баз данных, спецификации пользовательского интерфейса, перечень команд и сценарии работы пользователей, описание функций и процедур обработки пользовательских запросов. Исходное описание проекта подается на вход «Генератора проекта», который строит в памяти модель проекта, осуществляет ее анализ на предмет корректности и целостности, а затем на основании этой модели генерирует тексты программного кода для клиентских и серверных компонент системы, а также генерирует утилиты, необходимые для сборки, инсталляции и сопровождения системы.

На рис. 7 показана общая архитектура программной конструкции, связанной с применением технологии ГП.

В приведенной цепочке разработчик прикладной информационной системы имеет дело только с первым ее звеном — проектом системы. При этом он избавлен от необходимости иметь дело с системным программным окружением вычислительной среды, в которой должна функционировать разрабатываемая прикладная система. Все связи прикладных информационных процессов с конкретной системной вычислительной средой привносит в результирующую рабочую программу «Генератор проектов» на стадии анализа и генерации итогового программного кода. Естественно, что при этом объем описания проекта оказывается существенно короче программного кода, который создается автоматически. Экономия трудозатрат разработчика оказывается существенной. В частности, объем описания проекта АСВП на порядок меньше, чем объем сгенерированного программного кода. Даже если предположить, что написанный вручную программный код благодаря искусству программистов будет весьма экономным, то все равно трудоемкость разработки прикладных систем будет в разы меньше.

Но главное даже не в числе строк программ, а прежде всего в экономии интеллектуальных затрат разработчиков прикладных систем и, в итоге, автоматически созданные программы более надежны и свободны от нечаянных ошибок и опечаток. И кроме того, разрабатываемые в рамках технологии ГП прикладные системы обеспечивают



ся эффективными средствами сопровождения, т. е. достаточно простой процедурой внесения исправлений и развития программ в процессе их эксплуатации.

Прикладные программные комплексы в рамках технологии ГП разрабатываются как автономные системы и не требуют для своей работы специальной среды и дорогостоящих программных продуктов (кроме использующихся систем управления базами данных (СУБД) и общесистемного обеспечения). Разрабатываемые в рамках технологии ГП прикладные системы допускают масштабирование и портирование на различные вычислительные платформы и СУБД.

Как уже говорилось, система АСВП разрабатывалась в течение ряда лет, многие ее компоненты и версии были апробированы и использовались в реальном проектировании. Авторы выражают благодарность С. И. Скобелеву, М. К. Курьянскому, Д. Ю. Стрельцу, П. В. Плуноскому и К. Н. Ерасову за плодотворные обсуждения проблем весового проектирования самолетов, за постановку многих задач и за апробацию разработанных программ.

## Литература

1. *Шейнин В. М., Козловский В. И.* Весовое проектирование и эффективность пассажирских самолетов. — М.: Машиностроение, 1977. Т. 1. 343 с.
2. *Скобелев С. И., Широков Н. И.* Весовой анализ и контроль в САПР ЛА // Задачи и методы автоматизированного проектирования. — М.: ВЦ РАН, 1991. С. 92–100.
3. *Широков Н. И.* Автоматизированная система весовых расчетов в САПР ЛА // Автоматизация проектирования инженерных и финансовых информационных систем средствами Генератора проектов / Отв. ред. Ю. А. Флеров. — М.: ВЦ РАН, 2010. С. 55–66.
4. *Вышинский Л. Л., Широков Н. И.* Система автоматизации расчетов массово-инерционных характеристик ЛА с переменной массой // Развитие и применение инструментального комплекса Генератор проектов / Отв. ред. Ю. А. Флеров. — М.: ВЦ РАН, 2014. С. 20–31.
5. *Вышинский Л. Л., Гринев И. Л., Флеров Ю. А., Широков А. Н., Широков Н. И.* Генератор проектов — инструментальный комплекс для разработки «клиент-серверных» систем // Информационные технологии и вычислительные системы, 2003. № 1-2. С. 6–25.

Поступила в редакцию 24.05.17

## COMPUTER-AIDED SYSTEM OF AIRCRAFT WEIGHT DESIGN

L. L. Vyshinsky, Yu. A. Flerov, and N. I. Shirokov

A. A. Dorodnicyn Computing Centre, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The article is devoted to the problems of computer-aided weight design of aircraft. Weight and mass-inertial parameters are one of the basic values that affect the performance characteristics of aircraft. The informational basis of the system is the weight model of the aircraft. The paper describes the structure of the weight model and its individual components. The program implementation of the system, which is executed within the framework of the client-server architecture, is shown. The automated system of weight design is implemented using the software tool complex “Project Generator” (GP technology), which was developed at the Computing Centre of the Russian Academy of Sciences. The creation of information and computing systems within the framework of the GP technology is based on the so-called “project approach,” when the formal description of the system automatically generates code for the client and server components of the system.

**Keywords:** math modeling; design automation; aircraft; weight design; weighting model; design tree; project generator; code generation; client-server architecture

**DOI:** 10.14357/19922264180103

## References

1. *Sheynin, V. M., and V. I. Kozlovskiy.* 1977. *Vesovoe proektirovanie i effektivnost' passazhirskikh samoletov* [Weight design and efficiency of passenger aircraft]. Moscow: Mechanical Engineering. Vol. 1. 343 p.
2. *Skobelev, S. I., and N. I. Shirokov.* 1991. *Vesovoy analiz i kontrol' v SAPR LA* [Weight analysis and control in CAD of aircraft]. *Zadachi i metody avtomatizirovannogo proektirovaniya* [Tasks and methods of computer-aided design]. Moscow: Computing Centre of the USSR Academy of Sciences. 92–100.
3. *Shirokov, N. I.* 2010. *Avtomatizirovannaya sistema vesovyykh raschetov v SAPR LA* [Automated system weight calculations in CAD]. *Avtomatizatsiya proektirovaniya in-*

- zhenernykh i finansovykh informatsionnykh sistem sredstvami Generators proektov* [Computer aided design of engineering and financial information systems by the means of the Project Generator]. Moscow: Computing Centre of RAS. 55–66.
4. Vyshinskiy, L. L., and N. I. Shirokov. 2014. Sistema avtomatizatsii raschetov massovo-inertsionnykh kharakteristik LA s peremennoy massoy [CAD system of calculation aircraft mass-inertial characteristics with variable mass]. *Razvitiye i primeneniye instrumental'nogo kompleksa Generator proektov* [The development and application of a tool set Project Generator]. Moscow: Computing Centre of RAS. 20–31.
5. Vyshinskiy, L. L., I. L. Grinev, Yu. A. Flerov, A. N. Shirokov, and N. I. Shirokov. 2003. Generator proektov — instrumental'nyy kompleks dlya razrabotki “klient–servernykh” sistem [The project generator — tool complex for development of “client–server” systems]. *Informatsionnye tekhnologii i vychislitel'nye sistemy* [Information Technologies and Computer Systems] 1-2:6–25.

Received May 24, 2017

## Contributors

**Vyshinsky Leonid L.** (b. 1941) — Candidate of Sciences (PhD) in physics and mathematics, Head of Laboratory, A. A. Dorodnicyn Computing Centre, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; Wysh@ccas.ru

**Flerov Yuri A.** (b. 1942) — Corresponding Member of the Russian Academy of Science, Doctor of Science in physics and mathematics, professor, Deputy Director, A. A. Dorodnicyn Computing Centre, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; fler@ccas.ru

**Shirokov Nikolai I.** (b. 1963) — Candidate of Sciences (PhD) in physics and mathematics, senior scientist, A. A. Dorodnicyn Computing Centre, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; Wysh@ccas.ru

# КОНТИНУАЛЬНЫЙ КРИТЕРИЙ VaR НА СЦЕНАРНЫХ РЫНКАХ\*

Г. А. Агасандян<sup>1</sup>

**Аннотация:** Работа посвящена изучению проблем использования континуального критерия VaR (CC-VaR) на сценарных рынках — дискретном аналоге идеального теоретического однопериодного рынка опционов. Участие инвестора на рынке состоит в задании им прогноза распределения цены базового актива и формировании функции рискованных предпочтений (ф.р.п.). Предлагается дискретный алгоритм оптимизации как результат проецирования теоретического алгоритма, основанного на применении процедуры Неймана–Пирсона, на сценарный рынок. Приводится пример нарушения оптимальности для небольшого числа сценариев, однако такие нарушения происходят редко и они весьма незначительны. Предлагается использовать рандомизацию весов найденного портфеля как средство сглаживания функций распределения и повышения качества получаемых решений. Специальные алгоритмы предлагаются для расчетов, связанных с доходностью портфелей с рандомизацией. Изложение иллюстрируется графиками.

**Ключевые слова:** континуальный критерий VaR (CC-VaR); сценарии; прогнозная плотность; стоимостная плотность; функция рискованных предпочтений (ф.р.п.); оптимальный портфель; инвестиционная сумма; доход; доходность; рандомизация

**DOI:** 10.14357/19922264180104

## 1 Введение

В работе излагается методология применения инвестором в своих решениях на финансовых рынках так называемого *континуального критерия VaR* (CC-VaR). Сферой его приложения служат высоко развитые по многообразию торгуемого инструментария рынки, такие как рынки опционов со страйками, плотно заполняющими диапазон возможных будущих цен базового актива.

Критерий CC-VaR — естественное континуальное обобщение привычного *одноступенчатого* критерия VaR, он должен противостоять как его недостаткам, так и дисперсии в отношении проблем риска в задачах условной максимизации среднего дохода (см., например, [1–3]). Суть в том, что применение обычного критерия VaR на рынках опционов порождает доходы, обычно меньшие размера инвестиции, с вероятностью, близкой к единице (при частой решетке страйков), что едва ли может устроить инвестора.

Использование дисперсии, как в классическом подходе Марковица [1], основанной на свойствах случайности не выше второго порядка, не позволяет учитывать такие важные нюансы распределений, как, например, часто наблюдаемые на рынках тяжелые хвосты распределений, обусловленные свойствами случайности не ниже четвертого порядка (куртозис).

Предлагаемый *континуальный критерий VaR* должен обеспечить инвестору (во всяком случае, на теоретическом рынке) получение функции распределения доходов с заранее заданными им же свойствами.

В наиболее простой теоретической модели рынков предполагается однопериодным с одним базовым активом, континуальным по мощности инструментария и идеальным в том смысле, что цены продавца и покупателя считаются равными, а коммиссионные равны нулю [4, 5].

Вкратце приводятся основные результаты и формулы для такого рынка, но основное внимание уделяется проблемам их приложения к дискретным по инструментам рынкам, рассматриваемым как приближение к континуальным.

## 2 Континуальный $\delta$ -рынок и его инструменты

В начале периода цена базового актива  $X$  известна, а в конце периода она образует случайную величину  $X$ , принимающую значения  $x$  из континуального множества  $X \subset \mathbb{R}_+$  (или даже  $\mathbb{R}$ ). На рынке, называемом  *$\delta$ -рынком*, можно торговать любым инструментом  $G$  с доходом, представимым в виде произвольной неотрицательной измеримой

\*Работа выполнена при финансовой поддержке РФФИ (проект 17-01-00816).

<sup>1</sup>Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, agasand17@yandex.ru

функции  $g(x)$ ,  $x \in X$ . Ее называем *платежной функцией* инструмента и обозначаем  $\pi(x; \mathbf{G})$ ,  $x \in X$ , т. е.  $g(x) = \pi(x; \mathbf{G})$ ; в частности,  $\pi(x; \mathbf{X}) = x$ . Цена инструмента  $\mathbf{G}$  обозначается  $|\mathbf{G}|$ , а средний доход от него —  $\|\mathbf{G}\|$ .

На рынке обращаются так называемые  $\delta$ -инструменты  $\mathbf{D}(s)$ ,  $s \in X$ . Платежная функция  $\delta$ -инструмента равна  $\delta$ -функции относительно  $s$ :  $\pi(x; \mathbf{D}(s)) \equiv \delta(x - s)$ ; такой инструмент порождает нулевой доход, если  $x \neq s$ , и бесконечный, если  $x = s$ ,  $s \in X$ , притом интеграл от платежной функции по  $x \in X$  равен единице.

Нетрудно усмотреть в  $\delta$ -инструментах предельный аналог специально нормированных баттерфляев обычного рынка опционов при стремлении расстояния между соседними страйками к нулю.

Заданы *стоимостная*  $c(x)$  и *прогнозная*  $p(x)$  плотности,  $x \in X$ . Первая формируется рынком к началу периода, вторая задается инвестором на конец периода. Стоимость (цена)  $\delta$ -инструментов и их средний доход соответственно равны:

$$|\mathbf{D}(s)| = c(s); \quad \|\mathbf{D}(s)\| = p(s), \quad s \in X.$$

Портфель  $\mathbf{G}$  с  $g(x) = \pi(x; \mathbf{G})$ , его цена и средний доход имеют представления:

$$\begin{aligned} \mathbf{G} &= \int_X g(s) \mathbf{D}(s) ds; \\ |\mathbf{G}| &= \int_X g(s) |\mathbf{D}(s)| ds = \int_X g(s) c(s) ds; \\ \|\mathbf{G}\| &= \int_X g(s) \|\mathbf{D}(s)\| ds = \int_X g(s) p(s) ds. \end{aligned}$$

Имея в виду последующий переход к сценарным рынкам и сравнение с ними, введенный континуальный рынок будем называть *эталонным*, а его характеристики при сравнении с дискретными рынками помечать *звездочкой* (\*).

Особый интерес представляют такие континуальные комбинации  $\delta$ -инструментов, как *индикаторы*  $\mathbf{H}\{M\}$ ,  $M \subset X$ . Платежной функцией  $\mathbf{H}\{M\}$  служит характеристическая функция множества  $M$ , равная единице, если  $x \in M$ , и нулю, если  $x \notin M$ . В частности, безрисковый инструмент единичного объема  $\mathbf{U}$  является индикатором  $\mathbf{H}\{X\}$ . Его платежная функция тождественно равна единице. Имеют место соотношения:

$$\mathbf{H}\{M\} = \int_M \mathbf{D}(s) ds; \quad \mathbf{U} = \mathbf{H}\{X\} = \int_X \mathbf{D}(s) ds.$$

Очевидно, из специально нормированных индикаторов также можно строить сходящиеся к  $\delta$ -инструментам последовательности.

Критерий *СС-VaR* требует, чтобы порождаемый портфелем инвестора доход  $q$  удовлетворял неравенствам:

$$P\{q \geq \phi(\varepsilon)\} \geq 1 - \varepsilon \quad (1)$$

для всех  $\varepsilon \in [0, 1]$ . Здесь  $P\{M\}$  — вероятностная мера множества  $M$  в соответствии с прогнозом инвестора, а  $\phi(\varepsilon)$  — неотрицательная монотонно возрастающая и непрерывная *функция рискованных предпочтений* инвестора.

Алгоритм оптимизации приводится для задачи *СВ*, основной для прочих возможных постановок, в которой требуется минимизировать инвестиционную сумму лишь при выдерживании всех ограничений *СС-VaR* [5].

Применением известной из математической статистики процедуры Неймана–Пирсона [6] для всех  $\varepsilon \in [0, 1]$  находятся оптимальные множества  $X_\varepsilon \subset X$  прогнозной вероятности  $\varepsilon$ , а также прогнозная и стоимостная функции ( $x \in X$ ):

$$\begin{aligned} X_\varepsilon &= \{\rho(x) \leq \tau\}, \quad \tau \in [\tau', \tau''], \\ \tau' &= \min_{x \in X} \rho(x), \quad \tau'' = \max_{x \in X} \rho(x); \end{aligned}$$

$$f_P(\tau) = P\{\rho(x) \leq \tau\} = F_{P; \rho(X)}(\tau);$$

$$f_C(\tau) = C\{\rho(x) \leq \tau\} = F_{C; \rho(X)}(\tau)$$

(здесь  $F_{M; \zeta}(\cdot)$  — функция распределения случайной величины  $\zeta$  по мере  $M\{\cdot\}$ ).

Далее определяются диссонанта, ее производная, функция упорядочения, оптимальная весовая функция портфеля и сам портфель соответственно:

$$\gamma(\varepsilon) = f_C(f_P^{-1}(\varepsilon)), \quad \gamma'(\varepsilon) = \frac{1}{f_P^{-1}(\varepsilon)};$$

$$w(x) = f_P(\rho(x));$$

$$g(x) = \phi(w(x));$$

$$\mathbf{G} = \int_X g(x) \mathbf{D}(x) dx.$$

Наконец, выписываются представления для дохода, среднего дохода, инвестиционной суммы, дисперсии доходности и функции распределения дохода (здесь  $E_M$  и  $D_M$  — символы соответственно математического ожидания и дисперсии по вероятностной мере  $M\{\cdot\}$ ,  $\sigma$  — стандартное отклонение доходности):

$$\left. \begin{aligned} q &= g(X) = \phi(w(X)); \\ R &= \int_0^1 \phi(\varepsilon) d\varepsilon = \int_{\tau'}^{\tau''} \phi(f_P(\tau)) f_P'(\tau) d\tau (= E_P q); \end{aligned} \right\} (2)$$



$$\left. \begin{aligned}
 A &= \int_0^1 \phi(\varepsilon) d\gamma(\varepsilon) = \\
 &= \int_{\tau'}^{\tau''} \phi(f_P(\tau)) f'_C(\tau) d\tau (= E_C q) ; \\
 y &= \frac{R}{A} - 1 ;
 \end{aligned} \right\} (3)$$

$$\sigma^2 = \frac{D_P q}{A^2} = \int_X \frac{(g(x) - R)^2 p(x) dx}{A^2} = \frac{\mu_2 - R^2}{A^2},$$

$$\mu_2 = \int_0^1 \phi^2(\varepsilon) d\varepsilon ; \quad (4)$$

$$\left. \begin{aligned}
 F_{P;q}(z) &= \phi^{\leftarrow}(z), \quad z \in [\phi(0), \phi(1)] ; \\
 P\{q \geq \phi(\varepsilon)\} &\equiv 1 - \varepsilon \text{ для всех } \varepsilon \in [0, 1].
 \end{aligned} \right\} (5)$$

Конструкция теоретического  $\delta$ -рынка и алгоритм оптимизации портфеля используются теперь для изучения дискретного сценарного рынка.

### 3 Сценарный рынок и его агрегаты

Сценарная дискретизация эталонного рынка вводится разбиением множества  $X = [x_0, x_n]$  на  $n$  сценариев  $S_i = [x_{i-1}, x_i] \subset X$ ,  $x_{i-1} < x_i$ ,  $i \in I = \{1, \dots, n\}$ . Равномерное разбиение выделяется правилом  $x_i = x_0 + ih$ ,  $h = (x_n - x_0)/n$ ,  $i \in I$ .

На сценарном рынке базисными инструментами служат дискретные аналоги инструментов  $D(s)$  теоретического рынка — индикаторы сценариев  $D_i = \mathbf{H}\{S_i\}$ ,  $i \in I$ ; при этом  $\sum_{i \in I} D_i = \mathbf{U}$ . Их платежные функции  $\pi(x; D_i) = \chi_i(x)$ , где  $\chi_i(x)$  — характеристическая функция множества  $S_i$ , а стоимости — рыночные цены индикаторов — получаются из стоимостной плотности и образуют неотрицательный вектор  $\mathbf{c} \equiv \{c_i, i \in I\}$ , где

$$c_i = |D_i| = \int_{x_{i-1}}^{x_i} c(x) dx = \int_X \chi_i(x) c(x) dx, \quad i \in I.$$

Строго говоря, только сам этот вектор, а не плотность  $c(x)$  определяет рыночную картину текущих цен, а плотность в дискретном случае задается для удобства моделирования и сравнительного анализа. Этому вектору противостоит прогноз инвестора в форме получаемого из плотности  $p(x)$  вектора прогнозных вероятностей сценариев  $\mathbf{p} \equiv \{p_i\}$ , где

$$p_i = \|D_i\| = \int_{x_{i-1}}^{x_i} p(x) dx = \int_X \chi_i(x) p(x) dx, \quad i \in I,$$

т. е. вероятность сценария  $S_i$  совпадает со средним доходом от инструмента  $D_i$  (ее можно считать *справедливой* ценой индикатора  $D_i$ ).

Очевидно,  $c_i = E_C \chi_i(X)$ ,  $p_i = E_P \chi_i(X)$ ,  $i \in I$ .

Портфель инвестора  $\mathbf{G} = \sum_{i \in I} g_i D_i$  с  $g_i \geq 0$ ,  $i \in I$ , порождает вектор доходов (весов)  $\mathbf{g} = \{g_1, \dots, g_n\}$ . (Неотрицательность весов обусловлена выставлением инвестору маржевых требований во избежание необеспеченных потерь.) Рыночная цена портфеля  $\|\mathbf{G}\| = \sum_{i \in I} g_i c_i$ , а средний доход  $\|\mathbf{G}\| = \sum_{i \in I} g_i p_i$ .

В соответствии с (1) рисковые интересы инвестора на рынке задаются его ф.р.п.  $\phi(\varepsilon)$ ,  $\varepsilon \in [0, 1]$ , совместно с континуальным критерием VaR.

В связи с наличием некоторого разнобоя в формировании прогноза и ценовой картины рынка и в отсутствие иной информации вводится естественное

*Предположение С.* Стоимостная мера  $C\{\cdot\}$  (независимо от ее происхождения) трансформируется в меру с иной плотностью  $c(x)$ ,  $x \in X$ , так, чтобы выполнялось тождество  $\rho(x) \equiv p(x)/c(x) \equiv p_i/c_i$  по  $x \in S_i$ ,  $i \in I$ .  $\square$

Действие этого предположения меняет исходный *эталонный* теоретический рынок, основанный на изначальной паре  $\{p(x), c(x)\}$ , привязывая его к данному сценарному рынку, но и оставляя возможность считать его теоретическим. Будем называть новую схему *вариантом #0*. Отметим, что предлагаемая *предположением С* трансформация, пожалуй, самая простая из всех допускающих естественную упорядоченность элементов  $X$  по величине  $\rho(x)$  внутри сценариев.

### 4 Дискретный алгоритм оптимизации портфелей

Предлагается адаптировать континуальный алгоритм построения оптимального портфеля, рассмотренного в [5], к сценарному рынку. При этом адаптация возможна в нескольких вариантах назначения портфельных весов.

Дискретный алгоритм строится как результат проецирования континуального алгоритма на сценарный рынок и состоит из общей и специфической частей. При этом целью алгоритма, как и для эталонного рынка, служит решение *задачи СВ*, и также применяется процедура Неймана—Пирсона.

Итак, заданы векторы  $\mathbf{c}$  и  $\mathbf{p}$  размерности  $n$  для сценариев  $S_i$ ,  $i \in I$ . Дискретный алгоритм описывается следующей последовательностью обозначений и операций (при этом  $i, j \in I$ , все векторы имеют длину  $n$ , а матрицы — размер  $n \times n$ ):

$\rho = \mathbf{p}/\mathbf{c}$  — вектор относительных доходов  $\rho_i = p_i/c_i$ , дискретный аналог функции относительных доходов  $\rho(x)$ ;

$\xi = \mathbf{O}(\rho)$  — вектор, задающий на множестве сценариев позиции компонент вектора  $\rho$  в порядке возрастания: первый элемент указывает позицию (номер) наименьшего относительного дохода,  $n$ -й — наибольшего,  $\mathbf{O}$  — соответствующее преобразование;

$\eta = \mathbf{O}(\xi)$  — вектор, обратный к  $\xi$ : если  $\xi_i = j$ , то  $\eta_j = i$ ; компонента  $\eta_j$  означает номер дохода  $\rho_j$  в вариационном ряду для вектора  $\rho$ ; при этом также  $\xi = \mathbf{O}(\eta)$ ;

$\tau = \rho(\xi)$  — вектор относительных доходов с упорядоченными по возрастанию компонентами;

$\Xi = \{y_{ij}\}$ , где

$$y_{ij} = \begin{cases} 1, & \text{если } \xi_i = j; \\ 0, & \text{если } \xi_i \neq j; \end{cases}$$

эта матрица реализует преобразование  $\mathbf{O}$ , и имеют место равенства  $\eta = \Xi\xi$ ,  $\xi = \Xi^{-1}\eta$ ;

$\mathbf{T} = \{t_{ij}\}$ , где

$$t_{ij} = \begin{cases} 1, & \text{если } i \leq j; \\ 0, & \text{если } i > j. \end{cases}$$

Это треугольная матрица для последовательного суммирования компонент векторов, начиная с первой;

$\mathbf{d} = \Xi\mathbf{p} = \mathbf{p}(\xi)$  — подстановка вектора  $\mathbf{p}$ , компоненты которой упорядочены по возрастанию компонент вектора  $\rho$ ; обратно, верны равенства  $\mathbf{p} = \Xi^{-1}\mathbf{d} = \mathbf{d}(\eta)$ ;

$\varepsilon = \mathbf{T}\mathbf{d}$  — вектор кумулятивных вероятностей для вектора  $\mathbf{d}$ ;

$\mathbf{f} = \Xi\mathbf{c} = \mathbf{c}(\xi)$  — подстановка  $\mathbf{c}$ , компоненты которой упорядочены по возрастанию компонент вектора  $\rho$ ;  $\mathbf{c} = \Xi^{-1}\mathbf{f} = \mathbf{f}(\eta)$ ;

$\gamma = \mathbf{T}\mathbf{f}$  — вектор кумулятивных цен для вектора  $\mathbf{f}$ .

Считаем еще, что  $\varepsilon_0 = \gamma_0 = 0$ , но в векторы  $\varepsilon$  и  $\gamma$  эти величины не включаются. И справедливыми будут, например, формулы  $d_i = \varepsilon_i - \varepsilon_{i-1}$ ,  $f_i = \gamma_i - \gamma_{i-1}$ ,  $i, j \in I$ .

Можно добавить, что для сценарного рынка аналогами прогнозной функции  $f_P(\tau)$ , стоимостной функции  $f_C(\tau)$ , диссонанты  $\gamma(\varepsilon)$  и функции упорядочения  $w(x)$  [5], важных для теоретической модели характеристик, служат соответственно пары векторов  $\{\tau, \varepsilon\}$ ,  $\{\tau, \gamma\}$ ,  $\{\varepsilon, \gamma\}$  и суперпозиция  $\varepsilon(\eta)$ .

Все эти операции составляют *общую* часть алгоритма.

*Специфическая* часть алгоритма связана с назначением весов базисных инструментов в искомом портфеле — с векторами  $\mathbf{b}$  и  $\mathbf{g}$ . Первый из них показывает веса в порядке возрастания компонент вектора  $\rho$ , а второй — в исходном порядке. Вектору  $\mathbf{b}$  соответствует вектор вероятностей  $\mathbf{d}$ , а вектору  $\mathbf{g}$  — вектор  $\mathbf{p}$ .

В дискретном случае назначать веса можно не единственным способом. Обязательно лишь, чтобы выдерживался единый порядок возрастания относительных доходов от сценария к сценарию. Сначала назначается вектор  $\mathbf{b}$ , а на его основе уже однозначно определяется вектор  $\mathbf{g}$  по формуле:

$$\mathbf{g} = \Xi^{-1}\mathbf{b} = \mathbf{b}(\eta) \text{ (обратно, } \mathbf{b} = \Xi\mathbf{g} = \mathbf{g}(\xi)).$$

Портфель, получаемый применением дискретного алгоритма оптимизации для сценарного рынка, представляется в виде взвешенной суммы индикаторов:

$$\mathbf{G} = \sum_{i \in I} g_i \mathbf{D}_i.$$

Основными числовыми показателями инвестиции служат ее сумма  $A$ , средний доход  $R$  (сам доход —  $\zeta$ ), средняя доходность  $y$  и стандартное отклонение доходности  $\sigma$ . Они образуют запись  $\mathbf{J} = \langle A, R, y, \sigma \rangle$ . Для *дискретного* рынка имеем:

$$\left. \begin{aligned} A = (\mathbf{g}, \mathbf{c}) = (\mathbf{b}, \mathbf{f}); \quad R = (\mathbf{g}, \mathbf{p}) = (\mathbf{b}, \mathbf{d}); \\ y = \frac{R}{A} - 1; \quad \sigma = \left( \frac{(\mathbf{g} - R)^2}{A^2}, \mathbf{p} \right)^{1/2}. \end{aligned} \right\} (6)$$

Здесь  $(\mathbf{u}, \mathbf{v})$  означает скалярное произведение векторов  $\mathbf{u}$  и  $\mathbf{v}$ .

По векторам  $\mathbf{b}$  или  $\mathbf{g}$  можно определять и функцию распределения  $F(z)$  дохода. Ее представление с суммированием в порядке возрастания весов портфеля и в исходном порядке сценариев даются соответственно формулами:

$$F(z) = \sum_{i \in I} d_i \chi_{[b_i, \infty)}(z)$$

и

$$F(z) = \sum_{i \in I} p_i \chi_{[g_i, \infty)}(z).$$

Теперь обратимся к наиболее интересным вариантам назначения весов. Для *варианта* #0, введенного в разд. 3, производная диссонанты и инвестиционная сумма в терминах характеристик сценарного рынка приобретают вид:

$$\gamma'_0(z) = \gamma'_i = \frac{\gamma_i - \gamma_{i-1}}{\varepsilon_i - \varepsilon_{i-1}}, \quad z \in (\varepsilon_{i-1}, \varepsilon_i), \quad i \in I; \quad (7)$$

$$A_0 = \int_0^1 \phi(z) \gamma'_0(z) dz = \sum_{i \in I} \left( \gamma'_i \int_{\varepsilon_{i-1}}^{\varepsilon_i} \phi(z) dz \right). \quad (8)$$

Среди собственно дискретных наиболее естественным выглядит *вариант #1*, для которого, как нетрудно видеть, выдерживаются все ограничения CC-VaR:

$$\mathbf{b}_1 = \phi(\varepsilon); \quad \mathbf{g}_1 = \mathbf{b}_1(\eta). \quad (9)$$

В множестве вариантов, сохраняющих тот же порядок сценариев, выделяется и *вариант #2*, определяющий вектор  $\mathbf{b}$  интегральным осреднением:

$$\mathbf{b}_2 = \left\{ b_i, i \in I \right\}, \quad b_{2;i} = \int_{\varepsilon_{i-1}}^{\varepsilon_i} \frac{\phi(z) dz}{d_i}, \quad d_i = \varepsilon_i - \varepsilon_{i-1}; \quad (10)$$

$$\mathbf{g}_2 = \mathbf{b}_2(\eta).$$

Этот вариант представляет не только определенный теоретический интерес. Хотя в нем неравенства CC-VaR, как правило, нарушаются, но во многих иных отношениях *вариант #2* демонстрирует большее сходство с теоретическим *вариантом #0* и вполне может рассматриваться как приближение к нему.

Отметим еще, что для *варианта #0*, как для теоретического, справедлива формула для функции распределения дохода портфеля  $\zeta$  (ср. также (5)):

$$F_{\zeta;0}(z) = F_{\zeta}^*(z) = F_{P;q}(z) = \phi^+(z), \quad z \in [\phi(0), \phi(1)].$$

Сравнение записей  $\mathbf{J}_0$  и  $\mathbf{J}_2$  весьма показательно. Из (2), (5) и (10) имеем:

$$R_0 (= R^*) = \sum_{i \in I} \left( \int_{\varepsilon_{i-1}}^{\varepsilon_i} \phi(z) dz \right) = \sum_{i \in I} b_{0,i} (\varepsilon_i - \varepsilon_{i-1}) = \sum_{i \in I} b_{2,i} d_i = R_2. \quad (11)$$

Аналогичным образом, применяя к (3) соотношения (7) и (8) и вновь учитывая результаты дискретного алгоритма, получаем для инвестиционных сумм:

$$A_0 = \int_{\varepsilon_{i-1}}^{\varepsilon_i} \phi(z) \gamma'(z) dz = \sum_{i \in I} \left( \frac{\gamma_i - \gamma_{i-1}}{\varepsilon_i - \varepsilon_{i-1}} \int_{\varepsilon_{i-1}}^{\varepsilon_i} \phi(z) dz \right) = \sum_{i \in I} (b_{2,i} f_i) = A_2. \quad (12)$$

Таким образом, в *вариантах #0* и *#2* совпадают инвестиционные суммы и средние доходы,

а потому и средние доходности. Однако оптимальные портфели в континуальном и дискретном вариантах образуются по-разному, что проявляется в стандартном отклонении (и дисперсии) доходности, и, как правило,  $\sigma_0 \neq \sigma_2$ . Для *варианта #0* используются формулы (4), для *варианта #2* — (6).

Предложим еще краткую запись алгоритма (приводится для варианта #1):

$$\rho = \frac{\mathbf{P}}{\mathbf{c}}, \quad \xi = \mathbf{O}(\rho), \quad \eta = \mathbf{O}(\xi),$$

$$\mathbf{d} = \mathbf{p}(\xi), \quad \mathbf{T} = [t_{ij}], \quad t_{ij} = \begin{cases} 1, & \text{если } i \leq j; \\ 0, & \text{если } i > j, \end{cases}$$

$$\varepsilon = \mathbf{T}\mathbf{d}; \quad \mathbf{b} = \phi(\varepsilon), \quad \mathbf{g} = \mathbf{b}(\eta).$$

## 5 Проблемы оптимальности дискретного портфеля

Если для теоретического *варианта #0* гарантируется оптимальность получаемого портфеля, то для прочих вариантов она остается под вопросом.

Без труда доказываются простые утверждения.

**Лемма 1.** Если  $p_i \equiv 1/n$ ,  $i \in I$ , то решение, доставляемое дискретным алгоритмом в варианте #1, оптимально.

**Лемма 2.** При  $n = 2$  решение, доставляемое дискретным алгоритмом в варианте #1, оптимально.

Однако в общем случае гарантировать оптимальность не удастся, хотя пример неоптимальности портфеля обнаруживается непросто (в нем  $n = 3$ ).

Пусть  $\mathbf{c} = \{0,7; 0,2; 0,1\}$ ,  $\mathbf{p} = \{0,1; 0,599; 0,301\}$ . Имеем  $p_1/c_1 < p_2/c_2 \leq p_3/c_3$ , и потому  $\mathbf{d} = \mathbf{p}$ ,  $\mathbf{f} = \mathbf{c}$ . Применение дискретного алгоритма дает:

$$\rho = \{0,142857; 2,995; 3,01\}; \quad \xi = \{1, 2, 3\};$$

$$\mathbf{b} = \mathbf{g} = \{0,01; 0,488601; 1,0\};$$

$$A = 0,20472; \quad R = 0,594672; \quad y = 1,9048.$$

Однако в этом случае при  $\xi' = \{1, 3, 2\}$  доходность выше:

$$A' = 0,22308; \quad R' = 0,648401; \quad y' = 1,90658.$$

Нелишне отметить, что в данном примере область параметров неоптимальности алгоритма весьма узкая и содержится для  $p_2$  в интервале  $[0,598; 0,6]$ .

Можно предложить простую процедуру, позволяющую устанавливать оптимальность портфеля, но сводящуюся к полному перебору.

Пусть в результате применения дискретного алгоритма к вектору  $\rho$  получены  $\xi = \mathbf{O}(\rho)$  и средний

относительный доход  $r$ . К произвольной подстановке  $\xi'$  вектора  $\xi$ ,  $\xi' \neq \xi$ , можно применить тот же алгоритм (опуская операции с вектором  $\rho$ ) и определить  $\eta'$ ,  $\mathbf{b}'$  и  $\mathbf{g}' = \mathbf{b}'(\eta')$  и, наконец,  $A'$ ,  $R'$  и  $r'$ . Если для всех возможных подстановок оказывается, что  $r' \leq r$ , то искомый портфель — оптимальный. Однако такая проверка технически осуществима лишь при небольших значениях  $n$ , что, по существу, оправдывает проводимые в работе исследования.

Несмотря на свою ограниченную непосредственную применимость, лемма 1 позволяет проанализировать оптимальность получаемого алгоритмом решения при возрастании степени дробления сценариев на подсценарии.

Пусть  $\Delta$  — сколь угодно малая положительная величина. Вводится  $\Delta$ -рынок, образуемый разбиением каждого сценария  $S_i$ ,  $i \in I$ , на стандартные подсценарии вероятности  $\Delta$  и один нестандартный вероятности  $\Delta'_i < \Delta$ . Можно показать, что для ограниченной ф.р.п. портфель, получаемый дискретным алгоритмом на таком  $\Delta$ -рынке, при  $\Delta \rightarrow 0$  асимптотически оптимален.

## 6 Рандомизация портфеля на сценарном рынке

Вернемся к исходному дискретному рынку. Очевидно, платежная функция такого портфеля является ступенчатой функцией, при этом выбор весов в пределах допускаемых дискретным алгоритмом диапазонов сказывается на степени выполнения ограничений СС-VaR и близости функции распределения дохода к теоретической функции  $F_0(x)$ . Так, ограничения СС-VaR выдерживаются полностью лишь для *варианта* #1, но в отношении близости к функции  $F_0(x)$ , во всяком случае «на глаз», — для *варианта* #2.

Попытаемся улучшить качество функции распределения посредством рандомизации портфеля. Такую схему будем именовать *вариантом* #3. Рандомизацию проведем, придавая весам портфеля случайный характер:

$$\mathbf{G}_3 = \sum_{i \in I} \omega_i \mathbf{D}_i, \quad \omega_i = \phi(\theta_i), \quad (13)$$

где  $\theta_i \sim R\{E_i\}$ ,  $i \in I$ , — равномерно распределенные случайные величины на последовательных полуинтервалах:

$$E_i = [\underline{e}_i, \bar{e}_i), \quad \underline{e}_i = \varepsilon_{\eta_i-1} (= 0, \eta_i = 1), \\ \bar{e}_i = \varepsilon_{\eta_i}, \quad \bar{e}_i - \underline{e}_i = p_i, \quad i \in I.$$

Случайный доход для портфеля (13) представляется в виде:

$$\zeta_3 = \sum_{i \in I} \phi(\theta_i) \pi(x; \mathbf{D}_i),$$

а функции распределения  $F_{\omega_i}(z)$  весов  $\omega_i$  базисных индикаторов на интервалах их роста (для упрощения формулы приводятся лишь для таких интервалов):

$$F_{\omega_i}(z) = P\{\omega_i \leq z\} = \\ = P\{\phi(\theta_i) \leq z\} = P\{\theta_i \leq \phi^{-1}(z)\} = \frac{\phi^{-1}(z) - \underline{e}_i}{p_i}, \\ i \in I.$$

Поскольку  $\omega_i$  является условным доходом при реализации  $S_i$ , то

$$F_{\omega_i}(z) = F_{\zeta_i;3}(z|S_i), \quad i \in I,$$

и по формуле полной вероятности находим:

$$F_{\zeta;3}(z) = \sum_{i \in I} F_{\zeta_i;3}(z|S_i)p_i = \sum_{i \in I} F_{\zeta_i;3}(z|S_{\xi_i})d_i = \\ = \phi^{-1}(z), \quad z \in [0, 1].$$

Для *варианта* #3 функция распределения дохода портфеля  $F_{\zeta;3}(z)$  совпадает с функцией  $F_{P;q}(z)$  в формуле (5). Поэтому, очевидно, имеет место (с учетом (11)):

$$R_3 = R_0 (= R_2). \quad (14)$$

При этом также инвестиционная сумма  $A_3$  для рандомизированного портфеля является случайной величиной. Тем не менее ее среднее значение

$$A_3 = EA_3 = \sum_{i \in I} c_i E\omega_i,$$

где

$$E\omega_i = \int_{\underline{e}_i}^{\bar{e}_i} \frac{\phi(t) dt}{p_i},$$

и

$$A_3 = \sum_{i \in I} \frac{c_i}{p_i} \int_{\underline{e}_i}^{\bar{e}_i} \phi(t) dz = A_2 = A_0. \quad (15)$$

Такое совпадение средних имеет место при *любом* совместном распределении вероятностей компонент вектора  $\theta$  (и  $\omega$ ). Но оно не распространяется уже на средние значения *относительного* дохода. Даже в самом простом и естественном случае принимаемой далее *взаимной независимости* компонент вектора  $\theta$  функции распределения *относительного*



дохода, их математические ожидания и дисперсии в вариантах #0 и #3 могут различаться, а требования СС-VaR — не выполняться. Более того, в силу высокой кратности интегралов, возникающих при расчетах, понадобятся специальные численные методы (см. разд. 8).

### 7 Иллюстративный пример

Допустим, что плотности  $p(x)$  и  $c(x)$  подчиняются бета-распределению:  $p(x) \sim \text{Be}(\alpha_1, \alpha_2)$ ,  $c(x) \sim \text{Be}(\beta_1, \beta_2)$ , при этом  $\alpha_1 = 2; \alpha_2 = 1,5; \beta_1 = 3; \beta_2 = 2,5$ , т. е.

$$p(x) = \frac{x(1-x)^{0,5}}{B(2, 1,5)}; \quad c(x) = \frac{x^2(1-x)^{1,5}}{B(3, 2,5)},$$

$$x \in X = [0, 1], \quad (16)$$

где  $B(\alpha_1, \alpha_2)$  — бета-функция:  $B(2, 1,5) \approx 3,75$ ;  $B(3, 2,5) \approx 19,6875$ .

Также считаем, что  $\phi(\varepsilon) = \varepsilon^2$ ,  $\varepsilon \in [0, 1]$ , и потому функция распределения доходов оптимального теоретического портфеля (кстати, как и для эталонного)  $F_0(z) = \phi^{\leftarrow}(z) = z^{1/2}$ ,  $z \in [0, 1]$ .

Расчет записи  $\mathbf{J}^*$  для эталонного варианта (т. е. в соответствии с плотностями (16) и вне сферы действия предположения С) должен проводиться по формулам (2)–(5). Но для бета-распределенных плотностей точного аналитического решения получить не удастся. Поэтому потребуются либо приближенно вычислять трудоемкие параметрические интегралы, либо находить решение для дискретной модели с достаточно большим  $n$ . Второй способ определено предпочтительнее, и в результате при  $n = 5000$  получаем:

$$\mathbf{J}^* \approx \langle 0,239612; 0,333441; 0,391586; 1,24444 \rangle.$$

Заметим для сравнения, что теоретическое значение  $R^* = 1/(1 + \lambda) \approx 0,333333$ .

Сценарный рынок определяется равномерным разбиением диапазона цен (для бета-распределения)  $X = [0, 1]$  на  $n = 5$  сценариев. Вычисления дают:

$$\mathbf{p} = \{0,070; 0,187; 0,263; 0,284; 0,197\};$$

$$\mathbf{c} = \{0,041; 0,206; 0,341; 0,309; 0,104\}.$$

Затем на основе предположения С и плотности  $p(\cdot)$  из (16) рассчитывается вариант #0, для чего каждый из  $n = 5$  сценариев делится на  $k = 1000$  равных подсценариев. Результатом расчета с 5000 подсценариев будет запись:

$$\mathbf{J}_0 \approx \langle 0,253719; 0,333446; 0,314232; 1,1753 \rangle.$$

Как и должно быть, функция распределения дохода и средний доход для варианта #0 совпадают с эталонными. Однако большая (в данном случае) стоимость портфеля растягивает график функции распределения доходности по оси абсцисс в меньшей степени, а также снижает ее среднее и дисперсию.

Теперь применяется алгоритм оптимизации для  $n = 5$ . Получаем:

$$\xi = \{3, 2, 4, 1, 5\};$$

$$\eta = \{4, 2, 1, 3, 5\};$$

$$\varepsilon = \{0,263; 0,450; 0,733; 0,803; 1,0\};$$

$$\underline{e} = \{0,733; 0,263; 0; 0,450; 0,803\};$$

$$\bar{e} = \{0,803; 0,450; 0,263; 0,733; 1,0\}.$$

Веса портфеля для вариантов #1 и #2 назначаются по формулам (9) и (10):

$$\mathbf{b}_1 = \{0,069; 0,202; 0,538; 0,645; 1,0\};$$

$$\mathbf{g}_1 = \{0,645; 0,202; 0,069; 0,538; 1,0\};$$

$$\mathbf{J}_1 \approx \langle 0,361; 0,450; 0,246; 0,934 \rangle;$$

$$\mathbf{b}_2 = \{0,023; 0,130; 0,357; 0,591; 0,816\};$$

$$\mathbf{g}_2 = \{0,591; 0,130; 0,023; 0,357; 0,816\};$$

$$\mathbf{J}_2 \approx \langle 0,254; 0,333; 0,314; 1,141 \rangle.$$

В соответствии с соотношениями (4), (6), (11) и (12) первые три компоненты в записях  $\mathbf{J}_0$  и  $\mathbf{J}_2$  должны совпадать, но четвертые — различаться,  $\sigma_0 \neq \sigma_2$ .

Для вариантов #1 и #2 графики функций распределения доходов  $F_{\zeta;1}(\cdot)$  и  $F_{\zeta;2}(\cdot)$  совместно с  $F_{\zeta;0}(\cdot)$  приводятся на рис. 1 (сплошная, штриховая и пунктирная линии соответственно). Поскольку функция распределения  $F_{\zeta;3}(\cdot)$  совпадает с  $F_{\zeta;0}(\cdot)$ , она на рисунке уже не отображается.

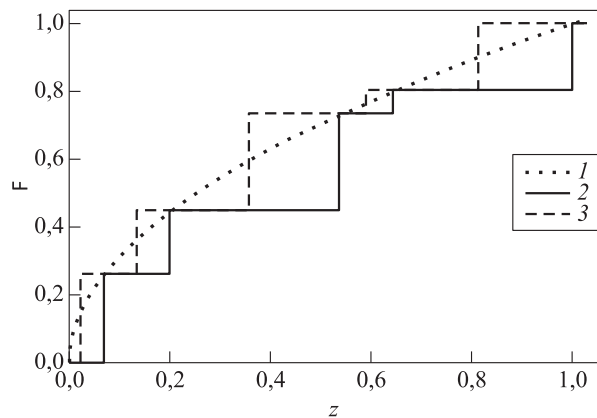


Рис. 1 Графики функций распределения доходов: 1 —  $F_{\zeta;0}$ ; 2 —  $F_{\zeta;1}$ ; 3 —  $F_{\zeta;2}$

Функции распределения для *относительного* дохода в вариантах #0, #1 и #2 получаются из аналогичных функций для *дохода* простым делением аргумента на разные (по вариантам) детерминированные стоимости. И лишь для варианта #3 из-за случайности стоимости требуются специальные вычисления.

## 8 Вычисление функции распределения относительного дохода

Относительный доход рандомизированного портфеля  $\psi_3 = \zeta_3/A_3$ , где  $\zeta_3$  — случайный доход портфеля;  $A_3$  — его стоимость (на этот раз *случайная* — отсюда и ее иное обозначение). Для этого варианта #3 имеем:

$$A_3(\theta_1, \theta_2, \dots, \theta_n) = \sum_{l \in I} c_l \phi(\theta_l).$$

Случайность стоимости и взаимозависимость величин  $\xi_3$  и  $A_3$  усложняют расчеты функции распределения  $F_{\psi;3}(\cdot)$ , которая в данном случае записывается в виде многократного интеграла с аргументом функции распределения в качестве его параметра. Для реализации приближенных вычислений предлагается применять разновидность метода Монте Карло с детерминированной решеткой значений переменных интегрирования.

Для вектора рандомизации  $\theta$  для каждого сценария  $S_i, i \in I$ , вводится равномерная решетка с  $n_J$  узлами (всего  $n_J^n$   $n$ -мерных узлов) по правилу

$$t_{i,j} = \underline{e}_i + \frac{p_i(j-1/2)}{n_J}, \quad j \in J = \{1, \dots, n_J\}, \quad i \in I,$$

порождающая решетки значений дохода и стоимости портфеля соответственно:

$$r_{i,j} = \phi(t_{i,j}), \quad j \in J, \quad i \in I, \\ a_{j_1, \dots, j_n} = \sum_{l \in I} c_l \phi(t_l, j_l), \quad j_i \in J, \quad i \in I.$$

Имеет место аппроксимирующее выборочное представление:

$$F_{\psi;3}(z) \approx \sum_{\substack{j_1, j_2, \dots, j_5 \in J; \\ i \in I}} \frac{p_i}{n_J^n} u\left(z - \frac{r_{i,j_i}}{\sum_{l \in I} c_l r_l, j_l}\right), \quad (17)$$

где  $u(\cdot)$  — характеристическая функция множества положительности аргумента.

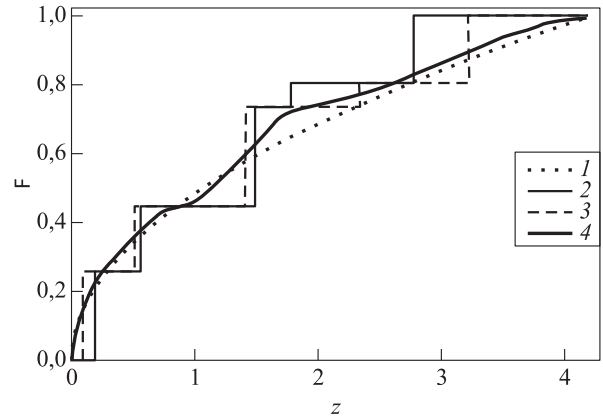


Рис. 2 Графики функций распределения относительных доходов: 1 —  $F_{\psi;0}$ ; 2 —  $F_{\psi;1}$ ; 3 —  $F_{\psi;2}$ ; 4 —  $F_{\psi;3}$

На рис. 2 изображен график функции  $F_{\psi;3}(\cdot)$  при  $n_J = 8$  (кривая 4) совместно с графиками функций  $F_{\psi;0}(\cdot)$ ,  $F_{\psi;1}(\cdot)$  и  $F_{\psi;2}(\cdot)$ .

Аналогично образуются и формулы приближенного вычисления основных числовых показателей  $\psi_3$  и  $A_3$  в варианте #3. Имеем:

$$\mu_m \approx \sum_{\substack{j_1, j_2, \dots, j_5 \in J; \\ i \in I}} \frac{p_i^S}{n_J^n} \frac{r_{i,j_i}^m}{\left(\sum_{l \in I} c_l^S r_{l,j_l}\right)^m}, \\ m = 1, 2, \quad \sigma_3^2 = \mu_2 - \mu_1^2;$$

$$E(A_3) \approx \sum_{j_1, j_2, \dots, j_5 \in J} \sum_{l \in I} \frac{c_l^S r_{l,j_l}}{n_J^n}, \quad n_J = 8; \quad n_X = 50;$$

$$E\left(\frac{1}{A_3}\right) \approx \sum_{\substack{j_1, j_2, \dots, j_5 \in J; \\ i \in I}} \frac{p_i^S / n_J^n}{\sum_{l \in I} c_l^S r_{l,j_l}}.$$

Здесь  $\mu_m$  —  $m$ -й начальный момент величины  $\psi_3$ ,  $E(A_3)$  и  $E(1/A_3)$  — средняя инвестиционная сумма и среднее число портфелей, которые можно приобрести на денежную единицу, соответственно. Формулы для них получаются очевидными трансформациями формулы для выборочного распределения (17).

Используя их, находим:

$$E\psi_3 = \mu_1 = 1,31971; \quad E\psi_3^2 = \mu_2 = 3,15321; \\ y_3 = \mu_1 - 1 = 0,31971; \quad \sigma_3 = 1,18809;$$

$$E(A_3) \approx 0,253541; \quad E\left(\frac{1}{A_3}\right) \approx 4,0136;$$

$$A_3 = E(A_3) \approx 0,253541.$$

Объединяя (14) и (15) (ср.  $A_3$ ), найденные значения  $y_3$  и  $\sigma_3$ , получаем запись:

$$J_3 \approx \langle 0,253541; 0,333446; 0,31971; 1,18809 \rangle.$$

Интересно обратить внимание на то, что при переходе к относительным доходам, как можно увидеть из рис. 1 и 2, наибольшие искажения графиков, отражающие масштаб их растяжения по оси абсцисс, наступают для варианта #1.

## Литература

1. Markowitz H. Portfolio selection // *J. Financ.*, 1952. Vol. 7. Iss. 1. P. 77–91.
2. Касимов Ю. Ф. Основы теории оптимального портфеля ценных бумаг. — М.: Филинь, 1998. 140 с.
3. Artzner P., Delbaen F., Eber J.-M., Heath D. Coherent measures of risk // *Math. Financ.*, 1999. Vol. 9. Iss. 3. P. 203–228.
4. Agasandian G. A. Optimal behavior of an investor in option market // Joint Conference (International) on Neural Networks. IEEE World Congress on Computational Intelligence. — Honolulu, Hawaii, 2002. P. 1859–1864.
5. Агасандян Г. А. Применение континуального критерия VaR на финансовых рынках. — М.: ВЦ РАН, 2011. 299 с.
6. Крамер Г. Математические методы статистики / Пер. с англ. — М.: Мир, 1975. 750 с. (*Cramer H. Mathematical methods of statistics.* — Princeton, NJ, USA: Princeton University Press, 1946. 575 p.)

Поступила в редакцию 17.04.17

---



---

## CONTINUOUS VaR-CRITERION IN SCENARIO MARKETS

G. A. Agasandyan

A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The paper investigates problems of using continuous VaR-criterion (CC-VaR) in scenario market as a discrete analog of ideal theoretical one-period option market. The participation of an investor in the market supposes that the investor prepares a forecast of future underlier’s price distribution and sets the risk-preferences function. A discrete optimization algorithm as the result of projecting the theoretical algorithm based on the Newman–Pearson procedure onto scenario market is suggested. An example of the market with three scenarios, for which the optimality can be broken, is adduced. However, such violations occur seldom and are insignificant. To improve the quality of solutions, randomization of portfolio weights as remedy of smoothing the distribution function is proposed. Special algorithms for calculations connected with yield of randomized portfolios are suggested. The exposition is illustrated by diagrams.

**Keywords:** continuous VaR-criterion (CC-VaR); scenario; forecast density; price density; investor’s risk-preferences function (r.p.f.); optimal portfolio; investment amount; income; yield; randomization

**DOI:** 10.14357/19922264180104

## Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 17-01-00816).

## References

1. Markowitz, H. 1952. Portfolio selection. *J. Financ.* 7(1):77–91.
2. Kasimov, Yu. F. 1998. *Osnovy teorii optimal'nogo portfelya tsennykh bumag* [Fundamentals of the theory of optimal security portfolio]. Moscow: Filin. 140 p.
3. Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. Coherent measures of risk. *Math. Financ.* 9(3):203–228.
4. Agasandian, G. A. 2002. Optimal behavior of an investor in option market. *Joint Conference (International) on Neural Networks. The IEEE World Congress on Computational Intelligence.* Honolulu, Hawaii. 1859–1864.
5. Agasandyan, G. A. 2011. *Primenenie kontinual'nogo kriteriya VaR na finansovykh rybnkakh* [Application of continuous VaR-criterion in financial markets]. Moscow: CC RAS. 299 p.
6. Cramer, H. 1946. *Mathematical methods of statistics.* Princeton, NJ: Princeton University Press. 575 p.

Received April 17, 2017

## Contributor

**Agasandyan Gennady A.** (b. 1941) — Doctor of Science in physics and mathematics, leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; agasand17@yandex.ru

# О ФОРМАЛЬНОЙ ПОСТАНОВКЕ ЗАДАЧ ПОИСКА СГУЩЕНИЙ В РАЗРЕЖЕННЫХ БУЛЕВЫХ МАТРИЦАХ\*

И. С. Алешин<sup>1</sup>

**Аннотация:** В огромном числе прикладных задач интеллектуального анализа данных, таких как исследование генной экспрессии и тканей, текстовой и веб-информации, рыночных корзин, клиентских сред, входная информация естественным образом представляется в виде двумерной матрицы «субъекты—объекты» («клиенты—сервисы»). Основной целью в рамках указанных областей является так называемая бикластеризация данных, т. е. выделение групп в определенном смысле схожих строк и столбцов. Небольшая часть таких задач характеризуется сильной разреженностью соответствующих матриц. Важным аспектом бикластеризации является поиск в некотором смысле плотных подматриц в булевых матрицах, что является основной целью данного исследования. В работе производится формализация предметной области в рамках алгебраического подхода, описаны системы универсальных и локальных ограничений, предложены и доказаны соответствующие критерии разрешимости рассматриваемых задач.

**Ключевые слова:** разреженные матрицы; плотные подматрицы; алгебраический подход; теоретико-множественные ограничения; бикластеризация

DOI: 10.14357/19922264180105

## 1 Введение

Данная работа выполнена в рамках алгебраического подхода, развиваемого академиком РАН Ю. И. Журавлевым [1–3] и его научной школой. Будем рассматривать булевы разреженные матрицы. Разреженные матрицы определяются как матрицы с преимущественно нулевыми элементами. Возникают ситуации, когда в исходной матрице ненулевые элементы распределены неравномерно, например существуют в некотором смысле плотные подмножества строк и столбцов. Понятия разреженности и плотности являются плохо формализованными, что делает целесообразным применение методов алгебраического подхода для поиска таких подматриц.

Выделение плотных подматриц играет важную роль во многих прикладных задачах машинного обучения и анализа данных, например анализе данных генной экспрессии и тканей в молекулярной биологии, тематического моделирования и построения рекомендательных систем. Во многих источниках используется термин «бикластеризация» (автором которого считается Миркин [4]), подразумевающий выделение групп объектов, обладающих схожими подмножествами признаков [5–7]. Значительная часть прикладных задач характеризуется сильной разреженностью данных.

Однако возможно, что данные матрицы могут иметь в некотором смысле «плохую» структуру, на-

пример ненулевые элементы могут быть распределены равномерно или входные данные могут быть сильно зашумлены, что затрудняет получение ответа на вопрос о наличии плотных подматриц. Тем не менее при решении многих прикладных задач предполагается, что положения точек соответствуют некоторым ярко выраженным, возможно зашумленным, подмножествам.

Данная работа устроена следующим образом. В разд. 2 проводится формализация предметной области, в подразд. 3.1 предложены возможные универсальные и локальные ограничения, в подразд. 3.2 описаны понятия разрешимости задач для случая одного сгущения, а также приведены и доказаны соответствующие критерии.

## 2 Постановка задачи

Рассматривается задача синтеза алгоритмов  $A$ , реализующих отображения из пространства возможных начальных информации  $\mathfrak{I}_i$  в пространство возможных финальных информации  $\mathfrak{I}_f$ . Множество всех отображений из  $\mathfrak{I}_i$  в  $\mathfrak{I}_f$  обозначим  $\mathfrak{M}_*$ :

$$\mathfrak{M}_* = \{A | A : \mathfrak{I}_i \rightarrow \mathfrak{I}_f\}.$$

В данном случае множество начальных информации представляет собой множество всевозможных  $\{0, 1\}$ -матриц размера  $m \times n$ :

$$\mathfrak{I}_i = \mathfrak{C}_{mn}\{0, 1\},$$

\* Работа выполнена при поддержке РФФИ (проект 17-20-02200).

<sup>1</sup> Факультет вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова, [iyaaln@yandex.ru](mailto:iyaaln@yandex.ru)



где  $\mathfrak{C}_{mn}\{0, 1\}$  обозначает множество матриц порядка  $m \times n$  с элементами из множества  $\{0, 1\}$ .

Далее будем рассматривать два случая возможных пространств финальных информаций. Наиболее тривиальной является ситуация, когда в матрице требуется найти не более одной подматрицы. В данном случае пространство финальных информаций (обозначим  $\mathfrak{J}_f^1$ ) выглядит следующим образом:

$$\mathfrak{J}_f^1 = \{(u, v), u \in \{0, 1\}^m, v \in \{0, 1\}^n\},$$

т. е. пространство финальных информаций представляет собой множество упорядоченных пар двоичных векторов длины  $m$  и  $n$ . Данные векторы характеризуют принадлежность строк и столбцов подматрице, т. е. строка с номером  $i$  (столбец с номером  $j$ ) принадлежит подматрице некоторой матрицы  $U \in \mathfrak{J}_f$ , если  $u_i = 1$  ( $v_j = 1$ ), и не принадлежит в противном случае. Заметим, что пара  $(u, v)$ , для которой

$$u = (\underbrace{0, \dots, 0}_m)$$

либо

$$v = (\underbrace{0, \dots, 0}_n),$$

соответствует пустой подматрице, что рассматриваться не будет. Финальную информацию, соответствующую случаю, когда искомая подматрица не найдена, обозначим  $\Delta$ . Также отметим, что заведомо не будем рассматривать нулевую матрицу, поскольку она не содержит ненулевых подматриц.

Рассмотрим произвольные  $U \in \mathfrak{C}_{mn}\{0, 1\}$ ,  $u \in \{0, 1\}^m$ ,  $v \in \{0, 1\}^n$ . Обозначим символом  $U(u, v)$  подматрицу матрицы  $U$ , образованную строками с номерами  $\{i | u_i = 1\}$  и столбцами с номерами  $\{j | v_j = 1\}$ .

Обозначим через  $\|u\| = \sum_{i=1}^m u_i$  число строк, входящих в подматрицу  $U(u, v)$ , а через  $\|v\| = \sum_{j=1}^n v_j$  соответствующее число столбцов.

**Определение 2.1.** *Площадью* подматрицы  $U(u, v)$  матрицы  $U$  будем называть величину  $\varepsilon(U(u, v))$ , равную  $\|u\| * \|v\|$ .

**Определение 2.2.** *Плотностью* подматрицы будем называть отношение числа ее ненулевых элементов к ее размеру. Обозначим плотность  $\rho(U(u, v))$ .

Для общего случая некоторого числа подматриц определим пространство финальных информаций  $\mathfrak{J}_f$  как множество векторов нефиксированного размера  $K$ , элементы которых представляют собой попарно различные упорядоченные пары булевых векторов принадлежностей строк и столбцов подматрицам:

$$\begin{aligned} \mathfrak{J}_f &= \{(\bar{u}, \bar{v}) = ((u^1, v^1), \dots, (u^K, v^K)), \\ u^i &\in \{0, 1\}^m, v^i \in \{0, 1\}^n, \text{ при } (u^i, v^i) \neq (u^j, v^j), \\ u^k &\neq (\underbrace{0, \dots, 0}_m), v^r \neq (\underbrace{0, \dots, 0}_n) \\ &\text{ для всех } i, j, k, r \in \{1, \dots, K\}, \end{aligned}$$

где  $K$  — число рассматриваемых подматриц некоторой матрицы  $U$ . Стоит заметить, что  $K \in \{1, \dots, (2^m - 1) * (2^n - 1)\}$ .

При данном представлении финальной информции может возникнуть неоднозначность. Рассмотрим следующий пример матрицы  $5 \times 5$ :

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Предположим, необходимо найти подматрицы, характеризующиеся максимальной плотностью и нерасширяемостью по размеру с неуменьшением плотности. Как можно легко видеть, исходная матрица содержит две такие подматрицы. Однако в данном случае существуют два соответствующих представления, отличающихся порядком подматриц:

$$\begin{aligned} &(((1, 1, 0, 0, 0), (1, 1, 1, 0, 0)), \\ &((0, 0, 1, 1, 1), (0, 0, 0, 1, 1))) \end{aligned}$$

и

$$\begin{aligned} &(((0, 0, 1, 1, 1), (0, 0, 0, 1, 1)), \\ &((1, 1, 0, 0, 0), (1, 1, 1, 0, 0))) . \end{aligned}$$

Для устранения данной неоднозначности будем считать, что все множества пар булевых векторов, соответствующих финальным информациям, упорядочены по размеру подматриц:

$$\|u^1\| * \|v^1\| \geq \|u^2\| * \|v^2\| \geq \dots \geq \|u^k\| * \|v^k\|.$$

При совпадении площадей подматриц упорядочим их таким образом, чтобы  $\rho(u^i, v^i) \geq \rho(u^j, v^j)$ . При совпадении плотностей будем считать, что первой идет подматрица, имеющая строку (при полном совпадении строк — столбец) с наименьшим номером.

Также стоит отметить, что при указанном выше упорядочении в рамках каждого элемента множества финальной информции

$$\begin{aligned} |\mathfrak{J}_f| &= \sum_{k=1}^{(2^m-1)(2^n-1)} \binom{(2^m-1)(2^n-1)}{k} = \\ &= 2^{(2^m-1)(2^n-1)} - 1. \end{aligned}$$

Перейдем к рассмотрению понятия алгоритма выделения сгущений.

**Определение 2.3.** Алгоритмом выделения подматриц будем называть всякое отображение  $A : \mathfrak{J}_i \rightarrow \mathfrak{J}_f$  (или  $\mathfrak{J}_i \rightarrow \mathfrak{J}_f^1$ ).

Согласно теме данной работы интерес представляют в некотором смысле плотные подматрицы. Плотные подматрицы также будем именовать сгущениями. Из интуитивных соображений следует, что далеко не любые подматрицы можно считать плотными. Например, в сгущения не должны входить «малоплотные» строки (столбцы) при наличии ненулевых строк (столбцов), не входящих в сгущение, с большей плотностью. Множество разумных требований к сгущениям накладывает определенные ограничения на алгоритмы. Таким образом, задачи определяются структурной информацией  $I_s$ , выделяющей из множества всех отображений  $\mathfrak{M}_*$  подмножества допустимых отображений, обозначаемых  $\mathfrak{M}[I_s]$ . Алгоритмы, реализующие данные отображения, будем называть корректными. Требования к алгоритмам будут подробно рассмотрены в следующем разделе.

### 3 Универсальные и локальные ограничения. Разрешимые задачи

В алгебраическом подходе синтез алгоритмов осуществляется на базе структурной информации, т.е. информации о требованиях к искомым алгоритмам [8–12]. Характерной частью структурной информации являются так называемые локальные ограничения, представляющие собой набор прецедентов.

**Определение 3.1.** Набором прецедентов называется произвольная совокупность  $(S_1, \dots, S_q)$ , где  $S_i = (U^i, (\bar{u}^i, \bar{v}^i))$ ,  $U^i \in \mathfrak{J}_i$ ,  $(\bar{u}^i, \bar{v}^i) \in \mathfrak{J}_f$  (или  $(u^i, v^i) \in \mathfrak{J}_f^1$ ) для некоторого  $q \in \mathbb{N}$ . Иными словами, множество прецедентов представляет собой набор матриц, для которых известны расположения подматриц, интерпретируемых как сгущения.

Однако во многих задачах явно недостаточно рассматривать только прецедентную информацию для построения корректного алгоритма. В рамках алгебраического подхода при недостаточно формализованной постановке задачи легко получить формально верные, но абсолютно бесполезные результаты. Например, в рамках задачи выделения плотных подматриц, если потребовать, чтобы для некоторого алгоритма  $A$  выполнялась в определенном смысле «точность на прецедентах», то фор-

мально правильным окажется алгоритм, такой что  $A(U^i) = (\bar{u}^i, \bar{v}^i) \forall U^i$ , принадлежащих прецедентной информации, и принимающий произвольное константное значение для всех остальных  $U^i$ . В данном случае возникает так называемое явление переобучения. Ясно, что данные алгоритмы являются «несодержательными» в текущей постановке.

Таким образом, в постановке задачи необходимо учитывать дополнительные ограничения на пары («начальная информация», «финальная информация»), выделяющие из множества всех отображений так называемые допустимые отображения. Данные требования получили название универсальных ограничений, которые будем обозначать  $I_s^u$ .

**Определение 3.2.** Универсальными ограничениями  $I_s^u$  будем называть систему требований, представляющую собой набор предикатов  $I_s^u = \{I_s^{u_1}, \dots, I_s^{u_N}, \dots\}$ , где

$$I_s^{u_i} : (\mathfrak{J}_i \times \mathfrak{J}_f) \rightarrow \{0, 1\} \text{ ( или } \mathfrak{J}_i \times \mathfrak{J}_f^1 \rightarrow \{0, 1\} \text{ ) .}$$

Описание конкретных примеров систем универсальных ограничений и будет одной из главных целей текущего раздела.

#### 3.1 Описание универсальных ограничений для случая одного сгущения

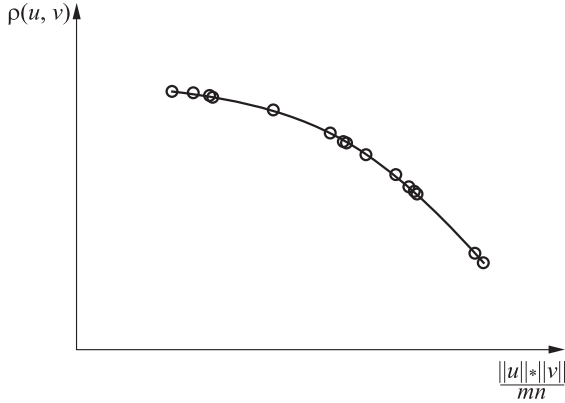
Приведем примеры некоторых «разумных» требований. Для начала рассмотрим ситуацию, когда матрица содержит не более одного сгущения.

##### Требование максимальности плотности/площади

Обозначим данное ограничение  $I_s^{u_{\max-\rho s}}$ . Пусть  $U$  — произвольная матрица из  $\mathfrak{C}_{mn}\{0, 1\}$ ,  $(u, v) \in \mathfrak{J}_f^1$  — соответствующая ей финальная информация. Будем говорить, что пара  $(U, (u, v))$  удовлетворяет требованию максимальности плотности/площади  $I_s^{u_{\max-\rho s}}$  (т.е.  $I_s^{u_{\max-\rho s}}(U, (u, v)) = 1$ ), если не существует финальной информации  $(u', v') \in \mathfrak{J}_f^1$ , такой что

$$((\rho(u', v') > \rho(u, v)) \& (\mathfrak{s}(u', v') \geq \mathfrak{s}(u, v))) \vee \vee ((\rho(u', v') \geq \rho(u, v)) \& (\mathfrak{s}(u', v') > \mathfrak{s}(u, v))) .$$

Таким образом, не должно существовать подматрицы, более плотной при данной площади или подматрицы большей плотности при данной площади. Если рассмотреть двумерную область, точки которой представляют собой подматрицы, характеризующиеся плотностью



Точки, характеризующие требование максимальной плотности/площади

и площадью  $(\rho(u, v), (||u|| * ||v||)/(mn))$ , то данное требование соответствует множеству точек  $(\rho(u, v), (||u|| * ||v||)/(mn))$ , которое определяет некоторую монотонно убывающую кривую (см. рисунок).

Стоит отметить тот факт, что для некоторых матриц из  $J_i$  ни одна подматрица некоторой площади  $m' * n'$ ,  $m' \leq m$ ,  $n' \leq n$ , не может считаться плотной при данном ограничении. Рассмотрим следующий пример матрицы  $3 \times 3$ :

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

В данной матрице не существует подматриц площади 1 и 3, удовлетворяющих требованию максимальной плотности.

### Требование максимальной плотности/размеров

Обозначим данное ограничение  $I_s^{u_{max-rc}}$ . Данное требование определяется аналогично ограничению  $I_s^{u_{max-ps}}$ , только вместо площади используются параметры числа строк и столбцов, т.е. для  $U$  — произвольной матрицы из  $\mathfrak{C}_{mn}\{0, 1\}$ ,  $(u, v) \in \mathfrak{J}_f^1$  не существует финальной информации  $(u', v') \in \mathfrak{J}_f^1$ , такой что

$$\begin{aligned} & ((\rho(u', v') > \rho(u, v)) \& \\ & \quad (||u'|| \geq ||u||) \& (||v'|| \geq ||v||)) \vee \\ & \quad \vee ((\rho(u', v') \geq \rho(u, v)) \& \\ & \quad \& (||u'|| > ||u||) \& (||v'|| \geq ||v||)) \vee \\ & \quad \vee ((\rho(u', v') \geq \rho(u, v)) \& \\ & \quad \& (||u'|| \geq ||u||) \& (||v'|| > ||v||)). \end{aligned}$$

Для описания остальных ограничений определим следующую величину:

$$\forall x^1, x^2 \in \{0, 1\}^l, \quad r(x^1, x^2) = \sum_{i=1}^l [x_i^1 \& x_i^2],$$

где  $[I]$  — индикатор события  $I$ .

### Требование сильной связности

Обозначим данное требование  $I_s^{u_{cons}}$ .

Интуитивно данное требование можно определить следующим образом. Если рассматривать матрицу  $U(u, v)$  как матрицу смежности некоторого графа, то требование сильной связности соответствует наличию пути длины 2 между любыми двумя вершинами этого графа.

Рассмотрим пример матрицы  $U^{6 \times 6}$ :

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Данное ограничение запрещает считать плотной подматрицу

$$(u, v) = ((1, 1, 1, 1, 1, 0), (1, 1, 1, 1, 1, 0)),$$

поскольку, например,

$$r(U(e^{6[3]}, v), U(e^{6[4]}, v)) = 0.$$

Формализуем данное требование. Пусть  $U \in \mathfrak{C}_{mn}\{0, 1\}$ ,  $(u, v) \in \mathfrak{J}_f^1$  — соответствующая ей финальная информация. Обозначим  $I = \{i : u(i) = 1\}$ ,  $J = \{j : v(j) = 1\}$ .

Обозначим  $e^{m[i]}$  вектор длины  $m$ , у которого  $i$ -я компонента равна 1, все остальные — 0. Тогда не существует  $i_1, i_2 \in I$ , что  $r(U(e^{m[i_1]}, v), U(e^{m[i_2]}, v)) = 0$ . Не существует  $j_1, j_2 \in J$ , что

$$r(U(u, e^{n[j_1]}), U(u, e^{n[j_2]})) = 0.$$

### Требование связности

Обозначим данное требование  $I_s^{u_{con}}$ .

Если рассматривать матрицу  $U(u, v)$  как матрицу смежности некоторого графа, то требование связности соответствует наличию пути между любыми двумя вершинами этого графа.

Формализуем данное требование. Пусть  $U \in \mathfrak{C}_{mn}\{0, 1\}$ ,  $(u, v) \in \mathfrak{J}_f^1$  — соответствующая ей финальная информация. Обозначим  $U' = U(u, v)$ . Пусть  $U' \in \mathfrak{C}_{m'n'}\{0, 1\}$ . Обозначим  $I' = \{1, \dots, m'\}$ ,  $J' = \{1, \dots, n'\}$ .

Тогда существуют  $I_1, \dots, I_l$ :

(1)  $\bigcup_{k=1}^l I_k = I'$ ,  $I_i \cap I_j = \emptyset$  при  $i \neq j$ , что для любого  $p$ , для любых  $i_1, i_2 \in I_p$

$$r\left(U\left(e^{m'[i_1]}, \{1\}^{n'}\right), U\left(e^{m'[i_2]}, \{1\}^{n'}\right)\right) \neq 0;$$

(2)  $\forall I_i, I_j \exists i_1, i_2, i_1 \in I_i, i_2 \in I_j$

$$r\left(U\left(e^{m'[i_1]}, \{1\}^{n'}\right), U\left(e^{m'[i_2]}, \{1\}^{n'}\right)\right) \neq 0.$$

Требование существования разбиения для столбцов определяется аналогично.

Рассмотренный выше пример также не удовлетворяет требованию связности. Рассмотрим следующий пример матрицы  $U^{6 \times 6}$ :

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Данное ограничение не запрещает считать плотной подматрицу  $(u, v) = ((1, 1, 1, 1, 1, 0), (1, 1, 1, 1, 1, 0))$ , однако ограничение сильной связности запрещает считать данную подматрицу плотной, поскольку  $r(U(e^{6[3]}, v), U(e^{6[5]}, v)) = 0$ .

Рассмотрим некоторые свойства предложенных ограничений.

**Утверждение 1.** Пусть  $(U, (u, v))$  — произвольная пара (начальная информация, финальная информация). Тогда из выполнения условия  $I_s^{u \text{ cons}}(U, (u, v)) = 1$  следует выполнение условия  $I_s^{u \text{ conl}}(U, (u, v)) = 1$ .

**Определение 3.3.** Система ограничений  $I_s^u$  называется покрывающей, если для любой матрицы  $U \in \mathfrak{J}_i$  существует пара  $(u, v) \in \mathfrak{J}_f^1$ , такая что  $I_s^u((U, (u, v))) = 1$ .

**Утверждение 2.** Системы ограничений  $I_s^{u \text{ max-} \rho \text{rc}}$  и  $I_s^{u \text{ max-} \rho \text{s}}$  — покрывающие.

Доказательство данного утверждения очевидно, поскольку для любой матрицы  $U \in \mathfrak{J}_i$  выполнены условия:

$$\begin{aligned} I_s^{u \text{ max-} \rho \text{rc}}((U, (\{1\}^m, \{1\}^n))) &= 1; \\ I_s^{u \text{ max-} \rho \text{s}}((U, (\{1\}^m, \{1\}^n))) &= 1. \end{aligned}$$

**Утверждение 3.** Системы ограничений  $I_s^{u \text{ cons}}$  и  $I_s^{u \text{ conl}}$  — покрывающие.

Для доказательства данного утверждения для любой матрицы  $U \neq O^{m \times n}$  можно выбрать  $(u, v) \in \mathfrak{J}_f^1$ , соответствующие подматрице из одной строки и одного столбца, содержащих ненулевой элемент. Поскольку случай нулевой матрицы  $O^{m \times n}$

не рассматривается в данной работе, утверждение доказано.

**Утверждение 4.** Системы ограничений  $\{I_s^{u \text{ cons}}, I_s^{u \text{ max-} \rho \text{s}}\}$ ,  $\{I_s^{u \text{ conl}}, I_s^{u \text{ max-} \rho \text{s}}\}$ ,  $\{I_s^{u \text{ cons}}, I_s^{u \text{ max-} \rho \text{rc}}\}$  и  $\{I_s^{u \text{ conl}}, I_s^{u \text{ max-} \rho \text{rc}}\}$  — покрывающие.

Для доказательства данного утверждения для любой матрицы  $U \neq O^{m \times n}$  можно выбрать  $(u, v) \in \mathfrak{J}_f^1$ , соответствующие подматрице из единиц максимального размера.

### 3.2 Разрешимые задачи для случая одного сгущения

Итак, задача  $Z$  определяется множеством начальных информации  $\mathfrak{J}_i$ , множеством финальных информации  $\mathfrak{J}_f^1$ , а также совокупностью универсальных ( $I_s^u$ ) и локальных ( $I_s^l$ ) ограничений. Локальные ограничения представляются в виде прецедентной информации  $(S_1, \dots, S_q)$ , где  $S_i = (U^i, (u^i, v^i))$ ,  $U^i \in \mathfrak{J}_i$ ,  $(u^i, v^i) \in \mathfrak{J}_f^1$ .

Таким образом,  $Z = (S, \mathfrak{J}_i, \mathfrak{J}_f^1, I_s^u)$ .

В качестве системы универсальных ограничений  $I_s^u$  рассмотрим систему  $\{I_s^{u \text{ cons}}, I_s^{u \text{ max-} \rho \text{s}}\}$  (аналогично рассматриваются системы  $\{I_s^{u \text{ conl}}, I_s^{u \text{ max-} \rho \text{s}}\}$ ,  $\{I_s^{u \text{ cons}}, I_s^{u \text{ max-} \rho \text{rc}}\}$ ,  $\{I_s^{u \text{ conl}}, I_s^{u \text{ max-} \rho \text{rc}}\}$ ,  $\{I_s^{u \text{ max-} \rho \text{s}}\}$  и  $\{I_s^{u \text{ max-} \rho \text{rc}}\}$ ).

**Определение 3.4.** Финальная информация  $(u, v) \in \mathfrak{J}_f^1$  называется допустимой для начальной информации  $U \in \mathfrak{C}_{mn}\{0, 1\}$  в рамках системы  $I_s^u$ , если  $I_s^u((U, (u, v))) = 1$ .

**Определение 3.5.** Алгоритм  $A : \mathfrak{J}_i \rightarrow \mathfrak{J}_f^1$  будем называть подходящим в рамках системы  $I_s^u$ , если для любой начальной информации  $A(U)$  — допустимая финальная информация.

**Определение 3.6.** Алгоритм  $A : \mathfrak{J}_i \rightarrow \mathfrak{J}_f^1$  будем называть корректным на прецедентах  $(S_1, \dots, S_q)$ , где  $S_i = (U^i, (u^i, v^i))$ , если для любого  $i$  из множества  $\{1, \dots, q\}$  верно, что либо  $\rho(U^i(A(U^i))) = \rho(U^i(u^i, v^i))$  и  $\|u^i\| * \|v^i\| = \mathfrak{s}(U^i(A(U^i)))$ , либо  $A(U^i) = (u^i, v^i) = \Delta$ .

**Определение 3.7.** Задача  $Z$ , определяемая множеством начальных информации  $\mathfrak{J}_i$ , множеством финальных информации  $\mathfrak{J}_f^1$ , универсальными ограничениями ( $I_s^u$ ) и набором прецедентов  $(S_1, \dots, S_q)$ , называется разрешимой, если для нее существует подходящий корректный алгоритм.



### 3.2.1 Разрешимые задачи для коммутирующих алгоритмов

Далее определим некоторые классы эквивалентности прецедентной информации.

Рассмотрим  $\mathfrak{S}_m$  и  $\mathfrak{S}_n$  — симметрические группы на множествах  $\{1, \dots, m\}$  и  $\{1, \dots, n\}$ , соответствующие перестановкам номеров строк (столбцов). Рассмотрим произвольные  $s^i \in \mathfrak{S}_m$  и  $s^j \in \mathfrak{S}_n$ . Для произвольной матрицы  $U \in \mathfrak{C}_{mn}\{0, 1\}$  определим матрицу  $s^i(s^j(U))$  как матрицу, полученную перестановкой строк (столбцов) матрицы  $U$  местами согласно подстановкам  $s^i \in \mathfrak{S}_m$  на множестве номеров строк,  $s^j \in \mathfrak{S}_n$  на множестве номеров столбцов.

**Определение 3.8.** Будем называть две матрицы  $U^1, U^2 \in \mathfrak{J}_i$  эквивалентными, если существуют  $s^i \in \mathfrak{S}_m, s^j \in \mathfrak{S}_n$  такие, что  $s^i(s^j(U^1)) = U^2$ .

**Определение 3.9.** Будем говорить, что алгоритм  $A : \mathfrak{J}_i \rightarrow \mathfrak{J}_f^1$  коммутирует с  $\mathfrak{S}_m$  ( $\mathfrak{S}_n$ ), если для любой матрицы  $U \in \mathfrak{J}_i$ , для любой  $s \in \mathfrak{S}_m$  ( $\mathfrak{S}_n$ ) либо  $\rho(A(U)) = \rho(A(s(U)))$  и  $\mathfrak{s}(A(U)) = \mathfrak{s}(A(s(U)))$ , либо  $A(U) = A(s(U)) = \Delta$ .

Необходимо подчеркнуть, что часто требование коммутации определяется как  $A(s(U)) = s(A(U))$ , однако в рамках рассматриваемой работы такое определение может быть некорректным. Рассмотрим пример матрицы  $U$  размера  $3 \times 3$ :

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Данная матрица имеет две плотные подматрицы размера 4:  $((1, 1, 0), (1, 1, 0))$  и  $((1, 0, 1), (1, 1, 0))$ . Так как требуется найти лишь одно сгущение, выберем, например, подматрицу, имеющую строку с наименьшим номером, т.е.  $A(U) = ((1, 1, 0), (1, 1, 0))$ . Рассмотрим подстановки на множестве номеров строк  $s^i \in \mathfrak{S}_3$ , в циклической записи имеющую вид  $(23)(1)$ , и множестве номеров столбцов  $s^j = (12)(3)$ . Обозначим через  $s$  последовательную перестановку столбцов матрицы  $U$ , соответствующую  $s^j$ , и строк, соответствующую  $s^i$ . Подействуем  $s$  на матрицу  $U$ . В результате получим, что  $s(U) \equiv U$ ,  $A(s(U)) = ((1, 1, 0), (1, 1, 0)) \neq ((1, 0, 1), (1, 1, 0)) = s(A(U))$ .

Далее будем рассматривать алгоритмы, коммутирующие с  $\mathfrak{S}_m$  и  $\mathfrak{S}_n$ . Стоит отметить, что условие коммутации может быть задано в виде универсального ограничения на алгоритмы.

Рассмотрим набор прецедентов  $(S_1, \dots, S_q) = ((U^1, (u^1, v^1)), \dots, (U^q, (u^q, v^q)))$ , где  $U^1, \dots, U^q$  — эквивалентные матрицы.

**Определение 3.10.** Набор прецедентов  $((U^1, (u^1, v^1)), \dots, (U^q, (u^q, v^q)))$  будем называть

противоречивым, если  $\exists i_1, i_2$ , что матрицы  $U^1, U^2$  эквивалентны, однако либо  $(u^{i_1}, v^{i_1}) = \Delta$ ,  $(u^{i_2}, v^{i_2}) \neq \Delta$ , либо  $\rho(U^{i_1}(u^{i_1}, v^{i_1})) \neq \rho(U^{i_2}(u^{i_2}, v^{i_2}))$ , либо  $\mathfrak{s}(U^{i_1}(u^{i_1}, v^{i_1})) \neq \mathfrak{s}(U^{i_2}(u^{i_2}, v^{i_2}))$ . В противном случае набор пар  $((U^1, (u^1, v^1)), \dots, (U^q, (u^q, v^q)))$  будем называть непротиворечивым.

**Теорема 1 (критерий разрешимости).** Задача  $Z$ , определяемая множеством начальных информации  $\mathfrak{J}_i$ , множеством финальных информации  $\mathfrak{J}_f^1$ , универсальными ограничениями  $(I_s^u)$  и набором прецедентов  $(S_1, \dots, S_q)$ , является разрешимой, если и только если набор прецедентов непротиворечив.

**Необходимость.** Пусть задача  $Z$ , определяемая множеством начальных информации  $\mathfrak{J}_i$ , множеством финальных информации  $\mathfrak{J}_f^1$ , универсальными ограничениями  $(I_s^u)$  и набором прецедентов  $(S_1, \dots, S_q)$ , является разрешимой, т.е. существует корректный подходящий алгоритм  $A : \mathfrak{J}_i \rightarrow \mathfrak{J}_f^1$ , коммутирующий с  $\mathfrak{S}_m$  и  $\mathfrak{S}_n$ . Пусть прецедентная информация противоречива, т.е. существует пара прецедентов  $S_i = (U^i, (u^i, v^i))$  и  $S_j = (U^j, (u^j, v^j))$ , таких что матрицы  $U^i$  и  $U^j$  эквивалентны, но либо  $(u^i, v^i) = \Delta$ ,  $(u^j, v^j) \neq \Delta$ , либо  $\rho(U^i(u^i, v^i)) \neq \rho(U^j(u^j, v^j))$ , либо  $\mathfrak{s}(U^i(u^i, v^i)) \neq \mathfrak{s}(U^j(u^j, v^j))$ . Поскольку алгоритм  $A$  корректный, то либо

$$\begin{aligned} \rho(U^i(A(U^i))) &= \rho(U^i(u^i, v^i)); \\ \rho(U^j(A(U^j))) &= \rho(U^j(u^j, v^j)); \\ \mathfrak{s}(U^i(A(U^i))) &= \mathfrak{s}(U^i(u^i, v^i)); \\ \mathfrak{s}(U^j(A(U^j))) &= \mathfrak{s}(U^j(u^j, v^j)), \end{aligned}$$

либо

$$\begin{aligned} A(U^i) &= (u^i, v^i) = \Delta; \\ \rho(U^j(A(U^j))) &= \rho(U^j(u^j, v^j)); \\ \mathfrak{s}(U^j(A(U^j))) &= \mathfrak{s}(U^j(u^j, v^j)). \end{aligned}$$

Однако возникает противоречие с требованием коммутации на алгоритм  $A$ .

**Достаточность.** Пусть прецедентная информация непротиворечива.

Определим алгоритм  $A$ , коммутирующий с  $\mathfrak{S}_m$  и  $\mathfrak{S}_n$ , для произвольной матрицы  $U \in \mathfrak{J}_i$  следующим образом:

$$A(U) = \begin{cases} (u^i, v^i), & \text{если } \exists i, \text{ что для} \\ S_i = (U^i, (u^i, v^i)) & \text{выполнено } U^i = U; \\ (u, v) : (U, (u, v)) & \text{— допустимая пара} \\ & \text{в рамках системы } I_s^u \text{ иначе.} \end{cases}$$

Алгоритм  $A$  является корректным и подходящим по построению; следовательно, задача  $Z$  является разрешимой.

### 3.2.2 Разрешимые задачи для монотонных прецедентов

Рассмотрим задачу  $Z$ , которая определяется множеством начальных информации  $\mathfrak{I}_i$ , множеством финальных информации  $\mathfrak{I}_f^1$ , а также совокупностью универсальных ( $I_s^u$ ) и локальных ( $I_s^l$ ) ограничений.

Пусть для  $Z$  существует корректный подходящий алгоритм  $A$ , коммутирующий с  $\mathfrak{S}_m$  и  $\mathfrak{S}_n$ .

Рассмотрим две произвольные допустимые пары  $(U^1, (u^1, v^1))$  и  $(U^2, (u^2, v^2))$ . Пусть  $(u^1, v^1) \neq \Delta$  и  $(u^2, v^2) \neq \Delta$ .

**Определение 3.11.** Будем говорить, что  $(U^2, (u^2, v^2))$  мажорирует  $(U^1, (u^1, v^1))$  (обозначим:  $(U^2, (u^2, v^2)) > (U^1, (u^1, v^1))$ ), если  $\rho(U^2, (u^2, v^2)) \geq \rho(U^1, (u^1, v^1))$  и  $\|u^2\| * \|v^2\| \geq \|u^1\| * \|v^1\|$ , причем хотя бы одно из этих неравенств строгое.

Стоит отметить, что данное определение приведено для системы, в которую входит ограничение  $I_s^{u_{max}, p, r, c}$ . Если же в нее входит ограничение  $I_s^{u_{max}, p, r, c}$ , то условие  $\|u^2\| * \|v^2\| \geq \|u^1\| * \|v^1\|$  следует заменить на пару условий  $(\|u^1\| \geq \|u^2\|) \wedge (\|v^1\| \geq \|v^2\|)$ .

Рассмотрим произвольную допустимую пару  $(U^1, (u^1, v^1))$  и произвольную матрицу  $U^2$ .

**Определение 3.12.** Будем говорить, что матрица  $U^2$  мажорирует пару  $(U^1, (u^1, v^1))$ , если существует допустимая в рамках  $I_s^u$  финальная информация  $(u^2, v^2)$ , такая что  $(U^2, (u^2, v^2)) > (U^1, (u^1, v^1))$ .

Приведем следующий пример. Рассмотрим две пары  $(U^1, (u^1, v^1))$  и  $(U^2, (u^2, v^2))$ :

$$U^1 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}; \quad U^2 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix};$$

$$(u^1, v^1) = ((1, 1, 1, 1, 0, 0), (1, 1, 1, 0, 0, 0));$$

$$(u^2, v^2) = ((1, 1, 1, 1, 1, 0), (1, 1, 1, 0, 0, 0)).$$

Как можно видеть,  $U^1 > (U^2, (u^2, v^2))$ , однако  $(U^1, (u^1, v^1)) < (U^2, (u^2, v^2))$ . Из интуитивных соображений может показаться, что указанные пары («начальная информация», «финальная информация») противоречат друг другу.

Рассмотрим произвольную допустимую пару  $(U^1, (u^1, v^1))$  и мажорирующую матрицу  $U^2$ .

**Определение 3.13.** Будем говорить, что пары  $(U^1, (u^1, v^1))$  и  $(U^2, (u^2, v^2))$ , где  $U^2$  мажорирует  $(U^1, (u^1, v^1))$ , несогласованны (монотонно

противоречивы), если не выполнено условие  $(U^2, (u^2, v^2)) > (U^1, (u^1, v^1))$ . В противном случае пары  $(U^1, (u^1, v^1))$  и  $(U^2, (u^2, v^2))$  будем называть согласованными (монотонно непротиворечивыми).

**Определение 3.14.** Будем говорить, что алгоритм  $A : \mathfrak{I}_i \rightarrow \mathfrak{I}_f^1$  удовлетворяет требованию согласованности (монотонной непротиворечивости), если для любых  $U^1, U^2 \in \mathfrak{I}_i$  пары  $(U^1, A(U^1))$  и  $(U^2, A(U^2))$  согласованные (монотонно непротиворечивые).

Далее рассмотрим согласованные (монотонно непротиворечивые) алгоритмы  $A : \mathfrak{I}_i \rightarrow \mathfrak{I}_f^1$ .

**Определение 3.15.** Набор прецедентов  $(S_1, \dots, S_q)$  будем называть согласованным (монотонно непротиворечивым), если любые две пары  $S_i, S_j, i, j = 1, \dots, q$ , являются согласованными. В противном случае набор прецедентов  $(S_1, \dots, S_q)$  будем называть несогласованным.

**Теорема 1\* (критерий разрешимости)** Задача  $Z$ , определяемая множеством начальных информации  $\mathfrak{I}_i$ , множеством финальных информации  $\mathfrak{I}_f^1$ , универсальными ограничениями ( $I_s^u$ ) и набором прецедентов  $(S_1, \dots, S_q)$ , является разрешимой, если и только если набор прецедентов является согласованным (монотонно непротиворечивым).

**Необходимость.** Пусть задача  $Z$ , определяемая множеством начальных информации  $\mathfrak{I}_i$ , множеством финальных информации  $\mathfrak{I}_f^1$ , универсальными ограничениями ( $I_s^u$ ) и набором прецедентов  $(S_1, \dots, S_q)$ , является разрешимой, т. е. существует корректный подходящий, монотонно непротиворечивый алгоритм  $A : \mathfrak{I}_i \rightarrow \mathfrak{I}_f^1$ , коммутирующий с  $\mathfrak{S}_m$  и  $\mathfrak{S}_n$ . Пусть прецедентная информация монотонно противоречива, т. е. существует пара прецедентов  $S_i = (U^i, (u^i, v^i))$  и  $S_j = (U^j, (u^j, v^j))$ , таких что  $U^i > (U^j, (u^j, v^j))$ , однако  $(U^i, (u^i, v^i)) < (U^j, (u^j, v^j))$ . Поскольку алгоритм  $A$  корректный, то

$$\rho(U^i (A(U^i))) = \rho(U^i (u^i, v^i));$$

$$\rho(U^j (A(U^j))) = \rho(U^j (u^j, v^j));$$

$$\mathfrak{s}(U^i (A(U^i))) = \mathfrak{s}(U^i (u^i, v^i));$$

$$\mathfrak{s}(U^j (A(U^j))) = \mathfrak{s}(U^j (u^j, v^j)).$$

Однако возникает противоречие с требованием монотонной непротиворечивости на алгоритм  $A$ .

**Достаточность.** Пусть прецедентная информация монотонно непротиворечива.

Определим алгоритм  $A$  для произвольной матрицы  $U \in \mathfrak{I}_i$  следующим образом:

$$A(U) = \begin{cases} (u^i, v^i), \text{ если } \exists i, \text{ что для} \\ S_i = (U^i, (u^i, v^i)) \text{ выполнено } U^i = U; \\ (u, v): (U, (u, v)) \text{ — допустимая пара} \\ \text{в рамках системы } I_s^u \text{ иначе,} \end{cases}$$

а также чтобы алгоритм  $A$  удовлетворял условию монотонной непротиворечивости для любых пар  $(U, A(U)), (U', A(U'))$ . Алгоритм  $A$  является корректным, подходящим и монотонно непротиворечивым по построению; следовательно, задача  $Z$  является разрешимой.

## 4 Заключение

В данной работе описана формальная постановка задачи поиска сгущений в разреженных булевых матрицах, а также доказаны критерии разрешимости для двух важных типов прецедентной информации. В рамках дальнейшей работы планируется исследование понятия регулярности, а также полноты семейств алгоритмических операторов, решающих правил и корректирующих операций.

Автор выражает благодарность своему научному руководителю академику РАН Константину Владимировичу Рудакову за постоянное внимание, советы и неоценимую помощь при выполнении данной работы.

## Литература

1. Журавлев Ю. И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть I // Кибернетика, 1977. № 4. С. 5–17.
2. Журавлев Ю. И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть II // Кибернетика, 1977. № 6. С. 21–27.
3. Журавлев Ю. И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть III // Кибернетика, 1978. № 2. С. 35–43.
4. Mirkin B. Mathematical classification and clustering. — Kluwer Academic Pubs., 1996. 439 p.
5. Hartigan J. A. Direct clustering of a data matrix // J. Am. Stat. Assoc., 1972. Vol. 67. No. 337. P. 123–129.
6. Cheng Y., Church G.M. Bicustering of expression data // Conference (International) on Intelligent Systems for Molecular Biology. — AAAI Press, 2000. P. 93–103.
7. Tanay A., Sharan R., Shamir R. Discovering statistically significant biclusters in gene expression data // Bioinformatics, 2002. Vol. 18 (Suppl. 1). P. 136–144.
8. Рудаков К. В. О некоторых универсальных ограничениях для алгоритмов классификации // ЖВМ МФ, 1986. Т. 26. № 11. С. 1719–1730.
9. Рудаков К. В. Универсальные и локальные ограничения в проблеме коррекции эвристических алгоритмов // Кибернетика, 1987. № 2. С. 30–35.
10. Рудаков К. В. О применении универсальных ограничений при исследовании алгоритмов классификации // Кибернетика, 1988. № 1. С. 1–5.
11. Рудаков К. В. Об алгебраической теории универсальных и локальных ограничений для задач классификации // Распознавание, классификация, прогноз. — М.: Наука, 1989. С. 176–201.
12. Рудаков К. В., Чехович Ю. В. О проблеме синтеза обучаемых алгоритмов выделения трендов (алгебраический подход) // Прикладная математика и информатика, 2001. № 8. С. 97–113.

Поступила в редакцию 08.06.17

# ON THE FORMALIZATION OF TASKS SEARCHING DENSE SUBMATRICES IN BOOLEAN SPARSE MATRICES

I. S. Aleshin

Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation

**Abstract:** In a significant part of data mining applications such as microbiology, gene expression data, text and web information, market baskets, customer environments, input information is represented as a two-dimensional matrix “subjects–objects” (“clients–services”). The main goal of such problems is biclustering of data, based on the selection of groups in a certain sense of similar rows and columns. A lot of such problems is characterized by strong sparseness of the corresponding matrices. An important aspect of biclustering is the search in some sense of dense submatrices in boolean matrices, which is the main purpose of this research. The author formalizes subject

area within the framework of algebraic approach, describes the systems of universal and local constraints, proposes and proves the corresponding criteria for solvability of the problems under consideration.

**Keywords:** sparse matrices; dense submatrices; algebraic approach; set-theoretic constraints; biclustering

**DOI:** 10.14357/19922264180105

## Acknowledgments

The investigation was supported by the Russian Foundation for Basic Research (project No. 17-20-02200).

## References

1. Zhuravlev, Yu. I. 1977. Korrektne algebrы nad mnozhestvami nekorrektnykh (evristicheskikh) algoritmov. Chast' I [Correct algebras over sets of incorrect (heuristic) algorithms. Part I]. *Kibernetika* [Cybernetics] 4:5–17.
2. Zhuravlev, Yu. I. 1977. Korrektne algebrы nad mnozhestvami nekorrektnykh (evristicheskikh) algoritmov. Chast' II [Correct algebras over sets of incorrect (heuristic) algorithms. Part II]. *Kibernetika* [Cybernetics] 6:21–27.
3. Zhuravlev, Yu. I. 1978. Korrektne algebrы nad mnozhestvami nekorrektnykh (evristicheskikh) algoritmov. Chast' III [Correct algebras over sets of incorrect (heuristic) algorithms. Part III]. *Kibernetika* [Cybernetics] 2:35–43.
4. Mirkin, B. 1996. *Mathematical classification and clustering*. Kluwer Academic Pubs. 439 p.
5. Hartigan, J. A. 1972. Direct clustering of a data matrix. *J. Am. Stat. Assoc.* 67(337):123–129.
6. Cheng, Y., and G. M. Church. 2000. Biclustering of expression data. *Conference (International) on Intelligent Systems for Molecular Biology*. AAAI Press. P. 93–103.
7. Tanay, Y., R. Sharan, and R. Shamir. 2002. Discovering statistically significant biclusters in gene expression data. *Bioinformatics* 18(Suppl. 1):136–144.
8. Rudakov, K. V. 1986. O nekotorykh universal'nykh ogranicheniyakh dlya algoritmov klassifikatsii [Some universal constraints for classification algorithms]. *ZhVM MF* [Comp. Math. Math. Phys.] 26(11):1719–1730.
9. Rudakov, K. V. 1987. Universal'nye i lokal'nye ogranicheniya v probleme korrektsii evristicheskikh algoritmov [Universal and local constraints in the problem of correction of heuristic algorithms]. *Kibernetika* [Cybernetics] 2:30–35.
10. Rudakov, K. V. 1988. O primeneniі universal'nykh ogranicheniy pri issledovanii algoritmov klassifikatsii [About the application of universal constraints in the research of classification algorithms]. *Kibernetika* [Cybernetics] 1:1–5.
11. Rudakov, K. V. 1989. Ob algebraicheskoy teorii universal'nykh i lokal'nykh ogranicheniy dlya zadach klassifikatsii [About the algebraic theory of universal and local constraints for classification problems]. *Raspoznavanie, klassifikatsiya, prognoz* [Recognition, classification, forecast]. Moscow: Nauka. 176–201.
12. Rudakov, K. V., and Yu. V. Chekhovich. 2002. Design of trend-identification algorithms with learning (algebraic approach). *Comput. Math. Modeling* 13(3):281–293.

Received June 8, 2017

## Contributor

**Aleshin Ilya S.** (b. 1993) — PhD Student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; ilyaaln@yandex.ru



# УЧЕТ ПРЕДВАРИТЕЛЬНЫХ ОЦЕНОК СКОРОСТИ ПОРОЖДЕНИЯ СХОДСТВ СПАРИВАЮЩЕЙ ЦЕПЬЮ МАРКОВА

Д. В. Виноградов<sup>1</sup>

**Аннотация:** В современном интеллектуальном анализе данных возрастает доля методов статистического машинного обучения. Для подхода, основанного на бинарной операции сходства, таковым является вероятностно-комбинаторный формальный метод (ВКФ-метод). Его основной алгоритм — спаривающая цепь Маркова. В статье предложен механизм учета длин траекторий (до склеивания) с формированием верхней границы, по которой следует останавливать излишне длинные траектории в дальнейшем. Теоретический результат, доказанный в статье, утверждает, что при учете достаточно большого числа предварительных траекторий вероятности изменяются экспоненциально малым образом в метрике тотальной вариации. Это предложение особенно полезно, когда имеется малая доля длинных траекторий относительно остальных, так как в этом случае обеспечивается баланс между величиной границы и изменением вероятностей.

**Ключевые слова:** сходство; спаривающая цепь Маркова; метрика тотальной вариации

**DOI:** 10.14357/19922264180106

## 1 Введение

Одним из современных методов интеллектуального анализа данных является ДСМ-метод [1] автоматического порождения гипотез. Этот метод возник как синтез нескольких когнитивных процедур в трудах проф. В. К. Финна, осмыслившего идеи Д. С. Милля, Ч. С. Пирса и К. Поппера.

Одна из процедур ДСМ-метода — индуктивное обобщение обучающих примеров — была исследована примерно в это же самое время группой зарубежных ученых под руководством проф. Рудольфа Вилле под названием теории формальных понятий [2]. Однако ДСМ-метод ввел в рассмотрение понятие контрпримера, чего нет в теории формальных понятий.

Следует признать, что имеются некоторые сложности в применении ДСМ-метода к анализу данных.

Во-первых, множество порождаемых ДСМ-гипотез может оказаться экспоненциально велико по сравнению с размером обучающей выборки.

Во-вторых, С. О. Кузнецовым [3], М. И. Забежайло и др. были доказаны пессимистические оценки сложности для многих ДСМ-процедур (так называемые NP-полнота и #P-полнота).

В-третьих, попытки присоединить дедуктивный вывод оказались неудачными: Д. П. Скворцовым [4] установлена неарифметичность стандартного подхода через кванторы по конечным множествам, а автором [5] — невыразимость этой

теории средствами логики предикатов первого порядка.

Наконец, автор [6] сумел обнаружить еще одну трудность: существование так называемых случайных ДСМ-гипотез. Эти случайные гипотезы возникают тогда, когда вычисляется сходство двух (или более) обучающих примеров, каждый из которых имеет свой механизм порождения целевого свойства. Это сходство оказывается фрагментом (набором общих признаков), случайно имеющимся в каждом из этих объектов. Возникновение таких сходств следует рассматривать как аналог «переобучения» для многих методов машинного обучения, когда максимальный учет информации из обучающей выборки приводит к модели, демонстрирующей плохую предсказательную способность.

Для преодоления указанных трудностей автором [7] был предложен вероятностно-комбинаторный подход. Так как некоторые ингредиенты заимствованы из теории формальных понятий, автор назвал его ВКФ-методом.

Ключевой процедурой ВКФ-метода является вероятностный алгоритм нахождения случайного ВКФ-кандидата с помощью спаривающей цепи Маркова. Автором показано, что это действительно цепь Маркова, которая останавливается с вероятностью единица. Однако хорошей оценки на длины траекторий получить пока не удалось.

Может так случиться, что малое число траекторий имеют длину (например, экспоненциальную),

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, krrguest@yandex.ru

значительно превосходящую (например, полиномиальную) длину остальных траекторий. Тогда можно предложить использовать  $r$  предварительных прогонов цепи Маркова для получения верхней границы ожидания спаривания (как суммы длин этих предварительно вычисленных траекторий). Если текущая траектория не завершается до этой границы, то текущие вычисления прекращаются, а цепь Маркова запускается заново.

Ясно, что при такой модификации изменяются вероятности появления результатов (эргодических состояний). В настоящей работе будет получена хорошая оценка, показывающая, что при выборе достаточно большого числа  $r$  предварительных прогонов изменение вероятностей может быть сделано сколь угодно малым.

## 2 Вспомогательные определения и факты

Сходство является бинарной операцией на множестве  $X$ , объемлющем множество объектов, т.е. представляет собой отображение  $\cap : X \times X \rightarrow X$ . Элементы множества  $X$  будем называть *фрагментами*.

Для независимости результата нахождения сходства нескольких объектов операция сходства должна удовлетворять аксиомам *нижней полурешетки*.

Для выражения тривиальности сходства добавляется специальный *пустой фрагмент*  $\emptyset$  со свойством наименьшего элемента.

Важнейшим примером будет нижняя полурешетка, состоящая из битовых строк фиксированной длины с побитовым умножением в качестве операции сходства. Пустым фрагментом будет служить строка, состоящая из одних нулей. Каждый бит может быть отождествлен с бинарным признаком. Тогда битовая строка соответствует множеству признаков, в которых встречаются единицы. При этом операция сходства соответствует пересечению множеств признаков, а пустой фрагмент — пустому множеству признаков.

Ясно, что в этом примере строка из одних единиц будет соответствовать наибольшему элементу.

Важность этого примера объясняется тем, что операция побитового умножения допускает эффективную реализацию на современных ЭВМ. Существуют специальные классы объектов (например, `dynamic_bitset` в C++), реализующие удобное оперирование битовыми строками.

С другой стороны, теорема Рудольфа Вилле [2] утверждает, что эта конструкция позволяет построить любую конечную решетку из нижней полу-

решетки битовых строк с операцией побитового умножения, если к ней добавить наибольший элемент.

Собирая вместе битовые строки, представляющие объекты, получаем прямоугольную таблицу  $I$ , которую будем называть *формальным контекстом* [2].

Формальный контекст можно воспринимать как бинарное отношение между элементами множества  $O$ , которые назовем *именами объектов*, и элементами множества  $F$ , которые назовем *признаками*. Если в строке, соответствующей объекту  $o \in O$ , и столбце, соответствующем фрагменту  $f \in F$ , стоит единица, то говорим, что *объект  $o$  обладает признаком  $f$* , и обозначаем это через  $oIf$ . В противном случае говорим, что *объект  $o$  не имеет признака  $f$* .

Для подмножества  $A \subseteq O$  объектов его *сходством* называется подмножество  $A' = \{f \in F : \forall o \in A[oIf]\} \subseteq F$ . Полагаем  $\emptyset' = F$ .

На самом деле это определение совпадает с последовательным вычислением побитового умножения строк, соответствующих отобранному во множество  $A$  объектам.

Для подмножества  $B \subseteq F$  признаков его *сходством* называется подмножество  $B' = \{o \in O : \forall f \in B[oIf]\} \subseteq O$ . Полагаем  $\emptyset' = O$ .

Легко проверить стандартные свойства операции сходства.

**Определение 1.** Пару  $\langle A, B \rangle$  назовем **ВКФ-кандидатом**, если  $A = B' \subseteq O$  и  $B = A' \subseteq F$ .

Обычно такие пары называют *формальными понятиями*, но предпочтительнее сменить название, так как оригинальное название подвергается обоснованной критике со стороны специалистов по философии и искусственному интеллекту.

Легко проверить, что **множество всех ВКФ-кандидатов** (для фиксированного формального контекста) **образует решетку**.

Теорема Р. Вилле [2, 8] гласит: **Любая конечная решетка изоморфна решетке всех ВКФ-кандидатов для подходяще выбранного формального контекста**.

Основываясь на операциях сходства, зададим ключевое понятие предлагаемого подхода:

**Определение 2.** Операция *закрывай-по-одному-вниз* на ВКФ-кандидате  $\langle A, B \rangle$  и объекте  $o \in O$  порождает пару  $\text{CьODown}(\langle A, B \rangle, o) = \langle (A \cup \{o\})', (A \cup \{o\})' \rangle$ . Операция *закрывай-по-одному-вверх* на ВКФ-кандидате  $\langle A, B \rangle$  и признаке  $f \in F$  порождает пару  $\text{CьOUp}(\langle A, B \rangle, f) = \langle (B \cup \{f\})', (B \cup \{f\})' \rangle$ .

Операция  $\text{CьODown}$  соответствует шагу алгоритма «Закрывай-по-одному», который был предложен С. О. Кузнецовым [9] для вычисления всех ВКФ-кандидатов перебором сверху вниз.

**Лемма 1.** Для любого ВКФ-кандидата  $\langle A, B \rangle$  и любого объекта  $o \in O$  пара  $\text{CbODown}(\langle A, B \rangle, o)$  является ВКФ-кандидатом.

Аналогично для любого ВКФ-кандидата  $\langle A, B \rangle$  и любого признака  $f \in F$  пара  $\text{CbOUp}(\langle A, B \rangle, f)$  является ВКФ-кандидатом.

**Определение 3.** Порядок на ВКФ-кандидатах:  $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ , если  $B_1 \subseteq B_2$ .

Это двойственное (с точки зрения теории формальных понятий) определение приводится в настоящем виде для согласованности с традицией отечественной школы.

**Лемма 2.** Для всякой упорядоченной пары ВКФ-кандидатов  $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$  и любого объекта  $o \in O$  имеем  $\text{CbODown}(\langle A_1, B_1 \rangle, o) \leq \text{CbODown}(\langle A_2, B_2 \rangle, o)$ .

Для всякой упорядоченной пары ВКФ-кандидатов  $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$  и любого признака  $f \in F$  имеем  $\text{CbOUp}(\langle A_1, B_1 \rangle, o) \leq \text{CbOUp}(\langle A_2, B_2 \rangle, o)$ .

Заметим, что состоянием изменяемых переменных в цикле (состоянием спаривающей цепи Маркова) является упорядоченная пара ВКФ-объектов  $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ .

Первоначально меньший ВКФ-кандидат совпадает с наименьшим ВКФ-кандидатом  $\text{Min} := \langle O, O' \rangle$ , а больший — с наибольшим  $\text{Max} := \langle F', F \rangle$ .

В цикле к обоим ВКФ-кандидатам применяется одна и та же операция  $\text{CbODown}$  с выбранным объектом или  $\text{CbOUp}$  с выбранным признаком.

Процесс останавливается, когда меньший ВКФ-кандидат совпадает с большим. Тогда этот общий ВКФ-кандидат и выдается алгоритмом 1.

Следующая теорема из статьи [7] объясняет, откуда здесь возникает цепь Маркова.

**Теорема 1.** Алгоритм 1 соответствует цепи Маркова.

Теперь переходим к вопросу о спаривании.

**Определение 4.** Состояние вида  $\langle A, B \rangle = \langle A, B \rangle$  спаривающей цепи Маркова для совпадающей пары ВКФ-кандидатов называется эргодическим. Состояние вида  $\langle A_1, B_1 \rangle < \langle A_2, B_2 \rangle$  называется невозвратным.

Теперь имеем классическую теорему о сходимости с вероятностью единица.

**Теорема 2.** Вероятность того, что состояние  $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$  спаривающей цепи Маркова окажется невозвратным, стремится к нулю, когда  $t \rightarrow \infty$ .

В статье [10] приводится доказательство этого результата, но это классический результат в теории цепей Маркова [11], поэтому здесь не будем приводить доказательство.

### 3 Основной результат

Так как спаривающая цепь Маркова может иметь траектории существенно разной длины, то возможно применение следующей техники остановки длинной траектории и запуска цепи заново.

**Определение 5.** Если  $T_1, \dots, T_r$  — независимые целочисленные случайные величины, имеющие распределение времени  $T$  склеивания, то верхняя граница склеивания по  $r$  предварительным прогонам определяется как  $\hat{T} = T_1 + \dots + T_r$ .

На практике предлагается сделать  $r$  прогонов спаривающей цепи Маркова с соответствующими временами склеивания  $t_1, \dots, t_r$  и взять оценку  $t_1 + \dots + t_r$  верхней границы склеивания.

Оценим, как изменяются вероятности попадания в эргодические состояния при остановке спаривающей цепи Маркова по  $r$  прогонам.

**Определение 6.** Для целочисленной случайной величины  $\hat{T}$ , независимой от целочисленной случай-

---

#### Алгоритм 1: Спаривающая цепь Маркова

---

**Data:** множество обучающих (+)-примеров; внешние функции  $\text{CbOUp}(\cdot, \cdot)$  и  $\text{CbODown}(\cdot, \cdot)$  операций «замыкай-по-одному»

**Result:** ВКФ-кандидат  $\langle A, B \rangle$

$O := (+)$ -примеры,  $F :=$  признаки;  $I \subseteq O \times F$  — формальный контекст для (+)-примеров;

$R := O \cup F$ ;  $\text{Min} := \langle O, O' \rangle$ ;  $\text{Max} := \langle F', F \rangle$ ;

**while** ( $\text{Min} \neq \text{Max}$ ) **do**

    Выбираем случайный элемент  $r \in R$ ;

**if** ( $r \in O$ ) **then**

$\text{Min} := \text{CbODown}(\text{Min}, r)$ ;  $\text{Max} := \text{CbODown}(\text{Max}, r)$ ;

**end**

**else**

$\text{Min} := \text{CbOUp}(\text{Min}, r)$ ;  $\text{Max} := \text{CbOUp}(\text{Max}, r)$ ;

**end**

**end**

---

ной величины  $T$ , условное распределение состояний относительно события  $B = \{T \leq \hat{T}\}$  есть распределение

$$\mu(\hat{T})_t = \frac{\mathbb{P}[X_T = i, T \leq \hat{T}]}{\mathbb{P}[T \leq \hat{T}]}$$

для любого эргодического состояния  $i$ .

**Определение 7.** Расстояние *тотального изменения* между распределениями вероятностей  $\mu = (\mu_i)_{i \in U}$  и  $\nu = (\nu_i)_{i \in U}$  на конечном пространстве  $U$  определяется правилом:

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{i \in U} |\mu_i - \nu_i|.$$

Это расстояние является половиной метрики  $l_1$ , следовательно, само является метрикой (в частности, симметрично).

Известна классическая и доказываемая прямо из определения 7

**Лемма 3.**  $\|\mu - \nu\|_{TV} = \max_{R \subseteq U} |\mu(R) - \nu(R)|$ .

В лемме 3 подмножество  $R$ , на котором достигается максимум, определяется так:

$$R = \{i \in U | \mu_i > \nu_i\}.$$

Следующая лемма является технической.

**Лемма 4.**  $\|\mu - \mu(\hat{T})\|_{TV} \leq \mathbb{P}[T > \hat{T}] / (1 - \mathbb{P}[T > \hat{T}])$ , где  $\mu(\hat{T})$  — распределение остановленной на верхней границе  $\hat{T}$  склеивания по  $r > 1$  испытаниям, а  $\mu$  — распределение выдачи неостановленной цепи.

**Доказательство.** По определению 6  $\mu(\hat{T})_i = \mathbb{P}[X_T = i, T \leq \hat{T}] / \mathbb{P}[T \leq \hat{T}]$ . Тогда

$$\begin{aligned} \mathbb{P}[T \leq \hat{T}] (\mu(\hat{T})_i - \mu_i) &= \\ &= \mathbb{P}[X_T = i, T \leq \hat{T}] - \mathbb{P}[T \leq \hat{T}] \mu_i = \\ &= \mathbb{P}[T > \hat{T}] \mu_i - \mathbb{P}[X_T = i, T > \hat{T}] \leq \\ &\leq \mathbb{P}[T > \hat{T}] \mu_i. \end{aligned}$$

Суммируя по множеству  $R = \{i \in U | \mu_i > \mu(\hat{T})_i\}$ , получим:

$$\mathbb{P}[T \leq \hat{T}] \|\mu - \mu(\hat{T})\|_{TV} \leq \mathbb{P}[T > \hat{T}],$$

что и приводит к утверждению леммы.

Теперь докажем основную лемму.

**Лемма 5.**  $\|\mu - \mu(\hat{T})\|_{TV} \leq 1/(2^r - 1)$ , где  $\mu(\hat{T})$  — распределение остановленной на верхней границе  $\hat{T}$  склеивания по  $r > 1$  испытаниям, а  $\mu$  — распределение выдачи неостановленной цепи.

**Доказательство.** Из-за леммы 4 достаточно доказать, что

$$\mathbb{P}[T > \hat{T}] \leq 2^{-r}.$$

Из определения  $T, T_1, \dots, T_r$  как независимых одинаково распределенных случайных величин следует, что  $\mathbb{P}[T > T_j] \leq 1/2$  для всех  $1 \leq j \leq r$ .

Докажем субмультипликативность:

$$\mathbb{P}\left[T > \sum_{j=1}^k T_j\right] \leq \mathbb{P}\left[T > \sum_{j=1}^{k-1} T_j\right] \cdot \mathbb{P}[T > T_k]$$

для всех  $1 < k \leq r$ .

Но это следует из формулы условной вероятности, так как если  $T > \sum_{j=1}^{k-1} T_j$ , то

$$\begin{aligned} \left\langle A_1 \left( \sum_{j=1}^{k-1} T_j \right), B_1 \left( \sum_{j=1}^{k-1} T_j \right) \right\rangle &< \\ &< \left\langle A_2 \left( \sum_{j=1}^{k-1} T_j \right), B_2 \left( \sum_{j=1}^{k-1} T_j \right) \right\rangle. \end{aligned}$$

Поэтому, применяя ко всем четырем ВКФ-кандидатам

$$\begin{aligned} \text{Min} &\leq \left\langle A_1 \left( 0 + \sum_{j=1}^{k-1} T_j \right), B_1 \left( 0 + \sum_{j=1}^{k-1} T_j \right) \right\rangle < \\ &< \left\langle A_2 \left( 0 + \sum_{j=1}^{k-1} T_j \right), B_2 \left( 0 + \sum_{j=1}^{k-1} T_j \right) \right\rangle \leq \text{Max} \end{aligned}$$

одинаковые операции  $\text{CbODown}$  и  $\text{CbOUp}$ , имеем, что если

$$\begin{aligned} \left\langle A_1 \left( t + \sum_{j=1}^{k-1} T_j \right), B_1 \left( t + \sum_{j=1}^{k-1} T_j \right) \right\rangle &< \\ &< \left\langle A_2 \left( t + \sum_{j=1}^{k-1} T_j \right), B_2 \left( t + \sum_{j=1}^{k-1} T_j \right) \right\rangle \end{aligned}$$

спаривается позднее момента  $T_k + \sum_{j=1}^{k-1} T_j = \sum_{j=1}^k T_j$ , то и спаривание  $\text{Min} < \text{Max}$  совершается позднее момента  $T_k$ , т.е.

$$\mathbb{P}\left(T > \sum_{j=1}^k T_j | T > \sum_{j=1}^{k-1} T_j\right) \leq \mathbb{P}[T > T_k].$$



Теперь результат леммы следует по индукции.

Соединяя результаты лемм 3 и 5, получим

**Теорема 3.** Для любого  $R$  с  $\mu(R) = \rho$  и  $r > \log_2(1-1/\rho)$  имеем  $\mu(\hat{T})(R) \geq \rho - 1/(2^r - 1)$  на верхней границе  $\hat{T}$  склеивания по  $r > 1$  испытаниям.

Доказательство.

$$\begin{aligned} \rho - \frac{1}{2^r - 1} &\leq \mu(R) - \|\mu - \mu(\hat{T})\|_{TV} = \\ &= \mu(R) - \max_{Q \subseteq U} \left| \mu(Q) - \mu(\hat{T})(Q) \right| \leq \\ &\leq \mu(R) - \left| \mu(R) - \mu(\hat{T})(R) \right| \leq \mu(\hat{T})(R). \end{aligned}$$

## 4 Заключение

В настоящей работе получена оценка изменения вероятностей появления ВКФ-кандидатов, если спаривающую цепь Маркова из ВКФ-метода остановить на основании  $r$  предварительных прогонов. Эта задача естественно возникает тогда, когда малое число траекторий имеют длину (например, экспоненциальную), значительно превосходящую (например, полиномиальную) длину остальных траекторий. В такой ситуации путем подбора большого числа предварительных прогонов можно получить достаточно малую верхнюю оценку на время спаривания так, чтобы вероятности появления ВКФ-кандидатов изменились как угодно мало.

Автор благодарит проф. В. К. Финна за внимание к работе, проф. Е. М. Бениаминова за полезные обсуждения и своих коллег по лаборатории 35 ФИЦ ИУ РАН за поддержку и полезные дискуссии.

## Литература

1. Финн В. К. Об интеллектуальном анализе данных // Новости искусственного интеллекта, 2004. № 3. С. 3–18.
2. Ganter B., Wille R. Formal concept analysis. — Berlin: Springer-Verlag, 1999. 284 p.
3. Кузнецов С. О. Быстрый алгоритм построения всех пересечений объектов из нижней полурешетки // Научная и техническая информация. Сер. 2, 1993. № 1. С. 17–20.
4. Скворцов Д. П. О некоторых способах построения логических языков с кванторами по кортежам // Семиотика и информатика, 1983. Вып. 20. С. 102–126.
5. Виноградов Д. В. Формализация правдоподобных рассуждений в логике предикатов // Научная и техническая информация. Сер. 2, 2000. № 11. С. 17–20.
6. Виноградов Д. В. Вероятность порождения случайного ДСМ-сходства при наличии контр-примеров // Научная и техническая информация. Сер. 2, 2015. № 3. С. 1–5.
7. Vinogradov D. V. VKF-method of hypotheses generation // Comm. Com. Inf. Sc., 2014. Vol. 436. P. 237–248.
8. Davey B. A., Priestley H. A. Introduction to lattices and order. — 2nd ed. — Cambridge: Cambridge University Press, 2002. 298 p.
9. Кузнецов С. О. Интерпретация на графах и сложные характеристики задач поиска закономерностей определенного вида // Научная и техническая информация. Сер. 2, 1989. № 1. С. 23–28.
10. Виноградов Д. В. Вероятностное порождение гипотез в ДСМ-методе с помощью простейших цепей Маркова // Научная и техническая информация. Сер. 2, 2012. № 9. С. 20–27.
11. Кемени Дж., Снелл Дж., Кнэпп А. Счетные цепи Маркова / Пер. с англ. — М.: Наука, 1987. 416 с. (Kemeny J. G., Snell J. L., Knapp A. W. Denumerable Markov chains. — Princeton, NJ, USA: Van Nostrand, 1966. 439 p.)

Поступила в редакцию 24.04.17

# INFLUENCE OF PRELIMINARY ESTIMATES ON THE SPEED OF SEARCH OF SIMILARITIES BY THE COUPLING MARKOV CHAIN

D. V. Vinogradov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** At present, Data Mining expands usage of statistical machine learning methods. The similarity-based approach uses the probabilistic combinatorial formal method (VKF (variational Kalman filter) method). The main algorithm is based on a coupling Markov chain. The authors propose a mechanism to convert lengths of preliminary trajectories (before coalescence) to an upper bound on which it is necessary to stop excessively long successive

runs. The main result claims that the change of probabilities is exponentially small with respect to total variation distance, if the chain uses sufficient number of preliminary runs. This proposal may be useful when there exists a small fraction of long trajectories with respect to the rest, because it provides a balance between the size of the bound and changes of probabilities.

**Keywords:** similarity; Markov chain; VKF candidate; total variation; coupling

**DOI:** 10.14357/19922264180106

## References

1. Finn, V. K. 2004. Ob intellektual'nom analize dannykh [On intelligent data analysis]. *Novosti iskusstvennogo intellekta* [Artificial Intelligence News] 3:3–18.
2. Ganter, B., and R. Wille. 1999. *Formal concept analysis*. Berlin: Springer-Verlag. 284 p.
3. Kuznetsov, S. O. 1993. Bystryy algoritm postroeniya vsexh persecheniy ob"ektov iz nizhney polureshetki [Fast algorithm for computation of all intersections of objects from a low-semilattice]. *Nauchnaya i tekhnicheskaya informatsiya. Ser. 2* [Scientific and technical information. Ser. 2] 1:17–20.
4. Skvortsov, D. P. 1983. O nekotorykh sposobakh postroeniya logicheskikh yazykov s kvantorami po kortezham [On some ways to construct logical languages with tuples quantifiers]. *Semiotika i informatika* [Semiotics and Informatics] 20:102–126.
5. Vinogradov, D. V. 2000. Formalizing plausible arguments in predicate logic. *Automatic Documentation Math. Linguistics* 34(6):6–10.
6. Vinogradov, D. V. 2015. The probability of encountering an accidental DSM similarity in the presence of counter examples. *Automatic Documentation Math. Linguistics* 49(2):43–46.
7. Vinogradov, D. V. 2014. VKF-method of hypotheses generation. *Comm. Comp. Inf. Sc.* 436:237–248.
8. Davey, B. A., and H. A. Priestley. 2002. *Introduction to lattices and order*. 2nd ed. Cambridge: Cambridge University Press. 298 p.
9. Kuznetsov, S. O. 1989. Interpretatsiya na grafakh i slozhnostnye kharakteristiki zadach poiska zakonomernostey opredelenogo vida [Graphs interpretation and complexity characteristics of tasks of search for patterns of special form]. *Nauchnaya i tekhnicheskaya informatsiya. Ser. 2* [Scientific and technical information. Ser. 2] 1:23–28.
10. Vinogradov, D. V. 2012. Random generation of hypotheses in the JSM method using simple Markov chains. *Automatic Documentation Math. Linguistics* 46(5):221–228.
11. Kemeny, J. G., J. L. Snell, and A. W. Knapp. 1966. *Denumerable Markov chains*. Princeton, NJ: Van Nostrand. 439 p.

Received April 24, 2017

## Contributor

**Vinogradov Dmitry V.** (b. 1964) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44/2 Vavilova Str., Moscow 119333, Russian Federation; krrguest@yandex.ru

# МЕТОД ИНТЕРПОЛЯЦИОННОГО АНАЛИТИЧЕСКОГО МОДЕЛИРОВАНИЯ ОДНОМЕРНЫХ РАСПРЕДЕЛЕНИЙ В СТОХАСТИЧЕСКИХ СИСТЕМАХ

И. Н. Синицын<sup>1</sup>

**Аннотация:** Среди известных методов аналитического моделирования одномерных распределений стохастических процессов (СтП) в дифференциальных и интегродифференциальных стохастических системах (СтС), основанных на прямом численном решении уравнения для одномерной характеристической функции (х.ф.), важное место занимают интерполяционные методы, восходящие к работам С. В. Мальчикова. При этом для интерполирования х.ф. использовалась теорема отсчетов В. А. Котельникова. В статье для двух типов скалярных нелинейных дифференциальных негауссовских СтС на базе уравнения для одномерной х.ф. разработан метод интерполяционного аналитического моделирования (МИАМ). Получены уравнения для функций влияния (чувствительности) параметров, входящих в уравнения СтС. Приведенный тестовый пример показывает эффективность метода для существенно нелинейной СтС при небольшом числе членов интерполяционного разложения. Формулируются некоторые обобщения метода.

**Ключевые слова:** одномерная плотность вероятности (п.в.); одномерная характеристическая функция (х.ф.); стохастическая система (СтС); стохастический процесс (СтП)

**DOI:** 10.14357/19922264180107

## 1 Введение

Среди известных методов аналитического моделирования СтП в дифференциальных и интегродифференциальных СтС, основанных на прямом численном решении уравнения для одномерной х.ф. [1], важное место занимают интерполяционные методы [2, 3]. При этом для интерполирования х.ф. используется теорема отсчетов В. А. Котельникова [4].

Рассмотрим обобщение результатов С. В. Мальчикова [2, 3] на случай скалярных нелинейных дифференциальных СтС с винеровскими и пуассоновскими компонентами. Особое внимание уделим вопросам чувствительности метода. Применение метода проиллюстрируем тестовым примером.

## 2 Уравнения скалярных нелинейных дифференциальных стохастических систем

Пусть скалярный СтП  $Y(t) = Y_t$  определяется скалярным стохастическим дифференциальным уравнением Ито следующего вида [1]:

$$dY_t = a(Y_t, \Theta, t) dt + b(Y_t, \Theta, t) dW_0(\Theta, t) + \int_{R_0^1} c(Y_t, \Theta, t, v) P^0(\Theta, dt, dv), \quad Y(t_0) = Y_0. \quad (1)$$

Здесь  $W_0 = W_0(\Theta, t)$  — скалярный винеровский СтП интенсивности  $\nu_0 = \nu_0(\Theta, t)$ ;  $\Theta$  —  $n^\Theta$ -мерный вектор параметров;  $P^0(\Theta, \mathcal{J}, \mathcal{A})$  представляет собой для любого множества  $\mathcal{A}$  простой пуассоновский СтП, причем

$$P^0(\Theta, \mathcal{J}, \mathcal{A}) = P(\Theta, \mathcal{J}, \mathcal{A}) - \mu_P(\Theta, \mathcal{J}, \mathcal{A}),$$

где  $\mu_P(\Theta, \mathcal{J}, \mathcal{A})$  — его математическое ожидание, равное

$$\mu_P = \mu_P(\Theta, \mathcal{J}, \mathcal{A}) = MP(\Theta, \mathcal{J}, \mathcal{A}) = \int_{\mathcal{J}} \nu_P(\Theta, \tau, \mathcal{A}) d\tau,$$

$\nu_P = \nu_P(\Theta, \mathcal{J}, \mathcal{A})$  — интенсивность соответствующего пуассоновского потока событий,  $\mathcal{J} = (t_1, t_2]$ ; интегрирование по  $v$  распространяется на пространство  $R^1$  с выколотым началом координат;  $a = a(Y_t, \Theta, t)$ ,  $b = b(Y_t, \Theta, t)$  и  $c = c(Y_t, \Theta, t, v)$  — скалярные нелинейные функции отмеченных аргументов.

Предположим, что при фиксированном  $\Theta$  выполнены условия существования и единственности

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sinitsin@dol.ru

СтП, определяемого (1) при соответствующем начальном условии [1, 5]. Кроме того, функции  $a$ ,  $b$  и  $c$  и характеристики СтП  $W_0$  и  $P^0$  будем считать дифференцируемыми по  $\Theta$ .

В дальнейшем для белого шума  $V = \dot{W}$  (понимаемого в строгом смысле и являющегося производной по времени от произвольного СтП с независимыми приращениями  $W(t)$ ) будем использовать уравнение

$$\dot{Y}_t = a(Y_t, \Theta, t) + b(Y_t, \Theta, t) V, \quad V = \dot{W}. \quad (2)$$

Для линейной СтС с параметрическим шумом уравнение (2) записывается в виде:

$$\dot{Y}_t = a(\Theta, t)Y_t + a_0(\Theta, t) + [b_0(\Theta, t) + b_1(\Theta, t)Y_t] V, \quad Y(t_0) = Y_0. \quad (3)$$

### 3 Уравнения для одномерной характеристической функции

Как известно [1], если существует одномерная плотность  $f = f(y; \Theta, t)$ , то уравнение для одномерной х.ф.  $g = g(\lambda; \Theta, t)$  при фиксированном  $\Theta$  имеет вид:

$$\frac{\partial g(\lambda; \Theta, t)}{\partial t} = \int_{-\infty}^{\infty} [i\lambda a(y, \Theta, t) + \chi(b(y, \Theta, t)\lambda; t)] e^{i\lambda y} f(y; \Theta, t) dy \quad (4)$$

при начальном условии

$$g(\lambda; \Theta, t_0) = g_0(\lambda; \Theta). \quad (5)$$

Здесь  $f = f(y; \Theta, t)$  — одномерная плотность вероятности (п.в.), связанная с х.ф.  $g(\lambda; \Theta, t)$  следующим преобразованием Фурье:

$$f_1(y; \Theta, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda y} g(\lambda; \Theta, t) d\lambda; \quad (6)$$

$$\chi(b(y, \Theta, t)\lambda; t) = -\frac{\lambda^2 \nu_0(\Theta, t)}{2} b^2(y, \Theta, t) + \int_{R'_0} \exp\{[i\lambda c(y, \Theta, t, v)] - 1 - i\lambda c(y, \Theta, t, v)\} \nu_P(\Theta, t, dv).$$

Для СтС (2) уравнения для х.ф. имеют вид (4)–(6). При этом предполагается задание функции  $\chi(\mu; t)$  в явном виде.

Так, например, для винеровского белого шума  $V = V_0 = \dot{W}_0$  имеем:

$$\chi(\mu; t) = -\frac{\mu^2 \nu_0(t)}{2}.$$

Если  $V(t) = V_P(t) = \dot{W}_P(t)$  — общий пуассоновский СтП  $W_P$ , то

$$\chi(\mu; t) = [g(\mu) - 1] \nu_C(t),$$

где  $g_C(\mu)$  и  $\nu_C(t)$  — х.ф. и интенсивность потока скачков.

Для белого шума  $V$ , представляющего собой линейную комбинацию независимых винеровского и общего пуассоновского СтП

$$V(t) = \dot{W}(t), \quad W(t) = W_0(t) + \sum_{k=1}^n c_k P_k(t),$$

имеем:

$$\chi(\mu; t) = -\frac{\mu^2}{2} \nu_0(t) \sum_{k=1}^n [g_k(c_k^T \mu) - 1] \nu_k(t).$$

Здесь  $\nu_0(t)$  — интенсивность винеровского СтП  $W_0$ ;  $g_{ck}(\lambda)$  — х.ф. скачков общего пуассоновского СтП  $P_k(t)$ ;  $\nu_k(t)$  — интенсивность потока скачков.

**Замечание 1.** Для стационарных дифференциальных СтС уравнения для х.ф.  $g^* = g^*(\lambda; \Theta)$  и п.в.  $f^*(y; \Theta)$  имеют вид:

$$\int_{-\infty}^{\infty} [i\lambda a^*(y, \Theta) + \chi(b^*(y, \Theta)\lambda)] e^{i\lambda y} f^*(y; \Theta) dy = 0;$$

$$f^*(y; \Theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda y} g^*(\lambda; \Theta) d\lambda.$$

Здесь звездочкой отмечены стационарные значения отмеченных функций.

### 4 Метод интерполяционного аналитического моделирования в скалярных системах

Для определения х.ф.  $g = g(\lambda; \Theta, t)$  для СтС (1) выразим ее через значение в фиксированных точках отсчета, применив для этого какую-нибудь интерполяционную формулу [6]:

$$g(\lambda; \Theta, t) = \sum_{l=-N}^N g_l(\lambda_l; \Theta, t) \alpha_l(\lambda) + R_N. \quad (7)$$



Здесь  $R_N$  — остаточный член;  $\lambda_l$  — некоторые фиксированные значения аргумента  $\lambda$  в точках отсчета;  $\alpha_l(\lambda)$  — известные функции, удовлетворяющие условию

$$\alpha_l(\lambda_r) = \delta_{lr} \quad (l, r = 0, \pm 1, \dots, \pm N),$$

где  $\delta_{lr}$  — символ Кронекера.

Таким образом, задача сводится к вычислению значений х.ф. в точках отсчета х.ф. После определения х.ф.  $g(\lambda; \Theta, t)$  п.в. определяется преобразованием Фурье (6). Подставляя в (6) выражение (7), найдем:

$$f(y; \Theta, t) = \sum_{l=-N}^N g_l(\lambda_l; \Theta, t) \beta_l(y); \quad (8)$$

$$\beta_l = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda y} \alpha_l(\lambda) d\lambda,$$

где функции  $\beta_l(y)$  являются преобразованием Фурье функций  $\alpha_l(\lambda)$  и, следовательно, известны, а значит, искомая п.в. выражается через значения х.ф. в точках отсчета. Полагая в (4)  $\lambda = \lambda_r$  и подставляя в него (8), получаем:

$$\dot{g}_r(\lambda_r; \Theta, t) = -\frac{\lambda_r^2}{2} \nu_0(\Theta, t) \gamma_{0r}(y, \Theta, t) g_r(\lambda_r; \Theta, t) + \gamma_{1r}(\Theta, t) g_r(\lambda_r; \Theta, t) + \sum_{l=-N}^N \varepsilon_{lr}(\Theta, t) g_l(\lambda_r; \Theta, t). \quad (9)$$

В уравнении (9) введены следующие обозначения:

$$\gamma_{0r}(y, \Theta, t) = \int_{-\infty}^{\infty} b^2(y, \Theta, t) \beta_r(y) e^{i\lambda y} dy; \quad (10)$$

$$\gamma_{1r}(y, \Theta, t) = \int_{-\infty}^{\infty} \int_{R_0^q} \{ \exp [i\lambda_y c(y, \Theta, t, v)] - 1 - i\lambda c(y, \Theta, t, v) \} e^{i\lambda y} \nu_P(\Theta, t, dv) dy; \quad (11)$$

$$\varepsilon_{lr}(\Theta, t) = i\lambda_r \int_{-\infty}^{\infty} a(y, \Theta, t) \beta_l(y) e^{i\lambda y} dy. \quad (12)$$

Интегралы (10)–(12) могут быть вычислены заранее, так как подынтегральные функции известны.

Из уравнения (9), вводя обозначения для эффективной интенсивности  $\nu_r^{\circ}(\Theta, t)$  соответствующего гауссовского белого шума, согласно определению

$$\frac{\lambda_r^2}{2} \nu_0(\Theta, t) \gamma_{0r}(\Theta, t) - \gamma_{1r}(\Theta, t) = \frac{\lambda_r^2}{2} \nu_r^{\circ}(\Theta, t),$$

перепишем систему линейных обыкновенных дифференциальных уравнений (9) в виде:

$$\dot{g}_r(\lambda_r; \Theta, t) = -\frac{\lambda_r^2}{2} \nu_r^{\circ}(\Theta, t) g_r(\lambda_r; \Theta, t) + \sum_{l=-N}^N \varepsilon_{lr}(\Theta, t) g_l(\lambda_r; \Theta, t) \quad (r = 0, \pm 1, \dots, \pm N). \quad (13)$$

Так как х. ф. и коэффициенты  $\varepsilon_{lr}$  в общем случае являются комплексными, то при численном решении уравнений (7) следует перейти к действительной форме записи.

Полагая

$$g_r(\lambda_r; \Theta, t) = g_{r1}(\lambda_r; \Theta, t) + i g_{r2}(\lambda_r; \Theta, t) \quad (i^2 = -1);$$

$$\beta_l(y) = \beta_{l1}(y) + i \beta_{l2}(y);$$

$$\beta_{l1}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \alpha_l(\lambda) \cos \lambda y d\lambda;$$

$$\beta_{l2}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \alpha_l(\lambda) \sin \lambda y d\lambda;$$

$$\varepsilon_{lr} = \varepsilon_{lr1}(\Theta, t) + i \varepsilon_{lr2}(\Theta, t);$$

$$\varepsilon_{lr1}(\Theta, t) = \lambda_r \int_{-\infty}^{\infty} [\beta_{l1}(y) \sin \lambda_r y + \beta_{l2}(y) \cos \lambda_r y] a(y, \Theta, t) dy,$$

$$\varepsilon_{lr2}(\Theta, t) = \lambda_r \int_{-\infty}^{\infty} [\beta_{l1}(y) \cos \lambda_r y + \beta_{l2}(y) \sin \lambda_r y] a(y, \Theta, t) dy,$$

после приравнивания в (13) действительных и мнимых частей получаем линейную систему обыкновенных дифференциальных уравнений для значений х.ф. в точках отсчета в действительной форме:

$$\left. \begin{aligned} \dot{g}_{r1}(\lambda_r; \Theta, t) &= -\frac{\nu_r^{\circ}(\Theta, t)}{2} \lambda_r^2 g_{r1}(\lambda_r; \Theta, t) + \\ &+ \sum_{l=-N}^N [\varepsilon_{lr1}(\Theta, t) g_{l1}(\lambda_l; \Theta, t) - \\ &- \varepsilon_{lr2}(\Theta, t) g_{l2}(\lambda_l; \Theta, t)]; \\ \dot{g}_{r2}(\lambda_r; \Theta, t) &= -\frac{\nu_r^{\circ}(\Theta, t)}{2} \lambda_r^2 g_{r2}(\lambda_r; \Theta, t) + \\ &+ \sum_{l=-N}^N [\varepsilon_{lr1}(\Theta, t) g_{l1}(\lambda_l; \Theta, t) + \\ &+ \varepsilon_{lr2}(\Theta, t) g_{l2}(\lambda_l; \Theta, t)]. \end{aligned} \right\} (14)$$

Упростим систему уравнений (14), используя, во-первых, известные свойства х.ф.

$$g(0; t) = 1, \quad g(-\lambda; t) = \overline{g(\lambda, t)}$$

и вытекающие отсюда соотношения

$$g_{01} = 1, \quad g_{02} = 0, \quad g_{-r1} = g_{r1}, \quad g_{-r2} = -g_{r2}, \quad (15)$$

во-вторых, выделим слагаемые, отвечающие условию  $l = 0$ , разбив оставшиеся части на две суммы и изменив в суммах по  $l$  индекс суммирования по  $l$  на противоположный по знаку. При этом нужно будет решать соответствующие уравнения для значений  $g_{r1}(\lambda_r; \Theta, t)$  и  $g_{r2}(\lambda_r; \Theta, t)$  при  $r > 0$ . В результате придем к следующей искомой системе уравнений:

$$\begin{aligned} \dot{g}_{r1}(\lambda_r; \Theta, t) &= G_{r1} = -\frac{\nu_r^3(\Theta, t)}{2} \lambda_r^2 g_r(\lambda_r; \Theta, t) + \\ &+ \varepsilon_{0r1}(\Theta, t) + \sum_{l=1}^N \{[\varepsilon_{lr1}(\Theta, t) + \\ &+ \varepsilon_{-lr1}(\Theta, t)] g_{l1}(\lambda_l; \Theta, t) - \\ &- [\varepsilon_{lr2}(\Theta, t) - \varepsilon_{-lr2}(\Theta, t)] g_{l2}(\lambda_l; \Theta, t)\}; \quad (16) \\ \dot{g}_{r2}(\lambda_r; \Theta, t) &= G_{r2} = \\ &= -\frac{\nu_r^3(\Theta, t)}{2} \lambda_r^2 g_r(\lambda_r; \Theta, t) + \varepsilon_{0r2}(\Theta, t) + \\ &+ \sum_{l=1}^N \{[\varepsilon_{lr2}(\Theta, t) - \varepsilon_{-lr2}(\Theta, t)] g_{l1}(\lambda_l; \Theta, t) + \\ &+ [\varepsilon_{lr1}(\Theta, t) + \varepsilon_{-lr1}(\Theta, t)] g_{l2}(\lambda_l; \Theta, t)\} \\ &\quad (r = 1, \dots, N). \quad (17) \end{aligned}$$

Теперь нетрудно из (8) получить выражения для п.в. Имеем:

$$\begin{aligned} f(y; \Theta, t) &= \sum_{r=-N}^N \{[g_{r1}(\lambda_r; \Theta, t) \beta_{r1}(y) - \\ &- g_{r2}(\lambda_r; \Theta, t) \beta_{r2}(y)] + \\ &+ i [g_{r1}(\lambda_r; \Theta, t) \beta_{r2}(y) - g_{r2}(\lambda_r; \Theta, t) \beta_{r1}(y)]\}. \quad (18) \end{aligned}$$

Преобразуем (18), учитывая, что, во-первых, п.в. является действительной функцией (поэтому мнимая часть при расчетах обращается в нуль), во-вторых, выделим в действительной части слагаемое, в котором  $l = 0$ , разбив оставшуюся сумму на две части, соответствующие положительным и отрицательным значениям индекса суммирования, в-третьих, примем во внимание (15). В результате получим искомое выражение для п.в.:

$$\begin{aligned} f(y; \Theta, t) &= \beta_{01}(y) + \\ &+ \sum_{r=1}^N \{[\beta_{r1}(y) + \beta_{-r1}(y)] g_{r1}(\lambda_r; \Theta, t) - \\ &- [\beta_{r2}(y) - \beta_{-r2}(y)] g_{r2}(\lambda_r; \Theta, t)\}. \quad (19) \end{aligned}$$

Таким образом, получены системы линейных обыкновенных дифференциальных (16), (17) и (19) для вычисления х.ф. и п.в. при соответствующих начальных условиях:

$$g_r(0; \Theta, t_0) = g_{r1}(0; \Theta, t_0) + i g_{r2}(0; \Theta, t_0). \quad (20)$$

В случае гауссовского начального распределения имеем:

$$\begin{aligned} g_r(0; \Theta, t_0) &= \exp \left[ i \lambda_r m(\Theta, t_0) - \frac{\lambda_r^2}{2} D(\Theta, t_0) \right]; \\ g_{r1}(0; \Theta, t_0) &= \exp \left[ -\frac{\lambda_r^2}{2} D(\Theta, t_0) \right] \cos \lambda_r m(t_0); \\ g_{r2}(0; \Theta, t_0) &= \exp \left[ -\frac{\lambda_r^2}{2} D(\Theta, t_0) \right] \sin \lambda_r m(t_0), \end{aligned}$$

где  $m(\Theta, t_0) = m_0$  и  $D(\Theta, t_0) = D_0$  — математическое ожидание и дисперсия  $Y(t_0)$ .

Так как х.ф. является преобразованием Фурье п.в., то теорема Котельникова [4] полностью определяется последовательностью ее значений в дискретном ряде точек. Поэтому если распределение  $Y_t$  сосредоточено на интервале  $2\Delta = 2\Delta_y$ , то  $\alpha_r = \alpha_r(\lambda)$  и  $\beta_l = \beta_l(y)$  можно выбрать согласно [4] в виде:

$$\begin{aligned} \alpha_r &= \alpha_r(\lambda) = fr \sin [\Delta (\lambda - \lambda^*)] \Delta (\lambda - \lambda^*), \\ \lambda^* &= \frac{\pi r}{\Delta}; \quad (21) \end{aligned}$$

$$\beta_r = \beta_r(y) = \begin{cases} \frac{1}{2\Delta} \exp(-i\lambda^*) & \text{при } |y| \leq \Delta; \\ 0 & \text{при } |y| > \Delta. \end{cases} \quad (22)$$

При этом п.в. будет определяться формулой:

$$\begin{aligned} f(y; \Theta, t) &= \\ &= \begin{cases} \frac{1}{\Delta} \sum_{r=1}^N [g_r(\lambda_r; \Theta, t) \cos \lambda^* y + \\ + g_{r2}(\lambda_r; \Theta, t) \sin \lambda^* y] & \text{при } |y| \leq \Delta; \\ 0 & \text{при } |y| > 0. \end{cases} \quad (23) \end{aligned}$$

Таким образом, в основе МИАМ одномерной скалярной дифференциальной СтС (1) лежит следующее утверждение.

**Теорема 1.** Пусть в СтС (1) существует единственный СтП с конечной одномерной п.в. Тогда при интерполяции х.ф. по Котельникову п.в. и х.ф. определяются линейными уравнениями (16) и (17) при начальных условиях (20).

Чувствительность МИАМ к параметрам  $\Theta$  оценивается согласно [7] путем вычисления первых векторных функций влияния (чувствительности):

$$\begin{aligned} \nabla^\Theta g_{r1} &= \nabla^\Theta g_{r1}(\lambda_r; \Theta, t); \\ \nabla^\Theta g_{r2} &= \nabla^\Theta g_{r2}(\lambda_r; \Theta, t), \end{aligned}$$

где  $\nabla^\Theta$  — символ вектора производных по векторному параметру  $\Theta$ . В частности, в силу (16) и (17) соответствующие уравнения для первых функций влияния имеют вид:

$$\dot{\nabla}^\Theta g_{r1} = \nabla^\Theta G_{r1}; \quad \dot{\nabla}^\Theta g_{r2} = \nabla^\Theta G_{r2}$$

при нулевых начальных условиях.

**Замечание 2.** Для стационарных СтС  $g_{r1}^*$  и  $g_{r2}^*$  находятся из линейных уравнений (16) и (17) при нулевых левых частях.

**Замечание 3.** Необходимое число  $N$  членов ряда, обеспечивающее достаточную малость остаточного члена  $R_N$  и чувствительность расчетов к параметрам  $\Theta$ , существенно зависит от выбора интервала  $\Delta$  в (21) и (22) и определяется аналитической природой задачи.

**Замечание 4.** Для СтС (2) теорема имеет аналогичный вид, если известна функция  $\chi(\mu; t)$ .

## 5 О методе интерполяционного моделирования в векторных системах

Следуя [2, 3], воспользуемся принципом эквивалентности векторной СтС скалярной СтС. В задачах практики обычно не требуется определения х.ф. и п.в. всех составляющих вектора  $Y_t$ , а достаточно определить одномерное распределение некоторых из них. Поэтому, следуя С. В. Мальчикову [2, 3], будем искать распределения отдельных составляющих. Для определения одномерных х.ф. и п.в. некоторой  $h$ -й составляющей вектора  $Y_t$  заменим исходную многомерную СтС эквивалентной одномерной СтС, описываемой стохастическим дифференциальным уравнением Ито (или уравнением с  $\theta$ -дифференциалом [1]), выходная переменная  $\tilde{Y}_{ht}$  которого имеет те же х.ф. и п.в., что и  $h$ -я составляющая  $Y_{ht}$ . Далее заменим негауссовский белый шум эквивалентным гауссовским, а затем воспользуемся теоремой 1. В результате придем к следующему утверждению.

**Теорема 2.** Пусть векторная дифференциальная система приводится к эквивалентной системе скалярных СтС Ито. Тогда в основе МИАМ лежат уравнения теоремы 1.

**Замечание 5.** Выходные переменные  $Y_{ht}, \tilde{Y}_{ht}$  в общем случае — различные СтП, поскольку  $\tilde{Y}_{ht}$  является марковским СтП и, следовательно, сам по себе может и не быть марковским СтП. Однако одномерные распределения у них одинаковы.

**Замечание 6.** Для квазилинейных СтС с аддитивным шумом замена многомерной системы одномерной допустима в случае асимптотической устойчивости линейной детерминированной части. Для комплексно-сопряженных корней эквивалентная система будет уже двумерной. Случаи кратных корней требуют специального рассмотрения.

## 6 Тестовый пример

Рассмотрим скалярную дифференциальную СтС с негауссовским (в общем случае) шумом  $V$  следующего вида [4]:

$$\dot{Y}_t = -A \operatorname{sgn} Y_t + V.$$

Согласно (12) имеем следующие выражения для коэффициентов  $\varepsilon_{lr}$ , определяющих значения х.ф. в точках отсчета:

$$\begin{aligned} \varepsilon_{lr} &= -i\lambda_r A \int_{-\Delta}^{\Delta} \operatorname{sgn} y \frac{1}{2\Delta} e^{i(\lambda_r - \lambda_l)y} dy = \\ &= \frac{i\lambda_r A}{2\Delta} \left\{ \int_{-\Delta}^0 \exp[i(\lambda_r - \lambda_l)y] dy - \right. \\ &\quad \left. - \int_0^{\Delta} \exp[i(\lambda_r - \lambda_l)y] dy \right\}. \end{aligned} \quad (24)$$

Учитывая, что

$$\lambda_r = \frac{r\pi}{\Delta}; \quad \lambda_l = \frac{l\pi}{\Delta}; \quad \varepsilon_{ll} = 0,$$

из (24) находим:

$$\begin{aligned} \varepsilon_{lr} &= \frac{A}{\Delta} \frac{\lambda_r}{\lambda_r - \lambda_l} [1 - \cos(\lambda_r - \lambda_l)\Delta] = \\ &= \frac{A}{r} \frac{r}{r-l} [1 - (-1)^{r-l}]. \end{aligned}$$

Таким образом, все  $\varepsilon_{lr}$  действительны, поэтому  $\varepsilon_{lr1} = \varepsilon_{lr}, \varepsilon_{lr2} = 0$ . Следовательно, уравнения (16) и (17) для действительной и мнимой частей х.ф. разделяются и принимают следующий вид:

$$\begin{aligned} \dot{g}_{r1} &= -\frac{\nu_r^2}{2} \left( \frac{r\pi}{\Delta} \right)^2 g_{r1} + \varepsilon_{0r1} + \sum_{l=1}^N g_{l1} (\varepsilon_{lr} + \varepsilon_{-lr}); \\ \dot{g}_{r2} &= -\frac{\nu_r^2}{2} \left( \frac{r\pi}{\Delta} \right)^2 g_{r2} + \varepsilon_{0r2} + \sum_{l=1}^N g_{l2} (\varepsilon_{lr} + \varepsilon_{-lr}). \end{aligned}$$

Здесь аргументы для краткости опущены.

В случае если начальное распределение имеет нулевое математическое ожидание,  $g_{rr} \equiv 0$ . В этом случае, согласно (23), одномерная плотность выражается формулой:

$$f(y; t) = \frac{1}{2\Delta} + \sum_{l=1}^N g_{l1} \cos\left(\frac{l\pi}{\Delta}y\right).$$

В стационарном случае известно точное выражение для стационарной плотности [1]:

$$f(y, \infty) = \frac{A}{\nu_r^2} \exp\left[-\frac{2A}{\Delta}|y|\right].$$

Как показали расчеты, согласно (19) при  $A = 1$ ,  $\nu_r^2 = 1$ ,  $N = 5$  и  $t \rightarrow \infty$  относительная ошибка при  $y = 0$  составляет 20%, при  $y = 1$  — менее 10%, а при  $y = 2$  — менее 1%.

## 7 Заключение

Для скалярных нелинейных дифференциальных гауссовских и негауссовских СтС (1) и (3) при условии существования одномерной п.в. СтП на базе уравнения для одномерной х.ф. разработан МИАМ. Получены уравнения для функций влияния вектора параметров, входящих в (1) и (3). Тестовый пример показывает эффективность метода уже при 3–5 членах разложения.

Результаты допускают обобщение на интегральные и интегродифференциальные СтС, приводимые к дифференциальным [1].

Особого интереса заслуживает обобщение результатов на некоторые классы векторных нелинейных дифференциальных СтС.

## Литература

1. Пугачёв В. С., Синицын И. Н. Теория стохастических систем. — М.: Логос, 2000; 2004. 1000 с.
2. Мальчиков С. В. Приближенный метод определения законов распределения фазовых координат нелинейных автоматических систем // Автоматика и телемеханика, 1970. № 5. С. 43–50.
3. Казаков И. Е., Мальчиков С. В. Анализ стохастических систем в пространстве состояний. — М.: Наука, 1983. 348 с.
4. Котельников В. А. Теория потенциальной помехоустойчивости. — М.: Госэнергоиздат, 1956. 412 с.
5. Справочник по теории вероятностей и математической статистике / Под ред. В. С. Королюка, Н. И. Портенко, А. В. Скорохода, А. Ф. Турбина. — М.: Наука, 1985. 640 с.
6. Корн Г., Корн Т. Справочник по математике (для научных работников и инженеров) / Пер с англ. — М.: Наука, 1974. 832 с. (Korn G., Korn T. Mathematical handbook for scientists and engineers. — New York — San Francisco — Toronto — London — Sydney: McGraw Hill Book Co., 1968. 1147 p.)
7. Евланов Л. Г., Константинов В. М. Системы со случайными параметрами. — М.: Наука, 1976. 585 с.

Поступила в редакцию 29.08.17

---



---

## METHOD OF INTERPOLATIONAL ANALYTICAL MODELING OF PROCESSES IN STOCHASTIC SYSTEMS

I. N. Sinitsyn

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** Among known methods of stochastic processes (StP) analytical modeling in differential and integrodifferential stochastic systems (StS) based on the direct numerical solution of equation for one-dimensional characteristic function (c.f.), it is necessary to distinguish interpolational S. V. Mal'chikov method. In this case, for c.f. interpolation, the Kotelnikov theorem was implemented. The paper contains the treatment of interpolational methods of StP analytical modeling for two classes of nonlinear non-Gaussian StS. Special attention is paid to sensitivity analysis. Test example for discontinuous nonlinearity confirms the method efficiency. Some generalizations are mentioned.

**Keywords:** one dimensional characteristic functions (c.f.); one-dimensional probability density (p.d.); stochastic processes (StP); stochastic system (StS)

**DOI:** 10.14357/19922264180107



## References

1. Pugachev, V.S., and I.N. Sinitsyn. 2000, 2004. *Teoriya stokhasticheskikh sistem* [Stochastic systems: Theory and applications]. Moscow: Logos. 1000 p.
2. Mal'chikov, S.V. 1970. Priblizhennyi metod opredeleniya zakonov raspredeleniya fazovykh koordinat nelineynykh avtomaticheskikh sistem [Approximate method for phase coordinates analysis in nonlinear control systems]. *Avtomatika i telemekhanika* [Automat. Rem. Contr.] 5:43–50.
3. Kazakov, I. E., and S. V. Mal'chikov. 1983. *Analiz stokhasticheskikh sistem v prostranstve sostoyaniy* [Stochastic systems analysis in state space]. Moscow: Nauka. 348 p.
4. Kotel'nikov, V.A. 1956. *Teoriya potentsial'noy pomekhostoychivosti* [Theory of potential noiseproof]. Moscow: Gosenergoizdat. 412 p.
5. Korolyuk, V.S., N.I. Portenko, A.V. Skorokhod, and A. F. Turbin, eds. 1985. *Spravochnik po teorii veroyatnostey v matematicheskoy statistike* [Handbook on probability theory and mathematical statistics]. Moscow: Nauka. 640 p.
6. Korn, G., and T. Korn. 1968. *Mathematical handbook for scientists and engineers*. New York – San Francisco – Toronto – London – Sydney: McGraw Hill Book Co. 1147 p.
7. Evlanov, L. G., and V.M. Konstantinov. 1976. *Sistemy so sluchaynymi parametrami* [Systems with random parameters]. Moscow: Nauka. 585 p.

*Received August 29, 2017*

## Contributor

**Sinitsyn Igor N.** (b. 1940) — Doctor of Science in technology, professor, Honored Scientist of RF, principal scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sinitsin@dol.ru

# О НЕКОТОРЫХ ВОЗМОЖНОСТЯХ УПРАВЛЕНИЯ РЕСУРСАМИ ПРИ ОРГАНИЗАЦИИ ПРОАКТИВНОГО ПРОТИВОДЕЙСТВИЯ КОМПЬЮТЕРНЫМ АТАКАМ\*

А. А. Грушо<sup>1</sup>, М. И. Забейло<sup>2</sup>, А. А. Зацаринный<sup>3</sup>, Е. Е. Тимонина<sup>4</sup>

**Аннотация:** Обсуждаются возможности рационального противодействия компьютерным атакам, описываемым как последовательности событий. В основе подхода — математическая техника обучения на прецедентах, использующая уточнение сходства объектов как бинарной алгебраической операции. Анализируются сходства цепочек событий. Порождаемые в процессе обучения классы сходства (толерантности) используются для идентификации атак рассматриваемого типа на ранних стадиях их развертывания. Предлагается технология рационального управления ресурсами при организации противодействия подобным вредоносным воздействиям.

**Ключевые слова:** информационная безопасность; анализ данных; сходство как алгебраическая операция; сходство цепочек; управление ресурсами

**DOI:** 10.14357/19922264180108

## 1 Введение

Весьма полезно организовать противодействие атакам на информационные технологии в проактивном режиме, позволяющем, в частности, идентифицировать предстоящие (прогнозируемые) вредоносные воздействия на ранних стадиях их развертывания и, как следствие, иметь возможность препятствовать их развертыванию в полном объеме [1].

Возможности проследить по первым наблюдаемым стадиям развития (потенциальной) атаки (согласуемые с уже накопленным опытом) варианты ее продолжения позволяют выделить те группы признаков, которые должны будут сопутствовать последовательно проявляющимся в ходе развертывания текущей атаки признакам. Это позволяет организовать противодействие атаке, фокусируя усилия в первую очередь на *сопутствующих* признаках, препятствуя таким образом формированию тех комбинаций признаков (КП), которые и вызывают появление соответствующего вредоносного эффекта (ВЭ).

Фокусировка на *сопутствующих уже наблюдаемым* факторам влияния (*признакам*), чтобы не допустить формирования уже известных КП, которые

ведут к конкретным ВЭ, дает основания охарактеризовать управление ресурсами в рассматриваемом подходе как рациональное действие. Тогда можно говорить о предлагаемом подходе как о составной части инструментария представления знаний средствами так называемых квазиаксиоматических теорий (см., например, [2, 3]) и построения теорий для открытых, пополняемых новыми данными предметных областей.

## 2 Цепочки состояний: основные понятия и определения

Обратимся к простейшему теоретико-множественному варианту описания процедурных конструкций предлагаемого типа. Будем рассматривать *событие* как последовательность *состояний* объекта защиты, где каждое *состояние* представлено некоторым множеством *признаков* — фактов из некоторого «базового» множества, наблюдаемых в текущем состоянии объекта защиты.

Пусть  $U = \{a_1, a_2, \dots, a_n\}$  — исходное множество признаков, которые будут использоваться для описания событий.

\* Работа выполнена при поддержке РФФИ (проект 15-29-07981).

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, gusho@yandex.ru

<sup>2</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, m.zabehailo@yandex.ru

<sup>3</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, alex250451@mail.ru

<sup>4</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

**Определение 1.**

1. *Состоянием*  $\mathbf{St}$  объекта защиты  $\mathbf{O}$  будем называть каждое непустое подмножество исходного множества признаков  $\mathbf{U} : \mathbf{St} \subseteq \mathbf{U} \neq \emptyset$ .
2. *Событием* (описанием события)  $\mathbf{E} = \langle \mathbf{St}_1, \mathbf{St}_2, \dots, \mathbf{St}_k \rangle$ , наблюдаемого в поведении  $\mathbf{O}$ , будем называть всякую конечную последовательность  $\langle \mathbf{St}_1, \mathbf{St}_2, \dots, \mathbf{St}_k \rangle$  состояний  $\mathbf{St}_i \in \mathbf{St} = (2^{\mathbf{U}} \setminus \{\emptyset\})\mathbf{St}$  для каждого  $i \in \{1, 2, \dots, k\}$ .

Пусть  $\mathbf{E}|_l = \langle \mathbf{St}_{l_1}, \mathbf{St}_{l_2}, \dots, \mathbf{St}_{l_m} \rangle$  есть подпоследовательность представляющей произвольным образом выбранное событие  $\mathbf{E}$  последовательности  $\langle \mathbf{St}_1, \mathbf{St}_2, \dots, \mathbf{St}_k \rangle$ , такая что в  $\mathbf{E}|_l$  сохраняется формирующий  $\mathbf{E}$  порядок вхождения соответствующих состояний  $\mathbf{St}_i$ . Выберем из каждого множества признаков  $\mathbf{St}_{l_i}$  ровно по одному элементу  $a_{l_i}$  и рассмотрим цепочку (последовательность) вида  $\vec{s} \langle a_{l_1}, a_{l_2}, \dots, a_{l_m} \rangle$ , элементы которой упорядочены в соответствии с уже имеющимся порядком состояний в последовательности  $\mathbf{E}$ . Формируемое таким путем отношение  $\prec_1$  частичного порядка на элементах цепочек вида  $\vec{s}$  будем интерпретировать как  $a_{l_i} \prec_1 a_{l_j}$ . Это означает, что в цепочку  $\vec{s}$  признак  $a_{l_i}$  входит раньше, чем признак  $a_{l_j}$ .

Пусть имеются два события:  $\mathbf{E}_1$ , задающее порядок  $\prec_1^1$  входящих в него состояний, и  $\mathbf{E}_2$  с соответствующим порядком  $\prec_1^2$ . Пусть также  $S(\mathbf{E}_1)$  и  $S(\mathbf{E}_2)$  — множества цепочек, представляющие соответственно события  $\mathbf{E}_1$  и  $\mathbf{E}_2$ , а  $\vec{s}_1$  и  $\vec{s}_2$  — цепочки из множеств  $S(\mathbf{E}_1)$  и  $S(\mathbf{E}_2)$  соответственно. Наложением цепочек  $\vec{s}_1$  и  $\vec{s}_2$  будем называть множество  $s_{12}$  максимальных по вложению общих подцепочек для  $\vec{s}_1$  и  $\vec{s}_2$ , каждый элемент которого сохраняет при этом имеющийся в каждой из сравниваемых исходных цепочек частичный порядок  $\prec_1^1$  и  $\prec_1^2$ .

Для любой цепочки  $\vec{s}_i$  результатом автоналожения, т. е. наложения  $\vec{s}_i$  на  $\vec{s}_i$ , будет множество  $s_{ii}$  всех максимальных по вложению в исходную цепочку  $\vec{s}_i$  подцепочек, сохраняющих уже имеющийся в  $\vec{s}_i$  порядок  $\prec_1^i$ . Пусть  $S(\mathbf{E})$  — множество цепочек, порождаемых событием  $\mathbf{E}$ . Сформируем для каждой цепочки  $\vec{s}_i \in S(\mathbf{E})$  множество  $s(\mathbf{E})$  всех автоналожений  $s_{ii}$  входящих в  $S(\mathbf{E})$  цепочек  $\vec{s}_i$ .

Нетрудно убедиться, что любая пара несовпадающих цепочек порождает несовпадающие автоналожения. Если эти цепочки имеют разные длины, то более длинная цепочка не может содержаться в автоналожении более короткой цепочки. При одинаковой длине цепочек каждая из них не может содержаться в автоналожении другой (они ведь не совпадают). Таким образом, каждое событие  $\mathbf{E}$  может быть однозначным образом представлено мно-

жеством  $s(\mathbf{E})$  автоналожений всех входящих в него цепочек вида  $\vec{s}_i$ .

Ситуацию, когда результат  $s_{12}$  наложения двух цепочек  $\vec{s}_1$  и  $\vec{s}_2$  не является одноэлементным множеством подцепочек, демонстрирует следующий пример.

**Пример.**

1. Пусть  $\mathbf{E}_1 = \langle \mathbf{St}_1, \mathbf{St}_2, \mathbf{St}_3, \mathbf{St}_4 \rangle = \langle \{a\}, \{b\}, \{a\}, \{c\} \rangle$ , а  $\mathbf{E}_2 = \langle \{a\}, \{b\}, \{c\} \rangle$ . В данном случае  $S(\mathbf{E}_1) = \{ \langle \{a\}, \{b\}, \{a\}, \{c\} \rangle \} = \{ \vec{s}_1 \}$ ,  $S(\mathbf{E}_2) = \{ \langle \{a\}, \{b\}, \{c\} \rangle \} = \{ \vec{s}_2 \}$ . Рассмотрим  $s_{12}$  — результат наложения цепочек  $\vec{s}_1$  и  $\vec{s}_2$ . Легко проверить, что

$$s_{12} = \{ \langle \{a^1\}, \{b\}, \{c\} \rangle, \langle \{a^2\}, \{c\} \rangle \},$$

где  $a^1$  и  $a^2$  соответственно первое и второе (в смысле  $\prec_1$ ) вхождение  $a$  в цепочку  $\vec{s}_1$ . Однако при этом цепочки, например,  $\langle \{a^1\}, \{c\} \rangle$  и  $\langle \{b\}, \{c\} \rangle$  не входят в  $s_{12}$ , так как могут быть расширены до вошедшей в  $s_{12}$  цепочки  $\langle \{a^1\}, \{b\}, \{c\} \rangle$ . Таким образом, наложение  $s_{12}$  содержит сразу две максимальные по вложению в  $\vec{s}_1$  и  $\vec{s}_2$  общие для них подцепочки.

2. Пусть  $\mathbf{E} = \langle \{a\}, \{b\}, \{a\}, \{b\} \rangle$ . Рассмотрим  $s_{11}$  — результат наложения цепочки  $\vec{s}_1$  на саму себя. Легко проверить, что

$$s_{11} = \{ \langle \{a\}, \{b\}, \{a\}, \{b\} \rangle, \langle \{a^{12}\}, \{b^{12}\} \rangle \},$$

где  $a^{12}$  порождено наложением  $a^1$  и  $a^2$  соответственно первого и второго (в смысле отношения  $\prec_1$ ) вхождения  $a$  в цепочку  $\vec{s}_1$ ,  $b^{12}$  порождено наложением  $b^1$  и  $b^2$  — первого и второго вхождения  $b$  в цепочку  $\vec{s}_1$ , порожденную событием  $\mathbf{E}$ .

Сходство пар событий формализуем как результат наложения представляющих эти события множеств (автоналожений) цепочек, определив для этого бинарную алгебраическую операцию на множествах цепочек и порождаемое ею отношение наложимости.

**Определение 2.** Пусть на общем множестве признаков  $\mathbf{U}$  заданы два события  $\mathbf{E}_1$  и  $\mathbf{E}_2$  и порождаемая ими пара множеств цепочек  $S_1 = S(\mathbf{E}_1)$  и  $S_2 = S(\mathbf{E}_2)$  соответственно. Результатом наложения множеств  $S_1$  и  $S_2$  будем считать множество цепочек  $S_{12}$ , образованное наложением каждого элемента из  $S_1$  на каждый элемент из  $S_2$ .

Пусть заданы множество признаков  $\mathbf{U}$  и множество  $\Omega = \{ \mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_r \}$  сформированных на  $\mathbf{U}$  событий  $\mathbf{E}_i$ ,  $i \in \{1, 2, \dots, r\}$ . Множество  $\text{Dom}(\mathbf{U}, \Omega)$  определим следующим образом.

**Определение 3.**

1. Для каждого события  $\mathbf{E}_i \in \Omega$  автоналожение  $s_{jj}^i$  каждой цепочки  $\vec{s}_j^i$  из множества цепочек  $S_i = S(\mathbf{E}_i)$  принадлежит  $\mathbf{Dom}(\mathbf{U}, \Omega)$ , т. е.

$$S_i = S(\mathbf{E}_i) \subseteq \mathbf{Dom}(\mathbf{U}, \Omega).$$

2. Для каждой пары  $S_i$  и  $S_j$  множеств из  $\mathbf{Dom}(\mathbf{U}, \Omega)$  их наложение (т. е. результат наложения каждой цепочки из  $S_i$  на каждую цепочку из  $S_j$ ) также принадлежит  $\mathbf{Dom}(\mathbf{U}, \Omega)$ .
3. Других элементов в множестве  $\mathbf{Dom}(\mathbf{U}, \Omega)$  нет.

**Определение 4.** Для каждой пары  $S_i$  и  $S_j$  множеств цепочек из  $\mathbf{Dom}(\mathbf{U}, \Omega)$  определена бинарная операция  $S_{ij} = S_i \otimes S_j$  поэлементного наложения каждой цепочки из  $S_i$  на каждую цепочку из  $S_j$ .

**Утверждение 1.** Для любых  $S_i, S_j, S_k \subseteq \mathbf{Dom}(\mathbf{U}, \Omega)$  операция  $\otimes$  обладает следующими свойствами:

- (а) *идемпотентность:*

$$S_i = S_i \otimes S_i,$$

- (б) *коммутативность:*

$$S_i \otimes S_j = S_j \otimes S_i$$

- (в) *ассоциативность:*

$$S_i \otimes (S_j \otimes S_k) = (S_i \otimes S_j) \otimes S_k. \quad (1)$$

Таким образом, алгебра  $\wp = \langle \mathbf{Dom}(\mathbf{U}, \Omega), \otimes \rangle$  есть нижняя полурешетка.

**Доказательство.** Идемпотентность является очевидным следствием представления в множестве  $\mathbf{Dom}(\mathbf{U}, \Omega)$  цепочек из  $\mathbf{E}_i$  их автоналожениями.

Коммутативность обеспечивается выделением в каждом двух накладываемых цепочках такого общего подмножества элементов, на котором обеспечивается одновременное сохранение соответствующих обеим цепочкам частичных порядков  $\prec_1^1$  и  $\prec_1^2$ .

**Ассоциативность.** Предположим противное: условие ассоциативности не выполняется. Тогда должны найтись три цепочки  $\vec{s}_i, \vec{s}_j$  и  $\vec{s}_k$  и некоторая максимальная по вложению в каждую из них подцепочка  $\vec{s}_0 = \langle a, b, \dots, c, d \rangle$  такие, что в одной из частей равенства (1)  $\vec{s}_0$  содержится, например, в левой части (1), а в правой нет. При этом

$$\begin{aligned} &(a \prec_1^i b), (a \prec_1^j b) \text{ и } (a \prec_1^k b); \\ &\dots \\ &(c \prec_1^i d), (c \prec_1^j d) \text{ и } (c \prec_1^k d). \end{aligned}$$

По определению процедуры наложения эти же упорядочения сохраняются на элементах множества

$\{a, b, \dots, c, d\}$  также и в правой части (1). Принимая во внимание ассоциативность операции  $\cap$  и наличие множества  $\{a, b, \dots, c, d\}$  одновременно в множествах признаков, образовавших соответственно цепочки  $\vec{s}_i, \vec{s}_j$  и  $\vec{s}_k$ , получим, что  $\{a, b, \dots, c, d\}$  войдет и в правую часть, т. е. в (1). Заметим, что в силу ассоциативности операции  $\cap$  и сохранения на  $\{a, b, \dots, c, d\}$  всех трех упорядочений  $\prec_1^i, \prec_1^j$  и  $\prec_1^k$  множество  $\{a, b, \dots, c, d\}$  нерасширяемо ни одним дополнительным элементом  $z$  из  $\mathbf{U}$ . Следовательно, оно представляет собою максимальное по вложению множество признаков, входящих и в правую часть (1) как общая часть цепочек  $\vec{s}_i, \vec{s}_j$  и  $\vec{s}_k$ , что противоречит предположению. Утверждение доказано.

Далее с помощью операции  $\otimes$  может быть определено отношение  $R^\otimes$  наложимости множеств цепочек из  $\mathbf{Dom}(\mathbf{U}, \Omega)$ .

**Определение 5.** Для каждой пары  $O_{i_1}$  и  $O_{i_2}$  элементов множества  $\mathbf{Dom}(\mathbf{U}, \Omega)$  определим отношение  $R^\otimes$  их взаимной наложимости следующим образом:  $O_{i_1} R^\otimes O_{i_2}$  тогда и только тогда, когда результат их наложения  $O_{i_1} \otimes O_{i_2} \neq \emptyset$ .

Принимая во внимание утверждение 1, можно заключить, что отношение наложимости  $R^\otimes$  представляет собой отношение типа *сходства* (толерантности — см., например, [4, 5]). Таким образом, на элементах множества  $\mathbf{Dom}(\mathbf{U}, \Omega)$  могут быть сформированы классы сходства вида

$$\mathbf{T}(O_{i_1}) = \{O_{i_2} | O_{i_1} R^\otimes O_{i_2}\}.$$

Далее, для фиксированного  $V = V_0 = A \otimes B$ ,  $A, B \in \mathbf{Dom}(\mathbf{U}, \Omega)$ , можно найти соответствующие подклассы формируемых классов сходства:

$$\mathbf{EQ}_{V_0} = \{\langle O_{i_1}, O_{i_2} \rangle | O_{i_1} \otimes O_{i_2} \otimes V_0 = V_0\}.$$

**Утверждение 2.** Пусть заданы множество  $\mathbf{Dom}(\mathbf{U}, \Omega)$ , операция  $\otimes$ , некоторый объект  $O_r \in \mathbf{Dom}(\mathbf{U}, \Omega)$ , а также некоторый подобъект  $V$ , характеризующий сходство объекта  $O_r$  по крайней мере с одним из элементов класса сходства  $\mathbf{T}_{\mathbf{Dom}(\mathbf{U}, \Omega)}(O_r)$ . Множество всех таких элементов  $O_s \in \mathbf{T}_{\mathbf{Dom}(\mathbf{U}, \Omega)}(O_r)$ , что  $O_r \otimes O_s \otimes V = V$ , представляет собой класс эквивалентности  $\mathbf{EQ}_{\mathbf{Dom}(\mathbf{U}, \Omega), v}(O_r)$ .

**Доказательство.** Рефлексивность и симметричность отношения  $E_v^\otimes$ , порождаемого из уже определенного отношения  $R^\otimes$  фиксацией непустого результата  $V$  применения операции сходства  $\otimes$ , определяется соответствующими свойствами отношения  $R^\otimes$ . Транзитивность следует из того, что, будучи общей частью одновременно пар  $\langle O_r, O_s \rangle$



и  $\langle O_s, O_t \rangle$ , заданный подобъект  $V$  оказывается также общей частью объектов  $O_r$  и  $O_t$ . Утверждение 2 доказано.

Таким образом, имея исходно заданное множество событий, можно корректным образом уточнить понимание сходства таких событий и получить возможность выделять классы эквивалентности множеств используемых цепочек. Другими словами, можно сформировать процедурную схему обучения на прецедентах (описаниях событий) и классификации вновь наблюдаемых событий [6], в которой диагностика вновь наблюдаемых событий будет опираться на возможность их корректного отнесения к уже построенным ранее на имеющейся обучающей выборке (и дополняемым при ее расширениях) классам эквивалентности описаний прецедентов.

При построении программной реализации предлагаемой процедурной схемы существенными оказываются оценки сложности вычислений. Для демонстрации особенностей рассматриваемой ситуации обратимся к двум характеризующим ее аспектам, а именно: к проблеме скорости роста размеров автоналожений с ростом длины исходных цепочек состояний, а также к оценке емкости результирующих множеств наложений, характеризующих сходства анализируемых событий.

При обсуждении скорости роста размеров автоналожений полезно учитывать также и содержательные особенности рассматриваемой задачи — анализ «треков» развертывания компьютерных атак. Заметим, что если оставить в стороне легко «узнаваемые» массовые атаки со значительными объемами повторяющихся «типовых» действий, то, сфокусировавшись на анализе небольших выборок однородных по вызываемым вредоносным эффектам целенаправленных воздействий, можно ожидать не очень «длинных» последовательностей состояний, которые характеризуются «не очень часто» повторяющимися вхождениями отдельных признаков. Таким образом, остается надежда на «погружаемость» вычислений как соответствующих автоналожений, так и сходств анализируемых цепочек в разумные ресурсы (необходимые объемы памяти и быстродействия) современных вычислительных устройств.

В части оценки емкости характерных классов объектов, порождаемых при формировании сходств описаний событий, в частности числа классов эквивалентности, порождаемых отношением наложимости  $R^\otimes$  и его сужением  $E_v^\otimes$ , ситуация оказывается, к сожалению, существенно хуже. При этом ситуация соответствует положению дел, известному и при работе с другими типами данных (см.,

например, [7]). Здесь можно предъявить некоторые переборные задачи, для которых доказуема принадлежность к классу #PC так называемых *перечислительно полных* задач (см., например, [8–11]).

Рассмотрим частный случай задачи о сходстве событий, представляемых последовательностями состояний, в котором для каждого входящего в исходную обучающую выборку  $\Omega = \{E_1, E_2, \dots, E_r\}$  события  $E_i = \langle St_1^i, St_2^i, \dots, St_k^i \rangle$  каждое из характеризующих его состояний  $St_r^i$  может иметь общие признаки из исходного множества  $U$  лишь с признаками из соответствующего (т. е. имеющего тот же номер  $r$ ) состояния  $St_r^j$  любого другого события  $E_j$  из  $\Omega$ .

**Утверждение 3.** *Задача о числе классов эквивалентности  $E_{Q_{V_0}}$ , формируемых в множестве  $Dom(U, \Omega)$  с помощью операции  $\otimes$  и порождаемых ею отношений  $R^\otimes$  и  $E_v^\otimes$ , принадлежит классу #PC перечислительно полных задач.*

**Доказательство.** Предложенная частная версия задачи о сходстве событий, представленных последовательностями состояний, может быть сведена к задаче о сходстве множеств признаков (см., например, [7]). Действительно, условие непустоты пересечения подмножеств признаков только в соответствующих, имеющих один и тот же номер  $r$ , состояниях  $St_r^i, \dots, St_r^j$  позволяет для каждого события  $E_i$  собрать все формирующие его признаки по всем состояниям в единое множество

$$U(E_i) = St_1^i \cup St_2^i \cup \dots \cup St_k^i,$$

не опасаясь утраты порядков  $\prec_1^i$  их вхождения в описание  $E_i$ . Такой общий для всех событий из  $\Omega$  порядок  $\prec_1$  может быть автоматически восстановлен с учетом вхождений для каждого  $s \in \{1, 2, \dots, k\}$  всех попарно непересекающихся объединений вида  $St_s = \cup_{i=1}^r St_s^i$  в исходное множество признаков  $U$ , т. е. порядок на признаках в данном случае оказывается несущественным, и анализу (см. задачу о числе классов эквивалентности в [7]) подлежат лишь общие части соответствующих множеств признаков  $U(E_i)$ , представляющих в данном случае события из обучающей выборки  $\Omega$ .

Для доказательства принадлежности задачи о числе классов эквивалентности классу #PC продемонстрируем полиномиальную сводимость к ней известной (см., например, [9]) переборной задачи о выполнимости монотонной булевой функции  $\Phi$ . Задача состоит в следующем. Для функции  $\Phi$ , представленной в виде 2-КНФ

$$\Phi(x_1, x_2, \dots, x_n) = \Phi_1 \& \Phi_2 \& \dots \& \Phi_m = \\ = (x_{11} \vee x_{12}) \& (x_{21} \vee x_{22}) \& \dots \& (x_{m1} \vee x_{m2}), \quad (2)$$

найти число наборов значений переменных  $x_1, x_2, \dots, x_n$ , выполняющих функцию  $\Phi$ .

Схема сведения одной задачи к другой:

- по заданной функции  $\Phi$  полиномиально быстро строится пара множеств  $U_\Phi, \Omega_\Phi$  такая, что классы эквивалентности объектов из  $\Omega_\Phi$  над множеством  $U_\Phi$  «перечисляют» все нули функции  $\Phi$ ;
- размер множества единиц функции  $\Phi$  (т. е. множества наборов, выполняющих эту функцию) есть размер дополнения множества нулей функции  $\Phi$  до множества всех возможных наборов значений переменных  $x_1, x_2, \dots, x_n$ .

Итак, пусть  $\Phi$  имеет вид (2). Не теряя общности, будем считать, что в записи функции  $\Phi$  все конъюнкты  $\Phi_i$  различны. В противном случае функция  $\Phi$  может быть приведена к нужному виду полиномиально сложной процедурой — сортировкой образующих ее конъюнктов.

Примитивным нулем номер  $i$  функции

$$\Phi(x_1, x_2, \dots, x_n) = \Phi_1 \& \Phi_2 \& \dots \& \Phi_i \& \dots \& \Phi_m$$

будем называть такой булевский вектор  $\sigma_i$ , что  $x_{i1} = 0, x_{i2} = 0$ , а все остальные переменные равны 1.

Первым сопряженным с примитивным нулем  $\sigma_i$  будем называть всякий такой булевский вектор  $\sigma_i^*$ , что ровно одна из  $n-2$  переменных множества  $X_i = \{x_1, x_2, \dots, x_n\} \setminus \{x_{i1}, x_{i2}\}$  принимает значение 0.

Положим  $U_\Phi = \{a^1, a^2, \dots, a^n\}$ , а множество  $\Omega_{\Phi,i}$  для каждого  $\sigma_i$  определим как

$$\Omega_{\Phi,i} = \{U_\Phi, A_0^i, A_1^i, A_2^i, \dots, A_n^i\},$$

где

$$A_0^i = U_\Phi \setminus \{a^{i1}, a^{i2}\};$$

$$A_1^i = U_\Phi \setminus \{a^{i1}, a^{i2}\} \setminus \{a^1\};$$

$$A_2^i = U_\Phi \setminus \{a^{i1}, a^{i2}\} \setminus \{a^2\};$$

...

$$A_n^i = U_\Phi \setminus \{a^{i1}, a^{i2}\} \setminus \{a^n\}.$$

При этом, как нетрудно убедиться, для каждого  $i$  имеет место равенство:

$$A_0^i = A_{i1}^i = A_{i2}^i = U_\Phi \setminus \{a^{i1}, a^{i2}\}.$$

В рассматриваемой ситуации всякий нуль функции  $\Phi$ , который сопряжен с заданным примитивным нулем  $\sigma_i$ , может быть представлен соответствующим сходством, которое будем называть его *исчерпывающим*, и порождаемым им классом эквивалентности по отношению эквивалентности  $E^\cap$ .

$E^\cap$  порождено операцией  $\cap$  пересечения множеств из  $\Omega_{\Phi,i}$ . Действительно, с точностью до совпадения  $A_0^i, A_{i1}^i$  и  $A_{i2}^i$  любое двух- и более элементное подмножество множества  $\Omega_{\Phi,i}$  порождает уникальное множество общих для всех входящих в  $\Omega_{\Phi,i}$  элементов из  $U_\Phi$ . Это обусловлено «диагональной» по вхождению в него элементов из  $U_\Phi$  структурой множества  $\Omega_{\Phi,i}$ .

При этом имеет место и обратное соответствие. Каждое исчерпывающее сходство элементов множества  $\Omega_{\Phi,i}$  может быть сопоставлено одному и только одному из нулей исходной булевой функции  $\Phi$ . Однозначность задействованного сопоставления есть следствие используемого здесь способа «кодировки» единиц каждого восстанавливаемого булевского вектора  $\sigma$ .

Пусть далее

$$\Omega_\Phi = \bigcup_{i=1}^m \Omega_{\Phi,i}. \quad (3)$$

Таким образом, нуль каждой конъюнкты  $\Phi_i$  из 2-КНФ-представления функции  $\Phi$  «кодируется»  $n+1$  множеством  $A^i = \{A_0^i, A_1^i, A_2^i, \dots, A_n^i\}$  по следующей схеме:

- $A_0^i$  представляет ситуацию, когда  $x_{i1} = x_{i2} = 0$ , а все остальные переменные функции  $\Phi$  принимают значение 1;
- каждое из множеств  $A_j^i$ , где  $j \in \{1, 2, \dots, n\}$ , характеризует ситуацию, когда одна отличная от  $x_{i1}$  и  $x_{i2}$  переменная  $x_j$  (и только она) обращается в 0 одновременно с  $x_{i1} = x_{i2} = 0$ ;
- по значениям индекса  $i$  подобных коллекций множеств вида  $A_j^i, j \in \{1, 2, \dots, n\}$ , будет ровно  $m$  по числу конъюнктов  $\Phi_i$  в 2-КНФ исходной заданной функции  $\Phi$ ;
- множество всех нулей функции  $\Phi$ , являющихся одновременно нулями конъюнкты  $\Phi_i$ , будет перечисляться множеством всех исчерпывающих пересечений (сходств) объектов из множества  $\Omega_{\Phi,i}$ . Можно показать, например индукцией по числу пересекаемых элементов из  $\Omega_{\Phi,i}$ , что с точностью до равенства  $A_0^i = A_{i1}^i = A_{i2}^i$  каждое пересечение объектов из множества  $\Omega_{\Phi,i}$  будет исчерпывающим.

В ситуации, когда  $\Phi$  обращается в нуль, по крайней мере один из ее конъюнктов  $\Phi_i$  непременно также обращается в нуль. Нетрудно убедиться, что каждый из нулей функции  $\Phi$  может быть рассмотрен как суперпозиция примитивных и сопряженных с ними нулей, т. е. в ситуации, когда одновременно реализуются соответствующие примитивные нули функции  $\Phi$ .

Разумеется, в объединенном множестве  $\Omega_{\Phi}$  будут встречаться одинаковые элементы. Такие повторы порождаются для сопряженных с примитивными нулями, каждый из которых формируется суперпозициями примитивных нулей. Очевидно, что наличие таких «повторов» не может расширить множество исчерпывающих сходств (пересечений) элементов сводного множества  $\Omega_{\Phi}$  в (3).

Итак, каждое исчерпывающее пересечение (сходство) объектов из множества  $\Omega_{\Phi}$  будет представлять какой-либо из нулей функции  $\Phi$ . И наоборот: каждый из нулей функции  $\Phi$  как соответствующая суперпозиция примитивных и сопряженных с ними нулей функции  $\Phi$  может быть представлен (исчерпывающим) сходством объектов из множества  $\Omega_{\Phi}$ . Таким образом, множество нулей функции  $\Phi$  и множество исчерпывающих сходств объектов из множества  $\Omega_{\Phi}$  могут быть взаимно однозначным образом сопоставлены друг другу, т.е. по числу исчерпывающих сходств  $\Omega_{\Phi}$  можно полиномиально быстро определить число наборов, выполняющих функцию  $\Phi$ .

Осталось убедиться, что множества  $U_{\Phi}$  и  $\Omega_{\Phi}$  полиномиально быстро порождаются по заданной функции  $\Phi$ . Действительно, в  $U_{\Phi}$  имеется  $n = |\{x_1, x_2, \dots, x_n\}|$  элементов, а в  $\Omega_{\Phi}$  содержится не более чем  $(n+1)m$  элементов, каждый из которых содержит не более чем  $n - 2$  элементов из  $U_{\Phi}$ .

### 3 Свертки цепочек состояний

Обратимся к анализу эффектов совместного вхождения признаков из исходного  $U$  в описания событий обучающей выборки  $\Omega = \{E_1, E_2, \dots, E_r\}$ .

**Определение 6.** Сверткой  $U(E)$  события  $E = \langle St_1, St_2, \dots, St_k \rangle$  будем называть объединение всех признаков  $U(E) = St_1 \cup St_2 \cup \dots \cup St_k$ , встречающихся в описаниях всех его состояний.

Соберем для  $\Omega = \{E_1, E_2, \dots, E_r\}$  множества  $U(E_i)$  в новую обучающую выборку  $\Omega^* = \{U(E_1), U(E_2), \dots, U(E_r)\}$ . Операцию сходства на элементах  $\Omega^*$  определим как теоретико-множественное пересечение  $\cap$ . Далее построим соответствующие классы эквивалентности и сформируем из них, используя отношение  $\subseteq$  взаимной вложимости соответствующих таким классам эквивалентности подмножеств объектов из  $\Omega^*$ , частично упорядоченные множества. Будем рассматривать диаграммы таких частично упорядоченных множеств, сфокусировав внимание на совместном вхождении признаков из исходного  $U$  одновременно в каждое из множеств объектов, порождающих какую-либо вершину такой диаграммы.

**Определение 7.** Замыканием признака  $a = \{a_1, a_2, \dots, a_n\} \in U$  относительно множества объектов  $\Omega^*$  будем называть множество  $[a]_{U, \Omega^*}$ , представляющее собой сходство (пересечение) всех тех и только тех объектов  $U(E)$  из  $\Omega^*$ , которые содержат данный признак  $a$ .

Используя отношение  $\subseteq$  взаимной вложимости элементов множества

$$GC^1(U, \Omega^*) = \{[a_1]_{U, \Omega^*}, [a_2]_{U, \Omega^*}, \dots, [a_n]_{U, \Omega^*}\} \quad (4)$$

как (непустых) подмножеств заданного множества  $U$ , сформируем частично упорядоченное множество  $\langle GC^1(U, \Omega^*), \subseteq \rangle$ , диаграмму  $D\_GC^1(U, \Omega^*)$  частичного порядка, которую будем (как и в [12]) называть *каркасом* для  $GC^1(U, \Omega^*)$ .

Порядок расположения элементов множества (4) на диаграмме  $D\_GC^1(U, \Omega^*)$  задает естественную последовательность первых вхождений признаков  $a_i$  из исходного множества  $U$  в объекты  $U(E)$  из  $\Omega^*$ . Будем обозначать это отношение частичного порядка на элементах множества  $U$  как  $\prec_2$ . Выражение  $a_{i_i} \prec_2 a_{i_j}$  означает, что на диаграмме  $D\_GC^1(U, \Omega^*)$  первое вхождение признака  $a_{i_i}$  доминируется первым вхождением признака  $a_{i_j}$ . Дополнительно каждая из вершин диаграммы  $D\_GC^1(U, \Omega^*)$  демонстрирует, какие признаки входят в описания событий из исходного множества  $\Omega$  вместе с порождающим путем формирования его замыкания  $[a_i]_{U, \Omega^*}$  эту вершину признаком  $a_i$ .

### 4 Проблема управления ресурсами при противодействии атакам

При организации противодействия компьютерным атакам представляется естественным задействовать любые возможности для рационального использования имеющихся ресурсов. В ситуации, когда атаки могут быть описаны как последовательности состояний объекта защиты, один из вариантов такого управления ресурсами противодействия может быть основан на совместном использовании введенных выше отношений  $\prec_1$  и  $\prec_2$ . Сделать это можно, в частности, по следующей схеме:

- описывается каждый факт атаки как последовательности состояний, образующих формализованное представление соответствующего события. Такие описания будем называть *примерами*;
- дополнительно формируются такие последовательности состояний, которые похожи по своей



структуре на факты примеров, но не являются отражением действий, ведущих к вредоносным эффектам. Эти описания будем называть *контрпримерами*. Примеры и контрпримеры будем называть *прецедентами*;

- используя порождаемое на описаниях событий отношение сходства, выделяется каждая такая цепочка признаков  $\vec{s}$ , которая является общей для всех событий, попадающих в соответствующий класс эквивалентности  $\mathbf{EQ}_s$  по отношению  $R^\otimes$  и  $E_v^\otimes$ . Каждую порожденную таким образом цепочку фиксируем в базе знаний (БЗ) системы противодействия анализируемым атакам. При этом при анализе текущего состояния  $\mathbf{O}$  можно при наблюдении появления признаков  $a_1^i, a_2^i, \dots$  идентифицировать в БЗ все такие цепочки вида  $\vec{s}^i = \langle a_1^i, a_2^i, \dots, a_t^i, \dots \rangle$ , которые позволяют предположить дальнейшее в смысле отношения  $\prec_1^i$  «появление» в рассматриваемой текущей ситуации каждого из соответствующих каждой цепочке признаков  $a_t^i$ . Далее принимается во внимание заранее порождаемая на свертках описаний исходной выборки прецедентов и сохраняемая в БЗ информация о совместном (с учетом отношения  $\prec_2$ ) вхождении признаков из  $\mathbf{U}$  в сходства событий исходной обучающей выборки прецедентов;
- последовательно противодействовать развивающейся атаке и ее продолжениям, известным по уже накопленной базе прецедентов, на тех направлениях, которые соответствуют появлению совместно с уже зафиксированными, а также далее ожидаемыми в смысле каждого отношения  $\prec_1^i$  признаками из каждой цепочки  $\vec{s}^i = \langle a_1^i, a_2^i, \dots, a_t^i, \dots \rangle$  сопутствующих им, т. е. совместно встречающихся с ними на диаграмме  $D\_GC^1(\mathbf{U}, \mathbf{\Omega}^*)$  признаков, а именно: сперва признакам  $b_{11}^i, b_{12}^i, \dots$ , сопутствующим признаку  $a_1^i$  в смысле отношения  $\prec_2$ , затем признакам  $b_{21}^i, b_{22}^i, \dots$ , сопутствующим признаку  $a_2^i$  в смысле отношения  $\prec_2$ , и т. д.

Очевидным аргументом в пользу рациональности именно такой тактики выбора приоритетов противодействия может быть следующее соображение. Если, как показывают результаты обучения на прецедентах исходной обучающей выборки, отраженные в отношениях  $\prec_1$  и  $\prec_2$  признаки  $a_{j_1}^i$  и  $b_{j_2}^i, b_{j_2}^i, \dots$  в ходе результативных атак проявляются вместе и тем более если это совместное вхождение имеет также и *ясную содержательную интерпретацию*, то при фиксации появления  $a_{j_1}^i$  представляется естественным в приоритетном порядке противостоять появлению  $b_{j_1}^i, b_{j_2}^i, \dots$ . При этом реакция на возможные разворачи-

вания различных вредоносных цепочек  $\vec{s}^i = \langle a_1^i, a_2^i, \dots, a_t^i, \dots \rangle$  на ранних стадиях их формирования оставляет больше надежд на эффективное использование ресурсов, имеющихся для организации противодействия.

Следует отметить существенную роль анализа контрпримеров в усечении артефактов, порождаемых при обучении на примерах. Порождение одной и той же цепочки при формировании классов эквивалентности как на примерах, так и на контрпримерах можно рассматривать как сигнал о ее несущественности при анализе имеющейся выборки прецедентов, т. е. о возможности ее отнесения к *артефактам* переобучения. Подобное игнорирование артефактов — один из способов борьбы с *эффектом* так называемого *переобучения* (см., например, [13]), характерным для использования различных методов машинного обучения.

## 5 Заключение

Представлен подход к организации управления ресурсами при формировании противодействия компьютерным атакам, характеризуемым разворачивающимися во времени последовательностями действий (событий).

В его основу положена техника машинного обучения на прецедентах, базирующаяся на формализации сходства как бинарной алгебраической операции и порождении на элементах описаний наблюдаемых атак двух отношений частичного порядка  $\prec_1$  и  $\prec_2$ . Отношение  $\prec_1$  отражает *последовательность* появления наблюдаемых активностей, отношение  $\prec_2$  отражает *совместное воздействие* на наблюдаемых активностей на объект защиты.

Приведенный теоретико-множественный вариант формализованного описания анализируемых атак может быть расширен за счет использования более сложных структур данных. Так, в частности, для описания стадий, последовательности которых представляют фиксируемые события-прецеденты, могут быть использованы не только собственно множества признаков, но и, например, множества признаков вместе с отношениями на них, т. е. графы, в том числе с числовыми метками на вершинах и ребрах.

Интересное расширение представленного подхода состоит в использовании тернарного отношения [14], связывающего фиксируемые вредоносные воздействия с возникающими негативными эффектами в условиях (т. е. с непременным учетом) тех или иных вариантов противодействия влиянию наблюдаемых вредоносных воздействий, т. е. отношения вида ( $\langle$ вариант вредоносных действий), ( $\langle$ вариант



противодействия))  $\Rightarrow$  ((возникающий в результате эффект)).

Для эффективной организации управления перебором вариантов при формировании соответствующих классов эквивалентности в рассматриваемой процедуре, как и в работах [12, 15], может быть использована техника целенаправленной навигации на последовательно порождаемых диаграммах взаимной вложимости классов эквивалентности.

## Литература

1. Grusho A., Levykin M., Timonina E., Piskovski V., Timonina A. Architecture of consecutive identification of attack to information resources // 7th Congress (International) on Ultra Modern Telecommunications and Control Systems Proceedings. — Piscataway, NJ, USA: IEEE, 2015. P. 265–268.
2. Финн В. К. Синтез познавательных процедур и проблема индукции // Научно-техническая информация. Сер. 2, 2009. № 6. С. 1–37.
3. Финн В. К. Об интеллектуальном анализе данных // Новости искусственного интеллекта, 2014. № 3. С. 3–18.
4. Шрейдер Ю. А. Равенство, сходство, порядок. — М.: Наука, 1971. 255 с.
5. Гусакова С. М., Финн В. К. Сходства и правдоподобный вывод // Известия АН СССР. Сер. Техническая кибернетика, 1987. № 5. С. 42–63.
6. Финн В. К. Индуктивные методы Д. С. Милля в системах искусственного интеллекта // Искусственный интеллект и принятие решений, 2010. Ч. I. № 3. С. 3–21; Ч. II. № 4. С. 14–40.
7. Забейло М. И. О некоторых оценках сложности вычислений в ДСМ-рассуждениях // Искусственный интеллект и принятие решений, 2015. Ч. I. № 1. С. 3–17; Ч. II. № 2. С. 3–17.
8. Simon J. On the difference between one and many (preliminary version) // Automata, languages and programming / Eds. A. Salomaa, M. Steinby. — Lecture notes in computer science ser. — Berlin–Heidelberg: Springer, 1977. Vol. 52. P. 480–491.
9. Valiant L. G. The complexity of enumeration and reliability problems // SIAM J. Comput., 1979. Vol. 8. P. 410–421.
10. Valiant L. G. The complexity of computing the permanent // Theor. Comput. Sci., 1979. Vol. 8. P. 189–201.
11. Гэри М., Джонсон Д. С. Вычислительные машины и трудно-решаемые задачи / Пер. с англ. — М.: Мир, 1982. 416 с. (Garey M. R., Johnson D. S. Computers and intractability: A guide to the theory of NP-completeness. — San Francisco, CA, USA: W. H. Freeman and Co., 1979. 338 p.)
12. Забейло М. И. О некоторых возможностях управления перебором в ДСМ-методе // Искусственный интеллект и принятие решений, 2014. Ч. I. № 1. С. 95–110; Ч. II. № 3. С. 3–21.
13. Everitt B. S. Cambridge dictionary of statistics. — Cambridge: Cambridge University Press, 2002. 410 p.
14. Финн В. К., Шестерникова О. П. О новом варианте обобщенного ДСМ-метода // Искусственный интеллект и принятие решений, 2016. № 1. С. 57–63.
15. Забейло М. И. Приближенный ДСМ-метод на примерах // Научно-техническая информация. Сер. 2, 2014. № 10. С. 1–12.

Поступила в редакцию 03.01.18

## ON SOME POSSIBILITIES OF RESOURCE MANAGEMENT FOR ORGANIZING ACTIVE COUNTERACTION TO COMPUTER ATTACKS

A. A. Grusho, M. I. Zabezhailo, A. A. Zatsarinny, and E. E. Timonina

Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** Rational counteraction to computer attacks, described as event sequences, is discussed. The approach is based on the mathematical technique of learning by precedents, formalizing similarity as a binary algebraic operation. Similarities of event sequences are analyzed. The learned classes of similarity (tolerance classes) are used to recognize computer attacks on initial steps of their life cycle. A problem-oriented resource management technology aimed at developing rational counteraction to attacks of the discussed type is presented.

**Keywords:** information security; data analysis; similarity as binary algebraic operation; similarity of sequences; resource management

**DOI:** 10.14357/19922264180108

## Acknowledgments

The paper was supported by the Russian Foundation for Basic Research (project 15-29-07981).

## References

1. Grusho, A., M. Levykin, E. Timonina, V. Piskovski, and A. Timonina. 2015. Architecture of consecutive identification of attack to information resources. *7th Congress (International) on Ultra Modern Telecommunications and Control Systems Proceedings*. Piscataway, NJ: IEEE. 265–268.
2. Finn, V. K. 2009. The synthesis of cognitive procedures and the problem of induction. *Automatic Documentation Math. Linguistics* 43(3):149–195.
3. Finn, V. K. 2014. Ob intellektual'nom analize dannyykh [On intelligent data analysis]. *Novosti iskusstvennogo intellekta* [Artificial Intelligence News] 3:3–18.
4. Shreyder, J. A. 1971. *Ravenstvo, skhodstvo, poryadok* [Equality, similarity, order]. Moscow: Nauka. 255 p.
5. Gusakova, S. M., and V. K. Finn. 1987. Skhodstvo i pravdopodobnyy vyvod [Similarity and plausible inference]. *Izvestiya AN SSSR. Ser. Tekhnicheskaya kibernetika* [News of Academy of Sciences of the USSR. Ser. engineering cybernetics] 5:42–63.
6. Finn, V. K. 2011–2012. Induktivnye metody D. S. Millya v sistemakh iskusstvennogo intellekta [J. S. Mill's inductive methods in artificial intelligence systems]. *Iskusstvennyy intellekt i prinyatie resheniy* [Scientific and Technical Information Processing]. Part I (2011). 38(6):385–402. Part II (2012). 39(5):241–260.
7. Zabezhalo, M. I. 2015. O nekotorykh otsenkakh slozhnosti vichisleniy v DSM-rassuzhdeniyakh [To the computational complexity of hypotheses generation in JSM-method]. *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial Intelligence and Decision Making]. Part I. 1:3–17. Part II. 2:3–17.
8. Simon, J. 1977. On the difference between one and many (preliminary version). *Automata, languages and programming*. Eds. A. Salomaa and M. Steinby. Lecture notes in computer science ser. Berlin–Heidelberg: Springer. 52:480–491.
9. Valiant, L. G. 1979. The complexity of enumeration and reliability problems. *SIAM J. Comput.* 8:410–421.
10. Valiant, L. G. 1979. The complexity of computing the permanent. *Theor. Comput. Sci.* 8:189–201.
11. Garey, M. R., and D. S. Johnson. 1979. *Computers and intractability: A guide to the theory of NP-completeness*. San Francisco, CA: W. H. Freeman and Co. 338 p.
12. Zabezhalo, M. I. 2014. O nekotorykh vozmozhnostyakh upravleniya pereborom v DSM-metode [On some possibilities to control computational complexity of hypotheses generation in JSM-method]. *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial Intelligence and Decision Making]. Part I. 1:95–110. Part II. 3:3–21.
13. Everitt, B. S. 2002. *Cambridge dictionary of statistics*. Cambridge: Cambridge University Press. 410 p.
14. Finn, V. K., and O. P. Shesternikova. 2016. O novom variante obobschennogo DSM-metoda [On the new variant of generalized JSM-method]. *Iskusstvennyy intellekt i prinyatie resheniy* [Artificial Intelligence and Decision Making] 1:57–63.
15. Zabezhalo, M. I. 2014. Priblizhennyy DSM-metod na primerakh [Approximate JSM-method by examples]. *Nauchno-tekhnicheskaya informatsiya* [Scientific and Technical Information Processing]. Ser. 2. 10:1–12.

Received January 3, 2018

## Contributors

**Grusho Alexander A.** (b. 1946) — Doctor of Science in physics and mathematics, professor, Head of Laboratory, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; grusho@yandex.ru

**Zabezhalo Michael I.** (b. 1956) — Doctor of Science in physics and mathematics, Head of Laboratory, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; m.zabezhalo@yandex.ru

**Zatsarinny Alexander A.** (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; alex250451@mail.ru

**Timonina Elena E.** (b. 1952) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; eltimon@yandex.ru

# РЕКОНСТРУКЦИЯ ОСЕЙ ГЛАВНЫХ КОМПОНЕНТ

М. П. Кривенко<sup>1</sup>

**Аннотация:** Анализ главных компонент (РСА — Principal Component Analysis) широко используется при исследовании данных, их сжатии и визуализации. Новые возможности открывает вероятностный РСА (РРСА — probabilistic PCA), реализуемый в рамках принципа максимального правдоподобия для гауссовской модели с латентными переменными. В рамках РРСА появились алгоритмы обработки данных, нацеленные на снижение размерности данных и обеспечивающие переход в пространство главных компонент, но не дающие в явном виде характеристики главных компонент. Статья посвящена деталям, углубляющим понимание особенностей РРСА, исправлениям выявленных ошибок в публикациях. Предложены и обоснованы два метода реконструкции характеристик главных компонент. Один из них основывается на пересчете ковариационной матрицы в сформированном пространстве главных компонент. Другой метод заключается в последовательном повторении одинаковых шагов: выявления первой главной компоненты и исключения ее из анализа данных.

**Ключевые слова:** анализ главных компонент; EM-алгоритм; реконструкция осей и дисперсий главных компонент

**DOI:** 10.14357/19922264180109

## 1 Введение

В сфере практического использования методов многомерного анализа данных исследователь сталкивается с трудностями описания сложных взаимосвязей между признаками и интерпретацией множества частных результатов о зависимостях. Эти трудности можно попытаться преодолеть, если свести систему заданных величин к совокупности независимых величин и/или снизить размерность исходного признакового пространства. Один из путей решения подобных задач заключается в применении анализа главных компонент (РСА); его суть заключается в следующем: с помощью линейного преобразования исходных величин осуществляется переход к новой системе уже некоррелированных признаков, которые называют главными компонентами. В случае нормального распределения данных подобное преобразование приведет к системе независимых признаков. Вопрос снижения размерности будет решаться в этой новой системе признаков в предположении, что информативность признака напрямую связана с его изменчивостью. Таким образом из совокупности исходных характеристик для последующего анализа оставляются лишь те, которым свойствен «существенно» больший разброс значений, чем остальным.

Популярность РСА определяется двумя важными свойствами. Во-первых, это оптимальная (по средней квадратичной ошибке) линейная схема для сжатия множества векторов высокой размерности

в множество векторов более низкой размерности, а затем их восстановления. Во-вторых, параметры модели могут быть вычислены непосредственно из данных.

Однако, несмотря на эти привлекательные особенности, модель РСА имеет несколько недостатков. Один из них заключается в том, что прямые (стандартные, привычные) методы нахождения направлений главных компонент с помощью разложения матрицы вторых моментов приводят к проблемам при большом объеме данных высокой размерности. Еще один недостаток стандартных подходов к РСА заключается в том, что неясно, как правильно обращаться с неполными данными. Обычно они либо отбрасываются, либо восполняются на основании имеющихся значений. Однако такой подход оказывается крайне неэффективным, когда доля пропусков значительна.

Дополнительные трудности при применении стандартных методов РСА доставляют отклонения от принимаемой модели данных: в частности, стандартный алгоритм РСА основан на предположении, что данные не испорчены выбросами.

Для нейтрализации этих недостатков стандартного РСА было предложено множество методов. Цель последующего материала — показать варианты улучшенного РСА, которые могут нейтрализовать «проклятие» размерности, допускают наличие пропусков и выбросов. В ходе изложения выявляются ошибки известных результатов, представля-

<sup>1</sup> Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, mkrivenko@ipiran.ru

ются два новых метода, обеспечивающих полноту применения PCA.

## 2 Стандартный анализ главных компонент

Ввести главные компоненты можно различными способами. В любом случае исходной информацией для построения искомого линейного преобразования служит ковариационная (или корреляционная) матрица, поэтому можно считать, что каждая компонента вектора данных центрирована своим средним. Формировать главные компоненты можно как в теоретическом случае, когда матрица вторых центральных моментов задана, так и в выборочном случае, когда она оценивается по наблюдаемым значениям.

Без потери общности ограничимся выборочным случаем центрированных данных —  $(d \times N)$ -матрицы «признак—объект»  $\mathbf{Y}$  — и рассмотрением ковариационной  $(d \times d)$ -матрицы  $\mathbf{S} = (1/N)\mathbf{Y}\mathbf{Y}^T$ . Для  $\mathbf{S}$  действует спектральное разложение  $\mathbf{S} = \mathbf{U}\mathbf{V}\mathbf{U}^T$ , где  $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$  и столбцы  $\mathbf{U}$  суть оси главных компонент;  $\mathbf{V} = \text{diag}(v_{11}, \dots, v_{dd})$  и  $v_{11} \geq \dots \geq v_{dd}$ .

Тогда основные положения PCA следующие:

- $(d \times k)$ -матрица преобразования данных  $\mathbf{W}_{\text{PCA}}$  состоит из столбцов, совпадающих с осями первых главных компонент, причем  $\mathbf{W}_{\text{PCA}}^T \mathbf{W}_{\text{PCA}} = \mathbf{I}$ ;
- преобразование  $\langle \mathbf{X} \rangle_{\text{PCA}}$  исходных данных  $\mathbf{Y}$  выражается как  $\langle \mathbf{X} \rangle_{\text{PCA}} = \mathbf{W}_{\text{PCA}}^T \mathbf{Y}$ , для  $k < d$  имеет место снижение размерности;
- восстановленные данные

$$\langle \mathbf{Y} \rangle_{\text{PCA}} = \mathbf{W}_{\text{PCA}} \langle \mathbf{X} \rangle_{\text{PCA}} = \mathbf{W}_{\text{PCA}} \mathbf{W}_{\text{PCA}}^T \mathbf{Y};$$

- ошибка восстановления

$$e_R = \sum_{n=1}^N \|\mathbf{y}_n - \langle \mathbf{y}_n \rangle_{\text{PCA}}\|^2 = \sum_{j=k+1}^d v_{jj},$$

$e_R = 0$  при  $k = d$ ; причем из всех ортогональных линейных проекций проекция главных компонент минимизирует квадратичную ошибку восстановления.

Общая теория PCA изложена в [1]. Обзор отдельных результатов и доказательство обобщающей теоремы, служащие обоснованием представления  $e_R = \sum_{j=k+1}^d v_{jj}$ , можно найти в [2].

## 3 Вероятностный анализ главных компонент

Описанный классический PCA основывается на идеях перехода к некоррелированным величинам и линейной проекции исходных данных в пространство меньшей размерности. Можно рассмотреть PCA с позиций правдоподобия; такой подход открывает новые возможности и дает ряд преимуществ:

- можно построить вариант EM (expectation-maximization) алгоритма для PCA, который является вычислительно эффективным в ситуации, когда требуются лишь несколько первых главных компонент, и позволяет избежать оценивания ковариационной матрицы исходных данных;
- вероятностный характер модели в сочетании с идеями EM-алгоритма позволяют справляться с обработкой пропущенных данных;
- модель может быть усложнена до смеси вероятностных моделей PCA, что создает предпосылки для описания данных более сложной и реалистичной структуры;
- вероятностная модель PCA предоставляет возможность генерировать выборки данных.

**Вероятностная модель.** Анализ главных компонент можно рассматривать как вариант обработки линейных гауссовских моделей. Целью таких моделей является представление ковариационной структуры наблюдаемой  $d$ -мерной переменной  $\mathbf{y}$  с использованием меньшего чем  $d(d+1)/2$  числа параметров, необходимого в случае полной ковариационной матрицы. Линейные гауссовские модели предполагают, что  $\mathbf{y}$  было получено как линейное преобразование некоторой  $k$ -мерной латентной переменной  $\mathbf{x}$  плюс шум. Как и прежде,  $\mathbf{W}$  —  $(d \times k)$ -матрица преобразования,  $\boldsymbol{\varepsilon}$  —  $d$ -мерный вектор шума (ошибки) с ковариационной матрицей  $\boldsymbol{\Psi}$ . Тогда модель имеет вид:

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}. \quad (1)$$

Принимается, что  $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$  и  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Psi})$ . Тогда  $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})$ . Чтобы «сэкономить» параметры при представлении исходных зависимостей в пространстве  $d$ , необходимо принять  $k < d$ , а также уточнить ковариационную структуру гауссовского шума  $\boldsymbol{\varepsilon}$  путем ограничения матрицы  $\boldsymbol{\Psi}$ .

Вероятностная модель PCA возникает в случае, когда  $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}$  или  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ . Предполагается независимость отдельных наблюдаемых данных,



поэтому важными становятся выражения для распределений  $\mathbf{y}$  и  $\mathbf{x}$ :

$$P(\mathbf{y}, \mathbf{x}) = N \left( \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I} & \mathbf{W} \\ \mathbf{W}^T & \mathbf{I} \end{pmatrix} \right).$$

Благодаря свойствам нормального многомерного распределения имеем:

$$P(\mathbf{y}|\mathbf{x}) = (2\pi\sigma^2)^{-d/2} \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{W}\mathbf{x} - \boldsymbol{\mu}\|^2 \right\},$$

откуда с использованием понятия ошибки восстановления становится понятна связь с PCA. Аналогично после упрощений можно получить, что

$$P(\mathbf{x}|\mathbf{y}) = N \left( (\mathbf{W}^T\mathbf{W} + \sigma^2\mathbf{I})^{-1} \mathbf{W}^T(\mathbf{y} - \boldsymbol{\mu}), \sigma^2 (\mathbf{W}^T\mathbf{W} + \sigma^2\mathbf{I})^{-1} \right). \quad (2)$$

Для матрицы, входящей в правую часть, обычно используется самостоятельное обозначение  $\mathbf{W}^T\mathbf{W} + \sigma^2\mathbf{I} = \mathbf{M}$ .

**Прямое оценивание параметров.** Согласно [3], оценки максимального правдоподобия для параметров модели следующие:

$$\left. \begin{aligned} \boldsymbol{\mu}_{ML} &= \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n; \\ \sigma_{ML}^2 &= \frac{1}{d-k} \sum_{j=k+1}^d v_{jj}; \\ \mathbf{W}_{ML} &= \mathbf{U}_k (\mathbf{V}_k - \sigma_{ML}^2 \mathbf{I})^{1/2} \mathbf{R}, \end{aligned} \right\} \quad (3)$$

где столбцы  $(d \times k)$ -матрицы  $\mathbf{U}_k$  суть оси первых  $k$  главных компонент;  $\mathbf{V}_k$  — диагональная  $(k \times k)$ -матрица соответствующих дисперсий (собственные векторы и собственные значения матрицы  $\mathbf{S}$ );  $\mathbf{R}$  — произвольная  $k \times k$  ортогональная матрица поворота. Для реализации рассмотренного варианта РРСА сначала необходимо факторизовать матрицу  $\mathbf{S}$ , после чего можно найти оценки максимального правдоподобия. Без потери общности далее будем считать, что данные центрированы и  $\boldsymbol{\mu} = \mathbf{0}$ .

**Редукция размерности и оптимальное восстановление.** В обычном PCA снижение размерности осуществляется с помощью преобразования  $\langle \mathbf{X} \rangle_{PCA} = \mathbf{W}_{PCA}^T \mathbf{Y}$ , а восстановление данных —  $\langle \mathbf{Y} \rangle_{PCA} = \mathbf{W}_{PCA} \langle \mathbf{X} \rangle_{PCA}$ . Это может быть реализовано и в рамках РРСА. Однако надо заметить, что в рассматриваемой вероятностной модели (1) осуществляется отображение из латентного пространства значений  $\mathbf{x}$  меньшей размерности в пространство данных  $\mathbf{y}$  большей размерности. В этом случае

каждой точке данных  $\mathbf{y}_n$  соответствует в латентном пространстве не один вектор, а множество точек с гауссовским распределением вида (2). В качестве альтернативы стандартной проекции PCA может рассматриваться удобная обобщающая характеристика этого распределения, например представление  $\mathbf{y}_n$  с помощью апостериорного среднего

$$\langle \mathbf{x}_n \rangle_{ML} = (\mathbf{W}_{ML}^T \mathbf{W}_{ML} + \sigma^2 \mathbf{I})^{-1} \mathbf{W}_{ML}^T \mathbf{y}_n,$$

являющегося одновременно и модой апостериорного распределения. Отметим при этом, что ковариация условного распределения задается величиной  $\sigma^2 (\mathbf{W}_{ML}^T \mathbf{W}_{ML} + \sigma^2 \mathbf{I})^{-1}$ , которая является постоянной для всех точек данных. Теперь с помощью приложения В [3] получаем, что оптимальное восстановление апостериорного среднего задается равенствами:

$$\begin{aligned} \langle \mathbf{y}_n \rangle_{ML} &= \mathbf{W}_{ML} (\mathbf{W}_{ML}^T \mathbf{W}_{ML})^{-1} \mathbf{M} \langle \mathbf{x}_n \rangle_{ML} = \\ &= \mathbf{W}_{ML} (\mathbf{W}_{ML}^T \mathbf{W}_{ML})^{-1} \mathbf{W}_{ML}^T \mathbf{y}_n. \end{aligned}$$

**EM-алгоритм оценивания параметров.** Рассмотрим версию EM-алгоритма, позволяющую получать для набора данных главные компоненты. Данный алгоритм не требует вычисления выборочных ковариаций, может обрабатывать данные высокой размерности более эффективно, чем традиционные PCA. Следуя [4], получаем шаги EM-алгоритма в следующем виде:

$$\begin{aligned} \mathbf{M} &= \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbf{I}; \\ \widetilde{\mathbf{W}} &= \mathbf{S} \mathbf{W} (\sigma^2 \mathbf{I} + \mathbf{M}^{-1} \mathbf{W}^T \mathbf{S} \mathbf{W})^{-1} = \\ &= \mathbf{S} \mathbf{W} (\sigma^2 \mathbf{M} + \mathbf{W}^T \mathbf{S} \mathbf{W})^{-1} \mathbf{M}; \\ \tilde{\sigma}^2 &= \frac{1}{d} \text{tr} \left( \mathbf{S} - \mathbf{S} \mathbf{W} \mathbf{M}^{-1} \widetilde{\mathbf{W}}^T \right), \end{aligned}$$

где  $\mathbf{W}$  и  $\sigma^2$  суть старые значения параметров модели;  $\widetilde{\mathbf{W}}$  и  $\tilde{\sigma}^2$  — новые. Приведена альтернативная форма для  $\widetilde{\mathbf{W}}$ , более удобная с точки зрения обращения матриц  $(\sigma^2 \mathbf{M} + \mathbf{W}^T \mathbf{S} \mathbf{W})$  — симметричная матрица). В результате итерационного процесса получаются значения  $\mathbf{W}_{EM}$  и  $\sigma_{EM}^2$ .

Для реализации итерационного EM-алгоритма необходимо решить вопросы выбора начальных приближений и критерия завершения пересчета оценок. Принимая во внимание, что никакой дополнительной информации о параметрах модели нет, приходится остановиться на произвольном выборе матрицы факторных нагрузок: чтобы избежать «неожиданностей» при матричных операциях, в качестве  $k$  столбцов  $\mathbf{W}$  берутся первые столбцы

единичной ( $d \times d$ )-матрицы. Что касается параметра  $\sigma^2$ , то здесь можно учесть дисперсии исходных данных и принять  $\sigma^2 = \text{tr}(\mathbf{S})/d$ .

В основу критерия завершения итерационного процесса можно положить контроль изменений следующих характеристик: ошибки восстановления  $e_R^{(t)}$ ,  $\mathbf{W}^{(t)}$  и  $(\sigma^2)^{(t)}$ , где  $t$  — шаг итерации. Первый показатель отражает поведение целевой функции, два последних — параметров модели.

Особого внимания требует формализация сходимости в случае  $\mathbf{W}$ : при максимизации правдоподобия действует представление (3), где конкретный вид  $\mathbf{R}$  зависит, например, от начальных приближений для параметров модели  $\mathbf{W}$  и  $\sigma^2$ . По этой причине объектом внимания становится соотношение:

$$\mathbf{W}^{(t)} \xrightarrow{t \rightarrow \infty} \mathbf{U}_k (\mathbf{V}_k - \sigma^2 \mathbf{I})^{1/2} \mathbf{R},$$

которое эквивалентно

$$\mathbf{B}\mathbf{W}^{(t)} \xrightarrow{t \rightarrow \infty} \mathbf{R},$$

где  $\mathbf{B} = ((\mathbf{V}_k - \sigma^2 \mathbf{I})^{1/2})^{-1} \mathbf{U}_k^T$ . Таким образом, необходимо следить за сходимостью  $(\mathbf{B}\mathbf{W}^{(t)})^T \mathbf{B}\mathbf{W}^{(t)} \xrightarrow{t \rightarrow \infty} \mathbf{I}$ , например путем измерения точности аппроксимации единичной матрицы с помощью  $(\mathbf{B}\mathbf{W}^{(t)})^T \mathbf{B}\mathbf{W}^{(t)}$ .

Анализ поведения EM-алгоритма в реальных ситуациях выявил, что чувствительность указанных показателей различна. Это может приводить к ошибкам: согласно одному показателю процесс сошелся, но предельных значений другими не достигнуто.

Ряд проведенных автором статьи экспериментов показал следующее: если целью исследований является обработка данных в пространстве сниженной размерности, то надо использовать относительную погрешность для оценок параметров  $\sigma^2$  и  $\mathbf{W}$ ; если же в этом нет необходимости (например, в задачах снижения уровня шума или передачи сжатых данных с последующим их восстановлением), то можно ограничиться только слежением за точностью восстановления.

**Вращательная неоднозначность и реконструкция осей.** Для результатов рассмотренных методов, которые максимизируют правдоподобие, действует представление (3). Если необходимо найти истинные оси главных компонент  $\mathbf{U}_k$  (а не только подпространство главных компонент), наличие произвольной матрицы поворота  $\mathbf{R}$  создает трудности. Однако в PPCA определить  $\mathbf{R}$  можно, так как соотношение

$$\mathbf{W}_{ML}^T \mathbf{W}_{ML} = \mathbf{R}^T (\mathbf{V}_k - \sigma^2 \mathbf{I}) \mathbf{R} \quad (4)$$

означает, что  $\mathbf{R}^T$  можно вычислить как матрицу из собственных векторов ( $k \times k$ )-матрицы  $\mathbf{W}_{ML}^T \mathbf{W}_{ML}$ . Пусть найдены матрица  $\mathbf{W}_{ML}$  и факторизация матрицы  $\mathbf{W}_{ML}^T \mathbf{W}_{ML} = \mathbf{L}_{ML} \mathbf{K}_{ML} \mathbf{L}_{ML}^T$ . С учетом (4) имеем:

$$\mathbf{R} = \mathbf{L}_{ML}^T; \quad (\mathbf{V}_k - \sigma^2 \mathbf{I})^{1/2} = \mathbf{K}_{ML}^{1/2}.$$

Отсюда  $\mathbf{W}_{ML} = \mathbf{U}_k \mathbf{K}_{ML}^{1/2} \mathbf{L}_{ML}^T$ . Следовательно, для PPCA имеем:

$$\mathbf{U}_k = \mathbf{W}_{ML} \tilde{\mathbf{L}}_{ML} \mathbf{K}_{ML}^{-1/2};$$

$$\sigma^2 = \sigma_{ML}^2;$$

$$\mathbf{V}_k = \mathbf{K}_{ML} + \sigma_{ML}^2 \mathbf{I}.$$

Однако оценка  $\mathbf{V}_k$  оказывается крайне неточной. Поэтому если получение дисперсий главных компонент важно, то необходимо реализовать следующие шаги:

- получить  $\tilde{\mathbf{U}}_k = \mathbf{W}_{ML} \mathbf{L}_{ML}$ ;
- найти  $\mathbf{U}_k$  с помощью нормализации  $\tilde{\mathbf{U}}_k$  по столбцам;
- сформировать проекцию  $\tilde{\mathbf{X}} = \mathbf{U}_k^T \mathbf{Y}$ ;
- найти  $\mathbf{V}_k$  путем оценивания дисперсий для отдельных компонент  $\tilde{\mathbf{X}}$ .

Подобный способ оценивания повышает точность нахождения  $\mathbf{V}_k$  на несколько порядков.

## 4 Развитие методов вероятностного анализа главных компонент

Интересным развитием EM-подхода в оценивании параметров PCA становится рассмотрение предельного случая в модели PPCA [5], когда дисперсии шума  $\epsilon$  становятся бесконечно малыми и равными во всех направлениях. Математически это означает, что  $\Psi = \lim_{\sigma^2 \rightarrow 0} \sigma^2 \mathbf{I}$ . Поскольку ошибка представления  $\mathbf{y}$  оказывается бесконечно малой, распределение сжимается до одной точки и с использованием  $\delta$ -функции принимает вид:

$$\begin{aligned} P(\mathbf{x}|\mathbf{y}) &= N\left((\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{y}, \mathbf{0}\right) = \\ &= \delta\left(\mathbf{x} - (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{y}\right). \end{aligned}$$

Это позволяет свести EM-алгоритм в предельном случае к простым действиям: с помощью E-шага оценивание неизвестного состояния  $\mathbf{x}$ , а затем

на М-шаге максимизация правдоподобия при полученных  $\mathbf{x}$  и наблюдаемых  $\mathbf{y}$ . Таким образом приходим к необходимости повторения следующих шагов итерационного процесса:

$$\mathbf{X} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{Y};$$

$$\widetilde{\mathbf{W}} = \mathbf{Y} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1},$$

где  $\mathbf{W}$  — старое значение параметров модели, а  $\widetilde{\mathbf{W}}$  — новое. В результате получаются оценка факторных нагрузок  $\mathbf{W}_{aEM}$ , а также соотношения:

$$\langle \mathbf{x}_n \rangle_{aEM} = (\mathbf{W}_{aEM}^T \mathbf{W}_{aEM})^{-1} \mathbf{W}_{aEM}^T \mathbf{Y}_n;$$

$$\langle \mathbf{y}_n \rangle_{aEM} = \mathbf{W}_{aEM} \langle \mathbf{x}_n \rangle_{aEM} =$$

$$= \mathbf{W}_{aEM} (\mathbf{W}_{aEM}^T \mathbf{W}_{aEM})^{-1} \mathbf{W}_{aEM}^T \mathbf{Y}_n.$$

В [5] декларировалась возможность получения характеристик первых главных компонент и давались расплывчатые рекомендации по тому, как это сделать. Все это дословно повторялось в ряде последующих работ, никак не обосновывалось и сразу же опровергалось проведенными автором статьи экспериментами.

В результате на первый взгляд получается элегантно решение задачи РРСА. Но оказывается, что оно существенно отличается от РРСА-решения: формируемая в результате работы численного EM-алгоритма оценка  $\mathbf{W}$  явно через оси главных компонент не выражается. Более того, в случае предельного РРСА возможности выбора различных оценок  $\mathbf{W}$  шире, чем в РРСА. Действительно, пусть

$$\mathbf{W} = \mathbf{U}_k \mathbf{A}, \quad (5)$$

где  $\mathbf{A}$  — произвольная  $k \times k$  невырожденная матрица. Тогда последовательными подстановками (5) в формулы для шагов алгоритма получаем, что EM-алгоритм для предельного РРСА завершит свою работу после первого же шага. В итоге он дает решение, лежащее в пространстве осей главных компонент и принимающее достаточно произвольный вид. Заметим, что в преобразование  $\mathbf{A}$ , в частности, входит и матрица перестановки, поэтому в результате работы алгоритма может получиться решение, просто не совпадающее с  $\mathbf{W}_{ML}$  по порядку следования осей главных компонент.

Попытка построить алгоритм, позволяющий не только осуществить переход в пространство первых главных компонент, но и сформировать оси главных компонент, была представлена в [6]. Идея состояла в последовательном применении ряда усложняющихся по числу главных компонент моделей, в результате чего решение  $\mathbf{W}_{cEM}$  задачи РРСА предлагалось искать с помощью последовательного выполнения следующих шагов:

$$\mathbf{X} = \{\mathcal{L}(\mathbf{W}^T \mathbf{W})\}^{-1} \mathbf{W}^T \mathbf{Y};$$

$$\widetilde{\mathbf{W}} = \mathbf{Y} \mathbf{X}^T \{\mathcal{U}(\mathbf{X} \mathbf{X}^T)\}^{-1},$$

где для произвольной матрицы  $\mathbf{A} = (a_{ij})$  верно следующее:

$$\mathcal{L}(\mathbf{A}) = \begin{cases} a_{ij}, & i \geq j; \\ 0, & i < j; \end{cases}$$

$$\mathcal{U}(\mathbf{A}) = \begin{cases} a_{ij}, & i \leq j; \\ 0, & i > j. \end{cases}$$

Заметим, что обоснование правомерности выводов о существовании решения задачи РРСА приводилось лишь для отдельной модели с определенным числом главных компонент. Во многих последующих публикациях (см. например, [7]) со ссылкой на [6] указывались необоснованные или вообще несуществующие результаты. В частности, бездоказательно утверждалось, что при исчезающем уровне шума матрица  $\mathbf{W}$  сходится к оценке максимального правдоподобия  $\mathbf{W}_{cEM} = \mathbf{U}_k \mathbf{V}_k^{1/2}$  (в обозначениях данной статьи). Таким образом алгоритм с выделением треугольных подматриц, по мысли его авторов и последователей, исключал вращательную неоднозначность.

Для построения контрпримеров автором статьи был проведен ряд экспериментов: генерирование случайной ковариационной матрицы и на ее основе выборки из 100 наблюдаемых  $d$ -мерных векторов, которые участвовали в РСА для различных значений  $k$ . Оценивалась доля случаев (общее число экспериментов — 100), когда матрица  $\mathbf{U}_k^T \mathbf{W}_{cEM}$  не совпадала с диагональной, где  $\mathbf{W}_{cEM}$  — результат работы описанного EM-алгоритма с ограничениями [6]. Соответствующие результаты сведены в таблицу и свидетельствуют об ошибочности выводов [6, 7]. Также с осторожностью надо подходить и к ряду итерационных алгоритмов для РСА, включающих оператор  $\mathcal{L}(\mathbf{A})$  (см., например, [8]).

Доля случаев (в %), когда  $\mathbf{U}_k^T \mathbf{W}$  не диагональная

$d$	$k$				
	1	2	3	4	5
10	0	0	29	56	59
50	0	0	28	36	54
100	0	0	25	37	52

## 5 Реконструкция характеристик главных компонент

Перепроверка утверждений, содержащихся в ряде публикаций (часть из них вошла в приведенный ниже список литературы), и выявление в них

ошибок привели автора данной работы к необходимости решать вопросы, можно ли и как построить оси и дисперсии главных компонент, зная результат EM-алгоритма для предельного PPCA.

Начнем с выделения характеристик  $k$  главных компонент после того, как найдено решение  $\mathbf{W}_{\text{EM}}$ . Столбцы  $\mathbf{W}_{\text{EM}}$  задают пространство главных компонент. После ортонормализации базиса полученное преобразование поворота  $\hat{\mathbf{W}}$  определяет пространство, содержащее такие же главные компоненты, что и  $\mathbf{y}$ . Таким образом,  $\mathbf{X} = \hat{\mathbf{W}}^T \mathbf{Y}$  есть оценка результата преобразования исходных данных для PCA. Тогда

$$\mathbf{X}\mathbf{X}^T = \hat{\mathbf{W}}^T \mathbf{Y}\mathbf{Y}^T \hat{\mathbf{W}} = \hat{\mathbf{W}}^T \mathbf{U}\mathbf{V}\mathbf{U}^T \hat{\mathbf{W}};$$

$$\left(\hat{\mathbf{W}}^T \mathbf{U}\right) \left(\hat{\mathbf{W}}^T \mathbf{U}\right)^T = \hat{\mathbf{W}}^T \mathbf{U}\mathbf{U}^T \hat{\mathbf{W}} = \mathbf{I}.$$

Но в силу того, что векторы  $\mathbf{u}_{k+1}, \dots, \mathbf{u}_d$  ортогональны подпространству главных компонент, т. е. столбцам  $\hat{\mathbf{W}}$ , имеем:

$$\mathbf{X}\mathbf{X}^T = \left(\hat{\mathbf{W}}^T \mathbf{U}_k\right) \mathbf{V}_k \left(\hat{\mathbf{W}}^T \mathbf{U}_k\right)^T;$$

$$\left(\hat{\mathbf{W}}^T \mathbf{U}_k\right)^T \left(\hat{\mathbf{W}}^T \mathbf{U}_k\right) = \mathbf{I},$$

т. е. получено спектральное разложение симметричной матрицы  $\mathbf{X}\mathbf{X}^T$ . Но если фактически найти  $\mathbf{X}\mathbf{X}^T$  и ее спектральное разложение  $\mathbf{X}\mathbf{X}^T = \mathbf{U}_x \mathbf{V}_x \mathbf{U}_x^T$ , то получаем равенства относительно искомого объектов:  $\hat{\mathbf{W}}^T \mathbf{U}_k = \mathbf{U}_x$  и, следовательно,  $\mathbf{U}_k = \hat{\mathbf{W}} \mathbf{U}_x$ ;  $\mathbf{V}_k = \mathbf{V}_x$ .

Предложенный метод основывается на двух важных моментах: EM-алгоритм для предельного PPCA формирует пространство главных компонент и переход к любому ортонормальному базису в этом пространстве не изменяет оси главных компонент. Поэтому ценой вычисления ковариационной матрицы данных в пространстве меньшей размерности становится возможным реконструкция первых главных компонент.

Иная идея использования результатов EM-алгоритма была положена в основу другого алгоритма. В силу того что предельный PPCA формирует пространство главных компонент, в случае единственной главной компоненты в результате применения EM-алгоритма будет получена именно она и, самое главное, будет доступна соответствующая ось. Раз так, то ее надо зафиксировать, исключить изменения данных вдоль нее и снова приступить к выявлению единственной главной компоненты (теперь уже второй для исходных данных). Таким образом последовательно можно выявить все нужные первые главные компоненты и их оси. Соответствующий алгоритм построения  $\mathbf{W}_{\text{EM}}$  и дисперсий первых главных компонент приобретает вид:

1. Положить  $\tilde{\mathbf{Y}}^{(1)} = \mathbf{Y}$ .
2. Для  $t = 1, \dots, k$  повторить последующие шаги:
  - 2.1. Найти для  $\tilde{\mathbf{Y}}^{(t)}$  ось единственной первой главной компоненты  $\tilde{\mathbf{w}}$ .
  - 2.2. Положить  $t$ -й столбец  $\mathbf{W}_{\text{EM}}$  равным  $\tilde{\mathbf{w}}$ ; спроецировать данные  $\mathbf{Y}$  на  $\tilde{\mathbf{w}}$  и найти дисперсию  $t$ -й главной компоненты.
  - 2.3. Сформировать базис  $\mathbf{B}^{(t)}$ , в качестве первого вектора которого взять найденную ось главной компоненты  $\tilde{\mathbf{w}}$ , все остальные векторы заполнить произвольно (например, с помощью базисных единичных векторов).
  - 2.4. Ортонормализовать  $\mathbf{B}^{(t)}$  с помощью алгоритма ортогонализации Грама–Шмидта.
  - 2.5. Модифицировать

$$\tilde{\mathbf{Y}}^{(t+1)} = \mathbf{B}^{(t)} \mathbf{D} \left(\mathbf{B}^{(t)}\right)^T \tilde{\mathbf{Y}}^{(t)}.$$

В данной записи  $\mathbf{D}$  — единичная матрица с  $d_{11} = 1$ ,  $d_{ii} = 0$ . Использование классического алгоритма ортогонализации Грама–Шмидта является принципиальным, так как важно при ортонормализации начинать с первого вектора.

## 6 Заключение

В работе известные результаты дополнены деталями, углубляющими понимание особенностей PPCA, имеющими прикладное значение, исправляющими выявленные ошибки в доступных публикациях. Предложены и обоснованы два метода реконструкции характеристик главных компонент. Они позволяют воспользоваться результатами предельного PPCA и проводить полный PCA, особо эффективный при малых значениях  $k$  и больших значениях  $d$ .

## Литература

1. Jolliffe I. T. Principal component analysis. — New York, NY, USA: Springer, 2002. 487 p.
2. Gabriel K. R. Least squares approximation of matrices by additive and multiplicative models // J. Roy. Stat. Soc. B, 1978. Vol. 40. No. 2. P. 186–196.
3. Tipping M. E., Bishop C. M. Mixtures of probabilistic principal component analyzers // Neural Comput., 1999. Vol. 11. No. 2. P. 443–482.
4. Rubin D. B., Thayer D. T. EM algorithms for ML factor analysis // Psychometrika, 1982. Vol. 47. No. 1. P. 69–76.



5. Roweis S. EM algorithm for PCA and SPCA // *Advances in neural information processing systems* / Eds. M. I. Jordan, M.J. Kearns, S. A. Solla. — MIT Press, 1997. Vol. 10. P. 626–632.
6. Ahn J. H., Oh J. H. A constrained EM algorithm for principal component // *Neural Comput.*, 2003. Vol. 15. P. 57–65.
7. Wang H., Hu Z. An unified EM algorithm for PCA and KPCA // *Neurocomputing*, 2007. Vol. 71. P. 459–462.
8. Choi S., Ahn J. H., Cichocki A. Constrained projection approximation algorithms for principal component analysis // *Neural Process. Lett.*, 2006. Vol. 24. P. 53–65.

*Поступила в редакцию 21.11.17*

---



---

## PRINCIPAL AXES RECONSTRUCTION

M. P. Krivenko

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** Principal component analysis (PCA) is a widely used technique for processing, compressing, and visualizing of data. New possibilities are opened by probabilistic PCA (PPCA), realized within the maximum likelihood principle for a Gaussian model with latent variables. Within the framework of PPCA, data processing algorithms have appeared, aimed at reducing the dimensionality of data and providing the transition to the space of the main components, but not explicitly giving the characteristics of the main components. The article is devoted to details that deepen the understanding of the features of PPCA and corrections of the errors revealed in publications. Two methods for reconstructing the characteristics of principal components are proposed and substantiated. One of them is based on recalculation of the covariance matrix in the formed space of main components. The other method consists in successively repeating the same steps: identifying the first main component and excluding it from data analysis.

**Keywords:** principal component analysis; EM-algorithm; reconstruction of axes and dispersion

**DOI:** 10.14357/19922264180109

### References

1. Jolliffe, I. T. 2002. *Principal component analysis*. New York, NY: Springer. 487 p.
2. Gabriel, K. R. 1978. Least squares approximation of matrices by additive and multiplicative models. *J. Roy. Stat. Soc. B* 40(2):186–196.
3. Tipping, M. E., and C. M. Bishop. 1999. Mixtures of probabilistic principal component analyzers. *Neural Comput.* 11(2):443–482.
4. Rubin, D. B., and D. T. Thayer. 1982. EM algorithms for ML factor analysis. *Psychometrika* 47(1):69–76.
5. Roweis, S. 1997. EM algorithm for PCA and SPCA. *Advances in neural information processing systems*. Eds. M. I. Jordan, M.J. Kearns, and S. A. Solla. MIT Press. 10:626–632.
6. Ahn, J. H., and J. H. Oh. 2003. A constrained EM algorithm for principal component. *Neural Comput.* 15:57–65.
7. Wang, H., and Z. Hu. 2007. An unified EM algorithm for PCA and KPCA. *Neurocomputing* 71:459–462.
8. Choi, S., J. H. Ahn, and A. Cichocki. 2006. Constrained projection approximation algorithms for principal component analysis. *Neural Process. Lett.* 24:53–65.

*Received November 21, 2017*

### Contributor

**Krivenko Michail P.** (b. 1946) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; mkrivenko@ipiran.ru

# СИСТЕМА СИТУАЦИОННОГО УПРАВЛЕНИЯ КАК МУЛЬТИСЕРВИСНАЯ ТЕХНОЛОГИЯ В ОБЛАЧНОЙ СРЕДЕ

А. А. Зацаринный<sup>1</sup>, А. П. Сучков<sup>2</sup>

**Аннотация:** Рассматриваются системотехнические подходы к созданию унифицированной системы управления в среде облачных технологий в виде совокупности сервисов. Предложенный подход на основе процессной пятистадийной модели позволяет определить общий вид (прототип) унифицированной системы ситуационного управления с учетом всех видов внешних и внутренних информационных взаимодействий в составе иерархической системы управления. Показано, что унифицированная система должна обладать средствами настройки (локализации) на конкретную область применения. Предложен перечень сервисов локализации и сервисов обеспечения основных функций системы управления. Ожидается, что такие сервисы будут востребованы в широком круге организационных систем, что должно привести к существенному уменьшению затрат на разработку и внедрение, обеспечить возможность реализации единой информационной, программной и технической политики в области межведомственного и внутриведомственного электронного взаимодействия.

**Ключевые слова:** унифицированная система ситуационного управления; облачный сервис; процессы локализации

**DOI:** 10.14357/19922264180110

## 1 Введение

В настоящее время подавляющее большинство организационных систем (ведомства, корпорации, предприятия в государственном секторе, бизнес-компании, банковский и торговый секторы экономики и т. п.) при реализации стратегических целей связывают свое развитие с созданием и эффективным применением информационных систем и технологий. При этом требования к ним сегодня неизмеримо возросли: речь идет не о компьютеризации отдельных служб или компании в целом; на первый план выходят системы с технологиями, реализующими функции поддержки принятия управляющих решений как в рамках текущих задач организационной системы, так и задач среднесрочного и перспективного планирования на основе прогнозирования. Более того, возросла актуальность задач информационного взаимодействия различных организационных систем между собой на основе единых регламентов взаимодействия. Такие задачи поставлены в рамках создания Системы распределенных ситуационных центров [1].

Это определяет тенденцию возрастания спроса на разработку и внедрение автоматизированных систем управления для широкого спектра органи-

заций. Очевидно, что затраты на разработку существенно снижаются при использовании доступных готовых типовых решений (инфраструктурных и программных), настраиваемых на конкретные приложения.

Предпосылками для создания системы управления как общедоступной мультисервисной технологии в облачной среде являются существующие модели обслуживания IaaS (Infrastructure as a Service — инфраструктура как услуга) и DaaS (Desktop as a Service), позволяющие реализовать определенную инфраструктуру (серверы, рабочие места, хранилища, базы данных (БД)) и специализированные рабочие места с прикладным программным обеспечением, реализующие необходимый функционал.

Услуга IaaS предоставляет возможность конструировать информационно-технологическую инфраструктуру и управлять ею. Самые известные IaaS-решения: Amazon CloudFormation [2], Google Compute Engine [3], Windows Azure [4].

При предоставлении услуги DaaS клиенты получают полностью готовые к работе (под ключ) специализированные виртуальные рабочие места, которые каждый пользователь имеет возможность дополнительно настраивать под свои задачи и получать доступ не к отдельной программе, а к необ-

<sup>1</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, AZatsarinny@ipiran.ru

<sup>2</sup>Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, ASuchkov@ipiran.ru

ходимому для полноценной работы программному комплексу. Такой подход используется в Amazon WorkSpaces [5] и Oracle Virtual Desktop Infrastructure [6].

Реализацию мультисервисной технологии можно осуществлять в гибридном облаке, сочетающем частные (для организаций) и общественные (для иерархической системы управления) реализации.

Специализированный облачный сервис в виде системы управления можно назвать CSaaS (Control System as a Service). Некоторый положительный опыт успешного применения облачных систем управления уже накоплен в области розничной, оптовой и интернет-торговли, гостиничного бизнеса, медицинских услуг (см., например, [7]).

В статье рассматриваются системотехнические подходы к созданию унифицированной системы управления в среде облачных технологий в виде совокупности сервисов. Ожидается, что такие сервисы будут востребованы в широком круге организационных систем, что должно привести к существенному уменьшению затрат на разработку и внедрение, обеспечить возможность реализации единой информационной, программной и технической политики в области межведомственного и внутриведомственного электронного взаимодействия.

## 2 Процессная модель унифицированной системы управления

В [8] обоснована процессная пятистадийная модель обобщенной системы управления, которая апробирована в ряде приложений [9]. Суть данной модели состоит в представлении системы управления в виде пяти групп процессов, взаимодействующих с внешней средой и между собой: целеполагание, мониторинг, анализ, решение и действие. При этом, как показано на рис. 1, выход каждого процесса является входом другого, что обеспечивает непрерывность управления.

Первая группа процессов обслуживает такое свойство системы управления, как целесообразность, и позволяет поддерживать формализованное представление целевых показателей (ЦП), устанавливать количественные и временные критерии их достижения (целеполагание и планирование). Сформированная система целей определяет совокупность параметров объектов мониторинга, подлежащих постоянному наблюдению, а также формы отчетности и других служебных документов.

Взаимодействие системы управления с внешней средой обслуживает ряд процессов сбора ин-

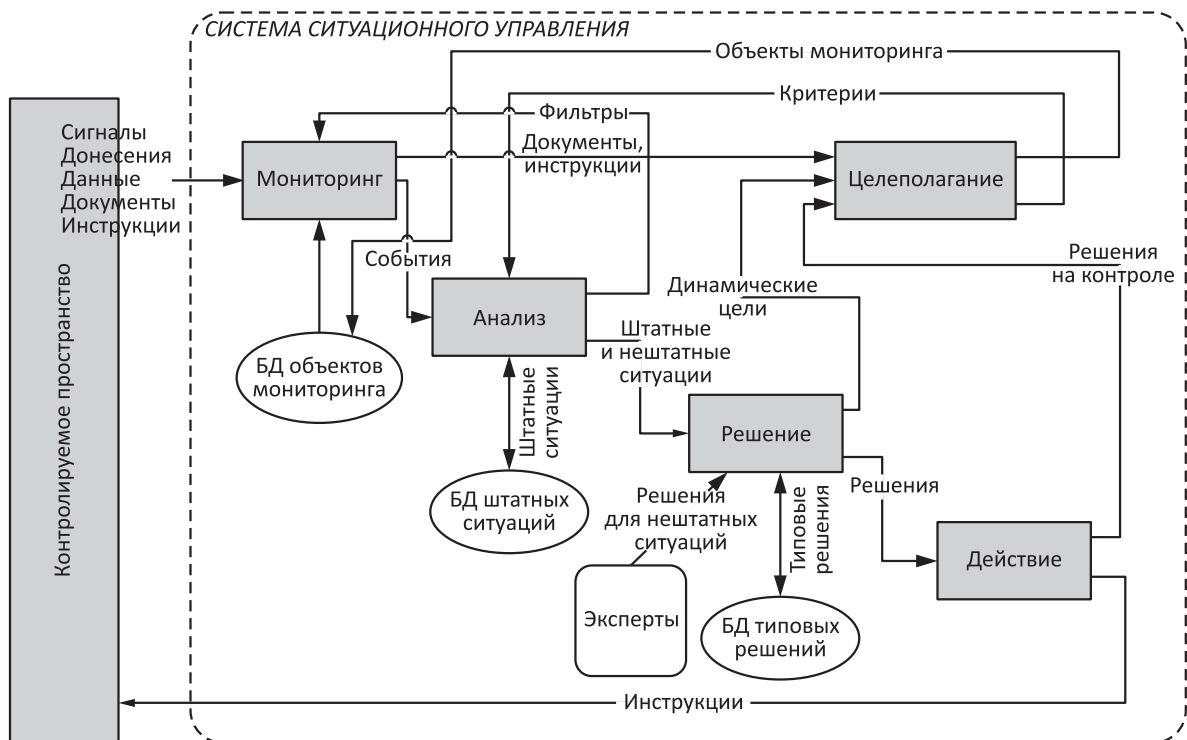


Рис. 1 Функциональная модель ситуационной системы управления

формации о состоянии контролируемых объектов, включая состояние ЦП, взаимодействие с сопряженными системами, а также сбор и анализ общедоступных данных по тематике предметной области. Состав данных мониторинга может быть достаточно разнообразен — сигналы сенсоров, отчеты и донесения нижестоящих и взаимодействующих систем управления, инструкции вышестоящих органов, разнородные данные общедоступных источников. В процессе мониторинга фиксируются события — изменения состояния контролируемых объектов внешней среды.

Третья группа процессов, получая на входе событийную информацию, осуществляет анализ и распознавание складывающихся пространственно-временных конфигураций событий (ситуации), имеется в виду пространство взаимосвязанных объектов мониторинга, изменяющихся во времени. При этом выявляются значимые ситуации, требующие быстрого реагирования, а также определяется тип ситуации (штатная, нештатная), анализируется существующий опыт нормализации ситуаций и вырабатываются альтернативы решений по управлению. Важнейшим компонентом третьей группы является анализ текущего состояния ЦП системы управления (планов).

На следующей стадии осуществляется поддержка процессов принятия решений, в общем случае это процесс многокритериального выбора (МКВ) среди альтернативных вариантов нормализации ситуации с учетом различных ресурсных ограничений и результатов сценарного прогнозирования последствий возможных управляющих воздействий. В случае штатной ситуации осуществляется адаптация типовых решений к складывающейся обстановке.

Пятая стадия реализует процессы выполнения выработанного решения путем применения управляющих воздействий, исходя из имеющихся ресурсов, а также контроль исполнения (путем установления динамических целей для осуществления контроля исполнения решения).

### 3 Общий вид (прототип) унифицированной системы ситуационного управления

Предлагаемая функциональная модель системы ситуационного управления может быть основой для описания унифицированной мультисервисной технологии высокой доступности. Для формирования полного облика такой технологии необходимо также учесть процессы взаимодействия в ие-

рархической многоуровневой системе управления, так как обычно любая система управления может обладать как внутренними подсистемами управления, например подразделениями, так и внешними подсистемами, например территориально распределенными системами управления.

Информационные взаимодействия, происходящие в иерархической многоуровневой системе ситуационного управления, подразделяются на внешние и внутренние (рис. 2). Внешние взаимодействия с окружающей средой происходят с помощью сенсорных подсистем, которые могут осуществлять прием сигналов, донесений, различных видов структурированных и неструктурированных данных, а также инструкций от вышестоящих органов, осуществляющих целеполагание и другое регулирование деятельности системы управления на основе нормативно-правовых документов.

В структуре информационных взаимодействий внутри иерархической системы управления можно выделить три вида данных: команды, отчеты и сведения.

Команды представляют собой управляющие воздействия в виде, как правило, формализованных сообщений (сигналов, целеуказаний, распоряжений, инструкций), передаваемых по иерархии подчиненности «сверху вниз» — от вышестоящих к нижестоящим подсистемам. С помощью команд выполняются действия по исполнению принятых решений, включая установление новых целей, определение сроков, ответственных исполнителей и выделение требуемых ресурсов. Такой функционал обычно реализуется средствами формализованного электронного документооборота (ФЭД).

Отчеты передаются от нижестоящих к вышестоящим подсистемам в виде формализованных донесений о реализации поставленных целей, которые находятся на контроле вышестоящего органа управления. Сюда могут включаться доклады о готовности и работоспособности, а также результаты деятельности по реализации решений.

Сведения представляют собой формализованные и неформализованные сообщения, передаваемые с целью информирования и координации деятельности взаимодействующих подсистем, а также обмена данными о результатах аналитики (прогнозирование, идентификация ситуаций, выявленные тренды, аномалии, артефакты). В [10] рассмотрены вопросы организации обмена аналитическими данными различных видов и обоснованы способы их формализации.

Таким образом, предложенный подход на основе процессной пятистадийной модели позволяет определить общий вид (прототип) унифицированной системы ситуационного управления с учетом



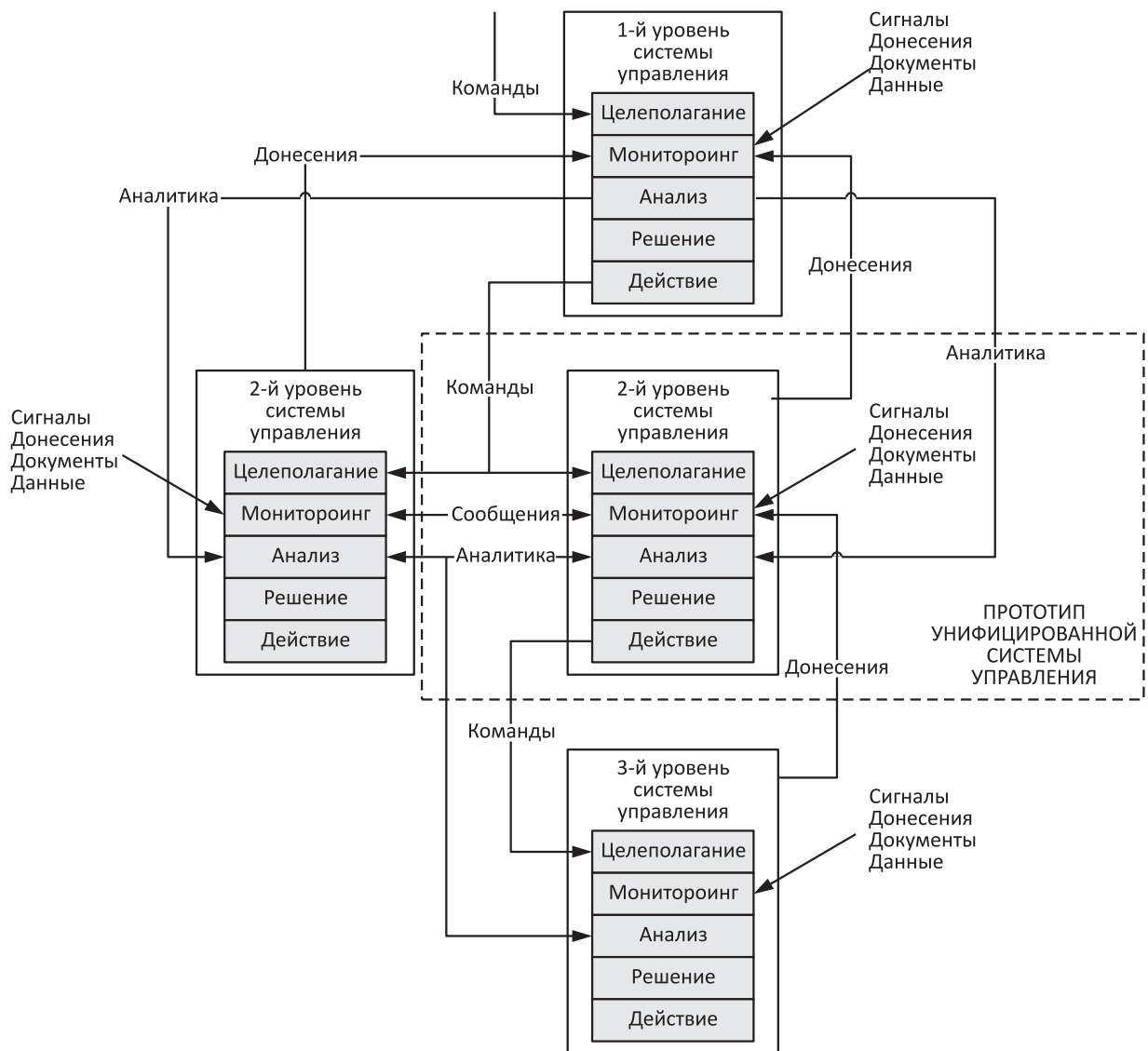


Рис. 2 Взаимодействие подсистем иерархической трехуровневой системы управления

всех видов внешних и внутренних информационных взаимодействий в составе иерархической системы управления (см. рис. 2).

#### 4 Прикладные вопросы локализации прототипа

Унифицированная система ситуационного управления высокой доступности в облачной среде должна обладать средствами настройки (локализации) на конкретную область применения. Следовательно, в ее структуре должны быть универсальные (неизменные для любого приложения или автоматически настраиваемые) компоненты и специфические (настраиваемые) компоненты.

К настраиваемым компонентам относятся: структура БД, инструментальные программные средства анализа неструктурированных и структурированных данных, программные средства адаптеров входных данных и прочие средства специального программного обеспечения (СПО), ФЭД.

Все эти компоненты опираются на проблемно-ориентированную модель данных, включающую в себя описание сущностей предметной области и их взаимосвязей, информационно-лингвистическое обеспечение (ИЛО) системы, формы служебных документов. Информационная модель формируется в процессе первоначальной локализации системы в основном по данным стадии целеполагания, на которой определяются ЦП системы управления и связанные с ними объекты мони-

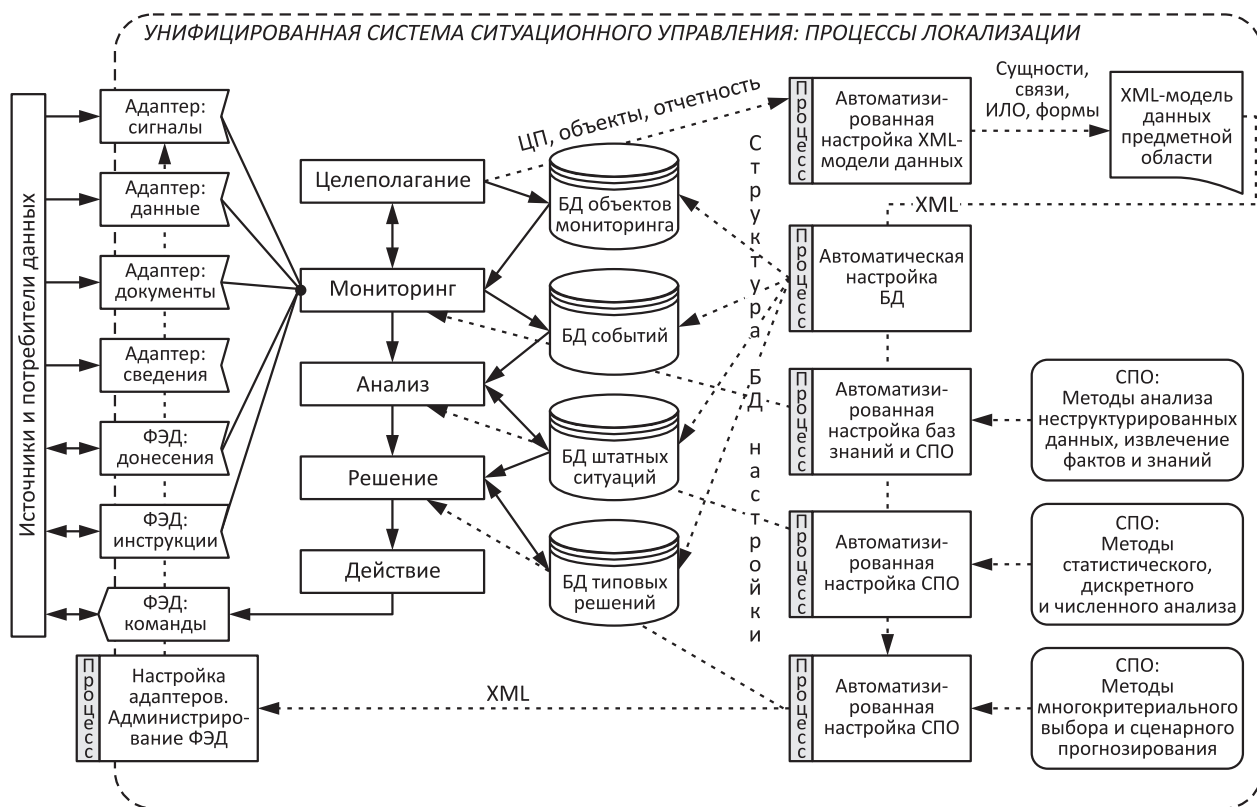


Рис. 3 Процессы настройки унифицированной системы ситуационного управления

торинга, формы отчетности и другие служебные документы.

Настройка унифицированной системы, а также ее модификация должны осуществляться с помощью ряда процессов, позволяющих сформировать информационную модель предметной области и на основе адаптированной модели данных осуществить локализацию системы в облачной среде (рис. 3).

В [8] систематизированы **основные** процессы управления: целеполагание, мониторинг, анализ данных мониторинга, поддержка процессов принятия решений в системе управления, реализация принятых решений, **вспомогательные** внутренние процессы управления, а также **обеспечивающие** процессы технологического управления, реализующие функции поддержания инфраструктуры: управление информационными ресурсами, организация сетевого взаимодействия, обеспечение информационной безопасности и управление доступом к ресурсам.

Все эти процессы подлежат определенной настройке, в статье будут рассмотрены процессы локализации основных и вспомогательных процессов управления.

1. Процесс автоматизированной настройки модели данных в формате XML или JSON.

К сожалению, в настоящее время большинство разработок в области прикладной информатики в нашей стране сопровождается созданием все новых и новых информационных моделей.

Общегосударственной системы обмена данными не существует, такие проекты, как, например, СМЭВ (Система межведомственного электронного взаимодействия) [11], не подкрепленные едиными моделями данных и документов и системами их ведения, столкнулись с большими трудностями по объемам перепрограммирования. Вместе с тем имеется положительный мировой опыт создания подобных моделей, например NIEM (National Information Exchange Model) — Национальная информационная XML-модель обмена данными в государственных органах США [12], в рамках которой разработаны и ведутся XML-модели многих направлений государственной деятельности, организованные в виде доменов, например правоохранительная и социальная сфера, транспорт, безопасность и др.

Используя такую модель, можно было бы собирать конкретные прикладные модели, как из «кубиков». В данном же случае необходима кропотливая работа прикладных аналитиков и использование имеющейся мозаики конкретных разработок и частично сформированного ИЛО государственных, ведомственных и корпоративных классификаторов.

С другой стороны, осуществление таких работ по единым технологиям может позволить последовательно осуществлять сборку единой модели. С этой целью необходимо предусмотреть сервисы формирования и ведения такой модели.

Основными разделами такой модели, в общем случае представляющей собой семантическую сеть, могут являться:

- классы сущностей, позволяющие формировать иерархические структуры (обеспечивающие наследование свойств в иерархии);
- сущности:
  - внешние объекты мониторинга (объекты, задаваемые целями управления);
  - внутренние объекты мониторинга (свои силы, материальные, технические, финансовые и другие средства);
  - противодействующие ресурсы (силы, средства);
  - элементы окружающей среды (природные, техногенные, социальные, политические и экономические факторы контролируемого пространства):
    - \* ИЛО (классификаторы, справочники, словари терминов);
    - \* система служебной документации.

2. Автоматическая настройка БД — формирование таблиц данных, входных и выходных форм при первичной локализации может осуществляться в автоматическом режиме с использованием специальных программных процедур и стандартных средств систем управления БД. Сложнее обстоит дело с корректировкой существующей структуры данных, где уже требуется применение нестандартных информационных технологий и зачастую вмешательство оператора. Это один из центральных вопросов поддержания жизнеспособности системы и, в свою очередь, один из сложнейших прикладных вопросов информатики.

3. Автоматизированная настройка баз знаний и СПО, реализующего методы анализа неструктурированных данных, извлечение фактов и знаний.

Основу процесса извлечения структурированной информации из неструктурированного текста составляет определение и идентификация сущностей из текста на естественном языке и выявление связей между этими сущностями. Данная проблема актуальна на стадии мониторинга данных из различных источников (СМИ, новостные ленты, социальные сети) и позволяет решить ряд прикладных задач в интересах системы управления: выявление новых сведений об объектах мониторинга, оценка социальных, политических и других последствий принимаемых решений, выявление тенденций и аномалий в развитии обстановки. Обычно такие методы анализа опираются на формализованные в том или ином виде знания о предметной области и специализированные методы обработки информации, реализованные в виде наукоемкого программного обеспечения [13, 14]. Как правило, такие программные продукты обладают инструментальными средствами настройки на конкретное приложение.

4. Автоматизированная настройка СПО, реализующего методы статистического, дискретного и численного анализа.

Целью анализа складывающейся обстановки является выявление штатных и нештатных ситуаций, декомпозиция нештатных ситуаций (упрощение до возможности применения типовых решений), оценка возможности применения типовых решений, выработка альтернатив решений.

Так как информационная модель контролируемого пространства формируется в виде семантической сети и отражает состояние контролируемых объектов и их взаимосвязей, для анализа данных применяются методы дискретной математики, связанные с теорией графов, математической логикой и лингвистическим анализом:

- идентификация и регистрация объектов, слияние подсетей;
- поиск подобных пространственно-временных конфигураций методами теории графов (изоморфизм и изоморфное вложение графов), декомпозиция конфигураций;

- логические выводы (поиск решения) на семантической сети;
- поиск прямых и ассоциативных связей (путей на графе);
- расчет интегральных показателей и ЦП на графах.

Декомпозиция ситуаций на более простые составляющие может позволить сформировать множество альтернативных решений. Например, если имеется несколько ЦП, то задача максимизации одного ЦП при фиксированных остальных может быть отнесена к типовой и решаемой. Как правило, таких альтернатив может быть бесконечное число (образуя так называемое множество Парето) и стоит вопрос о выборе одного решения, который должен быть решен на следующей стадии процесса управления.

Специальное программное обеспечение, реализующее указанные методы анализа, настраивается на сформированные БД и структуры данных, задаваемые информационной моделью предметной области.

#### 5. Автоматизированная настройка СПО, реализующего методы МКВ и сценарного прогнозирования.

Как правило, для выбора альтернативы из множества вариантов решений в условиях ограничения на ресурсы применяются технологии, обеспечивающие снижение размерности признакового пространства и построение интегрального показателя качества. При этом агрегируются исходные признаки с использованием знаний эксперта и/или предпочтений лица, принимающего решение (ЛПР) [15]. Например, можно использовать алгоритм сужения множества Парето на основе нечеткой информации об отношениях предпочтения ЛПР [16].

Оценка качества выбранных альтернатив решений может осуществляться с использованием метода сценарного прогнозирования, который позволяет учесть управляющие воздействия при прогнозировании развития обстановки. Для анализа обстановки и поддержки процессов принятия решений применяются математические методы статистического анализа:

- анализ временных рядов, характеризующих изменения количественных и качественных атрибутов узлов и связей (анализ трендов, сезонных колебаний, тенденций и аномалий);

- прогнозирование изменения параметров с учетом выявленных трендов и анализируемых сценариев развития обстановки;
- динамическое моделирование ситуаций;
- статистическая оценка количественных и качественных характеристик потоков событий.

После принятия решения формируется динамическая цель, которая должна быть конкретной, количественно измеримой, достижимой, обеспеченной ресурсами и привязанной к точке/интервалу времени.

Настройка СПО, реализующего указанные технологии, осуществляется также в соответствии со структурой данных, задаваемой информационной моделью предметной области.

#### 6. Настройка адаптеров. Администрирование ФЭД.

Система управления должна обладать возможностями настройки на необходимые источники данных (технические средства, информационные системы, СМИ, новостные сайты и др.) с помощью системы специализированных адаптеров, позволяющих преобразовывать получаемую информацию в соответствии с информационной моделью системы. Получение информации может осуществляться по регламенту, по запросу и по событию, поэтому технологии нормализации (преобразования в требуемые форматы) данных и запросов должны обладать возможностью настройки пользовательской среды поддержки процесса интеграции, реализованной, например, в виде средств создания персональных семантических цифровых библиотек как инструмента обработки и интеграции данных [17]. При отсутствии автоматизированного взаимодействия с источником данных адаптеры должны обеспечивать режим ручного ввода информации.

Существуют методы виртуальной интеграции разнородных источников, основанные на XML-модели данных, когда с использованием декларативных языков локальные источники единообразно отображаются на глобальную схему в терминах XML (информационную модель). При этом в системе виртуальной интеграции решаются вопросы отображения схем, оптимизации, декомпозиции и выполнения запросов [18].

Формализованный электронный документооборот предназначен для обмена структурированными данными внутри системы управления



(команды, доклады, сообщения, данные аналитики). Администрирование ФЭД подразумевает системные настройки комплекса, дизайн отчетов и других служебных документов, создание хранилищ на сервере БД, управление учетными записями и правами доступа, настройки протоколирования и мониторинга.

## 5 Унифицированная система ситуационного управления как мультисервисная технология

Совокупность процессов локализации и функционирования унифицированной системы ситуационного управления реализуется совокупностью взаимосвязанных облачных сервисов (рис. 4), позволяющих как интерактивную, так и их событийную активацию (рис. 5).

Сервисы, поддерживающие единую XML-модель, целесообразно реализовать в виде глобальной

портальной технологии, позволяющей формировать интернет-сообществу рассматриваемой предметной области единую информационную модель по образцу NIEM.

## 6 Использование унифицированной мультисервисной технологии для создания предметно-ориентированной системы ситуационного управления (например, ситуационного центра)

Реализация мультисервисной технологии предполагает наличие у провайдера типовой инфраструктуры системы управления, представляемой

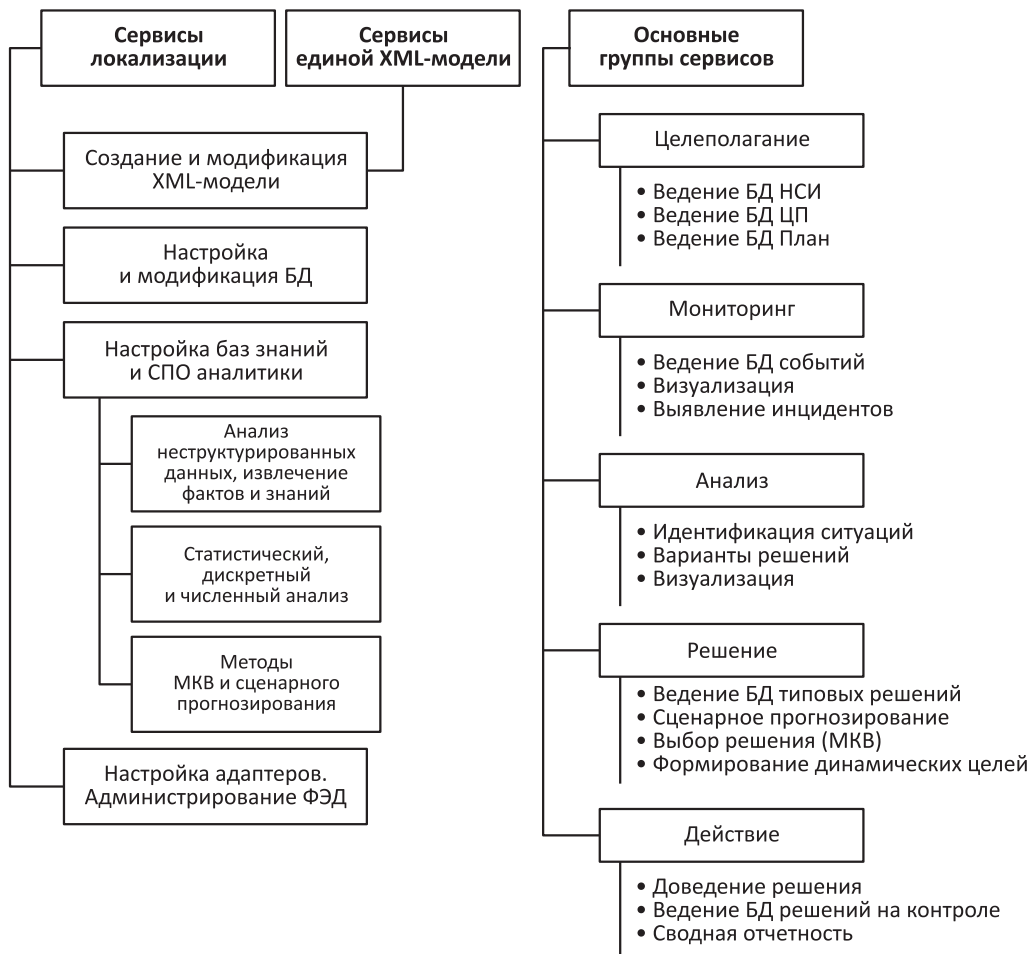
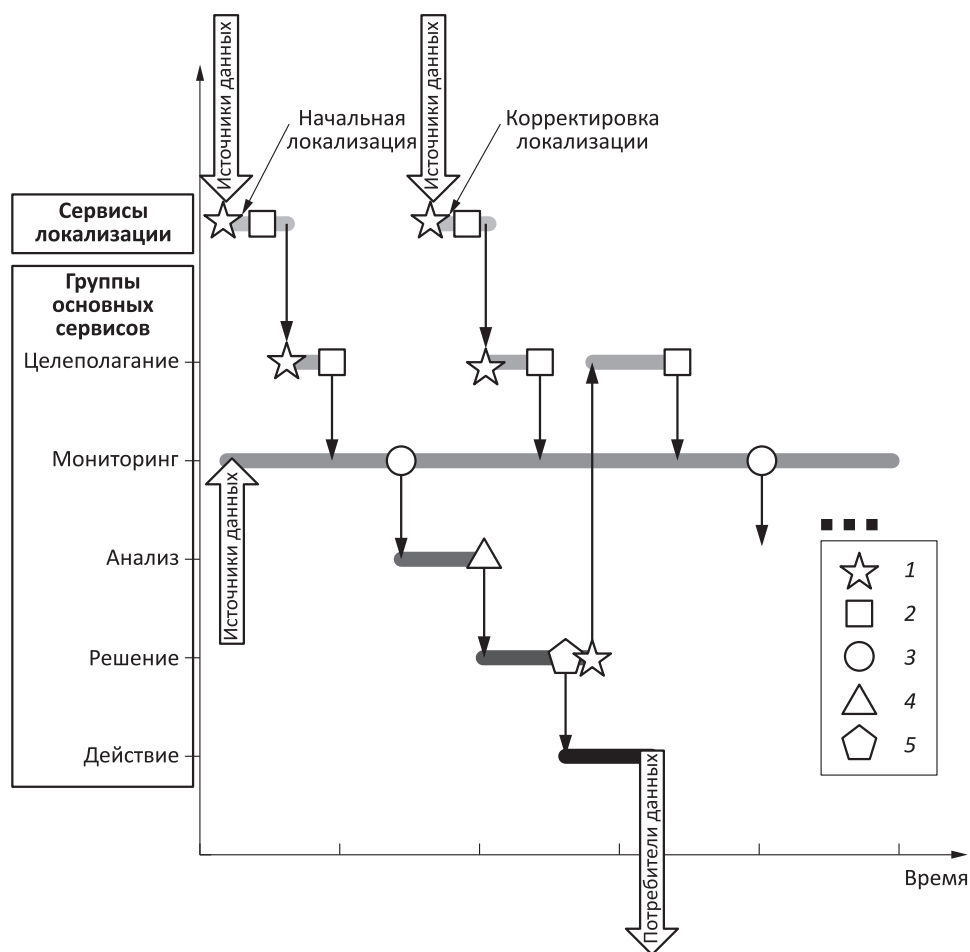


Рис. 4 Структура мультисервисной системы управления (НСИ — нормативно-справочная информация)



**Рис. 5** Интерактивная и событийная активация сервисов: 1 — изменение состава целей; 2 — изменение состава объектов мониторинга; 3 — возникновение ситуации; 4 — формирование альтернатив решений; 5 — принятие решения

пользователю в модели обслуживания IaaS, инструментального и прикладного программного обеспечения, размещаемого на специализированных рабочих местах, доступ пользователя к которым осуществляется в режиме «тонкого клиента».

С точки зрения реализации функционала облачной системы управления в рамках услуги DaaS целесообразно создание специализированных рабочих мест в соответствии с организационной структурой ситуационного центра [19]:

- сегмент руководства — ЛПР, реализующий группу сервисов целеполагания и решения;
- сегмент мониторинга состояния контролируемых объектов и окружающей среды (оперативная дежурная служба), реализующий группу сервисов мониторинга;
- сегмент ситуационного анализа и систематизации информации, реализующий группу сервисов анализа и действия;

– сегмент администрирования, реализующий группу сервисов локализации.

После создания и оснащения инфраструктуры запускаются интерактивные процессы локализации: осуществляется настройка информационной модели предметной области и компонентов системы под нужды заказчика (адаптеры, документооборот, аналитические приложения). Активация системы осуществляется с запуском сервисов мониторинга контролируемого пространства системы управления.

## Литература

1. Зацаринный А. А. О повышении эффективности информационно-аналитической поддержки принятия стратегических решений в органах государственной власти // Межотраслевая информационная служба, 2015. № 1(170). С. 11–22.

2. AWS CloudFormation. <https://aws.amazon.com/ru/cloudformation>.
3. Compute Engine: Scalable, High-Performance Virtual Machines. <https://cloud.google.com/compute>.
4. Глобальность. Надежность. Гибридность: Облако на ваших условиях. <https://azure.microsoft.com/ru-ru>.
5. Amazon WorkSpaces: Полностью управляемые и защищенные виртуальные облачные рабочие столы на AWS. <https://aws.amazon.com/ru/workspaces>.
6. Oracle Virtual Desktop Infrastructure. <http://www.oracle.com/technetwork/server-storage/virtualdesktop/overview/index.html>.
7. Облачная система управления бизнесом. [http://cloud-automation.ru/management\\_automation](http://cloud-automation.ru/management_automation).
8. Сучков А. П. Полнофункциональный процессный подход к реализации систем ситуационного управления // Системы и средства информатики, 2017. Т. 27. № 1. С. 85–99.
9. Федеральный исследовательский центр «Информатика и управление» Российской академии наук: Основные направления исследований. <http://www.frccsc.ru/expertise>.
10. Сучков А. П. Некоторые подходы к интеграции аналитических данных существующих и перспективных систем поддержки принятия решений // Системы и средства информатики, 2015. Т. 25. № 3. С. 201–211.
11. Электронное правительство: Госуслуги. <http://smev.gosuslugi.ru/portal>.
12. About NIEM (National Information Exchange Model). <https://www.niem.gov/about-niem>.
13. Осипов Г. С. Приобретение знаний интеллектуальными системами. — М.: Наука. 1997. 112 с.
14. Быстров И. И., Тарасов Б. В., Хорошилов А. А., Радоманов С. И. Основы применения онтологии и компьютерной лингвистики при проектировании перспективных автоматизированных информационных систем // Системы и средства информатики, 2015. Т. 25. № 4. С. 128–149.
15. Петровский А. Б., Лобанов В. Н. Многокритериальный выбор в пространстве признаков большой размерности: мультиметодная технология ПАКС-М // Искусственный интеллект и принятие решений, 2014. № 3. С. 92–104.
16. Басков О. В. Сужение множества Парето на основе нечеткой информации об отношении предпочтения ЛППР // Искусственный интеллект и принятие решений, 2014. № 1. С. 57–65.
17. Атаева О. М., Серебряков В. А. Основные понятия формальной модели семантических библиотек и формализация процессов интеграции в ней // Программные продукты и системы, 2015. № 4(112). С. 180–187.
18. Антипин К. В., Фомичев А. В., Гринев М. Н. и др. Оперативная интеграция данных на основе XML: системная архитектура BizQuery // Труды Института системного программирования РАН, 2004. Т. 5. С. 157–174.
19. Зацаринный А. А., Сучков А. П., Козлов С. В. Особенности проектирования и функционирования ситуационных центров // Системы высокой доступности, 2012. Т. 8. № 1. С. 12–22.

Поступила в редакцию 17.04.17

## THE SITUATIONAL MANAGEMENT SYSTEM AS A MULTISERVICE TECHNOLOGY IN THE CLOUD

A. A. Zatsarinny and A. P. Suchkov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The article discusses the approaches to creation of a unified management system as a set of services in cloud computing. The proposed approach is based on the five-stage processing model. It allows defining the general form (prototype) of a unified system of situational management, taking into account all types of internal and external information interactions in the structure of a hierarchical control system. It is shown that a unified system must have a means of configuration (localization) on a specific application field. The proposed list of services and localization services provides the main functions of the control system. It is expected that such services will be in demand in a wide range of organizational systems, which should lead to a significant reduction in the cost of development and implementation and ensure the possibility of implementation of uniform information, program, and technical policy in the field of interagency and interdepartmental electronic interaction.

**Keywords:** unified system of situational management; cloud service; process of localization

**DOI:** 10.14357/19922264180110

## References

- Zatsarinny, A. A. 2015. O povyshenii effektivnosti informatsionno-analiticheskoy podderzhki prinyatiya strategicheskikh resheniy v organakh gosudarstvennoy vlasti [Some problems of information-analytical support for strategic decision-making in government departments]. *Mezhotraslevaya informatsionnaya sluzhba* [Intersectoral Information Service] 1(170):11–22.
- AWS CloudFormation. Available at: <https://aws.amazon.com/ru/cloudformation> (accessed October 25, 2017).
- Compute Engine: Scalable, High-Performance Virtual Machines. Available at: <https://cloud.google.com/compute> (accessed October 25, 2017).
- Global'nost'. Nadezhnost'. Gibridnost': Oblako na vashikh usloviyakh [The Global. Reliability. Hybridity: The cloud on your terms]. Available at: <https://azure.microsoft.com/ru-ru> (accessed October 25, 2017).
- Amazon WorkSpaces: Fully managed, secure virtual cloud desktops running on AWS. Available at: <https://aws.amazon.com/workspaces> (accessed October 25, 2017).
- Oracle Virtual Desktop Infrastructure. Available at: <http://www.oracle.com/technetwork/server-storage/virtualdesktop/overview/index.html> (accessed October 25, 2017).
- Oblachnaya sistema upravleniya biznesom [Cloud-based business management system]. Available at: [http://cloud-automation.ru/management\\_automation](http://cloud-automation.ru/management_automation) (accessed October 25, 2017).
- Suchkov, A. P. 2017. Polnofunktional'nyy protsessnyy podkhod k realizatsii sistem situatsionnogo upravleniya [A fully functional process-based approach to the implementation of systems of situational management]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(1):85–99.
- Federal'nyy issledovatel'skiy tsentr “Informatika i upravlenie” Rossiyskoy akademii nauk: Osnovnye napravleniya issledovaniy [Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences: Main research areas]. Available at: <http://www.frccsc.ru/expertise> (accessed October 25, 2017).
- Suchkov, A. P. 2015. Nekotorye podkhody k integratsii analiticheskikh dannykh sushchestvuyushchikh i perspektivnykh sistem podderzhki prinyatiya resheniy [Some approaches to the analytical data integration of the existing and future decision support systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(3):201–211.
- Elektronnoe pravitel'stvo: Gosuslugi [E-government: Public services] Available at: <http://smev.gosuslugi.ru/portal/> (accessed October 25, 2017).
- About NIEM (National Information Exchange Model). Available at: <https://www.niem.gov/about-niem> (accessed October 25, 2017).
- Osipov, G. S. 1997. *Priobretenie znaniy intellektual'nymi sistemami* [Knowledge acquisition intelligent systems]. Moscow: Nauka. 112 p.
- Bystrov, I. I., B. V. Tarasov, A. A. Khoroshilov, and S. I. Radomanov. 2015. Osnovy primeneniya ontologii i komp'yuternoy lingvistiki pri proektirovaniy perspektivnykh avtomatizirovannykh informatsionnykh sistem [The application basis of ontology and computational linguistics in the design of advanced automated information systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(4):128–149.
- Petrovsky, A. B., and V. N. Lobanov. 2015. Multi-criteria choice in the attribute space of large dimension: Multi-method technology PAKS-M. *Scientific Technical Information Processing* 42(5):76–86.
- Baskov, O. V. 2015. An algorithm for Pareto set reduction using fuzzy information on decision-maker's preference relation. *Scientific Technical Information Processing* 42(5):382–387. doi: 10.3103/S0147688215050020.
- Ataeva, O. M., and V. A. Serebryakov. 2015. Osnovnye ponyatiya formal'noy modeli semanticheskikh bibliotek i formalizatsiya protsessov integratsii v ney [The basic concepts of a semantic libraries formal model and its integration process formalization]. *Programmnyye produkty i sistemy* [Software Systems] 4(112):180–187.
- Antipin, K. V., A. V. Fomichev, M. N. Grinev, et al.. 2004. Operativnaya integratsiya dannykh na osnove XML: Sistemnaya arkhitektura BizQuery [Operational data integration based on XML: The system architecture of BizQuery]. *Proceedings of the Institute for System Programming of the RAS* 5:157–174.
- Zatsarinny, A. A., A. P. Suchkov, and S. V. Kozlov. 2012. Osobennosti proektirovaniya i funktsionirovaniya situatsionnykh tsentrov [The features of design and operation of situational centers]. *Sistemy vysokoy dostupnosti* [Highly Available Systems] 8(1):12–22.

Received April 17, 2017

## Contributors

**Zatsarinny Alexander A.** (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; AZatsarinny@ipiran.ru

**Suchkov Alexander P.** (b. 1954) — Doctor of Science in technology, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; ASuchkov@frccsc.ru



## ЭКСПЛОРАТИВНЫЙ ПАТЕНТНЫЙ ПОИСК\*

И. В. Соченков<sup>1</sup>, Д. В. Зубарев<sup>2</sup>, И. А. Тихомиров<sup>3</sup>

**Аннотация:** Предложен эффективный метод поиска тематически похожих документов. Показана его применимость для решения задач эксплоративного патентного поиска. Метод позволяет снизить трудоемкость и повысить качество решения аналитических задач, связанных с патентным поиском и анализом. В качестве признаков для представления текстовых документов используются как отдельные лексемы, так и словосочетания, выделяемые синтаксически. Это позволяет решить проблему экспоненциального роста размерности признакового пространства и дает возможность эффективной индексации больших массивов текстовой информации. Представлены результаты экспериментов по решению одной из задач экспертизы заявок на изобретения и полезные модели. Сделаны выводы о перспективности использования метода при решении других задач.

**Ключевые слова:** эксплоративный поиск; патентный поиск; тематическое моделирование; поиск тематически похожих документов; поисково-аналитические системы

**DOI:** 10.14357/19922264180111

### 1 Введение

Активное развитие мировой науки и технологий ведет к увеличению накопленных объемов знаний, представленных, в том числе, в виде патентной информации. Практически любое прикладное научное исследование начинается с патентного поиска, это позволяет избежать повторного открытия уже известных технических решений и обеспечивает правильный выбор пути в достижении поставленных целей и задач исследования.

Вид патентного поиска, направленный на получение наиболее актуальной и полной информации по заданной теме, называется эксплоративным (exploratory search) [1, 2]. Его особенность заключается в том, что аналитик ищет не конкретный факт или документ, а собирает сведения о заданной проблеме в целом, делая акцент на полноту и актуальность. Зачастую аналитик может не вполне ясно представлять себе предмет поиска, внося уточнения по ходу изучения собираемой информации.

Проведение качественного эксплоративного патентного поиска является весьма трудоемкой задачей, которая не может быть решена без использования современных патентных поисково-аналитических систем. Традиционно сложились три основных подхода к патентному поиску [3]:

(1) поиск по классам МПК (международная патентная классификация);

(2) поиск по ключевым словам;

(3) тематический поиск по заданному документу-образцу.

Первый подход требует от аналитика априорного знания структуры применяемого рубрикатора для формирования области поиска в соответствии с его информационной потребностью. Это требование зачастую оказывается излишне жестким, например редакция МПК-2016.01 содержит более 73 тыс. подгрупп — классов нижнего уровня, что затрудняет подбор правильных кодов МПК для поиска и превращает этот процесс в долгую рутинную процедуру.

Второй подход позволяет эксперту сформулировать информационную потребность в виде ключевых слов, которые должны встречаться в текстах описаний патентов или их метаданных: в заголовке, аннотации, списке авторов и т.д. Составление списка ключевых слов, исчерпывающе описывающих объект поиска, является весьма трудоемкой задачей, требующей привлечения эксперта в предметной области. Таким образом, поиск по ключевым словам, с одной стороны, повышает полноту поиска, с другой — снижает его точность.

Третий подход является альтернативой вышеприведенным и заключается в автоматическом поиске документов, тематически похожих на заданный эталонный документ, который содержит

\* Работа выполнена при поддержке РФФИ (проект 16-29-12929).

<sup>1</sup> Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук; Сколковский институт науки и технологий, I.Sochenkov@skoltech.ru

<sup>2</sup> Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук; Российский университет дружбы народов, zubarev@isa.ru

<sup>3</sup> Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, tih@isa.ru

описание научно-технической задачи или изобретения по исследуемой теме. Такой подход лишен недостатков первых двух, однако требует качественной и эффективной реализации.

В настоящей работе представлены разработанный авторами метод поиска тематически похожих документов в приложении к задаче эксплоративного патентного поиска и результаты его экспериментального исследования.

## 2 Метод поиска тематически похожих документов

Задача поиска тематически похожих документов по заданному эталону в коллекциях полнотекстовых документов решается методами, которые условно делятся на две группы:

- (1) вероятностное тематическое моделирование [4];
- (2) построение и сопоставление ключевых дескрипторов документов.

Для обеих групп обычно применяется векторное представление документа. В качестве признаков используются отдельные лексемы. В последнее время также наблюдается тенденция к использованию биграмм и триграмм слов (см., например, работы [5, 6]). Для эффективной реализации применяют предварительную индексацию информации [7, 8].

Разработанный авторами метод поиска тематически похожих документов относится ко второй группе и является развитием метода, описанного в работах [9, 10]. Отличительными особенностями этого метода являются:

- учет двусловных сочетаний (выделяемых синтаксическими методами);
- способ оценки значимости признаков, позволяющий эффективно выделять наиболее важные из них;
- эффективная реализация с помощью инвертированных индексов, позволяющая решать задачу в коллекциях, содержащих десятки и сотни миллионов документов.

Учет словосочетаний, выделяемых синтаксически, позволяет избежать проблемы роста размерности признакового пространства (в отличие от методов, учитывающих коллокации в качестве признаков). В результате такого подхода достигаются высокие показатели качества результатов поиска тематически похожих документов. Рассмотрим метод более подробно.

Пусть  $C = \{d_i\}_{i=1}^N$  — коллекция текстовых документов,  $e$  — эталонный документ из  $C$ :  $e = d_k$ ,

$1 \leq k \leq N$ . Задача поиска тематически похожих документов заключается в формировании списка документов, упорядоченного по убыванию значений некоторой функции оценок  $S(e, d_i) \geq \text{minSim}$ ,  $i = 1, \dots, N$ , где  $\text{minSim}$  — пороговое значение сходства.

Каждому документу  $d_i$ ,  $1 \leq i \leq N$ , сопоставим его текст — множество словоупотреблений:

$$T(d_i) = \{x_j^i\}_{j=1}^{M_i}.$$

Лексическим дескриптором (ЛД) будем называть:

- отдельную лексему;
- словосочетание, состоящее из нескольких грамматически связанных лексем.

Положим  $L = \{w_m\}_{m=1}^Z$  — конечный лексикон (словарь) естественных языков. Пусть  $2^{T(d_i)}$  — булеан множества  $T(d_i)$ ;  $y \in 2^{T(d_i)}$  — некоторая совокупность словоупотреблений. Отображение  $D(y, d_i) : 2^{T(d_i)} \rightarrow L$  представляет «нормализацию» словоупотреблений текста  $d_i$  (приведение к канонической, т.е. словарной, форме). Множество  $G(d_i) = \cup_{y \in 2^{T(d_i)}} D(y, d_i)$  представляет лексикон документа ( $G(d_i) \subseteq L$ ). Обозначим:

$$D^{-1}(w_m, d_i) = \{y \in 2^{T(d_i)} | D(y, d_i) = w_m\}.$$

Тогда словоупотребления текста  $T(d_i)$ , представляющие в нем все употребления заданного ЛД  $w_m$ , описываются следующим образом:

$$G(w_m, d_i) = \bigcup_{y \in D^{-1}(w_m, d_i)} y.$$

В этом случае вес ЛД  $w_m$  в тексте  $T(d_i)$  документа  $d_i$  определяется как мощность соответствующего множества словоупотреблений:  $|G(w_m, d_i)|$ .

Совокупный вес лексикона документа:

$$g(d_i) = \sum_{w \in G(d_i)} |G(w_m, d_i)|.$$

Аналогом частоты встречаемости ЛД в тексте является величина:

$$\text{ITF}(w_m, d_i) = \log_{1+g(d_i)}(1 + |G(w_m, d_i)|).$$

Для оценки значимости ЛД в коллекции применяется величина:

$$\text{IDF}(w_m, C) = \log_{|C|} \frac{|C|}{|\{d_i \in C | w_m \in G(d_i)\}|}.$$

Таким образом, итоговая оценка значимости ЛД в тексте документа является модификацией классической схемы «взвешивания» лексики TF-IDF:

$$R(w_m, d_i, C) = \text{ITF}(w_m, d_i) \text{IDF}(w_m, C).$$

Редкие ЛД, для которых  $|\{d_i \in C | w_m \in G(d_i)\}| \leq 2$ , считаются незначимыми: для них полагаем  $IDF(w_m, C) = 0$  по определению. Это позволяет исключить слова-опечатки, редкие словосочетания, записи адресов электронной почты, веб-адресов и т. п.

Рассмотрим функции оценок сходства, эффективно вычисляемые с применением инвертированных индексов.

Пусть  $wID(w_m) : L \rightarrow \mathbb{N}$  — функционал, взаимно однозначно сопоставляющий каждому ЛД его числовой идентификатор. Документ  $d_i$ ,  $1 \leq i \leq N$  описывается с помощью упорядоченного (по возрастанию  $wID(w_m)$ ) множества пар  $DI(d_i) = \langle \langle wID(d_i), ITF(w_m, d_i) | 1 \leq m \leq Z, ITF(w_m, d_i) > 0 \rangle \rangle$  — прямого индекса документа.

Пусть  $dID(d_i) : C \rightarrow \mathbb{N}$  — функционал, взаимно однозначно сопоставляющий каждому документу его числовой идентификатор. Для поиска тематически похожих документов применяется инвертированный индекс, сопоставляющий каждому ЛД  $w_m$ ,  $1 \leq m \leq Z$ , упорядоченное по возрастанию идентификатора документа  $dID(d_i)$  конечное множество пар  $RI(w_m, C) = \langle \langle dID(d_i), ITF(w_m, d_i) | 1 \leq i \leq N, ITF(w_m, d_i) > 0 \rangle \rangle$ .

Инвертированный индекс коллекции строится на основе прямых индексов всех документов.

Для поиска тематически похожих документов с помощью инвертированного индекса производится вычисление оценки сходства  $S(e, d_i)$ ,  $i = 1, \dots, N$ , следующим образом. В качестве запроса выступает наиболее значимая часть лексикона эталонного документа

$$G_1(e) \subseteq G(e) : \forall w \in G_1(e) \\ \forall w' \in G(e) \setminus G_1(e) \gamma(w, e, C) \geq \gamma(w', e, C),$$

где в качестве функции  $\gamma$ , определяющей значимость ЛД в эталонном документе, может быть взята функция  $R$  или  $IDF$ . Информация, необходимая для расчета  $ITF$  ЛД берется из прямого индекса эталонного документа. Величина  $IDF$  рассчитывается на основе частот встречаемости ЛД в коллекции, которая хранится в счетчиках, инкрементально обновляемых на этапе индексации коллекции. В результате выборки данных из инвертированного индекса формируется множество

$$SR(e, C) = \{RI(w, C) | w \in G_1(e)\}.$$

Далее каждому  $RI(w, C)$  сопоставляется множество троек (дескрипторов)  $\beta_0(RI(w, C)) = \{\langle wID(w), dID(d_i), ITF(w, d_i) | 1 \leq i \leq N, ITF(w, d_i) > 0 \rangle\}$ , упорядоченное, как и исходное

множество, по возрастанию  $dID(d_i)$ . Полагая далее  $\beta_1(SR(e, C)) = \{\beta(RI(w, C)) | w \in G_1(e)\}$ , с помощью однопроходной операции слияния упорядоченных множеств получим множество троек:

$$U(e, C) = \bigcup_{u \in \beta_1(SR(e, C))} u.$$

При слиянии тройки упорядочиваются естественным образом по возрастанию  $dID(d_i)$ , а при их совпадении — по возрастанию  $wID(w_m)$ . Таким образом, в  $U(e, C)$  последовательно друг за другом содержатся дескрипторы, характеризующие употребления ЛД эталонного документа в документах коллекции. Это дает возможность вычислить следующие оценки тематического сходства за один проход последовательности  $U(e, C)$ :

$$S_{Ham}(e, d_i) = 1 - \frac{\sum_{w \in G_1(e) \cap G(d_i)} |ITF(w, e) - ITF(w, d_i)| IDF(w, C)}{\sum_{w \in G_1(e)} R(w, e, C)};$$

$$S_{cos}(e, d_i) = \frac{\sum_{w \in G_1(e) \cap G(d_i)} R(w, e, C) R(w, d_i, C)^2}{\left( \sum_{w \in G_1(e)} R(w, e, C) \sum_{w \in G_1(e) \cap G(d_i)} R(w, d_i, C) \right)}.$$

Обе величины нормированы на 1. Оценка тематического сходства для неполных дубликатов документов дает значения, близкие к 1. Для оценки тематического сходства документов, которые не являются полными дубликатами, важно выбирать такой размер множества  $G_1(e)$ , чтобы в него входили наиболее значимые ЛД. В противном случае учитывается и нетематическая лексика, что приводит к попаданию в результаты поиска также документов по другим темам. Заметим, что при реализации описанного подхода на практике, как правило, не требуется, чтобы эталонный документ был предварительно проиндексирован в коллекцию, т.е. допустимо  $e \notin C$ .

### 3 Экспериментальное исследование метода на примере экспертизы патентных заявок

В процессе экспертизы патентных заявок основанием для отказа в выдаче патента могут быть аналогичные изобретения или полезные модели,

**Таблица 1** Исходные данные — коллекция патентов

Вид документа	Число документов	Ретроспектива
Патент на изобретение	580 тыс.	1994–2015 гг.
Патент на полезную модель	120 тыс.	2004–2015 гг.
ВСЕГО	700 тыс.	

**Таблица 2** Исходные данные — коллекция отклоненных заявок

Вид документа	Число документов	Общее число ссылок на предшествующие аналоги
Заявка на выдачу патента	158	237

**Таблица 3** Результаты оценки качества обнаружения документов-аналогов

Метод	nDCG	Полнота	$D_{10}$	$D_{20}$	Медиана
Базовый*	0,141	0,414	0,198	0,262	200
Предложенный	0,411	0,751	0,451	0,502	20

\*Значения оценок вычислены на основе используемого в ФИПС метода по состоянию на август 2016 г.

запатентованные ранее. Авторами статьи выдвинута следующая гипотеза: чем выше тематическое сходство заявки и некоторого ранее запатентованного изобретения, тем больше оснований у эксперта отказать в выдаче патента из-за отсутствия новизны или изобретательского уровня. Для проверки данной гипотезы проведены эксперименты, суть которых представлена далее.

В качестве набора исходных данных использована коллекция полных текстов описаний российских патентов на изобретения и полезные модели (табл. 1) и коллекция заявок на выдачу патента, по которым был получен отказ из-за отсутствия новизны или изобретательского уровня с указанием ссылки на ранее запатентованный аналог (предшествующие аналоги) из первой коллекции (табл. 2).

Задача поиска тематически похожих документов в эксперименте заключается в обнаружении для заданной заявки всех предшествующих аналогов и упорядочении поисковой выдачи по убыванию оценки тематического сходства.

Тестовый набор данных разделялся случайным образом на обучающую (126 заявок) и контрольную (32 заявки) выборки для применения трехкратной кросс-валидации. На обучающей выборке подобраны значения параметров алгоритма поиска тематически похожих документов, обеспечивающие на этой выборке наибольшее среднее значение метрики nDCG. Наиболее значимая часть лексикона эталонного документа  $G_1(e)$  строилась на основе оценки  $R$  с ограничением  $|G_1(e)| = a|G(e)|$ ,  $0 < a \leq 1$ . Простым перебором параметра  $a$  в диапазоне  $[0,15; 0,5]$  с шагом 0,05 было получено значение  $a =$

$= 0,3$  для всех трех блоков валидации с проверкой по контрольной выборке. При этом использовалась оценка тематического сходства  $S_{\text{Ham}}$  (результаты для оценки  $S_{\text{cos}}$  были на несколько процентов ниже).

После этого выполнена оценка качества поиска на всем наборе тестовых данных. Использованы следующие оценки: nDCG; полнота;  $D_b$  — доля предшествующих аналогов, обнаруженных среди  $b$  первых документов выдачи ( $D_{10}$ ,  $D_{20}$ ); медиана — медианное значение позиции первого аналога в выдаче. В табл. 3 представлены результаты оценки разработанного метода и приведено его сравнение с базовым методом, который применяется в Федеральном институте промышленной собственности (ФИПС) для решения аналогичной задачи.

По сравнению с базовым методом предложенный метод позволяет в 3/4 случаев обнаружить документы, которые являются прямыми аналогами поданных заявок (против менее половины случаев у базового метода). При каждом втором поиске предшествующий аналог находится среди первых 20 документов выдачи (для базового метода это происходит только при одном поиске из четырех). По  $D_{10}$  предложенный метод демонстрирует двукратное преимущество.

## 4 Заключение

В статье предложен новый эффективный метод поиска тематически похожих документов по



заданному эталону. Метод позволяет реализовать поддержку эксперта при эксплоративном патентном поиске. Проведенные эксперименты показали работоспособность и качество предложенного метода. Его внедрение в процессы патентной экспертизы будет способствовать снижению трудоемкости анализа патентных заявок по существу.

Возможны и другие приложения разработанного метода. Например, его можно использовать для подбора экспертов на научные проекты, публикации или научно-технические отчеты, применять в качестве предварительного фильтра при поиске текстовых заимствований или подбора кодов МПК при формировании новой патентной заявки.

Поиск тематически похожих документов лежит в основе метода выделения научных направлений с использованием метрической тематической кластеризации [11]. Последний, в свою очередь, служит основой для выявления тенденций развития и обнаружения «точек роста» в областях знаний.

Кроме того, предложенный метод был успешно использован для обнаружения документов-кандидатов, которые потенциально являются источниками текстовых заимствований [12].

## Литература

1. *Marchionini G.* Exploratory search: From finding to understanding // *Commun. ACM*, 2006. Vol. 49. No. 4. P. 41–46.
2. *White R. W., Roth R. A.* Exploratory search: Beyond the query-response paradigm. — *Synthesis lectures on information concepts, retrieval, and services ser.* — Morgan & Claypool Pubs., 2009. Vol. 1. No. 1. 98 p.
3. *Osipov G., Smirnov I., Tikhomirov I., Sochenkov I., Shelmanov A., Shvets A.* Information retrieval for R&D support // *Professional search in the modern world / Eds. G. Paltoglou, F. Loizides, P. Hansen.* — Lecture notes in computer science ser. — Springer, 2014. Vol. 8830. P. 45–69.
4. *Воронцов К. В.* Аддитивная регуляризация тематических моделей коллекций текстовых документов // Докл. РАН, 2014. Т. 456. № 3. С. 268–271.
5. *Moloshnikov I., Sboev A., Gudovskikh D., Rybka R.* An algorithm of finding thematically similar documents with creating context-semantic graph based on probabilistic-entropy approach // *Procedia Comput. Sci.*, 2015. Vol. 66. P. 297–306.
6. *Nokel M., Loukachevitch N.* Accounting ngrams and multi-word terms can improve topic models // 54th Annual Meeting of the Association for Computational Linguistics. Proceedings of 12th Workshop on Multiword Expressions. — Stroudsburg, PA, USA: ACL, 2016. P. 44–49.
7. *Glauner P. O., Iwaszkiewicz J., Meur J. Y. L., Simko T.* Use of Solr and Xapian in the Invenio document repository software // *ArXiv*, 2013. <https://arxiv.org/pdf/1310.0250.pdf>.
8. *Grainger T., Potter T.* Solr in action. — New York, NY, USA: Manning Publications, 2014. 664 p.
9. *Ilyinsky S., Kuzmin M., Melkov A., Segalovich I.* An efficient method to detect duplicates of Web documents with the use of inverted index // 11th World Wide Web Conference (International) Proceedings. — New York, NY, USA: ACM, 2002. 4 p.
10. *Агеев М. С., Добров Б. В.* Метод эффективного расчета матрицы ближайших соседей для полнотекстовых документов // *Вестник СПб ун-та. Сер. 10: Прикладная математика. Информатика*, 2011. № 3. С. 72–84.
11. *Shvets A., Devyatkin D., Sochenkov I., Tikhomirov I., Popov K., Yarygin K.* Detection of current research directions based on full-text clustering // *Science and Information Conference.* — London: IEEE, 2015. P. 483–488.
12. *Zubarev D., Sochenkov I.* Using sentence similarity measure for plagiarism source retrieval // *CEUR Workshop Proceedings*, 2014. Vol. 1180: Working Notes for CLEF 2014 Conference. P. 1027–1034. <http://ceur-ws.org/Vol-1180/CLEF2014wn-Pan-ZubarevEt2014.pdf>.

Поступила в редакцию 04.04.17

## EXPLORATORY PATENT SEARCH

I. Sochenkov<sup>1,2</sup>, D. Zubarev<sup>1,3</sup>, and I. Tikhomirov<sup>1</sup>

<sup>1</sup>Institute for Systems Analysis, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

<sup>2</sup>Skolkovo Institute of Science and Technology, 3 Nobelya Str., Moscow 121205, Russian Federation

<sup>3</sup>Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

**Abstract:** The paper presents an effective method for topically similar document retrieval. The exploratory patent search based on this method is proposed. The developed method reduces complexity and time of patent expertise providing the computer assistance of patent search and analysis. The phrases extracted by the parser as well as single lexemes are used as descriptors for a document. This approach prevents exponential growth of the feature space

and provides effective indexing even for large text collections. The results of experiments show that the proposed method significantly outperforms the basic keyword-based approach. Conclusions are made about the prospects of using the method for solving other problems such as source retrieval for plagiarism detection and full-text clustering.

**Keywords:** exploratory search; patent search; topic modeling; topically similar document retrieval; search and analytical engines

**DOI:** 10.14357/19922264180111

## Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 16-29-12929).

## References

1. Marchionini, G. 2006. Exploratory search: From finding to understanding. *Commun. ACM* 49(4):41–46.
2. White, R. W., and R. A. Roth. 2009. *Exploratory search: Beyond the query-response paradigm*. Synthesis lectures on information concepts, retrieval, and services ser. Morgan & Claypool Publ. 1(1). 98 p.
3. Osipov, G., I. Smirnov, I. Tikhomirov, I. Sochenkov, A. Shelmanov, and A. Shvets. 2014. Information retrieval for R&D support. *Professional search in the modern world*. Eds. G. Paltoglou, F. Loizides, and P. Hansen. Lecture notes in computer science ser. Springer. 8830:45–69.
4. Vorontsov, K. V. 2014. Additive regularization for topic models of text collections. *Dokl. Math.* 89(3):301–304.
5. Moloshnikov, I., A. Sboev, D. Gudovskikh, and R. Rybka. 2015. An algorithm of finding thematically similar documents with creating context-semantic graph based on probabilistic-entropy approach. *Procedia Comput. Sci.* 66:297–306.
6. Nokel, M., and N. Loukachevitch. 2016. Accounting ngrams and multi-word terms can improve topic models. *54th Annual Meeting of the Association for Computational Linguistics. Proceedings of 12th Workshop on Multiword Expressions*. Stroudsburg, PA: ACL. 44–49.
7. Glauner, P. O., J. Iwaszkiewicz, J. Y. L. Meur, and T. Simko. 2013 Use of Solr and Xapian in the Invenio document repository software. *ArXiv preprint*. Available at: <https://arxiv.org/pdf/1310.0250.pdf> (accessed November 17, 2017).
8. Grainger, T., and T. Potter. 2014. *Solr in action*. New York, NY: Manning Publications. 664 p.
9. Ilyinsky, S., M. Kuzmin, A. Melkov, and I. Segalovich. 2002. An efficient method to detect duplicates of Web documents with the use of inverted index. *11th World Wide Web Conference (International) Proceedings*. New York, NY: ACM. 4 p.
10. Ageev, M. S., and B. V. Dobrov. 2011. Metod effektivnogo rascheta matritsy blizhayshikh sosedy dlya polnotekstovyykh dokumentov [An efficient nearest neighbors search algorithm for full-text documents]. *Vestnik SPb un-ta. Ser. 10. Prikladnaya matematika. Informatika* [Vestnik of Saint Petersburg University. Applied mathematics. Computer science. Control processes] 3:72–84.
11. Shvets, A., D. Devyatkin, I. Sochenkov, I. Tikhomirov, K. Popov, and K. Yarygin. 2015. Detection of current research directions based on full-text clustering. *Science and Information Conference*. London: IEEE. 483–488.
12. Zubarev, D., and I. Sochenkov. 2014. Using sentence similarity measure for plagiarism source retrieval. *CEUR Workshop Proceedings: CLEF 2014 (Working Notes)*. 1027–1034. Available at: <http://ceur-ws.org/Vol-1180/CLEF2014wn-Pan-ZubarevEt2014.pdf> (accessed November 17, 2017).

Received April 4, 2017

## Contributors

**Sochenkov Ilya V.** (b. 1985) — Candidate of Science (PhD) in physics and mathematics, senior scientist, Skolkovo Institute of Science and Technology, 3 Nobelya Str., Moscow 121205, Russian Federation; lead scientist, Institute for Systems Analysis, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; I.Sochenkov@skoltech.ru

**Zubarev Denis V.** (b. 1989) — PhD student, Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; researcher, Institute for Systems Analysis, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; zubarev@isa.ru

**Tikhomirov Ilya A.** (b. 1980) — Candidate of Science (PhD) in technology, Head of Laboratory, Institute for Systems Analysis, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; tih@isa.ru

# ТЕОРИЯ КАТЕГОРИЙ КАК МАТЕМАТИЧЕСКАЯ ПРАГМАТИКА МОДЕЛЬНО-ОРИЕНТИРОВАННОЙ СИСТЕМНОЙ ИНЖЕНЕРИИ

С. П. Ковалёв<sup>1</sup>

**Аннотация:** Развивается предложенный ранее математический аппарат на базе теории категорий, который позволяет формально описывать и строго исследовать процедуры применения моделей в инженерной деятельности, составляющие прагматику модельно-ориентированной системной инженерии (Model-Based Systems Engineering, MBSE). В основе аппарата лежит математическое представление сборочных чертежей (мегамоделей систем) диаграммами в категориях, объектами которых служат модели, а морфизмы представляют действия по сборке моделей систем из моделей компонентов. Предложены, исследованы и проиллюстрированы теоретико-категорные методы решения прямых и обратных прагматических задач сборки систем. Выявлена ключевая роль монады диаграмм. Особое внимание уделено задаче восстановления конфигурации заданной системы с учетом технологических ограничений, накладываемых на способы и процедуры сборки. Приведено сопоставление ряда ключевых понятий системной инженерии конструкциям теории категорий.

**Ключевые слова:** модельно-ориентированная системная инженерия; прагматика; мегамодель; теория категорий; задача восстановления конфигурации; монада диаграмм

**DOI:** 10.14357/19922264180112

## 1 Введение

Технологии MBSE в настоящее время интенсивно развиваются [1]. Они поддерживают работу инженеров с разнообразными моделями сложных изделий — абстрактными формальными представлениями, пригодными к автоматической обработке на компьютерах. Модели записываются на языках, имеющих хорошо проработанный синтаксис (форму) и семантику (смысл). Всеобъемлющий комплекс моделей образует электронно-цифровой макет (digital mock-up) — виртуальную копию изделия, на которой можно исследовать и оптимизировать все аспекты жизненного цикла «в цифре» прежде, чем воплощать «в железе», снижая затраты и сокращая сроки выполнения работ.

И все же внедрение технологий MBSE в инженерную деятельность происходит фрагментарно и часто не выходит за рамки «пилотных» проектов. Такое положение дел во многом обусловлено факторами неопределенности прагматического характера, которые снижают отдачу от вложений в перевод всех работ на модельно-ориентированные «рельсы» [2, 3]:

- слабая совместимость языков и инструментов моделирования от разных поставщиков;
- недостаточно внятное отражение принципов и приемов системного мышления в языках и инструментах;

- нехватка мощных типовых механизмов коллективной модульной разработки больших моделей;
- неполнота автоматизации жизненного цикла артефактов MBSE.

Эти факторы неопределенности известны более 10 лет, и в направлении их устранения ведутся активные исследования. Однако результаты в основном ограничиваются совершенствованием частных языков и технологий: лишь немногие вносят вклад в методологический базис MBSE в целом. Естественный, хотя и «трудный» подход к получению результатов общего характера, унифицирующих разнородные технологии, состоит в том, чтобы как можно строже формализовать процедуры применения моделей в инженерии, составляющие прагматику модельно-ориентированного подхода. Формализация, особенно с привлечением математического аппарата, позволит и совершенствовать процедуры MBSE, и передавать их на исполнение компьютеру без пробелов и искажений.

В настоящей работе, в развитие результатов работы [4], применяется аппарат, основанный на математическом представлении сборочных чертежей («мегамоделей» систем) ориентированными графами (диаграммами). Узлы такого графа помечаются обозначениями моделей частей, а ребра помечаются обозначениями действий (activities), посредством которых части собираются в систему. Пред-

<sup>1</sup>Институт проблем управления им. В. А. Трапезникова Российской академии наук, kovalyov@nm.ru

ставление структуры систем графами регламентируется стандартами, такими как IEC 81346 [5]. Естественным источником математических методов конструирования и анализа мегамоделей служит теория категорий (см., например, [6]): модели рассматриваются как объекты подходящих категорий, а действия формально описываются морфизмами. В работе строятся и исследуются теоретико-категорные конструкции, описывающие процедуры MBSE на абстрактном концептуальном уровне. Например, сборке системы из мегамодели отвечает построение копредела диаграммы — универсальной конструкции [6]. Актуальны в системной инженерии и обратные задачи сборки систем, решение которых состоит в реконструкции диаграмм по ребрам их копределов. Насколько известно автору, реконструкция такого рода ранее не рассматривалась в литературе.

Работа построена следующим образом. Раздел 2 посвящен прагматическим проблемам мегамоделирования и сборки систем. В разд. 3 и 4 вводятся конструкции теории категорий, позволяющие решать прямые и обратные задачи мегамоделирования соответственно. В заключении приводятся выводы и намечаются направления дальнейших исследований.

## 2 Прямые и обратные задачи сборки систем

Для анализа структуры систем и организации сборки в MBSE необходимо знать не столько внутреннюю структуру моделей, сколько ассортимент их возможностей соединяться с другими моделями в целях формирования моделей составных объектов. Иными словами, модели рассматриваются как «черные ящики» с известным поведением по отношению к другим моделям. Прагматический контекст сборки систем в условиях применения некоторой частной технологии MBSE можно представлять в виде каталога, состоящего из моделей и описаний действий по их соединению.

Напомним, что структуры систем и сборочные чертежи представляют собой частные случаи мегамоделей (megamodel) — моделей, состоящих из моделей и связей между ними [7]. Мегамодель, в которой связи описывают соединение моделей, образующих некоторую систему, называется конфигурацией этой системы [6]. Прямая задача мегамоделирования сборки систем формулируется следующим образом [4]. По мегамодели, представляющей конфигурацию некоторой системы, требуется сконструировать модель системы как це-

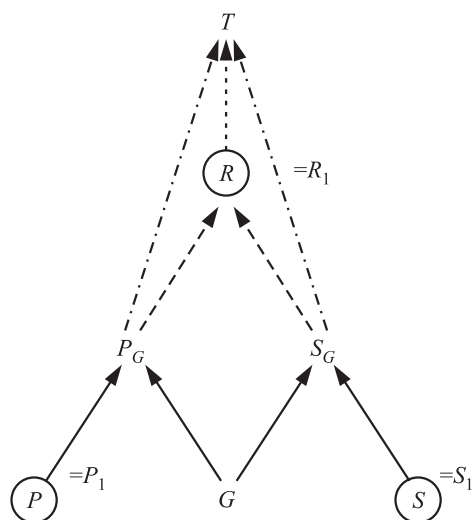


Рис. 1 Схема склеивания

лого. Например, рассмотрим сборку системы из двух объектов  $P$  и  $S$  с помощью объекта-клея  $G$ . Действие клея описывается конфигурацией следующего вида: объекты  $G$  и  $P$  порождают в результате соединения известный промежуточный комплексный объект  $P_G$ , содержащий их, а объекты  $G$  и  $S$  порождают объект  $S_G$ . Система  $R$ , полученная путем склеивания  $P$  с  $S$  при помощи  $G$ , отбирается среди объектов, содержащих  $P_G$  и  $S_G$ , по следующему структурному критерию: объект должен содержаться в любом объекте  $T$ , содержащем  $P_G$  и  $S_G$ . Схематически этот критерий изображен на рис. 1, где в кружки заключены элементы структурного представления системы по стандарту IEC 81346.

Хотя прямой расчет системы по конфигурации имеет большое значение, в прагматике MBSE он играет вспомогательную роль. Согласно стандарту IEC 81346 и практикам системной инженерии, система обычно проектируется сверху вниз — от корня структурной иерархии к составляющим [8]. Это означает, что технолог в основном решает не прямую, а обратную задачу: модель системы, которую нужно собрать, известна, а нужно построить (восстановить) конфигурацию, из которой такая система может быть получена путем сборки. При этом ситуация, когда известна только целевая система, встречается редко. Обычно конструктор заранее знает некоторые из составляющих ее объектов и способов их включения. В свою очередь, технолог априори знает ассортимент и возможности имеющихся в его распоряжении средств производства, поэтому выбирает конфигурацию целевой системы не произвольно, а из некоторого заранее известного класса. Учет этих знаний приводит к следующей формулировке обратной задачи мегамоделирова-



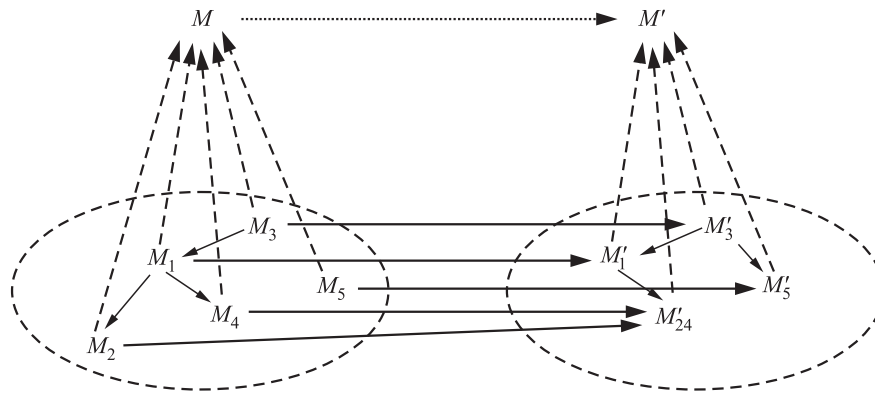


Рис. 2 Пример коллективной модификации сложной модели

ния. По заданному фрагменту мегамодели сборки, содержащему целевую систему, требуется восстановить мегамодель целиком, причем образующая ее конфигурация должна находиться в заданном классе, представляющем все технологически реализуемые конфигурации.

Например, обратная задача поиска клея ставится следующим образом. В соответствии со структурой целевой системы  $R$ , описанной согласно стандарту IEC 81346, дан фрагмент мегамодели, состоящий из трех объектов  $P$ ,  $S$  и  $R$  и двух стрелок  $P \rightarrow R$  и  $S \rightarrow R$ . Конфигурацией склеивания служит любая комбинация из пяти объектов и четырех стрелок, форма которой совпадает с частью рис. 1, образованной всеми сплошными стрелками. Требуется найти среди всех конфигураций указанного вида такую, для которой решение прямой задачи мегамоделирования порождает заданные стрелки  $P \rightarrow R$  и  $S \rightarrow R$  в качестве композитных включений двух «крайних» объектов конфигурации в целевую систему (в частности, крайними объектами искомой конфигурации должны быть  $P$  и  $S$ , а целевой системой —  $R$ ).

Обратные задачи мегамоделирования возникают также при организации управления моделями (model management) [9]. Так, на вход процедур коллективной модульной разработки поступает цельная модель, которую также нужно разделить на части. Однако критерием разделения выступает не доступность частей в виде деталей, а удобство модификации и локальной верификации каждой части. После модификации частей производится сборка новой версии цельной модели, т. е. решается прямая задача мегамоделирования. При этом требуется обеспечить структурную корректность всей процедуры: должна иметься возможность представить процедуры модификации частей как составляющие гипотетической общей процедуры модификации модели как целого. Такая общая процедура может

быть практически невыполнима ввиду чрезмерной ресурсоемкости, однако она должна существовать в принципе.

В простом случае, когда модификация каждой части может быть описана действием из каталога моделей, условие структурной корректности может выглядеть, как представлено на рис. 2. Здесь изображена процедура коллективной модификации некоторой большой модели  $M$  по частям  $M_1, \dots, M_5$ , превращающая ее в модель  $M'$ .

В MBSE предлагается ряд подходов к решению обратных задач моделирования, известных под общим названием исследования пространства проектирования (Design Space Exploration, DSE) [10]. В мегамоделировании пространством проектирования служит класс всех мегамоделей, из которого допускается выбирать искомые конфигурации. Методы DSE, по существу, предлагают перебирать пространство проектирования, для каждой выбранной из него конфигурации решать прямую задачу мегамоделирования и проверять, включает ли получившаяся мегамодель сборки исходно заданный фрагмент, содержащий целевую систему. Конечно, полный перебор может быть неосуществим в принципе, поскольку пространство проектирования может быть теоретически бесконечным (содержать мегамодели, включающие сколь угодно много составляющих). Поэтому предлагаются методы и автоматизированные инструменты DSE, основанные на представлении постановки обратной задачи набором ограничений, записанных на подходящем формальном языке мегамоделирования [11] (в данном случае требуется язык *метамегамоделирования*). В некоторых случаях возможна непосредственная генерация элементов пространства проектирования по такой метамегамодели, в других случаях — разделение пространства проектирования на обозримые части путем уточнения ограничений и т. п.

Если пространство проектирования достаточно велико, то решение некоторой обратной задачи мегамоделирования может дать большое число конфигураций. Выбор итоговой среди них осуществляется путем ранжирования по значениям различных количественных показателей, отражающих расход труда, материалов, финансовых средств, времени и других ресурсов на сборку системы. Такие показатели являются предметом стандартизации, в частности в рамках ГОСТов серии 14, регламентирующих технологическую подготовку производства. В фокусе внимания стандартов группы 2 серии 14 находится технологичность конструкции изделия — «совокупность свойств конструкции изделия, определяющих ее приспособленность к достижению оптимальных затрат при производстве, техническом обслуживании и ремонте для заданных показателей качества, объема выпуска и условий выполнения работ» [12]. Многие показатели технологичности конструкции изделия можно вычислять, основываясь на формальном представлении мегамоделей графами: подсчитывать число узлов, длину критического пути и т. д. Например, коэффициентом сборности называется отношение числа узлов, имеющих входящие ребра, к общему числу узлов графа мегамодели сборки. Коэффициент сборности мегамодели склеивания, изображенной на рис. 1, равен 0,5.

Существуют и менее формальные методы восстановления конфигурации систем, возникшие задолго до MBSE. В основе таких методов лежат разнообразные эвристики, направленные на сужение пространства проектирования до области применения того или иного технологического приема или образца. Широкий набор таких эвристик накапливается начиная с 1950-х гг. в рамках дисциплины, известной как теория решения изобретательских задач (ТРИЗ) [13].

### 3 Теория категорий в мегамоделировании

Как указывалось во введении, естественным источником математических методов конструирования и анализа мегамоделей служит теория категорий. Будем пользоваться теоретико-категорными конструкциями и обозначениями, введенными в работе [4]. Зафиксируем произвольную категорию  $\mathcal{C}$ , представляющую некоторый каталог моделей и описаний действий по их соединению. Примерами служат категория твердотельных геометрических моделей **MBS**, содержащаяся в категории множеств **Set**, и категория дискретно-со-

бытийных имитационных моделей **Pomset**, которая является конкретной категорией над **Set**. Задачи мегамоделирования рассматриваются в категории  $\mathcal{C}$ -диаграмм **DC**. Морфизмы диаграмм описывают структурные преобразования мегамоделей, выполняемые при помощи инструментов MBSE, поддерживающих сборку систем (например, морфизм диаграмм изображен на схеме коллективной модификации моделей на рис. 2). Таким образом, теория категорий предоставляет тот самый язык мегамоделирования, который необходим для проведения DSE в задачах восстановления конфигурации систем.

Для начала напомним, что прямая задача построения системы  $R$  по заданной конфигурации  $\Delta$  решается путем конструирования копредела — **DC**-морфизма  $\text{colim } \Delta : \Delta \rightarrow \ulcorner R \urcorner$ , универсального среди всех коконусов с областью  $\Delta$ . (Через  $\ulcorner R \urcorner$  обозначается одноточечная  $\mathcal{C}$ -диаграмма, состоящая из единственной вершины  $R$  и не имеющая ребер.) Конструирование копределов описано в работе [4] функтором  $\text{colim}$ , который переводит в  $\mathcal{C}$  подходящую подкатегорию в **DC** согласно следующей диаграмме:

$$\begin{array}{ccc}
 \Delta & \xrightarrow{\text{colim } \Delta} & \ulcorner \text{colim } (\Delta) \urcorner \\
 \theta \downarrow & & \downarrow \ulcorner \text{colim } (\theta) \urcorner \\
 \Xi & \xrightarrow{\text{colim } \Xi} & \ulcorner \text{colim } (\Xi) \urcorner
 \end{array}$$

Для прагматики MBSE функтор  $\text{colim}$  полезен, в частности, тем, что он позволяет обеспечить структурную корректность описанной в разд. 2 процедуры коллективной модификации большой модели. Действительно, если можно «упаковать» модификации частей модели в некоторый **DC**-морфизм  $\theta$ , то порожденная ими процедура модификации модели как целого описывается морфизмом  $\text{colim } (\theta)$ . Легко увидеть это из вышеприведенной диаграммы: если повернуть ее на  $90^\circ$  против часовой стрелки и изобразить копределы диаграммами соответствующего вида, то получится рис. 2. Предлагаются теоретико-категорные конструкции и для более сложных процедур управления моделями [7].

Обратим внимание на одно существенное отличие диаграммы, определяющей действие функтора  $\text{colim}$ , от диаграммы коллективной модификации модели из разд. 2: первая является **DC**-диаграммой, а вторая —  $\mathcal{C}$ -диаграммой. Чтобы проиллюстрировать приложение функтора  $\text{colim}$ , можно мысленно преобразовать **DC**-диаграмму в  $\mathcal{C}$ -диаграмму путем «отрисовки». В действительности это преобразование является строго формальным: отрисовка

произвольной диаграммы  $\Gamma : Z \rightarrow \mathbf{DC}$  заключается в замене каждой вершины  $A$  графа схемы  $Z$  графом  $C$ -диаграммы  $\Gamma(A)$ , а каждого ребра  $f : A \rightarrow B$  — совокупностью ребер, по одному для каждой вершины  $I$  графа диаграммы  $\Gamma(A)$ , направленному из  $I$  в вершину  $fd(I)$  графа диаграммы  $\Gamma(B)$  и помеченному  $C$ -морфизмом  $\varepsilon_I$ , где  $\langle \varepsilon, fd \rangle = \Gamma(f)$ . Ясно, что вместе с  $\mathbf{DC}$ -диаграммами можно отрисовывать и их морфизмы, причем так, что возникает функтор отрисовки  $\mathbf{K}_C : \mathbf{DDC} \rightarrow \mathbf{DC}$ . В свою очередь, процедура конструирования категории диаграмм  $\mathbf{DC}$  может быть задана функтором: любой функтор  $\text{fun} : C \rightarrow D$  порождает функтор

$$\text{fun} \circ - : \mathbf{DC} \rightarrow \mathbf{DD} : \Delta \mapsto \text{fun} \circ \Delta, \langle \varepsilon, fd \rangle \mapsto \langle \text{fun}(\varepsilon), fd \rangle,$$

так что имеется функтор:

$$\mathbf{D} : \mathbf{CAT} \rightarrow \mathbf{CAT} : C \mapsto \mathbf{DC}, \text{fun} \mapsto (\text{fun} \circ -).$$

Можно распространить действие этого функтора на естественные преобразования, а именно: произвольному естественному преобразованию  $\xi : \text{fun} \rightarrow \text{fun}'$ , где  $\text{fun}, \text{fun}' : C \rightarrow D$  — произвольные функторы, соответствует естественное преобразование  $\mathbf{D}\xi : \mathbf{D}\text{fun} \rightarrow \mathbf{D}\text{fun}'$  с компонентами вида

$$(\mathbf{D}\xi)_\Delta = \langle \xi_{\Delta(-)}, 1_X \rangle : \text{fun} \circ \Delta \rightarrow \text{fun}' \circ \Delta,$$

где  $\Delta : X \rightarrow C$  — произвольная диаграмма. Читатель, искушенный в теории категорий, заметит, что таким путем  $\mathbf{D}$  определен как 2-функтор. Более того, как установлено в [14], тройка  $\mathcal{D} = \langle \mathbf{D}, \ulcorner -, \urcorner, \mathbf{K} \rangle$  образует 2-монаду в  $\mathbf{CAT}$ , причем она по ряду свойств аналогична монаде степени  $\langle 2^-, \{-\}, \cup \rangle$  в категории **Set**. Стрелка  $\text{colim} : \mathbf{DC} \rightarrow C$  для любой копальной категории  $C$  является  $\mathcal{D}$ -алгеброй — объектом категории Эйленберга—Мура, ассоциированной с монадой  $\mathcal{D}$ . Если в категории  $C$  есть суммы, то имеется функтор  $\text{sum} : \mathbf{DC} \rightarrow C$ , сопоставляющий каждой  $C$ -диаграмме сумму всех ее вершин, и он также является  $\mathcal{D}$ -алгеброй. И кроме того, произвольная категория  $C$  порождает свободную  $\mathcal{D}$ -алгебру  $\mathbf{K}_C : \mathbf{DDC} \rightarrow \mathbf{DC}$  [15, предложение 20.7]. Например, категория **Pomset** выступает носителем попарно неизоморфных  $\mathcal{D}$ -алгебр всех трех вышеуказанных видов. В целом,  $\mathcal{D}$ -алгебры задают шаблоны системных процедур общего вида, состоящих в свертке мегамоделей в модели составных объектов с соблюдением условий естественности относительно модификаций. Монаду  $\mathcal{D}$  правомерно рассматривать как метамодель мегамоделирования, поскольку она включает в себя все средства языка теории категорий, применяемые для описания процедур, связанных со сборкой систем.

## 4 Теоретико-категорные методы восстановления конфигурации

Общее категорное решение обратных задач мегамоделирования неизвестно. Сформулируем на языке теории категорий постановку задачи восстановления конфигурации и исследуем ряд частных случаев. Для этого понадобится понятие подкокнуса. Коконус  $\delta$  называется подкокнусом в кокнусе  $\beta$ , если существует  $\mathbf{DC}$ -морфизм вида  $\langle \mu, iy \rangle : \text{dom } \delta \rightarrow \text{dom } \beta$  такой, что естественное преобразование  $\mu$  состоит из тождественных  $C$ -морфизмов, функтор  $iy$  инъективен и  $\beta \circ \langle \mu, iy \rangle = \delta$ .

Обратная задача мегамоделирования математически ставится следующим образом. Даны коконус  $\delta$ , представляющий известный фрагмент мегамодели сборки, и класс диаграмм  $\text{Cd}$ , представляющий пространство проектирования. Требуется найти в классе  $\text{Cd}$  диаграмму, обладающую копредделом, подкокнусом в котором является  $\delta$ .

Как упоминалось в разд. 2, большую пользу для решения такой задачи приносит сужение класса  $\text{Cd}$ , представляющего пространство проектирования, и упрощение структуры составляющих его диаграмм. В этом направлении получен следующий результат.

**Теорема 1.** Пусть  $\Delta : X \rightarrow C$  — произвольная диаграмма,  $Y$  — полная рефлексивная подкатегория в  $X$ ,  $\text{isc} : Y \hookrightarrow X$  — вложение. Для любого коконуса  $\theta : \Delta \circ \text{isc} \rightarrow \ulcorner S \urcorner$  существует единственный коконус  $\theta' : \Delta \rightarrow \ulcorner S \urcorner$  такой, что  $\theta = \theta' \circ \langle 1_{\Delta \circ \text{isc}}, \text{isc} \rangle$ . При этом коконус  $\theta$  является копредделом диаграммы  $\Delta \circ \text{isc}$  тогда и только тогда, когда коконус  $\theta'$  является копредделом диаграммы  $\Delta$ .

**Доказательство.** Напомним, что подкатегория  $Y$  называется рефлексивной в  $X$ , если вложение  $\text{isc} : Y \hookrightarrow X$  обладает левым сопряженным  $\text{isc}^* : X \rightarrow Y$  рефлексором [16, разд. 4.3]. Пусть  $\eta : 1_X \rightarrow \text{isc} \circ \text{isc}^*$  — единица рефлексии, она индуцирует  $\mathbf{DC}$ -морфизм  $\langle \Delta(\eta), \text{isc}^* \rangle : \Delta \rightarrow \Delta \circ \text{isc}$ . Согласно определению коконуса, для любого  $X$ -объекта  $I$  ребро  $f : \Delta(I) \rightarrow S$  искомого коконуса  $\theta'$  должно удовлетворять условию  $f = h \circ \Delta(\eta_I)$ , где  $h : \Delta(\text{isc}(\text{isc}^*(I))) \rightarrow S$  — ребро коконуса  $\theta$ , откуда  $\theta' = \theta \circ \langle \Delta(\eta), \text{isc}^* \rangle$ .

Предположим, что  $\theta$  — копредел, выберем произвольно коконус  $\rho : \Delta \rightarrow \ulcorner T \urcorner$ . Пусть  $w : S \rightarrow T$  — единственный  $C$ -морфизм, удовлетворяющий условию  $\rho \circ \langle 1_{\Delta \circ \text{isc}}, \text{isc} \rangle = \ulcorner w \urcorner \circ \theta$ . Тогда  $\ulcorner w \urcorner \circ \theta' = \rho \circ \langle 1_{\Delta \circ \text{isc}}, \text{isc} \rangle \circ \langle \Delta(\eta), \text{isc}^* \rangle = \rho$  и, кро-

ме того, если некоторый  $C$ -морфизм  $\tilde{w} : S \rightarrow T$  удовлетворяет условию  $\rho = \ulcorner \tilde{w} \urcorner \circ \theta'$ , то  $\rho \circ \langle 1_{\Delta \circ \text{isc}}, \text{isc} \rangle = \ulcorner \tilde{w} \urcorner \circ \theta' \circ \langle 1_{\Delta \circ \text{isc}}, \text{isc} \rangle = \ulcorner \tilde{w} \urcorner \circ \theta$ , откуда  $\tilde{w} = w$ . Следовательно,  $\theta'$  — копредел диаграммы  $\Delta$ .

Обратно, предположим, что  $\theta'$  — копредел, и выберем произвольно коконус  $\gamma : \Delta \circ \text{isc} \rightarrow \ulcorner U \urcorner$ . Пусть  $u : S \rightarrow U$  — единственный  $C$ -морфизм, удовлетворяющий условию

$$\gamma' = \ulcorner u \urcorner \circ \theta',$$

где  $\gamma' = \gamma \circ \langle \Delta(\eta), \text{isc}^* \rangle$ . Тогда

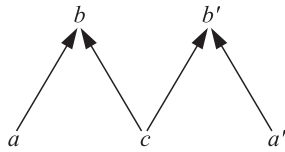
$$\ulcorner u \urcorner \circ \theta = \gamma \circ \langle \Delta(\eta), \text{isc}^* \rangle \circ \langle 1_{\Delta \circ \text{isc}}, \text{isc} \rangle = \gamma$$

и, кроме того, если некоторый  $C$ -морфизм  $\tilde{u} : S \rightarrow U$  удовлетворяет условию  $\gamma = \ulcorner \tilde{u} \urcorner \circ \theta$ , то

$$\gamma' = \ulcorner \tilde{u} \urcorner \circ \theta \circ \langle \Delta(\eta), \text{isc}^* \rangle = \ulcorner \tilde{u} \urcorner \circ \theta',$$

откуда  $\tilde{u} = u$ . Следовательно,  $\theta$  — копредел диаграммы  $\Delta \circ \text{isc}$ .  $\square$

Исследуем при помощи теоремы 1 задачу поиска клея из разд. 2. Обозначим через  $M$  схему конфигурации склеивания — категорию, порожденную графом следующего вида:



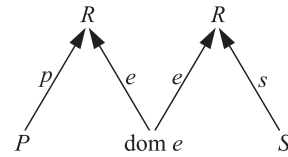
Все диаграммы со схемой  $M$  образуют пространство проектирования склеивания — класс  $\text{Cd}$  из общей постановки задачи восстановления конфигурации. Коконус  $\delta$ , фигурирующий в общей постановке, реконструируется по структуре системы, представленной согласно стандарту IEC 81346 на рис. 1: основанием коконуса служит диаграмма  $\Delta : a \mapsto P, a' \mapsto S$  с двухточечной дискретной схемой  $\{a, a'\} \subseteq M$ , а вершиной — собираемая система  $R$ .

Пусть  $V$  — полная подкатегория в  $M$ , состоящая из трех «центральных»  $M$ -объектов:  $V$  порождается графом  $b \leftarrow c \rightarrow b'$ . Ясно, что подкатегория  $V$  рефлексивна в  $M$ , причем ребра  $a \rightarrow b$  и  $a' \rightarrow b'$  графа категории  $M$  являются компонентами единицы рефлексии, соответствующими объектам  $a$  и  $a'$  соответственно. По теореме 1 копредел диаграммы со схемой  $M$  взаимно однозначно задается копределом диаграммы со схемой  $V$ , который хорошо известен в теории категорий под названием кокартова квадрата (pushout). С учетом этого получается общее решение задачи поиска клея.

**Предложение 1.** Для любых объектов  $P, S, R$  и пары морфизмов  $p : P \rightarrow R \leftarrow S : s$  существует диаграмма

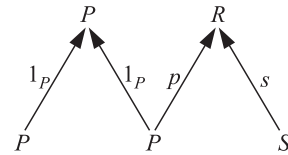
$\Delta : M \rightarrow C$ , имеющая копредел вида  $\langle \kappa, !_M \rangle : \Delta \rightarrow \ulcorner R \urcorner$  такой, что  $\kappa_a = p$  и  $\kappa_{a'} = s$ . Указанная диаграмма единственна тогда и только тогда, когда  $p = s = 1_R$  является единственным  $C$ -эпиморфизмом с кообластью  $R$ .

**Доказательство.** Произвольный  $C$ -морфизм  $e$  с кообластью  $R$  является эпиморфизмом тогда и только тогда, когда равенство  $1_R \circ e = 1_R \circ e$  задает кокартов квадрат в  $C$  [16, упражнение 3.4.4]. Поэтому если  $e$  — эпиморфизм, то ввиду теоремы 1 в качестве  $\Delta$  можно выбрать следующую диаграмму:



В частности, при  $e = 1_R$  получается решение задачи поиска клея в стиле ТРИЗ: если уже имеется система  $R$ , включающая в себя объекты  $P$  и  $S$ , то пусть она и послужит для склеивания этих объектов!

Если  $p \neq 1_R$ , то существует еще одно решение, которое выглядит следующим образом:



Если же  $p = s = 1_R$ , то решение задачи включает в себя нахождение разложений вида  $1_R = o \circ q$  (чтобы пометить морфизмом  $q$  внешнюю стрелку в искомой диаграмме  $\Delta$ ). Морфизм  $o$  в любом таком разложении является эпиморфизмом [15, предложение 7.42] и имеет кообласть  $R$ , поэтому если единственным эпиморфизмом с кообластью  $R$  является  $1_R$ , то искомая диаграмма  $\Delta$  единственна: она может быть составлена только из морфизмов  $1_R$ .  $\square$

Например, в категории множеств **Set** критерий единственности решения задачи поиска клея равносителен тому, что множество  $R$  пусто.

Простота решения общей задачи поиска клея обусловлена широким произволом в выборе объекта, выполняющего функцию склеивания. Однако на практике на выбор клея накладываются те или иные технологические ограничения. Например, часто требуется, чтобы клей не вносил существенных изменений в объекты  $P$  и  $S$ , т.е. чтобы крайние стрелки конфигурации склеивания были изоморфизмами. Так устроено склеивание в твердотельном моделировании: нанесение тонкого слоя клеящего вещества на участок поверхности тела не меняет его геометрии. В таком случае



можно упростить конфигурацию, отбросив крайние стрелки: составить класс  $\text{Cd}$ , фигурирующий в постановке обратной задачи, из всех диаграмм вида  $P \leftarrow G \rightarrow S$  со схемой  $V$ , а в качестве  $\delta$  выбрать коконус  $[\ulcorner p^\ulcorner, \lrcorner s^\lrcorner] : \Omega \rightarrow \lrcorner R^\lrcorner$ , где  $\Omega : b \mapsto P$ ,  $b' \mapsto S$  — диаграмма с двухточечной дискретной схемой  $\{b, b'\} \subseteq V$ . Решить такую задачу восстановления кодекартова квадрата гораздо труднее, чем предыдущую. Ее решение будет приведено в категории **Set**.

**Предложение 2.** Пусть даны произвольные множества  $P, S, R$  и пара отображений  $p : P \rightarrow R \leftarrow S : s$ . Множество  $G$  и пара отображений  $f : P \leftarrow G \rightarrow S : g$  такие, что соотношение  $p \circ f = s \circ g$  выполняется и задает кодекартов квадрат в категории **Set**, существуют тогда и только тогда, когда для любого  $r \in R$  выполняется следующее условие:

$$\begin{aligned} (|p^{-1}(r)| = 0 \Rightarrow |s^{-1}(r)| = 1) \wedge \\ \wedge (|s^{-1}(r)| = 0 \Rightarrow |p^{-1}(r)| = 1). \end{aligned}$$

При выполнении этого условия, искомые множество и пара отображений единственны тогда и только тогда, когда  $|p(P) \cap s(S)| = 0$ .

**Доказательство.** Напомним, что кодекартов квадрат в **Set** над некоторой парой отображений  $f : P \leftarrow G \rightarrow S : g$  строится следующим образом [16, разд. 3.3]. В качестве вершины  $R$  берется множество, полученное из раздельного объединения  $P \amalg S$  путем факторизации по отношению эквивалентности, порожденному множеством пар  $\{(f(x), g(x)) | x \in G\}$ . Ребро  $p : P \rightarrow R$  сопоставляет каждому элементу множества  $P$  класс эквивалентности элемента по этому отношению, а ребро  $s : S \rightarrow R$  действует аналогично на  $S$ . Поэтому если  $|p^{-1}(r)| = 0$  для некоторого  $r \in R$ , то элемент  $r$  «попал» в  $R$  из множества  $S \setminus g(G)$ : существует единственный  $t \in S$  такой, что  $s(t) = r$ . Аналогично рассматривается случай  $|s^{-1}(r)| = 0$ .

Обратно, предположим, что условие, приведенное в формулировке доказываемого утверждения, выполняется для некоторой пары отображений  $p : P \rightarrow R \leftarrow S : s$ ; в частности,  $p(P) \cup s(S) = R$ . Построим *декартов* квадрат (предел) указанной пары [16, разд. 3.4]: положим

$$\begin{aligned} G &= \{(x, y) | p(x) = s(y)\} \subseteq P \times S, \\ f : G &\rightarrow P : (x, y) \mapsto x, g : G \rightarrow S : (x, y) \mapsto y. \end{aligned}$$

Пусть пара отображений  $p' : P \rightarrow R' \leftarrow S : s'$  задает ребра кодекартова квадрата, построенного стандартным способом над парой отображений  $f : P \leftarrow G \rightarrow S : g$ . Построим отображение  $i : R \rightarrow R'$  следующим образом: если  $r \in R \setminus p(P)$ , то полагаем

$i(r) = s'(t)$ , где  $t$  — единственный элемент множества  $S$  такой, что  $s(t) = r$  (такой элемент существует по предположению), а в противном случае полагаем  $i(r) = p'(v)$  для произвольного  $v \in p^{-1}(r)$  (ясно, что значение  $i(r)$  не зависит от выбора  $v$ ). Легко видеть, что  $i$  — биекция,  $p = i^{-1} \circ p'$  и  $s = i^{-1} \circ s'$ , так что пара  $p : P \rightarrow R \leftarrow S : s$  действительно задает ребра кодекартова квадрата над парой  $f : P \leftarrow G \rightarrow S : g$ . Более того, если  $|p(P) \cap s(S)| = 0$ , то  $R$  совпадает с точностью до изоморфизма с раздельным объединением  $P \amalg S$ , а  $p$  и  $s$  — с инъекциями компонентов в него, поэтому пара  $p : P \rightarrow R \leftarrow S : s$  задает ребра кодекартова квадрата только над парой  $P \leftarrow \emptyset \rightarrow S$ . Если же  $|p(P) \cap s(S)| > 0$ , т.е. если множество  $G$  непусто, то искомым пар бесконечно много: очевидно, что для любой сюръекции  $e : G' \rightarrow G$  отображения  $p$  и  $s$  являются ребрами кодекартова квадрата над парой  $f \circ e : P \leftarrow G' \rightarrow S : g \circ e$ .  $\square$

Если отображения  $p$  и  $s$  инъективны, то отображения  $f$  и  $g$ , построенные в доказательстве предложения 2, также будут инъективными. Поэтому метод из этого доказательства позволяет в ряде случаев восстанавливать конфигурации склеивания в категории твердотельных моделей **MBS**.

Можно приспособить этот метод и для применения во многих известных категориях множеств со структурой. Здесь в условие существования решения добавляется требование совпадения структуры, индуцированной отображением  $p$  на подмножестве  $p(P) \subseteq R$ , с ограничением структуры объекта  $R$  на это подмножество и аналогичное требование для  $s$ .

Объект-клей  $G$ , как в **Set**, можно построить при помощи декартова квадрата (можно проверить, что и в произвольной категории, если для некоторой пары морфизмов с общей кообластью существует как решение задачи восстановления кодекартова квадрата, так и декартов квадрат, то последний является и кодекартовым, т.е. решением задачи). Например, так можно действовать в категории дискретно-событийных имитационных моделей **Pomset**. В ней задача восстановления кодекартова квадрата над парой  $p : P \rightarrow R \leftarrow S : s$  имеет решение тогда и только тогда, когда пара удовлетворяет условию из предложения 2 и, кроме того, каноническое отображение  $[p, s] : P \amalg S \rightarrow R$  является регулярным **Pomset**-эпиморфизмом. Клей представляет собой множество  $G$ , сконструированное в доказательстве предложения 2, порядок и разметка на котором индуцированы из произведения сценария-«клея» такого рода при сборке имитационных моделей больших производственных систем [17].

## 5 Заключение

Аппарат теории категорий обладает большим потенциалом в области повышения полезной отдачи от MBSE, в том числе путем математически строгого решения прямых и обратных задач мегамоделирования. Можно наглядно представить принципы применения аппарата в форме таблицы соответствия между ключевыми понятиями теории категорий и системной инженерии. Для физики, топологии, логики и информатики таблица такого рода построена в известном теоретико-категорном «розеттском камне» [18]. В настоящей работе установлены следующие соответствия, связанные с процедурами сборки систем в модельно-ориентированном подходе:

Теория категорий	Сборка систем в MBSE
Объект	Модель
Морфизм	Действие по сборке моделей систем
Диаграмма	Мегамодель
Копредел диаграммы	Сборка системы из конфигурации
Функтор $\text{colim}$	Коллективная модификация моделей
Монада диаграмм	Метамодель мегамоделирования

Конечно, чтобы передать аппарат в руки системных инженеров, недостаточно сопоставить понятия: необходимо реализовать категорные методы в программных инструментах [19]. В частности, целесообразно записать методы решения типовых прагматических задач MBSE на входном языке решателей — автоматизированных средств доказательства теорем. Существует ряд категорных решателей (см., например, [20]), которые можно взять за основу. На базе решателя можно создать программный процессор мегамоделей «общего назначения». Чтобы настроить такой процессор на конкретную технологию моделирования, нужно будет описать технологию категорией, представляющей каталог моделей и действий по их соединению. Таким путем может быть реализован перспективный универсальный язык системного моделирования [3]. Здесь открывается широкий спектр направлений для дальнейших исследований.

## Литература

1. Modeling and simulation-based systems engineering handbook / Eds. D. Gianni, A. D'Ambrogio, A. Tolk. — London: CRC Press, 2014. 513 p.
2. Selic B. The pragmatics of model-driven development // IEEE Software, 2003. Vol. 20. Iss. 5. P. 19–25.
3. Левенчук А. И. Системноинженерное мышление. — М.: TechInvestLab, 2015. 305 с.
4. Ковалёв С. П. Методы теории категорий в модельно-ориентированной системной инженерии // Информатика и её применения, 2017. Т. 11. Вып. 3. С. 42–50.
5. IEC 81346-1:2009. Industrial Systems, Installations and Equipment and Industrial Products — Structuring Principles and Reference Designations — Part 1: Basic Rules. — Geneva: ISO, 2009. 168 p.
6. Ginali S., Goguen J. A categorical approach to general systems // Applied general systems research: Recent development and trends / Ed. G. J. Klir. — NATO conference ser. — Boston, MA, USA: Springer U.S., 1978. Vol. 5. P. 257–270.
7. Diskin Z., Kokaly S., Maibaum T. Mapping-aware megamodeling: Design patterns and laws // Software language engineering: 6th Conference (International) Proceedings / Eds. M. Erwig, R. F. Paige, E. Van Wyk. — Lecture notes in computer science ser. — Springer, 2013. Vol. 8225. P. 322–343.
8. Косляков А., Свит У., Сеймур С., Бимер С. Системная инженерия. Принципы и практика / Пер. с англ. — М.: ДМК-Пресс, 2014. 636 с. (Kossiakoff A., Sweet W. N., Seymour S., Biemer S. M. Systems engineering principles and practice. — 2nd ed. — New York, NY, USA: John Wiley, 2011. 560 p.)
9. Bézivin J., Jouault F., Rosenthal P., Valduriez P. Modeling in the large and modeling in the small // Model driven architecture: European MDA Workshops on Foundations and Applications Proceedings / Eds. U. Aßmann, M. Aksit, A. Rensink. — Lecture notes in computer science ser. — Springer, 2005. Vol. 3599. P. 33–46.
10. Neema S., Szűtanovits J., Karsai G., Butts K. Constraint-based design-space exploration and model synthesis // 3rd Conference (International) on Embedded Software Proceedings / Eds. R. Alur, I. Lee. — Lecture notes in computer science ser. — Springer, 2003. Vol. 2855. P. 290–305.
11. Vanherpen K., Denil J., De Meulenaere P., Vangheluwe H. Design-space exploration in MDE: An initial pattern catalogue // 1st Workshop (International) on Combining Modelling with Search- and Example-Based Approaches Proceedings / Eds. R. Paige, M. Kessentini, P. Langer, M. Wimmer. — CEUR Workshop Proceedings ser. — Valencia, Spain, 2014. Vol. 1340. P. 42–51.
12. ГОСТ 14.205-83. Технологичность конструкции изделий. Термины и определения. — М.: Стандартинформ, 2009. 22 с.
13. Альтиуллер Г. С. Творчество как точная наука. — М.: Советское радио, 1979. 116 с.
14. Guitart R., van den Bril L. Décompositions et lax-complétions // Cah. Topologie Géométrie Différentielle Catégoriques, 1977. Vol. 18. No. 4. P. 333–407.
15. Adámek J., Herrlich H., Strecker G. E. Abstract and concrete categories. — New York, NY, USA: John Wiley, 1990. 507 p.

16. Маклейн С. Категории для работающего математика / Пер. с англ. — М.: Физматлит, 2004. 352 с. (*Mac Lane S. Categories for the working mathematician.* — New York, NY, USA: Springer, 1978. 317 p.)
17. Андриюшкевич С. К., Журавлев С. С., Золотухин Е. П., Ковалев С. П., Окольников В. В., Рудометов С. В. Разработка системы мониторинга с использованием имитационного моделирования // Проблемы информатики, 2010. № 4. С. 65–75.
18. Baez J., Stay M. Physics, topology, logic and computation: A Rosetta stone // *New structures for physics* / Ed. В. Coecke. — Lecture notes in physics ser. — Springer, 2011. Vol. 813. P. 95–172.
19. Ковалёв С. П. Системный анализ жизненного цикла больших информационно-управляющих систем // Автоматика и телемеханика, 2013. № 9. С. 98–118.
20. Gross J., Chlipala A., Spivak D. I. Experience implementing a performant category-theory library in Coq // 5th Conference (International) on Interactive Theorem Proving Proceedings / Eds. G. Klein, R. Gamboa. — Lecture notes in computer science ser. — Springer, 2014. Vol. 8558. P. 275–291.

Поступила в редакцию 11.12.17

## CATEGORY THEORY AS A MATHEMATICAL PRAGMATICS OF MODEL-BASED SYSTEMS ENGINEERING

S. Kovalyov

Institute of Control Sciences, Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation

**Abstract:** Mathematical device built upon the category theory is developed which was previously proposed to formally describe and rigorously explore procedures of employing models in engineering that constitute the pragmatics of model-based systems engineering. The essence of the device consists in mathematical representation of assembly drawings (megamodels of systems) as diagrams in categories whose objects are models, and morphisms represent actions associated with assembling system models from component models. Category-theoretical methods for solving direct and inverse pragmatic problems of assembling systems are proposed and explored. The key role of the diagram monad is revealed. Special attention is paid to the problem of recovering the configuration of a given system, taking into account technological limitations of the assembling means and procedures. A number of key systems engineering concepts are matched with relevant constructions of the category theory.

**Keywords:** model-based systems engineering; pragmatics; megamodel; category theory; configuration recovery problem; diagram monad

**DOI:** 10.14357/19922264180112

### References

1. Gianni, D., A. D’Ambrogio, and A. Tolk, eds. 2014. *Modeling and simulation-based systems engineering handbook*. London: CRC Press. 513 p.
2. Selic, B. 2003. The pragmatics of model-driven development. *IEEE Software* 20(5):19–25.
3. Levenchuk, A. I. 2015. *Sistemnoinzhenernoe myshlenie* [Systems engineering thinking]. Moscow: TechInvestLab. 305 p.
4. Kovalyov, S. P. 2017. Metody teorii kategoriy v model’no-orientirovannoy sistemnoy inzhenerii [Methods of category theory in model-based systems engineering]. *Informatika i ee Primeneniya — Inform. Appl.* 11(3):42–50.
5. IEC 81346-1:2009. 2009. Industrial Systems, Installations and Equipment and Industrial Products — Structuring Principles and Reference Designations — Part 1: Basic Rules. Geneva: ISO. 168 p.
6. Ginali, S., and J. Goguen. 1978. A categorical approach to general systems. *Applied general systems research: Recent development and trends*. Ed. G. J. Klir. NATO conference ser. Boston, MA: Springer U.S. 5:257–270.
7. Diskin, Z., S. Kokaly, and T. Maibaum. 2013. Mapping-aware megamodeling: Design patterns and laws. *Software language engineering: 6th Conference (International) Proceedings*. Eds. M. Erwig, R. F. Paige, and E. Van Wyk. Lecture notes in computer science ser. Springer. 8225:322–343.
8. Kossiakoff, A., W. N. Sweet, S. Seymour, and S. M. Biemer. 2011. *Systems engineering principles and practice*. 2nd ed. New York, NY: John Wiley. 560 p.
9. Bézivin, J., F. Jouault, P. Rosenthal, and P. Valduriez. 2005. Modeling in the large and modeling in the small. *Model driven architecture: European MDA Workshops on Foundations and Applications Proceedings*. Eds. U. Almann, M. Aksit, and A. Rensink. Lecture notes in computer science ser. Springer. 3599:33–46.
10. Neema, S., J. Sztipanovits, G. Karsai, and K. Butts. 2003. Constraint-based design-space exploration and model synthesis. *3rd Conference (International) on Embedded*

- Software Proceedings*. Eds. R. Alur and I. Lee. Lecture notes in computer science ser. Springer. 2855:290–305.
11. Vanherpen, K., J. Denil, P. De Meulenaere, and H. Vangheluwe. 2014. Design-space exploration in MDE: An initial pattern catalogue. *1st Workshop (International) on Combining Modelling with Search- and Example-Based Approaches Proceedings*. Eds. R. Paige, M. Kessentini, P. Langer, and M. Wimmer. CEUR Workshop Proceedings ser. Valencia, Spain. 1340:42–51.
  12. GOST 14.205-83. 2009. Tekhnologichnost' konstruksii izdeliy. Terminy i opredeleniya [Technological efficiency of products design. Terms and definitions]. Moscow: Standardinform Publ. 22 p.
  13. Altshuller, G. 1984. *Creativity as an exact science: The theory of the solution of inventive problems*. Amsterdam: Gordon and Breach Science Publ. 324 p.
  14. Guitart, R., and L. van den Bril. 1977. Décompositions et lax-complétions. *Cah. Topologie Géométrie Différentielle Catégoriques* 18(4):333–407.
  15. Adámek, J., H. Herrlich, and G. E. Strecker. 1990. *Abstract and concrete categories*. New York, NY: John Wiley. 507 p.
  16. Mac Lane, S. 1978. *Categories for the working mathematician*. New York, NY: Springer. 317 p.
  17. Andryushkevich, S. K., S. S. Zhuravlev, E. P. Zolotukhin, S. P. Kovalyov, V. V. Okol'nishnikov, and S. V. Rudometov. 2010. Razrabotka sistemy monitoringa s ispol'zovaniem imitatsionnogo modelirovaniya [Development of a monitoring system using simulation]. *Problemy informatiki [Problems of Informatics]* 4:65–75.
  18. Baez, J., and M. Stay. 2011. Physics, topology, logic and computation: A Rosetta stone. *New structures for physics*. Ed. B. Coecke. Lecture notes in physics ser. Springer. 813:95–172.
  19. Kovalev, S. P. 2013. Systems analysis of life cycle of large-scale information-control systems. *Automat. Rem. Contr.* 74(9):1510–1524.
  20. Gross, J., A. Chlipala, and D. I. Spivak. 2014. Experience implementing a performant category-theory library in Coq. *5th Conference (International) on Interactive Theorem Proving Proceedings*. Eds. G. Klein and R. Gamboa. Lecture notes in computer science ser. Springer. 8558:275–291.

Received December 11, 2017

## Contributor

**Kovalyov Sergey P.** (b. 1972) — Doctor of Science in physics and mathematics, leading scientist, Institute of Control Sciences, Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation; kovalyov@nm.ru



# БАЙЕСОВСКИЕ МОДЕЛИ ТЕСТИРОВАНИЯ БОЛЬШИХ ГРУПП ОБСЛУЖИВАЮЩИХ ПРИБОРОВ\*

А. А. Кудрявцев<sup>1</sup>, О. В. Шестаков<sup>2</sup>

**Аннотация:** Рассматривается новый подход к задачам верификации заявленных показателей функционирования больших групп сложных обслуживающих приборов в условиях невозможности (или высокой стоимости) тотальной проверки. Применяемый метод основан на изучении асимптотики статистик в рамках байесовского подхода к задачам теории массового обслуживания. На примере коэффициента готовности в модели  $M/M/1$  приводится утверждение, относящееся к классу теорем переноса, дающее возможность сравнивать реальные данные при стендовых испытаниях сложных агрегатов с теоретическими предсказаниями. Рассматривается специфика применения метода с точки зрения производителя приборов и заказчика большой партии однотипных приборов. Описываемый метод и результаты могут найти применение не только в области байесовских задач массового обслуживания, но и в других областях, использующих рандомизацию базовых параметров модели.

**Ключевые слова:** байесовский подход; системы массового обслуживания; смешанные распределения; теорема переноса; статистические методы

**DOI:** 10.14357/19922264180113

## 1 Введение

В настоящее время все более остро встает проблема анализа адекватности функционирования систем и сетей массового обслуживания. Это связано прежде всего с высокой сложностью современных систем и агрегатов, которая не дает возможности проверить функциональность на всех допустимых входных параметрах. Кроме того, при необходимости верификации больших партий обслуживающих приборов возникают естественные трудности, связанные с временными и финансовыми затратами на проверку возможности допуска всех элементов партии к эксплуатации. По этим причинам единственным реальным средством анализа пригодности больших групп сложных приборов, по всей видимости, является вероятностно-статистический. Если дополнительно учесть различные вариации, вызванные технологическими особенностями при производстве приборов и изменчивостью среды, в которой происходит функционирование, также разумно дополнить статистическую составляющую тестирования байесовским подходом [1], предполагающим рандомизацию базовых параметров [2].

Рассмотрим следующую модель. Заказчику требуется большая партия обслуживающих приборов,

соответствующая некоторому техническому заданию. Заказ поступает производителю. Требуется произвести две проверки адекватности приборов техническому заданию: одну на стороне производителя перед отправкой опытного образца в серию, другую на стороне заказчика при получении изготовленной партии. Поскольку партия мыслится большой, а приборы сложными, проводится выборочное тестирование на стенде, имитирующем условия реального функционирования. При этом у производителя и заказчика, естественно, существуют особенности и отличия методики проверки.

В данной работе рассматриваются теоретические статистические результаты на примере параметра загрузки  $\rho$  системы  $M/M/1$ , равного отношению интенсивности входящего потока  $\lambda$  к интенсивности обслуживания  $\mu$ , в предположении, что параметры модели независимы и имеют известные априорные распределения.

## 2 Стендовые испытания производителя

Рассматривается задача верификации удовлетворения тестового образца обслуживающего при-

\*Работа выполнена при частичной финансовой поддержке РФФИ (проект 17–07–00577).

<sup>1</sup>Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, pubigena@mail.ru

<sup>2</sup>Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, oshestakov@cs.msu.ru

бора требованиям технического задания заказчика для последующего запуска образца в серию. Для этого на испытательном стенде формируется набор интенсивностей входящего потока требований, являющийся конкретной реализацией выборки  $X_1, \dots, X_n$  параметра потока требований  $\lambda$  из априорного распределения, предварительно изученного заказчиком и описанного в тексте технического задания. Конкретный параметр обслуживания исследуемого прибора является реализацией случайной величины  $\mu$ , априорное распределение которой обусловлено техническим регламентом на производстве. Результаты стендовых испытаний прибора сравниваются с теоретическими статистическими прогнозами, что дает возможность сделать вывод о целесообразности запуска тестового образца в серию.

В данном случае теоретические статистические результаты, касающиеся параметра загрузки  $\rho$ , основываются на выборке  $X_1/\mu, \dots, X_n/\mu$ .

### 3 Стендовые испытания заказчика

Рассматривается задача стендовых испытаний большой партии обслуживающих приборов, поступивших для верификации удовлетворения требованиям заказчика. Поскольку партия представляется достаточно объемной, производить тотальную проверку приборов нецелесообразно ввиду высокой стоимости и больших временных затрат. В связи с этим на стенде фиксируется некоторая интенсивность входящего потока требований  $\lambda$  и формируется конкретная реализация выборки приборов из партии, которая подлежит исследованию. Каждый прибор обладает своим параметром обслуживания  $Y_1, \dots, Y_n$ , являющимся реализацией параметра  $\mu$  с известным распределением.

Предполагается, что заказчик предварительно исследовал условия, в которых должны функционировать приборы, и, таким образом, знает априорное распределение интенсивности входящего потока требований  $\lambda$ . Априорное распределение параметров обслуживания приборов задается техническим регламентом на производстве. Результаты работы каждого прибора при заданной интенсивности требований проходят статистический анализ, результатом которого является вывод о допустимости принятия всей партии в эксплуатацию.

В данном случае теоретические статистические результаты, касающиеся параметра загрузки  $\rho$ , основываются на выборке  $\lambda/Y_1, \dots, \lambda/Y_n$ .

## 4 Основные результаты

Из постановки задачи следует, что параметр входящего потока  $\lambda$  и параметр обслуживания  $\mu$  — неотрицательные собственные случайные величины, не имеющие атома в нуле. Требование  $P(\lambda = 0) = P(\mu = 0) = 0$  означает, что прибор не может с положительной вероятностью предназначаться для вечного хранения на складе и не должен игнорировать поступающие требования.

Предположим, что параметр входящего потока  $\lambda$  и одинаково распределенные параметры обслуживания  $Y_1, \dots, Y_n$  независимы. Обозначим

$$A(u) = P(\lambda < u); \quad Z_n = \frac{1}{n} \sum_{i=1}^n Y_i^{-1}.$$

**Теорема.** Пусть  $a = EY_i^{-1}$ ,  $0 < \sigma^2 = DY_i^{-1} < \infty$ . Тогда при  $n \rightarrow \infty$

$$\sup_{x \in \mathbb{R}} \left| P \left( \sqrt{n} \frac{\lambda(Z_n - a)}{\sigma} < x \right) - F(x) \right| \rightarrow 0, \quad (1)$$

где  $F(x)$  — функция распределения, соответствующая характеристической функции

$$f(t) = \int_0^{\infty} e^{-u^2 t^2 / 2} dA(u).$$

**Доказательство.** Обозначим через  $h_n(t)$  характеристическую функцию случайной величины  $\sqrt{n}(\lambda(Z_n - a))/\sigma$ , а через  $f_n(t)$  — характеристическую функцию случайной величины  $\lambda(Z_n - a)/\sigma$ . Для произвольного  $t \in \mathbb{R}$  и  $M > 0$

$$\begin{aligned} |f_n(t) - f(t)| &= \left| \int_0^{\infty} h_n(ut) dA(u) - \int_0^{\infty} e^{-u^2 t^2 / 2} dA(u) \right| \leq \\ &\leq \int_0^{\infty} |h_n(ut) - e^{-u^2 t^2 / 2}| dA(u) = \\ &= \int_0^M |h_n(ut) - e^{-u^2 t^2 / 2}| dA(u) + \\ &+ \int_M^{\infty} |h_n(ut) - e^{-u^2 t^2 / 2}| dA(u) \equiv I_1 + I_2. \end{aligned}$$

Далее

$$I_2 \leq 2 \int_M^{\infty} dA(u),$$

и для произвольного  $\varepsilon > 0$  можно выбрать такое  $M = M(\varepsilon)$ , что  $I_2 < \varepsilon/2$ . Поскольку в силу центральной предельной теоремы  $h_n(v) \rightarrow e^{-v^2/2}$  равномерно на любом конечном отрезке, найдется  $N = N(\varepsilon)$  такое, что  $|h_n(ut) - e^{-u^2t^2/2}| < \varepsilon/2$  для всех  $n > N$  и всех  $u \in [0, M]$ . Следовательно,  $I_1 < \varepsilon/2$  при  $n > N$  и, таким образом,  $|f_n(t) - f(t)| < \varepsilon$ . Так как функция распределения  $F(x)$  непрерывна, выполнено (1).

Теорема доказана.

**Замечание 1.** В теореме  $\lambda\sqrt{n}Z_n$  центрируется случайной величиной  $a\lambda\sqrt{n}$ . Для построения доверительных интервалов было бы желательно центрировать  $\sqrt{n}\lambda Z_n$  некоторой неслучайной возрастающей последовательностью чисел  $a_n$ . Однако сделать этого нельзя.

Действительно, предположим, что такая последовательность существует. Тогда

$$\sqrt{n}\lambda Z_n - a_n = \sqrt{n}\lambda(Z_n - a) + (\sqrt{n}\lambda a - a_n).$$

По доказанной теореме первое слагаемое имеет некоторое предельное распределение. Рассмотрим второе слагаемое. Если  $a_n = o(\sqrt{n})$  или наоборот  $\sqrt{n} = o(a_n)$ , то очевидно, что  $P(|\lambda a\sqrt{n} - a_n| > A) \rightarrow 1$  при  $n \rightarrow \infty$  для любого  $A > 0$ . Если же  $a_n = O(\sqrt{n})$ , т. е. существуют такие положительные числа  $c$  и  $C$ , что  $c\sqrt{n} < a_n < C\sqrt{n}$  для всех  $n > 1$ , то  $\sqrt{n}(\lambda a - C) < \sqrt{n}\lambda a - a_n < \sqrt{n}(\lambda a - c)$  п.в. и  $P(|\sqrt{n}(\lambda a - C)| > A) \rightarrow 1$ ,  $P(|\sqrt{n}(\lambda a - c)| > A) \rightarrow 1$  при  $n \rightarrow \infty$ , т. е. снова  $P(|\lambda a\sqrt{n} - a_n| > A) \rightarrow 1$ . Таким образом, предельного распределения у последовательности  $\sqrt{n}\lambda Z_n - a_n$  не существует.

**Замечание 2.** Теорема дает возможность исследовать асимптотическое поведение статистики, основанной на выборке  $\lambda/Y_1, \dots, \lambda/Y_n$ , в стендовых испытаниях заказчика. Очевидно, что теорема останется

верна при замене случайной величины  $\lambda$  на  $\mu^{-1}$ , а случайных величин  $Y_i^{-1}$  на  $X_i$ , что соответствует стендовым испытаниям производителя.

**Замечание 3.** Как видно, предельное распределение для испытаний заказчика однозначно определяется априорным распределением параметра входящего потока  $\lambda$  ( $\mu^{-1}$  для испытаний производителя), что дает возможность применять результаты теоремы для широкого класса распределений с конечными первыми моментами случайных величин  $Y_i^{-1}$  ( $X_i$  соответственно).

Для получения теоретических прогнозов можно использовать исследованные ранее вероятностные характеристики для некоторых априорных распределений  $Y_i$ , например weibullовского [3]. Однако математическое ожидание  $Y_i^{-1}$  может не существовать [4], например для показательного распределения параметра обслуживания.

## Литература

1. Sutton Ch., Jordan M. I. Bayesian inference for queueing networks and modeling of internet services // Ann. Appl. Stat., 2011. Vol. 5. No. 1. P. 254–282.
2. Кудрявцев А. А., Шоргин С. Я. Байесовские модели в теории массового обслуживания и надежности. — М.: ФИЦ ИУ РАН, 2015. 76 с.
3. Кудрявцев А. А., Титова А. И. Байесовские модели массового обслуживания и надежности: вырожденно-вейбулловский случай // Информатика и её применения, 2016. Т. 10. Вып. 4. С. 68–71.
4. Кудрявцев А. А., Шоргин С. Я. Байесовский подход к анализу систем массового обслуживания и показателей надежности // Информатика и её применения, 2007. Т. 1. Вып. 2. С. 76–82.

Поступила в редакцию 12.06.17

## BAYESIAN MODELS FOR TESTING LARGE GROUPS OF SERVICE DEVICES

A. A. Kudryavtsev<sup>1</sup> and O. V. Shestakov<sup>1,2</sup>

<sup>1</sup>Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

<sup>2</sup>Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

**Abstract:** The paper considers a new approach to the verification tasks for the declared performance indicators of large groups of complex servicing devices when the total verification is impossible or expensive. The method used is

based on the study of the asymptotics of statistics in the framework of the Bayesian approach to the problems of the queuing theory. By the example of the readiness coefficient in the  $M|M|1$  model, the authors prove the proposition relating to the class of transfer theorems which makes it possible to compare real data in bench tests of complex aggregates with theoretical predictions. The specifics of the application of the method from the point of view of a device manufacturer and a customer of a large batch of similar devices are considered. The described method and results can find application not only in the field of Bayesian queuing problems but also in other areas that use randomization of the basic parameters of the model.

**Keywords:** Bayesian approach; mass service systems; mixed distributions; transfer theorem; statistical methods

**DOI:** 10.14357/19922264180113

## Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 17-07-00577).

## References

1. Sutton, Ch., and M. I. Jordan. 2011. Bayesian inference for queueing networks and modeling of internet services. *Ann. Appl. Stat.* 5(1):254–282.
2. Kudryavtsev, A. A., and S. Ya. Shorgin. 2015. *Bayesovskie modeli v teorii massovogo obsluzhivaniya i nadezhnosti* [Bayesian models in mass service and reliability theories]. Moscow: FIC IU RAN. 76 p.
3. Kudryavtsev, A. A., and A. I. Titova. 2016. Bayesovskie modeli massovogo obsluzhivaniya i nadezhnosti: vyrozhdennno-veybullovskiy sluchay [Bayesian queuing and reliability models: Degenerate-Weibull case]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):68–71.
4. Kudryavtsev, A. A., and S. Ya. Shorgin. 2007. Bayesovskiy podkhod k analizu sistem massovogo obsluzhivaniya i pokazateley nadezhnosti [Bayesian approach to queueing systems and reliability characteristics]. *Informatika i ee Primeneniya — Inform. Appl.* 1(2):76–82.

Received June 12, 2017

## Contributors

**Kudryavtsev Alexey A.** (b. 1978) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; nubigena@mail.ru

**Shestakov Oleg V.** (b. 1976) — Doctor of Science in physics and mathematics, associate professor, Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; oshestakov@cs.msu.su



# РАЗРАБОТКА И ПРЕДВАРИТЕЛЬНОЕ ИССЛЕДОВАНИЕ СТОХАСТИЧЕСКОЙ ПОЛУМАРКОВСКОЙ МОДЕЛИ УПРАВЛЕНИЯ ЗАПАСОМ НЕПРЕРЫВНОГО ПРОДУКТА ПРИ ПОСТОЯННО ПРОИСХОДЯЩЕМ ПОТРЕБЛЕНИИ

П. В. Шнурков<sup>1</sup>, А. Ю. Егоров<sup>2</sup>

**Аннотация:** Проводится построение и предварительное исследование дискретной стохастической полумарковской модели, описывающей функционирование некоторой системы управления запасом непрерывного продукта при постоянно происходящем потреблении. Модель представляет собой пару случайных процессов  $(x(t), \zeta(t))$ , где основной процесс  $x(t)$  описывает объем запаса в системе в момент времени  $t$ , а сопровождающий полумарковский случайный процесс  $\zeta(t)$  определяется по состояниям основного непосредственно после пополнений. Задача оптимального управления ставится по отношению к стационарному показателю, имеющему характер средней удельной прибыли, полученной при эволюции исходной системы управления запасом.

**Ключевые слова:** управление запасом; полумарковский случайный процесс; стационарный стоимостный функционал; оптимальное управление стохастическими системами

**DOI:** 10.14357/19922264180114

## 1 Введение

Настоящая работа посвящена созданию и предварительному исследованию новой стохастической полумарковской модели управления запасом непрерывного продукта при постоянно происходящем потреблении. Несмотря на обилие работ, посвященных общим проблемам управления запасом, исследований стохастических полумарковских моделей как в отечественной, так и в зарубежной научной литературе немного. Краткая характеристика нескольких современных исследований, в которых рассматриваются проблемы управления запасом в различных стохастических полумарковских моделях, приведена в приложении [1].

Заметим, что необходимые сведения по общей теории полумарковских процессов можно найти в монографии [2]. Из современных изданий можно упомянуть работу [3]. Классические результаты по теории управления полумарковскими процессами изложены в работах [4, 5].

Целый ряд моделей управления запасом непрерывного продукта в форме регенерирующих и полумарковских случайных процессов был разработан и исследован в работах П. В. Шнуркова и его соавторов. Библиография этих работ также приведена в приложении [1]. Стохастическая полумарковская модель, наиболее близкая к исследуемой

в настоящей статье, рассмотрена П. В. Шнурковым и А. В. Ивановым в [6]. Отметим, что кроме непосредственного исследования полумарковской модели управления запасом в этой работе был доказан общий теоретический результат о представлении стационарного стоимостного показателя качества управления в виде дробно-линейного интегрального функционала от вероятностных распределений, задающих стратегию управления полумарковским процессом. Этот результат играет важную роль в решении задачи управления в рассматриваемой полумарковской модели.

Настоящая работа продолжает упомянутый цикл исследований регенерирующих и полумарковских моделей управления запасами. При этом полумарковская модель управления запасом, исследуемая в ней, является более глубоким и сложным вариантом модели, рассмотренной в [6]. Основные отличия данной модели от предшествующей заключаются в следующем:

- (1) в новой модели предполагается, что потребление продукта продолжается постоянно в течение всего времени жизни системы;
- (2) интенсивность потребления принимает два разных значения на различных периодах функционирования рассматриваемой системы.

<sup>1</sup>Национальный исследовательский университет «Высшая школа экономики», pshnurkov@hse.ru

<sup>2</sup>Национальный исследовательский университет «Высшая школа экономики», ayuegorov@hse.ru

Указанные особенности приводят к тому, что вероятностный анализ новой версии модели оказывается существенно более сложным. Отметим также, что структура рассматриваемой модели управления запасом имеет общие черты со стохастической моделью с плановыми переключениями, которая была предложена П. В. Шнурковым в работе [7].

Основная цель первой части проведенного исследования, изложенной в настоящей работе, заключается в нахождении явного аналитического представления для стационарного стоимостного показателя эффективности управления, который по своему экономическому содержанию представляет собой среднюю удельную прибыль, образующуюся при управлении данной системой.

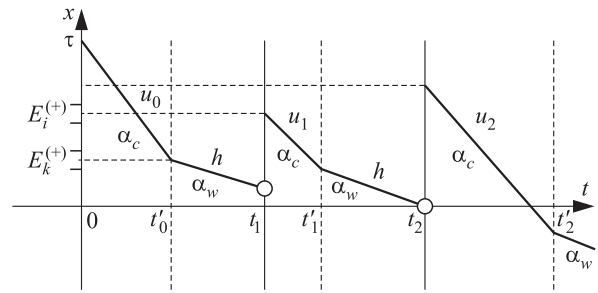
## 2 Общее описание и некоторые особенности функционирования рассматриваемой системы управления запасом непрерывного продукта

Рассмотрим систему, предназначенную для хранения и поставки потребителю запаса определенного вида непрерывного продукта. Предположим, что объем запаса продукта в данной системе в момент времени  $t$ , который будет обозначаться  $x(t)$ , представляет собой некоторый случайный процесс, формальное описание которого будет приведено в дальнейшем. Случайный процесс  $x(t)$  принимает значения в множестве  $(-\infty, \tau]$ , где  $\tau > 0$  — фиксированная максимальная вместимость хранилища. Как принято в классической теории управления запасами, отрицательная величина запаса означает отложенный спрос.

**Определение 1.** Периодом эволюции системы будем называть интервал времени  $[t_n; t_{n+1})$  между последовательными моментами пополнения запаса.

Каждый период эволюции естественным образом допускает разбиение на две части: период времени от момента последнего пополнения запаса до момента заказа на очередную поставку (период чистого потребления) и период времени от момента заказа до следующего пополнения (период задержки поставки). В течение периода задержки поставки потребление продолжается. Длительность задержки поставки является известной детерминированной величиной, обозначаемой через  $h$ .

Пусть потребление происходит непрерывно с постоянной скоростью  $\alpha_c > 0$  в течение периода



**Рис. 1** Возможная траектория процесса  $x(t)$ , характеризующая функционирование исследуемой системы

чистого потребления. В момент заказа скорость потребления меняется и становится равной величине  $\alpha_w = const$ ,  $\alpha_w \geq 0$ , причем  $\alpha_w \leq \alpha_c$ . В частном случае может быть  $\alpha_w = \alpha_c$ , что соответствует непрерывному потреблению с одной фиксированной скоростью на всем периоде жизни системы. Предполагается, что непосредственное пополнение запаса происходит мгновенно в момент окончания очередного периода эволюции системы  $t_n$ ,  $n = 1, 2, \dots$ . Возможная траектория процесса  $x(t)$  изображена на рис. 1.

Для удобства дальнейших рассуждений и аналитических выводов введем следующие предположения:  $\exists L > 0$  такое, что заданные величины  $\tau$  и  $(\alpha_w h)$  делятся нацело на  $L$ . При построении модели используем дискретизацию: разобьем множество значений объема запаса на конечное число подмножеств — полузамкнутых интервалов, длины которых, кроме одного бесконечного интервала, одинаковы и равны  $L$  (содержание данных предположений подробно прокомментировано в замечании 1 приложения [1]).

Будем использовать следующие обозначения для вводимых множеств возможных значений процесса  $x(t)$ :

$$E_0^{(-)} = (\tau_1^{(-)}, \tau_0^{(-)}) = (-L, 0), \dots$$

$$E_k^{(-)} = (\tau_{k+1}^{(-)}, \tau_k^{(-)}) = ((-k+1)L, -kL),$$

$$k \in \{1, \dots, N_1 - 1\};$$

$$E_{N_1}^{(-)} = (\tau_{N_1+1}^{(-)}, \tau_{N_1}^{(-)}) = (-\infty, -N_1 L);$$

$$E_k^{(+)} = [\tau_k^{(+)}, \tau_{k+1}^{(+)}] = [kL, (k+1)L),$$

$$k \in \{0, \dots, N_0 - 1\};$$

$$E_{N_0}^{(+)} = [\tau_{N_0}^{(+)}, \tau_{N_0+1}^{(+)}] = [N_0 L, \tau].$$

Зададим в каждом множестве  $E_i^{(+)}$  фиксированную точку  $\hat{\tau}_i^{(+)} \in E_i^{(+)} = [\tau_i^{(+)}, \tau_{i+1}^{(+)})$ , которую

назовем  $i$ -м уровнем планирования объема запаса,  $i \in \{0, \dots, N_0\}$ .

Для описания функционирования системы и построения ее математической модели необходимо задать исходные стоимостные характеристики. Предположим, что заданы следующие функции:

$c_1(x)$  — затраты, связанные с хранением продукта объема  $x$  в системе в единицу времени;  $c_1(x) = 0, x \leq 0$ ;

$c_2(x)$  — штрафы, связанные с дефицитом продукта объема  $x$  в системе в единицу времени;  $c_2(x) = 0, x \geq 0$ ;

$d(x)$  — доход, связанный с поставкой одной единицы продукта потребителю при условии, что объем имеющегося продукта равен  $x$ ; при этом представления данной функции при  $x \geq 0$  (реальный запас) и  $x \leq 0$  (дефицит) определяются выражениями  $d_1(x)$  и  $d_2(x)$  соответственно.

### 3 Формальное построение полумарковской стохастической модели, описывающей функционирование исследуемой системы

Зададим исходное вероятностное пространство  $(\Omega, \mathcal{F}, P)$ , на котором будут определены все необходимые для построения модели стохастические объекты. Отметим прежде всего, что основной стохастический объект  $x(t)$ , характеризующий уровень запаса в системе в момент времени  $t$ , представляет собой случайный процесс, заданный на данном вероятностном пространстве и множестве значений параметра времени:

$$x(t) = x(\omega, t), \quad \omega \in \Omega, \quad t \in T = [0, \infty).$$

Введем также две последовательности случайных величин:  $\{t_n = t_n(\omega), \omega \in \Omega\}_{n=0}^{\infty}$  — последовательность случайных моментов времени, в которые производится пополнение запаса, и  $\{t'_n = t'_n(\omega), \omega \in \Omega\}_{n=0}^{\infty}$  — последовательность случайных моментов времени, в которые производится заказ на пополнение запаса,  $t_n \leq t'_n < t_{n+1}$ ,  $n = 0, 1, 2, \dots$ ;  $t_0 = 0$ .

Предположим, что в момент очередного пополнения запаса  $t_n = t_n(\omega)$ ,  $\omega \in \Omega$ , состояние системы (уровень запаса) принимает фиксированное произвольное значение  $x(t_n) = x \in E_i^{(+)}$ . Дальнейшее описание модели будем проводить при условии, что

реализуется указанное событие ( $x(t_n) = x \in E_i^{(+)}$ ). Предполагается, что траектория процесса  $x(t)$  в моменты пополнения запаса  $t_n$ ,  $n = 0, 1, 2, \dots$ , непрерывна справа, т. е.  $\lim_{t \rightarrow t_n+0} x(t) = x(t_n)$ .

Зафиксируем номер  $i \in \{0, 1, \dots, N_0\}$  и введем на пространстве  $(\Omega, \mathcal{F}, P)$  последовательность независимых одинаково распределенных случайных величин  $\xi_{i,n} = \xi_{i,n}(\omega)$ ,  $n = 0, 1, 2, \dots$ ,  $\omega \in \Omega$ , принимающих значения в множестве  $U = [0, \infty)$ . Обозначим через  $G_i(t) = P(\xi_{i,n} < t)$  функции распределения случайных величин  $\{\xi_{i,n}\}_{n=0}^{\infty}$ . Кроме того, дополнительно обозначим  $G_i(0) = g_{i,0}$ ,  $0 \leq g_{i,0} < 1$ , где  $g_{i,0}$  — фиксированный параметр, вводимый для удобства дальнейших рассуждений.

Будем интерпретировать случайную величину  $\xi_{i,n}$  как время от момента пополнения  $t_n$  до момента следующего заказа на пополнение. Тогда момент заказа на пополнение есть  $t'_n = t_n + \xi_{i,n}$ , и с единичной вероятностью выполняется соотношение  $t_n \leq t'_n < t_{n+1}$ . Случайная величина  $\xi_{i,n}$  служит параметром управления в рассматриваемой стохастической модели, или решением, принимаемым в момент  $t_n$ , при условии, что  $x(t_n) = x \in E_i^{(+)}$ .

Зафиксируем реализацию случайной величины  $\xi_{i,n} : \xi_{i,n} = u_i$ . При этом условии траектория процесса  $x(t)$  на интервале времени  $[t_n; t'_n)$  определяется равенством:

$$x(t) = x - \alpha_c(t - t_n), \quad t_n \leq t < t'_n = t_n + u_i.$$

Предположим, что

$$x(t'_n) = x(t_n + u_i) = x' \in E_k^{(+)},$$

где  $k = 0, 1, \dots, i$ , т. е. в момент заказа  $t'_n$  объем продукта принимает значение  $x' \in E_k^{(+)}$ . Описание модели для случая, когда  $x' \in E_k^{(-)}$ , производится совершенно аналогично.

В момент  $t'_n$  начинается так называемый период задержки поставки. В течение данного периода поставщик готовит новую партию продукта для пополнения. Непосредственное пополнение запаса происходит мгновенно в конечный момент периода задержки.

В период задержки потребление не прекращается, а продолжается с заданной скоростью  $\alpha_w$ ,  $\alpha_w \leq \alpha_c$ . Поскольку длительность периода задержки не случайна и известна, объем потребления продукта в течение этого периода также известен и равен  $\alpha_w h$ . Таким образом, управляющему рассматриваемой системы и поставщику известен объем запаса в момент окончания периода задержки, т. е. непосредственно перед пополнением. Действительно, если выполняются принятые выше условия

$x(t_n) = x \in E_i^{(+)}$ ,  $\xi_{i,n} = t'_n - t_n = u_i$ , то траектория процесса  $x(t)$  в течение периода задержки  $[t'_n, t_{n+1})$  определяется равенством:

$$x(t) = x - \alpha_c u_i - \alpha_w(t - t'_n), \quad t'_n \leq t < t'_n + h = t_{n+1}. \quad (1)$$

Тогда значение процесса  $x(t)$  непосредственно перед пополнением формально определяется как предел слева функции  $x(t)$ , задаваемой формулой (1), в точке  $t_{n+1}$ , а именно:

$$x(t_{n+1} - 0) = x - \alpha_c u_i - \alpha_w h.$$

Таким образом, в рассматриваемой модели величину объема заказа на пополнение можно планировать непосредственно в момент заказа.

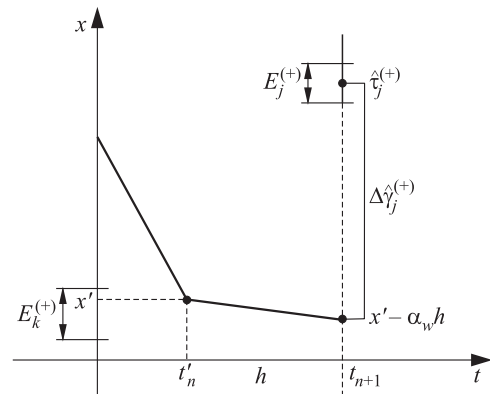
Перейдем к описанию процедуры пополнения запаса. При фиксированном условии  $x(t'_n) = x' \in E_k^{(+)}$  рассмотрим систему случайных событий  $\{B_{kj}^{(+)}\}_{j=0}^{N_0}$ , имеющих заданные вероятности

$$P(B_{kj}^{(+)} | x(t'_n) = x' \in E_k^{(+)}) = \beta_{kj}^{(+)}, \quad j = \{0, 1, \dots, N_0\}, \quad \sum_j \beta_{kj}^{(+)} = 1.$$

**Замечание 1.** Если  $x' - \alpha_w h \in E_m^{(+)}$ , где  $m = 0, 1, \dots, k$ , то предполагается, что  $\beta_{kj}^{(+)} = 0$  при  $j = \{0, 1, \dots, m\}$ . Это условие означает, что при пополнении запаса объективно не может планироваться уменьшение его объема.

Событие  $B_{kj}^{(+)}$  связано с траекторией процесса  $x(t)$  следующим образом. Если выполняется условие, что  $x(t'_n) = x' \in E_k^{(+)}$ , то реализация случайного события  $B_{kj}^{(+)}$  означает, что выбирается значение  $\hat{\tau}_j^{(+)} \in E_j^{(+)}$  из возможных значений уровней планирования. С точки зрения содержания модели в момент  $t'_n$  происходит планирование объема заказа. План имеет случайный характер: в соответствии с этим планом объем запаса непосредственно после пополнения должен достигнуть уровня  $\hat{\tau}_j^{(+)}$  с вероятностью  $\beta_{kj}^{(+)}$ . Как уже отмечалось, если  $x' - \alpha_w h \in E_m^{(+)}$ ,  $m = 0, 1, \dots, k$ , то  $\sum_{j=m}^{N_0} \beta_{kj}^{(+)} = 1$ . Обозначим также через  $\Delta \hat{\gamma}_j^{(+)}$  величину планируемого заказа, которая при указанном условии  $x(t'_n) = x' \in E_k^{(+)}$  равна  $\hat{\tau}_j^{(+)} - (x' - \alpha_w h)$ .

**Замечание 2.** Если состояние процесса  $x(t)$  в момент заказа соответствует дефициту продукта, т.е.  $x(t'_n) = x' \in E_k^{(-)}$ , то событие  $B_{kj}^{(-)}$  также заключается в выборе уровня планирования  $\tau_j^{(+)}$ . Данное условие отражает особенность модели, заключа-



**Рис. 2** Процедура пополнения запаса: планирование уровня пополнения

ющуюся в том, что при каждом пополнении запаса отложенный спрос полностью удовлетворяется и объем запаса становится положительным.

Иллюстрация к описанной процедуре пополнения запаса приведена на рис. 2. Для удобства начальный момент данного периода эволюции предполагается равным нулю:  $t_n = 0$ .

Продолжим описание процедуры пополнения запаса. Предположим, что реализовалось событие  $B_{kj}^{(-)}$  или  $B_{kj}^{(-)}$  и зафиксирован планируемый уровень пополнения  $\hat{\tau}_j^{(+)} \in E_j^{(+)}$ . Введем еще одну случайную величину  $\hat{\Delta}_j^{(+)}$ , которая характеризует отклонение от планируемого уровня запаса  $\hat{\tau}_j^{(+)}$ , возникающее при выполнении заказа. Данное отклонение обусловлено объективными случайными факторами, действующими на рынке продукта, которые влияют на выполнение заказа поставщиком.

Для описания процедуры пополнения с учетом случайных отклонений от уровней планирования зададим набор распределений  $B_j(x)$ , сосредоточенных на соответствующих интервалах  $E_j^{(+)}$ , причем

$$B_j(x) = \begin{cases} 0, & x < jL; \\ 1, & x \geq (j+1)L, \end{cases}$$

$j = 0, 1, \dots, N_0$ . Распределение  $B_j(x)$  описывает уровень запаса после пополнения с учетом отклонения от планируемого уровня  $\hat{\tau}_j^{(+)}$ , а именно: значение уровня запаса после пополнения представляет собой величину  $\hat{\gamma}_j^{(+)} = \hat{\tau}_j^{(+)} + \hat{\Delta}_j^{(+)}$ , имеющую распределение  $P(\hat{\gamma}_j^{(+)} < x) = B_j(x)$ . При этом функция распределения величины  $\hat{\Delta}_j^{(+)}$  определяется равенством:

$$\hat{B}_j(x) = P(\hat{\Delta}_j^{(+)} < x) = P(\hat{\gamma}_j^{(+)} - \hat{\tau}_j^{(+)} < x) = B_j(\hat{\tau}_j^{(+)} + x), \quad j = 0, 1, \dots, N_0.$$



Более полное формализованное описание траектории процесса  $x(t)$  и, в частности, процедуры пополнения содержится в приложении [1].

#### 4 Дополнительные стохастические объекты в полумарковской модели

В предыдущем разделе проведено формальное построение случайного процесса  $x(t) = x(\omega, t)$ ,  $\omega \in \Omega$ ,  $t \in [0, \infty)$ , описывающего уровень запаса в исследуемой системе в произвольный момент времени  $t$ . Теперь введем дополнительные стохастические объекты на вероятностном пространстве  $(\Omega, \mathcal{F}, P)$ , связанные со случайным процессом  $x(t)$ . Определим последовательность случайных величин  $\zeta_n = \zeta_n(\omega)$ ,  $\omega \in \Omega$ ,  $n = 0, 1, 2, \dots$ , следующим образом. Рассмотрим функцию  $\text{ind}(x)$ , заданную на множестве  $[0, \tau]$ , принимающую значения из конечного множества  $\{0, 1, \dots, N_0\}$  и определяемую равенством:

$$\text{ind}(x) = i, \text{ если } x \in [\tau_i^{(+)}; \tau_{i+1}^{(+)}) , i = 0, 1, \dots, N_0 .$$

Тогда случайная величина  $\zeta_n(\omega)$  определяется соотношением:

$$\zeta_n(\omega) = \zeta_n = \text{ind}(x(t_n)) , n = 0, 1, 2, \dots$$

Иначе говоря, если  $x(t_n) = x(t_n + 0) = x \in [\tau_i^{(+)}; \tau_{i+1}^{(+)})$ , то  $\zeta_n = i$ . Начальное значение последовательности  $\{\zeta_n\}$  определяется начальным значением процесса  $x(t)$ . Конкретно, если  $x(0) = x_0 \in [\tau_{i_0}^{(+)}; \tau_{i_0+1}^{(+)})$ , то полагаем  $\zeta_0 = i_0$ . Поскольку исходный случайный процесс  $x(t)$  обладает марковским свойством в моменты пополнения запаса, случайная последовательность  $\{\zeta_n, n = 0, 1, 2, \dots\}$  образует цепь Маркова с конечным множеством состояний  $\{0, 1, \dots, N_0\}$ .

Определим случайный процесс  $\zeta(t)$  следующим образом:

$$\zeta(t) = \zeta_n \text{ при } t_n \leq t < t_{n+1} , n = 0, 1, 2, \dots$$

Случайный процесс  $\zeta(t)$  является полумарковским с множеством состояний  $\{0, 1, \dots, N_0\}$ . Данный процесс будем называть сопровождающим полумарковским процессом. Случайная последовательность  $\{\zeta_n, n = 0, 1, 2, \dots\}$  является цепью Маркова, вложенной в данный полумарковский процесс. Обозначим через  $p_{ij} = P(\zeta_{n+1} = j | \zeta_n = i)$ ,  $i, j \in \{0, 1, \dots, N_0\}$ , переходные вероятности этой марковской цепи.

Теперь введем аддитивный стоимостный функционал  $V(t) = V(\omega, t)$ ,  $\omega \in \Omega$ , связанный с сопровождающим полумарковским процессом  $\zeta(t)$ . Этот функционал будет характеризовать прибыль, получаемую при функционировании системы. Общая схема задания такого функционала описана в классических работах [4, 5]. Отметим, что в рассматриваемой стохастической модели приращение этого функционала на периоде эволюции будет определяться исходными стоимостными характеристиками  $c_1(x)$ ,  $c_2(x)$ ,  $d_1(x)$  и  $d_2(x)$ . Обозначим через  $V(t)$  случайное значение этого функционала в момент времени  $t$ .

Пусть  $\Delta V_n = V(t_{n+1}) - V(t_n)$  — приращение функционала на периоде времени между моментами изменения состояний полумарковского процесса  $\zeta(t)$ .

Для постановки и решения задачи оптимального управления в рассматриваемой модели потребуются также следующие вероятностные характеристики:

$d_i = E[\Delta V_n | \zeta_n = i]$  — условное математическое ожидание прибыли на периоде между последовательными моментами изменения состояний процесса  $\zeta(t)$  при условии, что на этом периоде процесс находился в состоянии  $i$ ;

$T_i = E[t_{n+1} - t_n | \zeta_n = i]$  — условное математическое ожидание длительности периода между последовательными моментами изменения состояний процесса  $\zeta(t)$  при условии, что на этом периоде процесс находился в состоянии  $i$ ;

$\Pi = \{\pi_0, \dots, \pi_{N_0}\}$  — стационарное распределение цепи Маркова  $\{\zeta_n\}$ , вложенной в данный полумарковский процесс  $\zeta(t)$ .

Известно, что при достаточно общих условиях, основным из которых является существование стационарного распределения вложенной цепи Маркова  $\{\zeta_n\}$ , имеет место следующее утверждение (обычно называемое эргодической теоремой) о поведении стоимостного аддитивного функционала  $V(t)$ :

$$I = \lim_{t \rightarrow \infty} \frac{EV(t)}{t} = \frac{\sum_{i=0}^{N_0} d_i \pi_i}{\sum_{i=0}^{N_0} T_i \pi_i} . \tag{2}$$

Доказательство соотношения (2) для полумарковского процесса с конечным множеством состояний приведено в работе [5]. В настоящее время известно, что аналогичное утверждение имеет место для весьма общих полумарковских моделей с произвольными пространствами состояний и управлений [8].

По своему прикладному содержанию величина  $I$  представляет собой среднюю удельную прибыль, возникающую при длительной эволюции полумарковской модели. В дальнейшем величина  $I$  будет рассматриваться как показатель эффективности управления в данной модели.

Теперь приведем формулировки утверждений о явных аналитических представлениях вспомогательных характеристик, входящих в выражение для функционала (2).

## 5 Основные вероятностные характеристики модели

Зафиксируем состояние сопровождающего процесса  $\zeta(t)$  в момент очередного пополнения запаса, т. е. условие  $\zeta(t_n) = \zeta_n = i$ . Введем в множестве  $\Omega$  систему несовместных событий, связанных с этим условием:

$$A_k^{(+)} = (x(t_{n+1} - 0) \in E_k^{(+)}) , \quad k = 0, 1, \dots, i ;$$

$$A_k^{(-)} = (x(t_{n+1} - 0) \in E_k^{(-)}) , \quad k = 0, 1, \dots, N_1 .$$

Заметим, что система несовместных событий  $\{A_k^{(+)}, k = 0, 1, \dots, i; A_k^{(-)}, k = 0, 1, \dots, N_1\}$  образует разбиение множества элементарных исходов  $(\zeta_n = i) = (\omega \in \Omega : \zeta_n(\omega) = i)$ , т. е. множества, соответствующего принятому условию. Рассмотрим условные вероятности

$$\left. \begin{aligned} p_{ik}^{(+)} &= P(A_k^{(+)} | \zeta_n = i) = \\ &= P(x(t_{n+1} - 0) \in E_k^{(+)} | \zeta_n = i) , \\ &\quad k = 0, 1, \dots, i ; \\ p_{ik}^{(-)} &= P(A_k^{(-)} | \zeta_n = i) = \\ &= P(x(t_{n+1} - 0) \in E_k^{(-)} | \zeta_n = i) , \\ &\quad k = 0, 1, \dots, N_1 ; \end{aligned} \right\} (3)$$

Как уже отмечалось в разд. 3, в принятой модели значения процесса  $x(t)$  в момент заказа  $x(t'_n)$  и непосредственно перед пополнением  $x(t_{n+1} - 0)$  связаны однозначно:

$$x(t_{n+1} - 0) = x(t'_n) - \alpha_w h .$$

Отсюда следует, что можно описать переход процесса из произвольного подмножества состояний  $E_k^{(+)}$ ,  $k = 0, 1, \dots, i$ , или  $E_k^{(-)}$ ,  $k = 0, 1, \dots, N_1$ , на планируемый уровень  $\hat{\tau}_j^{(+)}$  при помощи вероятностей  $\beta_{kj}^{(+)}$ ,  $k = 0, 1, \dots, i$ , или  $\beta_{kj}^{(-)}$ ,  $k = 0, 1, \dots, N_1$ ,

соответственно. Учитывая, что в момент заказа  $t'_n$  процесс  $x(t)$  обладает марковским свойством, и используя указанные свойства системы событий  $\{A_k^{(+)}, k = 0, 1, \dots, i; A_k^{(-)}, k = 0, 1, \dots, N_1\}$ , можно применить формулу полной вероятности. Получаем:

$$p_{ij} = \sum_{k=0}^i p_{ik}^{(+)} \beta_{(k+r)j}^{(+)} + \sum_{k=0}^r p_{ik}^{(-)} \beta_{(r-k)j}^{(+)} +$$

$$+ \sum_{k=r+1}^{N_1} p_{ik}^{(-)} \beta_{(k-r)j}^{(-)} + \sum_{k=N_1+1}^{N_1+r} p_{iN_1}^{(-)} \beta_{(k-r)j}^{(-)} , \quad (4)$$

где  $r = \alpha_w h / L$ .

Формула (4) позволяет выразить вероятности перехода вложенной цепи Маркова  $\{\zeta_n\}$  через вспомогательные вероятностные характеристики (3) и заданные вероятности  $\{\beta_{kj}^{(+)}\}$  и  $\{\beta_{kj}^{(-)}\}$ . В дальнейшем будут найдены явные представления для указанных характеристик (3).

Обозначим:

$$d_{ik}^{(+)} =$$

$$= E[V_{n+1} - V_n ; x(t_{n+1} - 0) \in E_k^{(+)} | \zeta_n = i] ; \quad (5)$$

$$d_{ik}^{(-)} =$$

$$= E[V_{n+1} - V_n ; x(t_{n+1} - 0) \in E_k^{(-)} | \zeta_n = i] . \quad (6)$$

Аналогичным образом используем свойство системы событий  $\{A_k^{(+)}, k = 0, 1, \dots, i; A_k^{(-)}, k = 0, 1, \dots, N_1\}$  и формулу полного математического ожидания. Получим:

$$d_i = \sum_{k=0}^i d_{ik}^{(+)} + \sum_{k=0}^{N_1} d_{ik}^{(-)} . \quad (7)$$

Представление для условного математического ожидания длительности периода пребывания полумарковского процесса  $\zeta(t)$  в состоянии  $i$  может быть получено непосредственно (см. ниже теорему 2).

Теперь приведем формулировки утверждений о явных аналитических представлениях вспомогательных вероятностных характеристик, входящих в приведенные выше формулы (4) и (7).

**Теорема 1.** *Предположим, что в рассматриваемой стохастической модели реализуется событие  $x(t_n) = x \in E_i^{(+)}$ , где  $i$  — некоторое фиксированное целое число,  $i = 0, 1, \dots, N_0$ . Тогда аналитические представления условных вероятностей  $p_{ik}^{(+)}$  и  $p_{ik}^{(-)}$  имеют следующий вид:*

$$p_{ik}^{(+)} = \int_{iL}^{(i+1)L} \left[ \int_{\sigma_{k+1}^{(+)}(x)}^{\sigma_k^{(+)}(x)} dG_i(u) \right] dB_i(x)$$

при  $k = 0, \dots, (k_i^{(+)} - 1)$  ;

$$p_{ik_i^{(+)}}^{(+)} = \int_{iL}^{(i+1)L} \left[ \int_0^{\sigma_{k_i^{(+)}}^{(+)}(x)} dG_i(u) \right] dB_i(x)$$

при  $k = k_i^{(+)}$  ;

$$p_{ik}^{(+)} = 0 \text{ при } k = (k_i^{(+)} + 1), \dots, i ;$$

$$p_{ik}^{(-)} = 0 \text{ при } k = 0, \dots, (k_i^{(-)} - 1) ;$$

$$p_{ik_i^{(-)}}^{(-)} = \int_{iL}^{(i+1)L} \left[ \int_0^{\sigma_{k_i^{(-)}+1}^{(-)}(x)} dG_i(u) \right] dB_i(x)$$

при  $k = k_i^{(-)}$  ;

$$p_{ik}^{(-)} = \int_{iL}^{(i+1)L} \left[ \int_{\sigma_k^{(-)}(x)}^{\sigma_{k+1}^{(-)}(x)} dG_i(u) \right] dB_i(x)$$

при  $k = (k_i^{(-)} + 1), \dots, (N_1 - 1)$  ;

$$p_{iN_1}^{(-)} = \int_{iL}^{(i+1)L} [1 - G_i(\sigma_{N_1}^{(-)}(x))] dB_i(x)$$

при  $k = N_1, (i + N_1)L > \alpha_w h$  ;

$$p_{iN_1}^{(-)} = \int_{iL}^{(i+1)L} [1 - g_{i,0}] dB_i(x)$$

при  $k = N_1, (i + N_1)L \leq \alpha_w h$  ,

где  $k_i^{(\pm)}$  — граничные значения, которые зависят от известных фиксированных параметров системы  $\alpha_w$ ,  $h$  и  $L$  и конкретно определены в приложении [1];

$$\sigma_k^{(\pm)}(x) = \frac{x - \tau_k^{(\pm)}}{\alpha_c} - \frac{\alpha_w}{\alpha_c} h ;$$

$$\tau_{N_0+1}^{(+)} = \tau ;$$

$$\tau_k^{(+)} = kL, \quad k = 0, 1, \dots, N_0 ;$$

$$\tau_{N_1+1}^{(-)} = -\infty ;$$

$$\tau_k^{(-)} = -kL, \quad k = 0, 1, \dots, N_1 .$$

**Теорема 2.** Предположим, что выполнено вероятностное условие, сформулированное в теореме 1:

$x(t_n) = x \in E_i^{(+)}$ ,  $i = 0, 1, \dots, N_0$ . Тогда математическое ожидание времени пребывания полумарковского случайного процесса  $\zeta(t)$  в состоянии  $i$  представимо в виде:

$$T_i = \int_{iL}^{(i+1)L} \left[ \int_0^\infty u dG_i(u) \right] dB_i(x) + h .$$

**Теорема 3.** Предположим, что выполнено вероятностное условие, сформулированное в теореме 1:  $x(t_n) = x \in E_i^{(+)}$ ,  $i = 0, 1, \dots, N_0$ . Тогда условные математические ожидания приращений функционала прибыли в рассматриваемой стохастической полумарковской модели, определяемые соотношениями (5) и (6), допускают следующие аналитические представления:

$$d_{ik}^{(+)} = \int_{iL}^{(i+1)L} \left[ \int_{\max(0, \sigma_{k+1}^{(+)}(x))}^{\sigma_k^{(+)}(x)} D_{ik}^{(+)}(x, u) dG_i(u) \right] dB_i(x)$$

при  $k = 0, \dots, k_i^{(+)}$  ;

$$d_{ik}^{(-)} = \int_{iL}^{(i+1)L} \left[ \int_{\max(0, \sigma_k^{(-)}(x))}^{\sigma_{k+1}^{(-)}(x)} D_{ik}^{(-)}(x, u) dG_i(u) \right] dB_i(x)$$

при  $k = k_i^{(-)}, \dots, N_1, \alpha_w h - kL < 0$  ;

$$d_{ik}^{(-)} = \int_{iL}^{(i+1)L} \left[ \int_{\max(0, \sigma_k^{(-)}(x))}^{x/\alpha_c} D_{ik}^{(w-)}(x, u) dG_i(u) + \int_{x/\alpha_c}^{\sigma_{k+1}^{(-)}(x)} D_{ik}^{(c-)}(x, u) dG_i(u) \right] dB_i(x)$$

при  $k = k_i^{(-)}, \dots, N_1, \alpha_w h - kL \geq 0$  ,

$$d_{ik}^{(-)} = \int_{iL}^{(i+1)L} \left[ \int_{\max(0, \sigma_k^{(-)}(x))}^{\sigma_{k+1}^{(-)}(x)} D_{ik}^{(w-)}(x, u) dG_i(u) \right] dB_i(x)$$

при  $k = k_i^{(-)}, \dots, N_1, \alpha_w h - (k + 1)L < 0$  ;

$$d_{ik}^{(-)} = \int_{iL}^{(i+1)L} \left[ \int_{\max(0, \sigma_k^{(-)}(x))}^{\sigma_{k+1}^{(-)}(x)} D_{ik}^{(w-)}(x, u) dG_i(u) \right] dB_i(x)$$

при  $k = k_i^{(-)}, \dots, N_1, \alpha_w h - (k + 1)L \leq 0$  ,

где функции  $D_{ik}^{(+)}(x, u)$ ,  $D_{ik}^{(c-)}(x, u)$  и  $D_{ik}^{(w-)}(x, u)$  аналитически определяются через известные функции доходов  $d_1(x)$  и  $d_2(x)$  и затрат  $c_1(x)$  и  $c_2(x)$ , которые были введены в разд. 2. В связи с громоздкостью соответствующих формул они приводятся в приложении [1].

## 6 Заключение

Подведем краткие итоги первой части проведенного исследования, изложенной в настоящей работе. Построена и идейно обоснована новая стохастическая полумарковская модель управления запасом непрерывного продукта. Проблема управления формализуется в виде задачи безусловного экстремума стационарного стоимостного показателя  $I$ , определяемого соотношением (2), который представляет собой среднюю удельную прибыль. Для получения явного представления этого показателя необходимо найти формулы для ряда вероятностных характеристик разработанной полумарковской модели. Такие формулы были доказаны в данной работе.

Для завершения решения задачи оптимального управления запасом в предложенной полумарковской модели необходимо доказать, что стационарный стоимостный показатель эффективности управления  $I$  представим в виде дробно-линейного интегрального функционала от набора вероятностных распределений, задающих стратегию управления. Тогда решение задачи управления будет основано на использовании теоремы об экстремуме дробно-линейного интегрального функционала, доказанной П. В. Шнурковым [9, 10]. В результате проблема оптимального управления будет сведена к исследованию на глобальный экстремум некоторой заданной функции от конечного числа действительных неотрицательных переменных.

Соответствующие аналитические выводы и доказательства составляют вторую часть проведенного исследования, которую авторы планируют изложить в следующей статье.

## Литература

1. Шнурков П. В., Егоров А. Ю. Приложение к статье «Разработка и предварительное исследование стоха-

стической полумарковской модели управления запасом непрерывного продукта при постоянно происходящем потреблении», 2017. 49 с. [http://www.ipiran.ru/publications/Приложение\\_upd.pdf](http://www.ipiran.ru/publications/Приложение_upd.pdf).

2. Королюк В. С., Турбин А. Ф. Полумарковские процессы и их приложения. — Киев: Наукова думка, 1976. 184 с.
3. Janssen J., Manca R. Applied semi-Markov process. — New York, NY, USA: Springer, 2006. 309 p.
4. Джевелл В. Управляемые полумарковские процессы // Кибернетический сборник. Новая серия. — М.: Мир, 1967. Вып. 4. С. 97–134.
5. Майн Х., Осаки С. Марковские процессы принятия решений / Пер. с англ. — М.: Наука, 1977. 176 с. (Mine H., Osaki S. Markovian decision processes. — New York, NY, USA: Elsevier, 1970. 142 p.)
6. Шнурков П. В., Иванов А. В. Анализ дискретной полумарковской модели управления запасом непрерывного продукта при периодическом прекращении потребления // Дискретная математика, 2014. Т. 26. № 1. С. 143–154.
7. Шнурков П. В. Стохастическая модель планового технического обслуживания // Стохастические системы и их приложения. — Киев: Институт математики АН УССР, 1990. С. 98–105.
8. Luque-Vasquez F., Herndandez-Lerma O. Semi-Markov control models with average costs // Appl. Math., 1999. Vol. 26. No. 3. P. 315–331.
9. Шнурков П. В. О решении проблемы безусловного экстремума для дробно-линейного интегрального функционала на множестве вероятностных мер // Докл. РАН. Сер. Математика, 2016. Т. 470. № 4. С. 387–392.
10. Шнурков П. В., Горшенин А. К., Белоусов В. В. Аналитическое решение задачи оптимального управления полумарковским процессом с конечным множеством состояний // Информатика и её применения, 2016. Т. 10. Вып. 4. С. 72–88.

Поступила в редакцию 31.05.17



# DEVELOPMENT AND PRELIMINARY STUDY OF A STOCHASTIC SEMI-MARKOV MODEL OF CONTINUOUS SUPPLY OF PRODUCT MANAGEMENT UNDER THE CONDITION OF CONSTANT CONSUMPTION

P. V. Shnurkov and A. Y. Egorov

National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation

**Abstract:** The paper deals with a discrete semi-Markov stochastic model describing the operation of a control system of continuous supply of product with constant consumption. The model is a couple of random processes  $(x(t), \zeta(t))$  where the main process  $x(t)$  describes the amount of stock in the system at time  $t$  and the accompanying random process is a semi-Markov process with a finite set of states. The optimal control problem is put in relation to the stationary indicators related to the accompanying process. This indicator is the average of the specific nature of the profits earned in the evolution of the initial inventory control system. An explicit analytical representation for the probability characteristics of semi-Markov models is obtained. In the future, the results will allow to find an explicit representation of the Quality Score and solving the problem of optimal control.

**Keywords:** inventory management; semi-Markov stochastic process; stationary value functional; optimal control of stochastic systems

**DOI:** 10.14357/19922264180114

## References

1. Shnurkov, P.V., and A.Y. Egorov. 2017. Prilozhenie k stat'e "Razrabotka i predvaritel'noe issledovanie stokhasticheskoy polumarkovskoy modeli upravleniya zapasom nepreryvnogo produkta pri postoyanno proiskhodyashchem potreblenii" [Appendix to article "Development and preliminary study of stochastic semi-Markov model of continuous supply of product management at constantly happening consumption"]. 49 p. Available at: [http://www.ipiran.ru/publications/Приложение\\_upd.pdf](http://www.ipiran.ru/publications/Приложение_upd.pdf) (accessed December 22, 2017).
2. Korolyuk, V.S., and A.F. Turbin. 1976. *Polumarkovskie protsessy i ikh prilozheniya* [Semi-Markov processes and their applications]. Kiev: Naukova Dumka. 184 p.
3. Janssen, J., and R. Manca. 2006. *Applied semi-Markov process*. New York, NY: Springer. 309 p.
4. Jewell, W.S. 1963. Markov-renewal programming. *Oper. Res.* 11:938–971.
5. Mine, H., and S. Osaki. 1970. *Markovian decision processes*. New York, NY: Elsevier. 142 p.
6. Shnurkov, P.V., and A.V. Ivanov. 2015. Analysis of a discrete semi-Markov model of continuous inventory control with periodic interruptions of consumption. *Discrete Math. Appl.* 25(1):59–67.
7. Shnurkov, P.V. 1990. Stokhasticheskaya model' planovogo tekhnicheskogo obsluzhivaniya [Stochastic model of scheduled maintenance]. *Stokhasticheskie sistemy i ikh prilozheniya* [Stochastic systems and their applications]. Kiev: Institute of Mathematics of the Ukrainian Academy of Sciences. 98–105.
8. Luque-Vasquez, F., and O. Herndandez-Lerma. 1999. Semi-Markov control models with average costs. *Appl. Math.* 26(3):315–331.
9. Shnurkov, P.V. 2016. Solution of the unconditional extremum problem for a linear-fractional integral functional on a set of probability measures. *Doklady Mathematics* 94(2):550–554.
10. Shnurkov, P.V., A.K. Gorshenin, and V.V. Belousov. 2016. Analiticheskoe reshenie zadachi optimal'nogo upravleniya polumarkovskim protsessom s konechnym mnozhestvom sostoyaniy [An analytic solution of the optimal control problem for a semi-Markov process with a finite set of states]. *Informatika i ee Primeneniya — Inform. Appl.* 10(4):72–88.

Received May 31, 2017

## Contributors

**Shnurkov Peter V.** (b. 1953) — Candidate of Science (PhD) in physics and mathematics, associate professor, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; pshnurkov@hse.ru

**Egorov Artem Y.** (b. 1992) — Master student, National Research University Higher School of Economics, 34 Tallinskaya Str., Moscow 123458, Russian Federation; eyupimenova@edu.hse.ru

# ФИНАНСИРОВАНИЕ ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ: КОНЦЕПТУАЛЬНЫЙ ОБЛИК СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ НАУКОМЕТРИИ И АНАЛИЗА ДАННЫХ\*

О. В. Сютюренко<sup>1</sup>

**Аннотация:** Статья посвящена разработке концептуальной модели информационной системы, ориентированной на решение задач управления финансированием фундаментальных исследований с использованием методов наукометрии и анализа данных. Сформулированы базовые принципы и предложена методология создания и функционирования системы поддержки принятия решений (СППР) при распределении фонда госзаказа по тематическим научным направлениям. Представлена функциональная структура СППР. Рассматриваются перспективы использования нового многоцелевого информационного ресурса и аналитической постобработки информации. Показана взаимосвязь задач управления наукой с задачами инновационного развития отраслей экономики.

**Ключевые слова:** наукометрия; анализ данных; системы управления; научные приоритеты; мониторинг; научно-технический потенциал; постобработка данных

DOI: 10.14357/19922264180115

## 1 Введение

Эффективность управления и объемы финансирования фундаментальных исследований в наибольшей степени (сравнительно с другими факторами) влияют на их результативность [1]. В настоящее время для оценивания результативности научной деятельности совместно с экспертными заключениями все чаще используют и наукометрические показатели [2]. Эти показатели основаны на числе публикаций автора и на количестве ссылок на его работы. Возросший интерес к наукометрическим показателям вызван в первую очередь возможностью автоматизации процесса оценивания с использованием программных средств баз данных (БД) Web of Science, Scopus, РИНЦ (Российский индекс научного цитирования). Кроме того, можно использовать бесплатные программы, например Publish or Perish, работающие на данных поисковой системы научных публикаций Google Scholar. Доступность таких показателей связана с развитием электронных библиографических БД и, как следствие, возможностей автоматического расчета соответствующих индексов. Потенциально эти показатели могут использоваться на всех этапах процесса управления научно-исследовательской деятельностью, и прежде всего

- при формировании приоритетных направлений развития и при создании новых исследовательских центров и групп;
- при распределении финансирования между исследовательскими программами и проектами.

Ниже рассмотрена модель и концептуальный облик СППР при управлении бюджетным финансированием тематических направлений с использованием критериев и информационно-аналитических методов наукометрии и анализа данных. Нетривиальность задачи заключается в ее значительной размерности: распределение (и мониторинг) выделяемых бюджетных средств по ~ 8000 научных направлений, исследования по которым ведутся в ~ 400 научных организациях Российской академии наук (РАН).

## 2 Концептуальные основы информационной системы управления финансированием фундаментальных исследований

Планирование развития науки в целом или какой-либо ее отрасли, перераспределение средств и капитальных вложений требуют учета тенден-

\* Работа поддержана РФФИ (проект 17-07-00256).

<sup>1</sup> Всероссийский институт научной и технической информации Российской академии наук, olegasu@mail.ru

ций развития науки, прогнозирования появления и отмирания ее различных направлений. С развитием компьютерных технологий наиболее перспективным для решения задач управления является использование информационных методов наукометрии и анализа данных [3–7]. Результативность исследований зависит от уровня соответствия исследовательских задач и научно-технического потенциала (НТП) научной организации. Структурно-инфологическую модель НТП  $j$ -й научной организации можно представить как

$$E^j(t) = \Phi \left( P_{in}^j, P_{if}^j, P_m^j, P_{fn}^j, S^j, t \right),$$

где  $P_{in}^j$ ,  $P_{if}^j$ ,  $P_m^j$  и  $P_{fn}^j$  — интеллектуальные, информационные, материальные и финансовые ресурсы научной организации соответственно;  $S^j$  — уровень организации и управления ресурсами. Таким образом,  $E^j(t)$  — это обобщенная характеристика, определяющая совокупность основных видов ресурсов и факторов в период времени  $t$  и отражающая способность научной организации решать стоящие перед ней задачи. Результативность исследований является функцией максимизации соответствия научных задач и НТП  $[Z^j : E^j(t)] = R_{RQ} \Rightarrow \max$  научной организации. Уровень соответствия можно оценить с помощью проблемно-ориентированных оценок НТП. Результаты количественной оценки потенциала могут быть поставлены в зависимость от тех или иных сформулированных (выбранных) целей из множества  $\{Z\}$  [7].

Оптимальная стратегия управления научными исследованиями — серьезная методологическая проблема. Следует отметить, что новые подходы к решению проблем современного научно-технического развития с учетом их возрастающей масштабности, сложности, ресурсоемкости реализуются на основе:

- научного обоснования приоритетного ряда научных и научно-технических направлений;
- использования преимущественно информационных и экономических методов управления научными исследованиями и разработками (в том числе минимизации факторов неэффективного использования имеющихся и/или доступных ресурсов).

Наиболее перспективным и экономичным решением является использование проблемно-ориентированной СППР при управлении финансированием фундаментальных исследований. Интерактивная компьютерная система СППР путем сбора и анализа большого количества информации может влиять на процесс принятия решений (и моделировать решения) при распределении

фонда госзаказа по тематическим фундаментальным научным направлениям. В наборе показателей СППР доминируют агрегированные и производные показатели, такие как индексы и индикаторы. Системы такого класса предназначены для поддержки многокритериальных решений в сложной информационной среде. При этом под многокритериальностью понимается тот факт, что результаты принимаемых решений оцениваются не по одному, а по совокупности многих показателей (критериев), рассматриваемых одновременно. Информационная сложность определяется необходимостью учета большого объема данных, обработка которых без помощи современной вычислительной техники практически невыполнима. С помощью СППР может производиться выбор решений неструктурированных и слабоструктурированных задач, в том числе и многокритериальных. Система поддержки принятия решений позволяет решить две основные задачи:

- (1) выбор наилучшего решения из множества возможных (оптимизация);
- (2) упорядочение возможных решений по предпочтительности (ранжирование).

С системных позиций рассмотрим **концептуальные основы разработки (и функционирования) системы управления бюджетным финансированием** тематических фундаментальных научных направлений:

- А.** Использование неэкспертных количественных (количественных) наукометрических методов (и критериев) оценок уровня: научного тематического направления; научной организации.
- В.** Неэкспертное формирование комплексной оценки уровня научного направления на основе многомерного анализа данных и методов наукометрии. Распределение бюджетных средств осуществляется пропорционально вычисляемому рейтингу научных направлений, который синтезируется на основе критериев: совокупности наукометрических показателей научной организации (по данному направлению), ее НТП, уровня приоритетности направления.
- С.** Автоформализация совокупной оценки наукометрических показателей ученых  $i$ -го тематического направления  $j$ -й научной организации (соискателя бюджетного финансирования). Автоформализация интегральной оценки уровня научной организации с учетом ее НТП и приоритетности тематических направлений исследований.

- Д.** Использование экспертов и экспертных методов на этапе формирования критериев и системы показателей — на начальном этапе разработки системы. При сопоставительном анализе и моделировании принципиальным является сопоставимость параметров, прежде всего наукометрических.
- Е.** Автоматическое (автоматизированное) формирование исходных массивов данных тематических научных направлений и научных организаций (соискателей бюджетного финансирования). Постоянный мониторинг показателей и актуализация банка данных.
- Ж.** Автоматическое формирование (расчет) сопоставительного уровня (рейтинга) тематических научных направлений и распределение объемов бюджетного финансирования.
- З.** Поддержание статуса открытости системы и мультипликативного использования формируемого электронного информационного ресурса. Соответствие требованиям масштабируемости, многомерного и многовариантного представления данных, гибкости и адаптируемости к внешним изменениям, сетевая интеграция (прежде всего веб).

Распределение финансирования по тематическим научным направлениям из фонда госзаказа должно осуществляться с учетом следующих показателей:

- наукометрической оценки научного уровня разрабатываемого тематического направления с использованием таких показателей, как число публикаций, индекс цитируемости, индекс Хирша, импакт-фактор, ожидаемый отклик, индекс Прайса, ПРНД (показатель результативности научной деятельности, принятый в РАН) и др. [2, 8];
- оценки НТП научной организации — соискателя финансирования;
- принадлежности тематического направления к Перечню приоритетных направлений развития научно-технического комплекса России на 2014–2020 гг. [9].

Основная прагматическая задача — разработка сводной (весовой) эмпирической формулы количественной оценки уровня тематического направления (с учетом НТП научной организации и приоритетности направления). В определенной степени прототипом предлагаемой технологии является система оценки и отбора проектов Российского фонда фундаментальных исследований (РФФИ), при которой субъект — соискатель

грантового финансирования (ученый, группа ученых или научная организация) сам информационно обосновывает актуальность, важность, научный задел и уровень работ по выбранной тематике. Однако решения принимаются на основе экспертных заключений (в РФФИ > 3000 экспертов). Следует отметить, что РФФИ ежегодно поддерживает более 70 тыс. ученых.

**Уточним исходную постановку задачи.** Общее число выделенных финансируемых фундаментальных научных направлений  $\leq 8000$ . Количество научных организаций РАН  $\leq 500$ . Научные организации — соискатели финансирования по распоряжению (приказу) Федерального агентства научных организаций (ФАНО) РФ или Президиума РАН заблаговременно ( $\sim 6$  мес.) предоставляют в ВИНТИ РАН в электронном виде формализованные данные, характеризующие:

- (1) научное направление, на развитие которого запрашивается бюджетное финансирование, с авторизованной наукометрической оценкой (по РИНЦ), включая данные по бюджетному финансированию направления за последние 3–5 лет;
- (2) формализованные данные по НТП и эффективности научной организации.

Показатели эффективности: общее число публикаций сотрудников научной организации (за фиксированный период времени), отнесенное к численности научных сотрудников, в том числе в зарубежных научно-технических изданиях (Scopus, Web of Science), в отечественных изданиях, включенных в перечень ВАК.

На основе представленных в электронном виде заявок в определенном формате практически в автоматическом режиме (в ВИНТИ) формируются БД Системы. Логические структуры БД и анкеты-формы разрабатываются заранее и утверждаются в ФАНО. Логическая структура БД по тематическому направлению должна также включать в себя текстовое поле с расширенной аннотацией (объемом  $\leq 3500$  знаков). ВИНТИ может дополнять БД экспресс-аналитическими обзорами по актуальным темам, входящим в Перечень приоритетных направлений или критических технологий. Общая картина-характеристика *НТП и эффективности научной организации* определяется совокупностью таких показателей, как

- число публикаций по всем научным направлениям, по годам (временной лаг  $\sim 5$  лет), дифференцированно по отечественным и зарубежным журналам;



- численность и структура кадрового состава организации (с учетом возрастного сдвига и по годам);
- информационно-аналитические оценки имеющихся научно-технических результатов сопоставительно с российским и мировым уровнем;
- количество (и их финансовый объем) получаемых грантов (по направлениям) от российских фондов, зарубежных фондов и программ, спонсоров;
- состав (и объем) выполненных внебюджетных конкурсных (и заказных) исследований, проектов, разработок (временной лаг  $\sim 4$  года) и др.

Функциональная структура информационной системы распределения бюджетного фонда госзаказа по тематическим научным направлениям включает в себя:

- банк данных СППР, включающий в себя:
  - (а) БД с формализованными характеристиками тематических направлений (включая приоритеты);
  - (б) БД с унифицированными характеристиками, отражающими научный (и технический) потенциал организаций РАН;
  - (в) БД по исследованиям и инновационным разработкам, представляющим интерес для дальнейшей коммерциализации;
- подсистемы:
  - (а) сбора и регистрации данных по тематическим направлениям и научным организациям РАН;
  - (б) моделирования, расчета показателей и распределения фонда госзаказа по тематическим направлениям;
  - (в) взаимодействия с реферативным банком данных ВИНТИ;
  - (г) мониторинга, неэкспертного информационного анализа данных по промежуточным результатам, контроля расходования средств по целевому финансированию;
- блоки:
  - (а) сопоставительного анализа тематических направлений на предмет выявления дублируемых исследований;
  - (б) информационного, математического и нормативно-методического (регламентного) обеспечения;
  - (в) управления, интерактивного взаимодействия, визуализации и генерации отчетов.

Финальным результатом обработки данных научных организаций является автоматическое (автоматизированное) распределение фонда госзаказа по выделенным тематическим направлениям при выполнении общего условия сохранения баланса затрат.

Рассмотрим алгоритм базовой модели распределения бюджетного финансирования (получаемого в рамках госзаказа) по выделенным фундаментальным научным направлениям с простой логико-вычислительной структурой.

Общее число финансируемых научных направлений обозначим как  $M$  ( $\sim 8000$ ). С учетом множества  $\{N\}$ , где  $N$  — общее число научных организаций, получающих финансирование из фонда госзаказа (для РАН  $\sim 400$ ), приведенный оценочный уровень каждого из тематических направлений определяется эмпирическим выражением:

$$m'_{ij} = m_{ij} \left[ \sum_{i=1}^M \sum_{j=1}^N m_{ij} (100\%)^{-1} \right]^{-1},$$

где  $m_{ij}$  — наукометрическая оценка уровня  $i$ -го тематического направления (по публикациям) в  $j$ -й научной организации, где  $m_{ij} > 0$  (с некоторой долей упрощения  $m_{ij}$  может интерпретироваться как уровень научного задела по  $i$ -му тематическому направлению);  $m'_{ij}$  — «удельный вес» тематического направления (или относительный уровень  $i$ -го направления в  $j$ -й научной организации) в общей совокупности выделенных для финансирования научных направлений РАН (в %).

Априори полагаем, что успешность выполнения исследований по каждому тематическому направлению в общем случае зависит от уровня НТП организации, где эти исследования проводятся. Эмпирическая формула уровня НТП  $j$ -й организации  $W_j$  определяется как

$$W_j = k_j (a_j B_j). \quad (1)$$

Здесь  $k_j$  — коэффициент, учитывающий технический потенциал  $j$ -й организации, в первую очередь состояние инфраструктуры (возможность широкополосного телекоммуникационного доступа к отечественным и зарубежным БД с научно-технической информацией, парк измерительной аппаратуры, компьютерное оснащение, наличие супер-ЭВМ, уникального оборудования коллективного доступа, состояние специализированных служб и сервисов и т.п.), условно шкала  $k = 0-V$  ( $V \leq 100$ );  $B_j$  — эффективность научной деятельности организации — уровень публикационной активности научной организации (см. РИНЦ), определяемый как

$$B_j = L_j(t) (G_j)^{-1}, \quad (2)$$

где  $L_j$  — общее число публикаций за период  $t$ ,  $G_j$  — число исследователей;  $a_j$  — нивелирующий коэффициент публикационной активности для научных организаций с преобладающей долей экспериментальных исследований:

$$a_j = \begin{cases} 1, & \text{ординарное направление;} \\ 1 + \delta, & \text{экспериментальное направление,} \\ & 1 < \delta \leq 10. \end{cases}$$

Тогда  $m_{ij}^0$ , с учетом (1) и (2), — приведенный оценочный уровень  $i$ -го тематического направления с учетом НТП  $j$ -й научной организации:

$$m_{ij}^0 = W_j m_{ij}^1.$$

Далее, пусть  $Z_0$  — общий бюджетный фонд госзаказа (для РАН  $\leq 100$  млрд руб.). В общем случае для повышения операционной гибкости системы может предусматриваться функциональный бюджетный резерв (руководства РАН) —  $\Delta Z_0$  ( $\Delta Z_0 \leq \leq 0,5\% - 3\%$ ); тогда далее оперируем величиной  $Z'_0 = Z_0 - \Delta Z_0$ .

Из условия сохранения баланса

$$Z'_0 = \sum_{i=1}^M \sum_{j=1}^N Z_{ij} = \sum_{i=1}^M \sum_{j=1}^N (x m_{ij}^0)$$

вычисляем базисную величину  $x$ , на основе которой определяются вложения средств по тематическим направлениям в соответствии с принципом пропорциональности оценочного (рейтингового) уровня научного направления:

$$x = Z'_0 (S^{-1}).$$

Здесь  $S$  — суммарный уровень всех тематических направлений:

$$S = \sum_{i=1}^M \sum_{j=1}^N (d_{ij} m_{ij}^0),$$

где  $d_i$  — коэффициент, учитывающий вхождение  $i$ -го направления  $j$ -й организации в Перечень приоритетных направлений:

$$d_i = \begin{cases} 1, & \text{неприоритетное направление;} \\ 1 + \Delta, & \text{приоритетное научно-техническое} \\ & \text{направление, } \Delta = const. \end{cases}$$

В общем случае может иметь место ситуация, когда  $\Delta = var$ , однако это, при больших размерностях  $M$  и  $N$ , повлечет существенное усложнение

расчетов по распределению бюджетных средств. Таким образом, итоговая расчетная формула базисной величины  $x$  имеет вид:

$$x = Z'_0 \left( \sum_{i=1}^M \sum_{j=1}^N (d_{ij} m_{ij}^0) \right)^{-1} = Z'_0 \cdot 100^{-1}. \quad (3)$$

С учетом формулы (3) объем финансовых вложений из фонда госзаказа в каждое  $i$ -е тематическое направление определяется итоговым выражением:

$$Z_{ij} = x (d_{ij} m_{ij}^0). \quad (4)$$

Используя равенство (4), выполняется последний шаг — заключительная проверка сохранения баланса суммарных затрат:

$$\sum_{i=1}^M \sum_{j=1}^N Z_{ij} \leq Z'_0.$$

Здесь следует отметить два детерминирующих фактора.

1. Задача объективной оценки НТП еще весьма далека от своего завершения, к оценке НТП до сих пор нет единого подхода [10]. Однако в последнее десятилетие при оценке НТП научных организаций помимо прочих все шире используется наукометрический подход (на основе учета числа публикаций и индекса цитирования  $B_j$ ). Принципиально важно исходить из того, что оценка всегда есть функция цели, а следовательно, оценка НТП в разных случаях может осуществляться на основе разных наборов показателей. При наличии количественных значений каждого из показателей, характеризующих НТП, возможен подход, при котором производится «свертка» частных показателей в интегральный (агрегированный) показатель, принимаемый за численную оценку потенциала (индикатор потенциала). Выбор оптимальной операции «свертки», позволяющий получать интегральный показатель, достаточно объективно характеризующий оценку НТП, является непростой теоретической и практической задачей, включающей в себя как выбор вида функции агрегирования, так и «взвешивание» отдельных входящих в нее показателей.
2. Выбор приоритетов лежит в слабо формализуемой сфере целеобразования [8, 11, 12]. Не существует универсальной модели процесса формирования приоритетов. Приоритетные направления развития науки и техники России в последнее десятилетие связаны с фундаментальными исследованиями, информационными технологиями и электроникой, производственными технологиями, новыми материалами и химическими продуктами, технологиями

живых систем, транспортом, топливом и энергетикой, экологией и рациональным природопользованием. Данные приоритетные направления конкретизируются в подпрограммах и подкреплены разработкой критических технологий. Институциональное обеспечение разработки научно-технической политики и определения соответствующих национальных приоритетов осуществляется различными структурами на разных уровнях управления экономикой. Поскольку ресурсы всегда ограничены, а круг разнородных задач в сфере научно-технологического развития, требующих решения, чрезвычайно широк, проблема выбора национальных приоритетов научно-технологического развития приобретает первостепенную значимость. Бюджетные процедуры очень важны для реализации научно-технических приоритетов (для любой страны) и являются одной из функциональных задач управления наукой.

**Перечень задач, на решение которых ориентирована СППР, включает:**

- автоматизированный расчет и распределение бюджетных средств по выделенным научным направлениям (основная задача);
- выявление с помощью методов наукометрического анализа стагнирующих научных направлений (анализ на временном интервале ~ 5 лет);
- выявление «точек роста», анализ и оценка трендов развивающихся научных направлений;
- информационно-аналитическая поддержка разработки прогнозов научно-технологического развития (в РФ и за рубежом);
- информационно-аналитическая поддержка формирования национального перечня приоритетных направлений научно-технологического развития и критических технологий;
- информационная поддержка подготовки предложений по формированию государственных (и региональных) научно-технических программ;
- текущий мониторинг выполнения работ и расходования бюджетных средств по научным направлениям.

Пользователи Системы: Президиум РАН, ФАНО, Министерство образования и науки РФ и другие федеральные ведомства, научные фонды, экспертное сообщество РАН, научные организации РФ и СНГ, информационные центры, промышленные корпорации.

### 3 Потенциальные приложения интегрированной постобработки информации базы данных СППР и банка данных ВИНТИ

Использование наукометрических методов и методов анализа данных при совместной постобработке актуальной информации по выполняющимся в РАН исследованиям и систематизированных информационных ресурсов ВИНТИ является весьма перспективным для решения целого ряда задач [13, 14], таких как

- анализ структуры (и уровня) отечественной и мировой науки;
- определение тенденций и процессов, происходящих в мировой и региональной науке;
- выявление (на ранней стадии) наиболее актуальных или, напротив, теряющих свою актуальность научных направлений;
- отслеживание генезиса конкретных научных идей (или направлений) и истории их развития;
- определение продуктивности научных организаций и работы отдельных исследователей (научных групп) в конкретной научной области и эффективности материальных и иных затрат в этой области;
- анализ трендов развития инновационной деятельности в рамках отдельных научных организаций, направлений (или отделений РАН);
- анализ структуры научного сообщества и изучение науки как социального организма.

Постобработка больших массивов научно-технической (и технико-экономической) информации с использованием методов наукометрии и анализа данных (в том числе статистических методов) априори позволяет выявлять закономерности, выражающие зависимости между распределениями различных параметров исследуемых систем и процессов, и характер изменения распределений во времени. Совместная постобработка информации БД Системы, ВИНТИ РАН и данных Росстата (РФФИ, БД eLIBRARY и др.), таких как величина валового внутреннего продукта (ВВП), произведенной энергии, среднего годового дохода на душу населения, величины произведенного продукта, приходящейся на высокие технологии, и ряда других, — это перспективное множество

представляющих практический интерес (для управления и прогнозирования) показателей и распределений.

Рассмотрим несколько гипотетических примеров-вариантов аналитической постобработки информации. Определенный интерес могут представлять:

- данные и визуализация распределений и оценка корреляции изменений бюджета фундаментальных исследований (совокупный госзаказ) и роста числа публикаций по годам (с использованием данных Росстата);
- графики распределений и оценки зависимости числа публикаций от роста ВВП (структурно) по направлениям: энергетика, транспорт, цветная металлургия и т. д., — по годам (желательно для сопоставления использовать данные бюджета и ВВП предыдущего года);
- анализ распределений и корреляции роста индекса производства (по 10 основным отраслям промышленности) в процентах к предыдущему периоду и роста расходов на исследования и разработки (или, например, числа публикаций российских авторов) за тот же период;
- анализ сравнительного роста:
  - (а) ВВП;
  - (б) расходов на образование;
  - (в) расходов на исследования и разработки;
  - (г) объема публикаций российских авторов;
- анализ изменений объемов (и структуры) ВВП и распределения затрат по научно-техническим направлениям;
- анализ зависимости роста инвестиций в отрасли экономики и рост объемов публикаций (то же по отраслям народного хозяйства).

В связи с быстрым ростом цифрового пространства и сетевой среды целенаправленное использование методов и средств аналитической постобработки информационных ресурсов могло бы трансформироваться в новое направление информатики — сетевой анализ и сетевую наукометрию.

## 4 Выводы

1. Синтезирован концептуальный облик СППР при управлении бюджетным финансированием тематических направлений с использованием критериев и методов наукометрии и анализа данных, а также с учетом приоритетности направлений и НТП научных организаций. Основные достоинства проекта:

- (а) простота решения и относительно невысокая трудоемкость (и ресурсоемкость) реализации Системы;
- (б) отсутствие организационно-финансовых проблем по корпусу экспертов и организации экспертизы;
- (в) использование современных информационно-аналитических и наукометрических методов для моделирования и получения оценок (унифицированных и сопоставимых) и финишных результатов по распределению бюджетных средств госзаказа между ~ 8000 научных направлений.

2. Важным и перспективным является формирование нового многоцелевого информационного ресурса РАН для целей управления, анализа развития отечественной и мировой науки, научно-технического прогнозирования, развития экспертной деятельности, оптимизации процессов финансирования исследований и разработок, мониторинга текущего состояния.

3. Системы управления фундаментальными исследованиями могут быть разработаны на основе иных подходов, отличных от предложенного. Однако следует отметить, что по базовым положениям теории (и практики) управления не может быть эффективной и устойчивой система, если центр целеполагания и компетенции находится в одном месте, а центр управления и распределения финансовых ресурсов — в другом. Разделение управления ресурсами (ФАНО РФ) и исследованиями (РАН) существенно снижает эффективность науки, так как они имеют разные целевые критерии своей деятельности.

## 5 Заключение

Несколько выходя за рамки рассматриваемой проблематики, хотелось бы сделать два замечания более общего характера.

1. Основная задача науки состоит в научном обеспечении социально-экономического развития страны, поэтому наличие в структуре Системы базы совокупных данных научных организаций РАН по инновационным разработкам позволит в перспективе более объективно подойти к оценкам эффективности фундаментальной науки. В мировой практике оценка результатов фундаментальной науки, как правило, проводится на основе показателей публикационной



активности и цитирования, а также экспертных оценок научного сообщества. В настоящее время российским научным и экономическим сообществом не выработаны единые подходы к оценке эффективности фундаментальной науки. Очевидно, что уровень развития отечественной промышленности, особенно ее высокотехнологичного сектора, т. е. ее «платежеспособный спрос» на результаты исследований и разработок, будет существенно влиять на оценку эффективности фундаментальной науки. (По статистике РФФИ каждый десятый завершенный проект имеет прикладной инновационный потенциал и коммерческую, рыночную перспективу.)

2. Задача управления финансированием фундаментальных исследований является частью более общей задачи управления наукой как особым сегментом национальной экономики. Актуальная задача академического сообщества, и в первую очередь руководства РАН, — разработка пакета рабочих предложений для формирования научно обоснованной стратегии социально-экономического развития России. Это далеко не тривиальная задача, так как в российской экономике есть два существенных, если не сказать важнейших, фактора-детерминанта. Они взаимосвязаны и взаимозависимы.

*Во-первых*, это явная структурно-функциональная недостаточность существующего «промежуточного слоя» между фундаментальной наукой и промышленностью, необходимого для создания инновационных продуктов и трансфера технологий. До постсоветского периода «промежуточный слой» состоял из отраслевых прикладных научно-исследовательских и проектных организаций. В постсоветский период этот «промежуточный слой» практически деградировал, по отдельным направлениям необратимо деформирован и фактически утратил имевшийся НТП. Сейчас в разных отраслях экономики с разным уровнем эффективности функции «промежуточного слоя» выполняют технопарки, внедренческие центры, венчурные фонды, инжиниринговые центры, бизнес-инкубаторы, кластеры и отдельные сохранившиеся и приспособившиеся к новым условиям научно-исследовательские институты и конструкторские бюро (в основном в научно-производственных объединениях).

*Во-вторых*, несоответствие возможностей существующей национальной информационной инфраструктуры современным требованиям новой российской экономической институци-

ональной среды. Это важно, так как во многом основой успешного инновационного развития отраслей промышленности является использование информационных технологий. Новая парадигма экономического развития предполагает в качестве важнейшего фактора конкурентоспособности максимально широкое «вплетение» цифровых информационных технологий в ткань любых производственных, технологических и управленческих процессов.

Таким образом, вторая важная (и безальтернативная) задача — разработка эффективных механизмов (экономических, правовых) вовлечения, в той или иной форме, научных организаций РАН в модернизацию существующих отраслей российской экономики.

## Литература

1. Авдулов А. Н., Кулькин А. М. Финансирование науки в развитых странах мира. — М.: ИНИОН РАН, 2007. 114 с.
2. Калачихин П. А. Принципы построения государственной наукометрической системы // Научно-техническая информация. Сер. 2, 2016. № 7. С. 11–23.
3. Сюнтюрэнко О. В., Черепанов Е. В. Информатика: анализ данных и эконометрия // Средства связи, 1986. № 4. С. 39–44.
4. Коголовский М. Р., Паринов С. И. Новый источник данных для наукометрических исследований // Тр. XV Всеросс. научной конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». — Ярославль: ЯрГУ им. П. Г. Демидова, 2013. С. 107–117.
5. Антопольский А. Б. О целесообразности российского национального вебметрического индекса // Научно-техническая информация. Сер. 1, 2014. № 2. С. 14–18.
6. Месропян В. Р., Овсянников М. В. Перспективы использования наукометрических методов в прогнозировании // Научно-техническая информация. Сер. 1, 2014. № 2. С. 19–27.
7. Сюнтюрэнко О. В., Гиляревский Р. С. Использование методов наукометрии и сопоставительного анализа данных для управления научными исследованиями по тематическим направлениям // Научно-техническая информация. Сер. 2, 2016. № 12. С. 1–12.
8. Борисова Л. Ф., Сюнтюрэнко О. В. Реферативный банк данных ВИНТИ РАН: перспективы постобработки информации с использованием методов анализа данных // Научно-техническая информация. Сер. 1, 2007. № 11. С. 6–11.
9. О внесении изменений в Федеральную целевую программу «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014–2020 гг.:

- Постановление Правительства РФ от 25 сентября 2017 г. № 1156.
10. Ладный А. О. Анализ данных в задачах управления научно-техническим потенциалом. 2012. <http://it-claim.ru/Library/Books/ITS/wwwbook/ist6/ladni/ladni.htm>.
  11. Петровский А. Б., Бойченко В. С., Стернин М. Ю., Шепелев Г. И. Выбор приоритетов научно-технического развития: опыт зарубежных стран // Тр. Института системного анализа РАН, 2015. Т. 65. № 3. С. 13–26.
  12. Обоснование выбора приоритетов научно-технологического развития. Федеральный портал PROTOWN.RU, 2015. <http://www.protown.ru/information/hidden/4500.html>.
  13. Биктимиров М. Р., Гиляревский Р. С., Сюнтюренко О. В. Новая концептуальная основа развития информационной деятельности ВИНИТИ РАН // Научно-техническая информация. Сер. 1, 2016. № 1. С. 1–8.
  14. Сюнтюренко О. В. Производство информационно-аналитических продуктов и услуг с использованием методов наукометрии и анализа данных // Информация в современном мире: Мат-лы Междунар. конф. к 65-летию ВИНИТИ РАН. — М.: ВИНИТИ, 2017. С. 317–321.

Поступила в редакцию 26.12.17

---

---

## FINANCING OF BASIC RESEARCH: CONCEPTUAL SHAPE OF A SYSTEM OF SUPPORT OF DECISION-MAKING WITH USE OF METHODS OF SCIENTOMETRICS AND ANALYSIS OF DATA

O. V. Syuntyurenko

All Russian Institute of Scientific and Technical Information of the Russian Academy of Sciences, 20a Usievich Str., Moscow 125190, Russian Federation

**Abstract:** The article is devoted to development of a conceptual model of an information system focused on solution of the task of management of financing of basic research with use of methods of scientometrics and analysis of data. The basic principles and the methodology of creation are formulated. Functioning of the system of support of decision-making (SSDM) in the process of distribution of fund of the state order in thematic scientific directions is suggested. The functional structure of SPPR is presented. The prospects of use of a new multipurpose information resource and analytical postinformation processing are considered. The interrelation of tasks of management of science with problems of innovative development of branches of economy is shown.

**Keywords:** scientometrics; analysis of data; control systems; scientific priorities; monitoring; scientific and technical potential; postdata processing

**DOI:** 10.14357/19922264180115

### Acknowledgments

The work was supported by the Russian Foundation for Basic Research (projects 17-07-00256).

### References

1. Avdulov, A. N., and A. M. Kul'kin. 2007. *Finansirovanie nauki v razvitykh stranakh mira* [Science funding in the developed countries of the world]. Moscow: INION RAN. 114 p.
2. Kalachikhin, P. A. 2016. Printsipy postroeniya gosudarstvennoy nauko-metricheskoy sistemy [Principles of creation of the state scientometric system]. *Nauchno-tekhnicheskaya informatsiya* [Scientific and Technical Information]. Ser. 2. 7:11–23.
3. Syuntyurenko, O. V., and E. V. Cherepanov. 1986. Informatika: analiz dannykh i ekonometriya [Informatics: Data analysis and econometrics]. *Sredstva svyazi* [Means of Communications] 4:39–44.
4. Kogalovskiy, M. R., and S. I. Parinov. 2013. Novyy istochnik dannykh dlya nauko-metricheskikh issledovaniy [New data source for scientometric researches]. *Tr. 15-y Vseross. konf. "Elektronnye biblioteki: Perspektivnye metody i tekhnologii, elektronnye kollektzii"* [All-Russian Conference "Electronic Libraries and Collections"]. Yaroslavl': YarGU im. P. G. Demidova. 107–117.
5. Antopol'skiy, A. B. 2014. O tselesoobraznosti rossiyskogo natsional'nogo vebometricheskogo indeksa [About expediency of the Russian national vebometrical in-

- dex]. *Nauchno-tekhnicheskaya informatsiya* [Scientific and Technical Information]. Ser. 1. 2:14–18.
6. Mesropyan, V. R., and M. V. Ovsyannikov. 2014. Perspektivy ispol'zovaniya naukometricheskikh metodov v prognozirovanii [Perspectives of use of scientometric methods in forecasting]. *Nauchno-tekhnicheskaya informatsiya* [Scientific and Technical Information]. Ser. 1. 2:19–27.
  7. Syuntyurenko, O. V., and R. S. Gilyarevskiy. 2016. Ispol'zovanie metodov naukometrii i sopostavitel'nogo analiza dannykh dlya upravleniya nauchnymi issledovaniyami po tematicheskim napravleniyam [Use of scientometric methods and comparative data analysis for management of scientific research in the thematic directions] *Nauchno-tekhnicheskaya informatsiya* [Scientific and Technical Information]. Ser. 2. 12:1–12.
  8. Borisova, L. F., and O. V. Syuntyurenko. 2007. Referativnyy bank dannykh VINITI RAN: perspektivy postobrabotki informatsii s ispol'zovaniem metodov analiza dannykh [VINITI Abstract's databank: Perspectives of postinformation processing and use of data analysis]. *Nauchno-tekhnicheskaya informatsiya* [Scientific and Technical Information]. Ser. 1. 11:6–11.
  9. Postanovlenie Pravitel'stva RF ot 25 sentyabrya 2017 g. No. 1156 "O vnesenii izmeneniy v Federal'nyuyu tselevuyu programmu "Issledovaniya i razrabotki po prioritetyim napravleniyam razvitiya nauchno-tekhnologicheskogo kompleksa Rossii na 2014–2020 gg." [About modification of the Federal target program "Research and development in the priority directions of development of a scientific and technological complex of Russia for 2014–2020:" Resolution of the Government of the Russian Federation No. 1156 of September 25, 2017].
  10. Ladnyy, A. O. 2012. Analiz dannykh v zadachakh upravleniya nauchno-tekhnicheskimi potentsialom [Data analysis in tasks of management of scientific and technical potential]. Available at: <http://it-claim.ru/Library/Books/ITS/wwwbook/ist6/ladni/ladni.htm> (accessed February 5, 2018).
  11. Boychenko, V. S., A. B. Petrovsky, M. Yu. Sternin, and G. I. Shepelev. 2015. Vybory prioriteto nauchno-tekhnicheskogo razvitiya: opyt zarubezhnykh stran [Setting the priorities of scientific and technological development: The experience of foreign countries]. *Proceedings of the Institute for System Analysis of RAS* 65:13–26.
  12. Obosnovanie vybora prioriteto nauchno-tekhnologicheskogo razvitiya [Choice justification of priorities of scientific and technical development]. 2015. Available at: <http://www.protown.ru/information/hidden/4500.html> (accessed February 5, 2018).
  13. Biktimirov, M. R., R. S. Gilyarevskiy, and O. V. Syuntyurenko. 2016. Novaya kontseptual'naya osnova razvitiya informatsionnoy deyatelnosti VINITI RAN [New conceptual basis of the development of the activity of VINITI]. *Nauchno-tekhnicheskaya informatsiya* [Scientific and Technical Information]. Ser. 1. 1:1–8.
  14. Syuntyurenko, O. V. 2017. Proizvodstvo informatsionno-analiticheskikh produktov i uslug s ispol'zovaniem metodov naukometrii i analiza dannykh [Production of information and analytical products and services with use of scientometric methods and data analysis]. *Information in the modern world: VINITI 65 Anniversary Conference (International) Proceedings*. Moscow. 317–321.

Received December 26, 2017

## Contributor

**Syuntyurenko Oleg V.** (b. 1946) — Doctor of Science in technology, professor, leading scientist, All Russian Institute of Scientific and Technical Information of the Russian Academy of Sciences, 20a Usievich Str., Moscow 125190, Russian Federation; olegasu@mail.ru

**Агасандян Геннадий Аршавирович** (р. 1941) — доктор физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Алешин Илья Сергеевич** (р. 1993) — аспирант факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

**Атаева Ольга Муратовна** (р. 1978) — младший научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Виноградов Дмитрий Вячеславович** (р. 1964) — кандидат физико-математических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Вышинский Леонид Леонидович** (р. 1941) — кандидат физико-математических наук, заведующий сектором Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Грушо Александр Александрович** (р. 1946) — доктор физико-математических наук, профессор, заведующий лабораторией Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Егоров Артём Юрьевич** (р. 1992) — студент магистратуры Национального исследовательского университета «Высшая школа экономики»

**Забжайло Михаил Иванович** (р. 1956) — доктор физико-математических наук, доцент, заведующий лабораторией Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Зацаринный Александр Алексеевич** (р. 1951) — доктор технических наук, профессор, заместитель директора Федерального исследовательского центра

«Информатика и управление» Российской академии наук

**Зубарев Денис Владимирович** (р. 1989) — аспирант Российского университета дружбы народов, инженер-исследователь Института системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Ковалёв Сергей Протасович** (р. 1972) — доктор физико-математических наук, ведущий научный сотрудник Института проблем управления им. В. А. Трапезникова Российской академии наук

**Кривенко Михаил Петрович** (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Кудрявцев Алексей Андреевич** (р. 1978) — кандидат физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

**Малашенко Юрий Евгеньевич** (р. 1946) — доктор физико-математических наук, заведующий сектором Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Назарова Ирина Александровна** (р. 1966) — кандидат физико-математических наук, научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Новикова Наталья Михайловна** (р. 1953) — доктор физико-математических наук, профессор, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Серебряков Владимир Алексеевич** (р. 1946) — доктор физико-математических наук, профессор, заведующий отделом Вычислительного центра им. А. А. Дородницына Федерального исследовательского



ского центра «Информатика и управление» Российской академии наук

**Синицын Игорь Николаевич** (р. 1940) — доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Соченков Илья Владимирович** (р. 1985) — кандидат физико-математических наук, старший научный сотрудник Сколковского института науки и технологий; ведущий научный сотрудник Института системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Сучков Александр Павлович** (р. 1954) — доктор технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Сюнтюренко Олег Васильевич** (р. 1946) — доктор технических наук, профессор, ведущий научный сотрудник Всероссийского института научной и технической информации Российской академии наук

**Тимонина Елена Евгеньевна** (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Тихомиров Илья Александрович** (р. 1980) — кандидат технических наук, доцент, заведующий лабораторией Института системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Флеров Юрий Арсениевич** (р. 1942) — член-корреспондент Российской академии наук, доктор физико-математических наук, профессор, заместитель директора Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Шестаков Олег Владимирович** (р. 1976) — доктор физико-математических наук, доцент кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Широков Николай Иванович** (р. 1963) — кандидат физико-математических наук, старший научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

**Шнурков Петр Викторович** (р. 1953) — кандидат физико-математических наук, доцент Национального исследовательского университета «Высшая школа экономики»

---

## Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

---

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или  $\LaTeX$  (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 15 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на **русском и английском языках**:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:  
[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_rus/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_rus/authors.asp) и  
[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_eng/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp);
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

**Описание статьи из журнала:**

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

**Описание статьи из электронного журнала:**

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

**Описание статьи из продолжающегося издания (сборника трудов):**

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal”–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

**Описание материалов конференций:**

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursosberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

**Описание книги (монографии, сборники):**

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

**Описание переводной книги** (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — N.Y.: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. N.Y.: Wiley. 521 p.

**Описание неопубликованного документа:**

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004. Geology and production (NGT GiD). Certificate on official registration of the computer program No. 2004611198. (In Russian, unpubl.)

**Описание интернет-ресурса:**

Pravila tsitirovaniya istochnikov [Rules for the citing of sources]. Available at: <http://www.scribd.com/doc/1034528/> (accessed February 7, 2011).

**Описание диссертации или автореферата диссертации:**

Semenov, V. I. 2003. *Matematicheskoe modelirovanie plazmy v sisteme kompaktnyy tor* [Mathematical modeling of the plasma in the compact torus]. D.Sc. Diss. Moscow. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

**Описание ГОСТа:**

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichstva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

**Описание патента:**

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

**Адрес редакции журнала «Информатика и её применения»:**

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: [rust@ipiran.ru](mailto:rust@ipiran.ru) (Сейфуль-Мулюков Рустем Бадриевич)

<http://www.ipiran.ru/journal/issues/>



---

## Requirements for manuscripts submitted to Journal “Informatics and Applications”

---

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
- mathematical methods for studying complex systems and processes;
- information systems and networks;
- information technologies; and
- architecture and software of computational complexes and networks.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

***Agreement on the transfer of rights to publish:***

*“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.*

*Author(s) signature(s): (name(s), address(es), date).*

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or  $\LaTeX$  (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 15 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English**:

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format: (see site)

[http://www.ipiran.ru/journal/issues/2013\\_07\\_01/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp) and

[http://www.ipiran.ru/journal/issues/2013\\_07\\_01\\_eng/authors.asp](http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp);

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
  - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.  
Important! Keywords must not be sentences;
  - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.net/ru/bgn/>).

Please take into account the following examples of Russian references appearance:

**Article in journal:**

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Rus. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

**Journal article in electronic format:**

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

**Article from the continuing publication (collection of works, proceedings):**

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

**Conference proceedings:**

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursosberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazoidachi” [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]*. Moscow. 267–272.

**Books and other monographs:**

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem [Operation of turbine generators with direct cooling]*. Moscow: Energy Publs. 352 p.

**Dissertation and Thesis:**

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. PhD Thesis. Moscow: IPI RAN. 23 p.

**State standards and patents:**

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.

11. Submissions of files by e-mail must include:

- the journal title and author’s name in the “Subject” field;
- an article and additional materials have to be attached using the “attach” function;
- an electronic version of the article should contain the file with the text and a separate file with figures.

12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

**Editorial Board address:**

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia

Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05

e-mail: [rust@ipiran.ru](mailto:rust@ipiran.ru) (to Prof. Rustem Seyful-Mulyukov)

<http://www.ipiran.ru/english/journal.asp>