

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

**Научный журнал Российской академии наук
(издается под руководством Отделения нанотехнологий
и информационных технологий РАН)**

Издается с 1989 года
Журнал выходит ежеквартально

Учредитель:
**Федеральный исследовательский центр
«Информатика и управление» Российской академии наук**

РЕДАКЦИОННЫЙ СОВЕТ

академик РАН И. А. Соколов — председатель Редакционного совета
академик РАН Г. И. Савин
академик РАН А. Л. Слемпковский
член-корреспондент РАН Ю. Б. Зубарев
профессор Ш. Долев (S. Dolev, Beer-Sheva, Israel)
профессор Ю. Кабанов (Yu. Kabanov, Besancon, France)
профессор М. Никулин (M. Nikulin, Bordeaux, France)
профессор В. Ротарь (V. Rotar, San-Diego, USA)
профессор М. Финкельштейн (M. Finkelstein, Bloemfontein, South Africa)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

академик РАН И. А. Соколов — главный редактор
профессор, д.ф.-м.н. С. Я. Шоргин — заместитель главного редактора
д.т.н. В. Н. Захаров д.ф.-м.н. В. И. Сеницын
проф., д.ф.-м.н. А. И. Зейфман проф., д.т.н. И. Н. Сеницын
проф., д.т.н. В. Д. Ильин проф., д.ф.-м.н. В. Г. Ушаков
проф., д.т.н. К. К. Колин к.ф.-м.н. А. К. Горшенин — отв. секретарь
проф., д.ф.-м.н. В. Ю. Королев к.ф.-м.н. С. А. Христочевский
проф., д.г.-м.н. Р. Б. Сейфуль-Мулюков

Редакция

профессор, д.г.-м.н. Р. Б. Сейфуль-Мулюков
к.ф.-м.н. Е. Н. Арутюнов
С. Н. Стригина

© Федеральный исследовательский центр «Информатика
и управление» Российской академии наук, 2019

Журнал включен в базу данных Russian Science Citation Index (RSCI),
интегрированную с Web of Science

Журнал входит в систему Российского индекса научного цитирования (РИНЦ)
Журнал включен в базу данных CrossRef (систему DOI — Digital Object Identifier),
в базу данных Ulrich's periodicals directory

и в информационную систему «Общероссийский математический портал Math-Net.Ru»

Журнал реферируется в «Реферативном журнале» ВИНТИ
и в системе Google Scholar

Журнал включен в сформированный Минобрнауки России Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук

<http://www.ipiran.ru/journal/collected>

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 29 № 2 Год 2019

СОДЕРЖАНИЕ

Электронная бухгалтерская книга на базе ситуационных центров для цифровой экономики А. А. Грушо, А. А. Зацаринный, Е. Е. Тимонина	4
Модели обнаружения зависимостей во временных рядах в задачах построения прогностических моделей К. Р. Усманова, В. В. Стрижов	12
Сходимость распределения оценки риска пороговой обработки к смеси нормальных законов при случайном объеме выборки О. В. Шестаков	31
Априорное обобщенное распределение Фреше в байесовских моделях баланса А. А. Кудрявцев, С. И. Палионная, В. С. Шоргин	39
Интеллектуализация оперативно-диспетчерского управления сложной региональной динамической электроэнергетической системой А. В. Колесников, С. В. Листопад, Ф. Г. Майтаков	46
Информационные трансформации в процессах обработки геоданных Д. А. Никишин	60
Метод расчета карты глубин на основе мягких операторов М. В. Бобырь, А. Е. Архипов, Н. А. Милостная	71
Практичное определение «программно определяемого» В. Б. Егоров	85
Восстановление сигнала в системах реального времени по равномерным выборкам с уменьшением интервала Найквиста Б. Г. Майоров	95

СИСТЕМЫ И СРЕДСТВА ИНФОРМАТИКИ

Том 29 № 2 Год 2019

СОДЕРЖАНИЕ

Метаобъектный подход к моделированию бизнес-процессов предприятия в рамках единой ERP-системы	
<i>С. И. Спивак, Н. Д. Морозкин, Л. А. Лукьянов</i>	113
Метод поиска выпадающих объектов с использованием параметров неустойчивости обучения	
<i>И. С. Ожерельев, О. В. Сенько, Н. Н. Киселёва</i>	122
Оценка надежности гибридного высокопроизводительного вычислительного комплекса при решении научных задач	
<i>А. А. Зацаринный, А. И. Гаранин, В. А. Кондрашев, К. И. Волович, С. И. Мальковский</i>	135
Методология аннотирования в надкорпусных базах данных	
<i>А. А. Гончаров, О. Ю. Инькова, М. Г. Кружков</i>	148
Принципы организации данных для технологии поддержки конкретно-исторических исследований	
<i>И. М. Адамович, О. И. Волков</i>	161
Модель кооперативного решателя задач на основе цифровых двойников	
<i>В. Д. Ильин</i>	172
Об авторах	180
Правила подготовки рукописей статей	183
Requirements for manuscripts	187

ЭЛЕКТРОННАЯ БУХГАЛТЕРСКАЯ КНИГА НА БАЗЕ СИТУАЦИОННЫХ ЦЕНТРОВ ДЛЯ ЦИФРОВОЙ ЭКОНОМИКИ*

А. А. Грушо¹, А. А. Зацаринный², Е. Е. Тимонина³

Аннотация: Построена и исследована модель электронной бухгалтерской книги (ЭБК) для фиксации транзакций участников экономической деятельности цифровой экономики на региональном уровне. В качестве базовой инфраструктуры для реализации модели предложена сеть ситуационных центров (СЦ). В основу модели положен модернизированный вариант блокчейна — тангл (tangle). В РФ удобно использовать централизованный консенсус, позволяющий использовать частично созданную инфраструктуру СЦ и гарантии государства по безопасности участникам экономической деятельности цифровой экономики. Технология ЭБК на базе тангла решает основные задачи электронного взаимодействия для каждого участника экономической деятельности в цифровой экономике, в частности обеспечивает юридическую значимость информации о соглашениях и их выполнении, обеспечивает контроль непротиворечивости экономической деятельности и информирование органов власти о результатах этой деятельности. Использование для ЭБК инфраструктуры СЦ позволяет решать многие необходимые для цифровой экономики проблемы, связанные с обеспечением средствами, ресурсами и механизмами защиты информации.

Ключевые слова: информационная безопасность; блокчейн; тангл; цифровая экономика; ситуационный центр

DOI: 10.14357/08696527190201

1 Введение

Реализация программы цифровой экономики направлена на внедрение технологий, которые позволят достичь ряда результатов, многие из которых сегодня представляются немислимыми. Так, по прогнозам экспертов Давосского форума в 2015 г. среди результатов, ожидаемых к 2025 г., выделяются, например, такие как [1]:

- обеспечение хранения до 10% всемирного валового продукта по технологии распределенного реестра (блокчейн);
- сбор налогов с использованием технологии блокчейн.

* Работа выполнена при частичной поддержке РФФИ (проект 18-29-03124-мк).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, grusho@yandex.ru

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, AZatsarinny@ipiran.ru

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

В принятой программе цифровой экономики [2] системы распределенного реестра (блокчейн) отнесены к числу сквозных цифровых технологий. Вместе с тем именно эта технология сегодня вызывает множество противоречивых оценок. Это обусловлено недостаточной степенью теоретической проработки технологии блокчейна, а также практически полным отсутствием опыта его применения.

В настоящей статье рассматриваются некоторые методические и организационно-технические подходы к созданию и внедрению одного из важнейших компонентов технологии блокчейна — ЭБК.

Электронная бухгалтерская книга основана на концепции блокчейна [3, 4]. Традиционный блокчейн представляет собой последовательность блоков с записанными в них транзакциями. Для защиты данных от модификации используется принцип двусвязности, в соответствии с которым очередной блок содержит значение хеш-функции от предыдущего блока. Внутри блока транзакции подписаны электронной подписью участников транзакции, поэтому внесение искажений (подделка или изъятие данных из блокчейна) требует целенаправленных негативных воздействий на электронные подписи всех транзакций блока, а для блока в цепочке — на хеш-функции всех блоков цепочки.

Управление записями определяется консенсусом [5]. Первый консенсус в распределенной ЭБК назывался PoW (Proof of Work) и строился аналогично консенсусу биткойна.

Все консенсусы можно разделить на централизованные и децентрализованные. С технической точки зрения централизованные и децентрализованные консенсусы схожи, так как каждый узел сети индивидуально ответствен за безопасность и хранение совместно используемых для ЭБК данных. В обоих механизмах используются традиционные механизмы создания ЭБК (блокчейны). Кроме того, надежность децентрализованного блокчейна основана на согласовании и копировании внесенных в блокчейн блоков.

Централизованный консенсус предполагает концентрацию функций контроля в едином органе (далее — Центр). Это означает, что имеется единый контактный узел формирования и хранения блокчейнов. Примерами централизованных блокчейнов, по существу, служат технологии самой популярной в мире социальной сети Facebook и крупнейшей поисковой системы Google.

Развитием традиционных блокчейнов стал тангл (tangle) [6], который представляется ориентированным ациклическим графом (DAG — Directed Acyclic Graph) и который хранит транзакции. Вершины в DAG — это транзакции. Сеть состоит из узлов, которые являются сущностями, создающими транзакции и обеспечивающими валидацию транзакций.

Основная идея тангла состоит в том, что пользователи должны провести определенную работу для «одобрения» других ранее определенных транзакций. Таким образом, каждый пользователь, создающий транзакцию, вносит вклад в безопасность сети. Узлы проверяют, чтобы «одобренные» транзакции были неконфликтными. Если узел выясняет, что транзакция конфликтует с историей в тангле, то этот узел не «одобрит» конфликтную транзакцию.

Для создания транзакций узел выполняет следующие действия:

- (1) выбирает две или более транзакций для «одобрения»;
- (2) проверяет непротиворечивость выбранных для «одобрения» транзакций;
- (3) использует криптографические методы для формирования специальных сложно вычисляемых данных из «одобренных» транзакций и собственной информации.

В исходном варианте тангл строится на основе децентрализованного консенсуса [6]. В случае централизованного консенсуса криптографические методы для формирования транзакции можно заменить электронной подписью создателя транзакции совместно с электронной подписью Центра. Отметим, что в тангле уже не используются блоки.

В статье предлагается и исследуется модель ЭБК для фиксации транзакций участников экономической деятельности цифровой экономики на региональном уровне. В качестве базовой инфраструктуры для реализации модели предложена сеть СЦ [2, 7–9]. В основу модели положен модернизированный вариант блокчейна — тангл. При этом, как было отмечено выше, исходный вариант тангла построен для распределенного консенсуса. В то же время в работе [9] отмечено, что в РФ удобно пользоваться централизованным консенсусом, который позволяет использовать частично созданную инфраструктуру СЦ и гарантии государства по безопасности участникам экономической деятельности цифровой экономики.

2 Требования к информационной инфраструктуре цифровой экономики на базе ситуационных центров

Вместе с тем для внедрения предлагаемой модели ЭБК необходимо создание информационной инфраструктуры, соответствующей определенным требованиям. Такие требования должны формироваться на основе задач, решаемых с помощью ЭБК.

Представим перечень задач ЭБК в следующем виде:

- (1) создание инфраструктуры для электронных юридически значимых взаимодействий участников экономической деятельности цифровой экономики;
- (2) сохранение информации о транзакциях (информация о любых юридически значимых взаимодействиях участников экономической деятельности цифровой экономики);
- (3) обеспечение безопасности информации о транзакциях;
- (4) контроль легитимности транзакций и обеспечение федеральных органов исполнительной власти данными об экономической деятельности в регионе.

Основой обеспечения юридической значимости электронного взаимодействия участников экономической деятельности служат договоры (соглашения, контракты) и информация о выполнении договоров. Таким образом, электронное взаимодействие участников экономической деятельности цифровой экономики должно обеспечивать дистанционное заключение договора и юридическое подтверждение информации о выполнении или невыполнении договора. Тогда при СЦ создается структура, предоставляющая условия (услугу) выполнения функций Центра, через который происходит валидация всех договоров и валидация их исполнения или неисполнения.

Таким образом, все соглашения направляются в Центр при СЦ, куда также направляется вся информация о выполнении или невыполнении этих соглашений. Отсюда следует первое требование — обеспечение надежной связи участников экономической деятельности с Центром и возможность хранения транзакций в танглах, созданных на основе инфраструктуры СЦ [9].

Электронное взаимодействие участников экономической деятельности требует инфраструктуры изготовления и использования электронной подписи. Для юридической значимости транзакций и тангла по закону необходимо использовать электронную подпись [10]. Отсюда возникает требование создания систем удостоверяющих центров по изготовлению и отслеживанию сертификатов ключей. Для работы с электронной подписью участники и Центр должны иметь комплексы специального программного обеспечения (ПО).

Для электронного взаимодействия с участниками экономической деятельности других регионов необходимо использовать защищенные каналы связи. Здесь также возможно частично использовать возможности связи СЦ различных регионов. Предполагается, что необходимость в защищенной связи в цифровой экономике будет возникать нечасто. Однако всегда нужно иметь инфраструктуру изготовления и распределения ключей для защищенной передачи данных и выполнять требования по повышенной защищенности компьютерных систем.

Системы сбора и хранения информации о транзакциях создаются в виде отдельных танглов для каждого участника экономического взаимодействия. Взаимодействие участников с танглами требует специального протокола. Совокупность танглов должна безопасно храниться, а инфраструктура СЦ также может способствовать их хранению и резервированию.

Учитывая, что создание комплексов резервирования информации и восстановления является, как правило, дорогостоящим мероприятием, предлагается использовать аналогичные компоненты системы СЦ с учетом предъявления к ним дополнительных требований, прежде всего в части информационной безопасности.

Создание танглов требует ПО, которое можно создать централизованно и, следовательно, относительно недорого, пользуясь едиными стандартами оснащения СЦ. В функциональные обязанности СЦ должен входить контроль легитимности транзакций, а также всей экономической деятельности в регионе.

Информация для решения этих задач автоматически собирается за счет функций Центра и с использованием танглов.

3 Протоколы сбора и хранения информации о танглах

Поскольку тангл является развитием традиционного блокчейна, то он сохраняет с помощью механизма двусвязности целостность данных во внутренних вершинах DAG [6]. Это не позволяет вносить изменения ни злоумышленникам, ни коррумпированным руководителям региона.

Новые транзакции при своем создании проверяются на противоречивость, т. е. одновременно должны быть созданы две транзакции в танглах обоих участников (клиентов или владельцев) соглашения или выполнения этого соглашения. После выбора «одобрений» в каждом из узлов транзакция должна быть подписана электронной подписью участника этого тангла и электронной подписью Центра. Созданный и подписанный в Центре хеш о своей транзакции, конкатенированный с информацией в «одобренных» транзакциях, передается в виде уведомления каждому владельцу своего тангла. Тем самым каждый из участников соглашения получает юридически значимое подтверждение о внесении своей транзакции в свой тангл.

Выбор «одобрения» участником соглашения в своем тангле осуществляется на основании имеющегося у него экземпляра его тангла, бесконфликтность которого также подтверждена подписью Центра. Для получения копии своего тангла владелец может запросить Центр и приобрести специальное ПО для его чтения.

Кроме того, используя электронную подпись, владелец может получать подписанное Центром расширение своего графа транзакций.

В работе [6] рассмотрено 5 потенциальных атак на танглы. Во всех случаях автору удалось построить эффективную защиту от этих атак. Также в [6] доказано, что технология тангла устойчива к методам атак, которые можно реализовать на квантовых компьютерах.

4 Заключение

Предложенная в статье технология ЭБК на базе тангла позволяет решать основные задачи электронного взаимодействия для каждого участника экономической деятельности в цифровой экономике, в частности обеспечивает юридическую значимость информации о соглашениях и их выполнении, контроль непротиворечивости экономической деятельности, а также информирование органов власти о результатах этой деятельности.

Использование для ЭБК инфраструктуры СЦ позволяет решать многие необходимые для цифровой экономики проблемы, связанные с обеспечением средствами, ресурсами и механизмами защиты информации. Конечно, использование инфраструктуры СЦ для ЭБК возможно только при расширении базы

СЦ. Однако это значительно дешевле, чем создавать подобную инфраструктуру заново. Важно отметить, что построение ЭБК на инфраструктуре СЦ позволит обеспечить связью Центры в различных регионах, используя избыточность каналов связи, необходимых для СЦ.

Литература

1. Шваб К. Четвертая промышленная революция. — М.: Эксмо, 2016. 208 с.
2. Свон М. Блокчейн: Схема новой экономики / Пер. с англ. — М.: Олимп-Бизнес, 2017. 240 с. (*Swan M. Blockchain: Blueprint for a new economy. Sebastopol, CA, USA: O'Reilly Media, Inc., 2015. 149 p.*)
3. Лелу Л. Блокчейн от А до Я. Все о технологии десятилетия / Пер. с фр. — М.: Эксмо, 2018. 190 с. (*Leloup L. Blockchain: La révolution de la confiance. — Paris: Groupe Eyrolles, 2017. 160 p.*)
4. Memon A. W., Mahmood W. Survey of consensus protocols. 2018. https://www.researchgate.net/publication/328160285_Survey_of_Consensus_Protocols.
5. Попов С. The tangle. 2018. <https://assets.ctfassets.net/r1dr6vzfxhev/2t4uxvslqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/iota1.4.3.pdf>.
6. Цифровая экономика Российской Федерации: Программа, утвержденная распоряжением Правительства РФ от 28.07.2017 № 1632-р. <http://static.government.ru/media/files/9gFM4FHj4PsB79I5v7yLVuPgu4bvR7M0.pdf>.
7. О формировании системы распределенных ситуационных центров, работающих по единому регламенту взаимодействия: Указ Президента Российской Федерации от 25 июля 2013 года № 648.
8. Концепция создания системы распределенных ситуационных центров, работающих по единому регламенту взаимодействия, от 3 октября 2013 года № Пр-2308.
9. Грушо А. А., Зацаринный А. А., Тимонина Е. Е. Блокчейны цифровой экономики на базе системы ситуационных центров и централизованного консенсуса // Радиолокация, навигация, связь: Сб. трудов XXV Междунар. научн.-технич. конф. — Воронеж: Издательский дом ВГУ, 2019. Т. 6. С. 183–191.
10. Об электронной подписи: Федеральный закон от 6 апреля 2011 г. № 63-ФЗ. <https://base.garant.ru/12184522>.

Поступила в редакцию 15.03.19

THE ELECTRONIC LEDGER ON THE BASIS OF SITUATIONAL CENTERS FOR DIGITAL ECONOMY

A. A. Grusho, A. A. Zatsarinny, and E. E. Timonina

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The model of an electronic ledger for fixing transactions of participants of economic activity of digital economy at the regional level is constructed and

investigated. As basic infrastructure for realization of the model, a net of situational centers is suggested. The modernized variant of blockchain — tangle — is the basis for the model. In the Russian Federation, it is convenient to use centralized consensus allowing one to use partially created infrastructure of situational centers and guarantees of the state on security to participants of economic activity of digital economy. The technology of the electronic ledger on the basis of tangles solves the main problems of electronic interaction for each participant of economic activity in digital economy, in particular, provides the legal importance of information on agreements and their performance, provides control of consistency of economic activity and informing authorities on results of this activity. The use of infrastructure of situational centers for electronic ledgers allows solving many problems, necessary for digital economy, connected with providing tools, resources, and mechanisms of information security.

Keywords: information security; blockchain; tangle; digital economy; situational center

DOI: 10.14357/08696527190201

Acknowledgments

The paper was partly supported by the Russian Foundation for Basic Research (project 18-29-03124-mk).

References

1. Schwab, K. 2016. *Chetvertaya promyshlennaya revolyutsiya* [The fourth industrial revolution]. Moscow: Eksmo. 208 p.
2. Swan, M. 2015. *Blockchain: Blueprint for a new economy*. Sebastopol, CA: O'Reilly Media, Inc. 149 p.
3. Leloup, L. 2017. *Blockchain: La révolution de la confiance*. Paris: Groupe Eyrolles. 160 p.
4. Memon, A. W., and W. Mahmood. 2018. *Survey of consensus protocols*. Available at: https://www.researchgate.net/publication/328160285_Survey_of_Consensus_Protocols (accessed April 10, 2019).
5. Popov, S. 2018. The tangle. Available at: https://assets.ctfassets.net/r1dr6vzfxhev/2t4uxvslqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/iota1.4_3.pdf (accessed April 10, 2019).
6. Programma “Tsifrovaya ekonomika Rossiyskoy Federatsii,” utverzhennaya Rasporyazheniem Pravitel'stva RF ot 28.07.2017 No. 1632-r [The Program “Digital Economy of the Russian Federation” approved by Order 1632-r of the Government of the Russian Federation, July 28, 2017].
7. Ukaz Prezidenta Rossiyskoy Federatsii ot 25 iyulya 2013 goda No. 648 “O formirovaniy sistemy raspredelennykh situatsionnykh tsentrov, rabotayushchikh po edinomu reglamentu vzaimodeystviya” [The Decree of the President of the Russian Federation of July 25, 2013, No. 648 “About formation of a system of the distributed situational centers working according to uniform regulations of interaction”].

8. Kontseptsiya sozdaniya sistemy raspredelennykh situatsionnykh tse ntrov, rabotayushchikh po edinomu reglamentu vzaimodeystviya ot 3 oktyabrya 2013 goda No. Pr-2308. [The concept of creation of a system of the distributed situational centers working according to uniform regulations of interaction of October 3, 2013, No. Pr-2308].
9. Grusho, A. A., A. A. Zatsarinny, and E. E. Timonina. 2019. Blokcheyny tsifrovoy ekonomiki na baze sistemy situatsionnykh tse ntrov i tsentralizovannogo konsensusa [Blockchains of digital economy on the basis of the system of the situational centres and the centralized consensus]. *Radiolokatsiya, navigatsiya, svyaz': Sb. trudov XXV Mezhdunar. nauchn.-tekhnich. konf.* [Radarlocation, Navigation, Communication: 25th Scientific and Technical Conference (International) Proceedings]. Voronezh: VSU Pubs. 6:183–191.
10. Federal'nyy zakon ot 6 aprelya 2011 g. No. 63-FZ "Ob elektronnoy podpisi" [Federal law of April 6, 2011, No. 63-FZ "About the digital signature"]. Available at: <https://base.garant.ru/12184522/> (accessed April 10, 2019).

Received March 15, 2019

Contributors

Grusho Alexander A. (b. 1946) — Doctor of Science in physics and mathematics, professor, principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Sciences and Control" of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; grusho@yandex.ru

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 (FRC CSC RAS); principal scientist, Institute of Informatics Problems, FRC CSC RAS; 44-2 Vavilov Str., Moscow 119333, Russian Federation; AZatsarinny@ipiran.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 119133, Russian Federation; eltimon@yandex.ru

МОДЕЛИ ОБНАРУЖЕНИЯ ЗАВИСИМОСТЕЙ ВО ВРЕМЕННЫХ РЯДАХ В ЗАДАЧАХ ПОСТРОЕНИЯ ПРОГНОСТИЧЕСКИХ МОДЕЛЕЙ*

К. Р. Усманова¹, В. В. Стрижов²

Аннотация: При прогнозировании сложноорганизованных временных рядов, зависящих от экзогенных факторов и имеющих множественную периодичность, требуется решить задачу выявления связанных пар рядов. Предполагается, что добавление этих рядов в модель повышает качество прогноза. В данной работе для обнаружения связей между временными рядами предлагается использовать метод сходящегося перекрестного отображения. При таком подходе два временных ряда связаны, если существуют их траекторные подпространства, проекции на которые связаны. В свою очередь, проекции рядов на траекторные подпространства связаны, если окрестность фазовой траектории одного ряда отображается в окрестность фазовой траектории другого ряда. Ставится задача отыскания траекторных подпространств, обнаруживающих связь рядов. Решение этой задачи продемонстрировано на двух наборах рядов: потребление электроэнергии и температура воздуха, объем железнодорожных перевозок нефти и объем добычи нефти.

Ключевые слова: временные ряды; прогнозирование; траекторное подпространство; фазовая траектория; сходящееся перекрестное отображение

DOI: 10.14357/08696527190202

1 Введение

Работа посвящена обнаружению причинно-следственных связей между разнородными временными рядами. Например, прогноз железнодорожных грузоперевозок по различным группам грузов, связь сигналов электрокардиограммы и пульса [1].

Если прогноз временного ряда \mathbf{x} строится с использованием временных рядов $\mathbf{y}_1, \dots, \mathbf{y}_k$, то установление связей ряда \mathbf{x} с $\mathbf{y}_1, \dots, \mathbf{y}_k$ может повысить

*Работа выполнена при поддержке РФФИ (проекты 17-20-01212 и 19-07-0885). Настоящая статья содержит результаты проекта «Статистические методы машинного обучения», выполняемого в рамках реализации Программы Центра компетенций Национальной технологической инициативы «Центр хранения и анализа больших данных», поддерживаемого Министерством науки и высшего образования Российской Федерации по Договору МГУ им. М. В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 11.12.2018 № 13/1251/2018.

¹Московский физико-технический институт, karina.usmanova@phystech.edu

²Вычислительный центр имени А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский физико-технический институт, strijov@ccas.ru

качество прогноза и при этом упростить прогностическую модель. Если установлено, что ряд \mathbf{x} не зависит от ряда \mathbf{y}_i , то \mathbf{y}_i должен быть исключен из прогностической модели. В работе анализируются два подхода обнаружения связи между рядами: тест Гренджера [2, 3] и метод сходящегося перекрестного отображения [4, 5].

В основе теста Гренджера лежит следующий подход. Считаем, что ряд \mathbf{x} зависит от ряда \mathbf{y} (или следует из ряда \mathbf{y}), если использование истории ряда \mathbf{y} при построении прогностической модели статистически значимо повышает качество прогноза ряда \mathbf{x} [2, 3]. Тест Гренджера устанавливает связи между рядами и основан на сравнении качества прогноза, в котором используется история только прогнозируемого ряда, либо прогноза, который дополнительно использует историю других рядов. Если улучшение качества прогноза подтверждается статистически, то говорят, что прогнозируемый ряд связан с этими рядами. Тест Гренджера используется в тех задачах, где требуется исследовать взаимосвязь между развивающимися во времени процессами [6, 7]. Тест Гренджера применим к стационарным временным рядам, поэтому в случае нестационарных рядов их необходимо продифференцировать перед проведением теста.

Недостатком теста Гренджера является то, что при используемом в нем подходе невозможно точно определить структуру зависимости рядов. Например, два ряда могут следовать из третьего, но при отсутствии информации о третьем ряде тест Гренджера установит причинно-следственную связь между первым и вторым рядом, хотя она отсутствует. Проблема точного определения структуры зависимости рядов рассмотрена в работе [8].

В случае, когда тест Гренджера неприменим или не может обнаружить связь между рядами, применяется метод сходящегося перекрестного отображения (convergent cross mapping, CCM). Этот метод основан на оценке того, насколько хорошо один ряд может быть восстановлен с использованием второго. Считается [4, 5], что ряд \mathbf{x} восстанавливается по ряду \mathbf{y} , только если ряд \mathbf{y} влияет на ряд \mathbf{x} . Метод CCM основан на сравнении ближайших соседей в траекторном пространстве ряда \mathbf{x} , полученных с помощью ряда \mathbf{x} и с помощью ряда \mathbf{y} . Проверяется, насколько точно моменты времени, соответствующие ближайшим соседям вектора \mathbf{y}_t , определяют ближайших соседей вектора \mathbf{x}_t .

При построении линейной прогностической модели по временному ряду строится траекторная матрица, играющая роль матрицы объектов. Ответами являются значения ряда в последующие моменты времени. Когда размерность траекторного пространства избыточна, прогностическая модель становится неустойчивой. В этом случае необходимо производить отбор признаков [9, 10]. Метод проекций на латентные структуры (partial least squares, PLS) отбирает наиболее значимые признаки и строит новые признаки как их линейные комбинации [11, 12]. Таким образом, PLS находит подпространство траекторного пространства, проекция на которое наилучшим образом приближает исходный ряд. Снижение размерности применяется при изучении связей между рядами. Проекция на траекторные подпространства позволяют более детально изучить

связь между главными компонентами рядов и найти подпространство, в котором наблюдается связь между рядами.

В данной работе для построения прогноза одного временного ряда по нескольким рядам используется алгоритм многомерной гусеницы (multivariate singular spectrum analysis, MSSA-L) [13]. Этот алгоритм является обобщением на многомерный случай алгоритма анализа спектральных компонент (singular spectrum analysis, SSA) [14–16]. Метод SSA основан на разложении временного ряда в сумму интерпретируемых компонент. Он подразумевает четыре основных шага: запись ряда в виде траекторной матрицы, ее сингулярное разложение, группировка компонент, полученных при сингулярном разложении, по каждой сгруппированной матрице восстанавливается временной ряд. Таким образом исходный временной ряд представляется в виде суммы временных рядов.

Эксперимент проводился на трех парах временных рядов: искусственные ряды, среднесуточная температура и потребление электроэнергии, объем железнодорожных перевозок нефти и объем добычи нефти. Для каждой пары исследовалось наличие связи между компонентами входящих в нее рядов. В экспериментах на реальных данных строился прогноз рядов, использующий обнаруженные связанные компоненты рядов.

2 Обнаружение связей временных рядов

В данном разделе описываются метод сходящегося перекрестного отображения и тест Гренджера.

Метод сходящегося перекрестного отображения. Для заданного временного ряда $\mathbf{x} = [x_1, \dots, x_N]^T$ построим траекторную матрицу:

$$\mathbf{H}_x = \begin{bmatrix} x_1 & x_2 & \cdots & x_{n-1} & x_n \\ x_2 & x_3 & \cdots & x_n & x_{n+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N-n+1} & x_{N-n+2} & \cdots & x_{N-1} & x_N \end{bmatrix}, \quad (1)$$

где n — ширина окна. Обозначим t -ю строку матрицы \mathbf{H}_x как \mathbf{x}_t . Тогда матрица \mathbf{H}_x принимает вид:

$$\mathbf{H}_x = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_T \end{bmatrix}, \quad \mathbf{x}_t = [x_t, x_{t+1}, \dots, x_{t+n-1}], \quad T = N - n + 1. \quad (2)$$

Заметим, что все векторы \mathbf{x}_t принадлежат n -мерному траекторному пространству $\mathbb{H}_x \subseteq \mathbb{R}^n$ ряда \mathbf{x} .

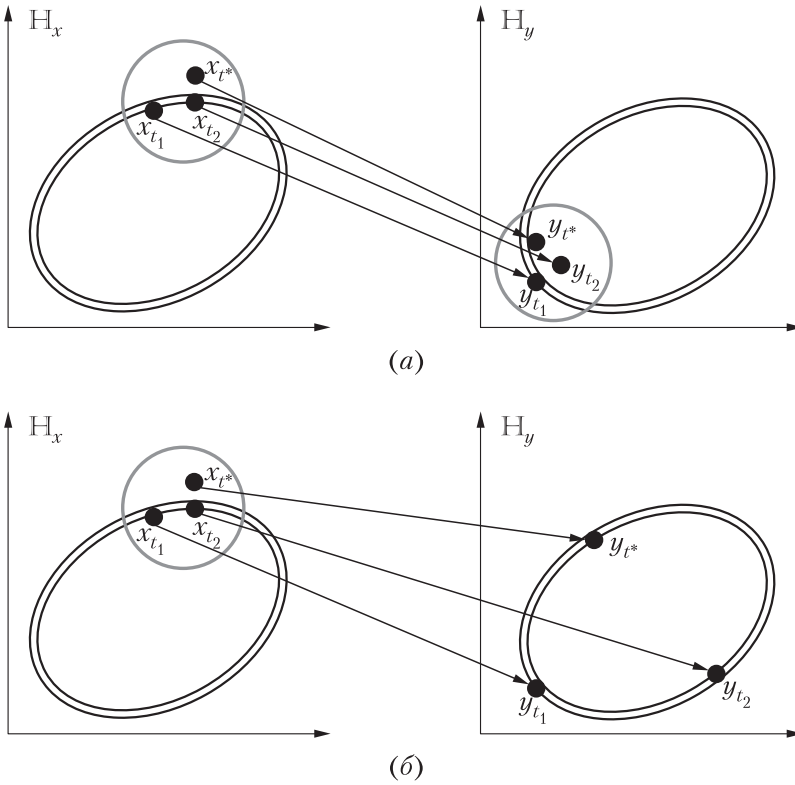


Рис. 1 Отображение из траекторного пространства ряда \mathbf{x} в траекторное пространство ряда \mathbf{y}

Обнаружение зависимости между рядами \mathbf{x} и \mathbf{y} осуществляется следующим образом. Выберем момент $t^* \in [1, T]$ и найдем k ближайших соседей вектора \mathbf{x}_{t^*} в \mathbb{H}_x . Обозначим их множество как $U_k(\mathbf{x}_{t^*}) = \{\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_k}\}$, где

$$\mathbf{x}_{t_i} = [x_{t_i}, x_{t_i+1}, \dots, x_{t_i+n-1}], \quad i = 1, \dots, k. \quad (3)$$

Так как оба ряда \mathbf{x} и \mathbf{y} определены на единой временной оси, отыщем $U_k(\mathbf{y}_{t^*})$ (рис. 1) в пространстве \mathbb{H}_y с целью анализа связей между этими двумя рядами. Для этого повторим в пространстве \mathbb{H}_y построения (1)–(3). Аналогично строим матрицу \mathbf{H}_y . Поставим в соответствие каждому вектору $\mathbf{x}_{t_i} \in U_k(\mathbf{x}_{t^*})$ вектор \mathbf{y}_{t_i} :

$$\varphi : \mathbf{x}_{t_i} \rightarrow \mathbf{y}_{t_i}, \quad i = 1, \dots, k.$$

Найденные векторы \mathbf{y}_{t_i} образуют множество $U_k(\mathbf{y}_{t^*})$. Утверждается, что ряды \mathbf{x} и \mathbf{y} связаны, отображение φ из пространства \mathbb{H}_x в пространство \mathbb{H}_y липшицево:

$$\rho_{\mathbb{H}_y}(\varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j)) \leq n\rho_{\mathbb{H}_x}(\mathbf{x}_i, \mathbf{x}_j), \quad \forall \mathbf{x}_i, \mathbf{x}_j \in \mathbb{H}_x,$$

где $\rho_{\mathbb{H}_x}(\cdot, \cdot)$ и $\rho_{\mathbb{H}_y}(\cdot, \cdot)$ — метрики в пространствах \mathbb{H}_x и \mathbb{H}_y соответственно. Проверим наличие такого отображения следующим образом. Введем меру близости векторов в окрестностях $U_k(\mathbf{x}_{t^*})$ и $U_k(\mathbf{y}_{t^*})$:

$$L(\mathbf{x}, \mathbf{y}) = \frac{R(U_k(\mathbf{x}_{t^*}))}{R(U_k(\mathbf{y}_{t^*}))}, \quad R(U_k(\mathbf{x}_{t^*})) = \frac{1}{k} \sum_{i=1}^k \rho_{\mathbb{H}_x}(\mathbf{x}_{t^*}, \mathbf{x}_{t_i}). \quad (4)$$

Если $L(\mathbf{x}, \mathbf{y})$ больше некоторого порога $L(n)$, то ряд \mathbf{y} зависит от ряда \mathbf{x} .

На рис. 1 показан описанный способ обнаружения связи между рядами. На рис. 1, а ряд \mathbf{y} зависит от ряда \mathbf{x} , на рис. 1, б — не зависит.

Прогноз \hat{x}_t , первого элемента вектора \mathbf{x}_t , строится следующим образом:

$$\hat{x}_t = \sum_{i=1}^k w_i x_{t_i}.$$

Здесь t_i — индексы ближайших соседей \mathbf{x}_t ;

$$w_i = \frac{u_i}{\sum_i u_i}, \quad u_i = \exp\left(-\frac{\rho(\mathbf{x}_t, \mathbf{x}_{t_i})}{\rho(\mathbf{x}_t, \mathbf{x}_{t_{n+1}})}\right),$$

где $\rho(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\|_2$ — метрика в пространстве \mathbb{H}_x .

Анализ связей подпространств траекторных пространств. Предлагается обнаружить связь не в траекторных пространствах \mathbb{H}_x и \mathbb{H}_y , а только в некоторых их подпространствах, натянутых на главные компоненты траекторных матриц \mathbf{H}_x и \mathbf{H}_y , необязательно первые. Сингулярное разложение матрицы \mathbf{H}_x :

$$\mathbf{H}_x = \mathbf{U}_x \mathbf{\Lambda}_x \mathbf{V}_x.$$

Обозначим $\mathcal{J}_x = \{1, \dots, n\}$ — индексы компонент ряда \mathbf{x} . Подпространство траекторного пространства зададим с помощью набора индексов компонент $\mathcal{A} \subseteq \mathcal{J}$, на которые оно будет натянуто. Построим проекцию ряда \mathbf{x} на подпространство, натянутое на компоненты с номерами из \mathcal{A}_x . Обозначим это подпространство $\mathbb{H}_{\mathcal{A}_x} \subseteq \mathbb{H}_x$. Заменим в матрице $\mathbf{\Lambda}_x$ элементы, находящиеся в строках с индексами не из \mathcal{A}_x , нулями. Обозначим полученную матрицу $\tilde{\mathbf{\Lambda}}_x$. Тогда проекция ряда \mathbf{x} на подпространство, натянутое на компоненты с индексами из \mathcal{A}_x , задается траекторной матрицей

$$\mathbf{P}_{\mathcal{A}_x} = \mathbf{U}_x \tilde{\mathbf{\Lambda}}_x \mathbf{V}_x.$$

Аналогично по некоторому набору \mathcal{A}_y строятся подпространство $\mathbb{H}_{\mathcal{A}_y} \subseteq \mathbb{H}_y$ и траекторная матрица $\mathbf{P}_{\mathcal{A}_y}$. Предлагается искать ближайших соседей не в полных траекторных пространствах \mathbb{H}_x и \mathbb{H}_y , задающихся траекторными матрицами \mathbf{H}_x и \mathbf{H}_y соответственно, а в подпространствах $\mathbb{H}_{\mathcal{A}_x}$ и $\mathbb{H}_{\mathcal{A}_y}$, задающихся матрицами $\mathbf{P}_{\mathcal{A}_x}$ и $\mathbf{P}_{\mathcal{A}_y}$. Предполагается, что переход к траекторным подпространствам меньшей размерности повышает устойчивость прогностической модели и позволяет более подробно изучить связь между рядами.

Рассмотрев различные подпространства, выбираем то, которое будет наилучшим образом описывать исследуемый временной ряд и иметь минимальную размерность. Перебор различных подпространств позволяет установить, между какими именно компонентами рядов \mathbf{x} и \mathbf{y} существует зависимость.

Будем перебирать различные комбинации индексов главных компонент и соответствующие им подпространства $\mathbb{H}_{\mathcal{A}_x}$ и $\mathbb{H}_{\mathcal{A}_y}$. Для каждой пары $(\mathcal{A}_x, \mathcal{A}_y)$ индексов главных компонент рядов \mathbf{x} и \mathbf{y} соответственно будем находить среднее расстояние между k ближайшими соседями для ряда \mathbf{x} и между ближайшими соседями для ряда \mathbf{y} . Введем меру близости векторов, аналогичную (4) и нормированную на размерность траекторных подпространств:

$$L(\mathbf{x}, \mathbf{y} | \mathcal{A}_x, \mathcal{A}_y) = \frac{R(U_k(\mathbf{x}_{t^*}) | \mathcal{A}_x) | \mathcal{A}_y |}{R(U_k(\mathbf{y}_{t^*}) | \mathcal{A}_y) | \mathcal{A}_x |}, \quad R(\mathbf{x} | \mathcal{A}_x) = \frac{1}{k} \sum_{i=1}^k \rho_{\mathbb{H}_x}(\mathbf{x}_{t^*}, \mathbf{x}_{t_i}). \quad (5)$$

Тогда задача поиска подпространств $\mathbb{H}_{\mathcal{A}_x}$ и $\mathbb{H}_{\mathcal{A}_y}$ эквивалентна поиску номеров главных компонент $(\mathcal{A}_x, \mathcal{A}_y)$ и имеет вид:

$$(\mathcal{A}_x, \mathcal{A}_y) = \arg \max_{\mathcal{A}_x, \mathcal{A}_y} L(\mathbf{x}, \mathbf{y} | \mathcal{A}_x, \mathcal{A}_y), \quad \mathcal{A}_x, \mathcal{A}_y \subseteq \{1, \dots, n\}.$$

Тест Гренджера. В качестве базового метода установления связей используется статистический тест Гренджера. Требуется проверить, зависит ли ряд \mathbf{x} от ряда \mathbf{y} . Выдвинем гипотезу о независимости ряда \mathbf{x} от ряда \mathbf{y} и проверим ее. Для этого строим прогноз $\hat{\mathbf{x}}$ ряда \mathbf{x} без использования ряда \mathbf{y} и находим невязку $\boldsymbol{\varepsilon}_x = \hat{\mathbf{x}} - \mathbf{x}$. Строим прогноз ряда \mathbf{x} с использованием ряда \mathbf{y} и находим невязку $\boldsymbol{\varepsilon}_{xy}$. Рассмотрим статистику:

$$\Theta(\mathbf{x}, \mathbf{y}) = \frac{N - 2n}{n} \frac{\|\boldsymbol{\varepsilon}_x\|_2^2 - \|\boldsymbol{\varepsilon}_{xy}\|_2^2}{\|\boldsymbol{\varepsilon}_{xy}\|_2^2},$$

где N — длина обучающей выборки; n — размерность регрессионной модели. Статистика Θ имеет распределение Фишера с параметрами $(n, N - 2n)$.

Если ряд \mathbf{x} не зависит от ряда \mathbf{y} , то значения $\boldsymbol{\varepsilon}_x$ и $\boldsymbol{\varepsilon}_{xy}$ будут близки, значение статистики $\Theta(\mathbf{x}, \mathbf{y})$ — мало. Значит, p -value гипотезы о независимости рядов \mathbf{x} и \mathbf{y} на уровне значимости α имеет вид:

$$p\text{-value} = \inf \{ \alpha : \Theta(\mathbf{x}, \mathbf{y}) > \theta(\alpha) \}.$$

Другими словами, для уровня значимости α утверждаем, что ряд \mathbf{x} зависит от ряда \mathbf{y} , если $\Theta(\mathbf{x}, \mathbf{y}) > \theta(\alpha)$. Критерий зависимости ряда \mathbf{x} от ряда \mathbf{y} размера α выглядит следующим образом:

Из $\Theta(\mathbf{x}, \mathbf{y}) > \theta(\alpha)$ следует, что ряд \mathbf{x} зависит от ряда \mathbf{y} .

Для приближенного, восстановленного с помощью модели MSSA-L ряда $\hat{\mathbf{y}}$ проверим зависимость от него ряда \mathbf{x} . Для этого используем статистику $\Theta(\mathbf{x}, \hat{\mathbf{y}})$.

Чтобы найти сдвиг по времени ряда \mathbf{y} относительно ряда \mathbf{x} , вычислим кросскорреляционную функцию

$$\gamma_{\mathbf{xy}}(h) = \frac{\mathbb{E}[(\mathbf{x}_t - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{y}_{t+h} - \boldsymbol{\mu}_{\mathbf{y}})]}{\sigma_{\mathbf{x}}^2 \sigma_{\mathbf{y}}^2}, \quad (6)$$

где \mathbb{E} — математическое ожидание; $\boldsymbol{\mu}$ — выборочное среднее; σ^2 — выборочная дисперсия; $h^* = \arg \max_h \gamma_{\mathbf{xy}}(h)$ — сдвиг по времени ряда \mathbf{y} относительно ряда \mathbf{x} . Использование ряда \mathbf{y} , сдвинутого на h отсчетов назад, при построении прогноза ряда \mathbf{x} повышает качество прогноза.

3 Постановка задачи прогнозирования

Поставим задачу прогноза набора временных рядов. Обозначим через

$$\mathbf{X} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(s)}]^\top \quad (7)$$

заданный набор из s временных рядов. Рассмотрим сегмент набора рядов $\mathbf{X}_{(t-n+1) \div t}$ от момента времени $(t - n + 1)$ до момента t :

$$\underbrace{\begin{pmatrix} x_1^{(1)} \cdots x_{t-n}^{(1)} & \begin{pmatrix} x_{t-n+1}^{(1)} \cdots \overbrace{x_t^{(1)}}^{\boldsymbol{\chi}_t} \end{pmatrix} & x_t^{(1)} \cdots \\ \vdots & \vdots & \vdots \\ x_1^{(s)} \cdots x_{t-n}^{(s)} & \begin{pmatrix} x_{t-n+1}^{(s)} \cdots x_t^{(s)} \end{pmatrix} & x_t^{(s)} \cdots \end{pmatrix}}_{\mathbf{X}_{(t-n+1) \div t}} \quad (8)$$

Пусть $\boldsymbol{\chi}_t = [x_t^{(1)}, \dots, x_t^{(s)}]^\top$ — столбец матрицы (8), значение ряда \mathbf{X} в момент времени t . Построим прогноз $\hat{\boldsymbol{\chi}}_t$ ряда \mathbf{X} в точке $\boldsymbol{\chi}_t$. Проделаем это l раз для сегментов ряда $\mathbf{X}_{(t-n+1+i) \div (t+i)}$, где $i = 1, \dots, l$. Получим l прогнозов $\hat{\boldsymbol{\chi}}_{t \div (t+l-1)} = [\boldsymbol{\chi}_t, \boldsymbol{\chi}_{t+1}, \dots, \boldsymbol{\chi}_{t+l-1}]$ ряда \mathbf{X} в точках $\boldsymbol{\chi}_t, \boldsymbol{\chi}_{t+1}, \dots, \boldsymbol{\chi}_{t+l-1}$. Построение прогнозов $\hat{\boldsymbol{\chi}}_{t \div (t+l-1)}$ изображено на рис. 2.

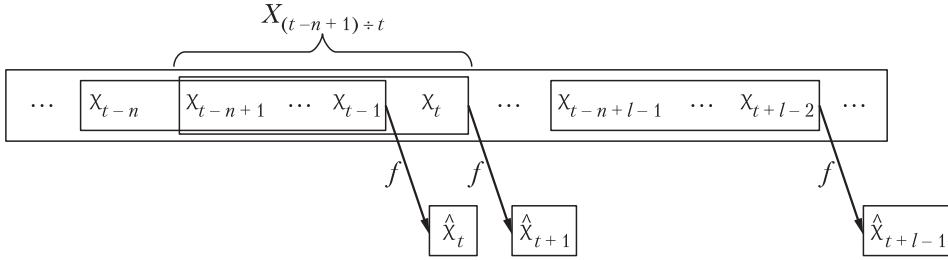


Рис. 2 Построение прогнозов для значений набора $\chi_t, \chi_{t+1}, \dots, \chi_{t+l-1}$ временных рядов в моменты времени $t, t+1, \dots, t+l-1$

Прогностическая модель имеет вид:

$$\hat{\chi}_t = \mathbf{f}(\hat{\mathbf{w}}, \chi_t, \chi_{t-1}, \dots, \chi_{t-n+1}),$$

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} S(\mathbf{w}, \mathbf{X}, \hat{\chi}_t, \hat{\chi}_{t+1}, \dots, \hat{\chi}_{t+n-1}) = S(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(s)}),$$

где функция потерь

$$S(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(s)}) = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(\chi_{t+i}, \hat{\chi}_{t+i}) = \frac{1}{n} \sum_{i=1}^n (x_{t+i}^{(1)} - \hat{x}_{t+i}^{(1)})^2.$$

В данной работе для построения прогноза набора из s временных рядов (7) используется алгоритм многомерной гусеницы (MSSA-L), представляющий собой обобщение на многомерный случай алгоритма гусеницы (SSA). Задача алгоритма MSSA-L состоит в представлении временного ряда в виде суммы интерпретируемых компонент. Это осуществляется в четыре шага: запись ряда в виде траекторной матрицы (1), сингулярное разложение этой матрицы (9), группировка компонент (10), полученных при сингулярном разложении, в интерпретируемые компоненты и восстановление временного ряда по каждой из интерпретируемых компонент.

По ряду $\mathbf{x}^{(i)}$ построим траекторную матрицу $\mathbf{H}^{(i)} \in \mathbb{R}^{T \times n}$, $T = N - n + 1$, согласно (1). Для набора временных рядов (8) построим объединенную траекторную матрицу

$$\mathbf{H} = [\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \dots, \mathbf{H}^{(s)}].$$

По траекторной матрице \mathbf{H} можно восстановить временной ряд \mathbf{X} . Метод многомерной гусеницы строит приближение $\hat{\mathbf{H}}$ матрицы \mathbf{H} меньшего ранга с по-

мощью сингулярного разложения этой матрицы и восстанавливает ряд по матрице $\hat{\mathbf{H}}$. Сингулярное разложение матрицы \mathbf{H} имеет вид:

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V} = \sum_{i=1}^d \lambda_i \mathbf{u}_i \mathbf{v}_i^T, \quad (9)$$

где $\lambda_1, \dots, \lambda_d > 0$ — сингулярные числа матрицы \mathbf{H} ; \mathbf{u}_i и \mathbf{v}_i — столбцы матриц \mathbf{U} и \mathbf{V} . Наилучшее приближение матрицы \mathbf{H} матрицей ранга $r < d$ имеет вид:

$$\hat{\mathbf{H}} = \sum_{i=1}^r \lambda_i \mathbf{u}_i \mathbf{v}_i^T. \quad (10)$$

По матрице $\hat{\mathbf{H}}$ восстанавливается временной ряд \mathbf{X} путем усреднения элементов, стоящих на антидиагоналях.

Алгоритм многомерной гусеницы строит прогноз временного ряда \mathbf{X} в момент t по $(n - 1)$ предыдущим значениям ряда. Алгоритм находит такой вектор коэффициентов $\mathbf{p} \in \mathbb{R}^{(n-1)}$, что значения набора рядов \mathbf{X} в момент t равны:

$$\chi_t = \begin{pmatrix} x_{t-n+1}^{(1)} \cdots x_{t-1}^{(1)} \\ x_{t-n+1}^{(2)} \cdots x_{t-1}^{(2)} \\ \vdots \\ x_{t-n+1}^{(s)} \cdots x_{t-1}^{(s)} \end{pmatrix} \cdot \mathbf{p}.$$

Заметим, что коэффициенты \mathbf{p} оказываются общими для всех компонент ряда \mathbf{X} . Вектор \mathbf{p} — это вектор коэффициентов в линейной комбинации первых $(n - 1)$ столбцов траекторной матрицы, наилучшим образом приближающей последний столбец матрицы. Вектор \mathbf{p} определяется методом наименьших квадратов.

Для каждого $i \in [1, r]$ обозначим через $\tilde{\mathbf{u}}_i$ первые $(n - 1)$ компонент столбца \mathbf{u}_i , π_i — последнюю компоненту столбца \mathbf{u}_i и $\nu = \sum_{i=1}^r \pi_i^2$. Тогда вектор коэффициентов \mathbf{p} вычисляется по формуле:

$$\mathbf{p} = \frac{1}{1 - \nu^2} \sum_{i=1}^r \pi_i \tilde{\mathbf{u}}_i.$$

4 Вычислительный эксперимент

Эксперимент проводился на трех парах временных рядов: искусственные ряды, потребление электроэнергии и среднесуточная температура воздуха, объем грузоперевозок нефти и цена на нефть. Для каждой пары рядов ставилась задача нахождения связанных траекторных подпространств с помощью метода сходящегося перекрестного отображения.

4.1 Сгенерированные данные

Эксперимент проводился на двух сгенерированных рядах:

$$\mathbf{x} = \sin t + 2 \sin \frac{t}{2} + \sigma_x^2 \varepsilon, \quad \sigma_x^2 = 0,3;$$

$$\mathbf{y} = \sin(2t + 5) + \sigma_y^2 \varepsilon, \quad \sigma_y^2 = 0,25,$$

где $\varepsilon \in \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Строим траекторную матрицу \mathbf{H}_x по ряду \mathbf{x} , взяв ширину окна $n = 250$. Для некоторого момента времени t^* рассмотрим вектор \mathbf{x}_{t^*} , равный t^* -й строке

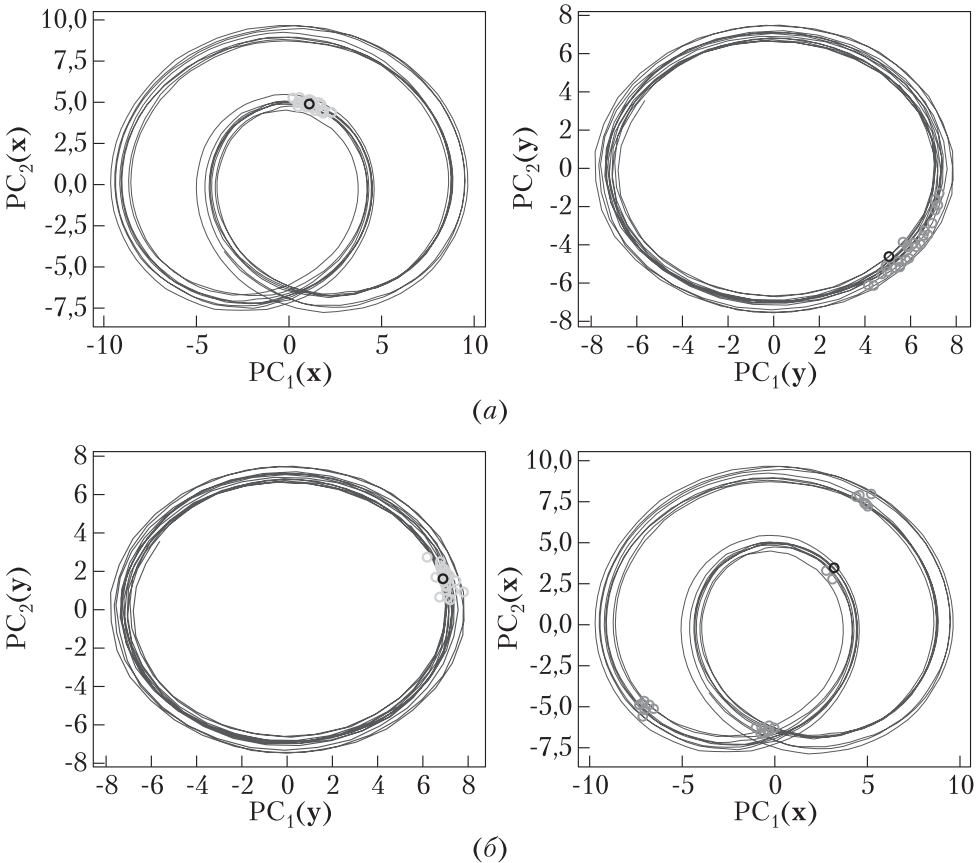


Рис. 3 Проекция $U_k(\mathbf{x}_{t^*})$ и $U_k(\mathbf{y}_{t^*})$ на фазовые траектории (25 соседей): (а) отображение из пространства \mathbb{H}_x в пространство \mathbb{H}_y ; (б) отображение из пространства \mathbb{H}_y в пространство \mathbb{H}_x

матрицы \mathbf{H}_x . Выберем k и найдем среди строк матрицы \mathbf{H}_x k ближайших (в смысле евклидовой нормы) соседей вектора \mathbf{x}_{t^*} . Обозначим индексы найденных векторов t_1, \dots, t_k , а сами найденные вектора — $\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_k}$. Обозначим i -ю строку траекторной матрицы \mathbf{H}_y ряда \mathbf{y} как \mathbf{y}_i . Тогда по найденным индексам t_1, \dots, t_k можно отобразить соответствующие $\mathbf{y}_{t_1}, \dots, \mathbf{y}_{t_k}$. Если ряд \mathbf{y} зависит от ряда \mathbf{x} , то векторы $\mathbf{y}_{t_1}, \dots, \mathbf{y}_{t_k}$, как и $\mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_k}$, будут находиться рядом в траекторном пространстве.

Аналогично для некоторого t^* находим ближайших соседей вектора \mathbf{y}_{t^*} . Обозначим их $\mathbf{y}_{t_1}, \dots, \mathbf{y}_{t_k}$. На рис. 3 изображены фазовые траектории рядов \mathbf{x} и \mathbf{y} , а также проекции векторов $\mathbf{x}_{t^*}, \mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_k}$ и $\mathbf{y}_{t^*}, \mathbf{y}_{t_1}, \dots, \mathbf{y}_{t_k}$ на фазовые траектории. Главные компоненты, образующие оси траекторного подпространства, обозначены $PC_1(\cdot)$ и $PC_2(\cdot)$.

Видим, что окрестность фазовой траектории ряда \mathbf{x} отображается в окрестность фазовой траектории ряда \mathbf{y} и наоборот. При отображении из траекторного пространства ряда \mathbf{y} в траекторное пространство ряда \mathbf{x} векторы $\mathbf{x}_{t^*}, \mathbf{x}_{t_1}, \dots, \mathbf{x}_{t_{15}}$ распадаются на четыре плотные группы. Это связано с тем, что период ряда \mathbf{y} в четыре раза меньше периода ряда \mathbf{x} .

4.2 Эксперимент на данных потребления электроэнергии и температуры

В эксперименте исследуется ряд объема потребления электроэнергии \mathbf{x} и ряд значений температуры \mathbf{y} в течение года. Так как эти ряды не являются стационарными, их необходимо продифференцировать и отнормировать перед тем, как исследовать зависимости между ними. Ряд температуры приведем к стационарной форме следующим образом. Рассмотрим ряд длины светового дня в течение года \mathbf{z} . С помощью вычисления кросскорреляционной функции (6)

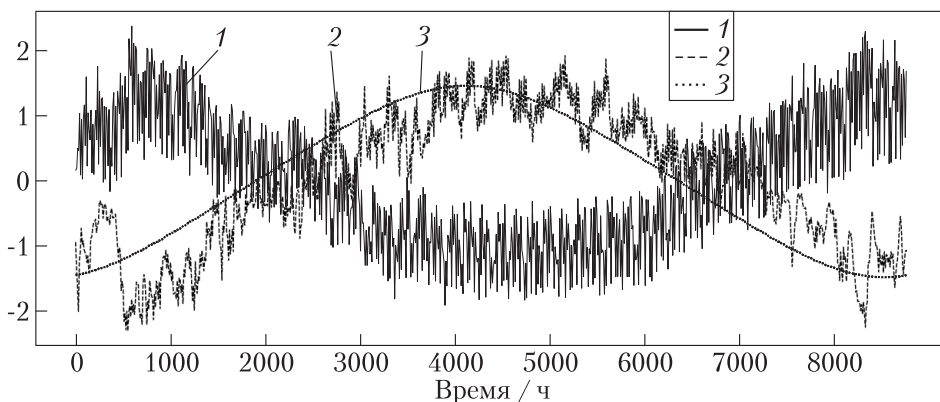


Рис. 4 Нормированные ряды потребления электроэнергии (1), температуры (2) и длины светового дня (3)

рядов \mathbf{y} и \mathbf{z} определим, насколько ряд \mathbf{z} опережает ряд \mathbf{y} , т. е. найдем такое h^* , что $\mathbf{y}(t + h^*) = \mathbf{z}(t)$. Вычтем из ряда \mathbf{y} ряд \mathbf{z} с учетом сдвига h^* . Полученный ряд $\mathbf{y}^*(t) = \mathbf{y}(t) - \mathbf{z}(t - h^*)$ будет стационарным рядом температуры.

Исходные ряды потребления электроэнергии \mathbf{x} , температуры \mathbf{y} и длины светового дня \mathbf{z} изображены на рис. 4.

Максимум модуля кросскорреляции $\gamma_{yz}(h)$ достигается при $h = 560$. Значит,

$$\mathbf{z}(t) = \mathbf{y}(t + 560).$$

И новый стационарный ряд температуры имеет вид:

$$\mathbf{y}^*(t) = \mathbf{y}(t) - \mathbf{z}(t - 560).$$

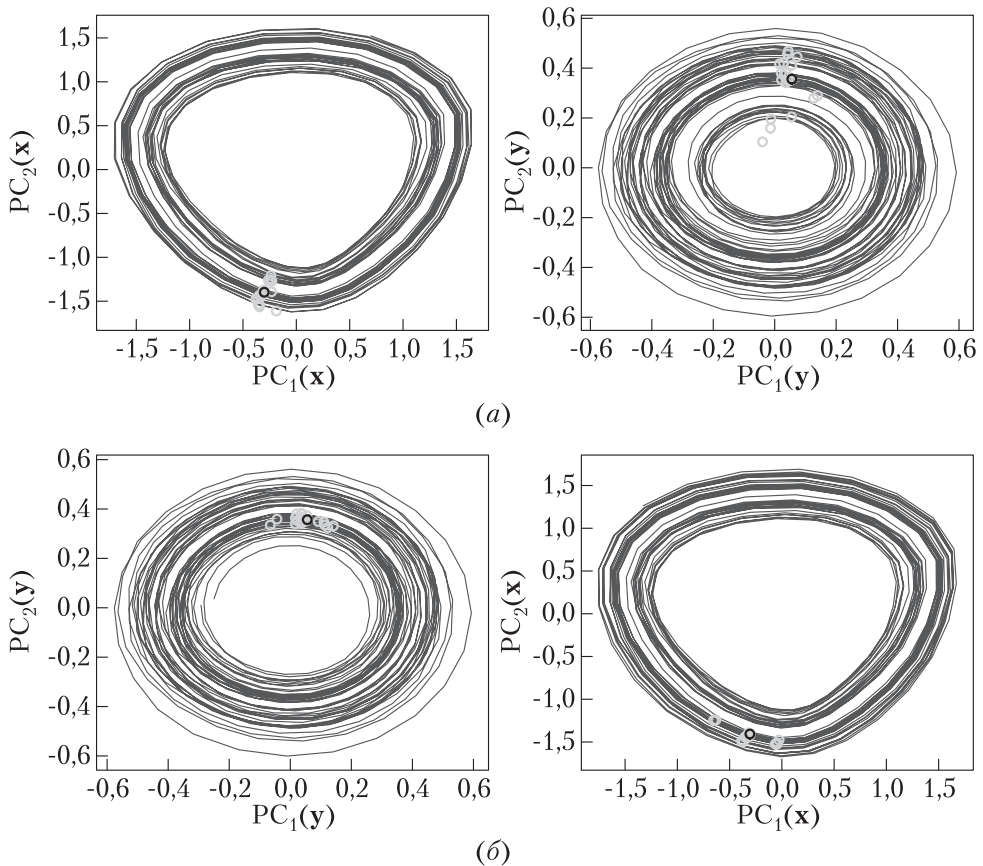


Рис. 5 Векторы $U_k(\mathbf{y}_{t^*})$ (ближайшие соседи вектора \mathbf{y}_{t^*}) и соответствующие векторы $U_k(\mathbf{x}_{t^*})$ на фазовых диаграммах с периодом 24 ч (20 соседей): (a) $\mathbb{H}_x \Rightarrow \mathbb{H}_y$; (б) $\mathbb{H}_y \Rightarrow \mathbb{H}_x$

Далее для удобства полученный ряд температуры \mathbf{y}^* будем обозначать \mathbf{y} , как и исходный.

Исследуем зависимость ряда температуры \mathbf{y} от ряда потребления электроэнергии \mathbf{x} . Делаем это аналогично эксперименту на искусственных данных. Возьмем $n = 170$, что соответствует периоду в семь дней, и $t^* = 400$. Найдем k ближайших соседей векторов \mathbf{x}_{t^*} и \mathbf{y}_{t^*} , их расположение в траекторном пространстве показано на рис. 5.

Перебор подпространств. Переберем траекторные подпространства рядов \mathbf{x} и \mathbf{y} размером не больше пяти. Для этого будем перебирать пары множеств индексов главных компонент $(\mathcal{A}_x, \mathcal{A}_y)$. Для каждой пары $(\mathcal{A}_x, \mathcal{A}_y)$ найдем $R(\mathbf{x}, \mathbf{y} | \mathcal{A}_x, \mathcal{A}_y)$, задающееся (5). Полученные значения $R(\mathbf{x}, \mathbf{y} | \mathcal{A}_x, \mathcal{A}_y)$ представлены на рис. 6.

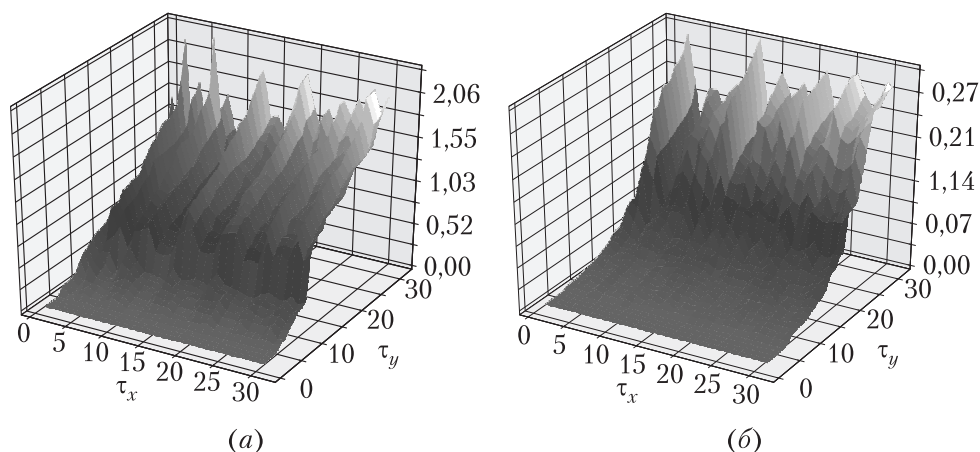


Рис. 6 Значение функционала (5) при отображении из траекторного пространства одного ряда в траекторное пространство другого: (а) $\mathbb{H}_x \Rightarrow \mathbb{H}_y$; (б) $\mathbb{H}_y \Rightarrow \mathbb{H}_x$

Построение прогноза. Построим прогноз ряда потребления электроэнергии \mathbf{x} , используя только его собственную историю, и сравним с прогнозом, строящимся с использованием ряда температуры \mathbf{y} . Рассмотрим также прогнозы, строящиеся при помощи только первых главных компонент ряда \mathbf{y} . На рис. 7 представлены графики зависимости среднеквадратичной ошибки прогноза от ширины окна n .

4.3 Эксперимент на данных об объеме перевозок нефти

В эксперименте проверяются связи между двумя временными рядами: объем грузоперевозок нефти \mathbf{x} и объем добычи нефти \mathbf{q} . Временные ряды заданы за период в 21 год по месяцам. Исходные ряды представлены на рис. 8.

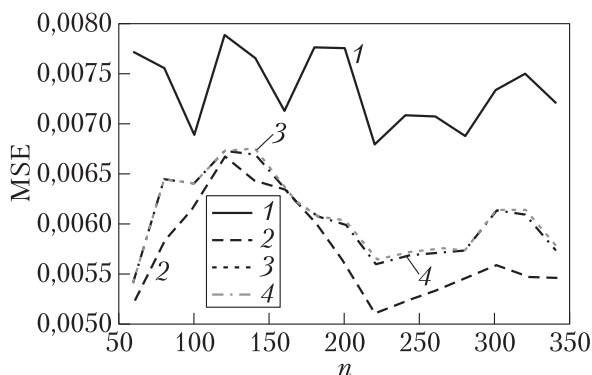


Рис. 7 Среднеквадратичная ошибка прогноза MSE в зависимости от ширины окна: 1 — прогноз по ряду X ; 2 — прогноз по рядам $X, Y(\text{all})$; 3 — прогноз по рядам $X, Y(0, 2)$; 4 — прогноз по рядам $X, Y(0, 1, 2, 3)$

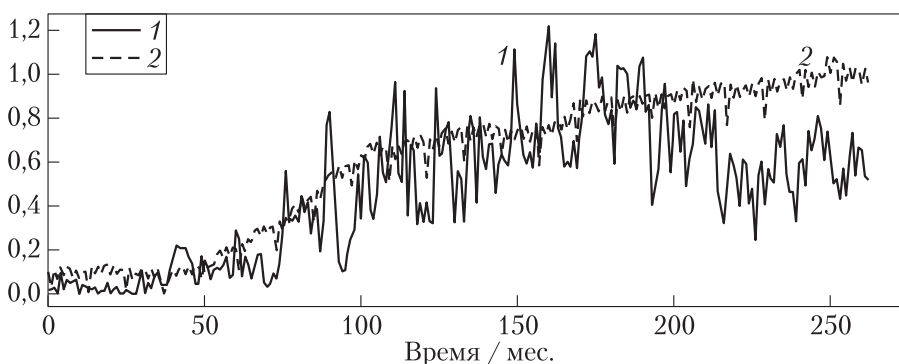


Рис. 8 Исследуемые временные ряды: 1 — импорт нефти, X ; 2 — добыча нефти, Q

Исследуем связи рядов в паре (x, q) . Аналогично предыдущим экспериментам, для проверки наличия связи между рядами смотрим, как точки, близкие в траекторном пространстве одного ряда, отображаются в траекторное подпространство другого ряда. Отображение между фазовыми траекториями рядов x и q показано на рис. 9.

Сравним ошибку прогноза ряда x с использованием ряда q и без него. Когда ряд q включен в прогностическую модель, рассмотрим два случая: для прогноза используются все компоненты ряда q и используются только первые компоненты ряда q . Аналогично сравним прогноз ряда q , построенный только по его собственной истории, с прогнозом, использующим историю ряда x . Результаты эксперимента представлены на рис. 10.

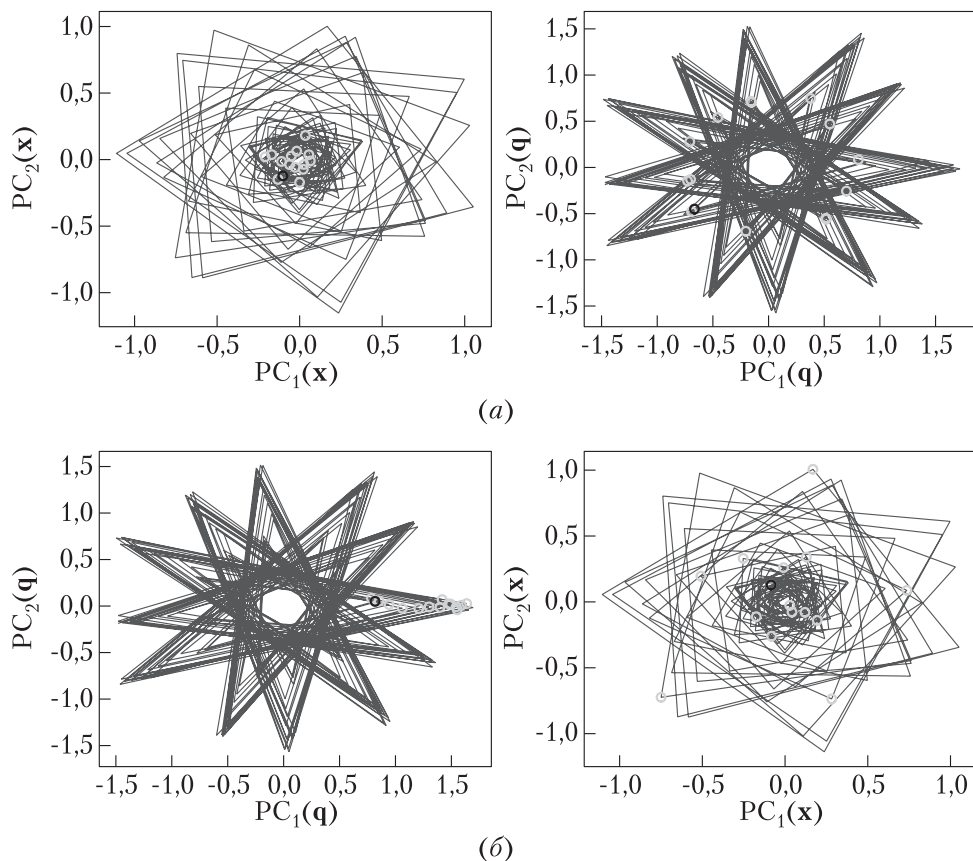


Рис. 9 Связь между траекторными подпространствами рядов \mathbf{x} и \mathbf{q} (15 соседей): (а) $\mathbb{H}_{\mathbf{x}} \Rightarrow \mathbb{H}_{\mathbf{q}}$; (б) $\mathbb{H}_{\mathbf{q}} \Rightarrow \mathbb{H}_{\mathbf{x}}$

5 Заключение

В работе решалась задача обнаружения связи между временными рядами, а также между их компонентами. Связи между временными рядами устанавливались с помощью метода сходящегося перекрестного отображения. Вывод о наличии связи между парой рядов сравнивался с результатом теста Гренджера, проведенного на этих же рядах.

Эксперимент проводился на трех наборах данных: искусственные данные, данные потребления электроэнергии и данные РЖД объема грузоперевозок нефти. Для каждого набора данных сделаны выводы о наличии связей между исследуемыми рядами и их проекциями на траекторные подпространства. На

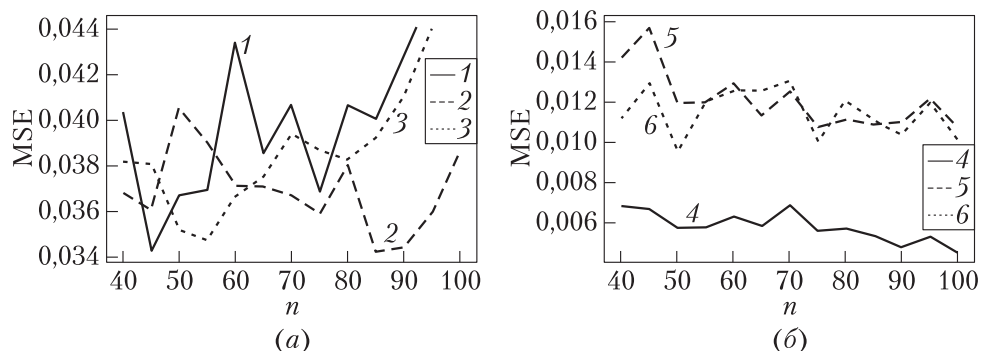


Рис. 10 Зависимость ошибки прогноза MSE от ширины окна n для пары рядов \mathbf{x} (а) и \mathbf{q} (б): прогноз по ряду \mathbf{X} (1), по рядам \mathbf{X}, \mathbf{Q} (all) (2), по рядам \mathbf{X}, \mathbf{Q} (1,2) (3), по ряду \mathbf{Q} (4), по рядам \mathbf{Q}, \mathbf{X} (all) (5) и по рядам \mathbf{Q}, \mathbf{X} (3,4) (6)

данных РЖД видно, что метод ССМ более чувствителен к наличию связи между рядами.

Литература

1. Данные работы по исследованию связи показателей ЭКГ и пульса. <http://ieeexplore.ieee.org/document/7185325>.
2. Granger C. W. J. Investigating causal relations by econometric models and cross-spectral methods // *Econometrica*, 1969. Vol. 37. No. 3. P. 424–438.
3. Barrett A. B., Barnett L., Seth. A. K. Multivariate Granger causality and generalized variance // *Phys. Rev.*, 2010. Vol. 81. No. 4. P. 041907.
4. Sugihara G., May R. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series // *Nature*, 1990. Vol. 344. No. 6268. P. 734–741.
5. Sugihara G., May R., Hao Ye, Chih-hao H., Deyle E., Fogarty M., Munch S. Detecting causality in complex ecosystems // *Science*, 2012. Vol. 338. P. 1227079.
6. Hiemstra C., Jones J. D. Testing for linear and nonlinear Granger causality in the stock price–volume relation // *J. Financ.*, 1994. Vol. 49. No. 5. P. 1639–1664.
7. Hoffmann R., Lee C.-G., Ramasamy B., Yeung M. FDI and pollution: A Granger causality test using panel data // *J. Int. Dev.*, 2005. Vol. 17. No. 3. P. 311–317.
8. White H., Xun L. Granger causality and dynamic structural systems // *J. Financ. Economet.*, 2010. Vol. 8. No. 2. P. 193–243.
9. Katrutsa A. M., Strijov V. V. Stress test procedure for feature selection algorithms // *Chemometr. Intell. Lab.*, 2015. Vol. 142. P. 172–183.
10. Li J., Cheng K., Wang S., Morstatter F., Trevino R. P., Tang J., Huan Liu. Feature selection: A data perspective // *ACM Comput. Surv.*, 2017. Vol. 50. No. 6. P. 1–45.
11. Geladi P. Notes on the history and nature of partial least squares modelling // *J. Chemometr.*, 1988. Vol. 2. No. 4. P. 231–246.

12. *Hoskuldsson A.* Pls regression methods // *J. Chemometr.*, 1988. Vol. 2. No.3. P. 211–228.
13. *Golyandina N., Stepanov D.* SSA-based approaches to analysis and forecast of multi-dimensional time series // 5th St. Petersburg Workshop on Simulation Proceedings. — St. Petersburg: NII Chemistry Saint Petersburg University Publ., 2005. P. 293–298.
14. *Golyandina N., Nekrutkin V., Zhigljavsky A. A.* Analysis of time series structure: SSA and related techniques. — Chapman and Hall, 2002. 320 p.
15. *Golyandina N., Zhigljavsky A.* Singular spectrum analysis for time series. — Springer Science & Business Media, 2013. 120 p.
16. *Elsner J. B., Tsonis A. A.* Singular spectrum analysis: A new tool in time series analysis. — Springer Science & Business Media, 2013. 164 p.

Поступила в редакцию 21.12.18

MODELS OF DETECTION RELATIONSHIP BETWEEN TIME SERIES IN FORECASTING PROBLEMS

K. R. Usmanova¹ and V. V. Strijov^{1,2}

¹Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region, 141701, Russian Federation

²Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The problem of forecasting multiple time series requires detection of relationship between them. Engagement of related time series in a forecast model boosts the forecast quality. This paper introduces the convergent cross mapping (CCM) method used to detect a relationship between time series. This method estimates accuracy of reconstruction of one time series using the other series. The CCM method detects relationship between series not only in full trajectory spaces, but also in some trajectory subspaces. The computational experiment was carried out on two sets of time series: electricity consumption and air temperature, oil transportation volume and oil production volume.

Keywords: time series; forecasting; trajectory subspace; phase trajectory; convergent cross mapping

DOI: 10.14357/08696527190202

Acknowledgments

This research was supported by the Russian Foundation for Basic Research (projects 17-20-01212 and 19-07-0885). This paper contains results of the project “Statistical methods of machine learning,” which is carried out within the framework of the

Program “Center of Big Data Storage and Analysis” of the National Technology Initiative Competence Center. It is supported by the Ministry of Science and Higher Education of the Russian Federation according to the Agreement between the M. V. Lomonosov Moscow State University and the Foundation of Project Support of the National Technology Initiative from 11.12.2018, No. 13/1251/2018.

References

1. Dannye raboty po issledovaniyu svyazi pokazateley EKG i pul'sa [Data from research about relationship between ECG indicators and pulse]. Available at <http://ieeexplore.ieee.org/document/7185325> (accessed May 17, 2019).
2. Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3):424–438.
3. Barrett, A. B., L. Barnett, and A. K. Seth. 2010. Multivariate Granger causality and generalized variance. *Phys. Rev.* 81(4):041907.
4. Sugihara, G., and R. May. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* 344(6268):734–741.
5. Sugihara, G., R. May, Y. Hao, H. Chih-hao, E. Deyle, M. Fogarty, and S. Munch. 2012. Detecting causality in complex ecosystems. *Science* 338(6106):496–500.
6. Hiemstra, C., and J. D. Jones. 1994. Testing for linear and nonlinear Granger causality in the stock price – volume relation. *J. Financ.* 49(5):1639–1664.
7. Hoffmann, R., C.-G. Lee, B. Ramasamy, and M. Yeung. 2005. FDI and pollution: A Granger causality test using panel data. *J. Int. Dev.* 17(3):311–317.
8. White, H., and L. Xun. 2010. Granger causality and dynamic structural systems. *J. Financ. Economet.* 8(2):193–243.
9. Katrutsa, A. M., and V. V. Strijov. 2015. Stress test procedure for feature selection algorithms. *Chemometr. Intell. Lab.* 142:172–183.
10. Li, J., K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and L. Huan. 2017. Feature selection: A data perspective. *ACM Comput. Surv.* 50(6):1–45.
11. Geladi, P. 1988. Notes on the history and nature of partial least squares modelling. *J. Chemometr.* 2(4):231–246.
12. Hoskuldsson, A. 1988. Pls regression methods. *J. Chemometr.* 2(3):211–228.
13. Golyandina, N., and D. Stepanov. 2005. SSA-based approaches to analysis and forecast of multidimensional time series. *5th St. Petersburg Workshop on Simulation Proceedings*. St. Petersburg: NII Chemistry Saint Petersburg University Publs. 293–298.
14. Golyandina, N., V. Nekrutkin, and A. A. Zhigljavsky. 2002. *Analysis of time series structure: SSA and related techniques*. Chapman and Hall. 320 p.
15. Golyandina, N., and A. Zhigljavsky. 2013. *Singular spectrum analysis for time series*. Springer Science & Business Media. 120 p.
16. Elsner, J. B., and A. A. Tsonis. 2013. *Singular spectrum analysis: A new tool in time series analysis*. Springer Science & Business Media. 164 p.

Received December 21, 2018

Contributors

Usmanova Karina R. (b. 1996) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; karina.usmanova@phystech.edu

Strijov Vadim V. (b. 1967) — Doctor of Science in physics and mathematics, leading scientist, Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; professor, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region, 141701, Russian Federation; strijov@ccas.ru

СХОДИМОСТЬ РАСПРЕДЕЛЕНИЯ ОЦЕНКИ РИСКА ПОРОГОВОЙ ОБРАБОТКИ К СМЕСИ НОРМАЛЬНЫХ ЗАКОНОВ ПРИ СЛУЧАЙНОМ ОБЪЕМЕ ВЫБОРКИ*

*О. В. Шестаков*¹

Аннотация: Популярность алгоритмов обработки сигналов, использующих методы вейвлет-анализа, значительно возросла за последние десятилетия. Объясняется это тем, что вейвлет-разложение представляет собой удобный математический аппарат, способный решать задачи, в которых применение традиционного Фурье-анализа оказывается неэффективным. Основные задачи, для решения которых используются методы вейвлет-анализа, — это компрессия сигналов и удаление шума. При этом чаще всего используется пороговая обработка коэффициентов вейвлет-разложения, которая обнуляет коэффициенты, не превышающие заданного порога. Наличие шума и процедуры пороговой обработки неизбежно приводят к погрешностям в оцениваемом сигнале. Свойства оценки таких погрешностей (среднеквадратичного риска) исследовались во многих работах. В частности, показано, что при определенных условиях оценка риска является сильно состоятельной и асимптотически нормальной. При использовании методов пороговой обработки обычно предполагается, что число вейвлет-коэффициентов фиксировано. Однако в некоторых ситуациях объем выборки заранее не известен и моделируется случайной величиной. В данной работе рассматривается модель со случайным числом наблюдений и описывается класс распределений, которые могут быть предельными для оценки среднеквадратичного риска.

Ключевые слова: пороговая обработка; случайный объем выборки; оценка среднеквадратичного риска

DOI: 10.14357/08696527190203

1 Введение

Задачи обработки и анализа сигналов возникают во многих прикладных областях, включая физику плазмы, медицину, геофизику, астрономию, радиолокацию, системы связи и передачи информации и т. д. В некоторых ситуациях объем данных (число отсчетов сигнала) заранее не известен. Время получения сигнала может быть ограничено, а моменты регистрации отсчетов могут быть

*Работа выполнена при финансовой поддержке Российского научного фонда (проект 18-11-00155).

¹Московский государственный университет им. М. В. Ломоносова, кафедра математической статистики факультета вычислительной математики и кибернетики; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, oshestakov@cs.msu.ru

случайными. Может отсутствовать точная информация о характеристиках используемого оборудования или возникать пропуски данных. В этом случае часто предполагается, что объем выборки представляет собой случайную величину с некоторым распределением вероятностей.

Часто одним из этапов анализа и обработки сигналов является преобразование, приводящее к его «экономному» представлению. Примером такого преобразования служит разложение по некоторому вейвлет-базису. Затем осуществляется пороговая обработка коэффициентов разложения. В результате часть коэффициентов, которая считается несущественной, обнуляется. Такие процедуры позволяют подавить большую часть шума, содержащегося в сигнале, и осуществить компрессию данных для более эффективной передачи по каналам связи. В моделях с фиксированным объемом выборки эти процедуры хорошо изучены и получены выражения для оптимальных параметров обработки, ориентированные на разные классы сигналов и функции потерь [1–6]. Также исследованы статистические свойства оценок среднеквадратичного риска, которые дают возможность получить представление о возникающих погрешностях, используя только наблюдаемые данные [7, 8]. В данной работе рассматривается модель со случайным числом вейвлет-коэффициентов функции сигнала и описывается класс распределений, к которым может сходиться распределение оценки риска при возрастающем объеме выборки.

2 Пороговая обработка вейвлет-коэффициентов

Рассмотрим следующую модель данных:

$$X_i = f_i + \varepsilon_i, \quad i = 1, \dots, 2^J,$$

где f_i — отсчеты функции сигнала; ε_i — независимые случайные величины, имеющие нормальное распределение с нулевым средним и дисперсией σ^2 . Предполагается, что $f \in \text{Lip}(\gamma)$, т. е. функция сигнала принадлежит классу функций, регулярных по Липшицу с некоторым показателем $\gamma > 0$.

После вейвлет-преобразования сигнала получается набор эмпирических вейвлет-коэффициентов, имеющих следующий вид:

$$Y_{j,k} = \mu_{j,k} + z_{j,k}, \quad j = 0, \dots, J-1, \quad k = 0, \dots, 2^j - 1,$$

где $\mu_{j,k}$ — вейвлет-коэффициенты «чистого» сигнала; $z_{j,k}$ — «шумовые» коэффициенты, которые имеют такую же структуру, как и ε_i .

Популярным методом удаления шума является пороговая обработка эмпирических вейвлет-коэффициентов, которая заключается в обнулении коэффициентов, чьи абсолютные значения не превышают заданного порога. Оценка $\hat{Y}_{j,k}$ вычисляется с помощью пороговой функции $\rho_T(Y_{j,k})$ с порогом T . Наиболее распространены функция жесткой пороговой обработки $\rho_T^{(h)}(x) = x\mathbf{1}(|x| > T)$

и мягкой пороговой обработки $\rho_T^{(s)}(x) = \text{sign}(x)(|x| - T)_+$. Смысл такой обработки заключается в том, что, поскольку большинство «чистых» коэффициентов малы по абсолютному значению, обнуление эмпирических коэффициентов должно удалить шум, не сильно затронув полезный сигнал.

Основным критерием качества методов пороговой обработки выступает среднеквадратичный риск:

$$r_J = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E} \left(\hat{Y}_{j,k} - \mu_{j,k} \right)^2. \quad (1)$$

Выражение (1) содержит значения «чистых» коэффициентов $\mu_{j,k}$, и поэтому в практических задачах его можно использовать только для вычисления величины риска обработки «модельных» (известных) функций сигналов. Когда функция сигнала неизвестна, можно построить статистическую оценку r_J . Если в слагаемом $|Y_{j,k}| > T$, то вклад этого слагаемого в риск составляет σ^2 в случае жесткой и $\sigma^2 + T^2$ в случае мягкой пороговой обработки, а если $|Y_{j,k}| \leq T$, то вклад составляет $\mu_{j,k}^2$ в обоих случаях. Поскольку $\mathbb{E}Y_{j,k}^2 = \sigma^2 + \mu_{j,k}^2$, величину $\mu_{j,k}^2$ следует оценить разностью $Y_{j,k}^2 - \sigma^2$. Таким образом, в качестве оценки среднеквадратичного риска можно использовать величину:

$$\hat{r}_J = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} F[Y_{j,k}],$$

где

$$F[Y_{j,k}] = \begin{cases} \left(Y_{j,k}^2 - \sigma^2 \right) \mathbf{1}(|Y_{j,k}| \leq T) + \sigma^2 \mathbf{1}(|Y_{j,k}| > T) & \text{в случае жесткой пороговой обработки;} \\ \left(Y_{j,k}^2 - \sigma^2 \right) \mathbf{1}(|Y_{j,k}| \leq T) + (\sigma^2 + T^2) \mathbf{1}(|Y_{j,k}| > T) & \text{в случае мягкой пороговой обработки.} \end{cases}$$

Для мягкой пороговой обработки эта оценка является несмещенной [9]. Кроме того, справедливо утверждение об асимптотической нормальности \hat{r}_J [10].

Теорема 1. Если $f \in \text{Lip}(\gamma)$ с $\gamma > 0$ при мягкой пороговой обработке и $\gamma > 1/2$ при жесткой пороговой обработке, то при выборе асимптотически оптимального порога

$$\mathbb{P} \left(\frac{\hat{r}_J - r_J}{\sigma^2 \sqrt{2^{J+1}}} < x \right) \rightarrow \Phi(x) \quad \text{при } J \rightarrow \infty,$$

где $\Phi(x)$ — функция распределения стандартного нормального закона.

Асимптотически оптимальный порог в теореме 1 при $J \rightarrow \infty$ удовлетворяет соотношению [4]:

$$T \simeq \sigma \sqrt{\frac{4\gamma}{2\gamma + 1} \ln 2^J}.$$

В следующем разделе показывается, что в модели со случайным числом вейвлет-коэффициентов предельное распределение оценки риска уже не обязано быть нормальным.

3 Оценка риска в модели со случайным числом вейвлет-коэффициентов

Пусть M — положительная целочисленная случайная величина. Тогда оценка среднеквадратичного риска принимает вид:

$$\widehat{R}_M = \sum_{j=0}^{M-1} \sum_{k=0}^{2^j-1} F[Y_{j,k}]. \quad (2)$$

Чтобы получить осмысленные асимптотические результаты для оценки (2), величина M должна быть «большой». Рассмотрим последовательность 2^{M_J} , $J = 1, \dots$, и исследуем поведение (2) при $J \rightarrow \infty$.

Теорема 2. Пусть выполнены условия теоремы 1. Пусть также $M_J \xrightarrow{P} \infty$ и для некоторой последовательности R_J совместное распределение пар

$$(U_J, V_J) \equiv \left(\frac{2^{M_J}}{2^J}, \frac{r_{M_J} - R_J}{\sigma^2 \sqrt{2^{J+1}}} \right)$$

сходится к распределению некоторого случайного вектора (U, V) при $J \rightarrow \infty$. Тогда

$$P \left(\frac{\widehat{R}_{M_J} - R_J}{\sigma^2 \sqrt{2^{J+1}}} < x \right) \rightarrow H(x) \quad \text{при } J \rightarrow \infty, \quad (3)$$

где $H(x)$ — функция распределения, соответствующая характеристической функции $h(t) = E\varphi(t\sqrt{U})e^{itV}$ ($\varphi(t)$ — характеристическая функция стандартного нормального закона).

Таким образом, предельные распределения в (3) представляют собой сдвиг-масштабные смеси нормальных законов.

Доказательство. Пусть $h_J(t)$ — характеристическая функция $(\widehat{R}_{M_J} - R_J)/(\sigma^2 \sqrt{2^{J+1}})$.

Тогда

$$h_J(t) = \sum_{j=1}^{\infty} P(M_J = j) \exp\left(it \frac{r_j - R_J}{\sigma^2 \sqrt{2^{J+1}}}\right) \varphi_j\left(t \frac{\sqrt{2^j}}{\sqrt{2^J}}\right),$$

где $\varphi_j(t)$ — характеристическая функция $(\hat{r}_j - r_j)/(\sigma^2 \sqrt{2^{j+1}})$. Обозначим

$$g_J(t) = E\varphi\left(t\sqrt{U_J}\right) e^{itV_J} = \sum_{j=1}^{\infty} P(M_J = j) \exp\left(it \frac{r_j - R_J}{\sigma^2 \sqrt{2^{J+1}}}\right) \varphi\left(t \frac{\sqrt{2^j}}{\sqrt{2^J}}\right).$$

Имеем

$$\begin{aligned} |h_J(t) - g_J(t)| &= \\ &= \left| \sum_{j=1}^{\infty} P(M_J = j) \exp\left(it \frac{r_j - R_J}{\sigma^2 \sqrt{2^{J+1}}}\right) \left(\varphi_j\left(t \frac{\sqrt{2^j}}{\sqrt{2^J}}\right) - \varphi\left(t \frac{\sqrt{2^j}}{\sqrt{2^J}}\right)\right) \right|. \end{aligned}$$

Фиксируем $t \neq 0$. Рассуждая, как в работе [11], для произвольного $\varepsilon > 0$ можно выбрать такое число $\kappa > 0$, что

$$\sum_{j > \kappa + J} P(M_J = j) = P(U_J > 2^\kappa) < \varepsilon$$

и, начиная с некоторого J , зависящего от ε и κ ,

$$\left| \sum_{j > \kappa + J} P(M_J = j) \exp\left(it \frac{r_j - R_J}{\sigma^2 \sqrt{2^{J+1}}}\right) \left(\varphi_j\left(t \frac{\sqrt{2^j}}{\sqrt{2^J}}\right) - \varphi\left(t \frac{\sqrt{2^j}}{\sqrt{2^J}}\right)\right) \right| < 2\varepsilon.$$

Далее в силу теоремы 1

$$\begin{aligned} &\left| \sum_{j \leq \kappa + J} P(M_J = j) \exp\left(it \frac{r_j - R_J}{\sigma^2 \sqrt{2^{J+1}}}\right) \left(\varphi_j\left(t \frac{\sqrt{2^j}}{\sqrt{2^J}}\right) - \varphi\left(t \frac{\sqrt{2^j}}{\sqrt{2^J}}\right)\right) \right| \leq \\ &\leq \sum_{j \leq \kappa + J} P(M_J = j) \sup_{\tau \leq 2^{\kappa/2} t} |\varphi_j(\tau) - \varphi(\tau)| < E \sup_{\tau \leq 2^{\kappa/2} t} |\varphi_{M_J}(\tau) - \varphi(\tau)| < \varepsilon, \end{aligned}$$

так как $M_J \xrightarrow{P} \infty$ [11].

Таким образом, для произвольного t

$$|h_J(t) - g_J(t)| \rightarrow 0 \quad \text{при} \quad J \rightarrow \infty.$$

Далее при любом фиксированном t

$$|g_J(t) - h(t)| = \left| \mathbf{E}\varphi(t\sqrt{U_J})e^{itV_J} - \mathbf{E}\varphi(t\sqrt{U})e^{itV} \right| \rightarrow 0 \quad \text{при } J \rightarrow \infty,$$

поскольку функция $q_t(u, v) = \varphi(t\sqrt{u})e^{itv}$ непрерывна и ограничена и $(U_J, V_J) \Rightarrow (U, V)$ при $J \rightarrow \infty$. Следовательно,

$$|h_J(t) - h(t)| \rightarrow 0 \quad \text{при } J \rightarrow \infty$$

и имеет место (3). Теорема доказана.

Замечание. Распределения, получаемые в качестве предельных в теореме 2, могут существенно отличаться по своим свойствам от нормального закона. В частности, они могут иметь гораздо более тяжелые хвосты. Кроме того, порядок теоретического среднеквадратичного риска в модели со случайным числом вейвлет-коэффициентов может быть существенно больше порядка риска в модели с фиксированным числом коэффициентов [12].

Литература

1. Donoho D., Johnstone I. M. Ideal spatial adaptation via wavelet shrinkage // *Biometrika*, 1994. Vol. 81. No. 3. P. 425–455.
2. Donoho D., Johnstone I. M., Kerkycharian G., Picard D. Wavelet shrinkage: Asymptopia? // *J. Roy. Stat. Soc. B*, 1995. Vol. 57. No. 2. P. 301–369.
3. Donoho D., Johnstone I. M. Minimax estimation via wavelet shrinkage // *Ann. Stat.*, 1998. Vol. 26. No. 3. P. 879–921.
4. Jansen M. Noise reduction by wavelet thresholding. — Lecture notes in statistics ser. — New York, NY, USA: Springer Verlag, 2001. Vol. 161. 196 p.
5. Jansen M. Minimum risk thresholds for data with heavy noise // *IEEE Signal Proc. Let.*, 2006. Vol. 13. No. 5. P. 296–299.
6. Кудрявцев А. А., Шестаков О. В. Асимптотическое поведение порога, минимизирующего усредненную вероятность ошибки вычисления вейвлет-коэффициентов // *Докл. Акад. наук*, 2016. Т. 468. № 5. С. 487–491.
7. Маркин А. В., Шестаков О. В. О состоятельности оценки риска при пороговой обработке вейвлет-коэффициентов // *Вестн. Моск. ун-та. Сер. 15: Вычисл. матем. и киберн.*, 2010. № 1. С. 26–34.
8. Шестаков О. В. Асимптотическая нормальность оценки риска пороговой обработки вейвлет-коэффициентов при выборе адаптивного порога // *Докл. Акад. наук*, 2012. Т. 445. № 5. С. 513–515.
9. Mallat S. A wavelet tour of signal processing. — New York, NY, USA: Academic Press, 1999. 857 p.
10. Шестаков О. В. Вероятностно-статистические методы анализа и обработки сигналов на основе вейвлет-алгоритмов. — М.: Аргамак-Медиа, 2016. 200 с.

11. Korolev V. Yu., Zeifman A. I. On convergence of the distributions of random sequences with independent random indexes to variance-mean mixtures // *Stoch. Models*, 2016. Vol. 32. Iss. 3. P. 414–432.
12. Шестаков О. В. Среднеквадратичный риск пороговой обработки при случайном объеме выборки // *Информатика и её применения*, 2018. Т. 12. Вып. 3. С. 14–17.

Поступила в редакцию 28.02.19

CONVERGENCE OF THE DISTRIBUTION OF THE THRESHOLD PROCESSING RISK ESTIMATE TO A MIXTURE OF NORMAL LAWS AT A RANDOM SAMPLE SIZE

O. V. Shestakov^{1,2}

¹Department 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

²Institute 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 popularity of signal processing algorithms using wavelet analysis methods has increased significantly over the past decades. This is explained by the fact that the wavelet decomposition is a convenient mathematical apparatus capable of solving problems in which the use of traditional Fourier analysis is ineffective. The main tasks for which the methods of wavelet analysis are used are signal compression and noise removal. In this case, the most commonly used method is threshold processing of wavelet expansion coefficients, which zeroes coefficients not exceeding a given threshold. The presence of noise and threshold processing procedures inevitably lead to errors in the estimated signal. The properties of estimates of such errors (mean square risk) have been studied in many papers. In particular, it has been shown that under certain conditions, the risk estimate is strongly consistent and asymptotically normal. When using threshold processing methods, it is usually assumed that the number of wavelet coefficients is fixed. However, in some situations, the sample size is not known in advance and is modeled by a random variable. In this paper, a model with a random number of observations is considered and a class of distributions is described that can be limiting for the mean-square risk estimate.

Keywords: threshold processing; random sample size; mean square risk estimate

DOI: 10.14357/08696527190203

Acknowledgments

This research is supported by the Russian Science Foundation (project No. 18-11-00155).

References

1. Donoho, D., and I. M. Johnstone. 1994. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81(3):425–455.
2. Donoho, D., I. M. Johnstone, G. Kerkyacharian, and D. Picard. 1995. Wavelet shrinkage: Asymptopia? *J. Roy. Stat. Soc. B* 57(2):301–369.
3. Donoho, D., and I. M. Johnstone. 1998. Minimax estimation via wavelet shrinkage. *Ann. Stat.* 26(3):879–921.
4. Jansen, M. 2001. *Noise reduction by wavelet thresholding*. Lecture notes in statistics ser. New York, NY: Springer Verlag. Vol. 161. 196 p.
5. Jansen, M. 2006. Minimum risk thresholds for data with heavy noise. *IEEE Signal Proc. Lett.* 13(5):296–299.
6. Kudryavtsev, A. A., and O. V. Shestakov. 2016. Asymptotic behavior of the threshold minimizing the average probability of error in calculation of wavelet coefficients. *Dokl. Math.* 93(3):295–299.
7. Markin, A. V., and O. V. Shestakov. 2010. Consistency of risk estimation with thresholding of wavelet coefficients. *Moscow Univ. Comput. Math. Cybern.* 34(1):22–30.
8. Shestakov, O. V. 2012. Asymptotic normality of adaptive wavelet thresholding risk estimation. *Dokl. Math.* 86(1):556–558.
9. Mallat, S. 1999. *A wavelet tour of signal processing*. New York, NY: Academic Press. 857 p.
10. Shestakov, O. V. 2016. *Veroyatnostno-statisticheskie metody analiza i obrabotki signalov na osnove veyvlet-algoritmov* [Probabilistic-statistical methods of signal analysis and processing based on wavelet algorithms]. Moscow: Argamak-Media Publ. 200 p.
11. Korolev, V. Yu., and A. I. Zeifman. 2016. On convergence of the distributions of random sequences with independent random indexes to variance-mean mixtures. *Stoch. Models* 32(3):414–432.
12. Shestakov, O. V. 2018. Srednekvadrachnyy risk porogovoy obrabotki pri sluchaynom ob”eme vyborki [Mean-square thresholding risk with a random sample size]. *Informatika i ee Primneniya — Inform. Appl.* 12(3):14–17.

Received February 28, 2019

Contributor

Shestakov Oleg V. (b. 1976) — Doctor of Science in physics and mathematics, 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

АПРИОРНОЕ ОБОБЩЕННОЕ РАСПРЕДЕЛЕНИЕ ФРЕШЕ В БАЙЕСОВСКИХ МОДЕЛЯХ БАЛАНСА*

А. А. Кудрявцев¹, С. И. Палионная², В. С. Шоргин³

Аннотация: Статья продолжает ряд работ авторов в области применения байесовского подхода к моделям массового обслуживания, надежности и баланса. В рамках данного подхода при рассмотрении сложного агрегата все параметры, оказывающие влияние на функционирование системы, разделяются на два класса: способствующие и препятствующие корректному функционированию системы. Изучаются вероятностные характеристики индекса баланса — отношения факторов, негативно влияющих на работу системы, к позитивно влияющим факторам — в предположении, что факторы суть случайные величины с априори известными исследователю распределениями. В работе изучаются вероятностные характеристики индекса баланса системы в случае, когда оба фактора имеют априорное обобщенное распределение Фреше. Результаты представлены в терминах специальной гамма-экспоненциальной функции.

Ключевые слова: байесовский подход; обобщенное распределение Фреше; гамма-экспоненциальная функция; модели баланса; смешанные распределения

DOI: 10.14357/08696527190204

1 Введение

В современном мире появляется все больше систем, координирующих различные сферы человеческой деятельности. Эти системы настолько сложны, что определять их характеристики путем детерминированного анализа практически не представляется возможным. В связи с этим возникают различные индексы, рейтинги и показатели, так или иначе характеризующие состояние системы без проведения затратных по времени и финансам всесторонних исследований.

На состояние системы чаще всего воздействует множество факторов, которые, в свою очередь, можно разделить на два условных класса: положительно и отрицательно влияющие на работу целевого объекта. Факторы, способствующие правильному функционированию системы, далее для удобства записи будем

*Работа выполнена при частичной финансовой поддержке РФФИ (проект 17-07-00577).

¹Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, nubigena@mail.ru

²Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, sofia.palioonnaya@gmail.com

³Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, VShorgin@ipiran.ru

называть *p*-факторами, а факторы, препятствующие функционированию — *n*-факторами. В данной статье для исследования качественных характеристик функционирования системы рассматривается индекс баланса, который определяется как отношение *n*- к *p*-фактору модели. Очевидно, что для достижения баланса в системе нужно избегать большого расхождения между величинами факторов и приближать значение индекса баланса к единице. Далее будем обозначать индекс баланса через $\rho = \lambda/\mu$, где λ и μ соответственно обозначают *n*- и *p*-факторы.

В силу непостоянства состояния окружающей среды, влияющей на функционирование системы, значения *n*- и *p*-факторов модели непредсказуемым образом меняются. Это дает основание для рассмотрения факторов, а следовательно и индексов баланса, как случайных величин [1]. При этом изменения, происходящие в окружающей среде, не столь глобальны, чтобы кардинальным образом изменить природу факторов, поэтому в рамках конкретной задачи можно предположить, что законы, оказывающие влияние на факторы и индексы, остаются неизменными. Такая постановка задачи позволяет применить байесовский метод с заданными априорными распределениями исходных параметров системы.

В статье в качестве априорного распределения факторов рассматривается обобщенное распределение Фреше; приводятся вероятностные характеристики индекса баланса.

2 Основные результаты

Рассмотрим гамма-экспоненциальную функцию [2]

$$\text{Ge}_{\alpha, \beta}(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \Gamma(\alpha k + \beta), \quad x \in \mathbb{R}, \quad 0 \leq \alpha < 1, \quad \beta > 0.$$

Для дальнейших вычислений понадобится следующее утверждение.

Лемма 1. Пусть $\alpha, \theta, r, u, v > 0$. Тогда

$$\int_0^{\infty} y^{-r-1} e^{-(\alpha/y)^u - (\theta/y)^v} dy = \begin{cases} \frac{\theta^{-r}}{v} \text{Ge}_{u/v, r/v} \left(- \left(\frac{\alpha}{\theta} \right)^u \right), & u < v; \\ \frac{\alpha^{-r}}{u} \text{Ge}_{v/u, r/u} \left(- \left(\frac{\theta}{\alpha} \right)^v \right), & u > v; \\ \frac{(\alpha^u + \theta^u)^{-r/u}}{u} \Gamma \left(\frac{r}{u} \right), & u = v. \end{cases}$$

Доказательство. Рассмотрим случай $u < v$. Используя теорему Лебега о предельном переходе, получаем:

$$\int_0^{\infty} y^{-r-1} e^{-(\alpha/y)^u - (\theta/y)^v} dy = \frac{\alpha^{-r}}{u} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_0^{\infty} t^{r/u+k-1} e^{-(\theta/\alpha t^s)^v} dt =$$

$$= \frac{\theta^{-r}}{v} \sum_{k=0}^{\infty} \frac{(-(\alpha/\theta)^u)^k}{k!} \int_0^{\infty} z^{(r+uk)/v-1} e^{-z} dz = \frac{\theta^{-r}}{v} \sum_{k=0}^{\infty} \frac{(-(\alpha/\theta)^u)^k}{k!} \Gamma\left(\frac{r+uk}{v}\right).$$

Случай $u > v$ рассматривается аналогично. Случай $u = v$ напрямую следует из определения гамма-функции. Лемма доказана.

Рассмотрим случайную величину ξ , имеющую обобщенное распределение Фреше $\text{GFr}(u, n, \alpha)$ с нулевым минимумом, плотность которого имеет вид:

$$f_{\xi}(x) = \frac{u\alpha^{nu} e^{-(\alpha/x)^u}}{(n-1)! x^{nu+1}}, \quad u > 0, \alpha > 0, n \in \mathbb{N}, x > 0.$$

Легко видеть, что имеет место следующее утверждение.

Лемма 2. Для случайной величины ξ , имеющей обобщенное распределение Фреше $\text{GFr}(u, n, \alpha)$, для $z \in \mathbb{R}$ выполняется соотношение:

$$E\xi^z = \frac{\alpha^z \Gamma(n - z/u)}{(n-1)!}, \quad z < nu.$$

Напрямую из лемм 1 и 2 следует справедливость следующего утверждения о вероятностных характеристиках индекса баланса ρ в байесовской модели.

Теорема 1. Пусть n -фактор λ и p -фактор μ имеют обобщенное распределение Фреше $\text{GFr}(u, n, \alpha)$ и $\text{GFr}(v, m, \theta)$ соответственно, причем λ и μ независимы. Тогда при $x > 0$ плотность и моменты индекса баланса $\rho = \lambda/\mu$ имеют вид:

$$f_{\rho}(x) = \begin{cases} \frac{v(\theta/\alpha)^{mv} x^{mv-1}}{(n-1)!(m-1)!} \text{Ge}_{v/u, mv/u+n} \left(-\left(\frac{\theta x}{\alpha}\right)^v \right), & u > v; \\ \frac{u(\alpha/\theta)^{nu} x^{-nu-1}}{(n-1)!(m-1)!} \text{Ge}_{u/v, nu/v+m} \left(-\left(\frac{\alpha}{\theta x}\right)^u \right), & u < v; \end{cases}$$

$$E\rho^z = \frac{(\alpha/\theta)^z \Gamma(n - z/u) \Gamma(m + z/v)}{(n-1)!(m-1)!}, \quad z < nu.$$

При $u = v$ распределение индекса баланса ρ совпадает с обобщенным бета-распределением второго рода [3] с плотностью

$$f_{\rho}(x) = \frac{u\alpha^{nu} \theta^{mu} x^{mu-1}}{B(n, m)(\alpha^u + \theta^u x^u)^{n+m}}, \quad x > 0.$$

Для исследования функции распределения индекса баланса введем в рассмотрение интегральную гамма-экспоненциальную функцию:

$$\text{Gi}(r, s, t; x) = \frac{1}{t\Gamma(s)\Gamma(t)} \int_0^x \text{Ge}_{r, tr+s} \left(-z^{1/t}\right) dz,$$

где $0 \leq r < 1, s > 0, t > 0$.

Теорема 2. Пусть n -фактор λ и p -фактор μ имеют обобщенное распределение Фреше $\text{GFr}(u, n, \alpha)$ и $\text{GFr}(v, m, \theta)$ соответственно, причем λ и μ независимы. Тогда при $x > 0$ функция распределения индекса баланса $\rho = \lambda/\mu$ имеет вид:

$$F_\rho(x) = \begin{cases} \text{Gi}\left(\frac{v}{u}, n, m; \left(\frac{\theta x}{\alpha}\right)^{mv}\right), & u > v; \\ 1 - \text{Gi}\left(\frac{u}{v}, m, n; \left(\frac{\alpha}{\theta x}\right)^{nu}\right), & u < v. \end{cases}$$

Доказательство. Согласно теореме 1 для функции распределения ρ при $u > v$ справедливо для $x > 0$:

$$\begin{aligned} F_\rho(x) &= \frac{v(\theta/\alpha)^{mv}}{(n-1)!(m-1)!} \int_0^x y^{mv-1} \text{Ge}_{v/u, mv/u+n} \left(-\left(\frac{\theta y}{\alpha}\right)^v\right) dy = \\ &= \frac{v(\theta/\alpha)^{mv}}{(n-1)!(m-1)!} \int_0^x y^{mv-1} \sum_{l=0}^{\infty} \frac{(-1)^l (\theta y)^{vl}}{\alpha^{vl} l!} \Gamma\left(\frac{vl}{u} + \frac{mv}{u} + n\right) dy = \\ &= \frac{1}{m(n-1)!(m-1)!} \int_0^{(\theta x/\alpha)^{mv}} \sum_{l=0}^{\infty} \frac{(-z^{1/m})^l}{l!} \Gamma\left(\frac{vl}{u} + \frac{mv}{u} + n\right) dz. \end{aligned}$$

В случае $u < v$ имеем:

$$\begin{aligned} F_\rho(x) &= \frac{u(\alpha/\theta)^{nu}}{(n-1)!(m-1)!} \int_0^x y^{-nu-1} \text{Ge}_{u/v, nu/v+m} \left(-\left(\frac{\alpha}{\theta y}\right)^u\right) dy = \\ &= \frac{u(\alpha/\theta)^{nu}}{(n-1)!(m-1)!} \int_0^x y^{-nu-1} \sum_{l=0}^{\infty} \frac{(-1)^l (\alpha/\theta)^{ul}}{y^{ul} l!} \Gamma\left(\frac{ul}{v} + \frac{nu}{v} + m\right) dy = \\ &= \frac{1}{n(n-1)!(m-1)!} \int_{(\alpha/(\theta x))^{nu}}^{\infty} \sum_{l=0}^{\infty} \frac{(-z^{1/p})^l}{l!} \Gamma\left(\frac{ul}{v} + \frac{nu}{v} + m\right) dz. \end{aligned}$$

Теорема доказана.

Следствие. Из теоремы 2 вытекает, в частности, что

$$\lim_{x \rightarrow \infty} G_i(r, s, t; x) = 1$$

при $0 \leq r < 1, s > 0, t > 0$.

Замечание. В результатах статьи требование $n \in \mathbb{N}$ для параметра обобщенного распределения Фреше является несущественным. Утверждения остаются справедливыми для любого $n \in \mathbb{R}$ при замене $(n - 1)!$ на $\Gamma(n)$.

3 Заключение

Результаты статьи могут применяться для исследования байесовских моделей баланса, когда априорные распределения факторов имеют вид частных случаев обобщенного распределения Фреше, например непосредственно распределения Фреше, обратного распределения Рэля или обратного экспоненциального распределения. Поскольку в доказательствах утверждений требование $n \in \mathbb{N}$ является несущественным, результаты могут быть легко переформулированы для таких априорных распределений, как обратное и масштабированное обратное χ - и χ^2 -распределения, распределение Леви, Пирсона (пятого типа) и других частных случаев обобщенного распределения Фреше. При совпадении параметров формы априорных распределений, как показано, например, в [4, 5], обобщенное бета-распределение (второго рода) индекса баланса принимает вид распределений Бурра, Дагума и Ломакса.

Литература

1. Кудрявцев А. А. Байесовские модели баланса // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 18–27.
2. Кудрявцев А. А., Титова А. И. Гамма-экспоненциальная функция в байесовских моделях массового обслуживания // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 104–108.
3. McDonald J. B., Butler R. J. Some generalized mixture distributions with an application to unemployment duration // Rev. Econ. Statistics, 1987. Vol. 69. No. 2. P. 232–240.
4. Кудрявцев А. А., Палионная С. И., Шоргин В. С. Априорное обратное гамма-распределение в байесовских моделях массового обслуживания // Системы и средства информатики, 2018. Т. 28. Вып. 4. С. 52–58.
5. Кудрявцев А. А., Палионная С. И., Шоргин В. С. Априорные Фреше и масштабированное обратное хи-распределение в байесовских моделях баланса // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 63–67.

Поступила в редакцию 12.03.19

A PRIORI GENERALIZED FRECHET DISTRIBUTION IN BAYESIAN BALANCE MODELS

A. A. Kudryavtsev¹, S. I. Palionnaia¹, and V. S. Shorgin²

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

²Institute 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: This article continues a series of authors’ works in the field of applying the Bayesian approach to queuing, reliability and balance models. In the framework of this approach, when complex aggregates are considered, all parameters affecting the functioning of the system are divided into two classes: contributing and preventing correct functioning of the system. The probabilistic characteristics of the balance index — the ratio of factors that negatively affect the operation of the system to positively influencing factors — are studied under the assumption that the factors are random variables with known *a priori* distributions. In this work, the probabilistic characteristics of the system’s balance index are concerned in the case when both factors have an *a priori* generalized Frechet distribution. The results are presented in terms of a special gamma exponential function.

Keywords: Bayesian approach; generalized Frechet distribution; gamma exponential function; balance models; mixed distributions

DOI: 10.14357/08696527190204

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 17-07-00577).

References

1. Kudryavtsev, A. A. 2018. Bayesovskie modeli balansa [Bayesian balance models]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):18–27.
2. Kudryavtsev, A. A., and A. I. Titova. 2017. Gamma-eksponentsial’naya funktsiya v bayesovskikh modelyakh massovogo obsluzhivaniya [Gamma-exponential function in Bayesian queuing models]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):104–108.
3. McDonald, J. B., and R. J. Butler. 1987. Some generalized mixture distributions with an application to unemployment duration. *Rev. Econ. Statistics* 69(2):232–240.

4. Kudryavtsev, A. A., S. I. Palionnaia, and V. S. Shorgin. 2018. Apriornoe obratnoe gamma-raspredelenie v bayesovskikh modelyakh massovogo obsluzhivaniya [*A priori* inverse gamma distribution in Bayesian queuing models]. *Sistemy i Sredstva Informatiki* — *Systems and Means of Informatics* 28(4):52–58.
5. Kudryavtsev, A. A., S. I. Palionnaia, and V. S. Shorgin. 2019. Apriornye Freshe i masshtabirovanoe obratnoe khi-raspredelenie v bayesovskikh modelyakh balansa [*A priori* Frechet and scaled inverse chi distribution in Bayesian balance models]. *Informatika i ee Primeneniya* — *Inform. Appl.* 13(1):63–67.

Received March 12, 2019

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

Palionnaia Sofia I. (b. 1995) — student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; sofiaalionnaya@gmail.com

Shorgin Vsevolod S. (b. 1978) — Candidate of Science (PhD) in technology, 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; VShorgin@ipiran.ru

ИНТЕЛЛЕКТУАЛИЗАЦИЯ ОПЕРАТИВНО-ДИСПЕТЧЕРСКОГО УПРАВЛЕНИЯ СЛОЖНОЙ РЕГИОНАЛЬНОЙ ДИНАМИЧЕСКОЙ ЭЛЕКТРОЭНЕРГЕТИЧЕСКОЙ СИСТЕМОЙ*

А. В. Колесников¹, С. В. Листопад², Ф. Г. Майтаков³

Аннотация: Проблемы оперативно-диспетчерского управления (ОДУ) региональными электросетями характеризуются неоднородностью, частичной наблюдаемостью объекта управления, а также его динамическим характером, обуславливающим взаимозависимость выполняемых действий и сложность исправления ошибочных решений. Применяемые способы отображения не релевантны ментальному образу, на который оператор ориентируется в своей деятельности, мешают ему работать и выполнять действия в уме, способствуют ошибкам интерпретации данных. Поэтому стало актуально создание информационно-коммуникационных технологий компьютерной имитации когнитивных образований для усиления человеческого интеллекта в оперативной работе путем дополнения естественных способностей оператора к работе с оперативно-технологической информацией программно-аппаратными средствами (механизмами), расширяющими мыслительные процессы человека.

Ключевые слова: когнитивные гибридные интеллектуальные системы; оперативно-диспетчерское управление; региональная динамическая электроэнергетическая система

DOI: 10.14357/08696527190205

1 Введение

Управление в динамической среде связано с неоднородностью и распределенностью в пространстве решаемых задачи или разрабатываемых систем. Принимаемые решения взаимосвязаны и накладывают ограничения друг на друга, не позволяя исправить ошибочное решение в будущем, так как даже небольшая начальная ошибка стремительно нарастает с течением времени. Вопросы обработки неоднородной информации за счет интеграции знаний экспертов исследованы в рамках гибридных интеллектуальных систем (ГиИС), синтезирующих над гетерогенным модельным полем метод решения сложной задачи.

*Работа выполнена при поддержке РФФИ (проект 19-07-00208а).

¹Балтийский федеральный университет им. И. Канта; Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, avkolesnikov@yandex.ru

²Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, ser-list-post@yandex.ru

³Балтийский федеральный университет им. И. Канта, maitakov@mail.ru

Традиционные ГиИС не задействуют правосторонние, визуально-образные рассуждения оператора, осложняя оперативную работу с непосредственно не воспринимаемым объектом, заставляя его мыслить логически, не обеспечивая интуитивное принятие решений. Преодолеть этот недостаток предлагается путем получения новых знаний об оперативно-технологическом, человеко-машинном управлении сложными динамическими системами, разработкой визуального событийно-ситуационного субъектно-ориентированного подхода оперативно-технологического управления сложными динамическими объектами и его реализации в новом классе систем искусственного интеллекта — когнитивных ГиИС (КГиИС).

2 Электроэнергетическое производство и процессы. Сложная система. Большая система. Техноценоз

Производство электроэнергии резко отличается от других отраслей экономики по целому ряду причин [1]:

- генерация электроэнергии, ее распределение и преобразование идут одновременно, между мощностями оперативно поддерживается точный баланс;
- скорость переходных процессов — 10^{-4} – 10^{-3} с; короткие замыкания, включения и отключения, качания, нарушения устойчивости длятся от долей до нескольких секунд; сложность и высокая скорость процессов и действий создают большие психофизиологические нагрузки на оперативный персонал и органы диспетчерского управления (ДУ);
- тесное взаимодействие с промышленностью, связью и транспортом;
- актуальность надежности и резерва мощности.

Применительно к производству, технике и технологиям в электроэнергетической системе (ЭС) в теории систем и системном анализе используется широкий спектр понятий: от «сложной системы» и «большой системы» до «техноценоза».

Согласно определению, данному Н. П. Бусленко в 1973 г., систему можно считать сложной, если она состоит из большого числа взаимосвязанных и взаимодействующих элементов и способна выполнять сложную функцию [2].

Термин «большая система» («система большого масштаба») вошел в научный лексикон после появления книги Р. Х. Гуда и Р. Э. Макола [3], применявших его к системам оборудования и аппаратуры. В работах У. Р. Эшби, М. Месаровича, Д. Мако и И. Такахары, Д. А. Поспелова, Ю. И. Черняка, А. А. Денисова, Д. Н. Колесникова, А. А. Вавилова, С. В. Емельянова, В. В. Калашникова, Б. В. Немчинова, Б. В. Фомина, А. И. Берга, Б. В. Бирюкова, Е. С. Геллерта, Г. Н. Поварова и Л. А. Растригина это понятие широко использовалось применительно к социальным, биологическим, экологическим, транспортным, промышленным, военным системам. Электроэнергетике как большой системе присущи:

- целостность и общая цель;
- «свобода воли»;
- стохастичность, нерегулярность внешних возмущений, невозможность точного предсказания нагрузки;
- составной характер, иерархичность;
- большая размерность по элементам, функциям и стоимости;
- сложность поведения и разнообразие взаимосвязей переменных, динамичность, нестационарность;
- уникальность;
- многокритериальность, отсутствие формализованной цели;
- слабая структурированность теоретических и фактических знаний о системе;
- неполнота описания;
- неоднородность подсистем и элементов системы;
- многофункциональность элементов;
- высокая степень автоматизации технологических процессов и ДУ;
- конкуренция внутри системы и во внешней среде.

Понятия «техника», «техническая реальность» и «техноценоз» занимают центральное место в исследованиях Б. И. Кудрина и его учеников, в частности В. И. Гнатьюка [4], понимающих техноценоз как системный объект, ограниченный в пространстве и времени, отдельные элементы которого, с одной стороны, достаточно независимы и не связаны между собой жестко (механически, электрически, гидравлически и т. д.), а с другой — объединены слабыми связями: единой системой управления, снабжения, эксплуатации, целью. Примером может служить региональный электротехнический комплекс. В профессиональном языке оперативно-диспетчерской работы в ЭС чаще применяют термин «сложная система».

3 Оперативно-диспетчерское управление региональной электроэнергетической системой

Суммарная мощность электростанций ЭС Калининградской области (КО) на декабрь 2018 г. — 1596,1 МВт. В ЭС КО входят 94 линии электропередачи (ЛЭП) класса напряжения 60–330 кВ общей протяженностью 2468,68 км, 75 трансформаторных подстанций и пять распределительных устройств электростанций. Электроэнергетическая система КО входит в состав Объединенной энергетической системы (ОЭС) Северо-Запада и состоит из двух энергорайонов — Западного и Восточного. Действуют электрические сети напряжением 15, 10 и 6 кВ. Оперативно-диспетчерское управление функционированием ЭС КО осуществляется диспетчерскими центрами (ДЦ) Филиала АО «СО ЕЭС» «Региональное ДУ энергосистемы КО» и АО «Янтарьэнерго». Вышестоящий

ДЦ по отношению к Балтийскому региональному ДУ — Филиал АО «СО ЕЭС» Объединенное диспетчерское управление Северо-Запада, смежный ДЦ — ДЦ Литовской Республики «LITGRIDAB».

Акционерные общества «Янтарьэнерго» и «Монитор Электрик» с 2017 г. разрабатывают автоматизированную систему оперативно-технологического управления (АСОТУ) «ОЛИМП» (рис. 1 и 2).

Автоматизированная система оперативно-технологического управления «ОЛИМП» включает подсистемы:

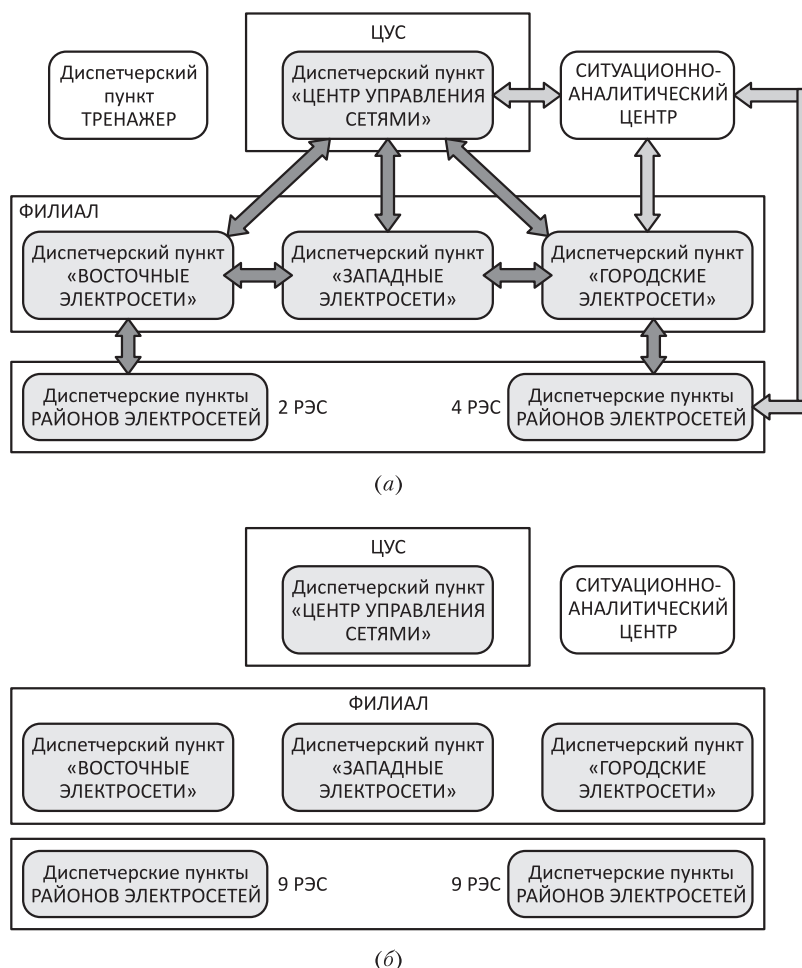


Рис. 1 Информационные модели ОДУ ЭС КО (РЭС — район электрических сетей): (а) модель 2016 г.; (б) целевая модель

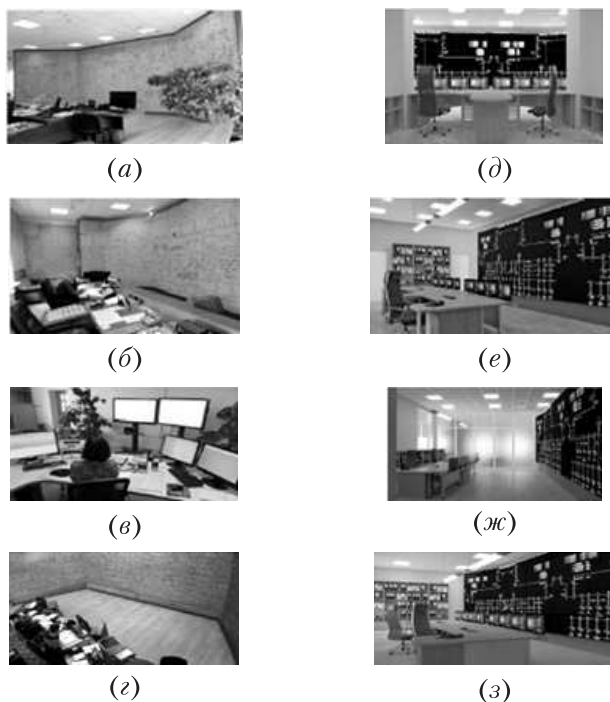


Рис. 2 Технические решения: (а)–(г) диспетчерского пункта «Центр управления сетями» (ЦУС), диспетчерских пунктов «Восточные электросети», «Западные электросети» и «Городские электросети» соответственно для модели 2016 г.; (д)–(з) аналогично для целевой модели АСОТУ «Олимп»

- сбора, обработки и отображения информации SCADA (Supervisory Control And Data Acquisition);
- управления общей информационной моделью;
- дистанционного управления элементами сети;
- выявления узких мест; автоматического восстановления электроснабжения;
- онлайн-анализа режимов и определения места повреждения;
- оценки состояния сети; оптимизации потокораспределения,

а также

- тренажер диспетчера;
- веб-интерфейс для мобильных бригад;
- элементы искусственного интеллекта.

На LCD-панели видеостены отображается схема сети.

4 Техника и технологии оперативно-диспетчерского управления в электроэнергетической системе

Тенденция технического оснащения ЭС — применение гибких (управляемых) систем электропередачи (*англ.* FACTS — Flexible Alternative Current Transmission Systems) и асинхронизированных синхронных машин (АСМ) [5] для повышения надежности и эффективности. В [6] как приоритетные для ЭС выделяют:

- интеллектуальные сети (*англ.* Smart Grids);
- электрический транзит ультравысокого напряжения;
- уменьшение затрат на сооружение и потерь в ЛЭП;
- распределенную генерацию;
- развитие силовой электроники и FACTS.

Применение интеллектуальных электронных устройств (ИЭУ) и FACTS улучшает оценку состояния ЭС. Перспективные области ИЭУ — автоматизированные системы управления технологическим процессом подстанций, распределенные системы (РС) автоматики (*англ.* Wide Area Control System — WACS), РС релейной защиты (*англ.* Wide Area Protection System — WAPS), комбинированные РС мониторинга, защиты и автоматики (*англ.* Wide Area Monitoring, Protection, and Control System — WAMPACS) и активно-адаптивные сети (ААС).

Из изученных по открытым источникам 41 организации-разработчика программно-технических комплексов (ПТК) перечислим базисные решения для интеллектуализации ОДУ в электроэнергетике.

Акционерное общество «РТСофт» (rtsoft.ru): «∀Платформа»; универсальная вычислительная платформа «Интегро»; ПТК Smart-SPRECON; ПТК «СМАРТ-КП2»; ПТК SMART-WAMS; программное обеспечение (ПО) PhasoGraph; ПТК систем сбора и передачи информации (ССПИ) противоаварийной автоматики (ПА) Smart-ССПИ ПА; ПТК сбора и передачи неоперативной технологической информации (ССПТИ) Smart-ССПТИ; ПТК «Гармоника-ПС»; центральная приемо-передающая станция Smart-FEP; ПТК Smart-ЦУС; информационно-управляющая система AMIGO.

Компания «АльтероПауэр» (alteropower.ru): РС хранения технологических данных CloudHistorian; шлюз-концентратор векторных измерений Advanced Phasor Data Concentrator сложных многоуровневых РС мониторинга.

Акционерное общество «Монитор Электрик» (monitel.ru): информационно-технологическая платформа «СК-11»; электронный оперативный журнал ёЖ-2; коммуникационный процессор СК-Прогу; режимный тренажер диспетчера «Финист»; программный комплекс (ПК) оперативного планирования MODES-Centre; система управления оповещениями Event Notifier (eNote).

Акционерное общество «НТЦ ФСК ЕАС» (ntc-power.ru): ПК интеллектуального адаптивного управления ЭС КУЭС; ПК РЕТРЕН 2.0; ПК визуализации оперативной информации КАСКАД — НТ 2.0; ПК ДИСКУЭС диспетчеризации умных сетей; интеллектуальный комплекс SMARTCOM анализа нештатных ситуаций; многоагентная система управления напряжением и реактивной мощностью «МАСУ НРМ» в ЭС.

Открытое акционерное общество «Институт энергосистем» (enersys.ru): облачный сервис Enersys Load Forecast (ELF); ПТК «Монитор»; ПТК «Synchrograf»; ПО «ReVizor»; ПТК «СДПМ».

Открытое акционерное общество «Созвездие–Сервис» (sozvezdie-service.ru): информационно-аналитическая система (ИАС) «Созвездие Навигатор 3.0»; ПК «Генератор 5.0».

СИМЕНС (SIEMENS) (smart-grid.siemens.ru): интеллектуальная система Spectrum Power; АСР 1703 Platforms, SICAM АК 3; интеллектуальные модули ввода-вывода SICAM TM, SICAM CMIC, SICAM A8000.

ABB (new.abb.com): Micro SCADA PRO SYS 600; SYS 600C; DMS 600; SDM 600; Relion; РСМ 600; IET 600; ITT 600; интеграция ABB Ability и IBM Watson.

Компания «ПиЭлСи Технолоджи» (plctech.ru): ПТК ТОРАZ IEC DAS; платформа ТОРАZ Graphics сложных информационных систем; среда визуализации баз данных ТОРАZ DBMS.

Закрытое акционерное общество «НПФ Прорыв» (proryv.com): ПК Телескоп+.

Группа компаний «Прософт-Системы» (prosoftsystems.ru): ПТК ARIS; ПО Redkit SCADA; система обмена технологической информацией «АССО»; ПК «Энергосфера».

Акционерное общество «Атомик Софт» (automiq.ru): Альфа-платформа; Alpha Developer Studio.

Инженерный центр ЭНТЕЛС (ersr.ru): платформы ЭНТЕК, EnLogic; ПТК SCADA ЭНТЕК TEL, комплектное распределительное устройство Etalon.

Ниже перечислены интеграторы по щитовому оборудованию и видеостенам для ОДУ ЭС: ПОИСК (poisk-company.ru), «Делайт 2000» (delight2000.com), ЮГ-СИСТЕМА ПЛЮС (yugsys.nt-rt.ru), MultiRu (multiru.ru), «СофтСистема» (video-panels.ru), «Форсайт» (<https://forsite-company.ru>); ООО «ТехПроМедиа» (displite.ru), ВГУП «Уральский электромеханический завод» (uemz.ru), Центр проекционных технологий «Викинг» (viking.ru).

5 Интеллектуализация в электроэнергетике

Понятие «интеллектуализация» применительно к ЭС по смыслу не совпадает с модернизацией и означает применение [7] новых:

- принципов и систем управления режимным состоянием ЕЭС;
- технологий и принципов управления перетоками мощности по ЛЭП FACTS;

- средств измерения параметров режима;
- эффективных алгоритмов управления ЭС в различных режимах;
- информационных технологий и ССПИ, представления и использования информации в устройствах управления;
- принципов формирования информации для визуализации переходных процессов в реальном времени.

Интеллектуальная ЭС [8] должна располагать системой управления с «мыслительными способностями» человеческого разума для функций ОДУ, но с более высоким быстродействием, т. е. «интеллектуальной» системой управления. В настоящее время ЭС — пассивный объект, так как перетоки мощности по ЛЭП определяются задаваемым в диспетчерском графике режимом нагрузки для электростанций. Только напряжение в узлах нагрузки поддерживается автоматикой регулирования под нагрузкой и регулируемых источников реактивной мощности на подстанциях. «Оживляет», т. е. изменяет и приспособливает ЭС к работе в условиях физической реальности, активная интеллектуальная деятельность ОДУ. Накопленные энергетиками знания актуальны для создания искусственного разума — оператора в структуре ДУ режимами ЭС и их объединений.

Тридцатого июля 2014 г. стартовала разработка национального проекта «Интеллектуальная энергетическая система России» — интеллектуальной ЭС с ААС нового поколения [8] — системы, в которой все субъекты электроэнергетического рынка принимают активное участие в процессах передачи и распределения электроэнергии. Дорожная карта интеллектуальной энергетики России, обнародованная в 2015 г. [9], предусматривает:

- (1) разработку систем управления и интеллектуального мониторинга технологических комплексов генерации;
- (2) разработку технологий систем распределенного управления режимами работы и технологий интеллектуального мониторинга ЭС;
- (3) создание систем управления и принятия решений с использованием мульти-агентных систем и нейронных сетей.

6 Интеллектуализация принятия оперативно-диспетчерских решений: когнитивные гибридные интеллектуальные системы

Оперативная работа с образом непосредственно не воспринимаемого объекта со сложной динамикой, проецирование на этот образ информации, поступающей извне, — трудная задача персонала ОДУ, а применяемые способы отображения не релевантны ментальному образу деятельности оператора, мешают работать, выполнять действия в уме, способствуют ошибкам интерпретации данных. Поэтому стало актуально создание информационно-коммуникационных технологий

компьютерной имитации когнитивных образований для усиления человеческого интеллекта в оперативной работе.

Один из подходов к преодолению данного противоречия лежит в плоскости нано-, био-, инфо-, когно- и социоконвергенции (НБИКС, *англ.* NBICS)— объединения и синергетического усиления достижений НБИКС-технологий полным слиянием в единую область знания, объединяющую методы решения интеллектуальных задач, в принципе самостоятельно не решаемых человеком-оператором. Такой информационно-целостный продукт был назван КГиИС [10].

Когнитивные гибридные интеллектуальные системы. Их идея высказана в 2009–2012 гг. основателем инженерии образов И. Б. Фоминых, формализовавшего ГиИС темпоральными моделями и методами инженерии образов [11].

В 2015 г. В. А. Данишевский и А. В. Колесников предложили архитектурную схему ГиИС с когнитивной компонентой для мониторинга и управления качеством электроэнергии [12]. В 2017 г. А. Н. Аверкин и С. А. Ярушев создали модели для поддержки принятия решений в динамических ситуациях со слабой структурированностью, основанной на гибридной системе, интегрирующей нечеткую иерархическую модель оценивания и нечеткую когнитивную модель ситуации [13]; С. В. Ульянов и др. разработали гибридные когнитивные нечеткие системы управления автономным роботом [14]; Л. В. Массель и В. Р. Кузьмин предложили семиотический подход к интеграции методов ситуационного управления, семантического моделирования и когнитивной графики применительно к ЭС [15].

Когнитивное моделирование. В 2002–2003 гг. А. А. Кулинич под руководством О. П. Кузнецова разработал методологию моделирования с применением когнитивного подхода при решении сложных проблем и системы «Компас» и «Канва» [16].

Когнитивная графика. В начале 1990-х гг. А. А. Зенкин и Д. А. Поспелов открыли эру когнитивной компьютерной графики языком цветомузыкальных «пифограмм». Под руководством В. М. Хачумова в ИПС РАН «интеллектуализируют» управление сложными динамическими объектами, разрабатывают методы и системы образного анализа данных, когнитивной графикой и отображения динамической информации. В 2012–2018 гг. А. А. Башлыков и А. П. Еремеев создали интегрированную интеллектуальную систему реального времени «СПРИНТ-РВ» с когнитивной составляющей «глобус Башлыкова».

Искусственные когнитивные системы решают проблемы в условиях реального мира, включая распознавание образов, понимание смысла и контекстуального значения информации, мышление, самостоятельное усвоение знаний из различных источников, обучение, субъективную оценку информации. В центре обучаемых наноматериалов и мемристорных систем Тюменского государственного университета реализуется проект «Искусственные когнитивные системы». Комплексную проблему мониторинга состояния железнодорожной инфраструктуры когнитивными сенсорами исследовали М. Н. Святкина и В. Б. Тарасов [17].

Когнитивные вычисления — работа с огромными массивами неструктурированной информации методами искусственного интеллекта на базе когнитивных вычислений. Пример — суперкомпьютер IBM Watson.

7 Заключение

Рассмотрены особенности ОДУ ЭС, характеризующиеся понятиями «сложная система», «большая система» и «техноценоз». Рассмотрены существующие на сегодняшний день подходы к интеллектуализации таких систем, в частности интеллектуальная ЭС с активно-адаптивной сетью. Предложена концепция КГиИС для интеллектуализации принятия оперативно-диспетчерских решений в ЭС. Когнитивные ГиИС позволят учитывать динамический характер сложных задач и синтезировать над гетерогенным визуальным полем релевантный на момент решения задачи интегрированный метод. Они также позволят оператору «увидеть» приближенное решение сложной задачи, которое впоследствии может быть обосновано и уточнено методами логико-математических рассуждений.

Литература

1. Басова Т. Ф., Е. И. Борисов, В. В. Бологова и др. Экономика и управление энергетическими предприятиями. — М.: Академия, 2004. 432 с.
2. Бусленко Н. П., Калашников В. В., Коваленко Н. Н. Лекции по теории больших систем. — М.: Советское радио, 1973. 439 с.
3. Good H., Machol R. System engineering. An introduction to the design of large-scale systems. — New York, NY, USA: McGraw-Hill Book Co., 1957. 551 p.
4. Гнатюк В. И. Философские основания техноценологического подхода. — Калининград: Техноценоз, 2014. 284 с.
5. Поспелова Т. Г. Эффекты применения FACTS и АСМ в повышении эффективности региональных и национальных энергосистем // Conference (International) “Energy of Moldova-2012. Regional Aspects of Development.” — Chisinau, 2012. P. 83–91.
6. Тухватуллин М. М., Ивекеев В. С., Ложкин И. А., Урманова Ф. Ф. Анализ современных устройств FACTS, используемых для повышения эффективности функционирования электроэнергетических систем России // Электротехнические системы и комплексы, 2015. № 3(28). С. 41–46.
7. Зеленохат Н. И. Интеллектуализация ЕЭС России: инновационные предложения. — М.: МЭИ, 2013. 192 с.
8. Интеллектуальная энергетика. <https://www.eriras.ru/data/788/rus>.
9. Волкова И. О. Дорожная карта интеллектуальной энергетики России в системе национальных проектов, программ и инициатив: поиск синергии. — М.: ВШЭ, 2016. <https://ipcrem.hse.ru/data/2016/04/25/1130065769/Доклад.И.О.%20Волкова.pdf>.
10. Колесников А. В. Визуальный событийно-ситуационный подход информационной подготовки решений и оперативно-технологического управления сложными динамическими объектами // Открытые семантические технологии проектирования

- интеллектуальных систем: Мат-лы IX Междунар. научн.-технич. конф. — Минск: БГУИР, 2019. С. 139–144.
11. *Фоминых И. Б.* Гибридные интеллектуальные системы на основе темпоральных логических моделей и инженерии образов // Интегрированные модели и мягкие вычисления в искусственном интеллекте: Сб. научных тр. V Междунар. научн.-практич. конф. — М.: Физматлит, 2009. Т. 1. С. 87–96.
 12. *Данишевский В. И., Колесников А. В.* Гибридная интеллектуальная система с когнитивной компонентой для диспетчерского управления региональной энергосистемой // Интегрированные модели и мягкие вычисления в искусственном интеллекте: Сб. научных тр. VIII Междунар. научн.-практич. конф. — М.: Физматлит, 2015. Т. 1. С. 391–399.
 13. *Аверкин А. Н., Ярушев С. Я.* Когнитивные гибридные системы поддержки принятия решений и прогнозирования // Программные продукты и системы, 2017. № 4. С. 632–642.
 14. *Ульянов С. В., Решетников А. Г., Мамаева А. А.* Гибридные когнитивные нечеткие системы управления автономным роботом на основе нейроинтерфейса и технологии мягких вычислений // Программные продукты и системы, 2017. № 3. С. 420–424.
 15. *Массель Л. В., Кузьмин В. Р.* Ситуационное исчисление как развитие семиотического подхода к построению интеллектуальной системы поддержки принятия решений // Вестник Новосибирского государственного ун-та. Сер. Информационные технологии, 2017. Т. 15. № 4. С. 43–52.
 16. *Кулинич А. А.* Когнитивная система поддержки принятия решений «КАНВА» // Программные продукты и системы, 2002. № 3. С. 25–28.
 17. *Святкина М. Н., Тарасов В. Б.* Логико-алгебраические методы построения когнитивных сенсоров // Открытые семантические технологии проектирования интеллектуальных систем: Мат-лы VI Междунар. научн.-технич. конф. — Минск: БГУИР, 2016. С. 331–348.

Поступила в редакцию 05.03.19

INTELLECTUALIZATION OF OPERATIONAL DISPATCH CONTROL OF COMPLEX REGIONAL DYNAMIC POWER SYSTEM

A. V. Kolesnikov^{1,2}, S. V. Listopad², and F. G. Maitakov¹

¹Immanuel Kant Baltic Federal University, 14 A. Nevskogo Str., Kaliningrad 236041, Russian Federation

²Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation

Abstract: The problems of operational dispatch management of regional power systems are characterized by heterogeneity, partial observability of the control object, as well as its dynamic nature, which determine the interdependence

of actions performed and the difficulty of correcting erroneous decisions. The applied mapping methods are not relevant to the mental image that the operator is guided by in his(her) activities, prevent him(her) from working and performing actions in his(her) mind, and contribute to errors in the interpretation of data. As a result, the creation of information and communication technologies of computer imitation of cognitive formations is relevant to enhance human intelligence in operational work, by supplementing the operator's natural abilities to work with operational technological information, software, and hardware (mechanisms) that expand human thought processes.

Keywords: cognitive hybrid intelligent systems; operational dispatch management; regional dynamic power system

DOI: 10.14357/08696527190205

Acknowledgments

The reported study was partially funded by the Russian Foundation for Basic Research according to the research project No. 19-07-00208A.

References

1. Basova, T. F., E. I. Borisov, V. V. Bologova, *et al.* 2004. *Ekonomika i upravlenie energeticheskimi predpriyatiyami* [Economics and management of energy enterprises]. Moscow: Akademiya Pubs. 432 p.
2. Buslenko, N. P., V. V. Kalashnikov, and N. N. Kovalenko. 1973. *Lektsii po teorii bol'shikh system* [Lectures on the theory of large systems]. Moscow: Soviet radio. 439 p.
3. Good, H., and R. Machol. 1957. *System engineering. An introduction to the design of large-scale systems*. New York, NY: McGraw-Hill Book Co. 551 p.
4. Gnatyuk, V. I. 2014. *Filosofskie osnovaniya tekhnotsenologicheskogo podkhoda* [Philosophical foundations of the technocenological approach]. Kaliningrad: Tekhocenoz CIC Pubs. 284 p.
5. Pospelova, T. G. 2012. Effekty primeneniya FACTS i ASM v povyshenii effektivnosti regional'nykh i natsional'nykh energosistem [Effects of using FACTS and ASM in increasing the efficiency of regional and national energy systems]. *Conference (International) "Energy of Moldova-2012. Regional Aspects of Development."* Chisinau. 83–91.
6. Tukhvatullin, M. M., V. S. Ivekeev, I. A. Lozhkin, and F. F. Urmanova. 2015. Analiz sovremennykh ustroystv FACTS, ispol'zuyemykh dlya povysheniya effektivnosti funktsionirovaniya elektroenergeticheskikh sistem Rossii [Analysis of modern FACTS devices, used for increasing the functioning efficiency of Russian electric power systems]. *Elektrotekhnicheskie sistemy i komplekсы* [Electrotechnical Systems and Complexes] 3(28):41–46.
7. Zelenokhat, N. I. 2013. *Intellektualizatsiya EES Rossii: innovatsionnye predlozheniya* [Intellectualization of the UES of Russia: Innovative proposals]. Moscow: MPEI Pubs. 192 p.

8. Intellektual'naya energetika [Intelligent power]. Available at: <https://www.eriras.ru/data/788/rus> (accessed February 26, 2019).
9. Volkova, I. O. 2016. Dorozhnaya karta intellektual'noy energetiki Rossii v sisteme natsional'nykh proektov, programm i initsiativ: poisk sinergii [The Road Map of intellectual energy in Russia in the system of national projects, programs, and initiatives: The search for synergy]. Available at: https://ipcrem.hse.ru/data/2016/04/25/1130065769/Доклад_И.О.%20Волкова.pdf (accessed February 26, 2019).
10. Kolesnikov, A. V. 2019. Vizual'nyy sobytiyno-situatsionnyy podkhod informatsionnoy podgotovki resheniy i operativno-tekhnologicheskogo upravleniya slozhnymi dinami-cheskimi ob"ektami [Visual event-situational approach of information preparation of solutions and operational-technological management of complex dynamic objects]. *Open Semantic Technologies for Intelligent Systems Conference (International) Proceedings*. Minsk: BGUIR. 139–144.
11. Fominykh, I. B. 2009. Gibridnye intellektual'nye sistemy na osnove temporal'nykh logicheskikh modeley i inzhenerii obrazov [Hybrid intelligent systems based on temporal logic models and image engineering]. *5th Scientific and Practical Conference (International) "Integrated Models and Soft Computing in Artificial Intelligence" Proceedings*. Moscow: Fizmatlit. 1:87–96.
12. Danishevskiy, V. I., and A. V. Kolesnikov. 2015. Gibridnaya intellektual'naya sistema s kognitivnoy komponentoy dlya dispetcherskogo upravleniya regional'noy energosistemoy [A hybrid intellectual system with a cognitive component for the dispatching control of the regional energy system]. *8th Scientific and Practical Conference (International) "Integrated Models and Soft Computing in Artificial Intelligence" Proceedings*. Moscow: Fizmatlit. 1:391–399.
13. Averkin, A. N., and S. Ya. Yarushev. 2017. Kognitivnye gibridnye sistemy podderzhki prinyatiya resheniy i prognozirovaniya [Cognitive hybrid decision support and forecasting systems]. *Software & Systems* 4:632–642.
14. Ul'yanov, S. V., A. G. Reshetnikov, and A. A. Mamayeva. 2017. Gibridnye kognitivnye nechetkie sistemy upravleniya avtonomnym robotom na osnove neyrointerfeysa i tekhnologii myagkikh vychisleniy [Hybrid cognitive fuzzy control systems of an autonomous robot based on a neural interface and soft computing technology]. *Software & Systems* 3:420–424.
15. Massel, L. V., and V. R. Kuzmin. 2017. Situatsionnoe ischislenie kak razvitie semioticheskogo podkhoda k postroeniyu intellektual'noy sistemy podderzhki prinyatiya resheniy [Situation calculus as development of semiotic approach to constructing intelligent decision-making support system]. *Vestnik NSU. Ser. Information Technologies* 15(4):43–52.
16. Kulinich, A. A. 2002. Kognitivnaya sistema podderzhki prinyatiya resheniy "KANVA" [Cognitive decision support system "KANVA"]. *Software & Systems* 3:25–28.
17. Svyatkina, M. N., and V. B. Tarasov. 2016. Logiko-algebraicheskie metody postroeniya kognitivnykh sensorov [Logic-algebraic methods for constructing cognitive sensors]. *6th Conference (International) "Open Semantic Technologies for Intelligent Systems" Proceedings*. Minsk: BGUIR. 331–348.

Received March 3, 2019

Contributors

Kolesnikov Alexander V. (b. 1948) — Doctor of Science in technology; professor, Institute of Physical and Mathematical Sciences and Information Technology, Immanuel Kant Baltic Federal University, 14 A. Nevskogo Str., Kaliningrad 236041, Russian Federation; leading scientist, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation, avkolesnikov@yandex.ru

Listopad Sergey V. (b. 1984) — Candidate of Science (PhD) in technology; senior scientist, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; ser-list-post@yandex.ru

Maitakov Fedor G. (b. 1973) — postgraduate student, Institute of Physical and Mathematical Sciences and Information Technology, Immanuel Kant Baltic Federal University, 14 A. Nevskogo Str., Kaliningrad 236041, Russian Federation, maitakov@mail.ru

ИНФОРМАЦИОННЫЕ ТРАНСФОРМАЦИИ В ПРОЦЕССАХ ОБРАБОТКИ ГЕОДАННЫХ

Д. А. Никишин¹

Аннотация: Анализируются процессы, характерные для сферы обработки геоданных. Эти процессы объединены в три специфические группы, в структуре которых выделены элементарные информационные трансформации и рассмотрены их особенности. На этой основе предложена двухуровневая типология основных видов информационных трансформаций, которая направлена на обобщение существующих методов конкретно-абстрактных трансформаций и исследование свойства обратимости этих отображений в информационных системах.

Ключевые слова: геоданные; информационные трансформации; генерализация пространственных данных; цифровая картография; геоинформатика

DOI: 10.14357/08696527190206

1 Введение

Целью данной работы ставился анализ структуры процессов информационных отображений, характерных для сферы обработки геоданных, и построение типологии информационных трансформаций, имеющих место в этих процессах. Актуальность этой задачи определяется необходимостью развития типологии информационных трансформаций [1] в парадигме полиадического компьютеринга [2].

Представленные в работе процессы информационного отображения объединены в три специфические группы. На основе анализа структуры этих групп выделены элементарные преобразования, которые могут быть охарактеризованы как информационные трансформации [3]. В контексте этого анализа рассмотрены особенности таких трансформаций и предложена двухуровневая типология основных видов информационных трансформаций. Предложенная типология позволит обобщить существующие методы конкретно-абстрактных трансформаций и исследовать свойство обратимости этих преобразований в информационных системах [4].

2 Основные модели геоданных и виды отображений

Ранее в работе [5] были представлены два основных вида пространственных данных: *геоизображение*, представляющее собой композитный образ предметной

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, dmnikishin@mail.ru

области (местности), и геоданные в виде набора аппроксимирующих моделей, соответствующих отдельным геообъектам, выделенным на местности. К геоизображениям помимо фотографических изображений (например, космический снимок) могут быть отнесены такие продукты, как цифровые матрицы рельефа, матрицы различных качеств и т. п. Геоданные предназначены для описания местности в виде структурированной совокупности целенаправленно выделяемых отдельных геообъектов с их типичными признаками и характерными особенностями. Этот класс может быть отождествлен с картой в векторном представлении [6], сюда же можно с определенной долей условности отнести такие специфические данные, как облака точек, получаемые посредством лазерной локации. Наглядное сопоставление этих моделей представлено на рис. 1, где в левой нижней части показано аэрокосмическое изображение, а в правой верхней — результат визуализации геоданных.

Можно выделить следующие виды информационных отображений, которые могут иметь место между предметной областью и различными формами пред-



Рис. 1 Комбинированная картина аэрокосмического геоизображения и визуализированных геоданных

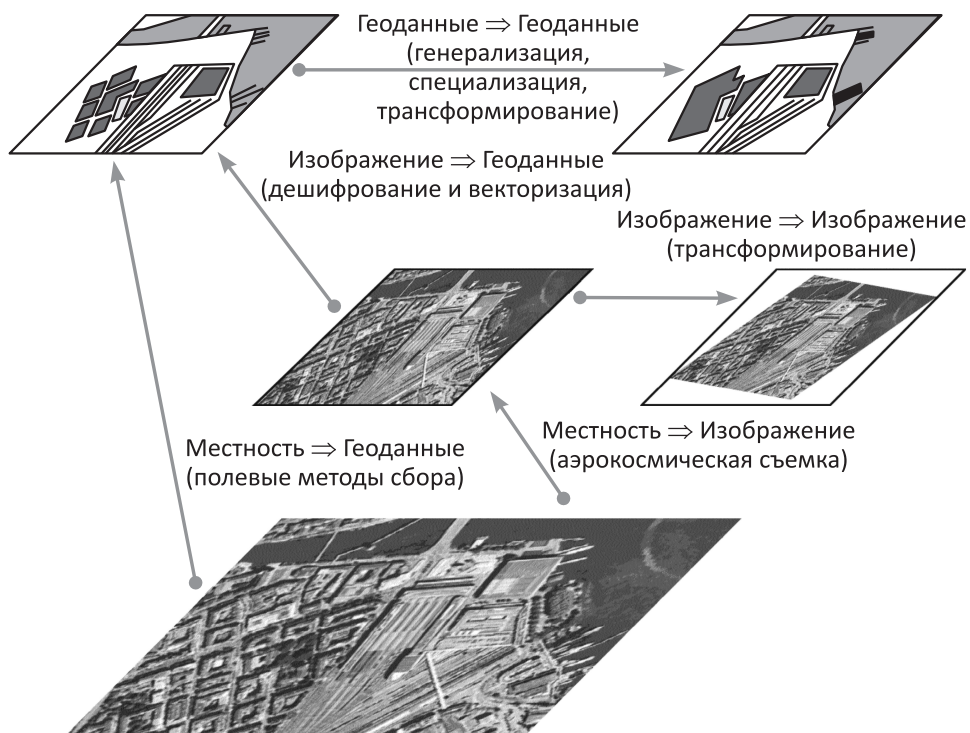


Рис. 2 Структура отображений между предметной областью и различными формами представления геоданных

ставления геоданных: «местность ⇒ изображение», «изображение ⇒ изображение», «изображение ⇒ геоданные», «местность ⇒ геоданные» и «геоданные ⇒ геоданные». Структура этих отображений показана на рис. 2, а сами они подробно рассматриваются далее в составе трех отдельных групп отображений.

3 Группа отображений с целью получения изображения

Первая группа включает в себя два информационных отображения, которые имеют целью получение изображения: «местность ⇒ изображение» (съемка местности) и «изображение ⇒ изображение» (трансформирование изображений).

Отображение «**местность ⇒ изображение**» осуществляется посредством аэрокосмической или наземной съемки. Это отображение обусловливается используемыми способами проецирования и формирования аэрокосмического изображения.

Проецирование изображения осуществляется различными способами и заключается в построении пространственного образа местности в разных проекциях (центральная, ортогональная и т. п.).

Формирование изображения заключается в регистрации поля значений интенсивности электромагнитного излучения приемником излучения съемочной системы [7]. При этом происходит преобразование (дискретизация) произвольных значений энергии излучения в значения, допустимые формализованной моделью приемника излучения [8].

Основные свойства получаемого в результате изображения — его геометрическое, радиометрическое, спектральное разрешение [7] — в значительной мере характеризуются явлением *оптической генерализации* (англ. optical generalization, или remote sensing generalization) [6] — геометрического и радиометрического (в том числе спектрального) обобщения изображения предметной области на получаемых снимках. Это обобщение проявляется как на этапе проецирования, так и на этапе формирования изображения и определяется комплексом технических факторов (метод выполнения съемки, ее высота, спектральный диапазон съемки и др.), а также условиями съемки (характер местности, атмосферные условия, освещенность и т. п.).

Стоит отметить, что независимо от того, является ли получаемое изображение аналоговым или цифровым, сущность его *формирования* принципиально не меняется и представляет собой величину яркости принимаемого излучения в пределах площади каждого элемента приемника излучения (пикселя или фотоэмульсионного зерна), а также дифференциацию этой яркости в соответствии с типом и спектральной чувствительностью используемого приемника излучения.

Отображение «**изображение** \Rightarrow **изображение**» (трансформирование изображения) может выполняться как с целью изменения проекции исходного изображения¹, так и с целью его генерализации. Данное отображение также обуславливается способами *проецирования* (в данном случае перепроецирования исходного изображения) [9] и *формирования* нового изображения.

В аналоговой форме это отображение по своему характеру представляет собой частный случай оптической генерализации. Цифровые методы обработки позволяют осуществлять более сложные методы проецирования, такие как аффинные и проективные преобразования, ортотрансформирование (на основе ЦМР), трансформирование типа «резиновый лист» и т. п. Цифровое представление также дает возможность применять различные по сложности методы формирования результирующего изображения².

¹ К данному виду информационных трансформаций можно отнести как частный случай и процесс сканирования (оцифровывания) аналогового изображения. Оно заключается в проецировании диффузной картины фотографического или полиграфического изображения на регулярную матрицу раstra.

² Формирование в данном случае заключается в вычислении значения трансформированного пикселя, которое выполняется на основе анализа соответствующей ему области исходного изображения (например, прореживание пиксельных строк и столбцов изображения методом «бли-

Нужно отметить, что независимо от конкретной цели (перепроецирования или генерализации) трансформирование в общем случае сопряжено с повторным проецированием, влекущим за собой определенные различия локальных масштабов отдельных частей проецируемого изображения. Вследствие этого на этапе формирования результирующего изображения (при его дискретизации) возникают изменения детальности изображения по полю изображения: на одних участках они будут носить характер генерализации, а на других будет проявляться обратный эффект, но не приводящий к увеличению детальности («разгон» изображения). Кроме того, присутствует эффект отображения дискретных элементов исходного изображения (его пикселей/зерна) на аналогичные элементы результирующего изображения. Все это обуславливает тот факт, что любое трансформирование в общем случае ухудшает информативные и изобразительные свойства, а в некоторых случаях может приводить даже к утрате существенных деталей изображения.

Подводя итог, можно отметить, что в структуре обоих рассмотренных отображений можно выделить схожий набор информационных трансформаций: проецирование изображения и его формирование.

Важно отметить, что в обоих случаях присутствует явление, носящее характер оптической генерализации, которое неразрывно связано с методами проецирования и формирования изображения.

Также надо отметить, что, в отличие от групп отображений, рассматриваемых далее, в данной группе информационные трансформации проецирования, формирования и генерализации выступают в качестве факторов формирования изображения и их выделение в отдельные процессы нецелесообразно.

4 Группа отображений с целью получения геоданных

Вторая группа включает в себя информационные отображения местности или ее изображения с целью получения структурированной модели в виде отдельных геообъектов (геоданных), представленных в рамках определенной геоонтологии. К таким отображениям относятся: «изображение \Rightarrow геоданные» (векторизация геоизображения) и «местность \Rightarrow геоданные» (наземная съемка местности).

Отображения данной группы также включают ряд новых информационных трансформаций, свойственных только этой группе:

- *декомпозиция* — выделение из предметной области (местности или ее изображения) отдельных элементов: объектов местности или элементов изображения (на цифровом изображении — в виде групп пикселей);
- *классификация* — анализ выделенных на предыдущем этапе элементов (геообъектов) и их отнесение к тому или иному классу (формальной типовой модели) в составе некоторой геоонтологии [8];

жайшего соседа» либо билинейная интерполяция), а также может затрагивать окрестности этой области (бикубическая интерполяция).

- *генерализация* — целенаправленный отбор из полученного набора классифицированных геобъектов тех, которые представляют интерес в контексте определенной цели создания (сбора) этих геоданных. Эта цель определяет тот или иной уровень генерализации и/или вариант тематической ориентации геоданных.

Здесь нужно отметить отличия такого вида генерализации от оптической генерализации. Во-первых, генерализация в данном случае носит качественно иной характер и представляет собой уже не механическое обобщение, а избирательное выделение из исходного множества как общего, определяющего, так и специфичного, но важного в контексте цели создания геоданных [10]. Во-вторых, этот вид генерализации выступает уже в качестве самостоятельной, явно эксплицируемой информационной трансформации. Поскольку генерализация здесь представляет собой особый вид информационной трансформации, ее предлагается именовать *первичной генерализацией*.

В структуре отображений данной группы также присутствуют информационные трансформации проецирования и формирования, но, в отличие от отображений рассмотренной выше группы, здесь эти трансформации носят характер явно эксплицируемых процессов.

Процесс формирования здесь также имеет существенные отличия по сравнению с формированием изображения. Формирование метрики заключается в аппроксимации групп пикселей или описании элементов местности соответствующими векторными моделями. Семантические свойства геобъектов формализуются посредством используемой геоонтологии с соответствующей дифференциацией значений этих свойств.

Также нужно отметить важную особенность всех информационных трансформаций, входящих в состав отображений данной группы, — в отличие от предыдущей группы, здесь они носят характер не методологий, а самостоятельных процессов. При этом, в отличие от методов формирования изображения, здесь полноценно задействован семантический аспект геоданных.

В совокупности все эти информационные трансформации могут рассматриваться как комплексный процесс сбора геоданных. Применительно к аэрокосмическим изображениям эти процессы традиционно объединяются в виде комплексного процесса дешифрирования и векторизации [11].

Отображение **«изображение ⇒ геоданные»** использует в качестве исходных данных имеющееся (полученное ранее) изображение местности и на его основе посредством процессов дешифрирования и векторизации формирует выходной набор геоданных.

В практике получения геоданных по изображениям процессы дешифрирования и векторизации геобъектов становятся одними из основных «камней преткновения» на пути к автоматизации этого способа получения данных о местности. Обусловлено это сложностью формализованного описания перечисленных выше информационных трансформаций даже для частных случаев, например одина-

ковых условий съемки, характера местности и перечня интересующих объектов, что делает их во многом субъективными.

Отображение «местность \Rightarrow геоданные» по своей задаче аналогично предыдущему и представляет собой непосредственный сбор геоданных на местности. Этот процесс можно рассматривать как аналог процессов дешифрирования и векторизации, и, как и эти процессы, процесс сбора данных также носит субъективный характер.

В связи с высокой стоимостью и низкой производительностью полевых работ по сравнению с аэрокосмическими методами такой вид отображения применяется ограниченно, в качестве дополняющего, но тем не менее полностью отказаться от него невозможно, поскольку никакой полноценный набор геоданных не может быть получен чисто дистанционными методами и, как правило, требуется полевое дообследование местности.

5 Группа отображений с целью изменения свойств существующего набора геоданных

Третья группа включает в себя только одно отображение «геоданные \Rightarrow геоданные», которое заключается в целенаправленном качественном и/или количественном преобразовании существующего набора геоданных в новый набор. Одной из его целей может быть обобщение набора геоданных, другой — изменение содержания набора данных, например его тематической направленности, т. е. его специализация. Обе цели независимы друг от друга и могут преследоваться одновременно.

Информационное отображение с целью обобщения представляет собой так называемое «узкое» понимание генерализации, которое фигурирует в соответствующей литературе, — *картографическую генерализацию* (англ. cartographic generalization) [7, 12–14]. Этот вид генерализации отличен от рассмотренных выше трансформаций оптической и первичной генерализации.

Частной целью, которая не вызывает количественных/качественных изменений в наборе данных, является трансформирование геоданных, заключающееся в преобразовании их метрики в другую проекцию или перекодировании их семантических характеристик в альтернативную классификационную схему.

Можно выделить следующие специфичные для данной группы информационные трансформации, которые применимы в контексте той или иной цели отображения: анализ объектов геоданных по отдельности и в их совокупности, их фильтрацию и, возможно, их обобщение (при наличии цели генерализации) [10]. Эти трансформации в совокупности с возможным перепроецированием и переструктурированием набора данных составляют процесс, который в зависимости от цели можно обозначить как специализацию и/или генерализацию набора геоданных.

Рассматривая данное отображение как группу, можно выделить следующие характерные особенности:

- (1) только это отображение предоставляет потенциальную возможность обеспечить его обратимость. Это становится возможным благодаря формализованному представлению данных в виде геообъектов, получаемому в результате отображений предыдущей группы;
- (2) этот вид отображения характеризуется большей методологической однозначностью, хотя и здесь высока доля субъективизма, определяемая свойствами конкретного набора данных и целями его преобразования. Тем не менее это отображение в наибольшей степени возможно автоматизировать, например в виде технологии автоматизированной генерализации масштабного ряда электронных карт [15, 16].

6 Заключение

Подводя итог сказанному, можно выделить следующие группы информационных отображений, характерных для сферы обработки геоданных.

1. Отображения с целью получения изображения включают схожий набор информационных трансформаций: проецирование изображения и его формирование. Информационные отображения этой группы сопровождаются явлением оптической генерализации, которая не может быть явно эксплицирована в структуре этих отображений.
2. Отображение с целью получения описания местности в виде набора геоданных (сбор геоданных) дополнительно включает следующие информационные трансформации: декомпозицию, классификацию и первичную генерализацию. В отличие от предыдущей группы отображений, здесь информационные трансформации носят характер уже не методологий, а самостоятельных процессов.
3. Отображение с целью обобщения, содержательной переработки или трансформирования существующего набора геоданных включает специфические информационные трансформации анализа отдельных информационных объектов геоданных и их совокупностей, их фильтрации и, возможно, их обобщения (при наличии цели генерализации). Частным случаем такого информационного отображения с целью обобщения выступает картографическая генерализация. Важной особенностью этого вида отображения, благодаря формализованному представлению геоданных, является потенциальная возможность его обратимости.

Относительно генерализации можно отметить, что она выступает в качестве попутной информационной трансформации во всех группах отображений, однако в каждой группе она проявляется в качестве специфической трансформации (оптическая, первичная и картографическая генерализация).

Представленная двухуровневая структура (уровень групп отображений и уровень информационных трансформаций) может быть охарактеризована как типология основных видов информационных трансформаций в процессах обработки геоданных.

Литература

1. Зацман И. М. Методология обратимой генерализации в контексте классификации информационных трансформаций // Системы и средства информатики, 2018. Т. 28. № 2. С. 128–144.
2. Rosenbloom P. S. On computing: The fourth great scientific domain. — Cambridge, MA, USA: MIT Press, 2013. 307 p.
3. Гончаров А. А., Зацман И. М. Информационные трансформации параллельных текстов в задачах извлечения знаний // Системы и средства информатики, 2019. Т. 29. № 1. С. 180–193.
4. Зацман И. М., Мамонова О. С., Щурова А. Ю. Обратимость и альтернативность генерализации моделей перевода коннекторов в параллельных текстах // Системы и средства информатики, 2017. Т. 27. № 2. С. 125–142.
5. Никишин Д. А. Сопоставление особенностей представления геоданных в картографии и геоинформатике // Системы и средства информатики, 2018. Т. 28. № 2. С. 60–74.
6. Баранов Ю. Б., Берлянт А. М., Капралов Е. Г., Кошкарев А. В., Серанинас Б. Б., Филиппов Ю. А. Геоинформатика: Толковый словарь основных терминов. — М.: ГИС-Ассоциация, 1999. 204 с.
7. Чандра А. М., Гош С. К. Дистанционное зондирование и географические информационные системы. — М.: Техносфера, 2008. 312 с.
8. Шовенгердт Р. А. Дистанционное зондирование. Модели и методы обработки изображений. — М.: Техносфера, 2010. 560 с.
9. Chang Kang-Tsung. Introduction to geographic information systems. — New York, NY, USA: McGraw-Hill Education, 2018. 461 p.
10. Никишин Д. А. Процессы генерализации в аналоговой и цифровой картографии // Системы и средства информатики, 2018. Т. 28. № 3. С. 204–216.
11. Книжников Ю. Ф., Кравцова В. И., Тутубалина О. В. Аэрокосмические методы географических исследований. — М.: Академия, 2004. 336 с.
12. Комков А. М., Николаев С. А., Шилов Н. И. Составление и редактирование карт. — М.: ВИА, 1958. 248 с.
13. Taillandier P., Duchêne C., Drogoul A. Knowledge revision in systems based on an informed tree search strategy: Application to cartographic generalisation // 5th Conference (International) on Soft Computing as Transdisciplinary Science and Technology Proceedings. — New York, NY, USA: ACM, 2008. P. 273–278.
14. Журкин И. Г., Шайтура С. В. Геоинформационные системы. — М.: КУДИЦ-Пресс, 2009. 272 с.
15. Технология автоматизированной генерализации ГИС «Панорама». <https://gisinfo.ru/projects/19.htm>.

16. Ware M. J., Wilson I. D., Ware J. A., Jones C. B. A tabu search approach to automated map generalisation // 10th ACM Symposium (International) on Advances in Geographic Information Systems Proceedings. — New York, NY, USA: ACM, 2002. P. 101–106.

Поступила в редакцию 13.03.19

INFORMATION TRANSFORMATION IN THE PROCESSING OF GEODATA

D. A. Nikishin

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The paper analyzes the processes typical for the sphere of geodata processing. These processes are grouped into three specific groups, in the structure of which elementary information transformations are identified and their features are considered. On this basis, a two-level typology of the main types of information transformations is proposed, which is aimed at generalizing the existing methods of concrete-abstract transformations and studying the reversibility properties of these maps in information systems.

Keywords: geodata; informational transformations; generalization of spatial data; digital cartography; geoinformatics

DOI: 10.14357/08696527190206

References

1. Zatsman, I. M. 2018. Metodologiya obratimoy generalizatsii v kontekste klassifikatsii informatsionnykh transformatsiy [Methodology of reversible generalization in the context of classification of information transformations]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(2):128–144.
2. Rosenbloom, P. S. 2013. *On computing: The fourth great scientific domain*. Cambridge, MA: MIT Press. 307 p.
3. Goncharov, A. A., and I. M. Zatsman. 2019. Informatsionnye transformatsii parallel'nykh tekstov v zadachakh izvlecheniya znaniy [Information transformations of parallel texts in problems of knowledge extraction]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):180–193.
4. Zatsman, I. M., O. S. Mamonova, and A. Yu. Shchurova. 2017. Obratimost' i al'ternativnost' generalizatsii modeley perevoda konnektorov v parallel'nykh tekstakh [Reversibility and alternativeness of generalization of connectives translations models in parallel texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(2):125–142.

5. Nikishin, D. A. 2018. Sopostavlenie osobennostey predstavleniya geodannykh v kartografii i geoinformatike [Comparison of characteristics of the representation of geodata in cartography and geoinformatics]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(2):60–74.
6. Baranov, Yu. B., A. M. Berlyant, E. G. Kapralov, A. V. Koshkarev, B. B. Serapinas, and Yu. A. Filippov. 1999. *Geoinformatik: Tolkovyy slovar' osnovnykh terminov* [Geoinformatics: Explanatory dictionary of basic terms]. Moscow: GIS-Association. 204 p.
7. Chandra, A. M., and S. K. Gosh. 2008. *Distsionnoe zondirovanie i geograficheskie informatsionnye sistemy* [Remote sensing and geographical information system]. Moscow: Tekhnosfera. 312 p.
8. Shovengerdt, R. A. 2010. *Distsionnoe zondirovanie. Modeli i metody obrabotki izobrazheniy* [Remote sensing. Models and methods for image processing.]. Moscow: Tekhnosfera. 560 p.
9. Chang, Kang-Tsung. 2018. *Introduction to geographic information systems*. New York, NY: McGraw-Hill Education. 461 p.
10. Nikishin, D. A. 2018. Protsessy generalizatsii v analogovoy i tsifrovoy kartografii [A generalization processes in analog and digital cartography]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(3):204–216.
11. Knizhnikov, Yu. F., V. I. Kravtsova, and O. V. Tutubalina. 2004. *Aerokosmicheskie metody geograficheskikh issledovaniy* [Aerospace methods of geographical researches]. Moscow: Akademiya. 336 p.
12. Komkov, A. M., S. A. Nikolaev, and N. I. Shilov. 1958. *Sostavlenie i redaktirovanie kart* [Drafting and editing maps]. Moscow: VIA. 248 p.
13. Taillandier, P., C. Duchêne, and A. Drogoul. 2008. Knowledge revision in systems based on an informed tree search strategy: Application to cartographic generalisation. *5th Conference (International) on Soft Computing as Transdisciplinary Science and Technology Proceedings*. New York, NY: ACM. 273–278.
14. Zhurkin, I. G., and S. V. Shaytura. 2009. *Geoinformatsionnye sistemy* [Geoinformation systems]. Moscow: KUDITS-Press. 272 p.
15. Tekhnologiya avtomatizirovannoy generalizatsii GIS “Panorama” [Technology of automated generalization of GIS “Panorama”]. Available at: <https://gisinfo.ru/projects/19.htm> (accessed March 13, 2019).
16. Ware, M. J., I. D. Wilson, J. A. Ware, and C. B. Jones. 2002. A tabu search approach to automated map generalisation. *10th ACM Symposium (International) on Advances in Geographic Information Systems Proceedings*. New York, NY: ACM. 101–106.

Received March 13, 2019

Contributor

Nikishin Dmitry A. (b. 1976)— Candidate of Science (PhD) 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; dmnikishin@mail.ru

МЕТОД РАСЧЕТА КАРТЫ ГЛУБИН НА ОСНОВЕ МЯГКИХ ОПЕРАТОРОВ*

М. В. Бобырь¹, А. Е. Архипов², Н. А. Милостная³

Аннотация: Рассмотрен нечеткий метод построения карты глубин по стереоизображениям на основе композиции алгоритма SAD (Sum of Absolute Differences) и нечетко-логического вывода. Особенность данного метода состоит в использовании мягких арифметических операторов при нечеткой импликации. Точность метода определялась на основе количественной оценки показателя RMSE (root mean square error). Наилучшим признавался тот мягкий оператор, у которого показатель RMSE стремится к минимуму. Предложенный метод показал, что при использовании мягких операторов нахождения минимума точность алгоритма SAD увеличивается на 20%. Данный вывод подтвержден результатами имитационного моделирования, представленного в статье.

Ключевые слова: стереозрение; карта глубин; мягкие вычисления; SAD; нечеткая логика

DOI: 10.14357/08696527190207

1 Введение

При разработке систем навигации мобильными роботами возникает задача, связанная с определением местоположения подвижных и/или неподвижных препятствий, возникающих на пути его следования [1, 2]. Одним из эффективных средств определения взаимного расположения объектов служит анализ данных о глубине сцены и препятствий, расположенных на пути следования робота. Для построения подобных сцен используются различные методы, например получение данных от сканирующих лазерных дальномеров или использование систем стереозрения [3]. При разработке систем стереозрения необходимо решить ряд проблем, связанных с нахождением объектов, располагающихся на стереоизображениях [4]. Под картой глубин понимается черно-белое изображение с градациями серого цвета, на котором яркость отдельных предметов указывает пропорциональное расстояние до них. Основными алгоритмами построения карт глубин являются алгоритмы SGM (Semi-Global Matching) [5] и SAD. Для расчета карты глубины с помощью SGM-алгоритма требуется несколько проходов, в то

*Работа выполнена при поддержке гранта Президента РФ (МД-707.2017.8) и Госзадания (соглашение № 2.3440.2017/4.6).

¹Юго-Западный государственный университет, Курск, fregat_mn@rambler.ru

²Юго-Западный государственный университет, Курск, alex.76_09@mail.ru

³Юго-Западный государственный университет, Курск, nat_mil@mail.ru

время как второй алгоритм использует только один проход вдоль изображения, что снижает время, необходимое для расчета величины смещения d (disparity). Следует заметить, что 1 кадр стереоинформации в формате HD (720 p — разрешение 1280×720 пикселей) должен обрабатываться за $1/30$ с т. е. за 33 мс. Следовательно, на обработку 1 пикселя требуется около 50 нс. Поэтому время принятия решений выступает одним из главных факторов, которые необходимо учитывать при разработке стереосистем управления мобильными роботами. Одним из путей повышения точности модели стереозрения является использование нечеткой логики [6]. Нечеткие модели имеют ряд системных ошибок, снижающих их точность [7]. Для устранения этих ошибок рекомендуется использовать при импликации мягкие арифметические операторы.

2 Постановка задачи

Поставим задачу повышения точности системы стереозрения для построения карт глубин на основе оценки показателя RMSE:

$$\text{RMSE} = \sqrt{\frac{q}{wh} \sum_{i=1}^{h-1} \sum_{j=1}^{w-1} (I'_{i,j} - I_{i,j})^2} \rightarrow \min, \quad (1)$$

где w и h — ширина и высота изображения; $I'_{i,j}$ — уровень яркости на эталонном изображении; $I_{i,j}$ — уровень яркости на выходном изображении.

Исследуемые мягкие операторы задаются формулами, представленными в табл. 1, где a и b — операнды, для которых осуществляется поиск минимального значения; δ — коэффициент мягкости поиска минимума ($\delta = 0,05$); n — число операндов, для которых осуществляется поиск минимума.

Наиболее точным признается тот метод, при реализации которого RMSE минимален.

Таблица 1 Непараметризованные t -нормы

Название оператора	Формула для импликации
MIN	$I = \min(a, b)$ (2)
PROD	$I = ab$ (3)
MEAN	$I = \frac{a + b}{2}$ (4)
Soft-min I	$\frac{a + b + \delta^2 - \sqrt{(a - b)^2 + \delta^2}}{2}$ (5)
Soft-min II	$I \frac{\sum_{i=1}^n a_i e^{-ka_i}}{\sum_{i=1}^n e^{-ka_i}}$, где $k = -100$ (6)

3 Нечеткий метод построения карты глубин по стереоизображениям

Для построения карты глубин используется композиция алгоритма SAD с мягкой нечетко-логической системой принятия управляющих решений. На основе алгоритма SAD необходимо для каждого пикселя рассчитать величину смещения d . Расчет выполняется в несколько действий.

Шаг 1. Сначала для каждого пикселя с координатами (x_l, y_l) левого стереоизображения рассчитывается разность уровней интенсивности в окне (например, размером 3×4 пикселя) левого и правого стереоизображений. После этого на правом изображении окно смещается на один пиксель и снова рассчитывается разность уровней интенсивностей левого и правого стереоизображений. Сдвиг пикселей на правом изображении и расчет разности уровней интенсивности продолжается до порогового значения $d = 0 \dots d_{\max}$:

$$\text{SAD}(x, y, s) = \sum_{\substack{m=-b; \\ n=-b}}^{\substack{m=b; \\ n=b}} R \sum_{\substack{m=-b; \\ n=-b}}^{\substack{m=b; \\ n=b}} G \sum_{\substack{m=-b; \\ n=-b}}^{\substack{m=b; \\ n=b}} B |I_l(x + m, y + n) - I_r(x + m + d, y + n)|, \quad (7)$$

где I_l и I_r — интенсивность цвета на левом и правом стереоизображениях; x и y — координаты на стереоизображениях; b — размер окна, определяемый по формуле $2b + 1$, при $b = 1$ размер окна 3×3 пикселя; R, G, B — компоненты трехканальной цветовой модели (если используется YUV-модель, преобразования осуществляются по формуле $Y = 0,299R + 0,587G + 0,114B$).

После расчета абсолютной разности уровней интенсивности для массива значений от 0 до порогового значения d_{\max} в алгоритме SAD находится минимальное значение $D(x, y)_{\text{SAD}}$. Это значение служит величиной смещения, которое формирует карту глубин для пикселя с координатами (x, y) :

$$D(x, y)_{\text{SAD}} = \underset{d \in [0 \dots d_{\max}]}{\operatorname{argmin}} \text{SAD}(x, y, d).$$

В графическом виде алгоритм SAD представлен на рис. 1.

Одна из основных проблем при реализации модели SAD заключается в поиске и соотнесении одинаковых точек на левом и правом стереоизображениях. Вследствие наличия на изображениях одинаковых участков большой площади, например фона, при сопоставлении появляются артефакты белого цвета. Введение трех входных переменных позволяет более точно определять соответствие одинаковых точек на левом и правом стереоизображениях. Входными переменными нечеткого метода построения карты глубин (НМПКГ) по стереоизображениям служат три уровня интенсивности, рассчитанные по формуле (7): в исходной точке с координатами (x, y) — $\text{SAD}(x, y, d)$; в точке, смещенной относительно

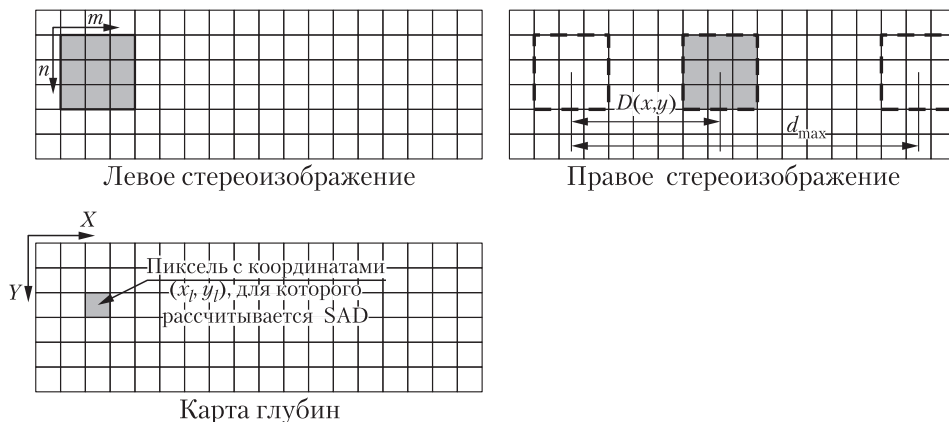


Рис. 1 Графическая интерпретация расчета алгоритма SAD

исходной на 3 пикселя влево, — $SAD(x - 3, y, d)$; в точке, смещенной относительно исходной на 3 пикселя вправо, — $SAD(x + 3, y, d)$. В качестве выходной величины используется величина смещения $D(x, y)$, вычисляемая по формуле (8) (см. с. 76). При этом для каждого пикселя карты глубин нечеткий вывод рассчитывается для массива значений в диапазоне от 0 до d_{\max} . Величине смещения $D(x, y)$ присваивается минимальная величина из этого диапазона значений.

Шаг 2. Расчет степеней входных функций принадлежности (ФП). С целью повышения быстродействия НМПКГ и реализации его на базе программируемых логических интегральных схем (ПЛИС) для управления мобильными роботами рекомендуется использовать тот вид ФП, для расчета степеней истинности которого понадобится меньшее число арифметических операций. Например, для расчета треугольных ФП необходимо выполнение только операций вычитания и деления. Операция вычитания в ПЛИС реализуется просто. Для реализации операции деления рекомендуется подбирать разность между метками ФП, кратную степени числа 2. Тогда операция деления заменяется на операцию сдвига вправо. Входные ФП в НМПКГ представлены на рис. 2, а. С целью повышения быстродействия предлагаемого метода все входные переменные имеют одинаковые ФП.

Расчет степеней функций принадлежности входных переменных осуществляется по следующим формулам:

$$A_{1i} = \begin{cases} \frac{5000 - SAD(x, y, d)}{5000}, & \text{если } SAD(x, y, d) < 5000; \\ 0 & \text{иначе;} \end{cases}$$

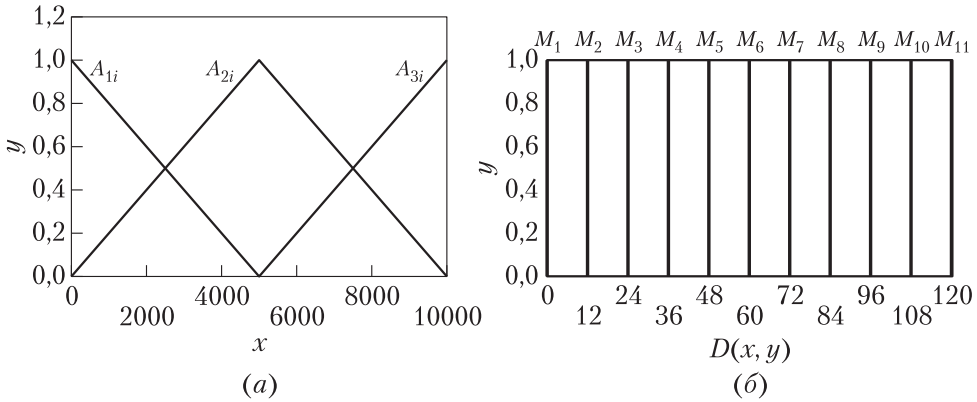


Рис. 2 Функции принадлежности входных (а) и выходной (б) переменных

$$A_{2i} = \begin{cases} \frac{\text{SAD}(x, y, d)}{5000}, & \text{если } \text{SAD}(x, y, d) < 5000; \\ \frac{10\,000 - \text{SAD}(x, y, d)}{5000}, & \text{иначе;} \end{cases}$$

$$A_{3i} = \begin{cases} \frac{\text{SAD}(x, y, d) - 5000}{5000}, & \text{если } \text{SAD}(x, y, d) > 5000; \\ 0, & \text{иначе,} \end{cases}$$

где i — номер входной переменной ($i = 1, 2, 3$).

Выбор такого числа входных переменных объясняется требованием повышения быстродействия обработки нечеткой информации в ПЛИС для построения карты глубин в режиме реального времени. Так как с ростом числа входных переменных увеличивается число нечетких правил, а следовательно и число арифметических операций, используемых в нечетком выводе, то при разработке нечеткой модели необходимо задавать разумные границы уровня ее сложности. Следует заметить, что число нечетких правил растет экспоненциально в зависимости от увеличения числа входных переменных.

Шаг 3. Расчет степени выполнения предпосылок нечетких правил. Число нечетких правил, использующихся для расчета величины смещения, составляет 27. В общем виде нечеткое правило имеет вид [8, 9]:

$$R_i : \text{IF } \text{SAD}(x, y, d) \text{ Is } A_{1i}, \text{SAD}(x - 3, y, d) \text{ Is } A_{2i}, \text{SAD}(x + 3, y, d) \text{ Is } A_{3i} \\ \text{THEN } D(x, y) \text{ Is } Y_k,$$

где Y_k — степень выполнения заключений нечетких правил ($k = 1 \dots 11$).

Нечеткие правила сведены в табл. 2.

Таблица 2 Степени выполнения предпосылок нечетких правил

Предпосылки нечеткого вывода		
$N_1 = \Theta(A_{11}, A_{12}, A_{13})$	$N_{10} = \Theta(A_{21}, A_{12}, A_{13})$	$N_{19} = \Theta(A_{31}, A_{12}, A_{13})$
$N_2 = \Theta(A_{11}, A_{12}, A_{23})$	$N_{11} = \Theta(A_{21}, A_{12}, A_{23})$	$N_{20} = \Theta(A_{31}, A_{12}, A_{23})$
$N_3 = \Theta(A_{11}, A_{12}, A_{33})$	$N_{12} = \Theta(A_{21}, A_{12}, A_{33})$	$N_{21} = \Theta(A_{31}, A_{12}, A_{33})$
$N_4 = \Theta(A_{11}, A_{22}, A_{13})$	$N_{13} = \Theta(A_{21}, A_{22}, A_{13})$	$N_{22} = \Theta(A_{31}, A_{22}, A_{13})$
$N_5 = \Theta(A_{11}, A_{22}, A_{23})$	$N_{14} = \Theta(A_{21}, A_{22}, A_{23})$	$N_{23} = \Theta(A_{31}, A_{22}, A_{23})$
$N_6 = \Theta(A_{11}, A_{22}, A_{33})$	$N_{15} = \Theta(A_{21}, A_{22}, A_{33})$	$N_{24} = \Theta(A_{31}, A_{22}, A_{33})$
$N_7 = \Theta(A_{11}, A_{32}, A_{13})$	$N_{16} = \Theta(A_{21}, A_{32}, A_{13})$	$N_{25} = \Theta(A_{31}, A_{32}, A_{13})$
$N_8 = \Theta(A_{11}, A_{32}, A_{23})$	$N_{17} = \Theta(A_{21}, A_{32}, A_{23})$	$N_{26} = \Theta(A_{31}, A_{32}, A_{23})$
$N_9 = \Theta(A_{11}, A_{32}, A_{33})$	$N_{18} = \Theta(A_{21}, A_{32}, A_{33})$	$N_{27} = \Theta(A_{31}, A_{32}, A_{33})$

Примечание. Знак Θ означает нечетко-логическую операцию нахождения жесткого или мягкого минимума по формулам (2)–(6).

Таблица 3 Степени выполнения заключений нечетких правил

Заключения нечеткого вывода		
$Y_1 = N_{27}$	$Y_2 = \max(N_{26}, N_{24})$	$Y_3 = \max(N_{25}, N_{23}, N_{21})$
$Y_4 = \max(N_{22}, N_{20}, N_{18})$	$Y_5 = \max(N_{19}, N_{17}, N_{15})$	$Y_6 = \max(N_{16}, N_{14}, N_{15})$
$Y_7 = \max(N_{13}, N_{11}, N_9)$	$Y_8 = \max(N_{10}, N_8, N_6)$	$Y_9 = \max(N_7, N_5, N_3)$
$Y_{10} = \max(N_4, N_2)$	$Y_{11} = N_1$	

Шаг 4. Расчет степени выполнения заключений нечетких правил. Число ФП выходной переменной составляет 11 (рис. 2, б). С учетом этого степени выполнения заключений будут рассчитываться по формулам, сведенным в табл. 3.

Шаг 5. Дефаззификация. Осуществляется для массива значений в диапазоне от исходного пикселя на карте глубин до d_{\max} по формуле упрощенного нечеткого вывода:

$$D(x, y)_{d=1}^{d_{\max}} \text{-defuz} = \frac{\sum_{k=1}^{11} Y_k M_k}{\sum_{k=1}^{11} Y_k}.$$

Использование упрощенного нечеткого вывода обоснованно тем, что для его реализации требуется в 4 раза меньше времени, чем при использовании метода центра тяжести.

Шаг 6. Расчет величины смещения на карте глубин. Величина смещения определяется минимальным значением в массиве значений величин дефаззификации в диапазоне от 0 до d_{\max} и рассчитывается по формуле:

$$D(x, y) = \underset{d \in [0 \dots d_{\max}]}{\operatorname{argmin}} \left(D(x, y)_{d=1}^{d_{\max}} \text{-defuz} \right). \quad (8)$$

4 Моделирование процесса построения карты глубин

Для моделирования процесса построения карты глубин использовались стереоизображения размером 450×375 пикселей [10] (рис. 3), позволяющие оценить визуально и количественно эффективность НМПКГ. Нечеткая система вывода была сформирована на основе ФП, представленных на рис. 2. Максимальная величина глубины $d_{\max} = 64$. Для преобразования в выходное значение рассчитанные значения величин смещения по формуле (8) умножались на 4, чтобы

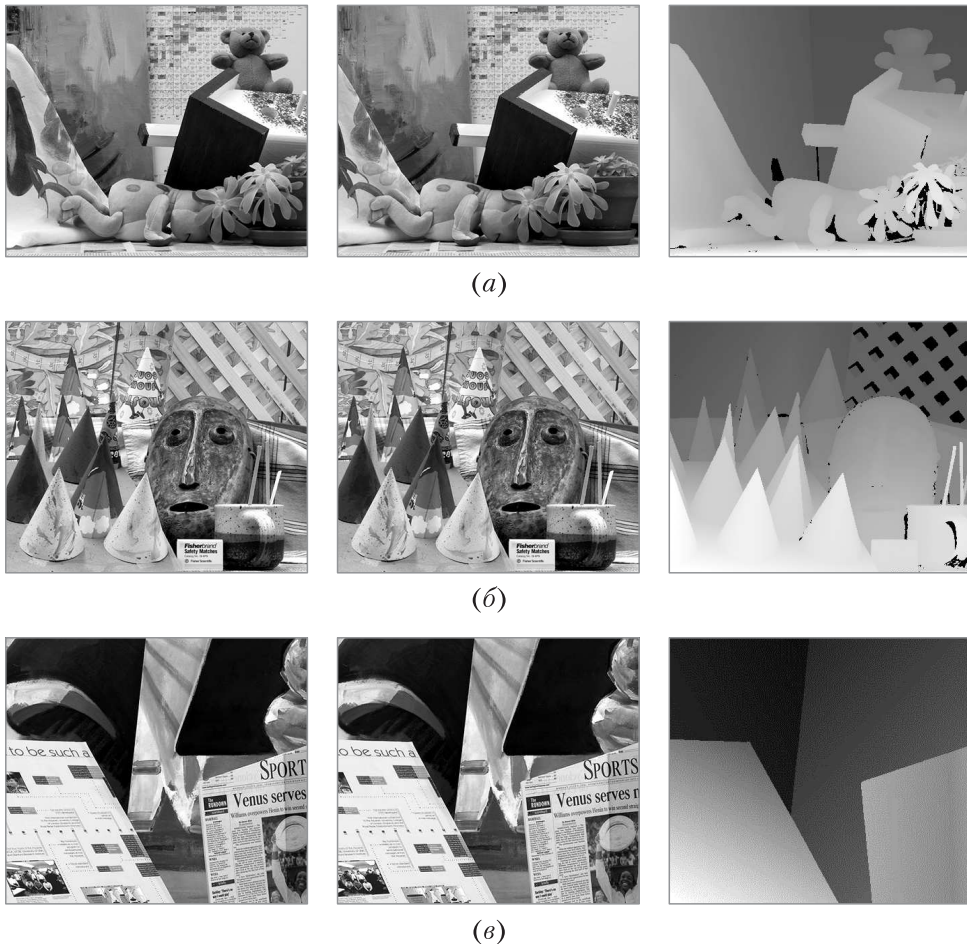


Рис. 3 Стереоизображения: (а) Teddy; (б) Cones; (в) Venus. Представлены левое и правое стереоизображения и эталонная карта глубин

Таблица 4 Значение показателя RMSE

Картинка	Метод	MIN	PROD	MEAN	Soft-MIN I	Soft-MIN II
Teddy	SAD	47,77				
	[11]	44,91	46,74	40,98	41,09	38,42
	НМПКГ	38,65	38,86	38,09	37,76	38,32
Cones	SAD	51,79				
	[11]	50,81	52,76	47,12	39,77	39,7
	НМПКГ	32,43	32,83	31,89	31,64	32,45
Venus	SAD	50,07				
	[11]	46,64	46,64	46,05	45,92	46,16
	НМПКГ	41,44	44,94	38,8	39,75	39,63

Примечание. Полу жирным шрифтом выделены минимальные значения показателя RMSE.

соответствовать 256 градациям уровня серого. Для проведения эксперимента в среде MS Visual Studio 2013 на языке программирования С# была разработана программная модель, позволяющая строить карты глубин и по формуле (1) осуществлять оценку их точности. Исследования проводились для метода SAD, нечетких операторов (2)–(6) по методике, представленной в работе [11], и с помощью НМПКГ по стереоизображениям. В табл. 4 сведены значения показателя RMSE.








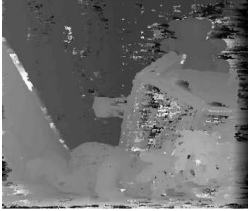

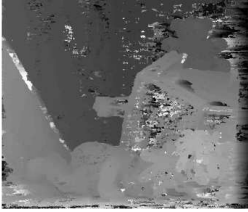
Следует отметить, что в методике [11] для построения карты глубин в композиционном правиле использовался только жесткий оператор.

Как показали данные, представленные в табл. 4, при использовании мягких операторов точность метода [11] повышается. В табл. 5 представлены видеоизображения карт глубин, полученных с помощью предложенного в статье метода. Также показана визуализация карты глубин с использованием мягких операторов (2)–(6).

Анализ табл. 5 показал, что на всех изображениях присутствуют артефакты в виде белых точек. На карте глубин белыми точками отображаются предметы, наиболее близко расположенные к видеокамерам, с помощью которых были получены стереоизображения. Например, при объезде препятствий мобильным роботом белые точки будут давать ложные объекты, находящиеся впереди траектории его движения.

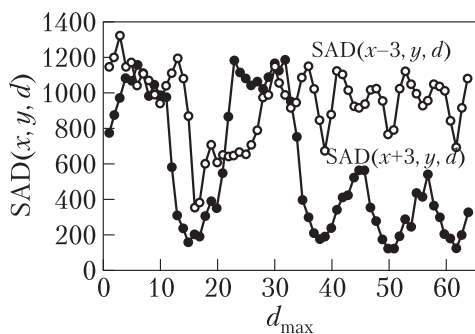
Для компенсации артефактов предлагается использовать следующий способ. Во-первых, необходимо оценить данные, которые поступают на вход нечеткой системы в зонах, имеющих артефакты. На рис. 4, а артефакт имеет координаты $x = 254$, $y = -68$. Величина смещения в данной точке равняется 62. Данная величина прогнозируется путем анализа входных данных, представленных на рис. 4, б. С учетом формулы (8) смещение равняется значению четвертого локального минимума. При масштабировании $62 \times 4 = 248$ получается светлый тон.

Таблица 5 Сравнительный анализ выходных изображений

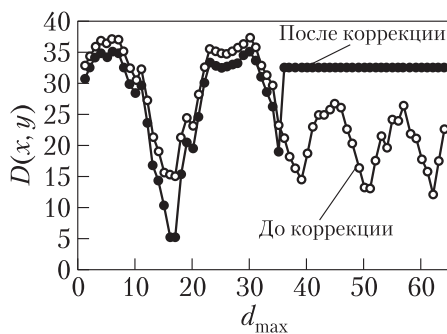
Оператор	Методика [11]	НМПКГ
MIN		
PROD		
MEAN		
Soft-MIN I		
Soft-MIN II		



(а)



(б)



(в)

Рис. 4 Карта глубин: (а) Teddy; (б) до коррекции $D(x, y)$; (в) после коррекции $D(x, y)$

Во-вторых, оценить реальное значение величины смещения на карте глубин. Визуальная оценка (см. рис. 4, б) показывает, что правильное значение смещения на карте глубин должно соответствовать первому минимуму и равняться 16.

В-третьих, исключить из расчета ложные зоны срабатывания нечеткого вывода. Для решения этой задачи необходимо ввести дополнительное условие расчета величины смещения. С учетом данных, представленных на рис. 4, б, таким условием может служить правило:

ЕСЛИ ($SAD(x, y, d) > 1200$) И ($D(x, y) \geq 32$), ТО $A_i = 1000$.

Данное компенсирующее правило позволяет сократить появление максимальных значений величины $D(x, y)$, определенной по формуле (8) (рис. 4, в). С использованием предложенного правила на выходе нечеткой системы формируется значение, равное 16.

Визуализация карты глубин с использованием компенсирующего правила представлена на рис. 5. Анализ рис. 5, в показывает, что белые точки отсутствуют на результирующем изображении.

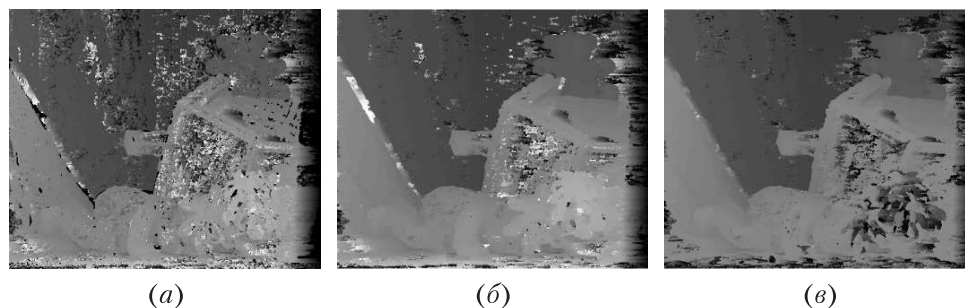


Рис. 5 Коррекция: (а) алгоритм SAD; (б) оператор Soft-min I без коррекционного правила; (в) оператор Soft-min I с коррекционным правилом

5 Обсуждение результатов моделирования и заключение

Анализ моделирования процесса построения карты глубин с использованием предложенного метода показал:

- (1) использование мягких операторов на основе анализа показателя RMSE увеличивает точность НМПКГ. Так, предложенный в статье НМПКГ для трех анализируемых изображений имеет преимущество по сравнению с алгоритмом SAD на 24%, по сравнению с методом-прототипом — на 11%;
- (2) при использовании в композиционном правиле мягких операторов точность нечеткой модели повышается. Лучшие значения коэффициента RMSE были получены при использовании следующих мягких операторов: MEAN, Soft-min I и Soft-min II;
- (3) для устранения артефактов необходимо в структуре нечеткого вывода ввести компенсирующее правило, которое ограничивает возможное значение входных величин нечеткой модели при наступлении пороговых значений величины интенсивности и максимума величины смещения;
- (4) наименьшее время работы нечеткой системы вывода было получено при использовании мягкого оператора MEAN.

Литература

1. *Algabri M., Mathkour H., Ramdane H., Alsulaiman M.* Comparative study of soft computing techniques for mobile robot navigation in an unknown environment // *Comput. Hum. Behav.*, 2015. Vol. 50. P. 42–56. doi: 10.1016/j.chb.2015.03.062.
2. *Бобырь М. В.* Адаптация системы управления мобильным роботом на основе нечеткой логики // *Мехатроника, автоматизация, управление*, 2015. Т. 16. № 7. С. 449–455. doi: 10.17587/mau.16.449-455.

3. *Delmerico J. A., David P., Corso J. J.* Building facade detection, segmentation, and parameter estimation for mobile robot stereo vision // *Image Vision Comput.*, 2013. Vol. 31. Iss. 11. P. 841–852. doi: 10.1016/j.imavis.2013.08.006.
4. *Ланге М. М., Ганебных С. Н., Ланге А. М.* Об эффективности иерархического алгоритма поиска приближенного ближайшего соседа в заданном наборе изображений // *Информатика и её применения*, 2017. Т. 11. Вып. 3. С. 51–59. doi: 10.14357/19922264170306.
5. *Zha D., Jin X., Xiang T.* A real-time global stereo-matching on FPGA // *Microprocess. Microsy.*, 2016. Vol. 47. P. 419–428. doi: 10.1016/j.micpro.2016.08.005.
6. *Pérez-Patricio M., Aguilar-González A., Arias-Estrada M., Hernández-De León H. R., Camas-Anzueto J. L., de Jesús Osuna-Coutiño J. A.* An FPGA stereo matching unit based on fuzzy logic // *Microprocess. Microsy.*, 2016. Vol. 42. P. 87–99. doi: 10.1016/j.micpro.2015.10.011.
7. *Bobyr M. V., Milostnaya N. A., Kulabuhov S. A.* A method of defuzzification based on the approach of areas' ratio // *Appl. Soft Comput.*, 2017. Vol. 59. Iss. 10. P. 19–32. doi: 10.1016/j.asoc.2017.05.040.
8. *Бобырь М. В., Кулабухов С. А., Милостная Н. А.* Обучение нейро-нечеткой системы на основе метода разности площадей // *Искусственный интеллект и принятие решений*, 2016. № 4. С. 15–26.
9. *Бобырь М. В.* Метод нелинейного обучения нейро-нечеткой системы вывода // *Искусственный интеллект и принятие решений*, 2018. № 1. С. 67–75.
10. Middlebury Stereo Datasets. <http://vision.middlebury.edu/stereo/data>.
11. *Georgoulas C., Andreadis I.* A real-time fuzzy hardware structure for disparity map computation // *J. Real-Time Image Pr.*, 2011. Vol. 6. Iss. 4. P. 257–273. doi: 10.1007/s11554-010-0157-6.

Поступила в редакцию 19.09.18

THE METHOD OF DEPTH MAP CALCULATION BASED ON SOFT OPERATORS

M. V. Bobyr, A. E. Arkhipov, and N. A. Milostnaya

Southwest State University, 94 50 Let Oktyabrya Str., Kursk 305040, Russian Federation

Abstract: The fuzzy method of depth map calculation using stereo images based on SAD (Sum of Absolute Differences) algorithm composition and fuzzy inference is considered. The feature of this method is soft arithmetic operators with fuzzy implication usage. The accuracy of the method of depth map construction is estimated by RMSE (root mean square error). The best soft operator has minimum RMSE. The method of depth map calculation which has 7 steps is presented. The proposed method showed that the accuracy of the SAD algorithm increases by 20% when soft operators are used. This conclusion is confirmed by the simulation results presented in the article.

Keywords: stereo vision; depth map; soft computing; SAD; fuzzy logic

DOI: 10.14357/08696527190207

Acknowledgments

The work was supported by the grant of the President of Russian Federation (MD-707.2017.8) and State Task (agreement No. 2.3440.2017/4.6).

References

1. Algabri, M., H. Mathkour, H. Ramdane, and M. Alsulaiman. 2015. Comparative study of soft computing techniques for mobile robot navigation in an unknown environment. *Comput. Hum. Behav.* 50:42–56. doi: 10.1016/j.chb.2015.03.062.
2. Bobyr, M. V. 2015. Adaptatsiya sistemy upravleniya mobil'nyim robotom na osnove nechetkoy logiki [Adaptation of the mobile robot control system based on fuzzy logic]. *Mekhatronika, avtomatizatsiya, upravlenie* [Mechatronics Automation Control] 16(7):449–455. doi: 10.17587/mau.16.449-455.
3. Delmerico, J. A., P. David, and J. J. Corso. 2013. Building facade detection, segmentation, and parameter estimation for mobile robot stereo vision. *Image Vision Comput.* 31(11):841–852. doi: 10.1016/j.imavis.2013.08.006.
4. Lange, M. M., S. N. Ganebnyh, and A. M. Lange. 2017. Ob effektivnosti ierarkhicheskogo algoritma poiska priblizhennogo blizhayshego soseda v zadannom nabore izobrazheniy [On efficiency of the hierarchical algorithm for searching approximate nearest neighbor in a given set of images]. *Informatika i ee Primeneniya — Inform. Appl.* 11(3):51–59. doi: 10.14357/19922264170306.
5. Zha, D., X. Jin, and T. Xiang. 2016. A real-time global stereo-matching on FPGA. *Microprocess. Microsy.* 47:419–428. doi: 10.1016/j.micpro.2016.08.005.
6. Pérez-Patricio, M., A. Aguilar-González, M. Arias-Estrada, H. R. Hernández-De León, J. L. Camas-Anzueto, and J. A. de Jesús Osuna-Coutiño. 2016. An FPGA stereo matching unit based on fuzzy logic. *Microprocess. Microsy.* 42:87–99. doi: 10.1016/j.micpro.2015.10.011.
7. Bobyr, M. V., N. A. Milostnaya, and S. A. Kulabuhov. 2017. A method of defuzzification based on the approach of areas' ratio. *Appl. Soft Comput.* 59(10):19–32. doi: 10.1016/j.asoc.2017.05.040.
8. Bobyr, M. V., S. A. Kulabuhov, and N. A. Milostnaya. 2016. Obuchenie neuro-nechetkoy sistemy na osnove metoda raznosti ploshchadey [Teaching of neuro-fuzzy system on the basis of the method of difference areas]. *Artificial Intelligence Decision Making* 4:15–26.
9. Bobyr, M. V. 2018. Metod nelineynogo obucheniya neuro-nechetkoy sistemy vyvoda [The method of non-linear learning the neuro-fuzzy inference system]. *Artificial Intelligence Decision Making* 1:67–75.
10. Middlebury Stereo Datasets. Available at: <http://vision.middlebury.edu/stereo/data> (accessed April 2, 2019).

11. Georgoulas, C., and I. Andreadis. 2011. A real-time fuzzy hardware structure for disparity map computation. *J. Real-Time Image Pr.* 6(4):257–273. doi: 10.1007/s11554-010-0157-6.

Received September 19, 2018

Contributors

Bobyr Maxim V. (b. 1978)— Doctor of Science in technology, professor, Southwest State University, 94 50 Let Oktyabrya Str., Kursk 305040, Russian Federation; fregat_mn@rambler.ru

Arkhipov Alexander E. (b. 1976) — Candidate of Science (PhD) in technology, senior scientist, Research Institute of Radio-electronic Systems, Southwest State University, 94 50 Let Oktyabrya Str., Kursk 305040, Russian Federation; alex.76_09@mail.ru

Milostnaya Natalya A. (b. 1975) — Candidate of Science (PhD) in technology, senior scientist, Southwest State University, 94 50 Let Oktyabrya Str., Kursk 305040, Russian Federation; nat_mil@mail.ru

ПРАКТИЧНОЕ ОПРЕДЕЛЕНИЕ «ПРОГРАММНО ОПРЕДЕЛЯЕМОГО»

*В. Б. Егоров*¹

Аннотация: В новом тысячелетии в информационные технологии (ИТ) вошло понятие программного определения, и сегодня предметом обсуждения становятся все новые категории программно определяемых (software-defined) объектов: сети, хранилища, серверы и, наконец, полностью программно определяемые центры обработки данных (ЦОДы). Однако, несмотря на расширяющуюся номенклатуру объектов и частую повторяемость в литературе и дискуссиях словосочетания «программно определяемый», уловить его подлинный смысл непросто. Более того, существует достаточно распространенное мнение, что это вообще не технический термин, а маркетинговый трюк. Нет ясного общепринятого определения этого феномена и в актуальных энциклопедиях. В настоящей статье сделана попытка выявить технический смысл «программно определяемого» и дать ему по возможности понятное практическое определение. На основе анализа особенностей инфраструктурных компонентов программно определяемых ЦОДов, в которых было декларировано программное определение, сделан вывод, что «программно определяемое» можно считать техническим термином, и для него сформулировано определение в известных общепринятых понятиях.

Ключевые слова: IaaS; policy-based; SDDC; автоматизация управления; инфраструктура как сервис (IaaS); программно определяемые сети (SDN); программно определяемые хранилища (SDS)

DOI: 10.14357/08696527190208

1 Введение

В начале тысячелетия в ИТ вошло в моду понятие программного определения и стали появляться все новые категории программно определяемых (software-defined) объектов. В середине 2000-х гг. в ИТ появились программно определяемые сети (SDN, software-defined networking), за ними с начала 2010-х гг. последовали программно определяемые хранилища (SDS, software-defined storage). Дальнейшее объединение SDN и SDS в рамках ЦОДа позволило говорить о программно определяемых центрах обработки данных (SDDC, software-defined data centers). Последующая экспансия понятия привела к тому, что сегодня программно определяемыми могут объявляться все компоненты инфраструктуры ЦОДа, причем не только основные функциональные — серверы, сети и храни-

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, VEgorov@ipiran.ru

лица, но также и вспомогательные, такие как питание и охлаждение. Завершением процесса универсализации программного определения стало «программно определяемое все» (SDE, software-defined everything), охватившее множество разнообразных продуктов и технологий, в которых декларировалось или в названии которых хотя бы присутствовало нечто «программно определяемое».

Однако, несмотря на моду и привычность бытования на слуху, подлинный смысл словосочетания «программно определяемое» уловить не так-то просто. Более того, может сложиться впечатление, что смысла в нем и вовсе нет. Например, существует достаточно распространенное мнение, что программное определение — «скорее маркетинговый трюк, чем технический термин» [1]. Присоединяясь к нему, сайт TechTarget саркастически замечает: «*Программно определяемое* — это термин, используемый для описания широкого круга продуктов, но он нередко применяется для привлечения внимания, а не описания функционирования продукта» [2]. Не увенчается успехом и попытка найти определение программно определяемого в интернете. Так, англоязычная Wikipedia предлагает десяток статей, содержащих термин software-defined, по конкретным программно определяемым объектам, но так и не дает для них обобщающего определения. Таким образом, активно распространяясь и завоевывая все новые сферы ИТ, понятие программного определения остается, по существу, не определенным.

Настоящая статья представляет собой попытку ответить на вопрос, есть ли смысл в программном определении, является ли оно техническим термином и, при положительном ответе, дать ему по возможности простое адекватное определение в известных понятиях.

2 Существующие трактовки программного определения

Поскольку само словосочетание «программно определяемый» весьма популярно, на самом деле в различных теоретических статьях и практических обсуждениях можно найти немало его трактовок и описательных дефиниций, но, как правило, ориентированных на конкретные обсуждаемые проблемы или предлагаемые продукты. Их обзор разумно начать с компании VMware, которая собственно и ввела в обращение сам термин.

У VMware программное определение — это унифицированная, основанная на линии поведения (policy-driven) технология автоматизированного управления сервисами в ЦОДах. Его задача — сделать сервисы такими же доступными в конфигурации и управлении, как и ставшие привычными виртуальные машины (ВМ). В число таких сервисов включается инфраструктура ЦОДа, также предоставляемая потребителю как услуга (IaaS, infrastructure as a service) [3].

Конкурент VMware в «облачных» технологиях компания Citrix Systems признает программное определение многогранной проблемой, связанной с виртуализацией, но не сводящейся к ней. Виртуализация должна быть дополнена технологией «оркестровки» (orchestration) и каталогами самообслуживания. При

этих условиях появляется возможность предоставления ресурсов по запросу и реализации технологии «облачных» сервисов [4].

Обсуждая вопросы, связанные с SDDC, в компании Cisco Systems предпочитают называть такого рода ЦОДы, избегая бессмысленных штампов, не программно определяемыми, а управляемыми в соответствии с правилами, делая упор на способах управления инфраструктурой [5].

Не вдаваясь в технические подробности, компания TechBeacom видит назначение программного определения в смещении акцентов с минимизации стоимости ИТ-услуг на оптимизацию скорости их доставки потребителю [6]. А в компании Nth Generation Computing SDDC представляют как ЦОД, конфигурируемый едва ли не мгновенно и по желанию кого угодно, исключая необходимость в специальных знаниях, умениях и опыте [4].

Двойственна позиция компании «Инфосистемы Джет». Полагая программное определение «ярлыком, который наклеивается на разное содержание в зависимости от контекста и особенностей портфолио поставщиков решений», ее ИТ-инженеры, тем не менее, согласны понимать его как подход, предусматривающий комплексное программное управление компонентами ЦОДа, в том числе виртуализированными [7].

Разобраться в этой пестроте мнений, порой противоречивых, непросто. Поэтому представляется разумным отвлечься от готовых определений и трактовок, зачастую порожденных интересами конкретных фирм и рыночной конъюнктурой, и обратиться непосредственно к предмету нашего интереса в компонентах инфраструктуры ЦОДов, в которых этот термин возник изначально, а именно программно определяемым сетям и хранилищам.

3 Программное определение в сетях

В контексте программного определения сетей прижился и регулярно повторяется и в рекламе, и в аналитических статьях целый ряд словесных штампов, которые вследствие регулярного повторения стали восприниматься как естественные атрибуты SDN, но на деле не имеют прямого отношения к программному определению как таковому.

В первую очередь, здесь следует назвать часто акцентируемое в характеристиках SDN разделение плоскостей данных и управления. Но такое разделение возникло в традиционных сетях задолго до появления концепции SDN, оно свойственно любым сетям, в том числе не претендующим на программное определение. Для SDN принципиальна лишь централизация управления сетью и существенно не разделение плоскостей, а вырождение плоскости управления в «точку» — логически единый контроллер сети [8, с. 111]. В роли сетевого контроллера, из которого осуществляется централизованное управление сетями SDN, как правило, предполагается высокопроизводительный сервер или работающая на нем ВМ. Известны примеры специально разработанного под сети SDN системного программного обеспечения (ПО), например операционной системы RUNOS, однако

в основном управление сетями SDN осуществляется прикладным ПО, в качестве которого наиболее часто упоминаются так называемые «оркестровщики» (orchestrators). При этом под «оркестровкой» понимается автоматизация системного менеджмента и управления инфраструктурой, причем автоматизация, принципиально основанная на заранее заданных правилах поведения или «политиках» (policy-based) [9].

Точно так же не имеет прямого отношения к сетям SDN, хотя регулярно упоминается в связи с ними, сетевая виртуализация. Связь между SDN и виртуализацией существует, но лишь косвенная — через «территориальную» общность: концепция SDN практически применима и на сегодняшний день демонстрируется немногими примерами реализации только в сетях крупных ЦОДов [8, с. 114], а такие ЦОДы неуклонно движутся к тотальной виртуализации всех ресурсов, в том числе сетевых. Таким образом, виртуализация сетей и их программное определение — параллельные процессы, протекающие в ЦОДах и interfering, причем не всегда во благо друг другу. Например, вопреки расхожему мнению виртуализация сети не упрощает, а в целом скорее усложняет централизованное управление ею. В случае виртуализированной сети SDN приходится управлять параллельно, причем в тесной взаимной связи, и виртуальными наложенными сетями, и физической сетью как основой, на которой функционируют наложенные виртуальные. Механизмы управления виртуальными и физическими сетями в общем случае разные и не обязательно легко согласуемы. Неслучайно поддержку виртуальных сетей в многосерверной конфигурации ЦОДа, принципиально требующей наличия аппаратной сетевой инфраструктуры, компания VMware начала обеспечивать лишь в относительно позднем расширении своей платформы виртуализации NSX: версии NSX Enterprise [10, с. 92]. Поэтому вряд ли уместно упоминание виртуализации среди атрибутов и безоговорочных преимуществ сетей SDN. Тем не менее объективное усложнение задачи управления виртуализированной сетевой инфраструктурой безусловно стимулирует внедрение любого рода его автоматизации, в том числе автоматизации, основанной на «политиках».

Таким образом, в рафинированном, очищенном от заблуждений и маркетинговых штампов виде программное определение сетей сводится:

- к централизации управления в логически едином сетевом контроллере;
- к основанной на «политиках» автоматизации управления сетью работающим в сетевом контроллере прикладным ПО («оркестровщиком»).

4 Программное определение в хранилищах

Появившись позже SDN, SDS переняли у них некоторые архитектурные принципы организации и соответствующую терминологию. В числе прочего это относится к разделению плоскостей данных и управления, хотя функции этих плоскостей в концепциях SDN и SDS не полностью идентичны вследствие объективных различий между сетями и хранилищами.

Плоскости данных в сетях и хранилищах базируются на принципиально разной физической аппаратуре. В SDN это унифицированные коммутаторы, работающие по единому протоколу, в частности OpenFlow. Унификация аппаратуры и стандартизация протокола позволяют помимо устранения зависимости ПО от поставщиков оборудования вынести управление SDN на прикладной уровень ПО, работающего в контроллере сети. В отличие от унифицированных коммутаторов SDN, SDS могут включать физически самые разные устройства: дисковую память серверов, сети хранения (SAN, storage-area networks), сетевые хранилища (NAS, network attached storage) и др. с разной организацией и различными методами доступа, в том числе специфическими, зависящими от поставщика этих устройств. Отсюда существенная разница в путях достижения схожих целей. В SDN плоскость данных в значительной степени реализуется на прикладном уровне, лишь затрагивая системное ПО сетевого контроллера, например в части поддержки OpenFlow. В SDS плоскость данных вынужденно опускается ниже и внедряется в системное ПО всех серверов, так или иначе имеющих отношение к физическим устройствам хранения. Также в противовес SDN, где функции плоскости данных пространственно локализованы в одном сервере — контроллере сети, в SDS эти функции вынужденно «размазываются» по множеству систем NAS/SAN и серверов. При таких условиях они могут реализовываться, только вовлекая все уровни ПО, включая самые нижние вплоть до гипервизора. Тем не менее принципиальная для SDS консолидация всех ресурсов хранения в общий виртуальный пул параллельно требует централизованного управления этим консолидированным ресурсом. Централизация может реализовываться, аналогично SDN, в физическом сервере или ВМ на уровне операционной системы и прикладного ПО, как это сделано, например, в платформе VMware vCenter Server.

Зато в плоскости управления сходства между SDN и SDS гораздо больше. Плоскость управления SDS обеспечивает общее управление различными уровнями хранения и динамическое управление уровнями обслуживания хранилища путем автоматизации на основе «политик» и с учетом виртуализации [11, с. 3]. Поскольку управление SDS осуществляется как единым целым, оно предполагается, как и в SDN, централизованным и также на основе наперед заданных «политик», т. е. в плоскости управления у SDN и SDS одни и те же цели достигаются схожими методами.

В итоге рафинированные особенности программного определения хранилищ предстают как:

- консолидация физических ресурсов хранения в логически едином, предпочтительно виртуализированном (но необязательно), пуле;
- централизация и основанная на «политиках» автоматизация управления этим пулом с помощью специального распределенного по физическим устройствам ПО, включая прикладное и системное.

5 Программное определение в серверах

Если для сетей и хранилищ программное определение обрело некие очертания и формы, то для главного компонента ЦОДов, вычислительного, оно долго оставалось где-то «за кадром». Возможно, это произошло из-за принимаемого по умолчанию положения, что в компьютерах и серверах все работы, т. е. в данном случае вычисления, выполняются, а значит, и могут определяться только программно. И все-таки чисто формальное отсутствие программно определяемых серверов вносило некий диссонанс в общую картину SDDC, тем более что параллельный программному определению процесс виртуализации уже завершился во всех трех главных инфраструктурных составляющих ЦОДов [12].

Может быть, именно желание избавиться от этого диссонанса послужило в конце концов толчком к появлению программно определяемых серверов. Произошло это относительно недавно, позже SDN и SDS, и у программно определяемых серверов (software-defined servers) еще не сложился общепринятый акроним. (Гут имеется объективная трудность из-за формального совпадения с устоявшейся аббревиатурой SDS для хранилищ.)

Программно определяемые серверы не обзавелись не только акронимом, но и общепринятым содержанием самого понятия. Например, Wikipedia под программным определением серверов понимает эффективное управление посредством специального ПО всеми компонентами сервера, включая питание и охлаждение. Близкое к этому понимание программного определения серверов демонстрирует компания Datacenterdynamics, распространяя его на весь ЦОД [13]. Действительно, на рынке уже имеются программно определяемые источники питания и охлаждения. В качестве примеров можно назвать «интеллекгентные» батареи бесперебойного питания компании Virtual Power Systems и управляемые системы охлаждения компании Inertech. Таким образом, все физические компоненты ЦОД становятся программно определяемыми, устраняя упомянутый выше диссонанс.

Однако есть и иные определения программного определяемого сервера. Так, компания TidalScale трактует программно управляемый сервер в русле концепции SDS как объединение множества физических компьютеров в одну виртуальную среду. Программное определение такой среды предполагается специальным ПО, в качестве которого TidalScale предлагает фирменный продукт HyperKernel Software. Однако если трактовка TidalScale и имеет отношение к программному определению, то лишь косвенное, так как тут речь, по существу, идет о классической виртуализации ЦОДа с типичным монитором ВМ, а еще более, надо полагать, о рекламе фирменного ПО.

Интересным и перспективным с точки зрения программного определения решением с объединением в одном продукте всех аппаратных компонентов SDDC, включая серверы, хранилища, сеть, питание и охлаждение, может стать гиперконвергенция. Гиперконвергентные инфраструктуры проще компонуются в ЦОДах и легче поддаются программному определению в силу своей изначально-

ной преконфигурированности. На сегодняшнем рынке аппаратуры ЦОДов доля гиперконвергентных систем быстро растет и во многом благодаря тому, что их поставщики предлагают с ними специализированное ПО, в том числе с отдельными готовыми модулями программного определения [3, 13].

Интегрируя различные имеющиеся подходы к программному определению серверов, последнее можно определить как комплексную автоматизацию централизованного программного управления всеми серверами ЦОДа и всеми работающими на них ВМ, которая может также распространяться на служебные системы серверов, такие как питание и охлаждение.

6 Заключение

Проведенный анализ программного определения в основных сферах его практического приложения показывает, что при всем разнообразии имеющихся его трактовок практически во всех лейтмотивом звучит тема комплексной автоматизации управления инфраструктурой ЦОДов во имя гибкого и быстрого предоставления пользователям «облачных» услуг, вплоть до IaaS. Также следует обратить внимание на доминанту этого лейтмотива: основание на правилах, так называемых «политиках». Сам термин «политики» тут неслучаен и является производным от парадигмы программирования *policy-based programming* (или *policy-based design*). Не отвлекаясь на подробности, для целей данной статьи достаточно констатировать суть этой доминанты: программирование неких «политик».

С учетом выявленного лейтмотива и его доминанты наиболее адекватными представляются следующие два взаимодополняющих определения программно определяемого ЦОДа:

- (1) «SDDC — это ЦОД, где все физические ресурсы виртуализированы. . . управляются из единой точки с помощью специализированного ПО и предоставляются как сервис» [14];
- (2) «помимо трех ключевых программно определяемых компонентов SDDC: вычислений, хранения и сетей — необходим еще уровень полной и сквозной автоматизации управления в соответствии с заданными правилами (так называемой «оркестровки»)» [5].

Объединяя их, можно определить программное определение через общепринятые понятия как комплексную автоматизацию централизованного управления инфраструктурой ЦОДов специализированным ПО.

Таким образом, программное определение имеет-таки смысл и является техническим термином. Правда, может возникнуть закономерный вопрос: зачем вводить новый термин с неясным смыслом, если то же самое можно определить общеизвестными очевидными понятиями? Вероятно, ответ на этот вопрос надо искать в маркетинговой психологии. В сегодняшнем мире ИТ все «программно определяемое», пусть и непонятное, воспринимается (может быть, именно в силу

своей непонятности) как нечто современное, гибкое и перспективное, в то время как набившая оскомину «комплексная автоматизация» кажется чем-то закостеневшим бюрократическим и неповоротливым. Однако на практике автоматизация управления на основе «политик», т. е. установленных правил поведения и принятия решений, может быть эффективной и прекрасно работать; надо только, чтобы кто-то не поленился правильно запрограммировать эти правила [15]. А чтобы победить инерцию и предубеждение, полезно назвать запрограммированные старые добрые правила новомодным термином, тем же программным определением, например, что уже сделали и продолжают делать многие фирмы.

Литература

1. *Taylor C.* Software-defined storage: It's real but defining it can be a real trick // *Enterprise Storage*, 2017. <http://www.enterprisestorageforum.com/storage-management/software-defined-storage-its-real-but-defining-it-can-be-a-real-trick-1.html>.
2. What does 'software defined' actually mean? // *TechTarget*. <https://searchservvirtualization.techtarget.com/feature/What-does-software-defined-actually-mean>.
3. The Software-Defined Data Center // *VMware*, 2018. <https://www.vmware.com/solutions/software-defined-datacenter.html>.
4. *Kovar J. F.* Software-defined data centers: Should you jump on the bandwagon? // *CRN*, 2013. <https://www.crn.com/news/data-center/240154576/software-defined-data-centers-should-you-jump-on-the-bandwagon.htm>.
5. *Ганьжа Д.* На пути к программно определяемому центру обработки данных // *Журнал сетевых решений/LAN*, 2016. № 7-8. С. 20–24. <https://www.osp.ru/lan/2016/07-08/13050118>.
6. *Bradley T.* Why you should embrace the software-defined data center // *TechBeacon*. <https://techbeacon.com/why-you-should-embrace-software-defined-data-center>.
7. Программно-определяемый ЦОД сегодня и завтра // *Инфосистемы Джет*, 2015. <https://jet.su/about/news/18464/>.
8. *Егоров В. Б.* Некоторые вопросы практической реализации концепции SDN // *Системы и средства информатики*, 2016. Т. 26. № 1. С. 109–120.
9. *Erl T.* Service-oriented architecture: Concepts, technology & design. — Upper Saddle River, NJ, USA: Prentice Hall, 2005. 792 p.
10. *Егоров В. Б.* Некоторые неявные вопросы сетевой виртуализации // *Системы и средства информатики*, 2017. Т. 27. № 3. С. 88–98.
11. Программное хранилище: ценность для ИТ-специалиста // *Открытые системы*, 2015. 5 с. https://www.osp.ru/netcat_files/1766/1964/h_56c345a97137b2fdc2b64e85d7897f2.
12. *Beal V.* SDDC — software-defined data center // *Webopedia*. https://www.webopedia.com/TERM/S/software_defined_data_center_SDDC.html.
13. *Robinson D.* SDDC is the answer — but how do we get there? // *Datacenterdynamics*, 2018. <https://www.datacenterdynamics.com/analysis/sddc-is-the-answer-but-how-do-we-get-there>.

14. *Владимирович И.* Программно-определяемый ЦОД. Грамотный выбор // Connect, 2016. № 10. <https://ланит.рф/press/smi/10073>.
15. *Asprey D.* Software defined software: The history of “software defined” // Trend Micro, 2013. <https://blog.trendmicro.com/software-defined-software-the-history-of-software-defined>.

Поступила в редакцию 11.01.19

A PRACTICAL DEFINITION OF “SOFTWARE-DEFINED”

V. B. Egorov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: In the new millennium, the “software-defining” concept has entered information technologies, with increasing number of software-defined objects being involved as discussion subjects: networks, storages, servers, and finally entirely software-defined datacenters. However, despite the multiplying subjects and frequent reiteration of the set expression “software-defined,” its real meaning is not so easy to perceive. Moreover, a sufficiently widespread opinion considers the expression rather a marketing trick than a technical term. Actual encyclopedias do not suggest a clear commonly accepted definition of this phenomenon as well. In the article, an attempt is made to uncover the technical essence of “software-defined,” and to formulate for it a plain practical definition. Basing on unbiased analysis of peculiarities of datacenter infrastructure components declared as software-defined, a conclusion was drawn that “software-defined” may be regarded as a technical term and provided with a simple definition in conventional generally accepted notions.

Keywords: control automation; infrastructure as a service (IaaS); policy-based; policy-driven; SDDC; SDN; SDS; software-defined

DOI: 10.14357/08696527190208

References

1. Taylor, C. 2017. Software-defined storage: It’s real but defining it can be a real trick. *Enterprise Storage*. Available at: <http://www.enterprisestorageforum.com/storage-management/software-defined-storage-its-real-but-defining-it-can-be-a-real-trick-1.html> (accessed January 10, 2019).
2. What does “software defined” actually mean? 2018. *TechTarget*. Available at: <https://searchservirtualization.techtarget.com/feature/What-does-software-defined-actually-mean> (accessed January 10, 2019).
3. The software-defined data center. 2017. *VMware*. Available at: <https://www.vmware.com/solutions/software-defined-datacenter.html> (accessed January 10, 2019).

4. Kovar, J. F. 2013. Software-defined data centers: Should you jump on the bandwagon? *CRN*. Available at: <https://www.crn.com/news/data-center/240154576/software-defined-data-centers-should-you-jump-on-the-bandwagon.htm> (accessed January 10, 2019).
5. Ganzha, D. 2016. Na puti k programmno opredelyaemomu tsentru obrabotki dannykh [On the way to a software-defined datacenter]. *Zhurnal setevykh resheniy/LAN* [Network/LAN Solutions Magazine] 7-8:20–24. Available at: <https://www.osp.ru/lan/2016/07-08/13050118> (accessed January 10, 2019).
6. Bradley, T. 2017. Why you should embrace the software-defined data center. *Tech-Beacom*. Available at: <https://techbeacon.com/why-you-should-embrace-software-defined-data-center> (accessed January 10, 2019).
7. Programmno-opredelyaemyy TsOD segodnya i zavtra [The software-defined datacenter today and tomorrow]. 2015. *Infosistemy Dzheta — Jet Infosystems*. Available at: <https://jet.su/about/news/18464/> (accessed January 10, 2019).
8. Egorov, V. B. 2016. Nekotorye voprosy prakticheskoy realizatsii kontseptsii SDN [Some issues of the SDN concept practical implementation]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(1):109–120.
9. Erl, T. 2005. *Service-oriented architecture: Concepts, technology & design*. Upper Saddle River, NJ: Prentice Hall. 792 p.
10. Egorov, V. B. 2017. Nekotorye neyavnye voprosy setevoy virtualizatsii [Some implicit issues of network virtualization]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 27(3):88–98.
11. Programmnoe khranilishche: tsennost' dlya IT-spetsialista [Software storage: Value for the IT-specialist]. 2015. *Otkrytye sistemy* [Open systems]. 5 p. Available at: https://www.osp.ru/netcat_files/1766/1964/h_56c345a97137b2fdcf2b64e85d7897f2 (accessed January 10, 2019).
12. Beal, V. SDDC — software-defined data center. *Webopedia*. Available at: https://www.webopedia.com/TERM/S/software_defined_data_center_SDDC.html (accessed January 10, 2019).
13. Robinson, D. 2018. SDDC is the answer — but how do we get there? *Datacenter-dynamics*. Available at: <https://www.datacenterdynamics.com/analysis/sddc-is-the-answer-but-how-do-we-get-there/> (accessed January 10, 2019).
14. Vladimirovich, I. 2016. Programmno-opredelyaemyy TsOD. Gramotnyy vybor [Software-defined datacenter. The competent choice]. *Connect* 10.
15. Asprey, D. 2013. Software defined software: The history of “software defined.” *Trend Micro*. Available at: <https://blog.trendmicro.com/software-defined-software-the-history-of-software-defined/> (accessed January 10, 2019).

Received January 11, 2019

Contributor

Egorov Vladimir B. (b. 1948)—Candidate of Science (PhD) 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 119133, Russian Federation; VEgorov@ipiran.ru

ВОССТАНОВЛЕНИЕ СИГНАЛА В СИСТЕМАХ РЕАЛЬНОГО ВРЕМЕНИ ПО РАВНОМЕРНЫМ ВЫБОРКАМ С УМЕНЬШЕНИЕМ ИНТЕРВАЛА НАЙКВИСТА

*Б. Г. Майоров*¹

Аннотация: Предложена математическая модель поведения сигнала между двумя соседними выборками из него. Получены новые общие формулы для нахождения синусоидального критерия наибольшего отклонения. Доказана теорема об уменьшении интервала Найквиста, которая связывает критерий наибольшего отклонения и величину постоянного интервала дискретизации. Результаты позволяют обоснованно определять величину этого интервала.

Ключевые слова: системы реального времени; синусоидальный критерий наибольшего отклонения сигнала; ступенчатая или линейная аппроксимация сигнала по выборкам; допустимая погрешность аппроксимации; постоянный интервал дискретизации времени

DOI: 10.14357/08696527190209

1 Введение

Область применения результатов настоящей статьи — проектирование систем управления жесткого реального времени (СУРВ) для объектов, функционирующих в автоматическом режиме. Напомним требования к дискретизации времени и восстановлению сигнала по выборкам, возникающие из особенностей функционирования СУРВ.

1. Необходимо определить постоянный интервал равномерной дискретизации сигнала по времени и задаться типом аппроксимации сигнала по выборкам. При этом адаптирующая аппроксимация с переменным шагом дискретизации не применяется.
2. В современных СУРВ используют аппроксимацию ступеньками или отрезками прямых линий по выборкам и существенно реже используют функцию $\text{sinc}(x)$ [1, 2].
3. Необходимо обработку информации и выдачу сигнала управления на исполнительные органы СУРВ проводить за один или два интервала дискретизации времени.
4. Необходимо обеспечить сохранение значений сигнала без искажений в точках осуществления выборок из сигнала при аппроксимации сигнала.

¹АО «Научно-производственное предприятие «Рубин»», г. Пенза, b.g.m.1947@rambler.ru

Аппроксимация сигнала по выборкам из него происходит с погрешностью аппроксимации Δ , равной наибольшему значению разности между первичным аналоговым значением сигнала на выходе датчика и его вторичным значением, восстановленным путем аппроксимации величин выборок. Постоянный интервал дискретизации времени T_s зависит от закона аппроксимации и максимально допустимой погрешности аппроксимации Δ от максимальных параметров аппроксимируемого сигнала $p(t)$: граничной частоты f в спектре частот, наибольшей скорости изменения $\max p^{(1)}(t) = V$, наибольшего ускорения изменения $\max p^{(2)}(t) = a$, наибольшей резкости изменения $\max p^{(3)}(t) = R$ и т. д.

Таким образом, можно записать функциональную зависимость от множества параметров:

$$T_s \leq F \max \left\{ f, \Delta, V, a, R, \dots, p^{(n)}(t) \right\}, \quad (1)$$

где (n) — наибольший порядок максимального значения производной $\max p^{(n)}(t)$ от сигнала $p(t)$ по времени, который учитывается при определении T_s .

Известны три основных метода определения частоты дискретизации $f_s = 1/T_s$ или постоянного по величине интервала дискретизации времени T_s в СУРВ.

Первый метод определен теоремой НКШ (Найквиста, Котельникова, Шеннона) [3–7].

Теорема НКШ (теорема отсчетов) [6, 7] доказана К. Шенноном на основе теоремы В. А. Котельникова [4] и работы Найквиста [3] и используется в теории и технике связи.

Теорема НКШ. *Если функция не содержит частот выше f Гц, она полностью определяется своими мгновенными значениями в моменты, отстоящие друг от друга на $1/2f$ секунд.*

Интервал времени $1/2f$ назван Шенноном [7] интервалом Найквиста [3].

Недостатки теоремы НКШ при решении задач в управляющих системах СУРВ следуют из процедур ее практического использования [7]:

- (1) восстановление функции происходит в течение большого времени существования T_c этой функции (звукового сигнала) с последовательным суммированием всех мгновенных значений (выборок), которые были сделаны из этой функции. При этом заранее известно, что справедливо неравенство $T_c \gg T_s$. В СУРВ большие задержки при обработке сигналов во времени недопустимы;
- (2) в устройстве восстановления функции по ее отсчетам в каждой точке отсчета на интервале времени T_c помещают импульс с амплитудой, равной величине каждого отсчета, на место этого отсчета, осуществляют последовательное преобразование этого импульса по передаточной характеристике фильтра нижних частот, результаты преобразований накапливают в течение интервала времени T_c и получают исходную функцию.

Математически процесс преобразования и восстановления сигнала $p(t)$ (звукового сигнала) описывается рядом В. А. Котельникова [3, 4]:

$$p(t) = \sum_{k=1}^{2T_c f} p_k \frac{\sin \pi(2ft - k)}{\pi(2ft - k)}, \quad (2)$$

где k — номер отсчета (выборки) на интервале времени T_c ; $2T_c f$ — число отсчетов на интервале времени T_c ; p_k — амплитуда отсчета с номером k .

Выражение (2) подобно гармоническому ряду Фурье и является сложным с точки зрения объема необходимых вычислительных операций для его реализации. Выражения для ступенчатой и линейной аппроксимации (для восстановления сигнала по выборкам из него) существенно проще в математическом описании и реализации;

- (3) применяемые в практике разработки и реализации СУРВ методы ступенчатой и линейной аппроксимации вообще не используются и не упоминаются при доказательстве теоремы НКШ [4, 7];
- (4) в теореме НКШ для определения T_s используется только граничная частота f в спектре частот сигнала, а остальные параметры, входящие в (1), не используются;
- (5) на практике [8] и в теории [9, 10] при использовании теоремы НКШ и ступенчатой или линейной аппроксимации сигнала по выборкам обычно используют эмпирическое соотношение $T_s \leq (1/2f)/(8 \div 10)$, которое является приблизительным и не имеет строгого теоретического обоснования. Это обусловлено большой популярностью и известностью теоремы НКШ и делает погрешность расчетов неопределенной величиной.

В связи с этими и прочими [11] недостатками нерационально определять интервал дискретизации сигнала T_s по теореме НКШ при осуществлении точных инженерных расчетов при использовании ступенчатой и линейной аппроксимации для восстановления сигнала $p(t)$. В [11] сделан справедливый вывод о необходимости использования теоремы НКШ только в теории связи (радиотехнике) для восстановления речевого сигнала, но не в системах управления реального времени.

Второй подход сводится к определению критерия наибольшего отклонения (КНО) или величины интервала дискретизации времени T_s . Критерий наибольшего отклонения предложил В. Н. Хлистунов [12, 13], который применил для этой цели наибольшее значение остаточного члена формулы Лагранжа. Достоинства критерия наибольшего отклонения заключаются в следующем.

1. Получены и обоснованы выражения параболического КНО или интервалов дискретизации времени $T_{p1} = \Delta/V$ и $T_{p2} = \sqrt{8\Delta/a}$, которые используют параметры Δ , V и a , входящие в выражение (1). Индекс « p » означает параболическую (parabola) зависимость.

2. Для восстановления сигнала $p(t)$ по выборкам из него через интервал времени $T_{p1} = \Delta/V$ используется ступенчатая аппроксимация. Для интервала дискретизации $T_{p2} = \sqrt{8\Delta/a}$ используется линейная аппроксимация.

Недостатки критерия наибольшего отклонения, которые не определены в [12, 13], следующие.

Недостаток 1. Отсутствует связь величины граничной частоты f в спектре сигнала $p(t)$ с максимальными значениями параметров V и a этого сигнала (1).

Недостаток 2. Совместное рассмотрение выражений КНО T_{p1} и T_{p2} показывает, что T_{p1} зависит только от максимальной скорости V и погрешности Δ ступенчатой аппроксимации сигнала по выборкам, а T_{p2} зависит только от максимального ускорения a и погрешности Δ линейной аппроксимации сигнала по выборкам. Зависимость T_{p2} одновременно от скорости и ускорения сигнала В. Н. Хлистуновым не получена. Это обусловлено тем, что его подход к получению КНО заключается в использовании остаточного члена формулы Лагранжа и не позволяет (принципиально) получить необходимый результат.

Недостаток 3. Отсутствует возможность определения величины T_{pn} (если $n \geq 3$) при условии восстановления сигнала при помощи ступенчатой или линейной аппроксимации, так как остаточный член формулы Лагранжа при $n = 3$ содержит максимальную величину третьей $\max p^{(3)}(t) = R$ производной от сигнала. При этом аппроксимация осуществляется квадратичными параболой по трем точкам, а не отрезками прямых линий или ступеньками. Подобный недостаток препятствует определению величины шага T_{pn} дискретизации сигнала $p(t)$ по времени, если $n \geq 3$.

Недостаток 2 КНО устранен автором в статье [14], где получено выражение в виде обобщенного критерия наибольшего отклонения для линейной аппроксимации сигнала $p(t)$ отрезком прямой линии по двум точкам

$$T_{p2c} = \frac{2\Delta}{V_{21}} + \frac{V_{21}}{a} = \sqrt{\frac{8\Delta}{a}} \left(\frac{m_V}{2} + \frac{1}{2m_V} \right) = T_{p2}\tau_{2c}. \quad (3)$$

Здесь $m_V = V/V_{21} \geq 1$ — коэффициент рассогласования скорости сигнала $p(t)$.

При $V_{21} = \sqrt{2\Delta a} = V$ (или $m_V = 1$) выражение (3) преобразуется в (2). Выражение $V_{21} = \sqrt{2\Delta a} = V$ названо условием согласования максимальных значений параметров $\max\{\Delta, V, a\}$ сигнала $p(t)$.

Добавка определения «обобщенный» к названию КНО вызвана полученным результатом в виде выражения (3), которое содержит в одной формуле все три параметра восстановленного сигнала $\max\{\Delta, V_{21}, a\}$, а не по два параметра, как в выражениях КНО В. Н. Хлистунова: $T_{p1} \leq \Delta/V$ и $T_{p2} \leq \sqrt{8\Delta/a}$.

Для использования (3) при аппроксимации сигнала $p(t)$ ступенькой по одной точке необходимо правую часть выражения (3) умножить на коэффициент 0,5:

$$T_{p2c} = \frac{\Delta}{V_{21}} + \frac{V_{21}}{2a}. \quad (4)$$

Частный случай $T_{p1} \leq \Delta/V$ зависимости (4) для ступенчатой аппроксимации сигнала $p(t)$ получается из (4) при выполнении над (4) операции нахождения предела ($a \rightarrow \infty$), если величина ускорения сигнала $p(t)$ неизвестна ($a = ?$).

Использование выражений (3) и (4) вместо КНО для расчета величины шага равномерной дискретизации T_{p2c} сигнала $p(t)$ по времени позволяет в разы увеличить эту величину ($\tau_{2c} = T_{p2c}/T_{p2} \gg 1$) без потери точности восстановления [14].

Математическое выражение (траектория), которое описывает возможное изменение сигнала $p(t)$ внутри интервала дискретизации, названа траекторией наибольшего отклонения сигнала $p(t)$ [14, 15] (далее — траектория отклонения). Здесь индекс « p » обозначает параболический закон траектории отклонения сигнала $p(t)$ от двух соседних выборок. Эта математическая модель исследована достаточно подробно и получены необходимые выражения для расчета величины шага равномерной дискретизации T_{pn} сигнала $p(t)$ по времени в общем случае [16, 17]. Недостаток 1 параболического КНО не устранен в явном виде и в этих работах. Для его устранения рассмотрим не параболическую, а синусоидальную траекторию наибольшего отклонения сигнала $p(t)$ [15] от двух соседних выборок. При этом получим необходимые общие выражения для нахождения шага равномерной дискретизации T_{sn} сигнала $p(t)$ для произвольных значений $n = 1, 2, 3, \dots$ или найдем синусоидальный КНО.

2 Синусоидальная траектория наибольшего отклонения

Уравнение синусоидальной траектории наибольшего отклонения [15, 18, 19]:

$$p(t) = \Delta \sin \omega t. \quad (5)$$

Здесь $0 \leq t \leq T_{sn}$, где $T_{sn} = T/2$; $\omega = 2\pi f = 2\pi/T = \pi/T_{sn}$, где f — граничная частота в спектре сигнала $p(t)$, T — период граничной частоты.

Выражение (5) описывает математически половину ($T_{sn} = T/2$) периода T гармоники частоты f в спектре частот сигнала $p(t)$, которая изображена на рис. 1.

Первая выборка из сигнала $p(t)$ осуществлена в момент времени $t = 0$, вторая — в момент времени $t = T_{sn} = T/2$. Линейная аппроксимация между двумя выборками изображена штрихпунктирной линией. Погрешность такой аппроксимации Δ достигает максимального значения в точке $(0,5T_{sn}, \Delta)$, которая является вершиной траектории отклонения или вершиной полуволны гармоники частоты f . Проходящая через вершину вертикальная линия $t = 0,5T_{sn} = 0,25T$ служит ее осью симметрии.

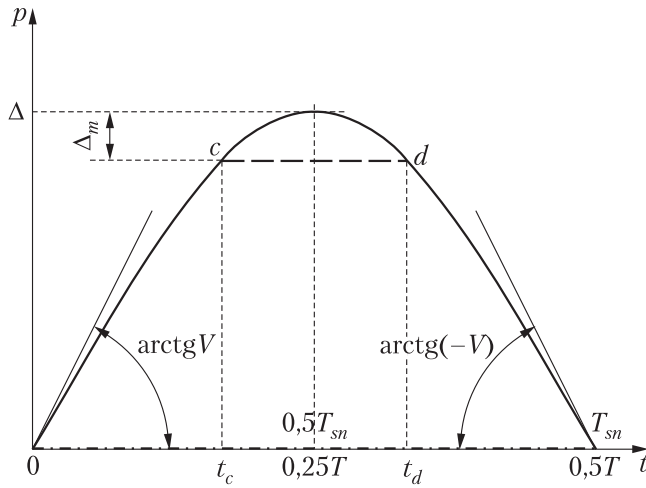


Рис. 1 Синусоидальная траектория наибольшего отклонения

В статье [18] дана формулировка аналога теоремы В. А. Котельникова [4] для случая восстановления сигнала $p(t)$ при помощи линейной аппроксимации по выборкам, следующим через интервал Найквиста $1/2f$.

Теорема 1. Любую функцию $p(t)$, состоящую из гармоник частотой от 0 до f , можно восстановить при помощи линейной аппроксимации с максимальной погрешностью Δ , равной амплитуде гармоники частоты f при помощи чисел, следующих друг за другом через $1/2f$ секунд.

На основании (5), рис. 1 и теоремы 1 запишем выражение для определения постоянного максимального допустимого шага дискретизации сигнала по времени T_{sn} :

$$T_{sn} = \frac{\pi}{\omega} = \frac{1}{2f}, \quad (6)$$

где $1/2f$ — интервал Найквиста [7, с. 437].

Определим шаг дискретизации времени T_{sn} сигнала, или синусоидальный КНО, через максимальные параметры сигнала, которые измерены при исследовании сигнала $p(t)$. Для этого найдем ряд производных от выражения (5) и приравняем их соответствующим максимальным значениям параметров сигнала $p(t)$ из выражения (1):

$$\begin{aligned} p^{(0)}(t) &= \Delta \sin \omega t, & \max p^{(0)}(t = 0,5T_s) &= \Delta & \text{или } \Delta &= \Delta; \\ p^{(1)}(t) &= \Delta \omega \cos \omega t, & \max p^{(1)}(t = 0) &= \Delta \omega = V & \text{или } \omega &= \frac{V}{\Delta}; \end{aligned}$$

$$p^{(2)}(t) = -\Delta\omega^2 \sin \omega t, \quad \max p^{(2)}(t = 0,5T_s) = -\Delta\omega^2 = -a \quad \text{или} \quad \omega = \sqrt{\frac{a}{\Delta}};$$

$$p^{(3)}(t) = -\Delta\omega^3 \cos \omega t, \quad \max p^{(3)}(t = 0) = -\Delta\omega^3 = -R \quad \text{или} \quad \omega = \sqrt[3]{\frac{R}{\Delta}}.$$

В электромеханических СУРВ затруднительно определить величину частоты граничной гармоники в спектре сигнала (или процесса) и сам спектр не определен. При этом величина интервала равномерной дискретизации времени процесса (6) зависит от максимальных параметров аппроксимируемого процесса (сигнала) (1).

Используем метод математической индукции и запишем общую цепочку равенств для определения круговой частоты:

$$\omega = \frac{V}{\Delta} = \sqrt{\frac{a}{\Delta}} = \sqrt[3]{\frac{R}{\Delta}} = \dots = \sqrt[n]{\frac{\max p^{(n)}(t)}{\Delta}},$$

где (n) — порядок старшей производной по времени от синусоидальной траектории отклонения.

Подставим круговую частоту ω в (6) и запишем цепочку равенств:

$$T_{si} = \pi \left[\frac{\Delta}{\max p^{(i)}(t)} \right]^{1/i}, \quad (7)$$

где $1 \leq i \leq n$.

Цепочка равенств (7) имеет важное свойство.

Свойство 1. Члены цепочки равенств (7), содержащей $(n-1)$ звено, повторяются в цепочке равенств (7), содержащей n звеньев, так как величина n входит только в ограничительную часть (7).

На основании свойства 1 запишем равенство $(i-1)$ -го и i -го членов выражения (7):

$$\pi \left[\frac{\Delta}{\max p^{(i-1)}(t)} \right]^{1/(i-1)} = \pi \left[\frac{\Delta}{\max p^{(i)}(t)} \right]^{1/i},$$

которое возведем в степень $i(i-1)$. Найдем параметр Δ и подставим его в $(i-1)$ -й член выражения (7). После необходимых преобразований получим цепочку равенств:

$$T_{si} = \pi \frac{\max p^{(i-1)}(t)}{\max p^{(i)}(t)}, \quad (8)$$

где $1 \leq i \leq n$.

Пример 1. В выражении (8) зададим $n = 1$ и $i = 1$ и получим одиночное выражение для синусоидального КНО: $T_{s1} = \pi\Delta/V$, которое подобно параболическому КНО В. Н. Хлистунова: $T_{p1} = \Delta/V$, что указывает на схожесть этих двух типов КНО.

Пример 2. В выражении (8) зададим $n = 2$ и $i = 1, 2$ и получим цепочку равенств:

$$T_{s1} = \pi \frac{\Delta}{V} = T_{s2} = \pi \frac{V}{a}.$$

Рассмотрим равенство:

$$\pi \frac{\Delta}{V} = \pi \frac{V}{a},$$

после преобразований получим условие согласования максимальных параметров сигнала: $V^2 = \Delta a$, или условие справедливости равенства $T_{s1} = T_{s2}$.

Пример 3. В выражении (8) зададим $n = 3$ и $i = 1, 2, 3$ и получим цепочку равенств:

$$T_{s1} = \pi \frac{\Delta}{V} = T_{s2} = \pi \frac{V}{a} = T_{s3} = \pi \frac{a}{R}.$$

Рассмотрим равенства

$$\pi \frac{\Delta}{V} = \pi \frac{V}{a}; \quad \pi \frac{V}{a} = \pi \frac{a}{R},$$

после преобразований получим два условия согласования параметров сигнала: $V^2 = \Delta a$ и $a^2 = VR$. Произведение последних дает $Va = \Delta R$, или условие справедливости равенств $T_{s1} = T_{s2} = T_{s3}$.

Из рассмотренных примеров и выражения (8) следует общее выражение для $(n - 1)$ -го условия согласования параметров сигнала, выполнение которых является условием справедливости выражений (7) и (8):

$$\max p^{(i)}(t) = \left[\max p^{(i-1)}(t) \cdot \max p^{(i+1)}(t) \right]^{0,5}, \quad (9)$$

где $1 \leq i \leq n - 1$, $n \geq 2$.

Выражению (9) соответствует правило:

Максимальное значение i -й производной от синусоидальной траектории отклонения по времени равно среднему геометрическому максимальных значений $(i - 1)$ -й и $(i + 1)$ -й производной от этой синусоидальной траектории отклонения.

Теорема 1 и выражения (5), (8) и (9) подтверждают факт наличия математических связей всех параметров сигнала $p(t)$ (1).

Недостаток выражений (7)–(9) заключается в том, они задают цепочки равенств T_{si} , которые неудобны для использования, поэтому получим их запись в виде одиночных выражений. Неудобство заключается в том, что на практике цепочка равенств нарушается при подстановке в нее максимальных параметров сигнала (1) из-за возникновения погрешностей при измерении этих параметров при обследовании объекта управления СУРВ.

3 Преобразование цепочек равенств

К одиночному компактному выражению, или синусоидальному КНО, предъявим требование наличия в его правой части всех известных максимальных значений параметров сигнала (1) в степени ± 1 и отсутствие многократных повторов параметров.

Теорема 2. *Максимально допустимая величина интервала равномерной дискретизации сигнала $p(t)$ по времени T_{sn} (вместо цепочки равенств (8)) определяется выражением:*

$$T_{sn} = \frac{\pi}{n} \sum_{i=1}^n \frac{\max p^{(i-1)}(t)}{\max p^{(i)}(t)}, \quad (10)$$

где $\max p^{(0)}(t = 0,5T_{s1}) = \Delta$; $\max p^{(i)}(t) = [\max p^{(i-1)}(t) \cdot \max p^{(i+1)}(t)]^{0,5}$.

Доказательство. Запишем сумму n членов левой части выражения (8) (для $1 \leq i \leq n$):

$$\sum_{i=1}^n T_{si} = nT_{sn},$$

где индекс i поменяем на индекс n , так как в соответствии со свойством 1 и выражением (8) каждый его член $T_{si} = T_{sn}$. Найдем сумму n членов правой части выражения (8):

$$\pi \sum_{i=1}^n \frac{\max p^{(i-1)}(t)}{\max p^{(i)}(t)}.$$

В результате явного равенства двух сумм получим доказательство существования компактного выражения (10) для определения величины T_{sn} .

Используем методику доказательства теоремы 2 и запишем компактное выражение для цепочки равенств (7):

$$T_{sn} = \frac{\pi}{n} \sum_{i=1}^n \frac{\Delta^{1/i}}{\max p^{(i)}(t)}, \quad (11)$$

где $\max p^{(i)}(t) = [\max p^{(i-1)}(t) \cdot \max p^{(i+1)}(t)]^{0,5}$.

Выражение (11) в соответствии со свойством 1 эквивалентно выражению (10), но имеет недостаток в виде необходимости определения $(n - 1)$ корня из повторяющегося параметра Δ .

Пример 4. В выражении (10) зададим $n = 2$ и $i = 1, 2$ и получим цепочку равенств:

$$T_{s1} = \pi \frac{\Delta}{V} = T_{s2} = \pi \frac{V}{a}.$$

Запишем сумму:

$$T_{s1} + T_{s2} = \pi \left(\frac{\Delta}{V} + \frac{V}{a} \right).$$

Учтем доказательство теоремы 2 и результат примера 2. Получим запись синусоидального КНО с учетом известных максимальных значений параметров сигнала (1) в степени ± 1 и с однократным повтором параметров:

$$T_{s2} = \frac{\pi}{2} \left(\frac{\Delta}{V} + \frac{V}{a} \right),$$

где $V^2 = \Delta a$. Полученное выражение подобно (3).

Пример 5. В выражении (10) зададим $n = 3$ и $i = 1, 2, 3$ и получим цепочку равенств:

$$T_{s1} = \pi \frac{\Delta}{V} = T_{s2} = \pi \frac{V}{a} = T_{s3} = \pi \frac{a}{R}.$$

Запишем сумму:

$$T_{s1} + T_{s3} = \pi \left(\frac{\Delta}{V} + \frac{a}{R} \right).$$

Учтем доказательство теоремы 2 и результат примера 3. Получим запись синусоидального КНО с учетом известных максимальных значений параметров сигнала (1) в степени ± 1 и с однократным повтором параметров:

$$T_{s3} = \frac{\pi}{2} \left(\frac{\Delta}{V} + \frac{a}{R} \right),$$

где $Va = \Delta R$.

Пример 6. Из цепочки равенств примера 5 запишем произведение:

$$T_{s1}T_{s3} = \pi^2 \frac{\Delta}{V} \frac{q}{R}.$$

Учтем доказательство теоремы 2 и результат примера 3. Получим запись синусоидального КНО с учетом всех известных максимальных значений параметров сигнала (1) в степени $\pm 0,5$ и с однократным повтором параметров:

$$T_{s3} = \pi \sqrt{\frac{\Delta}{V} \frac{a}{R}},$$

где $Va = \Delta R$.

Недостаток приведенной теоремы 1 и выражений (10) и (11) заключается в том, что величина максимальной погрешности Δ (после восстановления сигнала $p(t)$ по выборкам при помощи ступенчатой и линейной аппроксимации) равна амплитуде гармоники граничной частоты f в спектре частот сигнала $p(t)$. Этот недостаток не позволяет обосновать или уточнить эмпирическое соотношение $T_{sn} \leq (1/2f)/(8 \div 10)$, которое явно приводит к уменьшению интервала дискретизации по сравнению с интервалом Найквиста ($1/2f$) в $(8 \div 10)$ раз.

Получим выражение, которое связывает между собой заданную наперед величину максимальной погрешности ($\Delta_m < \Delta$), меньшую амплитуды Δ гармоники частоты f , и интервал дискретизации T_{ms} , меньший интервала Найквиста ($1/2f$).

4 Линейная аппроксимация траектории отклонения сигнала

Для решения этой задачи осуществим равномерную дискретизацию времени внутри половины периода граничной гармоники в спектре сигнала.

К. Шеннон дает важный комментарий к теореме НКШ [7, с. 435]: «*Это общеизвестный в теории связи факт. Интуитивное подтверждение состоит в том, что если функция $p(t)$ не содержит частот выше f , то она не может существенно изменить свое значение за время меньшее, чем половина периода наивысшей частоты, т. е. $1/2f$* ».

Траектория синусоидального отклонения в соответствии с выражением (5) и рис. 1 описывает экстремальное поведение функции сигнала $p(t)$ внутри половины периода наивысшей частоты ($1/2f$) (чем пренебрегает К. Шеннон). Тем самым учитываются отличия СУРВ от систем связи и передачи звуковых сигналов.

4.1 Доказательство теоремы об уменьшении интервала Найквиста

Докажем теорему об уменьшении интервала Найквиста для аппроксимации траектории синусоидального отклонения (первой полуволны гармоники граничной частоты f в спектре) отрезками прямых линий в количестве m штук.

На рис. 1 показан средний из m штук аппроксимирующий отрезок $[c, d]$, параллельный оси абсцисс, проекция которого на ось абсцисс представляет собой отрезок $[t_c, t_d]$.

В соответствии с рис. 1 запишем выражение для максимальной относительной погрешности линейной аппроксимации $\delta_m = \Delta_m/\Delta$.

В соответствии с рис. 1 и (6) запишем выражение для определения величины интервала дискретизации времени:

$$T_{ms} = \frac{T_{sn}}{m} = \frac{1}{2fm}, \quad (12)$$

где m — коэффициент уменьшения интервала Найквиста.

Докажем теорему об уменьшении интервала Найквиста ($1/2f$).

Теорема 3. Любую функцию $p(t)$, состоящую из гармоник частотой от 0 до f , с известной амплитудой Δ гармоники f можно восстановить при помощи линейной аппроксимации с максимальной относительной погрешностью δ_m по выборкам с интервалом времени, в

$$m = \frac{\pi}{2 \arccos(1 - \delta_m)} \quad (13)$$

раз меньшим, чем интервал Найквиста ($T_{sn} = 1/2f$).

Доказательство. Воспользуемся рис. 1 и рассмотрим отрезок прямой $[c, d]$, который аппроксимирует траекторию синусоидального отклонения по двум соседним выборкам из нее в моменты времени t_c и t_d .

Определим момент времени t_c начала горизонтального участка аппроксимации (см. точку c на рис. 1) с учетом (12):

$$t_c = 0,5T_{sn} - 0,5T_{ms} = 0,5T_{sn} \left(1 - \frac{1}{m}\right). \quad (14)$$

В выражение (5) подставим t_c (14), учтем $\omega = \pi/T_{sn}$ из (6) и получим выражение горизонтального отрезка прямой, проходящей через точки c и d (см. отрезок $[c, d]$ на рис. 1):

$$p(t_c) = \Delta \sin \left[\frac{\pi}{2} \left(1 - \frac{1}{m}\right) \right] = \Delta \cos \left(\frac{\pi}{2m} \right).$$

Зависимость погрешности аппроксимации Δ_m от времени t определим как разность выражения (5) и выражения для $p(t_c)$:

$$\Delta_m(t) = \Delta \sin \omega t - \Delta \cos \left(\frac{\pi}{2m} \right). \quad (15)$$

Выражение для максимального значения относительной погрешности восстановления $\delta_m = \Delta_m/\Delta$ получим подстановкой $t = 0,5T_{sn}$ и $\omega = \pi/T_{sn}$ в (15):

$$\delta_m = \frac{\Delta_m}{\Delta} = 1 - \cos \left(\frac{\pi}{2m} \right). \quad (16)$$

Из выражения (16) получим число m в виде выражения (13). Число m (13) входит в выражение (12) и делит интервал Найквиста $T_s = 1/2f$ на m частей, что и требовалось доказать.

Пример 7. Для проверки выражения (13) зададим $\Delta_m = \Delta$ (или $\delta_m = 1$), получим $m = 1$, из интервала выборки $T_{ms} = T_{sn}/m$ (12) получим $T_{ms} = T_{sn}$, или интервал Найквиста ($T_{sn} = 1/2f$), как частный случай.

Зададим $\Delta_m = 0$ (или $\delta_m = 0$), получим $m = \infty$, а из интервала выборки $T_{ms} = T_{sn}/m$ (12) получим $T_{ms} = 0$.

4.2 Два эквивалентных варианта итоговых формул

Выражение (12) для определения равномерного интервала дискретизации T_{ms} сигнала $p(t)$ при условии использования линейной аппроксимации для восстановления по выборкам из $p(t)$ состоит из двойного равенства, поэтому можно записать два эквивалентных выражения.

Из первого равенства (12) и выражений (13) и (10) запишем:

$$T_{ms} = \frac{T_{sn}}{m} = \frac{2}{n} [\arccos(1 - \delta_m)] \sum_{i=1}^n \frac{\max p^{(i-1)}(t)}{\max p^{(i)}(t)}. \quad (17)$$

Из второго равенства (12) и выражения (13) запишем:

$$T_{ms} = \frac{1}{2fm} = \frac{\arccos(1 - \delta_m)}{\pi f}. \quad (18)$$

В выражения (17) и (18) входят все параметры сигнала $p(t)$, как и в функциональную зависимость в общем виде (1), что показывает достижение цели исследования.

4.3 Исследование эмпирической формулы $T_{sn} \leq (1/2f)/(8 \div 10)$

По выражению (16) рассчитаны значения δ_m для $1 \leq m \leq 12$, сведены в таблицу и представлены в виде графика на рис. 2.

Из таблицы видно, что восстановление сигнала $p(t)$ при помощи линейной аппроксимации по ее выборкам с погрешностью $\delta_m \cdot 100\% = 1\%$ осуществимо при условии превышения граничной частоты f спектра частот сигнала $p(t)$ частотой выборки f_m в $22 \div 24$ раза, или при уменьшении интервала Найквиста в $11 \div 12$ раз, что существенно уточняет выражения, используемые в [20].

Ступенчатая аппроксимация сигнала $p(t)$ с указанной точностью требует увеличения частоты дискретизации в 2 раза по сравнению с частотами для линейной аппроксимации.

Эмпирическая формула $T_{sn} \leq (1/2f)/(8 \div 10)$ [20] позволяет осуществить линейную интерполяцию сигнала по выборкам из него с погрешностью $\delta_m \cdot 100\% \approx 1,9\% \div 1,2\%$ (см. строки 8–10 таблицы).

Значения функции $\delta_m = F(m)$	
m	δ_m
1	1,0000
2	0,2929
3	0,1340
4	0,0761
5	0,0489
6	0,0341
7	0,0251
8	0,0192
9	0,0152
10	0,0123
11	0,0102
12	0,0086

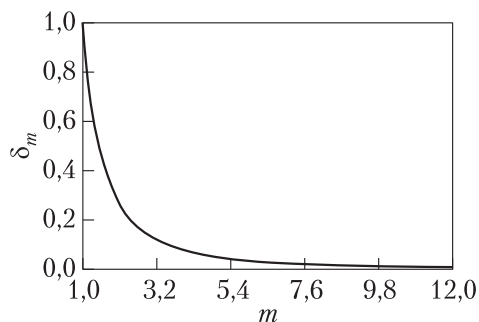


Рис. 2 График зависимости погрешности δ_m восстановления сигнала $p(t)$ по выборкам от коэффициента уменьшения интервала Найквиста m

5 Заключение

В СУРВ применяют в основном ступенчатую или линейную аппроксимацию сигнала по равномерным выборкам из него. Это ограничивает применение в СУРВ теоремы НКШ, по которой сигнал восстанавливают при помощи ряда В. А. Котельникова, что приводит к большим задержкам восстановления во времени.

Показано, что замена аппроксимирующего ряда Котельникова на аппроксимацию ступеньками или отрезками прямых приводит к абсолютной погрешности аппроксимации выборок, равной амплитуде граничной гармоники в спектре сигнала. Рассмотрены два пути уменьшения этой погрешности и больших задержек во времени:

- (1) использование общеизвестной эмпирической формулы уменьшает интервал Найквиста в 8–10 раз при использовании ступенчатой или линейной аппроксимации сигнала. При этом погрешность восстановления сигнала по выборкам ступеньками достигает 2,6%–3,8%;
- (2) предложена математическая модель: траектория наибольшего отклонения в виде полуволны граничной гармоники в спектре сигнала. Получен синусоидальный критерий наибольшего отклонения в виде выражений для нахождения величины интервала равномерной дискретизации времени сигнала, который подобен авторскому параболическому критерию наибольшего отклонения [14, 15].

Доказана теорема об уменьшении интервала Найквиста, которая позволяет определять величину интервала равномерной дискретизации времени сигнала через процедуру дискретизации экстремальной модели сигнала в виде полуволны граничной гармоники в спектре сигнала. При этом величина интервала равномер-

ной дискретизации времени сигнала (18) зависит от частоты граничной гармоники в спектре сигнала и необходимой величины относительной погрешности ступенчатой или линейной аппроксимации сигнала по выборкам.

В электромеханических СУРВ сложно определить величину частоты граничной гармоники в спектре сигнала (или процесса), да и сам спектр не определен. Для этого случая предлагается использовать (17) вместо (18). При этом величина интервала равномерной дискретизации времени сигнала (17) зависит от максимальных параметров аппроксимируемого процесса и необходимой величины относительной погрешности ступенчатой аппроксимации или линейной аппроксимации процесса по выборкам.

Литература

1. *Вторин В. А.* Основы АСУ ТП. — СПб.: ГЛТА, 2006. 152 с.
2. Контрольно-измерительное оборудование R&S 2019. — Каталоги Rohde & Schwarz, 2019. 256 с. https://www.rohde-schwarz.com/ru/top_topics_russia/catalogs/catalogs_231625.html.
3. *Nyquist H.* Certain topics in telegraph transmission theory // T. Am. Inst. Electr. Eng., 1928. Vol. 47. Iss. 2. P. 617–644.
4. *Котельников В. А.* О пропускной способности «эфира» и проволоки в электросвязи // Материалы к I Всесоюзному съезду по вопросам технической реконструкции дела связи и развития слаботочной промышленности. — М.: Редакция управления связи РККА, 1933. С. 4–5.
5. *Whittaker J.* Interpolatory function theory. — Cambridge tracts in mathematics and mathematical physics ser. — Cambridge: Cambridge University Press, 1935. No. 33. Ch. IV.
6. *Shannon C.* Communication in the presence of noise // P. IRE, 1949. Vol. 37. No. 1. P. 10–21.
7. *Шеннон К.* Работы по теории информации и кибернетике / Пер. с англ. — М.: ИЛ, 1963. 830 с.
8. *Lee I. I. H., Bang H. A.* A cooperative line-of-sight guidance law for a three-dimensional phantom track generation using unmanned aerial vehicles // J. Aerospace Eng., 2016. Vol. 227. No. 6. P. 897–915.
9. *Apte S. D.* Signals and systems: Principles and applications. — Cambridge: Cambridge University Press, 2016. 768 p.
10. *Liu X., Li D.* Analysis of cooperative jamming against pulse compression radar based on CFAR // EURASIP J. Adv. Sig. Pr., 2018. Art. No. 2018:69. P. 1–12. doi: 10.1186/s13634-018-0592-2.
11. *Хлистунов В. Н.* О применении теоремы Котельникова к дискретной измерительной технике // Измерительная техника, 1961. № 3. С. 15–28.
12. *Хлистунов В. Н.* О погрешности аппроксимации дискретных методов измерения // Приборостроение, 1960. № 5. С. 3–5.
13. *Хлистунов В. Н.* Основы цифровой электроизмерительной техники и цифровые преобразователи. — М.-Л.: Энергия, 1966. 345 с.

14. *Майоров Б. Г.* Обобщенный критерий наибольшего отклонения входных сигналов систем управления // Автоматика и телемеханика, 2005. № 10. С. 148–155.
15. *Майоров Б. Г.* Исследование теоремы Котельникова и применение результатов для определения величины дискретизации по времени входных сигналов систем управления // Высокопроизводительные вычислительные системы и микропроцессоры: Сб. научных трудов ИМВС РАН. — М.: ИМВС РАН, 2004. Вып. 7. С. 76–82.
16. *Майоров Б. Г.* Модель производной от траектории наибольшего отклонения сигнала (процесса) в виде степенной функции // Вопросы радиоэлектроники. Сер. СОИУ, 2016. № 12. С. 38–43.
17. *Майоров Б. Г.* Критерий наибольшего отклонения сигнала и определение интервала дискретизации времени // Вопросы радиоэлектроники, 2017. № 12. С. 111–119.
18. *Майоров Б. Г.* Дискретизация сигналов и процессов по теореме Котельникова с учетом максимальных значений их производных по времени // Успехи современной радиоэлектроники, 2010. № 1. С. 56–63.
19. *Майоров Б. Г.* Определение минимального интервала дискретизации времени при ступенчатой и линейной аппроксимации сигнала // Успехи современной радиоэлектроники, 2011. № 3. С. 71–75.
20. Микросхемы АЦП и ЦАП: Справочник. — М.: Додека-XXI, 2005. 512 с.

Поступила в редакцию 25.12.17

RECOVERY OF SIGNAL IN SYSTEMS OF REAL TIME ON UNIFORM SELECTIONS WITH REDUCTION OF NYQUIST INTERVAL

B. G. Mayorov

NPP “Rubin,” 2 Baydukova Str., Penza 440000, Russian Federation

Abstract: The mathematical behavior model of signal between two next selections of it is suggested. New general formulas for finding the sinusoidal criterion of the greatest deviation are received. The theorem of reduction of Nyquist interval which connects the criterion of the greatest deviation and the size of constant interval of sampling is proved. The results allow to determine the size of this interval reasonably.

Keywords: systems of real time; sinusoidal criterion of the greatest deviation of signal; step or linear approximation of signal on selections; admissible error of approximation; constant interval of sampling of time

DOI: 10.14357/08696527190209

References

1. Vtyurin, V. A. 2006. *Osnovy avtomatizirovannykh sistem upravleniya tekhnologicheskimi protsessami* [Fundamentals of automated process control systems]. St. Petersburg: GLTA Pubs. 152 p.

2. Rohde & Schwarz. 2019. Kontrol'no-izmeritel'noe oborudovanie R&S 2019: Katalog [Control and measuring equipment R&S 2019: Catalogue]. 256 p. Available at: https://www.rohde-schwarz.com/ru/top_topics_russia/catalogs/catalogs_231625.html (accessed April 10, 2019).
3. Nyquist, H. 1928. Certain topics in telegraph transmission theory. *T. Am. Inst. Electr. Eng.* 47(2):617–644.
4. Kotel'nikov, V. A. 1933. O propusknoy sposobnosti “efira” i provoloki v elektrosvyazi [About the flow capacity of “ether” and wire in telecommunication]. *Mat-ly k I Vsesoyuznomu s'ezdu po voprosam tekhnicheskoy rekonstruktsii dela svyazi i razvitiya slabotochnoy promyshlennosti* [Materials to the 1st All-Union Congress on Technical Reconstruction of Communication and Development of Low-Current Industry]. Moscow: Redaktsiya upravleniya svyazi RKKA [Editorial board of Communication Department of the Red Army]. 4–5.
5. Whittaker, J. 1935. *Interpolatory function theory*. Cambridge tracts in mathematics and mathematical physics ser. Cambridge: Cambridge University Press. No. 33. Ch. IV.
6. Shannon, C. 1949. Communication in the presence of noise. *P. IRE* 37(1):10–21.
7. Shannon, C. 1963. *Raboty po teorii informatsii i kibernetike* [Works on the theory of information and cybernetics]. Moscow: IL. 830 p.
8. Lee, I. I. H., and H. A. Bang. 2016. A cooperative line-of-sight guidance law for a three-dimensional phantom track generation using unmanned aerial vehicles. *J. Aerospace Eng.* 227(6):897–915.
9. Apte, S. D. 2016. *Signals and systems: Principles and applications*. Cambridge: Cambridge University Press. 768 p.
10. Liu, X., and D. Li. 2018. Analysis of cooperative jamming against pulse compression radar based on CFAR. *EURASIP J. Adv. Sig. Pr.* Art. No. 2018:69. P. 1–12. doi: 10.1186/s13634-018-0592-2.
11. Khlistunov, V. N. 1961. Applicability of Kotel'nikov's theorem to discrete measurement techniques. *Meas. Tech.* 4(3):203–207.
12. Khlistunov, V. N. 1960. O pogreshnosti approksimatsii diskretnykh metodov izmereniya [About error of approximation of discrete methods of measurement]. *Priborostroenie* [Instrument Making] 5:3–5.
13. Khlistunov, V. N. 1966. *Osnovy tsifrovoy elektroizmeritel'noy tekhniki i tsifrovye preobrazovateli* [Fundamentals of digital electric equipment and digitizers]. Moscow–Leningrad: Energiya. 345 p.
14. Mayorov, B. G. 2005. The generalized maximum deviation for input signals of control systems. *Automat. Rem. Contr.* 66(10):1666–1672.
15. Mayorov, B. G. 2004. Issledovanie teoremy Kotel'nikova i primeneniye rezul'tatov dlya opredeleniya velichiny diskretizatsii po vremeni vkhodnykh signalov sistem upravleniya [Research of the Kotel'nikov's theorem and application of results for determination of size of time sampling of control systems input signals] *Vysokoproizvoditel'nye vychislitel'nye sistemy i mikroprotsessory. Sb. nauchnykh trudov IMVS RAN* [High-performance computer systems and microprocessors. Scientific Works of IMVS of the Russian Academy of Sciences]. Moscow: IMVS RAS Pubs. 7:76–82.
16. Mayorov, B. G. 2016. Model' proizvodnoy ot traektorii naibol'shego otkloneniya signala (protsessa) v vide stepennoy funktsii [Model derivative of trajectory of the greatest deviation of signal (process) in the form of power function]. *Voprosy radioelektroniki*.

- SOIU ser.* [Radioelectronics Problems. Information display systems and control of special equipment ser.] 12:38–43.
17. Mayorov, B. G. 2017. Kriteriy naibol'shego otkloneniya signala i opredelenie intervala diskretizatsii vremeni [Criterion of the greatest deviation of signal and definition of interval of sampling of time]. *Voprosy radioelektroniki* [Radioelectronics Problems] 12:111–119.
 18. Majorov, B. G. 2010. Diskretizatsiya signalov i protsessov po teoreme Kotel'nikova s uchetom maksimal'nykh znacheniy ikh proizvodnykh po vremeni [Signals and processes sampling under the theorem of Kotelnikov subject to the maximum values of their derivatives on time]. *Achievements of Modern Radioelectronics* 1:56–63.
 19. Majorov, B. G. 2011. Opredelenie minimal'nogo intervala diskretizatsii vremeni pri stupenchatoy i lineynoy approksimatsii signala [Conditions definition of minimum range of time sampling interval acquiring at stepwise and linear approximation of the signal]. *Achievements of Modern Radioelectronics* 3:71–75.
 20. *Mikroskhemy ACP i CAP: Spravochnik* [ADC and DAC microchips: Handbook]. 2005. Moscow: Dodeka-XXI Publs. 512 p.

Received December 25, 2017

Contributor

Mayorov Boris G. (b. 1947) — Candidate of Science (PhD) in technology, associate professor, leading scientist, NPP “Rubin,” 2 Baydukova Str., Penza 440000, Russian Federation; b.g.m.1947@rambler.ru

МЕТАОБЪЕКТНЫЙ ПОДХОД К МОДЕЛИРОВАНИЮ БИЗНЕС-ПРОЦЕССОВ ПРЕДПРИЯТИЯ В РАМКАХ ЕДИНОЙ ERP-СИСТЕМЫ

С. И. Спивак¹, Н. Д. Морозкин², Л. А. Лукьянов³

Аннотация: Объединение бизнес-процессов предприятия внутри единой информационной системы носит массовый характер. Оптимизация деятельности предприятий в условиях гибкого рынка и стремительного технического прогресса невозможна без использования единой информационной системы. Существует огромное число ERP (Enterprise Resource Planning) систем, призванных решить в полной мере задачу автоматизации всех процессов на предприятии. Однако ни одна из них не способна в полной мере учитывать существующие механизмы автоматизации на предприятии. Эти механизмы обычно реализованы не только в виде документов с техническим заданием к тому или иному бизнес-процессу, но и как программные решения, обычно разработанные с помощью достаточно устаревших технологий. В данной работе рассматривается ERP-система собственной разработки, в которую заложен абсолютно новый подход внедрения, разработки и организации единого информационного пространства для всех подразделений предприятия. Предлагается метод внедрения, основанный на последовательном замещении существующей реализации бизнес-процессов подсистемами внедряемой платформы. Рассматривается метаобъектный подход к реализации бизнес-процессов, который может быть легко использован специалистами предметных областей, работающими непосредственно на предприятии. Предлагаемая ERP-система не изменяет логику существующих бизнес-процессов, но обеспечивает их интеграцию, прозрачность и единую модель данных.

Ключевые слова: MRP; MRP II; MES; CRUD; единое информационное пространство; планирование ресурсов; моделирование информационных систем; платформа; архитектура ERP

DOI: 10.14357/08696527190210

1 Введение

Как известно, существует огромное число информационных систем, призванных регулировать деятельность подразделений на предприятиях различной отраслевой принадлежности: промышленной, торговой, транспортной, финан-

¹Башкирский государственный университет, semen.spivak@mail.ru

²Башкирский государственный университет, rector@bsunet.ru

³Башкирский государственный университет, leonlukyanov@gmail.com

совой. Такие информационные системы помимо основной функциональности оптимизируют деятельность того или иного подразделения на предприятии. Очевидно, что отделы в организации постоянно взаимодействуют между собой в процессе создания товаров и услуг при реализации заказа. Именно по этой причине необходима интеграция информационных систем (особенно на крупном предприятии) для обеспечения прозрачности всех бизнес-процессов и оптимизации деятельности всей организации в целом [1].

В настоящее время понятие интегрированных подсистем заложено в маркетинговом термине «ERP-система». Широко распространено следующее определение этого термина: ERP — «это организационная стратегия интеграции производства и операций, управления трудовыми ресурсами, финансового менеджмента и управления активами, ориентированная на непрерывную балансировку и оптимизацию ресурсов предприятия посредством специализированного интегрированного пакета прикладного программного обеспечения, обеспечивающего общую модель данных и процессов для всех сфер деятельности» [2]. С общей точки зрения под ERP-системой понимают интегрированный пакет программного обеспечения, включающий в себя совокупность модулей (подсистем), задача которых состоит в математическом и информационном обеспечении определенного бизнес-процесса на предприятии. Исторически сложилось, что ERP-системы успешно решают задачу интеграции и оптимизации всех бизнес-процессов организации. Первые ERP-системы представляли собой системы планирования материалов. На сегодняшний день уже разработаны ERP-системы третьего поколения, организующие веб-сервисы для мобильных устройств [2]. Например, такая подсистема может автоматизировать работу отдела маркетинга, финансового отдела или отдела материально-технического обеспечения, особенно на предприятиях производственного типа. Цель ERP-системы состоит в автоматизации всех информационных процессов как внутри самой организации, так и внутри ее контрагентов. К недостаткам современных ERP-систем можно отнести сложность их внедрения и высокие финансовые затраты как при покупке, так и в процессе работы с данной системой [3]. А к достоинствам ERP-систем следует отнести прозрачность всех бизнес-процессов, простоту в управлении и устранение человеческого фактора.

Стоит отметить, что обычно при внедрении ERP-систем не производится интеграция с существующей системой; таким образом, весь накопленный опыт предприятия используется только на этапе внедрения новой системы посредством словесного или документального описания сотрудниками особенностей сложившихся на предприятии бизнес-процессов поставщикам данной ERP-системы. Можно наблюдать ситуации, когда бизнес-деятельность подстраивается под ERP-систему. Такой подход можно считать вполне адекватным и логичным в случае, когда организация небольшая или находится в процессе становления. Однако для крупных предприятий важно сохранить сложившиеся особенности деятельности всех подразделений, с тем чтобы не потерять накопленный опыт в вопросах управления, планирования и кооперации.

Авторы предлагают концептуально новый подход к моделированию бизнес-процессов любого предприятия с использованием инструмента собственной разработки — MOPEX (Meta Object Process EXecution). Моделирование включает как описательную, так и практическую части. Инструмент имеет гибкую архитектуру, множество вспомогательных сервисов и огромный потенциал к расширению и оптимизации ядра новыми возможностями. Математическая составляющая сервисов для решения задач планирования и управления представлена в виде подключаемых модулей различной структуры. Главная особенность разработанной платформы заключается в простоте моделирования новых, а также переработки существующих информационных процессов для автоматизации деятельности подразделений предприятия.

2 Архитектура MOPEX

Архитектура платформы MOPEX, представленная на рис. 1, имеет достаточно сложную структуру, обусловленную требованиями гибкости разработки,

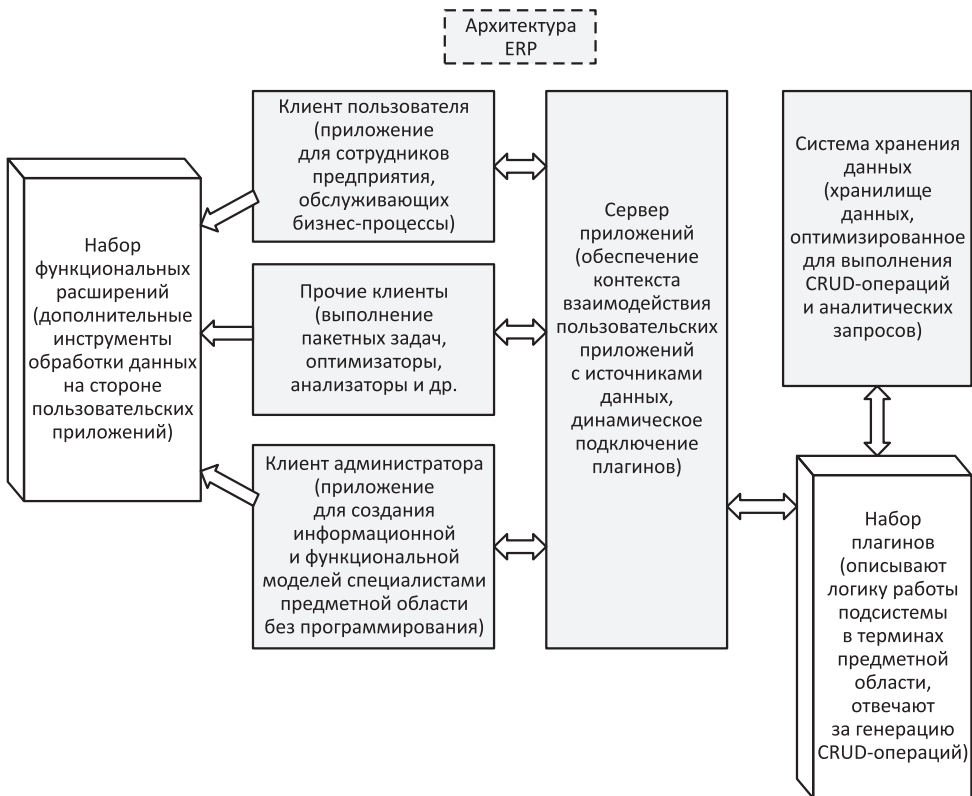


Рис. 1 Архитектура платформы MOPEX

безопасности и общей производительности. Взаимодействие основных компонентов архитектуры — сервера приложений, клиента пользователя, клиента администратора — осуществляется по сети согласно трехзвенной схеме [4]. Данная платформа позиционируется в качестве высокопроизводительной расширяемой ERP-системы, которая предоставляет инструмент описания бизнес-процессов. Стандартные решения, включаемые большинством ERP-систем, предлагаемых на рынке ИТ-индустрии, в данной платформе отсутствуют. Связано это с тем фактом, что предприятие при внедрении данной платформы должно самостоятельно моделировать бизнес-процессы по метаобъектной модели разработки.

Особенность архитектуры платформы заключается в наличии стека плагинов по управлению бизнес-логикой и стека расширений для графической обработки информации на стороне пользовательских приложений. Плагины, как и расширения, независимы друг от друга и могут подключаться в зависимости от роли пользователя системы. Например, если пользователь является разработчиком предметной области, то ему должен быть доступен инструмент построения отчетов и, возможно, инструмент создания пользовательского интерфейса. В противоположность этому, пользователь, выполняющий роль ответственного за ведение задачи, должен иметь доступ только к инструменту отображения отчетов. Набор функциональных возможностей строго определяется ролями пользователей и описывается дополнительными атрибутами (разрешен доступ, запись, выполнение, др.), позволяющими тонко настроить способ доступа к данным.

3 Метаобъектная модель разработки

Прежде чем перейти к рассмотрению архитектуры платформы, необходимо описать процесс разработки подсистемы для решения некоторой задачи предметной области. Следует отличать модуль ERP-системы от «подсистемы» предметной области. Модуль ERP-системы — это автономный программный продукт, предлагающий стандартное решение для автоматизации работы отдельного подразделения предприятия. Например, модулем ERP-системы является программа внутрицехового планирования [5]. В противоположность модулю, подсистема предметной области — это программный продукт, автоматизирующий некоторый бизнес-процесс. К подсистемам, например, можно отнести «Ведение маршрутов деталей и сборочных единиц (ДСЕ) внутри цеха», «Управление ДСЕ».

Для реализации подсистем на платформе MOPEX предлагается метаобъектная модель разработки, представленная на рис. 2. Данная модель представляет собой модификацию широко известной итеративной модели разработки [6]. Метаобъектная модель разработки осуществляется специалистами предметной области и включает в себя следующие этапы.

1. *Разработка информационной модели.* Формируется техническое задание (ТЗ) на разрабатываемую подсистему сотрудниками отдела. Осуществляется работа по описанию структуры данных разрабатываемой подсистемы.



Рис. 2 Метаобъектная модель разработки

Структура данных — это набор связанных таблиц, который регламентирует способ хранения данных подсистемы. Необходимо отметить, что на этапе создания информационной модели следует учитывать, что одни и те же данные могут быть использованы в разных подсистемах. Например, при создании подсистемы «Ведение маршрутов ДСЕ внутри цеха» структура данных включает в себя данные по деталям, операциям и оборудованию.

2. *Разработка функциональной модели.* Определяются функции разрабатываемой подсистемы, строятся диаграммы USE CASE, IDEF0, DFD. Например, в подсистеме «Ведение маршрутов ДСЕ внутри цеха» можно выделить следующие функции: просмотр справочника операций цеха, создание маршрута движения ДСЕ, корректировка загруженности оборудования и т. д.
3. *Реализация информационной модели.* Информационная модель создается в системе MOPEX в виде метаобъектов. Создание осуществляется с использованием клиента «Администратор» платформы MOPEX. Метаобъект —

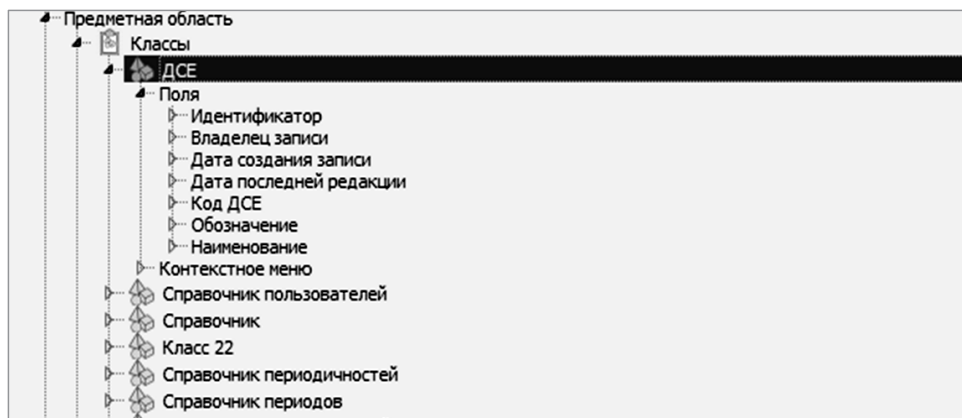


Рис. 3 Пример класса информационной модели

это класс и его окружение. Классом может выступать любой объект подсистемы, наделенный атрибутами. Окружение класса характеризует связи с другими классами подсистемы и иногда определяет встроенные функции, выполняемые при взаимодействии с данным классом. На рис. 3 приведен пример класса ДСЕ, используемого в подсистеме «Управление ДСЕ».

4. *Реализация функциональной модели.* На основе функциональной модели производится описание взаимодействия метаобъектов средствами платформы МОРЕХ, а именно выполняются следующие последовательные шаги:
 - (а) создание ролей подсистемы в соответствии с диаграммой USE CASE;
 - (б) создание пользователей подсистемы;
 - (в) назначение пользователям определенных ролей;
 - (г) создание пунктов и подпунктов меню подсистемы, отражающих основные функции подсистемы;
 - (д) определение контекстного меню, обеспечивающего выполнение некоторой функции, зависящей от текущего набора данных.

На рис. 4 приведен пример реализации функциональной модели подсистемы «Управление ДСЕ».

5. *Создание окружения подсистемы.* Производится создание дополнительных элементов разрабатываемой подсистемы: отчетных форм, экранных форм, функций по умолчанию, пакетных задач. Стоит отметить, что данные элементы не всегда являются обязательными в той или иной подсистеме. Например, в подсистеме «Управление ДСЕ» отчетные и экранные формы отсутствуют.

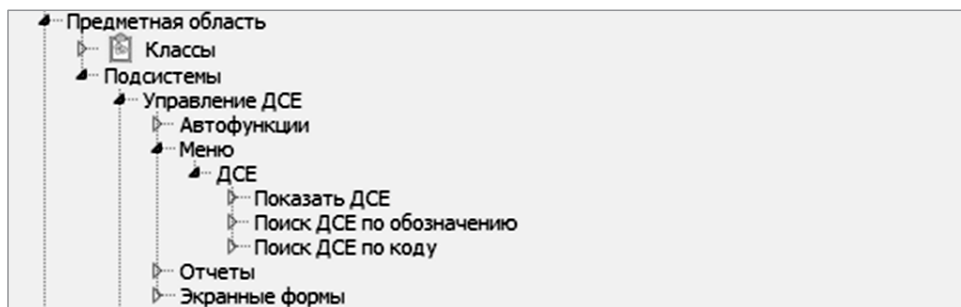


Рис. 4 Пример описания функциональной модели

6. *Апробация подсистемы.* Осуществляется тестирование подсистемы специалистами подразделения, для которого разрабатывалась данная подсистема. При выявлении недоработок подсистемы производится уточнение ошибок в форме дополнения к основному ТЗ. Затем проводится доработка подсистемы, начиная с первого этапа метаобъектной модели разработки.

4 Заключение

В работе были исследованы существующие на рынке платформы класса ERP; выявлены недостатки предлагаемых решений; предложена актуальная архитектура платформы для построения бизнес-приложений; предложена модель разработки, покрывающая все этапы создания бизнес-приложений — от проектирования до внедрения в эксплуатацию. Актуальность и новизна платформы среди всех существующих решений подкрепляется методом представления объекта информации и его контекста взаимодействия в виде метаобъекта, который при необходимости может быть использован простым внедрением в другом бизнес-процессе без копирования или преобразования со стороны разработчика задачи предметной области.

Литература

1. Григорьев А. А., Титов В. А. Характеристика, структура, организация систем управления ERP, ERP II и ERP III // *Фундаментальные исследования*, 2017. № 2. С. 48–51.
2. Abdullah A. M. A. Evolution of enterprise resource planning // *Excel J. Eng. Technology Management Sci.*, 2017. Vol. 1. No. 11. P. 1–6.
3. Елисеев В. Г., Фисенко Н. И. Анализ затрат на внедрение ERP систем // *Научная сессия МИФИ-2007: Сборник научных трудов.* — М.: МИФИ, 2007. Т. 2. С. 48–49.

4. Шмидт И. А. Архитектура платформы для разработки бизнес-приложений // Современные проблемы науки и образования, 2014. № 6. С. 348.
5. Васильева Ю. В. Планирование производственных ресурсов (MRP II) и планирование потребностей предприятия (ERP) // Современные тенденции в экономике и управлении: новый взгляд. — Новосибирск: Центр развития научного сотрудничества, 2015. С. 46–50.
6. Терехов А. Н. Технология программирования. — М.: НОУ ИНТУИТ, 2006. 77 с.

Поступила в редакцию 02.05.18

METAOBJECT APPROACH FOR MODELING ENTERPRISE BUSINESS PROCESSES INSIDE SINGLE ERP SYSTEM

S. I. Spivak, L. A. Lukyanov, and N. D. Morozkin

Bashkir State University, 32 Validy Str., Ufa 450076, Russian Federation

Abstract: Nowadays, combining enterprise business processes within a single information system has a widespread trend. Work optimization of enterprises in condition of flexible market and rapid technical progress is impossible without using a single information system. There are many ERP (Enterprise Resource Planning) systems designed to solve the main problem of automation of all enterprise business processes. However, neither one of them is able to take into account existing automation mechanisms of the enterprise. These mechanisms are often designed not only using documents with technical order for any type of business process but also using software developed with outdated technics. In this paper, the authors consider their own ERP system, which contains a completely new approach to the application, development, and organization of a single information space for all business units. The proposed approach is based on sequential replacement of the existing implementation of business processes by platform subsystems. The metaobject approach to the implementation of business processes is considered, which is applied by specialists of subject areas working directly in the enterprise. The proposed ERP system does not change the logic of existing business processes but ensures their integration, transparency, and common data model.

Keywords: MRP; MRPII; ERP; MES; CRUD; common information space; resource planning; information system modeling; platform; software architecture

DOI: 10.14357/08696527190210

References

1. Grigorev, A. A., and V. A. Titov. 2017. Kharakteristika, struktura, organizatsiya sistem upravleniya ERP, ERP II i ERP III [Characteristics, structure, organization systems ERP-management, ERP II and ERP III]. *Fundamental Research* 2:48–51.

2. Abdullah, A. M. A. 2017. Evolution of enterprise resource planning. *Excel J. Eng. Technology Management Sci.* 1(11):1–6.
3. Eliseev, V. G., and N. I. Fisenko. 2017. Analiz zatrat na vnedrenie ERP sistem [Cost analysis for the introduction of ERP systems]. *Nauchnaya sessiya MIFI-2007: Sbornik nauchnykh trudov* [Scientific session MEFPhI-2007: Collection of scientific papers]. 2:48–49.
4. Shmidt, I. A. 2014. Arkhitektura platformy dlya razrabotki biznes-prilozheniy [The architecture of a development platform for business applications]. *Modern Problems of Science and Education* 6:348.
5. Vasil'eva, Yu. V. 2015. Planirovanie proizvodstvennykh resursov (MRP II) i planirovanie potrebnostey predpriyatiya (ERP) [Resource planning (MRP II) and enterprise needs planning (ERP)]. *Sovremennye tendentsii v ekonomike i upravlenii: novyy vzglyad* [Current trends in economics and management: A modern point]. Novosibirsk: Center for the Development of Scientific Cooperation. 46–50.
6. Terekhov, A. N. 2006. *Tekhnologiya programmirovaniya* [Programming technology]. Moscow: NOU INTUIT. 77 p.

Received May 2, 2018

Contributors

Spivak Semen I. (b. 1945) — Doctor of Science in physics and mathematics, professor, Head of Department, Bashkir State University, 32 Validy Str., Ufa 450076, Russian Federation; semen.spivak@mail.ru

Morozkin Nikolay D. (b. 1953) — Doctor of Science in physics and mathematics, Chancellor, Bashkir State University, 32 Validy Str., Ufa 450076, Russian Federation; rector@bsunet.ru

Lukyanov Leonid A. (b. 1992) — PhD student, Bashkir State University, 32 Validy Str., Ufa 450076, Russian Federation; leonlukyanov@gmail.com

МЕТОД ПОИСКА ВЫПАДАЮЩИХ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ ПАРАМЕТРОВ НЕУСТОЙЧИВОСТИ ОБУЧЕНИЯ*

И. С. Ожерельев¹, О. В. Сенько², Н. Н. Киселёва³

Аннотация: Рассматривается метод поиска выпадающих объектов (ВО) в задачах распознавания, т. е. объектов, описания которых значительно отличаются от описаний объектов своего класса. Метод основан на одновременном использовании величин оценок принадлежности объекта к классам и интегральных искажений, вносимых объектом в формируемый в результате обучения алгоритм распознавания. Возможность использования разработанного метода при высокой размерности данных была продемонстрирована на задаче прогнозирования возможности образования неорганических соединений состава $A^{+3}B^{+3}C^{+2}O_4$ при обычных условиях. Метод может быть использован с целью выявления ошибочных наблюдений для повышения качества обучающей информации при решении задач распознавания.

Ключевые слова: выпадающие объекты; базы данных; распознавание; неустойчивость обучения; неорганические соединения

DOI: 10.14357/08696527190211

1 Введение

Под выпадающими объектами обычно понимаются объекты, описание которых заметно отклоняется от основной закономерности в данных. При этом отклонение должно быть настолько значительным, что оно не могло бы быть объяснено простой случайностью и требовало бы дополнительных предположений о механизме возникновения объекта. Данное определение фактически соответствует определению, приведенному в книге [1]. Выпадающие объекты достаточно часто встречаются в базах данных. Иногда выпадающие наблюдения связаны с какими-либо неизвестными особенностями исследуемого процесса или уникальностью объектов. Идентификация таких ВО может привести к получению новых знаний об исследуемом явлении. Однако чаще всего ВО возникают в связи с различного рода ошибками, включая экспериментальные ошибки (в том числе

* Работа выполнена при частичной поддержке РФФИ (проект 17-01-00634).

¹ Московский государственный университет им. М. В. Ломоносова, факультет вычислительной математики и кибернетики, ilya365365@gmail.com

² Федеральный исследовательский центр «Информатика и управление» Российской академии наук, senkoov@mail.ru

³ Институт металлургии и материаловедения им. А. А. Байкова Российской академии наук, kis@imet.ac.ru

и ошибки в определении класса объектов или ошибки в значениях признаков), ошибки при занесении информации в базу и др. При построении прогностических моделей с помощью методов машинного обучения ВО, связанные с ошибками, могут заметно повышать неустойчивость обучения и, как следствие, снижать обобщающую способность полученной модели. Естественно, что такие ВО следует удалять из обучающей выборки после их идентификации. Следует отметить, что экспертная оценка правильности часто противоречивых экспериментальных данных разных исследователей остается наиболее сложной и плохо формализуемой задачей. Оценка правильности классификации десятков тысяч веществ в базах данных по свойствам веществ и материалов — задача крайне дорогостоящая и практически нереальная. Даже экспертная оценка данных о нескольких сотнях веществ требует многих месяцев работы, что делает необходимой автоматизацию поиска ВО, позволяющего выявить потенциально ошибочные наблюдения и после экспертной оценки сделать соответствующие исправления. Поэтому ВО продолжают привлекать внимание исследователей, в том числе при решении задач выявления связи свойств соединений со свойствами их компонентов [2, 3]. В настоящее время существует достаточно большое число подходов к идентификации ВО. В их число входят одномерные статистические тесты, оценивающие возможность интерпретации экстремальных значений переменных как ВО [4]. В многомерных данных значения каждой из переменных для ВО могут оказаться неэкстремальными, что не позволяет выявлять такие ВО с помощью одномерных тестов. Для обнаружения подобных ВО может быть использован подход, основанный на вычислении скалярной меры отклонения отдельного наблюдения от всего массива данных. В качестве меры близости может выступать, например, расстояние Махаланобиса или его робастные модификации [5].

В рамках задачи восстановления зависимости переменной y от переменных $\{X_1, \dots, X_p\}$ под ВО может пониматься объект $s = (\mathbf{x}, y)$, для которого отмечается существенное отклонение оценки от ожидаемого значения y в точке \mathbf{x} пространства \mathbf{R}^p . Чтобы выявить ВО в указанном выше смысле, необходимо построить по обучающей выборке модель, связывающую ожидаемое значение y с вектором \mathbf{x} из \mathbf{R}^p . Например, это может быть модель

$$y = \hat{y}(\beta, \mathbf{x}) + \varepsilon. \quad (1)$$

Поиск вектора коэффициентов β может проводиться с помощью метода наименьших квадратов по обучающей выборке $\mathbf{S} = \{(\mathbf{x}_j, y_j) \mid j = 1, \dots, m\}$. Однако более робастные оценки могут быть получены с помощью методов LMS (least median of squares) или LTS (least trimmed squares) [6]. Для оценки отклонения произвольного объекта $s_j = (\mathbf{x}_j, y_j)$ из обучающей выборки от зависимости (1) может быть использован, например, индекс удаленных остатков:

$$\kappa_j^r = \frac{y_j - \hat{y}[\beta(\tilde{\mathbf{S}} \setminus s_j), \mathbf{x}_j]}{\hat{\sigma}_Y},$$

где $\beta(\tilde{\mathbf{S}} \setminus s_j)$ — вектор коэффициентов модели (1), рассчитанный по выборке $\tilde{\mathbf{S}}$ после исключения s_j ; $\hat{\sigma}_Y$ — выборочное стандартное отклонение.

В качестве альтернативной меры несоответствия объекта s_j преобладающей закономерности может быть использовано расстояние Кука [7, 8], показывающее, насколько s_j искажает регрессионную модель:

$$D_i = \frac{\sum_{j=1}^m \left\{ \hat{y}[\mathbf{x}_j, \beta(\tilde{\mathbf{S}})] - \hat{y}[\mathbf{x}_j, \beta(\tilde{\mathbf{S}} \setminus s_i)] \right\}^2}{p \hat{\sigma}_Y}.$$

Перечисленные индексы, очевидно, могут служить количественными оценками того, что объект s_j относится к выпадающим наблюдениям. Они позволяют ранжировать объекты по степени их отклонения от закономерности, описываемой моделью (1).

2 Метод, основанный на комбинировании параметров неустойчивости и величин отступов

При решении задачи распознавания естественно считать выпадающим объект из класса K_i с описанием \mathbf{x} , если для него велика разность $\max_{j=1, \dots, L} [\mathbf{P}(K_j | \mathbf{x}) - \mathbf{P}(K_i | \mathbf{x})]$ или, напротив, отрицательна и велика по модулю величина $\min_{j=1, \dots, L} [\mathbf{P}(K_j | \mathbf{x}) - \mathbf{P}(K_i | \mathbf{x})]$. На практике количественной оценкой того, что объект с описанием \mathbf{x} является выпадающим наблюдением, очевидно, служит $\max_{j=1, \dots, L} [\hat{\mathbf{P}}(K_j | \mathbf{x}) - \hat{\mathbf{P}}(K_i | \mathbf{x})]$, где $\hat{\mathbf{P}}(K_j | \mathbf{x})$ — оценка вероятности принадлежности объекта с описанием \mathbf{x} к классу K_j . Однако оценки вероятностей принадлежности к классам напрямую используют только статистические методы распознавания. В общем случае в этих целях могут быть использованы величины отступа $\Gamma(K_i, \mathbf{x}) = \max_{j=1, \dots, L} [\gamma_j(\mathbf{x}, \tilde{\mathbf{S}}) - \gamma_i(\mathbf{x}, \tilde{\mathbf{S}})]$, где $\gamma_i(\mathbf{x}, \tilde{\mathbf{S}})$ — оценка принадлежности объекта с описанием \mathbf{x} к классу K_i , которая рассчитана алгоритмом, обученным по $\tilde{\mathbf{S}}$ [9].

В зависимости от значений отступа обучающие объекты условно делятся на 5 типов в порядке убывания отступа: эталонные, неинформативные, пограничные, ошибочные, шумовые.

Эталонные объекты имеют большой положительный отступ, плотно окружены объектами своего класса и являются наиболее типичными его представителями.

Неинформативные объекты также имеют положительный отступ. Изъятие этих объектов из выборки (при условии, что эталонные объекты остаются) не влияет на качество классификации. Они не добавляют к эталонам никакой новой информации. Наличие неинформативных объектов характерно для выборок избыточного объема.

Пограничные объекты имеют отступ, близкий к нулю. Классификация таких объектов неустойчива в том смысле, что малые изменения метрики или состава

обучающей выборки могут изменять их классификацию. Например, в химических задачах такими объектами могут быть метастабильные при определенных внешних условиях (например, при комнатной температуре и атмосферном давлении) соединения или кристаллические модификации.

Гранично-ошибочные объекты имеют небольшие отрицательные отступы и близки к пограничным. Гранично-ошибочные объекты потенциально могут быть распознаны при совершенствовании алгоритма.

Шумовые объекты — это относительно небольшое число объектов, которые плотно окружены объектами чужих классов и удалены от основной массы объектов своего класса. Многие ВО представляют собой именно шумовые объекты. Для шумовых объектов характерны большие отрицательные величины отступа, по которым они легко могут быть идентифицированы.

В условиях высокой размерности ВО могут оказывать существенное влияние на процесс обучения, существенно искажая обученный распознающий алгоритм. При этом ВО превращаются в гранично-ошибочные или пограничные объекты. Информации о величинах отступа нередко оказывается недостаточно для достоверной идентификации таких ВО.

В данной работе предлагается подход к поиску ВО при решении задач распознавания принадлежности объектов к некоторым классам K_1, \dots, K_L по признакам X_1, \dots, X_p , основанный на ранжировании объектов согласно комбинированной оценке, учитывающей как величину отступа, так и величину вносимых искажений. Количественной оценкой того, что объект с описанием \mathbf{x} является выпадающим, естественно считать аналог упомянутого выше расстояния Кука, используемого для описания неустойчивости линейной регрессионной модели:

$$\delta_i = \frac{\sum_{i=1}^L \sum_{j=1}^m \left[\gamma_i(\mathbf{x}_j, \tilde{\mathbf{S}}) - \gamma_i(\mathbf{x}_j, \tilde{\mathbf{S}} \setminus s_j) \right]}{Lm}.$$

Коэффициенты $\Gamma(K_i, \mathbf{x})$ и δ_i по отдельности или в комбинациях могут быть использованы для ранжирования объектов обучающей выборки по степени отклонения от существующих в данных закономерностей.

Однако одного только ранжирования объектов по мере их отклонения от аппроксимируемой зависимости недостаточно для выявления ВО. Необходимо также найти тот порог отсечения, при превышении которого объект можно было бы считать выпадающим. Естественным критерием для выбора такого порога может служить эффективность распознавания, оцениваемая по одной из стандартных метрик. Наиболее полно эффективность распознавания характеризуется с помощью AUC (area under curve) — площади под ROC (receiver operating characteristics) кривой. Будем считать, что оценка принадлежности объекта к класс — величина, изменяющаяся в диапазоне $[0, 1]$. Если это не так, оценки можно спроецировать на отрезок $[0, 1]$.

Отбор ВО происходит следующим образом.

1. Получим оценки принадлежности к классам на полной выборке:
 - (а) обучим классификатор C_o на обучающей выборке $\tilde{\mathbf{S}}$;
 - (б) применим C_o к $\tilde{\mathbf{S}}$ и получим оценки вероятностей принадлежности объектов к классам $[\gamma_1(\mathbf{x}_1, \tilde{\mathbf{S}}), \dots, \gamma_L(\mathbf{x}_1, \tilde{\mathbf{S}})], \dots, [\gamma_1(\mathbf{x}_m, \tilde{\mathbf{S}}), \dots, \gamma_L(\mathbf{x}_m, \tilde{\mathbf{S}})]$.
2. Для каждого объекта выборки оценим, относится ли он к ВО:
 - (а) построим выборку $\tilde{\mathbf{S}}_i$, исключив из $\tilde{\mathbf{S}}$ пару (x_i, y_i) ;
 - (б) обучим классификатор C_i на выборке $\tilde{\mathbf{S}}_i$;
 - (в) применим C_i к объектам из $\tilde{\mathbf{S}}$ и получим оценки $[\gamma_1(\mathbf{x}_1, \tilde{\mathbf{S}}_i), \dots, \gamma_L(\mathbf{x}_1, \tilde{\mathbf{S}}_i)], \dots, [\gamma_1(\mathbf{x}_m, \tilde{\mathbf{S}}_i), \dots, \gamma_L(\mathbf{x}_m, \tilde{\mathbf{S}}_i)]$;
 - (г) вычислим

$$\delta(s_i) = \frac{\sum_{l=1}^L \sum_{j=1}^m [\gamma_l(\mathbf{x}_j, \tilde{\mathbf{S}}) - \gamma_l(\mathbf{x}_j, \tilde{\mathbf{S}}_i)]}{Lm}.$$

Величина $\delta(s_i)$ показывает, насколько изменились оценки объектов после исключения s_i из обучающей выборки;

- (д) вычислим

$$\omega(s_i) = \frac{\sum_{l \in \{1, \dots, L\} \setminus \lambda(s_i)} [\gamma_l(\mathbf{x}_i, \tilde{\mathbf{S}}_i)]}{L},$$

где $\lambda(s_i)$ — номер класса, к которому принадлежит объект s_i . Величина $\omega(s_i)$ — средняя оценка принадлежности s_i к классам, к которым он не относится, рассчитанная алгоритмом, обученным по выборке без этого объекта.

3. Отберем ВО, исходя из оценок $\delta(s_1), \dots, \delta(s_m)$ и $\omega(s_1), \dots, \omega(s_m)$. Для каждого объекта вычислим

$$E(s_i, a_1, a_2, p) = [a_1 |\delta(s_i)|^p + a_2 |\omega(s_i)|^p]^{1/p}.$$

Заметим, что $E(s_i, a_1, a_2, p)$ — модуль объекта в пространстве (δ, ω) со взвешенной метрикой Минковского.

4. Подвыборку из k объектов с наибольшими E исключаем из обучения.

3 Эксперименты

Вычислительный эксперимент реализован на языке программирования python версии 3.5.1 с использованием библиотеки scikit-learn версии 0.17.

3.1 Данные

В работе решалась задача выявления ВО при прогнозе возможности образования соединений состава $A^{+3}B^{+3}C^{+2}O_4$. Выборка состояла из 758 объектов двух классов. К первому классу принадлежали 695 объектов (существующие соединения), ко второму — 63 (химические системы $A-B-C-O$ без образования соединений вышеуказанного состава). Каждый объект описывали 108 непрерывных признаков. Пропусков в данных не было. В выборке содержались объекты с неверной меткой класса.

Требовалось обнаружить в базе данных объекты, которые потенциально могли бы быть ошибочно классифицированы на этапе их внесения в базу. Принципиальная возможность решения поставленной задачи в первую очередь связана с периодичностью изменения свойств неорганических соединений в зависимости от атомных номеров элементов — компонентов химических систем.

3.2 Методы распознавания, использованные в исследовании

Для решения задачи классификации использовался градиентный бустинг (GB, gradient boosting) над решающими деревьями [10]. Градиентный бустинг был выбран после сравнения с другими популярными алгоритмами классификации, такими как решающие деревья (DT, decision trees), метод опорных векторов (SVM, support vector machine), метод ближайших соседей (KNN, k-nearest neighbors) [9, 10]. Качество оценивалось на полной выборке при помощи десяти-фолдовой кроссвалидации с сохранением долевого содержания классов в выборках.

Исходные (до удаления ВО) результаты распознавания для разных методов представлены в табл. 1. Градиентный бустинг показал относительно неплохие результаты по сравнению с другими алгоритмами.

Таблица 1 Результаты распознавания

Метод	AUC ROC [11]
KNN	0,82
SVM	0,85
DT	0,77
GB	0,84

3.3 Связь параметров неустойчивости и величин отступа

Дополнительным аргументом комбинирования величин неустойчивости и отступа стали результаты исследования их взаимосвязи, представленные на рис. 1.

Из рис. 1 видно, что величины δ и ω не коррелируют между собой. Низкая корреляция оценок δ и ω свидетельствует о том, что они могут служить независимыми индикаторами ВО.

3.4 Влияние числа исключенных объектов на качество классификации

В работе было изучено, как на качество классификации влияют параметры взвешенной метрики Минковского a_1 , a_2 , p и число исключенных объектов k .

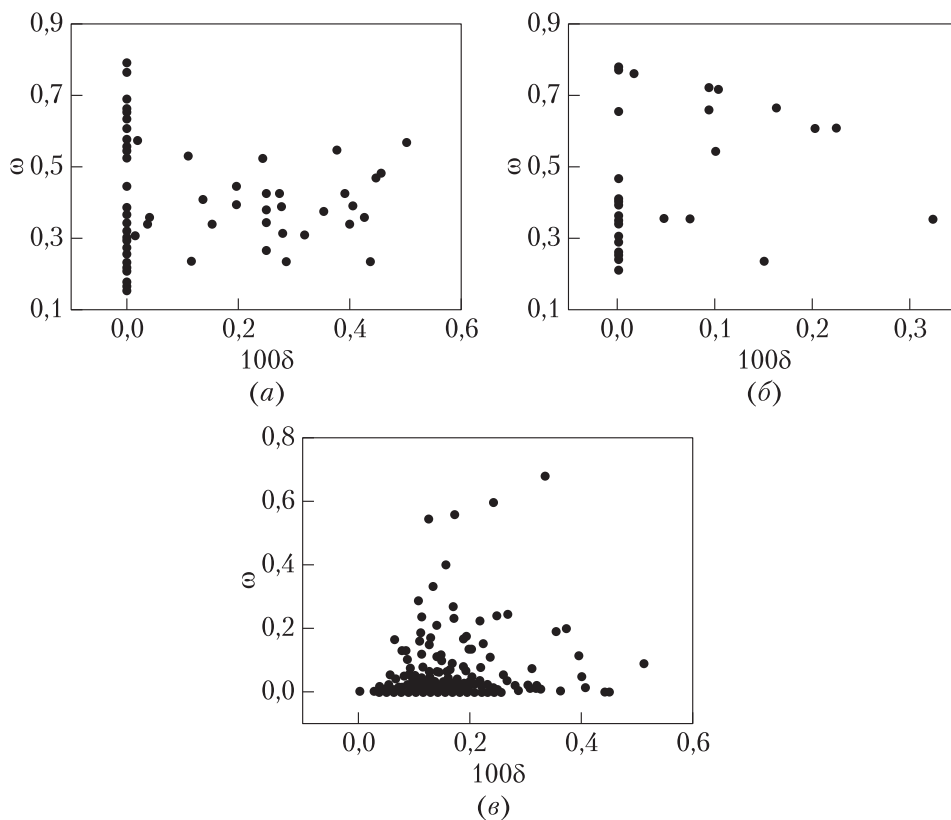


Рис. 1 Зависимость между величинами отступа ω и неустойчивости δ

Для оценки качества оптимизации процедуры отбора ВО и оценки влияния отбора на точность классификации использовались внешняя и внутренняя процедуры скользящего контроля. На каждом из шагов внешней процедуры формировались обучающая и контрольная выборки. Идентификация ВО внутри обучающих выборок производилась по показателям (δ, ω) , рассчитанным с использованием процедур внутреннего скользящего контроля. Внутренний скользящий контроль использовался для подбора параметров градиентного бустинга — оптимальной скорости и числа деревьев. Параметры отбора, включая число ВО k , степенной показатель метрики Минковского p и соотношение весов a_1/a_2 , также подбирались в ходе внутреннего скользящего контроля, исходя из требования максимизации точности распознавания, оцениваемой с помощью ROC AUC. Отметим, что поиск параметров отбора ВО осуществлялся при заранее найденных фиксированных оптимальных параметрах градиентного бустинга. После удаления из обучающей выборки выявленных ВО алгоритм распознавания обучался за-

Таблица 2 Сценарий экспериментов

№ эксперимента	Число блоков внутреннего скользящего контроля (N_{in})	Число блоков внешнего скользящего контроля (N_{out})
1	10	10
2	10	20
3	30	10

ново. Результаты распознавания оценивались на соответствующих контрольных выборках. Эксперименты проводились с различным числом блоков скользящего контроля (табл. 2).

При выборе p достаточно перебрать числа от одного до пяти, соотношение a_1/a_2 выбрать сложнее. Эмпирически установлено, что качество распознавания максимально, когда отношение $\text{med}_{x \in X} \delta(x) a_1 / \text{med}_{x \in X} \omega(x) a_1 \in [1, 2]$, где med — 0,5-квантиль. Таким образом, число выполняемых циклов внутреннего скользящего контроля составило $N_{in} N_{out} N_p N_{a_1/a_2} k_{\max}$, где N_p — число перебираемых степенных показателей; N_{a_1/a_2} — число перебираемых соотношений a_1/a_2 ; k_{\max} — предполагаемое максимальное число ВО. Заметим, что подбор параметров градиентного бустинга занял гораздо больше времени, чем вычисление значения метрики Минковского в двухмерном пространстве для каждого объекта выборки.

Из рис. 2 видно, что изменение a_1 значительно влияет на качество. При этом виден отчетливый экстремум при $a_1 = 300$.

Метод поиска ВО с использованием параметров неустойчивости обучения сравнивался с другими популярными методами поиска ВО. Первый метод основан на исключении объектов с максимальным по модулю отрицательным отступом. Предполагается, что такие наблюдения лежат в гуще объектов противоположного класса. Второй метод основан на исключении объектов с малым по модулю отступом. Предполагается, что объекты с малым отступом лежат на границе двух классов, не являются эталонами класса и снижают обобщающую способность алгоритма. Результаты использования этих алгоритмов продемонстрированы на рис. 3.

Метод поиска ВО с использованием параметров неустойчивости обучения показывал более высокие результаты при различных наборах гиперпараметров (рис. 4).

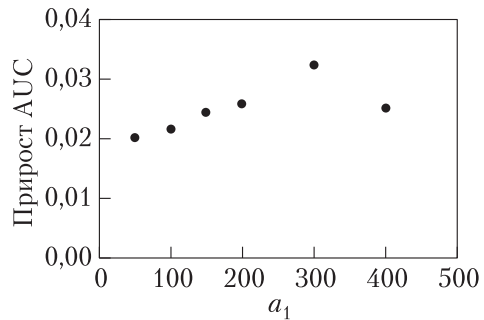


Рис. 2 График зависимости прироста AUC от a_1 ($a_2 = 1$; $p = 1$)

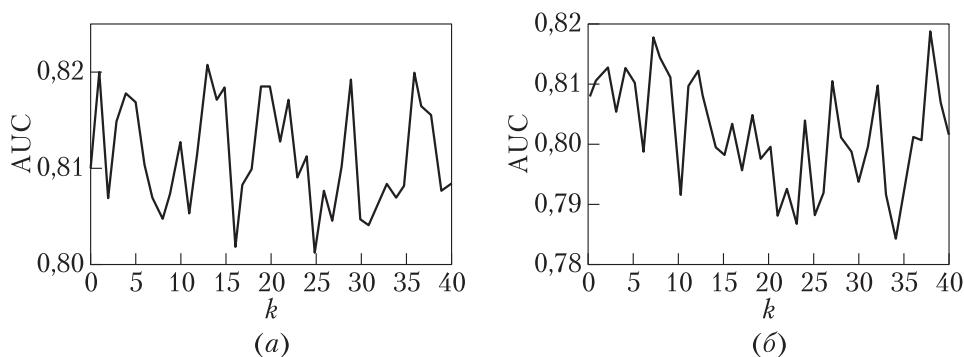


Рис. 3 AUC ROC после удаления k объектов с наибольшим (а) и с наименьшим (б) по модулю отступом. Вычисление AUC проводилось в режиме скользящего контроля с учетом исключенных объектов (эксперимент № 3). Видно, что зависимость неустойчива и нет выраженной тенденции изменения AUC: (а) прирост AUC составил 0,012; (б) прирост AUC составил 0,011

Описанный в работе алгоритм сильнее повышает AUC ROC, однако имеет большую вычислительную сложность. Из рис. 4 видно, что оптимальные результаты достигнуты при $p = 2$ и $a_1/a_2 = 0,0025$ и при $p = 3$ и $a_1/a_2 = 0,0025$. При $p = 4$ и $a_1/a_2 = 0,0025$ прирост AUC ROC ниже, чем в других случаях. При $p = 1$ и $a_1/a_2 = 0,005$ нет выраженной тенденции повышения точности при исключении объектов по порядку ранжирования. Кроме k необходимо подобрать p и верное соотношение a_1/a_2 . Основную сложность представляет подбор соотношения a_1/a_2 и числа исключенных объектов k . Поскольку нельзя сделать никаких предположений о зависимости параметров и качества обнаружения ВО, оптимальные значения параметров метода приходится подбирать при помощи процедуры скользящего контроля.

Из рис. 4 видно, что максимальное качество классификации ($AUC = 0,836$) достигается при исключении из полной обучающей выборки подвыборки из 21 объекта, которая была проанализирована экспертом. Оказалось, что восемь объектов имели ошибочную метку первого класса, один объект имел ошибочную метку второго класса, о принадлежности одного объекта не было найдено достоверных данных. Таким образом, продемонстрирована способность разработанного метода выявлять именно ошибочные наблюдения.

4 Заключение

В ходе работы был разработан алгоритм отбора ВО, основанный на исключении из выборки объектов, наиболее сильно искажающих разделяющую поверхность. Данный метод позволяет добиться большего улучшения качества, чем его аналоги, однако он требует довольно тщательного подбора гиперпара-

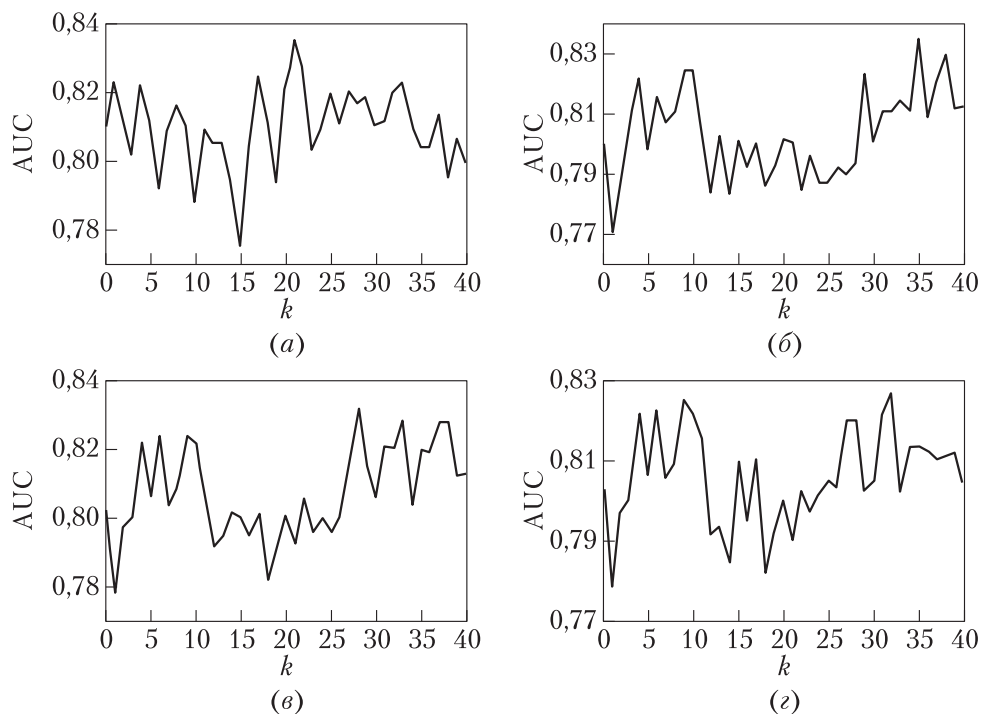


Рис. 4 Зависимость величины AUC от числа исключенных по порядку ранжирования объектов обучающей выборки по комбинированным оценкам, включающим отступ и неустойчивость ($a_2 = 1$): (а) $a_1 = 200$, $p = 1$ — прирост AUC на 0,026; (б) $a_1 = 400$, $p = 2$ — прирост AUC на 0,028; (в) $a_1 = 400$, $p = 3$ — прирост AUC на 0,024; (г) $a_1 = 400$, $p = 4$ — прирост AUC на 0,019. Видно, что, несмотря на сохранение неустойчивости, появляется выраженная тенденция повышения точности при исключении объектов по порядку ранжирования

метров, что создает сложности при использовании этого алгоритма для больших данных. С другой стороны, предложенный алгоритм может быть легко выполнен параллельно. Применение разработанного алгоритма при фильтрации ошибок в базах данных по свойствам неорганических соединений позволило значительно сократить время и трудозатраты на выявление ошибок в определении статуса химических объектов и повысить точность прогнозирования при конструировании новых неорганических соединений. Следует отметить, что анализ выявленных ВО дает стимул к дополнительному изучению соответствующих соединений.

Литература

1. Aggarwal C. C. Outlier analysis. — New York, NY, USA: Springer-Verlag, 2013. 446 p.

2. *Киселёва Н. Н.* Компьютерное конструирование неорганических соединений. — М.: Наука, 2005. 289 с.
3. *Киселева Н. Н., Столяренко А. В., Рязанов В. В., Сенько О. В., Докукин А. А.* Прогнозирование новых соединений состава $A^3+B^3+C^2+O_4$ // Ж. неорганической химии, 2017. Т. 62. № 8. С. 1068–1077.
4. *Grubbs F. E.* Procedures for detecting outlying observations in samples // Technometrics, 1969. Vol. 11. No. 1. P. 1–21.
5. *Rousseeuw P. J., Van Driessen K.* Computing LTS regression for large data sets // Data Min. Knowl. Discovery, 2006. Vol. 12. P. 29–45.
6. *Rousseeuw P. J.* Least median of squares regression // J. Acoust. Soc. Am., 1984. Vol. 79. P. 871–880.
7. *Cook R. D.* Influential observations in linear regression // J. Acoust. Soc. Am., 1979. Vol. 74. P. 169–174.
8. *Cao D. S., Liang Y. Z., Xu Q. S., Li H. D., Chen X.* A new strategy of outlier detection for QSAR/QSPR // J. Comput. Chem., 2010. Vol. 31. P. 592–602.
9. *Журавлев Ю. И., Рязанов В. В., Сенько О. В.* «Распознавание». Математические методы. Программная система. Практические применения. — М.: Фазис, 2006. 176 с.
10. *Hastie T., Tibshirani R., Friedman J. H.* The elements of statistical learning: Data mining, interference, and prediction. — 2nd ed. — New York, NY, USA: Springer, 2009. 767 p.
11. *Zweig M. H.* Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine // Clin. Chem., 1993. Vol. 39. P. 561–577.

Поступила в редакцию 29.03.18

METHOD FOR SEARCHING OUTLIER OBJECTS USING PARAMETERS OF LEARNING INSTABILITY

I. S. Ozhereliev¹, O. V. Senko², and N. N. Kiseleva³

¹Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation

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

³A. A. Baikov Institute of Metallurgy and Materials Science of the Russian Academy of Sciences, 49 Leninskiy Prosp., Moscow 119991, Russian Federation

Abstract: The paper describes a new method of outliers detection in pattern recognition tasks. The authors define an outlier as an object which deviates

significantly from the other objects of the same class. The method is based on simultaneous use of evaluated object estimates for classes and integral distortion of recognition algorithm that is caused by evaluated object. Usefulness of the developed technique was shown for the task of predicting if an inorganic compound of composition $A^{+3}B^{+3}C^{+2}O_4$ is formed under ordinary conditions. The method may be used for erroneous observations detection that is aimed to improve training information in different recognition tasks.

Keywords: outliers; data bases; recognition; instability of training; nonorganic compounds

DOI: 10.14357/08696527190211

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (project 17-01-00634).

References

1. Aggarwal, C. C. 2013. *Outlier analysis*. New York, NY: Springer-Verlag. 446 p.
2. Kiseleva, N. N. 2005. *Komp'yuternoe konstruirovaniye neorganicheskikh soedineniy* [Computer design of nonorganic compounds]. Moscow: Nauka. 289 p.
3. Kiseleva, N. N., A. V. Stolyarenko, V. V. Ryazanov, O. V. Sen'ko, and A. A. Dokukin. 2017. Prediction of new $A^{3+}B^{3+}C^{2+}O_4$ compounds. *Russ. J. Inorg. Chem.* 62:1058–1066.
4. Grubbs, F. E. 1969. Procedures for detecting outlying observations in samples. *Technometrics* 11(1):1–21.
5. Rousseeuw, P. J., and K. Van Driessen. 2006. Computing LTS regression for large data sets. *Data Min. Knowl. Discovery* 12:29–45.
6. Rousseeuw, P. J. 1984. Least median of squares regression. *J. Acoust. Soc. Am.* 79:871–880.
7. Cook, R. D. 1979. Influential observations in linear regression. *J. Acoust. Soc. Am.* 74:169–174.
8. Cao, D. S., Y. Z. Liang, Q. S. Xu, H. D. Li, and X. Chen. 2010. A new strategy of outlier detection for QSAR/QSPR. *J. Comput. Chem.* 31:592–602.
9. Zhuravlev, Yu. I., V. V. Ryazanov, and O. V. Sen'ko. 2006. “*Raspoznavanie.*” *Matematicheskie metody. Programmnaya sistema. Prakticheskie primeneniya* [“Recognition.” Mathematical methods. Program system. Applications]. Moscow: Fazis. 159 p.
10. Hastie, T., R. Tibshirani, and J. H. Friedman. 2009. *The elements of statistical learning: Data mining, interference, and prediction*. 2nd ed. New York, NY: Springer. 767 p.
11. Zweig, M. H. 1993. Receiver-operating Characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.* 39:561–577.

Received March 29, 2018

Contributors

Ozhereliev Ilya S. (b. 1994) — master student, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskiye Gory, GSP-1, Moscow 119991, Russian Federation; ilya365365@gmail.com

Senko Oleg V. (b. 1957) — Doctor of Science in physics and mathematics, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; senkoov@mail.ru

Kiseleva Nadezhda N. (b. 1949) — Doctor of Science in chemistry, head of laboratory, A. A. Baikov Institute of Metallurgy and Materials Science of the Russian Academy of Sciences, 49 Leninskiy Prosp., Moscow 119991, Russian Federation; kis@imet.ac.ru

ОЦЕНКА НАДЕЖНОСТИ ГИБРИДНОГО ВЫСОКОПРОИЗВОДИТЕЛЬНОГО ВЫЧИСЛИТЕЛЬНОГО КОМПЛЕКСА ПРИ РЕШЕНИИ НАУЧНЫХ ЗАДАЧ*

*А. А. Зацаринный¹, А. И. Гаранин², В. А. Кондрашев³, К. И. Волович⁴,
С. И. Мальковский⁵*

Аннотация: Обоснована необходимость применения гибридных решений при создании высокопроизводительных вычислительных систем. Дано краткое описание гибридного высокопроизводительного вычислительного комплекса (ГВВК) Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН) и представлена его укрупненная структурная схема. Предложен методический подход к оценке надежности ГВВК, на основе которого проведены расчеты надежности выделенных функциональных подсистем. Отдельно проведена оценка надежности «вычислительной инфраструктуры» ГВВК (без периферийных элементов). Даны рекомендации по повышению надежности функциональных подсистем.

Ключевые слова: гибридный высокопроизводительный вычислительный комплекс; функциональные подсистемы; надежность; отказ; эквивалентная схема для расчета надежности

DOI: 10.14357/08696527190212

1 Введение

В настоящее время одним из наиболее перспективных направлений развития высокопроизводительных вычислительных систем стали гибридные решения [1], предусматривающие в дополнение к центральным процессорам установку различных «ускорителей вычислений», или сопроцессоров. При подключении к центральному процессору по высокоскоростной шине такие сопроцессоры берут на себя большую часть вычислительной работы.

Идея гибридизации вычислительных систем на аппаратном уровне не нова. Еще в 1980-х гг. в некоторых вычислительных системах в дополнение

* Работа выполнена при частичной поддержке РФФИ (проекты 18-29-03100 и 18-29-03196).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, AZatsarinny@frccsc.ru

² Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, algaranin@mail.ru

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, VKondrashev@frccsc.ru

⁴ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, KVolovich@frccsc.ru

⁵ Вычислительный центр Дальневосточного отделения Российской академии наук, sergey.malkovsky@gmail.com

к центральным процессорам (CPU, central processing units) стали устанавливать сопроцессоры, которые брали на себя часть вычислительных задач. В качестве примера можно привести сопроцессор Intel 8087, выпущенный в 1980 г. Он использовался вместе с центральными процессорами Intel 8086 для ускорения расчетов с плавающей точкой. Еще раньше применялись сопроцессоры, ускоряющие операции ввода–вывода.

Новый этап развития гибридных вычислительных систем начался в середине 2000-х гг., когда для вычислительных целей стали использовать графические процессоры (GPU, graphic processing units) [2]. Интерес к гибридным схемам обусловлен высокой производительностью GPU, которая объясняется особенностями их архитектуры. Если современные CPU содержат несколько ядер (на большинстве современных систем от 2 до 4), то графический процессор изначально создавался как многоядерная структура, в которой число ядер измеряется сотнями. Разница в архитектуре обуславливает и разницу в принципах работы. Если архитектура CPU изначально предполагала последовательную обработку информации, то GPU предназначался для обработки компьютерной графики, так как рассчитан на массивно-параллельные вычисления [2]. Графические процессоры уже достигли той точки развития, когда многие вычислительные задачи могут с легкостью решаться на них, причем быстрее, чем на многоядерных системах классической архитектуры.

Каждая из рассматриваемых архитектур имеет свои достоинства. Центральный процессор лучше работает с задачами, имеющими небольшую степень параллелизма (большое число проверок условий, ветвления, необходимость частой синхронизации потоков и т. д.). При отсутствии таких ограничений вместе с необходимостью обработки больших объемов информации очевидное преимущество приобретают GPU.

Впервые появившись в рейтинге top500 самых мощных вычислительных систем в 2006 г., в настоящий момент (на ноябрь 2018 г.) гибридные вычислительные системы обеспечивают 41% суммарной производительности указанного мирового рейтинга. При этом в первой десятке мирового рейтинга число подобных систем составляет уже 70% и будет в дальнейшем неизбежно расти.

Среди вычислительных платформ подобного класса можно выделить архитектуру POWER, на основе которой создан ряд специализированных вычислительных систем, включающих самые мощные в настоящее время суперкомпьютерные установки в мире — Sierra и Summit с пиковой производительностью более 125 и 200 Пфлопс соответственно.

2 Краткое описание гибридного высокопроизводительного вычислительного комплекса Федерального исследовательского центра «Информатика и управление» Российской академии наук

В ФИЦ ИУ РАН в рамках программы развития предусмотрено формирование современной исследовательской инфраструктуры для повышения эффектив-

ности проводимых фундаментальных научных исследований, а также решения актуальных научно-практических задач.

Для предоставления вычислительных ресурсов заинтересованным научным коллективам в ФИЦ ИУ РАН на базе центра обработки данных (ЦОД) ФИЦ ИУ РАН создан ГВВК, предоставляющий вычислительную инфраструктуру научным подразделениям РАН.

Гибридный высокопроизводительный вычислительный комплекс создан на основе серверов архитектуры POWER различных модификаций (серверы Power 9 укомплектованы четырьмя графическими ускорителями Tesla V100, а также имеются серверы, оснащенные только классическими процессорами CPU).

Средствами ЦОД было организовано сетевое взаимодействие компонентов ГВВК между собой, с пользователями системы и вспомогательными сервисами, предоставляемыми облачной инфраструктурой ЦОД ФИЦ ИУ РАН.

Структурная схема гибридного высокопроизводительного вычислительного комплекса представлена на рис. 1.

3 Методический подход к оценке надежности гибридного высокопроизводительного вычислительного комплекса

При решении вопросов, связанных с оценкой надежности систем, аналогичных ГВВК, предлагается учитывать следующие особенности [3]:

- ГВВК является многофункциональной системой, функции которой имеют существенно различную значимость и, соответственно, характеризуются разным уровнем требований к надежности их выполнения;
- в ГВВК возможно возникновение некоторых исключительных (аварийных, критических) ситуаций, представляющих сочетание отказов либо ошибок функционирования комплекса и способных привести к значительным нарушениям процессов функционирования (авариям);
- в функционировании ГВВК участвуют различные виды ее обеспечения и персонал ГВВК, которые могут в той или иной степени влиять на уровень надежности ГВВК;
- на процесс функционирования ГВВК оказывают влияние большое число разнородных факторов (технические средства, программное обеспечение, человеческий фактор и др.), при этом в выполнении одной функции ГВВК обычно участвуют несколько различных элементов, а один и тот же элемент может участвовать в выполнении нескольких функций системы.

С учетом этих особенностей при решении вопросов надежности ГВВК ее количественное описание, анализ, оценка и обеспечение надежности предлагается проводить по каждой функции ГВВК в отдельности [3].

Совокупность технических и программных элементов ГВВК, выделяемая из всего состава ГВВК по признаку участия в выполнении некоторой (i -й) функции системы, образует i -ю функциональную подсистему (ФП) ГВВК.

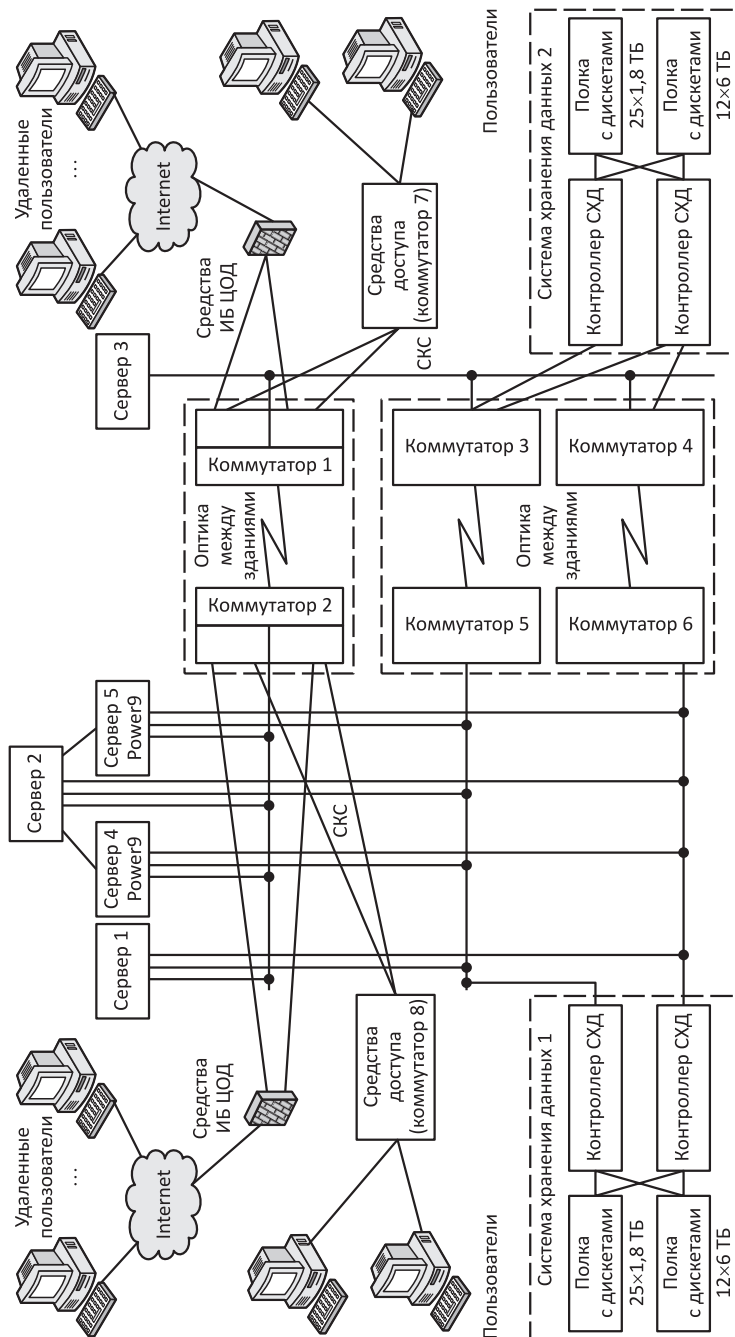


Рис. 1 Структурная схема ГВБК

Анализ надежности ГВВК в реализации его функций проводится по каждой ФП ГВВК в отдельности с учетом уровня надежности и других свойств входящих в нее технических и программных средств, а также человеческого фактора.

Выбор состава показателей надежности ГВВК осуществляется на основе установленного перечня функций системы. Расчет надежности технических систем по безотказности обычно проводится в предположении, что вся система и каждый ее элемент могут находиться только в одном из двух возможных состояний — работоспособном и неработоспособном, а отказы элементов независимы друг от друга.

Проведем структурно-логический анализ [4, 5] процессов функционирования и взаимосвязей элементов в процессе функционирования ГВВК, структурная схема которого представлена на рис. 1.

Доступ удаленных пользователей осуществляется через сеть Интернет с использованием средств обеспечения информационной безопасности (ИБ) ЦОД. Локальные пользователи подключаются через «Средства доступа» (коммутаторы 7 и 8). Средства ИБ и средства доступа подключаются к коммутаторам 1 или 2.

При оценке надежности ГВВК выделим три ФП:

- (1) ФП для решения «информационных» задач;
- (2) ФП для решения задач «подготовки данных»;
- (3) ФП для решения «расчетных» задач.

Серверы 1 и 3 совместно с системами хранения данных (СХД) 1 и 2 используются для решения задач «информационного» типа. При решении задач «подготовки данных» указанные серверы взаимодействуют только с СХД 2.

Серверы 4 и 5 под управлением сервера 2 во взаимодействии с СХД 1 образуют высокопроизводительный вычислительный комплекс и используются для решения расчетных задач.

Коммутаторы 1 и 2 относятся к локальной вычислительной сети (LAN, Local Area Network), коммутаторы 3–6 — к сети хранения данных (SAN, Storage Area Network).

Принято, что если в составе сервера имеется неисправный компонент, то такой сервер выводится из эксплуатации для ремонта.

Сформулируем понятие «отказ». Под **отказом** ГВВК будем понимать событие, заключающееся в прекращении решения задач «информационного» типа, задач «подготовки данных» и расчетных задач по причине неисправности аппаратно-программных средств и требующее для восстановления процесса функционирования проведения ремонтных работ с привлечением обслуживающего персонала.

Используя результаты структурно-логического анализа структурной схемы, представленной на рис. 1, и определение понятия «отказа» системы, составим эквивалентные схемы для расчета надежности указанных выше ФП.

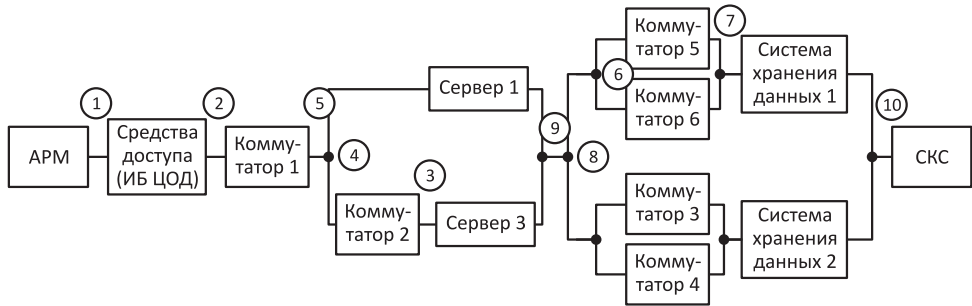


Рис. 2 Эквивалентная схема для расчета надежности ГВВК при решении «информационных» задач

Под эквивалентной схемой для расчета надежности («надежностной структурой», «структурной схемой надежности») системы понимается логическое соединение элементов в таком виде, из которого следовало бы определение отказа системы [4, 5].

При построении эквивалентных схем будем исходить из расчета на «наихудший случай», т. е. выбираем такой вариант построения, при котором для решения пользователем конкретной задачи задействуется наибольшее число элементов системы.

Эквивалентная схема для расчета надежности ГВВК при решении «информационных» задач изображена на рис. 2, где под СКС понимается структурированная кабельная система, обеспечивающая соединение между собой элементов ГВВК; АРМ — автоматизированное рабочее место.

Эта эквивалентная схема представляет собой последовательно-параллельное соединение элементов ГВВК, участвующих в решении «информационных» задач.

При последовательном соединении n элементов системы результирующая вероятность работоспособности системы P_p определяется [4, 6–8] как произведение вероятностей работоспособности входящих в нее элементов:

$$P_p = \prod_{i=1}^n p_i.$$

При параллельном соединении n элементов системы результирующая вероятность работоспособности системы P_p определяется через произведение вероятностей неработоспособности входящих в нее элементов:

$$P_p = 1 - \prod_{i=1}^n q_i,$$

где $q_i = 1 - p_i$ — вероятность неработоспособности i -го элемента.

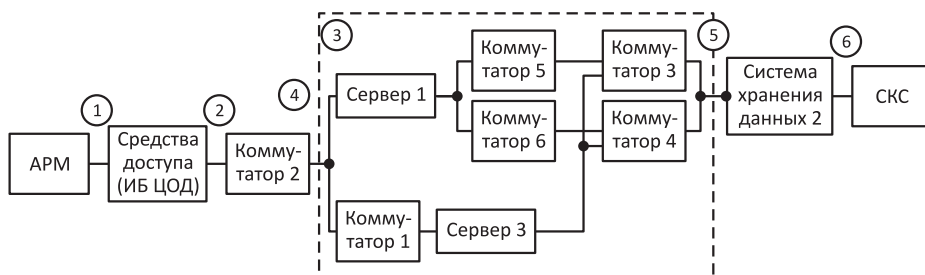


Рис. 3 Эквивалентная схема для расчета надежности ГВБК при решении задач «подготовки данных»

В случае, когда $T_o \gg T_b$, все элементы одинаковые ($\lambda_k = \lambda$) и отказы элементов независимы между собой, для оценки среднего времени безотказной работы ($T_{ор}$) системы, состоящей из n параллельно соединенных элементов, можно воспользоваться выражением [6]:

$$T_{ор} = \frac{1}{\lambda} \sum_{k=1}^n \frac{1}{k} = T_o \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \right), \quad (1)$$

где T_o — среднее время безотказной работы одного элемента; T_b — среднее время восстановления работоспособности отказавшего элемента; $\lambda = 1/T_o$ — интенсивность отказов элемента в каждый данный момент времени.

Последовательность действий при вычислениях указана цифрами в кружках.

Сложнее с эквивалентной схемой для расчета надежности ГВБК при решении задач «подготовки данных» (рис. 3). На рисунке пунктиром выделен фрагмент, который не может быть представлен в виде последовательно-параллельного соединения элементов. Такие системы относятся к системам с произвольной структурой [7, 8].

Для расчета показателей надежности систем произвольной структуры разработан ряд точных и приближенных методов [4–8]. Наиболее простым из точных методов является метод прямого перебора состояний элементов системы.

Его суть состоит в следующем: если определен критерий отказа системы, то все множество ее состояний можно разделить на два подмножества: подмножество состояний работоспособности F и подмножество состояний отказа Ω . Чтобы найти вероятность безотказной работы системы с произвольной структурой (P_Φ), необходимо сложить вероятности всех работоспособных состояний системы.

Оценим надежность выделенного фрагмента указанным методом.

На рис. 4 для удобства вычислений представлен несколько преобразованный вариант эквивалентной схемы фрагмента, выделенного на рис. 3. На схеме элемент 1 заменяет сервер 1 с коэффициентом готовности p_1 , элементы 3–6 заменяют однотипные коммутаторы 3–6, обозначим их коэффициент готовности p_2 ,

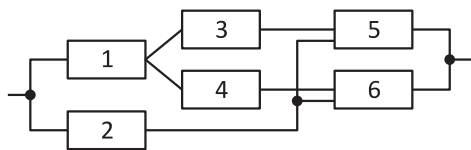


Рис. 4 Эквивалентная схема фрагмента, выделенного на рис. 3

а элемент 2 заменяет сервер 3 и коммутатор 1 с результирующей надежностью двух последовательно соединенных элементов, обозначим его коэффициент готовности p_3 .

В табл. 1 показаны состояния работоспособности выделенного фрагмента в зависимости от состояния его элементов. Символами x_1 – x_6 обозначены элементы 1–6, $x_i = 1$ означает, что i -й элемент системы исправен, а $x_i = 0$ — что он неисправен.

Просуммировав вероятности исправных состояний системы, получим:

$$P_{\Phi} = p_1 p_2 p_3^4 + q_1 p_2 p_3^4 + p_1 q_2 p_3^4 + 4 p_1 p_2 q_3 p_3^3 + 4 q_1 p_2 q_3 p_3^3 + 4 p_1 q_2 q_3 p_3^3 + 5 p_1 p_2 q_3^2 p_3^2 + 5 q_1 p_2 q_3^2 p_3^2 + 4 p_1 q_2 q_3^2 p_3^2 + 2 p_1 p_2 q_3^3 p_3 + 2 q_1 p_2 q_3^3 p_3.$$

Таблица 1 Работоспособные состояния элементов системы

Индекс состояния	Состояние элементов						Вероятность состояния	Индекс состояния	Состояние элементов						Вероятность состояния
	x_1	x_2	x_3	x_4	x_5	x_6			x_1	x_2	x_3	x_4	x_5	x_6	
0	1	1	1	1	1	1	$p_1 p_2 p_3^4$	36	1	1	0	1	1	0	$p_1 p_2 q_3^2 p_3^2$
1	0	1	1	1	1	1	$q_1 p_2 p_3^4$	45	1	1	1	0	0	1	$p_1 p_2 q_3^2 p_3^2$
2	1	0	1	1	1	1	$p_1 q_2 p_3^4$	46	1	1	1	0	1	0	$p_1 p_2 q_3^2 p_3^2$
3	1	1	0	1	1	1	$p_1 p_2 q_3 p_3^3$	134	0	1	0	0	1	1	$q_1 p_2 q_3^2 p_3^2$
4	1	1	1	0	1	1	$p_1 p_2 q_3 p_3^3$	135	0	1	0	1	0	1	$q_1 p_2 q_3^2 p_3^2$
5	1	1	1	1	0	1	$p_1 p_2 q_3 p_3^3$	136	0	1	0	1	1	0	$q_1 p_2 q_3^2 p_3^2$
6	1	1	1	1	1	0	$p_1 p_2 q_3 p_3^3$	145	0	1	1	0	0	1	$q_1 p_2 q_3^2 p_3^2$
13	0	1	0	1	1	1	$q_1 p_2 q_3 p_3^3$	146	0	1	1	0	1	0	$q_1 p_2 q_3^2 p_3^2$
14	0	1	1	0	1	1	$q_1 p_2 q_3 p_3^3$	235	1	0	0	1	0	1	$p_1 q_2 q_3^2 p_3^2$
15	0	1	1	1	0	1	$q_1 p_2 q_3 p_3^3$	236	1	0	0	1	1	0	$p_1 q_2 q_3^2 p_3^2$
16	0	1	1	1	1	0	$q_1 p_2 q_3 p_3^3$	245	1	0	1	0	0	1	$p_1 q_2 q_3^2 p_3^2$
23	1	0	0	1	1	1	$p_1 q_2 q_3 p_3^3$	246	1	0	1	0	1	0	$p_1 q_2 q_3^2 p_3^2$
24	1	0	1	0	1	1	$p_1 q_2 q_3 p_3^3$	345	1	1	0	0	0	1	$p_1 p_2 q_3^3 p_3$
25	1	0	1	1	0	1	$p_1 q_2 q_3 p_3^3$	346	1	1	0	0	1	0	$p_1 p_2 q_3^3 p_3$
26	1	0	1	1	1	0	$p_1 q_2 q_3 p_3^3$	1345	0	1	0	0	0	1	$q_1 p_2 q_3^3 p_3$
34	1	1	0	0	1	1	$p_1 p_2 q_3^2 p_3^2$	1346	0	1	0	0	1	0	$q_1 p_2 q_3^3 p_3$
35	1	1	0	1	0	1	$p_1 p_2 q_3^2 p_3^2$								

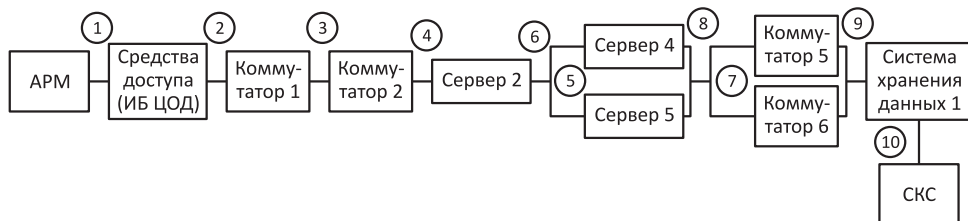


Рис. 5 Эквивалентная схема для расчета надежности ГВВК при решении «расчетных» задач

В остальном расчеты проводятся так же, как и в предыдущем случае (оценка надежности ГВВК при решении «информационных» задач). Последовательность действий при вычислениях указана цифрами в кружках.

Эквивалентная схема для расчета надежности ГВВК при решении «расчетных» задач представляет собой последовательно-параллельное соединение элементов ГВВК, участвующих в решении «расчетных» задач (рис. 5). Расчет надежности этой схемы проводится аналогично расчету для случая решения «информационных» задач.

Исходные данные для расчетов представлены в табл. 2.

Таблица 2 Исходные данные по надежности элементов ГВВК

Устройство	$T_o, ч$	$1 - K_r$
Серверы 1–3	220 110	$1,7 \cdot 10^{-7}$
Сервер Power 9	340 200	$1,5 \cdot 10^{-7}$
Коммутаторы 1, 2	525 600	$3,4 \cdot 10^{-6}$
Коммутаторы 3–6	613 200	$2,9 \cdot 10^{-6}$
Коммутаторы 7, 8	408 800	$1,2 \cdot 10^{-6}$
СХД	700 000	$2,9 \cdot 10^{-6}$
Средства ИБ	344 200	$1,5 \cdot 10^{-6}$
СКС	20 000	$2,5 \cdot 10^{-5}$
АРМ	12 000	$4,2 \cdot 10^{-5}$

4 Оценка надежности и анализ результатов расчетов

Используя предложенную методику и представленные исходные данные, проведем расчеты надежности рассмотренных ФП.

В результате расчетов получены следующие результаты:

- для ФП, решающей задачи «информационного» типа, вероятность отказа $G_{И} = 7,2 \cdot 10^{-5}$;
- для ФП, решающей задачи типа «подготовка данных», $G_{ПД} = 7,5 \cdot 10^{-5}$;
- для ФП, решающей задачи «расчетного» типа, $G_P = 7,8 \cdot 10^{-5}$.

Кроме того, представляет интерес оценка надежности непосредственно «вычислительной инфраструктуры» ГВВК (без периферийных элементов), элементы которой в основном резервируют друг друга (серверы 1 и 3, серверы 4 и 5,

СХД 1 и 2, коммутаторы 3, 5 и 4, 6). Результаты расчетов вероятности отказа «вычислительной инфраструктуры» показали, что она изменяется в пределах $G_{\text{ВИ}} = 3,8 \cdot 10^{-6} \dots 2,3 \cdot 10^{-6}$ в зависимости от конфигурации ее элементов при решении задач различных ФП.

В процессе решения научных и научно-практических задач решаемая задача последовательно использует определенные ресурсы элементов ГВВК (АРМ, средства доступа, коммутаторы, серверы и др.). Из теории надежности известно [4, 6–8], что при последовательном соединении элементов результирующая надежность системы будет ниже надежности «наихудшего» с точки зрения надежности элемента. В данном случае надежность любой из рассмотренных ФП не может быть выше надежности АРМ.

С другой стороны, ГВВК функционирует в интересах большого числа подключенных к нему пользователей, которых можно рассматривать как параллельно соединенные элементы, и с точки зрения «равнопрочности» надежность «вычислительной инфраструктуры» ГВВК должна быть не ниже надежности параллельно соединенных АРМ (т. е. если функционирует хотя бы один пользователь, то «вычислительная инфраструктура» также должна функционировать). Расчеты, проведенные с использованием выражения (1), показывают, что при подключении к ГВВК от 50 до 100 пользователей результирующая вероятность отказа системы, состоящей из n параллельно соединенных АРМ, изменяется в пределах $G_{\Sigma \text{АРМ}} \approx 9,3 \cdot 10^{-6} \dots 8,1 \cdot 10^{-6}$, что выше вероятности отказа «вычислительной инфраструктуры», и с большой долей уверенности можно утверждать, что равнопрочность этих двух составных частей ГВВК выполняется.

Некоторое различие результатов оценки надежности для различных ФП объясняется тем, что при построении эквивалентных схем в расчете на «наихудший случай» в разных вариантах построения ФП задействуется разное число элементов системы (с увеличением числа последовательно соединенных элементов результирующая надежность системы снижается).

Выводы

В статье проведена оценка надежности ГВВК ФИЦ ИУ РАН при решении научных задач. По результатам исследований можно сделать следующие выводы:

- (1) как показали расчеты, среднее время наработки на отказ «вычислительной инфраструктуры» ГВВК (≈ 15 лет) соизмеримо со средним сроком службы ГВВК. Это обеспечивается за счет использования в структуре ГВВК высоконадежных элементов, а также за счет структурного резервирования наиболее важных элементов, участвующих в решении задач, в режиме «горячего резервирования»;
- (2) проведенная оценка вероятности отказа «вычислительной инфраструктуры» ГВВК показала, что она ниже вероятности отказа совокупности подключа-

емых к ней АРМ, поэтому с большой долей уверенности можно утверждать, что если функционирует хотя бы один АРМ пользователя, то «вычислительная инфраструктура» также будет функционировать;

- (3) при необходимости обеспечить повышение надежности ФП наибольший прирост может быть получен при повышении надежности наименее надежных («наихудших» с точки зрения надежности) элементов (в данном случае АРМ и СКС).

Литература

1. *Shan A.* Heterogeneous processing: A strategy for augmenting Moore's law // *Linux J.*, 2006. <https://www.linuxjournal.com/article/8368>.
2. *Соколов И. А., Зацаринный А. А., Дивеев А. И., Захаров В. Н., Посыткин М. А., Абгарян К. К.* Гибридный высокопроизводительный вычислительный комплекс (ГВВК), 2018. <http://www.frccsc.ru/hhpcc>.
3. ГОСТ 24.701-86. Единая система стандартов автоматизированных систем управления. Надежность автоматизированных систем управления. Основные положения. <http://docs.cntd.ru/document/1200022035>.
4. *Каштанов В. А., Медведев А. И.* Теория надежности сложных систем. — М.: Физматлит, 2010. 608 с.
5. *Зацаринный А. А., Гаранин А. И., Козлов С. В.* Научно-практические аспекты обеспечения надежности информационно-телекоммуникационных сетей. — М.: ФИЦ ИУ РАН, 2017. 246 с.
6. *Гнеденко Б. В., Беляев Ю. К., Соловьев А. Д.* Математические методы в теории надежности. — М.: Наука, 1965. 524 с.
7. *Козлов Б. А., Ушаков И. А.* Справочник по расчету надежности аппаратуры радиоэлектроники и автоматики. — М.: Сов. радио, 1975. 462 с.
8. *Беляев Ю. К., Богатырев В. А., Болотин В. В. и др.* Надежность технических систем: Справочник / Под ред. И. А. Ушакова. — М.: Радио и связь, 1985. 608 с.

Поступила в редакцию 15.03.19

EVALUATION OF RELIABILITY OF THE HYBRID HIGH-PERFORMANCE COMPUTING COMPLEX IN SOLUTION OF SCIENTIFIC PROBLEMS

*A. A. Zatsarinny¹, A. I. Garanin¹, V. A. Kondrashev¹, K. I. Volovich¹,
and S. I. Malkovsky²*

¹Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

²Computing Center, Far Eastern Branch of the Russian Academy of Sciences, 65 Kim U Chen Str., Khabarovsk 680000, Russian Federation

Abstract: The necessity of using hybrid solutions in creation of high-performance computing systems is substantiated. A brief description of the hybrid high-performance computing complex (HHPCC) of FRC CSC RAS is given and its enlarged block diagram is presented. The paper offers the methodical approach to estimation of reliability of HHPCC on the basis of which calculations of reliability of the allocated functional subsystems are carried out. Separately, reliability of the “computing infrastructure” of HHPCC (without peripheral elements) was evaluated. Recommendations for improving reliability of functional subsystems are given.

Keywords: hybrid high-performance computing system; reliability; functional subsystems; failure; equivalent circuit for reliability calculation

DOI: 10.14357/08696527190212

Acknowledgments

The work was partly supported by the Russian Foundation for Basic Research (projects 18-29-03100 and 18-29-03196).

References

1. Shan, A. 2006. Heterogeneous processing: A strategy for augmenting Moore’s law. *Linux J.* Available at: <https://www.linuxjournal.com/article/8368> (accessed April 4, 2019).
2. Sokolov, I. A., A. A. Zatsarinny, A. I. Diveev, V. N. Zakharov, M. A. Posypkin, and K. K. Abgarjan. 2018. Gibridnyy vysokoproizvoditel’nyy vychislitel’nyy kompleks (GVVK) [Hybrid high-performance computing system (HHPCC)]. Available at: <http://www.frccsc.ru/hhpcc> (accessed April 4, 2019).
3. GOST 24.701-86. 1986. Edinaya sistema standartov avtomatizirovannykh sistem upravleniya. Nadyozhnost’ avtomatizirovannykh sistem upravleniya. Osnovnye polozheniya [State Standard “An uniform system of standards for automated control systems. Reliability of automated control systems. Fundamentals”]. Available at: <http://docs.cntd.ru/document/1200022035> (accessed April 4, 2019).

4. Kashtanov, V. A., and A. I. Medvedev. 2010. *Teoriya nadezhnosti slozhnykh sistem* [The reliability theory for complex systems]. Moscow: Fizmatlit. 608 p.
5. Zatsarinnyy, A. A., A. I. Garanin, and S. V. Kozlov. 2017. *Nauchno-prakticheskie aspekty obespecheniya nadezhnosti informatsionno-telekommunikatsionnykh setey* [Scientific and practical aspects of ensuring reliability of information and telecommunication networks]. Moscow: FIC IU RAN. 246 p.
6. Gnedenko, B. V., Y. K. Belyaev, and A. D. Soloviev. 1965. *Matematicheskie metody v teorii nadezhnosti* [Mathematical methods in the theory of reliability]. Moscow: Nauka. 524 p.
7. Kozlov, B. A., and I. A. Ushakov. 1975. *Spravochnik po raschetu nadezhnosti apparatury radioelektroniki i avtomatiki* [Handbook on calculation of reliability of radioelectronics and automation equipment]. Moscow: Sovetskoe Radio. 462 p.
8. Belyaev, Yu. K., V. A. Bogatyrev, V. V. Bolotin, et al. 1985. *Nadezhnost' tekhnicheskikh sistem: Spravochnik* [Reliability of technical systems: Handbook]. Ed. I. A. Ushakov. Moscow: Radio i svyaz'. 608 p.

Received March 15, 2019

Contributors

Zatsarinny Alexander A. (b. 1951) — Doctor of Science in technology, professor, Deputy Director, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44- 2 Vavilov Str., Moscow 119133, Russian Federation; AZatsarinny@frccsc.ru

Garanin Alexander I. (b. 1951) — Candidate of Science (PhD) in technology, senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; algaranin@mail.ru

Kondrashev Vadim A. (b. 1963) — senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; vd@ipi.ac.ru

Volovich Konstantin I. (b. 1970) — senior scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; kv@ipi.ac.ru

Malkovsky Sergey I. (b. 1983) — scientist, Computing Center, Far Eastern Branch of the Russian Academy of Sciences, 65 Kim U Chen Str., Khabarovsk 680000, Russian Federation; sergey.malkovsky@gmail.com

МЕТОДОЛОГИЯ АННОТИРОВАНИЯ В НАДКОРПУСНЫХ БАЗАХ ДАННЫХ

А. А. Гончаров¹, О. Ю. Инькова², М. Г. Кружков³

Аннотация: Рассматриваются методологические принципы аннотирования языковых единиц (ЯЕ) в рамках кросслингвистических надкорпусных баз данных (НБД). Надкорпусные базы данных представляют собой новый вид лингвистических информационных ресурсов, позволяющий сохранять результаты лингвистического анализа корпусных данных в виде аннотаций, структурированных в соответствии с задачами исследования. Выделяются 4 основных этапа аннотирования: поиск объектов исследования, выбор ЯЕ (в оригинале и в переводе) и формирование контекста, выбор дополнительных признаков ЯЕ (в оригинале и в переводе) и, наконец, формирование переводного соответствия (ПС) и выбор его признаков. Рассматривается также функциональный потенциал НБД и обобщаются описанные ранее частные методики аннотирования, на основе чего делается вывод о возможности применения разработанной методологии к другим объектам исследования.

Ключевые слова: надкорпусные базы данных; фасетные классификации; лингвистическое аннотирование; методология аннотирования; контрастивная лингвистика

DOI: 10.14357/08696527190213

1 Введение

При проведении современных лингвистических исследований, особенно тех, что относятся к области контрастивной лингвистики, сложно обойтись без специально разработанных средств информатики. В их число в первую очередь входят типологические базы данных (подробнее см. [1]) и электронные корпуса текстов [2]. Однако помимо этих инструментов существуют и другие, сочетающие в себе функции базы данных и корпуса текстов. К ним относятся НБД, разрабатываемые в ИПИ ФИЦ ИУ РАН с 2013 г. Первой из них стала НБД глагольных форм, принципы построения которой описаны в [3]. Термин «надкорпусная база данных» впервые употреблен в 2015 г. в [2] и с тех пор используется в работах как на русском, так и на английском языке (supracorpora database).

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, a.gonch48@gmail.com

²Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, olyainkova@yandex.ru

³Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, magnit75@yandex.ru

Надкорпусные базы данных представляют собой «надстроенный» над корпусом информационный ресурс, позволяющий последовательно фиксировать в рамках этого корпуса все или некоторые употребления исследуемых ЯЕ, снабжая их описаниями, структурированными в соответствии с задачами исследования, а также получать статистические данные по любым параметрам, включенным в описание и, шире, в структуру НБД, или же по сочетаниям этих параметров.

Принимая во внимание тот факт, что на сегодняшний день созданы и непрерывно совершенствуются уже шесть НБД, обеспечивающих возможность исследования ЯЕ различных категорий — глагольных форм, лингвоспецифичной лексики, коннекторов (для которых анализировались не только переводы, выполненные человеком, но и переводы, сгенерированные автоматическим переводчиком), безличных глагольных конструкций, дискурсивных слов и немецких модальных глаголов, в данной статье предпринята попытка обобщить накопленный опыт использования НБД и описать методологические принципы аннотирования, актуальные для самых разных категорий ЯЕ [4, 5]. В качестве иллюстрации поисковых возможностей и различных этапов описанной методологии используются примеры из всех НБД, однако не все описанные функции были реализованы в каждой из баз в силу особенностей исследуемых ЯЕ.

2 Поиск объектов для аннотирования

Аннотирование объектов в НБД в настоящее время осуществляется на основе параллельных корпусов Национального корпуса русского языка (НКРЯ), т. е. все существующие НБД являются кросслингвистическими. Объем используемых корпусов, как правило, слишком велик, чтобы их можно было просматривать вручную в поисках объектов, которые предполагается аннотировать. Например, русско-французский параллельный корпус насчитывает почти 4 млн словоупотреблений (около 134 тыс. выровненных пар предложений), немецко-русский корпус — 2,6 млн словоупотреблений (около 83 тыс. пар предложений). Поэтому прежде чем приступить к аннотированию, необходимо осуществить отбор пар, в которых должны содержаться интересные исследователей ЯЕ.

В работах [6, 7] рассматривается морфологическая разметка корпусов НКРЯ (в [8] — применительно к НБД), в результате которой каждой словоформе приписывается лемма (базовая форма слова) и набор морфологических характеристик, таких как частеречная принадлежность, лицо, число, время и т. д. При этом важно отметить, что в параллельных корпусах НКРЯ омонимия обычно не снимается: это означает, что одной словоформе может соответствовать несколько вариантов такого разбора. Например, в текстовом фрагменте «... *лежал утром в постели, на своей квартире, Илья Ильич Обломов*» словоформе *постели* приписано семь вариантов разбора, из которых верным является только один. Ниже представлены три из них:

- (1) «постель» — сущ., жен. род, неодуш., мн. ч., имен. падеж;
- (2) «постель» — сущ., жен. род, неодуш., ед. ч., предл. падеж;
- (3) «постелить» — глаг., соверш. вид, действ. залог, 2 л., ед. ч., повелит. наклонение.

Такая разметка позволяет осуществлять поиск по лексическим и грамматическим характеристикам словоформ, однако из-за неснятой омонимии в результатах поиска могут присутствовать нерелевантные текстовые фрагменты («шум»). Процедура поиска не является частью непосредственно аннотирования, а предваряет его. В НБД используются следующие виды поиска объектов аннотирования (выбор конкретного вида поиска зависит от объекта исследования):

- лексический поиск в точной форме по одной или нескольким словоформам (например, для неизменяемых лингвоспецифичных единиц, дискурсивных слов, коннекторов);
- лексический поиск по одной или нескольким леммам (например, для изменяемой лингвоспецифичной лексики, немецких модальных глаголов);
- грамматический поиск (например, для поиска глагольных форм или безличных глагольных конструкций);
- лексико-грамматический поиск (например, для поиска глагольных форм с *бы*, *было*, *чтобы*);
- сочетание вышеперечисленных видов с дополнительными функциями:
 - указание расстояния между элементами запроса;
 - учет пунктуации;
 - исключение некоторых словоформ из результатов поиска (в частности, подобные функции используются для уменьшения шума при поиске безличных конструкций).

Поиск объектов аннотирования может осуществляться на основе заранее подготовленных запросов для исследуемых ЯЕ: пользователь выбирает нужную ему ЯЕ из фиксированного списка, и при запуске поиска выполняется заранее подготовленная поисковая процедура. Кроме того, можно использовать другой вид поиска объектов аннотирования, при котором пользователю дается возможность самостоятельно составить запрос на поиск ЯЕ или языковых конструкций, в которые они входят. Для этого используется специально разработанный язык запросов, который позволяет определять составляющие искомой языковой конструкции и их взаиморасположение, специфицировать единицы, которые следует исключить из выдачи, и т. д. Например, в НБД коннекторов запрос **не** – {**успеть**} [1, 19] **как** задает конструкцию, в которой непосредственно за отрицательной частицей **не** следует глагол **успеть** в любой форме, а слово **как** отстоит от **успеть** не более чем на 19 словоформ. Ниже показано несколько примеров, которые будут отобраны при выполнении данного запроса:

- (1) *Илья Ильич уселся на стуле, подобрал под себя ноги и **не успел** задуматься, **как** раздался звонок.* [И. А. Гончаров. Обломов (1848–1859)]
- (2) ***Не успел** будочник, закрывши пальцем свою правую ноздрю, потянуть левою полгорсти, **как** мертвец чихнул так сильно, что совершенно забрызгал им всем троим глаза.* [Н. В. Гоголь. Шинель (1842)]
- (3) — *Все сделаю в лучшем виде! Оглянуться **не успеют**, как шашка прыгнет в дамки!* [А. и Г. Вайнеры. Эра милосердия (1975)]

Созданные до сегодняшнего времени НБД, как правило, используются для исследования ЯЕ одного языка в зеркале их переводов на другой язык (или языки). Поэтому в параллельных корпусах обычно используется монопольный поиск объектов аннотирования, а текст на втором языке при поиске не учитывается, поскольку исследование многообразия вариантов перевода нередко является одной из поставленных задач (так, в [9] разнообразие моделей перевода выступает в качестве одного из параметров определения степени лингвоспецифичности коннекторов). Однако иногда у исследователей возникает потребность подробнее рассмотреть определенные варианты перевода для некоторой ЯЕ. В связи с этим НБД также поддерживает возможность при поиске объектов аннотирования задавать критерии поиска одновременно для языка оригинала и языка перевода. Например, если пользователь хочет исследовать определенную модель перевода, то он может, задав соответствующие запросы для двух языков, отобразить все выровненные пары, в которых, например, в русской части встречается *хотя*, а во французской — *malgré*, чтобы затем аннотировать подходящие употребления.

Иногда встает задача выявления «нестандартных» моделей перевода для некоторых ЯЕ. Чтобы было легче обнаружить такие модели перевода, при поиске объектов аннотирования предусмотрена возможность отсеять выровненные пары, в переводной части которых содержатся определенные лексические единицы. Например, если в базе данных немецких модальных глаголов задать поиск по лемме **{wollen}**, будет найдена 3541 пара с какой-либо из форм данного глагола. Если же при этом в строку поиска по русской части ввести значение **-{хотеть, желать}**, то из результатов поиска будут исключены пары, содержащие словоформы этих глаголов, и останется 2105 пар, удовлетворяющих заданным условиям. Таким образом, из первоначальных результатов поиска исключается более 40% пар, иллюстрирующих стандартные модели перевода, что значительно облегчает и ускоряет процесс исследования.

Представленные выше виды поиска объектов аннотирования резюмируются на рис. 1.

Наконец, отметим, что НБД позволяет осуществлять поиск объектов аннотирования как по всему корпусу, так и в отдельных произведениях. Если поиск осуществляется по всему корпусу, особенно когда речь идет о высокочастотных единицах, то необходимо равномерно отбирать объекты аннотирования из различных текстов, входящих в корпус, чтобы уменьшить влияние индивидуальных



Рис. 1 Виды поиска объектов аннотирования

особенностей того или иного текста на общие результаты анализа (о важности этого фактора см., например, [10, с. 13–15]). В связи с этим на странице поиска реализована функция, позволяющая выводить найденные объекты аннотирования и их контексты в случайном порядке вместо порядка, используемого по умолчанию, при котором контексты сгруппированы по произведениям.

3 Выбор языковой единицы и формирование контекста

После того как при помощи инструментов поиска были отобраны нужные для исследования объекты, можно приступать непосредственно к последовательному формированию аннотаций. В данном разделе будут рассмотрены два тесно связанных друг с другом этапа аннотирования: выбор аннотируемой ЯЕ и формирование ее контекста. В качестве отправной возьмем ситуацию, когда, во-первых, для анализа функционирования выбранной ЯЕ не нужно выходить за рамки той пары, в которой эта ЯЕ содержится, и, во-вторых, когда аннотируемая ЯЕ является однословной. В таких случаях контекст выбранной пары не нуждается в расширении, а для классификации может быть выбрана начальная форма слова (например, существительное в именительном падеже единственного числа, глагол в инфинитиве и т. п.). Однако подобные ситуации представляют собой лишь один из возможных вариантов, поскольку на данном этапе исследователи сталкиваются с двумя основными сложностями. Первая из них заключается в том, что контекст выбранной пары может оказаться слишком узким для анализа выбранной ЯЕ. Вторая же сложность возникает из-за того, что к исследуемой категории ЯЕ могут относиться единицы, имеющие разную структуру (например, неоднословные или имеющие ряд вариантов).

Что касается первой сложности, то в НБД предусмотрена возможность записи расширенного контекста. В [8] был описан механизм расширения контекста путем объединения двух и более пар. Следует, впрочем, отметить, что при этом необязательно ориентироваться на полные текстовые фрагменты, которые содержатся в каждой из пар. Исследователь имеет возможность выбрать лишь те части текста (в рамках как одной выбранной пары, так и соседних пар), которые необходимы и достаточны для полного представления функционирования ЯЕ. Пропуск частей текста отмечается в записанных контекстах при помощи знака «[...]», как это делается при обычном цитировании.

Вторая сложность сводится к тому, что как принципы классификации аннотируемых ЯЕ, так и принципы выбора базовых форм этих единиц могут различаться. Исследуемая ЯЕ способна выходить за пределы орфографического слова (в соответствии с графическим критерием выделения слова под этим термином понимается слово, отделенное от соседних пробелами), которое не всегда может рассматриваться как самостоятельный объект исследования. Иногда приходится иметь дело не просто с последовательностями слов в фиксированной форме (например, «в то же время»), но и с последовательностями, допускающими варьирование формы входящих в них слов (например, «какой (какая, какое, какие...) бы то ни было») или факультативность некоторых элементов (например, «(ну) как же», «как (это) ни странно»), а также с единицами, части которых могут располагаться на удалении друг от друга (например, «как... так и»). Такие ЯЕ могут определяться как «сочетания, эквивалентные слову», и представляют собой объект самостоятельных работ по лексикографии [11], где, несмотря на ориентированность именно на этот тип единиц, в рамках алфавитного перечисления проблема вариативности единицы и выбора ее базовой формы встает особенно остро.

Надкорпусная база данных позволяет решить эту проблему за счет отнесения однословных ЯЕ в процессе классификации к нескольким кластерам одновременно. Это означает, что при пополнении списка ЯЕ нет необходимости выбирать какую бы то ни было «основную» форму для каждой новой обнаруженной однословной единицы. Так, в НБД коннекторов ЯЕ «да еще и» отнесена к 5 кластерам: «многоэлементные» (коннекторы, состоящие из нескольких слов), «да», «да и», «и», «еще». Таким образом, благодаря системе перекрестных кластеров вопрос об основной форме оказывается нерелевантным при аннотировании, в то время как внутренняя структура единицы снабжается структурированным описанием. Более того, возможность использования перекрестных кластеров в рамках НБД для описания различных категорий ЯЕ позволяет исследовать те единицы, принципы описания которых нуждаются в уточнении или не были разработаны ранее, а также производить аннотирование частей однословных ЯЕ (например, коннекторов, для аннотирования частей которых могут использоваться различные категории двуязычных кортежей, подробно описанные в [12]).

Помимо записи контекста объекта аннотирования и выбора ЯЕ на данном этапе аннотирования используется шрифтовое оформление. Так, исследуемая

единица, как правило, выделяется полужирным шрифтом, а функционально значимые слова (непосредственно влияющие на функционирование исследуемой ЯЕ, например «*именно потому, что*» для коннектора «*потому, что*») — курсивом.

4 Выбор дополнительных признаков

После того как был записан необходимый контекст, выбрана исследуемая ЯЕ и с помощью шрифтов выделены основные и функционально-значимые элементы контекста, выбираются дополнительные признаки ЯЕ. Важно, что в их число входят признаки, характеризующие выбранную ЯЕ именно в данном контексте (не зависящие от контекста характеристики относятся к структуре ЯЕ и описаны выше).

Как правило, лингвисты при описании используют неоднородные параметры, которые могут быть сгруппированы в отдельные «кластеры дополнительных признаков». Вся совокупность таких кластеров представляет собой фасетную классификацию [13], где кластеры можно рассматривать как фасеты, а дополнительные признаки — как рубрики фасетов. Это дает возможность получать информацию о функционировании ЯЕ за счет различных комбинаций разнородных релевантных признаков. К примеру, для коннекторов выделяются кластеры признаков, позволяющие фиксировать логико-семантические отношения, выражаемые коннектором, синтаксическую структуру вводимого коннектором фрагмента текста, позицию коннектора в этом фрагменте, порядок следования связываемых коннектором фрагментов текста и т. д. (подробнее см. [9]). Для немецких модальных глаголов аннотируются значение глагола, его грамматическая форма, характеристики предложения, в котором он употреблен, и т. д.

Таким образом, фасетная классификация может применяться для аннотирования ЯЕ разных категорий. Более того, применение фасетной классификации в рамках НБД обеспечивает два основных варианта работы с типологиями. Во-первых, оно позволяет проверить некоторую уже существующую типологию (например, классификацию логико-семантических отношений, перечень словарных значений модальных глаголов), выявить в ней лакуны или иные недостатки и усовершенствовать таким образом эту типологию. Во-вторых, типология может быть разработана с нуля в процессе работы с языковым материалом (например, именно так был создан описанный в [14, 15] фасет, позволяющий рубрицировать ошибки машинного перевода).

Следует отметить, что и внутри одного кластера признаки могут классифицироваться с использованием перекрестных кластеров так же, как это было описано выше применительно к ЯЕ. Например, в НБД коннекторов логико-семантическое отношение причины, устанавливаемое на уровне высказывания, внутри кластера «отношения» следует относить к кластерам «причина» и «уровень высказывания» одновременно. Все это делает возможным детальное многоуровневое описание разнообразных ЯЕ и обеспечивает удобство аналитической работы с материалом уже после аннотирования.

5 Формирование переводных соответствий и выбор их признаков

Как отмечалось выше, существующие НБД, как правило, используются для исследования ЯЕ некоторого языка в зеркале их переводов на некоторый «референтный» язык¹ (или языки). Поэтому после формирования аннотации употребления исследуемой ЯЕ также формируется аннотация соответствующего фрагмента перевода, содержащего некоторую ЯЕ на референтном языке, которая в данном контексте признается соответствием или эквивалентом исходной ЯЕ.

Аннотирование ЯЕ на референтном языке в целом осуществляется по той же схеме, что и аннотирование исходной единицы (см. разд. 3 и 4), хотя присутствуют и некоторые отличия, связанные, главным образом, с тем, что ЯЕ референтного языка, соответствующие исследуемым ЯЕ исходного языка, в значительной степени разнородны. В качестве примера можно привести НБД немецких модальных глаголов, описанную в [19].

Основным информационным объектом кросслингвистической НБД является аннотация ПС, которая объединяет в себе две моноязычных аннотации: аннотацию употребления исходной ЯЕ и аннотацию соответствующей ей ЯЕ на референтном языке. Количественный и качественный анализ аннотаций ПС позволяет исследователям формулировать и проверять гипотезы, выявлять и заполнять лакуны в системе лингвистических знаний.

Важно отметить, что аннотации ПС, в свою очередь, также могут сопровождаться признаками, которые характеризуют ПС в целом и не релевантны для аннотаций ЯЕ, входящих в состав ПС. Например, признаки ПС могут иметь отношение к трансформациям, совершенным в процессе перевода (перефразирование, смена грамматического субъекта, смена модальности).

Другие признаки ПС используются в качестве служебных помет, позволяющих выделить некоторые аннотации ПС в общем массиве. Например, помета NB используется для выделения аннотаций, которые могут представлять особый интерес; помета Exr используется для выделения аннотаций, для проверки которых предлагается привлечь новых экспертов, и т. д.

Несмотря на описанные выше возможные различия в принципах аннотирования каких-либо классов ЯЕ, общие принципы остаются в целом едиными и резюмируются в виде схемы, представленной на рис. 2.

¹В работах по переводоведению обычно используется пара терминов *source language* и *target language*, которые могут переводиться на русский язык, например, как «исходный язык» и «язык перевода». Однако в рамках рассматриваемой методологии ЯЕ могут исследоваться не только в зеркале переводов на другие языки, но и в зеркале стимулов перевода, т. е. иноязычных конструкций, которые могут вызывать появление исследуемых ЯЕ при переводе с другого языка. Поэтому для обозначения языков, в зеркале которых исследуются ЯЕ исходного языка, предлагается использовать термин «референтный язык» [16]. Данный термин (*reference language*) в аналогичных значениях используется, например, в работах [17, 18].

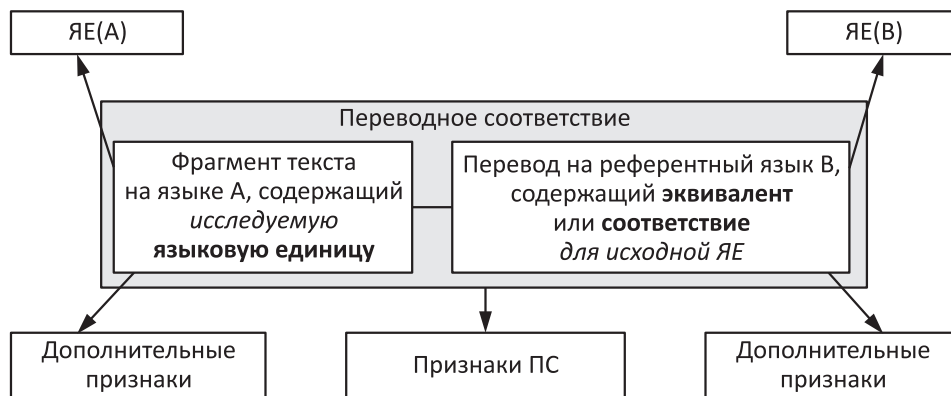


Рис. 2 Общая схема аннотации ПС

6 Заключение

Таким образом, рассмотренная методология аннотирования в НБД может быть успешно применена для анализа различных ЯЕ и их переводов, поскольку она позволяет исследовать не только лексические единицы, но и грамматические категории. Хотя общая структура методологии остается неизменной, различные ее аспекты варьируются и развиваются по мере аннотирования новых лингвистических объектов и совершенствования НБД. Это особенно важно потому, что основное отличие функционала НБД от НКРЯ состоит именно в возможности лингвистического аннотирования параллельных текстов и последующей работы с полученными аннотациями [2, 12]. Все вышеперечисленное говорит о большом потенциале и перспективности использования НБД как для лингвистических исследований (теоретическая грамматика и семантика, лексикография, теория перевода и др.), так и в исследованиях, направленных на совершенствование работы автоматических переводчиков.

Литература

1. Кружков М. Г. Информационные ресурсы контрастивных лингвистических исследований: типологические базы данных // Системы и средства информатики, 2015. Т. 25. № 1. С. 198–212.
2. Кружков М. Г. Информационные ресурсы контрастивных лингвистических исследований: электронные корпуса текстов // Системы и средства информатики, 2015. Т. 25. № 2. С. 140–159.
3. Бунтман Н. В., Зализняк Анна А., Зацман И. М., Кружков М. Г., Лоцилова Е. Ю., Сичинава Д. В. Информационные технологии корпусных исследований: принципы построения кросслингвистических баз данных // Информатика и её применения, 2014. Т. 8. Вып. 2. С. 98–110.

4. *Buntman N., Minel J. L., Le Pesant D., Zatsman I.* Typology and computer modeling of translation difficulties // Информатика и её применения, 2010. Т. 4. Вып. 3. С. 77–83.
5. Handbook of linguistic annotation / Eds. N. Ide, J. Pustejovsky. — Dordrecht, The Netherlands: Springer Science + Business Media, 2017. 1468 p.
6. *Segalovich I.* A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine // Conference (International) on Machine Learning: Models, Technologies and Applications Proceedings. — Las Vegas, NV, USA: CSREA Press, 2003. P. 273–280.
7. *Зобнин А. И., Носырев Г. В.* Морфологический анализатор MyStem 3.0 // Труды Института русского языка им. В. В. Виноградова, 2015. № 6. С. 300–310.
8. *Гончаров А. А., Зацман И. М.* Информационные трансформации параллельных текстов в задачах извлечения знаний // Системы и средства информатики, 2019. № 1. С. 180–193.
9. *Инькова О. Ю.* Лингвоспецифичность коннекторов: методы и параметры описания // Семантика коннекторов: контрастивное исследование. — М.: ТОРУС ПРЕСС, 2018. С. 5–23.
10. *Johansson S.* Seeing through Multilingual Corpora: On the use of corpora in contrastive studies. — Amsterdam: John Benjamins B. V., 2007. 355 p.
11. *Рогожникова Р. П.* Толковый словарь сочетаний, эквивалентных слову: Около 1500 устойчивых сочетаний русского языка. — М.: Астрель, АСТ, 2003. 416 с.
12. *Зацман И. М., Инькова О. Ю., Кружков М. Г., Попкова Н. А.* Представление кроссязыковых знаний о коннекторах в надкорпусных базах данных // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 106–118.
13. *Зацман И. М., Инькова О. Ю., Нуриев В. А.* Построение классификационных схем: методы и технологии экспертного формирования // Научно-техническая информация. Сер. 2. Информационные процессы и системы, 2017. № 1. С. 8–22.
14. *Nuriev V., Buntman N., Inkova O.* Machine translation of Russian connectives into French: Errors and quality failures // Информатика и её применения, 2018. Т. 12. Вып. 2. С. 105–113.
15. *Бунтман Н. В., Гончаров А. А., Зацман И. М., Нуриев В. А.* Количественный анализ результатов машинного перевода с использованием надкорпусных баз данных // Информатика и её применения, 2018. Т. 12. Вып. 4. С. 96–105.
16. *Зацман И. М., Кружков М. Г.* Надкорпусная база данных коннекторов: развитие системы терминов проектирования // Системы и средства информатики, 2018. Т. 28. № 4. С. 156–167.
17. *Veninca P., Haiman J.* The Rhaeto-Romance languages. — London: Routledge, 1992. 260 p.
18. *Van Dyk J.* Language learning through sight translation // Translation in second language learning and teaching. — Bern: Peter Lang, 2009. P. 203–214.
19. *Добровольский Д. О., Зализняк Анна А.* Немецкие конструкции с модальными глаголами и их русские соответствия: проект надкорпусной базы данных // Компьютерная лингвистика и интеллектуальные технологии: По мат-лам Междунар. конф. «Диалог». — М.: РГГУ, 2018. Вып. 17(24). С. 172–184.

Поступила в редакцию 15.03.19

ANNOTATION METHODOLOGY OF SUPRACORPORA DATABASES

A. A. Goncharov, O. Yu. Inkova, and M. G. KruzHKov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: The paper considers methodological principles of annotating linguistic units in parallel corpora using supracorpora databases. Supracorpora databases are a novel information resource in linguistics that allows researchers to save the results of linguistic analysis of corpus data in the form of annotations structured according to the research objectives. When dealing with parallel corpora, the annotation procedure consists of 4 basic stages: annotation objects lookup; definition of the linguistic unit and its context (both in original and translated texts); definition of the linguistic unit’s attributes (both in original and translated texts); and combination of two linguistic units into a translation correspondence and definition of its attributes. The paper summarizes the previously described annotation techniques, examines functional potential of supracorpora databases, and concludes that it is possible to apply the developed methodology to a wide variety of research objects.

Keywords: supracorpora databases; faceted classifications; linguistic annotation; annotation methodology; contrastive linguistics

DOI: 10.14357/08696527190213

References

1. KruzHKov, M. G. 2015. Informatsionnye resursy kontrastivnykh lingvisticheskikh issledovaniy: tipologicheskie bazy dannykh [Information resources for contrastive studies: Typological databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(1):198–212.
2. KruzHKov, M. G. 2015. Informatsionnye resursy kontrastivnykh lingvisticheskikh issledovaniy: elektronnye korpusa tekstov [Information resources for contrastive studies: Electronic text corpora]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(2):140–159.
3. Buntman, N. V., Anna A. Zaliznyak, I. M. Zatsman, M. G. KruzHKov, E. Yu. Loshchilova, and D. V. Sichinava. 2014. Informatsionnye tekhnologii korpusnykh issledovaniy: printsipy postroeniya krosslingvisticheskikh baz dannykh [Information technologies for corpus studies: Underpinnings for cross-linguistic database creation]. *Informatika i ee Primeneniya — Inform. Appl.* 8(2):98–110.
4. Buntman, N., J. L. Minel, D. Le Pesant, and I. Zatsman. 2010. Typology and computer modeling of translation difficulties. *Informatika i ee Primeneniya — Inform. Appl.* 4(3):77–83.
5. Ide, N., and J. Pustejovsky, eds. 2017. *Handbook of linguistic annotation*. Dordrecht, The Netherlands: Springer Science + Business Media. 1468 p.

6. Segalovich, I. 2003. A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine. *Conference (International) on Machine Learning: Models, Technologies and Applications Proceedings*. Las Vegas, NV: CSREA Press. 273–280.
7. Zobnin, A. I. and G. V. Nosyrev. 2015. Morfologicheskii analizator MyStem 3.0 [Morphological analyzer MyStem 3.0]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [Proceedings of the V. V. Vinogradov Russian Language Institute] 6:300–310.
8. Goncharov, A. A., and I. M. Zatsman. 2019. Informatsionnye transformatsii parallel'nykh tekstov v zadachakh izvlecheniya znaniy [Information transformations of parallel texts in knowledge extraction]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):180–193.
9. Inkova, O. Yu. 2018. Lingvospetsifichnost' konnektorov: metody i parametry opisaniya [Language specificity of connectives methods and parameters of description]. *Semantika konnektorov: kontrastivnoe issledovanie* [Semantics of connectives: A contrastive study]. Moscow: TORUS PRESS. 5–23.
10. Johansson, S. 2007. *Seeing through Multilingual Corpora: On the use of corpora in contrastive studies*. Amsterdam: John Benjamins B. V. 355 p.
11. Rogozhnikova, R. P. 2003. *Tolkovyy slovar' sochetaniy, ekvivalentnykh slovu: Okolo 1500 ustoychivyykh sochetaniy russkogo yazyka* [Dictionary of word-equivalent combinations: Around 1500 Russian fixed word combinations]. Moscow: Astrel': AST. 416 p.
12. Zatsman, I. M., O. Yu. Inkova, M. G. Kruzhkov, and N. A. Popkova. 2016. Predstavlenie krossyazykovykh znaniy o konnektorakh v nadkorporusnykh bazakh dannykh [Representation of cross-lingual knowledge about connectors in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):106–118.
13. Zatsman, I. M., O. Yu. Inkova, and V. A. Nuriev. 2017. The construction of classification schemes: Methods and technologies of expert formation. *Automatic Documentation Math. Linguistics* 51(1):27–41.
14. Nuriev, V., N. Buntman, and O. Inkova. 2018. Machine translation of Russian connectives into French: Errors and quality failures. *Informatika i ee Primeneniya — Inform. Appl.* 12(2):105–113.
15. Buntman, N. V., A. A. Goncharov, I. M. Zatsman, and V. A. Nuriev. 2018. Kolichestvennyy analiz rezul'tatov machinnogo perevoda s ispol'zovaniem nadkorporusnykh baz dannykh [Using supracorpora databases for quantitative analysis of machine translations]. *Informatika i ee Primeneniya — Inform. Appl.* 12(4):96–105.
16. Zatsman, I. M., and M. G. Kruzhkov. 2018. Nadkorporusnaya baza dannykh konnektorov: razvitie sistemy terminov proektirovaniya [Supracorpora database of connectives: Design-oriented evolution of the term system]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 28(4):156–167.
17. Beninca, P., and J. Haiman. 1992. *The Rhaeto-Romance languages*. London: Routledge. 260 p.
18. Van Dyk, J. 2009. Language learning through sight translation. *Translation in second language learning and teaching*. Bern: Peter Lang. 203–214.
19. Dobrovol'skiy, D. O., and Anna A. Zaliznyak. 2018. Nemetskie konstruksii s modal'nymi glagolami i ikh russkie sootvetstviya: proekt nadkorporusnoy bazy dannykh

[German constructions with modal verbs and their Russian correlates: A supracorpora database project]. *Computational Linguistics and Intellectual Technologies: Conference (International) "Dialogue" Proceedings*. Moscow: RGGU. 17(24):172–184.

Received March 15, 2019

Contributors

Goncharov Alexander A. (b. 1994) — junior 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; a.gonch48@gmail.com

Inkova Olga Yu. (b. 1965) — Doctor of Science in philology, 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; olyainkova@yandex.ru

KruzHKov Mikhail G. (b. 1975) — 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; magnit75@yandex.ru

ПРИНЦИПЫ ОРГАНИЗАЦИИ ДАННЫХ ДЛЯ ТЕХНОЛОГИИ ПОДДЕРЖКИ КОНКРЕТНО-ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЙ

И. М. Адамович¹, О. И. Волков²

Аннотация: Статья продолжает серию работ, посвященных технологии поддержки конкретно-исторических исследований (ПКИИ). Технология ПКИИ построена на принципах сотворчества и краудсорсинга с опорой на систему автоматического извлечения фактов из текстов историко-биографической направленности T-парсер, а также инструментальный комплекс π -Фактограф, обеспечивающий поддержку исследовательского процесса, и ориентирована на широкий круг не являющихся профессиональными историками и биографами пользователей. Данная статья посвящена описанию и обоснованию принципов физического хранения информации в распределенной среде технологии ПКИИ, предотвращающих неограниченный рост объемов центрального хранилища документов без потерь важной информации. Проверка эффективности предложенных мер осуществляется методом имитационного моделирования. Обосновывается проверка адекватности модели за счет сопоставления и анализа результатов ее прогонов с различными значениями атрибутов документов, отражающих их общественную ценность. По результатам проверки делается обоснованный вывод об эффективности предложенных принципов организации данных для технологии ПКИИ.

Ключевые слова: конкретно-историческое исследование; распределенная технология; частично реплицированное размещение данных; модель распределенной системы; центральное хранилище

DOI: 10.14357/08696527190214

1 Введение

Поддержка конкретно-исторических исследований стала одной из актуальных задач современности в связи с вовлечением в исследовательский процесс не только членов профессионального исторического сообщества, но и самых широких слоев непрофессионалов, проявляющих растущий интерес к частной, семейной истории [1].

В ФИЦ ИУ РАН была разработана концепция комбинированной распределенной среды для ПКИИ, описана соответствующая информационная тех-

¹Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, Adam@amsd.com

²Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, Volkov@amsd.com

нология (далее — технология ПКИИ) и создан ряд программных средств ее поддержки [2].

Технология была построена на принципах сотворчества и краудсорсинга (мобилизации ресурсов широкого круга добровольцев посредством информационных технологий). Такая новая, но перспективная форма вовлечения общественности в процессы решения научных задач получила название «краудсайенс» (гражданская наука), в основе которой лежит научное волонтерство [3].

В основу технологии ПКИИ исходно была положена система Т-парсер [4], осуществляющая автоматическое, адаптированное к специфике биографического поиска извлечение фактов из текстов историко-биографической направленности и решающая задачи фактографического индексирования источников и их семантической разметки (вычленения значимых фрагментов и оснащения их метаданными). Соответственно, под данными в технологии ПКИИ на первом этапе подразумевались фрагменты текстов историко-биографической направленности. По причине компактности текстовой информации суммарный объем данных технологии ПКИИ, несмотря на предполагаемый значительный по интенсивности поток добавляемых в систему документов, исходящий от широкого круга не являющихся профессиональными историками и биографами пользователей, предполагался достаточно скромным и допускающим его централизованное хранение без значительных финансовых и организационных затрат на его содержание и администрирование. Это означает, что на первом этапе развития технологии ПКИИ в качестве принципа организации данных было обоснованно выбрано их централизованное размещение.

Но на втором этапе развития технологии ПКИИ в нее был интегрирован инструментальный комплекс π -Фактограф [5], ориентированный на поддержку исследовательского процесса и решающий следующие задачи:

- фиксация не только четко установленных фактов, но и предположений, неточных значений, вопросов;
- соотнесение высказывания с ресурсами-источниками или/и цепочками вывода, с одной стороны, и целевыми информационными ресурсами, в которых они публикуются, с другой;
- поддержка сочетания формализованного и неформализованного представления знаний;
- отслеживание процессов накопления данных, выявления дефектов, выдвижения/опровержения гипотез по источникам, времени, исполнителям, аргументации.

Для эффективного решения этих задач в основу π -Фактографа положена π -модель данных, опирающаяся на концепцию информационного поля, включающего также документы-источники, целевые материалы, промежуточные данные. Формализованные данные играют роль метаданных, сопровождающих основной массив документов, с их помощью может быть организован эффективный поиск. Факты, отражающие как действия исследователя, так и сведения

о предметной области, сопровождаются неформальным комментарием, в качестве которого может выступать не только текст, но и рисунок или произвольный мультимедийный объект. Этот комментарий может содержать цитату или изложение источника, а также отсылки на документ или конкретное место в документе, который, в свою очередь, может содержать отсылки к фактам. Таким образом формируется многосвязное информационное поле, содержащее мультимедийные данные самой разной природы, что подразумевает их значительный суммарный объем. Это означает существенную модификацию технологии ПКИИ в части организации данных.

Была разработана и обоснована единая логическая модель данных технологии ПКИИ [6]. Принципы же физического хранения информации в связи со значительным увеличением ее суммарного объема также должны быть полностью пересмотрены.

2 Возможные подходы к организации хранения информации

Подсистема хранения данных технологии ПКИИ в условиях ее одновременного пополнения и использования со стороны множества независимых исследователей представляет собой род распределенной базы данных (РБД), поскольку данные связаны единой логической структурой, а доступ к ним обеспечивается единым высокоуровневым интерфейсом в рамках единой технологии. Для РБД выделяют следующие типы стратегии размещения данных [7]:

- централизованное размещение данных — вся БД хранится на одном узле;
- секционированное размещение данных — каждый фрагмент БД хранится на определенном узле;
- реплицированное размещение данных — одна или более копий фрагментов БД хранятся на нескольких узлах.

Наиболее простым с организационной точки зрения решением представляется централизованное размещение данных. Но суммарный объем данных в централизованном хранилище (ЦХ) в этом случае может быть значительным. Так, архивный фонд одного пользователя технологии может содержать десятки тысяч графических объектов (фотографий, сканов документов и т. п.) [8], а число волонтеров, участвующих в крупном сетевом проекте, может достигать сотен тысяч [9], что позволяет грубо оценить возможный объем БД в 1 ПБайт. Организация хранения, резервирования и администрирования единой БД такого объема требует значительных затрат, что вряд ли согласуется с некоммерческим характером деятельности волонтеров. Таким образом, использование централизованного размещения данных в технологии ПКИИ представляется нецелесообразным.

Секционированное размещение, при котором данные каждого пользователя хранятся исключительно локально и предоставляются другим пользователям в периоды их сетевой доступности по запросу, а центральная БД служит исключительно для регистрации пользователей и хранения отложенных запросов,

практически не требует затрат для своей организации. Но при таком подходе задача хранения и администрирования локальной БД полностью возлагается на пользователя. Поскольку пользователь технологии ПКИИ является волонтером, участвующим в проекте на некоммерческой основе, невозможно требовать от него поддержания каких-либо специальных мер, обеспечивающих сохранность и постоянную доступность информации для других пользователей, если это по каким-либо причинам не соответствует его личным интересам. Неизбежно пользователь будет время от времени выключать свой компьютер, содержащий ценную историческую информацию, а также удалять информацию, потерявшую ценность лично для него, после завершения исследования.

Таким образом, поскольку технология ПКИИ должна обеспечивать хранение документов, представляющих ценность для многих пользователей и их доступность по информационным запросам, организация хранения информации в ней должна базироваться на принципах частичной репликации. В качестве критерия эффективности организации хранения будем использовать суммарное среднее время выполнения запросов [10].

3 Описание принципов организации данных в технологии поддержки конкретно-исторических исследований

Поскольку, как показано выше, централизованное хранение копий всех документов в технологии ПКИИ невозможно, необходимо обеспечить репликацию той их части, которая представляет интерес не только для пользователя, разместившего данный документ в БД, но и для других пользователей. Наличие в БД документов различной общественной ценности вытекает из характера личных фондов [11]. Проблема состоит в том, что средствам технологии ПКИИ заранее не может быть известна степень интереса ко вновь добавленному документу со стороны прочих пользователей. Поэтому представляется целесообразным каждый новый документ воспринимать как представляющий общественный интерес и производить его репликацию в ЦХ. Далее по результатам анализа информационных запросов корректировать эту оценку и при необходимости принимать решение об удалении реплики из ЦХ.

В результате предлагается ввести в технологию ПКИИ параметр Δt — срок хранения документа в ЦХ после его последнего использования (первичной загрузки в ЦХ или информационного запроса). Любой документ, загруженный пользователем, остается у пользователя на локальном компьютере, а также временно помещается в ЦХ. Информационный запрос на выдачу документа и создание его локальной копии на компьютере автора запроса в случае наличия этого документа в ЦХ выполняется немедленно либо осуществляется с локальной копии одного из пользователей, уже имеющих этот документ. Поскольку пользователи могут периодически выключать локальный компьютер, информационный запрос не всегда может быть выполнен немедленно, и в этом случае он становится

отложенным запросом, который будет исполнен, как только компьютер любого из пользователей, имеющих локальную копию данного документа, станет доступен по сети для средств технологии ПКИИ.

Если число локальных копий документа превышает некоторое пороговое значение M , то его хранение в ЦХ не имеет смысла, поскольку вероятность его получения из какой-либо локальной копии весьма высока.

Период времени, в течение которого некоторый исследователь является пользователем технологии ПКИИ, неизбежно ограничено. Рано или поздно его исследование будет закончено и его интерес к технологии может пропасть. К сожалению, невозможно требовать от человека, участвующего в общем процессе на исключительно добровольных началах, соблюдения каких-либо обязательных процедур для корректного завершения сотрудничества, как то: передачи локальных файлов в ЦХ, оповещения других пользователей о своем выходе из проекта и т. д. Поэтому в технологию ПКИИ вводится параметр ΔT — предельный период времени неактивности пользователя, при превышении которого его следует считать выбывшим из проекта, а его локальные копии документов — более не доступными для средств технологии ПКИИ.

Данные меры могут позволить предотвратить неограниченный рост объема ЦХ, избегая при этом потерь важной информации, но это предположение требует проверки.

4 Модель организации данных технологии поддержки конкретно-исторических исследований

Проверка была осуществлена методами имитационного моделирования. Модель реализована на принципах дискретно-событийного моделирования.

Моделируется стационарная стадия развития проекта, когда число активных участников примерно постоянно. При этом учитывается, что даже в самых устойчивых больших виртуальных коллективах за год персональный состав обновляется как минимум на 50%–60% [12].

В качестве параметров модели заданы:

T_{exp} — продолжительность всего эксперимента в днях модельного времени;

N — текущее число участников проекта;

T_p — средняя длительность периода участия в проекте для пользователя;

P_{act} — средняя доля периода активности в сутках для одного участника;

D — среднее количество добавляемых документов в день для одного участника;

Q — среднее число запросов одного участника в день;

M — минимальное число локальных копий документа для его исключения из ЦХ;

Δt — максимальный срок хранения документа в ЦХ после его последнего использования;

ΔT — длительность периода неактивности пользователя, при превышении которого его следует считать выбывшим из проекта;

ΔT_q — срок актуальности отложенного запроса, после истечения которого запрос считается неисполненным.

Каждый добавляемый документ D имеет атрибут I_D , выражающий его «интересность» — параметр, определяющий вероятность составления информационного запроса пользователем на поиск данного документа. Многие исследователи отмечали, что ранговое распределение документов по популярности следует закону Зипфа, т. е. убывает степенным образом с номером ранга документа. Как показано в [13], распределение по популярности соответствует зипфоподобному ранговому распределению с показателем α :

$$f_r \approx r^{-\alpha}.$$

При этом значение α очень близко к 1.

Целью моделирования ставилось выявление динамики следующих показателей:

$Q_D(t)$ — количество документов, хранящихся в ЦХ;

$D_L(t)$ — доля «списанных» документов, хранение которых прекращено в связи с невостребованностью;

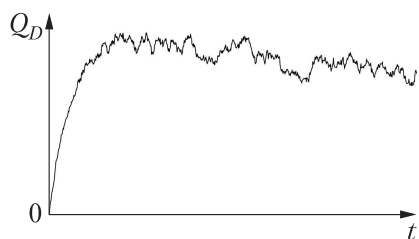
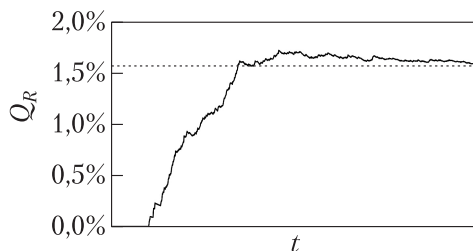
$Q_R(t)$ — доля запросов, признанных неисполненными;

$T_R(t)$ — среднее время выполнения для исполненных запросов.

5 Результаты проверки

5.1. Проверка показала, что показатель $Q_D(t)$ (количество документов, хранящихся в ЦХ) к сроку примерно $6\Delta t$, т. е. довольно быстро, достигает стадии насыщения. На рис. 1 приведен график, демонстрирующий динамику показателя $Q_D(t)$. Из него видно, что основная задача предложенных мер достигнута: поддержка функционирования технологии не сопряжена со значительными финансовыми и организационными затратами на содержание и администрирование центрального хранилища информации.

5.2. Показатель $D_L(t)$ (доля «списанных» документов, хранение которых прекращено в связи с невостребованностью) вычислялся с целью оценки адекватности концепции, положенной в основу модели, т. е. степени соответствия модели тому реальному явлению или объекту, для описания которого она строится. Допущение, требующее проверки на адекватность, в концепции модели технологии ПКИИ только одно: атрибут документа I_D с зипфоподобным ранговым распределением с показателем степени, близким к 1, реально отражает степень интереса к документу


Рис. 1 Динамика показателя $Q_D(t)$

Рис. 2 Динамика показателя $Q_R(t)$

со стороны участников, следствием чего является тот факт, что в «списанные» документы попадают преимущественно документы, неинтересные для большинства пользователей.

Для проверки этого факта была проведена серия прогонов модели с различающимися вдвое значениями Δt как с учетом атрибута I_D при формировании запроса (эксперимент 1), так и без его учета (эксперимент 2). Для каждого эксперимента вычислялось уменьшение количества списанных документов при удвоении Δt .

Проверка показала, что во втором эксперименте уменьшение составило на 51,14% бóльшую величину по сравнению с первым. Это говорит о том, что факт прекращения или продолжения хранения документа в системе зависит прежде всего от его ценности для сообщества исследователей, а не от искусственно выбранных параметров системы, отвечающих за срок нахождения документа в ЦХ.

- 5.3. Проверка показала, что показатель $Q_R(t)$ (доля запросов, признанных неисполненными) не превышает 2%. На рис. 2 приведен график, демонстрирующий динамику показателя $Q_R(t)$. Из него видно, что доля неисполненных запросов незначительна, быстро достигает точки насыщения и далее стабилизируется. Это доказывает эффективность выбранных мер по организации данных для технологии ПКИИ.
- 5.4. Проверка показала, что показатель $T_R(t)$ (среднее время выполнения для исполненных запросов) достигает значения 0,504 ч модельного времени и далее не имеет тенденции к росту по мере развития системы и добавления в нее новых документов. Это наглядно представлено на графике динамики показателя $T_R(t)$ (рис. 3). Получасовое ожидание доступа к историческому документу, обеспечиваемое технологией ПКИИ, существенно меньше обычного времени ожидания заказанных документов в архивах и библиотеках и считается психологически приемлемым. Это также доказывает эффективность выбранных мер по организации данных для технологии ПКИИ.

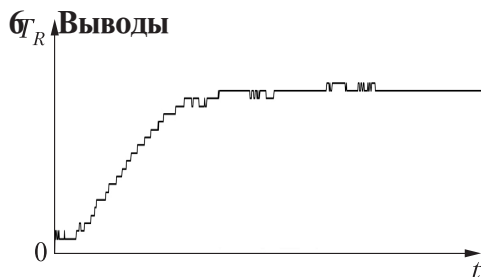


Рис. 3 Динамика показателя $T_R(t)$

Экспериментальная проверка показала, что предложенные принципы организации данных для рассматриваемой технологии, учитывающие большой объем разнородной информации, интенсивно добавляемой в систему сообществом исследователей, участвующих в некоммерческом проекте по ПККИ на принципах сотворчества и краудсорсинга, могут быть весьма

эффективны. Они обеспечивают возможность интенсивного обмена информацией между исследователями, не требуя при этом существенных затрат на поддержку централизованной инфраструктуры технологии ПККИ и не накладывая на членов исследовательского сообщества, участвующих в проекте на добровольных началах, обременительных требований организационного характера. Это важный фактор применимости технологии ПККИ, поскольку она предназначена для широкого круга не являющихся профессиональными историками и биографами пользователей.

Литература

1. Грибач С. В. Исследование семейных кризисов посредством психолингвистического эксперимента // Сб. конф. НИЦ Социосфера, 2010. № 6. С. 45–54.
2. Адамович И. М., Волков О. И. Технология распределенного автоматизированного анализа исторических текстов // Системы и средства информатики, 2016. Т. 26. № 3. С. 148–161. doi: 10.14357/08696527160311.
3. Егерев С. В. От краудсорсинга простых операций к «науке граждан» // Социология науки и технологий, 2016. № 4. Т. 7. С. 74–85.
4. Адамович И. М., Волков О. И. Система извлечения биографических фактов из текстов исторической направленности // Системы и средства информатики, 2015. Т. 25. № 3. С. 235–250. doi: 10.14357/08696527150315.
5. Маркова Н. А. Технология поддержки конкретно-исторических исследований на основе модели фактоподобных высказываний // Программная инженерия, 2015. № 5. С. 43–48.
6. Адамович И. М., Волков О. И. Единая технология поддержки конкретно-исторических исследований // Системы и средства информатики, 2019. Т. 29. № 1. С. 194–205. doi: 10.14357/08696527190116.
7. Новосельский В. Б., Павловская Т. А. Выбор и обоснование критерия эффективности при проектировании распределенных баз данных // Научно-технический вестник СПбГУ ИТМО, 2009. № 2(60). С. 76–82.
8. Адамович И. М., Волков О. И., Маркова Н. А. Технология работы с электронным семейным архивом // Вестник Воронежского государственного университета. Сер. Системный анализ и информационные технологии, 2012. № 1. С. 182–187.

9. *Маркова Н. А.* Формализация фактоподобных высказываний в конкретно-исторических исследованиях // Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL'2014): Тр. XVI Всеросс. научн. конф. — Дубна: ОИЯИ, 2014. С. 98–103.
10. *Телятников А. О.* Моделирование и оптимизация распределенных баз данных // Научные труды Донецкого национального технического университета. Сер. Вычислительная техника и автоматизация, 2006. Вып. 107. С. 95–103.
11. *Шпагина М. П.* Личные фонды: трансформация феномена // Документ в системе социальных коммуникаций: Сб. мат-лов III Всеросс. научн.-практич. конф. с международным участием. — Томск: ТГУ, 2008. С. 152–155.
12. *Нестеров В. Ю.* К вопросу о динамике сетевых сообществ, 2000. <http://flogiston.ru/articles/netpsy/groupdyn>.
13. *Крашаков С. А., Теслюк А. Б., Щур Л. Н.* Об универсальности рангового распределения популярности веб-серверов // Вестник РФФИ, 2004. № 1(35). С. 46–66.

Поступила в редакцию 15.10.18

THE PRINCIPLES OF DATA ORGANIZATION FOR THE TECHNOLOGY OF CONCRETE HISTORICAL RESEARCH SUPPORT

I. M. Adamovich and O. I. Volkov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation

Abstract: This article continues the series on technology of concrete historical investigation support which was built on the principles of co-creation and crowdsourcing and based on the system of automatic facts extraction from the historical and biographical texts T-parser and program complex π -Factograph oriented to the research process support and designed for a broad range of users which are not professional historians and biographers. The article is devoted to the description and reasoning of the data physical storage using the distributed technology environment principles which prevent unlimited increase of central storage volume without losses of important information. The efficiency of proposed measures was checked with the use of the simulation modeling method. The method of model adequacy checked by the results of comparison and analysis of its runs with the different values of such document's attribute as its public value was rationalized. As a result of the above tests, the reasonable conclusion about the efficiency of the proposed data physical storage principles for the technology of concrete historical research support was made.

Keywords: concrete historical research; distributed technology; partially replicated data placement; model of distributed system; central storage

DOI: 10.14357/08696527190214

References

1. Gribach, S. V. 2010. Issledovanie semeynykh krizisov posredstvom psikholingvisticheskogo eksperimenta [The study of family crises through a psycholinguistic experiment]. *Sborniki konf. NITs Sotsiosfera* [Conference NIC Sociosfera Proceedings] 6:45–54.
2. Adamovich, I. M., and O. I. Volkov. 2016. Tekhnologiya raspredelenogo avtomatizirovannogo analiza istoricheskikh tekstov [The distributed automated technology of historical texts analysis]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(3):148–161. doi: 10.14357/08696527160311.
3. Egerev, S. V. 2016. Ot kraudsoringa prostykh operatsiy k “nauke grazhdan” [From SciSourcing of simple operations to the citizen science]. *Sotsiologiya nauki i tekhnologii* [Sociology Science Technology] 7(4):74–85.
4. Adamovich, I. M., and O. I. Volkov. 2015. Sistema izvlecheniya biograficheskikh faktov iz tekstov istoricheskoy napravlenosti [The system of facts extraction from historical texts]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 2(25):235–250. doi: 10.14357/08696527150315.
5. Markova, N. A. 2015. Tekhnologiya podderzhki konkretno-istoricheskikh issledovaniy na osnove modeli faktopodobnykh vyskazyvaniy [Support technology for specific historical studies on the base of fact-like propositions model]. *Programmaya inzheneriya* [Software Engineering] 5:43–48.
6. Adamovich, I. M., and O. I. Volkov. 2019. Edinaya tekhnologiya podderzhki konkretno-istoricheskikh issledovaniy [Unified technology of concrete historical research support]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(1):194–205. doi: 10.14357/08696527190116.
7. Novoselsky, V. B., and T. A. Pavlovskaya. 2009. Vybory i obosnovaniye kriteriya effektivnosti pri proektirovaniy raspredelennykh baz dannykh [Choice and justification of effectiveness criterion in distributed databases design]. *Nauchno-tekhnicheskiiy vestnik SPbGU ITMO* [Scientific Technical J. Information Technologies Mechanics Optics] 2(60):76–82.
8. Adamovich, I. M., O. I. Volkov, and N. A. Markova. 2012. Tekhnologiya raboty s elektronnym semeynym arkhivom [Family archival fund]. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Ser. Sistemnyy analiz i informatsionnyye tekhnologii* [Proceedings of Voronezh State University. Ser. Systems Analysis and Information Technologies] 1:182–187.
9. Markova, N. A. 2014. Formalizatsiya faktopodobnykh vyskazyvaniy v konkretno-istoricheskikh issledovaniyakh [Formalization of the fact-like propositions in specific historical studies]. *Elektronnyye biblioteki: perspektivnyye metody i tekhnologii, elektronnyye kolleksii: Tr. 16-y Vseross. nauchn. konf.* [Digital Libraries: Advanced Methods and Technologies, Digital Collections: 16th All-Russian Scientific Conference Proceedings]. Dubna: JINR. 98–103.
10. Telyatnikov, A. O. 2006. Modelirovaniye i optimizatsiya raspredelennykh baz dannykh [Modeling and optimization of the distributed databases]. *Nauchnyye trudy Donetskogo natsional'nogo tekhnicheskogo universiteta. Ser. Vychislitel'naya tekhnika i avtomatizatsiya* [Proceedings of Donetsk National Technical University. Ser. Computing Technology and Automation] 107:95–103.
11. Shpagina, M. P. 2008. Lichnyye fondy: transformatsiya fenomena [Personal funds: Phenomenon transformation]. *Dokument v sisteme sotsial'nykh kommunikatsiy: Sb.*

- mat-lov 3-y Vseross. nauchn.-praktich. konf. s mezhdunarodnym uchastiem* [Document in the system of social communications: 3rd All-Russian Academic and Practical Conference with the International Participation Proceedings]. Tomsk: TSU. 152–155.
12. Nesterov, V. Yu. 2000. K voprosu o dinamike setevykh soobshchestv [On the issue of the dynamics of network communities]. Available at: <http://flogiston.ru/articles/netpsy/groupdyn> (accessed October 7, 2018).
 13. Krashakov, S. A., A. B. Teslyuk, and L. N. Shchur. 2004. Ob universal'nosti rangovogo raspredeleniya populyarnosti veb-serverov [On the universality of rank distributions of website popularity]. *Vestnik RFFI* [Bull. RFBR] 1(35):46–66.

Received October 15, 2018

Contributors

Adamovich Igor M. (b. 1934) — Candidate of Science (PhD) 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 119133, Russian Federation; Adam@amsd.com

Volkov Oleg I. (b. 1964) — leading programmer, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; Volkov@amsd.com

МОДЕЛЬ КООПЕРАТИВНОГО РЕШАТЕЛЯ ЗАДАЧ НА ОСНОВЕ ЦИФРОВЫХ ДВОЙНИКОВ

В. Д. Ильин¹

Аннотация: Предложена модель кооперативного решателя задач на основе цифровых двойников. Участники кооперации (названные dt-инфами) рассматриваются как обучающиеся машины для решения задач. Dt-инф, наделенный конечным множеством состояний и целью (понимаемой как множество разрешимых задач предметной области), в процессах решения задач взаимодействует с другими членами кооперации и элементами окружения путем обмена унифицированными сообщениями. Предметная область (для каждого dt-инфа) представлена задачным графом, рассчитанным на поиск разрешающих структур и формализацию знаний о задачах. Множество вершин графа составлено из задачных конструктивных объектов. Каждая вершина имеет память, представленную памятью задачи (простой или составной) или задачной области. Ребро задачного графа — пара вершин с непустым пересечением по памяти. Нагрузка ребра определяется множеством всех пар элементов памяти, входящих в пересечение. Предложенная модель рассчитана на применение при разработках информационных технологий и онлайн-сервисов.

Ключевые слова: цифровые двойники; машина для кооперативного решения задач; задачный конструктивный объект; задачный граф; информационная технология; онлайн-сервис

DOI: 10.14357/08696527190215

1 Введение

В наши дни для решения задач в различных сферах деятельности успешно применяются М2М-технологии (*англ.* Machine-to-Machine) [1, 2], технологии облачных вычислений (*англ.* cloud computing) и электронных сервисов [3–8], Интернета вещей (*англ.* Internet of Things, IoT) [9, 10] и цифровых двойников (*англ.* digital twins) [11].

Осознание актуальной цели в процессе некоторой деятельности не всегда сопровождается ясным представлением о задачах, которые должны быть решены на пути к цели. При достижении нетривиальных целей (спроектировать машину, разработать информационную технологию и т. д.) наиболее продуктивным оказывается символическое представление замысла в *человеко-машинной среде решения задач (s-среде)* [12]. Такой подход имеет ряд проверенных практикой достоинств. Во-первых, анализируя символическую модель замысла (чертеж, спецификацию задачи и пр.), можно проверить, соответствует ли модель замыслу, и, если не

¹Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, vdilyin@yandex.ru

соответствует, внести исправления в модель. Во-вторых, на модели, признанной соответствующей замыслу, можно провести проверку состоятельности самого замысла. И если проверка завершилась успехом, принять решение о целесообразности реализации замысла. А в противном случае — заняться изменением замысла.

В 1989 г. в монографии [13] была предложена *методология символьного воплощения замысла в s-среде* применительно к порождению программ (автоматизированному конструированию программ на основе системы знаний о программируемых задачах). Идея достижения цели с использованием символических моделей воплощена во многих технологиях для различных видов деятельности. Наиболее успешными становятся те технологии, которые исключают поведение, не соответствующее правилам (попытки недопустимого поведения имеют столько же шансов на успех, сколько попытки сыграть в покер с шахматной программой).

В наши дни существует огромное число задач, запрограммированных на различных языках и реализованных в составе системных и прикладных комплексов. Какие это задачи, к каким классам отнесены? Где можно найти сведения об их формулировках, методах и алгоритмах решения и тестовых примерах? Как использовать этот арсенал? Усовершенствованный подход к представлению знаний о программируемых задачах и поиску разрешающих структур на задачных графах был предложен в [14].

1.1 Выделение фрагментов текста статьи

Для выделения определений, замечаний и примеров используются следующие средства:

□ ⟨фрагмент описания⟩ □ ≈ утверждение (определение, аксиома и др.) (здесь и далее символ ≈ заменяет слово «означает»);

◇ ⟨фрагмент описания⟩ ◇ ≈ замечание;

☞ ⟨фрагмент описания⟩ ☞ ≈ пример.

Курсивом выделены первые вхождения названий понятий и фрагменты описания, к которым автор хочет привлечь внимание.

1.2 Цифровой двойник

□ *Цифровой двойник* (далее — dt) — это *символьная модель некоторого объекта* (технического устройства, деятельности или организации), *предназначенная для анализа и совершенствования его поведения*. □

Для построения dt используют значения параметров, характеризующих поведение моделируемого объекта (☞ для технических устройств — данные, поступающие от соответствующих датчиков ☞). С момента создания dt (как образа объекта) он накапливает знания о своем прообразе, обмениваясь с ним сообщениями и обновляясь.

Различают цифровые двойники-прототипы (Digital Twin Prototype, DTP), цифровые двойники-экземпляры (Digital Twin Instance, DTI) и цифровые аг-

регированные двойники (Digital Twin Aggregate, DTA): DTP используется при создании физической версии моделируемого объекта; DTI — модель существующего объекта, с которым DTI предназначена взаимодействовать; DTA — комплекс средств обработки данных, предназначенный для взаимодействия с заданной совокупностью DTI.

◇ Технология *Digital Twin* построения цифровой *символьной модели* производного объекта обязательно предполагает возможность получения данных о его поведении (путем измерения значений некоторой совокупности параметров объекта). ◇

2 Dt-инф — обучаемая машина для решения задач

Будем рассматривать dt-инф с двух точек зрения. Во-первых, как информационный автомат, взаимодействующий с себе подобными автоматами и с *элементами окружения* (далее — enels) путем посылки сообщений. Такое представление нужно, чтобы, не вникая в механизмы решения задач, работать с dt-инфом как с активным символьным объектом, реализованным в s-среде.

Второй облик dt-инфа используем, чтобы раскрыть его предназначение как средства решения задач определенной предметной области. Реализация модели предметной области, интерпретация запросов на модели — такие вопросы будем рассматривать на основе второго представления.

□ *Информационная модель* dt-инфа — машина с конечным множеством состояний S , обладающая *целью* (понимаемой как множество задач, разрешимых на *задачном графе* G [14]). □

Множество T пар состояний, принадлежащих декартову произведению $S * S$, задает отношение допустимых переходов. Пара (s_i, s_j) принадлежит T , если из состояния s_i разрешен переход в состояние s_j ; $i, j = 1, \dots, n_S$ (n_S — число состояний).

Задачный граф G служит формальным представлением *задачной области*, рассчитанным на поиск *разрешающих структур* и формализацию знаний о задачах. Множество вершин графа составлено из *задачных конструктивных объектов*. Каждая вершина графа имеет память, представленную памятью задачи (или *задачной области*). *Ребро задачного графа* — это пара вершин с непустым пересечением по памяти. *Нагрузка ребра* определяется множеством всех пар элементов памяти, входящих в это пересечение [14].

Будем называть dt-инф *вычислительным*, если каждому s_i из множества S состояний алгоритмически поставлено в соответствие не более одного s_j ($i \neq j$). Если же каждому s_i из множества S может быть поставлено в соответствие более одного s_j ($i \neq j$), dt-инф будем называть *исчислительным*.

Поведение dt-инфа определяется следующими требованиями:

- принимать и отправлять сообщения, взаимодействуя с другими dt-инфами и enels;

- интерпретировать сообщения, выдавая сведения об ошибках и допуская коррекцию принятых сообщений;
- выполнять работу, запрошенную в сообщении, допуская диалог, паузы и отмену;
- оформлять результат выполненной работы для отправки корреспонденту;
- образовывать объединения dt-инфов, предназначенные для решения комплексов задач.

Dt-инф работает как дискретная машина для обработки сообщений, безотносительно к содержанию работы, выполняемой в соответствии с поступившим запросом.

Такой взгляд на dt-инф соответствует намерению исследовать проблему создания кооперации dt-инфов.

Dt-инф как машина для обработки сообщений состоит из следующих компонент, составляющих конвейер (которые могут работать асинхронно):

- BEFORE — инициализация, запуск, обучение;
- START — прием, буферизация, интерпретация сообщений;
- STEP — шаг в процессе выполнения работы, запрошенной в сообщении;
- END — завершение работы с текущим сообщением;
- AFTER — завершение работы dt-инфа.

Одновременно с тем, как компонента BEFORE занята (☞ обучением пользователя ☞), компонента START может интерпретировать очередное сообщение, а STEP — выполнять работу, запрошенную в предыдущем сообщении.

Каждая компонента dt-инфа может находиться в одном из состояний, принадлежащих множеству

$$\text{State} = \{\text{head}, \text{tail}, \text{dial}, \text{intr}, \text{paus}, \text{canc}, \text{resl}, \text{fail}, \text{null}\},$$

где head — состояние готовности; tail — работа закончена; dial — взаимодействие с enels; intr — взаимодействие с другими dt-инфами; paus — пауза; canc — отмена; resl — результат получен; fail — неудача: результат получить не удалось; null — состояние, требующее непредусмотренного вмешательства отправителя сообщения (enel или dt-инфа).

Отношение перехода из одного состояния в другое задано для каждой компоненты соответствующей таблицей перехода состояний. Таблицы для всех компонент имеют одинаковую структуру: число строк и число столбцов равны числу состояний. Строки и столбцы имеют те же имена, что и соответствующие состояния. Если в клетке на пересечении строки st_i и столбца st_j стоит буква f , это означает, что переход из состояния st_i в состояние st_j невозможен. Если в клетке размещена ссылка на правило rule, определяющее переход ($st_i \rightarrow st_j$), то это означает, что разрешенный переход выполняется в соответствии с этим правилом.

Компоненты взаимодействуют в соответствии с правилами, хранящимися в базе правил взаимодействия компонент dt-инфа.

Dt-инф как решатель задач из некоторой предметной области должен располагать средствами для интерпретации сообщений и выполнения запрашиваемых работ. В состав dt-инфа, решающего задачи из заданной предметной области, входят:

- анализатор сообщений (запросов на решение задач);
- построитель разрешающих структур на задачном графе;
- система управления базами данных;
- формоориентированный процессор ввода–вывода;
- средства расширения множества разрешимых задач;
- функциональная компонента.

Язык сообщений лаконичен (необязательные синтаксические конструкции заключены в фигурные скобки "{ " и " } "):

⟨сообщение⟩ ::= ⟨конверт⟩ { ⟨содержимое⟩ }

⟨конверт⟩ ::= ⟨получатель⟩, ⟨отправитель⟩

⟨содержимое⟩ ::= найти ⟨выход⟩ по ⟨вход⟩ { при ⟨зависимость выход-вход⟩ } |
⟨разрешающая структура⟩ ⟨адрес входа⟩ ⟨адрес выхода⟩ }

Получатель и отправитель представлены именами dt-инфов с указанием координат портов сообщений. Выход и вход задаются именами понятий предметной области. Зависимость выход–вход задается в тех случаях, когда запрос представлен в виде формулировки задачи.

Разрешающая структура представлена именем; адрес входа указывает место, откуда следует прочитать входные данные; адрес выхода — куда следует записать полученный результат. Если один из адресов (или оба) не указан, это означает, что работа выполняется с использованием входных и выходных портов (для данных и результатов).

Обнаружив сообщение во входном порту сообщений, анализатор начинает синтаксический разбор. Если сообщение синтаксически верно, оно направляется построителю разрешающих структур. В противном случае отправителю выдаются сведения об ошибке (с предложением исправить сообщение). Когда получен искомый результат, он направляется отправителю по адресу, указанному на конверте.

Построитель разрешающих структур отыскивает на задачном графе разрешающую структуру, позволяющую по заданному входу найти искомый выход.

Система управления базами данных (СУБД) играет роль памяти dt-инфа. Все функции, свойственные СУБД, присущи памяти dt-инфа.

Формоориентированный процессор ввода–вывода представляет собой объединение структурно-ориентированных редакторов ввода, формирователей отчетов и средств вывода во внешнюю (по отношению к dt-инфу) среду.

Расширение множества разрешимых задач достигается увеличением числа вершин и ребер задачного графа. Такая работа выполняется в специальном режиме «расширение» на копии расширяемого задачного графа.

Целевое назначение dt-инфа представляет его функциональная компонента. Содержание задач, служащих вершинами задачного графа, связи по памяти между задачами, представленные ребрами графа, — все это определяет задачный облик функциональной компоненты.

3 Заключение

Предложенная модель кооперативного решателя задач основана на результатах, полученных при создании методологии порождения программ [13] и методологии построения человеко-машинной среды решения задач [12].

Каждый участник кооперации специализируется на определенном классе задач из заданной предметной области. Это могут быть задачи построения системы знаний, автоматизированного конструирования программ и пр.

Цифровые двойники служат средством совершенствования поведения кооперации. Они специализируются на анализе поведения отдельных участников и кооперации в целом, получая данные и отправляя результаты путем обмена сообщениями со своими прообразами. Для анализа используются обновляемые знания о прообразах.

Модель может служить методологическим основанием для проектирования кооперативного решателя задач на основе цифровых двойников.

Литература

1. Kim R. Y. Efficient wireless communications schemes for machine to machine communications // *Comm. Com. Inf. Sc.*, 2011. Vol. 181. No. 3. P. 313–323.
2. Lien S. Y., Liaw T. H., Kao C. Y., et al. Cooperative access class barring for machine-to-machine communications // *IEEE T. Wirel. Commun.*, 2012. Vol. 11. No. 1. P. 27–32.
3. Armbrust M., Fox A., Griffith R., et al. A view of cloud computing // *Commun. ACM*, 2010. Vol. 53. No. 4. P. 50–58. doi: 10.1145/1721654.1721672.
4. Wang L., Laszewski G., Younge A., et al. Cloud computing: A perspective study // *New Generat. Comput.*, 2010. No. 28. P. 137–146. doi: 10.1007/s00354-008-0081-5.
5. Wei Y., Blake M. B. Service-oriented computing and cloud computing: Challenges and opportunities // *IEEE Internet Comput.*, 2010. No. 14. P. 72–75.
6. Rogers O., Cliff D. A financial brokerage model for cloud computing // *J. Cloud Comput.*, 2012. Vol. 1. No. 1. P. 1–12. doi: 10.1186/2192-113X-1-2.
7. Ильин А. В. Интернет-сервис планирования расходов // *Системы и средства информатики*, 2015. Т. 25. № 2. С. 111–122.

8. Jede A., Teuteberg F. Understanding socio-technical impacts arising from software-as-a-service usage in companies // *Bus. Inf. Syst. Eng.*, 2016. Vol. 58. No. 3. P. 161–176. doi: 10.1007/s12599-016-0429-1.
9. Perera C., Liu C. H., Jayawardena S. The emerging Internet of Things marketplace from an industrial perspective: A survey // *IEEE T. Emerging Topics Comput.*, 2015. Vol. 3. No. 4. P. 585–598. doi: 10.1109/TETC.2015.2390034.
10. «Интернет вещей» (IoT) в России. Технология будущего, доступная уже сейчас // PWC, 2017. https://www.pwc.ru/ru/publications/iot/loT-inRussia-research_rus.pdf.
11. The Digital Twin. General Electric, 2018. https://www.ge.com/digital/sites/default/files/The-Digital-Twin_Compressing-Time-to-Value-for-Digital-Industrial-Companies.pdf.
12. Ильин А. В., Ильин В. Д. Создание человеко-машинной среды решения задач // *Системы и средства информатики*, 2016. Т. 26. № 4. С. 149–161.
13. Ильин В. Д. Система порождения программ. — М.: Наука, 1989. 264 с.
14. Ильин А. В. Конструирование разрешающих структур на задачных графах системы знаний о программируемых задачах // *Информационные технологии и вычислительные системы*, 2007. № 3. С. 30–36.

Поступила в редакцию 10.03.19

THE MODEL OF THE COOPERATIVE PROBLEM SOLVER BASED ON DIGITAL TWINS

V. D. Ilyin

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

Abstract: The model of cooperative problem solver based on digital twins is proposed. Participants of cooperation (called “dt-infs”) are considered as learnable machines for solving problems. In the processes of solving problems, such machine endowed with a finite set of states and a goal (understood as a set of solvable problems in the subject area) interacts with other members of cooperation and environmental elements through the exchange of unified messages. For each dt-inf, the subject area is represented by a task graph, designed to search for resolving structures and formalize knowledge of the tasks. The set of graph vertices is composed of task constructive objects. Each vertex has a memory represented by the task memory of the task (simple or composite) or the task area. An edge of the task graph is a pair of vertices with nonempty intersection by memory. Load of the edge is determined by the set of all pairs of memory elements included in the intersection. The proposed model is designed for use in the development of information technologies and online services.

Keywords: digital twins; cooperative problem solver; task constructive object; task graph; information technology; online service

DOI: 10.14357/08696527190215

References

1. Kim, R. Y. 2011. Efficient wireless communications schemes for machine to machine communications. *Comm. Com. Inf. Sc.* 181(3):313–323.
2. Lien, S. Y., T. H. Liau, C. Y. Kao, *et al.* 2012. Cooperative access class barring for machine-to-machine communications. *IEEE T. Wirel. Commun.* 11(1):27–32.
3. Armbrust, M., A. Fox, R. Griffith, *et al.* 2010. A view of cloud computing. *Commun. ACM* 53(4):50–58. doi: 10.1145/1721654.1721672.
4. Wang, L., G. Laszewski, A. Younge, *et al.* 2010. Cloud computing: A perspective study. *New Generat. Comput.* 28:137–146. doi: 10.1007/s00354-008-0081-5.
5. Wei, Y., and M. B. Blake. 2010. Service-oriented computing and cloud computing: challenges and opportunities. *IEEE Internet Comput.* 14:72–75.
6. Rogers, O., and D. Cliff. 2012. A financial brokerage model for cloud computing. *J. Cloud Comput.* 1(1):1–12. doi: 10.1186/2192-113X-1-2.
7. Ilyin, A. V. 2015. Internet-servis planirovaniya raskhodov [The online service for cost planning]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(2):111–122.
8. Jede, A., and F. Teuteberg. 2016. Understanding socio-technical impacts arising from software-as-a-service usage in companies. *Bus. Inf. Syst. Eng.* 58(3):161–176. doi: 10.1007/s12599-016-0429-1.
9. Perera, C., C. H. Liu, and S. Jayawardena. 2015. The emerging Internet of Things marketplace from an industrial perspective: A survey. *IEEE T. Emerging Topics Comput.* 3(4):585–598. doi: 10.1109/TETC.2015.2390034.
10. PWC. 2017. “Internet veshchey” (IoT) v Rossii. Tekhnologiya budushchego, dostupnaya uzhe seychas [“Internet of Things” (IoT) in Russia. Technology of the future, available now]. Available at: https://www.pwc.ru/ru/publications/iot/loT-inRussia-research_rus.pdf (accessed March 10, 2019).
11. General Electric. 2018. The digital twin. Available at: https://www.ge.com/digital/sites/default/files/The-Digital-Twin_Compressing-Time-to-Value-for-Digital-Industrial-Companies.pdf (accessed March 10, 2019).
12. Ilyin, A. V., and V. D. Ilyin. 2016. Sozdanie cheloveko-mashinnoy sredy resheniya zadach [Creation of a human–machine environment for problem solving]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 26(4):149–161.
13. Ilyin, V. D. 1989. *Sistema porozhdeniya programm* [The system for programs generating]. Moscow: Nauka. 264 p.
14. Ilyin, A. V. 2007. Konstruirovaniye razreshayushchikh struktur na zadachnykh grafakh sistemy znaniy o programmiruemykh zadachakh [Construction of solving structures in task graphs of knowledge system of programmable tasks]. *Informatsionnye tekhnologii i vychislitel'nye sistemy* [Information Technology and Computing Systems] 3:30–36.

Received March 10, 2019

Contributor

Ilyin Vladimir D. (b. 1937)— Doctor of Science in technology, professor, leading scientist, Dorodnicyn Computing Center of the Russian Academy of Sciences, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; vdilyin@yandex.ru

ОБ АВТОРАХ

Адамович Игорь Михайлович (р. 1934) — кандидат технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Архипов Александр Евгеньевич (р. 1976) — кандидат технических наук, старший научный сотрудник Научно-исследовательского института радиоэлектронных систем Юго-Западного государственного университета

Бобырь Максим Владимирович (р. 1978) — доктор технических наук, профессор Юго-Западного государственного университета

Волков Олег Игоревич (р. 1964) — ведущий программист Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Волович Константин Иосифович (р. 1970) — старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гаранин Александр Иванович (р. 1951) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Гончаров Александр Анатольевич (р. 1994) — младший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Грушо Александр Александрович (р. 1946) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Егоров Владимир Борисович (р. 1948) — кандидат технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацаринный Александр Алексеевич (р. 1951) — доктор технических наук, заместитель директора Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН); главный научный сотрудник Института проблем информатики ФИЦ ИУ РАН

Ильин Владимир Дмитриевич (р. 1937) — доктор технических наук, профессор, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Российской академии наук Федерального исследовательского центра «Информатика и управление» Российской академии наук

Инькова Ольга Юрьевна (р. 1965) — доктор филологических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Киселёва Надежда Николаевна (р. 1949) — доктор химических наук, руководитель лаборатории Института металлургии и материаловедения им. А. А. Байкова Российской академии наук

Колесников Александр Васильевич (р. 1948) — доктор технических наук, профессор Института физико-математических наук и информационных технологий Балтийского федерального университета имени Иммануила Канта; ведущий научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кондрашев Вадим Адольфович (р. 1963) — старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кружков Михаил Григорьевич (р. 1975) — старший 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

Листопад Сергей Викторович (р. 1984) — кандидат технических наук, старший научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Лукиянов Леонид Аркадьевич (р. 1992) — аспирант Башкирского государственного университета

Майоров Борис Геннадьевич (р. 1947) — кандидат технических наук, доцент, ведущий научный сотрудник АО «Научно-производственное предприятие «Рубин», г. Пенза

Майтаков Федор Георгиевич (р. 1973) — аспирант Института физико-математических наук и информационных технологий Балтийского федерального университета имени Иммануила Канта

Мальковский Сергей Иванович (р. 1983) — научный сотрудник Вычислительного центра Дальневосточного отделения Российской академии наук

Милостная Наталья Анатольевна (р. 1975) — кандидат технических наук, старший научный сотрудник Юго-Западного государственного университета

Морозкин Николай Данилович (р. 1953) — доктор физико-математических наук, ректор Башкирского государственного университета

Никишин Дмитрий Александрович (р. 1976) — кандидат технических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ожерельев Илья Сергеевич (р. 1994) — студент магистратуры факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Палионная Софья Игоревна (р. 1995) — студентка факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова

Сенько Олег Валентинович (р. 1957) — доктор физико-математических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Спивак Семен Израилевич (р. 1945) — доктор физико-математических наук, заведующий кафедрой Башкирского государственного университета

Стрижов Вадим Викторович (р. 1967) — доктор физико-математических наук, ведущий научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор Московского физико-технического института

Тимонина Елена Евгеньевна (р. 1952) — доктор технических наук, профессор, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Усманова Карина Равилевна (р. 1996) — студентка Московского физико-технического института

Шестаков Олег Владимирович (р. 1976) — доктор физико-математических наук, профессор 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

Шоргин Всеволод Сергеевич (р. 1978) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Правила подготовки рукописей статей для публикации в журнале «Системы и средства информатики»

Журнал «Системы и средства информатики» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информационных технологий.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- информационно-телекоммуникационные системы и средства их построения;
- архитектура и программное обеспечение вычислительных машин, комплексов и сетей;
- методы и средства защиты информации.

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 представленной 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 L^AT_EX (.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 и 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/publications/collected/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 города и страны и электронного адреса каждого автора;
 - сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/collected/2012.22.02_rus/authors.asp и
http://www.ipiran.ru/journal/collected/2012.22.02_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 (ссылка 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”.

Примеры ссылок на различные виды публикаций в списке “References”:

Описание статьи из журнала:

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 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 resursoberegayushchie tekhnologii nedropol’zovaniya i povyshe-niya 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 Pubs. 352 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 slovarya 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 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]. Moscow: Standardinform Pubs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Spособ 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 «Системы и средства информатики» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Системы и средства информатики»:

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 (Сейфуль-Мулюков Рустем Бадриевич)
<http://www.ipiran.ru/journal/collected>

Requirements for manuscripts submitted to Journal “Systems and Means of Informatics”

Journal “Systems and Means of Informatics” publishes theoretical, review, and discussion articles on the research and development in the field of information technology.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

Topics covered include the following areas:

- information and communication systems and tools of their design;
 - architecture and software of computational complexes and networks; and
 - methods and tools of information protection.
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 “Systems and Means of Informatics” 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).

The Editorial Board has the right to request from the authors an official expert conclusion that the submitted article has no classified data prohibited for publication.

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 authors receive their article for correction after reviewing, it does not mean that the article is approved to be published. 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 L^AT_EX (.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/publication/collected/template.doc>

8. 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/collected/2012_22_02_rus/authors.asp and
http://www.ipiran.ru/journal/collected/2012_22_02_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 in 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).

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 resursoberegayushchie tekhnologii nedropol'zovaniya i povyshe-niya neftegazootdachi"* [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 turbogenera-torov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publ. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovaryaya informat-sionnogo 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 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]. M.: Standardinform Publ. 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. "System and Means of Informatics" 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 (to Prof. Rustem Seyful-Mulyukov)

http://www.ipiran.ru/english/journal_systems.asp

SYSTEMS AND MEANS OF INFORMATICS (SISTEMY I SREDSTVA INFORMATIKI)

SCIENTIFIC JOURNAL

Volume 29 No.2 Year 2019

Editor-in-Chief and Chair of Editorial Council
Academician I. A. Sokolov

I N T H I S I S S U E :

THE ELECTRONIC LEDGER ON THE BASIS OF SITUATIONAL CENTERS FOR DIGITAL ECONOMY <i>A. A. Grusho, A. A. Zatsarinny, and E. E. Timonina</i>	4
MODELS OF DETECTION RELATIONSHIP BETWEEN TIME SERIES IN FORECASTING PROBLEMS <i>K. R. Usmanova and V. V. Strijov</i>	12
CONVERGENCE OF THE DISTRIBUTION OF THE THRESHOLD PROCESSING RISK ESTIMATE TO A MIXTURE OF NORMAL LAWS AT A RANDOM SAMPLE SIZE <i>O. V. Shestakov</i>	31
<i>A PRIORI</i> GENERALIZED FRECHET DISTRIBUTION IN BAYESIAN BALANCE MODELS <i>A. A. Kudryavtsev, S. I. Palionnaia, and V. S. Shorgin</i>	39
INTELLECTUALIZATION OF OPERATIONAL DISPATCH CONTROL OF COMPLEX REGIONAL DYNAMIC POWER SYSTEM <i>A. V. Kolesnikov, S. V. Listopad, and F. G. Maitakov</i>	46
INFORMATION TRANSFORMATION IN THE PROCESSING OF GEODATA <i>D. A. Nikishin</i>	60
THE METHOD OF DEPTH MAP CALCULATION BASED ON SOFT OPERATORS <i>M. V. Bobyr, A. E. Arkhipov, and N. A. Milostnaya</i>	71
A PRACTICAL DEFINITION OF "SOFTWARE-DEFINED" <i>V. B. Egorov</i>	85
RECOVERY OF SIGNAL IN SYSTEMS OF REAL TIME ON UNIFORM SELECTIONS WITH REDUCTION OF NYQUIST INTERVAL <i>B. G. Mayorov</i>	95