

Информатика и её применения

Том 14 Выпуск 4 Год 2020

СОДЕРЖАНИЕ

О вероятностных оценках достоверности эмпирических выводов А. А. Грушо, М. И. Забежайло, Д. В. Смирнов, Е. Е. Тимонина	3
Влияние параметров изоляции на разделение ресурсов при нарезке сети Ф. А. Москалева, Ю. В. Гайдамака, В. С. Шоргин	9
Система массового обслуживания с орбитами для анализа совместного обслуживания трафика с малыми задержками URLLC и широкополосного доступа eMBB в беспроводных сетях пятого поколения П. А. Харин, Е. Д. Макеева, И. А. Кочеткова, Д. В. Ефросинин, С. Я. Шоргин	17
Стационарные характеристики системы Geo/G/1/∞ с неординарным входящим потоком, управляющим размером очереди С. И. Матюшенко, Р. В. Разумчик	25
О распределении отношения суммы элементов выборки, превосходящих некоторый порог, к сумме всех элементов выборки. II В. Ю. Королев	33
О марковских и рациональных потоках случайных событий. II В. А. Наумов, К. Е. Самуйлов	37
Методы детерминированных и рандомизированных энтропийных проекций для редукции размерности матрицы данных Ю. С. Попков, А. Ю. Попков, Ю. А. Дубнов	47
Оптимизация структуры сетей глубокого обучения М. С. Потанин, К. О. Вайсер, В. А. Жолобов, В. В. Стрижов	55
Повышение сбоеустойчивости самосинхронных схем И. А. Соколов, Ю. А. Степченков, Ю. Г. Дьяченко, Ю. В. Рождественский	63
Об одном подходе к формированию в условиях высокой неопределенности маркеров конфиденциальности в системах интенсивного использования данных В. И. Будзко, В. В. Ядринцев, И. В. Соченков, В. И. Королёв, В. Г. Беленков	69
Метод визуального представления конфликтов в гибридных интеллектуальных многоагентных системах С. Б. Румовская, И. А. Кириков	77
Оценка стоимости опционов на основе моделей ARIMA-GARCH с ошибками, распределенными по закону S_U Джонсона А. Р. Данилишин, Д. Ю. Голембиовский	83

Информатика и её применения

Том 14 Выпуск 4 Год 2020

СОДЕРЖАНИЕ

Применение многомасштабного подхода и методов анализа данных для моделирования теплопроводности в слоистых структурах К. К. Абгарян, И. С. Колбин	91
О цифровой грамотности и средах ее формирования В. Б. Бетелин, А. Г. Кушниренко, А. Л. Семенов, С. Ф. Сопрунов	100
Эволюция классификаций в надкорпусных базах данных А. А. Гончаров, И. М. Зацман, М. Г. Кружков	108
Об авторах	117
Авторский указатель за 2020 г.	120
2020 Author Index	125
Правила подготовки рукописей	130
Requirements for manuscripts	133

О ВЕРОЯТНОСТНЫХ ОЦЕНКАХ ДОСТОВЕРНОСТИ ЭМПИРИЧЕСКИХ ВЫВОДОВ*

А. А. Грушо¹, М. И. Забейло², Д. В. Смирнов³, Е. Е. Тимонина⁴

Аннотация: Работа посвящена некоторым особенностям анализа данных в задачах поиска инсайдеров. Обсуждаются возможности использования различных подходов к описанию диагностики действий инсайдеров при анализе больших эмпирических данных. В задачах этого типа необходимо установить (спрогнозировать, диагностировать и др.) наличие или отсутствие целевых свойств у каких-либо пользователей из заданного множества. Оценка правильности правдоподобных рассуждений проверяется на основе оценок вероятностей случайного появления найденных закономерностей в простейших вероятностных моделях. Рассмотренные примеры показывают, при каких соотношениях параметров возможно эффективное выявление корреляционных связей между событиями, с помощью которых можно выявлять инсайдеров. Указаны два способа управления соотношениями между параметрами, позволяющие получать содержательную информацию. Первый способ основан на разделении периода наблюдений на промежутки, в течение которых искомая корреляция может проявиться. Второй способ связан со способами сокращения множества пользователей, которые потенциально могут стать инсайдерами, т. е. речь идет о формировании кластеров, в которых вероятностные оценки становятся работоспособными. Исковые соотношения между параметрами для поиска корреляций можно определять с помощью предельных теорем в схеме серий.

Ключевые слова: враждебный инсайдер; каузальный анализ; вероятностные оценки случайного появления свойств

DOI: 10.14357/19922264200401

1 Введение

Угроза информационной безопасности со стороны враждебного инсайдера (далее — инсайдера) считается второй в мире по значимости после хакеров. Обнаружить враждебные действия инсайдера намного сложнее, чем вредоносное воздействие хакера, потому что инсайдеры знакомы с системой безопасности организации и имеют доступ к ее части, при этом объем вредоносных действий, которые можно зафиксировать, мал по сравнению с объемом легальных действий, совершаемых пользователями в процессе работы с информационной системой (ИС) организации. Разработанные алгоритмы поиска инсайдеров часто не способны выявить новые методы атак инсайдера, а лучшие результаты показывают на шаблонных входных данных.

Необходимо отметить, что, как правило, поиск инсайдеров связан с поиском аномалий в различных информационных пространствах, где могут

появляться признаки инсайдеров. Например, Министерство обороны США провело исследования под названием Anomaly Detection at Multiple Scales (ADAMS) [1], основная цель которого — разработать и внедрить технологию классификации и обнаружения аномалий в больших объемах данных с целью поиска следов инсайдеров.

В настоящее время множество статей посвящено проблеме поиска инсайдеров. Отметим некоторые работы, имеющие связь с исследованиями данной статьи.

Пример корреляционного подхода в поиске инсайдера можно найти в работе [2], в которой рассматриваются описания связей между событиями и инсайдерами. В работе [3] прямо ставится проблема верификации вывода о результатах поиска инсайдера в нескольких информационных пространствах, но методы верификации не рассматриваются. Вероятностные методы рассмотрены в [4], где показано, как использовать скрытые марковские модели

* Работа выполнена при частичной финансовой поддержке РФФИ (проект 18-29-03081).

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

² Вычислительный центр Федерального исследовательского центра «Информатика и управление» Российской академии наук, m.zabeylo@yandex.ru

³ ПАО Сбербанк России, dvlsmirnov@sberbank.ru

⁴ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, eltimon@yandex.ru

для оценок нормального поведения пользователей ИС. Обзор [5] рассматривает различные подходы к анализу угроз, порождаемых инсайдерами.

Приведем несколько российских работ, посвященных проблеме поиска инсайдеров. В статьях [6–8] построены методы работы с несколькими информационными пространствами в задачах поиска инсайдера и поиска аномалий, связанных с инсайдером.

В данной работе рассматривается задача верификации найденных признаков инсайдеров вероятностными методами.

Пусть задано множество пользователей ИС, содержащей профессиональную информацию и ценную информацию (в том числе дорогостоящую). Допустим, что среди пользователей есть один инсайдер. Относительно его выявления возможны различные частные постановки задачи.

- А. Если произошло компьютерное преступление, в котором виновен инсайдер, то надо провести расследование и определить его личность среди множества пользователей, собрать доказательную базу.
- В. Если есть множество конкретных должностных лиц, то надо определить, есть ли в этом множестве инсайдер.

Признаком возможных действий инсайдера могут служить аномалии в технологических информационных процессах. Тогда надо выявлять аномалии различных типов и на основе найденных аномалий определить, существует ли инсайдер.

Перечислим некоторые подходы к выявлению искомой информации.

1. Поиск связи какого-нибудь пользователя u и свойства, которое наблюдается в некоторых событиях, связанных с возможным появлением инсайдера. Например, появление у клиентов банка больших сумм на счету сопровождается дорогими покупками пользователя — служащего банка. Гипотеза состоит в том, что этот пользователь продает информацию о появлениях указанного свойства и, получив деньги, тратит на заранее запланированные покупки или участие в азартных играх. Связь пользователя с появлением данного свойства связана с задачами А и В.
2. В ходе расследования компьютерного преступления в банке выяснилось, что преступления подобного типа уже совершались и проходили по определенной схеме. Если пользователь участвует в схеме, то должны появляться другие элементы схемы при проявлении действия из этой схемы. Тогда подход состоит в выявлении действия схемы.

3. На фоне стандартной деятельности пользователей появились вкрапления нестандартных действий пользователя, которые следует рассматривать как аномалию, если эти действия удается заметить.

4. Одним из важных свойств, используемых для выявления инсайдеров, являются противоречия, например размер зарплаты и размеры расходов, круг профессиональных задач и широта интересов и др.

5. Подходом, близким по эффективности, но отличающимся от метода поиска противоречий, является поиск некоторых характеристик пользователя, порождающих желание стать инсайдером. К таким характеристикам относятся, например, знакомства с криминальными элементами, жадность, игромания и др.

Выше перечислены некоторые свойства, но этот набор свойств далеко не полный. Во всех случаях возможно построение вероятностной модели и проверки согласия с ней. Но надо заметить, что выявленные свойства, которые назовем эмпирическими закономерности, могут проистекать от случайного стечения обстоятельств, поэтому надо иметь методы, позволяющие фильтровать возможности случайных сочетаний свойств, которые можно принять за искомые закономерности. Отметим, что чем сложнее эти методы, тем меньше они применимы на практике. С другой стороны, упрощение моделей и методов может привести к потере реальных закономерностей. Выскажем гипотезу, что пропущенные или отброшенные кандидаты на закономерности в случае реального инсайдера должны повторяться и при запоминании отброшенных кандидатов могут получить уверенное подтверждение при вторичном анализе.

Работа посвящена построению простых методов фильтрации кандидатов на закономерности в подходах 1-го типа или фильтрации аномалий.

2 Модель корреляции важных событий и действий пользователей

Пусть $U = \{u_1, u_2, \dots, u_m\}$ — пользователи ИС (сотрудники организации), $V = \{v_1, v_2, \dots, v_n\}$ — клиенты, на которых работают пользователи, и эти множества не пересекаются.

Простейшие модели и логику работы с ними построим для подходов 1-го типа.

Пусть C — события, связанные с появлением и доступностью ценной информации. Предположим, что события C возникают в соответствии с пуассоновским процессом с параметром λ . Траты (большие) произвольного пользователя u из множества U также происходят в соответствии с пуассоновским процессом с параметром μ . Все процессы независимы. Введем параметр T — время возможного появления зависимости пользовательских трат и появления ценной информации. Вероятность того, что данный пользователь в промежуток времени T не проводил больших трат, равна $e^{-\mu T}$.

Рассмотрим возможность появления трат в фиксированный промежуток времени T , связанный с фиксированным событием C . Вероятность того, что все пользователи в промежуток T не проводили больших трат, равна $e^{-\mu T m}$, т.е. можно окружить каждое появление события C отрезком длины T , начинающимся с момента появления события C . По предположению, события C появляются в соответствии с пуассоновским процессом, и можно выбрать промежуток времени T таким образом, чтобы отрезки длины T не пересекались с вероятностью, близкой к 1 (для этого надо рассматривать события C в процессе с $\lambda \ll \mu$). Вероятность того, что ни в одном промежутке времени при появлении k событий C не было больших трат, не превосходит $k e^{-\mu T m}$.

Если $\mu T m$ — большая величина, то вероятность непопадания больших трат на какой-нибудь отрезок длины T вокруг событий C мала. Тогда появление таких событий указывает на множество подозрительных пользователей. Можно выписать распределение числа таких пользователей, но вопрос, когда их будет мало или не будет вовсе, с большой вероятностью в данном случае более интересен, т.е. поставим вопрос о выборе параметров целесообразного поиска. При каких соотношениях параметров асимптотически в результате корреляционной проверки проявляется только инсайдер? Для решения этой задачи используем параметр k — число появлений события C .

Предположим, что все пользователи честные. Пусть φ — случайная величина, равная числу появлений события C за время наблюдения t за поведением и тратами пользователей из множества $U = \{u_1, u_2, \dots, u_m\}$. События C появляются в соответствии с законом простого пуассоновского процесса:

$$P(\varphi = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Предположим, что выброшены все пересечения промежутков T после событий C . Распределение

Пуассона при больших k можно аппроксимировать нормальным распределением. Отсюда получим

$$P(\varphi = \lambda t(1 + o(1))) = 1 + o(1).$$

Тогда время t наблюдения для получения k событий C приближенно равно k/λ .

Для определения необходимых значений k и определения невозможного события для всех честных пользователей воспользуемся следующими оценками: $p = 1 - e^{-\mu T}$ — вероятность появления большой траты в данный промежуток T для фиксированного пользователя u . По предположению промежутки T не пересекаются. Тогда вероятность появления большой траты в каждый промежуток T для данного пользователя u равна p^k . Отсюда среднее число таких пользователей равно mp^k . При $T = const$ и $k = (\ln m + \ln \ln m)e^{\mu T}$ получим оценку среднего числа таких пользователей $1/(\ln m)$.

В этих условиях проявление инсайдера с вероятностью, близкой к 1, возможно при появлении пользователя, который делает большие траты после каждого появления события C не менее k раз. Отметим, что время наблюдения оценивается величиной

$$t = \frac{k}{\lambda} (1 + o(1)).$$

Такие параметры поиска значительно лучше, чем при ожидании полного отсутствия трат.

3 Модель машинного обучения

В предыдущем разделе показано, что предложенный метод работает, но требует выполнения двух условий:

- (1) наблюдения за всеми тратами;
- (2) появления трат после наступления событий C .

Наблюдения за всеми большими тратами возможно, например, когда в игровом клубе находится инсайдер, который докладывает в службу безопасности обо всех сотрудниках организации, играющих на крупные суммы. В таком случае в предыдущем разделе найдены условия однозначного выявления инсайдера.

Рассмотрим схожую задачу поиска зависимости больших трат и появления ценной информации в базе данных организации в несколько других предположениях. В рассматриваемом случае возможность выявления инсайдера основана на сокращении числа подозреваемых пользователей (2-й подход).

Пусть за время t в множестве U выделено подмножество U_1 пользователей с очень большими

единовременным тратами (покупка машины, квартиры, дачи и др.). Под эти покупки, возможно, взяты кредиты или займы у родственников или знакомых. Ясно, что объем множества U_1 значительно меньше, чем объем множества U . Воспользуемся этим обстоятельством. Каждый участник множества U_1 имеет основания использовать свое служебное положение для дополнительного заработка. Соберем информацию о доступах пользователей из U_1 за время t к ценной информации. Пусть $C(1, \varphi), C(2, \varphi), \dots, C(k(\varphi), \varphi)$ — случаи получения ценной информации пользователем φ из U_1 за время t (за счет расширения множества запросов в базе данных и других способов).

Сравним полученные данные с доступами к ценной информации у каждого из пользователей в множестве $U \setminus U_1$ (машинное обучение). Пусть $v(i)$ — число пользователей, получившие i раз за время t данные, которые можно считать ценной информацией. Рассмотрим относительные частоты чисел $v(i)$ к мощности множества $U \setminus U_1$. Эти числа можно рассматривать как оценки вероятностей P того, что данный (честный) пользователь i раз получит ценную информацию за время t . Тогда при малых значениях $P(k(\varphi))$ для φ из U_1 и небольших множествах U_1 получаем оценку вероятности того, что пользователь φ злоупотреблял своим служебным положением и может быть инсайдером, который продает ценную информацию.

4 Заключение

Рассмотренные в работе методы выявления инсайдеров служат простейшими примерами задач поиска и анализа причинности (идентификации каузальных оснований) [9]. В самом деле, предполагается, что наступление события C служит причиной появления больших трат пользователя, торговца ценной информацией, т. е. предполагается, что продажа информации о событиях C влечет возможность осуществления больших трат. Поскольку сделанное предположение может оказаться ложным, то необходимо найти критерии фильтрации ложных каузальных оснований. Наоборот, должны быть подтверждения правильности предполагаемых каузальных оснований. Для решения этих задач подходят вероятностно-статистические методы [10]. Однако применения этих методов явно отличаются от традиционных методов математической статистики [11]. Основная идея применения вероятностных методов — это построение событий, вероятности которых либо равны 0, либо при выбранных соотношениях параметров асимптотически стремятся к 0. При равенстве 0 мы имеем

дело с запретами вероятностных мер в конечных вероятностных пространствах [12], при стремлении вероятностей к 0 имеем дело с состоятельными процедурами статистических решений [13] (чаще всего в схеме серий, что отличает эти подходы от классической математической статистики). Но такие подходы соответствуют задаче фильтрации кандидатов на каузальные основания. Возникает вопрос об истинности предположений о рассматриваемых вероятностных мерах и точности решений. Здесь также видны отличия традиционных задач математической статистики от фильтрации каузальных оснований [9, 14]. Эти отличия состоят в избыточности поиска каузальных оснований, которая заключается в том, что, как правило, можно рассматривать несколько информационных пространств [6] и использовать для анализа различные каузальные основания различных исходных характеристик.

Литература

1. Anomaly Detection at Multiple Scales (ADAMS). 2011. <https://info.publicintelligence.net/DARPAADAMS.pdf>.
2. Memory A., Goldberg H. G., Senator T. E. Context-aware insider threat detection // Activity Context-Aware System Architectures: Papers from the AAAI 2013 Workshop, 2013. P. 44–47. <https://pdfs.semanticscholar.org/04aa/e6d97900ba62e90b07ac682fb7bd8c2e1029.pdf>.
3. Ruttenberg B. E., Blumstein D., Druce J., et al. Probabilistic modeling of insider threat detection systems // Graphical models for security / Eds. P. Liu, S. Mauw, K. Stølen. — Lecture notes in computer science ser. — Springer, 2018. Vol. 10744. P. 91–98. doi: 1007/978-3-319-74860-3_6.
4. Rashid T., Agrafiotis I., Nurse J. R. C. A new take on detecting insider threats: Exploring the use of hidden Markov models // 8th ACM CCS Workshop (International) on Managing Insider Security Threats Proceedings. — ACM, 2016. P. 47–56.
5. Gheyas I. A., Abdallah A. E. Detection and prediction of insider threats to cyber security: A systematic literature review and meta-analysis // Big Data Anal., 2016. Vol. 1. No. 6. P. 1–29. doi: 10.1186/s41044-016-0006-0.
6. Грушо А. А., Забейайло М. И., Смирнов Д. В., Тимонина Е. Е. Модель множества информационных пространств в задаче поиска инсайдера // Информатика и её применения, 2017. Т. 11. Вып. 4. С. 65–69.
7. Грушо А. А., Грушо Н. А., Забейайло М. И., Смирнов Д. В., Тимонина Е. Е. Параметризация в прикладных задачах поиска эмпирических причин // Информатика и её применения, 2018. Т. 12. Вып. 3. С. 62–66.
8. Grusho A., Grusho N., Timonina E. Method of several information spaces for identification of anomalies // Intelligent distributed computing XIII / Eds. I. Kottenko, C. Badica, V. Desnitsky, D. ElBaz, and M. Ivanovic. — Studies in computational intelligence ser. — Springer, 2020. Vol. 868. P. 515–520.

9. Грушо А. А., Забейжайло М. И., Тимонина Е. Е. О каузальной репрезентативности обучающих выборок прецедентов в задачах диагностического типа // Информатика и её применения, 2020. Т. 14. Вып. 1. С. 80–86.
10. Грушо А. А., Грушо Н. А., Тимонина Е. Е. Методы выявления «слабых» признаков нарушений информационной безопасности // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 3–8.
11. Ширяев А. Н. Вероятность: В 2 кн. — 3-е изд. — М.: МЦНМО, 2004. 521 с.
12. Грушо А. А., Тимонина Е. Е. Запреты в дискретных вероятностно-статистических задачах // Дискретная математика, 2011. Т. 23. № 2. С. 53–58.
13. Grusho A., Kniazev A., Timonina E. Detection of illegal information flow // Computer network security / Eds. V. Gorodetsky, I. Kotenko, V. Skormin. — Lecture notes in computer science ser. — Springer, 2005. Vol. 3685. P. 235–244.
14. Забейжайло М. И., Трунин Ю. Ю. К проблеме доказательности медицинского диагноза: интеллектуальный анализ данных о пациентах в выборках ограниченного размера // Научно-техническая информация. Сер. 2: Информационные процессы и системы, 2019. № 12. С. 12–18.

Поступила в редакцию 12.10.20

ON PROBABILISTIC ESTIMATES OF THE VALIDITY OF EMPIRICAL CONCLUSIONS

A. A. Grusho¹, M. I. Zabezhailo², D. V. Smirnov³, and E. E. Timonina¹

¹Institute of Informatics Problems, Federal Research Center “Computer Sciences and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation

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

³Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation

Abstract: The work focuses on some features of data analysis in insider search problems. The possibilities of using different approaches to describe the diagnosis of insider actions in the analysis of large empirical data are discussed. In tasks of this type, it is necessary to establish (predict, diagnose, etc.) the presence or the absence of target properties in any users from a given set. The assessment of the correctness of plausible reasoning is checked on the basis of estimates of the probabilities of the random appearance of the found laws in the simplest probabilistic models. The examples discussed show at what ratios of parameters it is possible to effectively identify correlations between events with which insiders can be identified. Two methods of controlling relations between parameters are indicated, allowing to obtain content information. The first method is based on dividing the observation period at the intervals during which the desired correlation may appear. The second method relates to the ways to reduce the set of users that could potentially become insiders, i. e., the authors are talking about the formation of clusters in which probabilistic estimates become operational. The desired relationships between the parameters for finding correlations can be determined using limit theorems in the series scheme.

Keywords: hostile insider; causal analysis; probabilistic estimates of random appearance of properties

DOI: 10.14357/19922264200401

Acknowledgments

The paper was partially supported by RFBR, project No. 18-29-03081.

References

1. Anomaly Detection at Multiple Scales. 2011. Available at: <https://info.publicintellgence.net/DARPAADAMS.pdf> (accessed October 11, 2020).
2. Memory, A., H.G. Goldberg, and T.E. Senator. 2013. Context-aware insider threat detection. *Activity Context-Aware System Architectures: Papers from the AAAI 2013 Workshop*. 44–47. Available at: <https://pdfs.semanticscholar.org/04aa/e6d97900ba62e90b07ac682fb7bd8c2e1029.pdf> (accessed August 13, 2020).
3. Ruttenberg, B. E., D. Blumstein, J. Druce, *et al.* 2018. Probabilistic modeling of insider threat detection systems. *Graphical models for security*. Eds. P. Liu, S. Mauw, and K. Stølen. Lecture notes in computer science ser. Springer. 10744:91–98. doi: 1007/978-3-319-74860-3_6.
4. Rashid, T., I. Agrafiotis, and J. R. C. Nurse. 2016. A new take on detecting insider threats: Exploring the use of

- hidden Markov models. *8th ACM CCS Workshop (International) on Managing Insider Security Threats Proceedings*. ACM. 47–56.
5. Gheyas, I., and A. Abdallah. 2016. Detection and prediction of insider threats to cyber security: A systematic literature review and meta-analysis. *Big Data Anal.* 1(6):1–29. doi: 10.1186/s41044-016-0006-0.
 6. Grusho, A. A., M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2017. Model' mnozhestva informatsionnykh prostranstv v zadache poiska insaydera [The model of the set of information spaces in the problem of insider detection]. *Informatika i ee Primeneniya — Inform. Appl.* 11(4):65–69.
 7. Grusho, A. A., N. A. Grusho, M. I. Zabezhailo, D. V. Smirnov, and E. E. Timonina. 2018. Parametrizatsiya v prikladnykh zadachakh poiska empiricheskikh prichin [Parametrization in applied problems of search of the empirical reasons]. *Informatika i ee Primeneniya — Inform. Appl.* 12(3):62–66.
 8. Grusho, A., N. Grusho, and E. Timonina. 2020. Method of several information spaces for identification of anomalies. *Intelligent distributed computing XIII*. Eds. I. Kotenko, C. Badica, V. Desnitsky, D. ElBaz, and M. Ivanovic. Studies in computational intelligence ser. Springer. 868:515–520.
 9. Grusho, A. A., M. I. Zabezhailo, and E. E. Timonina. 2020. O kausal'noy reprezentativnosti obuchayushchikh vyborok pretsedentov v zadachakh diagnosticheskogo tipa [On causal representativeness of training samples of precedents in diagnostic type tasks]. *Informatika i ee Primeneniya — Inform. Appl.* 14(1):80–86.
 10. Grusho, A. A., N. A. Grusho, and E. E. Timonina. 2019. Metody vyyavleniya “slabykh” priznakov narusheniy informatsionnoy bezopasnosti [Methods of identification of “weak” signs of violations of information security]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):3–8.
 11. Shiryaev, A. N. 2004. *Veroyatnost'* [Probability]. Moscow: MTsNMO. 521 p.
 12. Grusho, A. A., and E. E. Timonina. 2011. Prohibitions in discrete probabilistic statistical problems. *Discrete Mathematics Applications* 21(3):275–281.
 13. Grusho, A., A. Kniazev, and E. Timonina. 2005. Detection of illegal information flow. *Computer network security*. Eds. V. Gorodetsky, I. Kotenko, and V. Skormin. Lecture notes in computer science ser. Springer. 3685:235–244.
 14. Zabezhailo, M. I., and Y. Y. Trunin. 2019. On the problem of medical diagnostic evidence: Intelligent analysis of empirical data on patients in samples of limited size. *Automatic Documentation Mathematical Linguistics* 53:322–328. doi: 10.3103/S0005105519060086.

Received October 12, 2020

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

Zabezhailo Michael I. (b. 1956) — Doctor of Science in physics and mathematics, principal scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; m.zabezhailo@yandex.ru

Smirnov Dmitry V. (b. 1984) — business partner for IT Security Department of Sberbank of Russia, 19 Vavilov Str., Moscow 117999, Russian Federation; dvlsmirnov@sberbank.ru

Timonina Elena E. (b. 1952) — Doctor of Science in technology, professor, leading scientist, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences; 44-2 Vavilov Str., Moscow 119133, Russian Federation; eltimon@yandex.ru

ВЛИЯНИЕ ПАРАМЕТРОВ ИЗОЛЯЦИИ НА РАЗДЕЛЕНИЕ РЕСУРСОВ ПРИ НАРЕЗКЕ СЕТИ*

Ф. А. Москалева¹, Ю. В. Гайдамака², В. С. Шоргин³

Аннотация: Технология нарезки радиоресурсов сети определяется как один из основных компонентов пятого поколения мобильных коммуникаций, способных решить проблему колоссального роста объема трафика данных в сотовых сетях. Ключевая особенность нарезки радиоресурсов сети, или сетевого слайсинга, позволяющая ограничить влияние одного слайса на другой, заключается в обеспечении изолированных гарантий производительности для предоставления высокого качества обслуживания (QoS, Quality of Service). В статье с помощью аппарата теории массового обслуживания построена модель разделения ресурсов при нарезке сети, позволяющая исследовать разделение ресурсов в соответствии с различными стратегиями справедливости. Задача разделения ресурсов сформулирована в форме задачи оптимизации относительно зависящей от параметра изоляции весовой функции ресурса системы, занятого заявками каждого слайса. Проведенный численный анализ показал существенное влияние параметра изоляции на изменение характеристик производительности системы.

Ключевые слова: нарезка сети; справедливое разделение ресурсов; изоляция слайсов; параметр изоляции

DOI: 10.14357/19922264200402

1 Введение

Нарезка радиоресурсов сети (*англ.* network slicing) — ключевая технология, позволяющая операторам сети предоставлять свою физическую инфраструктуру для поддержки услуг с различными требованиями [1]. Определенный набор услуг может быть связан с логически независимой сквозной сетью, т. е. слайсом. Слайс (*англ.* slice) — логическая сеть, обеспечивающая определенные функциональные возможности и сетевые характеристики [2]. Слайсы настраиваются и управляются арендаторами (*англ.* tenants), например виртуальными операторами (Virtual Network Operator, VNO), которым мобильный оператор, владеющий инфраструктурой, делегирует контроль над использованием ресурсов и качеством предоставления услуг внутри слайса.

Концепция нарезки сети подразумевает автоматизацию создания и настройки слайса, изоляцию слайсов (независимость показателей качества обслуживания в слайсе от трафика в других слайсах, а также безопасность и т. п.), эластичность нарезки (справедливое [3] и эффективное использование ресурсов, адаптация к условиям), иерархию управления (самоуправление в слайсе), возможность назначать приоритетные слайсы [4].

Гарантии качества обслуживания в логической сети обеспечиваются изоляцией слайса, так что никакие изменения в других слайсах не могут повлиять на показатели качества обслуживания (QoS). При этом выбор стратегии изоляции и разделения ресурсов на радиointерфейсе — достаточно сложная задача [5].

Из-за стохастической природы беспроводной среды и высокой изменчивости трафика во времени и пространстве идеальную изоляцию можно обеспечить лишь в случае резервирования ресурсов в соответствии с наихудшими ожидаемыми условиями, что ведет к неэффективному использованию ресурсов в большинстве случаев. Попытка учесть стохастическую природу беспроводной среды и ее особенности с использованием аппарата цепей Маркова сделана в [6].

Как показано в [7], устанавливая взаимосвязь параметров сети, можно достичь баланса между изоляцией и эффективностью, что позволяет приоритизировать и настраивать каждый слайс в соответствии с конкретными задачами, для которого он используется. При этом управление ресурсами как внутри слайса, так и между разными слайсами должно гарантировать не только изоляцию слайса, но и справедливость разделения ресурсов между пользователями [8].

* Публикация выполнена при поддержке Программы стратегического академического лидерства РУДН и при финансовой поддержке РФФИ (проекты 19-07-00933 и 20-07-01064).

¹ Российский университет дружбы народов, moskaleva-fa@rudn.ru

² Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, gaidamaka-yuv@rudn.ru

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

2 Математическая модель системы с двумя слайсами

В [9] механизмы нарезки сети с учетом гарантий для различных типов трафика исследованы при фиксированных значениях параметров изоляции.

В статье построена математическая модель разделения радиоресурсов соты между двумя виртуальными операторами (далее — операторами), которая иллюстрирует влияние параметров изоляции на метрику производительности сети. Положим, что на базовой станции сети связи пятого поколения с технологией радиодоступа New Radio активированы два слайса, принадлежащие разным операторам, и модуль нарезки делит между ними общий ресурс емкостью C единиц ресурса. Примером единицы ресурса может быть герц для полосы радиочастот, бит в секунду для скорости или ресурсный блок LTE/NR (long-term evolution / new radio). Каждый оператор осуществляет предоставление некоторой услуги связи, предполагающей непрерывную передачу пользователю потокового трафика на выделенном ресурсе не менее $b_{\min}, d_{\min} > 0$ и не более $b_{\max}, d_{\max} > 0$ единиц ресурса для первого и второго слайсов соответственно. Длительность предоставления услуги пользователю (длительность пользовательской сессии) определяется объемом выделенного для обслуживания сессии ресурса, который зависит от числа активных сессий в каждом слайсе. Считаем, что ресурсы каждого слайса делятся между его пользователями поровну. Для обеспечения справедливого разделения ресурсов введены параметры изоляции, определяющие число пользователей в каждом слайсе, которым оператор обязуется предоставить услугу с минимальным качеством. При низкой загрузке соты каждой сессии выделяется ресурс, достаточный для получения пользователем услуги на максимальной скорости. С увеличением нагрузки, создаваемой запросами пользователей обоих операторов, выделяемый каждой сессии ресурс снижается, пока не достигнет уровня минимальной скорости, требуемой для предоставления услуги. При дальнейшем росте нагрузки начинает работать концепция нарезки сети, согласно которой для приема в систему заявок слайса, не достигшего заданного параметром изоляции предела, должно быть прервано обслуживание одной или нескольких заявок второго слайса, т.е. «нарушителя» (*англ.* violator), который пользовался простаивающим ресурсом недогруженного слайса.

В следующих разделах построена модель в виде системы массового обслуживания (СМО), позволяющая исследовать зависимость показателей качества обслуживания от параметров изоляции, предложен алгоритм разделения радиоресурсов с помощью методов теории оптимизации, приведен пример численного анализа полученных результатов.

Пусть в многолинейную СМО поступают два пуассоновских потока заявок с интенсивностями λ_1 и λ_2 (рис. 1), соответствующие запросам на установление сессии от пользователей двух операторов. Длительности обслуживания заявок 1-го и 2-го потоков распределены по экспоненциальному закону с параметрами μ_1 и μ_2 соответственно. Число единиц ресурса, выделяемых заявке, принятой на обслуживание, зависит от общей загруженности системы и варьируется в диапазонах $[b_{\min}, b_{\max}]$ и $[d_{\min}, d_{\max}]$ для 1-го и 2-го потоков соответственно.

В системе предусмотрены параметры изоляции \bar{M} и \bar{N} , характеризующие максимальные значения числа заявок 1-го и 2-го потока, для которых обслуживание гарантировано.

Состояние системы описывает случайный процесс $X(t) = (M(t), N(t))$, где $M(t)$ — число заявок 1-го потока; $N(t)$ — число заявок 2-го потока в момент t , с пространством состояний

$$\mathbb{X} = \left\{ (m, n) : m = 0, \dots, \left\lfloor \frac{C}{b_{\min}} \right\rfloor, \right. \\ \left. n = 0, \dots, \left\lfloor \frac{C}{d_{\min}} \right\rfloor, mb_{\min} + nd_{\min} \leq C \right\}.$$

Введем величины $b(m, n)$ и $d(m, n)$, обозначающие число единиц ресурса, выделенных для обслуживания одной заявки 1-го и 2-го потоков соответственно в состоянии $(m, n) \in \mathbb{X}$. Заметим, что $b_{\min} \leq b(m, n) \leq b_{\max}$, $d_{\min} \leq d(m, n) \leq d_{\max}$, при этом значения $b(m, n)$ и $d(m, n)$ являются решением задачи оптимизации, которая сформулирована в разд. 3 статьи.

Интенсивности обслуживания заявок 1-го и 2-го потока определяются как $mb(m, n)\mu_1$ и $nd(m, n)\mu_2$ соответственно.

Параметры изоляции \bar{M} и \bar{N} управляют приемом в систему поступающих заявок и прерывани-

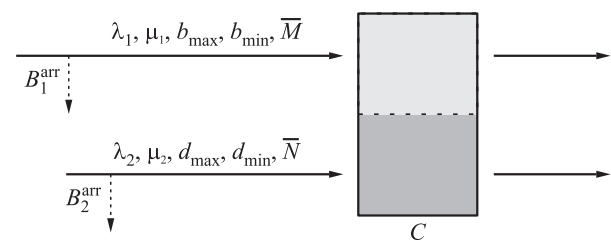


Рис. 1 Модель СМО

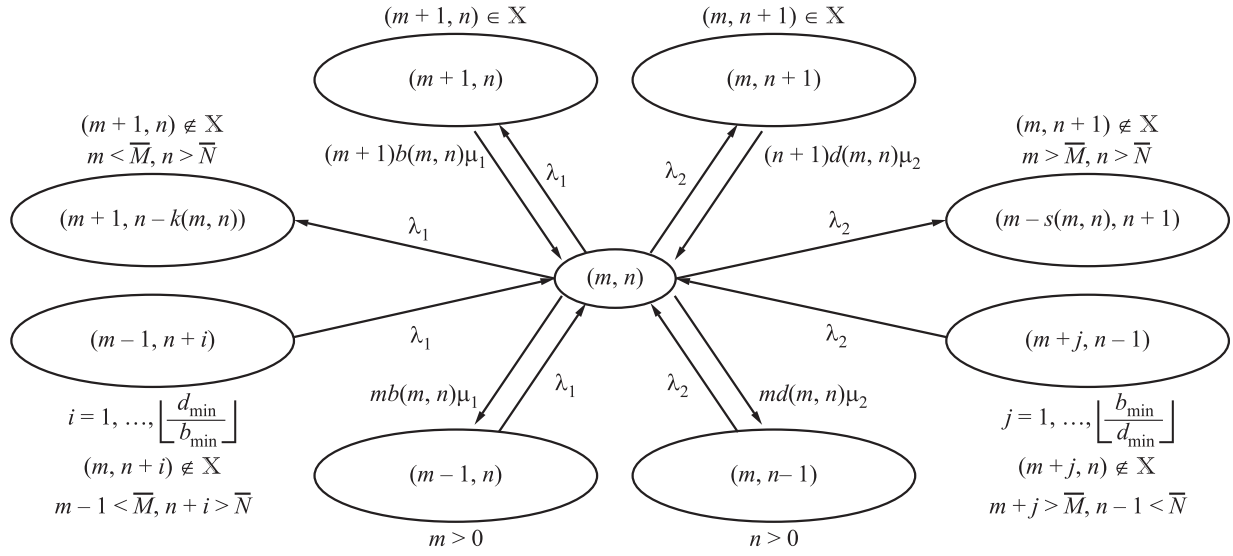


Рис. 2 Центральное состояние диаграммы интенсивностей переходов случайного процесса $\mathbb{X}(t)$

ем обслуживания ранее принятых заявок следующим образом. Если при поступлении заявки 1-го потока в состоянии $(m, n) \in \mathbb{X}$, где $m < \bar{M}$, $n > \bar{N}$, в системе недостаточно свободного ресурса для обслуживания $(m+1)$ заявок 1-го потока с выделением каждой минимального числа b_{\min} единиц ресурса, поступающая заявка 1-го потока вытесняет одну или несколько заявок 2-го потока, чтобы встать на обслуживание (рис. 2). Число $k(m, n)$ таких заявок, которые после прерывания обслуживания покинут систему, не оказывая влияния на ее дальнейшее функционирование, вычисляется следующим образом:

$$k(m, n) = \left\lceil \frac{(m+1)b_{\min} + nd_{\min} - C}{d_{\min}} \right\rceil, \\ (m, n) \in \mathbb{X}, (m+1, n) \notin \mathbb{X}, k(m, n) \leq n.$$

Аналогично для случая $m > \bar{M}$, $n > \bar{N}$ для приема заявки 2-го потока будет прервано обслуживание $s(m, n)$ заявок 1-го потока:

$$s(m, n) = \left\lceil \frac{mb_{\min} + (n+1)d_{\min} - C}{b_{\min}} \right\rceil, \\ (m, n) \in \mathbb{X}, (m, n+1) \notin \mathbb{X}, s(m, n) \leq m.$$

Для случая $m < \bar{M}$, $n < \bar{N}$ и для случая $m > \bar{M}$, $n > \bar{N}$ поступающая заявка теряется, если в системе недостаточно свободного ресурса для ее приема в систему.

Таким образом, множества потери $\mathbb{B}_1^{\text{arr}}$ заявок 1-го потока и $\mathbb{B}_2^{\text{arr}}$ заявок 2-го потока при поступлении (англ. arrival) и множества прерывания обслуживания \mathbb{B}_1^{pr} заявок 1-го потока и \mathbb{B}_2^{pr} заявок

2-го потока при вытеснении (англ. preemption) [10] имеют следующий вид:

$$\mathbb{B}_1^{\text{arr}} = \{(m, n) \in \mathbb{X} : (m+1, n) \notin \mathbb{X}, m \geq \bar{M}\}; \\ \mathbb{B}_2^{\text{arr}} = \{(m, n) \in \mathbb{X} : (m, n+1) \notin \mathbb{X}, n \geq \bar{N}\}; \\ \mathbb{B}_1^{\text{pr}} = \{(m, n) \in \mathbb{X} : (m, n+1) \notin \mathbb{X}, m > \bar{M}, n < \bar{N}\}; \\ \mathbb{B}_2^{\text{pr}} = \{(m, n) \in \mathbb{X} : (m+1, n) \notin \mathbb{X}, m < \bar{M}, n > \bar{N}\}.$$

Распределение стационарных вероятностей \mathbf{p} получаем путем решения системы линейных уравнений:

$$\mathbf{p}\mathbf{A} = \mathbf{0}; \\ \mathbf{p}\mathbf{1} = \mathbf{1},$$

где \mathbf{A} — инфинитезимальная матрица, элементы которой $a((m, n), (m', n'))$ записаны ниже:

$$a((m, n), (m', n')) = \begin{cases} \lambda_1, & m+1, n' = n, (m', n') \in \mathbb{X}; \\ \lambda_1, & m' = m+1, n' = n-k, \\ & (m+1, n) \notin \mathbb{X}, m < \bar{M}, n \geq \bar{N}; \\ \lambda_2, & m' = m, n' = n+1, (m', n') \in \mathbb{X}; \\ \lambda_2, & m' = m-s, n' = n+1, \\ & (m, n+1) \notin \mathbb{X}, m \geq \bar{M}, n < \bar{N}; \\ mb(m, n)\mu_1, & m' = m-1, n' = n, m' > 0; \\ nd(m, n)\mu_2, & m' = m, n' = n-1, n' > 0; \\ A, & m' = m, n' = n; \\ 0 & \text{иначе.} \end{cases}$$

Здесь диагональные элементы $a((m, n), m, n) = A$ имеют следующий вид:

$$A = -\lambda_1 I((m+1, n) \in \mathbb{X}, n \geq N) - \lambda_2 I((m, n+1) \in \mathbb{X}, m \geq M) - mb(m, n)\mu_1 I(m > 0) - nd(m, n)\mu_2 I(n > 0).$$

Получив распределение вероятностей, можно найти некоторые характеризующие производительность системы метрики для заявок 1-го и 2-го потока:

- вероятность потери заявки при поступлении и вероятность прерывания обслуживания заявки при вытеснении ($s = 1, 2$)

$$B_s^{\text{arr}} = \sum_{(m,n) \in \mathbb{B}_s^{\text{arr}}} p(m, n);$$

$$B_s^{\text{pr}} = \sum_{(m,n) \in \mathbb{B}_s^{\text{pr}}} p(m, n);$$

- среднее время обслуживания заявки

$$S_1 = N_1 \left(\lambda_1 (1 - B_1^{\text{arr}}) - \lambda_2 \sum_{(m,n) \in \mathbb{B}_1^{\text{arr}}} s(m, n) p(m, n) \right)^{-1};$$

$$S_2 = N_2 \left(\lambda_2 (1 - B_2^{\text{arr}}) - \lambda_1 \sum_{(m,n) \in \mathbb{B}_2^{\text{arr}}} k(m, n) p(m, n) \right)^{-1};$$

- вероятность нарушения

$$V_1 = \sum_{m > \bar{M}} p(m, n); \quad V_2 = \sum_{n > \bar{N}} p(m, n).$$

В следующем разделе сформулирована задача оптимизации для вычисления значений величин $b(m, n)$ и $d(m, n)$, обеспечивающая эффективное использование ресурса системы.

3 Решение задачи разделения ресурсов

Для состояний $(m, n) \in \mathbb{X}$, в которых $mb_{\max} + nd_{\max} > C$, необходимо определить значения величин $b(m, n)$ и $d(m, n)$, позволяющие не только полностью использовать ресурс системы, но и максимизировать некоторую функцию полезности. Для примера выберем стратегию max-min справедливости при разделении ресурсов (англ. max-min fairness), функция полезности при кото-

рой [11] относится к логарифмическому типу, является возрастающей, строго вогнутой и непрерывно дифференцируемой и определяется как $U(x) = \ln x$, $\min(b_{\min}, d_{\min}) \leq x \leq \max(b_{\max}, d_{\max})$.

Разделение ресурса между заявками в системе соответствует решению следующей задачи оптимизации:

$$\left. \begin{aligned} &w_1(m, n)mU(b(m, n)) + \\ &+ w_2(m, n)nU(d(m, n)) \rightarrow \max \\ &\text{s.t. } mb(m, n) + nd(m, n) = C; \\ &b_{\min} \leq b(m, n) \leq b_{\max}; \\ &d_{\min} \leq d(m, n) \leq d_{\max}. \end{aligned} \right\} \quad (1)$$

Весовые функции $w_1(m, n)$ и $w_2(m, n)$ будем вычислять по формулам:

$$w_1(m, n) = \begin{cases} 1, & m \leq \bar{M}; \\ \frac{1}{m - \bar{M} + 1}, & m > \bar{M}; \end{cases}$$

$$w_2(m, n) = \begin{cases} 1, & n \leq \bar{N}; \\ \frac{1}{n - \bar{N} + 1}, & n > \bar{N}. \end{cases}$$

Такой выбор весовых функций обеспечивает max-min-справедливое распределение ресурсов для пользователей до тех пор, пока их количество в соответствующих слайсах не превышает зарезервированное, и «штрафует» среззы-нарушители уменьшением их веса.

Таким образом, стационарная точка задачи оптимизации имеет координаты

$$b(m, n) = \frac{w_1 C}{w_1 m + w_2 n}; \quad d(m, n) = \frac{w_2 C}{w_1 m + w_2 n} \quad (2)$$

и расположена на пересечении прямых $C = mb + nd$ и $w_1/b(m, n) = w_2/d(m, n)$.

Решение (2) задачи оптимизации (1) обеспечивает одновременно как изоляцию слайсов с помощью параметров \bar{M} и \bar{N} , так и эластичность нарезки для эффективного использования ресурса системы [4].

4 Пример численного анализа

Проиллюстрируем зависимость метрик, характеризующих производительность системы, от параметров изоляции при нарезке сети. Предполагаем, что два оператора делят между собой 50 Мбит/с ($C = 50$) согласно решению задачи оптимизации (1). Минимальные скорости передачи данных равны 5 Мбит/с ($b_{\min} = 5$) и 1 Мбит/с ($d_{\min} = 1$),

максимальные — 8 Мбит/с ($b_{\max} = 8$) и 50 Мбит/с ($d_{\max} = 50$), эти диапазоны показаны на рис. 3 темно-серым цветом. Параметры изоляции: $\bar{M} = 5$ и $\bar{N} = 25$ заявок, интенсивности поступления: $\lambda_1 = 1/150$ и $\lambda_2 = 1/120$, интенсивности обслуживания: $\mu_1 = 1/1920$ и $\mu_2 = 1/4000$, средние размеры файлов: 1,2 ГБ и 500 МБ для услуг 1-го и 2-го операторов соответственно. Указанные значения параметров системы близки к реальным и соответствуют услуге буферизуемого потокового видео в высоком разрешении для 1-го оператора, и услуге загрузки файлов, например при обновлении программного обеспечения, для 2-го оператора.

На рис. 3 штриховыми линиями показано решение (2) задачи оптимизации (1) для max-min-справедливого разделения ресурсов: на рис. 3, а всем пользователям обоих операторов выделен одинаковый ресурс 6,2 Мбит/с; на рис. 3, б все 5 пользователей 1-го оператора получили по $b_{\max} = 8$ Мбит/с, а единственный пользователь 2-го оператора — оставшиеся 10 Мбит/с; на рис. 3, в все пользо-

ватели обоих операторов получили минимальный требуемый ресурс.

На рис. 4 показаны графики вероятностей блокировки, потери при поступлении и прерывания обслуживания заявки при вытеснении для обоих операторов. Вероятность блокировки заявки вычисляется по формуле: $B_s = B_s^{\text{arr}} + B_s^{\text{pr}}$, $s = 1, 2$. Сравнение рис. 4, а и рис. 4, б показывает, что потери при поступлении вносят основной вклад в блокировку заявки, которая соответствует отказу пользователю в получении услуги. При этом с ростом параметра изоляции первого слайса, т.е. гарантированного числа \bar{M} принимаемых заявок 1-го потока, вероятность прерывания обслуживания заявки 1-го потока B_1^{pr} падает, а вероятность потери заявки 2-го потока при поступлении B_2^{arr} растет попарно симметрично. Переломным значением оказывается $\bar{M} = 5$, при котором ресурса системы становится недостаточно, чтобы удовлетворять одновременно гарантиям 1-го и 2-го слайса, т.е. $\bar{M}b_{\min} + \bar{N}d_{\min} > C$.

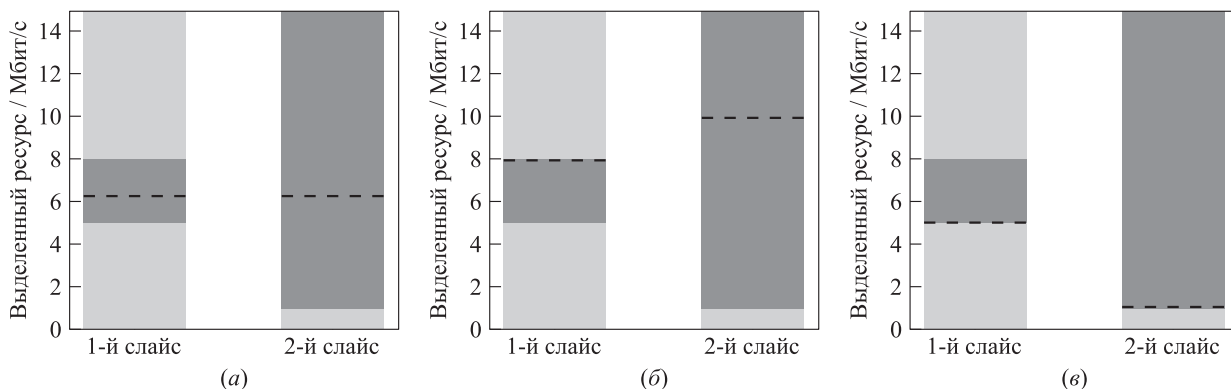


Рис. 3 Разделение ресурсов между заявками: (а) $X(t) = (4, 4)$; (б) $X(t) = (5, 1)$; (в) $X(t) = (9, 5)$

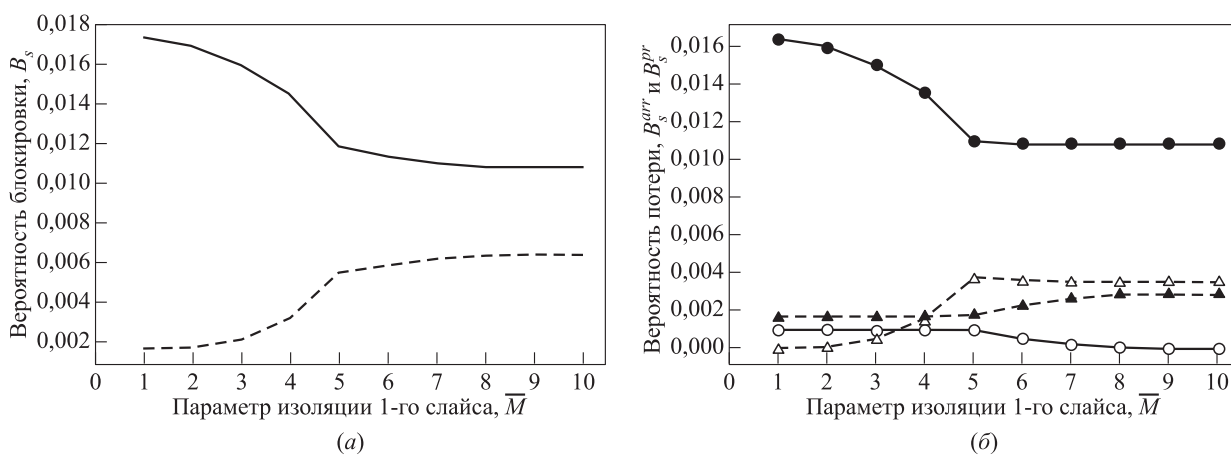


Рис. 4 Вероятности блокировки B_s (а) и потери при поступлении B_s^{arr} (затитые значки) и прерывания обслуживания при вытеснении B_s^{pr} (пустые значки) (б), $s = 1$ (сплошные кривые) и 2 (штриховые кривые)

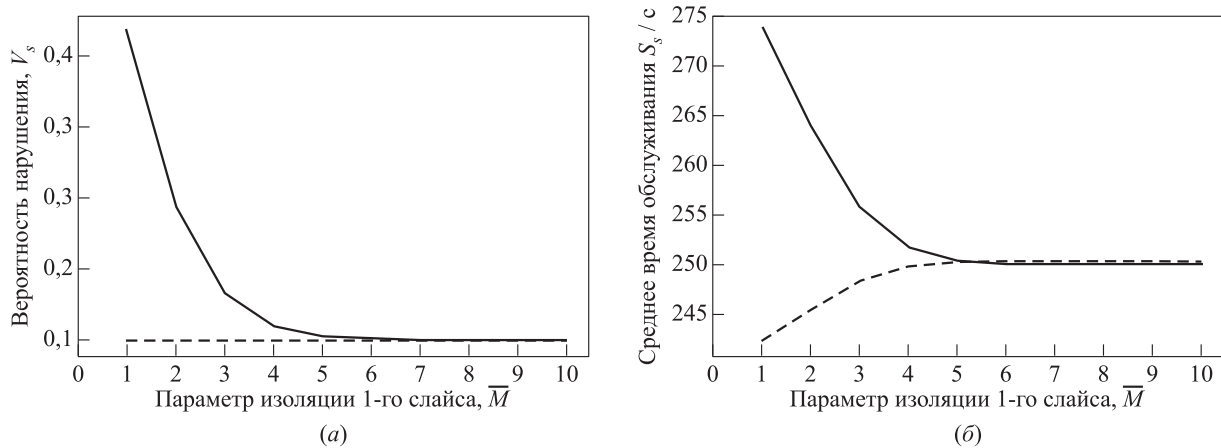


Рис. 5 Вероятность нарушения V_s (а) и среднее время обслуживания S_s (б), $s = 1$ (сплошные кривые) и 2 (штриховые кривые)

На рис. 5, а показано, что существенное влияние параметр изоляции оказывает на вероятность V_s пребывания слайса в состоянии нарушителя ($m > \bar{M}$ для первого слайса, $n > \bar{N}$ для второго). С ростом параметра изоляции \bar{M} вероятность V_1 для 1-го слайса стремится к нулю, а вероятность V_2 для 2-го слайса меняется незначительно, что свидетельствует об обеспечении изоляции. С увеличением гарантии для 1-го слайса до значения $\bar{M} = 5$ среднее время S_1 обслуживания заявок 1-го потока (рис. 5, б) падает, после чего стабилизируется, поскольку при переходе через значение $\bar{M} = 5$ параметр изоляции перестает оказывать влияние на вероятность приема заявок, а следовательно, и на среднее время обслуживания. Противоположный характер поведения кривой наблюдается для второго слайса.

Таким образом, можно сделать вывод, что изменение параметра изоляции оказывает существенное влияние на характеристики производительности системы, при этом построенная модель способна обеспечить изоляцию слайсов.

5 Заключение

Построенная в работе модель разделения ресурсов при нарезке радиоресурса сети с обеспечением изоляции слайсов и разработанный алгоритм разделения ресурсов, включающий решение задачи оптимизации, позволяют справедливо разделить ресурсы сети для их эффективного использования двумя виртуальными сетевыми операторами, арендующими радиоресурс у мобильного оператора. Заметим, что замена max-min-стратегии справедливости на стратегию пропорциональной справедливости или более общий случай α -справедливости повлияет лишь на вид целевой функции в зада-

че оптимизации. Модель может быть применена при поиске комбинации параметров изоляции, удерживающих систему в целевой области значений метрик ее производительности. Целью дальнейших исследований модели разделения ресурсов при нарезке сети может стать развитие построенной модели на случай отсутствия минимальных требований к ресурсам, что позволит исследовать эластичный трафик, а также расширение модели до произвольного числа виртуальных сетевых операторов.

Авторы благодарят Н. В. Ярину и Е. Ю. Лисовскую за полезные обсуждения в ходе работы над статьей.

Литература

1. NGMN 5G White Paper. 2015. <http://www.ngmn.de/5gwhite-paper.html>.
2. 5G: System Architecture for the 5G System (Release 15). Version 15.2.0. ETSI 3GPP TS 23.501, 2018. https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.02.00_60/ts_123501v150200p.pdf.
3. Jain R., Chiu D.-M., Hawe W. R. A quantitative measure of fairness and discrimination for resource allocation in shared computer system // arXiv.org, 1998. arXiv: cs/9809099 [cs.NI].
4. Network slicing — use case requirements. — GSMA, 2018. <https://www.gsma.com/futurenetworks/wp-content/uploads/2018/07/Network-Slicing-Use-Case-Requirements-fixed.pdf>.
5. Richart M., Baliosian J., Serrat J., Gorricho J.-L. Resource slicing in virtual wireless networks: A survey // IEEE T. Netw. Serv. Man., 2016. Vol. 13. Iss. 3. P. 462–476.
6. Vila I., Pérez-Romero J., Sallent O., Umbert A. Characterisation of radio access network slicing scenarios with 5G

- QoS provisioning // *IEEE Access*, 2020. Vol. 8. P. 51414–51430. doi: 10.1109/access.2020.2980685.
7. Marabissi D., Fantacci R. Highly flexible RAN slicing approach to manage isolation, priority, efficiency // *IEEE Access*, 2019. Vol. 7. P. 97130–97142. doi: 10.1109/access.2019.2929732.
 8. Lieto A., Malanchini I., Capone A. Enabling dynamic resource sharing for slice customization in 5G networks // *Conference and Exhibition on Global Telecommunications Proceedings. — IEEE*, 2018. P. 1–7. doi: 10.1109/GLOCOM.2018.8647249.
 9. Агеев К. А., Сопин Э. С., Яркина Н. В., Самуйлов К. Е., Шоргин С. Я. Анализ механизмов нарезки сети с учетом гарантий для различных типов трафика // *Информатика и её применения*, 2020. Т. 14. Вып. 3. С. 94–100.
 10. Yarkina N., Gaidamaka Y., Correia L. M., Samouylov K. An analytical model for 5G network resource sharing with flexible SLA-oriented slice isolation // *Mathematics*, 2020. Vol. 8. Iss. 7. Art. No. 1177.
 11. Kelly F. Charging and rate control for elastic traffic // *Eur. T. Telecommun.*, 1997. Vol. 8. P. 33–37.

Поступила в редакцию 11.10.20

IMPACT OF THE ISOLATION PARAMETERS ON RESOURCE ALLOCATION IN THE NETWORK SLICING MODEL

F. A. Moskaleva¹, Yu. V. Gaidamaka^{1,2}, and V. S. Shorgin²

¹Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, 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: Network slicing technology is defined as one of the main components of the fifth generation of mobile communications, capable of solving the problem of the colossal growth of data traffic in cellular networks. A key feature of slicing that limits the impact of one slice on another is to provide isolated performance guarantees to deliver high quality of service. In this article, a model of resource allocation during slicing is developed using the queuing theory. The main task of the work is to determine how network resources should be fairly shared between two slices in the system. The resource allocation problem is formulated as an optimization problem. For the constructed model, a numerical analysis was carried out showing the significant effect of the isolation parameters on the performance characteristics of the system.

Keywords: network slicing; fairness resource allocation; isolation of slices; isolation parameter

DOI: 10.14357/19922264200402

Acknowledgments

The paper has been supported by the RUDN University Strategic Academic Leadership Program and funded by the Russian Foundation for Basic Research according to the research projects No. 19-07-00933 and No. 20-07-01064.

References

1. NGMN 5G white paper. 2015. Available at: <http://www.ngmn.de/5gwhite-paper.html> (accessed October 22, 2020).
2. TS 23.501. 2018. Technical Specification: System architecture for the 5G system (Release 15). Version 15.2.0. Available at: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.02.00_60/ts_123501v150200p.pdf (accessed October 22, 2020).
3. Jain, R., D.-M. Chiu, and W.R. Hawe. 1998. A quantitative measure of fairness and discrimination for resource allocation in shared computer system. arXiv:cs/9809099 [cs.NI]. Available at: <https://arxiv.org/abs/cs/9809099> (accessed October 22, 2020).
4. GSMA. 2018. Network slicing — use case requirements. Available at: <https://www.gsma.com/futurenetworks/wp-content/uploads/2018/07/Network-Slicing-Use-Case-Requirements-fixed.pdf> (accessed October 22, 2020).
5. Richart, M., J. Baliosian, J. Serrat, and J.-L. Gorricho. 2016. Resource slicing in virtual wireless networks: A survey. *IEEE T. Netw. Serv. Man.* 13(3):462–476.
6. Vila, I., J. Pérez-Romero, O. Sallent, and A. Umberto. 2020. Characterisation of radio access network slicing scenarios with 5G QoS provisioning. *IEEE Access* 8:51414–51430. doi: 10.1109/access.2020.2980685.

7. Marabissi, D., and R. Fantacci. 2019. Highly flexible RAN slicing approach to manage isolation, priority, efficiency. *IEEE Access* 7:97130–97142. doi: 10.1109/access.2019.2929732.
8. Lieto, A., I. Malanchini, and A. Capone. 2018. Enabling dynamic resource sharing for slice customization in 5G networks. *Conference and Exhibition on Global Telecommunications Proceedings*. IEEE. 1–7. doi: 10.1109/GLOCOM.2018.8647249.
9. Ageev, K. A., E. S. Sopin, N. V. Yarkina, K. E. Samuylov, and S. Ya. Shorgin. 2020. Analiz mekhanizmov narezki seti s uchetom garantii dlya razlichnykh tipov trafika [Analysis of network slicing mechanisms with guarantees for various types of traffic]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):94–100.
10. Yarkina, N., Yu. Gaidamaka, L. M. Correia, and K. Samouylov. 2020. An analytical model for 5G network resource sharing with flexible SLA-oriented slice isolation. *Mathematics* 8(7):1177.
11. Kelly, F. 1997. Charging and rate control for elastic traffic. *Eur. T. Telecommun.* 8:33–37.

Received October 11, 2020

Contributors

Moskaleva Faina A. (b. 1996) — PhD student, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; moskaleva-fa@rudn.ru

Gaidamaka Yuliya V. (b. 1971) — Doctor of Science in physics and mathematics, professor, Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, 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; gaidamaka-yuv@rudn.ru

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

СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С ОРБИТАМИ ДЛЯ АНАЛИЗА СОВМЕСТНОГО ОБСЛУЖИВАНИЯ ТРАФИКА С МАЛЫМИ ЗАДЕРЖКАМИ URLLC И ШИРОКОПОЛОСНОГО ДОСТУПА eMBB В БЕСПРОВОДНЫХ СЕТЯХ ПЯТОГО ПОКОЛЕНИЯ*

П. А. Харин¹, Е. Д. Макеева², И. А. Кочеткова³, Д. В. Ефросинин⁴, С. Я. Шоргин⁵

Аннотация: Для современных беспроводных сетей пятого поколения (5G) характерны три сценария их использования — крупномасштабные системы межмашинной связи (massive machine-type communication, mMTC), сверхнадежная передача данных с малой задержкой (ultrareliable low-latency communication, URLLC) и усовершенствованная подвижная широкополосная связь (enhanced mobile broadband, eMBB). Требования к качеству предоставления услуг и их параметры в рамках данных сценариев существенно различаются: так, для URLLC характерна сверхнизкая, до 1 мс, задержка, а для eMBB — сверхвысокая скорость передачи данных. Возникает актуальная задача организации совместного предоставления таких услуг. В статье построена схема совместного обслуживания трафика URLLC и eMBB, исходя из того что данные URLLC имеют малый объем и могут занимать менее одного ресурсного блока (physical resource block, PRB). Анализ схемы проведен при помощи разработанной системы массового обслуживания (СМО) с двумя орбитами, предназначенными для временного хранения прерываемых и ожидающих начала обслуживания менее приоритетных запросов eMBB. Получен матричный алгоритм для расчета распределения вероятностей и формулы для вероятностных характеристик системы.

Ключевые слова: 5G; сверхнадежная передача данных с малой задержкой (URLLC); усовершенствованная подвижная широкополосная связь (eMBB); система массового обслуживания с орбитами; прерывание обслуживания

DOI: 10.14357/19922264200403

1 Введение

С развитием технологий беспроводной передачи данных увеличивается объем генерируемой и передаваемой информации. Согласно прогнозам Cisco Systems, рост числа мобильных устройств к 2021 г. составит до 11,6 млрд при 8 млрд в 2016 г.

Для сетей 5G выделяют три сценария их использования в соответствии со следующими характеристиками: сверхвысокая скорость передачи данных, спектральная эффективность, плотность подключения устройств, сверхнизкая задержка и сверхвысокая надежность передачи данных. Сценарии — крупномасштабные системы межмашинной связи (mMTC), сверхнадежная передача данных с малой задержкой (URLLC) и усовер-

шенствованная подвижная широкополосная связь (eMBB) [1].

Для mMTC специфична высокая плотность подключения устройств, что необходимо для приложений умного города, умного дома, промышленного интернета вещей.

Сценарий URLLC используется для передачи небольшого объема данных, но с высокой надежностью и очень низкими задержками для обеспечения мониторинга в реальном времени, например сети беспилотных автомобилей или беспилотных летательных аппаратов [2, 3].

Сценарий eMBB используется для предоставления пользователю высоких скоростей передачи, которые необходимы, например, для потоковой пе-

* Публикация подготовлена при поддержке Программы РУДН «5-100». Исследование выполнено при финансовой поддержке РФФИ (проекты 18-00-01555(18-00-01685) и 20-37-70079).

¹ Российский университет дружбы народов, pxarin@mail.ru

² Российский университет дружбы народов, len16730637@yandex.ru

³ Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, gudkova-ia@rudn.ru

⁴ Университет имени Иоганна Кеплера, Австрия; Российский университет дружбы народов, dmitry.efrosinin@jku.at

⁵ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, sshorgin@ipiran.ru

редачи видео в реальном времени с высокой четкостью, виртуальной реальности, дополненной реальности, телемедицины и дистанционного обучения.

Требования к качеству обслуживания для каждого сценария существенно различаются, поэтому возникает задача организации совместного обслуживания соответствующего трафика [4, 5]. В данной работе построена модель совместного обслуживания трафика URLLC и eMBB в условиях, когда высший приоритет имеет трафик URLLC [6, 7]. Ввиду приоритетного обслуживания URLLC, допустимо прерывание eMBB, как, например, в работах [8, 9].

Авторы продолжают исследование [9], добавив более гибкий механизм в отношении трафика eMBB — допустимо возобновление его обслуживания после прерывания.

2 Системная модель обслуживания трафика URLLC и eMBB

Рассмотрим выделенную соту беспроводной сети, в которой реализованы сценарии URLLC и eMBB. Предположим, что запросы на предоставление услуг eMBB и URLLC — передачу соответствующего трафика — поступают согласно пуассоновскому закону с интенсивностями λ_m и λ_u соответственно. Трафик URLLC имеет высокие требования к задержке и надежности передачи данных, поэтому является приоритетным по отношению к eMBB.

Организованы два буфера для устройств eMBB. Первый буфер Q_1 предназначен для устройств eMBB с задержкой передачи данных, т.е. для устройств eMBB, которые не смогли поступить на обслуживание в связи с отсутствием свободных ресурсов. Второй буфер Q_2 предназначен для прерванных eMBB-устройств новыми поступившими URLLC-устройствами в систему. Буферы Q_1 и Q_2 системы ограничены C_1 и C_2 соответственно. Пример такой соты показан на рис. 1.

Фрагмент канала такой системы показан на рис. 2. Данный канал состоит из N физических ресурсных блоков (PRB), которые, в свою очередь, делятся на b ресурсных единиц. Длительность одного PRB равна 1 слоту ($\sim 0,5$ мс). Для удовлетворения требований устройств URLLC к задержке передачи данных каждый слот делится на минислоты.

Так как трафик eMBB предназначен для передачи данных большого объема, а трафик URLLC — для передачи небольшого объема данных, то планирование передачи данных устройств eMBB

проводится относительно PRB, а устройств URLLC — относительно ресурсных единиц (resource unit, RU). Совместное планирование обслуживания проводится на разных частотах и в разное время для каждого типа трафика согласно ортогональному множественному доступу (orthogonal multiple access, OMA). Для рассматриваемой модели сделаем следующее предположение по нагрузке: каждое устройство eMBB требует 1 PRB, т.е. b ресурсных единиц, а каждое устройство URLLC — 1 ресурсную единицу. Для того чтобы принять устройство eMBB, необходимо наличие хотя бы одного свободного PRB, при этом буферы должны быть пусты. Таким образом, если данные условия выполнены, планировщик назначает один из свободных PRB, b ресурсных единиц, новому устройству eMBB. Иначе, при наличии хотя бы одного свободного места в первом буфере, планировщик записывает это устройство в буфер Q_1 , в противном случае это устройство будет заблокировано.

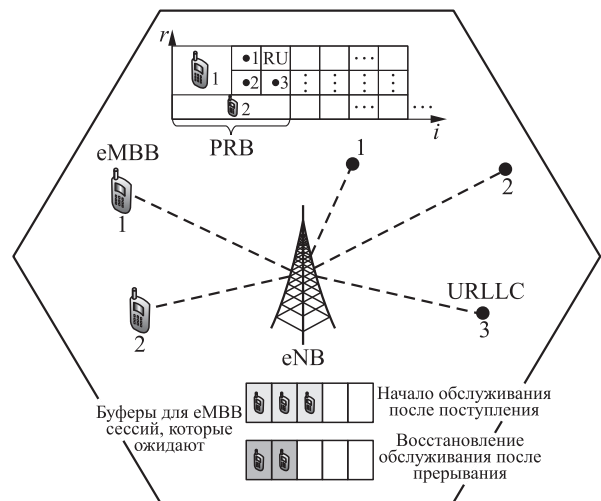


Рис. 1 Схема соты с устройствами URLLC и eMBB

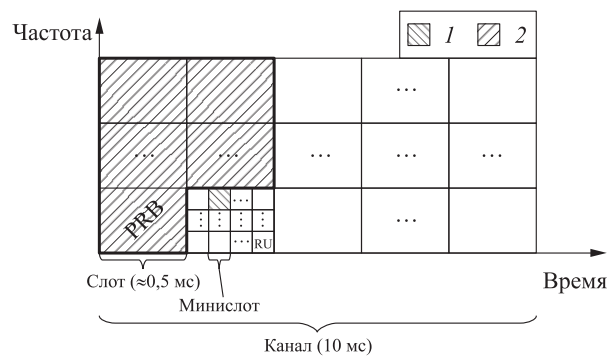


Рис. 2 Схема занятия радиоресурсов устройствами URLLC (1) и eMBB (2)

В случае же устройства URLLC для принятия к обслуживанию достаточно наличия хотя бы одной ресурсной единицы в системе, иначе возможны два исхода событий. Если в системе обслуживается хотя бы одно устройство eMBB, то планировщик приостанавливает передачу данных случайно выбранного устройства eMBB, таким образом освободив один PRB, и назначает одну ресурсную единицу освобожденного PRB новому устройству URLLC. А прерванный сеанс записывается во второй буфер для прерванных устройств eMBB для последующего восстановления соединения. Если все ресурсные единицы заняты под устройства URLLC, то новое устройство URLLC блокируется. После окончания обслуживания каждого из устройств освобождаются ресурсные единицы и больше ничего не происходит, т.е. устройства в буферах продолжают свое ожидание обслуживания.

Устройства из первого буфера Q_1 и второго буфера Q_2 пытаются перезапустить или возобновить обслуживание через экспоненциально распределенное время с параметрами $\gamma_1 > 0$ и $\gamma_2 > 0$ соответственно. Поскольку прерванные сеансы eMBB должны быть возобновлены как можно скорее, буфер Q_2 считается приоритетным. Таким образом, устройство eMBB с задержкой из буфера Q_1 может

установить соединение только в том случае, когда буфер Q_2 полностью свободен и в системе есть хотя бы один свободный PRB. А для восстановления сеанса прерванного устройства eMBB достаточно наличия хотя бы одного свободного PRB в системе, несмотря на состояние первого буфера. В случае когда устройства с буферов не могут быть приняты к обслуживанию, они продолжают ожидание.

3 Система массового обслуживания с орбитами

Рассматриваемую системную модель можно представить в виде СМО с двумя орбитами с исходными параметрами, описанными выше. Схема такой СМО показана на рис. 3.

Функционирование данной СМО опишем четырехмерным марковским случайным процессом (СП):

$$\vec{X}(t)_{t \geq 0} = (N_m(t), N_u(t), Q_1(t), Q_2(t))_{t \geq 0},$$

где $N_m(t)$ и $N_u(t)$ — число сессий соответственно типа eMBB и URLLC в зоне обслуживания в момент времени t ; $Q_1(t)$ — число новых сессий eMBB на

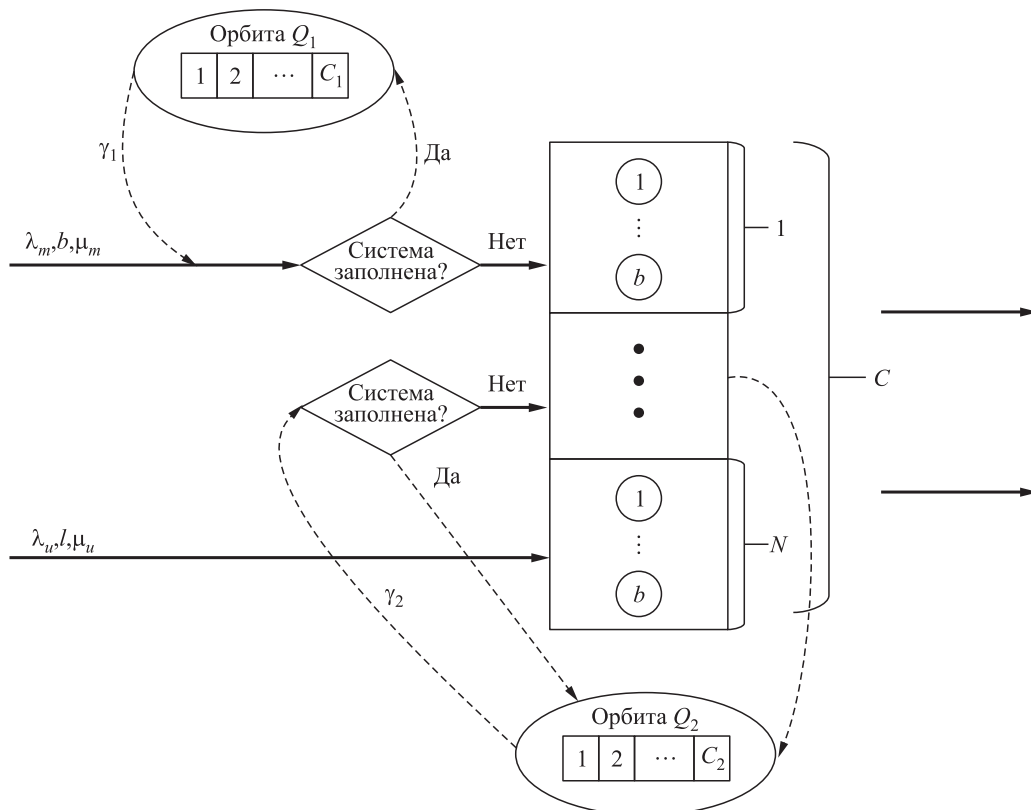


Рис. 3 Схема СМО с двумя орбитами

орбите 1; $Q_2(t)$ — число прерванных сессий eMBB на орбите 2 в момент времени t .

Пространство состояний СП имеет следующий вид:

$$\mathcal{N} = \{(n_m, n_u, q_1, q_2) : n_m \geq 0, n_u \geq 0, 0 \leq q_1 \leq C_1, 0 \leq q_2 \leq C_2, bn_m + n_u \leq C, n_m + q_2 \leq N\}, \quad (1)$$

где n_m и n_u — соответственно число eMBB- и URLLC-сессий на обслуживании; q_1 и q_2 — число eMBB-сессий соответственно на орбите Q_1 и Q_2 . Обозначим через $\pi(n_m, n_u, q_1, q_2)$ вероятность нахождения системы в состоянии (n_m, n_u, q_1, q_2) .

4 Алгоритм расчета распределения вероятностей

Получим рекуррентный алгоритм для расчета финального распределения вероятностей состояний рассматриваемой системы. Матрица интенсивностей переходов представима в блочно-трехдиагональном виде, представленном ниже в виде утверждения 1.

Утверждение 1. Если на множестве (1) введен лексикографический порядок

$$(n'_m, n'_u, q'_1, q'_2) > (n_m, n_u, q_1, q_2) \Leftrightarrow q'_1 > q_1 \text{ или } (q'_1 = q_1, (q'_2 > q_2 \text{ или } (q'_2 = q_2, (n'_m > n_m \text{ или } (n'_m = n_m, n'_u > n_u))))),$$

то матрица интенсивностей переходов СП представима в следующем блочном трехдиагональном виде:

$$\Lambda = \begin{pmatrix} \Lambda_{00} & \Lambda_{11} & 0 & 0 & 0 & \cdots & \cdots & 0 \\ \Lambda_{21} & \Lambda_{01} & \Lambda_{12} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \Lambda_{22} & \Lambda_{02} & \Lambda_{13} & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \Lambda_{2C_1} & \Lambda_{0C_1} \end{pmatrix};$$

размерности блоков матрицы Λ представимы в виде:

$$\begin{aligned} \Lambda_{0k} &: \dim(N(k)) \times \dim(N(k)), \quad k = 0, \dots, C_1; \\ \Lambda_{1k} &: \dim(N(k-1)) \times \dim(N(k)), \quad k = 1, \dots, C_1; \\ \Lambda_{2k} &: \dim(N(k)) \times \dim(N(k-1)), \quad k = 1, \dots, C_1, \end{aligned}$$

где

$$\dim(N(k)) = \sum_{q_2=0}^{C_2} \sum_{n_m=0}^{N-q_2} \sum_{n_u=0}^{C-bn_m} 1;$$

ненулевые элементы блоков Λ_{1k} , Λ_{2k} и Λ_{0k} матрицы Λ выглядят следующим образом:

$$\begin{aligned} \Lambda_{1k}((n_m, n_u, q_1, q_2), (n'_m, n'_u, q'_1, q'_2)) = \\ = \{\lambda_m, q'_1 = q_1 + 1, q_1 \neq 0 \text{ или } q_2 \neq 0; \end{aligned}$$

$$\begin{aligned} \Lambda_{2k}((n_m, n_u, q_1, q_2), (n'_m, n'_u, q'_1, q'_2)) = \\ = \{\gamma_1, n'_m = n_m + 1, q'_1 = q_1 - 1, \\ bn_m + n_u < C - b, q_2 = 0, q_1 \neq 0; \end{aligned}$$

$$\begin{aligned} \Lambda_{0k}((n_m, n_u, q_1, q_2), (n'_m, n'_u, q'_1, q'_2)) = \\ = \begin{cases} \lambda_u, & n'_m = n_m, n'_u = n_u + 1, q'_1 = q_1, \\ & q'_2 = q_2; \\ \text{или} \\ n'_m = n_m - 1, n'_u = n_u + 1, \\ & q'_2 = q_2 + 1; \\ n_u \mu_u, & n'_m = n_m, n'_u = n_u - 1, q'_1 = q_1, \\ & q'_2 = q_2; \\ \lambda_m, & n'_m = n_m + 1, n'_u = n_u, q'_1 = q_1 = 0, \\ & q'_2 = q_2 = 0; \\ n_m \mu_m, & n'_m = n_m - 1, n'_u = n_u, q'_1 = q_1, \\ & q'_2 = q_2; \\ \gamma_2, & n'_m = n_m + 1, n'_u = n_u, q'_1 = q_1, \\ & q'_2 = q_2 - 1; \\ *, & n'_m = n_m, n'_u = n_u, q'_1 = q_1, q'_2 = q_2, \end{cases} \end{aligned}$$

где * равно сумме всех ненулевых элементов матрицы Λ по строке со знаком минус.

После упорядочения пространства состояний согласно утверждению выше, получим алгоритм нахождения финальных вероятностей в форме утверждения 2.

Утверждение 2. Компоненты π_k , $k = \overline{0, C_1 - 1}$, вектора финальных вероятностей π можно вычислить, используя подход «блок — вперед — исключение — обратное замещение», согласно формуле

$$\pi_k = \pi_{C_1} \prod_{j=1}^{N-k} M_{C_1-j}, \quad k = \overline{0, C_1 - 1},$$

где компонента π_{C_1} является единственным решением системы уравнений

$$\begin{aligned} \sum_{k=0}^{C_1} \left(\pi_{C_1} \prod_{j=1}^{C_1-k} M_{C_1-j} \right) \mathbf{e} = \mathbf{1}; \\ \pi_{C_1} (\Lambda_{0C_1} + \Lambda_{2C_1} M_{C_1-1}) = \mathbf{0}, \end{aligned}$$

где \mathbf{e} — вектор-столбец из единиц размерностью $\dim N(k)$, а матрицы M_{C_1-j} вычисляются следующим образом:

$$\begin{aligned} M_0 &= -\Lambda_{11} \cdot \Lambda_{00}^{-1}; \\ M_k &= -\Lambda_{1k+1} (\Lambda_{0k} + M_{k-1} \cdot \Lambda_{2k})^{-1}, \\ & \quad k = \overline{1, C_1 - 1}. \end{aligned}$$

5 Алгоритм расчета вероятностных характеристик модели

Получив финальное распределение вероятностей по утверждению 2, можно рассчитать характеристики рассматриваемой модели.

1. Вероятность блокировки новой сессии URLLC:

$$B_u = \sum_{q_1=0}^{C_1} \sum_{q_2=0}^{C_2} \pi(0, C, q_1, q_2).$$

2. Вероятность задержки передачи новой сессии eMBB:

$$D = \sum_{(n_m, n_u, q_1, q_2) \in \mathcal{N}_D} \pi(n_m, n_u, q_1, q_2),$$

где $\mathcal{N}_D = (n_m, n_u, q_1, q_2) \in \mathcal{N} : (b(n_m + 1) + n_u \leq C, q_1 < C_1, q_2 < C_2, ((q_1 \neq 0) \text{ или } (q_2 \neq 0)))$, или $(b(n_m + 1)n_u > C, q_1 < C_1)$ — множество состояний, где новое устройство eMBB будет отправлено на орбиту Q_1 .

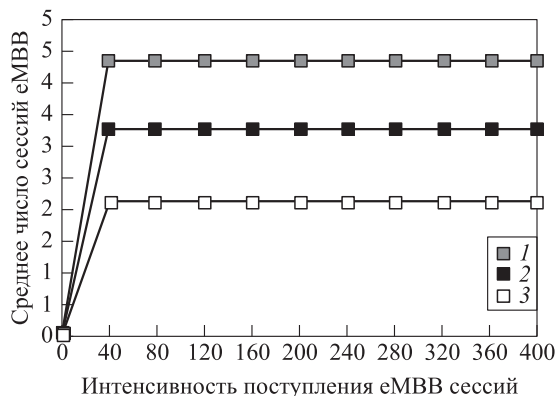
3. Вероятность прерывания сессии eMBB:

$$I = \sum_{n_m=1}^N \sum_{q_1=0}^{C_1} \sum_{q_2=0}^{C_2} \left(\frac{\pi(n_m, C - bn_m, q_1, q_2)}{n_m} \lambda_u / (\lambda_u + n_m \mu_m + n_u \mu_u 1(A) + \lambda_m + \gamma_1 1(B) + \gamma_2 1(C)) \right),$$

где $A = n_u > 0, B = b(n_m + 1) + n_u \leq C, q_1 > 0, q_2 = 0; C = b(n_m + 1) + n_u \leq C, q_2 > 0$.

4. Вероятность блокировки новой сессии eMBB:

$$B_m = \sum_{n_m=0}^N \sum_{n_u=0}^C \sum_{q_2=0}^{C_2} \pi(n_m, n_u, q_1, q_2).$$



(а)

5. Среднее число обслуживаемых устройств URLLC в системе (активных сессий URLLC):

$$\bar{n}_u = \sum_{n_u=1}^C n_u \sum_{n_m=0}^{[(C-n_u)/b]} \sum_{q_1=0}^{C_1} \sum_{q_2=0}^{C_2} \pi(n_m, n_u, q_1, q_2).$$

6. Среднее число обслуживаемых устройств eMBB в системе (активных сессий eMBB):

$$\bar{n}_m = \sum_{n_m=1}^N n_m \sum_{n_u=0}^{C-bn_m} \sum_{q_1=0}^{C_1} \sum_{q_2=0}^{C_2} \pi(n_m, n_u, q_1, q_2).$$

7. Среднее число устройств eMBB с задержкой передачи данных, т. е. среднее число сессий eMBB на орбите Q_1 :

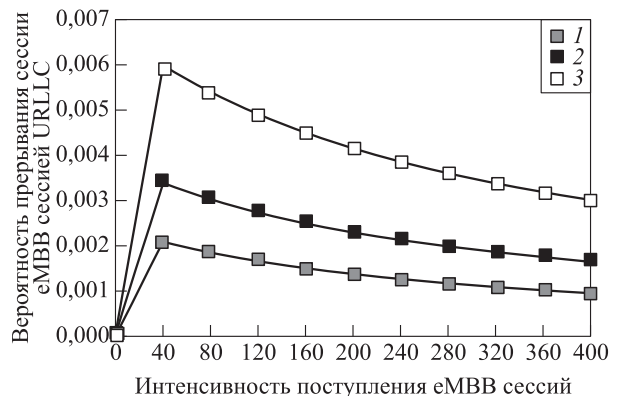
$$\bar{q}_1 = \sum_{q_1=1}^{C_1} q_1 \sum_{n_m=0}^N \sum_{n_u=0}^{C-bn_m} \sum_{q_2=0}^{C_2} \pi(n_m, n_u, q_1, q_2).$$

8. Среднее число прерванных устройств eMBB, т. е. среднее число сессий eMBB на орбите Q_2 :

$$\bar{q}_2 = \sum_{q_2=1}^{C_2} q_2 \sum_{n_m=0}^N \sum_{n_u=0}^{C-bn_m} \sum_{q_1=0}^{C_1} \pi(n_m, n_u, q_1, q_2).$$

6 Пример численного анализа

Для проведения численного анализа рассмотрим следующие исходные данные: $N = 6$ ресурсных блоков по $b = 12$ ресурсных единиц; $C_1 = C_2 = 6$ мест на орбитах Q_1 и Q_2 . Анализ проведем для двух сценариев: в первом сценарии рассматривается зависимость от изменения интенсивности поступления сессий eMBB λ_m с шагом 40 (рис. 4),



(б)

Рис. 4 Первый сценарий: (а) среднее число активных сессий eMBB; (б) вероятность прерывания сессии eMBB; 1 — $\lambda_u = 40$; 2 — 80; 3 — $\lambda_u = 120$

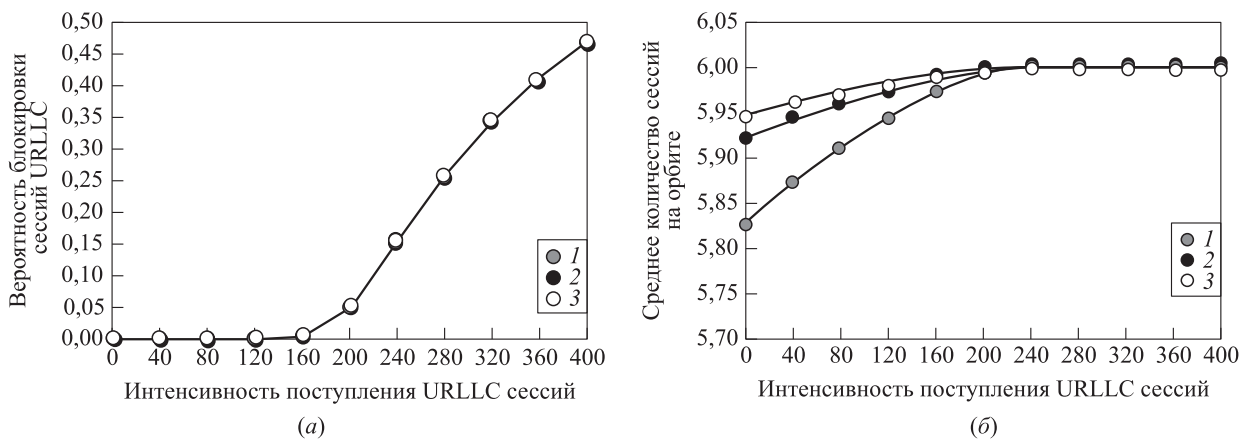


Рис. 5 Второй сценарий: (а) вероятность блокировки новой сессии URLLC; (б) среднее число сессий eMBB на орбите Q_1 ; 1 — $\lambda_m = 40$; 2 — 80; 3 — $\lambda_m = 120$

во втором — от изменения интенсивности поступления сессий URLLC λ_u с тем же шагом (рис. 5). Среднее время обслуживания сессий фиксированное: $\mu_m^{-1} = 1^{-1}$, $\mu_u^{-1} = 3^{-1}$, интенсивности повторных попыток установления соединения сессиями eMBB с орбит Q_1 и Q_2 равны $\gamma_1 = \gamma_2 = 100$.

Как можно видеть на рис. 5, а, интенсивность поступления в систему сессий eMBB не оказывает влияния на вероятность блокировки новых сессий URLLC. Этот результат объясняется абсолютным приоритетом сессий URLLC по отношению к сессиям eMBB. Что касается сессий eMBB в системе на рис. 4, а, наибольшее влияние на их число оказывает интенсивность поступления сессий URLLC. Следовательно, чем ниже скорость поступления устройств URLLC, тем больше устройств eMBB может быть обслужено.

Отметим, что на рис. 5, б видно, что орбита Q_1 почти все время занята, т.е. вероятность блокировки новой сессии eMBB близка к 100%. Заметим, графики вероятности прерывания сессий eMBB (см. рис. 4, б) сначала возрастают, затем резко убывают. Это можно объяснить снижением среднего числа сессий eMBB в системе (рис. 4, а), что вполне логично: чем меньше число обслуживаемых сессий, тем меньше вероятность их прерывания. Также отметим, что рассмотренные характеристики орбит Q_1 и Q_2 зависят от двух рассмотренных интенсивностей поступления сессий.

7 Заключение

В данной работе представлена модель совместного обслуживания трафика eMBB и URLLC в виде СМО с двумя орбитами. Разработан алгоритм расчета финальных вероятностей в матричной форме.

Проведен численный анализ характеристик такой системы.

Литература

1. Popovski P., Trillingsgaard K. F., Simeone O., Durisi G. 5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view // IEEE Access, 2018. Vol. 6. P. 55765–55779.
2. Cheng S., Wang L., Hwang C., Chen J., Cheng L. On-device cognitive spectrum allocation for coexisting URLLC and eMBB users in 5G systems // IEEE T. Cognitive Communications Networking, 2020. 13 p. doi: 10.1109/TCCN.2020.3007890. (This article has been accepted for publication in a future issue of this journal, but has not been fully edited.)
3. Tun Y. K., Kim D. H., Alsenwi M., Tran N. H., Han Z., Hong C. S. Energy efficient communication and computation resource slicing for eMBB and URLLC coexistence in 5G and beyond // IEEE Access, 2020. Vol. 8. P. 136024–136035.
4. Dos Santos E. J., Jr., Souza R. D., Rebelatto J. L., Alves H. Network slicing for URLLC and eMBB with max-matching diversity channel allocation // IEEE Commun. Lett., 2020. Vol. 24. Iss. 3. P. 658–661.
5. Kim Y., Park S. Calculation method of spectrum requirement for IMT-2020 eMBB and URLLC with puncturing based on M/G/1 priority queuing model // IEEE Access, 2020. Vol. 8. P. 25027–25040.
6. Anand A., De Veciana G., Shakkottai S. Joint scheduling of URLLC and eMBB traffic in 5G wireless networks // IEEE ACM T. Network., 2020. Vol. 28. Iss. 2. P. 477–490.
7. Li J., Zhang X. Deep reinforcement learning-based joint scheduling of eMBB and URLLC in 5G networks // IEEE Wirel. Commun. Lett., 2020. Vol. 9. Iss. 9. P. 1543–1546.
8. Gudkova I. A., Samouylov K. E. Modelling a radio admission control scheme for video telephony service in wireless networks // Internet of things, smart spaces, and next generation networking / Eds. S. Andreev, S. Balandin,

Y. Koucheryavy. — Lecture notes in computer science ser. — Springer, 2012. Vol. 7469. P. 208–215.

9. Makeeva E., Polyakov N., Kharin P., Gudkova I. Probability model for performance analysis of joint URLLC and eMBB transmission in 5G networks // Internet

of things, smart spaces, and next generation networks and systems / Eds. O. Galinina, S. Andreev, S. Balandin, Y. Koucheryavy. — Lecture notes in computer science ser. — Springer, 2019. Vol. 11660. P. 635–648.

Поступила в редакцию 15.10.20

RETRIAL QUEUING MODEL FOR ANALYZING JOINT URLLC AND eMBB TRANSMISSION IN 5G NETWORKS

P. A. Kharin¹, E. D. Makeeva¹, I. A. Kochetkova^{1,2}, D. V. Efrosinin^{1,3}, and S. Ya. Shorgin²

¹Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

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

³Johannes Kepler Universitaet Linz, 69 Altenbergerstrasse, Linz 4040, Austria

Abstract: Fifth-generation (5G) networks are characterized by three use cases — massive machine-type communication (mMTC), ultra-reliable low-latency communication (URLLC), and enhanced mobile broadband (eMBB). Quality of service requirements as well as service parameters within these use cases vary significantly, e. g., URLLC is characterized by an ultralow latency of up to 1 ms, while eMBB has an ultrahigh data transfer rate. The task is to organize the joint provision of such services. The paper proposes a scheme for joint URLLC and eMBB traffic transmission. It is based on the fact that URLLC data has a small volume and can occupy less than one physical resource block. The authors analyze the scheme using the developed retrial queuing system with two orbits — one for temporary storage of interrupted eMBB users and the other for eMBB users waiting for service to start. The authors propose the matrix geometric algorithm for calculating the probability distribution as well as formulas for probabilistic characteristics.

Keywords: 5G; ultra-reliable low latency communication (URLLC); enhanced mobile broadband (eMBB); retrial queuing system; interruption

DOI: 10.14357/19922264200403

Acknowledgments

The publication has been prepared with the support of the “RUDN University Program 5-100.” The reported study was funded by RFBR, projects Nos. 18-00-01555(18-00-01685) and 20-37-70079.

References

1. Popovski, P., K. F. Trillingsgaard, O. Simeone, and G. Durisi. 2018. 5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view. *IEEE Access* 6:55765–55779.
2. Cheng, S., L. Wang, C. Hwang, J. Chen, and L. Cheng. 2020. On-device cognitive spectrum allocation for coexisting URLLC and eMBB users in 5G systems. *IEEE T. Cognitive Communications Networking*. 13 p. doi: 10.1109/TCCN.2020.3007890. (This article has been accepted for publication in a future issue of this journal, but has not been fully edited.)
3. Tun, Y. K., D. H. Kim, M. Alsenwi, N. H. Tran, Z. Han, and C. S. Hong. 2020. Energy efficient communication and computation resource slicing for eMBB and URLLC coexistence in 5G and beyond. *IEEE Access* 8:136024–136035.
4. Dos Santos, E. J., R. D. Souza, Jr., J. L. Rebelatto, and H. Alves. 2020. Network slicing for URLLC and eMBB with max-matching diversity channel allocation. *IEEE Commun. Lett.* 24(3):658–661.
5. Kim, Y., and S. Park. 2020. Calculation method of spectrum requirement for IMT-2020 eMBB and URLLC with puncturing based on M/G/1 priority queuing model. *IEEE Access* 8:25027–25040.
6. Anand, A., G. De Veciana, and S. Shakkottai. 2020. Joint scheduling of URLLC and eMBB traffic in 5G wireless networks. *IEEE ACM T. Network.* 28(2):477–490.
7. Li, J., and X. Zhang. 2020. Deep reinforcement learning-based joint scheduling of eMBB and URLLC in 5G networks. *IEEE Wirel. Commun. Lett.* 9(9):1543–1546.
8. Gudkova, I. A., and K. E. Samouylov. 2012. Modelling a radio admission control scheme for video telephony service

in wireless networks. *Internet of things, smart spaces, and next generation networking*. Eds. S. Andreev, S. Balandin, and Y. Koucheryavy. Lecture notes in computer science ser. Springer. 7469:208–215.

9. Makeeva, E., N. Polyakov, P. Kharin, and I. Gudkova. 2019. Probability model for performance analy-

sis of joint URLLC and eMBB transmission in 5G networks. *Internet of things, smart spaces, and next generation networks and systems*. Eds. O. Galinina, S. Andreev, S. Balandin, and Y. Koucheryavy. Lecture notes in computer science ser. Springer. 11660:635–648.

Received October 15, 2020

Contributors

Kharin Petr A. (b. 1993) — PhD student, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; pxarin@mail.ru

Makeeva Elena D. (b. 1996) — PhD student, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; len16730637@yandex.ru

Kochetkova Irina A. (b. 1985) — Candidate of Science (PhD) in physics and mathematics, associate professor, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, 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 119133, Russian Federation; gudkova-ia@rudn.ru

Efrosinin Dmitry V. (b. 1977) — Doctor of Science in physics and mathematics, associate professor, Johannes Kepler Universitaet Linz, 69 Altenbergerstrasse, Linz 4040, Austria; associate professor, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; dmitry.efrosinin@jku.at

Shorgin Sergey Ya. (b. 1952) — Doctor of Science in physics and mathematics, professor, principal 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; sshorgin@ipiran.ru

СТАЦИОНАРНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМЫ $Geo/G/1/\infty$ С НЕОРДИНАРНЫМ ВХОДЯЩИМ ПОТОКОМ, УПРАВЛЯЮЩИМ РАЗМЕРОМ ОЧЕРЕДИ*

С. И. Матюшенко¹, Р. В. Разумчик²

Аннотация: Рассматривается функционирующая в дискретном времени система массового обслуживания (СМО) с одним прибором, очередь неограниченной емкости и неординарным геометрическим потоком заявок. В системе реализован специальный механизм управления очередью: в момент поступления в систему новой группы заявок ее размер сравнивается с текущим общим числом заявок в системе и, если число заявок в новой группе превышает общее число заявок в системе, новая группа целиком принимается в систему, вытесняя при этом все прежде находившиеся в ней заявки; в противном случае новая группа покидает систему, не оказывая на нее никакого воздействия. Заявки обслуживаются прибором по одной. В предположении, что заявки в группе независимы, а распределения чисел заявок в группе и времени обслуживания являются произвольными дискретными, найдены основные стационарные характеристики функционирования.

Ключевые слова: дискретное время; неординарный поток; управление очередью; выходящий поток

DOI: 10.14357/19922264200404

1 Введение

Дисциплины обслуживания очередей, которые позволяют повысить эффективность работы СМО путем использования доступной информации (известной либо точно, либо приближенно) о размерах (временах обслуживания и т. п.) поступающих в них заявок, продолжают оставаться предметом активных научных исследований [1–5]. При этом внимание часто сосредоточено вокруг наиболее известной из всех специальных дисциплин обслуживания — дисциплины преимущественного обслуживания заявки минимальной остаточной длины (shortest remaining time first, SRPT).

В недавней работе [6] авторами рассмотрена новая однолинейная СМО с групповым потоком и специальным механизмом обработки очереди, по которому поступающая группа всегда вытесняет из системы все находящиеся в ней заявки, если размер группы достаточно велик. Побудительным мотивом³ к изучению этого механизма (в отличие от

SRPT) послужил поиск путей максимизации загрузки системы.

В предположениях, что поток групп пуассоновский, заявки в группе независимы, число заявок в группе имеет произвольное, а времена обслуживания — экспоненциальное распределение, в [6] найдены с помощью метода обращения времени основные стационарные характеристики системы. В данной статье рассматривается аналогичная [6] СМО, но функционирующая в дискретном времени и при этом в более общих предположениях о распределении времени обслуживания (оно допускается произвольным дискретным). Предложен основанный на вероятностных соображениях из [9] метод⁴ нахождения совместного стационарного распределения числа заявок в системе и остаточных длин заявок в очереди. Стандартными методами изучен и ряд других стационарных характеристик: период занятости, время пребывания заявки в системе, выходящий поток потерянных заявок.

* Исследование выполнено при финансовой поддержке РФФИ (проект 20-07-00804) и в соответствии с программой Московского центра фундаментальной и прикладной математики.

¹ Российский университет дружбы народов, matyushenko.si@pfur.ru

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

³ Появление описанного механизма, по-видимому, также связано с результатами работ [7, 8], в которых показано, что одна его более простая разновидность приводит к мультипликативному виду совместного стационарного распределения в соответствующим образом образованных сетях массового обслуживания. Отметим, что сети из СМО, рассмотренных в [6] и в этой работе, таким свойством уже не обладают.

⁴ Здесь же необходимо отметить, что в ряде частных случаев (например, для геометрического распределения времени обслуживания) для вывода стационарного распределения, как и в непрерывном времени, может быть применен метод обращения времени.

Прежде чем переходить к подробному описанию системы, отметим, что отличительная особенность рассмотренной СМО заключается в том, что управляющие размером очереди решения принимаются по результатам сравнения не остаточных времен обслуживания (см., например, [10–12]), а остаточных размеров групп заявок. Несмотря на то что в такой СМО поступление заявок может привести к потере уже находящихся в системе заявок, она не относится к типу СМО с отрицательными заявками/сигналами (см., например, [13]).

2 Описание системы

Рассматривается функционирующая в дискретном времени¹ однолинейная СМО с очередью неограниченной емкости, в которую поступает неординарный геометрический поток заявок, определяемый следующим образом. На каждом такте (далее будем называть тактом как интервал времени между соседними изменениями состояния системы, так и сами моменты, в которые происходят эти изменения) с вероятностью a приходит группа заявок случайного размера, не зависящего от всего процесса функционирования системы. При этом в каждой поступившей группе имеется $i \geq 1$ заявок с вероятностью l_i . Заявки обслуживаются прибором по одной, причем время обслуживания заявки становится известным в момент ее поступления на прибор. Распределение времени обслуживания заявки является произвольным дискретным с вероятностью $b_i, i \geq 0$, того, что обслуживание заявки продлится i тактов (предполагается, что $b_0 = 0$).

Будем использовать следующие обозначения:

$\bar{a} = 1 - a$ — вероятность непоступления заявки на такте;

$B_i = \sum_{j=i}^{\infty} b_j, i \geq 0$, — вероятность того, что обслуживание заявки продлится не менее i тактов;

$E b^k = \sum_{i=1}^{\infty} i^k b_i$ — k -й момент времени обслуживания;

$L_i = \sum_{j=i}^{\infty} l_j, i \geq 1$, — вероятность того, что в поступившей группе окажется не менее i заявок.
Очевидно, $L_1 = 1$;

$E l^k = \sum_{i=1}^{\infty} i^k l_i$ — k -й момент размера группы;

$\beta(z) = \sum_{j=1}^{\infty} z^j b_j$ — производящая функция (ПФ) времени обслуживания заявки.

В системе реализован следующий механизм управления очередью. В момент поступления в систему новой группы заявок ее размер x сравнивается

с текущим общим числом заявок в системе y . Та из групп заявок, длина которой больше, остается в системе, а другая покидает систему. Другими словами, если $x > y$, то все y заявок мгновенно уходят из системы; новая группа заявок размера x целиком помещается в очередь, и одна заявка из группы немедленно занимает прибор. Если же $x \leq y$, то поступающая группа заявок теряется, не оказывая на систему никакого воздействия.

Примем, что все изменения состояния СМО происходят в конце такта в следующем порядке²:

- если на этом такте завершилось обслуживание заявки на приборе, то она покидает систему и на прибор сразу же поступает следующая заявка из очереди;
- затем с вероятностью a в систему поступает группа заявок и, если система оказалась непустой, происходит потеря либо поступившей группы, либо всех тех заявок, которые находились в системе (до момента поступления).

Далее будем предполагать, что $L_i > 0$ и $B_i > 0$ при всех i и выполнено условие (5) (см. разд. 4), необходимое и достаточное для существования стационарного режима.

3 Период занятости

Рассмотрим случайный процесс $\{\eta(t) = (\nu(t), \xi(t)), t \geq 0\}$, где $\nu(t)$ — общее число заявок в системе, а $\xi(t)$ — остаточное время обслуживания (далее — длина) заявки на приборе непосредственно после такта t . При $\nu(t) = 0$ координата $\xi(t)$ не определяется. Процесс $\{\eta(t), t \geq 0\}$ является цепью Маркова, причем множество ее состояний \mathcal{X} имеет вид:

$$\mathcal{X} = \{0\} \cup \{(n, i), n \geq 1, i \geq 1\},$$

где n — число заявок в системе; i — остаточная длина заявки на приборе.

Пусть после очередного такта в системе оказалось $n \geq 1$ заявок и обслуживание заявки только началось. Обозначим через $\mathcal{U}_n(z), 0 < z \leq 1$, ПФ числа тактов до того момента, когда в системе впервые окажется $n - 1$ заявок. Если в момент начала функционирования в системе находится $n \geq 1$ заявок, то распределение ее периода занятости (ПЗ) в терминах ПФ имеет вид $\prod_{j=1}^n \mathcal{U}_j(z)$; иначе — $\sum_{n=1}^{\infty} l_n \mathcal{U}_n(z)$. Воспользовавшись формулой полной вероятности, получаем систему уравнений для $\mathcal{U}_n(z)$:

¹Дискретное время вводится обычным образом (см., например, [14]).

²В зарубежной литературе это схема EAS-IA (см., например, [15, с. 2–3]).

$$\mathcal{U}_n(z) = \mathcal{D}_n(z)\mathcal{U}_{n-1}(z) + \mathcal{E}_n(z)\mathcal{U}_n(z) + \mathcal{F}_n(z) \sum_{j=n+1}^{\infty} a_l j \mathcal{U}_j(z), \quad n \geq 1. \quad (1)$$

Здесь используется соглашение $\mathcal{U}_0(z) \equiv 1$ и обозначения:

$$\mathcal{D}_n(z) = \frac{z_{n-1}}{z_n} \beta(z z_n); \quad \mathcal{E}_n(z) = \frac{a l_n}{z_n} \beta(z z_n);$$

$$\mathcal{F}_n(z) = z \frac{1 - \beta(z z_n)}{1 - z z_n},$$

где $z_n = 1 - a L_{n+1}$.

Заметим, что

$$\mathcal{D}_n(z) + \mathcal{E}_n(z) + \mathcal{F}_n(z) \frac{1 - z z_n}{z} = 1.$$

Системе (1) можно придать следующий вид:

$$\mathcal{U}_n(z) = \sum_{j=1}^{\infty} \mathcal{T}_{nj}(z) \mathcal{U}_j(z) + \mathcal{B}_n(z), \quad n \geq 1,$$

где $\mathcal{B}_1(z) = \mathcal{D}_1(z)$, $\mathcal{B}_n(z) = 0$, $n \geq 2$; $\mathcal{T}_{nj}(z)$ — соответствующим образом подобранные по (1) коэффициенты. Поскольку при $0 < z < 1$ в каждой строке $\sum_{j=1}^{\infty} \mathcal{T}_{nj}(z) < 1$ и, очевидно, свободные члены удовлетворяют условию $\mathcal{B}_n(z) \leq K(1 - \sum_{j=1}^{\infty} \mathcal{T}_{nj}(z))$ при некоторой постоянной $K > 0$, то система (1) имеет ограниченное решение, которое может быть найдено методом последовательных приближений (см., например, [16, теоремы Ia, IVa] или [17, теорема 1]). При $z = 1$ единственное решение¹ (1) — это $\mathcal{U}_n(1) = 1$ при всех n , если параметры системы таковы, что стационарный режим функционирования существует (т. е. выполняется (5), см. разд. 4).

4 Стационарное распределение очереди

Введем обозначения:

$P_0 = \lim_{t \rightarrow \infty} P(\nu(t) = 0)$ — стационарная вероятность того, что непосредственно после очередного такта система будет пуста;

$p_{ni} = \lim_{t \rightarrow \infty} P(\nu(t) = n, \xi(t) = i)$, $n \geq 1, i \geq 1$, — стационарная вероятность того, что непосредственно после очередного такта в системе будет n заявок и до окончания обслуживания заявки на приборе останется i тактов.

Положим

$$P_n = \sum_{i=1}^{\infty} p_{ni}, \quad n \geq 1; \quad \bar{P}_n = \sum_{i=0}^n P_i, \quad n \geq 0.$$

Из системы уравнений равновесия (СУР) стандартным образом находится двойная ПФ $\mathcal{P}(u, v) = P_0 + \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} u^n v^i p_{ni}$, $0 < u, v \leq 1$:

$$\mathcal{P}(u, v) = \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} p_{ni} \times$$

$$\times \left((1 - a L_{n+1}) u^n v^{i-1} + \beta(v) \sum_{j=n+1}^{\infty} a_l j u^j \right) +$$

$$+ \sum_{n=1}^{\infty} p_{n1} \left((1 - a L_n) u^{n-1} (\beta(v) - u) + \right.$$

$$\left. + a l_n u^n (\beta(v) - 1) \right) + P_0 \left(\beta(v) \sum_{j=1}^{\infty} a_l j u^j + 1 - a \beta(v) \right).$$

Однако ее вид малопривлекателен для проведения анализа. Поэтому поступим следующим образом. Введем новую СМО с конечным числом n мест для ожидания, отличающуюся от исходной только тем, что если в очереди находится n заявок и поступает новая группа заявок размера больше n , то все находящиеся в системе заявки покидают ее, а (любые) n заявок из новой группы принимаются в систему. Воспользовавшись приемом, введенном в [9] и подробно изложенном в [14], можно показать, что стационарные вероятности состояний в исходной и новой СМО отличаются лишь на постоянный множитель. Это дает возможность записать следующую СУР:

$$p_{ni} = (1 - a L_{n+1}) p_{n,i+1} + a l_n b_i p_{n1} + a l_n b_i \bar{P}_{n-1} +$$

$$+ \bar{P}_n \sum_{j=n+1}^{\infty} a_l j q_{n,i,j}, \quad n \geq 1, \quad i \geq 1, \quad (2)$$

где $q_{n,i,j}$ — условная вероятность того, что, когда в системе впервые окажется n заявок, остаточное время обслуживания заявки на приборе будет равно i при условии, что после очередного такта в системе оказалось $j \geq n + 1$ заявок и обслуживание заявки на приборе только началось. Если система находится в стационарном режиме, то $q_{n,i,j} = b_i$, $n \geq 1, j \geq n + 1$.

Переходя к ПФ $\mathcal{P}_n(z) = \sum_{i=1}^{\infty} z^i p_{ni}$, $0 < z \leq 1$, из (2) получаем:

$$P_n(z) \frac{z - z_n}{z} = (a l_n \beta(z) - z_n) p_{n1} +$$

$$+ a L_n \beta(z) \bar{P}_{n-1} + a L_{n+1} \beta(z) P_n, \quad n \geq 1. \quad (3)$$

¹Ввиду громоздкости выкладок лишь заметим, что это можно показать и прямыми вычислениями, если воспользоваться представлением решения (1), данным в [18, соотн. (7)], и выписать явный вид входящих в него слагаемых.

Подставляя $z = 1$, находим

$$p_{n1} = \bar{P}_{n-1} \frac{1 - z_{n-1}}{z_{n-1}}.$$

Теперь, воспользовавшись теоремой Руше, с учетом найденного вида p_{n1} из (3) имеем:

$$P_n = c_n \bar{P}_{n-1},$$

$$c_n = \frac{z_n}{1 - z_n} \frac{1 - z_{n-1}}{z_{n-1}} \frac{1 - \beta(z_n)}{\beta(z_n)}, \quad n \geq 1.$$

Отсюда, с учетом соотношения $\bar{P}_n = \bar{P}_{n-1} + P_n$, следует, что

$$\bar{P}_n = P_0 \prod_{i=1}^n (1 + c_i), \quad n \geq 1.$$

Используя теперь условие нормировки $\lim_{n \rightarrow \infty} \bar{P}_n = 1$, окончательно получаем, что

$$\left. \begin{aligned} P_0 &= \left(\prod_{i=1}^{\infty} (1 + c_i) \right)^{-1}; \\ P_n &= \frac{c_n}{\prod_{i=n}^{\infty} (1 + c_i)}, \quad n \geq 1. \end{aligned} \right\} \quad (4)$$

Из эргодической теоремы Фостера следует, что необходимым и достаточным условием существования стационарного режима является сходимость произведения $\prod_{i=1}^{\infty} (1 + c_i)$, которая эквивалентна условию

$$\sum_{i=1}^{\infty} \frac{z_i}{1 - z_i} \frac{1 - z_{i-1}}{z_{i-1}} \frac{1 - \beta(z_i)}{\beta(z_i)} < \infty. \quad (5)$$

Для выполнения (5) достаточно¹, чтобы $E b E l < \infty$. При расчете моментов стационарного распределения по (4) необходимо быть уверенным, что соответствующие ряды сходятся; достаточным условием существования $E l^k$ является существование соответствующего момента размера группы. Для расчета же совместного стационарного распределения числа заявок в системе и остаточного времени обслуживания заявки на приборе можно воспользоваться формулой²:

$$p_{ni} = \bar{P}_{n-1} \left(\frac{1 - z_{n-1}}{z_{n-1} z_n^{i-1}} - \left(\frac{a l_n}{z_{n-1}} + a L_{n+1} (1 + c_n) \right) \sum_{j=1}^{i-1} \frac{b_{i-j}}{z_n^j} \right), \quad n, i \geq 1,$$

которая получается путем обращения ПФ (3). Отсюда, поскольку времена обслуживания заявок в группе предполагаются независимыми, немедленно следует совместное стационарное распределение общего числа заявок в системе, остаточного времени обслуживания заявки на приборе и каждой заявки в очереди.

5 Некоторые характеристики производительности

Остановимся на выводе формул для вероятностей потери заявки. Обозначим через π_1 и π_2 соответственно вероятность потери произвольной заявки при поступлении и во время пребывания в системе. Для этого необходимо перейти от стационарных вероятностей P_n по тактам к стационарным вероятностям по моментам поступления заявок в систему (которые будем обозначать P_n^*), а также зафиксировать порядок выбора заявок на обслуживание из очереди.

Нетрудно видеть, что

$$P_0^* = P_0 + p_{11}; \quad P_n^* = P_n - p_{n1} + p_{n+1,1}, \quad n \geq 1.$$

Поскольку случайно выбранная заявка с вероятностью $kl_k/E l$ принадлежит группе размера k , по формуле полной вероятности получаем³:

$$\pi_1 = \sum_{n=1}^{\infty} P_n^* \sum_{j=1}^n \frac{j l_j}{E l}.$$

Предположим, что заявки обслуживаются из очереди в порядке поступления. Обозначим через $\pi_{2,k,j}$, $k \geq 1$, $1 \leq j \leq k$, условную вероятность того, что заявка будет потеряна, при условии что она принята в систему в группе размера k и оказалась в группе на j -м месте. Величины $\pi_{2,k,j}$ могут быть вычислены рекуррентно по формулам:

$$\pi_{2,k,1} = 1 - \frac{\beta(z_k)}{z_k}, \quad k \geq 1;$$

$$\pi_{2,k,j} = \pi_{2,k,1} + (1 - \pi_{2,k,1})(a L_k + z_{k-1} \pi_{2,k-1,j-1}),$$

$$1 \leq j \leq k, \quad k \geq 2.$$

Усредняя $\pi_{2,k,j}$ по распределению размера принятой в систему группы, содержащей случайно

¹ Действительно, так как $\beta(z_n) \geq 1 - (1 - z_n) E b$ и $\beta(z_n) < \beta(z_{n+1})$, то $c_n \leq a L_n E b / (\beta(z_1)(1 - a L_n))$, а ряд $\sum_{i=1}^{\infty} a L_i / (1 - a L_i)$ сходится, если $\sum_{i=1}^{\infty} a L_i = a E l < \infty$.

² Здесь и далее используется соглашение $\sum_{j=1}^0 = 0$.

³ А вероятность потери поступающей группы равна $\sum_{n=1}^{\infty} P_n^* (1 - L_{n+1})$.

выбранную заявку, и предполагая, что, поступая в группе размера k , заявка может равновероятно оказаться на любом из k мест, находим:

$$\pi_2 = \sum_{n=1}^{\infty} \frac{l_n \bar{P}_{n-1}^*}{\sum_{k=1}^{\infty} k l_k \bar{P}_{k-1}^*} \sum_{j=1}^n \pi_{2,n,j}, \quad \bar{P}_n^* = \sum_{i=0}^n P_i^*.$$

Остановимся теперь на стационарных распределениях времен пребывания в системе обслуженной и потерянной заявки. Обозначим через $V_{1,j,k}(z)$ ПФ условных вероятностей того, что заявка будет обслужена и время ее пребывания в системе равно i тактам, при условии что она принята в систему в группе размера k и оказалась в группе на j -м месте. При $j = 1$ заявка будет обслужена, если за время ее пребывания на приборе в систему не поступит группа размера больше k . Поэтому

$$V_{1,1,k}(z) = \sum_{j=1}^{\infty} z^j b_j z_k^{j-1} = \frac{\beta(z z_k)}{z_k}.$$

Если выделенная заявка оказалась не на первом месте в группе, то ее время пребывания зависит от времени обслуживания находящихся перед ней заявок. Так как эти времена по предположению независимы, то в терминах ПФ имеем:

$$V_{1,j,k}(z) = V_{1,1,k}(z) z_{k-1} V_{1,j-1,k-1}(z), \quad 2 \leq j \leq k.$$

Воспользовавшись теперь формулой полной вероятности, получаем следующее выражение для ПФ $V_1(z)$ стационарного распределения времени пребывания обслуженной заявки в системе:

$$V_1(z) = \sum_{n=1}^{\infty} \frac{l_n \bar{P}_{n-1}^*}{\sum_{k=1}^{\infty} k l_k \bar{P}_{k-1}^*} \sum_{j=1}^n V_{1,n,j}(z). \quad (6)$$

Вводя $V_{2,j,k}(z)$ — ПФ условных вероятностей того, что заявка не будет обслужена и время ее пребывания в системе равно i тактам, при условии что она принята в систему в группе размера k и оказалась в группе на j -м месте, — и рассуждая аналогичным образом, нетрудно по формуле полной вероятности получить следующие соотношения:

$$V_{2,1,k}(z) = \frac{1 - z_k}{z_k} \frac{z z_k - \beta(z z_k)}{1 - z z_k};$$

$$V_{2,j,k}(z) = V_{2,1,k}(z) + V_{1,1,k}(z) (1 - z_{k-1} + z_{k-1} V_{2,j-1,k-1}(z)), \quad 2 \leq j \leq k.$$

Безусловная ПФ $V_2(z)$ стационарного распределения времени пребывания в системе принятой, но в итоге потерянной заявки рассчитывается по формуле (6) с заменой $V_{1,n,j}(z)$ на $V_{2,n,j}(z)$.

6 Выходящий поток потерянных заявок

При изучении рассмотренной системы в связке с другими СМО важны характеристики выходящего из нее потока потерянных заявок. В общем случае, очевидно, он не является ординарным. Не является он и геометрическим: числа заявок, покидающих систему на соседних интервалах, представляют собой зависимые случайные величины.

Рассмотрим последовательные моменты τ_n^- , $n \geq 1$, потерь заявок и введем вложенную цепь Маркова $\nu_n = \nu(\tau_n^-)$ — общее число заявок в системе непосредственно после момента τ_n^- . Положим $l_n = \tau_{n+1}^- - \tau_n^-$ и обозначим через $h_{i,t_1,t_2}(n) = P(\nu_n = i, l_n = t_1, l_{n+1} = t_2)$, $i, t_1, t_2 \geq 1$, вероятность того, что после n -й потери общее число заявок в системе будет равно i и длины интервалов между последующими двумя потерями равны t_1 и t_2 тактам соответственно. Положим

$$p_i^-(n) = \sum_{t_1=1}^{\infty} \sum_{t_2=1}^{\infty} h_{i,t_1,t_2}(n);$$

$$h_t(n) = \sum_{i=1}^{\infty} \sum_{t_2=1}^{\infty} h_{i,t,t_2}(n);$$

$$h_{t_1,t_2}(n) = \sum_{i=1}^{\infty} h_{i,t_1,t_2}(n).$$

При выполнении условия существования стационарного режима существуют и стационарные вероятности $p_i^- = \lim_{n \rightarrow \infty} p_i^-(n)$, $i \geq 1$, того, что непосредственно после момента потери в системе будет i заявок, а также стационарные вероятности

$$h_t = \lim_{n \rightarrow \infty} h_t(n), \quad t \geq 1;$$

$$h_{t_1,t_2} = \lim_{n \rightarrow \infty} h_{t_1,t_2}(n), \quad t_1, t_2 \geq 1.$$

Система уравнений Колмогорова—Чепмена для вероятностей $p_i^-(n)$, $n \geq 2$, $i \geq 1$, имеет вид:

$$p_i^-(n+1) = p_i^-(1) \sum_{j=1}^{\infty} p_j^-(n) \frac{\beta(\bar{a})^j}{\bar{a}} +$$

$$+ \left(a l_i \sum_{j=1}^{i-1} p_j^-(n) + p_i^-(n) \sum_{j=1}^i a l_j \right) \left(1 - \frac{\beta(\bar{a})}{\bar{a}} \right) +$$

$$+ \sum_{k=1}^{\infty} \left(a l_i \sum_{j=k+1}^{k+i-1} p_j^-(n) + p_{k+i}^-(n) \sum_{j=1}^i a l_j \right) \times$$

$$\times \frac{\beta(\bar{a})^k (1 - \beta(\bar{a}))}{\bar{a}}, \quad n \geq 1, \quad i \geq 1, \quad (7)$$

к которому необходимо добавить условие нормировки $\sum_{i=1}^{\infty} p_i^-(n+1) = 1$. Вероятности $p_i^-(1)$, $i \geq 1$, того, что после первой потери в системе окажется i заявок, также могут быть найдены из (7), если зафиксировать общее число заявок в системе в начальный момент функционирования. Так, если изначально система пуста, для нахождения $\{p_i^-(1), i \geq 1\}$ достаточно положить в (7) $n = 0$ и $p_i^-(0) = a\bar{a}^{i-1}$, $i \geq 1$. Устремляя в (7) $n \rightarrow \infty$, получаем систему уравнений для стационарных вероятностей $p_i^-, i \geq 1$, решение которой при сделанных предположениях о распределениях $\{b_i, i \geq 0\}$ и $\{l_i, i \geq 1\}$ (см. разд. 2) может быть найдено численно.

Перейдем к нахождению распределений $\{h_t, t \geq 1\}$ и $\{h_{t_1, t_2}, t_1, t_2 \geq 1\}$. Обозначим через $h_{t,i,j}^{(1)}$, $t, i, j \geq 1$, условную вероятность того, что очередная потеря произойдет через t тактов, при условии что изначально в системе находится i заявок и остаточное время обслуживания заявки на приборе равно j тактам. Положим

$$h_{t,i}^{(1)} = \sum_{j=1}^{\infty} b_j h_{t,i,j}^{(1)}.$$

Воспользовавшись формулой полной вероятности, находим

$$h_{t,1,j}^{(1)} = \mathbf{1}_{(1 \leq j \leq t-1)} \sum_{k=0}^{t-j-1} \bar{a}^{k+j-1} \sum_{m=1}^{\infty} a l_m h_{t-j-k,m}^{(1)} + \mathbf{1}_{(j \geq t+1)} \bar{a}^{t-1} a, \quad j \geq 1; \quad (8)$$

$$h_{t,i,j}^{(1)} = \mathbf{1}_{(1 \leq j \leq t-1)} \bar{a}^j h_{t-j,i-1}^{(1)} + \mathbf{1}_{(j \geq t)} \bar{a}^{t-1} a, \quad i \geq 2, \quad j \geq 1, \quad (9)$$

где $\mathbf{1}_{(A)}$ — индикатор множества A . Соотношения (8) и (9) позволяют последовательно по t , начиная с $t = 1$, определять¹ вероятности $h_{t,i,j}^{(1)}$ и $h_{t,i}^{(1)}$. Усредняя $h_{t,i}^{(1)}$ по распределению $\{p_i^-, i \geq 1\}$, получаем выражение для стационарного распределения длины интервала между последовательными потерями:

$$h_t = \sum_{i=1}^{\infty} h_{t,i}^{(1)} p_i^-, \quad t \geq 1. \quad (10)$$

Формулы для совместного стационарного распределения длин $k, k \geq 2$, последовательных интервалов между потерями могут быть получены аналогичным образом. Остановимся на случае $k = 2$. Обозначим через $h_{t_1, t_2, i}^{(2)}$, $t_1, t_2, i \geq 1$, условную вероятность того, что очередная потеря произойдет через t_1 тактов, а последующая — через t_2 так-

тов, при условии что сразу после очередной потери в системе находится i заявок. Вводя обозначение

$$\Delta_{t,i,j} = h_{t,i,j-1}^{(1)} \sum_{m=1}^i a l_m + \sum_{m=i+1}^{\infty} a l_m h_{t,m}^{(1)}$$

и применяя формулу полной вероятности, получаем, что $h_{t_1, t_2, i}^{(2)}$ могут быть рассчитаны на основе $h_{t,i}^{(1)}$ рекуррентно по следующим формулам:

$$\begin{aligned} h_{t_1, t_2, 1}^{(2)} &= \sum_{j=1}^{t_1-1} b_j \sum_{k=0}^{t_1-j-1} \bar{a}^{k+j-1} \sum_{m=1}^{\infty} a l_m h_{t_1-j-k, t_2, m}^{(2)} + \\ &\quad + \sum_{j=t_1+1}^{\infty} b_j \bar{a}^{t_1-1} \Delta_{t_2, 1, j-t_1+1}; \\ h_{t_1, t_2, i}^{(2)} &= \sum_{j=1}^{t_1-1} b_j \bar{a}^{j-1} h_{t_1-j, t_2, i-1}^{(2)} + \\ &\quad + b_{t_1} \bar{a}^{t_1-1} \sum_{m=1}^{\infty} a l_m h_{t_2, \max(m, i-1)}^{(1)} + \\ &\quad + \sum_{j=t_1+1}^{\infty} b_j \bar{a}^{t_1-1} \Delta_{t_2, i, j-t_1+1}. \end{aligned}$$

Усредняя $h_{t_1, t_2, i}^{(2)}$, как в (10), получаем безусловное распределение длин последовательных двух интервалов между потерями в стационарном режиме.

7 Заключение

В связи с найденным видом стационарного распределения встает вопрос о точности вычислений. При расчете P_0 по (4) нельзя сказать, когда нужно оборвать вычисления, чтобы гарантировать заданную точность. Этот вопрос, встающий особенно остро, когда распределение $\{b_i, i \geq 0\}$ имеет тяжелый хвост, требует дополнительных исследований. Некоторым ориентиром на практике могут служить двусторонние оценки для P_0 , например $e^{-\sum_{i=1}^{\infty} c_i} \leq P_0 \leq (1 + \sum_{i=1}^{\infty} c_i)^{-1}$ (см. далее, например, [19]). Полезными могут оказаться и приближенные формулы (см., например, (8) в [6]). В плане дальнейших исследований несомненный интерес представляет обобщение использованного метода на несколько СМО, связанных рассмотренной дисциплиной обслуживания, а также снятие наложенных на входящий поток ограничений.

Литература

1. Schroeder B., Harchol-Balter M. Web servers under overload: How scheduling can help // ACM T. Internet Techn., 2006. Vol. 6. Iss. 1. P. 20–52.

¹ Расчет необходимо вести в следующем порядке: $h_{1,1,1}^{(1)}, h_{1,1,2}^{(1)}, \dots, h_{1,1,i}^{(1)}, h_{1,2}^{(1)}, \dots$

2. Pradhan S., Gupta U.C. Modeling and analysis of an infinite-buffer batch-arrival queue with batch-size-dependent service // Perform. Evaluation, 2017. Vol. 108. P. 16–31.
3. Groszof I., Scully Z., Harchol-Balter M. SRPT for multi-server systems // Perform. Evaluation, 2018. Vol. 127–128. P. 154–175.
4. Marin A., Mitrani I., Elahi B.M., Williamson C. Control and optimization of the SRPT service policy by frequency scaling // Conference (International) on Quantitative Evaluation of Systems / Eds. A. McIver, A. Horvath. — Lecture notes in computer science ser. — Springer, 2018. Vol. 11024. P. 257–272.
5. Scully Z., Harchol-Balter M., Scheller-Wolf A. Simple near-optimal scheduling for the M/G/1 // SIGMETRICS Perform. Eval. Rev., 2019. Vol. 47. Iss. 2. P. 24–26.
6. Marin A., Rossi S. A queueing model that works only on biggest jobs // European Workshop on Performance Engineering / Eds. M. Gribaudo, M. Iacono, T. Phung-Duc, R. Razumchik. — Lecture notes in computer science ser. — Springer, 2020. Vol. 12039. P. 118–132.
7. Pittel B.G. Closed exponential networks of queues with saturation: The Jackson-type stationary distribution and its asymptotic analysis // Math. Oper. Res., 1979. Vol. 4. Iss. 4. P. 357–378.
8. Balsamo S., Harrison P., Marin A. A unifying approach to product-forms in networks with finite capacity constraints // SIGMETRICS Perform. Eval. Rev., 2010. Vol. 38. Iss. 1. P. 25–36.
9. Печинкин А.В. Об одной инвариантной системе массового обслуживания // Math. Operationsforsch. Statist. Ser. Optimization, 1983. Vol. 14. No. 3. P. 433–444.
10. Таташев А.Г. Многоканальная система массового обслуживания с потерями кратчайших требований // Автоматика и телемеханика, 1991. № 7. С. 187–189.
11. Милованова Т.А. Система ВМАР/G/1 с инверсионным порядком обслуживания и вероятностным приоритетом // Автоматика и телемеханика, 2009. № 5. С. 155–168.
12. Мейханаджян Л.А. Стационарные вероятности состояний в системе обслуживания конечной емкости с инверсионным порядком обслуживания и обобщенным вероятностным приоритетом // Информатика и её применения, 2016. Т. 10. Вып. 2. С. 123–131.
13. Бочаров П.П., Гаврилов Е.В., Печинкин А.В. Экспоненциальная сеть массового обслуживания с зависимым обслуживанием, отрицательными заявками и изменением типа заявок // Автоматика и телемеханика, 2004. № 7. С. 35–59.
14. Печинкин А.В., Разумчик Р.В. Системы массового обслуживания в дискретном времени. — М.: Физматлит, 2018. 432 с.
15. Nobel R. Retrial queueing models in discrete time: A short survey of some late arrival models // Ann. Oper. Res., 2015. Vol. 247. Iss. 1. P. 37–63.
16. Канторович Л.В., Крылов В.И. Приближенные методы высшего анализа. — 5-е изд. — М.—Л.: Физматлит, 1962. 708 с.
17. Shivakumar P.N., Williams J.J. An iterative method with truncation for infinite linear systems // J. Comput. Appl. Math., 1988. Vol. 24. P. 199–207.
18. Carmichael R.D. On non-homogeneous equations with an infinite number of variables // Am. J. Math., 1914. Vol. 36. Iss. 1. P. 13–20.
19. Klamkin M.S., Newman D.J. Extensions of the Weierstrass product inequalities // Math. Mag., 1970. Vol. 43. Iss. 3. P. 137–141.

Поступила в редакцию 15.10.20

STATIONARY CHARACTERISTICS OF DISCRETE-TIME Geo/G/1/∞ QUEUE WITH BATCH ARRIVALS AND ONE QUEUE SKIPPING POLICY

S. I. Matyushenko¹ and R. V. Razumchik²

¹Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, 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: Consideration is given to the discrete-time single-server system with one queue of infinite capacity and the geometric (Bernoulli) input flow. Customers are homogeneous, arrive in batches, and are served one by one in FIFO (first in, first out) manner. The sizes of arriving batches as well as the service times are assumed to be independent and identically distributed random variables with arbitrary discrete distributions. The queue skipping policy is implemented in the system: upon arrival of a batch, its size is compared with the current total number of customers in the system. If the size of the batch is larger than the system content, all customers residing in the

system (including the one in server) are lost and the arrived batch enters the system; otherwise, the new batch leaves the system having no effect on it. Main stationary system performance characteristics, including those of the flow of lost customers, are obtained.

Keywords: discrete-time; queueing system; batch arrivals; queue skipping policy

DOI: 10.14357/19922264200404

Acknowledgments

The reported study was funded by RFBR (project number 20-07-00804) and conducted in accordance with the Program of Moscow Center for Fundamental and Applied Mathematics.

References

1. Schroeder, B., and M. Harchol-Balter. 2006. Web servers under overload: How scheduling can help. *ACM T. Internet Techn.* 6(1):20–52.
2. Pradhan, S., and U. C. Gupta. 2017. Modeling and analysis of an infinite-buffer batch-arrival queue with batch-size-dependent service. *Perform. Evaluation* 108:16–31.
3. Grosz, I., Z. Scully, and M. Harchol-Balter. 2018. SRPT for multiserver systems. *Perform. Evaluation* 127–128:154–175.
4. Marin, A., I. Mitrani, B. M. Elahi, and C. Williamson. 2018. Control and optimization of the SRPT service policy by frequency scaling. *Conference (International) on Quantitative Evaluation of Systems*. Eds. A. McIver, and A. Horvath. Lecture notes in computer science ser. Springer. 11024:257–272.
5. Scully, Z., M. Harchol-Balter, and A. Scheller-Wolf. 2019. Simple near-optimal scheduling for the $M/G/1$. *SIGMETRICS Perform. Eval. Rev.* 47(2):24–26.
6. Marin, A., and S. Rossi. 2020. A queueing model that works only on biggest jobs. *European Workshop on Performance Engineering*. Eds. M. Gribaudo, M. Iacono, T. Phung-Duc, and R. Razumchik. Lecture notes in computer science ser. Springer. 12039:118–132.
7. Pittel, B. G. 1979. Closed exponential networks of queues with saturation: The Jackson-type stationary distribution and its asymptotic analysis. *Math. Oper. Res.* 4(4):357–378.
8. Balsamo, S., P. Harrison, and A. Marin. 2010. A unifying approach to product-forms in networks with finite capacity constraints. *SIGMETRICS Perform. Eval. Rev.* 38(1):25–36.
9. Pechinkin, A. V. 1983. Ob odnoy invariantnoy sisteme massovogo obsluzhivaniya [On an invariant queueing system]. *Math. Operationsforsch. Statist. Ser. Optimization* 14(3):433–444.
10. Tatashev, A. G. 1991. A queueing system with invariant discipline. *Automat. Rem. Contr.* 52(7):1034–1037.
11. Milovanova, T. A. 2009. BMAP/G/1/∞ system with last come first served probabilistic priority. *Automat. Rem. Contr.* 70(5):885–896.
12. Meykhanadzhyan, L. A. 2016. Stacionarnye veroyatnosti sostoyaniy v sisteme obsluzhivaniya konechnoy emkosti s inversionnym poryadkom obsluzhivaniya i obobshchennym veroyatnostnym prioriteto [Stationary characteristics of the finite capacity queueing system with inverse service order and generalized probabilistic priority]. *Informatika i ee Primeneniya — Inform. Appl.* 10(62):123–131.
13. Bocharov, P. P., E. V. Gavrilov, and A. V. Pechinkin. 2004. Exponential queueing network with dependent servicing, negative customers, and modification of the customer type. *Automat. Rem. Contr.* 65(7):1066–1088.
14. Pechinkin, A. V., and R. V. Razumchik. 2018. *Sistemy massovogo obsluzhivaniya v diskretnom vremeni* [Discrete time queueing systems]. Moscow: Fizmatlit. 432 p.
15. Nobel, R. 2015. Retrial queueing models in discrete time: A short survey of some late arrival models. *Ann. Oper. Res.* 247(1):37–63.
16. Kantorovich, L. V., and V. I. Krylov. 1962. *Priblizhennyye metody vysshego analiza*. Moscow–Saint-Petersburg: Fizmatlit. 708 p.
17. Shivakumar, P. N., and J. J. Williams. 1988. An iterative method with truncation for infinite linear systems. *J. Comput. Appl. Math.* 24:199–207.
18. Carmichael, R. D. 1914. On non-homogeneous equations with an infinite number of variables. *Am. J. Math.* 36(1):13–20.
19. Klamkin, M. S., and D. J. Newman. 1970. Extensions of the Weierstrass product inequalities. *Math. Mag.* 43(3):137–141.

Received October 15, 2020

Contributors

Matyushenko Sergey I. (b. 1963) — Candidate of Science (PhD) in physics and mathematics, associate professor, Department of Applied Informatics and Probability Theory, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; matyushenko_si@pfur.ru

Razumchik Rostislav V. (b. 1984) — Candidate of Science (PhD) in physics and mathematics, 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; rrazumchik@ipiran.ru

О РАСПРЕДЕЛЕНИИ ОТНОШЕНИЯ СУММЫ ЭЛЕМЕНТОВ ВЫБОРКИ, ПРЕВОСХОДЯЩИХ НЕКОТОРЫЙ ПОРОГ, К СУММЕ ВСЕХ ЭЛЕМЕНТОВ ВЫБОРКИ. II*

В. Ю. Королев¹

Аннотация: Рассматривается задача описания распределения доли суммы независимых случайных величин, которая составлена из слагаемых, превосходящих некоторый заданный порог. В отличие от известных вариантов такой задачи, в которых фиксируется число суммируемых крайних порядковых статистик, особенность рассматриваемой здесь задачи заключается в том, что заданный порог может быть превзойден не предсказуемым заранее числом элементов выборки. Для случая, когда порог неограниченно возрастает с увеличением объема выборки, показано, что распределение указанного отношения может быть приближено обобщенным пуассоновским распределением, в котором обобщающим служит обобщенное распределение Парето.

Ключевые слова: сумма независимых случайных величин; случайная сумма; биномиальное распределение; смесь распределений вероятностей; экстремальная порядковая статистика; теорема Балкема – Де Хаана – Пикандса; обобщенное распределение Парето; обобщенное пуассоновское распределение

DOI: 10.14357/19922264200405

В 1897 г. итальянский экономист и социолог Вильфридо Парето выявил эмпирическую закономерность, заключающуюся в том, что 80% дохода страны аккумулируются в 20% семей [1]. Эту закономерность многие пытались обосновать или опровергнуть. В данной заметке предпринята попытка рассмотреть связанную с этим принципом вероятностно-статистическую задачу о том, какую долю суммы наблюдений составляют наблюдения, превосходящие заданный порог.

Эта задача имеет большое значение не только для экономики, но и для других областей знания. Например, прослеживая изменение во времени параметров распределения отношения суммы элементов выборки, превосходящих некоторый порог, к сумме всех элементов выборки при исследовании метеорологических данных (температура, осадки, теплообмен между атмосферой и океаном) во времени (например, в скользящем режиме, когда выборка — это «окно», сдвигающееся в направлении астрономического времени при исследовании соответствующего временного ряда), можно получить информацию об особенностях проявления процесса изменения климата, в просторечии называемого «глобальным потеплением».

Данная статья продолжает исследования, начатые в работе [2]. С математической точки зрения

эта задача тесно связана со статистикой цензурированных выборок. Известны разные варианты этой задачи. Некоторые из них упомянуты в [2], где рассмотрена задача описания распределения доли суммы независимых случайных величин, которая составлена из слагаемых, превосходящих некоторый заданный *фиксированный* порог. В отличие от известных вариантов такой задачи, в которых фиксируется число суммируемых крайних порядковых статистик, особенность рассмотренной в [2] задачи заключается в том, что заданный порог может быть превзойден не предсказуемым заранее числом элементов выборки.

В указанной статье в терминах функции распределения отдельного слагаемого формально представлен явный вид распределения отношения суммы элементов выборки, превосходящих заданный порог, к сумме всех наблюдений. На эвристическом уровне выведены асимптотические и предельные распределения этого отношения при фиксированном пороге, удобные для использования в качестве асимптотических аппроксимаций в практических вычислениях. Рассмотрены ситуации, в которых распределение слагаемых имеет легкий хвост (конечны вторые моменты), и ситуации, в которых распределение слагаемых имеет тяжелый хвост (принадлежит к области притяжения устойчивого

*Исследования выполнены при поддержке РФФИ (проект 19-07-00914) и в соответствии с программой Московского центра фундаментальной и прикладной математики.

¹Факультет вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, vkorolev@cs.msu.ru

закона). Во всех случаях описана нормировка отношения, гарантирующая невырожденность предельного (при неограниченном увеличении числа слагаемых) распределения, и сами предельные распределения (нормальное в случае легких хвостов и устойчивое в случае тяжелых хвостов). В настоящей работе рассмотрен случай, когда порог неограниченно возрастает с увеличением объема выборки. Показано, что распределение указанного отношения может быть приближено обобщенным пуассоновским распределением, в котором обобщающим служит обобщенное распределение Парето.

Обозначим $S_n = X_1 + \dots + X_n$. Индикатор множества (события) A обозначим $\mathbb{I}(A)$.

Пусть $u > 0$ таково, что $0 < F(u) < 1$. Очевидно, $X_j = X_j \mathbb{I}(X_j < u) + X_j \mathbb{I}(X_j \geq u)$. Тогда

$$S_n = \sum_{j=1}^n X_j \mathbb{I}(X_j < u) + \sum_{j=1}^n X_j \mathbb{I}(X_j \geq u) \equiv \equiv S_n^{(<u)} + S_n^{(\geq u)}.$$

Основным объектом изучения будет распределение отношения $R(u) = S_n^{(\geq u)} / S_n$, а в первую очередь — распределение случайной величины $S_n^{(\geq u)}$ при условии, что порог возрастает при увеличении объема выборки.

Как было показано в статье [2], если порог u фиксирован, а объем n доступной выборки настолько большой (неограниченно возрастает), что обе суммы $S_n^{(<u)}$ и $S_n^{(\geq u)}$ содержат много слагаемых, то для распределения случайной величины $S_n^{(<u)}$ можно применять нормальную аппроксимацию. Если хвосты распределения случайных величин X_j убывают достаточно быстро (так, что конечен второй момент), то нормальная аппроксимация также справедлива для распределения случайных величин $S_n^{(\geq u)}$ и R_u . Если же функция $F(u)$ имеет столь тяжелый хвост, что у случайной величины X_j отсутствует дисперсия, то для распределений случайных величин $S_n^{(\geq u)}$ и R_u справедлива аппроксимация устойчивым законом.

Однако иногда интерес представляет ситуация, в которой порог u столь велик, что сумма $S_n^{(\geq u)}$ содержит лишь умеренное число слагаемых. В таком случае для распределения случайной величины $S_n^{(\geq u)}$ оказывается возможным применить обобщенную пуассоновскую аппроксимацию.

Формально предположим, что порог u зависит от n : $u = u_n$. При этом будем считать, что существует число $\lambda \in (0, \infty)$ такое, что

$$n(1 - F(u_n)) \longrightarrow \lambda$$

при $n \rightarrow \infty$. Например, если для простоты предположить, что $n(1 - F(u_n)) = \lambda$, то $u_n = F^{-1}(1 - \lambda/n)$. При этом порог u_n не обязан неограниченно возрастать при $n \rightarrow \infty$, например, если носитель распределения F конечен, т. е. $\text{rext } F \equiv \sup\{x : F(x) < 1\} < \infty$.

В сделанных предположениях

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \left| \mathbb{P}(N_n(u_n) = k) - e^{-\lambda} \frac{\lambda^k}{k!} \right| = 0.$$

Пусть N_λ — случайная величина, имеющая распределение Пуассона с параметром λ . Предположим, что при каждом $n \geq 1$ случайная величина N_λ независима от последовательности $X_1^{(\geq u_n)}, X_2^{(\geq u_n)}$.

Имеем

$$\begin{aligned} & \sup_x \left| \mathbb{P}(S_n^{(\geq u)} < x) - \mathbb{P}\left(\sum_{k=0}^{N_\lambda} X_j^{(\geq u_n)} < x\right) \right| = \\ & = \sup_x \left| \sum_{k=0}^{\infty} \mathbb{P}(N_n(u_n) = k) \mathbb{P}\left(\sum_{j=0}^k X_j^{(\geq u_n)} < x\right) - \right. \\ & \quad \left. - \sum_{k=0}^{\infty} \mathbb{P}(N_\lambda = k) \mathbb{P}\left(\sum_{j=0}^k X_j^{(\geq u_n)} < x\right) \right| \leq \\ & \leq \sum_{k=0}^{\infty} |\mathbb{P}(N_n(u_n) = k) - \mathbb{P}(N_\lambda = k)| \leq \\ & \leq 2(1 - F(u_n)) \min\{1, \lambda\} \end{aligned}$$

(см., например, [3]).

Характеристическую функцию случайной величины $X_j^{(\geq u_n)}$ обозначим $f_{u_n}(t)$:

$$f_{u_n}(t) = \frac{1}{1 - F(u_n)} \int_{u_n}^{\infty} e^{itx} dF(x), \quad t \in \mathbb{R}.$$

Из сказанного выше вытекает, что в сделанных предположениях распределение случайной величины $S_n^{(\geq u_n)}$ может быть аппроксимировано обобщенным пуассоновским распределением, задаваемым характеристической функцией

$$g_n(t; \lambda) = \exp\{\lambda [f_{u_n}(t) - 1]\}, \quad t \in \mathbb{R}.$$

Более того, если $\text{rext } F = \infty$, то в сделанных предположениях $u_n \rightarrow \infty$. Это позволяет для распределения случайной величины $S_n^{(\geq u_n)}$ использовать аппроксимацию, не зависящую от вида функции распределения F , а именно справедлива теорема Балкемы — Де Хаана — Пикандса [4, 5], согласно которой, если при некоторой линейной нормировке с помощью числовых последовательностей c_n и $b_n > 0$ распределения случайных величин $(X_{(n)} - c_n)/b_n$ имеют слабым пределом при

$n \rightarrow \infty$ невырожденную функцию распределения (обязательно при этом принадлежащую к одному из трех возможных типов предельных распределений экстремальных значений, см., например, [6]), то существуют такие числа $\alpha, \beta > 0$ и γ , что

$$\begin{aligned} \lim_{u \rightarrow \infty} P(X_j - u < x | X_j > u) &= \\ &= \lim_{u \rightarrow \infty} \frac{F(x+u) - F(u)}{1 - F(u)} = \\ &= H_{\alpha, \beta, \gamma}(x) \equiv 1 - \left(1 + \gamma \cdot \frac{x - \alpha}{\beta}\right)^{-1/\gamma}, \quad x > 0. \end{aligned}$$

Распределение, соответствующее функции распределения $H_{\alpha, \beta, \gamma}(x)$, называется *обобщенным распределением Парето*, при этом γ — параметр формы, α — параметр положения (сдвига), β — параметр масштаба. При $\gamma > 0$ $H_{\alpha, \beta, \gamma}(x)$ — это распределение Парето, при $\gamma = 0$ $H_{\alpha, \beta, \gamma}(x)$ — это показательное распределение, при $\gamma < 0$ $H_{\alpha, \beta, \gamma}(x)$ — это бета-распределение.

Характеристическую функцию обобщенного распределения Парето $H_{\alpha, \beta, \gamma}(x)$ обозначим $h_{\alpha, \beta, \gamma}(t)$:

$$\begin{aligned} h_{\alpha, \beta, \gamma}(t) &= \int_0^{\infty} e^{itx} dH_{\alpha, \beta, \gamma}(x) = \\ &= \frac{1}{\beta} \int_0^{\infty} e^{itx} \left(1 + \gamma \cdot \frac{x - \alpha}{\beta}\right)^{-(\gamma+1)/\gamma} dx, \quad t \in \mathbb{R}. \end{aligned}$$

Пусть Y_1, Y_2, \dots — независимые случайные величины, имеющие одно и то же обобщенное распределение Парето $H_{\alpha, \beta, \gamma}(x)$. Тогда из теоремы Балкемы – Де Хаана – Пикандса и вида распределения случайной величины $X_j^{(\geq u)}$ вытекает, что при большом u справедливо представление

$$X_j^{(\geq u)} \approx u + Y_j,$$

в котором параметры α, β и γ обобщенного распределения Парето случайных величин Y_j зависят от вида функции распределения F .

Тогда из сказанного выше вытекает, что если $u_n \rightarrow \infty$ и $n(1 - F(u_n)) \rightarrow \lambda$ при $n \rightarrow \infty$, то

$$P(S_n^{(\geq u)} < x) \approx P\left(\sum_{j=0}^{N_\lambda} (u_n + Y_j) < x\right),$$

т.е. в таком случае для аппроксимации распределения случайной величины $S_n^{(\geq u)}$ можно использовать обобщенное пуассоновское распределение, соответствующее характеристической функции

$$g_n(t; \alpha, \beta, \gamma, \lambda) = \exp\{\lambda [e^{itu_n} h_{\alpha, \beta, \gamma}(t) - 1]\}, \quad t \in \mathbb{R}, \quad (1)$$

при этом параметры α, β, γ и λ оцениваются статистически.

Что же касается распределения отношения $S_n^{(\geq u_n)} / S_n^{(< u_n)}$ при пороге, растущем описанным выше образом, то можно заметить, что, тогда как при $n \rightarrow \infty$ сумма $S_n^{(\geq u_n)}$ содержит умеренное число слагаемых, число слагаемых в сумме $S_n^{(< u_n)}$ неограниченно возрастает, и, поскольку эти слагаемые ограничены, можно считать, что

$$\frac{S_n^{(< u_n)}}{n} \approx EX_1^{(< u_n)} \equiv \underline{a}_{u_n}.$$

Тогда при указанных выше условиях

$$n \frac{S_n^{(\geq u_n)}}{S_n^{(< u_n)}} \approx \frac{Z_n(\alpha, \beta, \gamma, \lambda)}{\underline{a}_{u_n}},$$

где $Z_n(\alpha, \beta, \gamma, \lambda)$ — случайная величина с обобщенной пуассоновской характеристической функцией $g_n(t; \alpha, \beta, \gamma, \lambda)$ (см. (1)).

Литература

1. *Koch P.* Принцип 80/20 / Пер. с англ. — М.: Эксмо, 2012. 352 с. (*Koch R.* The 80/20 principle: The secret of achieving more with less. — London: Nicholas Brealey Publishing, 1998. 302 p.)
2. *Королев В. Ю.* О распределении отношения суммы элементов выборки, превосходящих некоторый порог, к сумме всех элементов выборки. I // Информатика и её применения, 2020. Т. 14. Вып. 3. С. 26–34.
3. *Barbour A. D., Hall P.* On the rate of Poisson convergence // P. Camb. Philos. Soc., 1984. Vol. 95. P. 473–480.
4. *Balkema A., de Haan L.* Residual life time at great age // Ann. Probab., 1974. Vol. 2. P. 792–804.
5. *Pickands J.* Statistical inference using extreme order statistics // Ann. Stat., 1975. Vol. 3. P. 119–131.
6. *Галамбош Я.* Асимптотическая теория экстремальных порядковых статистик / Пер. с англ. — М.: Наука, 1984. 314 с. (*Galambos J.* The asymptotic theory of extreme order statistics. — New York, NY, USA: Wiley, 1978. 352 p.)

Поступила в редакцию 28.11.19

ON THE DISTRIBUTION OF THE RATIO OF THE SUM OF SAMPLE ELEMENTS EXCEEDING A THRESHOLD TO THE TOTAL SUM OF SAMPLE ELEMENTS. II

V. Yu. Korolev^{1,2}

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

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

Abstract: The problem of description of the distribution of the ratio of the sum of sample elements exceeding a threshold to the total sum of sample elements is considered. Unlike other versions of this problem in which the number of summed extreme order statistics and the threshold are fixed, here the specified threshold can be exceeded by an unpredictable number of sample elements. The situation is considered where the threshold infinitely increases as the sample size grows. It is demonstrated that in this case, the distribution of the ratio mentioned above can be approximated by the compound Poisson distribution in which the compounding law is the generalized Pareto distribution.

Keywords: sum of independent random variables; random sum; binomial distribution; Poisson approximation; extreme order statistic; Balkema – De Haan – Pickands theorem; generalized Pareto distribution; compound Poisson distribution

DOI: 10.14357/19922264200405

Acknowledgments

The research was supported by the Russian Foundation for Basic Research (project 19-07-00914). The research was conducted in accordance with the program of the Moscow Center for Fundamental and Applied Mathematics.

References

1. Koch, R. 1998. *The 80/20 principle: The secret of achieving more with less*. London: Nicholas Brealey Publishing. 302 p.
2. Korolev, V. Yu. 2020. O raspredelenii otnosheniya summy elementov vyborki, prevoskhodyashchikh nekotoryy porog, k summe vseh elementov vyborki. I [On the distribution of the ratio of the sum of sample elements exceeding a threshold to the total sum of sample elements. I]. *Informatika i ee Primeneniya — Inform.Appl.* 14(3):26–34.
3. Barbour, A. D., and P. Hall. 1984. On the rate of Poisson convergence. *P. Camb. Philos. Soc.* 95:473–480.
4. Balkema, A., and L. de Haan. 1974. Residual life time at great age. *Ann. Probab.* 2:792–804.
5. Pickands, J. 1975. Statistical inference using extreme order statistics. *Ann. Stat.* 3:119–131.
6. Galambos, J. 1978. *The asymptotic theory of extreme order statistics*. New York, NY: Wiley. 352 p.

Received November 28, 2019

Contributor

Korolev Victor Yu. (b. 1954) — Doctor of Science in physics and mathematics, professor, Head of Department, Faculty of Computational Mathematics and Cybernetics, and principal scientist, Moscow Center for Fundamental and Applied Mathematics, Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow 119991, Russian Federation; leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vkorolev@cs.msu.ru

О МАРКОВСКИХ И РАЦИОНАЛЬНЫХ ПОТОКАХ СЛУЧАЙНЫХ СОБЫТИЙ. II*

В. А. Наумов¹, К. Е. Самуйлов²

Аннотация: Статья представляет собой вторую часть обзора, выполненного в рамках проекта РФФИ № 19-17-50126. Цель обзора — ознакомление заинтересованных читателей с основами теории марковских потоков событий для более подробного изучения и облегчения применения этих моделей на практике. В первой части приведены свойства общих марковских потоков событий и показана их связь с марковскими аддитивными процессами и процессами марковского восстановления. Во второй части обзора рассмотрены важные для приложений частные случаи таких потоков — подклассы марковских потоков событий, а именно: простые и групповые потоки однородных и неоднородных событий. Показано, как свойства марковских потоков событий связаны с мультипликативностью стационарных распределений марковских систем. Обсуждаются матрично-экспоненциальные распределения и рациональные потоки событий, расширяющие возможности марковских потоков для моделирования сложных систем, при этом сохраняющие удобство их анализа с помощью вычислительной техники.

Ключевые слова: марковские процессы; марковские аддитивные процессы; потоки без последствия; МС-потоки

DOI: 10.14357/19922264200406

1 Введение

Настоящий обзор, состоящий из двух частей, включает изложение основ теории марковских потоков и снабжен ссылками на большое число работ, посвященных марковским и рациональным потокам событий. Он начался с рассмотрения в первой части случайных величин фазового типа, определения марковских потоков общего вида и их связи с марковскими аддитивными процессами и процессами марковского восстановления. Во второй части обзора перейдем к важным для приложений подклассам марковских потоков однородных и неоднородных событий в разд. 2, а в завершение в разд. 3 обсудим матрично-экспоненциальные распределения и в разд. 4 рациональные потоки событий, которые расширяют возможности марковских потоков для моделирования сложных систем и при этом сохраняют удобство их анализа.

Как и в первой части обзора, далее в работе жирные строчные буквы обозначают векторы, а жирные прописные буквы обозначают матрицы. Кроме того, используются следующие обозначения:

$$\delta(i, j) = \begin{cases} 1, & \text{если } i = j; \\ 0 & \text{в противном случае;} \end{cases}$$

у вектора e_i i -я координата равна единице, а остальные равны нулю; $\mathbf{I} = [\delta(i, j)]$ — единичная матрица; \mathbf{u} — вектор-столбец из единиц; \mathcal{N}^K — множество неотрицательных целочисленных векторов длины K , $\mathcal{N}_0^K = \mathcal{N}^K \setminus \{0\}$. Для краткости вместо «наступило n_1 событий типа 1, n_2 событий типа 2, \dots , n_K событий типа K » будем писать «наступило \mathbf{n} событий», где $\mathbf{n} = (n_1, n_2, \dots, n_K)$.

2 Важные для приложений частные случаи марковских потоков событий

2.1 Простой марковский поток однородных событий

Рассмотрим некоторый поток случайных неоднородных событий и обозначим через $N_k(t)$ число событий типа k , наступивших за время t , $\mathbf{N}(t) = (N_1(t), N_2(t), \dots, N_K(t))$. Поток случайных событий называется марковским, если для некоторого случайного процесса $X(t)$ с конечным множеством состояний $\mathcal{X} = \{1, 2, \dots, L\}$ процесс

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-17-50126.

¹ Исследовательский институт инноваций, г. Хельсинки, Финляндия, valeriy.naumov@pfu.fi

² Российский университет дружбы народов; Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, samouylov-ke@rudn.ru

$\xi(t) = (X(t), \mathbf{N}(t))$ является марковским процессом, однородным во времени и по второй компоненте, т. е. если для любых $t, h > 0$ справедливы равенства

$$\begin{aligned} P(X(h+t) = j, \mathbf{N}(h+t) = \mathbf{k} + \mathbf{n} | X(h) = i, \\ \mathbf{N}(h) = \mathbf{k}) = p_{\mathbf{n}}(i, j, t), \quad \mathbf{k}, \mathbf{n} \in \mathcal{N}^K, \quad i, j \in \mathcal{X}. \end{aligned}$$

Матрицы вероятностей переходов $\mathbf{P}_{\mathbf{n}}(t) = [p_{\mathbf{n}}(i, j, t)]$ однозначно определяются матрицами интенсивностей переходов $\mathbf{A}_{\mathbf{n}} = [a_{\mathbf{n}}(i, j)]$, $\mathbf{n} \geq \mathbf{0}$, где

$$\begin{aligned} a_{\mathbf{0}}(i, j) &= \lim_{t \rightarrow 0} \frac{1}{t} (p_{\mathbf{0}}(i, j, t) - \delta(i, j)), \quad i, j \in \mathcal{X}; \\ a_{\mathbf{n}}(i, j) &= \lim_{t \rightarrow 0} \frac{1}{t} p_{\mathbf{n}}(i, j, t), \quad i, j \in \mathcal{X}, \quad \mathbf{n} \in \mathcal{N}_0^K, \end{aligned}$$

при этом фазовый процесс $X(t)$ является однородным марковским процессом с матрицей интенсивностей переходов $\mathbf{A} = \sum_{\mathbf{n} \in \mathcal{N}^K} \mathbf{A}_{\mathbf{n}}$.

В первой части обзора определен процесс марковского восстановления (X_l, σ_l, τ_l) , где $X_l = X(t_l)$ — состояния фазового процесса $X(t)$ марковского потока в моменты после наступления событий потока, $X(t) \in \mathcal{X} = \{1, 2, \dots, L\}$, $0 < t_1 < t_2 < \dots$ — моменты наступления событий, также называемые вызывающими моментами; $\tau_l = t_l - t_{l-1}$ — длины интервалов между моментами наступления событий; σ_l — вектор, $\sigma_l = (\sigma_{l,1}, \dots, \sigma_{l,K})$, в котором $\sigma_{l,k}$ есть размер группы событий типа k , наступивших в момент t_l , $l = 1, 2, \dots$. Матрицы $\mathbf{G}_{\mathbf{n}}(x) = [G_{\mathbf{n}}(i, j, x)]$, описывающие связанный с марковским потоком процесс марковского восстановления (X_l, σ_l, τ_l) , и их преобразования Лапласа–Стилтьеса имеют следующий вид:

$$\begin{aligned} \mathbf{G}_{\mathbf{n}}(x) &= \int_0^x \exp(z\mathbf{A}_0)\mathbf{A}_{\mathbf{n}} dz = \\ &= (\exp(x\mathbf{A}_0) - \mathbf{I})\mathbf{A}_0^{-1}\mathbf{A}_{\mathbf{n}}, \quad \mathbf{n} \in \mathcal{N}_0^K; \quad (1) \end{aligned}$$

$$\int_0^x e^{-\nu x} d\mathbf{G}_{\mathbf{n}}(x) = (\nu\mathbf{I} - \mathbf{A}_0)^{-1}\mathbf{A}_{\mathbf{n}}, \quad \mathbf{n} \in \mathcal{N}_0^K. \quad (2)$$

Используя матрицы $\mathbf{G}_{\mathbf{n}}(x)$, можно найти совместное распределение числа σ_l наступивших событий и длин τ_l интервалов между вызывающими моментами

$$\begin{aligned} F_{\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m}(x_1, x_2, \dots, x_m) &= \\ &= P(\sigma_l = \mathbf{k}_l, \tau_l < x_l, l = 1, 2, \dots, m) = \\ &= \alpha \mathbf{G}_{\mathbf{k}_1}(x_1) \mathbf{G}_{\mathbf{k}_2}(x_2) \cdots \mathbf{G}_{\mathbf{k}_m}(x_m) \mathbf{u}, \quad (3) \end{aligned}$$

а также плотность этого распределения

$$\begin{aligned} f_{\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m}(x_1, x_2, \dots, x_m) &= \\ &= \alpha \exp(x_1 \mathbf{A}_0) \mathbf{A}_{\mathbf{k}_1} \exp(x_2 \mathbf{A}_0) \mathbf{A}_{\mathbf{k}_2} \cdots \\ &\cdots \exp(x_m \mathbf{A}_0) \mathbf{A}_{\mathbf{k}_m} \mathbf{u}, \quad \mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_m \in \mathcal{N}_0^K, \\ &x_0, x_1, \dots, x_m > 0, \quad m = 1, 2, \dots \quad (4) \end{aligned}$$

где α — начальное распределение фазового процесса.

Простой марковский поток однородных событий — это марковский поток событий одного типа, причем в каждый вызывающий момент наступает ровно одно событие. Он характеризуется двумя матрицами интенсивностей переходов $\mathbf{S} = \mathbf{A}_0$ и $\mathbf{R} = \mathbf{A}_1$, а остальные матрицы \mathbf{A}_k , $k \geq 2$, для такого потока — нулевые. Первыми работами, посвященными простым марковским потокам однородных событий, стали [1–5]. Их применение к решению задач теории телеграфика рассматривается в [6, 7]. Поток вызывающих моментов любого марковского потока — это простой марковский поток, характеризуемый матрицами $\mathbf{S} = \mathbf{A} - \mathbf{R}$ и $\mathbf{R} = \sum_{\mathbf{n} \in \mathcal{N}_0^K} \mathbf{A}_{\mathbf{n}}$. К простым марковским потокам относятся также процессы восстановления фазового типа [8]. Для таких потоков ранг матрицы \mathbf{R} равен единице и она имеет вид $\mathbf{R} = \mathbf{s}\mathbf{q}$, где $\mathbf{s} = -\mathbf{S}\mathbf{u}$. Верно и обратное [7]. В англоязычной литературе простые марковские потоки называют Markovian arrival process и используют для их обозначения сокращение MAP или MArP.

Простой марковский поток однородных событий является полумарковским, поскольку последовательность (X_l, τ_l) , $l = 1, 2, \dots$, — процесс марковского восстановления. Из (1) и (2) вытекают следующие формулы для полумарковской матрицы $\mathbf{G}(x) = [G(i, j, x)]$ процесса (X_l, τ_l) марковского восстановления с элементами

$$G(i, j, x) = P(X_l = j, \tau_l < x | X_{l-1} = i)$$

и для ее преобразования Лапласа–Стилтьеса:

$$\left. \begin{aligned} \mathbf{G}(x) &= (\exp(x\mathbf{S}) - \mathbf{I})\mathbf{S}^{-1}\mathbf{R}; \\ \int_0^x e^{-\nu x} d\mathbf{G}(x) &= (\nu\mathbf{I} - \mathbf{S})^{-1}\mathbf{R}. \end{aligned} \right\} \quad (5)$$

Из (4) вытекает следующее выражение для плотности функции распределения длин интервалов τ_l между моментами наступления событий простого марковского потока однородных событий:

$$\begin{aligned} f(x_1, x_2, \dots, x_m) &= \\ &= \alpha \exp(x_1 \mathbf{S}) \mathbf{R} \exp(x_2 \mathbf{S}) \mathbf{R} \cdots \exp(x_m \mathbf{S}) \mathbf{R} \mathbf{u}, \\ &x_0, x_1, \dots, x_m > 0, \quad m = 1, 2, \dots \quad (6) \end{aligned}$$

Поскольку простой марковский поток является полумарковским, при анализе систем массового

обслуживания с такими поступающими потоками можно использовать результаты, полученные для систем с полумарковским входящим потоком, например [9–12].

В первом разделе обзора указано, что стационарные распределения $\mathbf{q} = [q(i)]$ и $\mathbf{q}_n = [q_n(i)]$, $\mathbf{n} \in \mathcal{N}_0^K$, вложенных цепей Маркова X_l и (X_l, σ_l) связаны со стационарным распределением \mathbf{p} фазового процесса $X(t)$ следующими равенствами:

$$\mathbf{q} = \frac{1}{\lambda} \mathbf{p} \mathbf{A}, \quad \mathbf{p} = -\lambda \mathbf{q} \mathbf{A}_0^{-1}, \quad \mathbf{q} = \sum_{\mathbf{n} \in \mathcal{N}_0^K} \mathbf{q}_n,$$

$$\mathbf{q}_n = \frac{1}{\lambda} \mathbf{p} \mathbf{A}_n, \quad \mathbf{n} \in \mathcal{N}_0^K.$$

Если вектор из единиц \mathbf{u} является правым собственным вектором каждой из матриц \mathbf{A}_n и выполняются равенства

$$\mathbf{A}_n \mathbf{u} = \lambda_n \mathbf{u}, \quad \mathbf{n} \in \mathcal{N}_0^K, \quad (7)$$

то из (3) следует, что при любом начальном распределении \mathbf{s} марковский поток будет стационарным потоком без последствия. Аналогично, если вектор стационарных вероятностей \mathbf{p} является левым собственным вектором матриц \mathbf{A}_n и выполняются равенства

$$\mathbf{p} \mathbf{A}_n = \lambda_n \mathbf{p}, \quad \mathbf{n} \in \mathcal{N}_0^K. \quad (8)$$

Условия (7) и (8), достаточные для того чтобы марковский поток был пуассоновским, для простого марковского потока приобретают вид $\mathbf{R} \mathbf{u} = \lambda \mathbf{u}$ и $\mathbf{p} \mathbf{R} = \lambda \mathbf{p}$ соответственно, где $\lambda = \mathbf{p} \mathbf{R} \mathbf{u}$ — интенсивность потока. Проверка необходимых и достаточных условий пуассоновости простого марковского потока более сложна и требует знания собственных векторов матрицы \mathbf{S} [13].

Считающий процесс $N(t)$ стационарной версии простого марковского потока является асимптотически нормальным с математическим ожиданием $M(t) = \lambda t$ и дисперсией

$$D(t) = (2\mathbf{d}_1 \mathbf{s} - \lambda) t + 2(\mathbf{d}_2 \mathbf{s} - \lambda) + o(1),$$

где векторы-столбцы \mathbf{d}_1 и \mathbf{d}_2 — единственные решения систем линейных уравнений [2]:

$$\mathbf{d}_1 \mathbf{A} = \mathbf{p}(\lambda \mathbf{I} - \mathbf{R}), \quad \mathbf{d}_1 \mathbf{u} = 1;$$

$$\mathbf{d}_2 \mathbf{A} = \mathbf{d}_1 - \mathbf{p}, \quad \mathbf{d}_2 \mathbf{u} = 1.$$

2.2 Простой марковский поток неоднородных событий

Простой марковский поток неоднородных событий — это марковский поток событий нескольких типов, в каждый вызывающий момент которого наступает ровно одно событие. Такой поток

характеризуется $K + 1$ матрицами интенсивностей переходов $\mathbf{S} = \mathbf{A}_0$ и $\mathbf{R}_k = \mathbf{A}_{\mathbf{e}_k}$, $k = 1, 2, \dots, K$, а остальные матрицы \mathbf{A}_n — нулевые. При этом поток событий одного типа, например типа i , является простым марковским потоком однородных событий, описываемым матрицами $\mathbf{S}_i = \mathbf{A} - \mathbf{A}_{\mathbf{e}_i}$ и \mathbf{R}_i . Первыми работами, посвященными простым марковским потокам неоднородных событий, считаются [14–16]. В англоязычной литературе такой поток называют Markovian Arrival Process with marked arrivals и используют для его обозначения сокращение ММАР. Из (5) вытекает следующее выражение для плотности совместного распределения $P(\omega_l = k_l, \tau_l < x_l, l = 1, 2, \dots, m)$ типов ω_l событий, наступивших в момент t_l , и длин τ_l интервалов между вызывающими моментами:

$$f_{k_1, k_2, \dots, k_m}(x_1, x_2, \dots, x_m) =$$

$$= \alpha \exp(x_1 \mathbf{S}) \mathbf{R}_{k_1} \exp(x_2 \mathbf{S}) \mathbf{R}_{k_2} \dots$$

$$\dots \exp(x_m \mathbf{S}) \mathbf{R}_{k_m} \mathbf{u}, \quad 1 \leq k_1, k_2, \dots, k_m \leq K,$$

$$x_0, x_1, \dots, x_m > 0, \quad m = 1, 2, \dots \quad (9)$$

2.3 Марковский поток групп однородных событий

Марковский поток групп однородных событий — это марковский поток событий одного типа, в каждый вызывающий момент которого может наступить несколько событий. Такие марковские потоки впервые исследовались в [8, 17, 18], а их описание с помощью матриц \mathbf{A}_n впервые появилось в [19]. В англоязычной литературе такой поток сейчас называют batch Markovian arrival process и используют для его обозначения сокращение ВМАР. В [20] получены формулы и асимптотики для первых двух моментов считающего процесса $N(t)$, а в [21] — для старших моментов $N(t)$.

3 Матрично-экспоненциальные распределения

Функция распределения $F(t)$ неотрицательной случайной величины называется матрично-экспоненциальной, если $F(0) < 1$ и она представима в виде

$$F(t) = 1 - \mathbf{q} \exp(t \mathbf{S}) \mathbf{u} \quad (10)$$

с некоторым вектором \mathbf{q} и матрицей \mathbf{S} , имеющей собственные числа лишь с отрицательными действительными частями. Для того чтобы функция распределения $F(t)$ неотрицательной случайной величины была матрично-экспоненциальной,

необходимо и достаточно, чтобы она имела рациональное преобразование Лапласа–Стилтьеса $\tilde{F}(\nu)$. Минимальный порядок матрицы \mathbf{S} в матрично-экспоненциальном представлении (10) равен числу полюсов функции $\tilde{F}(\nu)$ с учетом их кратности. Представление с матрицей \mathbf{S} минимального порядка называется минимальным.

В некоторых работах по матрично-экспоненциальным функциям распределения [22–24], а также в книгах [25, 26], чтобы подчеркнуть аналогию с экспоненциальными функциями распределения, вместо (10) использовалось представление $F(t) = 1 - \mathbf{q} \exp(-t\mathbf{B})\mathbf{u}$ со знаком минус перед t и матрицей \mathbf{B} , имеющей собственные числа с положительными действительными частями. В настоящее время используются только представления вида (10). Иногда допускается, что вектор \mathbf{u} в (10) может быть любым, а не состоящим из единиц, как в рассматриваемом случае. Однако в [24, 27] было показано, что всегда можно подобрать матрично-экспоненциальное представление с $\mathbf{u} = (1, 1, \dots, 1)$.

Идея матрично-экспоненциальных функций распределения восходит к работе [28], в которой показано, что рациональные преобразования Лапласа–Стилтьеса неотрицательных функций распределения представимы в виде:

$$\tilde{F}(s) = p_0 + \sum_{l=1}^L q_0 \cdots q_{l-1} p_l \prod_{i=1}^l \frac{\lambda_i}{\lambda_i + s},$$

где $p_i + q_i = 1$, $i = 1, \dots, L$, $p_L = 1$, и $-\lambda_i$, $i = 1, \dots, L$, — полюсы $\tilde{F}(s)$. Такое представление можно записать в матрично-экспоненциальном виде (10), полагая

$$\mathbf{q} = (1, 0, \dots, 0);$$

$$\mathbf{S} = \begin{bmatrix} -\lambda_1 & q_1 \lambda_1 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & q_2 \lambda_2 & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & -\lambda_{L-1} & q_{L-1} \lambda_{L-1} \\ 0 & \cdots & 0 & 0 & -\lambda_L \end{bmatrix},$$

при этом элементы матрицы \mathbf{S} могут быть комплексными. В [22] показано, что вектор \mathbf{q} и матрица \mathbf{S} в матрично-экспоненциальном представлении (10) всегда могут быть выбраны действительными.

Из (10) вытекают формулы для начальных моментов

$$\int_0^{\infty} t^n dF(t) = n! \mathbf{q}(-\mathbf{S})^{-n} \mathbf{u}, \quad n = 1, 2, \dots$$

и для преобразования Лапласа–Стилтьеса функции распределения $F(t)$

$$\begin{aligned} \tilde{F}(\nu) &= \int_0^{\infty} e^{-\nu t} dF(t) = \\ &= 1 - \mathbf{q}\mathbf{u} + \mathbf{q}(\nu\mathbf{I} - \mathbf{S})^{-1}\mathbf{s} = 1 - \nu\mathbf{q}(\nu\mathbf{I} - \mathbf{S})^{-1}\mathbf{u}, \end{aligned}$$

где $\mathbf{s} = -\mathbf{S}\mathbf{u}$. Кроме того, матрично-экспоненциальные функции распределения обладают следующими свойствами [24].

1. Пусть $F_i(t) = 1 - \mathbf{q}_i \exp(t\mathbf{S}_i)\mathbf{u}$, $i = 1, 2$, — матрично-экспоненциальные функции распределения и $p_1 + p_2 = 1$. Тогда

$$\begin{aligned} p_1 F_1(t) + p_2 F_2(t) &= \\ &= 1 - (p_1 \mathbf{q}_1, p_2 \mathbf{q}_2) \exp\left(t \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix}\right) \mathbf{u}; \\ (F_1 * F_2)(t) &= \\ &= 1 - (\mathbf{q}_1, F_1(0)\mathbf{q}_2) \exp\left(t \begin{bmatrix} \mathbf{S}_1 & -\mathbf{S}_1 \mathbf{u} \mathbf{q}_2 \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix}\right) \mathbf{u}. \end{aligned}$$

2. Пусть τ и γ — независимые неотрицательные случайные величины с функциями распределения $F(t)$ и $G(t)$ соответственно, причем $F(t)$ имеет матрично-экспоненциальное представление (10). Тогда функция распределения $H(t)$ случайной величины $(\tau - \gamma)^+$ имеет матрично-экспоненциальное представление

$$H(t) = 1 - \mathbf{q}\mathbf{U} \exp(t\mathbf{S})\mathbf{u},$$

где

$$\mathbf{U} = \int_0^{\infty} e^{t\mathbf{S}} dG(t). \quad (11)$$

3. Пусть $F(t)$ имеет матрично-экспоненциальное представление (10), а у квадратной матрицы \mathbf{V} все собственные числа имеют неотрицательные вещественные части. Тогда

$$\begin{aligned} \int_0^{\infty} e^{-t\mathbf{V}} dF(t) &= (1 - \mathbf{q}\mathbf{u})\mathbf{I} + (\mathbf{q} \otimes \mathbf{I})\Psi(\mathbf{S}\mathbf{u} \otimes \mathbf{I}) = \\ &= \mathbf{I} - (\mathbf{q} \otimes \mathbf{I})\Psi(\mathbf{u} \otimes \mathbf{V}), \end{aligned}$$

где $\Psi = (\mathbf{I} \otimes \mathbf{V} - \mathbf{S} \otimes \mathbf{I})^{-1}$.

Последнее свойство можно использовать для вычисления матриц \mathbf{U} в (11) для матрично-экспоненциальных функций распределения $G(t)$.

Ясно, что функции распределения фазового типа являются матрично-экспоненциальными. Однако их матрично-экспоненциальные представления

$$F(t) - 1 - \mathbf{q} \exp(t\mathbf{S})\mathbf{u}, \quad F(0) = 1 - \mathbf{q}\mathbf{u},$$

$$\frac{d}{dt} F(t) = \mathbf{q} \exp(t\mathbf{S})\mathbf{s}, \quad t > 0,$$

с ограничениями

$$\left. \begin{aligned} 0 < \sum_{j \in \mathcal{X}} q(j) \leq 1, \quad q(i) \geq 0, \quad i \in \mathcal{X}; \\ \sum_{j \in \mathcal{X}} s(i, j) \leq 0, \quad s(i, j) \geq 0, \quad i \neq j, \quad i, j \in \mathcal{X}, \end{aligned} \right\} (12)$$

где $\mathbf{S} = [s(i, j)]$, следует отличать от матрично-экспоненциальных представлений этих же функций, но без ограничений (12). Порядок матрично-экспоненциального представления, удовлетворяющего ограничениям (12), будем называть числом этапов этого представления, а порядок матрично-экспоненциального представления, не удовлетворяющего этим ограничениям, следуя [28], будем называть числом *фиктивных* этапов. Необходимые и достаточные условия того, чтобы для функции распределения с рациональным преобразованием Лапласа–Стилтьеса существовало представление, удовлетворяющее ограничениям (12), получены в [29]. Для этого надо, чтобы (а) функция распределения имела непрерывную положительную плотность на правой полуоси и (б) ее преобразование Лапласа–Стилтьеса имело единственный полюс с максимальной вещественной частью.

4 Рациональные потоки событий

Рациональный поток групп неоднородных событий (t_l, σ_l) , $l = 1, 2, \dots$, можно определить как поток, для которого совместное распределение числа σ_l наступивших событий и длин τ_l интервалов между моментами t_l наступления событий дается формулами (1) и (3) с матрицами \mathbf{A}_n , $\mathbf{n} \in \mathcal{N}^K$, обладающими следующими свойствами:

- (1) действительные части собственных чисел матрицы \mathbf{A}_0 отрицательны;
- (2) действительные части собственных чисел матрицы $\mathbf{A} = \sum_{n \in \mathcal{N}^K} \mathbf{A}_n$ неположительны;
- (3) $\mathbf{A}\mathbf{u} = \mathbf{0}$.

Для стационарных версий рациональных потоков дополнительно требуется, чтобы начальный вектор α совпадал с решением \mathbf{p} системы линейных уравнений $\mathbf{p}\mathbf{A} = \mathbf{0}$, $\mathbf{p}\mathbf{u} = 1$.

Простой рациональный поток однородных событий, также называемый матрично-экспоненциальным потоком [30], — это поток событий одного типа, в каждый вызывающий момент которого наступает ровно одно событие и для которого плотность совместного распределения длин τ_l интервалов между моментами наступления событий дается формулой (6) с матрицами \mathbf{S} и \mathbf{R} , обладающими следующими свойствами [31]:

- (а) вещественные части собственных чисел матрицы \mathbf{S} отрицательны;
- (б) вещественные части собственных чисел матрицы $\mathbf{S} + \mathbf{R}$ неположительны;
- (в) $(\mathbf{S} + \mathbf{R})\mathbf{u} = \mathbf{0}$.

Примерами простых рациональных потоков однородных событий могут служить полумарковские потоки [22] и процессы восстановления [27] с матрично-экспоненциальными функциями распределения длин интервалов между наступлениями событий.

Рациональный поток неоднородных событий — это поток событий нескольких типов, в каждый вызывающий момент которого наступает ровно одно событие. Для такого потока совместное распределение типов наступивших событий ω_l и длин τ_l интервалов между моментами наступления событий дается формулой (9), а на матрицы \mathbf{S} и $\mathbf{R} = \mathbf{R}_1 + \mathbf{R}_2 + \dots + \mathbf{R}_K$ накладываются перечисленные выше ограничения (а)–(в) [32].

5 Заключение

Метод этапов Эрланга [33] более 100 лет применяется при анализе стохастических систем. К его широкому распространению привело открытие матрично-экспоненциального представления для функций распределения фазового типа [34] и моделей марковских потоков событий [1, 17]. Эти модели хорошо подходят для анализа стохастических систем с помощью вычислительной техники, приспособленной к обработке векторов и матриц, что привело к развитию специальных матричных методов анализа стохастических систем.

Метод фиктивных этапов, предложенный в [28], позволил распространить метод Эрланга на любые распределения с рациональным преобразованием Лапласа–Стилтьеса. Использование матрично-экспоненциальных представлений для функций распределения [22, 23, 25] и потоков случайных событий [31] с произвольными рациональными преобразованиями Лапласа–Стилтьеса упрощает применение метода фиктивных этапов. Формальное применение метода фиктивных этапов приводит

к решению, в котором вероятности, соответствующие фиктивным этапам, могут оказаться отрицательными, большими единицы или даже комплексными. Однако вероятности, соответствующие нефиктивным состояниям, будут неотрицательными числами, не превосходящими единицы. Существуют различные интерпретации понятий отрицательных вероятностей и интенсивностей переходов [35–38]. Более детально ознакомиться с марковским и рациональным потоками событий, а также с матричными методами анализа стохастических систем можно в обзорах [39–44] и монографиях [18, 25, 26, 45–57].

Литература

1. Наумов В. А. О независимой работе подсистем сложной системы // Тр. III Всесоюзной школы-семинара по теории массового обслуживания. — М.: МГУ, 1976. № 2. С. 169–177.
2. Бочаров П. П., Наумов В. А. Анализ гиперэкспоненциальной двухфазной системы с ограниченным накопителем // Информационные сети и их структура. — М.: Наука, 1976. С. 168–180.
3. Наумов В. А. Об обслуженной и избыточной нагрузках полнодоступного пучка с ограниченной очередью // Численные методы решения задач математической физики и теории систем. — М.: УДН, 1977. С. 51–55.
4. Наумов В. А. Исследование некоторых многофазных систем массового обслуживания: Дис. . . . канд. физ.-мат. наук. — М.: УДН, 1978. 98 с.
5. Lucantoni D. M., Meier-Hellstern K., Neuts M. F. A single-server queue with server vacations and a class of non-renewal arrival processes // Adv. Appl. Probab., 1990. Vol. 22. Iss. 3. P. 676–705.
6. Башарин Г. П., Кокотушкин В. А., Наумов В. А. О методе эквивалентных замен расчета фрагментов сетей связи для ЦВМ // Изв. АН СССР. Техническая кибернетика, 1979. № 6. С. 92–99.
7. Basharin G., Naumov V. Simple matrix description of peaked and smooth traffic and its applications // 3rd ITC Specialist Seminar on Fundamentals of Teletraffic Theory. — М.: VINITI, 1984. P. 38–44.
8. Neuts M. F. Renewal processes of phase type // Nav. Res. Logist. Q., 1978. Vol. 25. Iss. 3. P. 445–454.
9. Cinlar E. Queues with semi-Markovian arrivals // J. Appl. Probab., 1967. Vol. 4. Iss. 2. P. 365–379.
10. Franken P. Erlangsche Formeln für semimarkowschen Eingang // Elektronische Informationsverarbeitung Kybernetik, 1968. Vol. 4. Iss. 3. P. 197–204.
11. Franken P., Kerstan J. Bedienungssysteme mit unendlich vielen Bedienungsapparaten // Operationsforschung Mathematische Statistik. — Berlin: Akademie-Verlag, 1968. Vol. I. P. 67–76.
12. Neuts M. F., Chen S.-Z. The infinite server queue with semi-Markovian arrivals and negative exponential services // J. Appl. Probab., 1972. Vol. 9. Iss. 1. P. 178–184.
13. Bean N. G., Green D. A., Taylor P. G. When is a MAP poisson? // 2nd Australia–Japan Workshop on Stochastic Models in Engineering, Technology and Management Proceedings / Eds. J. Wilson, D. N. P. Murthy, S. Osaki. — Brisbane: Technology Management Center, University of Queensland, 1996. P. 34–43.
14. Наумов В. А. Матричный аналог формулы Эрланга // Модели распределения информации и методы их анализа. — М.: ВИНТИ, 1988. С. 39–43.
15. He Q.-M. Queues with marked customers // Adv. Appl. Probab., 1996. Vol. 28. Iss. 2. P. 567–587.
16. He Q.-M., Neuts M. F. Markov chains with marked transitions // Stoch. Proc. Appl., 1998. Vol. 74. P. 37–52.
17. Neuts M. F. A versatile Markovian point process. — Newark, DE: University of Delaware, Department of Statistics and Computer Science, 1977. Technical Report 77/13. 29 p.
18. Neuts M. F. Structured stochastic matrices of $M/G/1$ type and their applications. — New York, NY, USA: Marcel Dekker, 1989. 512 p.
19. Lucantoni D. M. New results on the single server queue with a batch Markovian arrival process // Communications Statistics. Stochastic Models, 1991. Vol. 7. Iss. 1. P. 1–46.
20. Narayana S., Neuts M. F. The first two moment matrices of the counts for the Markovian arrival processes // Communications Statistics. Stochastic Models, 1992. Vol. 8. Iss. 3. P. 459–477.
21. Nielsen B. F., Nilsson L. A. F., Thygesen U. H., Beyer J. E. Higher order moments and conditional asymptotics of the batch Markovian arrival process // Stoch. Models, 2007. Vol. 23. Iss. 1. P. 1–26.
22. Бочаров П. П., Наумов В. А. О некоторых системах массового обслуживания конечной емкости // Проблемы передачи информации, 1977. Т. 13. № 4. С. 96–104.
23. Наумов В. А. Об однолинейной системе с ограниченным накопителем и заявками нескольких видов // Модели систем распределения информации и их анализ. — М.: Наука, 1982. С. 77–82.
24. Наумов В. А. О функциях распределения с рациональным преобразованием Лапласа–Стилтьеса // Анализ информационно-вычислительных систем. — М.: УДН, 1986. С. 47–56.
25. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. 528 с.
26. Bocharov P. P., D'Apice C., Pechinkin A. V., Salerno S. Queueing theory. — Utrecht–Boston: VSP, 2004. 446 p.
27. Asmussen S., Bladt M. Renewal theory and queueing algorithms for matrix-exponential distributions // Matrix-analytic methods in stochastic models / Eds. A. S. Alfa, S. Chakravarty. — New York, NY, USA: Marcel Dekker, 1996. P. 313–341.

28. Cox D. R. A use of complex probabilities in the theory of stochastic processes // *Math. Proc. Cambridge*, 1955. Vol. 51. Iss. 2. P. 313–319.
29. O’Cinneide C. A. Characterization of phase-type distributions // *Communications Statistics. Stochastic Models*, 1990. Vol. 6. Iss. 1. P. 1–57.
30. Bodrog L., Horváth A., Telek M. On the properties of moments of matrix exponential distributions and matrix exponential processes // *Dagstuhl Seminar Proceedings*, 2008. Vol. 07461. Paper 1394.
31. Asmussen S., Bladt M. Point processes with finite-dimensional conditional probabilities // *Stoch. Proc. Appl.*, 1999. Vol. 82. Iss. 1. P. 127–142.
32. Horvath G., Telek M. Acceptance-rejection methods for generating random variables from matrix exponential distribution and rational arrival processes // *Matrix-analytic methods in stochastic models* / Eds. G. Latouche, V. Ramaswami, J. Sethuraman, *et al.* — New York, NY, USA: Springer, 2012. P. 123–144.
33. Erlang A. K. Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler // *Elektroteknikereren*, 1917. Iss. 13. P. 5–13.
34. Neuts M. F. Probability distribution of phase type // *Liber Amicorum Professor Emeritus H. Florin.* — Ottignies-Louvain-la-Neuve, Belgium: University of Louvain, Department of Mathematics, 1975. P. 173–206.
35. Bartlett M. S. Negative probability // *Math. Proc. Cambridge*, 1945. Vol. 41. Iss. 1. P. 71–73.
36. Cox D. R. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables // *Math. Proc. Cambridge*, 1955. Vol. 51. Iss. 3. P. 433–441.
37. Bladt M., Neuts M. F. Matrix-exponential distributions: Calculus and interpretations via flows // *Stoch. Models*, 2003. Vol. 19. Iss. 1. P. 113–124.
38. Khrennikov A. Interpretations of probability. — 2nd ed. — Berlin: Walter de Gruyter, 2009. 237 p.
39. Наумов В. А. Марковские модели потоков требований // *Системы массового обслуживания и информатика.* — М.: УДН, 1987. С. 67–73.
40. Asmussen S. Matrix-analytic models and their analysis // *Scand. J. Stat.*, 2000. Vol. 27. Iss. 2. P. 193–226.
41. Bladt M. A review on phase-type distributions and their use in risk theory // *ASTIN Bull.*, 2005. Vol. 35. No. 1. P. 145–161.
42. Artalejo J. R., Gómez-Corral A. Markovian arrivals in stochastic modelling: A survey and some new results // *SORT — Stat. Oper. Res. T.*, 2010. Vol. 34. Iss. 2. P. 101–144.
43. Вишневский В. М., Дудин А. Н. Системы массового обслуживания с коррелированными входными потоками и их применение для моделирования телекоммуникационных сетей // *Автоматика и телемеханика*, 2017. № 8. С. 3–59.
44. Basharin G., Naumov V., Samouylov K. On Markovian modelling of arrival processes // *Stat. Pap.*, 2018. Vol. 59. Iss. 4. P. 1533–1540.
45. Neuts M. F. Matrix-geometric solutions in stochastic models: An algorithmic approach. — Baltimore, MA, USA: The John Hopkins University Press, 1981. 332 p.
46. Latouche G., Ramaswami V. Introduction to matrix analytic methods in stochastic modeling. — Philadelphia, PA, USA: ASA & SIAM, 1999. 334 p.
47. Asmussen S. Applied probability and queues. — New York, NY, USA: Springer, 2003. 438 p.
48. Breuer L., Baum D. An introduction to queueing theory and matrix-analytic methods. — Dordrecht: Springer, 2005. 272 p.
49. Bini D. A., Latouche G., Meini B. Numerical methods for structured Markov chains. — New York, NY, USA: Oxford University Press, 2005. 336 p.
50. Asmussen S., O’Cinneide C. A. Matrix-exponential distributions // *Encyclopedia of statistical sciences* / Eds. S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic, N. L. Johnson. — Hoboken, NJ, USA: John Wiley & Sons, 2006. Vol. 3. P. 1–5. doi: 10.1002/0471667196.ess1092.
51. Li Q.-L. Constructive computation in stochastic models with applications. — Berlin: Springer-Verlag, 2009. 650 p.
52. Lipsky L. Queueing theory: A linear algebraic approach. — 2nd ed. — New York, NY, USA: Springer, 2009. 548 p.
53. Alfa A. S. Queueing theory for telecommunications. — New York, NY, USA: Springer, 2010. 238 p.
54. He Q.-M. Fundamentals of matrix-analytic methods. — New York, NY, USA: Springer, 2014. 349 p.
55. Buchholz P., Kriege J., Felko I. Input modeling with phase-type distributions and Markov models. Theory and applications. — New York, NY, USA: Springer, 2014. 127 p.
56. Наумов В. А., Самуйлов В. А., Гайдамака Ю. В. Мультипликативные решения конечных цепей Маркова. — М.: РУДН, 2015. 159 с.
57. Bladt M., Nielsen B. F. Matrix-exponential distributions in applied probability. — Boston, MA, USA: Springer, 2017. 736 p.

Поступила в редакцию 02.07.20

ON MARKOVIAN AND RATIONAL ARRIVAL PROCESSES. II

V. A. Naumov¹ and K. E. Samouylov^{2,3}

¹Service Innovation Research Institute, 8A Annankatu, Helsinki 00120, Finland

²Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

³Institute 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 is the second part of the review carried out within the framework of the RFBR project No. 19-17-50126. The purpose of this review is to get the interested readers familiar with the basics of the theory of Markovian arrival processes to facilitate the application of these models in practice and, if necessary, to study them in detail. In the first part of the review, the properties of the general Markovian arrival processes are presented and their relationship with Markov additive processes and Markov renewal processes is shown. In the second part of the review, the important for applications subclasses of Markovian arrival processes, i. e., simple and batch arrival processes of homogeneous and heterogeneous arrivals, are considered. It is shown how the properties of Markovian arrival processes are associated with the product form of stationary distributions of Markov systems. In conclusion, matrix-exponential distributions and rational arrival processes are discussed that expand the capabilities of Markovian arrival processes for modeling complex systems, while preserving the convenience of analyzing them using computations.

Keywords: Markov chain; Markovian arrival process; Markov additive process; MAP; MARP

DOI: 10.14357/19922264200406

Acknowledgments

The reported study was funded by RFBR, project No. 19-17-50126.

References

1. Naumov, V. A. 1976. O nezavisimoy rabote podsystem slozhnoy sistemy [About independent operation of sub-systems of a complex system]. *Tr. III Vsesoyuznoy shkoly-seminara po teorii massovogo obsluzhivaniya* [3th All-Russian School-Seminar of Queuing Theory Proceedings]. Moscow: Nauka. 2:169–177.
2. Bocharov, P. P., and V. A. Naumov. 1976. Analiz gipereksponentsial'noy dvukhfaznoy sistemy s ogranichennym nakopitelem [Analysis of a hyperexponential two-phase system with a limited storage]. *Informatsionnye seti i ikh struktura* [Information networks and their structure]. Moscow: Nauka. 168–180.
3. Naumov, V. A. 1977. Ob obsluzhennoy i izbytochnoy nagruzkakh polnodostupnogo puchka s ogranichennoy ochered'yu [About serviced and excessive loads of a fully accessible bundle with a limited queue]. *Chislennyye metody resheniya zadach matematicheskoy fiziki i teorii system* [Numerical methods for solving problems of mathematical physics and systems theory]. Moscow: UDN. 51–55.
4. Naumov, V. A. 1978. Issledovanie nekotorykh mnogofaznykh sistem massovogo obsluzhivaniya [Research of some multiphase queuing systems]. Moscow: UDN. PhD Thesis. 98 p.
5. Lucantoni, D. M., K. Meier-Hellstern, and M. F. Neuts. 1990. A single-server queue with server vacations and a class of non-renewal arrival processes. *Adv. Appl. Probab.* 22(3):676–705.
6. Basharin, G. P., V. A. Kokotushkin, and V. A. Naumov. 1979. O metode ekvivalentnykh zamen rascheta fragmentov setey svyazi dlya TsVM [On the method of equivalent substitutions for calculating fragments of communication networks for a central computer]. *Engineering Cybernetics* 6:92–99.
7. Basharin, G. P., and V. A. Naumov. 1984. Simple matrix description of peaked and smooth traffic and its applications. *3rd ITC Specialist Seminar on Fundamentals of Teletraffic Theory*. Moscow: VINITI. 38–44.
8. Neuts, M. F. 1978. Renewal processes of phase type. *Nav. Res. Logist. Q.* 25(3):445–454.
9. Cinlar, E. 1967. Queues with semi-Markovian arrivals. *J. Appl. Probab.* 4(2):365–379.
10. Franken, P. 1968. Erlangsche Formeln für semi-markowschen Eingang. *Elektronische Informationsverarbeitung Kybernetik* 4(3):197–204.
11. Franken, P., and J. Kerstan. 1968. Bedienungssysteme mit unendlich vielen Bedienungsapparaten. *Operationsforschung Mathematische Statistik* 1:67–76.
12. Neuts, M. F., and S.-Z. Chen. 1972. The infinite server queue with semi-Markovian arrivals and negative exponential services. *J. Appl. Probab.* 9(1):178–184.
13. Bean, N. G., D. A. Green, and P. G. Taylor. 1996. When is a MAP poisson? *2nd Australia–Japan Workshop on*

- Stochastic Models in Engineering, Technology and Management Proceedings*. Eds. J. Wilson, D. N. P. Murthy, and S. Osaki. Brisbane: Technology Management Center, University of Queensland. 34–43.
14. Naumov, V. A. 1988. Matrichnyy analog formuly Erlanga [The matrix analogue of a formula of Erlang]. *Modeli raspredeleniya informatsii i metody ikh analiza* [Information distribution models and methods for their analysis]. Moscow: VINITI. 39–43.
 15. He, Q.-M. 1996. Queues with marked customers. *Adv. Appl. Probab.* 28(2):567–587.
 16. He, Q.-M., and M. F. Neuts. 1998. Markov chains with marked transitions. *Stoch. Proc. Appl.* 74:37–52.
 17. Neuts, M. F. 1977. A versatile Markovian point process. Newark, DE: University of Delaware, Department of Statistics and Computer Science. Technical Report 77/13. 29 p.
 18. Neuts, M. F. 1989. *Structured stochastic matrices of M/G/1 type and their applications*. New York, NY: Marcel Dekker. 512 p.
 19. Lucantoni, D. M. 1991. New results on the single server queue with a batch Markovian arrival process. *Communications Statistics. Stochastic Models* 7(1):1–46.
 20. Narayana, S., and M. F. Neuts. 1992. The first two moment matrices of the counts for the Markovian arrival processes. *Communications Statistics. Stochastic Models* 8(3):459–477.
 21. Nielsen, B. F., L. A. F. Nilsson, U. H. Thygesen, and J. E. Beyer. 2007. Higher order moments and conditional asymptotics of the batch Markovian arrival process. *Stoch. Models* 23(1):1–26.
 22. Bocharov, P. P., and V. A. Naumov. 1977. O nekotorykh sistemakh massovogo obsluzhivaniya konechnoy emkosti [On some queueing systems of finite capacity]. *Problemy peredachi informatsii* [Problems of Information Transmission] 13(4):96–104.
 23. Naumov, V. A. 1982. Ob odnolineynoy sisteme s ogranichennym nakopitelem i zayavkami neskol'kikh vidov [About a single-line system with limited storage and multiple types of requests]. *Modeli sistem raspredeleniya informatsii i ikh analiz* [Models of information distribution systems and methods for their analysis]. Moscow: Nauka. 77–82.
 24. Naumov, V. A. 1986. O funktsiyakh raspredeleniya s ratsional'nym preobrazovaniem Laplasy–Stilt'esa [On distribution functions with rational Laplace–Stieltjes transformation]. *Analiz informatsionno-vychislitel'nykh sistem* [Analysis of information and computing systems]. Moscow: UDN. 47–56.
 25. Bocharov, P. P., and A. V. Pechinkin. 1995. *Teoriya massovogo obsluzhivaniya* [Queueing theory]. Moscow: RUDN. 528 p.
 26. Bocharov, P. P., C. D'Apice, A. V. Pechinkin, and S. Salerno. 2004. *Queueing theory*. Utrecht–Boston: VSP. 446 p.
 27. Asmussen, S., and M. Bladt. 1996. Renewal theory and queueing algorithms for matrix-exponential distributions. *Matrix-analytic methods in stochastic models*. Eds. A. S. Alfa and S. Chakravarty. New York, NY: Marcel Dekker. 313–341.
 28. Cox, D. R. 1955. A use of complex probabilities in the theory of stochastic processes. *Math. Proc. Cambridge* 51(2):313–319.
 29. O'Conneide, C. A. 1990. Characterization of phase-type distributions. *Communications Statistics. Stochastic Models* 6(1):1–57.
 30. Bodrog, L., A. Horváth, and M. Telek. 2008. On the properties of moments of matrix exponential distributions and matrix exponential processes. *Dagstuhl Seminar Proceedings* 07461:1394.
 31. Asmussen, S., and M. Bladt. 1999. Point processes with finite-dimensional conditional probabilities. *Stoch. Proc. Appl.* 82(1):127–142.
 32. Horvath, G., and M. Telek. 2012. Acceptance-rejection methods for generating random variables from matrix exponential distribution and rational arrival processes. *Matrix-analytic methods in stochastic models*. Eds. G. Latouche, V. Ramaswami, J. Sethuraman, et al. New York, NY: Springer. 123–144.
 33. Erlang, A. K. 1917. Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler. *Elektroteknikeren* 13:5–13.
 34. Neuts, M. F. 1975. Probability distribution of phase type. *Liber Amicorum Professor Emeritus H. Florin*. Ottignies-Louvain-la-Neuve, Belgium: University of Louvain, Department of Mathematics. 173–206.
 35. Bartlett, M. S. 1945. Negative probability. *Math. Proc. Cambridge* 41(1):71–73.
 36. Cox, D. R. 1955. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Math. Proc. Cambridge* 51(3):433–441.
 37. Bladt, M., and M. F. Neuts. 2003. Matrix-exponential distributions: Calculus and interpretations via flows. *Stoch. Models* 19(1):113–124.
 38. Khrennikov, A. 2009. *Interpretations of probability*. 2nd ed. Berlin: Walter de Gruyter. 237 p.
 39. Naumov, V. A. 1987. Markovskie modeli potokov trebovaniy [Markov models of demand flows]. *Sistemy massovogo obsluzhivaniya i informatika* [Queueing systems and computer science]. Moscow: UDN. 67–73.
 40. Asmussen, S. 2000. Matrix-analytic models and their analysis. *Scand. J. Stat.* 27(2):193–226.
 41. Bladt, M. 2005. A review on phase-type distributions and their use in risk theory. *ASTIN Bull.* 35(1):145–161.
 42. Artalejo, J. R., and A. Gómez-Corral. 2010. Markovian arrivals in stochastic modelling: A survey and some new results. *SORT — Stat. Oper. Res. T.* 34(2):101–144.
 43. Vishnevskiy, V. M., and A. N. Dudin. 2017. Queueing systems with correlated arrival flows and their applications to modeling telecommunication networks. *Automat. Rem. Contr.* 78(8):1361–1403.
 44. Basharin, G., V. Naumov, and K. Samouylov. 2018. On Markovian modelling of arrival processes. *Stat. Pap.* 59(4):1533–1540.

45. Neuts, M. F. 1981. *Matrix-geometric solutions in stochastic models: An algorithmic approach*. Baltimore, MA: The John Hopkins University Press. 332 p.
46. Latouche, G., and V. Ramaswami. 1999. *Introduction to matrix analytic methods in stochastic modeling*. Philadelphia, PA: ASA & SIAM. 334 p.
47. Asmussen, S. 2003. *Applied probability and queues*. New York, NY: Springer. 438 p.
48. Breuer, L., and D. Baum. 2005. *An introduction to queueing theory and matrix-analytic methods*. Dordrecht: Springer. 272 p.
49. Bini, D. A., G. Latouche, and B. Meini. 2005. *Numerical methods for structured Markov chains*. New York, NY: Oxford University Press. 336 p.
50. Asmussen, S., and C.A. O’Cinneide. 2006. Matrix-exponential distributions. *Encyclopedia of statistical sciences*. Eds. S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic, and N. L. Johnson. Hoboken, NJ: John Wiley & Sons. 3:1–5. doi: 10.1002/0471667196.ess1092.pub2.
51. Li, Q.-L. 2009. *Constructive computation in stochastic models with applications*. Berlin: Springer-Verlag. 650 p.
52. Lipsky, L. 2009. *Queueing theory: A linear algebraic approach*. 2nd ed. New York, NY: Springer. 548 p.
53. Alfa, A. S. 2010. *Queueing theory for telecommunications*. New York, NY: Springer. 238 p.
54. He, Q.-M. 2014. *Fundamentals of matrix-analytic methods*. New York, NY: Springer. 349 p.
55. Buchholz, P., J. Krieger, and I. Felko. 2014. *Input modeling with phase-type distributions and Markov models. Theory and applications*. New York, NY: Springer. 127 p.
56. Naumov, V.A., K. E. Samouylov, and Yu. V. Gaidamaka. 2015. *Mul’tiplikativnye resheniya konechnykh tsepey Markova* [Multiplicative solutions of finite Markov chains]. Moscow: RUDN. 159 p.
57. Bladt, M., and B. F. Nielsen. 2017. *Matrix-exponential distributions in applied probability*. Boston, MA: Springer. 736 p.

Received July 2, 2020

Contributors

Naumov Valeriy A. (b. 1950) — Candidate of Science (PhD) in physics and mathematics, scientific director, Service Innovation Research Institute, 8A Annankatu, Helsinki 00120, Finland; valeriy.naumov@pfu.fi

Samouylov Konstantin E. (b. 1955) — Doctor of Science in technology, professor, Head of Department, Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, 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; samuylov-ke@rudn.university

МЕТОДЫ ДЕТЕРМИНИРОВАННЫХ И РАНДОМИЗИРОВАННЫХ ЭНТРОПИЙНЫХ ПРОЕКЦИЙ ДЛЯ РЕДУКЦИИ РАЗМЕРНОСТИ МАТРИЦЫ ДАННЫХ*

Ю. С. Попков¹, А. Ю. Попков², Ю. А. Дубнов³

Аннотация: Предложены методы детерминированного и рандомизированного проектирования, ориентированные на решение задачи понижения размерности. В случае детерминированного проектирования развивается параллельная процедура сжатия матрицы данных, минимизирующая кросс-энтропию Кульбака–Лейблера с учетом ограничения на информационную емкость, основанная на методе проекции градиента. Для рандомизированного проектирования рассматривается задача понижения размерности признакового пространства. Идея применения процедур проектирования для сжатия матрицы данных реализуется в предлагаемом методе рандомизированного энтропийного проектирования, где используется принцип сохранения среднего расстояния между многомерными и маломерными точками в соответствующих пространствах. Задача поиска оптимальных проекторов сводится к поиску распределения вероятностей, максимизирующего информационную энтропию Ферми при ограничении на среднее расстояние между точками многообразия, которые отображаются матрицами данных и оптимальной проекции.

Ключевые слова: понижение размерности; кросс-энтропия Кульбака–Лейблера; энтропия

DOI: 10.14357/19922264200407

1 Введение

Во многих прикладных задачах обработки данных последние присутствуют в виде прямоугольных матриц $U_{(m \times s)}$. Например, в задачах машинного обучения элементами строк матрицы данных выступают признаки объекта (s), а строки служат характеристиками m объектов в признаковом s -пространстве. По разным причинам возникает необходимость «сжать» матрицу данных, т.е. использовать для обучения матрицу размерности $(m \times r)$ или $(n \times r)$, $n < m$, $r < s$. Содержательно это сводится к уменьшению числа признаков или уменьшению числа признаков и объектов, на массиве которых проводится обучение.

Данная проблема вложена в более общую: приближение заданного набора многомерных точек маломерным аффинным многообразием [1]. Здесь следует отметить метод главных компонент (МГК) [2, 3] и его робастные версии [4], а также метод случайных проекций [5, 6].

В [7] был предложен энтропийный метод одномерного (столбцы или строки) детерминированного сжатия матрицы данных (EDR-метод), основан-

ный на «прямом» и «обратном» проектировании. Матрицы-проекторы определяются путем минимизации кросс-энтропийного функционала.

В данной работе EDR-метод развивается для детерминированного параллельного сжатия матрицы данных на основе процедуры минимизации кросс-энтропийного функционала специального вида при ограничениях. Последние связаны с информационной емкостью матрицы-проекции.

Идея применения процедур проектирования для сжатия матрицы данных реализуется в предлагаемом методе рандомизированного энтропийного проектирования (REDR-метод). Здесь используется принцип сохранения среднего расстояния между многомерными и маломерными точками в соответствующих пространствах.

2 Параллельное проектирование с ограничениями (EDR-метод)

Параллельная реализация процедуры «прямого» и «обратного» проектирования, примененная

* Работа выполнена при поддержке РФФИ (проекты 17-29-03119 и 20-07-00470).

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Институт проблем управления им. В. А. Трапезникова Российской академии наук; ОРТ Брауде Колледж, Кармиель, Израиль, popkov@isa.ru

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

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Национальный исследовательский университет «Высшая школа экономики», yury.dubnov@phystech.edu

к матрице данных $U > 0$, приводит к следующей цепочке матричных равенств:

– «прямая» проекция

$$\left. \begin{aligned} U_{(m \times s)} Q_{(s \times r)} &= Y_{(m \times r)}; \\ B_{(n \times m)} Y_{(m \times r)} &= Z_{(n \times r)}, \end{aligned} \right\} \quad (1)$$

– «обратная» проекция

$$\left. \begin{aligned} Z_{(n \times r)} W_{(r \times s)} &= D_{(n \times s)}; \\ E_{(m \times n)} D_{(n \times s)} &= X_{(m \times s)}. \end{aligned} \right\} \quad (2)$$

Матрицы-проекторы Q, B, W и E — неотрицательные. Равенства (1) преобразуют матрицу $U_{(m \times s)}$ в «сжатую» матрицу $Z_{(n \times r)}$, где $n < m, r < s$. Равенство (2) преобразует матрицу $Z_{(n \times r)}$ в матрицу $X_{m \times s}$ той же размерности, что и исходная матрица данных $U_{(m \times s)}$.

Из равенств (1) и (2) имеем:

$$X_{(m \times s)} = E_{(m \times n)} \{ [B_{(n \times m)} (U_{(m \times s)} Q_{(s \times r)})] W_{(r \times s)} \} > 0.$$

Скобки в этом равенстве указывают на последовательность операций проектирования:

$$(\bullet) \rightarrow [\bullet] \rightarrow \{\bullet\}.$$

Элементы матрицы-проекции $Z_{(n \times s)}$ имеют вид:

$$z_{\mu, \nu} = \sum_{\beta=1}^m b_{\mu, \beta} \sum_{\alpha=1}^r u_{\beta, \alpha} q_{\alpha, \nu}, \quad \mu = \overline{1, n}, \quad \nu = \overline{1, r}.$$

Элементы матрицы $X_{(m \times s)}$ имеют вид:

$$x_{ij} = \sum_{\mu=1}^n e_{i, \mu} \sum_{\nu=1}^r w_{\nu, j} \sum_{\beta=1}^m b_{\mu, \beta} \sum_{\alpha=1}^s u_{\beta, \alpha} q_{\alpha, \nu} > 0, \quad i = \overline{1, m}, \quad j = \overline{1, s}. \quad (3)$$

Для измерения отклонения преобразованной матрицы $X_{(m \times s)}$ от исходной $U_{(m \times s)}$ воспользуемся информационной кросс-энтропией [8]:

$$\mathcal{H}(X|U) = \sum_{i=1}^m \sum_{j=1}^s s_{ij}(X|U), \quad (4)$$

где

$$s_{ij} = x_{ij} \ln \frac{x_{ij}}{u_{ij}}.$$

С учетом равенства (3) нетрудно видеть, что информационная кросс-энтропия (4) есть скалярная

функция от матрицы данных $U > 0$ и матриц-проекторов $(Q, B, W, E) \geq 0$, т. е.

$$\mathcal{H} = \mathcal{H}(U|Q, B, W, E).$$

Важным показателем качества процедуры редукции служит оптимальное снижение информационной емкости редуцированной матрицы $Z_{(n \times r)}$ по сравнению с информационной емкостью исходной матрицы данных $U_{(m \times s)}$ [9].

Информационная емкость измеряется в энтропийных терминах:

$$\begin{aligned} \mathcal{I}_Z &= \sum_{(i,j)=1}^{n,r} z_{ij}(Q, B) \ln z_{ij}(Q, B) + e^{-1}nr; \\ \mathcal{I}_U &= \sum_{(i,j)=1}^{m,s} u_{ij} \ln u_{ij} + e^{-1}ms. \end{aligned}$$

Различие в указанных информационных емкостях будем характеризовать квадратичным функционалом:

$$\mathcal{J}(Q, B) = \left(\sum_{(i,j)=1}^{n,r} z_{ij}(Q, B) \ln z_{ij}(Q, B) - A \right)^2,$$

где

$$A = e^{-1}(ms - nr) + \sum_{(i,j)=1}^{m,s} u_{ij} \ln u_{ij}.$$

Образуем обобщенный функционал

$$\mathcal{F}(U|Q, B, W, E) = \mathcal{H}(U|Q, B, W, E) + \mathcal{J}(Q, B)$$

и оптимальные значения неотрицательных элементов матриц-проекторов будем определять, минимизируя функционал $\mathcal{F}(U|Q, B, W, E)$:

$$\begin{aligned} (Q^*, B^*, W^*, E^*) = \\ + \arg \min_{(Q, B, W, E) \geq 0} \mathcal{F}(U|Q, B, W, E). \quad (5) \end{aligned}$$

Замечание. В задаче (5) условие близости информационных емкостей матрицы данных и редуцированной матрицы может быть реализовано ввиду соответствующего ограничения. Тогда оптимальные значения неотрицательных элементов матриц-проекторов определяются решением следующей задачи:

$$(Q^*, B^*, W^*, E^*) = \arg \min_{(Q, B, W, E) \in \Omega} \mathcal{H}(U|Q, B, W, E),$$

где

$$\begin{aligned} \Omega = \{ (Q, B, W, E) : (Q, B, W, E) \geq 0; \\ \mathcal{I}_Z(Q, B) \geq \delta \mathcal{I}_U \}, \quad \delta \in (0, 1). \end{aligned}$$

Допустимый уровень снижения информационной емкости редуцированной матрицы регулируется параметром δ .

3 Алгоритм решения задачи (5)

Задача (5) представляет собой задачу минимизации функционала на неотрицательном ортанте. Для ее решения можно применить метод проекций градиента, предварительно осуществив векторизацию соответствующих матриц.

Введем блочные векторы \mathbf{v} и \mathbf{c} , каждый размерности $N = (sr + nm)$, блоками которых являются векторы \mathbf{q} и \mathbf{b} для вектора \mathbf{v} и векторы \mathbf{w} и \mathbf{e} для вектора \mathbf{c} . Все векторы — результат векторизации матриц-проекторов Q, B, W и E соответственно.

Представим (5) в следующем виде:

$$\mathcal{F}(\mathbf{v}, \mathbf{c} | \mathbf{u}) = \mathcal{H}(\mathbf{v}, \mathbf{c} | \mathbf{u}) + \mathcal{J}(\mathbf{v}) \Rightarrow \min,$$

где

$$\mathcal{H}(\mathbf{v}, \mathbf{c} | \mathbf{u}) = \langle \mathbf{x}(\mathbf{v}, \mathbf{c}), \mathbf{y}(\mathbf{v}, \mathbf{c} | \mathbf{u}) \rangle_{R^{ms}};$$

$$\mathcal{J}(\mathbf{v}) = \langle \mathbf{z}(\mathbf{v}), \mathbf{g}(\mathbf{v}) \rangle_{R^{nr}};$$

$$\mathbf{v} \geq \mathbf{0}; \quad \mathbf{c} \geq \mathbf{0}.$$

Здесь приняты следующие обозначения [10]:

- вектор \mathbf{u} — результат векторизации матрицы данных U , его размерность (ms); вектор \mathbf{x} размерности (ms) с компонентами (3);
- вектор \mathbf{y} размерности (ms) с компонентами

$$y_k = \ln \frac{x_k}{u_k}, \quad k = \overline{1, ms};$$

- вектор \mathbf{g} размерности (nr) с компонентами

$$g_k = \ln z_k, \quad k = \overline{1, nr}.$$

Для численного решения этой задачи применим покоординатную схему метода проекций градиента.

В параллельной процедуре вектор \mathbf{v} объединяет элементы матриц Q и B , с помощью которых проводится «сжатие» матрицы данных по одному измерению. В вектор \mathbf{c} входят элементы матриц W и E , с помощью которых проводится «сжатие» по второму измерению. Такое разделение векторов удобно для применения покоординатного алгоритма.

Итерационный шаг покоординатной схемы метода проекций градиента состоит из двух последовательно реализуемых этапов: на одном осуществляется итерация по \mathbf{v} -проециям градиента, а на другом — по \mathbf{c} -проециям градиента функционала $\mathcal{F}(\mathbf{v}, \mathbf{c} | \mathbf{u})$. Обозначим градиенты по этим векторам:

$$\nabla_{\mathbf{v}} \mathcal{F} = \nabla_{\mathbf{v}} \mathcal{H} + \nabla_{\mathbf{v}} \mathcal{J};$$

$$\nabla_{\mathbf{c}} \mathcal{F} = \nabla_{\mathbf{c}} \mathcal{H}.$$

Алгоритм минимизации функционала \mathcal{F} имеет следующий вид:

(а) начальный шаг

$$\mathbf{v}^0 > \mathbf{0}, \quad \mathbf{c}^0 > \mathbf{0};$$

(б) i -й итерационный шаг

$$X^i = E^i \{ [B^i (UQ^i)] W^i \};$$

$$\mathcal{F}^i = \mathcal{H}^i(\mathbf{v}^i, \mathbf{c}^i | \mathbf{u}) + \mathcal{J}^i(\mathbf{v}^i);$$

$$\mathbf{v}^{(i+1)} = \begin{cases} \mathbf{v}^i + \gamma_{\mathbf{v}} (\nabla_{\mathbf{v}} \mathcal{H}^i(\mathbf{v}^i, \mathbf{c}^i | \mathbf{u}) + \nabla_{\mathbf{v}} \mathcal{J}^i(\mathbf{v}^i)), & \text{если } \mathbf{v}^{(i+1)} \geq \mathbf{0}; \\ \mathbf{v}^i, & \text{если } \mathbf{v}^{(i+1)} < \mathbf{0}; \end{cases}$$

$$\mathbf{v}^{(i+1)} \Rightarrow Q^{(i+1)}, B^{(i+1)};$$

$$\mathbf{c}^{(i+1)} = \begin{cases} \mathbf{c}^i + \gamma_{\mathbf{c}} \nabla_{\mathbf{c}} \mathcal{H}(\mathbf{v}^i, \mathbf{c}^i | \mathbf{u}), & \text{если } \mathbf{c}^{(i+1)} \geq \mathbf{0}, \\ \mathbf{c}^i, & \text{если } \mathbf{c}^{(i+1)} < \mathbf{0}; \end{cases}$$

$$\mathbf{c}^{(i+1)} \Rightarrow W^{(i+1)}, E^{(i+1)};$$

$$X^{(i+1)} = E^{(i+1)} \{ [B^{(i+1)} (UQ^{(i+1)})] W^{(i+1)} \};$$

$$\mathcal{F}^{(i+1)} = \mathcal{H}^{(i+1)} + \mathcal{J}^{(i+1)};$$

(в) условие остановки:

$$\text{если } \mathcal{F}^{(i+1)} - \mathcal{F}^{(i)} \leq \Delta, \text{ то STOP.}$$

4 Энтропийно-рандомизированное проектирование (REDR-метод)

Рассмотрим матрицу данных $U_{(m \times s)}$. В пространстве R^s ее отображает множество точек $\mathcal{U} = \{\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}\}$. Будем придерживаться использованной ранее интерпретации «объекты (m) — признаки (s)». Объекты обычно выбираются из одного класса, что позволяет выдвинуть гипотезу о том, что расстояние между точками в множестве \mathcal{U} флуктуирует несильно, т.е. точки образуют достаточно «компактную» группу.

Определим *индикатор* этой группы (матрицы данных) в виде:

$$\rho_U = \frac{2}{m(m-1)} \sum_{(\alpha, \beta)=1}^m \varrho(\mathbf{u}^\alpha, \mathbf{u}^\beta). \quad (6)$$

Матрицу данных $U_{(m \times s)}$ трансформируем в матрицу-проецию $Z_{(n \times r)}$, $n < m$, $r < s$, с помощью левых и правых матриц-проекторов:

$$Z_{(n \times r)} = B_{(n \times m)} U_{(m \times s)} Q_{(s \times r)}.$$

Матрицы B и Q — случайные, интервального типа:

$$Q \in \mathcal{Q} = [Q^-, Q^+]; \quad B \in \mathcal{B} = [B^-, B^+].$$

Совместная функция плотности распределения вероятности (ПРВ) $P(Q, B)$ определена на носителе \mathcal{Z} :

$$(Q, B) \in \mathcal{Z} = \mathcal{Q} \cap \mathcal{B}.$$

Элементы матрицы-проекции \mathcal{Z} имеют вид:

$$z_{\mu, \nu}(Q, B) = \sum_{\beta=1}^m b_{\mu, \beta} \sum_{\alpha=1}^r u_{\beta, \alpha} q_{\alpha, \nu},$$

$$\mu = \overline{1, n}, \nu = \overline{1, r}.$$

По аналогии с (6) определим индикатор матрицы-проекции $Z_{(n \times s)}$ в виде:

$$\rho_Z(Q, B) = \frac{2}{n(n-1)} \sum_{(\eta, \kappa)=1}^n \varrho(\mathbf{z}^{(\eta)}(Q, B), \mathbf{z}^{(\kappa)}(Q, B)).$$

Поскольку элементы матриц-проекторов — случайные, индикатор $\rho_Z(Q, B)$ является функцией случайных переменных. Его математическое ожидание:

$$G[P(Q, B)] = \int_{\mathcal{Z}} P(Q, B) \rho_Z(Q, B) dQdB.$$

Для определения функции ПРВ $P(Q, B)$ будем использовать оценку максимальной энтропии [11]

$$\mathcal{H}[P(Q, B)] = - \int_{\mathcal{Z}} P(Q, B) \ln P(Q, B) dQdB \Rightarrow \max \quad (7)$$

при ограничениях:

$$\int_{\mathcal{Z}} P(Q, B) dQdB = 1; \quad G[P(Q, B)] = \delta\rho_U,$$

$$0 < \varepsilon \leq \delta \leq \theta < 1. \quad (8)$$

Задача (7)–(8) относится к классу ляпуновских задач [12], для которых условия оптимальности формулируются в терминах стационарности функционала Лагранжа:

$$\mathcal{L}[P(Q, B), \lambda] = \mathcal{H}[P(Q, B)] + \lambda (\delta\rho_U - G[P(Q, B)]),$$

где λ — скалярный множитель Лагранжа.

Получим:

$$P^*(Q, B) = \frac{\exp(-\lambda \rho_Z(Q, B))}{\mathcal{P}(\lambda)}, \quad (9)$$

где

$$\mathcal{P}(\lambda) = \int_{\mathcal{Z}} \exp(-\lambda \rho_Z(Q, B)) dQdB.$$

Множитель Лагранжа λ определяется из следующего уравнения:

$$\frac{\int_{\mathcal{Z}} \exp(-\lambda \rho_Z(Q, B)) \rho_Z(Q, B) dQdB}{\mathcal{P}(\lambda)} = \delta\rho_U.$$

Таким образом, энтропийно-оптимальная функция ПРВ $P^*(Q, B)$ (9) позволяет, путем ее семплирования, генерировать матрицы-проекторы Q и B , сохраняющие «в среднем» расстояние между точками (векторами $\mathbf{z}^{(\alpha)}$) матрицы-проекции \mathcal{Z} .

5 Энтропийные случайные матрицы-проекторы с заданными значениями элементов

Рассмотрим матрицу данных $U_{(m \times s)}$, которую нужно «сжать» по переменной s до размера r :

$$Y_{(m \times r)} = U_{(m \times s)} Q_{(s \times r)}. \quad (10)$$

Матрица $U_{(m \times s)} \geq 0$ и имеет нормированные элементы ($0 \leq u_{ij} \leq 1$). Введем полезные обозначения:

- вектор-столбец $\bullet^{(\circ)}$, вектор-строка $\bullet^{(\circ)}$;
- векторы-строки: $\mathbf{u}^{(i)} = \{u_{i,1}, \dots, u_{i,s}\}$, $i = \overline{1, m}$;
- векторы-строки $\mathbf{y}^{(k)} = \{y_{k,1}, \dots, y_{k,r}\}$, $k = \overline{1, m}$;
- векторы-столбцы ${}^{(l)}\mathbf{q} = \{q_{1,l}, \dots, q_{s,l}\}$, $l = \overline{1, r}$.

Тогда равенство (10) представим в виде:

$$\mathbf{y}^{(i)}(Q) = \left\{ {}^{(1)}\mathbf{q}^\top \mathbf{u}^{(i)}, \dots, {}^{(r)}\mathbf{q}^\top \mathbf{u}^{(i)} \right\} \in R^r,$$

$$i = \overline{1, m}.$$

Определим индикатор матрицы-проекции $Y_{(m \times r)}$ в виде:

$$\rho_Y(Q) = \frac{2!(m-2)!}{m!} \sum_{(i,j)=1}^m \|\mathbf{y}^{(i)}(Q) - \mathbf{y}^{(j)}(Q)\|^2. \quad (11)$$

Рассмотрим случай, когда элементы матрицы $Q_{(s \times r)}$ могут принимать значения 0 или 1 и размещение их в матрице — случайное. Заменяем матрицу строкой длины sr . Число различных последовательностей из 0 и 1 равно $N = 2^{sr}$. В качестве примера для $s = 3$ и $r = 1$ таких реализаций будет 8:

$$000, 100, 010, 001, 110, 011, 101, 111;$$

для $s = 4$ и $r = 1$ их будет 16:

0000, 1000, 0100, 0010, 0001, 1100, 0110, 0011,
1001, 0101, 1010, 1111, 1110, 0111, 1011, 1101.

Для матрицы-проектора $Q_{(s \times r)}$ существует конечное число ее $(0, 1)$ -реализаций:

$$Q^{(1)}, \dots, Q^{(N)}, \quad N = 2^{sr}.$$

Полагая, что реализации — случайные, их вероятностные свойства будем характеризовать функцией распределения вероятностей (ДРВ) с дискретным носителем $W(\alpha)$, $\alpha = \overline{1, N}$, где

$$W(\alpha) = w_\alpha, \quad 0 \leq w_\alpha \leq 1, \quad l = \overline{1, N}.$$

Математическое ожидание индикатора (11):

$$\mathcal{M}_W = \sum_{\alpha=1}^N w_\alpha \rho(Q^{(\alpha)}).$$

Функцию ДРВ $W(\alpha)$ будем искать в классе функций, максимизирующих функцию информационной энтропии Ферми [13]:

$$F(W) = - \sum_{\alpha=1}^N w_\alpha \ln w_\alpha + (1 - w_\alpha) \ln(1 - w_\alpha) \Rightarrow \Rightarrow \max, \quad (12)$$

при ограничении математического ожидания индикатора (11):

$$G[\mathbf{w} | Q^{(1)}, \dots, Q^{(N)}] = \sum_{\alpha=1}^N w_\alpha \rho(Q^{(\alpha)}) = \delta \rho_U, \quad 0 < \varepsilon \leq \delta \leq 1. \quad (13)$$

Задача (12)–(13) представляет собой конечномерную задачу на условный экстремум с вогнутой целевой функцией и квадратичным ограничением. Хотя последнее требует специального исследования, но, поскольку оно одно, решение несложно найти численно.

Рассмотрим функцию Лагранжа:

$$L(\mathbf{w}, \lambda) = F(\mathbf{w}) + \lambda \left(\delta \rho_U - \sum_{\alpha=1}^N w_\alpha \rho(Q^{(\alpha)}) \right).$$

Условия стационарности этой функции имеют вид ($\alpha = \overline{1, N}$):

$$\frac{\partial L}{\partial w_\alpha} = - \ln \frac{w_\alpha}{1 - w_\alpha} - \lambda \rho(Q^{(\alpha)}) = 0;$$

$$\frac{\partial L}{\partial \lambda} = \left(\delta \rho_U - \sum_{\alpha=1}^N w_\alpha \rho(Q^{(\alpha)}) \right) = 0.$$

Отсюда получаем, что энтропийно-оптимальное распределение вероятностей имеет вид:

$$w_\alpha^* = \frac{\exp(-\lambda \rho(Q^{(\alpha)}))}{1 + \exp(-\lambda \rho(Q^{(\alpha)}))}, \quad \alpha = \overline{1, N}, \quad (14)$$

где параметр λ определяется из следующего уравнения:

$$\sum_{\alpha=1}^N \frac{\exp(-\lambda \rho(Q^{(\alpha)})) \rho(Q^{(\alpha)})}{1 + \exp(-\lambda \rho(Q^{(\alpha)}))} = \delta \rho_U.$$

Таким образом, равенство (14) определяет распределение вероятностей матриц-проекторов с элементами $\{0, 1\}$. Имеет смысл выбрать матрицу-проектор:

$$Q^{(\alpha^*)} \Rightarrow \alpha^* = \max_{1 \leq \alpha \leq N} w_\alpha^*,$$

хотя возможны и другие стратегии.

Алгоритм

Шаг 0. Нормировка матрицы данных:

$$u_{ij} := \frac{u_{ij} - u_{\min}}{u_{\max} - u_{\min}}, \quad i = \overline{1, m}; j = \overline{1, s}.$$

Шаг 1. Вычисление индикатора матрицы данных:

$$\rho_U := \frac{2!(m-2)!}{m!} \sum_{\beta, \gamma=1, \gamma \neq \beta}^m \varrho(\mathbf{u}^{(\beta)}, \mathbf{u}^{(\gamma)}).$$

Шаг 2. Генерация множества Ω матриц-проекторов с элементами 0, 1:

$$Q^{(\alpha)}, \quad \alpha = \overline{1, N}, \quad N = 2^{rs}.$$

Шаг 3. Формирование множества \mathfrak{U} матриц-проекции с элементами

$$y_{i,k}^{(\alpha)} := \sum_{\nu=1}^s u_{i,\nu} q_{\nu,k}^{(\alpha)}, \quad i = \overline{1, m}, k = \overline{1, r},$$

и векторами-строками

$$\mathbf{y}_\nu^\alpha := \{y_{\nu,1}^{(\alpha)}, \dots, y_{\nu,r}^{(\alpha)}\}, \quad \nu = \overline{1, m}.$$

Шаг 4. Вычисление индикатора матриц-проекции:

$$\rho_Y^{(\alpha)} := \frac{2!(m-2)!}{m!} \sum_{\nu, \mu=1, \nu \neq \mu}^m \varrho(\mathbf{y}^{(\nu)}, \mathbf{y}^{(\mu)}), \quad \alpha = \overline{1, N}.$$

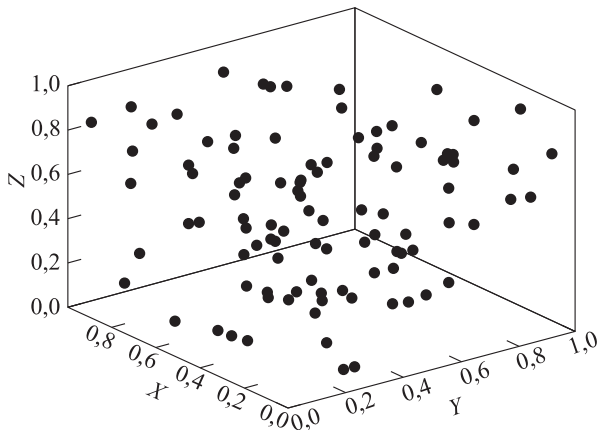


Рис. 1 Множество \mathcal{U} трехмерных точек

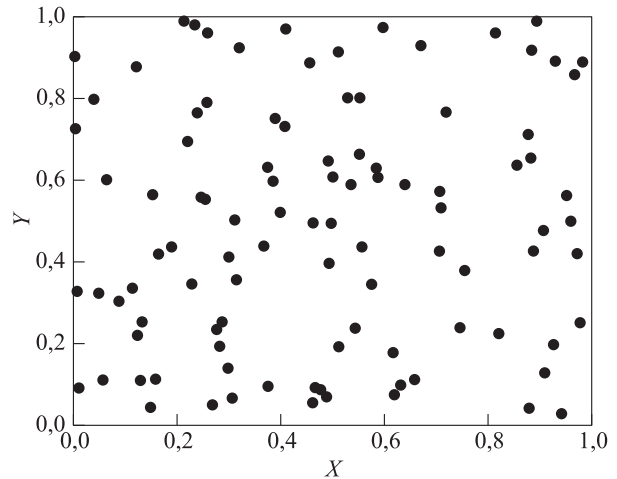


Рис. 2 Множество \mathcal{Q} двумерных точек с $\rho_Y = \rho_U$

Шаг 5. Определение значения множителя Лагранжа λ^* из уравнения:

$$\sum_{\alpha=1}^N \frac{\exp(-\lambda \rho_Y^{(\alpha)}) \rho_Y^{(\alpha)}}{1 + \exp(-\lambda \rho_Y^{(\alpha)})} = \delta \rho_U.$$

Шаг 6. Определение в множестве \mathcal{Q} наиболее вероятной матрицы-проектора:

$$Q^{(\alpha^*)} : \alpha^* = \arg \max_{\alpha} w(\alpha | \lambda^*).$$

Шаг 7. Определение элементов наиболее вероятной матрицы-проекции:

$$Y_{(m \times r)}^{(\alpha^*)} = U_{(m \times s)} Q_{(s \times r)}^{(\alpha^*)}.$$

СТОП

Пример. В качестве иллюстрации предлагаемого алгоритма рассмотрим матрицу данных $U_{(100 \times 3)}$, элементами которой служат случайные числа из интервала $[0, 1]$. В пространстве R^3 эта матрица отображается в множество \mathcal{U} трехмерных точек, характеризуемых векторами-строками $\mathbf{u}^{(i)}, i = \overline{1, 100}$. Это множество изображено на рис. 1. Расстояния между точками — евклидовы, и индикатор матрицы данных $\rho_U = 0,677$.

Применяя рассмотренный выше алгоритм с параметром $\delta = 1,0$, генерируем множество \mathcal{Q} двумерных точек с $\rho_Y = \rho_U$, показанное на рис. 2. Множество \mathcal{Q} локализовано в квадрате со стороной 1,7.

6 Заключение

Предложены процедуры детерминированного и рандомизированного проектирования, ориенти-

рованные на редукцию размерности матрицы данных. В случае детерминированного проектирования развивается параллельная процедура сжатия матрицы данных, минимизирующая кросс-энтропию Кульбака–Лейблера с учетом ограничения на информационную емкость. Предложен алгоритм условной минимизации, использующий метод проекций градиента.

Для рандомизированного проектирования рассмотрена задача редукции матрицы данных по одному измерению и с заданными элементами матриц-проекторов, в частности со случайными $(0, 1)$ -элементами. Задача сводится к поиску распределения вероятностей, максимизирующего информационную энтропию Ферми при ограничении на среднее расстояние между точками многообразия, которые отображаются матрицами данных и оптимальной проекции. Предложен алгоритм для решения этой задачи.

Литература

1. Bruckstein A. M., Donoho D. L., Elad M. From sparse solutions of systems of equations to sparse modeling of signals and images // SIAM Rev., 2009. Vol. 51. Iss. 1. P. 34–81.
2. Кендалл М., Стьюарт А. Статистические выводы и связи / Пер. с англ. — М.: Наука, 1973. 896 с. (Kendall M. G., Stuart A. The advanced theory of statistics. — London: Charles Griffin, 1961. Vol. 2. 676 p.)
3. Jolliffe I. Principal component analysis. — New York, NY, USA: Springer, 2011. 488 p. doi: 10.1007/b98835.
4. Поляк Б. Т., Хлебников М. В. Метод главных компонент: робастные версии // Автоматика и телемеханика, 2017. № 3. С. 130–148.

5. Bingham E., Mannila H. Random projection in dimensionality reduction: Applications to image and text data // 7th ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings. — New York, NY, USA: ACM, 2001. P. 245–250. doi: 10.1145/502512.502546.
6. Vempala S.S. The random projection method. — DIMACS ser. in discrete mathematics and theoretical computer science. — Providence, RI, USA: American Mathematical Society, 2004. Vol. 65. 105 p.
7. Попков Ю. С., Дубнов Ю. А., Попков А. Ю. Энтропийная редукция размерности в задачах рандомизированного машинного обучения // Автоматика и телемеханика, 2018. № 11. С. 106–122.
8. Kullback S., Leibler R. A. On information and sufficiency // Ann. Math. Stat., 1951. Vol. 22. Iss. 1. P. 79–86.
9. Попков Ю. С., Попков А. Ю. Кросс-энтропийная оптимальная редукция размерности матрицы данных с ограничением информационной емкости // Докл. Акад. наук, 2019. Т. 488. С. 21–23. doi: 10.31857/S0869-5652488121-23.
10. Magnus J. R., Neudecker H. Matrix differential calculus with applications in statistics and econometrics. — Chichester – New York – Brisbane – Toronto – Singapore: John Wiley & Sons, 1988. 393 p.
11. Попков Ю. С. Асимптотическая эффективность оценок максимальной энтропии // Докл. Акад. наук, 2020. Т. 493. С. 104–107. doi: 10.31857/S2686954320040165.
12. Иоффе А. Д., Тихомиров В. М. Теория экстремальных задач. — М.: Наука, 1984. 481 с.
13. Popkov Yu. S. Macrosystems theory and its applications. — Lecture notes in control and information sciences ser. — Berlin–Heidelberg: Springer-Verlag, 1995. Vol. 203. 327 p.

Поступила в редакцию 25.12.19

DETERMINISTIC AND RANDOMIZED METHODS OF ENTROPY PROJECTION FOR DIMENSIONALITY REDUCTION PROBLEMS

Y. S. Popkov^{1,2,3}, A. Y. Popkov¹, and Y. A. Dubnov^{1,4}

¹Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation

³ORT Braude College, Karmiel 2161002, Israel

⁴National Research University Higher School of Economics, 20 Myasnitskaya Str., Moscow 101000, Russian Federation

Abstract: The work is devoted to development of methods for deterministic and randomized projection aimed at dimensionality reduction problems. In the deterministic case, the authors develop the parallel reduction procedure minimizing Kullback–Leibler cross-entropy target to condition on information capacity based on the gradient projection method. In the randomized case, the authors solve the problem of reduction of feature space. The idea of application of projection procedures for reduction of data matrix is implemented in the proposed method of randomized entropy projection where the authors use the principle of keeping average distances between high- and low-dimensional points in the corresponding spaces. The problem leads to searching of a probability distribution maximizing Fermi entropy target to average distance between points.

Keywords: dimensionality reduction; Kullback–Leibler cross-entropy; entropy

DOI: 10.14357/19922264200407

Acknowledgments

This work was supported by RFBR, projects Nos. 17-29-03119 and 20-07-00470.

References

1. Bruckstein, A. M., D. L. Donoho, and M. Elad. 2009. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.* 51(1):34–81.
2. Kendall, M. G., and A. Stuart. 1961. *The advanced theory of statistics*. London: Charles Griffin. Vol. 2. 676 p.
3. Jolliffe, I. 2011. *Principal component analysis*. New York, NY: Springer. 488 p. doi: 10.1007/b98835.
4. Polyak, B. T., and M. T. Khlebnikov. 2017. Principal component analysis: Robust versions. *Automat. Rem. Contr.* 78:490–506.
5. Bingham, E., and H. Mannila. 2001. Random projection in dimensionality reduction: Applications to image and

- text data. *7th ACM SIGKDD Conference (International) on Knowledge Discovery and Data Mining Proceedings*. ACM. 245–250. doi: 10.1145/502512.502546.
6. Vempala, S. S. 2004. *The random projection method*. DIMACS ser. in discrete mathematics and theoretical computer science. Providence, RI: American Mathematical Society. Vol. 65. 105 p.
 7. Popkov, Y. S., Y. A. Dubnov, and A. Y. Popkov. 2018. Entropy dimension reduction method for randomized machine learning problems. *Automat. Rem. Contr.* 79(11): 2038–2051.
 8. Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. *Ann. Math. Stat.* 22(1):79–86.
 9. Popkov, Y. S., and A. Y. Popkov. 2019. Cross-entropy optimal dimensionality reduction with a condition on information capacity. *Dokl. Math.* 100:420–422.
 10. Magnus, J. R., and H. Neudecker. 1988. *Matrix differential calculus with applications in statistics and econometrics*. Chichester – New York – Brisbane – Toronto – Singapore: John Wiley & Sons. 393 p.
 11. Popkov, Y. S. 2020. Asymptotic efficiency of maximum entropy estimates. *Dokl. Math.* 102:350–352. doi: org/10.1134/S106456242004016X.
 12. Joffe, A. D., and V. M. Tikhomirov. 1984. *Teoriya ekstremal'-nykh zadach* [Theory of extreme problems]. Moscow: Nauka. 481 p.
 13. Popkov, Yu. S. 1995. *Macrosystems theory and its applications*. Lecture notes in control and information sciences ser. Berlin–Heidelberg: Springer-Verlag. Vol. 203. 327 p.

Received December 25, 2019

Contributors

Popkov Yuri S. (b. 1937) — Doctor of Science in technology, professor, Academician of RAS, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; principal scientist, V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 65 Profsoyuznaya Str., Moscow 117997, Russian Federation; senior research fellow, ORT Braude College, Karmiel 2161002, Israel; popkov@isa.ru

Popkov Alexey Y. (b. 1978) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; apopkov@isa.ru

Dubnov Yuri A. (b. 1990) — scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; senior lecturer, National Research University Higher School of Economics, 20 Myasnitskaya Str., Moscow 101000, Russian Federation; yury.dubnov@phystech.edu

ОПТИМИЗАЦИЯ СТРУКТУРЫ СЕТЕЙ ГЛУБОКОГО ОБУЧЕНИЯ*

М. С. Потанин¹, К. О. Вайсер², В. А. Жолобов³, В. В. Стрижов⁴

Аннотация: Исследуется проблема выбора оптимальной структуры модели. Моделью служит суперпозиция обобщенных линейных моделей, элементами которой являются линейная регрессия, логистическая регрессия, метод главных компонент, автоэнкодер и нейросеть. Под структурой модели понимаются значения структурных параметров модели, задающих вид итоговой суперпозиции. Исследуются свойства алгоритма выбора структуры модели. Исследуется зависимость точности, сложности и устойчивости модели от способа задания структуры. Создан алгоритм выбора оптимальной структуры нейронной сети. Проведен вычислительный эксперимент с использованием реальных и синтетических данных. В результате эксперимента существенно снижена структурная сложность моделей с сохранением точности аппроксимации.

Ключевые слова: выбор моделей; линейные модели; автокодировщик; нейронные сети; структура; генетический алгоритм

DOI: 10.14357/19922264200408

1 Введение

Решается задача аппроксимации выборки нейронными сетями. Нейронная сеть служит универсальной моделью [1, 2], так как приближает произвольную непрерывную функцию многих переменных с любой точностью. Нейрон, или однослойная нейронная сеть, представляет собой суперпозицию двух функций — функции активации и линейной комбинации признаков объекта. Но однослойные сети применимы только для линейно разделимых выборок. Для аппроксимации выборок общего вида требуется универсальная модель, оптимизация структуры которой и исследуется в данной работе.

Теорема 1 (Колмогоров, 1961) в [3] утверждает, что функция от n аргументов представима в виде комбинации $n(2n + 1)$ функций одного аргумента. Какими именно должны быть функции σ_i и g_{ij} , не указывается. Теорема об универсальной аппроксимации 2 (Цыбенко, 1989) в [3] утверждает, что искусственная нейронная сеть прямой связи, в которой связи не образуют циклов, с одним скрытым слоем аппроксимирует любую непрерывную функцию многих переменных с любой точностью. Однако за-

труднительно выбрать такую структуру нейронной сети, чтобы размеры скрытого слоя не были велики. В теореме 3 (Ханин, 2017) оценивается оптимальная размерность скрытых слоев и обосновывается возможность замены нейронной сети с функциями активации ReLU [3] с входным слоем размерности n и одним скрытым слоем размерности k на эквивалентную с глубиной $k + 2$ и размерностями скрытых слоев $n + 2$. Эти три теоремы и определяют исследуемую структуру суперпозиций сети глубокого обучения.

Исследуется зависимость ошибки от суперпозиции автокодировщиков [4] и многослойной нейронной сети. Ошибка состоит из двух слагаемых: ошибки восстановления элементов выборки после кодирования и восстановления зависимых переменных. Слагаемые используют одни и те же признаки объектов, которые являются независимыми переменными, но разные зависимые переменные. Для автокодировщика зависимые переменные — это сами признаки объекта, для нейронной сети, следующей за ним, зависимая переменная — ответ y на объекте. Точка разделения — это место в суперпозиции, где автокодировщик, имеющий оптимальные параметры, передает преобразован-

*Работа выполнена при поддержке РФФИ (проекты 19-07-1155, 19-07-0885) и правительства РФ (соглашение 05.Y09.21.0018). 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 имени М. В. Ломоносова с Фондом поддержки проектов Национальной технологической инициативы от 11.12.2018 № 13/1251/2018.

¹Московский физико-технический институт, mark.potinin@phystech.edu

²Московский физико-технический институт, vajser.ko@phystech.edu

³Московский физико-технический институт, zholobov.va@phystech.edu

⁴Вычислительный центр имени А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский физико-технический институт, strizov@ccas.ru

ный вектор признаков в нейросеть. Необходимо найти оптимальное расположение разделения автокодировщика и сети, которое минимизирует ошибку аппроксимации выборки. Под структурой такой модели понимаются величины, задающие вид итоговой суперпозиции, т.е. число слоев автокодировщика и нейросети, а также число нейронов в слоях. Процедура минимизации ошибки аппроксимации выборки следующая: сначала максимизируется точность реконструкции кодировщиков, затем оптимизируются параметры нейросети. Вместе с точностью оптимизируется сложность модели. Под сложностью понимается структурная сложность модели — число параметров модели.

В данном исследовании для выбора оптимальной структуры используется генетический алгоритм. Задается множество случайных начальных значений структурных параметров. Затем вычисляется значение функции ошибки аппроксимации, которое характеризует качество модели в наборе. Согласно этой функции выбираются модели, которые обмениваются структурными параметрами, образуя новую структуру. Многократное повторение этой операции позволяет получить оптимальную структуру модели.

Алгоритмы прореживания OBD (optimal brain damage) [5] и OBS (optimal brain surgeon) [6] используют производные второго порядка функции ошибки по параметрам для выбора удаляемых параметров. В [7] авторы предлагают новый метод прореживания для глубоких нейронных сетей. Параметры каждого слоя независимо прореживаются на основе производных второго порядка функции послыонной ошибки по соответствующим параметрам. В [8] используются производные первого порядка для снижения сложности сверточных нейронных сетей. Использование [5] позволило в [9] уменьшить число структурных параметров рекуррентной нейронной сети на 60% и снизить ошибку на валидационной выборке на 30% по сравнению с исходной моделью. Автоматизированные методы поиска нейросетевой архитектуры [10] являются частью парадигмы автоматического машинного обучения [11]. Система поиска получает на вход набор данных и тип решаемой задачи. Результат — оптимизированная архитектура нейронной сети.

2 Постановка задачи выбора модели

Задана выборка (\mathbf{x}_i, y_i) , $\mathbf{x}_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}^1$, $i = 1, \dots, m$, где \mathbf{x} — описание объекта, вектор из n элементов признаков; y — зависимая переменная.

Моделью называется отображение $f : (\mathbf{x}, \mathbf{w}) \mapsto y$. Требуется построить аппроксимирующую модель $f(\mathbf{x})$ вида:

$$f = \sigma_k \circ \mathbf{w}_k^T \sigma_{k-1} \circ \mathbf{W}_{k-1} \sigma_{k-2} \circ \dots \circ \mathbf{W}_2 \sigma_1 \circ \mathbf{W}_1 \mathbf{x}. \quad (1)$$

Эта модель рассматривается как суперпозиция линейной модели, глубокой нейросети и автоэнкодера. Рассмотрим различные модели как частные случаи (1). Линейная, или логистическая, регрессия и один нейрон имеют вид $f(\mathbf{x}, \mathbf{w}) = \sigma(\mathbf{w}^T \mathbf{x})$, где σ — функция активации, непрерывная монотонная дифференцируемая функция (2); \mathbf{w} — вектор параметров; \mathbf{x} — объект, вектор с присоединенным элементом единица, соответствующим аддитивному параметру w_0 . При использовании линейной функции активации получаем линейную регрессию $f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \mathbf{x}$.

Такую функцию активации обозначим $\sigma = \text{id}$. При использовании сигмоидной функции активации получаем модель логистической регрессии:

$$f(\mathbf{x}, \mathbf{w}) = \sigma(\mathbf{w}^T \mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w}^T \mathbf{x})}. \quad (2)$$

Двухслойная нейронная сеть, состоящая из линейной комбинации нейронов, однослойных нейронных сетей:

$$\begin{aligned} f(\mathbf{x}, \mathbf{w}) &= \\ &= \sigma^{(2)} \left(\sum_{i=1}^{n_2} w_i^{(2)} \sigma^{(1)} \left(\sum_{j=1}^n w_{ij}^{(1)} x_j + w_{i0}^{(1)} \right) + w_0^{(2)} \right) = \\ &= \sigma \circ \mathbf{w}^T \sigma \circ \mathbf{W} \mathbf{x}. \end{aligned}$$

Метод главных компонент. Модель допускает вращения признакового пространства, т.е. объекты преобразуются только с помощью поворотов, $\mathbf{h} = \mathbf{W} \mathbf{x}$, где \mathbf{W} — матрица поворота. Она ортогональна: $\mathbf{W} \mathbf{W}^T = \mathbf{I}_n$. Полученное пространство образов \mathbf{h} называется скрытым. Происходит преобразование без потерь.

При удалении нескольких строк оптимальной [12] матрицы \mathbf{W} , например их число $u < n$, полученный вектор \mathbf{h} имеет размер $u \times 1$. Получается проекция \mathbf{h} вектора \mathbf{x} . Согласно теореме С. Р. Рао [12], первые u главных компонент восстанавливают \mathbf{h} оптимальным способом, $\mathbf{r}(\mathbf{x}) = \mathbf{W}^T \mathbf{h}$.

Автокодировщик \mathbf{h} — это монотонное нелинейное отображение входного вектора свободных переменных $\mathbf{x} \in \mathbb{R}^n$ в скрытое представление $\mathbf{h} \in \mathbb{R}^u$ вида:

$$\mathbf{h}(\mathbf{x}) = \sigma(\mathbf{W} \mathbf{x} + \mathbf{b}).$$

В случае $\sigma = \text{id}$ и ортогональной матрицы \mathbf{W} автокодировщик тождествен методу главных компонент. Скрытое представление \mathbf{h} реконструирует вектор \mathbf{x} линейно:

$$\mathbf{r}(\mathbf{x}) = \mathbf{W}'\mathbf{h} + \mathbf{w}'_0.$$

3 Задача выбора оптимальной структуры модели

Решается задача выбора оптимальной структуры модели

$$f = \sigma_k \circ \Gamma_k \otimes \mathbf{w}_k^T \sigma_{k-1} \circ \Gamma_{k-1} \otimes \mathbf{W}_{k-1} \sigma_{k-2} \circ \dots \dots \circ \Gamma_2 \otimes \mathbf{W}_2 \sigma_1 \circ \Gamma_1 \otimes \mathbf{W}_1 \mathbf{x}, \quad (3)$$

где Γ — матрица, задающая структуру модели; \otimes — адамарово произведение, определяемое как поэлементное умножение. Если элемент $\gamma \in \{0, 1\}$ матрицы Γ равен нулю, то соответствующий элемент матрицы параметров \mathbf{W} обнуляется и не участвует в работе модели. Множество индексов, соответствующих ненулевым элементам матрицы Γ , обозначается \mathcal{A} . Требуется найти такое подмножество индексов \mathcal{A}^* , которое доставляет минимум функции

$$\mathcal{A}^* = \arg \min_{\mathcal{A} \subseteq \mathcal{I}} S(f_{\mathcal{A}} | \mathbf{w}^*, \mathcal{D}_{\mathcal{C}}) \quad (4)$$

на разбиении выборки \mathcal{D} , определенной множеством индексов \mathcal{C} . Здесь $\mathcal{I} = \mathcal{C} \sqcup \mathcal{L}$ — все индексы всех матриц Γ . Таким образом, требуется снизить число признаков и повысить устойчивость модели. При этом параметры \mathbf{w}^* модели доставляют минимум ошибки

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} S(\mathbf{w} | \mathcal{D}_{\mathcal{L}}, f_{\mathcal{A}}) \quad (5)$$

на разбиении выборки, определенной множеством \mathcal{L} . Процедура разбиения описана в вычислительном эксперименте.

Генетический алгоритм. Для решения задачи оптимизации структуры (4) используется генетический алгоритм. Структура нейронной сети (3) включает в себя k слоев, l -й слой содержит N_l нейронов, $\sum_{l=1}^k N_l = L$. Каждому слою соответствует матрица $\Gamma_l \in \{0, 1\}^{N_l}$. Это означает, что параметры, которые умножаются поэлементно на ноль, не будут учитываться. Составляется вектор

$$\gamma = [\gamma_1, \gamma_2, \dots, \gamma_L] = \text{vec}[\Gamma_1, \Gamma_2, \dots, \Gamma_k],$$

соответствующий (3). Процедура оптимизации структуры:

1. Задается множество начальных значений $\mathcal{G} = \{\gamma_1, \gamma_2, \dots, \gamma_R\}$, где случайным образом задаются элементы бинарного вектора γ .
2. Для каждого $\gamma_i \in \mathcal{G}$ вычисляется значение функции ошибки S (7) (см. разд. 4).
3. Для каждого γ_i оценивается вероятность выбора его как структуры для скрещивания с помощью функции

$$P_i = \frac{1/S_i}{\sum_{i=1} 1/S_i}.$$

Выбирается пара структур γ_p, γ_q с максимальной вероятностью.

4. Выбирается случайный индекс точки разделения $\nu \in \{1, \dots, L-1\}$.
5. Структуры разделяются на две части, происходит обмен элементами, следующими за ν :

$$\begin{aligned} [\gamma_{p,1}, \dots, \gamma_{p,\nu}, \gamma_{q,\nu+1}, \dots, \gamma_{q,L}] &\rightarrow \gamma'_p, \\ [\gamma_{q,1}, \dots, \gamma_{q,\nu}, \gamma_{p,\nu+1}, \dots, \gamma_{p,L}] &\rightarrow \gamma'_q. \end{aligned}$$

6. Выбираются случайные номера $\eta_1, \dots, \eta_Q \in \{1, \dots, L\}$.
7. У векторов γ'_p, γ'_q инвертируются позиции с номерами η_1, \dots, η_Q .
8. Пункты 4–8 повторяются $R/2$ раз. Множество \mathcal{G} содержит на каждой итерации R структур, которым соответствует наименьшее ошибка.

Здесь R и Q — фиксированные параметры алгоритма. В эксперименте производится настройка Γ по частям, т. е. алгоритм запускается отдельно для каждого слоя. Результатом работы становится вектор, нулевые элементы которого соответствуют нейронам, исключаемым из структуры.

4 Функция ошибки и критерии качества модели

Для оптимизации структуры предлагается использовать композитную функцию ошибки (7). Она состоит из двух слагаемых. Первое слагаемое соответствует точности восстановления зависимой переменной. Второе слагаемое — это точность реконструкции независимой переменной автокодировщиком. Задача (5) представляет собой задачу минимизации функции S . Она включает слагаемые (7) и (9) для оптимизации параметров модели (1):

$$f = \sigma_k \circ \Gamma_k \mathbf{w}_k^T \sigma_{k-1} \circ \Gamma_{k-1} \mathbf{W}_{k-1} \sigma_{k-2} \circ \dots \underbrace{\dots \circ \Gamma_2 \mathbf{W}_2 \sigma_1 \circ \Gamma_1 \mathbf{W}_1 \mathbf{x}}_{E_x} \underbrace{\dots}_{S} \quad (6)$$

Первое слагаемое E_x — это функция ошибки реконструкции объекта стеком автокодировщиков. Второе слагаемое S — это функция ошибки нейросети. При выборе моделей используются три вида критериев качества: точность, устойчивость и сложность.

Точность. В задаче восстановления регрессии функция ошибки имеет вид:

$$S = \sum_{i \in \mathcal{I}} (y_i - f(\mathbf{x}_i))^2. \quad (7)$$

При включении в модель (1) метода главных компонент или автокодировщика метки объектов не используются. Функция ошибки штрафует невязки восстановленного объекта:

$$E_x = \sum_{i \in \mathcal{I}} \|\mathbf{x}_i - \mathbf{r}(\mathbf{x}_i)\|_2^2, \quad (8)$$

где $\mathbf{r}(\mathbf{x})$ — линейная реконструкция объекта \mathbf{x} . Функция (8) с аддитивной регуляризацией:

$$E_x = \sum_{i=1}^m \|\mathbf{r}(\mathbf{x}_i, \mathbf{W}_{AE}) - \mathbf{x}_i\|^2 + \lambda^2 \|\mathbf{W}\|_{\text{Frobenius}}^2, \quad (9)$$

где m — число элементов в обучающей выборке. Параметры автокодировщика $\mathbf{W}_{AE} = \{\mathbf{W}', \mathbf{W}, \mathbf{b}', \mathbf{b}\}$ оптимизированы таким образом (8), чтобы приблизить реконструкцию $\mathbf{r}(\mathbf{x})$ к исходному вектору \mathbf{x} .

Процедура оптимизации параметров композитной функции (6):

- (1) оптимизируются параметры модели согласно (9);
- (2) заданные параметры фиксируются;
- (3) оптимизируются параметры согласно (7).

Сложность — это число параметров модели, $\sum_{l=1}^k \|\Gamma_k\|_1 \rightarrow \min$.

Устойчивость — это минимум дисперсии функции ошибки (7): $D(S) \rightarrow \min$. При вычислении устойчивости выборка считается фиксированной и изменение устойчивости считается зависящим только от структуры и параметров модели.

5 Вычислительный эксперимент

Исследуется процедура оптимизации структуры нейросети с сохранением качества аппроксимации. Структура оптимизируется с помощью генетического алгоритма. Цель вычислительного эксперимента состоит в определении оптимальной позиции разделения автокодировщиков и нейронной сети, а также исследовании зависимости точности, сложности и устойчивости модели от способа задания структуры. Исходный код находится на Github [3].

Наборы данных. Качество предложенного подхода к построению модели оценивается на нескольких реальных наборах данных и одном синтетическом наборе. Выборки взяты из открытого репозитория данных для машинного обучения [13]. Описание всех выборок представлено в таблице. Синтетический набор данных состоит из признаков с различными свойствами ортогональности и коррелированности друг с другом и с целевой переменной. Процедура генерации синтетических данных описана в работе [14]. Возможны следующие конфигурации синтетических данных: неполный и скоррелированный; адекватный и случайный; адекватный и избыточный; адекватный и скоррелированный.

Каждый набор данных разбивается на три части. Обучающая выборка — 60% от исходного набора. На этой выборке модель тренируется и фиксируются значения параметров. Валидационная выборка — 20% от исходного набора. На этой выборке применяется генетический алгоритм, который ищет оптимальную структуру. Тестовая выборка — 20% от исходного набора. Она никак не участвует

Результат применения генетического алгоритма для прореживания сети

Выборка \mathcal{D}	m	n	Ошибка сети с прореживанием	Ошибка сети без прореживания	Сложность без прореживания	Сложность после прореживания
Credit Card	30000	35	$0,3204 \pm 0,0032$	$0,2681 \pm 0,0034$	68	25
Protein	45730	9	$4,4968 \pm 0,0238$	$4,4968 \pm 0,0238$	16	1
Airbnb	10498	16	$135,0773 \pm 0,5909$	$33,9163 \pm 0,5978$	32	12
Wine quality	4898	11	$0,5818 \pm 0,0147$	$0,5941 \pm 0,0149$	20	4
Synthetic, 10^{-3}	2000	30	$0,3005 \pm 0,0081$	$0,303 \pm 0,0079$	60	12

в оптимизации структуры модели. Эта выборка используется только для контроля качества — сравнение модели исходной и оптимизированной структуры, а также сравнение с другими алгоритмами прореживания сетей. Решается задача восстановления регрессии, т. е. зависимой переменной служит $y \in \mathbb{R}$. В процессе работы были рассмотрены два подхода к решению задачи и, соответственно, две структуры нейронной сети.

Первый подход. Автокодировщик преобразует входные векторы x , которые затем подаются на вход полносвязной нейронной сети.

Второй подход. Оптимизируются параметры автокодировщика, его параметры фиксируются и предпоследний слой соединяется с полносвязной нейросетью, оптимизируются параметры модели. Предпоследний слой содержит меньшее число параметров по сравнению с размерностью входного пространства признаков. Число параметров определяет число нейронов в этом слое, т. е. полносвязная сеть получает на вход скрытое представление исходной независимой переменной.

Параметры в обеих сетях инициализированы нормальным распределением с нулевым средним и смещением $\sqrt{(1/2)(N_{in} + N_{out})}$, где N_{in} — число входных признаков; N_{out} — число признаков на выходе слоя. Такая инициализация параметров была предложена в [15]. Она выбрана экспериментальным путем как показывающая наилучший результат точности аппроксимации. Каждая из сетей обучалась в течение 500 итераций обновления пара-

метров, и размер пакета обучения равен 128. В качестве функции активации в слоях автокодировщика используется $\text{Relu}(x) = \max(0, x)$, для полносвязной сети тоже Relu, но в последнем слое id.

Для вычисления ошибки (7) используются выходные значения полносвязной сети. Варьируется число промежуточных слоев автокодировщика и полносвязной сети от 1 до 5. Рассматривается декартово произведение двух множеств: $[1, 2, 3, 4, 5] \times [1, 2, 3, 4, 5]$. Число нейронов в каждом скрытом слое одинаково для любой сети и равно десяти. Для каждой конфигурации считается ошибка (7). Качество оценивается на синтетическом наборе данных. Полученный результат представлен на рис. 1. Размер пузыря пропорционален полученной ошибке. Каждая конфигурация сети обучалась на наборе, полученном с помощью бутстреп-метода из данных, взятых для обучения. Число итераций процедуры бутстреп-метода равно 10. Ошибка считалась на отложенном наборе данных для тестирования. Видно, что при увеличении числа слоев полносвязной сети ошибка в основном падает. Минимальная ошибка достигается при конфигурации: четыре слоя автокодировщика и два слоя полносвязной сети для первого подхода; один слой автокодировщика и четыре слоя полносвязной сети для второго подхода. Данную конфигурацию возьмем для дальнейшего исследования параметров модели, используя, соответственно, второй подход.

На рис. 2 представлена дисперсия и значение ошибки (7) в зависимости от числа слоев в автоко-

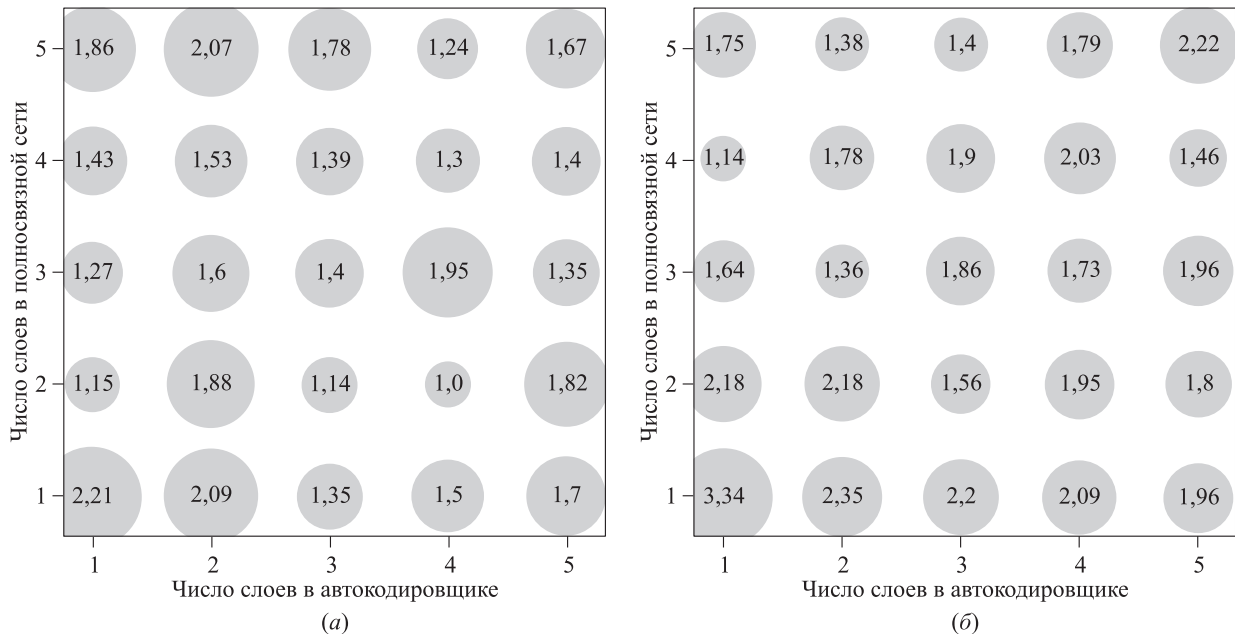


Рис. 1 Ошибка (7) в зависимости от конфигурации модели: (а) первый подход; (б) второй подход

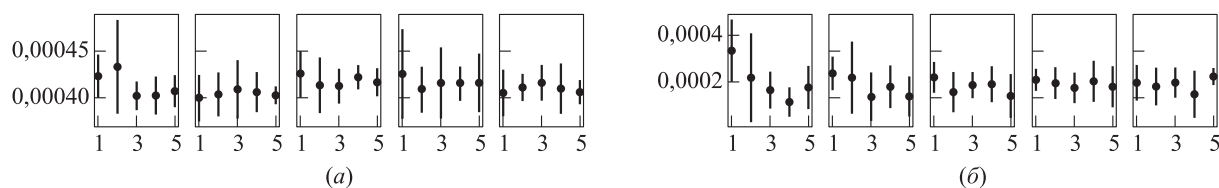


Рис. 2 Ошибка и ее дисперсия в зависимости от структуры модели: (а) первый подход; (б) второй подход. Порядковый номер графика и точки на графике соответствуют разному числу слоев соответствующей модели

дировщике и полносвязной сети. Дисперсия была получена с помощью десяти итераций бутстреппирования обучающей выборки для каждой конфигурации. С увеличением числа слоев в автокодировщике при использовании первого подхода снижаются ошибка и дисперсия ошибки.

С помощью описанных двух подходов получается оптимальная архитектура соединения автокодировщика и полносвязной сети. Далее применяется описанный ранее генетический алгоритм для прореживания сети и уменьшения ее сложности. Алгоритм применяется на нейронах каждого слоя сети. В таблице приведены результаты применения генетического алгоритма для исследуемых наборов данных, качество алгоритма оценивается по тестовой выборке. В качестве ошибки выступает $MAE = (1/m) \sum_{i=1}^m |y_i - f(\mathbf{x}_i)|$, а в качестве сложности алгоритма выступает число ненулевых нейронов. Под нулевым нейроном понимается нейрон, все параметры которого равны нулю.

6 Заключение

В представленной работе исследованы два подхода к построению модели, состоящей из автокодировщика и нейронной сети и имеющей композиционную функцию ошибки. Представлены подходы к поиску оптимальной точки разделения автокодировщика и нейронной сети. Исследовано применение генетического алгоритма для оптимизации структуры и снижения сложности. Работа предложенного алгоритма исследовалась на пяти различных наборах данных. Как показано в таблице, предложенный алгоритм выбора структуры существенно снижает сложность модели без потери качества аппроксимации.

Литература

1. *Cybenko G. V.* Approximation by superpositions of a sigmoidal function // *Math. Control Signal.*, 1989. Vol. 2. Iss. 4. P. 303–314.

2. *Бахтеев О. Ю., Стрижов В. В.* Выбор моделей глубокого обучения субоптимальной сложности // *Автоматика и телемеханика*, 2018. Вып. 8. С. 129–147.
3. *Потанин М. С., Вайсер К. О., Жолобов В. А., Стрижов В. В.* Приложение к статье: вычислительный эксперимент по выбору универсальной модели и базовые теоремы, 2019. <https://github.com/MarkPotanin/GeneticOpt>.
4. *Hinton G. E., Salakhutdinov R. R.* Reducing the dimensionality of data with neural networks // *Science*, 2006. Vol. 313. Iss. 5786. P. 504–507.
5. *LeCun Y., Denker J. S., Solla S., Howard R. E., Jackel L. D.* Optimal brain damage // *Adv. Neur. In.*, 1989. Vol. 2. P. 598–605.
6. *Hassibi B., Stork D. G.* Second order derivatives for network pruning: Optimal brain surgeon // *Adv. Neur. In.*, 1992. Vol. 5. P. 164–171.
7. *Dong X., Chen S., Pan S.* Learning to prune deep neural networks via layer-wise optimal brain surgeon // *Adv. Neur. In.*, 2017. Vol. 30. P. 4857–4867.
8. *Molchanov P., Tyree S., Karras T., Aila T., Kautz J.* Pruning convolutional neural networks for resource efficient transfer learning // *ArXiv.org*, 2016. ArXiv:1611.06440.
9. *Chaber P., Lawrynczuk M.* Pruning of recurrent neural models: An optimal brain damage approach // *Nonlinear Dynam.*, 2018. Vol. 92. Iss. 2. P. 763.
10. *Elsken T., Metzen J. H., Hutter F.* Neural architecture search: A survey // *ArXiv.org*, 2018. ArXiv:1808.05377.
11. *Hutter F., Kotthoff L., Vanschoren J.* Automated machine learning – methods, systems, challenges. — Springer, 2019. 223 p.
12. *Pao C. P.* Линейные статистические методы и их применение / Пер. с англ. — М.: Наука, 1968. 548 с. (*Rao C. R.* Linear statistical inference and its application. — New York, NY, USA: Wiley, 1968. 548 p.)
13. *UCI Machine Learning Repository*, 2007. <https://archive.ics.uci.edu/ml>.
14. *Katrusa A. M., Strijov V. V.* Stress test procedure for feature selection algorithms // *Chemometr. Intell. Lab.*, 2015. Vol. 142. P. 172–183.
15. *Glorot X., Bengio Y.* Understanding the difficulty of training deep feedforward neural networks // 13th Conference (International) on Artificial Intelligence and Statistics Proceedings. — Sardinia, Italy, 2010. P. 249–256.

Поступила в редакцию 02.12.19

DEEP LEARNING NEURAL NETWORK STRUCTURE OPTIMIZATION

M. S. Potanin¹, K. O. Vajser¹, V. A. Zholobov¹, and V. V. Strijov^{1,2}

¹Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation

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

Abstract: The paper investigates the optimal model structure selection problem. The model is a superposition of generalized linear models. Its elements are linear regression, logistic regression, principal components analysis, autoencoder and neural network. The model structure refers to values of structural parameters that determine the form of final superposition. This paper analyzes the model structure selection method and investigates dependence of accuracy, complexity and stability of the model on it. The paper proposes an algorithm for selection of the neural network optimal structure. The proposed method was tested on real and synthetic data. The experiment resulted in significant structural complexity reduction of the model while maintaining accuracy of approximation.

Keywords: model selection; linear models; autoencoders; neural networks; structure; genetic algorithm

DOI: 10.14357/19922264200408

Acknowledgments

This research was supported by RFBR (projects 19-07-1155 and 19-07-0885) and by the Government of the Russian Federation (agreement 05.Y09.21.0018). 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. Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Math. Control Signal.* 2(4):303–314.
2. Bakhteev, O. Yu, and V. V. Strijov. 2018. Deep learning model selection of suboptimal complexity. *Automat. Rem. Contr.* 79(8):1474–1488.
3. Potanin, M. S., K. O. Vajser, V. A. Zholobov, and V. V. Strijov. 2019. Prilozhenie k stat'e: vychislitel'nyy eksperiment po vyboru universal'noy modeli i bazovye teoremy [Appendix to the paper: Computational experiment and basic theorems]. Available at: <https://github.com/MarkPotanin/GeneticOpt> (accessed November 5, 2020).
4. Hinton, G. E., and R. R. Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507.
5. LeCun, Y., J. Denker, and S. Solla. 1989. Optimal brain damage. *Adv. Neur. In.* 2:598–605.
6. Hassibi, B., and D. G. Stork. 1992. Second order derivatives for network pruning: Optimal brain surgeon. *Adv. Neur. In.* 5:164–171.
7. Dong, X., S. Chen, and S. Pan. 2017. Learning to prune deep neural networks via layer-wise optimal brain surgeon. *Adv. Neur. In.* 30:4857–4867.
8. Molchanov, P., S. Tyree, T. Karras, T. Aila, and J. Kautz. 2016. Pruning convolutional neural networks for resource efficient transfer learning. arXiv:1611.06440 [cs.LG]. Available at: <https://arxiv.org/abs/1611.06440> (accessed November 5, 2020).
9. Chaber, P., and M. Lawryńczuk. 2018. Pruning of recurrent neural models: an optimal brain damage approach. *Nonlinear Dynam.* 92(2):763–780.
10. Elsken, T., J. H. Metzen, and F. Hutter. 2018. Neural architecture search: A survey. arXiv:1808.05377 [stat.ML]. Available at: <https://arxiv.org/abs/1808.05377> (accessed November 5, 2020).
11. Hutter, F., L. Kotthoff, and J. Vanschoren. 2019. *Automated machine learning-methods, systems, challenges*. Springer. 223 p.
12. Rao, C. R. 1973. *Linear statistical inference and its applications*. Vol. 2. New York, NY: Wiley. 548 p.
13. UCI Machine Learning Repository. Available at: <http://archive.ics.uci.edu/ml> (accessed November 5, 2020).
14. Katrutsa, A. M., and V. V. Strijov. 2015. Stress test procedure for feature selection algorithms. *Chemometr. Intell. Lab.* 142:172–183.
15. Glorot, X., and Yo. Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. *13th Conference (International) on Artificial Intelligence and Statistics Proceedings*. 249–256.

Received December 2, 2019

Contributors

Potanin Mark St. (b. 1994) — PhD student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; mark.potanin@phystech.edu

Vayser Kirill O. (b. 2000) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; vajser.ko@phystech.edu

Zholobov Vladimir Al. (b. 1998) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141701, Russian Federation; zholobov.va@phystech.edu

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

ПОВЫШЕНИЕ СБОЕУСТОЙЧИВОСТИ САМОСИНХРОННЫХ СХЕМ*

И. А. Соколов¹, Ю. А. Степченков², Ю. Г. Дьяченко³, Ю. В. Рождественский⁴

Аннотация: Анализируется проблема устойчивости самосинхронных (СС) схем, изготовленных по технологии комплементарный металл–диэлектрик–полупроводник (КМДП), к кратковременным логическим сбоям (ЛС), генерируемым внешними воздействиями: ядерными частицами, космическими лучами, электромагнитными наводками. Практические СС-схемы реализуются в виде конвейера с запрос-ответным взаимодействием между его ступенями и двухфазной дисциплиной работы с чередованием рабочей фазы и спейсера. Комбинационная часть ступени конвейера использует парафазное со спейсером кодирование информационных сигналов. Индикаторная подсхема ступени конвейера подтверждает окончание переключения всех элементов ступени, возбужденных в текущей фазе работы, и формирует сигналы управления запрос-ответным взаимодействием ступеней конвейера. Рассмотрены физические причины появления ЛС и проанализированы типы сбоев, возможных в КМДП-СС-схемах с проектными нормами 65 нм и ниже. Сравниваются характеристики сбоеустойчивости разных вариантов СС-регистров хранения. Предлагаются схемотехнические и топологические методы повышения сбоеустойчивости СС-конвейера. Даются оценки сбоеустойчивости СС-конвейера в зависимости от места появления ЛС.

Ключевые слова: самосинхронная схема; сбоеустойчивость; конвейер; рабочая фаза; спейсер

DOI: 10.14357/19922264200409

1 Введение

Широкое использование цифровых микросхем в условиях неблагоприятной радиационной обстановки и переход к субмикронным технологиям их изготовления сделали актуальной задачу повышения устойчивости цифровых микросхем к ЛС. Логический сбой — это изменение уровня сигнала в узле схемы из-за кратковременной причины — пролета через тело полупроводника микросхемы одинокой ядерной частицы (ЯЧ), мощного электромагнитного импульса, сильной помехи по шинам питания или по сигнальным линиям и т. д.

В комбинационных схемах, находящихся в статическом состоянии, ЛС прекращается сам собой по окончании действия физической причины его появления. Но в конвейерных синхронных схемах даже кратковременный ЛС может успеть записаться в выходной регистр и испортить результат обработки данных. Повышение быстродействия цифровых схем усугубляет эту проблему.

Элементы с памятью (триггеры, ячейки памяти) более чувствительны к ЛС, поскольку ЛС в них может инвертировать хранимый бит данных, который самостоятельно не восстановится после исчезновения причины сбоя.

Самосинхронные схемы [1] обладают высокой устойчивостью к ЛС [2] за счет избыточного кодирования информационных сигналов и запрос-ответной дисциплины взаимодействия функциональных СС-блоков. Однако индикаторная часть СС-схем традиционно реализуется на элементах с памятью — гистерезисном триггере (Г-триггере [1]). Кроме того, практические СС-схемы имеют конвейерную реализацию, аналогично синхронным аналогам, с регистром в каждой ступени конвейера. В результате уровень сбоеустойчивости СС-схем зависит и от устойчивости к ЛС Г-триггера и разряда регистра хранения данных. Поэтому задача повышения устойчивости СС-конвейера к кратковременным одиночным ЛС является актуальной.

Данная статья анализирует естественную сбоеустойчивость СС-конвейера в КМДП-базисе с проектными нормами не более 65 нм и предлагает схемотехнические методы ее повышения.

2 Типы логических сбоев в самосинхронных схемах

Физической причиной ЛС в КМДП-схемах служит индуцирование избыточных носителей заряда

* Исследование выполнено при финансовой поддержке Минобрнауки (проект 075-15-2020-799).

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ISokolov@ipiran.ru

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

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук, diaura@mail.ru

⁴Федеральный исследовательский центр «Информатика и управление» Российской академии наук, YRogdest@ipiran.ru

Таблица 1 Возможные изменения ПФС сигнала из-за ЛС

№ п/п	Спейсер «00»		Спейсер «11»	
	До ЛС	После ЛС	До ЛС	После ЛС
1	00	01	11	01
2	00	10	11	10
3	00	11	11	00
4	01	11	01	00
5	10	11	10	00
6	01	00	01	11
7	10	00	10	11

(электронно-дырочных пар) в теле полупроводника и сигнальных трассах из-за внешнего воздействия или сильных помех. Под действием электрического поля электроны и дырки в полупроводнике разлетаются в противоположных направлениях, порождая ток ионизации (ТИ) в узле схемы. В первом приближении импульс ТИ описывается формулой [3]:

$$I_{ТИ}(t) = \frac{Qk}{\tau_{сп} - \tau_{н}} \left(e^{-t/\tau_{сп}} - e^{-t/\tau_{н}} \right),$$

где Q — интегральный заряд, образовавшийся в объеме полупроводника; $\tau_{н}$ и $\tau_{сп}$ — постоянные времени нарастания и спада импульса ТИ; k — коэффициент, характеризующий часть общего заряда Q , попавшего в данный узел схемы. Интегральный заряд Q оценивается по формуле:

$$Q = \frac{q\rho L_{тр}\theta LET}{E_{eh}},$$

где q — заряд электрона; ρ — плотность полупроводника; $L_{тр}$ — длина трека ЯЧ в полупроводнике; θ — угол падения частицы; LET — потери энергии частицы; E_{eh} — энергия образования электронно-дырочной пары.

В наихудшем случае ТИ приводит к кратковременной инверсии логического уровня на выходе сбойного элемента на время рассасывания заряда Q , но следующая часть СС-схемы может воспринять ЛС и зафиксировать в регистре.

Сильные помехи по сигнальным линиям и шинам питания и земли за счет паразитных емкостных связей наводят положительные или отрицательные импульсы напряжения на соседние трассы. При определенных условиях (амплитуде помехи, соотношении паразитных емкостей трасс-«агрессоров» и трассы-«жертвы») перепад напряжения может инвертировать логический уровень на трассе-«жертве».

В комбинационных СС-схемах информационные сигналы представлены в парафазном коде со спейсером [1]. Парафазный сигнал (ПФС) формируется парой дуальных логических ячеек. В работе [2] было показано, что в СС-схемах, изготовленных по объемной КМДП-технологии с проектными нормами 65 нм и ниже, при надлежащем размещении в топологии дуальных ячеек и трасс ПФС ЛС может привести к изменению текущего состояния ПФС с нулевым («00») или единичным («11») спейсером в соответствии с табл. 1.

Однако свойства СС-схем и индикация состояния, инверсного спейсера (антиспейсера, АС), как второго спейсера [2] маскируют часть ЛС.

3 Сбоеустойчивость самосинхронных схем

Вероятность распространения ЛС в СС-схеме зависит от ряда факторов: типа ЛС; текущей фазы (рабочая фаза или спейсер); времени появления ЛС до момента срабатывания индикаторных выходов схемы; маскирования сбойного ПФС остальными сигналами СС-схемы; длительности ЛС; места появления ЛС.

На практике СС-схемы реализуются в виде конвейера, пример которого показан на рис. 1. Здесь Ст1, Ст2 и Ст3 — ступени конвейера; Г — Г-триггер, формирующий фазовый сигнал управления регистром.

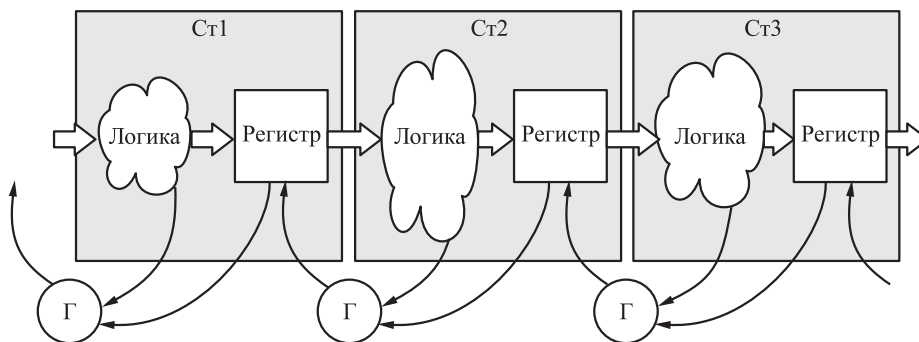


Рис. 1 Схема СС-конвейера

Вероятность распространения кратковременного одиночного ЛС в СС-конвейере в первом приближении рассчитывается по формуле:

$$P_{\text{ЛС}} = P_{\text{РФ}} [(P_{\text{вх}}P_{\text{зн1}} + P_{4\text{-}7}P_{\text{зн2}}) P_{\text{ЗР1}} + P_{\text{И1}}] + P_{\text{СФ}} [(P_{\text{вх}}P_{\text{зн3}} + P_{1\text{-}3}P_{\text{зн4}}) P_{\text{ЗР2}} + P_{\text{И2}}], \quad (1)$$

где $P_{\text{РФ}}$ и $P_{\text{СФ}}$ — вероятности пребывания СС-схемы в рабочей фазе и в спейсере в момент появления ЛС; $P_{\text{вх}}$ — вероятность того, что данный ЛС появился в выходном регистре предыдущей ступени; $P_{\text{ЗР1}}, \dots, P_{\text{ЗР4}}$ — вероятности того, что данный ЛС не будет замаскирован логикой ступени; $P_{1\text{-}3}$ и $P_{4\text{-}7}$ — вероятности принадлежности ЛС к типам 1–3 или 4–7 из табл. 1 соответственно; $P_{\text{ЗР1}}$ и $P_{\text{ЗР2}}$ — вероятности записи ЛС в выходной регистр ступени; $P_{\text{И1}}$ и $P_{\text{И2}}$ — вероятности того, что данный ЛС появился в индикаторной части ступени.

Формула (1) дает оценку сверху. Она учитывает вклад трех частей ступени СС-конвейера: комбинационной части (КЧ), индикаторной части (ИЧ) и выходного регистра (ВР), — но не учитывает времени появления ЛС в рабочем цикле СС-схемы и соотношений длительностей ЛС, рабочей и спейсерной фаз разных ступеней конвейера.

Сбоеустойчивость СС-конвейера зависит от сбоеустойчивости всех его частей: КЧ, ИЧ и ВР. Рассмотрим каждую из них, оценивая их устойчивость с помощью вероятностного подхода [2]. Критическими считаются ЛС, приводящие к искажению обрабатываемых данных или остановке конвейера.

3.1 Комбинационная часть ступени самосинхронного конвейера

Анализ влияния одиночного кратковременного ЛС, возникшего в КЧ ступени СС-конвейера, на работоспособность конвейера показал, что благодаря двухфазной дисциплине работы, парафазному со спейсером кодированию информационных сигналов и индикации окончания переключения всех элементов схемы, возбужденных на данной фазе, КЧ конвейера устойчива к 85,5% ЛС [2]. Схематехнические и топологические методы, предложенные в работе [2], повышают устойчивость КЧ СС-конвейера к одиночным ЛС, приведенным в табл. 1, до уровня 98,9% при условии индикации АС как второго спейсера.

3.2 Регистр ступени самосинхронного конвейера

В СС-конвейерах с ПФС разряд регистра ступени традиционно реализуется на двух Г-триггерах

и индикаторном элементе [4], обеспечивая хранение рабочего состояния и спейсера ПФС при минимальных аппаратных затратах. Но такая реализация обладает низкой сбоеустойчивостью.

Возможные варианты разряда СС-регистра хранения представлены на рис. 2. Все они имеют парафазные вход (R, S) и выход (Q, QV) с нулевым спейсером, сигнал разрешения записи (E), регулирующий фазовые переходы регистра, и индикаторный выход (I). Рисунок 3 иллюстрирует реализацию Г-триггера G12AT (из разряда регистра на рис. 2, ж) с нулевым спейсером [5], устойчивого к АС на входе.

Таблица 2 содержит оценки устойчивости разрядов СС-регистра хранения, приведенных на рис. 2, к ЛС из табл. 1 и показатели сложности их реализации в КМДП-транзисторах. Сравнение характеристик вариантов разряда СС-регистра хранения показывает, что максимально защищенным от ЛС оказывается разряд вида рис. 2, а. Наилучшим отношением показателя сбоеустойчивости к сложности реализации обладает разряд вида рис. 2, ж.

Таким образом, классический Г-триггер не рекомендуется использовать для реализации разряда регистра СС-конвейера, предназначенного для эксплуатации в условиях активных неблагоприятных внешних воздействий.

3.3 Индикаторная часть ступени самосинхронного конвейера

Индикаторная часть — наиболее чувствительная часть СС-конвейера. Критическая ситуация возможна при преждевременном переключении Г-триггера, формирующего сигнал управления регистром ступени. В результате регистр может раньше времени перейти в спейсер (рабочее состояние) и тем самым помешать правильному переключению следующей ступени конвейера.

Однако использование сбоеустойчивой DICE-подобной реализации Г-триггеров [2] обеспечивает их абсолютную устойчивость к ЛС, указанным в табл. 1, и предотвращает появление критической ситуации в конвейере. В результате ЛС может привести лишь к временной приостановке конвейера: по окончании ЛС конвейер продолжит нормальную работу.

3.4 Общая сбоеустойчивость самосинхронного конвейера

В реальных СС-схемах КЧ ступени конвейера обычно имеет площадь топологической реализации в несколько раз больше, чем регистр ступени. Пусть, например, это соотношение равно двум. Площадь ИЧ также меньше площади КЧ примерно

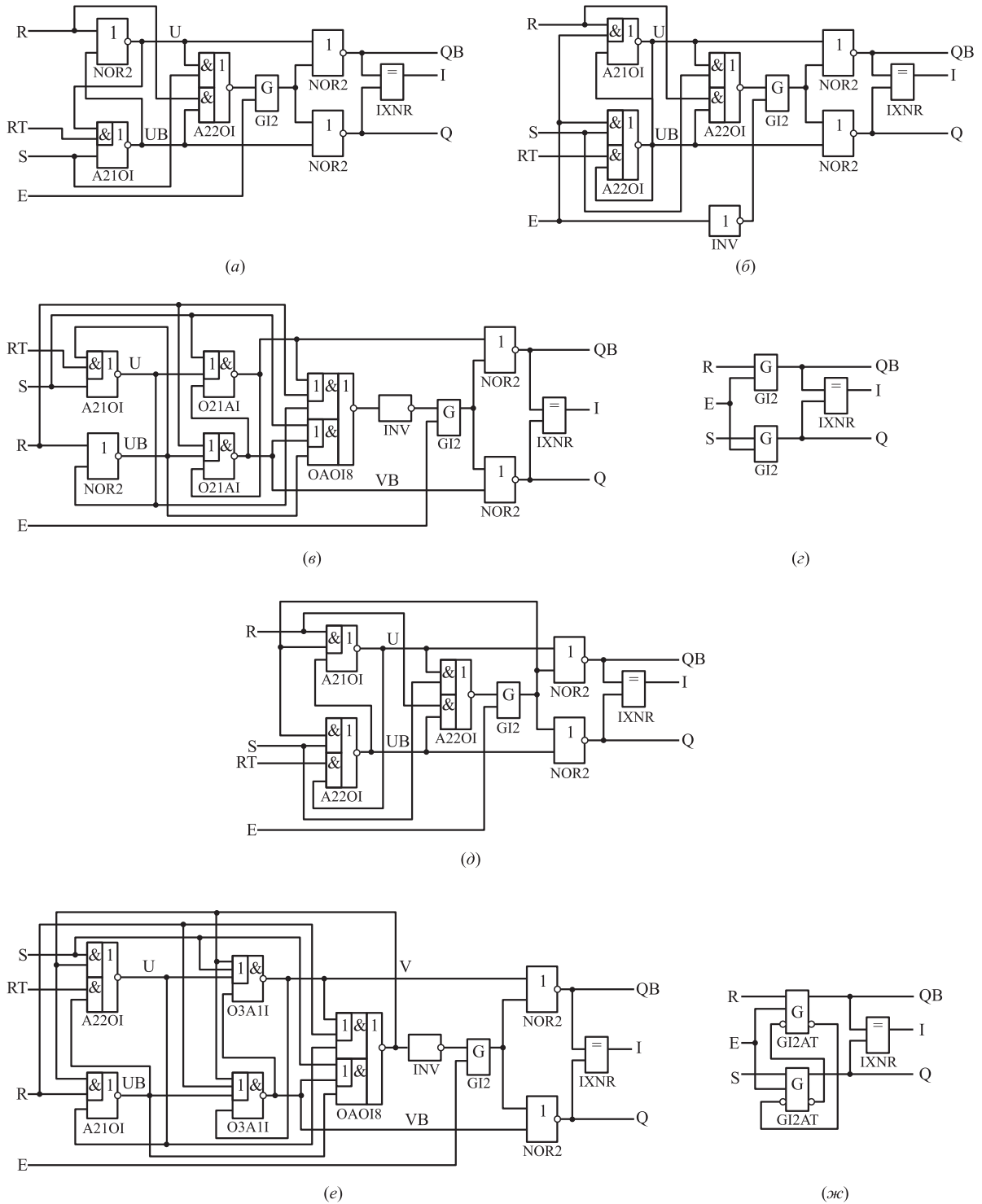


Рис. 2 Варианты сбоеустойчивого разряда регистра СС-конвейера

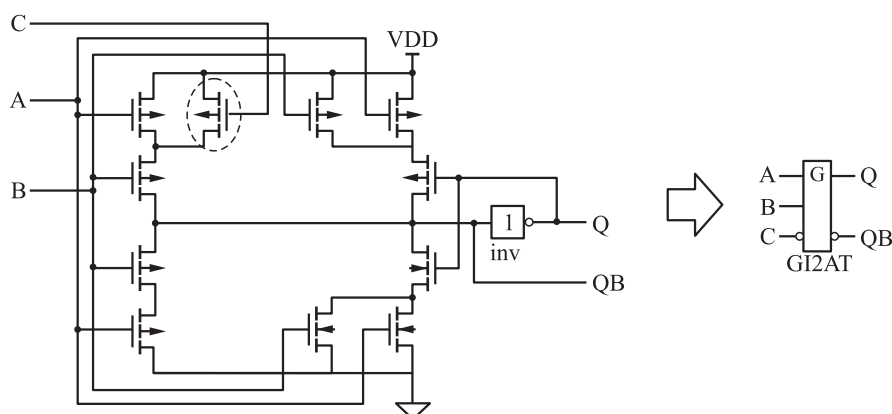


Рис. 3 Схема Г-триггера, защищенного от АС на входе

Таблица 2 Вероятность внесения ошибки в обрабатываемые данные из-за ЛС в ступени конвейера

№ п/п	Тип разряда регистра	Устойчивость к ЛС, %			Число транзисторов
		В КЧ	В регистре	Общая	
1	Рис. 2, а	98,76	93,62	97,78	44
2	Рис. 2, б	97,66	96,27	98,04	54
3	Рис. 2, в	89,84	89,99	92,42	62
4	Рис. 2, г	86,36	58,59	82,83	30
5	Рис. 2, д	97,92	94,53	97,59	48
6	Рис. 2, е	98,18	92,43	97,20	70
7	Рис. 2, ж	97,88	91,41	96,76	32

в 2 раза. Тогда при равномерном распределении сбоев по площади СБИС вероятность их появления в КЧ в 2 раза больше вероятности их появления в регистре и ИЧ. С учетом отсутствия критических ситуаций при ЛС в ИЧ вероятность распространения ЛС по СС-конвейеру для разных вариантов реализации разряда регистра будет соответствовать результатам, приведенным в табл. 2 в графе «общая устойчивость».

4 Заключение

Индикация антиспейсера ПФС как второго спейсера и использование DICE-подобного Г-триггера с «самолечением» АС обеспечивают существенное повышение сбоеустойчивости комбинационной и индикаторной частей СС-конвейера.

Классический вариант разряда регистра СС-конвейера на обычных Г-триггерах имеет сравнительно низкую сбоеустойчивость — на уровне 83%.

Наилучшую сбоеустойчивость по отношению к кратковременным одиночным ЛС демонстрирует вариант СС-конвейера, регистр в котором реализуется на одноканальном RS-триггере с разрешени-

ем записи. Он устойчив к 98% ЛС в ступени конвейера.

Литература

1. Kishinevsky M., Kondratyev A., Taubin A., Varshavsky V. Concurrent hardware: The theory and practice of self-timed design. — New York, NY, USA: J. Wiley & Sons, 1994. 368 p.
2. Stepchenkov Y. A., Kamenskih A. N., Diachenko Y. G., Rogdestvenski Y. V., Diachenko D. Y. Improvement of the natural self-timed circuit tolerance to short-term soft errors // Advances Science Technology Engineering Systems J., 2020. Vol. 5. Iss. 2. P. 44–56.
3. Nicolaidis M. Soft errors in modern electronic systems. — New York, NY, USA: Springer, 2011. 316 p.
4. Степченко Ю. А., Дьяченко Ю. Г., Рождественский Ю. В., Морозов Н. В., Степченко Д. Ю., Рождественская А. В., Сурков А. В. Самосинхронный умножитель с накоплением: варианты реализации // Системы и средства информатики, 2014. Т. 24. № 3. С. 63–77.
5. Соколов И. А., Захаров В. Н., Степченко Ю. А., Дьяченко Ю. Г. Устройство сбоеустойчивого разряда самосинхронного регистра хранения: Заявка на изобретение № 2020140031 от 28.02.20.

Поступила в редакцию 13.03.2020

IMPROVEMENT OF SELF-TIMED CIRCUIT SOFT ERROR TOLERANCE

I. A. Sokolov, Yu. A. Stepchenkov, Yu. G. Diachenko, and Yu. V. Rogdestvenski

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

Abstract: The paper considers a tolerance of self-timed (ST) circuits fabricated with complementary metal–oxide–semiconductor (CMOS) process to short-term soft errors generated by external causes, namely, nuclear particles, cosmic rays, electromagnetic pulses, and noises. Pipeline implementation is usual for practical ST-circuits. Its control bases on handshake between pipeline stages and two-phase operation discipline with a sequence of the working phase and spacer one. Combinational part of the pipeline stage uses dual-rail information signal coding with a spacer. The pipeline stage indication part acknowledges a switching completion of all stage cells, fired at the current operation phase, and generates handshake signals in ST-pipeline stages control. The paper discusses the physical causes of the short-term soft errors. It analyzes soft error types that may appear in CMOS ST-circuits fabricated with 65-nanometer and below standard bulk process. The tolerance level of the proposed soft error hardened ST-register bits is discussed and compared. The paper suggests circuitry and layout techniques improving ST-pipeline soft error tolerance and estimates soft error immunity level for all pipeline parts depending on soft error location.

Keywords: self-timed circuit; tolerance; pipeline; working phase; spacer

DOI: 10.14357/19922264200409

Acknowledgments

The research was supported by the Ministry of Science and Higher Education of the Russian Federation, project No. 075-15-2020-799.

References

1. Kishinevsky, M., A. Kondratyev, A. Taubin, and V. Varshavsky. 1994. *Concurrent hardware: The theory and practice of self-timed design*. New York, NY: J. Wiley & Sons. 368 p.
2. Stepchenkov, Y. A., A. N. Kamenskih, Y. G. Diachenko, Y. V. Rogdestvenski, and D. Y. Diachenko. 2020. Improvement of the natural self-timed circuit tolerance to short-term soft errors. *Advances Science Technology Engineering Systems J.* 5(2):44–56.
3. Nicolaidis, M. 2011. *Soft errors in modern electronic systems*. New York, NY: Springer. 316 p.
4. Stepchenkov, Y., Y. Diachenko, Y. Rogdestvenski, N. Morozov, D. Stepchenkov, A. Rogdestvenskene, and A. Surkov. 2014. Samosinkhronny umnozhitel' s nakopleniem: varianty realizatsii [Self-timed fused multiply-add unit: Implementation variants]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(3):63–77.
5. Sokolov, I. A., V. N. Zakharov, Yu. A. Stepchenkov, and Yu. G. Diachenko. 2020. Ustroystvo sboeustoychivogo razryada samosinkhronnogo registra khraneniya [Fault-tolerance bit device for self-timed storage register]. Application for Patent RF No. 2020140031.

Received March 13, 2020

Contributors

Sokolov Igor A. (b. 1954) — Doctor of Science in technology, Academician of RAS, director, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; isokolov@ipiran.ru

Stepchenkov Yuri A. (b. 1951) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119133, Russian Federation; YStepchenkov@ipiran.ru

Diachenko Yuri G. (b. 1958) — Candidate of Science (PhD) in technology, senior scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44- 2 Vavilov Str., Moscow 119133, Russian Federation; diaura@mail.ru

Rogdestvenski Yuri V. (b. 1952) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation, Moscow 119333, Russian Federation; YRogdest@ipiran.ru

ОБ ОДНОМ ПОДХОДЕ К ФОРМИРОВАНИЮ В УСЛОВИЯХ ВЫСОКОЙ НЕОПРЕДЕЛЕННОСТИ МАРКЕРОВ КОНФИДЕНЦИАЛЬНОСТИ В СИСТЕМАХ ИНТЕНСИВНОГО ИСПОЛЬЗОВАНИЯ ДАННЫХ*

В. И. Будзко¹, В. В. Ядринцев², И. В. Соченков³, В. И. Королёв⁴, В. Г. Беленков⁵

Аннотация: Основные задачи, результаты решения которых отражены в статье, связаны с формированием маркеров конфиденциальности (МК) при их применении в системах интенсивного использования данных (СИИД) в условиях, когда состав и структура охраняемых сведений не может быть заранее определена в связи с отсутствием данных или с высокой динамикой их изменения либо их определение нецелесообразно в связи с большим (или неограниченным) количеством сущностей, сведения о которых подлежат охране. В данной работе предложен подход к формированию в указанных условиях МК текстовых материалов. Сформулирована логика процесса семантической обработки текста, позволяющего в условиях высокой неопределенности состава и структуры охраняемых сведений формировать МК при их применении для обеспечения информационной безопасности в СИИД. Полученные экспериментальные результаты позволяют перейти к практической реализации рассмотренного подхода в СИИД.

Ключевые слова: маркер конфиденциальности; информационная безопасность; интенсивное использование данных; кластер; семантика; технические средства обеспечения безопасности; интеллектуальные задачи безопасности; текстовая классификация; обнаружение текстовых заимствований

DOI: 10.14357/19922264200410

1 Введение

Рассмотрение проблематики формирования МК текстовых материалов при их применении для обеспечения информационной безопасности (далее — формирование МК) в СИИД (Data Intensive Domains — DID) продолжает прикладные исследования, результаты которых отражены авторами в работах [1–7].

Рассматриваемый в работе возможный подход к формированию МК основан на использовании методов анализа содержания текстов и не зависит от конфиденциальности материала. Также выполнено экспериментальное исследование, в рамках которого промоделирована применимость методов анализа текстов к решению задач предотвращения утечек информации ограниченного доступа.

При рассмотрении подхода используются следующие основные определения и понятия:

Системы интенсивного использования данных — автоматизированные информационные системы,

обеспечивающие анализ и управление данными, обработку информации и решение функциональных исследовательских и прикладных задач в различных областях с DID.

Артефакт — обособленный каким-либо образом контент или фрагмент контента, рассматриваемый как единое целое совместно с его атрибутикой: учетными и иными сопровождающими данными.

Сущность — именование конкретной области деятельности, либо принадлежащего ей объекта, либо показателя/характеристики/вида сведений, упоминание или конкретная информация о которых подлежат защите.

Маркер конфиденциальности — метка какого-либо рода (например, слово), в явном виде определяющая необходимость защиты и уровень конфиденциальности сведений фрагмента или конкретного артефакта в целом.

Частный маркер конфиденциальности — метка какого-либо рода (например, слово, текст, агрегат

*Работа выполнена при частичной финансовой поддержке РФФИ (проект 18-29-03215), экспериментальные исследования выполнены при поддержке Программы РУДН «5-100».

¹Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vbudzko@ipiran.ru

²Федеральный исследовательский центр «Информатика и управление» Российской академии наук; Российский университет дружбы народов, vvyadrincev@gmail.com

³Федеральный исследовательский центр «Информатика и управление» Российской академии наук, sochenkov@isa.ru

⁴Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vkorolev@ipiran.ru

⁵Федеральный исследовательский центр «Информатика и управление» Российской академии наук, vbelenkov@ipiran.ru

текстов и/или данных), в явном виде определяющая необходимость защиты и уровень конфиденциальности сведений, относящихся к конкретному упоминанию сущности в контенте артефакта.

Маркер конфиденциальности относится к результату обработки запроса в целом или к его разделам — объектам доступа для пользователя, выдавшего запрос (далее — объекты доступа). Они отражают наличие в результате или в его разделах следующих сведений, которые относятся или могут относиться к охраняемым сведениям в этих областях деятельности:

- информация о конкретных областях деятельности;
- информация о конкретном объекте;
- информация о показателях/характеристиках конкретного объекта или конкретной области деятельности.

Маркер конфиденциальности также отражает уровни доступности сведений, содержащихся в результате обработки запроса или в его разделах, относящихся:

- к конкретным областям деятельности или к конкретным объектам;
- в целом к результату обработки запроса или к его разделу.

Далее в статье будет представлен процесс семантической обработки текста, позволяющий в условиях высокой неопределенности состава и структуры охраняемых сведений формировать МК при их применении для обеспечения информационной безопасности в СИИД. Для решения задач выявления конфиденциального содержимого предлагается использовать методы поиска тематически похожих документов на основе автоматически формируемых наборов МК, а также методов поиска семантически сходных фрагментов текстов. Оба предложенных подхода экспериментально проверены в ходе настоящего исследования.

2 Формирование маркеров и определение уровня конфиденциальности материала на основе анализа его текста

Последовательность действий по формированию маркеров и определению уровня конфиденциальности материала на основе анализа его текста

приведена в [6]. Она включает следующие основные шаги:

- (1) приведение артефакта к единому формату представления (например, txt);
- (2) определение в артефакте или его фрагменте позиций именованных сущностей;
- (3) приведение артефакта или его фрагмента к единому именованию сущностей;
- (4) замена личных местоимений на именованные сущности;
- (5) проверка принадлежности и формирование перечня сущностей, упоминания или конкретные сведения о которых подлежат защите;
- (6) определение отнесения фрагментов текстов к именованным/упоминаниям конкретных показателей/характеристик/видов сведений;
- (7) определение уровней конфиденциальности сведений об объектах и областях деятельности;
- (8) определение уровня конфиденциальности артефакта/его фрагмента;
- (9) формирование МК артефакта/его фрагмента.

Указанный процесс не позволяет определить МК текста в условиях, когда состав и структура охраняемых сведений:

- не может быть заранее определена в связи с отсутствием данных;
- их определение нецелесообразно в связи с высокой динамикой их изменения либо в связи с большим (или неограниченным) количеством сущностей, сведения о которых подлежат охране.

В указанных условиях для формирования маркеров конфиденциальности текстовых материалов должен быть использован иной подход, например основанный на смысловом сопоставлении конфиденциальных материалов с текстами, передаваемыми в сообщениях. Рассмотрим далее эти подходы более подробно.

3 Формирование маркеров конфиденциальности текстов на основе определения их близости к одному из кластеров конфиденциальных материалов

Суть предлагаемого подхода состоит в прямом (не пошаговом) формировании МК текста в целом

без разрешения в явном виде неопределенности состава и структуры сведений, охраняемых в конкретной области (областях) деятельности.

Для этого заблаговременно в каждой рассматриваемой области деятельности:

- по каждой коллекции, содержащей охраняемые сведения одного уровня конфиденциальности, осуществляется формирование кластеров. С каждым кластером связывается набор ключевых слов и словосочетаний (НКСС) и МК, соответствующий уровню конфиденциальности;
- при определении МК текста осуществляется проверка его близости или принадлежности к одному из кластеров (близости к его НКСС); тексту присваивается МК, определенный для кластера, к которому принадлежит /наиболее близок этот текст; при отсутствии однозначности в определении принадлежности текста к одному из кластеров тексту присваивается МК, отвечающий НКСС с наивысшим уровнем конфиденциальности.

Подход хорошо согласуется с условиями обработки сведений, не относящихся к высшим уровням конфиденциальности (например, с обработкой сведений, не составляющих государственную тайну).

В работе [8] описан способ обучения классификатора, предназначенного для определения категории документа в задаче предотвращения утечек информации (*англ.* DLP — Data Leak Prevention); DLP-система должна быть способна определить принадлежность документа к классу конфиденциальных документов (или других, в зависимости от области применения системы). В патенте на изобретение [9] описаны система и способ определения текста, содержащего конфиденциальные данные. В основе способа лежит поиск ключевых слов и вычисление их плотности. Способ предназначен для выявления шаблонных конфиденциальных документов.

Также отметим, что DLP-системам важно уметь предотвращать утечку дубликатов документов. Для выявления таких документов предлагается использовать методы поиска текстовых заимствований [10, 11], дополненные методами машинного обучения для тематической классификации текстов.

Основная идея применения методов машинного обучения и анализа текстовой информации заключается в следующем. Предполагается, что организация имеет определенное множество конфиденциальных документов. Это множество может быть динамически пополняемым, при этом DLP-

система дополнительно индексирует вновь регистрируемые конфиденциальные документы.

На этапе фильтрации система сопоставляет с помощью эффективных методов анализа текстов те материалы, которые передаются сотрудниками организации по каналам связи, контролируемых DLP-системой. При выявлении конфиденциальной по тематике и содержанию информации в исходящем сообщении регистрируется соответствующее нарушение политики конфиденциальности и создается оповещение. Само сообщение в таком случае может быть заблокировано (не передано).

Для *определения близости или принадлежности текста* к одному или нескольким кластерам целесообразно использовать подход машинного обучения, описанный в разд. 4. В разд. 5 будут приведены результаты экспериментов по применению методов выявления текстовых заимствований (в том числе перефразированных) для предотвращения утечек конфиденциальных документов. В модельных экспериментах будет предполагаться, что конфиденциальная информация передается фрагментами (не обязательно в форме целых документов без модификаций).

4 Методы обработки естественного языка для формирования маркеров конфиденциальности

В данном разделе рассматриваются методы обработки естественного языка для формирования МК, применимые для выявления конфиденциальной информации по тематике. Информация, использованная в проведенных модельных экспериментах, не относится к конфиденциальной. Соответственно, маркеры или признаки документов, выделяемых для коллекций, индицируют неконфиденциальные охраняемые сведения. Но отметим, что рассматриваемые методы применимы и к текстовым документам, содержащим конфиденциальную информацию.

Случайным образом были выбраны по 10 тыс. случайных текстов для следующих 4 «жанров» (классов): патентные документы; авторефераты; российские журналы; российские средства массовой информации. Для проведения экспериментов применены известные подходы на основе следующих открытых библиотек языка программирования Python: NumPy, Pandas, sklearn, NLTK и др. Используются следующие подходы к формированию обучающей матрицы:

Таблица 1 Результаты классификации

Метод	Наиболее важные признаки (в порядке убывания)	NB	LR
Count Vectors	исследования, г, фиг, м, н, л, работы, развития, диссертации, е, т, р, деятельности, п, изобретения, д, наук, устройство, анализ, способ, проблемы, научно, конференции, основные, результаты	0,861	0,975
WordLevel TF-IDF	фиг, новости, news, компании, компания, адрес, исследования, диссертации, изобретения, способ, устройство, изобретение, р, н, м, л, е, г, отличающийся, словам, масс, т, рис, v, диссертационного, наук, ит, году, корпуса, д	0,908	0,976
N-Gram Vectors	адрес новости, адрес новости news, новости news, изобретение относится, п л, п отличающийся, читайте также, меньшей мере, ключевые слова, технический результат, диссертационного совета, изобретения является	0,926	0,963
CharLevel Vectors	ii, on, s, at, mna, es, en, in, n, ion, io, e, al, tio, re, iss, фиг, зоб, er, пан, у, nt, ati, st, l, p, te, иг, ed, an, s, ic, mc, al	0,809	0,948

- подсчет слов (**Count Vectors**), число признаков не ограничено;
- TF-IDF (максимальное число признаков — 5 000) на уровнях:
 - Слов — **WordLevel TF-IDF**;
 - N-грамм (от 2 до 3) — **N-Gram Vectors**;
 - символьных N-грамм (от 2 до 3) — **CharLevel Vectors**.

Для оценки моделей использована мера f_1 с макроусреднением (скользящая средняя при $K = 5$). В табл. 1 приведены результаты классификации: в первом столбце приведен метод формирования матрицы признаков, во втором — наиболее важные признаки (выделенные с помощью метода главных компонент) в порядке убывания, в третьем и четвертом столбцах — оценки для следующих обучаемых моделей соответственно: Naïve Bayes (**NB**) и логистическая регрессия (**LR**).

Таким образом, для первичного выявления попыток нарушения конфиденциальности применимы методы текстовой классификации, которые обеспечивают высокую полноту и точность на уровне выше 97% (такой результат показал подход **WordLevel TF-IDF** с обучаемой моделью **LR**).

Завершая данный раздел, отметим следующее:

- методы текстовой классификации успешно решают задачу *определения близости или принадлежности текста*;
- при помощи анализа главных компонент можно получить неявные маркеры (слова, словосочетания, последовательность символов и др. в зависимости от метода), которые, в том числе, могут быть *маркерами конфиденциальных данных*.

5 СМЫСЛОВОЙ ПОИСК ЗАИМСТВОВАНИЙ ДЛЯ ПРЕДОТВРАЩЕНИЯ УТЕЧКИ КОНФИДЕНЦИАЛЬНОЙ ИНФОРМАЦИИ

Рассмотренный в данном разделе эксперимент на практике соответствует ситуации, когда сущность какого-либо изобретения передается сотрудником организации не непосредственным копированием материалов (выявление этого случая не представляет сложности), а излагается в свободной форме путем переписывания или изложения собственными словами.

В качестве набора, содержащего «конфиденциальную» информацию, используется коллекция «Российские журналы» (далее — тестируемая коллекция) объемом 1,85 млн документов; в качестве «исходного» набора документов (содержимое которых раскрывается в неизменном или перефразированном виде) — коллекция патентных документов Федерального института промышленной собственности (далее — исходная коллекция) объемом около 800 тыс. документов.

Анализ выполнен в режиме сплошной фильтрации тестируемой коллекции, т. е. для каждого документа из нее проводился поиск по предварительно проиндексированной исходной коллекции. Далее описывается метод измерения сходства предложений, который лежит в основе алгоритма смыслового сопоставления текстов. Предложения s_e и s_t из документов d_e и d_t соответственно представляются в виде множества пар слов $N(s_e, s_t)$, где первый элемент взят из предложения s_e , а второй — из s_t . Сопоставление двух предложений основано на парах слов из $N(s_e, s_t)$. Оценка сходства вычисляется

с учетом следующих показателей: I_1 (IDF-мера), I_2 (показатель на основе TF-IDF) и I_3 (сопоставление синтаксических структур предложений):

- $I_1(s_e, s_t) = \sum_{(w_e, w_t) \in N(s_e, s_t)} \text{IDF}(w_e)$.
 - Здесь $\text{IDF}(w_e) = \log_{|D|}(|D|/m(w_e, D))$, где D — набор документов; $m(w_e, D)$ — количество документов, содержащих слово w_e ;
- $I_2(s_e, s_t) = \sum_{(w_e, w_t) \in N(s_e, s_t)} f(w_e, w_t) \text{IDF}(w_e) \text{TF}(w_t, d_t)$.
 - Здесь $\text{TF}(w_t, d_t) = \log_{|d_t|}(k(w_t, d_t))$, где $|d_t|$ — число слов в документе d_t ; $k(w_t, d_t)$ — число встречаемых слов w_t в d_t ; $f(w_e, w_t)$ служит штрафом за несоответствие w_e и w_t ;
- $I_3(s_e, s_t) = \sum_{(w_h, \sigma, w_d) \in (\text{Syn}(s_e) \cap \text{Syn}(s_t))} \text{IDF}(w_h) / \sum_{(w_h, \sigma, w_d) \in \text{Syn}(s_e)} \text{IDF}(w_h)$.
 - Каждое предложение представляется в виде набора $\text{Syn}(s_e)$ триплетов (w_h, σ, w_d) , где w_h и w_d — нормализованные главное и зависимое слова соответственно; σ — тип синтаксического отношения (например, именная группа).

Общее сходство предложений определяется как линейная комбинация вышеприведенных мер:

$$\text{Sim}(s_e, s_t) = \text{WIdf} \cdot I_1(s_e, s_t) + \text{WTfIdf} \cdot I_2(s_e, s_t) + \text{WSynt} \cdot I_3(s_e, s_t),$$

где WIdf , WTfIdf и WSynt — относительный вклад каждой меры. Полное описание метода индексации и смыслового сопоставления текстов приведено в работах [10, 11]. Там же приведены метрики качества реализованного метода.

В табл. 2 представлен результат в виде кумулятивной статистики поиска заимствований. Во втором столбце отражено распределение количе-

ства документов по общему уровню заимствований в документе (т.е. по общей доле заимствованного текста), а в третьем столбце — по уровню заимствований из документа (т.е. по доле текста, которая заимствована из одного документа).

Анализ данных, приведенных в табл. 2, показывает, что:

- 27 документов имеют общий уровень заимствований более 90%; свыше 600 документов — более 50%.
- 81 раз встречаются документы, из которых заимствовано более 90%; свыше 500 раз — более 50%.

Отметим, что задача анализа характера заимствований с точки зрения академической этики или с юридической точки зрения не ставилась. Отметим лишь, что у большинства пар документов (заимствующий документ и документ-источник) с большой долей заимствования наблюдаются пересечения авторов. Это говорит о том, что один и тот же текст одного автора (или научной группы) использовался как для патентного документа, так и для публикации статьи, возможно с некоторыми изменениями.

6 Заключение

В статье рассмотрен возможный подход к формированию МК при их использовании в СИИД в условиях, когда состав и структура охраняемых сведений не может быть заранее определена в связи с отсутствием данных или высокой динамикой их изменения, либо их определение нецелесообразно в связи большим (или неограниченным) количеством сущностей, сведения о которых подлежат охране. В основе подхода к формированию МК лежит:

- предварительное формирование кластеров, к каждому из которых относятся документы, содержащие охраняемые сведения, и НКСС этих кластеров;
- определение близости или принадлежности конкретного текста к одному из таких кластеров;
- реализация на регулярной основе процедуры актуализации состава и структуры кластеров и их НКСС.

Предлагаемый подход не зависит от конфиденциальности обрабатываемых материалов и позволяет осуществить формирование МК в указанных условиях. Инструменты, использованные для получения результатов, описанных в разд. 4 и 5, применимы, например, для решения следующих задач:

Таблица 2 Распределение документов по уровню заимствований

Заимствования	в документе	из документа
> 0,1	17 279	18 990
> 0,2	4 385	5 863
> 0,3	2 066	2 647
> 0,4	1 109	1 011
> 0,5	612	512
> 0,6	301	279
> 0,7	122	147
> 0,8	54	99
> 0,9	27	81

- классификация документов для определения потенциальных документов, подлежащих дальнейшему анализу;
- выявление важных характеристик текстовых документов;
- поиск возможной утечки информации (при условии наличия заранее подготовленной коллекции конфиденциальных данных);
- поиск похожих или дублирующих документов для использования в автоматизированных системах документооборота.

Литература

1. Будзко В. И., Беленков В. Г., Борохов С. В. Проблемы обеспечения информационной безопасности при интенсивном использовании данных // Информационные технологии и математическое моделирование систем: Тр. Междунар. научн.-технич. конф. — Одинцово: ЦИТП РАН, 2017. С. 122–124.
2. Беленков В. Г., Борохов С. В., Будзко В. И., Кейер П. А., Королев В. И. Вопросы обеспечения информационной безопасности информационных систем, реализующих интенсивное использование данных // Аналитика и управление данными в областях с интенсивным использованием данных: Сб. научных тр. XIX Междунар. конф. DAMDID/RCDL. — М: ФИЦ ИУ РАН, 2017. С. 155–158.
3. Будзко В. И., Беленков В. Г., Королев В. И. Элементы конфиденциальности и перспективы их использования в системах, реализующих интенсивное использование данных // Системы высокой доступности, 2018. Т. 14. № 4. С. 55–60.
4. Будзко В. И., Беленков В. Г., Королев В. И. Об одном концептуальном подходе к защите информации в системах, реализующих DID // Информационные технологии и математическое моделирование систем: Тр. Междунар. научн.-технич. конф. — Одинцово: ЦИТП РАН, 2018. С. 43–46.
5. Будзко В. И., Беленков В. Г., Королев В. И. Об особенностях использования средств и методов ОИБ в системах, реализующих DID // Информационные технологии и математическое моделирование систем: Тр. Междунар. научн.-технич. конф. — Одинцово: ЦИТП РАН, 2018. С. 47–50.
6. Будзко В. И., Королев В. И., Беленков В. Г. Особенности использования маркеров конфиденциальности в системах интенсивного использования данных // Системы высокой доступности, 2019. Т. 15. № 2. С. 57–65.
7. Будзко В. И., Королев В. И., Беленков В. Г. Архитектура инструментального комплекса извлечения информации с учетом встроенных экстрактов конфиденциальности и интеграции извлеченных данных // Системы высокой доступности, 2020. Т. 16. № 2. С. 5–21.
8. Дорогой Д. С., Шаров А. В., Тузовский А. А., Терещенко И. А. Способ обучения классификатора, предназначенного для определения категории документа. Патент на изобретение № 2672395 с приоритетом от 29.09.2017. Оpubл. 14.11.2018, бюл. № 32.
9. Дорогой Д. С. Система и способ определения текста, содержащего конфиденциальные данные. Патент на изобретение № 2665915 с приоритетом от 16.06.2017. Оpubл. 04.09.2018, бюл. № 25.
10. Zubarev D. V., Sochenkov I. V. Using sentence similarity measure for plagiarism source retrieval // CEUR Workshop Proceedings: Cross Language Evaluation Forum, 2014. Vol. 1180. P. 1027–1034.
11. Zubarev D. V., Sochenkov I. V., Tikhomirov I. A., Grigoriev O. G. Double funding of scientific projects: Similarity and plagiarism detection // Аналитика и управление данными в областях с интенсивным использованием данных: Сб. научных тр. XIX Междунар. конф. DAMDID/RCDL. — М.: ФИЦ ИУ РАН, 2017. С. 282–285.

Поступила в редакцию 23.06.20

EXTRACTION OF CONFIDENTIALITY MARKERS FROM TEXTS UNDER CONDITIONS OF HIGH UNCERTAINTY IN SYSTEMS WITH DATA INTENSIVE USAGE

V. I. Budzko¹, V. V. Yadrinsev^{1,2}, I. V. Sochenkov¹, V. I. Korolev¹, and V. G. Belenkov¹

¹Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

²Peoples’ Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation

Abstract: The main tasks, the results of the solution of which are reflected in the article, are associated with the formation of confidentiality markers when they are used in data-intensive systems under conditions when the composition and structure of the protected information cannot be determined in advance due to the lack of data

or the high dynamics of their change, or their definition is not advisable due to the large number of entities whose information is subject to protection. In this paper, an approach is proposed for the formation of confidentiality markers for text materials in the indicated conditions. The article presents the semantic text analysis, which forms confidentiality markers when used to ensure information security in data-intensive systems under high uncertainty in the composition and structure of protected information. The obtained experimental results show that practical implementation of the considered approach in data-intensive systems is promising.

Keywords: confidentiality marker; information security; data-intensive domains; topical cluster; semantics; data leak prevention; intelligent security tasks; text classification; detection of text reuse

DOI: 10.14357/19922264200410

Acknowledgments

The research was supported in part by the Russian Foundation for Basic Research in the framework of scientific project No. 18-29-03215, and experimental research was supported by the “RUDN University Program 5-100.”

References

1. Budzko, V. I., V. G. Belenkov, and S. V. Borokhov. 2017. Problemy obespecheniya informatsionnoy bezopasnosti pri intensivnom ispol'zovanii dannykh [Problems of ensuring information security with intensive-data use]. *Scientific and Technical Conference (International) “Information Technologies and Mathematical Modeling of Systems” Proceedings*. Odintsovo. 122–124.
2. Belenkov, V. G., S. V. Borokhov, V. I. Budzko, P. A. Keyer, and V. I. Korolev. 2017. Voprosy obespecheniya informatsionnoy bezopasnosti informatsionnykh sistem, realizuyushchikh intensivnoe ispol'zovanie dannykh [Issues of ensuring information security of information systems that implement intensive data use]. *19th Conference (International) “Data Analytics and Management in Data Intensive Domains” Proceedings*. Moscow: FRC CSC RAS. 155–158.
3. Budzko, V. I., V. I. Korolev, and V. G. Belenkov. 2018. Elementy konfidentsial'nosti i perspektivy ikh ispol'zovaniya v sistemakh, realizuyushchikh intensivnoe ispol'zovanie dannykh [Privacy elements and prospects of their use in data intensive systems]. *Highly Available Systems* 14(4):55–60.
4. Budzko, V. I., V. G. Belenkov, and V. I. Korolev. 2018. Ob odnom kontseptual'nom podkhode k zashchite informatsii v sistemakh, realizuyushchikh DID [About one conceptual approach to information security in DID-systems]. *Scientific and Technical Conference (International) “Information Technologies and Mathematical Modeling of Systems” Proceedings*. Odintsovo. 43–46.
5. Budzko, V. I., V. G. Belenkov, and V. I. Korolev. 2018. Ob osobennostyakh ispol'zovaniya sredstv i metodov OIB v sistemakh, realizuyushchikh DID [On the features of the use of tools and methods of information security in DID-systems]. *Scientific and Technical Conference (International) “Information Technologies and Mathematical Modeling of Systems” Proceedings*. Odintsovo. 47–50.
6. Budzko, V. I., V. I. Korolev, and V. G. Belenkov. 2019. Osobennosti ispol'zovaniya markerov konfidentsial'nosti v sistemakh intensivnogo ispol'zovaniya dannykh [Features of use privacy tokens in systems that implement intensive use of data]. *Highly Available Systems* 15(2):57–65.
7. Budzko, V. I., V. I. Korolev, and V. G. Belenkov. 2020. Arkhitektura instrumental'nogo kompleksa izvlecheniya informatsii s uchetom vstroennykh ekstraktov konfidentsial'nosti i integratsii izvlechennykh dannykh [Architecture of the information extraction tool complex with built-in privacy elements and integration of extracted data]. *Highly Available Systems* 16(2):5–21.
8. Dorogoy, D. S., A. V. Sharov, A. A. Tuzovskiy, and I. A. Tereshchenko. 2018. Sposob obucheniya klassifikatora, prednaznachennogo dlya opredeleniya kategorii dokumenta [A method of training a classifier designed to determine the category of a document]. Patent RF No. 2672395.
9. Dorogoy, D. S. 2017. Sistema i sposob opredeleniya teksta, soderzhashchego konfidentsial'nye dannye [System and method for determining text containing confidential data]. Patent RF No. 2665915.
10. Zubarev, D. V., and I. V. Sochenkov. 2014. Using sentence similarity measure for plagiarism source retrieval. *CEUR Workshop Proceedings: Cross Language Evaluation Forum*. 1180:1027–1034.
11. Zubarev, D. V., I. V. Sochenkov, I. A. Tikhomirov, and O. G. Grigoriev. 2017. Double funding of scientific projects: Similarity and plagiarism detection. *19th Conference (International) “Data Analytics and Management in Data Intensive Domains” Proceedings*. Moscow: FRC CSC RAS. 282–285.

Received June 23, 2020

Contributors

Budzko Vladimir I. (b. 1944) — Doctor of Science in technology, Academician of the Academy of Cryptography of the Russian Federation, principal scientist, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vbudzko@ipiran.ru

Yadrintsev Vasiliy V. (b. 1993) — PhD student, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Str., Moscow 117198, Russian Federation; engineer-researcher, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vyadrincev@gmail.com

Sochenkov Ilya V. (b. 1985) — Candidate of Science (PhD) in physics and mathematics, Head of Department, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; sochenkov@isa.ru

Korolev Vadim I. (b. 1943) — Doctor of Science in technology, professor, leading scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; professor, Financial University under the Government of the Russian Federation, 49 Leningradskiy Prosp., Moscow 125993, Russian Federation; vkorolev@ipiran.ru

Belenkov Viktor G. (b. 1952) — Candidate of Science (PhD) in technology, leading scientist, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; vbelenkov@ipiran.ru

МЕТОД ВИЗУАЛЬНОГО ПРЕДСТАВЛЕНИЯ КОНФЛИКТОВ В ГИБРИДНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ МНОГОАГЕНТНЫХ СИСТЕМАХ

С. Б. Румовская¹, И. А. Кириков²

Аннотация: Малые коллективы экспертов, включающие специалистов различных направлений, эффективно решают сложные проблемы благодаря их анализу с различных точек зрения и получению более качественного интегрированного решения. Конфликт в малых коллективах экспертов может как завести в тупик процесс принятия решения, так и породить позитивные изменения: развитие группы, диагностику отношений, сплочивание группы. Конфликт порождает дискуссии, позволяющие получить более продуманные и согласованные решения. Подобные коллективы эффективно решают проблемы, и моделирование их работы, в частности возможной конфликтной ситуации и процесса управления ею, позволяет вырабатывать метод решения, релевантный сложной задаче. Визуализация конфликтной ситуации делает возникшие противоречия контрастными, видимыми. В работе коллектив агентов-экспертов представляется в виде неориентированного взвешенного графа и рассматриваются методы визуализации (укладки) графов. Для визуализации проблемно- и процессно-ориентированных конфликтов в рамках гибридных интеллектуальных многоагентных систем (ГиИМАС) предложен метод, разработанный на базе пружинной модели укладки графов.

Ключевые слова: коллектив экспертов; конфликт агентов; визуализация конфликта

DOI: 10.14357/19922264200411

1 Введение

Малые коллективы экспертов, включающие специалистов различных направлений, эффективно решают проблемы благодаря их всестороннему анализу, получая интегрированное компромиссное качественное решение [1]. Как следствие, исследование методов коллективного решения проблем и их моделирование — важное направление научных исследований в области системного анализа, имеющее большое практическое значение для медицины, транспорта и логистики, экономики и т.д. В [2] разработана и описана модель самоорганизации в коллективе агентов (экспертов) гибридными интеллектуальными многоагентными системами, алгоритм функционирования которых динамически перестраивается, вырабатывая релевантный проблеме метод решения.

С другой стороны, при решении проблем малыми коллективами снижается скорость принятия решений из-за процессов распределения задач и интеграции частных решений, а также возникновения конфликтов, которые могут носить как деструктивный, так и конструктивный характер.

Конструктивные конфликты позволяют получить более продуманные и согласованные решения [3], обеспечивают уникальность и автономность каждого из взаимодействующих субъектов, а также развитие отношений между ними, предоставляют информацию о возможностях противодействующих субъектов, высвобождают накапливающееся внутреннее напряжение, сохраняя связи, актуализируют разные позиции и мнения по поводу возникающих проблем, способствуют поиску оптимальных способов их решения, усиливают групповую идентичность и сплоченность. В связи с этим для повышения релевантности и качества принимаемых решений гибридного и синергетического искусственного интеллекта, моделирующего работу малого коллектива экспертов, в [4] предложена модель ГиИМАС с проблемно- и процессно-ориентированными конфликтами, а в [5] описан метод идентификации конфликтов между агентами в рамках предложенной в [4] модели.

Управление конфликтами в ГиИМАС позволяет по аналогии с реальными малыми коллективами экспертов подавлять деструктивные проявления конфликта и стимулировать конструктивные [3].

¹ Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, sophiyabr@gmail.com

² Калининградский филиал Федерального исследовательского центра «Информатика и управление» Российской академии наук, baltbipiran@mail.ru

В [3] рассмотрены методы моделирования и визуального представления конфликта в коллективе экспертов при решении проблем, по результатам которого было решено использовать опыт моделирования конфликтов логико-структурным методом [6, 7] с помощью графов.

Цель настоящей работы — разработка метода визуализации конфликтов на базе предложенной модели [4] и метода их идентификации [5], который сделает возникшие противоречия между агентами ГиИМАС контрастными, видимыми для пользователя системы. При визуализации конфликтов в ГиИМАС будем рассматривать ее как неориентированный взвешенный полный граф без петель, множество вершин которого взаимно однозначно соответствует множеству агентов ГиИМАС, ребра представляют отношения между ними, а вес ребер — напряженность возникающего конфликта [5].

2 Методы и модели визуализации графов

К наиболее известным методам рисования (визуализации, укладки) графов относятся:

- основанные на физических аналогиях (П. Эйдэс [8], Т. Камада и С. Каваи [9], Т. Фрухтерман и Э. Рейнгольд [10]) — обладают наибольшим потенциалом;
- поуровневый подход (К. Сигуяма с соавторами [11]) и восходящее или нисходящее представление [12] для ориентированных графов;
- рисование деревьев (Э. Рейнгольд и Дж. Тилфорд [13]);
- планарные укладки графов без пересечения ребер (В. Т. Татт [14]);
- ортогональные изображения графов для снижения числа пересечений ребер (Ди Батиста с соавторами [15]) — ребра изображаются прямыми, параллельными осям координат;
- произвольное представление графов [12];
- прямолинейное (ребра представляются отрезками), сеточное и полигональное (для отображения ребер используются ломаные), главная цель которых — снижение числа пересечений ребер [12].

Для укладки неориентированных взвешенных полных графов без петель хорошо зарекомендовали себя [12] алгоритмы, основанные на физических аналогиях [8–10], в которых строится специальная модель — вершины и ребра графа соответствуют «реальным» физическим взаимодействующим объектам, вводится функция энергии. Лучшая укладка

графа соответствует минимуму энергии системы. Выделяют силовой (force-directed) алгоритм рисования графов и пружинный (spring), который эквивалентен методу многомерного шкалирования (MDS, multidimensional scaling) [16].

Силовая и пружинная модели при определенном наборе параметров дают совпадение минимизируемых функций энергии и, соответственно, похожие укладки (совпадают с точностью до поворотов и масштабирования). Поэтому иногда их не различают, называя общим термином «методы, основанные на физических аналогиях» (force-directed techniques). Тем не менее силовая модель более «гибкая» за счет большего числа настраиваемых параметров: возможность регулировать веса объектов позволяет учитывать дополнительные атрибуты вершин и ребер. Также силовую модель проще интегрировать и модифицировать для учета пересечения ребер, размера доступного пространства для рисования и т. д. Однако пружинная модель проще и обладает большей производительностью.

Для дальнейшей работы выбираем пружинную модель [17], так как для достижения поставленной цели нет необходимости в учете множества дополнительных параметров, связанных со свойствами вершин и ребер, помимо весов последних. Ребра графа заменяют пружинами, при растяжении и сжатии которых возникают силы упругости, действующие по закону Гука и стремящиеся вернуть пружине ее первоначальную длину. Энергия системы пружин прямо пропорциональна расстоянию между вершинами $|p_i - p_j|$:

$$E : \sum_{(i,j) \in n} |p_i - p_j|^2,$$

где $p_i = (x_i, y_i) \in R^2$ и $p_j = (x_j, y_j) \in R^2$ — образы i -й и j -й вершин в R^2 , их позиции (координаты), $i, j \in [1, n]$ (ребра графа отображаются на прямые, соединяющие соответствующие вершины).

Для неориентированных взвешенных графов, число вершин и ребер которых не превышает 200 (малые коллективы, моделируемые ГиИМАС, содержат не более 20 экспертов, преимущественно менее 10), наиболее часто применяется пружинная модель Т. Камада и С. Каваи [9], на базе которой описан предлагаемый метод визуализации конфликтов (МВК).

3 Метод визуализации конфликтов между агентами

В [4] было введено понятие процесса управления конфликтами:

$$\text{cnfm} = \langle \text{CNF}, \text{cnfcl}, \text{cmkb}, \text{act}_{\text{cnfm}}, \text{ACT}_{\text{agcr}} \rangle.$$

Здесь CNF — матрица, описывающая конфликт между каждой парой агентов кортежем, представленным выражением:

$$cnf_{ij\ cnft} = \langle id_i, id_j, cnfin, cnft, ACT_{agcr\ i}, ACT_{agcr\ j} \rangle,$$

где id_i и id_j — идентификаторы агентов-субъектов конфликта, $cnfin \in [0, 1]$ — напряженность конфликта, $cnft$ — символьная переменная «тип конфликта», определенная на множестве $CNFT = \{\text{«проблемно-ориентированный»}, \text{«процессно-ориентированный»}\}$, $ACT_{agcr\ i}, ACT_{agcr\ j} \subseteq \subseteq ACT_{agcr}$ — множества допустимых действий агентов ag_i и ag_j по разрешению противоречий; $cnfcl$ — классификатор конфликтов агентов, идентифицирующий их характер и оценивающий напряженность, т.е. формирующий для каждой пары агентов значение элемента матрицы CNF [5]; $cmkb$ — база знаний об эффективности методов управления конфликтами в зависимости от характеристик проблемы и конфликтов между агентами; act_{cnfin} — функция агента-фасилитатора «управление конфликтом»; ACT_{agcr} — множество допустимых действий агентов по разрешению противоречий.

Матрица CNF вместе с пороговым минимальным значением напряженности визуализируемого конфликта η (задается пользователем, по умолчанию $\eta = 0$) — входные данные МВК.

Первый шаг МВК — вычисление промежуточных матриц CP , CPR и D .

Элементы $cp_{ij} = Pr_3(cnf_{ij\ prob})$ матрицы CP описывают величину напряженности проблемно-ориентированного конфликта между агентами.

Элементы $cpr_{ij} = Pr_3(cnf_{ij\ prob})$ матрицы CPR описывают величину напряженности процессно-ориентированного конфликта между агентами.

Элементы матрицы D расстояний графа конфликтов рассчитываются в соответствии с выражением:

$$d_{ij} = \begin{cases} 0, & \text{если } i = j; \\ 0,00001, & \text{если } cpr_{ij} = cp_{ij} = 0; \\ (0,5 (cp_{ij}^2 + cpr_{ij}^2))^{0,5}, & \text{в противном случае.} \end{cases}$$

Второй шаг МВК — запуск алгоритма Камада–Каваи [9]:

- (1) вершины графа, представляющие агентов ГиИМАС, помещаются в случайные координаты p_i ;
- (2) выбирается вершина m , на которую действует максимальная сила;
- (3) остальные вершины фиксируются, энергия системы минимизируется двумерным методом Ньютона–Рафсона, вычисляется смещение для вершины m ;

- (4) шаги 2 и 3 повторяются до достижения одного из признаков останова: либо заданного числа итераций, либо порога силы, действующей на вершины, ниже которого алгоритм не запускается.

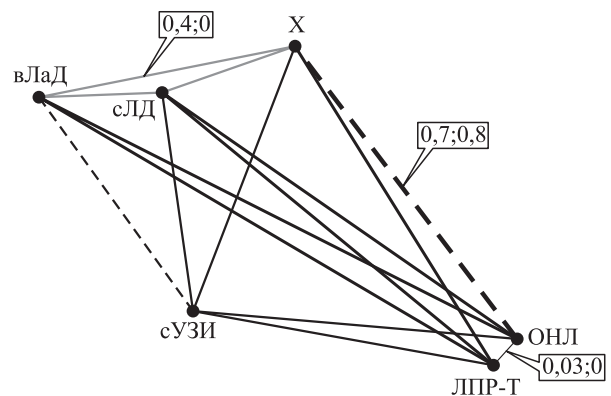
Алгоритм Камада–Каваи на выходе даст оптимальную укладку графа, описывающего ГиИМАС (соответствует состоянию с минимальной суммарной энергией системы). Энергия всей системы рассчитывается как

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^n 0,5k_{ij} (|p_i - p_j| - l_{ij})^2.$$

Здесь n — число вершин; $k_{ij} = (d_{ij})^{-2}$ — сила пружины между вершинами; p_i и p_j — положение на плоскости вершин i и j соответственно; $l_{ij} = L_0 d_{ij} (\max_{i < j} d_{ij})^{-1}$ — идеальная длина пружины, где L_0 — длина стороны квадратной области дисплея.

Третий шаг МВК — получение результирующей визуализации (см. рисунок) конфликтующих агентов. Корректируем укладку графа:

- с учетом матриц CP и CPR : если превалирует проблемно-ориентированный конфликт между агентами ($cp_{ij} \geq cpr_{ij}$), то ребро между двумя вершинами (агентами) прорисовывается сплошной линией, а если процессно-ориентированный ($cp_{ij} < cpr_{ij}$) — штриховой. Толщина и цвет (от светло-серого до черного на рисунке) линии указывают на величину среднего квадратического напряженностей конфликтов между агентами;



Визуализация конфликта (на примере коллектива агентов, решающего задачу диагностики рака поджелудочной железы): Х — хирург; ОНЛ — онколог по нехирургическому лечению; ЛПР-Т — лицо, принимающее решение (терапевт); сУЗИ — специалист по ультразвуковому исследованию; вЛад — врач лабораторной диагностики; сЛД — специалист по лучевой диагностике

- к вершинам добавляем подписи — идентификаторы агентов;
- если $\eta > 0$, то для агентов с напряженностью конфликта ниже η ребра на результирующем графе не будут отображены.

При наведении указателя мыши на ребро отображаются напряженности конфликтов между соответствующей парой агентов. При каждой фиксации изменений напряженности конфликтов в матрице CNF будет запускаться MBK. Для каждой сессии работы системы весь визуальный ряд конфликта между агентами сохраняется, обеспечивая возможность более детального изучения пользователем. Таким образом, пользователь может отслеживать развитие конфликта с самого начала работы системы и до момента получения решения. Визуализация дает быстро воспринимаемое знание о том, между какими агентами возникают конфликты при решении проблемы, какого они типа, как меняется их напряженность в процессе работы системы. С помощью данных знаний можно предотвратить и/или быстрее разрешить развитие конфликта в естественных малых коллективах экспертов, решающих подобную проблему.

4 Заключение

В работе представлены результаты обзора методов укладки графов, на основе анализа которого для разработки метода визуализации конфликтов в коллективе агентов выбран метод рисования графа, основанный на физических аналогиях, а именно: эффективная пружинная модель укладки графов Т. Камада — С. Каваи, зарекомендовавшая себя для неориентированных взвешенных графов малой и средней размерности. Результирующая визуализация конфликта агентов предоставляет пользователю быстрое понимание того, между какими членами коллектива и на каких этапах возникает конфликт, какого он типа и напряженности, а также на каком этапе он минимизируется/нивелируется. Разработанный подход к визуализации повышает прозрачность работы ГиИМАС для пользователя, не зависит от численности коллектива агентов и легко реализуем.

Литература

1. Колесников А. В. Гетерогенные естественные и искусственные системы // Интегрированные модели и мягкие вычисления в искусственном интеллекте. — М.: Физматлит, 2013. Т. 1. С. 86–103.
2. Колесников А. В., Кириков И. А., Листопад С. В. Гибридные интеллектуальные системы с самоорганизацией: координация, согласованность, спор. — М.: ИПИ РАН, 2014. 189 с.
3. Румовская С. Б., Кириков И. А. Методы моделирования и визуального представления конфликта в малом коллективе экспертов, решающих проблемы (обзор) // Информатика и её применения, 2019. Т. 13. Вып. 3. С. 122–130. doi: 10.14357/19922264190317.
4. Листопад С. В., Кириков И. А. Моделирование конфликтов агентов в гибридных интеллектуальных многоагентных системах // Системы и средства информатики, 2019. Т. 29. № 3. С. 139–148. doi: 10.14357/08696527190312.
5. Листопад С. В., Кириков И. А. Метод идентификации конфликтов агентов в гибридных интеллектуальных многоагентных системах // Системы и средства информатики, 2020. Т. 30. № 1. С. 56–65. doi: 10.14357/08696527200105.
6. Готин С. В., Калоша Л. П. Логико-структурный подход и его применение для анализа и планирования деятельности. — М.: Вариант, 2007. 118 с.
7. Новиков Д. А. Иерархические модели военных действий // Управление большими системами, 2012. № 37. С. 25–62.
8. Eades P. A heuristic for graph drawing // Congressus Numerantium, 1984. Vol. 42. P. 149–160.
9. Kamada T., Kawai S. An algorithm for drawing general undirected graphs // Inform. Process. Lett., 1989. Vol. 31. Iss. 1. P. 7–15. doi: 10.1016/0020-0190(89)90102-6.
10. Fruchterman T., Reingold E. Graph drawing by force-directed placement // Software Pract. Exper., 1991. Vol. 21. Iss. 11. P. 1129–1164. doi: 10.1002/spe.4380211102.
11. Sugiyama K., Tagawa S., Toda M. Methods for visual understanding of hierarchical system structures // IEEE T. Syst. Man Cyb., 1981. Vol. 11. Iss. 2. P. 109–125. doi: 10.1109/TSMC.1981.4308636.
12. Tamassia R., Battista G. D., Ioannis G., Eades P. Graph drawing: Algorithms for the visualization of graphs. — Englewood Cliffs, NJ, USA: Prentice Hall, 1999. 397 p.
13. Reingold E., Tilford J. Tidier drawing of trees // IEEE T. Software Eng., 1981. Vol. SE-7. Iss. 2. P. 223–228. doi: 10.1109/TSE.1981.234519.
14. Tutte W. T. How to draw a graph // P. Lond. Math. Soc., 1963. Vol. S3-13. Iss. 1. P. 743–767. doi: 10.1112/plms/s3-13.1.743.
15. Battista G. D., Liotta G., Vargiu F. Spirality of orthogonal representations and optimal drawings of series-parallel graphs and 3-planar graphs // Algorithms and data structures / Eds. F. K. H. A. Dehne, J.-R. Sack, N. Santoro, S. Whitesides. — Lecture notes in computer science ser. — Springer, 1993. Vol. 709. P. 151–162. doi: 10.1007/3-540-57155-8_244.
16. Kruskal J. B., Seery J. B. Designing network diagrams // 1st General Conference (International) on Social Graphs Proceedings, 1980. P. 22–50.
17. Пунтырев С. Н. Модели, алгоритмы и программный комплекс визуализации сложных сетей: Дис. ... канд. физ.-мат. наук. — Екатеринбург, 2010. 136 с.

Поступила в редакцию 30.09.20

CONFLICT VISUAL REPRESENTATION METHOD IN HYBRID INTELLIGENT MULTIAGENT SYSTEMS

S. B. Rumovskaya and I. A. Kirikov

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

Abstract: Small collectives of experts, including specialists from different fields, effectively solve complex problems by analyzing them from different points of view and obtaining a better-integrated solution. A conflict in small collectives of experts can both lead to a deadlock in the decision-making process and generate positive changes: development of the group, diagnostics of relations, and consolidation of the group. A conflict breeds debate, the depth of which allows for more thoughtful and coordinated solutions. Such collectives of experts solve problems effectively. Thus, modeling of their work and possible conflict situation with managing it allows developing a decision-support method that is relevant to solving a complex problem. Visualization of a conflict situation makes appeared contradictions contrast and observable. In the research, the authors represent a collective of agents-experts in the form of an undirected weighted graph. The methods of graph visualization are considered. To visualize problem- and process-oriented conflicts within hybrid intelligent multiagent systems, the authors propose a method based on the spring model of graph drawing.

Keywords: collective of experts; conflict; visualization of conflict

DOI: 10.14357/19922264200411

References

- Kolesnikov, A. V. 2013. Geterogennyye estestvennyye i iskusstvennyye sistemy [Natural and artificial heterogeneous systems]. *Integrirrovannyye modeli i myagkie vychisleniya v iskusstvennom intellekte* [Integrated models and soft computing in artificial intelligence]. Moscow: Fizmatlit. 1:86–103.
- Kolesnikov, A. V., I. A. Kirikov, and S. V. Listopad. 2014. *Gibridnyye intellektual'nye sistemy s samoorganizatsiyey: koordinatsiya, soglasovannost', spor* [Hybrid artificial systems with self-organization: Coordination, conformance, row]. Moscow: IPI RAN. 189 p.
- Rumovskaya, S. B., and I. A. Kirikov. 2019. Metody modelirovaniya i vizual'nogo predstavleniya konflikta v malom kollektive ekspertov, reshayushchikh problemy (obzor) [Methods of modeling and visual representation of a conflict in small collective of experts solving problems (review)]. *Informatika i ee Primeneniya — Inform. Appl.* 13(3):122–130. doi: 10.14357/19922264190317.
- Listopad, S. V., and I. A. Kirikov. 2019. Modelirovanie konfliktov agentov v gibridnykh intellektual'nykh mnogoagentnykh sistemakh [Modeling of agent conflicts in hybrid intelligent multiagent systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(3):139–148. doi: 10.14357/08696527190312.
- Listopad, S. V., and I. A. Kirikov. 2020. Metod identifikatsii konfliktov agentov v gibridnykh intellektual'nykh mnogoagentnykh sistemakh [Agent conflict identification method in hybrid intelligent multiagent systems]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 30(1):56–65. doi: 10.14357/08696527200105.
- Gotin, S. V., and L. P. Kalosha. 2007. *Logiko-strukturnyy podkhod i ego primeneniye dlya analiza i planirovaniya deyatelnosti* [Logical-structural approach and its application for the analysis and planning of activities]. Moscow: Variant. 118 p.
- Novikov, D. A. 2012. Ierarkhicheskie modeli voennykh deystviy [Hierarchical models of combat]. *Upravlenie bol'shimi sistemami* [Control of Large Systems] 37:25–62.
- Eades, P. 1984. A heuristic for graph drawing. *Congressus Numerantium* 42:149–160.
- Kamada, T., and S. Kawai. 1989. An algorithm for drawing general undirected graphs. *Inform. Process. Lett.* 31(1):7–15. doi: 10.1016/0020-0190(89)90102-6.
- Fruchterman, T., and E. Reingold. 1991. Graph drawing by force-directed placement. *Software Pract. Exper.* 21(11):1129–1164. doi: 10.1002/spe.4380211102.
- Sugiyama, K., S. Tagawa, and M. Toda. 1981. Methods for visual understanding of hierarchical system structures. *IEEE T. Syst. Man Cyb.* 11(2):109–125. doi: 10.1109/TSMC.1981.4308636.
- Tamassia, R., G. D. Battista, G. Ioannis, and P. Eades. 1999. *Graph drawing: Algorithms for the visualization of graphs*. Englewood Cliffs, NJ: Prentice Hall. 397 p.
- Reingold, E., and J. Tilford. 1981. Tidier drawing of trees. *IEEE T. Software Eng.* SE-7(2):223–228. doi: 10.1109/TSE.1981.234519.
- Tutte, W. T. 1963. How to draw a graph. *P. Lond. Math. Soc.* S3-13(1):743–767. doi: 10.1112/plms/s3-13.1.743.

15. Di Battista, G., G. Liotta, and F. Vargiu. 1993. Spirality of orthogonal representations and optimal drawings of series-parallel graphs and 3-planar graphs. *Algorithms and data structures*. Eds. F. K. H. A. Dehne, J.-R. Sack, N. Santoro, and S. Whitesides. Lecture notes in computer science ser. Springer. 709:151–162. doi: 10.1007/3-540-57155-8_244.
16. Kruskal, J. B., and J. B. Seery. 1980. Designing network diagrams. *1st General Conference (International) on Social Graphics Proceedings*. 22–50.
17. Pupyrev, S. N. 2010. Modeli, algoritmy i programnyy kompleks vizualizatsii slozhnykh setey [Models, algorithms and software for visualization of complex networks]. Ekaterinburg. PhD Diss. 136 p.

Received September 30, 2020

Contributors

Rumovskaya Sophiya B. (b. 1985) — Candidate of Science (PhD) in technology, 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; sophiyabr@gmail.com

Kirikov Igor A. (b. 1955) — Candidate of Science (PhD) in technology, director, Kaliningrad Branch of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 5 Gostinaya Str., Kaliningrad 236000, Russian Federation; baltbipiran@mail.ru

ОЦЕНКА СТОИМОСТИ ОПЦИОНОВ НА ОСНОВЕ МОДЕЛЕЙ ARIMA–GARCH С ОШИБКАМИ, РАСПРЕДЕЛЕННЫМИ ПО ЗАКОНУ S_U ДЖОНСОНА

А. Р. Данилишин¹, Д. Ю. Голембиовский²

Аннотация: В продолжение статьи «Риск-нейтральная динамика для модели ARIMA–GARCH с ошибками, распределенными по закону S_U Джонсона» в данной работе приводятся результаты экспериментов для моделей ARIMA–GARCH (autoregressive integrated moving average – generalized autoregressive conditional heteroskedasticity) с нормальными (N), экспоненциальными бета второго типа (EGB2) и S_U Джонсона (JSU) распределениями ошибок. Стоимость европейских опционов оценивается методом Монте-Карло на основе результатов, полученных в указанной статье при помощи расширенного принципа Гирсанова. Параметры моделей ARIMA–GARCH-N, ARIMA–GARCH-EGB2 и ARIMA–GARCH-JSU были найдены методом квазимаксимального правдоподобия. Эффективность полученных риск-нейтральных моделей исследовалась на примере биржевых европейских опционов PUT и CALL на базовые активы DAX (Deutscher Aktienindex) и Light Sweet Crude Oil.

Ключевые слова: ARIMA; GARCH; риск-нейтральная мера; расширенный принцип Гирсанова; распределение S_U Джонсона; ценообразование опционов

DOI: 10.14357/19922264200412

1 Введение

Данная работа является продолжением статьи [1]. В указанной статье была введена модель ARIMA(p, d, q)–GARCH(P, Q)–JSU($\xi, \lambda, \gamma, \delta$) для доходности базового актива в следующем виде (для распределения JSU $\tilde{Y}_t = S_t/S_{t-1} - 1$, для N- и EGB2-распределений $Y_t = \ln(S_t/S_{t-1})$) [2–5]:

$$\left. \begin{aligned} \Delta^d Y_t &= m_t + \sqrt{h_t} \varepsilon_t, \quad \varepsilon_t | \mathcal{F}_{t-1} \sim \text{JSU}(\xi, \lambda, \gamma, \delta); \\ m_t &= \mathbb{E}[\Delta^d Y_t | \mathcal{F}_{t-1}] = \phi_0 + \dots + \phi_p \Delta^d Y_{t-p} + \\ &\quad + \theta_1 \sqrt{h_{t-1}} \varepsilon_{t-1} + \dots + \theta_q \sqrt{h_{t-q}} \varepsilon_{t-q}; \\ h_t &= \text{Var}[\Delta^d Y_t | \mathcal{F}_{t-1}] = \alpha_0 + \dots + \alpha_P h_{t-P} + \\ &\quad + \beta_1 h_{t-1} \varepsilon_{t-1}^2 + \dots + \beta_Q h_{t-Q} \varepsilon_{t-Q}^2 \end{aligned} \right\} (1)$$

с ограничениями на математическое ожидание и дисперсию ошибки

$$\begin{aligned} \mathbb{E}[\varepsilon_t] &= \xi - \lambda e^{1/(2\delta^2)} \sinh\left(\frac{\gamma}{\delta}\right) = 0; \\ \text{Var}[\varepsilon_t] &= \frac{\lambda^2}{2} \left(e^{1/\delta^2} - 1 \right) \times \\ &\quad \times \left(e^{1/\delta^2} \cosh\left(\frac{2\gamma}{\delta}\right) + 1 \right) = 1. \end{aligned}$$

Далее к разностному (стационарному) ряду (1) применялся расширенный принцип Гирсанова [6, 7], а в случае распределения JSU — его найденная модификация. Для распределения JSU полученные коэффициенты модели, обеспечивающие риск-нейтральную динамику процесса даются следующими соотношениями:

$$\left. \begin{aligned} Y_t &= r + \delta_t \frac{1+r}{1+m_t} \varepsilon_t; \\ \varepsilon_t | \mathcal{F}_{t-1} &\sim \text{JSU}(\tilde{\xi}, \tilde{\lambda}, \gamma, \delta); \\ \tilde{\xi} &= \tilde{\lambda} e^{1/(2\delta^2)} \sinh\left(\frac{\gamma}{\delta}\right); \\ \tilde{\lambda} &= \sqrt{2} \left(\left(e^{1/\delta^2} - 1 \right) \times \right. \\ &\quad \left. \times \left(e^{1/\delta^2} \cosh\left(\frac{2\gamma}{\delta}\right) + 1 \right) \right)^{-1/2}. \end{aligned} \right\} (2)$$

Риск-нейтральные коэффициенты для моделей ARIMA–GARCH-EGB2 и ARIMA–GARCH-N представлены соотношениями [7]:

$$Y_t = r - \ln \frac{B(\alpha + \delta_t \bar{\delta}, \beta - \delta_t \bar{\delta})}{B(\alpha, \beta)} + \delta_t \bar{\delta} \bar{\omega}(\alpha, \beta) + \delta_t \varepsilon_t; \quad (3)$$

¹Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики, danilishin-artem@mail.ru

²Московский государственный университет имени М. В. Ломоносова, факультет вычислительной математики и кибернетики; Московский финансово-промышленный университет «Синергия», golemb@cs.msu.su

$$\left. \begin{aligned} \varepsilon_{t|\mathcal{F}_{t-1}} &\sim \text{EGB2}(\alpha, \beta, \bar{\delta}, \bar{\mu}); \\ \bar{\delta} &= \frac{1}{\sqrt{l(\alpha, \beta)}}, \quad \bar{\mu} = -\frac{\bar{w}(\alpha, \beta)}{\sqrt{l(\alpha, \beta)}}; \\ Y_t &= r - \frac{1}{2}\delta_t^2 + \delta_t\varepsilon_t; \quad \varepsilon_{t|\mathcal{F}_{t-1}} \sim N(0, 1). \end{aligned} \right\} (4)$$

В данной статье приводятся результаты численных экспериментов, подтверждающие корректность теоретических результатов первой работы и эффективность полученных риск-нейтральных моделей ARIMA–GARCH-N, ARIMA–GARCH-EGB2 и ARIMA–GARCH-JSU.

Работа построена следующим образом. В разд. 2 приводятся формулы для оценки справедливой стоимости опционов CALL и PUT методом Монте-Карло. В разд. 3 описываются два набора данных для проведения численных экспериментов: цены закрытия торгов по опционам на индекс немецких акций DAX и на нефть Light Sweet Crude Oil, а также соответствующие ряды цен базовых активов. В разд. 4 даются формулы оценки параметров моделей ARIMA–GARCH методом квазимаксимального правдоподобия [8]. В разд. 5 приводятся тесты, определяющие спецификации моделей и результаты оценок параметров моделей. Раздел 6 содержит результаты оценки справедливой стоимости опционов, полученные при использовании различных моделей динамики базовых активов. В заключении формулируются выводы исследования.

2 Оценка справедливой стоимости опциона методом Монте-Карло

Оценка справедливой стоимости опционов проводится по методу Монте-Карло [9]. Европейские опционы CALL и PUT с ценой исполнения X и стоимостью базового актива S_T в день экспирации T характеризуются функциями выплат $b_c(S_T, X) = \max(S_T - X, 0)$ и $b_p(S_T, X) = \max(X - S_T, 0)$. Стоимость опционов определяется как среднее значение соответствующей функции выплаты относительно риск-нейтральной меры \mathbb{Q} , приведенное к текущему моменту времени [10]:

$$\frac{p_{\text{call}}}{p_{\text{put}}} = \frac{\mathbb{E}^{\mathbb{Q}}[b_{c/p}(S_T, X)]}{(1+r)^T} = \frac{\mathbb{E}^{\mathbb{P}}[b_{c/p}(S_T, X)d\mathbb{Q}/d\mathbb{P}]}{(1+r)^T}, \quad (5)$$

где $d\mathbb{Q}/d\mathbb{P}$ — производная Радона–Никодима риск-нейтральной меры (в рамках данной работы это мера, полученная на основе расширенного принципа Гирсанова либо его модификации) относительно физической меры \mathbb{P} [11, 12]. Метод Монте-Карло позволяет по реализациям построенного процесса

ARIMA–GARCH оценить среднее значение относительно риск-нейтральной меры \mathbb{Q} :

$$\frac{1}{M} \sum_{m=1}^M b_{c/p}(S_T, X) \frac{d\mathbb{Q}}{d\mathbb{P}}(m) \xrightarrow{M \rightarrow \infty} \frac{P}{M \rightarrow \infty} \mathbb{E}^{\mathbb{P}} \left[b_{c/p}(S_T, X) \frac{d\mathbb{Q}}{d\mathbb{P}} \right]. \quad (6)$$

3 Описание данных

Данные для проведения численных экспериментов состоят из двух однотипных наборов цен закрытия торгов по опционам на 3 июня 2019 г. По дате экспирации опционы были поделены на самые ближние и дальние с учетом ликвидности.

Первый набор данных представлен европейскими опционами на фондовый индекс DAX. Индекс отражает суммарный доход по капиталу, поэтому при его расчете учитываются полученные дивидендные доходы по акциям, которые, как предполагается, реинвестируются в акцию, по которой был получен дивиденд. Рассматриваются 19 опционов CALL и 19 опционов PUT, величина страйка варьируется от 9400 до 13000 с шагом 200. Базовым активом выступает фьючерс с датой экспирации, соответствующей дате экспирации самого опциона. Дата экспирации ближних опционов — 22 июня 2019 г., а дальних — 22 декабря 2023 г. Валюта опционов — евро. Данные взяты с сайта www.eurexchange.com.

Во второй набор данных вошли 10 европейских опционов CALL и PUT на фьючерс, базовым активом которого выступает нефть (Light Sweet Crude Oil). Величина страйка варьируется от 51,0 до 55,5 с шагом 0,5. Дата экспирации ближних опционов — 20 июня 2019 г., дальних — 22 июня 2020 г. Валюта — доллар США. Источник данных: www.cmegroup.com.

В представленных данных фигурируют два вида валют; соответственно, при расчете справедливой стоимости опционов использовались две безрисковые процентные ставки. В качестве таковых были взяты соответствующие ставки LIBOR. На дату расчета (03 июня 2019 г.) ставка USD LIBOR равнялась 2,36075%, а ставка EUR LIBOR составляла 0,46614%. Источник данных: www.global-rates.com.

4 Калибровка моделей ARIMA–GARCH

Пусть $\ln(L_n(v))$ — логарифм функции правдоподобия модели ARIMA–GARCH для доходности базового актива с вектором параметров $v \in \Theta$.

В случае модели ARIMA (p, d, q) –GARCH (P, Q) –JSU $(\xi, \lambda, \gamma, \delta)$ (1) имеется вектор параметров

$$v = (\gamma, \delta, \phi_0, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \alpha_0, \alpha_1, \dots, \alpha_P, \beta_1, \dots, \beta_Q).$$

Число параметров равно $n = p + 1 + q + P + Q + 2$. Оптимальные параметры определяются исходя из максимума функции правдоподобия [8]:

$$\begin{aligned} \hat{v}_n &= \arg \max_{v \in \Theta} \ln(L_n(v)) = \\ &= \arg \max_{v \in \Theta} \sum_{t=0}^T \left(\ln(\delta) - \ln \left(\sqrt{\frac{2h_t}{A}} \right) - \right. \\ &\quad \left. - \frac{1}{2} \ln \left(1 + \left(\frac{\varepsilon_t}{\sqrt{2h_t/A}} - B \right)^2 \right) - \right. \\ &\quad \left. - \frac{1}{2} \left(\gamma + \delta \sinh^{-1} \left(\frac{\varepsilon_t}{\sqrt{2h_t/A}} - B \right) \right)^2 \right), \\ &\quad \delta, \alpha_0 > 0, \alpha_1, \dots, \alpha_P, \beta_1, \dots, \beta_Q \geq 0, \quad (7) \end{aligned}$$

где

$$\begin{aligned} A &= (e^{1/\delta^2} - 1) \left(e^{1/\delta^2} \cosh \left(\frac{2\gamma}{\delta} \right) + 1 \right); \\ B &= e^{1/(2\delta^2)} \sinh \left(\frac{\gamma}{\delta} \right); \quad \sum_{i=1}^P \alpha_i + \sum_{j=1}^Q \beta_j < 1. \end{aligned}$$

Таблица 1 Результаты оценивания моделей ARIMA(0,0,1)–GARCH(1,1) для $Y_t^{N/EG\text{B}2} = \ln(S_t/S_{t-2})$ и $\tilde{Y}_t^{\text{JSU}} = S_t/S_{t-2} - 1$ фондового индекса DAX

Распределение ошибок	N	EGB2	JSU
$L_n(\hat{v})$	1652,609	1658,26	1658,657
ϕ_0	-0,000100	-0,00006	-0,000074
Std. error	(0,000798)	(0,000786)	(0,000792)
t-value	-0,12480	-0,076381	-0,093104
θ_1	0,950806	0,948445	0,949900
Std. error	(0,013106)	(0,015226)	(0,013229)
t-value	72,54849	62,292279	71,806222
α_0	0,000003	0,000003	0,000003
Std. error	(0,000005)	(0,000004)	(0,000004)
t-value	0,55453	0,687438	0,718274
α_1	0,071761	0,067895	0,067271
Std. error	(0,031491)	(0,027377)	(0,030275)
t-value	2,27881	2,480017	2,222013
β_1	0,894232	0,902903	0,900226
Std. error	(0,050380)	(0,037822)	(0,041630)
t-value	17,74964	23,872549	21,624622
AIC	-6,5773	-6,5919	-6,5934
BIC	-6,5352	-6,5369	-6,5385
ξ	—	-0,224916	-0,543876
κ	—	2,897165	2,298773

Для случая EGB2-распределения число параметров $v = (\alpha, \beta, \phi_0, \dots, \phi_p, \theta_1, \dots, \theta_q, \alpha_0, \alpha_1, \dots, \alpha_P, \beta_1, \dots, \beta_Q)$ такое же, как в предыдущем случае. Оптимизационная задача имеет следующий вид:

$$\begin{aligned} \hat{v}_n &= \arg \max_{v \in \Theta} \ln(L_n(v)) = \\ &= \arg \max_{v \in \Theta} \sum_{t=0}^T \left(\ln \left(\sqrt{l(\alpha, \beta)} \right) - \ln(B(\alpha, \beta)) + \right. \\ &\quad \left. + \alpha \bar{\omega}(\alpha, \beta) + \frac{\alpha \varepsilon_t \sqrt{l(\alpha, \beta)}}{\sqrt{h_t}} - \ln \left(\sqrt{h_t} \right) - \right. \\ &\quad \left. - (\alpha + \beta) \ln \left(1 + \exp \left(\frac{\varepsilon_t \sqrt{l(\alpha, \beta)}}{\sqrt{h_t}} + \bar{\omega}(\alpha, \beta) \right) \right) \right), \\ &\quad \alpha, \beta, \alpha_0 > 0, \alpha_1, \dots, \alpha_P, \beta_1, \dots, \beta_Q \geq 0, \quad (8) \end{aligned}$$

где $\Gamma(c)$ — гамма-функция;

$$\begin{aligned} \bar{\omega}(\alpha, \beta) &= \frac{d \ln \Gamma(c)}{dc} \Big|_{c=\alpha} - \frac{d \ln \Gamma(c)}{dc} \Big|_{c=\beta}; \\ l(\alpha, \beta) &= \frac{d^2 \ln \Gamma(c)}{dc^2} \Big|_{c=\alpha} + \frac{d^2 \ln \Gamma(c)}{dc^2} \Big|_{c=\beta}; \\ &\quad \sum_{i=1}^P \alpha_i + \sum_{j=1}^Q \beta_j < 1. \end{aligned}$$

Таблица 2 Результаты оценивания моделей ARIMA(2,0,0)–GARCH(1,1) для $Y_t^{N/EG\text{B}2} = \ln(S_t/S_{t-2})$ и $\tilde{Y}_t^{\text{JSU}} = S_t/S_{t-2} - 1$ (Light Sweet Crude Oil)

Распределение ошибок	N	EGB2	JSU
$L_n(\hat{v})$	1261,086	1267,017	1268,022
ϕ_1	0,667116	0,657088	0,659745
Std. error	(0,043985)	(0,043122)	(0,043152)
t-value	15,1668	15,2379	15,2888
ϕ_2	-0,313186	-0,323465	-0,322464
Std. error	(0,043636)	(0,042236)	(0,042439)
t-value	-7,1773	-7,6585	-7,5983
α_0	0,000015	0,000013	0,000013
Std. error	(0,000002)	(0,000001)	(0,000001)
t-value	8,341500	16,4056	16,1163
α_1	0,049707	0,042324	0,041834
Std. error	(0,007002)	(0,005701)	(0,005506)
t-value	7,0988	7,424	7,5984
β_1	0,913128	0,92728	0,928020
Std. error	(0,013747)	(0,011196)	(0,011169)
t-value	66,424300	82,8231	83,0925
AIC	-4,9944	1267,752	-5,0140
BIC	-4,9524	-4,9542	-4,9553
ξ	—	-0,70877	-0,27232
κ	—	4,405384	2,663401

Для случая нормального распределения использовались функции библиотеки rugarch среды R [8].

5 Калибровка параметров моделей ARIMA–GARCH

В табл. 1 и 2 представлены результаты оценки параметров моделей ARIMA–GARCH. Для обоих временных рядов был проведен Q-тест Льюнга–Бокса для разного числа лагов (нулевая гипотеза заключается в отсутствии автокорреляций для первых k лагов), который показал, что автокорреляционные связи для рядов $\ln(S_t/S_{t-1})$ и $S_t/S_{t-1} - 1$ отсутствуют с вероятностью 99% (значение статистики составило 30,123 для индекса DAX и 24,576 для Light Sweet Crude Oil, в то время как критическое значение для $k = 30$ лагов и уровня значимости 99% равно 50,892). В результате были рассмотрены ряды $\ln(S_t/S_{t-2})$ и $S_t/S_{t-2} - 1$ и на основе их коррелограмм ACF и PACF (рис. 1 и 2)

был сделан вывод о спецификации ARIMA части моделей ARIMA–GARCH.

Для коррелограмм индекса DAX характерна модель ARIMA(0, 0, 1) (ненулевое значение автокорреляции первого лага и затухающая динамика значений частных автокорреляций). Коррелограммы цен на нефть определяют модель ARIMA(2, 0, 0) (ненулевые значения частных автокорреляций двух первых лагов и затухающее поведение автокорреляций).

Коэффициенты всех трех моделей имеют одинаковый знак и порядок. Все полученные коэффициенты моделей статистически значимы на уровне значимости 99% (следует из t -критерия). Можно также заметить, что коэффициенты, соответствующие моделям с распределениями ошибок EGB2 и JSU, почти идентичны.

Предсказательная сила моделей оценивалась с помощью информационных критериев Акаике (AIC) и Байеса (BIC), статистики которых рассчитываются по следующим формулам:

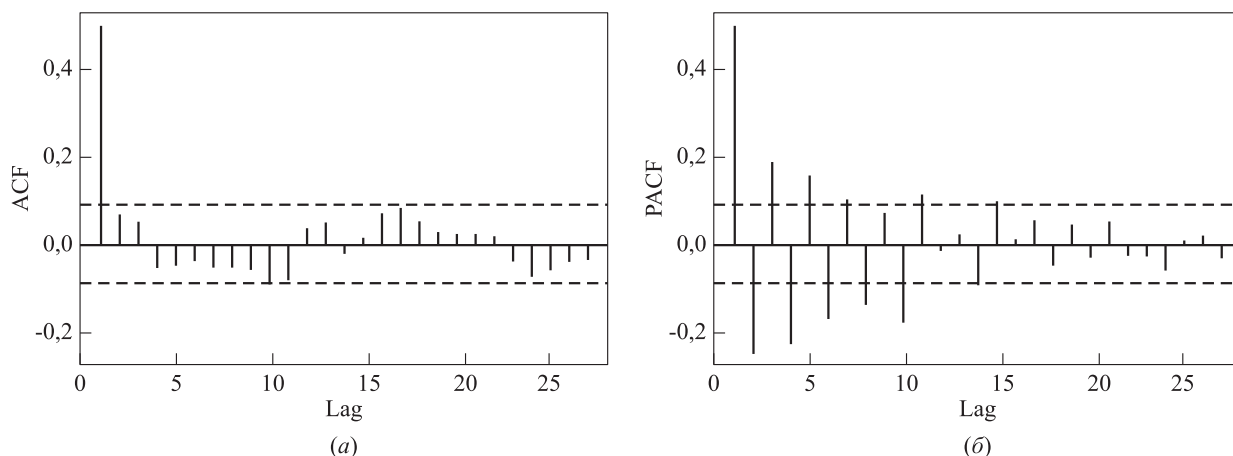


Рис. 1 Графики ACF (а) и PACF (б) для $Y_t = \ln(S_t/S_{t-2})$ (DAX)

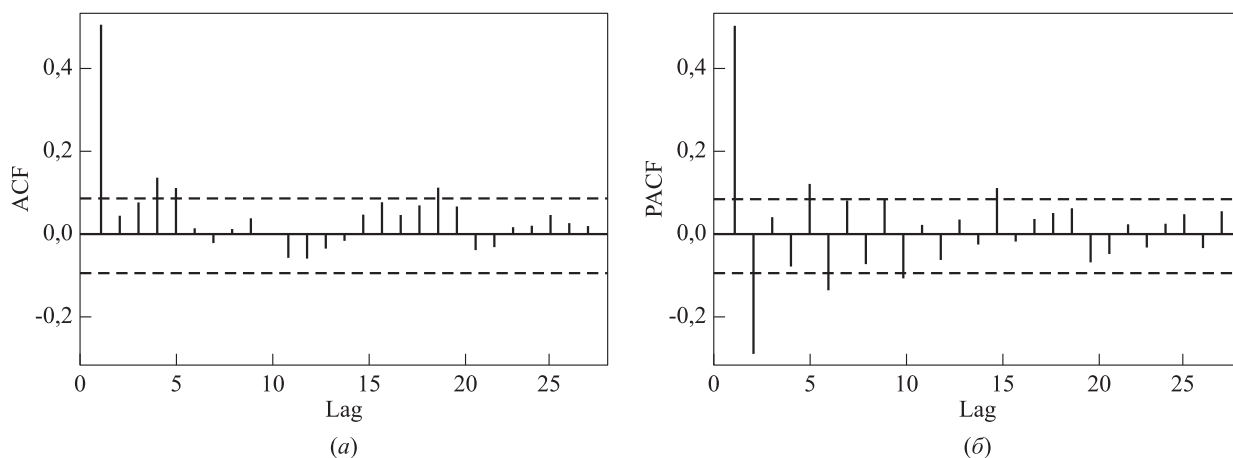


Рис. 2 Графики ACF (а) и PACF (б) для $Y_t = \ln(S_t/S_{t-2})$ (Light Sweet Crude Oil)

$$AIC = 2k - 2 \ln(L_n(\hat{v})) ;$$

$$BIC = k \ln(N) - 2 \ln(L_n(\hat{v})) ,$$

где N — объем выборки; k — число параметров; $L_n(\hat{v})$ — значение функции правдоподобия для найденных оптимальных параметров \hat{v} . Таблицы 1 и 2 показывают, что асимметричные распределения EGB2 и JSU лучше моделируют ряд, чем нормальное распределение.

6 Ценообразование опционов

В рамках данной работы справедливая стоимость опционов оценивалась с помощью метода Монте-Карло в соответствии с формулами (5) и (6) по риск-нейтральным траекториям моделей ARIMA–GARCH (2)–(4), где число реализаций доходности базового актива $M = 10\,000$. Эффектив-

ность каждой модели ARIMA–GARCH оценивалась по абсолютной ошибке (AE):

$$AO(\text{Moneyness}) = |p_{\text{call}}^m / p_{\text{put}}^m(\text{Moneyness}) - p_{\text{call}} / p_{\text{put}}(\text{Moneyness})|,$$

где $p_{\text{call}}^m / p_{\text{put}}^m$ — рыночные котировки опционов; $\text{Moneyness} = X / s_0$.

Абсолютные ошибки оценивания опционов иллюстрируются на рис. 3–6. Для опционов на индекс DAX наиболее близкие значения к рыночным котировкам дает модель с распределением ошибок JSU, модель с нормальным распределением хуже всего находит стоимость опционов DAX. При этом стоит также отметить, что для опционов с датой экспирации 22 декабря 2023 г. расхождение между моделями становится существенней, показывая неэффективность использования моделей с нормальным распределением ошибок на длинных временных горизонтах.

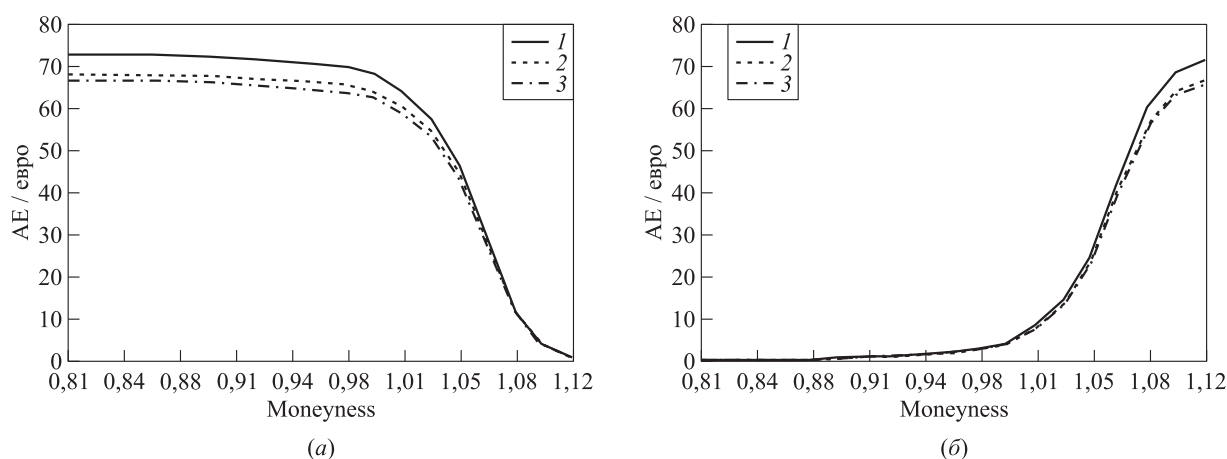


Рис. 3 Абсолютные ошибки цен опционов CALL (а) и PUT (б) на индекс DAX со сроком 22 июня 2019 г.: 1 — ARIMA–GARCH–N; 2 — ARIMA–GARCH–EGB2; 3 — ARIMA–GARCH–JSU

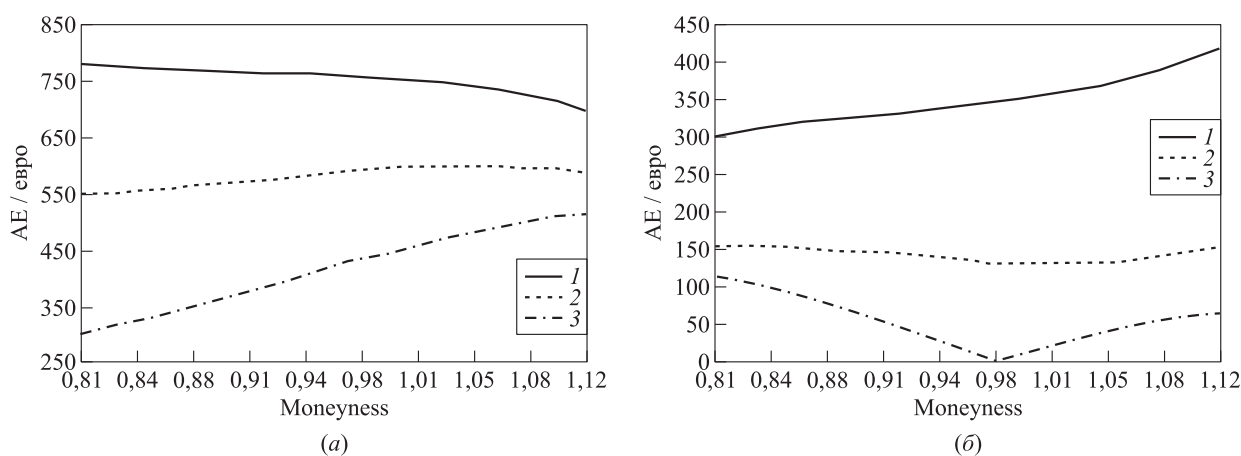


Рис. 4 Абсолютные ошибки цен опционов CALL (а) и PUT (б) на индекс DAX со сроком 22 декабря 2023 г.: 1 — ARIMA–GARCH–N; 2 — ARIMA–GARCH–EGB2; 3 — ARIMA–GARCH–JSU

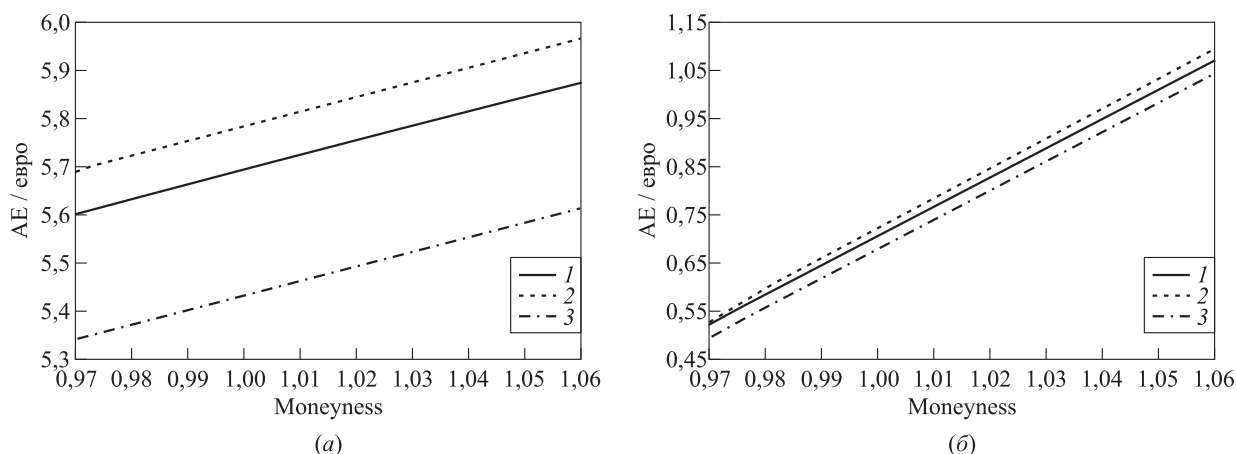


Рис. 5 Абсолютные ошибки цен опционов CALL (а) и PUT (б) Light Sweet Crude Oil (20 июня 2019 г.): 1 – ARIMA–GARCH–N; 2 – ARIMA–GARCH–EGB2; 3– ARIMA–GARCH–JSU

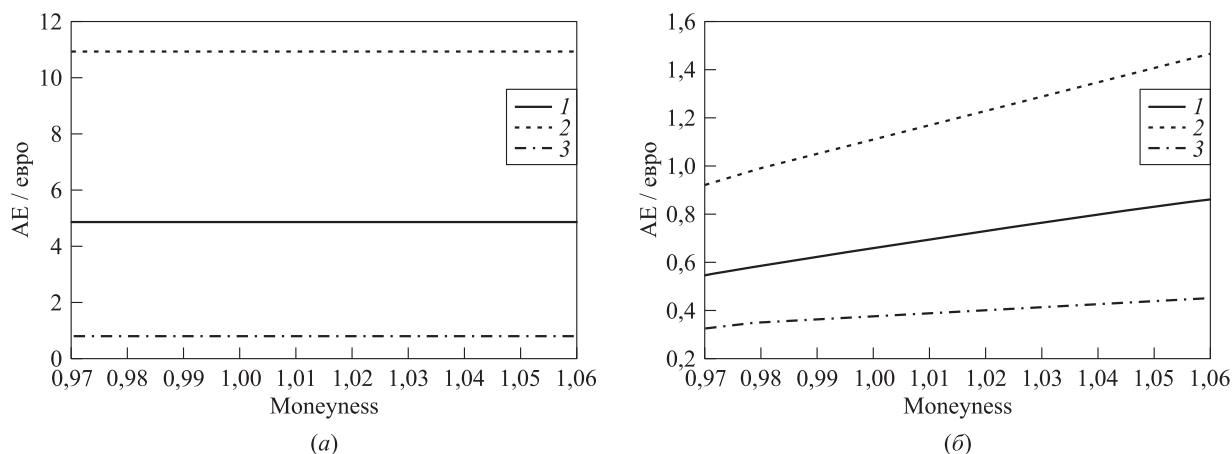


Рис. 6 Абсолютные ошибки цен опционов CALL (а) и PUT (б) Light Sweet Crude Oil (22 июня 2020 г.): 1 – ARIMA–GARCH–N; 2 – ARIMA–GARCH–EGB2; 3– ARIMA–GARCH–JSU

Результаты оценки справедливой стоимости опционов на индекс DAX аналогичны результатам работы [7], в которой среди рассматриваемых опционов присутствовали опционы на индекс S&P 500 и делался вывод о том, что модели ARIMA–GARCH с ошибками, распределенными по закону EGB2, дают лучшие оценки стоимости опционов в деньгах (для опционов CALL Moneyness < 1, для PUT Moneyness > 1) по сравнению с моделями, где ошибки распределены нормально.

Опционы на базовый актив Light Sweet Crude Oil также характеризуются существенным преимуществом модели JSU. Однако теперь уже модель с распределением EGB2 показывает результаты хуже, чем модель с нормальным распределением.

7 Заключение

В данной статье были рассмотрены три альтернативные модели временных рядов с условным

средним ARIMA и условной дисперсией GARCH структуры. Оценка справедливой стоимости опционов проводилась по методу Монте-Карло (5), (6) на основе риск-нейтральной динамики моделей в соответствии с формулами (2)–(4). Калибровка моделей осуществлялась методом квазикасимального правдоподобия в соответствии с соотношениями (7) и (8). Определение порядка моделей было выполнено на основе Q-теста Льюнга–Бокса и графиков коррелограмм ACF и PACF.

Результаты эмпирических исследований показывают, что на небольших временных промежутках модели обеспечивают близкие значения справедливой стоимости опционов. Для опционов с дальней датой экспирации (больше года) модели показывают существенно разные результаты. Модель, построенная для распределения JSU, дает оценки стоимости, максимально близкие к ценам закрытия биржевых торгов для всех рассматриваемых опционных контрактов.

Литература

1. Данилишин А. Р., Голембиовский Д. Ю. Риск-нейтральная динамика для модели ARIMA–GARCH с ошибками, распределенными по закону S_U Джонсона // Информатика и её применения, 2020. Т. 14. Вып. 1. С. 48–55.
2. Bollerslev T. A conditionally heteroskedastic time series model for speculative prices and rates of return // Rev. Econ. Stat., 1987. Vol. 69. Iss. 3. P. 542–547. doi: 10.2307/1925546.
3. Akgiray V. Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts // J. Bus., 1989. Vol. 62. Iss.1. P. 55–80. doi: 10.1086/296451.
4. Follmer H., Schied A. Stochastic finance: An introduction in discrete time. — Berlin: Walter de Gruyter, 2002. 422 p.
5. Terasvirta T. An introduction to univariate GARCH models // Handbook of financial time series / Eds. T. G. Andersen, R. A. Davis, J.-P. Kreiss, Th. V. Mikosch. — Berlin–Heidelberg: Springer, 2009. Vol. 10. P. 17–42. doi: 10.1007/978-3-540-71297-8_1.
6. Elliott R. J., Madan D. B. A discrete time equivalent martingale measure // Math. Financ., 1998. Vol. 8. Iss. 2. P. 127–152. doi: 10.1111/1467-9965.00048.
7. Yi Xi. Comparison of option pricing between ARMA-GARCH and GARCH-M models. — London, Ontario, Canada: University of Western Ontario, 2013. MoS Thesis. 73 p.
8. Christian F., Francq M. GARCH models: Structure, statistical inference and financial applications. — New York, NY, USA: Wiley, 2019. 504 p.
9. Boyle T. A Monte Carlo approach // J. Financ. Econ., 2012. Vol. 4. P. 323–338. doi: 10.1016/0304-405x(77)90005-8.
10. Hull J. Options, futures, and other derivatives. — 10th ed. — Pearson, 2018. 896 p.
11. Williams D. Probability with martingales. — Cambridge: Cambridge University Press, 1991. 251 p.
12. Bell D. Transformations of measure on an infinite dimensional vector space // Seminar on stochastic processes, 1990 / Eds. E. Çinlar, P. J. Fitzsimmons, R. J. Williams. — Progress in probability book ser. — Boston, MA, USA: Birkhäuser, 1991. Vol. 24. P. 15–25. doi: 10.1007/978-1-4684-0562-0_3.

Поступила в редакцию 01.10.19

ESTIMATING THE FAIR VALUE OF OPTIONS BASED ON ARIMA–GARCH MODELS WITH ERRORS DISTRIBUTED ACCORDING TO THE JOHNSON’S S_U LAW

A. R. Danilishin¹ and D. Yu. Golembiovsky^{1,2}

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

²Department of Banking, Sinergy University, 80-G Leningradskiy Prosp., Moscow 125190, Russian Federation

Abstract: In continuation of the article “Risk-neutral dynamics for the ARIMA–GARCH (autoregressive integrated moving average – generalized autoregressive conditional heteroskedasticity) random process with errors distributed according to the Johnson’s S_U law,” this paper presents the experimental results for the ARIMA–GARCH (autoregressive integrated moving average – generalized autoregressive conditional heteroskedasticity) models with normal (N), exponential beta of the second type (EGB2), and S_U Johnson (JSU) error distributions. The fair value of European options is estimated by the Monte-Carlo method based on the results obtained in the specified article by using the extended Girsanov principle. The parameters of the ARIMA–GARCH-N, ARIMA–GARCH-EGB2, and ARIMA–GARCH-JSU models were found by the quasi-maximum likelihood method. The efficiency of the resulting risk-neutral models was studied using the example of European exchange-traded options PUT and CALL on basic assets DAX and Light Sweet Crude Oil.

Keywords: ARIMA; GARCH; risk-neutral measure; Girsanov extended principle; Johnson’s S_U distribution; option pricing

DOI: 10.14357/19922264200412

References

1. Danilishin, A. R., and D. Y. Golembiovsky. 2020. Risk-neutral’naya dinamika dlya ARIMA–GARCH modeli s oshibkami, raspredelennymi po zakonu S_U Dzhonsona. [Risk-neutral dynamics for ARIMA–GARCH random process with errors distributed according to the Johnson’s S_U law]. *Informatika i ee Primeneniya — Inform. Appl.* 14(1):56–62.
2. Bollerslev, T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of re-

- turn. *Rev. Econ. Stat.* 69(3):542–547. doi: 10.2307/1925546.
3. Akgiray, V. 1989. Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. *J. Bus.* 62(1):55–80. doi: 10.1086/296451.
 4. Follmer, H., and A. Schied. 2002. *Stochastic finance: An introduction in discrete time*. Berlin: Walter de Gruyter. 422 p.
 5. Terasvirta, T. 2009. An introduction to univariate GARCH models. *Handbook of financial time series*. Eds. T.G. Andersen, R.A. Davis, J.-P. Kreiss, and Th. V. Mikosch. Berlin–Heidelberg: Springer. 10:17–42. doi: 10.1007/978-3-540-71297-8_1.
 6. Elliott, R. J., and D. B. Madan. 1998. A discrete time equivalent martingale measure. *Math. Financ.* 8(2):127–152. doi: 10.1111/1467-9965.00048.
 7. Yi, Xi. 2013. Comparison of option pricing between ARMA–GARCH and GARCH–M models. London, Ontario, Canada: University of Western Ontario. MoS Thesis. 73 p.
 8. Christian, F., and M. Francq. 2019. *GARCH models: Structure, statistical inference and financial applications*. New York, NY: Wiley. 504 p.
 9. Boyle, P. 2012. Options: A Monte Carlo approach. *J. Financ. Econ.* 4(3):323–338. doi: 10.1016/0304-405x(77)90005-8.
 10. Hull, J. 2018. *Options, futures, and other derivatives*. 10th ed. Pearson. 896 p.
 11. Williams, D. 1991. *Probability with martingales*. Cambridge: Cambridge University Press. 251 p.
 12. Bell, D. 1991. Transformations of measure on an infinite dimensional vector space. *Seminar on stochastic processes, 1990*. Eds. E. Çinlar, P.J. Fitzsimmons, and R. J. Williams. Progress in probability book ser. Boston, MA: Birkhäuser. 24:15–25. doi: 10.1007/978-1-4684-0562-0_3.

Received October 1, 2019

Contributors

Danilishin Artem R. (b. 1992) — PhD student, Department of Operations Research, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; danilishin-artem@mail.ru

Golembiovsky Dmitry Y. (b. 1960) — Doctor of Science in technology, professor, Department of Operation Research, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, 1-52 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; professor, Department of Banking, Sinergy University, 80-G Leningradskiy Prosp., Moscow 125190, Russian Federation; golem@cs.msu.su

ПРИМЕНЕНИЕ МНОГОМАСШТАБНОГО ПОДХОДА И МЕТОДОВ АНАЛИЗА ДАННЫХ ДЛЯ МОДЕЛИРОВАНИЯ ТЕПЛОПРОВОДНОСТИ В СЛОИСТЫХ СТРУКТУРАХ*

К. К. Абгарян¹, И. С. Колбин²

Аннотация: Моделирование тепловых свойств слоистых структур в настоящее время стало востребованным направлением научных исследований. Это связано с постоянно растущей скоростью работы микроэлектронных элементов на основе слоистых структур, выделяющих при работе все большее количество энергии в виде тепла, которое требуется отводить, чтобы избежать перегрева и потери функциональных свойств устройств. В работе представлен интеграционный подход, позволяющий объединить методы многомасштабного моделирования и анализа данных. Применение данного подхода дает возможность получить новое качество при решении задачи построения модели теплопереноса в двухслойной структуре GaAs/AlAs. Показана эффективность применения методов машинного обучения для анализа зависимости эффективного коэффициента теплопроводности слоистых материалов от структурных особенностей и внешних факторов. Развитие предложенного подхода сможет обеспечить формирование информации для обоснованного подбора материалов слоистых структур для микроэлектронных устройств.

Ключевые слова: многомасштабное моделирование; интеграционный подход; слоистые структуры; предсказательное моделирование; кинетическое уравнение Больцмана; коэффициент теплопроводности; методы анализа данных

DOI: 10.14357/19922264200413

1 Введение

В настоящее время стремительное развитие машинного обучения в качестве мощного метода интеграции данных с множественной точностью и выявления корреляций между взаимосвязанными явлениями дает возможность за ограниченное время находить решение сложных задач в разных предметных областях. Технологии машинного обучения получили существенный импульс в развитии за последнее десятилетие. В настоящее время ведутся активные исследования в области применения алгоритмов машинного обучения в задачах материаловедения [1], например при исследовании критических явлений или внутренних корреляций квантовой системы методом искусственных нейронных сетей, а также при определении параметров потенциалов межатомного взаимодействия и моделировании динамики решеток с использованием машинно-обученных (полу)эмпирических потенциалов [2]. Однако классические методы машинного обучения часто игнорируют фундаментальные законы физики, что приводит к некорректным задачам или нефизичным решениям.

Сегодня можно говорить о том, что многомасштабное моделирование — это успешная стратегия интеграции мультимасштабных, многофизических данных, которая позволяет раскрыть механизмы, объясняющие появление функциональных зависимостей при изучении физических явлений и процессов [1]. Применение технологии математического многомасштабного моделирования [3, 4], согласно которой расчеты на каждом уровне проводятся с использованием соответствующих математических моделей и вычислительных алгоритмов, позволяет:

- объяснить многие явления и свойства объектов, включая исследование структурных особенностей физических явлений и процессов на нескольких масштабах;
- получать качественно новые результаты в области прогнозирования свойств новых объектов;
- решать задачи оптимизации состава и структуры многомасштабных объектов, выстраивать взаимосвязи между структурой и свойствами, что дает возможность синтезировать компози-

* Работа выполнена при частичной финансовой поддержке РФФИ (проекты 19-29-03051 мк и 19-08-01191 А).

¹ Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; Московский авиационный институт (национальный исследовательский университет), kristal83@mail.ru

² Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, iskolbin@gmail.com

ционные структуры, обладающие заданным набором свойств.

При этом необходимо учитывать, что одно многомасштабное моделирование часто не дает возможности эффективно комбинировать большие наборы данных из разных источников и с разных масштабных уровней. Авторы последних научных публикаций, включая [1], показывают, как машинное обучение и многомасштабное моделирование могут дополнять друг друга, создавая надежные прогностические модели. Они базируются на подходах, основанных на теоретическом физико-математическом моделировании с применением математического аппарата обыкновенных дифференциальных уравнений, уравнений в частных производных, а также на подходах, основанных на методах анализа данных. Для достижения поставленных целей используется опыт, накопленный в прикладной математике, информатике, вычислительной биологии, биохимии, биофизике, медицине и в других областях.

Таким образом, как отмечается в [1], междисциплинарный подход предполагает, что интеграция машинного обучения и многомасштабного моделирования может дать новое представление о механизмах создания новых материалов, процессах разработки новых лекарств, о причинах заболеваний, помочь в определении новых целей и стратегий лечения, а также в принятии решений на благо здоровья человека.

Использование интеграции методов машинного обучения и технологий многомасштабного моделирования (интеграционный подход) стало на сегодня одним из наиболее перспективных и быстроразвивающихся трендов в области материаловедения.

Необходимо отметить, что при физико-математическом моделировании структурных свойств материалов необходимо получить максимально точные характеристики, используя при этом минимальный объем эмпирических данных. Подобными свойствами обладают квантово-механические расчеты из первых принципов [5]. Однако они отличаются существенной вычислительной сложностью, в связи с чем рассматриваемые системы, как правило, ограничены размером в сотни, иногда тысячи атомов. В данной работе продемонстрировано, что подходы машинного обучения дают возможность проводить целенаправленные расчеты с точностью, приближенной к первопринципным расчетам для систем из миллионов атомов на существующем в настоящий момент оборудовании за приемлемое время [6].

В данной работе для решения задачи о теплопереносе в слоистых структурах была разработана

многомасштабная модель и на ее основе многомасштабная композиция, в которой были задействованы методы машинного обучения. Показано, что такой интеграционный подход позволяет существенно ускорить вычислительно затратные расчеты по получению функциональных зависимостей между внешними параметрами среды, а также внутренними параметрами слоистых структур и значением эффективного коэффициента теплопроводности.

2 Модель теплопереноса в слоистых структурах

Моделирование тепловых свойств слоистых структур в настоящее время стало востребованным направлением научных исследований, что связано с постоянно растущей скоростью работы микроэлектронных элементов на основе слоистых структур, выделяющих все большее количество энергии в виде тепла, которое требуется отводить, чтобы избежать перегрева и потери функциональных свойств устройств [7].

Цель настоящей работы заключается в применении интеграционного подхода к задаче моделирования зависимости эффективного коэффициента теплопроводности слоистых структур, а именно сверхрешеток, от варьируемых параметров материала и внешней среды. Для этого применялось прямое многомасштабное физико-математическое моделирование, проводилась генерация выборки на основе модели молекулярного подавления [8] с дальнейшим обучением нейронных сетей на ней. На выходе алгоритма были получены компактные нейросетевые модели, которые сравнивались между собой по точности работы на тестовом наборе данных для определения лучшей сети.

Основными переносчиками тепла в полупроводниках и диэлектриках служат квазичастицы-фононы, описывающие колебания кристаллической ячейки. Стоит отметить, что применение классических подходов к решению задач теплопроводности на основе закона Фурье для рассматриваемых слоистых структур дает неудовлетворительные результаты, так как при подобном подходе игнорируются квантово-механические эффекты в материалах, что дает сильное рассогласование с экспериментальными данными [7]. Для построения моделей теплопереноса в слоистых структурах свою эффективность показали методы на основе решения кинетического уравнения Больцмана для фононов.

В отсутствие температурного градиента и иных термодинамических сил система находится в тепло-

вом равновесии и распределение фононов подчинено закону Бозе–Эйнштейна:

$$f_0(\omega, T) = \frac{1}{e^{\hbar\omega/(kT)} - 1}.$$

При наличии теплового градиента распределение фононов может быть описано с помощью кинетического уравнения Больцмана:

$$\frac{df}{dt} = \frac{df}{dt}(\text{diffusion}) + \frac{df}{dt}(\text{scattering}) = 0,$$

где

$$\frac{df}{dt}(\text{diffusion}) = -\nabla T v \frac{df}{dT}.$$

Кинетическое уравнение Больцмана относится к сложным интегродифференциальным уравнениям. Для достаточно небольшого температурного градиента распределение фононов может быть выражено в приближении времени релаксации:

$$\frac{f - f_0}{\tau^0} = -\nabla T v \frac{df_0}{dT}.$$

При этом одна из проблем при построении вычислительных алгоритмов связана с учетом рассеяния фононов. Однако во многих случаях при решении данного уравнения достаточно учитывать лишь приближения времени релаксации, что существенно упрощает задачу [9]. В такой постановке необходимо решить вопрос, связанный с расчетом данных по параметрам релаксации. Ранее они вычислялись полужемпирически с учетом согласования модельных расчетов с результатами экспериментов.

Несмотря на определенные успехи данный подход весьма трудоемок. В связи с этим его в основном применяли для моделирования структур однокомпонентных материалов (как правило, кремния и германия) [10, 11].

Значительный прорыв в данном вопросе произошёл при комбинировании методов с использованием кинетического уравнения с первопринципными расчетами [8, 9, 12]. При этом требуемые характеристики фононов могут быть получены не из аппроксимации экспериментальных данных, а из первопринципных квантово-механических расчетов, что значительно повышает точность вычислений, открывает возможности эффективного предсказания свойств моделируемых материалов, минимизируя различные допущения. Данные, которые получают, — это координаты базисных атомов кристаллической ячейки, межатомные силовые константы для моделирования двухфононных взаимодействий, межатомные силовые константы для моделирования трехфононных

взаимодействий, диэлектрический тензор и эффективные заряды Борна для неаналитического поправочного члена.

Более точное моделирование можно получить с использованием методов молекулярной динамики, однако это сопряжено с высокой вычислительной сложностью и нетривиальной задачей подбора оптимального потенциала [13, 14] и значений его параметров под конкретный материал.

Рассмотрим вычисление эффективного коэффициента теплопроводности бинарной гетероструктуры на примере сверхрешетки GaAs/AlAs. Для этого применим подход, описанный в работе [8]. Расчет ведется в приближении виртуального кристалла [15, 16], коэффициент релаксации имеет вид:

$$\frac{1}{\tau_\lambda^0} = \frac{1}{\tau_\lambda^{3\text{ph}+}} + \frac{1}{\tau_\lambda^{3\text{ph}-}} + \frac{1}{\tau_\lambda^a} + \frac{1}{\tau_\lambda^b}.$$

Здесь λ — обобщенный индекс (фононная мода), который объединяет информацию по поляризации фононов p и по волновому вектору \mathbf{q} [16]; $\tau_\lambda^{3\text{ph}+}$ описывает процессы адсорбции: один фонон из двух падающих [15, 16]; $\tau_\lambda^{3\text{ph}-}$ описывает процессы эмиссии, в ходе которых один падающий фонон разделяется на два [15, 16], τ_λ^a — сплавной член, зависит от послойного распределения компонентов [17]; τ_λ^b — барьерный член [17].

Рассматривается послойное распределение материала. Для моделирования распределения материалов при росте сверхрешетки использовалась модель Мураки [18]:

$$\tilde{X}(i) = \begin{cases} 1, & i < 1; \\ 1 - \phi(1 - R^i), & 1 \leq i \leq n_{\text{ML}}; \\ 1 - \phi(1 - R^{n_{\text{ML}}})R^{i-n_{\text{ML}}}, & i \geq n_{\text{ML}}. \end{cases}$$

Здесь R — варьируемый параметр модели; n_{ML} — число монослоев первого материала в периоде сверхрешетки.

Схема расчета параметров релаксации представлена на рис. 1.

Для расчета эффективного коэффициента теплопроводности использовалась модель модального подавления [18]:

$$\kappa(L) = \sum_{\lambda} S_{\lambda} C_{\lambda} \|v_{\lambda}\| \Lambda_{\lambda} \cos^2(\theta_{\lambda}).$$

Здесь

$$S_{\lambda} = \frac{1}{1 + 2K_{\lambda}}; \quad \Lambda_{\lambda} = \|v_{\lambda}\| \tau_{\lambda}^0; \quad K_{\lambda} = \frac{\Lambda_{\lambda}(\cos(\theta_{\lambda}))}{L};$$

$$C_{\lambda} = \frac{k_B}{N\Omega} \frac{\hbar\omega_{\lambda}}{k_B T} f_0(f_0 + 1), \quad f_0 = f_0(\omega_{\lambda}, T);$$

θ_{λ} — угол между групповой скоростью v_{λ} фононной моды λ и осью теплопереноса.

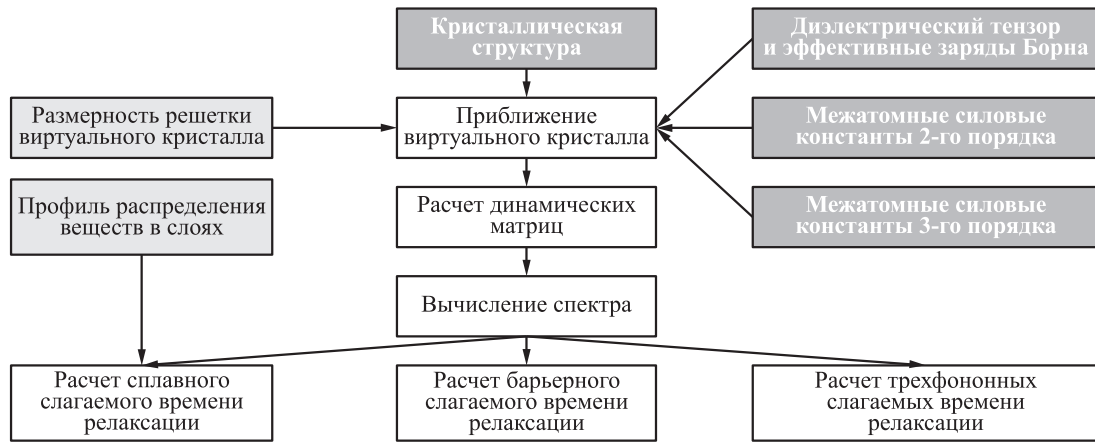


Рис. 1 Схема расчета параметров релаксации сверхрешетки

3 Формирование обучающей выборки

Обучающая выборка сформирована по результатам расчетов эффективного коэффициента теплопроводности в пакете Alma VTE с варьированием следующих параметров:

- R — параметр модели Мураки, отвечающий за по- слойное распределение материалов в периоде сверхрешетки, варьировался от 0 до 0,9;
- x — число монослоев первого материала (GaAs), варьировалось от 1 до 20;
- y — число монослоев второго материала (AlAs), варьировалось от 1 до 20;
- T — температура окружающей среды, варьировалась от 100 до 500 К;
- L — толщина сверхрешетки, варьировалась от 1 нм до 100 мкм.

4 Многомасштабная композиция для моделирования теплопроводности в слоистых структурах

В данной работе при построении многомасштабной композиции для моделирования теплопроводности в слоистых структурах используются подходы, разработанные ранее и представленные в [3, 4].

Покажем, как с помощью данной технологии при решении задачи о применении интеграционных подходов к моделированию теплопроводности в слоистых структурах из конкретных базовых моделей-композиций (БК) составляются

многомасштабные композиции (МК) — вычислительные аналоги многомасштабных моделей, передающие информационную сущность многомасштабных вычислительных процессов [3, 4]. При описании многомасштабной композиции и базовых моделей-композиций, например $MC_{1,1}^{A_i}$: $\{V_0^{A_i}, X_0^{A_i}, MA_0^{A_i}\}$, представляющих собой однопараметрическое семейство множеств разного структурного типа, объединенных в общем вычислительном процессе, где $V_0^{A_i}$ — множество входных данных; $X_0^{A_i}$ — множество выходных данных; $MA_0^{A_i}$ — множество, состоящее из моделей и алгоритмов, используем упрощенную форму записи:

$$MC_{1,1}^{A_i} : \{V_0^{A_i}, X_0^{A_i}, MA_0^{A_i}\} = MC_{1,1}^{A_i}.$$

Рассмотрим расчет теплопроводности бинарной гетероструктуры GaAs/AlAs для разных периодов сверхрешетки. При построении многомасштабной композиции

$$MK_{0,1,2}^{(A_{a_i}^i A_{a_j}^j / A_{a_k}^k A_{a_j}^j)} = MK_{0,1,2}^{(GaAs/AlAs)}$$

выделим три масштабных уровня (0, 1, 2) и используем следующие обозначения.

Базовая модель-композиция — $MC_0^{A_i}$ (i — номер элемента в таблице Менделеева), поставим ей в соответствие:

- EI_0^{31} — экземпляр для описания характеристик химического элемента Ga ($i = 31$);
- EI_0^{33} — экземпляр для описания характеристик химического элемента As ($i = 33$);
- EI_0^{13} — экземпляр для описания характеристик химического элемента Al ($i = 13$);

$MC_{1,1}^{GaAs}$ — экземпляр соответствующей БК для расчета кристаллохимической структуры GaAs;

$MC_{1,1}^{AlAs}$ — экземпляр БК для расчета кристаллохимической структуры AlAs;

$MC_{1,2}^{GaAs}$ — экземпляр БК для квантово-механического расчета GaAs;

$MC_{1,2}^{AlAs}$ — экземпляр БК для квантово-механического расчета AlAs;

$MC_{2,3}^{GaAs/AlAs}$ — экземпляр БК для решения уравнения Больцмана;

$K_{2,4}^{GaAs/AlAs}$ — экземпляр композиции (формируется за счет объединения БК с одного масштабного уровня, в данном случае — 2-го) для определения регрессионной функции.

Общее представление о распределении БК по масштабным уровням, задействованным в вычислительном процессе, показано в таблице.

Покажем, как организован вычислительный процесс.

На нулевом масштабном уровне с помощью БК MC_0^1 «АТОМ» задаются основные данные по химическим элементам, входящим в состав соединений, участвующих в вычислительном процессе (атомный номер химического элемента, масса атома, заряд ядра, радиус атома, электронная конфигурация, структура решетки и др.). В данном случае формируются экземпляры БК MC_0 «АТОМ Ga_0^{31} », MC_0 «АТОМ As_0^{33} » и MC_0 «АТОМ Al_0^{13} ».

К первому масштабному уровню отнесены БК MC_1^1 «КРИСТАЛЛОХИМИЧЕСКАЯ ФОРМУЛА» и MC_1^2 «КВАНТОВО-МЕХАНИЧЕСКАЯ ЯЧЕЙКА». С их помощью данные, полученные с нулевого масштабного уровня из MC_0 «АТОМ Ga_0^{31} », MC_0 «АТОМ As_0^{33} » и MC_0 «АТОМ Al_0^{13} », передаются на квантово-механический уровень, первоначально в MC_1^1 «КРИСТАЛЛОХИМИЧЕСКАЯ

ФОРМУЛА», где, используя знания о химическом составе и кристаллографической структуре, определяется кристаллохимическая структура соединения (метрические параметры кристаллической решетки, координаты базисных атомов и др.). Данная БК программно реализована в двух расчетных модулях (модель «Плотная упаковка» и программный комплекс Materials Studio (<https://www.3dsbiovia.com/products/collaborative-science/biovia-materials-studio>)).

Далее полученные в ходе вычислительного процесса данные передаются в БК MC_1^2 «КВАНТОВО-МЕХАНИЧЕСКАЯ ЯЧЕЙКА». Данная БК программно реализована в двух расчетных модулях (программный комплекс VASP (<https://www.vasp.at>) и пакет программ с открытым кодом Quantum Espresso (<https://www.quantum-espresso.org>)). Здесь на базе квантово-механической теории по каждому слою бинарной гетероструктуры уточняются параметры кристаллической решетки, рассчитывается электронная плотность [7], полная энергия для заданной конфигурации базисных атомов, значения диэлектрических тензоров, эффективные заряды Борна, силовые константы 2-го и 3-го порядка рассчитываются как производные соответствующих порядков от энергии.

На втором масштабном уровне, используя данные, полученные на предыдущем уровне, применяется модель «виртуального кристалла», в которой двухслойная структура рассматривается как сплав, однако при этом дополнительно учитывается послойное распределение материалов (модель Мураки) [17]. Для расчета теплопроводности используется уравнение Больцмана с добавлением барьерного члена для учета приближения времени релаксации. Базовая модель-композиция MC_2^3 «УРАВНЕНИЕ БОЛЬЦМАНА» программно реализована в расчетном модуле, в основе которого лежит комплекс AlmaBTE (<https://almabte.bitbucket.io>).

Базовые модели-композиции, входящие в состав $MK_{0,1,2}^{(A_{a_i}^i A_{a_j}^j / A_{a_k}^k A_{a_j}^j)}$

№ масштабного уровня	Обозначение и название БК	Название масштабного уровня
0	$MC_0^{A_i} = MC_0$ «АТОМ A_0^i »	Уровень химических элементов
1	$MC_{1,1}^{A_i A_j} = MC_1^1$ «КРИСТАЛЛОХИМИЧЕСКАЯ ФОРМУЛА» $MC_{1,2}^{A_i A_j} = MC_1^2$ «КВАНТОВО-МЕХАНИЧЕСКАЯ ЯЧЕЙКА»	Квантово-механический уровень
2	$MC_{2,3}^{GaAs/AlAs} = MC_{2,3}^{A_i A_j / A_{sk}^k A_j^j} = MC_2^3$ «УРАВНЕНИЕ БОЛЬЦМАНА» $K_{2,4}^{GaAs/AlAs} = K_{2,4}^{A_i A_j / A_{sk}^k A_j^j} = K_2^4$ «АППРОКСИМАЦИОННАЯ МОДЕЛЬ»	Наноуровень

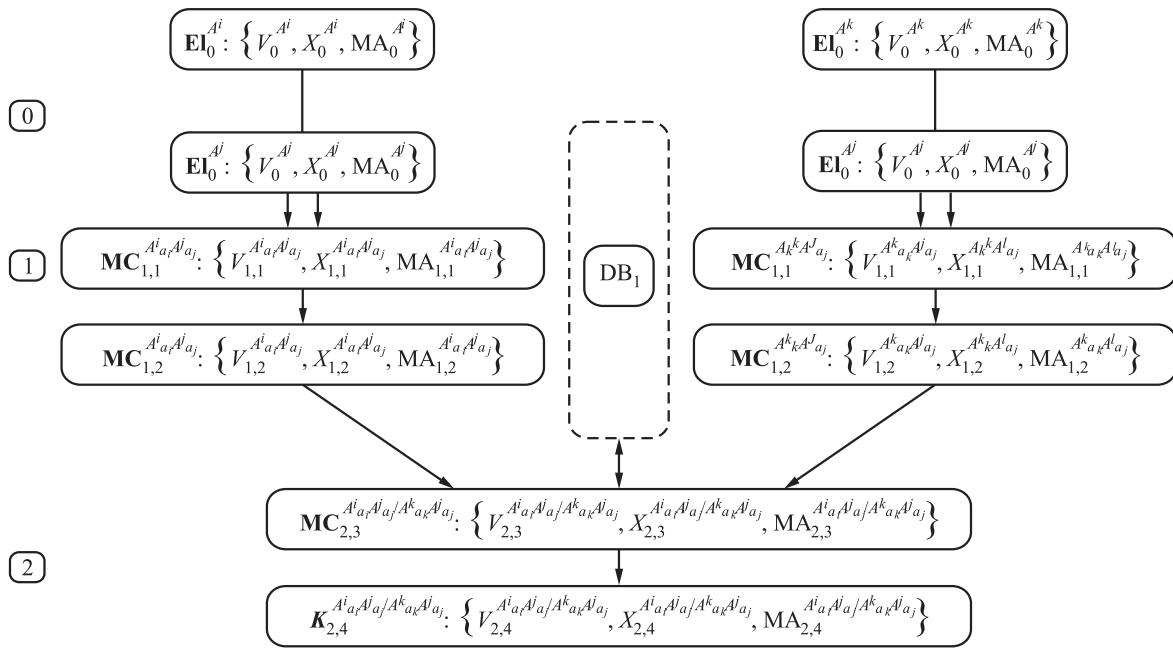


Рис. 2 Схематическое представление многомасштабной композиции $МК_{0,1,2}^{(A^i_{a_i} A^j_{a_j} / A^k_{a_k} A^j_{a_j})}$ для моделирования теплопереноса в слоистой структуре GaAs/AlAs

В результате расчетов получаем значения эффективного коэффициента теплопроводности для двухслойной структуры в зависимости от периода (числа слоев GaAs и AlAs, т.е. распределения вещества) бинарной структуры, от ширины образца, от температуры внешней среды. Формируем набор расчетных данных, варьируя внешнюю температуру (например, от 100 до 500 К), меняем размеры периода, распределение веществ в наногетероструктуре GaAs/AlAs, общую ширину сверхрешетки. Расчеты проводятся во вложенных циклах, таким образом собирается выборка.

Полученные наборы данных передаются в экземпляр композиции $K_{24}^{GaAs/AlAs}$, которая формируется за счет объединения БК с одного масштабного уровня, в данном случае 2-го, и служит для определения регрессионной функции, описывающей функциональную зависимость изменения эффективного коэффициента теплопроводности бинарной структуры от варьируемых параметров. Определяется область допустимых значений, которые может принимать искомая функция. Используются программные модули, в основе которых лежат пакеты TensorFlow (<https://www.tensorflow.org/>) и PyTorch (<https://pytorch.org>). На рис. 2 представлена архитектурная схема расчетных модулей многомасштабной композиции $МК_{0,1,2}^{(A^i_{a_i} A^j_{a_j} / A^k_{a_k} A^j_{a_j})}$ и основные потоки данных.

5 Результаты вычислений, выводы

Были построены нейросетевые модели для расчета эффективного коэффициента теплопроводности для слоистых структур — сверхрешеток GaAs/AlAs с разными периодами слоев. Данные для обучения были сгенерированы в программном пакете AlmaBTE 1.3.2, параметры материалов получены из открытой базы данных проекта. Выборка формировалась для различных комбинаций содержания GaAs и AlAs, толщин пленок, периодов сверхрешетки. Полученный массив данных был разделен на 3 части: 60% для обучения нейросетей, 20% для валидации (во избежание переобучения) и 20% как тестовая выборка для оценки результирующей точности обученных моделей. Оптимизация нейросетей велась с использованием алгоритма RMSprop с шагом 0,0001 в среде Tensorflow 2.3.

В качестве модели для расчета использовались многослойные нейронные сети прямого распространения. В работе были рассмотрены сети с различным числом скрытых слоев, также проведены сравнения для варьирующего числа нейронов и распространенных активационных функций. Полученные в результате обучения сети сравнивались по среднеквадратичной ошибке. Относительная сред-

неквадратичная ошибка при этом составила порядка 3%–5%, что позволяет говорить о достаточно хорошей точности выбранного подхода.

Разработанные подходы могут быть использованы при решении обратных задач для предсказательного моделирования структурных характеристик слоистых материалов с заданными значениями эффективного коэффициента теплопроводности.

Литература

1. Mark A., Tepole A. B., Cannon W. R., et al. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences // *NPJ Digital Medicine*, 2019. Vol. 2. Iss. 1. Art. No. 115. 11 p. doi: 10.1038/s41746-019-0193-y.
2. Ladygin V., Korotaev P., Yanilkin A., Shapeev A. Lattice dynamics simulation using machine learning interatomic potentials // *Comp. Mater. Sci.*, 2020. Vol. 172. Art. ID: 109333.
3. Абгарян К. К. Многомасштабное моделирование в задачах структурного материаловедения. — М.: МАКС Пресс, 2017. 284 с.
4. Абгарян К. К. Информационная технология построения многомасштабных моделей в задачах вычислительного материаловедения // *Системы высокой доступности*, 2018. Т. 15. № 2. С. 9–15.
5. Kohn W., Sham L. J. Self-consistent equations including exchange and correlation effects // *Phys. Rev.*, 1965. Vol. 140. Iss. 4A. P. A1133–A1138. doi: 10.1103/PhysRev.140.A1133.
6. Jia Weile, Wang Han, Chen Mohan, et al. Pushing the limit of molecular dynamics with *ab initio* accuracy to 100 million atoms with machine learning. — Cornell University, 2020. arXiv:2005.00223 [physics.comp-ph]. <https://arxiv.org/pdf/2005.00223.pdf>.
7. Хвесьюк В. И., Скрябин А. С. Теплопроводность наноструктур // *Теплофизика высоких температур*, 2017. Т. 55. № 3. С. 446–471. doi: 10.7868/S0040364417030127.
8. Vermeersch B., Carrete J., Mingo N. Cross-plane heat conduction in thin films with *ab-initio* phonon dispersions and scattering rates // *Appl. Phys. Lett.*, 2016. Vol. 108. Iss. 19. Art. ID: 193104. doi: 10.1063/1.4948968.
9. Carrete J., Vermeersch B., Katre A., Roekeghem A., Wang T., Madsen G., Mingo N. AlmaBTE: A solver of the space–time dependent Boltzmann transport equation for phonons in structured materials // *Comput. Phys. Commun.*, 2017. Vol. 220C. P. 351–362. doi: 10.1016/j.cpc.2017.06.023.
10. Chung J. D., McGaughey A. J. H., Kaviani M. Role of phonon dispersion in lattice thermal conductivity modeling // *J. Heat Transf.*, 2004. Vol. 126. Iss. 3. P. 376–380. doi: 10.1115/1.1723469.
11. Loy J. M., Murthy J. Y., Singh D. A. Fast hybrid Fourier–Boltzmann transport equation solver for nongray phonon transport // *J. Heat Transf.*, 2012. Vol. 135. Iss. 1. Art. ID: 011008. doi: 10.1115/1.4007654.
12. Broido D. A., Malorny M., Birner G., Mingo N., Stewart D. A. Intrinsic lattice thermal conductivity of semiconductors from first principles // *Appl. Phys. Lett.*, 2007. Vol. 91. Iss. 23. Art. ID: 231922. doi: 10.1063/1.2822891.
13. Powell D. Elasticity, lattice dynamics and parameterisation techniques for the Tersoff potential applied to elemental and type III–V semiconductors. — Sheffield, U.K.: University of Sheffield, 2006. PhD Thesis. 252 p.
14. Abgaryan K. K., Mutigullin I. V., Uvarov S. I., Uvarova O. V. Multiscale modeling of clusters of point defects in semiconductor structures // *CEUR Workshop Proceedings / Eds. S. I. Smagin, A. A. Zatsarinnyy*. — Khabarovsk, 2019. Vol. 2426. P. 43–51.
15. Li W., Lindsay L., Broido D. A., Stewart D. A., Mingo N. Thermal conductivity of bulk and nanowire Mg₂SiSn_{1–x} alloys from first principles // *Phys. Rev. B*, 2012. Vol. 86. Iss. 17. Art. ID: 174307. doi: 10.1103/PhysRevB.86.174307.
16. Li W., Carrete J., Katcho N. A., Mingo N. A solver of the Boltzmann transport equation for phonons // *Comput. Phys. Commun.*, 2014. Vol. 185. Iss. 6. P. 1747–1758. doi: 10.1016/j.cpc.2014.02.015.
17. Carrete J., Vermeersch B., Thumfart L., Kakodkar R. R., Trevisi G., Frigeri P., Seravalli L., Feser J. P., Rastelli A., Mingo N. Predictive design and experimental realization of InAs/GaAs superlattices with tailored thermal conductivity // *J. Phys. Chem. C*, 2018. Vol. 122. Iss. 7. P. 4054–4062. doi: 10.1021/acs.jpcc.7b11133.
18. Muraki K., Fukatsu S., Shiraki Y., Ito R. Surface segregation of In atoms during molecular beam epitaxy and its influence on the energy levels in InGaAs/GaAs quantum wells // *Appl. Phys. Lett.*, 1992. Vol. 61. Iss. 5. P. 557–559. doi: 10.1063/1.107835.

Поступила в редакцию 05.10.20

APPLICATION OF MULTISCALE APPROACH AND DATA SCIENCES FOR MODELING THERMAL CONDUCTIVITY IN LAYERED STRUCTURES

K. K. Abgaryan^{1,2} and I. S. Kolbin¹

¹A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

²Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125080, Russian Federation

Abstract: Modeling thermal properties of layered structures is currently a popular area of scientific research. This is due to the constantly growing speed of operation of microelectronic elements often based on layered structures that release more and more energy during operation in the form of heat which must be removed to avoid overheating and loss of functional properties of devices. The paper presents an integration approach that allows one to combine the methods of multiscale modeling and data analysis. It is shown that application of this approach makes it possible to obtain a new quality when solving the problem of constructing a model of heat transfer in a two-layer GaAs/AlAs structure. The effectiveness of use of machine learning methods for analyzing the dependence of the effective thermal conductivity coefficient of laminated materials on structural features and external factors is shown. The development of the proposed approach will be able to provide formation of information for reasonable selection of materials for layered structures for microelectronic devices.

Keywords: multiscale modeling; integration approach; layered structures; predictive modeling; kinetic Boltzmann equation, thermal conductivity coefficient; data analysis methods

DOI: 10.14357/19922264200413

Acknowledgments

The work was partially supported by the Russian Foundation for Basic Research, projects 19-29-03051 mk and 19-08-01191 A).

References

1. Mark, A., A. B. Tepole, W. R. Cannon, *et al.* 2019. Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ Digital Medicine* 2(1):115. 11 p. doi: 10.1038/s41746-019-0193-y.
2. Ladygin, V., P. Korotaev, A. Yanilkin, and A. Shapeev. 2020. Lattice dynamics simulation using machine learning interatomic potentials. *Comp. Mater. Sci.* 172:109333.
3. Abgaryan, K. K. 2017. *Mnogomasshtabnoe modelirovanie v zadachakh strukturnogo materialovedeniya* [Multiscale modeling in material science problems]. Moscow: MAKSS Press. 284 p.
4. Abgaryan, K. K. 2018. Informatsionnaya tekhnologiya postroeniya mnogomasshtabnykh modeley v zadachakh vychislitel'nogo materialovedeniya [Information technology in the construction of multiscale models in problems of computational materials science]. *Sistemy vysokoy dostupnosti* [High Availability Systems] 14(2):9–15.
5. Kohn, W., and L. J. Sham. 1965. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A* 140(4A):A1133–A1138. doi: 10.1103/PhysRev.140.A1133.
6. Jia, Weile, Han Wang, Mohan Chen, *et al.* 2020. Pushing the limit of molecular dynamics with *ab initio* accuracy to 100 million atoms with machine learning. Cornell University. arXiv:2005.00223 [physics.comp-ph]. Available at: <https://arxiv.org/pdf/2005.00223.pdf> (accessed November 9, 2020).
7. Khvesyuk, V. I., and A. S. Skryabin. 2017. Heat conduction in nanostructures. *High Temp.* 55(3):434–456. doi: 10.1134/S0018151X17030129.
8. Vermeersch, B., J. Carrete, and N. Mingo. 2016. Cross-plane heat conduction in thin films with *ab-initio* phonon dispersions and scattering rates. *Appl. Phys. Lett.* 108(19):193104. doi: 10.1063/1.4948968.
9. Carrete, J., B. Vermeersch, A. Katre, A. Roekeghem, T. Wang, G. Madsen, and N. Mingo. 2017. AlmaBTE: A solver of the space-time dependent Boltzmann transport equation for phonons in structured materials. *Comput. Phys. Commun.* 220C:351–362. doi: 10.1016/j.cpc.2017.06.023.
10. Chung, J. D., A. J. H. McGaughey, and M. Kaviany. 2004. Role of phonon dispersion in lattice thermal conductivity modeling. *J. Heat Transf.* 126(3):376–380. doi: 10.1115/1.1723469.

11. Loy, J. M., J. Y. Murthy, and D. A. Singh. 2012. Fast hybrid Fourier–Boltzmann transport equation solver for nongray phonon transport. *J. Heat Transf.* 135(1):011008. doi: 10.1115/1.4007654.
12. Broido, D. A., M. Malorny, G. Birner, N. Mingo, and D. A. Stewart. 2007. Intrinsic lattice thermal conductivity of semiconductors from first principles. *Appl. Phys. Lett.* 91(23):231922. doi: 10.1063/1.2822891.
13. Powell, D. 2006. Elasticity, lattice dynamics and parameterization techniques for the Tersoff potential applied to elemental and type III–V semiconductors. Sheffield: University of Sheffield. PhD Thesis. 252 p.
14. Abgaryan, K. K., I. V. Mutigullin, S. I. Uvarov, and O. V. Uvarova. 2019. Multiscale modeling of clusters of point defects in semiconductor structures. *CEUR Workshop Proceedings*. Eds. S. I. Smagin and A. A. Zatsarinnyy. Khabarovsk. 2426:43–51.
15. Li, W., L. Lindsay, D. A. Broido, D. A. Stewart, and N. Mingo. 2012. Thermal conductivity of bulk and nanowire $\text{Mg}_2\text{SiSn}_{1-x}$ alloys from first principles. *Phys. Rev. B* 86(17):174307. doi: 10.1103/PhysRevB.86.174307.
16. Li, W., J. Carrete, N. A. Katcho, and N. Mingo. 2014. A solver of the Boltzmann transport equation for phonons. *Comput. Phys. Commun.* 185(6):1747–1758. doi: 10.1016/j.cpc.2014.02.015.
17. Carrete, J., B. Vermeersh, L. Thumfart, R. R. Kakodkar, G. Trevisi, P. Frigeri, L. Seravalli, J. P. Feser, A. Rastelli, and N. Mingo. 2018. Predictive design and experimental realization of InAs/GaAs superlattices with tailored thermal conductivity. *J. Phys. Chem. C* 122(7):4054–4062. doi: 10.1021/acs.jpcc.7b11133.
18. Muraki, K., S. Fukatsu, Y. Shiraki, and R. Ito. 1992. Surface segregation of In atoms during molecular beam epitaxy and its influence on the energy levels in InGaAs/GaAs quantum wells. *Appl. Phys. Lett.* 61(5):557–559. doi: 10.1063/1.107835.

Received October 15, 2020

Contributors

Abgaryan Karine K. (b. 1963) — Doctor of Science in physics and mathematics, principal scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; Head of Department, Moscow Aviation Institute (National Research University), 4 Volokolamskoe Shosse, Moscow 125080, Russian Federation; kristal83@mail.ru

Kolbin Ilya S. (b. 1986) — Candidate of Science (PhD) in physics and mathematics, scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; iskolbin@gmail.com

О ЦИФРОВОЙ ГРАМОТНОСТИ И СРЕДАХ ЕЕ ФОРМИРОВАНИЯ*

В. Б. Бетелин¹, А. Г. Кушниренко², А. Л. Семенов³, С. Ф. Сопрунов⁴

Аннотация: Цифровая грамотность становится ключевой характеристикой личности человека XXI в. Ее можно и нужно формировать с раннего возраста. Цифровые среды могут быть средами достижения предметных, метапредметных и личностных образовательных результатов. В работе анализируется отечественный и международный опыт в данном направлении начиная с 1960-х гг., приводится система принципов, обеспечивших эффективность для образовательных целей цифровых сред и систем, разработанных под руководством и при участии авторов. Эти среды и системы эффективно используются сегодня в РФ в формировании у детей раннего возраста цифровой грамотности и основы для *computational thinking* — системного мышления XXI в. В статье приводится обзор результатов, полученных в данном направлении. Специальное внимание уделяется проблеме визуализации и представления в реальном мире алгоритмических процессов и задающих их программ.

Ключевые слова: информатика; цифровая грамотность; *computational thinking*; робот; язык программирования; Лого; ПервоЛого; ПиктоМир; визуализация

DOI: 10.14357/19922264200414

1 Формирование идеи цифровой грамотности и конструирование сред ее достижения

Успешные попытки обучения школьников программированию начались в СССР в начале 1960-х гг. Это обучение велось в специально набранных старших классах, которые позже назвали «математическими», декларировалось как профессиональная подготовка и сочеталось с углубленной подготовкой по математике. Вели занятия не школьные учителя, а внешние специалисты [1]. О переносе подобной методики в начальную школу не могло быть и речи.

Другие подходы к освоению детьми цифрового мира, о развитии которых пойдет речь в настоящей статье, возникли в конце 1960-х гг. Тогда Симор Паперт, математик, программист, психолог (ученик Ж. Пиаже) и педагог, работал в Массачусетском технологическом институте над проблемами искусственного интеллекта. Вместе с Марвином Минским он написал книгу «Перцептроны» о машинном обучении, когда для этой технологии еще не было материальной базы. В 1967 г. коллеги Паперта (при его участии) из большой корпорации BBN

(живущей, в основном, оборонными заказами) создали для развития детей язык программирования Лого (на базе языка Лисп). Паперт включил в язык команды управления рисующим на полу роботом, «экранный аватар», который потом прославился как Черепашка Лого. Это задавало подход к преподаванию программирования детям как элемента их общего развития. В мемориальной статье [2] говорится: «Добавленная (Папертом) к Лого черепашка, управляемая компьютером на экране и на полу, обеспечила визуализацию и овеществление процессов программирования и осмысленность их результатов. Тем самым Лого превратился в уникальную среду для освоения алгоритмического мышления, которой пользуются миллионы детей в десятках стран мира».

С середины 1970-х гг. А. П. Ершов с коллегами в ВЦ СО АН начал обучение школьников, в том числе — младших классов, программированию в микромирах «Дежурик» и «Маляр» [3]. К 1981 г. его представления о роли программирования (сегодня бы сказали — цифровой грамотности) оформились в виде лозунга «Программирование — вторая грамотность», ставшего названием его исторического доклада в Лозанне [4, 5].

*Работа выполнена при поддержке РФФИ (проект № 19-29-14199 — С. Ф. Сопрунов), Российского научного фонда (проект № 17-11-01377 — А. Л. Семенов, разделы 3 и 5), госзадания 2020 года в ФГУ ФНИЦ НИИСИ РАН по теме 0065-2019-0010 (В. Б. Бетелин и А. Г. Кушниренко).

¹Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук, betelin@niisi.msk.ru

²Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук, agk@mail.ru

³Московский государственный университет имени М. В. Ломоносова; Институт кибернетики и образовательной информатики им. А. И. Берга Федерального исследовательского центра «Информатика и управление» Российской академии наук; НИУ Московский физико-технический институт, alsemno@yandex.ru

⁴Центр педагогического мастерства, soprunov@mail.ru

С. Паперт был убежден, что использование компьютеров совершит переворот в образовании. Его книга *Mindstorms* (1980 г.) оказала огромное влияние на образовательное сообщество всего мира, не исключая и СССР (приблизительный русский перевод издан в 1988 г.) [6]. В нашу страну по приглашению А. Л. Семенова Паперт приехал в конце 1980-х гг.; с ним авторов настоящей публикации связала многолетняя дружба.

За истекшие годы в мире были разработаны сотни реализаций Черепашки. В России распространение Лого началось с работы Ю. А. Первина в Новосибирске. Большую популярность получили системы программирования LogoWriter и ЛогоМиры — российские адаптации англоязычных продуктов компании Logo Computer Systems Inc., разработанные под руководством А. Семенова и С. Сопрунова в Институте новых технологий (ИНТ). ЛогоМиры продолжают активно использоваться и сегодня [7, 8]. Одной из самых распространенных сред для раннего изучения программирования стал Scratch, разработанный учениками и коллегами Паперта [9].

Черепашка Паперта «живет» в своем «микромире», где ученик управляет ей и ее учит. В 1970–1980-х гг. в нескольких образовательных центрах мира началось использование разнообразных «микромиров» учения. В частности, в Стенфорде Ричард Паттис изобрел исполнителя *Karel The Robot* и назвал его в честь Карела Чапека [10]. Роботы, «живущие» в клетчатой среде, с 1980 г. использовались в курсе программирования на мехмате МГУ [11] (первая версия — «Путник»).

Новый этап в развитии цифровой грамотности в нашей стране начался историческим Постановлением ЦК КПСС и Совета Министров СССР от 28 марта 1985 г. № 271 «О мерах по обеспечению компьютерной грамотности учащихся средних учебных заведений и широко внедрения электронно-вычислительной техники в учебный процесс» [12], позволившим нашей стране опередить весь мир в тотальной цифровизации школы. В подготовке Постановления А. П. Ершов сыграл ключевую роль. Во всех старших классах СССР началось изучение обязательного предмета «Основы информатики и электронно-вычислительной техники». Система программирования «КуМир», основанная на микромире «Робот в лабиринте» и др., была реализована усилиями МГУ и Академии наук СССР (ВНТК «Школа-1», руководитель Е. П. Велихов, зам. руководителя А. Л. Семенов) и стала центром основополагающего учебника информатики для старших классов СССР [13], затем использовалась в учебниках для учащихся 5–11 классов России [14–16]. В начале 2010-х гг. по инициативе чле-

нов Научно-методического совета ФИПИ по информатике (А. Л. Семенов, А. Ю. Уваров, С. Г. Андреев, В. А. Галкин, Г. И. Савин, председатель НМС В. Б. Бетелин) в варианты государственной итоговой аттестации по информатике стали включаться задачи, использующие исполнителя «Робот», доступные школьникам с минимальной подготовкой по программированию. Реализация в НИИСИ РАН свободно распространяемой версии проекта «КуМир» [17] сделала возможным решение подобных задач на любых школьных и домашних компьютерах.

2 Естественный возраст формирования цифровой грамотности

Развитие информационных технологий, цифровизация быта, образования, науки и экономики привели к общемировой тенденции дальнейшего понижения возраста знакомства детей с информатикой и программированием вплоть до дошкольного возраста [18]. Именно об этом говорили С. Паперт и А. П. Ершов более 40 лет назад.

В России эта тенденция проявляется как раннее знакомство детей с обрамляющими программирование информатикой и математикой. Центральное место при этом занимают не числовые, а символические и комбинаторные аспекты объектов и процессов окружающего мира. В частности, в действующем федеральном государственном образовательном стандарте начального общего образования [19] описываются «предметные результаты освоения основной образовательной программы начального общего образования по математике и информатике: овладение основами логического и алгоритмического мышления... уметь действовать в соответствии с алгоритмом и строить простейшие алгоритмы, исследовать, распознавать и изображать геометрические фигуры, работать с таблицами, схемами, графиками и диаграммами, цепочками, совокупностями».

Общий взгляд на проблемы реализации в школьной программе *computational thinking* — основ информатики, алгоритмики и программирования — изложен в статье А. Л. Семенова [20]. Под его руководством за последние 30 лет для начальной школы были подготовлены программы и учебно-методические комплексы для курсов «Информатика» и «Информатика и математика». Очередная версия учебника «Информатика. 1–4 классы» издана в 2019 г. [21]. В начальной школе России, как правило, один учитель ведет большую часть пред-

метов, он может для изучения математики, информатики и технологии использовать общий ресурс учебного времени, для освоения компьютера использовать уроки русского языка и искусства и т. п. Возможны различные варианты работы, в том числе «бескомпьютерный». При интеграции программирования в визуальных средах типа обсуждаемых ниже ПервоЛого и ПиктоМира, модуль информатики занимает, в среднем, около 2 ч в неделю, что дает вариант целостного освоения языковой, числовой и цифровой грамотности в начальной школе.

Осознание программирования уже не как «второй» грамотности, а как важнейшего элемента комплексной грамотности XXI в. приводит к идее микромира, работа с которым не предполагает «текстовой» грамотности. А. Л. Семенов в 1990-е гг. предложил создать версию Лого, не использующую словесной и числовой грамотности, как они обычно понимаются, а помогающую ребенку осваивать различные грамотности одновременно и параллельно. Эта идея была реализована коллективом С. Ф. Сопрунова в проекте ПервоЛого (LogoFirst, IconLogo), высоко оцененном С. Папертом [22]. В ПервоЛого все команды задаются пиктограммами (icons). Например, поворот Черепашки вправо на 50° , который в Лого представляется как: **направо 50**, в ПервоЛого задается поворотом штурвала, одна из ручек которого помечена красным. Число в центре штурвала возникает автоматически при повороте или может задаваться учеником, и тогда штурвал повернется в соответствии с числом (см. рисунок).

Когда ребенок использует эту команду для управления Черепашкой, он одновременно осваивает плоскую геометрию и начинает сопоставлять числа с геометрическими величинами. Использование ПервоЛого в школах России показало правильность исходной идеи параллельного освоения

различных видов грамотности: словесной, числовой, алгоритмической, пространственно-геометрической.

В последнее десятилетие идея визуальной, бестекстовой среды освоения алгоритмического мышления была реализована группой учебной информатики А. Г. Кушниренко – А. Г. Леонова под научным руководством академика В. Б. Бетелина в системе ПиктоМир как развитие в направлении младших возрастов идеи КуМира [23].

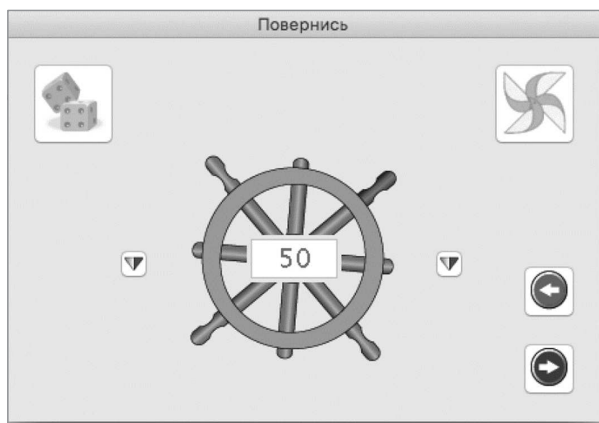
Еще одним шагом стало создание «овеществленных» сред программирования, где и Черепашка может перемещаться по полу, и программа может составляться из «кубиков» тоже на полу (или на столе), рисоваться на бумаге и т. д. Той же группой А. Г. Кушниренко – А. Г. Леонова был разработан и прошел апробацию в детских садах учебно-методический комплекс «Алгоритмика для дошколят», в основе которого лежит управление реальным роботом-машинкой Ползуном. Программу ребенок составляет на столе из кубиков с нарисованными на гранях пиктограммами команд и конструкций языка. Собранная из таких кубиков программа фотографируется, фотография распознается компьютером и переводится в звуковые команды, которые слышит и выполняет Ползун [24].

О разнообразии подобных исполнителей можно судить, например, по примерам реального использования программируемых роботов в российских детских садах на портале maam.ru [25]. Один из самых популярных детских роботов Bee-Bot [26] программируется на корпусе робота-«пчелки». Роботы Ozobot [27] следуют нарисованной линии, при этом на самой линии могут цветом кодироваться действия.

В работе [24], подготовленной с участием ряда авторов настоящей статьи, подчеркивается еще одно важное свойство роботоподобных овеществленных исполнителей — их использование позволяет построить четкую, наглядную, доступную шестилеткам-семилеткам систему научных понятий (больших идей, см. далее) программирования.

3 Задачи цифровой грамотности

Основная цель общего образования сегодня — формирование *навыков XXI в.* и общая ориентация в мире, позволяющая осознавать потребность в конкретных знаниях (включая умения, навыки) и получать эти знания (при необходимости формируя у себя умения и навыки). Большая часть «*навыков XXI в.*» оказывается более древней и присущей образованному человеку системой, чем остальная часть результатов образования века XX. К таким



Команда «направо 50» в ПервоЛого

навыкам относится само умение учиться, понимать другого человека, ставить цели и анализировать неудачи и т. п. Ориентация в мире меняется быстрее, и особенно быстро — сейчас. К ней относятся так называемые *большие идеи* разного высокого уровня общности, например общее понятие энергии, наследственности, возможности математического описания физических процессов и т. п. Концепция *большой идеи* (Big Idea) возникла в естественно-научном образовании как оппозиция к «горé фактов». *Большая идея* — это часть представления человека о мире, без которой представление в целом, поведение в мире становятся другими. Все более необходимыми для ориентации становятся большие идеи *цифровой грамотности*. Способность использовать ориентацию, вместе с формирующейся в информатике способностью решать совершенно новые задачи, составляют основу пре-адаптивности (А. Г. Асмолов).

И *навыки XXI в.*, и ориентация в мире должны формироваться в конкретной интересной и важной для ученика деятельности — созидании, коммуникации и т. п. Конкретные среды, освоение конкретных умений и навыков, приобретение необходимых знаний важны в первую очередь не сами по себе, а как элементы пути формирования более общих результатов. «Образование — это то, что остается, когда мы уже забыли все, чему нас учили», как, возможно, сказал Джордж Галифакс (XVII в.) и, судя по цитатам, многократно повторяли крупнейшие физики XX в. (например, Макс фон Лауэ).

Особенно этот подход актуален сегодня, когда учится, общается, действует и живет не сам по себе человек (взрослый или ребенок), а *человек расширенный*, способный в мире что-то делать, обращаясь, кроме ресурсов собственного организма, к цифровым ресурсам (источникам, инструментам, средам и сервисам) и через них присваивающий культуру человечества [28].

На взгляд авторов, основой для *computational thinking*, как и для цифровой компетенции (даже если различать эти понятия), должны быть осознание себя (как и других) как человека расширенного и эффективная деятельность в качестве такового. И в процессе образования следует адресоваться к такому *расширенному человеку*, к его *расширенному сознанию*.

Современный контекст информатики — программирования в широком, «ершовском» смысле — позволяет:

- (1) формировать многие из ключевых навыков XXI века, в том числе относящихся к коллективной работе;

- (2) интегрировать для *расширенной личности* цифровую грамотность, прежде всего — ее *большие идеи*, в общую грамотность и систему ориентации в мире;
- (3) делать это на конкретном материале разнообразных, интересных, посильно трудных задач с высоким уровнем индивидуальной новизны для ученика.

При этом конкретные среды, в которых идет достижение указанных целей (1) и (2), могут быть различными. Условие (3) может достигаться по-разному, в комбинации различных факторов:

- интеллектуальный вызов — понятная новая задача с неочевидным, неожиданным решением;
- эмоциональная сторона — эстетическая привлекательность картин, разворачивающихся на экране, вовлеченность в игровые ситуации, отождествление с персонажами;
- связь со взрослой жизнью — сходство твоего программирования с профессиональным, профессиональная ориентация, польза в будущем, а иногда и оплачиваемая работа сейчас.

4 Роль среды. Язык программирования

Представленная перспектива стояла перед авторами и тогда, когда под руководством Андрея Петровича Ершова проектировались уже упомянутые в связи с Постановлением ЦК и СМ процессы. Одним из принципиальных решений было максимальное упрощение синтаксиса языка программирования, приближение его по экономности средств к обычному математическому языку, бóльшая, чем в математическом языке, логичность. Реализацией языка занималась группа В. Б. Бетелина и А. Г. Кушнirenко. Рабочая гипотеза тогда состояла в том, что именно такой минимальный, базовый, логично построенный Школьный алгоритмический язык и должен стать важнейшим компонентом цифровой грамотности.

В последующие десятилетия ядро языка оставалось стабильным, развивалась система программирования на его основе, включающая структурный редактор, средства пошагового исполнения и отладки. Сегодня эта система обеспечивает учащегося максимальным комфортом в разработке и отладке программ. Жизнь показала плодотворность такого подхода. Обладая сформированной алгоритмической грамотностью, можно осваивать новые конструкции и языки других систем программирования. Но при этом, как и для других видов грамотности, **решается задача охвата грамотностью**

всех детей, что и подразумевается в лозунге Ершова. На завершающем этапе общего образования — в старшей школе, которая строится как профильная, — дальнейшее формирование цифровой компетенции идет различными путями. Для учащихся с уже сформированной цифровой грамотностью, освоивших большие идеи информатики, этого достаточно для дальнейшего образования и ориентации в мире. При подготовке будущих математиков основой служит Школьный алгоритмический язык и его реализация в КуМире, где могут строиться алгоритмы работы с математическими объектами, при необходимости с добавлением новых структур данных и исполнителей. Одновременно формируется культура работы с цифровыми инструментами в математике, в том числе с системами компьютерной алгебры (пример — Wolfram Mathematica [29]). В ИТ-профилях идет освоение какого-нибудь «производственного» языка, например одного из языков C, C Sharp, C++, Python и т. п.

Остается вопрос: можно ли при формировании начальной цифровой грамотности **после вводного дошкольного нетекстового этапа освоения программирования** использовать не минимальный алгоритмический язык, а какой-то иной конкретный язык программирования? Это возможно, но при этом следует учитывать следующие факторы, в порядке важности:

- (1) простота освоения — грамотность для всех;
- (2) наличие исполнителей, сразу используемых для мотивирующего решения задач, выполнения проектов;
- (3) освоение конструкций популярных сегодня языков профессионального программирования.

Первые два фактора реализованы в проекте Лого — «языка без порога и потолка». Возможны и другие среды, специально созданные для школы, в которых можно делать что-то интересное для школьника, ценное вне программирования. И осваивая какой-то язык, используемый вне школы, можно сформировать базовые представления цифровой грамотности, получить опыт работы с ними, возможно, жертвуя факторами (1) и (2).

Ситуация с исполнителями заслуживает отдельного рассмотрения. Безусловно, Робот, как и Черепашка, не относятся к *большим идеям*, ценным вне контекста первоначального обучения программированию. Однако внутри этого контекста они занимают особое положение. То, что Робота изобрели, видимо, независимо в противоположных точках глобуса, служит этому подтверждением.

5 Большие идеи цифровой грамотности в ее системе целей

Анализ и попытки применения самого понятия *большой идеи* показывают возможность и необходимость классификации больших идей по широте их применимости. Некоторые из больших идей информатики значимы и вне контекста цифровой грамотности, другие универсальны в цифровой сфере, третьи, все еще оставаясь большими идеями, приложимы в более специальных контекстах.

Видимо, наиболее актуальной проблемой в парадигме больших идей образования становится **выделение больших идей для сферы искусственного интеллекта и подбор учебных ситуаций, задач, проектов, где эти идеи будут осваиваться.**

6 Выводы

Опыт раннего использования программирования для формирования цифровой грамотности в нашей стране и за рубежом показывает критическую важность применения визуальных и материальных (в том числе бестекстовых) сред для развития такой грамотности как важного элемента общего развития современного ребенка.

Литература

1. *Шварцбург С. И.* Из опыта работы с учащимися 9 класса, овладевающими специальностью лаборантов-программистов // Математика в школе, 1960. № 5. С. 9–19.
2. *Семенов А. Л.* Симор Паперт и мы. Конструкционизм — образовательная философия XXI века // Вопросы образования, 2017. № 1. С. 269–294.
3. *Звенигородский Г. А.* Первые уроки программирования / Под ред. А. П. Ершова. — М.: Наука, 1985. 208 с.
4. *Ershov A. P.* Programming, the second literacy // Computer and Education: IFIP TC-3 3rd World Conference on Computer in Education Proceedings. — Lausanne, Switzerland: North-Holland Pub. Co., 1981. P. 1–17.
5. *Ершов А. П.* Программирование — вторая грамотность: Русская версия доклада // 3-й Всемирный конгресс по обучению математике. — Лозанна, Швейцария, 1981. http://ershov.iis.nsk.su/ru/second_literacy/article.
6. *Пейперт С.* Переворот в сознании: дети, компьютеры и плодотворные идеи / Пер. с англ. — М.: Педагогика, 1989. 224 с. (*Papert S.* Mindstorms: Children, computers and powerful ideas. — New York, NY, USA: Basic Books Inc. Publ., 1980. 252 p.)
7. *Сопрунов С. Ф., Ушакова А. С., Яковлева Е. И.* Перво-Лого 4.0: Справочное пособие. — М.: Институт новых технологий, 2012. 144 с.

8. ЛогоМиры 3.0. Интегрированная творческая среда. <http://www.int-edu.ru/content/logomiry-30-integrirovannaya-tvorcheskaya-sreda>.
9. Scratch. Создавай истории, игры и мультфильмы. Делись с другими по всему миру. <https://scratch.mit.edu>.
10. *Pattis R. E. Karel The Robot: A gentle introduction to the art of programming.* — New York, NY, USA: John Wiley & Sons, 1981. 176 p.
11. *Кушниренко А. Г., Лебедев Г. В.* Программирование для математиков. — М.: Наука, 1988. 384 с.
12. О мерах по обеспечению компьютерной грамотности учащихся средних учебных заведений и широкого внедрения электронно-вычислительной техники в учебный процесс: Постановление ЦК КПСС и Совета Министров СССР от 28 марта 1985 г. № 271 // Вопросы образования, 2005. № 3. С. 341–346. <https://vo.hse.ru/data/2015/04/20/1095612939/22post0.pdf>.
13. *Ершов А. П., Кушниренко А. Г., Лебедев Г. В., Семенов А. Л., Шень А. Х.* Основы информатики и вычислительной техники: Пробный учебник для средних учебных заведений / Под ред. А. П. Ершова. — М.: Просвещение, 1988. 207 с.
14. *Кушниренко А. Г., Лебедев Г. В., Сворень Р. А.* Основы информатики и вычислительной техники. — М.: Просвещение, 1990. 224 с.
15. *Звонкин А. К., Ландо С. К., Семенов А. Л., Вялый Н. М.* Информатика. Алгоритмика. 6–7 классы. — М.: Просвещение, 2006–2008.
16. *Кушниренко А. Г., Семенов А. Л., Зайдельман Я. Н., Тарасова В. В.* Информатика. 7–9 классы. — М.: Дрофа, 2017.
17. Проект «КуМир». <https://www.niisi.ru/kumir>.
18. *Richtel M.* Reading, writing, arithmetic, and lately coding // New York Times, May 10, 2014. <https://www.nytimes.com/2014/05/11/us/reading-writing-arithmetic-and-lately-coding.html>.
19. Об утверждении и введении в действие федерального государственного образовательного стандарта начального общего образования: Приказ министерства образования и науки РФ от 6 октября 2009 г. № 373. 20 с. <http://base.garant.ru/197127>.
20. *Семенов А. Л.* Концептуальные проблемы информатики, алгоритмики и программирования в школе // Вестник кибернетики, 2016. № 2(22). С. 12–16.
21. *Рудченко Т. А., Семенов А. Л.* Информатика. 1–4 классы. — М.: Просвещение, Институт новых технологий, 2011–2019.
22. *Papert S.* The connected family: Bringing the digital generation gap. — Taylor Trade Pubs., 1996. 211 p.
23. Стартовая страница проекта «ПиктоМир». <https://www.niisi.ru/piktomir>.
24. *Бетелин В. Б., Кушниренко А. Г., Семенов А. Л.* Основные понятия программирования в изложении для дошкольников // Информатика и её применения, 2020. Т. 14. Вып. 3. С. 56–62.
25. Портал maam.ru. <https://www.maam.ru/obrazovanie/robototehnika>.
26. Возможности мини-робота Bee-Bot: Мастер-класс. <https://www.maam.ru/detskijasad/master-klas-vozmozhnosti-mini-robot-bee-bot.html>.
27. Robots to teach coding and creativity. <https://ozobot.com/stem-education>.
28. *Семенов А. Л.* Цели общего образования в цифровом мире // Информатизация образования и методика электронного обучения: Мат-лы III Междунар. конф.: в 2 ч. — Красноярск: СФУ, 2019. Ч. 2. С. 383–388.
29. Wolfram Mathematica. <https://www.wolfram.com/mathematica>.

Поступила в редакцию 21.11.2019 (последняя правка 11.09.2020)

ABOUT DIGITAL LITERACY AND ENVIRONMENTS FOR ITS DEVELOPMENT

V. B. Betelin¹, A. G. Kushnirenko¹, A. L. Semenov^{2,3,4}, and S. F. Soprunov⁵

¹Federal State Research Center Scientific Research Institute of System Development, Russian Academy of Sciences, 36-1 Nakhimovsky Prosp., Moscow 117218, Russian Federation

²M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation

³Alex Berg Institute of Cybernetics and Educational Computing of Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

⁴Moscow Institute of Physics and Technology (National Research University), 9 Institutskiy Per., Dolgoprudny, Moscow 141701, Russian Federation

⁵Center for Pedagogical Excellence, 15-1 Bol’shaya Spasskaya Str., Moscow 129090, Russian Federation

Abstract: Digital literacy is becoming a key characteristic of a XXI century person. It can and should be formed from an early age. Digital environments can be environments for achieving subject, metasubject, and personal

educational outcomes. The work analyzes domestic and international experience in this direction, beginning in 1960s, offers a general view of the “big ideas” mastered in digital environments and provides a system of principles that ensure the effectiveness for educational purposes of digital environments and systems developed under the guidance with the participation of the authors, they are effectively used in the Russian Federation in the formation of digital literacy in young children, the basis for computational thinking — systemic thinking of the XXI century. The article gives an overview of the results obtained in this direction over the last years. Special attention is paid to the problem of visualization and representation in the real world of algorithmic processes and the programs that set them.

Keywords: informatics; computer science; digital literacy; computational thinking; robot; programming language; Logo; PervoLogo; LogoFirst; PictoMir; learning environment; visualization

DOI: 10.14357/19922264200414

Acknowledgments

The work was supported by the State Assignment at the FSRC Scientific Research Institute of System Development, RAS of the 2020, topic 0065-2019-0010 (V. B. Betelin and A. G. Kushnirenko), Russian Science Foundation (grant No. 17-11-01377 — A. L. Semenov, parts 3 and 5), and Russian Foundation for Basic Research (grant No. 19-29-14199 — S. F. Soprunov).

References

1. Shvartsburd, S. I. 1960. Iz opyta raboty s uchashchimisya 9 klassa, ovladevayushchimi spetsial'nost'yu laborantov-programmistov [From the experience of working with students in grade 9, mastering the specialty of laboratory assistants-programmers]. *Matematika v shkole* [Mathematics in School] 5:9–19.
2. Semenov, A. L. 2017. Simor Papert i my. Konstruktsionizm — obrazovatel'naya filosofiya XXI veka [Seymour Papert and us. Constructionism as the educational philosophy of the 21st century]. *Voprosy obrazovaniya* [Education Studies] 1:269–294.
3. Zvenigorodskiy, G. A. 1985. *Pervye uroki programirovaniya* [First programming lessons]. Moscow: Nauka. 208 p.
4. Ershov, A. P. 1981. Programming, the second literacy. *Computer and Education: IFIP TC-3 3rd World Conference on Computer in Education Proceedings*. — Lausanne, Switzerland: North-Holland Pub. Co. 1–17.
5. Ershov, A. P. 1981. Programirovanie — vtoraya gramotnost'. Russkaya versiya doklada [Programming, the second literacy. Russian version of lecture]. *3-y Vsemirnyy kongress po obucheniyu matematike* [3rd World Conference on Computer Education]. Lausanne, Switzerland. Available at: http://ershov.iis.nsk.su/ru/second_literacy/article (accessed October 26, 2020).
6. Papert, S. 1980. *Mindstorms: Children, computers and powerful ideas*. New York, NY: Basic Books Inc. Publs. 252 p.
7. Soprunov, S. F., A. S. Ushakova, and E. I. Yakovleva. 2012. *PervoLogo 4.0. Spravochnoe posobie* [PervoLogo 4.0. Reference manual]. Moscow: Institut novykh tekhnologiy. 144 p.
8. LogoMiry 3.0. Integrirovannaya tvorcheskaya sreda [LogoMiry 3.0. Integrated Creative Environment]. Available at: <http://www.int-edu.ru/content/logomiry-30-integrirovannaya-tvorcheskaya-sreda> (accessed October 26, 2020).
9. Scratch: Sozdavay istorii, igry i mul'tfil'my. Delis' s drugimi po vsemu miru [Scratch: Create stories, games, and animations. Share with other around the world]. Available at: <https://scratch.mit.edu/> (accessed October 26, 2020).
10. Pattis, R. E. 1981. *Karel the robot: A gentle introduction to the art of programming*. New York, NY: John Wiley & Sons. 176 p.
11. Kushnirenko, A. G., and G. V. Lebedev. 1988. *Programirovanie dlya matematikov* [Programming for mathematicians]. Moscow: Nauka. 384 p.
12. TsK KPSS i Sovet Ministrov SSSR. March 28, 1985. O merakh po obespecheniyu komp'yuternoy gramotnosti uchashchikhsya srednikh uchebnykh zavedeniy i shirokogo vnedreniya elektronno-vychislitel'noy tekhniki v uchebnyy protsess: Postanovlenie No.271 [On measures to ensure computer literacy of secondary school students and the widespread introduction of electronic computing technology in the educational process: Decree No.271]. Available at: <https://vo.hse.ru/data/2015/04/20/1095612939/22post0.pdf> (accessed October 26, 2020).
13. Ershov, A. P., A. G. Kushnirenko, G. V. Lebedev, A. L. Semenov, and A. Kh. Shen'. 1988. *Osnovy informatiki i vychislitel'noy tekhniki: Probnyy uchebnik dlya srednikh uchebnykh zavedeniy* [Fundamentals of computer science and computer technology: A trial textbook for secondary schools]. Moscow: Prosveshchenie. 207 p.
14. Kushnirenko, A. G., G. V. Lebedev, and R. A. Svoren'. 1990. *Osnovy informatiki i vychislitel'noy tekhniki* [Fundamentals of computer science and computer engineering]. Moscow: Prosveshchenie. 224 p.
15. Zvonkin, A. K., S. K. Lando, A. L. Semenov, and N. M. Vyalyy. 2006–2008. *Informatika. Algoritmika. 6–7 klassy* [Informatics. Algorithmics. Grades 6–7]. Moscow: Prosveshchenie.
16. Kushnirenko, A. G., A. G. Leonov, Ya. N. Zaydel'man, and V. V. Tarasova. 2017. *Informatika. 7–9 klassy* [Informatics. 7–9 grades]. Moscow: Drofa.

17. Proekt “KuMir” [Project “KuMir”]. Available at: <https://www.niisi.ru/kumir/> (accessed October 26, 2020).
18. Richtel, M. May 10, 2014. Reading, writing, arithmetic, and lately, coding. *New York Times*. Available at: <https://www.nytimes.com/2014/05/11/us/reading-writing-arithmetic-and-lately-coding.html> (accessed October 26, 2020).
19. Minobrnauki Rossii. October 6, 2009. Ob utverzhdenii i vvedenii v deystvie federal'nogo gosudarstvennogo obrazovatel'nogo standarta nachal'nogo obshchego obrazovaniya: Prikaz No.373 [On the approval and implementation of the federal state educational standard of primary general education: Order No. 373]. 20 p. Available at: <http://base.garant.ru/197127/> (accessed October 26, 2020).
20. Semyonov, A. L. 2016. Kontseptual'nye problemy informatiki, algoritmiki i programmirovaniya v shkole [Conceptual problems of teaching computer science, algorithm studies, and programming at school]. *Vestnik kibernetiki* [Proceedings in Cybernetics] 2(22):12–16.
21. Rudchenko, T. A., and A. L. Semenov. 2011–2019. *Informatika. 1–4 klassy* [Informatics. Grades 1–4]. Moscow: Prosveshchenie, Institut novykh tekhnologiy.
22. Papert, S. 1996. *The connected family: Bringing the digital generation gap*. Taylor Trade Publs. 211 p.
23. PictoMir: Startovaya stranitsa proekta [The start page of the PictoMir project]. Available at: <https://www.niisi.ru/piktomir/> (accessed October 26, 2020).
24. Betelin, V. B., A. G. Kushnirenko, and A. G. Leonov. 2020. Osnovnye ponyatiya programmirovaniya v izlozhenii dlya doshkol'nikov [Basic programming concepts for preschoolers]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):56–62.
25. Portal maam.ru. Available at: <https://www.maam.ru/obrazovanie/robototehnika> (accessed October 26, 2020).
26. Vozmozhnosti mini-robota Bee-bot. Master-klass [Features of the Bee-bot mini-robot. Master class]. Available at: <https://www.maam.ru/detskijasad/masterklas-vozmozhnosti-mini-robota-bee-bot.html> (accessed October 26, 2020).
27. Robots to teach coding and creativity. Available at: <https://ozobot.com/stem-education> (accessed October 26, 2020).
28. Semenov, A. L. 2019. Tseli obshchego obrazovaniya v tsifrovom mire [The goals of general education in the digital world]. *Mat-ly III Mezhdunar. konf. “Informatizatsiya obrazovaniya i metodika elektronnoy obucheniya”* [3rd Conference (International) “Informatization of Education and the Methods of Electronic Learning” Proceedings]. Krasnoyarsk: SFU. 2:383–388.
29. Wolfram Mathematica. Available at: <https://www.wolfram.com/mathematica/> (accessed October 26, 2020).

Received November 21, 2019 (last revision September 11, 2020)

Contributors

Betelin Vladimir B. (b. 1946) — professor, Academician of RAS, scientific director, Federal State Research Center Scientific Research Institute of System Development, Russian Academy of Sciences, 36-1 Nakhimovsky Prosp., Moscow 117218, Russian Federation; betelin@niisi.msk.ru

Kushnirenko Anatoliy G. (b. 1944) — Candidate of Science (PhD) in physics and mathematics, Head of Department of Educational Informatics, Federal State Research Center Scientific Research Institute of System Development, Russian Academy of Sciences, 36-1 Nakhimovsky Prosp., Moscow 117218, Russian Federation; agk_@mail.ru

Semenov Alexei L. (b. 1950) — professor, Academician of RAS, Academician of the Russian Academy of Education, Head of Department, M. V. Lomonosov Moscow State University, 1 Leninskie Gory, GSP-1, Moscow 119991, Russian Federation; director, Alex Berg Institute of Cybernetics and Educational Computing, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; principal scientist, Moscow Institute of Physics and Technology (National Research University), 9 Institutskiy Per., Dolgoprudny, Moscow 141701, Russian Federation; alsemno@ya.ru

Soprunov Sergey F. (b. 1949) — Candidate of Science (PhD) in physics and mathematics, methodologist, Center for Pedagogical Excellence, 15-1 Bol'shaya Spasskaya Str., Moscow 129090, Russian Federation; soprunov@mail.ru

ЭВОЛЮЦИЯ КЛАССИФИКАЦИЙ В НАДКОРПУСНЫХ БАЗАХ ДАННЫХ*

А. А. Гончаров¹, И. М. Зацман², М. Г. Кружков³

Аннотация: Рассматривается задача фиксации изменений, вносимых в описания значений немецких модальных глаголов в процессе аннотирования параллельных немецко-русских текстов с использованием надкорпусной базы данных (НБД). На примере этой задачи проанализирована специфика использования динамической классификационной системы (ДКС) в информационных системах. Специфика ДКС состоит в том, что смысловое содержание ее рубрик может меняться в процессе аннотирования, а это нередко влечет за собой потребность в реклассификации ранее сформированных аннотаций с измененными впоследствии рубриками. Основная цель статьи состоит в поиске ответов на два вопроса: (1) от каких факторов зависит необходимость редактирования и реклассификации аннотаций, сформированных до изменения рубрик; (2) с помощью каких операций можно вносить изменения в дефиниции рубрик ДКС. В статье определены семь типов возможных изменений дефиниций и описаны соответствующие им операции, применяемые к рубрикам ДКС в процессе аннотирования. Операции распределены на три группы в зависимости от того, требуют ли они программной или экспертной реклассификации ранее сформированных аннотаций.

Ключевые слова: динамическая классификация; фасетная классификация; реклассификация; надкорпусная база данных

DOI: 10.14357/19922264200415

1 Введение

Современные компьютерные технологии и информационные ресурсы служат основой развития многих направлений лингвистической науки, в частности корпусной лингвистики [1–4]. Одним из таких ресурсов стали НБД (см. [5], где впервые было использовано это понятие), создаваемые в ИПИ ФИЦ ИУ РАН.

Надкорпусные базы данных можно определить как «надстроенный» над корпусом [текстов] информационный ресурс, позволяющий последовательно фиксировать... употребления исследуемых языковых единиц (ЯЕ), снабжая их описаниями, структурированными в соответствии с задачами исследования» [6]. Процесс описания употреблений ЯЕ называется лингвистическим аннотированием [7], а результатом аннотирования одного употребления ЯЕ становится аннотация, спектр используемых видов которых описан в [8]. Аннотация содержит информацию о том, к каким рубрикам фасетной классификации [9, 10] относится некоторое употребление ЯЕ. Структурированность

аннотаций позволяет использовать их для получения разнообразных статистических данных [11, 12].

Одна из НБД была разработана для исследования немецких модальных глаголов (НМГ), которые характеризуются развитой полисемией — имеют три и более значений, согласно [13]. Соответствующая база данных далее обозначается как НБД НМГ. Она обеспечивает аннотирование употреблений НМГ, их переводов на русский язык и переводных соответствий в целом. Материалом исследования стали тексты параллельного немецкого подкорпуса Национального корпуса русского языка (<https://ruscorpora.ru/new/index.html>) для направления перевода «немецкий–русский» общим объемом более 2,6 млн словоупотреблений.

В процессе аннотирования (1) отбираются пары фрагментов параллельных текстов, где в оригинале употреблен модальный глагол; (2) отобранные фрагменты оригинала и его перевода копируются в аннотацию (табл. 1, первый и третий столбцы) и снабжаются рубриками фасетной классификации для оригинала, перевода и переводного соответствия (табл. 1, второй, четвертый и пятый столбцы);

* Работа выполнена в Институте проблем информатики ФИЦ ИУ РАН при поддержке РФФИ (проекты 18-07-00192 и 20-012-00166).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, a.gonch48@gmail.com

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

³ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, magnit75@yandex.ru

Таблица 1 Пример аннотации, сформированной в НБД НМГ

Контекст НМГ в оригинале	Коды рубрик оригинала	Перевод на русский язык	Коды рубрик перевода	Коды рубрик переводного соответствия
Soll ich Gefühle, einen Glauben <i>heucheln</i> , die ich nicht habe? [Н. Böll. Ansichten eines Clowns (1963)]	sollen <sollen-10> <Praes> <1sg> <+Inf I> <Inversion> <Interrog>	— Хотите , чтобы я притворялся верующим, изображая чувства, которых у меня нет? [Г. Белль. Глазами клоуна (пер. Л. Б. Черная, 1964)]	<input checked="" type="checkbox"/> <хотеть> <input type="checkbox"/> <Praes> <input type="checkbox"/> <Вы> <input type="checkbox"/> <+Subord + «чтобы»> <input type="checkbox"/> <Interrog>	SubjCh

Примечания.

Расшифровка кодов рубрик оригинала:

- **sollen** — в данном контексте исследуемым модальным глаголом немецкого языка является «sollen»;
- <sollen-10> — глагол «sollen» употреблен в десятом значении согласно порядку описания значений в словарной статье в [13];
- <Praes> — глагол употреблен в форме настоящего времени изъявительного наклонения;
- <1sg> — глагол употреблен в форме первого лица единственного числа;
- <+Inf I> — глаголу подчинен инфинитив настоящего времени (см. слово «heucheln» в оригинале, выделенное курсивом);
- <Inversion> — инверсия подлежащего и сказуемого, т.е. сказуемое — в данном случае модальный глагол — занимает начальную позицию в предложении;
- <Interrog> — глагол употреблен в вопросительном предложении.

Расшифровка кодов рубрик перевода:

- <хотеть> — носителем модального значения, передающего значение немецкого модального глагола в переводе на русский язык, служит глагол «хотеть» (в аннотации символом «» отмечаются носители модального значения в русском языке);
- <Praes> — глагол употреблен в форме настоящего времени изъявительного наклонения;
- <Вы> — глагол употреблен в форме второго лица множественного числа для вежливого обращения к одному лицу;
- <+Subord + «чтобы»> — глаголу подчинено придаточное предложение, присоединенное с помощью союза «чтобы» («чтобы я притворялся верующим...»);
- <Interrog> — глагол употреблен в вопросительном предложении.

Расшифровка кодов рубрик переводного соответствия:

- SubjCh — при переводе подлежащее при модальном глаголе было изменено (в оригинале — «должен ли я», в переводе — «хотите [ли Вы]»).

(3) затем сформированная аннотация записывается в НБД НМГ (подробнее о методологии аннотирования в НБД см. [6]).

Примерами фасетов, которые используются для аннотирования немецкоязычного оригинала, служат:

- (1) грамматическое время, например: <Praes> — настоящее время изъявительного наклонения; <Praet> — простое прошедшее время изъявительного наклонения и т. д.;
- (2) форма лица и числа, например: <1sg> — 1-е лицо ед. числа; <1pl> — 1-е лицо множ. числа; <2sg> — 2-е лицо ед. числа и т. д.;
- (3) значение НМГ: в качестве отправной точки в этот фасет были внесены все описания значений рассматриваемых модальных глаголов в соответствии с немецко-русским словарем [13], например: <sollen-01> — первое

значение глагола «sollen»; <müssen-01> — первое значение глагола «müssen»; <müssen-02> — второе значение глагола «müssen» и т. д.

Классификационная система, используемая в процессе аннотирования, отражает состояние знания лингвистов на определенные моменты времени. В процессе работы их знания могут меняться, что иногда влечет и перераспределение классифицируемых объектов по классам. Все это необходимо отражать в самой классификационной системе.

Цель настоящей статьи заключается в описании (1) специфики динамических классификаций; (2) типов возможных изменений рубрик этих классификаций на примере фасета «Значения немецких модальных глаголов» («Значения НМГ»); (3) связанных с такими изменениями потребностей в редактировании ранее сформированных аннотаций и их реклассификации¹. Средства отражения динамики этих изменений в НБД обсуждаются в мини-

¹О классификации и, соответственно, реклассификации аннотаций представляется возможным говорить потому, что совокупности рубрик, присваиваемых аннотациям, представляют собой классификационные коды (подобные кодам Универсальной десятичной классификации), отражающие содержание оригинала, перевода и переводного соответствия этих аннотаций. Так, классификационный код аннотации из табл. 1 будет включать коды рубрик из столбцов 2, 4 и 5 этой таблицы. Следовательно, при изменении дефиниции одной из рубрик соответствующим образом изменятся содержание аннотации и, возможно, ее код.

мальной степени, так как их планируется детально рассмотреть в отдельной статье.

2 Эволюция классификационных систем и реклассификация

Классификационные системы делятся на две категории — *статические* и *динамические* — в зависимости от потенциальной возможности изменения рубрик на рассматриваемом интервале времени. Рассмотрим различия между ними на примере лингвистического аннотирования ЯЕ, выполняемого в течение заданного интервала времени.

Статическая классификационная система (СКС) остается неизменной на протяжении всего заданного интервала времени и должна содержать полный перечень необходимых рубрик до начала аннотирования. Если же при аннотировании возникает потребность включить в СКС новые рубрики и/или изменить существующие, то это можно сделать только по окончании заданного интервала времени. Поэтому до его окончания неполнота СКС может привести к появлению незавершенных аннотаций или аннотаций с нерелевантными рубриками, которые не позволяют адекватно представить аннотируемый материал.

Чтобы избежать появления незавершенных и нерелевантных аннотаций, можно использовать *динамическую классификационную систему*, которая допускает внесение изменений в существующие рубрики и добавление новых рубрик в течение интервала времени аннотирования. Использование ДКС обусловлено динамикой знания лингвистов в процессе проведения ими исследований в области корпусной семантики. Поскольку НБД ориентированы именно на отражение динамики, то концепция, лежащая в основе проектирования НБД, в общем случае предполагает использование фасетных динамических классификаций, которые можно редактировать в процессе лингвистического аннотирования [8, 10].

Эксперимент по аннотированию значений НМГ показал, что знание лингвистов об этих значениях действительно достаточно часто меняется в процессе работы с НБД [14]. Хотя большинство используемых при аннотировании фасетов классификации стабильны и не изменяются в процессе работы (грамматическое время, лицо, число и т. д.), фасет, обеспечивающий аннотирование семантики употребления и значений НМГ, таковым не является. Рубрики именно этого фасета находятся в центре внимания проводимых с использованием НБД исследований, так как «уточнение номенклатуры значений немецких модальных глаголов и условий

их реализации» обозначено в качестве одной из целей создания НБД НМГ [15, с. 173] (подробнее об аннотировании НМГ см. [15, с. 175–181]).

Лингвисты в ходе аннотирования текстов параллельного немецкого подкорпуса нередко сталкиваются в этих текстах с примерами использования модальных глаголов, которые не соотносятся ни с одним из ранее описанных значений этих глаголов. Поэтому в интересах завершенности аннотирования они вырабатывают новые или уточняют существующие дефиниции значений прямо в процессе семантического анализа таких примеров, используя ДКС. Иными словами, знание лингвистов об исследуемых ЯЕ эволюционирует в процессе аннотирования, что влечет за собой изменение дефиниций рубрик ДКС. Результатом ее применения становятся не только более точные и завершенные аннотации, но и обновленная ДКС, отражающая обнаруженное в ходе аннотирования и эксплицированное новое лингвистическое знание.

Существенная особенность использования ДКС заключается в том, что изменения дефиниций рубрик могут потребовать редактирования соответствующих полей тех аннотаций, которые включали рубрики ДКС с предыдущими версиями их дефиниций, что неизбежно замедлит процесс аннотирования. Такое редактирование по сути представляет собой *реклассификацию* ранее сформированных аннотаций, так как меняются их классификационные коды.

Реклассификация может быть определена как *процесс перехода от одной классификационной системы к другой системе или к новой версии той же системы*. Проблема реклассификации была впервые осознана, видимо, в информационно-библиотечной науке. Она активно обсуждалась уже в первой половине XX в., тогда как «в том или ином виде реклассификация существовала с появлением каталогизации и самой классификации» [16, с. 249]. Причинами перехода на новую классификацию или новую версию старой классификации могут быть «недостаточная приемлемость используемой классификационной системы и появление более приемлемой классификационной системы» [17, с. 83].

Сегодня проблема реклассификации вышла далеко за рамки информационно-библиотечной науки. Наглядным примером ДКС может служить Международная патентная классификация (МПК). Чтобы избежать реклассификации изобретений из-за ускорения роста и динамики технического знания, начиная с 2006 г. при простановке рубрики МПК в описаниях изобретений одновременно указывается дата (год и месяц), когда она была согласована экспертами и утверждена Международным бюро Всемирной организации интеллектуальной

собственности (МБ ВОИС) [18]. Однако в лингвистическом аннотировании применение такого подхода вряд ли возможно, так как в лингвистике и лексикографии нет международного директивно-го органа, аналогичного МБ ВОИС.

В 2008 г. К. Ньюли опубликовал статью «Ten Long-Term Research Questions in Knowledge Organization», где были сформулированы 10 актуальных вопросов, относящихся к проблематике организации знания, поиск ответов на которые, с его точки зрения, важен в долгосрочной перспективе. Седьмой из поставленных Ньюли вопросов — «Как справиться с проблемой изменения знания при организации знания?» («How can KO deal with changes in knowledge?») [19, с. 142–143]. Поскольку одним из средств организации знания служат классификационные системы, вопрос, заданный Ньюли, предполагает и поиск ответа на вопрос: «Как справиться с проблемой изменения знания при его отражении с использованием классификационных рубрик?»

Проблема роста и изменения знания, влекущая за собой пересмотр классификационных систем и предполагающая разработку методов реклассификации, актуальна для широкого спектра отраслей знания, в которых создаются и используются классификационные системы и онтологии для индексирования и рубрицирования объектов исследования и/или их описаний. В информатике с этой проблемой сталкиваются, например, при создании баз данных и баз знаний с динамическими классификационными системами или онтологиями [20]. Эти системы и онтологии устаревают со временем, и поэтому они, как правило, регулярно обновляются, чтобы отразить рост и динамику научного знания. Подобное обновление нередко

влечет необходимость реклассификации тех объектов исследования и/или их описаний, которым индексы или рубрики были присвоены ранее.

Как было отмечено выше, необходимость в обновлении используемой классификационной системы проявилась и при развитии НБД НМГ. При этом ряд операций по изменению дефиниций рубрик ДКС НБД, а следовательно, по обновлению фасетной классификации может требовать реклассификации ранее сформированных аннотаций, в которых использовались рубрики, измененные впоследствии. Рассмотрим типы возможных изменений рубрик ДКС НБД, которая, как отмечено выше, является фасетной классификацией.

3 Типы изменений рубрик

Каждая рубрика ДКС НБД имеет: (1) уникальный идентификатор; (2) код; (3) дефиницию. В табл. 2 приводится фрагмент списка рубрик фасета «Значения НМГ», соответствующих первым пяти значениям глагола «sollen» согласно словарю [13]. Столбец 1 содержит уникальные неизменяемые идентификаторы рубрик. Столбец 2 содержит коды рубрик: для фасета «Значения НМГ» они отражают порядок, в котором описания значений слова приводятся в рамках словарной статьи в [13]. Эти коды могут меняться при изменении порядка описания значений. Столбец 3 содержит дефиниции рубрик. Так, исходные версии дефиниций рубрик фасета «Значения НМГ» были сформулированы на основании описаний значений модальных глаголов в [13] и внесены в НБД до начала аннотирования в сокращенной (по сравнению с [13]) и структурированной форме. Структура

Таблица 2 Фрагмент списка рубрик фасета «Значения НМГ» в НБД (на 23.04.2020)

Id рубрики	Код рубрики	Дефиниция рубрики
482	sollen-01	Обязанность что-л. делать по чьему-л. указанию, по закону, по правилам и т. п.: должен. Моральный запрет (под отрицанием): нельзя
483	sollen-02	Совет или требование, исходящее от других лиц, а также пожелание самого говорящего (в придаточных дополнительных, также в формах <i>conj</i> ; часто в неполных синтаксических конструкциях, как правило, с опущенным глаголом перемещения)
484	sollen-03	Желательность по мнению говорящего (в формах <i>praet conj</i> и <i>ppr conj</i>): следовало (бы), нужно было (бы), должно было (бы). Совет, рекомендация (только в формах <i>praet conj</i>). Нежелательность (под отрицанием): не следовало (бы), нельзя
485	sollen-04	Передача чужого мнения, приводимого со слов других лиц (в формах <i>praes</i>): говорят, полагают
486	sollen-05	Отнесенность действия к будущему, снятие категоричности (в формах <i>praes ind</i> ; также в вопросительных предложениях). Косвенный императив. Неуверенное согласие. Запрос информации для выполнения последующего действия (в вопросительных предложениях без вопросительного слова). Угроза и возмущение

дефиниции рубрики фасета «Значения НМГ» такова: ⟨стилистические пометы⟩, ⟨описание значения НМГ⟩, ⟨комментарий к употреблению НМГ⟩, ⟨варианты перевода на русский язык⟩. Другие зоны, присутствующие в описании значения в словаре, — зона примеров, зона фразеологии и т. д. — в НБД на данный момент не отражаются.

Так, дефиниция рубрики 482 из табл. 2 состоит из двух блоков, и ее структура может быть эксплицирована следующим образом.

Блок 1:

- 1.1 Стилистические пометы: — (здесь отсутствуют).
- 1.2 Описание значения НМГ: *Обязанность что-л. делать по чьему-л. указанию, по закону, по правилам и т. п.*
- 1.3 Комментарий к употреблению НМГ: — (здесь отсутствует).
- 1.4 Варианты перевода на русский язык: *должен.*

Блок 2:

- 2.1 Стилистические пометы: — (здесь отсутствуют).
- 2.2 Описание значения НМГ: *Моральный запрет.*
- 2.3 Комментарий к употреблению НМГ: *под отрицанием.*
- 2.4 Варианты перевода на русский язык: *нельзя.*

В результате наблюдения за процессом лингвистического аннотирования было выделено 7 типов изменений рубрик фасетной классификации. Чтобы отражать эти изменения в НБД, была предусмотрена возможность применения соответствующих 7 операций к рубрикам, входящим в фасет «Значения НМГ». Для их описания введем следующие обозначения:

- X, Y, Z, \dots — рубрики фасета «Значения НМГ»;
- $def_X, def_Y, def_Z, \dots$ — дефиниции рубрик фасета «Значения НМГ» (def — сокр. от англ. *definition*);
- $S_{def_X}, S_{def_Y}, S_{def_Z}, \dots$ — смысловое содержание дефиниций или семантика рубрик фасета «Значения НМГ» (S — сокр. от англ. *semantics*).

Операции, применяемые к рубрикам фасета «Значения НМГ»:

- (1) CREATE — создание новой рубрики X ;
- (2) REORDER — изменение кода рубрики X ;

- (3) REVISE — изменение def_X без сужения S_{def_X} , при условии что изменение def_X не связано с изменением дефиниций других рубрик;
- (4) MERGE — слияние def_X и def_Y , при котором def_Y включается в def_X , после чего рубрика Y удаляется;
- (5) DELETE — удаление рубрики X ;
- (6) SPLIT — разделение def_X на две части — def_{X1} и def_{X2} , в результате которого def_{X1} становится новой дефиницией рубрики X и создается новая рубрика Y , дефиницией которой становится def_{X2} ;
- (7) REDISTR (сокр. от англ. *redistribute*) — изменение def_X и def_Y , при котором происходит перераспределение смыслового содержания между S_{def_X} и S_{def_Y} .

В рамках НБД НМГ операции 4, 6 и 7 могут быть применены максимум к двум рубрикам. Если требуется объединить более двух дефиниций, разделить одну дефиницию более чем на две или перераспределить смысловое содержание более двух дефиниций, то это можно сделать путем последовательного выполнения нескольких операций (соответственно MERGE, SPLIT или REDISTR).

Отражение в НБД динамики изменения рубрик фасетной классификации преследует как минимум две цели.

Первая из них — обеспечить возможность определения объема работ по редактированию и реклассификации аннотаций при изменении ДКС.

Вторая — обеспечить лингвисту возможность увидеть, соответствует ли конкретная аннотация последней версии ДКС или нет. В этом случае при работе с НБД (например, при получении статистики) пользователь сможет исключить из рассмотрения те примеры, которые еще не были приведены в соответствие с последней версией ДКС методом реклассификации¹. На данный момент функции отражения истории изменений реализованы в НБД только для фасета «Значения НМГ» как наиболее динамичного.

Перечисленные выше 7 операций рассмотрим с точки зрения того, требуется ли после их выполнения реклассификация аннотаций, сформированных с использованием рубрик более ранней версии ДКС. На этом основании операции могут быть поделены на три группы:

¹Следует отметить, что в работе [17] при описании примеров библиотек, в которых проводилась реклассификация, особо отмечены те случаи, когда библиотека не прекращала функционировать во время работ по реклассификации и ее фонды были в полном объеме доступны для посетителей. По этой причине представляется важным обеспечение возможности полноценного использования НБД вне зависимости от того, подлежат ли некоторые аннотации редактированию и реклассификации.

Таблица 3 Пример выполнения операции REVISE

Id рубрики	Код рубрики	Дефиниция рубрики	Время записи
485	sollen-04	Передача <i>чужого</i> мнения, <i>приводимого со слов</i> других лиц (в формах <i>praes</i>): говорят, полагают.	13/04/2020 18:02:32
485	sollen-04	Передача мнения других лиц (в формах <i>praes</i>): говорят, полагают <i>и т. п.</i>	29/05/2020 13:37:21

- (1) реклассификация не осуществляется (CREATE, REORDER, REVISE)

В результате выполнения операции CREATE(X) в НБД не появляется таких аннотаций, которые требовали бы реклассификации.

При выполнении операций REORDER(X) или REVISE(X) реклассификация также не осуществляется, а в аннотациях, в которых X была проставлена ранее, начинает отображаться результат выполнения соответственно REORDER(X) — меняется код рубрики — или REVISE(X) — меняется дефиниция рубрики (см. пример выполнения операции REVISE в табл. 3; несовпадающие фрагменты дефиниций выделены полужирным курсивом);

- (2) реклассификация осуществляется программно (MERGE)

При выполнении MERGE(X, Y) def_Y включается в def_X, после чего Y удаляется. Лингвист, который вносит данное изменение в НБД, может самостоятельно определить поглощающую рубрику X и поглощаемую рубрику Y. В результате выполнения операции MERGE(X, Y) во всех аннотациях, где ранее была проставлена Y, автоматически проставляется идентификатор и код обновленной рубрики X;

- (3) реклассификация осуществляется лингвистами экспертно (DELETE, SPLIT, REDISTR)

При выполнении DELETE(X) аннотации, которые ранее были отнесены к X, автоматически помечаются рубрикой «TBR-D» (от англ. *To Be Reclassified because of Deletion*), указывающей, что требуется их реклассификация лингвистами, а сама X удаляется.

При выполнении SPLIT(X) аннотации, которые ранее были отнесены к рубрике X, автоматически помечаются рубрикой «TBR-S» (от англ. *To Be Reclassified because of Split*), указывающей, что требуется их реклассификация лингвистами, а также создается новая рубрика Y.

Операция REDISTR необходима потому, что нередко приходится иметь дело со случаями,

когда одновременно меняются def_X и def_Y, причем эти изменения оказываются взаимосвязанными. Дело в том, что описания разных значений НМГ в рамках одной и той же словарной статьи не автономны: они должны позволять разделить все употребления НМГ на непересекающиеся семантические классы, определения которых указаны в столбце 3 табл. 1, т. е. каждой дефиниции соответствует один класс. Все аннотации, которые подлежат реклассификации лингвистами после выполнения этой операции, автоматически помечаются рубрикой «TBR-R» (от англ. *To Be Reclassified because of Redistribution*).

4 Заключение

На примере задачи фиксации изменений, вносимых в описания значений НМГ, определены семь операций, необходимых для использования ДКС в НБД. Этот набор операций выявлен в результате эксперимента, в рамках которого с 2018 г. по настоящее время регистрировались все изменения, вносимые в дефиниции рубрик фасета «значения НМГ». Если вернуться к упомянутому в статье вопросу К. Ньюли: «Как справиться с проблемой изменения знания при организации знания?» — то можно сказать, что предлагаемый набор операций представляется одним из возможных методов решения этой проблемы. Из-за ускорения роста и динамики знания, которое необходимо учитывать в задачах классификации, расширяется сфера применения ДКС в информационных системах. Поэтому важной задачей становится разработка методов и операций изменения ДКС как инструментов организации знания, учитывающих его быстрый рост и динамику. В продолжение исследования операций изменения ДКС планируется рассмотреть следующие вопросы: (1) отражение истории изменений дефиниций рубрик ДКС в НБД; (2) определение объема аннотаций, которые должны быть реклассифицированы после внесения изменений в дефиниции ранее использованных рубрик.

Литература

1. Захаров В. П., Богданова С. Ю. Корпусная лингвистика. — Иркутск: ИГЛУ, 2011. 161 с.
2. McEnery T., Hardie A. Corpus linguistics: Method, theory and practice. — Cambridge: Cambridge University Press, 2012. 310 p.
3. Koneev M. Введение в корпусную лингвистику. — Прага: Animedia Co., 2014. 196 с.
4. Kübler S., Zinsmeister H. Corpus linguistics and linguistically annotated corpora. — London/New York: Bloomsbury, 2015. 320 p.
5. Кружков М. Г. Информационные ресурсы контрастных лингвистических исследований: электронные корпуса текстов // Системы и средства информатики, 2015. Т. 25. № 2. С. 140–159.
6. Гончаров А. А., Инькова О. Ю., Кружков М. Г. Методология аннотирования в надкорпусных базах данных // Системы и средства информатики, 2019. Т. 29. № 2. С. 148–160.
7. Handbook of linguistic annotation / Eds. N. Ide, J. Pustejovsky. — Dordrecht: Springer Science + Business Media, 2017. 1568 p.
8. Зализняк А. А., Зацман И. М., Инькова О. Ю. Надкорпусная база данных коннекторов: построение системы терминов // Информатика и её применения, 2017. Т. 11. Вып. 1. С. 100–108.
9. Зацман И. М., Инькова О. Ю., Кружков М. Г., Попкова Н. А. Представление кросязыковых знаний о коннекторах в надкорпусных базах данных // Информатика и её применения, 2016. Т. 10. Вып. 1. С. 106–118.
10. 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.
11. Inkova O., Popkova N. Statistical data as information source for linguistic analysis of Russian connectors // Информатика и её применения, 2017. Т. 11. Вып. 3. С. 123–131.
12. Зацман И., Кружков М., Лощилова Е. Методы и средства информатики для описания структуры неоднословных коннекторов // Структура коннекторов и методы ее описания / Под ред. О. Ю. Иньковой. — М.: ТОРУС ПРЕСС, 2019. С. 205–230.
13. Немецко-русский словарь: актуальная лексика / Под ред. Д. О. Добровольского. — М.: Лексрус, 2020 (в печати).
14. Гончаров А. А., Зацман И. М., Кружков М. Г. Темпоральные данные в лексикографических базах знаний // Информатика и её применения, 2019. Т. 13. № 4. С. 90–96.
15. 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.
16. Bentz D. M., Cavender T. P. Reclassification and recataloging // Libr. Trends, 1953. Vol. 2. Iss. 2. P. 249–263.
17. Kumbhar R. Library classification trends in the 21st century. — Oxford: Chandos Publishing, 2012. 186 p.
18. Зацман И. М., Косарик В. В., Курчавова О. А. Задачи представления личностных и коллективных концептов в цифровой среде // Информатика и её применения, 2008. Т. 2. Вып. 3. С. 54–69.
19. Gnoli C. Ten long-term research questions in knowledge organization // Knowl. Organ., 2008. Vol. 35. Iss. 2/3. P. 137–149.
20. Зацман И. М. Проблемно-ориентированная верификация полноты темпоральных онтологий и заполнение понятийных лакун // Информатика и её применения, 2020. Т. 14. Вып. 3. С. 119–128.

Поступила в редакцию 05.10.20

EVOLUTION OF CLASSIFICATIONS IN SUPRACORPORA DATABASES

A. A. Goncharov, I. M. Zatsman, and M. G. Kruzchkov

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

Abstract: The paper examines the task of recording changes to descriptions of meanings of German modal verbs in the process of annotating parallel German-Russian texts within a supracorpora database. This task was used as a case study to analyze the specifics of using dynamic classification systems (DCS) in information systems. The distinctive feature of a DCS is that semantic content of its concepts may change in the process of annotation which often entails the need to reclassify previously annotated data according to the changes made. This paper aims to answer the following questions: (i) What factors may have an impact on the need to edit and/or reclassify the annotations created prior to the concept changes? and (ii) What kind of operations may be used to represent

the changes to concepts in the DCS? The paper describes seven types of possible changes and enumerates the corresponding operations applied to the DCS concepts in the process of annotation. The operations are grouped in three categories depending on how they affect the need to reclassify the previously created annotations.

Keywords: dynamic classification; faceted classification; reclassification; supracorpora databases

DOI: 10.14357/19922264200415

Acknowledgments

The study has been conducted at the Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences with financial aid from the Russian Foundation for Basic Research (grants Nos. 18-07-00192 and 20-012-00166).

References

- Zakharov, V. P., and S. Yu. Bogdanova. 2011. *Korpusnaya lingvistika* [Corpus linguistics]. Irkutsk: IGLU. 161 p.
- McEnery, T., and A. Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press. 310 p.
- Kopotev, M. 2016. *Vvedenie v korpusnuyu lingvistiku* [Introduction to corpus linguistics]. Praha: Animedia Company. 196 p.
- Kübler, S., and H. Zinsmeister. 2015. *Corpus linguistics and linguistically annotated corpora*. London/New York: Bloomsbury. 320 p.
- 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.
- Goncharov, A. A., O. Yu. Inkova, and M. G. Kruzhkov. 2019. Metodologiya annotirovaniya v nadkorpusnykh bazakh dannykh [Annotation methodology of supracorpora databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 29(2):148–160.
- Ide, N., and J. Pustejovsky, eds. 2017. *Handbook of linguistic annotation*. Dordrecht: Springer Science + Business Media. 1568 p.
- Zaliznyak, A. A., I. M. Zatsman, and O. Yu. Inkova. 2017. Nadkorpusnaya baza dannykh konnektorov: postroenie sistemy terminov [Supracorpora database on connectives: Term system development]. *Informatika i ee Primeneniya — Inform. Appl.* 11(1):100–108.
- Zatsman, I. M., O. Yu. Inkova, M. G. Kruzhkov, and N. A. Popkova. 2016. Predstavlenie krossyazykovykh znaniy o konnektorakh v nadkorpusnykh bazakh dannykh [Representation of cross-lingual knowledge about connectors in supracorpora databases]. *Informatika i ee Primeneniya — Inform. Appl.* 10(1):106–118.
- Zatsman, I. M., O. Yu. Inkova, and V. A. Nuriev. 2017. The construction of classification schemes: Methods and technologies of expert formation. *Automatic Documentation Mathematical Linguistics* 51(1): 27–41.
- Inkova, O., and N. Popkova. 2017. Statistical data as information source for linguistic analysis of Russian connectors. *Informatika i ee Primeneniya — Inform. Appl.* 11(3):123–131.
- Zatsman, I., M. Kruzhkov, and E. Loshchilova. 2019. Metody i sredstva informatiki dlya opisaniya struktury neodnoslovnnykh konnektorov [Methods and means of informatics for multiword connectives structure description]. *Struktura konnektorov i metody ee opisaniya* [Connectives structure and methods of its description]. Ed. O. Yu. Inkova. Moscow: TORUS PRESS. 205–230.
- Dobrovol'skiy, D. O., ed. 2020 (in press). *Nemetsko-russkiy slovar': aktual'naya leksika* [German–Russian dictionary: Actual vocabulary]. Moscow: Leksrus.
- Goncharov, A. A., I. M. Zatsman, and M. G. Kruzhkov. 2019. Temporal'nye dannye v leksikograficheskikh bazakh znaniy [Temporal data in lexicographic databases]. *Informatika i ee Primeneniya — Inform. Appl.* 13(4):90–96.
- Dobrovol'skiy, D. O., and Anna A. Zalizniak. 2018. Nemetskie konstruksii s modal'nymi glagolami i ikh russkie sootvetstviya: proekt nadkorpusnoy bazy dannykh [German constructions with modal verbs and their Russian correlates: A supracorpora database project]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: po mat-lam Mezhdunar. konf. “Dialog”* [Computer Linguistic and Intellectual Technologies: Conference (International) “Dialog” Proceedings]. Moscow. 17(24):172–184.
- Bentz, D. M., and T. P. Cavender. 1953. Reclassification and recataloging. *Libr. Trends* 2(2):249–263.
- Kumbhar, R. 2012. *Library classification trends in the 21st century*. Oxford: Chandos Publishing. 186 p.
- Zatsman, I. M., V. V. Kosarik, and O. A. Kurchavova. 2008. Zadachi predstavleniya lichnostnykh i kollektivnykh kontseptov v tsifrovoy srede [Representation of individual and collective concepts in digital medium]. *Informatika i ee Primeneniya — Inform. Appl.* 2(3):54–69.
- Gnoli, C. 2008. Ten long-term research questions in knowledge organization. *Knowl. Organ.* 35(2/3):137–149.
- Zatsman, I. M. 2020. Problemno-orientirovannaya verifikatsiya polnoty temporal'nykh ontologiy i zapolnenie ponyatiynykh lakun [Problem-oriented verifying the completeness of temporal ontologies and filling conceptual lacunas]. *Informatika i ee Primeneniya — Inform. Appl.* 14(3):119–128.

Received October 5, 2020

Contributors

Zatsman Igor M. (b. 1952) — Doctor of Science in technology, Head of Department, 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; izatsman@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

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

Абгарян Каринэ Карленовна (р. 1963) — доктор физико-математических наук, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук; заведующая кафедрой Московского авиационного института (национального исследовательского университета)

Беленков Виктор Геннадьевич (р. 1952) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Бетелин Владимир Борисович (р. 1946) — доктор физико-математических наук, профессор, академик РАН, научный руководитель Федерального научного центра «Научно-исследовательский институт системных исследований Российской академии наук»

Будзко Владимир Игоревич (р. 1944) — доктор технических наук, академик Академии криптографии Российской Федерации, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Вайсер Кирилл Олегович (р. 2000) — студент Московского физико-технического института

Гайдамака Юлия Васильевна (р. 1971) — доктор физико-математических наук, профессор кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Голембиовский Дмитрий Юрьевич (р. 1960) — доктор технических наук, профессор, профессор кафедры исследования операций факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова; профессор кафедры банковского дела университета «Синергия»

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

Данилишин Артём Ростиславович (р. 1992) — аспирант кафедры исследования операций факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова

Дубнов Юрий Андреевич (р. 1990) — научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; старший преподаватель Национального исследовательского университета «Высшая школа экономики»

Дьяченко Юрий Георгиевич (р. 1958) — кандидат технических наук, старший научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ефросинин Дмитрий Владимирович (р. 1977) — доктор физико-математических наук, профессор Университета имени Иоганна Кеплера (Австрия, Линц); доцент Российского университета дружбы народов

Жолобов Владимир Александрович (р. 1998) — студент Московского физико-технического института

Забейхайло Михаил Иванович (р. 1956) — доктор физико-математических наук, доцент, главный научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Зацман Игорь Моисеевич (р. 1952) — доктор технических наук, заведующий отделом Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Кириков Игорь Александрович (р. 1955) — кандидат технических наук, директор Калининградского филиала Федерального исследовательского центра

«Информатика и управление» Российской академии наук

Колбин Илья Сергеевич (р. 1986) — кандидат физико-математических наук, научный сотрудник Вычислительного центра им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук

Королев Виктор Юрьевич (р. 1954) — доктор физико-математических наук, профессор, заведующий кафедрой математической статистики факультета вычислительной математики и кибернетики, главный научный сотрудник Московского центра фундаментальной и прикладной математики Московского государственного университета имени М. В. Ломоносова; ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Королёв Вадим Иванович (р. 1943) — доктор технических наук, профессор, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; профессор Финансового университета при Правительстве Российской Федерации

Кочеткова Ирина Андреевна (р. 1985) — кандидат физико-математических наук, доцент Российского университета дружбы народов; 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

Кушниренко Анатолий Георгиевич (р. 1944) — кандидат физико-математических наук, заведующий отделом Федерального научного центра «Научно-исследовательский институт системных исследований Российской академии наук»

Макеева Елена Дмитриевна (р. 1996) — аспирант кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Матюшенко Сергей Иванович (р. 1963) — кандидат физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Москалева Фаина Александровна (р. 1996) — аспирант кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Наумов Валерий Арсентьевич (р. 1950) — кандидат физико-математических наук, научный руководи-

тель Исследовательского института инноваций (Финляндия, Хельсинки)

Попков Алексей Юрьевич (р. 1978) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Попков Юрий Соломонович (р. 1937) — доктор технических наук, профессор, академик РАН, главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук; главный научный сотрудник Института проблем управления Российской академии наук; старший научный сотрудник ОРТ Брауде Колледжа, Кармиель, Израиль

Потанин Марк Станиславович (р. 1994) — аспирант Московского физико-технического института

Разумчик Ростислав Валерьевич (р. 1984) — кандидат физико-математических наук, ведущий научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Рождественский Юрий Владимирович (р. 1952) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

Румовская София Борисовна (р. 1985) — кандидат технических наук, научный сотрудник Калининградского филиала Федерального исследовательского центра «Информатика и управление» Российской академии наук

Самуйлов Константин Евгеньевич (р. 1955) — доктор технических наук, профессор, заведующий кафедрой прикладной информатики и теории вероятностей Российского университета дружбы народов; старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Семенов Алексей Львович (р. 1950) — доктор физико-математических наук, профессор, академик РАН, академик РАО, заведующий кафедрой Московского государственного университета имени М. В. Ломоносова; директор Института кибернетики и образовательной информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Смирнов Дмитрий Владимирович (р. 1984) — бизнес-партнер по ИТ Департамента безопасности ПАО «Сбербанк России»

Соколов Игорь Анатольевич (р. 1954) — доктор технических наук, академик РАН, директор Феде-

рального исследовательского центра «Информатика и управление» Российской академии наук

Сопрунов Сергей Федорович (р. 1949) — кандидат физико-математических наук, методист Центра педагогического мастерства

Соченков Илья Владимирович (р. 1985) — кандидат физико-математических наук, заведующий отделом Федерального исследовательского центра «Информатика и управление» Российской академии наук

Степченко Юрий Афанасьевич (р. 1951) — кандидат технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук

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

Харин Петр Александрович (р. 1993) — аспирант кафедры прикладной информатики и теории вероятностей Российского университета дружбы народов

Шоргин Всеволод Сергеевич (р. 1978) — кандидат технических наук, старший научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Шоргин Сергей Яковлевич (р. 1952) — доктор физико-математических наук, профессор, главный научный сотрудник Института проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук

Ядринцев Василий Владимирович (р. 1993) — аспирант кафедры информационных технологий Российского университета дружбы народов, инженер-исследователь Федерального исследовательского центра «Информатика и управление» Российской академии наук

АВТОРСКИЙ УКАЗАТЕЛЬ ЗА 2020 г.

	Вып.	Стр.
Абгарян К. К., Гаврилов Е. С. Интеграционная платформа для многомасштабного моделирования нейроморфных систем	2	104–110
Абгарян К. К., Колбин И. С. Применение многомасштабного подхода и методов анализа данных для моделирования теплопроводности в слоистых структурах	4	91–99
Агаларов Я. М. Оптимизация емкости основного накопителя в системе массового обслуживания типа $G/M/1/K$ с дополнительным накопителем	2	72–79
Агасандян Г. А. Вычислительные аспекты применения $CC-VaR$ на совокупности рынков	3	62–70
Агеев К. А., Сопин Э. С., Яркина Н. В., Самуйлов К. Е., Шоргин С. Я. Анализ механизмов нарезки сети с учетом гарантий для различных типов трафика	3	94–100
Адамова К. А. см. Шнурков П. В.		
Базилевский М. П. Многофакторные модели полносвязной линейной регрессии без ограничений на соотношения дисперсий ошибок переменных	2	92–97
Бахтеев О. Ю. см. Грабовой А. В.		
Беленков В. Г. см. Будзко В. И.		
Бетелин В. Б., Кушниренко А. Г., Леонов А. Г. Основные понятия программирования в изложении для дошкольников	3	55–61
Бетелин В. Б., Кушниренко А. Г., Семенов А. Л., Сопрунов С. Ф. О цифровой грамотности и средах ее формирования	4	100–107
Борисов А. В. Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям II: случай аддитивных шумов	1	17–23
Борисов А. В. Численные схемы фильтрации марковских скачкообразных процессов по дискретизованным наблюдениям III: случай мультипликативных шумов	2	10–18
Босов А. В. Управление выходом стохастической дифференциальной системы по квадратичному критерию. V. Случай неполной информации о состоянии	2	19–25
Босов А. В., Мартюшова Я. Г., Наумов А. В., Сапунова А. П. Байесовский подход к построению индивидуальной траектории пользователя в системе дистанционного обучения	3	86–93
Босов А. В., Стефанович А. И. Управление выходом стохастической дифференциальной системы по квадратичному критерию. IV. Альтернативное численное решение	1	24–30
Брюхов Д. О., Ступников С. А., Ковалёв Д. Ю., Шанин И. А. Нейрофизиология как предметная область для решения задач с интенсивным использованием данных	1	40–47
Будзко В. И., Ядринцев В. В., Соченков И. В., Королёв В. И., Беленков В. Г. Об одном подходе к формированию в условиях высокой неопределенности маркеров конфиденциальности в системах интенсивного использования данных	4	69–76
Вайсер К. О. см. Потанин М. С.		
Вохминцев А. В., Мельников А. В., Пачганов С. А. Метод навигации и составления карты в трехмерном пространстве на основе комбинированного решения вариационной подзадачи точка–точка ICP для аффинных преобразований	1	101–112
Гаврилов Е. С. см. Абгарян К. К.		
Гайдамака Ю. В. см. Москалева Ф. А.		
Голембиовский Д. Ю. см. Данилишин А. Р.		
Голембиовский Д. Ю. см. Данилишин А. Р.		
Гончаров А. А., Зацман И. М., Кружков М. Г. Эволюция классификаций в надкорпусных базах данных	4	108–116
Гончаров А. В., Стрижов В. В. Выравнивание декартовых произведений упорядоченных множеств	1	31–39

	Вып.	Стр.
Горшенин А. К., Королев В. Ю. Аппроксимация распределений размеров частиц лунного реголита на основе метода статистической симуляции выборок	2	50–57
Горшенин А. К., Королев В. Ю., Щербинина А. А. Статистическое оценивание распределений случайных коэффициентов стохастического дифференциального уравнения Ланжевена	3	3–12
Горшенин А. К., Кузьмин В. Ю. Анализ конфигураций LSTM-сетей для построения среднесрочных векторных прогнозов	1	10–16
Грабовой А. В., Бахтеев О. Ю., Стрижов В. В. Введение отношения порядка на множестве параметров аппроксимирующих моделей	2	58–65
Грушо А. А., Забейайло М. И., Смирнов Д. В., Тимонина Е. Е. О вероятностных оценках достоверности эмпирических выводов	4	3–8
Грушо А. А., Забейайло М. И., Смирнов Д. В., Тимонина Е. Е., Шоргин С. Я. Методы математической статистики в задаче поиска инсайдера	3	71–75
Грушо А. А., Забейайло М. И., Тимонина Е. Е. О каузальной репрезентативности обучающих выборок прецедентов в задачах диагностического типа	1	80–86
Грушо А. А., Тимонина Е. Е., Грушо Н. А., Терехина И. Ю. Выявление аномалий с помощью метаанных	3	76–80
Грушо А. А. см. Грушо Н. А.		
Грушо Н. А., Грушо А. А., Забейайло М. И., Тимонина Е. Е. Методы нахождения причин сбоев в информационных технологиях с помощью метаанных	2	33–39
Грушо Н. А. см. Грушо А. А.		
Данилишин А. Р., Голембиовский Д. Ю. Оценка стоимости опционов на основе моделей ARIMA–GARCH с ошибками, распределенными по закону S_u Джонсона	4	83–90
Данилишин А. Р., Голембиовский Д. Ю. Риск-нейтральная динамика для модели ARIMA–GARCH с ошибками, распределенными по закону S_U Джонсона	1	48–55
Диментов А. В. см. Краснов Ф. В.		
Донской В. И. Извлечение оптимизационных моделей из данных	3	109–118
Дубнов Ю. А. см. Попков Ю. С.		
Дулин С. К., Дулина Н. Г., Ермаков П. В. Информационный синтез документов	1	128–135
Дулина Н. Г. см. Дулин С. К.		
Дьяченко Ю. Г. см. Соколов И. А.		
Ермаков П. В. см. Дулин С. К.		
Ефросинин Д. В. см. Харин П. А.		
Жолобов В. А. см. Потанин М. С.		
Забейайло М. И. см. Грушо А. А.		
Забейайло М. И. см. Грушо А. А.		
Забейайло М. И. см. Грушо А. А.		
Забейайло М. И. см. Грушо Н. А.		
Захаров В. Н. см. Френкель С. Л.		
Зацман И. М. Проблемно-ориентированная верификация полноты темпоральных онтологий и заполнение понятийных лакун	3	119–128
Зацман И. М. см. Гончаров А. А.		
Зацман И. М. см. Нуриев В. А.		
Зейфман А. И. см. Сатин Я. А.		
Кириков И. А. см. Румовская С. Б.		
Кириллюк И. Л., Сенько О. В. Выбор моделей оптимальной сложности методами Монте-Карло (на примере моделей производственных функций регионов Российской Федерации)	2	111–118
Ковалёв Д. Ю. см. Брюхов Д. О.		
Козеренко Е. Б., Михеев М. Ю., Сомин Н. В., Эрлих Л. И., Кузнецов К. И. Аналитическая текстология в системах интеллектуальной обработки неструктурированных данных	1	113–120
Колбин И. С. см. Абгарян К. К.		

	Вып.	Стр.
Королев В. Ю. О распределении отношения суммы элементов выборки, превосходящих некоторый порог, к сумме всех элементов выборки. I	3	35–43
Королев В. Ю. О распределении отношения суммы элементов выборки, превосходящих некоторый порог, к сумме всех элементов выборки. II	4	33–36
Королев В. Ю. см. Горшенин А. К.		
Королев В. Ю. см. Горшенин А. К.		
Королёв В. И. см. Будзко В. И.		
Костина А. А., Мирин А. Ю., Молдовян Д. Н., Фахрутдинов Р. Ш. Метод задания конечных некоммутативных ассоциативных алгебр произвольной четной размерности для построения постквантовых криптосхем	1	94–100
Кочеткова И. А. см. Харин П. А.		
Краснов Ф. В., Диментов А. В., Шварцман М. Е. Использование тематических моделей для парного сравнения коллекций научных статей	3	129–135
Кривенко М. П. Последовательный анализ серий данных на основе многомерных референсных регионов	2	86–91
Кружков М. Г. см. Гончаров А. А.		
Кудрявцев А. А., Шестаков О. В. Метод логарифмических моментов для оценивания параметров гамма-экспоненциального распределения	3	49–54
Кузнецов К. И. см. Козеренко Е. Б.		
Кузьмин В. Ю. см. Горшенин А. К.		
Кушниренко А. Г. см. Бетелин В. Б.		
Кушниренко А. Г. см. Бетелин В. Б.		
Леонов А. Г. см. Бетелин В. Б.		
Макеева Е. Д. см. Харин П. А.		
Малашенко Ю. Е., Назарова И. А. Аппроксимация множества достижимых потоков многопользовательской сети	3	81–85
Мартюшова Я. Г. см. Босов А. В.		
Матюшенко С. И., Разумчик Р. В. Стационарные характеристики системы $Geo/G/1/\infty$ с неординарным входящим потоком, управляющим размером очереди	4	25–32
Мейханаджян Л. А., Разумчик Р. В. Стационарные характеристики системы $M/G/2/\infty$ с одним частным случаем дисциплины инверсионного порядка обслуживания с обобщенным вероятностным приоритетом	2	66–71
Мельников А. В. см. Вохминцев А. В.		
Мельников С. Ю., Самуйлов К. Е. Статистические свойства двоичных неавтономных регистров сдвига с внутренним суммированием	2	80–85
Милованова Т. А., Разумчик Р. В. Однолинейная система массового обслуживания с инверсионным порядком обслуживания с вероятностным приоритетом, групповым пуассоновским потоком и фоновыми заявками	3	26–34
Мирин А. Ю. см. Костина А. А.		
Михеев М. Ю. см. Козеренко Е. Б.		
Молдовян Д. Н. см. Костина А. А.		
Москалева Ф. А., Гайдамака Ю. В., Шоргин В. С. Влияние параметров изоляции на разделение ресурсов при нарезке сети	4	9–16
Назарова И. А. см. Малашенко Ю. Е.		
Наумов А. В. см. Босов А. В.		
Наумов В. А., Самуйлов К. Е. О марковских и рациональных потоках случайных событий. I	3	13–19
Наумов В. А., Самуйлов К. Е. О марковских и рациональных потоках случайных событий. II	4	37–46
Новиков Д. А. см. Шнурков П. В.		
Нуриев В. А., Зацман И. М. Редуцирование спектра моделей перевода в надкорпусных базах данных	2	119–126
Пачганов С. А. см. Вохминцев А. В.		

	Вып.	Стр.
Попков А. Ю. см. Попков Ю. С.		
Попков Ю. С., Попков А. Ю., Дубнов Ю. А. Методы детерминированных и рандомизированных энтропийных проекций для редукции размерности матрицы данных	4	47–54
Попов Г. А., Симаворян С. Ж., Симонян А. Р., Улитина Е. И. Моделирование процесса мониторинга систем информационной безопасности на основе систем массового обслуживания	1	71–79
Попов М. В., Посыпкин М. А. Аппроксимация множества решений систем нелинейных неравенств с использованием графических ускорителей	3	20–25
Посыпкин М. А. см. Попов М. В.		
Потанин М. С., Вайсер К. О., Жолобов В. А., Стрижов В. В. Оптимизация структуры сетей глубокого обучения	4	55–62
Разумчик Р. В. см. Матюшенко С. И.		
Разумчик Р. В. см. Мейханаджян Л. А.		
Разумчик Р. В. см. Милованова Т. А.		
Рождественский Ю. В. см. Соколов И. А.		
Румовская С. Б., Кириков И. А. Метод визуального представления конфликтов в гибридных интеллектуальных многоагентных системах	4	77–82
Самуйлов К. Е. см. Агеев К. А.		
Самуйлов К. Е. см. Мельников С. Ю.		
Самуйлов К. Е. см. Наумов В. А.		
Самуйлов К. Е. см. Наумов В. А.		
Сапунова А. П. см. Босов А. В.		
Сатин Я. А., Зейфман А. И., Шилова Г. Н. О подходах к построению предельных режимов для некоторых моделей массового обслуживания	2	3–9
Севастьянов Л. А., Щетинин Е. Ю. О методах повышения точности многоклассовой классификации на несбалансированных данных	1	63–70
Семенов А. Л. см. Бетелин В. Б.		
Сенько О. В. см. Кирилук И. Л.		
Серебрянский С. М., Тырсин А. Н. Повышение точности решения обратных задач за счет уточнения граничных условий	1	56–62
Симаворян С. Ж. см. Попов Г. А.		
Симонян А. Р. см. Попов Г. А.		
Смирнов Д. В. см. Грушо А. А.		
Смирнов Д. В. см. Грушо А. А.		
Соколов И. А., Степченков Ю. А., Дьяченко Ю. Г., Рождественский Ю. В. Повышение сбоеустойчивости самосинхронных схем	4	63–68
Сомин Н. В. см. Козеренко Е. Б.		
Сопин Э. С. см. Агеев К. А.		
Сопрунов С. Ф. см. Бетелин В. Б.		
Соченков И. В. см. Будзко В. И.		
Степченков Ю. А. см. Соколов И. А.		
Стефанович А. И. см. Босов А. В.		
Стрижов В. В. см. Гончаров А. В.		
Стрижов В. В. см. Грабовой А. В.		
Стрижов В. В. см. Потанин М. С.		
Ступников С. А. см. Брюхов Д. О.		
Терехина И. Ю. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо А. А.		
Тимонина Е. Е. см. Грушо Н. А.		
Тырсин А. Н. см. Серебрянский С. М.		
Улитина Е. И. см. Попов Г. А.		

	Вып.	Стр.
Фахрутдинов Р. Ш. см. Костина А. А.		
Френкель С. Л., Захаров В. Н. Совместная оценка предсказуемости данных и качества предикторов	2	40–49
Харин П. А., Макеева Е. Д., Кочеткова И. А., Ефросинин Д. В., Шоргин С. Я. Система массового обслуживания с орбитами для анализа совместного обслуживания трафика с малыми задержками URLLC и широкополосного доступа eMBB в беспроводных сетях пятого поколения	4	17–24
Хусаинов А. А. Производительность ограниченного конвейера	1	87–93
Шанин И. А. см. Брюхов Д. О.		
Шварцман М. Е. см. Краснов Ф. В.		
Шестаков О. В. Асимптотика оценки среднеквадратичного риска в задаче обращения преобразования Радона по проекциям, регистрируемым на случайной сетке	2	26–32
Шестаков О. В. Асимптотическая регулярность вейвлет-методов обращения линейных однородных операторов по наблюдениям, регистрируемым в случайные моменты времени	1	3–9
Шестаков О. В. О статистических свойствах оценки риска в задаче обращения преобразования Радона при случайном объеме проекционных данных	3	44–48
Шестаков О. В. см. Кудрявцев А. А.		
Шилова Г. Н. см. Сатин Я. А.		
Шихиев Ф. Ш. см. Шихиев Ш. Б.		
Шихиев Ш. Б., Шихиев Ф. Ш. Инкапсуляция семантических представлений в элементы грамматики	1	121–127
Шнурков П. В., Адамова К. А. Решение задачи безусловного экстремума для дробно-линейного интегрального функционала, зависящего от параметра	2	98–103
Шнурков П. В., Новиков Д. А. О концепции стохастической модели с управлением в моменты выхода процесса на границу заданного подмножества множества состояний	3	101–108
Шоргин В. С. см. Москалева Ф. А.		
Шоргин С. Я. см. Агеев К. А.		
Шоргин С. Я. см. Грушо А. А.		
Шоргин С. Я. см. Харин П. А.		
Щербинина А. А. см. Горшенин А. К.		
Щетинин Е. Ю. см. Севастьянов Л. А.		
Эрлих Л. И. см. Козеренко Е. Б.		
Ядринцев В. В. см. Будзко В. И.		
Яркина Н. В. см. Агеев К. А.		

2020 AUTHOR INDEX

	Issue	Page
Abgaryan K. K. and Gavrilov E. S. Integration platform for multiscale modeling of neuromorphic systems	2	104–110
Abgaryan K. K. and Kolbin I. S. Application of multiscale approach and data sciences for modeling thermal conductivity in layered structures	4	91–99
Adamova K. A. see Shnurkov P. V.		
Agalarov Ya. M. Optimization of the capacity of the main storage in $G/M/1/K$ queueing system with an additional storage device	2	72–79
Agasandyan G. A. Computational aspects of optimization on CC-VaR in a complex of markets	3	62–70
Ageev K. A., Sopin E. S., Yarkina N. V., Samouylov K. E., and Shorgin S. Ya. Analysis of the network slicing mechanisms with guaranteed allocated resources for various traffic types	3	94–100
Bakhteev O. Yu. see Grabovoy A. V.		
Bazilevskiy M. P. Multifactor fully connected linear regression models without constraints to the ratios of variables errors variances	2	92–97
Belenkov V. G. see Budzko V. I.		
Betelin V. B., Kushnirenko A. G., and Leonov A. G. Basic concepts of programming expounded for preschoolers	3	55–61
Betelin V. B., Kushnirenko A. G., Semenov A. L., and Soprunov S. F. About digital literacy and environments for its development	4	100–107
Borisov A. V. Numerical schemes of Markov jump process filtering given discretized observations II: Additive noise case	1	17–23
Borisov A. V. Numerical schemes of Markov jump process filtering given discretized observations III: Multiplicative noises case	2	10–18
Bosov A. V. Stochastic differential system output control by the quadratic criterion. V. Case of incomplete state information	2	19–28
Bosov A. V., Martyushova Ya. G., Naumov A. V., and Sapunova A. P. Bayesian approach to the construction of an individual user trajectory in the system of distance learning	3	86–93
Bosov A. V. and Stefanovich A. I. Stochastic differential system output control by the quadratic criterion. IV. Alternative numerical decision	1	24–30
Briukhov D. O., Stupnikov S. A., Kovalev D. Yu., and Shanin I. A. Neurophysiology as a subject domain for data intensive problem solving	1	40–47
Budzko V. I., Yadrintsev V. V., Sochenkov I. V., Korolev V. I., and Belenkov V. G. Extraction of confidentiality markers from texts under conditions of high uncertainty in systems with data intensive usage	4	69–76
Danilishin A. R. and Golembiovsky D. Yu. Estimating the fair value of options based on ARIMA–GARCH models with errors distributed according to the Johnson’s S_u law	4	83–90
Danilishin A. R. and Golembiovsky D. Yu. Risk-neutral dynamics for the ARIMA-GARCH random process with errors distributed according to the Johnson’s S_u law	1	48–55
Diachenko Yu. G. see Sokolov I. A.		
Dimentov A. V. see Krasnov F. V.		
Donskoy V. I. Optimization models extraction from data	3	109–118
Dubnov Y. A. see Popkov Y. S.		
Dulin S. K., Dulina N. G., and Ermakov P. V. Information fusion of documents	1	128–135
Dulina N. G. see Dulin S. K.		
Efrosinin D. V. see Kharin P. A.		
Ehrlich L. I. see Kozerenko E. B.		
Ermakov P. V. see Dulin S. K.		

	Issue	Page
Fahrutdinov R. Sh. see Kostina A. A.		
Frenkel S. L. and Zakharov V. N. Joint assessment of data predictability and quality predictors	2	40–49
Gaidamaka Yu. V. see Moskaleva F. A.		
Gavrilov E. S. see Abgaryan K. K.		
Golembiovsky D. Yu. see Danilishin A. R.		
Golembiovsky D. Yu. see Danilishin A. R.		
Goncharov A. V. and Strijov V. V. Alignment of ordered set Cartesian product	1	31–39
Goncharov A. A., Zatsman I. M., and Kruzhkov M. G. Evolution of classifications in supracorpora databases	4	108–116
Gorshenin A. K. and Korolev V. Yu. Approximation of particle size distributions of lunar regolith based on the resampling	2	50–57
Gorshenin A. K., Korolev V. Yu., and Shcherbinina A. A. Statistical estimation of distributions of random coefficients in the Langevin stochastic differential equation	3	3–12
Gorshenin A. K. and Kuzmin V. Yu. Analysis of configurations of LSTM networks for medium-term vector forecasting	1	10–16
Grabovoy A. V., Bakhteev O. Yu., and Strijov V. V. Ordering the set of neural network parameters	2	58–65
Grusho A. A., Timonina E. E., Grusho N. A., and Teryokhina I. Yu. Identifying anomalies using metadata	3	76–80
Grusho A. A., Zabezhailo M. I., Smirnov D. V., and Timonina E. E. On probabilistic estimates of the validity of empirical conclusions	4	3–8
Grusho A. A., Zabezhailo M. I., and Timonina E. E. On causal representativeness of training samples of precedents in diagnostic type tasks	1	80–86
Grusho A. A. see Grusho N. A.		
Grusho N. A., Grusho A. A., Zabezhailo M. I., and Timonina E. E. Methods of finding the causes of information technology failures by means of metadata	2	33–39
Grusho N. A., Zabezhailo M. I., Smirnov D. V., Timonina E. E., and Shorgin S. Ya. Mathematical statistics in the task of identifying hostile insiders	3	71–75
Grusho N. A. see Grusho A. A.		
Kharin P. A., Makeeva E. D., Kochetkova I. A., Efrosinin D. V., and Shorgin S. Ya. Retrial queuing model for analyzing joint URLLC and eMBB transmission in 5G networks	4	17–24
Khusainov A. A. Performance of the bounded pipeline	1	87–93
Kirikov I. A. see Rumovskaya S. B.		
Kirilyuk I. L. and Sen'ko O. V. Selection of optimal complexity models by methods of nonparametric statistics (on the example of production function model of regions of the Russian Federation)	2	111–118
Kochetkova I. A. see Kharin P. A.		
Kolbin I. S. see Abgaryan K. K.		
Korolev V. I. see Budzko V. I.		
Korolev V. Yu. On the distribution of the ratio of the sum of sample elements exceeding a threshold to the total sum of sample elements. I	3	35–43
Korolev V. Yu. On the distribution of the ratio of the sum of sample elements exceeding a threshold to the total sum of sample elements. II	4	33–36
Korolev V. Yu. see Gorshenin A. K.		
Korolev V. Yu. see Gorshenin A. K.		
Kostina A. A., Mirin A. Yu., Moldovyan D. N., and Fahrutdinov R. Sh. Method for defining finite noncommutative associative algebras of arbitrary even dimension for development of the postquantum cryptoschemes	1	94–100
Kovalev D. Yu. see Briukhov D. O.		
Kozerenko E. B., Mikheev M. Y., Somin N. V., Ehrlich L. I., and Kuznetsov K. I. Analytical textology in intelligent processing systems for unstructured data	1	113–120
Krasnov F. V., Dimentov A. V., and Shvartsman M. E. Using topic models for pairwise comparison of collections of scientific papers	3	129–135

	Issue	Page
Krivenko M. P. Sequential analysis of serial measurements based on multivariate reference regions	2	86–91
Kruzhkov M. G. see Goncharov A. A.		
Kudryavtsev A. A. and Shestakov O. V. Method of logarithmic moments for estimating the gamma-exponential distribution parameters	3	49–54
Kushnirenko A. G. see Betelin V. B.		
Kushnirenko A. G. see Betelin V. B.		
Kuzmin V. Yu. see Gorshenin A. K.		
Kuznetsov K. I. see Kozerenko E. B.		
Leonov A. G. see Betelin V. B.		
Makeeva E. D. see Kharin P. A.		
Malashenko Yu. E. and Nazarova I. A. Approximation of the multiuser network feasible flows set	3	81–85
Martyushova Ya. G. see Bosov A. V.		
Matyushenko S. I. and Razumchik R. V. Stationary characteristics of discrete-time Geo/G/1/∞ queue with batch arrivals and one queue skipping policy	4	25–32
Melnikov A. V. see Vokhmintcev A. V.		
Melnikov S. Yu. and Samouylov K. E. Statistical properties of binary nonautonomous shift registers with internal xor	2	80–85
Meykhanadzhyan L. A. and Razumchik R. V. Stationary characteristics of M/G/2/∞ queue with identical servers, LIFO service, and resampling policy	2	66–71
Mikheev M. Y. see Kozerenko E. B.		
Milovanova T. A. and Razumchik R. V. A single-server queueing system with LIFO service, probabilistic priority, batch Poisson arrivals, and background customers	3	26–34
Mirin A. Yu. see Kostina A. A.		
Moldovyan D. N. see Kostina A. A.		
Moskaleva F. A., Gaidamaka Yu. V., and Shorgin V. S. Impact of the isolation parameters on resource allocation in the network slicing model	4	9–16
Naumov A. V. see Bosov A. V.		
Naumov V. A. and Samouylov K. E. On Markovian and rational arrival processes. I	3	13–19
Naumov V. A. and Samouylov K. E. On Markovian and rational arrival processes. II	4	37–46
Nazarova I. A. see Malashenko Yu. E.		
Novikov D. A. see Shnurkov P. V.		
Nuriev V. A. and Zatsman I. M. Reducing the spectrum of translation models in supracorpora databases	2	119–126
Pachganov S. A. see Vokhmintcev A. V.		
Popkov A. Y. see Popkov Y. S.		
Popkov Y. S., Popkov A. Y., and Dubnov Y. A. Deterministic and randomized methods of entropy projection for dimensionality reduction problems	4	47–54
Popov G. A., Simavoryan S. Zh., Simonyan A. R., and Ulitina E. I. Modeling of monitoring of information security process on the basis of queuing systems	1	71–79
Popov M. V. and Posypkin M. A. Approximation of the set of solutions of systems of nonlinear inequalities using graphic accelerators	3	20–25
Posypkin M. A. see Popov M. V.		
Potantin M. S., Vayser K. O., Zholobov V. A., and Strijov V. V. Deep learning neural network structure optimization	4	55–62
Razumchik R. V. see Matyushenko S. I.		
Razumchik R. V. see Meykhanadzhyan L. A.		
Razumchik R. V. see Milovanova T. A.		
Rogdestvenski Yu. V. see Sokolov I. A.		
Rumovskaya S. B. and Kirikov I. A. Conflict visual representation method in hybrid intelligent multiagent systems	4	77–82
Samouylov K. E. see Ageev K. A.		

	Issue	Page
Samouylov K. E. see Melnikov S. Yu.		
Samouylov K. E. see Naumov V. A.		
Samouylov K. E. see Naumov V. A.		
Sapunova A. P. see Bosov A. V.		
Satin Ya. A., Zeifman A. I., and Shilova G. N. On approaches to constructing limiting regimes for some queuing models	2	3–9
Semenov A. L. see Betelin V. B.		
Sen'ko O. V. see Kirilyuk I. L.		
Serebryanskii S. M. and Tyrsin A. N. Improvement of the accuracy of solution of tasks for the account of the construction of boundary conditions	1	56–62
Sevastianov L. A. and Shchetinin E. Yu. On methods for improving the accuracy of multiclass classification on imbalanced data	1	63–70
Shanin I. A. see Briukhov D. O.		
Shcherbinina A. A. see Gorshenin A. K.		
Shchetinin E. Yu. see Sevastianov L. A.		
Shestakov O. V. Asymptotic regularity of the wavelet methods of inverting linear homogeneous operators from observations recorded at random times	1	3–9
Shestakov O. V. Asymptotics of the mean-square risk estimate in the problem of inverting the Radon transform from projections registered on a random grid	2	29–32
Shestakov O. V. On the statistical properties of risk estimate in the problem of inverting the Radon transform with a random volume of projection data	3	44–48
Shestakov O. V. see Kudryavtsev A. A.		
Shihiev F. Sh. see Shihiev Sh. B.		
Shihiev Sh. B. and Shihiev F. Sh. Incapsulation of semantic representations into elements of a grammar	1	121–127
Shilova G. N. see Satin Ya. A.		
Shnurkov P. V. and Adamova K. A. Solution of the unconditional extremal problem for a linear-fractional integral functional dependent on the parameter	2	98–103
Shnurkov P. V. and Novikov D. A. On the concept of a stochastic model with control at the moments of the process at the border of a presented subset of multiple states	3	101–108
Shorgin S. Ya. see Ageev K. A.		
Shorgin S. Ya. see Grusho N. A.		
Shorgin S. Ya. see Kharin P. A.		
Shorgin V. S. see Moskaleva F. A.		
Shvartsman M. E. see Krasnov F. V.		
Simavoryan S. Zh. see Popov G. A.		
Simonyan A. R. see Popov G. A.		
Smirnov D. V. see Grusho A. A.		
Smirnov D. V. see Grusho N. A.		
Sochenkov I. V. see Budzko V. I.		
Sokolov I. A., Stepchenkov Yu. A., Diachenko Yu. G., and Rogdestvenski Yu. V. Improvement of self-timed circuit soft error tolerance	4	63–68
Somin N. V. see Kozerenko E. B.		
Sopin E. S. see Ageev K. A.		
Soprunov S. F. see Betelin V. B.		
Stefanovich A. I. see Bosov A. V.		
Stepchenkov Yu. A. see Sokolov I. A.		
Strijov V. V. see Goncharov A. V.		
Strijov V. V. see Grabovoy A. V.		
Strijov V. V. see Potanin M. S.		
Stupnikov S. A. see Briukhov D. O.		
Teryokhina I. Yu. see Grusho A. A.		
Timonina E. E. see Grusho A. A.		

	Issue	Page
Timonina E. E. see Grusho A. A.		
Timonina E. E. see Grusho A. A.		
Timonina E. E. see Grusho N. A.		
Timonina E. E. see Grusho N. A.		
Tyrsin A. N. see Serebryanskii S. M.		
Ulitina E. I. see Popov G. A.		
Vayser K. O. see Potanin M. S.		
Vokhmintcev A. V., Melnikov A. V., and Pachganov S. A. Simultaneous localization and mapping method in three-dimensional space based on the combined solution of the point–point variation problem ICP for an affine transformation	1	101–112
Yadrintsev V. V. see Budzko V. I.		
Yarkina N. V. see Ageev K. A.		
Zabekhailo M. I. see Grusho A. A.		
Zabekhailo M. I. see Grusho A. A.		
Zabekhailo M. I. see Grusho N. A.		
Zabekhailo M. I. see Grusho N. A.		
Zakharov V. N. see Frenkel S. L.		
Zatsman I. M. Problem-oriented verifying the completeness of temporal ontologies and filling conceptual lacunas	3	119–128
Zatsman I. M. see Goncharov A. A.		
Zatsman I. M. see Nuriev V. A.		
Zeifman A. I. see Satin Ya. A.		
Zholobov V. A. see Potanin M. S.		

Правила подготовки рукописей для публикации в журнале «Информатика и её применения»

Журнал «Информатика и её применения» публикует теоретические, обзорные и дискуссионные статьи, посвященные научным исследованиям и разработкам в области информатики и ее приложений.

Журнал издается на русском языке. По специальному решению редколлегии отдельные статьи могут печататься на английском языке.

Тематика журнала охватывает следующие направления:

- теоретические основы информатики;
- математические методы исследования сложных систем и процессов;
- информационные системы и сети;
- информационные технологии;
- архитектура и программное обеспечение вычислительных комплексов и сетей.

1. В журнале печатаются статьи, содержащие результаты, ранее не опубликованные и не предназначенные к одновременной публикации в других изданиях.

Публикация предоставленной автором(ами) рукописи не должна нарушать положений глав 69, 70 раздела VII части IV Гражданского кодекса, которые определяют права на результаты интеллектуальной деятельности и средства индивидуализации, в том числе авторские права, в РФ.

Ответственность за нарушение авторских прав, в случае предъявления претензий к редакции журнала, несут авторы статей.

Направляя рукопись в редакцию, авторы сохраняют свои права на данную рукопись и при этом передают учредителям и редколлегии журнала неисключительные права на издание статьи на русском языке (или на языке статьи, если он отличен от русского) и на перевод ее на английский язык, а также на ее распространение в России и за рубежом. Каждый автор должен представить в редакцию подписанный с его стороны «Лицензионный договор о передаче неисключительных прав на использование произведения», текст которого размещен по адресу <http://www.ipiran.ru/publications/licence.doc>. Этот договор может быть представлен в бумажном (в 2-х экз.) или в электронном виде (отсканированная копия заполненного и подписанного документа).

Редколлегия вправе запросить у авторов экспертное заключение о возможности публикации предоставленной статьи в открытой печати.

2. К статье прилагаются данные автора (авторов) (см. п. 8). При наличии нескольких авторов указывается фамилия автора, ответственного за переписку с редакцией.

3. Редакция журнала осуществляет экспертизу присланных статей в соответствии с принятой в журнале процедурой рецензирования.

Возвращение рукописи на доработку не означает ее принятия к печати.

Доработанный вариант с ответом на замечания рецензента необходимо прислать в редакцию.

4. Решение редколлегии о публикации статьи или ее отклонении сообщается авторам.

Редколлегия может также направить авторам текст рецензии на их статью. Дискуссия по поводу отклоненных статей не ведется.

5. Редактура статей высылается авторам для просмотра. Замечания к редакции должны быть присланы авторами в кратчайшие сроки.

6. Рукопись предоставляется в электронном виде в форматах MS WORD (.doc или .docx) или \LaTeX (.tex), дополнительно — в формате .pdf, на дискете, лазерном диске или электронной почтой. Предоставление бумажной рукописи необязательно.

7. При подготовке рукописи в MS Word рекомендуется использовать следующие настройки.

Параметры страницы: формат — А4; ориентация — книжная; поля (см): внутри — 2,5, снаружи — 1,5, сверху — 2, снизу — 2, от края до нижнего колонтитула — 1,3.

Основной текст: стиль — «Обычный», шрифт — Times New Roman, размер — 14 пунктов, абзацный отступ — 0,5 см, 1,5 интервала, выравнивание — по ширине.

Рекомендуемый объем рукописи — не свыше 10 страниц указанного формата. При превышении указанного объема редколлегия вправе потребовать от автора сокращения объема рукописи.

Сокращения слов, помимо стандартных, не допускаются. Допускается минимальное количество аббревиатур.

Все страницы рукописи нумеруются.

Шаблоны примеров оформления представлены в Интернете: <http://www.ipiran.ru/journal/template.doc>

8. Статья должна содержать следующую информацию на **русском и английском языках**:

- название статьи;
- Ф.И.О. авторов, на английском можно только имя и фамилию;
- место работы, с указанием почтового адреса организации и электронного адреса каждого автора;
- сведения об авторах, в соответствии с форматом, образцы которого представлены на страницах:
http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp и
http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;
- аннотация (не менее 100 слов на каждом из языков). Аннотация — это краткое резюме работы, которое может публиковаться отдельно. Она является основным источником информации в информационных системах и базах данных. Английская аннотация должна быть оригинальной, может не быть дословным переводом русского текста и должна быть написана хорошим английским языком. В аннотации не должно быть ссылок на литературу и, по возможности, формул;
- ключевые слова — желательно из принятых в мировой научно-технической литературе тематических тезаурусов. Предложения не могут быть ключевыми словами;
- источники финансирования работы (ссылки на гранты, проекты, поддерживающие организации и т. п.).

9. Требования к спискам литературы.

Ссылки на литературу в тексте статьи нумеруются (в квадратных скобках) и располагаются в каждом из списков литературы в порядке первых упоминаний.

Списки литературы представляются в двух вариантах:

- (1) **Список литературы к русскоязычной части.** Русские и английские работы — на языке и в алфавите оригинала;
- (2) **References.** Русские работы и работы на других языках — в латинской транслитерации с переводом на английский язык; английские работы и работы на других языках — на языке оригинала.

Необходимо для составления списка “References” пользоваться размещенной на сайте <http://www.translit.net/ru/bgn/> бесплатной программой транслитерации русского текста в латиницу.

Список литературы “References” приводится полностью отдельным блоком, повторяя все позиции из списка литературы к русскоязычной части, независимо от того, имеются или нет в нем иностранные источники. Если в списке литературы к русскоязычной части есть ссылки на иностранные публикации, набранные латиницей, они полностью повторяются в списке “References”.

Ниже приведены примеры ссылок на различные виды публикаций в списке “References”.

Описание статьи из журнала:

Zagurenko, A. G., V. A. Korotovskikh, A. A. Kolesnikov, A. V. Timonov, and D. V. Kardymon. 2008. Tekhniko-ekonomicheskaya optimizatsiya dizayna gidrorazryva plasta [Technical and economic optimization of the design of hydraulic fracturing]. *Neftyanoe hozyaystvo [Oil Industry]* 11:54–57.

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S10231935080077.

Описание статьи из электронного журнала:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Описание статьи из продолжающегося издания (сборника трудов):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental'noe issledovanie prochnosti soedineniy “stal”–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Описание материалов конференций:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma "Novye resursosberegayushchie tekhnologii nedropol'zovaniya i povysheniya neftegazootdachi"* [6th Symposium (International) "New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact" Proceedings]. Moscow. 267–272.

Описание книги (монографии, сборники):

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem* [Operation of turbine generators with direct cooling]. Moscow: Energy Publs. 352 p.

Latyshev, V. N. 2009. *Tribologiya rezaniya. Kn. 1: Friksionnye protsessy pri rezanii metallov* [Tribology of cutting. Vol. 1: Frictional processes in metal cutting]. Ivanovo: Ivanovskii State Univ. 108 p.

Описание переводной книги (в списке литературы к русскоязычной части необходимо указать: / Пер. с англ. — после названия книги, а в конце ссылки указать оригинал книги в круглых скобках):

1. В русскоязычной части:

Тимошенко С. П., Янг Д. Х., Уивер У. Колебания в инженерном деле / Пер. с англ. — М.: Машиностроение, 1985. 472 с. (*Timoshenko S. P., Young D. H., Weaver W. Vibration problems in engineering. — 4th ed. — New York, NY, USA: Wiley, 1974. 521 p.*)

2. В англоязычной части:

Timoshenko, S. P., D. H. Young, and W. Weaver. 1974. *Vibration problems in engineering*. 4th ed. New York: Wiley. 521 p.

Описание неопубликованного документа:

Latypov, A. R., M. M. Khasanov, and V. A. Baikov. 2004 (unpubl.). *Geologiya i dobycha (NGT GiD)* [Geology and production (NGT GiD)]. Certificate on official registration of the computer program No. 2004611198.

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]. Moscow. D.Sc. Diss. 272 p.

Kozhunova, O. S. 2009. *Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa* [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.

Описание ГОСТа:

GOST 8.586.5-2005. 2007. *Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichstva zhidkostey i gazov s pomoshch'yu standartnykh suzhayushchikh ustroystv* [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. Moscow: Standardinform Publs. 10 p.

Описание патента:

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. *Sposob orientirovaniya po krenu letatel'nogo apparata s opticheskoy golovkoy samonavedeniya* [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

10. Присланные в редакцию материалы авторам не возвращаются.

11. При отправке файлов по электронной почте просим придерживаться следующих правил:

- указывать в поле subject (тема) название журнала и фамилию автора;
- использовать attach (присоединение);
- в состав электронной версии статьи должны входить: файл, содержащий текст статьи, и файл(ы), содержащий(е) иллюстрации.

12. Журнал «Информатика и её применения» является некоммерческим изданием. Плата за публикацию не взимается, гонорар авторам не выплачивается.

Адрес редакции журнала «Информатика и её применения»:

Москва 119333, ул. Вавилова, д. 44, корп. 2, ФИЦ ИУ РАН

Тел.: +7 (499) 135-86-92 Факс: +7 (495) 930-45-05

e-mail: iier@frccsc.ru (Стригина Светлана Николаевна)

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

Requirements for manuscripts submitted to Journal “Informatics and Applications”

Journal “Informatics and Applications” (Inform. Appl.) publishes theoretical, review, and discussion articles on the research and development in the field of informatics and its applications.

The journal is published in Russian. By a special decision of the editorial board, some articles can be published in English.

The topics covered include the following areas:

- theoretical fundamentals of informatics;
- mathematical methods for studying complex systems and processes;
- information systems and networks;
- information technologies; and
- architecture and software of computational complexes and networks.

1. The Journal publishes original articles which have not been published before and are not intended for simultaneous publication in other editions. An article submitted to the Journal must not violate the Copyright law. Sending the manuscript to the Editorial Board, the authors retain all rights of the owners of the manuscript and transfer the nonexclusive rights to publish the article in Russian (or the language of the article, if not Russian) and its distribution in Russia and abroad to the Founders and the Editorial Board. Authors should submit a letter to the Editorial Board in the following form:

Agreement on the transfer of rights to publish:

“We, the undersigned authors of the manuscript “. . .”, pass to the Founder and the Editorial Board of the Journal “Informatics and Applications” the nonexclusive right to publish the manuscript of the article in Russian (or in English) in both print and electronic versions of the Journal. We affirm that this publication does not violate the Copyright of other persons or organizations.

Author(s) signature(s): (name(s), address(es), date).

This agreement should be submitted in paper form or in the form of a scanned copy (signed by the authors).

2. A submitted article should be attached with **the data on the author(s)** (see item 8). If there are several authors, the contact person should be indicated who is responsible for correspondence with the Editorial Board and other authors about revisions and final approval of the proofs.
3. The Editorial Board of the Journal examines the article according to the established reviewing procedure. If the authors receive their article for correction after reviewing, it does not mean that the article is approved for publication. The corrected article should be sent to the Editorial Board for the subsequent review and approval.
4. The decision on the article publication or its rejection is communicated to the authors. The Editorial Board may also send the reviews on the submitted articles to the authors. Any discussion upon the rejected articles is not possible.
5. The edited articles will be sent to the authors for proofread. The comments of the authors to the edited text of the article should be sent to the Editorial Board as soon as possible.
6. The manuscript of the article should be presented electronically in the MS WORD (.doc or .docx) or \LaTeX (.tex) formats, and additionally in the .pdf format. All documents may be sent by e-mail or provided on a CD or diskette. A hard copy submission is not necessary.
7. The recommended typesetting instructions for manuscript.

Pages parameters: format A4, portrait orientation, document margins (cm): left — 2.5, right — 1.5, above — 2.0, below — 2.0, footer 1.3.

Text: font — Times New Roman, font size — 14, paragraph indent — 0.5, line spacing — 1.5, justified alignment.

The recommended manuscript size: not more than 10 pages of the specified format. If the specified size exceeded, the editorial board is entitled to require the author to reduce the manuscript.

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

All pages of the manuscript should be numbered.

The templates for the manuscript typesetting are presented on site: <http://www.ipiran.ru/journal/template.doc>.

8. The articles should enclose data both in **Russian and English**:

- title;
- author’s name and surname;
- affiliation — organization, its address with ZIP code, city, country, and official e-mail address;
- data on authors according to the format: (see site)

http://www.ipiran.ru/journal/issues/2013_07_01/authors.asp and

http://www.ipiran.ru/journal/issues/2013_07_01_eng/authors.asp;

- abstract (not less than 100 words) both in Russian and in English. Abstract is a short summary of the article that can be published separately. The abstract is the main source of information on the article and it could be included in leading information systems and data bases. The abstract in English has to be an original text and should not be an exact translation of the Russian one. Good English is required. In abstracts, avoid references and formulae;
 - indexing is performed on the basis of keywords. The use of keywords from the internationally accepted thematic Thesauri is recommended.
Important! Keywords must not be sentences;
 - Acknowledgments.
9. References. Russian references have to be presented both in English translation and Latin transliteration (refer <http://www.translit.net/ru/bgn/>).

Please take into account the following examples of Russian references appearance:

Article in journal:

Zhang, Z., and D. Zhu. 2008. Experimental research on the localized electrochemical micromachining. *Russ. J. Electrochem.* 44(8):926–930. doi:10.1134/S1023193508080077.

Journal article in electronic format:

Swaminathan, V., E. Lepkoswka-White, and B. P. Rao. 1999. Browsers or buyers in cyberspace? An investigation of electronic factors influencing electronic exchange. *JCMC* 5(2). Available at: <http://www.ascusc.org/jcmc/vol5/issue2/> (accessed April 28, 2011).

Article from the continuing publication (collection of works, proceedings):

Astakhov, M. V., and T. V. Tagantsev. 2006. Eksperimental’noe issledovanie prochnosti soedineniy “stal’–kompozit” [Experimental study of the strength of joints “steel–composite”]. *Trudy MGTU “Matematicheskoe modelirovanie slozhnykh tekhnicheskikh sistem” [Bauman MSTU “Mathematical Modeling of Complex Technical Systems” Proceedings]*. 593:125–130.

Conference proceedings:

Usmanov, T. S., A. A. Gusmanov, I. Z. Mullagalin, R. Ju. Muhametshina, A. N. Chervyakova, and A. V. Sveshnikov. 2007. Osobennosti proektirovaniya razrabotki mestorozhdeniy s primeneniem gidrorazryva plasta [Features of the design of field development with the use of hydraulic fracturing]. *Trudy 6-go Mezhdunarodnogo Simpoziuma “Novye resursosberegayushchie tekhnologii nedropol’zovaniya i povysheniya neftegazoidachi” [6th Symposium (International) “New Energy Saving Subsoil Technologies and the Increasing of the Oil and Gas Impact” Proceedings]*. Moscow. 267–272.

Books and other monographs:

Lindorf, L. S., and L. G. Mamikonians, eds. 1972. *Ekspluatatsiya turbogeneratorov s neposredstvennym okhlazhdeniem [Operation of turbine generators with direct cooling]*. Moscow: Energy Publs. 352 p.

Dissertation and Thesis:

Kozhunova, O. S. 2009. Tekhnologiya razrabotki semanticheskogo slovary informatsionnogo monitoringa [Technology of development of semantic dictionary of information monitoring system]. Moscow: IPI RAN. PhD Thesis. 23 p.

State standards and patents:

GOST 8.586.5-2005. 2007. Metodika vypolneniya izmereniy. Izmerenie raskhoda i kolichestva zhidkostey i gazov s pomoshch’yu standartnykh suzhayushchikh ustroystv [Method of measurement. Measurement of flow rate and volume of liquids and gases by means of orifice devices]. M.: Standardinform Publs. 10 p.

Bolshakov, M. V., A. V. Kulakov, A. N. Lavrenov, and M. V. Palkin. 2006. Sposob orientirovaniya po krenu letatel’nogo apparata s opticheskoy golovkoy samonavedeniya [The way to orient on the roll of aircraft with optical homing head]. Patent RF No. 2280590.

References in Latin transcription are presented in the original language.

References in the text are numbered according to the order of their first appearance; the number is placed in square brackets. All items from the reference list should be cited.

10. Manuscripts and additional materials are not returned to Authors by the Editorial Board.

11. Submissions of files by e-mail must include:

- the journal title and author’s name in the “Subject” field;
- an article and additional materials have to be attached using the “attach” function;
- an electronic version of the article should contain the file with the text and a separate file with figures.

12. “Informatics and Applications” journal is not a profit publication. There are no charges for the authors as well as there are no royalties.

Editorial Board address:

FRC CSC RAS, 44, block 2, Vavilov Str., Moscow 119333, Russia

Ph.: +7 (499) 135 86 92, Fax: +7 (495) 930 45 05

e-mail: iiep@frccsc.ru (to Svetlana Strigina)

<http://www.ipiran.ru/english/journal.asp>