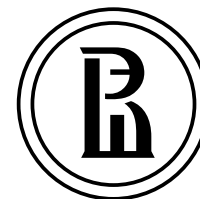


БИЗНЕС- ИНФОРМАТИКА

НАУЧНЫЙ ЖУРНАЛ НИУ ВШЭ



Издатель:
Национальный
исследовательский университет
«Высшая школа экономики»

СОДЕРЖАНИЕ

Е.В. Румянцева, К.К. Фурманов

Использование вневыборочных остатков
Кокса–Снелл при прогнозировании
наступления событий 7

*Е.А. Исаев, Д.В. Первухин, Г.О. Рытиков,
Е.К. Филюгина, Д.А. Айрапетян*

Оценка эффективности информационных
систем с учетом рисков 19

Ю.А. Зеленков, Е.А. Анисичкина

Динамика исследований в области
интеллектуального анализа данных:
тематический анализ публикаций за 20 лет 30

*В.И. Ананьин, К.В. Зимин, М.И. Лугачев,
Р.Д. Гимранов*

Статистическая устойчивость
цифровой организации 47

Е.В. Кислицын, В.В. Городничев

Имитационное моделирование развития
отдельных отраслей тяжелой промышленности 59

H. Nozari, M. Fallah, H. Kazemipoor, S.E. Najafi

Big data analysis of IoT-based supply chain
management considering FMCG industries 78

Выпускается ежеквартально

Журнал включен в Перечень
российских рецензируемых
научных журналов,
в которых должны быть
опубликованы основные научные
результаты диссертаций
на соискание ученых степеней
доктора и кандидата наук

Главный редактор
Е.А. Кучерявый

Заместитель главного редактора
Е.П. Зараменских

Компьютерная верстка
О.А. Богданович

Администратор веб-сайта
И.И. Хрусталева

Адрес редакции:
119049, г. Москва,
ул. Шаболовка, д. 28/11, стр. 4
Тел./факс: +7 (495) 772-9590 *26311
<http://bijournal.hse.ru>
E-mail: bijournal@hse.ru

За точность приведенных сведений
и содержание данных,
не подлежащих открытой публикации,
несут ответственность авторы

**При перепечатке ссылка на журнал
«Бизнес-информатика» обязательна**

Тираж:
русскоязычная версия – 150 экз.,
англоязычная версия – 150 экз.,
онлайн-версии на русском и английском –
свободный доступ

Отпечатано в типографии НИУ ВШЭ
г. Москва, Кочновский проезд, 3

© Национальный
исследовательский университет
«Высшая школа экономики»

«**Б**изнес-информатика» — рецензируемый междисциплинарный научный журнал, выпускаемый с 2007 года Национальным исследовательским университетом «Высшая школа экономики» (НИУ ВШЭ). Администрирование журнала осуществляется Высшей школой бизнеса НИУ ВШЭ. Журнал выпускается ежеквартально.

Миссия журнала — развитие бизнес-информатики как новой области информационных технологий и менеджмента. Журнал осуществляет распространение последних разработок технологического и методологического характера, способствует развитию соответствующих компетенций, а также обеспечивает возможности для дискуссий в области применения современных информационно-технологических решений в бизнесе, менеджменте и экономике.

Журнал публикует статьи по следующей тематике: моделирование социальных и экономических систем, цифровая трансформация бизнеса, управление инновациями, информационные системы и цифровые технологии в бизнесе, анализ данных и системы бизнес-интеллекта, математические методы и алгоритмы бизнес-информатики, моделирование и анализ бизнес-процессов, поддержка принятия управленческих решений.

В соответствии с решением президиума Высшей аттестационной комиссии Российской Федерации журнал включен в Перечень российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней доктора

и кандидата наук. Журнал «Бизнес-информатика» включен в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней кандидата и доктора наук (Перечень ВАК) по следующим научным специальностям и соответствующим отраслям науки: 05.13.10 — Управление в социальных и экономических системах (технические науки); 05.13.11 — Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей (технические науки); 05.13.18 — Математическое моделирование, численные методы и комплексы программ (технические науки); 05.25.05 — Информационные системы и процессы (технические науки); 08.00.05 — Экономика и управление народным хозяйством (по отраслям и сферам деятельности) (экономические науки); 08.00.13 — Математические и инструментальные методы экономики (экономические науки).

Журнал входит в базы Web of Science Emerging Sources Citation Index (WoS ESCI), Russian Science Citation Index на платформе Web of Science (RSCI), EBSCO.

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор), свидетельство ПИ № ФС77-66609 от 08 августа 2016 г.

Международный стандартный серийный номер (ISSN): 1998-0663 (на русском), 2587-814X (на английском).

Главный редактор: Кучерявый Евгений Андреевич.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

ГЛАВНЫЙ РЕДАКТОР

Кучерявый Евгений Андреевич

Университет Тампере, Тампере, Финляндия

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА

Зараменских Евгений Петрович

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

ЧЛЕНЫ РЕДКОЛЛЕГИИ

Абдульраб Абиб

Национальный институт прикладных наук, Руан, Франция

Авдошин Сергей Михайлович

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

Акопов Андриани Сумбатович

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

Алескерев Фуад Тагиевич

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

Афанасьев Александр Петрович

Институт проблем передачи информации им. А.А. Харкевича РАН,
Москва, Россия

Афанасьев Антон Александрович

Центральный экономико-математический институт РАН,
Москва, Россия

Бабкин Эдуард Александрович

Национальный исследовательский университет
«Высшая школа экономики», Нижний Новгород, Россия

Баландин Сергей Игоревич

Ассоциация FRUCT, Хельсинки, Финляндия

Баранов Александр Павлович

Главный научно-исследовательский вычислительный центр
Федеральной налоговой службы, Москва, Россия

Барахнин Владимир Борисович

Федеральный исследовательский центр информационных
и вычислительных технологий, Новосибирск, Россия

Беккер Йорг

Университет Мюнстера, Мюнстер, Германия

Белов Владимир Викторович

Рязанский государственный радиотехнический университет,
Рязань, Россия

Вестнер Маркус

Регенсбургский университет прикладных наук,
Регенсбург, Германия

Гаврилова Татьяна Альбертовна

Санкт-Петербургский государственный университет,
Санкт-Петербург, Россия

Глотен Эрве

Тулонский университет, Ла-Гард, Франция

Голосов Алексей Олегович

Компания «ФОРС — Центр разработки», Москва, Россия

Грибов Андрей Юрьевич

Компания «КиберПлат», Москва, Россия

Громов Александр Игоревич

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

Гурвич Владимир Александрович

Раттерский университет (Университет Нью-Джерси),
Раттерс, США

Демидова Лилия Анатольевна

Рязанский государственный радиотехнический университет,
Рязань, Россия

Джейкобс Лоренц

Университет Цюриха, Цюрих, Швейцария

Дискин Иосиф Евгеньевич

Всероссийский центр изучения общественного мнения,
Москва, Россия

Ефимушкин Владимир Александрович

Центральный научно-исследовательский институт связи,
Москва, Россия

Зандкуль Курт

Университет Росток, Росток, Германия

Иванников Александр Дмитриевич

Институт проблем проектирования в микроэлектронике РАН,
Москва, Россия

Ильин Николай Иванович

Федеральная служба охраны Российской Федерации, Москва, Россия

Исаев Дмитрий Валентинович

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

Калягин Валерий Александрович

Национальный исследовательский университет
«Высшая школа экономики», Нижний Новгород, Россия

Кравченко Татьяна Константиновна

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

Кузнецов Сергей Олегович

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

Лугачев Михаил Иванович

Московский государственный университет
им. М.В. Ломоносова, Москва, Россия

Лин Квей-Жей

Технологический институт Нагоя, Нагоя, Япония

Мальцева Светлана Валентиновна

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

Мейор Питер

Комиссия ООН по науке и технологиям, Женева, Швейцария

Миркин Борис Григорьевич

Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия

Моттль Вадим Вячеславович

Тульский государственный университет, Тула, Россия

Назаров Дмитрий Михайлович

Уральский государственный экономический университет,
Екатеринбург, Россия

Пальчунов Дмитрий Евгеньевич

Новосибирский государственный университет, Новосибирск, Россия

Пардалос Панайот (Панос)

Университет Флориды, Гейнсвилл, США

Пастор Оскар

Политехнический университет Валенсии, Валенсия, Испания

Посегга Йохим

Университет Пассау, Пассау, Германия

Самуйлов Константин Евгеньевич

Российский университет дружбы народов, Москва, Россия

Сюняев Али Рашидович

Технологический институт Карлсруэ, Карлсруэ, Германия

Таратухин Виктор Владимирович

Университет Мюнстера, Мюнстер, Германия

Триболе Жозе

Университет Лиссабона, Лиссабон, Португалия

Ульянов Михаил Васильевич

Институт проблем управления им В.А. Трапезникова РАН,
Москва, Россия

Ускенбаева Раиса Кабиевна

Международный университет информационных технологий,
Алматы, Казахстан

Цуканова Ольга Анатольевна

Санкт-Петербургский национальный исследовательский
университет информационных технологий, механики и оптики,
Санкт-Петербург, Россия

Чхартишвили Александр Гедванович

Институт проблем управления им В.А. Трапезникова РАН,
Москва, Россия

Шмидт Юрий Давыдович

Дальневосточный федеральный университет, Владивосток, Россия

Штраус Кристина

Университет Вены, Вена, Австрия

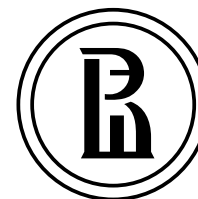
ISSN 1998-0663 (print), ISSN 2587-8166 (online)

English version: ISSN 2587-814X (print), ISSN 2587-8158 (online)

Vol. 15 No 1 – 2021

BUSINESS INFORMATICS

HSE SCIENTIFIC JOURNAL



Publisher:

National Research University
Higher School of Economics

The journal is published quarterly

The journal is included
into the list of peer reviewed
scientific editions established
by the Supreme Certification
Commission of the Russian Federation

Editor-in-Chief:

Y. Koucheryav

Deputy Editor-in-Chief

E. Zaramenskikh

Computer Making-up:

O. Bogdanovich

Website Administration:

I. Khrustaleva

Address:

28/11, build. 4, Shablovka Street
Moscow 119049, Russia

Tel./fax: +7 (495) 772-9590 *26311

<http://bijournal.hse.ru>

E-mail: bijournal@hse.ru

Circulation:

English version – 150 copies,
Russian version – 150 copies,
online versions in English and Russian –
open access

Printed in HSE Printing House
3, Kochnovsky Proezd, Moscow,
Russia

© National Research University
Higher School of Economics

CONTENTS

E.V. Rumyantseva, K.K. Furmanov

Using out-of-sample Cox–Snell residuals
in time-to-event forecasting7

*E.A. Isaev, D.V. Pervukhin, G.O. Rytikov,
E.K. Filyugina, D.A. Hayrapetyan*

Risk-based efficiency assessment
of information systems..... 19

Yu.A. Zelenkov, E.A. Anisichkina

Trends in data mining research:
A two-decade review using topic analysis.....30

*V.I. Ananyin, K.V. Zimin, M.I. Lugachev,
R.D. Gimranov*

Statistical sustainability
of a digital organization.....47

E.V. Kislitsyn, V.V. Gorodnichev

Simulation of development of individual
heavy industry sectors59

H. Nozari, M. Fallah, H. Kazemipoor, S.E. Najafi

Big data analysis of IoT-based supply chain
management considering FMCG industries78

ABOUT THE JOURNAL

Business Informatics is a peer reviewed interdisciplinary academic journal published since 2007 by National Research University Higher School of Economics (HSE), Moscow, Russian Federation. The journal is administered by HSE Graduate School of Business. The journal is published quarterly.

The mission of the journal is to develop business informatics as a new field within both information technologies and management. It provides dissemination of latest technical and methodological developments, promotes new competences and provides a framework for discussion in the field of application of modern IT solutions in business, management and economics.

The journal publishes papers in the areas of, but not limited to: modeling of social and economic systems, digital transformation of business, innovation management, information systems and technologies in business, data analysis and business intelligence systems, mathematical methods and algorithms of business informatics, business processes modeling and analysis, decision support in management.

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Russian Federation.

The journal is included into Web of Science Emerging Sources Citation Index (WoS ESCI), Russian Science Citation Index on the Web of Science platform (RSCI), EBSCO.

International Standard Serial Number (ISSN): 2587-814X (in English), 1998-0663 (in Russian).

Editor-in-Chief: Dr. Yevgeni A. Koucheryavy.

EDITORIAL BOARD

EDITOR-IN-CHIEF

Yevgeni A. Koucheryav
Tampere University, Finland

DEPUTY EDITOR-IN-CHIEF

Evgeny P. Zaramenskikh
National Research University Higher School of Economics,
Moscow, Russia

EDITORIAL BOARD

Habib Abdulrab
National Institute of Applied Sciences, Rouen, France

Sergey M. Avdoshin
National Research University Higher School of Economics,
Moscow, Russia

Andranik S. Akopov
National Research University Higher School of Economics,
Moscow, Russia

Fuad T. Aleskerov
National Research University Higher School of Economics,
Moscow, Russia

Alexander P. Afanasyev
Institute for Information Transmission Problems (Kharkevich
Institute), Russian Academy of Sciences, Moscow, Russia

Anton A. Afanasyev
Central Economics and Mathematics Institute, Russian Academy
of Sciences, Moscow, Russia

Eduard A. Babkin
National Research University Higher School of Economics,
Nizhny Novgorod, Russia

Sergey I. Balandin
Finnish-Russian University Cooperation in Telecommunications
(FRUCT), Helsinki, Finland

Vladimir B. Barakhnin
Federal Research Center of Information and Computational
Technologies, Novosibirsk, Russia

Alexander P. Baranov
Federal Tax Service, Moscow, Russia

Jörg Becker
University of Munster, Munster, Germany

Vladimir V. Belov
Ryazan State Radio Engineering University, Ryazan, Russia

Alexander G. Chkhartishvili
V.A. Trapeznikov Institute of Control Sciences, Russian Academy
of Sciences, Moscow, Russia

Vladimir A. Efimushkin
Central Research Institute of Communications, Moscow, Russia

Tatiana A. Gavrilo
Saint-Petersburg University, St. Petersburg, Russia

Hervé Glotin
University of Toulon, La Garde, France

Alexey O. Golosov
FORS Development Center, Moscow, Russia

Andrey Yu. Gribov
CyberPlat Company, Moscow, Russia

Alexander I. Gromoff
National Research University Higher School of Economics,
Moscow, Russia

Vladimir A. Gurvich
Rutgers, The State University of New Jersey, Rutgers, USA

Laurence Jacobs
University of Zurich, Zurich, Switzerland

Liliya A. Demidova
Ryazan State Radio Engineering University, Ryazan, Russia

Iosif E. Diskin
Russian Public Opinion Research Center, Moscow, Russia

Nikolay I. Ilyin
Federal Security Guard of the Russian Federation,
Moscow, Russia

Dmitry V. Isaev
National Research University Higher School of Economics,
Moscow, Russia

Alexander D. Ivannikov
Institute for Design Problems in Microelectronics, Russian Academy
of Sciences, Moscow, Russia

Valery A. Kalyagin
National Research University Higher School of Economics,
Nizhny Novgorod, Russia

Tatiana K. Kravchenko
National Research University Higher School of Economics,
Moscow, Russia

Sergei O. Kuznetsov
National Research University Higher School of Economics,
Moscow, Russia

Kwei-Jay Lin
Nagoya Institute of Technology, Nagoya, Japan

Mikhail I. Lugachev
Lomonosov Moscow State University, Moscow, Russia

Svetlana V. Maltseva
National Research University Higher School of Economics,
Moscow, Russia

Peter Major
UN Commission on Science and Technology for Development,
Geneva, Switzerland

Boris G. Mirkin
National Research University Higher School of Economics,
Moscow, Russia

Vadim V. Mottl
Tula State University, Tula, Russia

Dmitry M. Nazarov
Ural State University of Economics, Ekaterinburg, Russia

Dmitry E. Palchunov
Novosibirsk State University, Novosibirsk, Russia

Panagote (Panos) M. Pardalos
University of Florida, Gainesville, USA

Óscar Pastor
Polytechnic University of Valencia, Valencia, Spain

Joachim Posegga
University of Passau, Passau, Germany

Konstantin E. Samouylov
Peoples' Friendship University, Moscow, Russia

Kurt Sandkuhl
University of Rostock, Rostock, Germany

Yuriy D. Shmidt
Far Eastern Federal University, Vladivostok, Russia

Christine Strauss
University of Vienna, Vienna, Austria

Ali R. Sunyaev
Karlsruhe Institute of Technology, Karlsruhe, Germany

Victor V. Taratukhin
University of Munster, Munster, Germany

José M. Tribolet
Universidade de Lisboa, Lisbon, Portugal

Olga A. Tsukanova
Saint-Petersburg National Research University of Information
Technologies, Mechanics and Optics, St. Petersburg, Russia

Mikhail V. Ulyanov
V.A. Trapeznikov Institute of Control Sciences, Russian Academy
of Sciences, Moscow, Russia

Raissa K. Uskenbayeva
International Information Technology University, Almaty, Kazakhstan

Markus Westner
Regensburg University of Applied Sciences, Regensburg, Germany

Использование вневыборочных остатков Кокса–Снелл при прогнозировании наступления событий

Е.В. Румянцева

E-mail: evrumyantseva@hse.ru

К.К. Фурманов 

E-mail: kfurmanov@hse.ru

Национальный исследовательский университет «Высшая школа экономики»

Адрес: 101000, г. Москва, ул. Мясницкая, д. 20

Аннотация

В статье рассматривается задача оценивания прогнозной силы модели наступления события по вневыборочным данным. Данные о времени наступления событий, как правило, цензурированы справа: ожидаемое событие часто не успевает произойти за время наблюдения, из-за чего фиксируется только минимальное возможное значение прогнозируемой величины. В результате стандартные меры точности прогноза, такие как средняя абсолютная или средняя квадратическая ошибка, оказываются неприменимыми, а для измерения качества применяются коэффициенты ранговой корреляции: С-индекс Харрелла, коэффициенты Уно и Сомерса. Эти меры не отражают близости прогнозов к действительным значениям, а характеризуют только согласованность ранжировок – способность модели отличать наблюдения, в которых ожидаемое событие происходит относительно быстро, от тех наблюдений, в которых время ожидания относительно велико, из-за чего коэффициенты ранговой корреляции могут принимать высокие значения даже при сколь угодно большой систематической ошибке прогноза. Кроме того, сведение качества прогноза к корреляции или даже близости прогнозируемого и действительного значений малоудовлетворительно: время наступления редко удается оценить с определенностью, и при прогнозировании интерес представляет не только точечная оценка момента наступления, но и оценка закона распределения объясняемой величины целиком. В настоящей статье при выборе прогнозной модели предлагается дополнять сравнение коэффициентов ранговой корреляции анализом остатков Кокса–Снелл, рассчитанных для вневыборочных данных (контрольных или валидационных). Для визуального анализа предлагается применять график оценки интегрального риска остатков, а в качестве численной характеристики согласованности модели с вневыборочными данными – расстояние Колмогорова между наблюдаемым распределением остатков и экспоненциальным распределением с единичным средним, которое соответствует идеально специфицированной модели. Предлагаемый подход иллюстрируется примером выбора прогнозной модели для времени досрочного погашения договоров ипотечного кредитования.

Ключевые слова: прогнозирование; анализ наступления событий; остатки Кокса—Снелл; цензурирование.

Цитирование: Румянцева Е.В., Фурманов К.К. Использование вневыборочных остатков Кокса—Снелл при прогнозировании наступления событий // Бизнес-информатика. 2021. Т. 15. № 1. С. 7–18.
DOI: 10.17323/2587-814X.2021.1.7.18

Введение

В ряде задач статистики и анализа данных требуется оценить время ожидания определенного события. В финансовых приложениях это может быть время наступления дефолта по договору кредитования, в медицинских – время смерти или выздоровления пациента, в социальных и демографических – возраст матери на момент рождения ребенка или вступления в брак. Область прикладной статистики, занимающаяся такими задачами, называется анализом наступления событий (event-history analysis) и имеет существенные особенности, которые препятствуют применению многих устоявшихся методов регрессионного анализа. Одна из этих особенностей – цензурирование (censoring).

Продолжительность всякого исследования конечна, и за это время ожидаемое событие может не произойти. Более того, для некоторых объектов оно вообще никогда не наступит: часть заемщиков завершит выплаты по кредиту, не доходя до дефолта, часть женщин не выйдет замуж. В результате об этих объектах известно лишь то, что время ожидания для них превышает некоторое значение – длительность от начала ожидания события до конца периода наблюдения. Таким образом, наблюдения за этими объектами оказываются цензурированными справа.

Методы оценивания статистических моделей по цензурированным данным известны достаточно хорошо, классический труд в этой области – книга [1]. Измерение точности прогноза наступления событий – менее разработанная область. По-видимому, этот пробел объясняется тем, что модели наступления разрабатывались обычно не для прогноза, а с академическими целями – для определения эффективности лечения или мер социальной политики, выявления детерминант продолжительности безработицы и т.п.

За последние десять лет интерес к прогнозированию существенно возрос. С одной стороны, модели наступления событий стали чаще находить чисто прагматическое применение: примерами

могут служить анализ финансовых рисков [2, 3] и продолжительность сбора средств при краудфандинге [4]. С другой стороны, распространение машинного обучения и, в частности, процедур кросс-валидации обусловило интерес к оцениванию качества моделей по их способности делать точный вневыборочный прогноз [5, 6]. При этом стандартные метрики, такие как среднеквадратическая ошибка прогноза или средняя абсолютная процентная ошибка, оказываются неприменимыми из-за цензурирования.

В настоящей статье предлагается подход к выбору прогнозной модели, опирающийся на совместное использование коэффициентов согласованности фактического и прогнозируемого времени наступления и остатков Кокса—Снелл, рассчитываемых по контрольной выборке. В качестве примера рассматривается задача построения прогнозной модели для времени досрочного погашения ипотечного кредита.

В первом разделе описываются основные понятия, которые требуют определения (либо упоминания) из-за обособленности предметной области и специфической терминологии. Во втором разделе приводится обзор характеристик точности прогноза времени наступления события. Третий раздел посвящен остаткам Кокса—Снелл, которые авторы предлагают использовать для измерения качества прогноза. Пример использования остатков для выбора наилучшей прогнозной модели приведен в четвертом разделе, за которым следует заключение.

1. Вероятностная модель наступления события

Время ожидания события моделируется неотрицательной случайной величиной, которая по усмотрению исследователя может быть как непрерывной, так и дискретной. В настоящей статье мы рассматриваем случай непрерывного времени. Распределение времени ожидания характеризуется следующими функциями, играющими ведущую роль в анализе наступления событий.

Функция дожития (survival function) $S(t)$ задает вероятность того, что время ожидания события превысит произвольный срок t :

$$S(t) = P(T > t).$$

Термин соответствует медицинским и актуарным приложениям, в которых анализируемое событие — это смерть пациента или застрахованного лица.

Функция риска (hazard function) $h(t)$ отражает изменение вероятности наступления события с течением времени:

$$h(t) = \lim_{\substack{\Delta \rightarrow 0 \\ \Delta > 0}} \frac{P(t < T \leq t + \Delta | T > t)}{\Delta}.$$

Интегральная функция риска (integrated hazard function) $H(t)$ не имеет столь ясной интерпретации, но играет важную роль в настоящей статье:

$$H(t) = \int_0^t h(s) ds.$$

Терминология зависит от области приложения, поэтому те же характеристики можно встретить и под другими названиями. В частности, функция дожития еще известна как функция надежности, а функция риска — как сила смертности или функция опасности отказа.

Внимание исследователей, как правило, сосредоточено на двух аспектах моделирования.

Первый аспект — связь вероятности наступления события со временем его ожидания. Для описания этой связи особенно удобна функция риска.

Второй аспект — связь вероятности наступления события с объясняющими переменными (ковариатами). Существует множество способов увязать распределение времени ожидания с объясняющими переменными, изложение которых можно найти в книгах [1, 7]. Здесь мы ограничимся рассмотрением четырех регрессионных моделей.

Логнормальная и обобщенная гамма-регрессия относятся к моделям ускоренного времени (accelerated failure-time models), то есть представляемы в линейном виде:

$$\ln T = x'\beta + \varepsilon.$$

Здесь x' — вектор-строка объясняющих переменных, β — вектор-столбец коэффициентов при них, а ε — случайная ошибка. В вектор x' включается единичный элемент, соответствующий свободно-му члену регрессии, так что $x'\beta = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k$.

Логнормальная и гамма-регрессия отличаются только распределением случайной ошибки.

Регрессия Гомперца и модель Кокса относятся к моделям пропорциональных рисков (proportional hazards), то есть предполагают, что функция риска наступления события пропорциональна объясняющим переменным:

$$h(t; x', \beta) = h_0(t) \exp(x'\beta).$$

Здесь $h_0(t)$ — так называемая опорная функция риска (baseline hazard). В модели Гомперца предполагается экспоненциальное убывание или возрастание риска с течением времени: $h_0(t) = e^{\gamma t}$, где γ — параметр модели, оцениваемый наряду с коэффициентами β . Модель Кокса не накладывает ограничений на опорную функцию риска, эта функция оценивается непараметрически. Таким образом, регрессия Кокса строго постулирует характер связи риска наступления события с объясняющими переменными, как и прочие рассматриваемые модели, но в отличие от них не ограничивает зависимость риска от времени.

Методы оценивания моделей наступления событий реализованы в статистических пакетах, их описание содержится в книгах [1, 7]. Для дальнейшего изложения важно лишь то, что на основании каждой из этих моделей можно рассчитать оценки функций дожития и интегрального риска для любых значений объясняющих переменных x' и таким образом получить прогноз наступления события в заданных условиях.

2. Меры точности прогноза: Обзор

Меры средней ошибки прогноза. Под этим названием можно объединить самые распространенные характеристики точности для моделей с количественным объясняемым признаком: среднюю абсолютную ошибку (MAE), среднюю квадратическую ошибку (RMSE), среднюю абсолютную процентную ошибку (MAPE) и другие. Есть примеры их применения и в анализе наступления событий [2, 5, 8], однако эти примеры — скорее исключение. В большинстве случаев данные подвержены цензурированию, из-за которого точная величина расхождения между временем наступления события и его прогнозом неизвестна и усреднению не поддается. Проблему можно легко решить, если предположить, что распределение расхождений относится к некоторому параметрическому семейству, тогда

параметры этого распределения и математическое ожидание расхождений можно оценить методом максимального правдоподобия. Однако примеры применения такого подхода авторам статьи неизвестны (возможно, это объясняется естественным желанием исследователей избежать дополнительных предпосылок).

В работах [2, 8] рассматривается подход, при котором средняя абсолютная ошибка рассчитывается только для нецензурированных наблюдений, содержащих точные сведения о времени наступления события. Этот подход имеет существенный недостаток. Цензурирование зависит от времени наступления: чем дольше ожидание события, тем скорее наблюдение окажется цензурированным, потому что событие не наступит в течение периода наблюдения. Тогда точное значение объясняемой переменной будет известно для тех случаев, когда оно мало, а поскольку средняя абсолютная ошибка рассчитывается именно по этим наблюдениям, то предпочтение будет отдаваться моделям, занижающим время наступления события. Формально говоря, здесь цензурирование заменяется на усечение — не менее существенный вид неполноты данных [1].

Коэффициенты согласованности ранжировок. Пожалуй, самая распространенная мера качества прогноза при анализе наступления событий — коэффициент конкордации Харрелла [9], он же *S*-индекс, который можно определить следующим образом. Пусть случайные величины T_1 и T_2 отражают фактическое время наступления события в двух случайно отобранных независимых наблюдениях, а $\hat{T}_1$ и $\hat{T}_2$ представляют собой соответствующие прогнозы. Тогда коэффициент Харрелла C определяется выражением:

$$C = P(\hat{T}_1 < \hat{T}_2 | T_1 < T_2). \quad (1)$$

Одно из достоинств коэффициента — его ясная интерпретация. Пусть рассматриваются два случая с различным временем наступления события. Тогда модель с вероятностью C прогнозирует более быстрое наступления события именно в том случае, когда оно и в самом деле наступило быстрее. Наибольшее значение C равно единице и наблюдается при полном соответствии фактических и прогнозируемых ранжировок по времени (чем быстрее событие наступает согласно модели, тем меньше и действительное время наступления), наименьшее значение равно нулю и

означает полное расхождение ранжировок (чем меньше предсказание модели, тем больше истинное время).

Существуют различные оценки коэффициента конкордации, учитывающие цензурирование. Помимо оригинальной статистики Харрелла [9], в последнее время стал широко применяться коэффициент Уно [10], оценивающий тот же параметр. Кроме коэффициента конкордации (1) также используется коэффициент корреляции Сомерса [11, 12].

Общее слабое место перечисленных характеристик — то, что они отражают лишь согласованность ранжировок или, иначе говоря, способность модели выявлять те случаи, в которых событие наступает относительно быстрее или медленнее. Это не есть точность прогноза в обычном понимании. Представим, что предсказанное моделью время в каждом наблюдении оказывается ровно в десять раз больше истинного. Вряд ли можно назвать десятикратно завышенный прогноз точным, однако коэффициент конкордации (и любой коэффициент корреляции рангов) будет равен единице, потому что ранжировки прогнозов и истинных значений совпадут.

Иногда именно способность модели выделять случаи относительно короткого или длительного ожидания оказывается в фокусе исследования, например, если речь идет об определении признаков скорого выздоровления или смерти [13]. Однако во многих задачах представляет интерес именно абсолютное время наступления события. Такие задачи возникают и в традиционной области применения — медицине [14], но еще более характерны для финансовых приложений, где от дефолтов и досрочных погашений зависит поток денежных средств [2, 3, 15].

Еще одна группа характеристик прогнозной силы — **показатели классифицирующей способности**, измеряющие точность бинарного прогноза (событие наступило или не наступило к определенному сроку). Это направление стало активно развиваться в последнее десятилетие [16, 17], оно заслуживает упоминания, но подробно рассматриваться в настоящей статье не будет, так как представляет собой содержательно иной подход к задаче прогнозирования. Впрочем, коэффициент Харрелла также может рассматриваться как характеристика классифицирующей способности [18].

3. Остатки Кокса–Снелл и их применение при изучении прогностной способности

Пусть имеется выборка $(T_1, x'_1), \dots, (T_n, x'_n)$, где T_i – время наступления события, а x'_i – вектор объясняющих переменных в наблюдении i . Пусть $\hat{H}(t; x')$ – оцененная функция интегрального риска случайных величин T_i (ее можно получить из какой-либо регрессионной модели). Остатком Кокса–Снелл [19] в наблюдении i называется величина $r_i^{CS} = \hat{H}(T_i; x'_i)$.

Если оценка $\hat{H}(t; x')$ совпадает с истинной интегральной функцией риска $H(t; x')$, то остатки Кокса–Снелл распределены по экспоненциальному закону со средним 1. В этом случае интегральная функция риска для остатков имеет вид $H^{CS}(t) = t$.

В регрессионной диагностике применяется визуальный тест, который опирается на остатки Кокса–Снелл и состоит из следующих шагов.

1. Оценивается регрессионная модель и рассчитываются остатки Кокса–Снелл для каждого наблюдения.

2. Рассчитывается оценка интегральной функции риска для остатков $\hat{H}^{CS}(t)$. При цензурировании данных T_i о времени наступления событий остатки тоже оказываются цензурированными, что требует соответствующего метода оценивания. Мы используем метод Нельсона–Аалена [20, 21].

3. Строится график зависимости оцененного интегрального риска $\hat{H}^{CS}(r_i^{CS})$ от остатков r_i^{CS} , в дальнейшем будем называть его графиком остатков Кокса–Снелл. В случае верно специфицированной модели наступления событий график выстраивается вдоль линии $H(t) = t$ (рисунк 1а). Пример графика для неверно подобранной модели представлен на рисунке 1б.

Этот тест известен и отражен в учебной литературе, соответствующие графики приводятся исследователями в статьях в подтверждение корректности предлагаемых моделей [7, 22, 23]. Мы не встречали примеров использования остатков Кокса–Снелл при изучении прогностной способности моделей, возможные причины этого пробела указаны ниже в заключении. Далее в настоящей статье остатки Кокса–Снелл, рассчитанные по контрольной выборке, называются вневыборочными остатками в том смысле, что они характеризуют поведение прогностной модели вне той выборки, по которой она оценивалась.

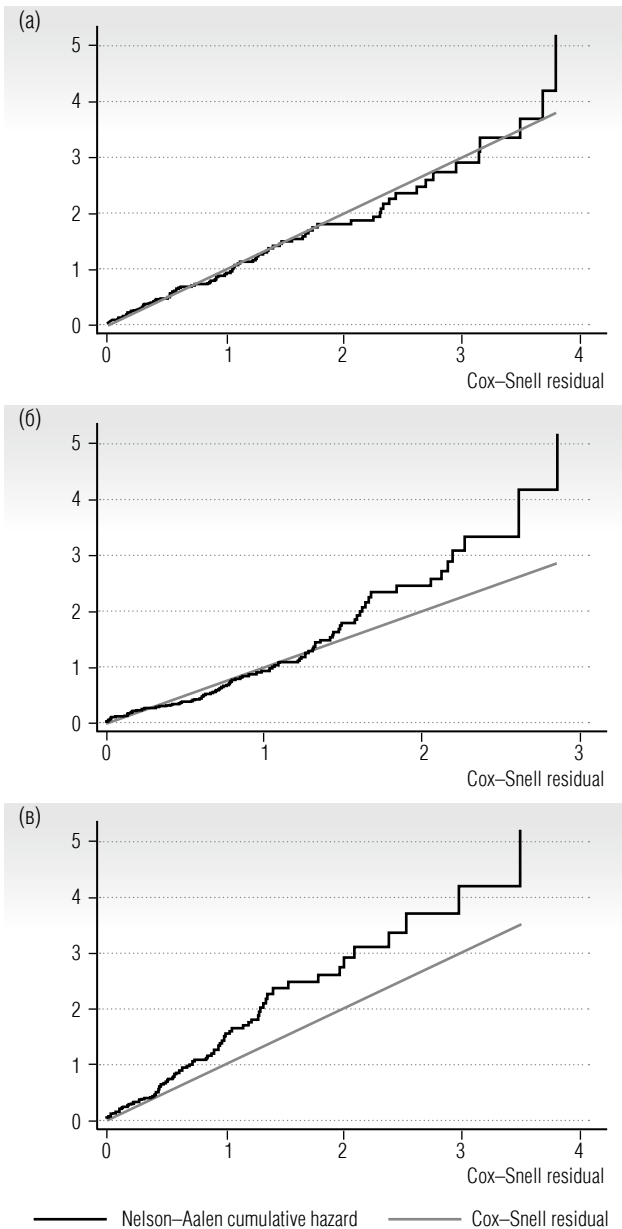


Рис. 1. Примеры графиков остатков Кокса–Снелл при
(а) верной спецификации оцениваемой модели;
(б) ошибочно выбранном распределении;
(в) систематической ошибке прогноза

Расчет вневыборочных остатков позволяет выявить систематическую ошибку прогноза. Представим себе случай, когда модельная функция риска оказывается в c раз выше истинной: $h(t; x') = c h(t; x')$. Тогда остатки Кокса–Снелл будут завышать истинную интегральную функцию риска также в c раз: $r_i^{CS} = \hat{H}(T_i; x'_i) = c H(T_i; x'_i)$. В результате график остатков проходит выше или ниже (в зависимости от c), чем линия $H(t) = t$ (рисунк 1в). Такая ситуация практически невозможна при диагностике регрес-

сии по обучающей выборке и маловероятна даже при изучении контрольной выборки, если наблюдения для контроля отбирались случайно. Такое смещение прогноза скорее можно наблюдать при внешней валидации, когда прогнозная сила модели проверяется на новых данных.

Визуальный анализ предпочтителен при выборе модели «вручную», для задач машинного обучения удобнее иметь числовую характеристику. Ниже в качестве такой характеристики используется расстояние Колмогорова между оценкой функции дожития для остатков Кокса–Снелл $S^{CS}(t) = \exp(-\hat{H}^{CS}(t))$ и соответствующей функцией для экспоненциального распределения $S(t) = e^{-t}$:

$$KD = \sup_{t \in [0; t_{\max}]} |\hat{S}^{CS}(t) - e^{-t}|. \quad (2)$$

Здесь $t_{\max}$ — наибольшее время ожидания события в выборке (может соответствовать и цензурированному наблюдению). Мы рассматриваем только множество $[0; t_{\max}]$, так как оценка Нельсона–Аалена не позволяет восстановить правый хвост распределения остатков Кокса–Снелл. Соответствующая функция дожития в точке $t_{\max}$ все еще отлична от нуля, и нет данных, позволяющих оценить ее для значений аргумента $t > t_{\max}$.

В следующем параграфе приводится пример применения предлагаемой метрики при построении прогнозной модели.

4. Пример: Моделирование досрочного погашения кредита

Пример опирается на данные крупной компании, занимающейся ипотечным кредитованием. Данные содержат сведения о более чем 280 тысячах договоров, заключенных в период 2001–2013 гг. Объясняемая переменная — время от заключения договора до досрочного погашения кредита. Наблюдения за объясняемой переменной цензурированы справа по следующим причинам:

- ◆ Прекращение периода наблюдения: пример опирается на срез данных на 1 января 2014 года, поэтому точная дата погашения действовавших на тот момент договоров не известна.

- ◆ Завершение срока договора: если досрочного погашения в действительности не произошло, то наблюдение за временем погашения считается цензурированным, как если бы было возможно досрочное погашение после истечения срока. Это своего рода математическая уловка, удобная при

построении модели и применяемая в случаях, когда с изучаемыми объектами могут случиться взаимоисключающие события разного рода (в данном случае — досрочное погашение и истечение срока). Удобно считать, что происходят оба события, но наблюдается только первое из них.

- ◆ Наступление дефолта по договору. Это еще одно событие, исключающее возможность досрочного погашения.

По этим данным оцениваются модели наступления события, отличающиеся (а) предпосылками о законе распределения объясняемой величины и его связи с объясняющими переменными (логнормальная и гамма-регрессия, модели Кокса и Гомперца) и (б) набором объясняющих переменных («короткая» и «длинная» модели). «Короткая» и «длинная» модели учитывают характеристики самого кредита, основного заемщика, а также предмета ипотеки, находящегося под залогом. «Короткая» модель учитывает кредитную ставку, срок кредита, коэффициент «платеж/доход», возраст основного заемщика, тип занятости основного заемщика и количество комнат в предмете ипотеки. «Длинная» модель включает, помимо вышеперечисленных, следующие признаки: пол, семейное положение и уровень образования основного заемщика, число созаемщиков, местонахождение предмета ипотеки (положение региона в рейтинге социально-экономического положения субъектов РФ, используемом в компании, согласно которому все регионы поделены на три группы: с низким, умеренным и высоким уровнем развития), тип жилья под залогом (дом, таунхаус или квартира), отношение жилой и общей площади в предмете ипотеки, отношение совокупных денежных выплат за плановый срок жизни кредита к стоимости жилья и коэффициент «кредит/зalog».

Все наблюдения случайным образом разбиты на обучающую и контрольную выборки в соотношении 60:40 соответственно.

Графики остатков Кокса–Снелл для восьми оцененных моделей для контрольной части данных приведены в Приложении. Рассчитанные по формуле (2) расстояния между наблюдаемым и теоретическим распределениями остатков, а также значения коэффициента Харрелла приведены в *таблице 1*.

Из таблицы видно, что коэффициент Харрелла практически не зависит от выбора распределения времени досрочного погашения, но меняется при добавлении в модель объясняющих переменных: «длинная» модель оказывается стабильно лучше, обе-

Таблица 1.

**Характеристики точности вневыборочного прогноза
моделей наступления досрочного погашения кредита**

	«Короткая» модель		«Длинная» модель	
	Коэффициент Харрелла	Расстояние Колмогорова	Коэффициент Харрелла	Расстояние Колмогорова
Логнормальная	0,593	0,078	0,612	0,066
Гамма	0,593	0,015	0,610	0,009
Гомперца	0,592	0,059	0,608	0,063
Кокса	0,593	0,007	0,609	0,014

спечивая вероятность согласия прогнозов и ранжировок примерно на 0,02 больше, чем у «короткой». Конечно, исследователь может считать это расхождение несущественным, но оно устойчиво и проявляется и при прочих вариантах разбиения данных на обучающую и контрольную части. Напротив, расстояние Колмогорова между идеальным и наблюдаемым распределениями остатков Кокса–Снелл заметно меняется в зависимости от распределения: обобщенная гамма-регрессия и регрессия Кокса обеспечивают меньшее расхождение наблюдаемого и предполагаемого распределений, по сравнению с логнормальной регрессией и регрессией Гомперца, причем как в случае «короткой», так и в случае «длинной» модели.

Расхождения в прогнозах, полученных при разных предположениях о виде распределения, представлены на *рисунке 2*. Здесь представлены оценки функций дожития для кредита с часто встречающимся набором характеристик:

- ◆ пол основного заемщика – мужчина;
- ◆ семейный статус основного заемщика – женат;
- ◆ образование основного заемщика – высшее образование или ученая степень;
- ◆ тип занятости основного заемщика – наемный работник;
- ◆ количество созаемщиков, включая основного заемщика – 2;
- ◆ возраст основного заемщика – менее 35 лет;
- ◆ коэффициент «платеж/доход» – от 20% до 35%;
- ◆ тип предмета ипотеки – квартира;
- ◆ местонахождение предмета ипотеки – умеренно развитый регион;
- ◆ количество комнат в предмете ипотеки – 2;
- ◆ годовая кредитная ставка – менее 11,5% (низкорисковая кредитная ставка);
- ◆ соотношение жилой и общей площади в предмете ипотеки – между 50% и 70%;

- ◆ коэффициент «кредит/залог» – от 50% до 70%;
- ◆ срок кредита – свыше 180 месяцев (долгосрочный кредит);
- ◆ отношение совокупных денежных выплат за плановый срок жизни кредита к стоимости жилья – от 1 до 1,82.

На *рисунке 2а* отражена вероятность непогашения в течение ближайших лет для только что выданного кредита, на *рисунке 2б* – для кредита, выданного пять лет назад (условная функция дожития при условии $\{T > 5\}$). По горизонтальной оси отложено число дней с момента выдачи кредита, по вертикальной – вероятность дожития. На обоих графиках оценки, полученные из модели Кокса (линия «S_sox») и гамма-регрессии («S_gamma»), практически совпадают. Линия, соответствующая регрессии Гомперца («S_gomp»), отклоняется от прочих с самого начала, предсказывая больший риск погашения. Логнормальная регрессия (линия «S_ln»), наоборот, предсказывает наименьший риск погашения, причем существенное расхождение с моделями Кокса и гамма начинается не сразу, оно заметно только для «пожилого» кредита. Графики вневыборочных остатков, приведенные в Приложении, подтверждают, что гамма-регрессия верно описывает распределение времени погашения, модель Кокса ей почти не уступает, а вот остатки регрессии Гомперца, как и остатки логнормальной модели, отклоняются от идеального экспоненциального распределения, причем в последнем случае отклонение заметно именно на правом хвосте распределения («пожилые» кредиты).

Таким образом, если при выборе прогнозной модели опираться на коэффициент Харрелла, то будет выбрана логнормальная регрессия, недооценивающая риск досрочного погашения. Принимая в расчет вневыборочные остатки Кокса–Снелл, аналитик скорее сделает выбор в пользу обобщенной

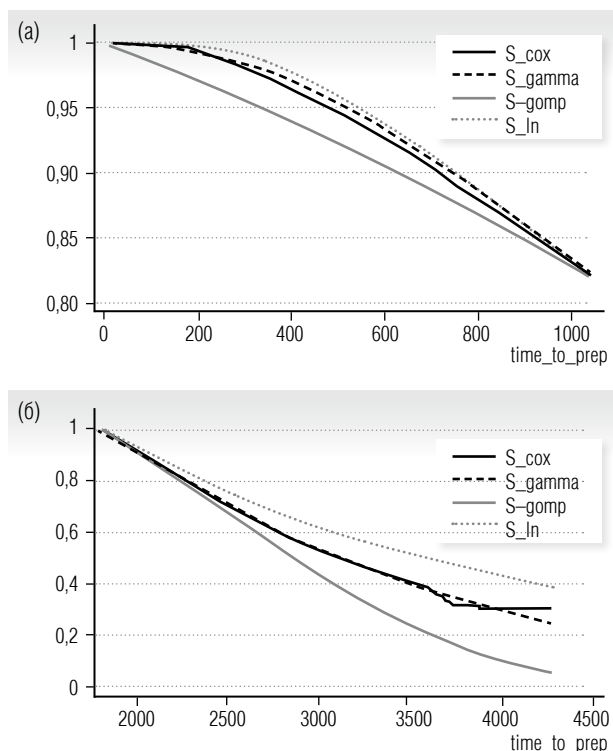


Рис. 2. Оценки функции дожития для
(а) только что выданного кредита;
(б) пятилетнего кредита

гамма-регрессии или модели Кокса, которые дают близкие прогнозы, хорошо согласующиеся с распределением времени погашения в контрольной выборке. Как видно из рисунка 2, оцененная вероятность погашения заметно зависит от выбора прогнозной модели. Это особенно заметно на дальнем горизонте прогнозирования (линии на рисунке 2б расходятся более чем на 20 процентных пунктов), но уже при расчете на один–два года вперед расхождение прогнозов превышает 5 процентных пунктов.

Заключение

Остатки Кокса–Снелл и связанный с ними визуальный тест хорошо известны и освещены не только в исследовательской, но и в учебной литературе, однако широкого применения при анализе прогнозной силы на контрольной выборке не нашли. Это важный момент, и он вызывает естественное сомнение: действительно ли от них есть польза? Приведенный пример был призван показать, что польза есть, сейчас же время сделать важную оговорку.

Сами по себе остатки Кокса–Снелл не помогают отличить точный прогноз от неточного. Они лишь показывают, насколько прогнозируемое распределение близко к тому, что наблюдается в выборке

(обучающей — при обычном использовании, контрольной — если использовать так, как предлагается в настоящей статье). Модель вообще без объясняющих переменных, которая дает одинаковый прогноз во всех наблюдениях, вполне может оказаться лучше по остаткам, чем регрессия, учитывающая важные и действительно улучшающие прогноз факторы, если в первой удачнее выбрать семейство распределений (что, кстати, без объясняющих переменных сделать проще).

Коэффициенты ранговой корреляции/конкордации, наоборот, позволяют отобрать модель с удачным набором объясняющих переменных, но никак не характеризуют способность предсказать в целом распределение времени наступления события. В классических задачах регрессионного анализа этого часто и не требуется, аналитика интересуется в первую очередь точечный прогноз — оценка математического ожидания или медианы. Однако время наступления событий почти никогда не удается точно охарактеризовать одним числом, и сведение прогнозирования к оценке единственного параметра непродуктивно. Здесь не меньший интерес, чем математическое ожидание, представляют квантили различных порядков, то есть, по сути, распределение целиком. Выбору распределения часто не уделяется внимание: исследователи предпочитают опираться на модель Кокса, не требующую параметризации распределения. Приведенный в настоящей статье пример показывает, что даже на выборке весьма большого объема относительно простая параметрическая регрессия может превзойти модель Кокса, склонную к переобучению. Кроме того, модель Кокса ограничивает функциональную форму связи риска наступления события с объясняющими переменными, параметрические методы позволяют ослабить это ограничение.

Комбинация коэффициента конкордации и остатков Кокса–Снелл позволяет оценить качество прогнозной модели и в аспекте выбора объясняющих переменных, и в аспекте выбора семейства распределений. Конечно, удобнее было бы иметь единственную характеристику, однозначно указывающую на модель с лучшей прогнозной силой, но о такой характеристике пока, по-видимому, говорить не приходится. ■

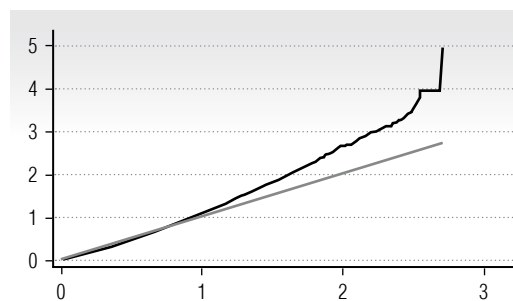
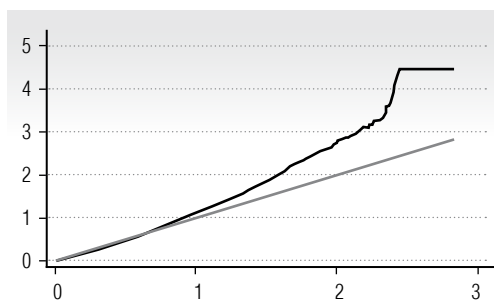
Приложение

Графики вневыборочных остатков Кокса–Снелл для моделей досрочного погашения ипотечного кредита:

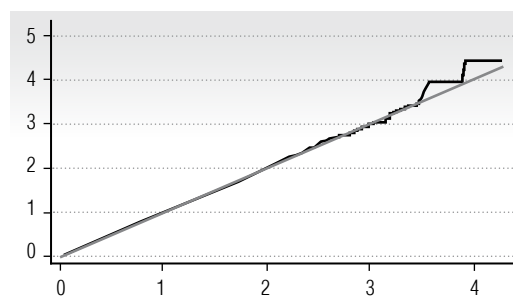
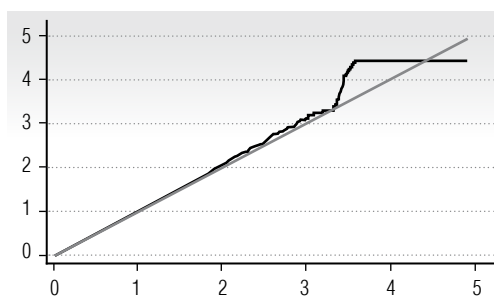
«Короткая» модель

«Длинная» модель

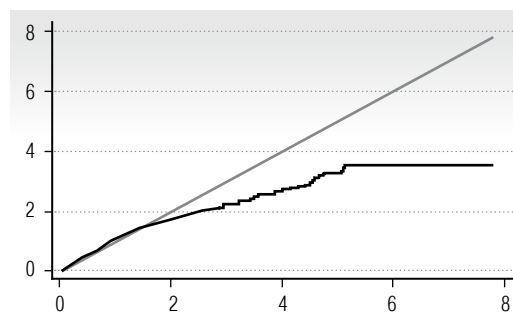
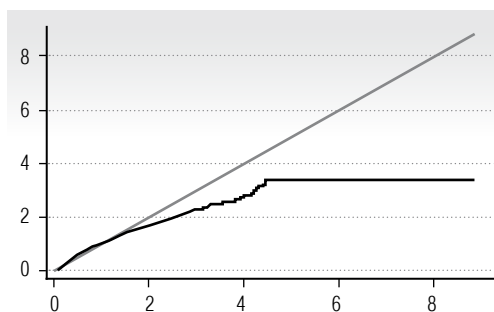
Логнормальная регрессия



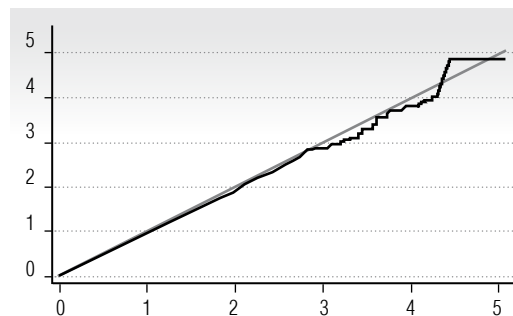
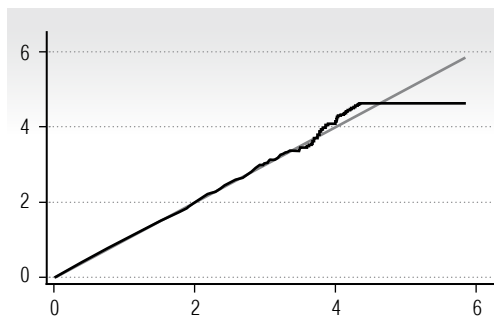
Гамма-регрессия



Регрессия Гомперца



Регрессия Кокса



— Nelson-Aalen cumulative hazard — Cox-Snell residual

Литература

1. Klein J.P., Moeschberger M.L. Survival analysis: Techniques for censored and truncated data. Second edition. Springer. 2005.
2. Zhang J., Thomas L.C. Comparison of linear regression and survival analysis using single and mixture distribution approaches in modelling LGD // *International Journal of Forecasting*. 2012. Vol. 28. No 1. P. 204–215. DOI: 10.1016/j.ijforecast.2010.06.002.
3. Замисный П., Козлов А. Новый подход к использованию данных БКИ при оценке вероятности дефолта и досрочного погашения // *Банковское кредитование*. 2018. №4. С. 4–11.
4. Rakesh V., Lee W.-C., Reddy C.K. Probabilistic group recommendation model for crowdfunding domains // *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM 2016)*. San Francisco, California, USA. 22–25 February 2016. P. 257–266. DOI: 10.1145/2835776.2835793.
5. Ameri S., Fard M.J., Chinnam R.B., Reddy C.K. Survival analysis based framework for early prediction of student dropouts // *Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM 2016)*. Indianapolis, Indiana, USA. 24–28 October 2016. P. 903–912. DOI: 10.1145/2983323.2983351.
6. Wang P., Li Y., Reddy C.K. Machine learning for survival analysis: A survey // *ACM Computing Surveys*. February 2019. Article no 110. DOI: 10.1145/3214306.
7. Cleves M.A., Gould W.W., Gutierrez R.G., Marchenko Y.U. An introduction to survival analysis using Stata. Third Edition. College Station, Texas: Stata Press, 2010.
8. Dirick L., Claeskens G., Baesens B. Time to default in credit scoring using survival analysis: a benchmark study // *Journal of Operational Research Society*. 2017. Vol. 68. No 6. P. 652–665. DOI: 10.1057/s41274-016-0128-9.
9. Evaluating the yield of medical tests / F.E. Harrell Jr. [et al.] // *Journal of the American Medical Association*. 1982. Vol. 247. No 18. P. 2543–2546. DOI: 10.1001/jama.1982.0332043004703.
10. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data / H. Uno [et al.] // *Statistics in Medicine*. 2011. Vol. 30. No 10. P. 1105–1117. DOI: 10.1002/sim.4154.
11. Newson R.B. Comparing the predictive powers of survival models using Harrell's C or Somers' D // *The Stata Journal*. 2010. Vol. 10. No 3. P. 339–358.
12. Somers R.H. A new asymmetric measure of association for ordinal variables // *American Sociological Review*. 1962. Vol. 27. No 6. P. 799–811.
13. Survival marker genes of colorectal cancer derived from consistent transcriptomic profiling / J. Martinez-Romero [et al.] // *BMC Genomics*. 2018. No 19. Article no 857. DOI: 10.1186/s12864-018-5193-9.
14. Калинин М.Н., Хасанова Д.Р., Ибатуллин М.М. Возможные сроки начала антикоагулянтной терапии у больных с ишемическим инсультом и фибрилляцией предсердий: последующий анализ индекса геморрагической трансформации (Hemorrhagic Transformation Index) // *Неврология, нейропсихиатрия, психосоматика*. 2019. Т. 11. № 2. С. 12–21. DOI: 10.14412/2074-2711-2019-2-12-21.
15. Румянцева Е.В., Фурманов К.К. Моделирование времени жизни ипотечного кредита // *Прикладная эконометрика*. 2016. Т. 41. № 1. С. 123–143.
16. Hung H., Chiang C.T. Optimal composite markers for time-dependent receiver operating characteristic curves with censored survival data // *Scandinavian Journal of Statistics*. 2010. Vol. 37. No 4. P. 664–679.
17. Kamarudin A.N., Cox T., Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications // *BMC Medical Research Methodology*. 2017. No 17. Article no 53. DOI: 10.1186/s12874-017-0332-6.
18. Heagerty P.J., Zheng Y. Survival model predictive accuracy and ROC curves // *Biometrics*. 2005. Vol. 61. No 1. P. 92–105. DOI: <https://doi.org/10.1111/j.0006-341X.2005.030814.x>.
19. Cox D.R., Snell E.J. A general definition of residuals // *Journal of the Royal Statistical Society. Series B (Methodological)*. 1968. Vol. 30. No 2. P. 248–275.
20. Nelson W. Theory and applications of hazard plotting for censored failure data // *Technometrics*. 1972. Vol. 14. No 4. P. 945–966. DOI: 10.1080/00401706.1972.10488991.
21. Aalen O. Nonparametric inference for a family of counting processes // *Annals of Statistics*. 1978. Vol. 6. No 4. P. 701–726.
22. Арженовский С. Социально-экономические детерминанты курения в России // *Квантиль*. 2006. № 1. С. 81–100.
23. Рапаков Г.Г., Горбунов В.А. Исследование методов анализа времени до события при обработке демографических данных // *Вестник ВГУ. Серия: Системный анализ и информационные технологии*. 2015. № 4. С. 110–120.

Об авторах

Румянцева Екатерина Владимировна

кандидат физико-математических наук;

старший преподаватель департамента прикладной экономики, факультет экономических наук, Национальный исследовательский университет «Высшая школа экономики», 101000, г. Москва, ул. Мясницкая, д. 20;

E-mail: evrumyantseva@hse.ru

Фурманов Кирилл Константинович

кандидат экономических наук;

доцент департамента прикладной экономики, факультет экономических наук, Национальный исследовательский университет «Высшая школа экономики», 101000, г. Москва, ул. Мясницкая, д. 20;

E-mail: kfurmanov@hse.ru

ORCID: 0000-0002-3433-9497

Using out-of-sample Cox–Snell residuals in time-to-event forecasting

Ekaterina V. Rumyantseva

E-mail: evrumyantseva@hse.ru

Kirill K. Furmanov

E-mail: kfurmanov@hse.ru

National Research University Higher School of Economics

Address: 20, Myasnitskaya Street, Moscow 101000, Russia

Abstract

The problem of assessing out-of-sample forecasting performance of event-history models is considered. Time-to-event data are usually incomplete because the event of interest can happen outside the period of observation or not happen at all. In this case, only the shortest possible time is observed and the data are right censored. Traditional accuracy measures like mean absolute or mean squared error cannot be applied directly to censored data, because forecasting errors also remain unobserved. Instead of mean error measures, researchers use rank correlation coefficients: concordance indices by Harrell and Uno and Somers' Delta. These measures characterize not the distance between the actual and predicted values but the agreement between orderings of predicted and observed times-to-event. Hence, they take almost "ideal" values even in presence of substantial forecasting bias. Another drawback of using correlation measures when selecting a forecasting model is undesirable reduction of a forecast to a point estimate of predicted value. It is rarely possible to predict the timing of an event precisely, and it is reasonable to consider the forecast not as a point estimate but as an estimate of the whole distribution of the variable of interest. The article proposes computing Cox–Snell residuals for the test or validation dataset as a complement to rank correlation coefficients in model selection. Cox–Snell residuals for the correctly specified model are known to have unit exponential distribution, and that allows comparison of the observed out-of-sample performance of a forecasting model to the ideal case. The comparison can be done by plotting the estimate of integrated hazard function of residuals or by calculating the Kolmogorov distance between the observed and the ideal distribution of residuals. The proposed approach is illustrated with an example of selecting a forecasting model for the timing of mortgage termination.

Key words: forecasting; event-history analysis; Cox–Snell residuals; censoring.**Citation:** Rumyantseva E.V., Furmanov K.K. (2021) Using out-of-sample Cox–Snell residuals in time-to-event forecasting. *Business Informatics*, vol. 15, no 1, pp. 7–18. DOI: 10.17323/2587-814X.2021.1.7.18**References**

1. Klein J.P., Moeschberger M.L. (2005) *Survival analysis: Techniques for censored and truncated data*. Second edition. Springer.
2. Zhang J., Thomas L.C. (2012) Comparison of linear regression and survival analysis using single and mixture distribution approaches in modelling LGD. *International Journal of Forecasting*, vol. 28, no 1, pp. 204–215. DOI: 10.1016/j.ijforecast.2010.06.002.

3. Zamisniy P., Kozlov A. (2018) Using credit history data for estimating the probabilities of default and early termination. *Bank Crediting*, no 4. pp. 4–11 (in Russian).
4. Rakesh V., Lee W.-C., Reddy C.K. (2016) Probabilistic group recommendation model for crowdfunding domains. Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM 2016), San Francisco, California, USA, 22–25 February 2016, pp. 257–266. DOI: 10.1145/2835776.2835793.
5. Ameri S., Fard M.J., Chinnam R.B., Reddy C.K. (2016) Survival analysis based framework for early prediction of student dropouts. Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM 2016), Indianapolis, Indiana, USA, 24–28 October 2016, pp. 903–912. DOI: 10.1145/2983323.2983351.
6. Wang P., Li Y., Reddy C.K. (2019) Machine learning for survival analysis: A survey. *ACM Computing Surveys*, February, article no 110. DOI: 10.1145/3214306.
7. Cleves M.A., Gould W.W., Gutierrez R.G., Marchenko Y.U. (2010) *An introduction to survival analysis using Stata*. Third Edition. College Station, Texas: Stata Press.
8. Dirick L., Claeskens G., Baesens B. (2017) Time to default in credit scoring using survival analysis: a benchmark study. *Journal of Operational Research Society*, vol. 68, no 6, pp. 652–665. DOI: 10.1057/s41274-016-0128-9.
9. Harrell F.E. Jr., Califf R.M., Pryor D.B., Lee K.L., Rosati R.A. (1982) Evaluating the yield of medical tests. *Journal of the American Medical Association*, vol. 247, no 18, pp. 2543–2546. DOI: 10.1001/jama.1982.0332043004703.
10. Uno H., Cai T., Pencina M.J., D’Agostino R.B., Wei L.J. (2011) On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, vol. 30, no 10, pp. 1105–1117. DOI: 10.1002/sim.4154.
11. Newson R.B. (2010) Comparing the predictive powers of survival models using Harrell’s C or Somers’ D. *The Stata Journal*, vol. 10, no 3, pp. 339–358.
12. Somers R.H. (1962) A new asymmetric measure of association for ordinal variables. *American Sociological Review*, vol. 27, no 6, pp. 799–811.
13. Martinez-Romero J., Bueno-Fortes S., Martín-Merino M., de Molina A.R., De Las Rivaz J. (2018) Survival marker genes of colorectal cancer derived from consistent transcriptomic profiling. *BMC Genomics*, no 19, article no 857. DOI: 10.1186/s12864-018-5193-9.
14. Kalinin M.N., Khasanova D.R., Ibatullin M.M. (2019) Possible timing for anticoagulation therapy initiation in ischemic stroke patients with atrial fibrillation: further analysis of the hemorrhagic transformation index. *Neurology, Neuropsychiatry, Psychosomatics*, vol. 11, no 2, pp. 12–21 (in Russian). DOI: 10.14412/2074-2711-2019-2-12-21.
15. Rumyantseva E.V., Furmanov K.K. (2016) Modeling mortgage survival. *Applied Econometrics*, vol. 41, no 1, pp. 123–143 (in Russian).
16. Hung H., Chiang C.T. (2010) Optimal composite markers for time-dependent receiver operating characteristic curves with censored survival data. *Scandinavian Journal of Statistics*, vol. 37, no 4, pp. 664–679.
17. Kamarudin A.N., Cox T., Kolamunnage-Dona R. (2017) Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Medical Research Methodology*, no 17, article no 53. DOI: 10.1186/s12874-017-0332-6.
18. Heagerty P.J., Zheng Y. (2005) Survival model predictive accuracy and ROC curves. *Biometrics*, vol. 61, no 1, pp. 92–105. DOI: <https://doi.org/10.1111/j.0006-341X.2005.030814.x>.
19. Cox D.R., Snell E.J. (1968) A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, no 2, pp. 248–275.
20. Nelson W. (1972) Theory and applications of hazard plotting for censored failure data. *Technometrics*, vol. 14, no 4, pp. 945–966. DOI: 10.1080/00401706.1972.10488991.
21. Aalen O. (1978) Nonparametric inference for a family of counting processes. *Annals of Statistics*, vol. 6, no 4, pp. 701–726.
22. Arzhenovsky S. (2006) Socioeconomic determinants of smoking in Russia. *Quantile*, no 1, pp. 81–100 (in Russian).
23. Rapakov G.G., Gorbunov V.A. (2015) Time-to-event analysis methods for demographics processing. *Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*, no 4. pp. 110–120 (in Russian).

About the authors

Ekaterina V. Rumyantseva

Cand. Sci. (Phys.-Math.);

Senior Lecturer, Department of Applied Economics, National Research University Higher School of Economics,

20, Myasnitskaya Street, Moscow 101000, Russia;

E-mail: evrumyantseva@hse.ru

Kirill K. Furmanov

Cand. Sci. (Econ.);

Assistant Professor, Department of Applied Economics, National Research University Higher School of Economics,

20, Myasnitskaya Street, Moscow 101000, Russia;

E-mail: kfurmanov@hse.ru

ORCID: 0000-0002-3433-9497

DOI: [10.17323/2587-814X.2021.1.19.29](https://doi.org/10.17323/2587-814X.2021.1.19.29)

Оценка эффективности информационных систем с учетом рисков

Е.А. Исаев^a 

E-mail: is@itaec.ru

Д.В. Первухин^b 

E-mail: dvperv@gmail.com

Г.О. Рытиков^{b,c} 

E-mail: GR-yandex@yandex.ru

Е.К. Филюгина^d 

E-mail: ekaterina.filyugina@mail.ru

Д.А. Айрапетян^e 

E-mail: hayrapetyandiana@gmail.com

^a Институт математических проблем биологии, Российская академия наук (филиал Института прикладной математики им. М.В. Келдыша РАН)

Адрес: 142290, Московская область, г. Пущино, ул. Профессора Виткевича, д. 1

^b Государственный университет управления

Адрес: 109542, г. Москва, Рязанский проспект, д. 99

^c Московский политехнический университет

Адрес: 107023, г. Москва, ул. Большая Семеновская, 38

^d ООО «Импакт Электроникс»

Адрес: 127055, г. Москва, ул. Новослободская, д. 14/19, стр. 8

^e ООО «Эпоха Возрождения»

Адрес: 107078, г. Москва, ул. Русаковская, д. 13, стр. 2

Аннотация

Внедрение информационных систем направлено на повышение финансовых показателей компании, создание прозрачной системы отчетности и улучшение многих других конкурентно значимых факторов. Однако получение данных преимуществ не отрицает сложность принятия решения о реализации или об отказе от реализации конкретного ИТ-проекта. Общая стоимость владения информационной системой на протяжении всего жизненного цикла, как правило, не рассматривается в сопоставлении с предполагаемыми выгодами от использования системы, в связи с неопределенностью таких выгод. Относительная определенность подходов и методов присутствует только в части затрат, как для априорной (плановой), так и апостериорной (фактической) оценки.

Возможно достаточно точное определение как капитальных, так и операционных затрат. Косвенное определение положительного влияния информационной системы на деятельность организации также представляется возможным. Однако общепринятых методов анализа предполагаемого положительного эффекта от реализации ИТ-проекта в настоящее время не существует. При этом крупные компании, в соответствии с требованиями соответствующих регуляторов и/или в связи с внутренними управленческими соображениями выстраивают систему управления рисками для определения уровня возможностей, потерь и предотвращения наступления неблагоприятных событий. В данном исследовании рассматривается возможность применения подхода к анализу эффективности внедряемой информационной системы на основе снижения рисков компании, ведущих к уменьшению экономических выгод. При этом принимаются во внимание внутренние риски информационной системы, возникающие при установке системы, ее эксплуатации и при завершении работы с системой.

Ключевые слова: оценка рисков; управление ИТ-проектами; информационные системы; внедрение информационных систем.

Цитирование: Исаев Е.А., Первухин Д.В., Рытиков Г.О., Филюгина Е.К., Айрапетян Д.А. Оценка эффективности информационных систем с учетом рисков // Бизнес-информатика. 2021. Т. 15. № 1. С. 19–29. DOI: 10.17323/2587-814X.2021.1.19.29

Введение

Современные условия ведения бизнеса предполагают режим жесткой конкуренции и все более сокращающееся время на принятие решений, что подчеркивается во многих научных работах [1]. Проведено большое количество эмпирических исследований, подтверждающих влияние таких факторов, как организационная архитектура, производственная инфраструктура и связанные с ними бизнес-процессы на способность предприятия к выживанию и эффективному функционированию [2]. При этом одним из ключевых аспектов успешного управления компанией становится использование информационных технологий и современных программных средств, а также соответствующих методов и моделей (например, сверточных нейронных сетей, значительно ускоряющих процесс обработки больших массивов данных [3]).

В продолжение авторских работ в области оценки эффективности реализации ИТ-проектов [4], стоит упомянуть современную иерархическую систему существующих классов информационных систем (ИС). С точки зрения архитектуры предприятия, при классификации ИС «сверху вниз» сначала следует упомянуть группу систем, предна-

значенных для предоставления оперативной аналитики (примерами могут служить SAP HANA, Lumira, Predictive Analytics). За данным блоком следуют ERP-системы, автоматизирующие отдельные бизнес-процессы и обеспечивающие поддержку финансово-хозяйственного управления. Оперативная трансляция потребностей рынка в конкретные производственные задания обеспечивается блоком систем MES, PLM и SCADA, где первая отвечает за производство в целом, вторая — за управление жизненным циклом продукции, а третья — за качество отдельных производственных итераций. Внедрение каждой из этих систем предполагает планирование и реализацию инвестиционного проекта [5]. Способность оценить целесообразность проекта и его релевантность становится критически важной задачей для компании [6, 7]. При этом внедрение информационных технологий, наравне с определенными ожиданиями, связано с определенными рисками¹ [8]. К числу положительных последствий внедрения ИС относятся упорядоченная организационная структура с прозрачной и единообразной отчетностью, ускорение процесса анализа деятельности компании с дальнейшим принятием стратегически важных управленческих решений, а также автоматизация многих бизнес-процессов.

¹ Здесь и далее термины «риск» и «управление рисками» используются в рамках терминологии, устоявшейся в области риск-менеджмента, и определенной, например, в таких стандартах, как ISO 31000, FERMA, Basel и др.

Отрицательный же эффект на ключевые показатели компании оказывают затраты на внедрение ИС и необходимую структурную реорганизацию [9], поддержка работоспособности системы в течение периода ее эксплуатации, а также соответствующие обновления программного продукта.

На сегодняшний день вопрос оценки эффективности внедрения ИС все еще остается открытым. Наряду с общеизвестными подходами, такими как IE (Information Economics), TEI (Total Economic Impact), REJ (Rapid Economic Justification) и BSC (Balanced Scorecard), используются стандартные для рассмотрения проектов численные показатели ROI (return of investment), NPV (net present value), IRR (internal rate of return), EVA (economic value added), ROV (real options valuations). Также необходим правильный выбор методологии управления ИТ-проектами [10]. Что же касается настоящей работы, то здесь предлагается подход к анализу эффективности внедряемой информационной системы на основе оценки рисков компании до и после реализации ИТ-проекта. Важной особенностью является то, что при этом принимаются во внимание внутренние риски, связанные с самой информационной системой и вероятностью критических сбоев при ее эксплуатации [11].

1. Влияние информационных систем на риски компании

Система риск-менеджмента крупных компаний предусматривает формирование карты рисков, которая может быть использована, в том числе, для определения положительного эффекта от внедрения информационных систем [12]. При определении влияния рисков на показатели деятельности компании принимается во внимание как сила отрицательного воздействия отдельного события на те или иные показатели, так и частота возможных случаев проявления тех или иных неблагоприятных событий.

Влияние рисков на показатели компании могут быть оценены в контексте двух ситуаций — «как есть», т.е. до реализации тех или иных мероприятий, направленных на снижение рисков (например, до внедрения информационной системы), и «как будет», т.е. после реализации соответствующих мероприятий. Таким образом, эффект от реализации мероприятия (внедрения ИС) может быть выражен в виде разности значений одних и тех же показателей в ситуациях «как есть» и «как будет».

Есть множество примеров, подтверждающих вывод о том, что информационные системы способны уменьшить риски компании. В частности, одним из наиболее заметных факторов, влияющих на ключевые показатели компании, является принятие взвешенных и обоснованных решений, в том числе, на уровне высшего руководства [13]. Именно управленческие ошибки могут привести к нарушению существующих бизнес-процессов. Возможными источниками экономически необоснованных управленческих решений может стать как некорректность данных, на основании которых принимается решение, так и человеческий фактор [14]. Внедряемая система позволяет в значительной мере усовершенствовать методы сбора и анализа данных, а также снизить (вплоть до полного исключения) их ручную обработку сотрудниками компании. Результаты анализа информации с применением алгоритмов, заложенных в ИС, обеспечивают непредвзятость и надежность при формировании отчетности, что повышает достоверность показателей, на основе которых будут приниматься решения.

На примере истории успеха компаний, справившихся с проблемой неликвидного товара (или значительно уменьшивших ее) за счет изменения структуры производства и последующего сокращения складских площадей [4] можно оценить эффективность внедренных информационных технологий, используя описываемый в данной работе метод. Риски, проявление которых возможно при совершении ошибок в области управления запасами, сокращаются, поскольку снижается необходимость в закупке существенного количества материалов. В то же время повышается потребность в оперативном взаимодействии внутри компании и с ее контрагентами. Получение данных в реальном времени становится критически важным фактором, обеспечить который позволяют информационные системы.

Другим важным аспектом, имеющим вероятность негативного исхода, является технологическая сторона производства. ИТ-решения предоставляют возможность фиксировать цепочки производственных операций внутри компании [4]. При этом происходит отслеживание и стандартизация деятельности по взаимодействию с контрагентами. Это позволяет в автоматическом режиме генерировать сигналы по управлению производственными мощностями.

Подобные примеры приведены с целью расширения функционала систем риск-менеджмента компаний. Рассмотрение проблемных вопросов, решение которых связано с внедрением ИС, мо-

жет быть оценено на основе изменения ключевых показателей в результате внедрения той или иной информационной системы. Такой подход закрывает вопрос об оценке эффективности ИТ-проекта. Однако при этом возникает новая проблема: дело в том, что риски существуют не только вне информационной системы, их источником может стать и сам реализуемый проект. Таким образом, для усовершенствования оценки проекта необходимо делать поправку на риски, возникающие, например, в случаях возникновения проблем при внедрении ИС, критических сбоев во время использования системы, а также при выводе ее из эксплуатации [15].

2. Внутренние риски информационных систем

Согласно имеющимся исследованиям эффективности ERP-систем, в 51% проектов внедрения наблюдаются непредвиденные сложности, возникающие в процессе установки ИС, в 53% случаев фактический бюджет проекта превышает первоначально утвержденный [16], 83% проектов не укладываются в срок, 42% проектов оказываются незавершенными, исходя из ожидаемых характеристик внедряемого технического решения [17, 18], 40% проектов не решают поставленную бизнес-задачу после их завершения и вводу ИС в эксплуатацию [19, 20]. Данная статистика свидетельствует о том, что менеджеры и инвесторы признают наличие существенных рисков при внедрении ИС. Обычно проекты внедрения (а также последующая поддержка системы) поручаются внешним консультантам, но в такой бизнес-модели возникают вопросы, связанные с конфликтами интересов [21] и размытием ответственности за результаты, что усугубляет потенциальный общий негативный эффект от проектных рисков [22].

Для уменьшения вероятности возникновения рисков, связанных с программным обеспечением, внедрением системы, ее использованием и выводом из эксплуатации, был разработан ряд подходов, направленных на выявление проблемных аспектов [23]. Некоторые из этих подходов учтены в информационных системах класса ALM (например, SAP Cloud ALM, Solution Manager), которые автоматизируют процесс управления жизненным циклом. В основном риски ИТ-проектов носят технический характер и связаны с формулировками требований, их изменчивостью и скоростью

выхода обновлений. При этом наибольшую значимость имеют технические аспекты, без реализации которых нет возможности произвести внедрение информационной системы [24]. Однако следует отметить группу рисков, идентификация которых влияет на бизнес-процессы компании. Это интеграция ИС с производством, часто сопровождающаяся резким падением эффективности труда [25, 26]. Таким образом, требуется разработка адекватных стратегий и мероприятий для решения данной проблемы [27].

Определение рисков ИТ-проекта является важной работой, которая была проведена ассоциацией ISACA. Результатом этой деятельности стал документ COBIT5 [28], в котором существующие на практике риски распределены по 111 категориям. Примерами могут служить риски утечки информации [29], а также риски, связанные с жизненным циклом проекта, архитектурой ИС, инфраструктурой компании, выбором вендора программного обеспечения, а также сотрудниками и их компетенциями.

Описание рисков предполагает их разделение по сферам влияния. Первым считается стратегический тип, описывающий упущенные возможности использования информационных технологий для повышения эффективности деятельности компании [30]. Второй тип рисков непосредственно связан с технической реализацией ИТ-проекта. Последним является операционный тип, который соотносится с эксплуатацией системы и предоставлением услуг технической поддержки. Эти риски, в свою очередь, делятся по степени влияния на деятельность компании в случае их реализации.

Таким образом, при оценке рисков в ситуации «как будет», т.е. после внедрения системы, следует предусмотреть внутренние риски информационной системы, непосредственно связанные с ее внедрением и использованием.

3. Оценка влияния информационных систем на риски и показатели эффективности компании

На примере процессного управления рассмотрим вопросы оценки влияния информационных систем на риски компании (и, соответственно, на ее ключевые показатели), с учетом внутренних рисков, связанных с самими информационными системами.

Будем рассматривать процесс, состоящий из нескольких последовательных этапов, по мере реали-

зации которых происходит повышение некоторого выбранного показателя эффективности. Если показатель эффективности s является единственным, то его значения на разных стадиях процесса можно представить в виде вектора $\bar{s} = \{s_i\}$, $i = \overline{0, N}$, где N – число этапов процесса. При этом s_0 представляет собой начальное значение показателя (т.е. имеющее место до начала первого этапа), а s_i – значение показателя после завершения i -го этапа процесса. Соответственно, s_N – это значение показателя после завершения последнего, N -го этапа процесса, т.е. всего процесса целиком.

Теперь введем понятие коэффициента динамики, характеризующего положительную динамику рассматриваемого показателя эффективности по мере реализации тех или иных этапов процесса. Для произвольного i -го этапа процесса значение коэффициента динамики определяется следующим образом: $k_i = s_i / s_{i-1}$, где s_i – значение показателя эффективности после завершения i -го этапа процесса, а s_{i-1} – значение того же показателя перед началом этапа (т.е. после завершения предшествующего этапа). По сути, коэффициент динамики характеризует соотношение «входа» и «выхода» процесса. Учитывая то, что процесс состоит из нескольких этапов, значения коэффициента динамики представляют собой вектор $\bar{k} = \{k_i\}$, $i = \overline{1, N}$, где k_i – значение коэффициента на i -м этапе процесса, N – число этапов процесса. Поскольку рассмотренный коэффициент динамики не учитывает ни риски, ни влияние информационной системы, будем называть его базовым или безрисковым.

Также несложно вывести соотношение между начальным значением показателя эффективности (перед началом процесса) и его конечным значением (после полного завершения всего процесса):

$$s_N = s_0 \cdot \prod_{i=1}^N k_i,$$

где s_0 и s_N – соответственно начальное и конечное значения показателя эффективности;

k_i – значение базового коэффициента динамики на i -м этапе процесса;

N – число этапов процесса.

Теперь сделаем допущение относительно рисков общего характера, связанных с реализацией процесса (информационные системы, их влияние на процесс и их внутренние риски пока не рассматриваются).

Определим вероятность реализации риска i -го этапа p_i как вероятность того, что этот этап процесса не будет выполнен и, соответственно, ожидаемые результаты не будут достигнуты. Тогда вероятность того, что результат i -го этапа будет получен, составит $(1 - p_i)$. Например, если доля производственного брака при исполнении i -й технологической операции по всем возможным причинам составляет 1,3%, то $p_i = 0,013$, а вероятность успешного выполнения данной операции составит $(1 - p_i) = 0,987$.

Сказанное позволяет определить вариацию рассмотренного выше коэффициента динамики, скорректировав его значения с учетом рисков (будем называть эту вариацию коэффициентом динамики первого рода). Его связь с базовым (безрисковым) коэффициентом динамики выглядит следующим образом:

$$k_i^{(1)} = k_i \cdot (1 - p_i),$$

где k_i – значение базового (безрискового) коэффициента динамики на i -м этапе процесса;

$k_i^{(1)}$ – значение коэффициента динамики первого рода на i -м этапе процесса;

p_i – вероятность реализации риска на i -м этапе процесса.

Заметим, что значения коэффициента динамики первого рода, соответствующие определенным этапам процесса, всегда меньше значений базового (безрискового) коэффициента, относящихся к тем же этапам.

Поскольку внедрение информационной системы должно способствовать уменьшению вероятностей реализации внутреннего риска, введем такой показатель, как влияние информационной системы на риски. Значение этого показателя, относящееся к i -му этапу процесса (r_i), находится в интервале $[0, 1]$ и показывает, что после внедрения информационной системы вероятность реализации риска на i -м этапе снизится и будет равна $p_i \cdot (1 - r_i)$. В частности, если $r_i = 1$, то внедрение информационной системы должно полностью устранить риск неудачи i -го этапа процесса, а если $r_i = 0$, то это означает, что система не оказывает влияния на риск данного этапа.

Таким образом, можно ввести еще одну вариацию коэффициента динамики, приняв во внимание влияние информационной системы на риски (коэффициент динамики второго рода). Его связь с базовым (безрисковым) коэффициентом динамики выглядит следующим образом:

$$k_i^{(2)} = k_i \cdot (1 - p_i \cdot (1 - r_i)),$$

где k_i – значение базового (безрискового) коэффициента динамики на i -м этапе процесса;

$k_i^{(2)}$ – значение коэффициента динамики второго рода на i -м этапе процесса;

p_i – вероятность реализации риска на i -м этапе процесса;

r_i – влияние информационной системы на риск i -го этапа процесса.

Наконец, примем во внимание, что сама информационная система может давать сбои в своей работе. Такие риски представляют собой внутренние риски системы. Их наличие означает, что реальное влияние информационной системы на риск i -го этапа процесса может быть меньше значения показателя r_i . Если обозначить вероятность внутреннего риска информационной системы на i -м этапе как q_i , то влияние системы на риск данного этапа составит $r_i \cdot (1 - q_i)$. Несложно заметить, что при нулевой вероятности внутреннего риска системы ее влияние на риск этапа не изменится, однако ненулевые значения данного параметра приведут к снижению такого влияния, вплоть до нуля.

Таким образом, еще одна вариация коэффициента динамики (коэффициент динамики третьего рода) будет учитывать все рассмотренные параметры, включая вероятности внутренних рисков информационной системы. Связь коэффициента динамики третьего рода с базовым (безрисковым) коэффициентом выглядит следующим образом:

$$k_i^{(3)} = k_i \cdot (1 - p_i \cdot (1 - r_i \cdot (1 - q_i))),$$

где k_i – значение базового (безрискового) коэффициента динамики на i -м этапе процесса;

$k_i^{(3)}$ – значение коэффициента динамики третьего рода на i -м этапе процесса;

p_i – вероятность реализации риска на i -м этапе процесса;

r_i – влияние информационной системы на риск i -го этапа процесса;

q_i – вероятность реализации внутреннего риска информационной системы на i -м этапе процесса.

Таким образом, оценка эффективности информационной системы, направленной на снижения риска i -го этапа проекта, может быть выражена либо в виде разности между значениями коэффициентов динамики третьего рода (после внедрения информационной системы, с учетом ее влияния

на риск этапа и внутренних рисков самой системы) и первого рода (до внедрения информационной системы), либо в виде отношения этих значений.

Для оценки результата от внедрения идеальной (не содержащей внутренних рисков) и реальной (с внутренними рисками) информационной системы можно использовать следующие формулы:

$$\Delta s^{ideal} = s_0 \cdot \left(\prod_{i=1}^N k_i^{(2)} - \prod_{i=1}^N k_i^{(1)} \right),$$

$$\Delta s^{real} = s_0 \cdot \left(\prod_{i=1}^N k_i^{(3)} - \prod_{i=1}^N k_i^{(1)} \right),$$

где Δs^{ideal} и Δs^{real} – эффективность соответственно идеальной и реальной информационной системы;

s_0 – начальное значение показателя эффективности;

$k_i^{(1)}, k_i^{(2)}, k_i^{(3)}$ – значения коэффициента динамики соответственно первого, второго и третьего рода на i -м этапе процесса;

N – число этапов процесса.

Приведенные рассуждения можно распространить на случай, когда для процесса рассматриваются не один, а несколько показателей эффективности. Тогда вместо вектора значений показателей будем иметь матрицу $\mathbf{S} = \{s_{ij}\}$, $i = 0, N, j = 1, M$, где s_{0j} – начальное значение j -го показателя, s_{ij} – значение j -го показателя после i -го этапа процесса, M – число показателей, N – число этапов процесса.

Значения коэффициентов динамики будут определяться отдельно для каждого из показателей эффективности и тоже будут представлять собой матрицы. Например, значениям базовых коэффициентов динамики будет соответствовать матрица $\mathbf{K} = \{k_{ij}\}$, $i = 1, N, j = 1, M$, где s_{ij} – значение j -го коэффициента на i -м этапе процесса, M – число показателей, N – число этапов процесса. Для коэффициентов динамики первого, второго и третьего рода будут существовать аналогичные матрицы.

Заключение

Повышение эффективности деятельности компании может достигаться путем внедрения информационных систем, что сопровождается соответствующими изменениями в организационной структуре и организации бизнес-процессов. Однако риски, связанные с данными действиями, дают

повод задуматься о целесообразности проекта. При этом проблемы со стороны производственной и управленческой деятельности до внедрения информационной системы также представляют собой определенную угрозу для компании. Таким образом, появляется возможность получить оценку ИТ-проекта путем учета как положительных, так и отрицательных последствий внедрения системы. ■

Благодарности

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (Государственное задание «Структура и свойства полимерных материалов, полученных с применением системы методов химически, термически и/или механически индуцированной поверхностной и объемной модификации», номер темы FZRR-2020-0024, мнемокод 0699-2020-0024).

Литература

1. Khanfar A.A., Mavi R.K., Jie F. Prioritizing critical failure factors of IT projects with fuzzy analytic hierarchy process // *Proceedings of the International Conference on Mathematics, Engineering and Industrial Applications (ICoMEIA)*. Kuala Lumpur, Malaysia, 24–26 July 2018. Vol. 2013. No 020058. DOI: 10.1063/1.5054257.
2. Sorooshian S., Mun S.Y. Literature review: Critical risk factors affecting information-technology projects // *Quality – Access to Success*. 2020. Vol. 21. No 175. P. 157–161.
3. Исаев Е.А., Самодуров В.А., Первухин Д.В., Филюгина Е.К. Применение сверточных нейронных сетей (CNN) для поиска паттернов в рядах данных различной природы // *Промышленные АСУ и контроллеры*. 2019. № 12. С. 24–32. DOI: 10.25791/asu.12.2019.1069.
4. Анализ положительного эффекта от внедрения ИТ решения, основанный на оценке рисков / Д.В. Первухин и [др.] // *Приборы и системы. Управление, контроль, диагностика*. 2019. № 7. С. 45–54. DOI: 10.25791/pribor.07.2019.742.
5. Ayuso S., Rodriguez M.A., Garcia-Castro R., Arino M.A. Maximizing stakeholders' interests: An empirical analysis of the stakeholder approach to corporate governance // *Business & Society*. 2014. Vol. 53. No 3. P. 414–439. DOI: 10.1177/0007650311433122.
6. Mishra A., Sinha K.K., Thirumalai S., Van de Ven A. Sourcing structures and the execution efficiency of information technology projects: A comparative evaluation using stochastic frontier analysis // *Journal of Operations Management*. 2020. Vol. 66. No 3. P. 281–309. DOI: 10.1002/joom.1064.
7. Klaus-Rosinska A. Concept of measuring the performance of IT projects // *Proceedings of the 25th International Conference on Systems Engineering (ICSEng)*. Las Vegas, US, 22–24 August 2017. P. 412–417. DOI: 10.1109/ICSEng.2017.63.
8. Rodriguez A., Ortega D., Concepcion R. An intuitionistic method for the selection of a risk management approach to information technology projects // *Information Sciences*. 2017. Vol. 375. P. 202–218. DOI: 10.1016/j.ins.2016.09.053.
9. Shafiee S., Kristjansdottir K., Hvam L., Forz C. How to scope configuration projects and manage the knowledge they require // *Journal of Knowledge Management*. 2018. Vol. 22. No 5. P. 982–1014. DOI: 10.1108/JKM-01-2017-0017.
10. Сравнительный анализ теоретических моделей каскадных, итеративных и гибридных подходов к управлению жизненным циклом ИТ-проекта / Д.В. Первухин [др.] // *Бизнес-информатика*. 2020. Т. 14. № 1. С. 32–40. DOI: 10.17323/2587-814X.2020.1.32.40.
11. Boehm B.W. Software risk management: Principles and practices // *IEEE Software*. 1991. Vol. 8. No 1. P. 32–41. DOI: 10.1109/52.62930.
12. Gregory P.H. CISA certified information systems auditor all-in-one exam guide. New York: McGraw-Hill, 2010.
13. Ako-Nai A., Singh A.M. Information technology governance framework for improving organizational performance // *South African Journal of Information Management*. 2019. Vol. 21. No 1. Article No a1010. DOI: 10.4102/sajim.v21i1.1010.
14. Ul Haq S., Gu D., Liang C., Abdullah I. Project governance mechanisms and the performance of software development projects: Moderating role of requirements risk // *International Journal of Project Management*. 2019. Vol. 37. No 4. P. 533–548. DOI: 10.1016/j.ijproman.2019.02.008.
15. Almutairi M., Riddle S. A framework for managing security risks of outsourced IT projects: An empirical study // *Proceedings of the International Conference on Software Engineering and Information Management (ICSIM 2018)*. Casablanca, Morocco, 4–6 January 2018. P. 40–44. DOI: 10.1145/3178461.3178476.
16. Implementation of risk mitigation among IT governance practitioners in Malaysia / A. Bokolo (Jr.) [et al.] // *Advanced Science Letters*. 2018. Vol. 24. No 2. P. 1344–1347. DOI: 10.1166/asl.2018.10746.
17. Lee J.S., Keil M., Shalev E. Seeing the trees or the forest? The effect of IT project managers' mental construal on IT project risk management activities // *Information System Research*. 2019. Vol. 30. No 3. P. 1051–1072. DOI: 10.1287/isre.2019.0853.
18. Lai S.-T., Leu F.-Y. A critical quality measurement model for managing and controlling big data project risks // *Proceedings of the 12th IEEE International Conference on Broadband Wireless Computing, Communication and Applications (BWCCA)*. Barcelona, Spain, 8–10 November 2017. Vol. 12. P. 777–787. DOI: 10.1007/978-3-319-69811-3_69.
19. Pike G. Supporting business innovation while reducing technology risk. White paper. Walldorf, Germany: SAP AG, 2006.

20. Wu D.J., Ding M., Hitt L.M. IT implementation contract design: Analytical and experimental investigation of IT value, learning, and contract structure // *Information Systems Research*. 2013. Vol. 24. No 3. P. 787–801. DOI: 10.1287/isre.1120.0448.
21. Ghribi S., Hudon P.A., Mazouz B. Risk factors in IT public–private partnership projects // *Public Works Management & Policy*. 2019. Vol. 24. No 4. P. 321–343. DOI: 10.1177/1087724X18823009.
22. Didraga O., Brandas C., Batagan L., Alecu F. Characteristics of effective IT project risk management in Romanian IT companies // *Economic Computation and Economic Cybernetics Studies and Research*. 2019. Vol. 53. No 4. P. 176–193. DOI: 10.24818/18423264/53.4.19.11.
23. Pimchangthong D., Boonjing V. Effects of risk management practices on IT project success // *Management and Production Engineering Review*. 2017. Vol. 8. No 1. P. 30–37. DOI: 10.1515/mper-2017-0004.
24. Neumeier A., Radszuwill S., Garizy T.Z. Modeling project criticality in IT project portfolios // *International Journal of Project Management*. 2018. Vol. 36. No 6. P. 833–844. DOI: 10.1016/j.ijproman.2018.04.005.
25. Maruping L.M., Venkatesh V., Thong J.Y.L., Zhang X. A risk mitigation framework for information technology projects: A cultural contingency perspective // *Journal of Management Information Systems*. 2019. Vol. 36. No 1. P. 120–157. DOI: 10.1080/07421222.2018.1550555.
26. Lee J.S., Keil M. The effects of relative and criticism-based performance appraisals on task-level escalation in an IT project: a laboratory experiment // *European Journal of Information Systems*. 2018. Vol. 27. No 5. P. 551–569. DOI: 10.1080/0960085X.2017.1408752.
27. Gloria I. (Jr.), Chaves M.S. Identification and mitigation of risks in IT projects: a case study during the merger period in the telecommunications industry // *Revista de Gestao e Projetos*. 2017. Vol. 8. No 3. P. 1–17. DOI: 10.5585/gep.v8i3.581.
28. COBIT5. A business framework for the governance and management of enterprise IT. ISACA, 2012.
29. Almutairi M., Riddle S. Managing outsourced IT projects' security risks: A case study // *Proceedings of the 10th International Conference on Information Management and Engineering (ICIME 2018)*. Manchester, England, 22–24 September 2018. P. 21–26. DOI: 10.1145/3285957.3285986.
30. Thirasakthana M., Kiattisin S. Identifying standard testing time for estimation improvement in IT project management // *Proceedings of the 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*. Bangkok, Thailand, 12–14 December 2018. P. 1–5. DOI: 10.1109/TIMES-iCON.2018.8621787.

Об авторах

Исаев Евгений Анатольевич

кандидат технических наук;

старший научный сотрудник, Институт математических проблем биологии, Российская академия наук (филиал Института прикладной математики им. М.В. Келдыша РАН), 142290, Московская область, г. Пушкино, ул. Профессора Виткевича, д. 1;
E-mail: is@itaec.ru

ORCID: 0000-0002-3703-447X

Первухин Дмитрий Васильевич

старший преподаватель, Государственный университет управления, 109542, г. Москва, Рязанский проспект, д. 99;

E-mail: dvperv@gmail.com

ORCID: 0000-0001-6500-035X

Рытиков Георгий Олегович

кандидат физико-математических наук;

доцент, Государственный университет управления, 109542, г. Москва, Рязанский проспект, д. 99;

доцент, Московский политехнический университет, 107023, г. Москва, ул. Большая Семеновская, 38;

E-mail: GR-yandex@yandex.ru

ORCID: 0000-0001-5521-8662

Филогина Екатерина Константиновна

сотрудник ООО «Импакт Электроникс», 127055, г. Москва, ул. Новослободская, д. 14/19, стр. 8;

E-mail: ekaterina.filyugina@mail.ru

ORCID: 0000-0001-6461-7235

Айрапетян Диана Ареновна

экономист, ООО «Эпоха возрождения», 107078, г. Москва, ул. Русаковская, д. 13, стр. 2;

E-mail: hayrapetyandiana@gmail.com

ORCID: 0000-0001-6646-1748

Risk-based efficiency assessment of information systems

Eugeni A. Isaev^a

E-mail: is@itaec.ru

Dmitry V. Pervukhin^b

E-mail: dvperv@gmail.com

Georgy O. Rytikov^{b,c}

E-mail: GR-yandex@yandex.ru

Ekaterina K. Filyugina^d

E-mail: ekaterina.filyugina@mail.ru

Diana A. Hayrapetyan^e

E-mail: hayrapetyandiana@gmail.com

^a Institute of Mathematical Problems of Biology, Russian Academy of Sciences (Branch of the Keldysh Institute of Applied Mathematics, Russian Academy of Sciences)

Address: 1, Professor Vitkevich Street, Pushino, Moscow Region 142290, Russia

^b State University of Management

Address: 99, Ryazansky Prospect, Moscow 109542, Russia

^c Moscow Polytechnic University

Address: 38, Bolshaya Semyonovskaya Street, Moscow 107023, Russia

^d Impact Electronics Ltd.

Address: 14/19 build. 8, Novoslobodskaya Street, Moscow 127055, Russia

^e Epokha Vozrozhdeniya Ltd.

Address: 13 build. 2, Rusakovskaya Street, Moscow 107078, Russia

Abstract

The implementation of information systems is aimed at improving the financial performance of a company, creating a transparent reporting system and improving many other competitive factors. However, the acquisition of these benefits does not negate the complexity of making a decision whether or not to implement a particular IT project. The total cost of ownership of the information system throughout the life cycle is usually not considered in comparison with the expected benefits from the use of the system, due to the uncertainty of such benefits. Comparative certainty of approaches and methods is present only in terms of costs, both for a priori (planned) and a posteriori (actual) assessment. It is possible to determine both capital and operating costs accurately enough. Indirect definition of the positive influence of an information system on the activity of the organization also seems possible. However, there are currently no generally recognized methods for analyzing the expected positive effect of an IT project. At the same time, large companies, in accordance with the requirements of the respective regulators and / or due to internal management considerations, build a risk management system to determine the level of capabilities, losses and to prevent adverse events. This study considers the feasibility of an approach to analyze the effectiveness of the implementation of the information system on the basis of the company's risk reduction, leading to a decrease in economic benefits. It takes into account the internal risks of the information system that occur during the installation of the system, its operation and the termination of work with the system.

Key words: risk assessment; IT project management; information system; implementation of information systems.

Citation: Isaev E.A., Pervukhin D.V., Rytikov G.O., Filyugina E.K., Hayrapetyan D.A. (2021) Risk-based efficiency assessment of information systems. *Business Informatics*, vol. 15, no 1, pp. 19–29.

DOI: 10.17323/2587-814X.2021.1.19.29

References

1. Khanfar A.A., Mavi R.K., Jie F. (2018) Prioritizing critical failure factors of IT projects with fuzzy analytic hierarchy process. Proceedings of the *International Conference on Mathematics, Engineering and Industrial Applications (ICoMEIA)*, Kuala Lumpur, Malaysia, 24–26 July 2018, vol. 2013, no 020058. DOI: 10.1063/1.5054257.
2. Sorooshian S., Mun S.Y. (2020) Literature review: Critical risk factors affecting information-technology projects. *Quality – Access to Success*, vol. 21, no 175, pp. 157–161.
3. Isaev E.A., Samodurov V.A., Pervukhin D.V., Filyugina E.K. (2019) The application of convolutional neural networks (CNN) to search for patterns of various nature in data series. *Industrial Automatic Control Systems and Controllers*, no 12, pp. 24–32 (in Russian). DOI: 10.25791/asu.12.2019.1069.
4. Pervukhin D.V., Isaev E.A., Rytikov G.O., Filyugina E.K., Hayrapetyan D.A. (2019) Analysis of the positive effect of the IT solutions implementation based on risk assessment. *Instruments and Systems: Monitoring, Control, and Diagnostics*, no 7, pp. 45–54 (in Russian). DOI: 10.25791/pribor.07.2019.742.
5. Ayuso S., Rodriguez M.A., Garcia-Castro R., Arino M.A. (2014) Maximizing stakeholders' interests: An empirical analysis of the stakeholder approach to corporate governance. *Business & Society*, vol. 53, no 3, pp. 414–439. DOI: 10.1177/0007650311433122.
6. Mishra A., Sinha K.K., Thirumalai S., Van de Ven A. (2020) Sourcing structures and the execution efficiency of information technology projects: A comparative evaluation using stochastic frontier analysis. *Journal of Operations Management*, vol. 66, no 3, pp. 281–309. DOI: 10.1002/joom.1064.
7. Klaus-Rosinska A. (2017) Concept of measuring the performance of IT projects. Proceedings of the *25th International Conference on Systems Engineering (ICSEng)*, Las Vegas, US, 22–24 August 2017, pp. 412–417. DOI: 10.1109/ICSEng.2017.63.
8. Rodriguez A., Ortega D., Concepcion R. (2017) An intuitionistic method for the selection of a risk management approach to information technology projects. *Information Sciences*, vol. 375, pp. 202–218. DOI: 10.1016/j.ins.2016.09.053.
9. Shafiee S., Kristjansdottir K., Hvam L., Forz C. (2018) How to scope configuration projects and manage the knowledge they require. *Journal of Knowledge Management*, vol. 22, no 5, pp. 982–1014. DOI: 10.1108/JKM-01-2017-0017.
10. Pervukhin D.V., Isaev E.A., Rytikov G.O., Filyugina E.K., Hayrapetyan D.A. (2020) Theoretical comparative analysis of cascading, iterative, and hybrid approaches to IT project life cycle management. *Business Informatics*, vol. 14, no 1, pp. 32–40. DOI: 10.17323/2587-814X.2020.1.32.40.
11. Boehm B.W. (1991) Software risk management: Principles and practices. *IEEE Software*, vol. 8, no 1, pp. 32–41. DOI: 10.1109/52.62930.
12. Gregory P.H. (2010) *CISA certified information systems auditor all-in-one exam guide*. New York: McGraw-Hill.
13. Ako-Nai A., Singh A.M. (2019) Information technology governance framework for improving organizational performance. *South African Journal of Information Management*, vol. 21, no 1, article no a1010. DOI: 10.4102/sajim.v21i1.1010.
14. Ul Haq S., Gu D., Liang C., Abdullah I. (2019) Project governance mechanisms and the performance of software development projects: Moderating role of requirements risk. *International Journal of Project Management*, vol. 37, no 4, pp. 533–548. DOI: 10.1016/j.ijproman.2019.02.008.
15. Almutairi M., Riddle S. (2018) A framework for managing security risks of outsourced IT projects: An empirical study. Proceedings of the *International Conference on Software Engineering and Information Management (ICSIM 2018)*, Casablanca, Morocco, 4–6 January 2018, pp. 40–44. DOI: 10.1145/3178461.3178476.
16. Bokolo A. (Jr.), Pa N.C., Nor R.N.H., Jusoh Y.Y., Aris T.N.M. (2018) Implementation of risk mitigation among IT governance practitioners in Malaysia. *Advanced Science Letters*, vol. 24, no 2, pp. 1344–1347. DOI: 10.1166/asl.2018.10746.
17. Lee J.S., Keil M., Shalev E. (2019) Seeing the trees or the forest? The effect of IT project managers' mental construal on IT project risk management activities. *Information System Research*, vol. 30, no 3, pp. 1051–1072. DOI: 10.1287/isre.2019.0853.
18. Lai S.-T., Leu F.-Y. (2017) A critical quality measurement model for managing and controlling big data project risks. Proceedings of the *12th IEEE International Conference on Broadband Wireless Computing, Communication and Applications (BWCCA)*, Barcelona, Spain, 8–10 November 2017, vol. 12, pp. 777–787. DOI: 10.1007/978-3-319-69811-3_69.
19. Pike G. (2006) *Supporting business innovation while reducing technology risk*. White paper. Walldorf, Germany: SAP AG.
20. Wu D.J., Ding M., Hitt L.M. (2013) IT implementation contract design: Analytical and experimental investigation of IT value, learning, and contract structure. *Information Systems Research*, vol. 24, no 3, pp. 787–801. DOI: 10.1287/isre.1120.0448.
21. Ghribi S., Hudon P.A., Mazouz B. (2019) Risk factors in IT public–private partnership projects. *Public Works Management & Policy*, vol. 24, no 4, pp. 321–343. DOI: 10.1177/1087724X18823009.
22. Didraga O., Brandas C., Batagan L., Alecu F. (2019) Characteristics of effective IT project risk management in Romanian IT companies. *Economic Computation and Economic Cybernetics Studies and Research*, vol. 53, no 4, pp. 176–193. DOI: 10.24818/18423264/53.4.19.11.
23. Pimchangthong D., Boonjing V. (2017) Effects of risk management practices on IT project success. *Management and Production Engineering Review*, vol. 8, no 1, pp. 30–37. DOI: 10.1515/MPER-2017-0004.
24. Neumeier A., Radszuwill S., Garizy T.Z. (2018) Modeling project criticality in IT project portfolios. *International Journal of Project Management*, vol. 36, no 6, pp. 833–844. DOI: 10.1016/j.ijproman.2018.04.005.
25. Maruping L.M., Venkatesh V., Thong J.Y.L., Zhang X. (2019) A risk mitigation framework for information technology projects: A cultural contingency perspective. *Journal of Management Information Systems*, vol. 36, no 1, pp. 120–157. DOI: 10.1080/07421222.2018.1550555.
26. Lee J.S., Keil M. (2018) The effects of relative and criticism-based performance appraisals on task-level escalation in an IT project: a laboratory experiment. *European Journal of Information Systems*, vol. 27, no 5, pp. 551–569. DOI: 10.1080/0960085X.2017.1408752.

27. Gloria I. (Jr.), Chaves M.S. (2017) Identification and mitigation of risks in IT projects: a case study during the merger period in the telecommunications industry. *Revista de Gestao e Projetos*, vol. 8, no 3, pp. 1–17. DOI: 10.5585/gep.v8i3.581.
28. ISACA (2012) *COBIT5. A business framework for the governance and management of enterprise IT*. ISACA.
29. Almutairi M., Riddle S. (2018) Managing outsourced IT projects' security risks: A case study. Proceedings of the *10th International Conference on Information Management and Engineering (ICIME 2018)*, Manchester, England, 22–24 September 2018, pp. 21–26. DOI: 10.1145/3285957.3285986.
30. Thirasakthana M., Kiattisin S. (2018) Identifying standard testing time for estimation improvement in IT project management. Proceedings of the *3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, Bangkok, Thailand, 12–14 December 2018, pp. 1–5. DOI: 10.1109/TIMES-iCON.2018.8621787.

About the authors

Eugeni A. Isaev

Cand. Sci. (Tech.);

Senior Researcher, Institute of Mathematical Problems of Biology, Russian Academy of Sciences (Branch of the Keldysh Institute of Applied Mathematics, Russian Academy of Sciences), 1, Professor Vitkevich Street, Pushino, Moscow Region 142290, Russia;

E-mail: is@itaec.ru

ORCID: 0000-0002-3703-447X

Dmitry V. Pervukhin

Senior Lecturer, State University of Management, 99, Ryazansky Prospekt, Moscow 109542, Russia;

E-mail: dvperv@gmail.com

ORCID: 0000-0001-6500-035X

Georgy O. Rytikov

Cand. Sci. (Phys.-Math.);

Associate Professor, State University of Management, 99, Ryazansky Prospekt, Moscow 109542, Russia;

Associate Professor, Moscow Polytechnic University, 38, Bolshaya Semyonovskaya Street, Moscow 107023, Russia;

E-mail: GR-yandex@yandex.ru

ORCID: 0000-0001-5521-8662

Ekaterina K. Filyugina

Impact Electronics Ltd., 14/19 build. 8, Novoslobodskaya Street, Moscow 127055, Russia;

E-mail: ekaterina.filyugina@mail.ru

ORCID: 0000-0001-6461-7235

Diana A. Hayrapetyan

Economist, Epokha Vozrozhdeniya Ltd., 13 build. 2, Rusakovskaya Street, Moscow 107078, Russia;

E-mail: hayrapetyandiana@gmail.com

ORCID: 0000-0001-6646-1748

Динамика исследований в области интеллектуального анализа данных: тематический анализ публикаций за 20 лет

Ю.А. Зеленков 

E-mail: yzelenkov@hse.ru

Е.А. Анисичкина

E-mail: eaanisichkina@edu.hse.ru

Национальный исследовательский университет «Высшая школа экономики»

Адрес: 101000, г. Москва, ул. Мясницкая, д. 20

Аннотация

Цель работы состоит в выявлении тенденций развития интеллектуального анализа данных (data mining) как научной дисциплины. На основе метода латентного размещения Дирихле (latent Dirichlet allocation, LDA) построена тематическая модель трудов Международной конференции по интеллектуальному анализу данных (International Conference on Data Mining, ICDM) за 2001–2019 годы и выделено девять основных направлений (тем) исследований. Для каждой темы исследованы динамика ее популярности (количество публикаций) и влияния (количество цитирований). Центральная тема, которая объединяет все прочие направления, — это общие вопросы обучения (general learning), включающее алгоритмы машинного обучения. За рассматриваемый период 20% усилий научного сообщества было потрачено на развитие именно этого направления, однако в последнее время его влияние снижается. Также снижается внимание к таким темам, как обнаружение ассоциаций (pattern mining) и разделение объектов (segmentation), включая кластеризацию. В то же время растет популярность исследований, связанных с рекомендательными системами (recommender systems), анализ сетей различной природы, в том числе социальных (network analysis) и анализ поведения человека, в частности поведения потребителей (human behavior analysis), что, скорее всего, связано с увеличением доступности данных и практической ориентацией этих тем. Направление исследований, связанных с приложениями интеллектуального анализа данных (applications), также имеет тенденцию к росту. Две последние темы — анализ текстовой информации (text mining) и прогнозирование потоков данных (data streams) привлекают относительно постоянный интерес исследователей. Полученные результаты проливают свет на структуру и динамику интеллектуального анализа данных как научной дисциплины за последние двадцать лет. Они также свидетельствуют, что за последние пять лет сформировалась новая повестка, характеризующаяся сдвигом интереса от алгоритмов к практическим приложениям, влияющим на все аспекты человеческой деятельности.

Ключевые слова: интеллектуальный анализ данных; датамайнинг; тематический анализ; наукометрия.

Цитирование: Зеленков Ю.А., Анисичкина Е.А. Динамика исследований в области интеллектуального анализа данных: тематический анализ публикаций за 20 лет // Бизнес-информатика. 2021. Т. 15. № 1. С. 30–46. DOI: 10.17323/2587-814X.2021.1.30.46

Введение

Термин «интеллектуальный анализ данных» (data mining, DM) используется с 1960-х годов для описания поиска корреляций без выдвижения априорных гипотез [1]. Согласно широко принятому определению, которое сейчас используется во многих учебниках, интеллектуальный анализ данных — это извлечение неявной, ранее неизвестной и потенциально полезной информации из данных [2, 3]. В учебнике [4] интеллектуальный анализ данных определяется как комбинация трех концепций, к числу которых относятся:

- ♦ статистика, которая включает классические описательные модели и инструменты, например, степени свободы, F -статистики и p -значения, но не рассматривает вывод гипотез;

- ♦ большие данные, как общий термин для наборов данных любого размера, но с акцентом на большой объем, поскольку доступность данных влияет практически на все аспекты нашей жизни;

- ♦ машинное обучение, то есть инструменты для создания компьютерных программ, которые анализируют базы данных в поисках закономерностей или шаблонов [2].

Статистика и машинное обучение создают техническую основу интеллектуального анализа данных. Некоторые авторы также рассматривают DM как часть процесса извлечения знаний (knowledge data discovery). Этот процесс может включать такие методы, как предварительная обработка данных (очистка и интеграция), хранение данных, аналитическая обработка данных в режиме онлайн, кубы данных и т.д. [3].

Как следует из приведенных определений, интеллектуальный анализ данных — это научная дисциплина, сочетающая достижения в нескольких областях исследований. Структура любой научной дисциплины может быть представлена как набор эволюционирующих тем, то есть значимых неявных ассоциаций, скрытых во фрагментированных областях знаний. Динамика этих тем (например, изменение количества публикаций и их цитирова-

ния) отражает сдвиг интересов научного сообщества. В частности, изучение этой динамики позволяет определить наиболее актуальные направления исследований, имеющие место в настоящее время, и экстраполировать их на ближайшее будущее. Кроме того, понимание фундаментальной динамики интересов исследователей помогает определить место изучаемой дисциплины в общей совокупности человеческих знаний, ее взаимодействие с другими дисциплинами и вклад в научно-технический прогресс.

Традиционным методом изучения структуры научной дисциплины является обзор литературы. Однако из-за междисциплинарного характера интеллектуального анализа данных практически нет обзоров, рассматривающих DM как отдельную дисциплину. Тем не менее, следует отметить, что обзоры более узких тем, например, таких как рекомендательные системы, публикуются постоянно.

К. Янг и С. Ву в обзорной статье [5], опубликованной в 2006 году, констатировали, что интеллектуальный анализ данных достиг огромных успехов. Однако, они также отметили проблемы с эффективным и своевременным обменом важными темами в научном сообществе. Кроме того, авторы выделили десять наиболее важных задач DM [5]:

- ♦ разработка единой теории интеллектуального анализа данных;
- ♦ масштабирование для данных высокой размерности и высокоскоростных потоков данных;
- ♦ анализ последовательностей данных и временных рядов;
- ♦ извлечение сложных знаний из сложных данных;
- ♦ анализ графов;
- ♦ распределенный и мультиагентный анализ данных;
- ♦ анализ данных в биологии и экологии;
- ♦ процесс интеллектуального анализа данных;
- ♦ безопасность, конфиденциальность и целостность данных;
- ♦ работа с нестационарными, несбалансированными и чувствительными к стоимости ошибки данными.

Перечисленные задачи разделяют общий поток исследований DM на более мелкие и более сфокусированные сегменты. В 2010 году С. Ву предоставил дополнительные комментарии по этим вопросам [6], и они стали предметом обсуждения на специальной панели на 10-й Международной конференции по интеллектуальному анализу данных (ICDM).

К. Янг и С. Ву [5] рассматривали разработку единой теории DM как наиболее важную цель. Это должна быть теоретическая основа, объединяющая различные методы, разработанные для отдельных задач, включая кластеризацию, классификацию, ассоциативные правила и т.д., а также различные технологии интеллектуального анализа данных, такие как статистика, машинное обучение и системы баз данных. По мнению авторов, такая теория должна обеспечить основу для будущих исследований.

Отметим, что большинство перечисленных проблем относится к алгоритмам работы с новыми типами данных, которые стали актуальными в 2000-х годах (данные сверхвысокой размерности, высокоскоростные потоки данных, временные ряды, сети и другие сложные структуры). Кроме того, авторы работ [5, 6] отмечают экологическую информатику как важнейшую область приложений DM.

Помимо анализа критических проблем, в работе [6] представлен список наиболее важных тем интеллектуального анализа данных (таблица 1). Этот список составлен на основании опроса экспертов; следовательно, он может служить ссылкой на структуру

научной дисциплины. Однако экспертный подход не позволяет определить количественные показатели, которые определяют относительную важность различных тем и их изменение с течением времени.

В работе [7] представлен обзор литературы по методам и приложениям интеллектуального анализа данных с января 2000 г. по август 2011 г. Авторы выбрали 216 статей из 159 научных журналов, используя такие ключевые слова, как “data mining”, “decision tree”, “artificial neural network”, “clustering” и т.д. На основании выбранных документов они определили девять категорий методов DM (оптимизация систем, системы, основанные на знаниях, моделирование, архитектура алгоритмов, нейронные сети и т.д.). Также авторы [7] выявили основные тенденции в области интеллектуального анализа данных. Согласно представленным результатам, наиболее важной тенденцией являются ассоциативные правила (ранг 5), за ними следуют нейронные сети (ранг 4), а затем – классификация и метод опорных векторов (оба направления имеют ранг 3). Авторы не описывают метод ранжирования, однако можно предположить, что он основан на подсчете количества ссылок на каждую тему в анализируемом корпусе публикаций.

Насколько нам известно, процитированные выше публикации – единственные, в которых исследуется динамика интеллектуального анализа данных как единой научной дисциплины. Как уже отмечалось, они основаны на субъективных (экспертных) оценках.

Идея нашей работы – применить формальные методы тематического анализа к публикациям в области DM. В качестве объекта анализа мы используем материалы Международной конференции по интеллектуальному анализу данных (International Conference on Data Mining, ICDM), которая проводится ежегодно с 2001 года.

1. Данные

Международная конференция по интеллектуальному анализу данных (International Conference on Data Mining, ICDM) – это ведущая конференция, которая, наряду с SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), ACM International Conference on Web Search and Data Mining (WSDM) и некоторыми другими, образует сеть основных форумов в области интеллектуального анализа данных. База данных Web of Science (WoS) содержит информацию о 5120 публикациях основных треков ICDM и связанных с ними се-

Таблица 1.

Десять главных тем интеллектуального анализа данных [6]

№	Тема
1	Классификация (включая C4.5, CART, kNN и Naive Bayes)
2	Статистическое обучение (SVM и mixture models)
3	Анализ ассоциаций
4	Обнаружение связей (например, алгоритм PageRank)
5	Кластерный анализ
6	Бэджинг и бустинг
7	Шаблоны последовательностей
8	Интегрированный анализ (например, объединение классификации и ассоциативных правил)
9	Приближенные множества (rough sets)
10	Анализ графов

минаров за 2001–2019 годы. На *рисунке 1* показано распределение этих публикаций по времени.

База данных WoS содержит информацию, необходимую для нашего исследования, включая сведения об авторах, названия публикаций, аннотации и количество цитирований.

2. Метод исследования

Одним из наиболее популярных методов анализа библиографических сетей является установление связей между терминами пропорционально частоте их совместного появления в документах (term-level coupling), реализованное в программном пакете VOSviewer [8]. Этот подход позволяет идентифицировать кластеры слов, которые можно рассматривать как более или менее устойчивые неявные структуры, формирующие научную дисциплину. Авторы обзора [9] перечисляют основные вычислительные методы, автоматизирующие процесс обнаружения знаний в публикациях. Они отмечают, что тематическое моделирование, которое позволяет наблюдать как информация на уровне темы распространяется среди документов, обеспечивает более глубокое понимание корпуса документов, чем анализ на уровне терминов. Однако тематическое моделирование все еще относительно редко используется при анализе научной литературы [9, 10].

В статье [11] используется сочетание тематического моделирования и анализа цитирования для оценки динамики влияния тем и тематического разноо-

бразия публикаций по информатике. Авторы статьи [12] использовали тематическое моделирование для анализа содержания материалов Международной конференции по принципам и практике многоагентных систем (International Conference on Principles and Practice of Multi-Agent Systems, PRIMA). Среди недавних публикаций, в работе [10] тематический анализ применяется к области управления знаниями. В последней статье особое внимание уделяется динамике тем, то есть тому, как количество публикаций и цитирований по каждой теме меняется во времени. Это позволяет пролить свет на изменение исследовательского интереса и выявить критические тенденции настоящего времени.

Еще одно приложение тематического анализа представлено в работе [13], где он используется для количественной оценки сходства и эволюции научных дисциплин. В работе [14] авторы предлагают эволюционную тематическую модель на базе деревьев, генерируемых на основе неоднородной библиографической сети.

Определим тему как множество слов, которые часто совместно встречаются в текстах, относящихся к определенной предметной области. Вероятностное тематическое моделирование основывается на идее, что документы представляют собой смесь тем, а каждая тема представляет собой распределение вероятностей по терминам.

Пусть корпус документов D содержит множество тем T , которое неизвестно. Каждое использование

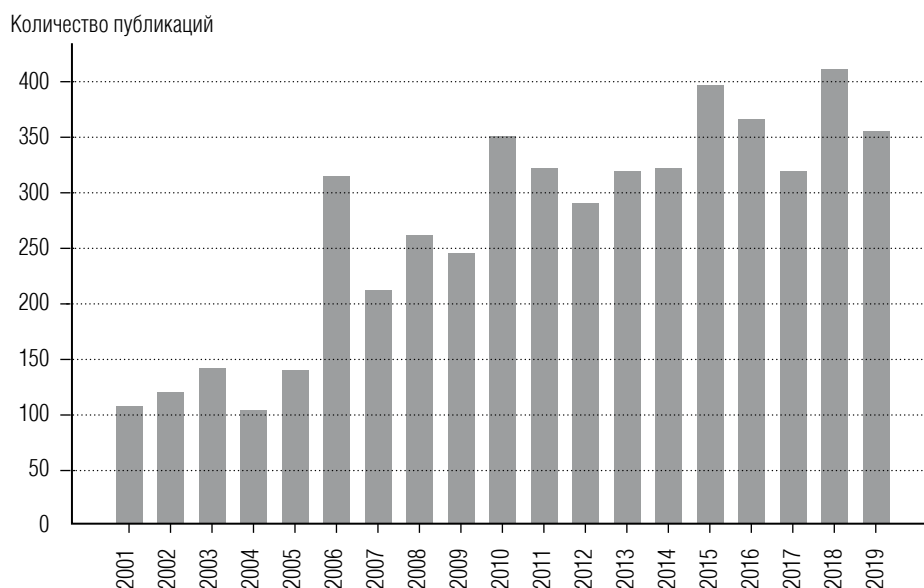


Рис. 1. Количество публикаций в трудах ICDM

термина w в документе d ассоциируется с некоторой темой $t \in T$. Таким образом, коллекцию документов можно рассматривать как множество триплетов (d, w, t) , выбранных случайно и независимо из распределения, определенного на конечном множестве $D \times W \times T$. Документы $d \in D$ и термины $w \in W$ – наблюдаемые переменные. Темы $t \in T$ – это скрытые переменные, которые необходимо определить.

Тематическая модель автоматически определяет скрытые темы по наблюдаемым частотам слов в документах:

$$p(w|d) = \sum_{t \in T} p(t|d) p(w|t).$$

Входными данными алгоритма является матрица $D \times W$, ячейки которой содержат количества слов w в документе d .

Для того, чтобы построить матрицу $D \times W$, мы использовали аннотации 5120 статей, загруженных из базы данных Web of Science и описанных в предыдущем разделе. Согласно [15], различия между данными, подготовленными на основе аннотаций и полных текстов статей, более очевидны в небольших (размером несколько сотен) коллекциях документов. Поэтому в качестве объекта анализа мы выбрали аннотации.

В соответствии с общей методикой интеллектуального анализа текста, аннотации были токенизированы, а полученные термы преобразованы в стандартную форму. Затем были удалены слова, принадлежащие расширенному списку стоп-слов. Расширенный список стоп-слов включает стандартные английские стоп-слова и специфические для корпуса слова, которые встречаются менее чем в 5% и более чем в 60% документов. Мы также создали биграммы для объединения термов, которые часто встречаются рядом. В результате мы получили разреженную матрицу $D \times W$ размером 5120×1000 , только 1,62% ячеек которой содержат значения, отличные от нуля.

Для вычисления тем мы использовали алгоритм латентного размещения Дирихле (latent Dirichlet allocation, LDA), который основан на дополнительном предположении, что распределение Θ документов θ_d и распределение Φ тем φ_t порождаются распределениями Дирихле [16]. Для построения модели следует определить количество тем $|T|$; алгоритм LDA вычисляет распределения Θ и Φ . В результате каждая тема представлена взвешенным списком слов, а вес слова соответствует его важности в определении темы. Каждый доку-

мент представляется взвешенным списком тем, а вес темы соответствует ее значимости в документе.

Определение количества тем – ключевой вопрос тематического анализа. Многие авторы используют различные виды поиска по сетке для оптимизации определенной метрики [10]. В нашем исследовании использовалась байесовская оптимизация [17]. Этот подход позволяет оптимизировать одновременно не только количество тем, но и параметры распределений Θ и Φ , а также другие параметры алгоритма. Оптимизируемая метрика при этом – перплексия $P(D)$, которая измеряет сходимость модели с заданным словарем W :

$$P(D) = \exp \left[-\frac{1}{N} \sum_{d \in D} \sum_{w \in d} n_{dw} \ln p(w|d) \right].$$

Перплексия коллекции D является мерой качества языка и часто используется в компьютерной лингвистике. В нашем случае язык – это распределение слов в документах $p(w|d)$. Чем меньше перплексия, тем меньше вероятность, что это распределение является случайным.

В целом представленный подход удовлетворяет рекомендациям для пользователей LDA, представленным в работе [18].

Дополнительным показателем, который мы использовали для оценки качества модели, является разнообразие, то есть энтропия распределения слов, характеризующих тему:

$$H_t = -\frac{1}{\ln n_w} \sum_i^{n_w} p_i(w_i) \ln p_i(w_i), \quad (1)$$

где n_w – количество слов, описывающих тему;

$p_i(w_i)$ – вес i -го слова в теме t .

Поскольку значение этой метрики нормализовано по количеству признаков (слов), ее возможные значения находятся в интервале $[0; 1]$. Значение 0 соответствует максимальной сфокусированности, когда тема описывается единственным словом. Значение 1 определяет ситуацию, когда все признаки, представленные в описании темы, имеют одинаковые веса, т.е. она не идентифицирована. Очевидно, что в более-менее пригодной для использования модели значения этой метрики должны быть небольшими и примерно одинаковыми для всех тем.

Когда оптимальное количество тем и соответствующее распределение тем для каждого документа найдено, мы можем изучать динамику тем. Пусть θ_{dt} – это вес темы t в документе d ($0 \leq \theta_{dt} \leq 1$). Таким

образом, общую популярность темы в корпусе документов можно определить следующим образом [10]:

$$\hat{\theta}_t = \frac{1}{|D|} \sum_{d \in D} \theta_{dt}. \quad (2)$$

Для вычисления популярности темы в заданный год y достаточно в формуле (2) положить $D = D_y$, где D_y – множество статей, опубликованных в год y .

Пусть C_d – количество цитирований документа d и $C = \sum_{d \in D} C_d$. Тогда влияние темы можно определить следующим образом [10]:

$$\hat{i}_t = \frac{1}{C} \sum_{d \in D} \theta_{dt} C_d. \quad (3)$$

Аналогично, чтобы вычислить влияние темы в год необходимо установить $D = D_y$ в уравнении (3).

3. Результаты и обсуждение

После выполнения всех операций предварительной обработки, описанных в предыдущем разделе, и 100 итераций байесовской оптимизации модели LDA, было найдено оптимальное количество тем,

равное 9. Соответствующее значение перплексии составило 568,75.

Анализируя доминирующие термины по темам (рисунк 2), можно сделать вывод, что каждая тема представляет собой некоторую согласованную область исследования. При этом веса тем в документах либо велики (т.е. документ сильно связан с темой), либо близки к нулю (т.е. документ не относится к данной теме).

Чтобы определить краткие наименования тем, мы проанализировали распределение терминов и наиболее репрезентативные статьи по каждой теме. Чтобы выбрать наиболее репрезентативные статьи, мы отсортировали публикации по весу темы, а затем по количеству цитирований, обе сортировки в порядке убывания. На рисунке 2 представлены присвоенные краткие наименования, а в таблице 2 приведено описание тем.

В таблице 2 также представлены значения разнообразия, популярности и влияния для каждой темы во всей коллекции D , рассчитанные по формулам (1), (2) и (3) соответственно. Поскольку сумма как



Рис. 2. Визуализация тематической модели с помощью облака слов (каждое облако представляет одну тему, размер шрифта слова соответствует его весу в данной теме)

Таблица 2.

Темы интеллектуального анализа данных

Тема	Комментарии	Разнообразие	Популярность	Влияние
Text mining	Обнаружение паттернов в текстах	0,779	0,107	0,110
General learning	Алгоритмы машинного обучения и связанные методы, такие как отбор признаков, разметка классов и др.	0,826	0,213	0,211
Segmentation	Методы разделения объектов: кластеризация, обнаружение выбросов и т.д.	0,777	0,084	0,080
Applications	Практическое использование методов DM	0,826	0,097	0,095
Data streams	Модели данных, зависящих от времени	0,805	0,097	0,102
Recommender systems	Рекомендательные системы	0,799	0,076	0,079
Pattern mining	Общие вопросы поиска корреляций между элементами данных	0,750	0,110	0,114
Network analysis	Обнаружение сообществ и потоков влияния в различных сетях	0,762	0,093	0,111
Human behavior analysis	Выявление и прогнозирование закономерностей в поведении людей: отток клиентов, сегментация рынка, мошенничество, угрозы безопасности и т.д.	0,844	0,121	0,096

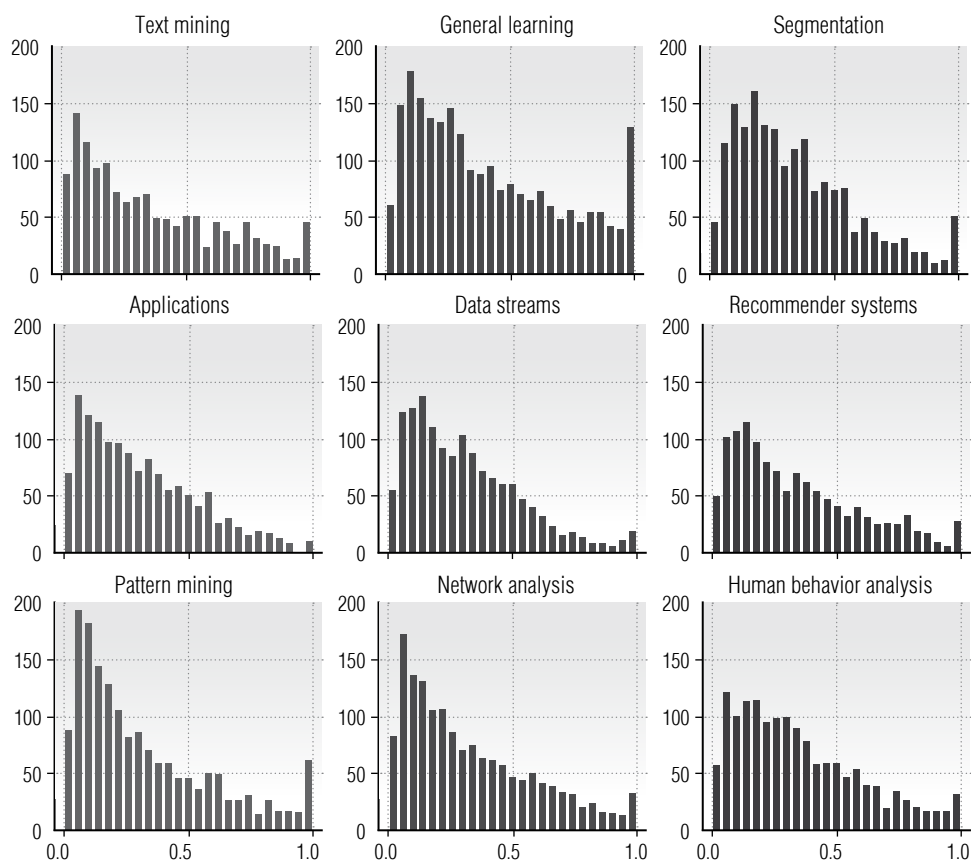


Рис. 3. Распределение весов тем по корпусу документов
(горизонтальная ось – веса тем, вертикальная – соответствующее количество документов)

популярности, так и воздействия равна единице, представленные значения можно рассматривать как долю определенной темы в общем потоке исследований интеллектуального анализа данных, то есть ее общий вес.

Дополнительную информацию о структуре DM можно получить из анализа распределения весов тем в корпусе документов (*рисунок 3*). Для более наглядного представления мы исключили из графика для каждой темы документы, в которых вес этой темы равен нулю.

Как следует из *таблицы 2*, тема, которая привлекала наибольшее внимание за последние 20 лет, — это общие вопросы обучения (general learning). На развитие этого направления потрачено более 20% усилий исследователей в области интеллектуального анализа данных ($\hat{\theta}_t = 0,213$). Работы в этой области охватывают широчайший спектр проблем машинного обучения, например, отбор признаков [19] (вес доминирующей темы в этом документе $\theta_{dt} = 0,974$), мульти-классовую классификацию с использованием ансамблей [20] ($\theta_{dt} = 0,963$), градиентные методы [21] ($\theta_{dt} = 0,925$) и т.д. Эти примеры статей выбраны, поскольку все они имеют большое количество цитирований (более 90, согласно WoS). Интересно, что работы с максимальным весом этой темы в основном посвящены методам на основе ядер (например, [22]) с $\theta_{dt} = 0,990$. Согласно *рисунок 3*, эта тема имеет вес, близкий к единице в наибольшем количестве документов (более 100). Эти статьи посвящены исключительно методам машинного обучения и не пересекаются с другими тематиками DM.

Мы определили вторую по важности тему ($\hat{\theta}_t = 0,121$) как «Анализ поведения человека» (human behavior analysis). Она фокусируется на обнаружении и прогнозировании закономерностей в деятельности групп людей и систем, на которые эти группы влияют. В этой области изучаются такие вопросы, как продвижение за счет снижения цен [23] ($\theta_{dt} = 0,988$), обнаружение подозрительных финансовых транзакций [24] ($\theta_{dt} = 0,985$), волатильность биткойна [25] ($\theta_{dt} = 0,985$) и другие. Отметим, что значительная часть этих работ представлена на семинарах, сопровождающих основную конференцию.

Следующая тема, — обнаружение ассоциаций (pattern mining), — фокусируется на извлечении ассоциаций (правил), то есть на задаче поиска корреляций между элементами в наборе данных. Исследователи изучают как практическое применение ассоциативных правил (например, данных о

рыночной корзине), так и общие принципы обнаружения связей в больших базах данных. С одной стороны, это может быть идентификация наиболее часто встречающихся комбинаций элементов данных [26] ($\theta_{dt} = 0,960$). С другой стороны, это может быть паттерн, состоящий из редких, но сильно коррелированных элементов [27] ($\theta_{dt} = 0,987$). Отметим также, что это наиболее сфокусированная тема с наименьшим значением H_t .

В наиболее представительных публикациях по теме анализа текстовой информации (text mining) рассматриваются такие вопросы, как идентификация и ранжирование авторов [28] ($\theta_{dt} = 0,974$), тематическое моделирование [29] ($\theta_{dt} = 0,969$) и кластеризация текста с использованием моделей на основе семантики [30] ($\theta_{dt} = 0,988$). Это область исследований с четкими границами, которая включает обнаружение закономерностей только в текстах и не рассматривает другие типы неструктурированных данных.

Направление исследований, который мы определили как потоки данных (data streams), касается моделей, зависящих от времени. Он включает в себя как более или менее традиционный анализ и прогнозирование временных рядов [31] ($\theta_{dt} = 0,981$), так и более экзотические модели, например, основанные на причинности по Грейнджеру [32] ($\theta_{dt} = 0,987$).

Тема «Приложения» (applications) объединяет работы, в основном посвященные практическому использованию методов интеллектуального анализа данных, которые не относятся к другим направлениям, выделенным выше. Примерами являются обнаружение событий с использованием со-локации мобильных пользователей [33] ($\theta_{dt} = 0,987$) и биометрическая модель безопасности для медицинского интернета вещей [34] ($\theta_{dt} = 0,983$).

Направление исследований «Сетевой анализ» (network analysis) посвящено моделям на основе графов, позволяющим восстановить пространственную структуру или топологию исследуемого объекта. Наиболее популярной темой такого рода исследований является обнаружение сообществ с использованием различных методов сетевого анализа [35] ($\theta_{dt} = 0,983$). Следующим вопросом, внимание к которому в последнее время возрастает, является анализ потоков влияния [36] ($\theta_{dt} = 0,985$), в том числе прогнозирование популярности сообщений в социальных сетях.

Мы определили следующую тему как разделение объектов (segmentation), поскольку она включает не

только широкий спектр алгоритмов кластеризации [37] ($\theta_{dt} = 0,977$), но и приложения, основанные на методах разделения объектов, например, обнаружение выбросов [38] ($\theta_{dt} = 0,981$). Согласно нашей модели, эта тема преобладает в работах, связанных с неструктурированными данными (изображения, видео, звук). Однако в большинстве случаев вес этой темы в таких приложениях не превышает веса других тем.

Наконец, последняя, но не менее важная тема, обнаруженная нашей моделью, — это рекомендательные системы (recommender systems). Это направление в дополнительных комментариях не нуждается. Стоит только отметить, что в последнее время исследователи уделяют особое внимание генерации интерпретируемых рекомендаций [39] ($\theta_{dt} = 0,986$).

Как следует из рисунка 3, помимо “general learning”, только в трех других областях существует относительно большое количество статей с весом темы близким к единице: это “segmentation”, “pattern mining” и “text mining”. Эти направления также относятся к машинному обучению, поэтому здесь публикуется относительно большое количество статей, посвященных исключительно алгоритмам. С другой стороны, такая тема, как “applications”, практически не содержит статей с весом темы близким к единице. Данная тема также имеет высокое значение метрики разнообразия (1). Этот факт можно объяснить тем, что в работах,

посвященных практическим приложениям DM, как правило, представляются и новые модификации алгоритмов.

Наша модель не выделяет искусственные нейронные сети (ИНС) как отдельное направление интеллектуального анализа данных. Это противоречит [7], но согласуется с работами [5, 6]. Согласно нашим результатам, публикации, в которых используются модели ИНС, чаще всего относятся к областям “general learning” и “segmentation”.

Следующий вопрос, который необходимо рассмотреть, это взаимодействие тем, которое можно рассматривать как частоту совместного появления тем в документах. Пусть θ_{di} и θ_{dj} — веса тем i и j в документе d . Таким образом, мы можем определить взаимодействие тем в этом документе как произведение $\theta_{di} \theta_{dj}$. Максимальное возможное значение взаимодействия двух тем в документе $\theta_{di} \theta_{dj} = 0,25$ при $\theta_{di} = \theta_{dj} = 0,5$. Отсюда максимальное возможное значение взаимодействия тем в коллекции документов $0,25 \cdot |D|$.

Таким образом, взаимодействие двух тем i и j по корпусу документов может быть вычислено как

$$c_{ij} = \sum_{d \in D} \theta_{di} \theta_{dj}.$$

Соответствующие данные представлены на рисунке 4. “General learning” можно рассматривать

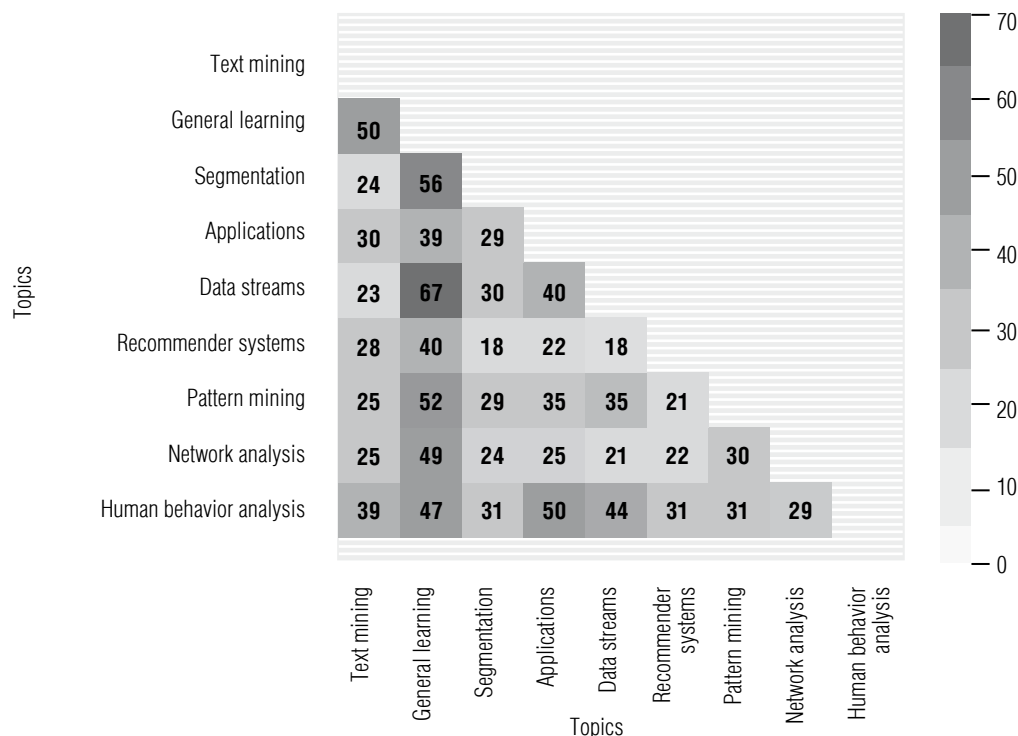


Рис. 4. Взаимодействие тем

как центральную тему, поскольку она наиболее тесно связана с другими областями исследований. “Human behavior analysis” также имеет относительно сильные связи с другими направлениями. “Recommender systems”, “network analysis” и “text mining” – более изолированные темы в нашей модели, поскольку они основаны на специализированных алгоритмах.

Следующий этап анализа – исследование динамики популярности и влияния выявленных тем. Популярность является производной от количества публикаций, а влияние рассчитывается с использованием количества цитирований. На *рисунке 5* представлена динамика популярности тем (сплошная линия) согласно выражению (2). Пунктирная линия показывает тренд. На *рисунке 6* представлены те же данные по влиянию тем, рассчитанному по формуле (3). Эти данные отражают дрейф интересов сообщества интеллектуального анализа данных в каждой области исследований.

Отметим, что популярность и влияние многих тем подвержены значительным флуктуациям. С одной стороны, это можно объяснить кратковременным смещением внимания исследователей. С другой стороны, ICDM, хотя и является одним из наиболее представительных форумов в области интеллектуального анализа данных, может не полностью отражать реальную динамику этой дисциплины. Например, формулировка названий треков конференции программным комитетом (call for papers) может повлиять на исследователей, представляющих работы. Однако мы считаем, что, несмотря на обнаруженные флуктуации, анализ публикаций за 2001–2019 гг. позволяет выявить глобальные тенденции, что и является целью нашего исследования.

Представленные данные показывают, что внимание исследователей к “pattern mining” резко снижается как с точки зрения популярности, так и воздействия. Такая же, но менее выраженная тенденция характерна и для “general learning” и “segmentation”.

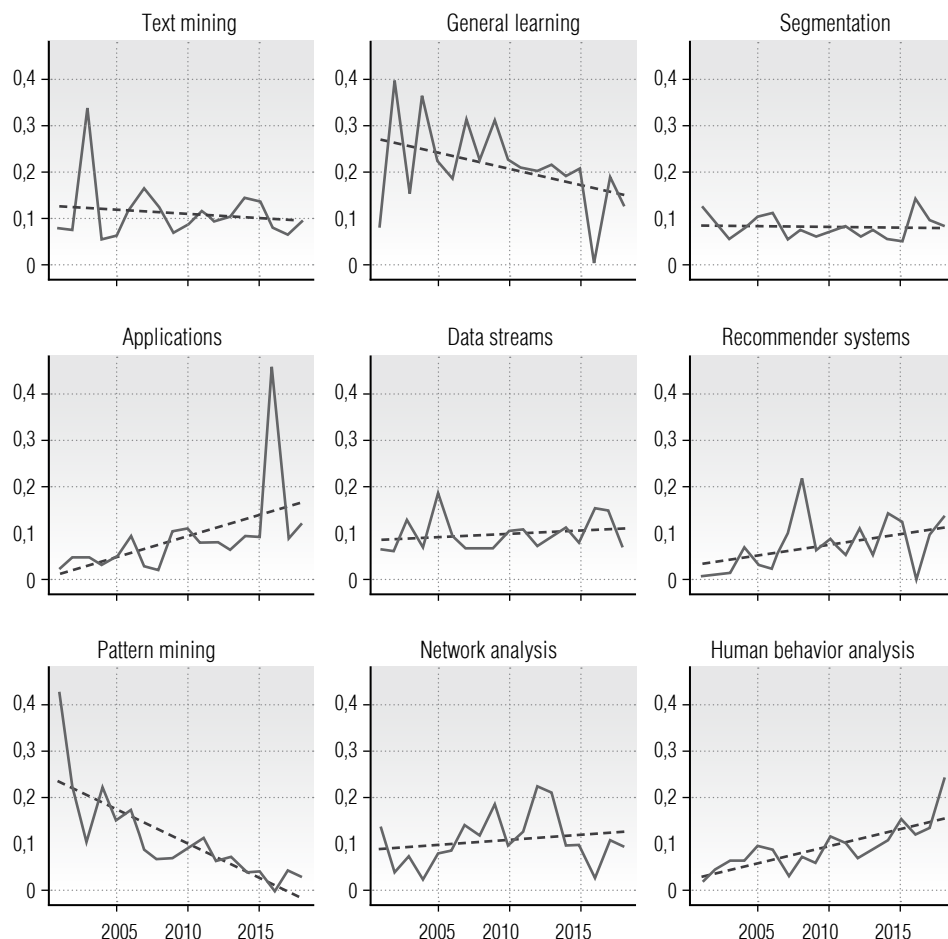


Рис. 5. Динамика популярности тем (сплошная линия) и тренд (пунктирная линия)

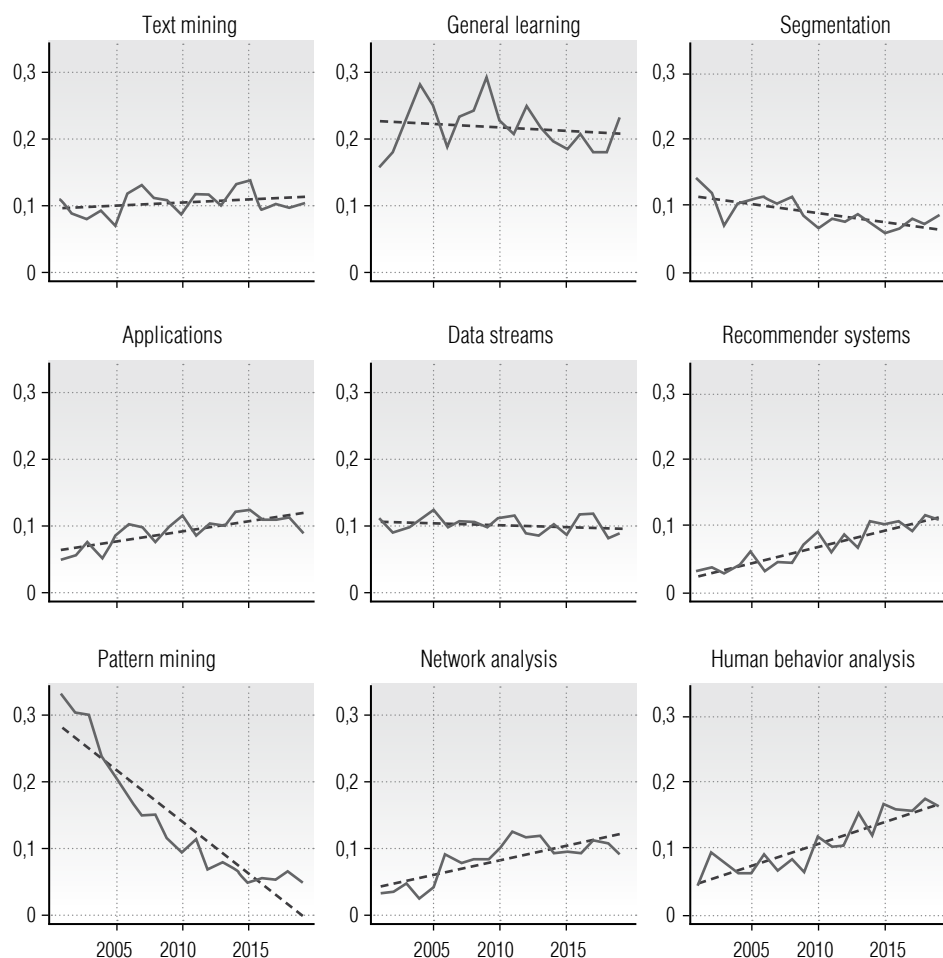


Рис. 6. Динамика влияния тем (сплошная линия) и тренд (пунктирная линия)

Эти тенденции требуют более глубокого изучения. Во-первых, можно предположить, что это связано с тем, что достижения в области алгоритмов машинного обучения и связанных с ними методов, включая анализ ассоциаций, уже настолько велики, что дальнейшее продвижение требует серьезных усилий. Сегодня наибольшая активность наблюдается в области глубокого обучения, но объем таких публикаций в трудах ICDM еще относительно невелик.

В то же время растет популярность исследований, связанных с рекомендательными системами (recommender systems), сетевым анализом (network analysis) и анализом поведения человека (human behavior analysis), что, скорее всего, связано с увеличением доступности данных и практической направленностью этих тем. Направление исследований, связанных с практическим применением интеллектуального анализа данных (applications), также имеет тенденцию к росту. Это также может объяснить снижение интереса к основополагаю-

щим алгоритмам, поскольку значительная часть исследовательского сообщества сосредоточена на более актуальных практических вопросах.

Последние две темы, “text mining” и “data streams”, вызывают относительно постоянный интерес исследователей. Представленные результаты проливают свет на структуру и динамику интеллектуального анализа данных за последние двадцать лет и позволяют расширить понимание этой научной дисциплины.

Следующий вопрос, который представляет интерес при анализе содержания публикаций, – это тематическое разнообразие документов. По аналогии с (1) мы можем определить разнообразие документа через энтропию его тем:

$$H_d = -\sum_i^T \theta_{di} \ln \theta_{di} ,$$

где θ_{di} – вес темы i в документе d ;

T – количество тем.

Разнообразие публикаций

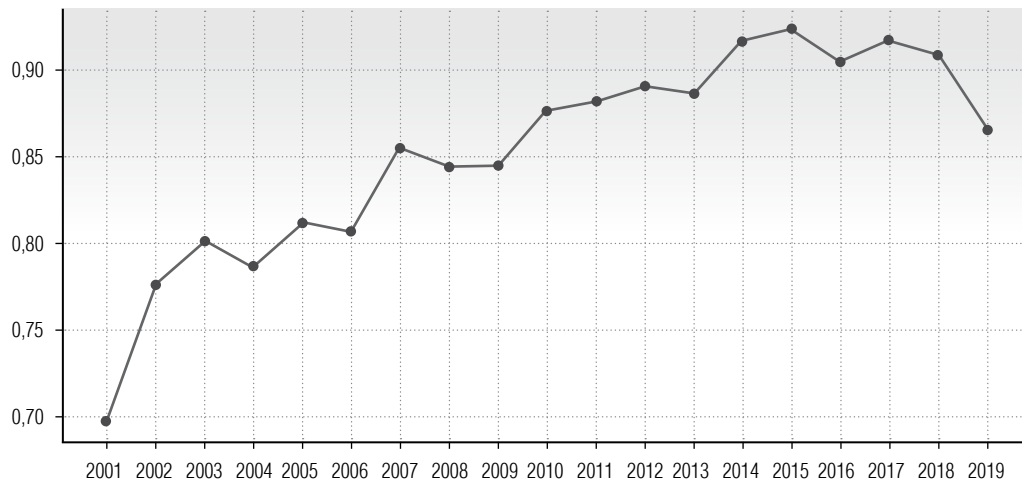


Рис. 7. Разнообразие публикаций ICDM в 2001–2019

На рисунке 7 представлено среднее разнообразие публикаций в трудах ICDM за 2001–2019 гг. Мы видим, что тематическое разнообразие документов неуклонно росло с момента первой конференции и достигло пика в 2015 году. За последние четыре года среднее количество тем, охватываемых одним документом, сократилось.

Мы полагаем, что это можно объяснить следующим образом. В начале 2000-х годов основной интерес исследователей был сосредоточен на алгоритмах обнаружения знаний, которые представлены списком критических направлений, выделенных в [5] и подтвержденных в [6] (таблица 1). По мере развития этих алгоритмов область их практического применения расширялась. Следовательно, и набор тем, освещаемых в одной научной публикации, становился все более широким. Это можно рассматривать как поиск в тематическом пространстве, пик которого пришелся на 2015 год. После 2015 года сформировалась новая исследовательская повестка дня. Как показано выше, алгоритмы общего обучения, а также связанные с ними области, такие как идентификация паттернов и сегментация, отодвигаются на второй план, хотя и продолжают играть важную роль. На первый план выходят более практические приложения, связанные с анализом человеческого поведения, системами рекомендаций, анализом сетевых сообществ и т.д.

В таблице 3 представлено сравнение тем интеллектуального анализа данных, выделенных в нашей работе и в статьях [5, 6]. Большинство тем 2010 года сосредоточено в направлении “general

learning”. Мы также включили в эту категорию приближенные множества (rough sets), поскольку данная теория применяется к классическим задачам извлечения знаний, например, для обнаружения закономерностей в неполных данных [40].

Таблица 3.

**Соответствие тем
интеллектуального анализа данных**

Темы DM в 2020 г. (настоящая работа)	Темы DM в 2005-2010 гг. [5, 6]
Text mining	Анализ связей (например, алгоритм PageRank)
General learning	Классификация
	Статистическое обучение
	Бэджинг и бустинг
	Интегрированный анализ
	Приближенные множества
Segmentation	Кластеризация
Applications	Не представлено
Data streams	Шаблоны последовательностей
Recommender systems	Не представлено
Pattern mining	Анализ ассоциаций
Network analysis	Анализ графов
Human behavior analysis	Не представлено

Новые темы, стремительный рост которых обнаружила наша модель, в 2010 году вообще не рассматривались. Отметим также, что выделенная

в нашей работе тема “text mining” включает гораздо больший спектр технологий и приложений, чем поиск взаимосвязей между документами.

В середине 2010-х годов исследователи в области экономики и управления зафиксировали рост интереса к принятию решений на основе данных (data driven decision-making, DDD) [41]. Это практика принятия решений, опирающаяся на анализ данных, а не исключительно на человеческие знания и интуицию. Авторы работы [42] сообщают, что использование DDD в производственном секторе США за период с 2005 по 2010 гг. почти утроилось (с 11% до 30% компаний). Более поздние исследования подтверждают возрастающую роль DDD как одной из лучших практик управления [43].

Определение DDD, которое приводится в литературе по менеджменту, полностью совпадает с определением “data mining”, рассмотренным в начале нашей работы. DDD также включает инжиниринг и процессинг данных, а также обнаружение полезных шаблонов. Однако, если DM рассматривает эту деятельность с технологической точки зрения, то DDD подходит к данному вопросу в контексте организационных процессов, включая деятельность, осуществляемую исключительно человеком. Тем не

менее, мы можем рассматривать растущий интерес к DDD как один из ключевых факторов, способствующих сдвигу в исследованиях DM, показанному на *рисунке 7*.

Заключение

Мы представили исследование интеллектуальной структуры “data mining” как научной дисциплины, проведенное с использованием тематического анализа. Такой подход позволил выделить девять основных направлений интеллектуального анализа данных и изучить их динамику.

Главный результат работы заключается в том, что мы обнаружили смещение интересов от алгоритмов машинного обучения к более практическим приложениям. По нашим данным, такая смена фокуса наметилась в середине 2010-х годов. Мы объясняем этот сдвиг сочетанием трех факторов. Во-первых, базовые алгоритмы интеллектуального анализа данных достигли высокого уровня зрелости. Во-вторых, с развитием социальных сетей стало доступно большое количество данных. В-третьих, существует устойчивый спрос со стороны бизнеса на принятие решений на основе данных. ■

Литература

1. Piatetsky-Shapiro G., Fayyad U. An introduction to SIGKDD and a reflection on the term ‘data mining’ // ACM SIGKDD Explorations Newsletter. 2012. Vol. 13. No 2. P. 102–103. DOI: 10.1145/2207243.2207269.
2. Witten I.H., Frank E., Hall M., Pal C. Data mining: Practical machine learning tools and techniques. Cambridge, MA: Morgan Kaufmann, 2017.
3. Han J., Kamber M., Pei J. Data mining: Concepts and techniques. Waltham, MA: Morgan Kaufmann, 2012. DOI: 10.1016/C2009-0-61819-5.
4. Rather B. Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data. Sound Parkway, NJ: CRC Press, 2011.
5. Yang Q., Wu X. 10 challenging problems in data mining research // International Journal of Information Technology & Decision Making. 2006. Vol. 5, No 4. P. 597–604. DOI: 10.1142/S0219622006002258.
6. Wu X. 10 years of data mining research: retrospect and prospect // Proceedings of the 10th IEEE International Conference on Data Mining (ICDM). Sydney, Australia, 13–17 December 2010. P. 7. DOI: 10.1109/ICDM.2010.172.
7. Liao S.H., Chu P.H., Hsiao P.Y. Data mining techniques and applications – A decade review from 2000 to 2011 // Expert Systems with Applications. 2012. Vol. 39. No 12. P. 11303–11311. DOI: 10.1016/j.eswa.2012.02.063.
8. Van Eck N.J., Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping // Scientometrics. 2009. Vol. 84. No 2. P. 523–538. DOI: 10.1007/s11192-009-0146-3.
9. Thilakaratne M., Falkner K., Atapattu T. A systematic review on literature-based discovery: general overview, methodology, & statistical analysis // ACM Computing Surveys. 2019. Vol. 52. No 6. Article no 129. DOI: 10.1145/3365756.
10. Zelenkov Y. The topic dynamics in knowledge management research // Proceedings of the 14th International Conference on Knowledge Management in Organizations (KMO 2019). Zamora, Spain, 15–18 July 2019. P. 324–335. DOI: 10.1007/978-3-030-21451-7_28.
11. Mann G.S., Mimno D., McCallum A. Bibliometric impact measures leveraging topic analysis // Proceedings of the 6th ACM/IEEE Joint Conference on Digital Libraries (JCDL '06). Chapel Hill, NC, USA, 11–15 June 2006. P. 65–74. DOI: 10.1145/1141753.1141765.
12. Dam H.K., Ghose A. Analyzing topics and trends in the PRIMA literature // Proceedings of the 19th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA 2016). Phuket, Thailand, 22–26 August 2016. P. 216–229. DOI: 10.1007/978-3-319-44832-9_13.
13. Dias L., Gerlach M., Scharloth J., Altman E.G. Using text analysis to quantify the similarity and evolution of scientific disciplines // Royal Society Open Science. 2018. Vol. 5. No 1. Article no 171545. DOI: 10.1098/rsos.171545.
14. Jensen S., Liu X., Yu Y., Milojevic S. Generation of topic evolution trees from heterogeneous bibliographic networks // Journal of Informetrics. 2016. Vol. 10. No 2. P. 606–621. DOI: 10.1016/j.joi.2016.04.002.

15. Syed S., Spruit M. Full-text or abstract? Examining topic coherence scores using latent Dirichlet allocation // Proceedings of the 4th IEEE International Conference on Data Science and Advanced Analytics (DSAA 2017). Tokyo, Japan, 19–21 October 2017. P. 165–174. DOI: 10.1109/DSAA.2017.61.
16. Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet allocation // Journal of Machine Learning Research. 2003. No 3. P. 993–1022.
17. Mockus J. Bayesian approach to global optimization: Theory and applications. Heidelberg: Springer, 2012. DOI: 10.1007/978-94-009-0909-0.
18. Understanding the limiting factors of topic modeling via posterior contraction analysis / J. Tang [et al.] // Proceedings of Machine Learning Research. 2014. Vol. 32. No 1. P. 190–198.
19. Molina L.C., Belanche L., Nebot A. Feature selection algorithms – A survey and experimental evaluation // Proceedings of the 2nd IEEE International Conference on Data Mining (ICDM 2002). Maebashi City, Japan, 9–12 December 2002. P. 306–313. DOI: 10.1109/ICDM.2002.1183917.
20. Read J., Pfahringer B., Holmes G. Multi-label classification using ensembles of pruned sets // Proceedings of the 8th IEEE International Conference on Data Mining (ICDM). Pisa, Italy, 15–19 December 2008. P. 995–1000. DOI: 10.1109/ICDM.2008.74.
21. Chen X., Pan W., Kwok J.T., Carbonell J.G. Accelerated gradient method for multi-task sparse learning problem // Proceedings of the 9th IEEE International Conference on Data Mining (ICDM). Miami Beach, FL, USA, 6–9 December 2009. P. 746–751. DOI: 10.1109/ICDM.2009.128.
22. Shin K. Partitionable kernels for mapping kernels // Proceedings of the 11th IEEE International Conference on Data Mining (ICDM). Vancouver, BC, Canada, 11–14 December 2011. P. 645–654. DOI: 10.1109/ICDM.2011.115.
23. Li Z., Yada K. Why do retailers end price promotions – A study on duration and profit effects of promotion // Proceedings of the 15th IEEE International Conference on Data Mining Workshop (ICDMW). Atlantic City, NJ, USA, 14–17 November 2015. P. 328–335. DOI: 10.1109/ICDMW.2015.56.
24. Camino R.D., State R., Montero L., Valtchev P. Finding suspicious activities in financial transactions and distributed ledgers // Proceedings of the 17th IEEE International Conference on Data Mining Workshops (ICDMW). New Orleans, LA, USA, 18–21 November 2017. P. 787–796. DOI: 10.1109/ICDMW.2017.109.
25. Guo T., Bifet A., Antulov-Fantulin N. Bitcoin volatility forecasting with a glimpse into buy and sell orders // Proceedings of the 18th IEEE International Conference on Data Mining (ICDM). Singapore, 17–20 November 2018. P. 989–994. DOI: 10.1109/ICDM.2018.00123.
26. Gouda K., Zaki M.J. Efficiently mining maximal frequent itemsets // Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM). San Jose, CA, USA, 29 November – 2 December 2001. P. 163–170. DOI: 10.1109/ICDM.2001.989514.
27. Ma S., Hellerstein J.L. Mining mutually dependent patterns // Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM). San Jose, CA, USA, 29 November – 2 December 2001. P. 409–416. DOI: 10.1109/ICDM.2001.989546.
28. Zhou D., Orshanskiy S.A., Zha H., Gees C.L. Co-ranking authors and documents in a heterogeneous network // Proceedings of the 7th IEEE International Conference on Data Mining (ICDM). Omaha, NE, USA, 28–31 October 2007. P. 739–744. DOI: 10.1109/ICDM.2007.57.
29. Tang J., Jin R., Zang J. A topic modeling approach and its integration into the random walk framework for academic search // Proceedings of the 8th IEEE International Conference on Data Mining (ICDM). Pisa, Italy, 15–19 December 2008. P. 1055–1060. DOI: 10.1109/ICDM.2008.71.
30. Shehata S., Karray F., Kamel M. Enhancing text clustering using concept-based mining model // Proceedings of the 6th IEEE International Conference on Data Mining (ICDM). Hong Kong, China, 18–22 December 2006. P. 1043–1048. DOI: 10.1109/ICDM.2006.64.
31. Yang D., Li B., Rettig L., Cudre-Mauroux P. HistoSketch: Fast similarity-preserving sketching of streaming histograms with concept drift // Proceedings of the 17th IEEE International Conference on Data Mining (ICDM). New Orleans, LA, USA, 18–21 November 2017. P. 545–554. DOI: 10.1109/ICDM.2017.64.
32. Dhurandhar A. Learning maximum lag for grouped graphical Granger models // Proceedings of the 10th IEEE International Conference on Data Mining Workshops (ICDMW). Sydney, Australia, 13 December 2010. P. 217–224. DOI: 10.1109/ICDMW.2010.9.
33. Wang H., Li Z., Lee W.-C. PGT: Measuring mobility relationship using personal, global and temporal factors // Proceedings of the 14th IEEE International Conference on Data Mining (ICDM). Shenzhen, China, 14–17 December 2014. P. 570–579. DOI: 10.1109/ICDM.2014.111.
34. Pirbhulal S., Wu W., Li G. A biometric security model for wearable healthcare // Proceedings of the 18th IEEE International Conference on Data Mining Workshops (ICDMW). Singapore, 17–20 November 2018. P. 136–143. DOI: 10.1109/ICDMW.2018.00026.
35. Yang J., Leskovec J. Defining and evaluating network communities based on ground-truth // Proceedings of the 12th IEEE International Conference on Data Mining (ICDM). Brussels, Belgium, 10–13 December 2012. P. 745–754. DOI: 10.1109/ICDM.2012.138.
36. Shi L., Tong H., Tang J., Lin C. Flow-based influence graph visual summarization // Proceedings of the 14th IEEE International Conference on Data Mining (ICDM). Shenzhen, China, 14–17 December 2014. P. 983–988. DOI: 10.1109/ICDM.2014.128.
37. Hung M.-C., Yang D.-L. An efficient fuzzy c-means clustering algorithm // Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM). San Jose, CA, USA, 29 November – 2 December 2001. P. 225–232. DOI: 10.1109/ICDM.2001.989523.
38. Pei Y., Zaiane O.R., Gao Y. An efficient reference-based approach to outlier detection in large datasets // Proceedings of the 6th IEEE International Conference on Data Mining (ICDM). Hong Kong, China, 18–22 December 2006. P. 478–487. DOI: 10.1109/ICDM.2006.17.
39. A reinforcement learning framework for explainable recommendation / X. Wang [et al.] // Proceedings of the 18th IEEE International Conference on Data Mining (ICDM). Singapore, 17–20 November 2018. P. 587–596. DOI: 10.1109/ICDM.2018.00074.
40. Wang H., Wang S. Discovering patterns of missing data in survey databases: an application of rough sets // Expert Systems with Applications. 2009. Vol. 36. No 3. Part 2. P. 6256–6260. DOI: 10.1016/j.eswa.2008.07.010.

41. Provost F., Fawcett T. Data science and its relationship to big data and data-driven decision making // *Big Data*. 2013. Vol. 1. No 1. P. 51–59. DOI: 10.1089/big.2013.1508.
42. Brynjolfsson E., McElheran K. The rapid adoption of data-driven decision-making // *American Economic Review*. 2016. Vol. 106. No 5. P. 133–139. DOI: 10.1257/aer.p20161016.
43. Song P., Zheng C., Zhang C., Yu X. Data analytics and firm performance: An empirical study in an online B2C platform // *Information & Management*. 2018. Vol. 55, No 5. P. 633–642. DOI: 10.1016/j.im.2018.01.004.

Об авторах

Зеленков Юрий Александрович

доктор технических наук;

профессор департамента бизнес-информатики, Высшая школа бизнеса, Национальный исследовательский университет «Высшая школа экономики», 101000, г. Москва, ул. Мясницкая, д. 20;

E-mail: yzelenkov@hse.ru

ORCID: 0000-0002-2248-1023

Анисичкина Екатерина Алексеевна

студентка бакалавриата, образовательная программа «Бизнес-информатика», Высшая школа бизнеса,

Национальный исследовательский университет «Высшая школа экономики», 101000, г. Москва, ул. Мясницкая, д. 20;

E-mail: eaanisichkina@edu.hse.ru

Trends in data mining research: A two-decade review using topic analysis

Yuri A. Zelenkov

E-mail: yzelenkov@hse.ru

Ekaterina A. Anisichkina

E-mail: eaanisichkina@edu.hse.ru

National Research University Higher School of Economics

Address: 20, Myasnitskaya Street, Moscow 101000, Russia

Abstract

This work analyses the intellectual structure of data mining as a scientific discipline. To do this, we use topic analysis (namely, latent Dirichlet allocation, DLA) applied to the proceedings of the International Conference on Data Mining (ICDM) for 2001–2019. Using this technique, we identified the nine most significant research flows. For each topic, we analyse the dynamics of its popularity (number of publications) and influence (number of citations). The central topic, which unites all other direction, is General Learning, which includes machine learning algorithms. About 20% of the research efforts were spent on the development of this direction for the entire time under review, however, its influence has declined most recently. The analysis also showed that attention to topics such as Pattern Mining (detecting associations) and Segmentation (object separation algorithms such as clustering) is decreasing. At the same time, the popularity of research related to Recommender Systems, Network Analysis, and Human Behaviour Analysis is growing, which is most likely due to the increasing availability of data and the practical value of these topics. The research direction related to practical Applications of data mining also shows a tendency to grow. The last two topics, Text Mining and Data Streams have attracted steady interest from researchers. The results presented here shed light on the

structure and trends of data mining over the past twenty years and allow us to expand our understanding of this scientific discipline. We can argue that in the last five years a new research agenda has been formed, which is characterized by a shift in interest from algorithms to practical applications that affect all aspects of human activity.

Key words: data mining topics, topic analysis, scientometrics.

Citation: Zelenkov Yu.A., Anisichkina E.A. (2021) Trends in data mining research: A two-decade review using topic analysis. *Business Informatics*, vol. 15, no 1, pp. 30–46. DOI: 10.17323/2587-814X.2021.1.30.46

References

1. Piatetsky-Shapiro G., Fayyad U. (2012) An introduction to SIGKDD and a reflection on the term 'data mining'. *ACM SIGKDD Explorations Newsletter*, vol. 13, no 2, pp. 102–103. DOI: 10.1145/2207243.2207269.
2. Witten I.H., Frank E., Hall M., Pal C. (2017) *Data mining: Practical machine learning tools and techniques*. Cambridge, MA: Morgan Kaufmann.
3. Han J., Kamber M., Pei J. (2012) *Data mining: Concepts and techniques*. Waltham, MA: Morgan Kaufmann. DOI: 10.1016/C2009-0-61819-5.
4. Rather B. (2011) *Statistical and machine-learning data mining: Techniques for better predictive modeling and analysis of big data*. Sound Parkway, NW: CRC Press.
5. Yang Q., Wu X. (2006) 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making*, vol. 5, no 4, pp. 597–604. DOI: 10.1142/S0219622006002258.
6. Wu X. (2010) 10 years of data mining research: retrospect and prospect. Proceedings of the *10th IEEE International Conference on Data Mining (ICDM)*, Sydney, Australia, 13–17 December 2010, p. 7. DOI: 10.1109/ICDM.2010.172.
7. Liao S.H., Chu P.H., Hsiao P.Y. (2012) Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications*, vol. 39, no 12, pp. 11303–11311. DOI: 10.1016/j.eswa.2012.02.063.
8. Van Eck N.J., Waltman L. (2009) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, vol. 84, no 2, pp. 523–538. DOI: 10.1007/s11192-009-0146-3.
9. Thilakarathne M., Falkner K., Atapattu T. (2019) A systematic review on literature-based discovery: general overview, methodology, & statistical analysis. *ACM Computing Surveys*, vol. 52, no 6, article no 129. DOI: 10.1145/3365756.
10. Zelenkov Y. (2019) The topic dynamics in knowledge management research. Proceedings of the *14th International Conference on Knowledge Management in Organizations (KMO 2019)*, Zamora, Spain, 15–18 July 2019, pp. 324–335. DOI: 10.1007/978-3-030-21451-7_28.
11. Mann G.S., Mimno D., McCallum A. (2006) Bibliometric impact measures leveraging topic analysis. Proceedings of the *6th ACM/IEEE Joint Conference on Digital Libraries (JCDL '06)*, Chapel Hill, NC, USA, 11–15 June 2006, pp. 65–74. DOI: 10.1145/1141753.1141765.
12. Dam H.K., Ghose A. (2016) Analyzing topics and trends in the PRIMA literature. Proceedings of the *19th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA 2016)*, Phuket, Thailand, 22–26 August 2016, pp. 216–229. DOI: 10.1007/978-3-319-44832-9_13.
13. Dias L., Gerlach M., Scharloth J., Altman E.G. (2018) Using text analysis to quantify the similarity and evolution of scientific disciplines. *Royal Society Open Science*, vol. 5, no 1, article no 171545. DOI: 10.1098/rsos.171545.
14. Jensen S., Liu X., Yu Y., Milojevic S. (2016) Generation of topic evolution trees from heterogeneous bibliographic networks. *Journal of Informetrics*, vol. 10, no 2, pp. 606–621. DOI: 10.1016/j.joi.2016.04.002.
15. Syed S., Spruit M. (2017) Full-text or abstract? Examining topic coherence scores using latent Dirichlet allocation. Proceedings of the *4th IEEE International Conference on Data Science and Advanced Analytics (DSAA 2017)*, Tokyo, Japan, 19–21 October 2017, pp. 165–174. DOI: 10.1109/DSAA.2017.61.
16. Blei D.M., Ng A.Y., Jordan M.I. (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research*, no 3, pp. 993–1022.
17. Mockus J. (2012) *Bayesian approach to global optimization: Theory and applications*. Heidelberg: Springer. DOI: 10.1007/978-94-009-0909-0.
18. Tang J., Meng Z., Nguyen X., Mei Q., Zhang M. (2014) Understanding the limiting factors of topic modeling via posterior contraction analysis. *Proceedings of Machine Learning Research*, vol. 32, no 1, pp. 190–198.
19. Molina L.C., Belanche L., Nebot A. (2002) Feature selection algorithms – A survey and experimental evaluation. Proceedings of the *2nd IEEE International Conference on Data Mining (ICDM 2002)*, Maebashi City, Japan, 9–12 December 2002, pp. 306–313. DOI: 10.1109/ICDM.2002.1183917.
20. Read J., Pfahringer B., Holmes G. (2008) Multi-label classification using ensembles of pruned sets. Proceedings of the *8th IEEE International Conference on Data Mining (ICDM)*, Pisa, Italy, 15–19 December 2008, pp. 995–1000. DOI: 10.1109/ICDM.2008.74.
21. Chen X., Pan W., Kwok J.T., Carbonell J.G. (2009) Accelerated gradient method for multi-task sparse learning problem. Proceedings of the *9th IEEE International Conference on Data Mining (ICDM)*, Miami Beach, FL, USA, 6–9 December 2009, pp. 746–751. DOI: 10.1109/ICDM.2009.128.
22. Shin K. (2011) Partitionable kernels for mapping kernels. Proceedings of the *11th IEEE International Conference on Data Mining (ICDM)*, Vancouver, BC, Canada, 11–14 December 2011, pp. 645–654. DOI: 10.1109/ICDM.2011.115.
23. Li Z., Yada K. (2015) Why do retailers end price promotions – A study on duration and profit effects of promotion. Proceedings of the *15th IEEE International Conference on Data Mining Workshop (ICDMW)*, Atlantic City, NJ, USA, 14–17 November 2015, pp. 328–335. DOI: 10.1109/ICDMW.2015.56.

24. Camino R.D., State R., Montero L., Valtchev P. (2017) Finding suspicious activities in financial transactions and distributed ledgers. Proceedings of the *17th IEEE International Conference on Data Mining Workshops (ICDMW)*, New Orleans, LA, USA, 18–21 November 2017, pp. 787–796. DOI: 10.1109/ICDMW.2017.109.
25. Guo T., Bifet A., Antulov-Fantulin N. (2018) Bitcoin volatility forecasting with a glimpse into buy and sell orders. Proceedings of the *18th IEEE International Conference on Data Mining (ICDM)*, Singapore, 17–20 November 2018, pp. 989–994. DOI: 10.1109/ICDM.2018.00123.
26. Gouda K., Zaki M.J. (2001) Efficiently mining maximal frequent itemsets. Proceedings of the *2001 IEEE International Conference on Data Mining (ICDM)*, San Jose, CA, USA, 29 November – 2 December 2001, pp. 163–170. DOI: 10.1109/ICDM.2001.989514.
27. Ma S., Hellerstein J.L. (2001) Mining mutually dependent patterns. Proceedings of the *2001 IEEE International Conference on Data Mining (ICDM)*, San Jose, CA, USA, 29 November – 2 December 2001, pp. 409–416. DOI: 10.1109/ICDM.2001.989546.
28. Zhou D., Orshanskiy S.A., Zha H., Gees C.L. (2007) Co-ranking authors and documents in a heterogeneous network. Proceedings of the *7th IEEE International Conference on Data Mining (ICDM)*, Omaha, NE, USA, 28–31 October 2007, pp. 739–744. DOI: 10.1109/ICDM.2007.57.
29. Tang J., Jin R., Zang J. (2008) A topic modeling approach and its integration into the random walk framework for academic search. Proceedings of the *8th IEEE International Conference on Data Mining (ICDM)*, Pisa, Italy, 15–19 December 2008, pp. 1055–1060. DOI: 10.1109/ICDM.2008.71.
30. Shehata S., Karray F., Kamel M. (2006) Enhancing text clustering using concept-based mining model. Proceedings of the *6th IEEE International Conference on Data Mining (ICDM)*, Hong Kong, China, 18–22 December 2006, pp. 1043–1048. DOI: 10.1109/ICDM.2006.64.
31. Yang D., Li B., Rettig L., Cudre-Mauroux P. (2017) HistoSketch: Fast similarity-preserving sketching of streaming histograms with concept drift. Proceedings of the *17th IEEE International Conference on Data Mining (ICDM)*, New Orleans, LA, USA, 18–21 November 2017, pp. 545–554. DOI: 10.1109/ICDM.2017.64.
32. Dhurandhar A. (2010) Learning maximum lag for grouped graphical Granger models. Proceedings of the *10th IEEE International Conference on Data Mining Workshops (ICDMW)*, Sydney, Australia, 13 December 2010, pp. 217–224. DOI: 10.1109/ICDMW.2010.9.
33. Wang H., Li Z., Lee W.-C. (2014) PGT: Measuring mobility relationship using personal, global and temporal factors. Proceedings of the *14th IEEE International Conference on Data Mining (ICDM)*, Shenzhen, China, 14–17 December 2014, pp. 570–579. DOI: 10.1109/ICDM.2014.111.
34. Pirbhulal S., Wu W., Li G. (2018) A biometric security model for wearable healthcare. Proceedings of the *18th IEEE International Conference on Data Mining Workshops (ICDMW)*, Singapore, 17–20 November 2018, pp. 136–143. DOI: 10.1109/ICDMW.2018.00026.
35. Yang J., Leskovec J. (2012) Defining and evaluating network communities based on ground-truth. Proceedings of the *12th IEEE International Conference on Data Mining (ICDM)*, Brussels, Belgium, 10–13 December 2012, pp. 745–754. DOI: 10.1109/ICDM.2012.138.
36. Shi L., Tong H., Tang J., Lin C. (2014) Flow-based influence graph visual summarization. Proceedings of the *14th IEEE International Conference on Data Mining (ICDM)*, Shenzhen, China, 14–17 December 2014, pp. 983–988. DOI: 10.1109/ICDM.2014.128.
37. Hung M.-C., Yang D.-L. (2001) An efficient fuzzy c-means clustering algorithm. Proceedings of the *2001 IEEE International Conference on Data Mining (ICDM)*, San Jose, CA, USA, 29 November – 2 December 2001, pp. 225–232. DOI: 10.1109/ICDM.2001.989523.
38. Pei Y., Zaiane O.R., Gao Y. (2006) An efficient reference-based approach to outlier detection in large datasets. Proceedings of the *6th IEEE International Conference on Data Mining (ICDM)*, Hong Kong, China, 18–22 December 2006, pp. 478–487. DOI: 10.1109/ICDM.2006.17.
39. Wang X., Chen Y., Yang J., Wu L., Wu Z., Xie X. (2018) A reinforcement learning framework for explainable recommendation. Proceedings of the *18th IEEE International Conference on Data Mining (ICDM)*, Singapore, 17–20 November 2018, pp. 587–596. DOI: 10.1109/ICDM.2018.00074.
40. Wang H., Wang S. (2009) Discovering patterns of missing data in survey databases: an application of rough sets. *Expert Systems with Applications*, vol. 36, no 3, part 2, pp. 6256–6260. DOI: 10.1016/j.eswa.2008.07.010.
41. Provost F., Fawcett T. (2013) Data science and its relationship to big data and data-driven decision making. *Big Data*, vol. 1, no 1, pp. 51–59. DOI: 10.1089/big.2013.1508.
42. Brynjolfsson E., McElheran K. (2016) The rapid adoption of data-driven decision-making. *American Economic Review*, vol. 106, no 5, pp. 133–139. DOI: 10.1257/aer.p20161016.
43. Song P., Zheng C., Zhang C., Yu X. (2018). Data analytics and firm performance: An empirical study in an online B2C platform. *Information & Management*, vol. 55, no 5, pp. 633–642. DOI: 10.1016/j.im.2018.01.004.

About the authors

Yury A. Zelenkov

Dr. Sci. (Tech.);

Professor, Department of Business Informatics, Graduate School of Business, National Research University Higher School of Economics, 20, Myasnitskaya Street, Moscow 101000, Russia;

E-mail: yzelenkov@hse.ru

ORCID: 0000-0002-2248-1023

Ekaterina A. Anisichkina

Student, BSc Program “Business Informatics”, Graduate School of Business, National Research University Higher School of Economics, 20, Myasnitskaya Street, Moscow 101000, Russia;

E-mail: eaanisichkina@edu.hse.ru

Статистическая устойчивость цифровой организации

В.И. Ананьин^a

E-mail: v.ananiin@gmail.com

К.В. Зимин^b

E-mail: konst.zimin@gmail.com

М.И. Лугачев^c 

E-mail: mil@econ.msu.ru

Р.Д. Гимранов^d

E-mail: gimranov_rd@mail.ru

^a Российская академия народного хозяйства и государственной службы при Президенте РФ
Адрес: 119571, г. Москва, проспект Вернадского, д. 82, стр. 1

^b Союз директоров по ИТ России
Адрес: 123056, Москва, ул. Селезневская, д. 34

^c Московский государственный университет им. М.В. Ломоносова
Адрес: 119991, г. Москва, ГСП-1, Ленинские горы, д. 1, стр. 46

^d ПАО «Сургутнефтегаз»
Адрес: 628415, г. Сургут, ул. Григория Кукуевецкого, д. 1, корп. 1

Аннотация

Важным свойством цифровой организации является ее способность к быстрым изменениям. Для того чтобы организация сохраняла способность к быстрым изменениям, она должна находиться на грани устойчивости, поскольку устойчивая организация всегда сопротивляется изменениям. В статье рассматривается пограничное состояние организации, находящееся на грани ее устойчивости и неустойчивости. В этом состоянии организация начинает терять предсказуемость в деталях поведения, но еще сохраняет предсказуемость в целом. Такое пограничное состояние авторы назвали статистической устойчивостью организации. Явление статистической устойчивости организации очень похоже на свойство стабильности частотности массовых событий и средних величин, описываемых в математической статистике аналогичным термином. Для анализа природы статистической устойчивости организации авторы использовали идеи странных аттракторов и режимов с обострением из теории сложных систем. Странный аттрактор — это такая область поведения организации, которая вне этой области для организации является областью устойчивости, а внутри — областью полной непредсказуемости. Теория сложных систем показала, что именно в областях странных аттракторов создаются условия изменчивости систем, а теория режимов с обострением показывает условия, при которых эта изменчивость может привести к самоорганизации, то есть спонтанному появлению новых структур. В

статье показано, что систематическая цифровизация объективно приводит к формированию статистической устойчивости организации и создает предпосылки для поддержания способности организации к быстрым изменениям. В традиционном менеджменте статистическая устойчивость организации рассматривается как угроза и источник рисков. Поэтому в условиях систематической цифровизации традиционные подходы к менеджменту должны быть существенно уточнены.

Ключевые слова: цифровая организация; изменчивость; устойчивость/неустойчивость организации; экспоненциальная организация; самоорганизация; странный аттрактор; разнообразие; рассеяние; цифровая трансформация; статистическая устойчивость.

Цитирование: Ананьин В.И., Зимин К.В., Лугачев М.И., Гимранов Р.Д. Статистическая устойчивость цифровой организации // Бизнес-информатика. 2021. Т. 15. № 1. С. 47–58. DOI: 10.17323/2587-814X.2021.1.47.58

Введение

В первой половине 2000-х годов в США появились успешные стартапы, построенные на новых информационных технологиях. Бизнес-модели этих организаций радикально отличались от существующих, так как были построены на совместном использовании данных. Такие организации получили название цифровых. Многие цифровые организации обладали удивительным свойством: их бизнес имел взрывной рост и хорошо описывался экспоненциальной кривой. Такие быстро растущие цифровые организации получили название экспоненциальных организаций [1]. Их феноменальные успехи во второй половине 2000-х годов инициировали волну создания новых цифровых организаций. Включение в эту волну мировых лидеров бизнеса запустило мощную тенденцию цифровой трансформации самих крупных компаний. Волна формирования цифровых организаций приобрела грандиозный масштаб, когда в середине 2010-х годов многие государства взяли курс на построение цифровой экономики и создали мощные стимулы для проведения цифровой трансформации в своих национальных предприятиях [2–5].

К 2020 году образовалось большое количество цифровых организаций и организаций, находящихся в процессе цифровой трансформации. Изучение природы экспоненциальных организаций началось в начале 2010-х годов, и уже в 2014 году в США вышла монументальная работа коллектива авторов Университета сингулярности по анализу закономерностей формирования экспоненциальных организаций [1]. В этой работе авторы показали, что всем экспоненциальным организациям удалось открыть не только новые формы организационного, человеческого и информационного

капитала, но и новые формы их синтеза [6]. Авторы также показали, что в этих организациях радикально сократилось время как принятия решений, так и их реализации. Фактически, экспоненциальные организации начали трансформироваться в сторону предприятия реального времени (real time enterprise, RTE) [7], в котором решения принимаются и реализуются со скоростью развития складывающихся управленческих ситуаций. В этом случае бизнес организации получает на рынке мощные конкурентные преимущества за счет того, что быстрее других находит решение и адаптируется к новым изменившимся условиям.

Лидерства важно не только достигнуть, но и суметь его удержать. Организация будет иметь на рынке конкурентные преимущества до того момента, пока ее конкуренты или партнеры не осваивают аналогичные практики принятия решений. Тогда скорость изменения ситуаций на рынке будет определяться не скоростью каких-то экономических рыночных факторов, а скоростью принятия решений и изменений конкурентов или даже партнеров. Если вдруг окажется, что наши конкуренты или партнеры меняются быстрее, чем мы сами, то в один момент наш бизнес теряет лидерство: для конкурентов мы становимся легкой мишенью, а для партнеров — обузой.

Цифровизация создает поддержку принятия не только краткосрочных и среднесрочных решений, но и стратегических долгосрочных решений, когда нужно распознать потенциального «черного лебедя» [8] и успеть подготовиться к встрече с ним. Эта проблема сильно волнует всех высших руководителей компаний — мировых лидеров. Питеру Диямандису в предисловии к книге [1] удалось емко ее сформулировать: «Сегодня вы вынуждены конкурировать не только и не столько с устоявшими-

ся мультинациональными корпорациями. Вашим конкурентом может стать любой парень из Кремниевой долины или округа Бандра в Мумбае, который сидит в своем гараже и использует новейшие онлайн-инструменты для разработки и распространения своих инновационных продуктов» [1]. Есть «черные лебеди», которые приходят вообще безо всяких предвестников, например, мировая пандемия коронавируса COVID-19 в 2020 году¹. В этом случае способность быстро меняться и адаптироваться к новым условиям становится для бизнеса уже просто вопросом жизни или смерти.

Если для бизнеса быстрая адаптивность, то есть способность быстро измениться и приспособиться к новым условиям, так важна, то почему так мало компаний, способных это делать? Какого нового качества не хватает бизнесу, чтобы информационные технологии стали драйвером его реальной трансформации? Наш взгляд, это новое качество — неустойчивость организации. Именно анализу неустойчивости организации посвящена данная статья.

1. Неустойчивость организации как угроза и возможность

Неустойчивость прочно ассоциируется с кризисом организации и поэтому в менеджменте традиционно пользуется дурной славой. Неустойчивость грозит потерей предсказуемости поведения организации и потерей ее управляемости. Менеджмент чувствует в ней угрозу и всеми силами с ней борется. Для управления рисками и антикризисного управления неустойчивость — это маргинальное состояние организации, которого надо либо избежать, либо быстро ликвидировать последствия разразившегося в организации кризиса [9–11].

Тем не менее, неустойчивость может оказаться полезной, когда менеджмент хочет проводить быстрые изменения в своей организации. Чем более устойчива организация, тем больше она будет сопротивляться изменениям, и тем меньше будет возможностей для проведения быстрых и масштабных изменений. И это не чей-то злой умысел, а объективное проявление все той же устойчивости. Для того, чтобы проводить такие изменения, организация должна быть выведена на грань устойчивости, когда она еще окончательно не потеряла

предсказуемости своего поведения. Тот, кто освоил это искусство, получает мощное конкурентное преимущество — высокую степень адаптивности и инновационность своего бизнеса.

Как мы видим, неустойчивость несет не только угрозы, но и возможности. Это хорошо можно наблюдать как в технических, так и природных системах. В инженерной практике хорошо известно, что высокий уровень маневренности современного истребителя достигается тем, что в его конструкцию изначально заложена аэродинамическая неустойчивость. Предсказуемость управления истребителем обеспечивается работой бортовой ЭВМ системы управления, которая на основе математической модели динамики полета истребителя в реальном режиме времени обеспечивает пилоту предсказуемость управления движением. Аэродинамическая неустойчивость истребителя дает ему большое конкурентное преимущество в маневренности и господство в воздухе, но именно эта неустойчивость порождает для него и большие риски: отказ бортовой электроники делает его поведение хаотическим и неуправляемым. Если для истребителя неустойчивость — это возможность, то для транспортного или пассажирского самолета — это угроза, с которой у инженеров и пилотов идет постоянная борьба.

В медицинской практике известно, что при пересадке органа необходимо ослабить иммунитет больного. Иммунитет человека — важнейший механизм обеспечения устойчивости организма, обеспечивающий свою целостность и индивидуальность путем распознавания и удаления чужеродных веществ и клеток. Сильный иммунитет приведет к отторжению чужого органа, поэтому иммунитет надо сначала ослабить, затем позволить ему аккуратно, целевым образом принять чужой орган как свой, и уже потом — снова восстановить иммунитет. В период ослабления иммунитета очень важно защитить организм от посторонних инфекций. Приживление чужого органа — это сложный процесс самоорганизации всего организма. Здесь организм человека сам решает, как он поступит в такой ситуации, а врачи могут лишь вовремя понять его и помочь ему.

В отличие от технических и природных систем, организация — социальная система, которая состоит

¹ Johns Hopkins University, COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE), 2020, <https://coronavirus.jhu.edu/map.html>

из людей, имеющих свои представления и историю отношений, свои намерения и ожидания, мотивы и нормы поведения. Люди в организации выстраивают различного рода взаимные отношения, поэтому организация не может быть сведена к простой совокупности людей, а представляет собой некоторое новое качество, которое проявляет себя через объективные свойства организации в целом. Одним из таких свойств является ее устойчивость/неустойчивость.

2. Статистическая устойчивость организации

Понятие устойчивости/неустойчивости организации наиболее близко к области знаний об управлении. В теории управления [12] традиционно выделяют объект и субъект управления. В нашем случае объектом управления является действующая организация, а субъектом — руководитель (команда руководителей) этой организации. Команда руководителей управляет организацией, то есть воздействует на организацию так, что это позволяет им достигать ожидаемых результатов. Управление всегда осуществляется в условиях возмущений, то есть неожиданных (незапланированных) отклонений от нормальных условий функционирования организации и самого управления. Возмущения могут иметь как внешнее происхождение (например, изменение условий деятельности организации), так и внутреннее (например, появление внутренних проблем, конфликтов или инициатив). Возмущения требуют от команды руководителей особых управленческих действий.

Организацию будем считать устойчивой, если для команды руководителей любая ситуация, вызванная возмущением, является предсказуемой, то есть команда обладает достаточным знанием и опытом как в части развития ситуации, так и в способах ее разрешения.

Организацию будем считать неустойчивой, если для команды руководителей развитие ситуации, связанной с возмущением, оказывается непредсказуемыми.

Рассмотрим устойчивость организации на примере отдельного бизнес-процесса реальной промышленной компании.

В крупной промышленной компании существует процесс согласования договоров. У каждого договора есть куратор, который отвечает за разработку, согласование и контроль исполнения договора с каким-либо поставщиком. На этапе совместной с

поставщиком разработки договора и в ходе согласования со службами компании в договор вносятся множество изменений, которые часто требуют повторных согласований с поставщиком и другими участниками. Несмотря на то, что в компании разработаны четкие регламенты и нормативы проведения согласования, движение большинства договоров по процессу оказывается слабо предсказуемым. Это связано с высоким уровнем изменчивости рыночной конъюнктуры, изменениями в самой организации, а также множеством противоречий между подразделениями компании, порождающих низкий уровень взаимного доверия, который выливается в недоверие куратору договора. В результате согласование каждого договора для куратора превращается из «забега по четким правилам по заранее проложенным дорожкам» в «забег по пересеченной местности с меняющимися правилами и ландшафтом».

Все участники процесса в общем заинтересованы, чтобы договоры согласовывались как можно быстрее, но также не хотят брать на себя дополнительную ответственность, связанную с неопределенностью последствий возникающих изменений. Всем участникам каждый раз как-то удается договариваться со скоростью развития ситуации. Интересно, что средние показатели времени согласования договоров по процессу уже несколько лет остаются стабильными, но при этом движение каждого договора оказывается труднопредсказуемым.

Естественно, процесс согласования договоров подвергается возмущениям, как внешним, так и внутренним. В результате мы видим признаки как устойчивости, так и неустойчивости.

В начале договорной компании, когда договоров еще мало и преобладают типовые договоры, процесс приобретает устойчивость: движение каждого договора становится предсказуемым и в значительной степени соответствует регламентам процесса.

В прошлом году по распоряжению генерального директора с целью наведения порядка в процесс была введена еще одна контрольная служба. В результате координация участников процесса сильно замедлилась, процесс стал неустойчивым и начались неожиданные срывы согласования. Вышшие руководители с изумлением обнаружили, что теперь им самим приходится вмешиваться в «разруливание» коллизий согласования. После того, как все участники вновь «притерлись» друг к другу, процесс опять вышел в старый режим и его средние показатели вновь стали теми же, что и ранее.

В приведенном примере мы можем наблюдать эволюцию режима функционирования процесса от классической устойчивости, то есть полной предсказуемости, до классической неустойчивости, то есть полной непредсказуемости процесса. Интересно, что между этими крайностями есть еще режим, когда движение каждого договора оказывается мало-предсказуемым, но, тем не менее, средние показатели согласования договоров по процессу уже несколько лет остаются весьма предсказуемыми. Такой режим функционирования процесса будем называть «статистической устойчивостью». Явление статистической устойчивости можно часто наблюдать как на уровне отдельных управленческих ситуаций, проектов операционных процессов, так и на уровне организации в целом.

3. Статистическая устойчивость и самоорганизация

Эффекты, аналогичные статистической устойчивости, также широко распространены в технических и природных системах и хорошо описываются в теории сложных систем [13–17]. Теория сложных систем описывает поведение технических и природных систем в форме динамических математических моделей, где поведение системы представляется как траектории ее движения в фазовом пространстве. Теория выделяет особые состояния сложной системы, которые являются своеобразными областями притяжения всех траекторий ее математической модели. Такую область притяжения траекторий в математической модели называют аттрактором [13,16]. В области аттрактора поведение системы становится абсолютно предсказуемым. Примером такого аттрактора является центр воронки, образованный вихрем воды, уходящей в отверстие раковины. Аттрактор является свойством самой системы и отражает режимы ее устойчивого поведения, независимо от произошедших в ней возмущений.

Теория сложных систем также говорит о том, что при определенных условиях у системы могут появляться аттракторы, которые представляют собой некоторое ограниченное множество состояний системы (замкнутую область фазового пространства системы), имеющее особые свойства. Извне к этому множеству состояний (фазовой области) также притягиваются траектории, описывающие поведение системы, но внутри этой области поведение системы оказывается принципиально непредсказуемо. В математике такие области получили название

«странные аттракторы» [16]. Интересно, что внутри такого странного аттрактора поведение системы в каждый момент времени непредсказуемо, но, тем не менее, поведение системы в среднем оказывается вполне предсказуемым.

Странные аттракторы появляются только в системах, которые являются открытыми, то есть через которые проходит поток энергии и информации, и которые находятся в состоянии повышенного уровня динамики. Странные аттракторы обладают удивительным свойством: при изменении условий деятельности системы они могут распадаться, порождая хаотическое поведение всей системы, но могут и трансформироваться в другие странные аттракторы, и даже порождать устойчивые структуры, являющиеся классическими аттракторами. Такое порождение устойчивых структур, происходящее не по чьей-то воле, а в силу естественных законов поведения сложной системы, называют самоорганизацией [13–16].

Рассмотренный выше процесс согласования договоров, находящийся в состоянии статистической устойчивости, удивительным образом похож на странный аттрактор. Постоянно воспроизводящийся комплекс практик согласования договоров явно проявляет себя как аттрактор. И этот аттрактор странный, так как при реализации согласования каждого договора участники каждый раз находят свое специфическое решение. Помимо принятых практик, они включают межличностные отношения и «на живую нитку» каждый раз находят уникальные решения. При этом показатели процесса по каждому договору можно считать случайными величинами, но средние показатели по процессу в целом оказываются вполне предсказуемыми. В этом случае деятельность по согласованию больше похожа не на забег по заранее проложенным дорожкам (процесс), а на спортивное ориентирование по пересеченной местности, да еще и с изменяющимся рельефом.

В теории сложных систем при формировании структур выделяются два фундаментальных процесса [15, 18] – режим, порождающий рост разнообразия, и режим, снижающий (размывающий, рассеивающий) разнообразие в системе.

Режим, порождающий рост разнообразия в системе, получил название «режима с обострением» (S-режим). Это режим взрывного роста каких-либо параметров процесса, например, процесс горения, переходящий во взрыв. Режим с обострением порождает быстрый рост разнообразия образующихся в системе структур. Аналогом режима с обострени-

ем для организации может являться лавинообразное нарастание внутренних инициатив или конфликтов.

Режим, порождающий снижение разнообразия в системе, получил название рассеивания или диссипации (от англ. *dissipation*). Это режим размывания, рассеивания возникших в системе структур, например, угасание горения за счет выгорания топлива и рассеяния выделяемой энергии. Аналогом режима снижения разнообразия в организации может являться координация участников, в ходе которой они между собой коммуницируют, делятся информацией и знаниями, приходят к соглашениям и находят общие решения.

По представлениям теории сложных систем динамику самоорганизации определяет соотношение этих двух процессов.

Доминирование роста разнообразия над рассеиванием. Это соотношение получило название LS-режим. Рассеивание менее интенсивно (L – lower), чем процесс роста разнообразия (S-режима). Аналогом LS-режима в организации является ситуация, когда на участников обрушивается поток из инициатив и конфликтов, который для них становится все менее предсказуемым и/или на который они не успевают скоординировано отвечать. В этих условиях появляющиеся новые структуры крайне неустойчивы: они быстро образуются и быстро распадаются.

Доминирование рассеивания над ростом разнообразия. Это соотношение получило название HS-режим. Рассеивание более интенсивно (H – higher), чем процесс роста разнообразия (S-режима). Аналогом HS-режима в организации является ситуация, когда среди ее участников любая инициатива тонет в согласованиях, порождая сопротивление изменениям устоявшихся практик. В HS-режиме организация проявляет свойства классической устойчивости.

Паритет рассеивания и роста разнообразия – вот условие формирования странных аттракторов и явления, которое в сложных системах получило название самоорганизации. Самоорганизация – процесс спонтанного упорядочивания, возникновения пространственных, временных, пространственно-временных или функциональных структур, протекающий в открытых нелинейных системах [15, 18, 19]. Самоорганизация – понятие, выражающее способность сложных систем к упорядочению

своей внутренней структуры не по чьей-то воле, а в силу объективных законов эволюции сложных систем. В сложных системах смещение паритета в сторону режима роста разнообразия приводит к образованию множества новых зародышей структур. Последующее смещение паритета в сторону рассеивания запускает процесс естественного отбора наиболее жизнеспособных из вновь образовавшихся зародышей структур.

Что же это за новые структуры, которые появляются в нашем процессе согласования договоров, находящегося в режиме статистической устойчивости? Новые структуры – это практики, которые начинают формироваться как неформальный индивидуальный или коллективный опыт участников процесса согласования. Начинается он как интуитивные частные решения, без претензий на обобщение. Далее, если опыт выживет (то есть если он будет повторно использоваться), то он сможет быть осознан как общее знание. Если это общее знание подтвердит свою ценность, то оно может трансформироваться в рациональные организационные, методические, кадровые и технические решения.

В нашем процессе согласования договоров его участники успевают найти решение и договориться в изменяющихся условиях согласования. Изменения в процесс приходят не только извне, но и порождаются внутри уникальными решениями самих участников. Процесс согласования каждого договора становится труднопредсказуемым, но при этом достигается паритет процессов порождения изменений (рост разнообразия) и координации действий участников (рассеяния). В условиях статистической устойчивости процесса смещение паритета в сторону LS-режима приводит к появлению разнообразия практик, а смещение паритета в сторону HS-режима запускает естественный отбор наиболее жизнеспособных практик.

В режиме статистической устойчивости организация требует от руководителя большого внимания и уважения к себе. Тому, кто начнет ее понимать, она сама «подскажет» возможные направления своего развития. Бессмысленно навязывать организации свои рациональные представления и волю. В этом случае в управлении большую роль начинает играть интуиция, но при этом интуитивный поиск должен идти рука об руку с постоянным рациональным анализом. Если в устойчивом состоянии организации руководитель диктует ей свою волю, то в кризисном управлении он ищет с ней союза. Для многих руководителей это может стать мировоззренческим шо-

ком. «Ее величество» статистическая устойчивость никогда не прощает неуважения, невнимательности и медлительности. Но за понимание и такт она может наградить «по-царски»: организация получает мощный генератор новых практик. Именно в этих ситуациях появляется эффект самоорганизации, который является основой инноваций и прорывных решений.

4. Статистическая устойчивость и цифровизация

В традиционных организациях основой ценности создаваемого продукта являлись физические свойства самого продукта (например, уровень потребительского качества) или организация деятельности бизнеса (например, поставка «точно в срок»). В цифровых организациях продукт становится цифровым, его ценность основана на данных. При этом чем больше данных и программного кода связано с цифровым продуктом, тем больше появляется возможностей извлечь из него ценность [6]. В этом случае и бизнес-модели взаимодействия с другими организациями становятся цифровыми, то есть они уже принципиально основаны на данных. Цифровыми становятся и цепочки создания ценности, в которые организация оказывается включенной наряду с другими цифровыми организациями. Данные становятся «воздухом», которым «дышат» цифровые организации, а новые технологии (например, большие данные, интернет вещей, машинное обучение и искусственный интеллект, виртуальная и дополненная реальность, блокчейн) помогают им его производить, усваивать и использовать для построения новых форм организации бизнеса.

Систематическое использование информации и новых технологий оказывает сильное влияние на паритет процессов порождения изменений (рост разнообразия) и координации действий участников (рассеяние) (таблица 1).

Анализ влияния цифровизации на баланс процессов порождения изменений и координации действий участников (таблица 1) показывает, что цифровизация способствует росту изменчивости и разнообразия организации. При этом одним из ограничивающих факторов роста изменчивости является развитость процессов координации, то есть способности организации координировать свои действия в нарастающем потоке изменений. Увеличивая изменчивость и разнообразие, цифровизация смещает паритет в сторону LS-режима и делает

вновь образовавшиеся практики нежизнеспособными. Они распадаются, и в организации вновь восстанавливается паритет процессов порождения изменений и координации. Такой паритет проявляет себя как статистическая устойчивость организации. Рассмотрим пример из реальной практики.

Во многих производственных компаниях классическая реализация логистической цепи поставок материально-технических ресурсов (МТР) включает этапы планирования производственной потребности в МТР, подготовку и выполнение закупки, транспортировку, хранение, отпуск в производство, использование. Информатизация процесса реализована классическими ERP-системами по модели MRP II. В связи с тем, что имеется разрыв между реальным состоянием дел (поставкой, объемом МТР на складах, отпуском) и информацией в системе, а также в связи с задержками в поступлении и обработке информации в ERP-системе в логистической цепи образуются запасы, демпфирующие как области неопределенности, так и издержки оппортунистического поведения. Именно с помощью этих демпферов организация устойчива в части обеспечения потребности производства в МТР. Для каждой заявки МТР фиксируются ключевые показатели, которые отслеживает менеджмент при принятии решений. А для процесса в целом фиксируются некоторые параметры — средняя продолжительность закупки, допустимый объем страхового запаса и т.п., которые определяют организационно-распорядительные решения по процессу.

Цифровая трансформация меняет этот процесс как в части обеспечения достоверности информации, так и в части снижения издержек оппортунистического поведения. В процессе трансформации осуществляется 100% маркировка и автоматическое (без ручного ввода) отслеживание МТР (штрихкоды или RFID-метки, считыватели), выполняется переход на полностью роботизированные склады, обеспечивается электронный документооборот, запускаются инфроботы по выполнению всех рутинных и стандартных операций. В каждый момент времени обеспечивается полная и достоверная информация о текущем состоянии выполнения заявки на поставку МТР и их движении, которая выполняет роль обратной связи для контуров управления процессов. На основе этой информации работают системы принятия управленческих решений, многие из которых реализуются с использованием искусственного интеллекта.

Таблица 1.

**Влияние цифровизации на паритет
процессов порождения изменений (рост разнообразия)
и координации действий участников (рассеяние)**

Изменения, связанные с цифровизацией	Смещение паритета в сторону LS-режима	Смещение паритета в сторону HS-режима
Рост глубины и масштаба аналитики	Интеллектуальная аналитика на больших данных позволяет увидеть то, чего раньше видеть не могли: чаще открываются новые угрозы и возможности, которые требуют все более сложной реакции. Фактически цифровизация открывает большое количество новых потенциальных внешних и внутренних возмущений и увеличивает скорость их развития	Та же интеллектуальная аналитика позволяет в сложной ситуации быстрее находить приемлемое для всех участников решение. Она создает единую картину, ускоряющую координацию участников
Рост масштаба и глубины интеграции	Информационное пространство выходит за границы самой цифровой организации и открывается (в рамках прав доступа) другим участникам цепочек создания ценности. При этом увеличивается не только масштаб, но и глубина интеграции информации. Информационные связи в цепочках становятся более сильными, что приводит к тому, что изменения в одной организации становятся общими для организаций всей цепочки. Цифровая организация становится более чувствительной к внешним изменениям	Именно та же интеграция информационного пространства является основой для быстрой синхронизации происходящих изменений в цепочках создания ценности (оркестровка цепочек создания ценности). Более сильные информационные связи в цепочках способны передавать не только изменения, но и варианты синхронизации усилий участников цепочки
Рост количества участников координации	Управление цепочками создания ценности резко расширяет количество участников координации, а значит происходит резкий рост разнообразия и изменчивости их интересов, ожиданий и намерений	Современные информационные технологии позволяют повысить эффективность координации: быстро найти необходимых участников и сформировать рабочую группу, быстро создать рабочее информационное пространство и организовать групповую работу, привлекать и повторно использовать накопленные знания, принимать решения и контролировать их исполнение. При этом надо учитывать, что эффективная координация требует постоянного роста квалификации и развития экспертной и лидерской мотивации участников
Рост сложности моделей опережающий рост интеллекта	Работа с большими данными приводит к появлению большого разнообразия моделей операционной деятельности, а также к усложнению причинно-следственных связей в этих моделях. Использование этих моделей для принятия решений потребует больших интеллектуальных ресурсов. Если рост интеллекта будет постоянно отставать от роста сложности моделей, то принятие решений будет постоянно сопровождаться ошибками, их исправлениями и дополнительными изменениями	Постоянный рост сложности и разнообразия моделей деятельности будет стимулировать развитие наших интеллектуальных способностей, как это всегда было в истории. Режим статистической устойчивости — это не хаос, а другой порядок, которому соответствует и новое мышление, основанное на синтезе интуитивного и рационального мышления. Систематическое использование этого нового мышления способно усилить наши интеллектуальные возможности
Рост скорости процессов и их индивидуализация	Цифровизация позволяет радикально повысить производительность операционных процессов, выходя на уровень цепочек создания ценности, которые выходят за границы организации. При этом реализация каждого заказа в процессе становится все более индивидуальной. Это приводит к тому, что скорость реализации изменений растет и сама реализация изменений становится все более сложной и разнообразной	Часть решений на нижнем уровне, особенно в условиях роста скорости и индивидуализации процессов, возьмут на себя системы искусственного интеллекта. Принятие решений на более верхних уровнях в условиях роста скорости и разнообразия останется за человеком. Новое мышление «в сложности» способно уменьшить влияние этих изменений

Таким образом, с точки зрения исполнения каждой конкретной заявки на МТР система ведет себя совершенно непредсказуемым образом, принимая оперативные решения на любом этапе. Например, если фиксируется задержка поставки по рамочному договору, система может автоматически генерировать заявки на закупку в розничной сети. Когда предыдущему производственному этапу потребовалось меньше МТР, чем было запланировано, вносятся изменения по всей цепочке поставок. Наиболее умные системы могут учитывать и другие факторы. Например, снижение температуры окружающего воздуха приведет к увеличению потребления электроэнергии для котельных и дополнительный расход материалов на техобслуживание и ремонт оборудования, которое обслуживается по наработке. Таким образом, реализуется режим статистической устойчивости, когда неустойчивость каждого экземпляра процесса позволяет реализовать устойчивое обеспечение производства МТР. Управление этим процессом больше похоже на устойчивость/неустойчивость истребителя, чем устойчивость транспортного самолета.

Мы видим, что планирование и исполнение заявок на МТР становится все более индивидуальным и менее предсказуемым, то есть происходит рост разнообразия в потоке заявок на МТР (LS-режим). При этом система планирования успевает не только оперативно спланировать исполнение новой заявки, но и при необходимости оперативно скорректировать планы уже исполняющихся заявок, то есть рассеивает разнообразие в потоке заявок (HS-режим). Такая система планирования МТР обеспечивает паритет роста и рассеивания разнообразия потока заявок и делает его статистически устойчивым. Интересно, что именно в таком режиме статистической устойчивости начинают проявляться элементы самоорганизации: система начинает «подсказывать» планировщикам новые возможные методы планирования, специфичные для данного бизнеса и сильно отличающиеся от широко распространенных моделей MRP II.

Существование режима статистической устойчивости цифровой организации — это не проблема, а объективное свойство, с которым бессмысленно бороться. Со статистической устойчивостью надо научиться жить, и, похоже, она может стать новой нормой цифровых организаций.

В исследовании природы экспоненциальных организаций [1] явно показано, что их взрывной рост всегда сопровождается ростом изменчивости, который постоянно выводит их на грань устойчивости своего

бизнеса. Экспоненциальные организации научились жить в режиме статистической устойчивости, который позволяет им балансировать на грани устойчивости и поддерживать высокий уровень инновационности своего бизнеса. Именно в режиме статистической устойчивости методом проб и ошибок эти компании сумели нащупать источники своего взрывного роста. Интересно, что экспоненциальные организации сознательно поддерживают режим статистической устойчивости даже тогда, когда они превращаются из стартапов в большие компании. Даже тогда, когда экспоненциальные организации стабилизируют свои отдельные операционные бизнес-процессы, они все равно сохраняют высокую готовность к переходу в режим статистической устойчивости и готовности к изменениям.

Заключение

Систематическая цифровизация выводит организацию в режим статистической устойчивости, когда предсказуемость действий ее отдельных участников начинает исчезать, но предсказуемость в целом еще сохраняется. С точки зрения традиционного менеджмента, ориентированного на долгосрочную предсказуемость и классическую устойчивость форм организации, статистическая устойчивость — это хаос и источник рисков, с которым необходимо бороться. С точки зрения менеджмента цифровых организаций, ориентированного на высокий уровень изменчивости и быстрый рост, статистическая устойчивость — это генератор новых практик и источник новых возможностей. В этом режиме организация может балансировать на грани устойчивости и поддерживать высокий уровень изменчивости и инновационности.

Часто цифровизацию воспринимают как традиционный реинжиниринг бизнес-процессов, но только на базе новых технологий. В этом случае статистическая устойчивость традиционно рассматривается как источник рисков и с ней начинают бороться. На самом деле статистическая устойчивость — это необходимое условие как проведения трансформации в цифровую организацию, так и ее дальнейшей жизнедеятельности, поэтому ее надо стимулировать и поддерживать.

Цифровизация стимулирует развитие неустойчивости организации, и как бы подталкивает ее к краю обрыва. Но именно здесь, на краю обрыва, открывается ее новая природа: оказывается, организация — это не машина, а организм, к которому менеджменту предстоит адаптироваться. Если природа за нас, то кто против нас? Горе тому, кто борется с природой. ■

Литература

1. Исмаил С., Мэлоун М., Геест Ю. Взрывной рост: Почему экспоненциальные организации в десятки раз продуктивнее вашей (и что с этим делать) / Пер. с англ. М.: Альпина Паблишер, 2017.
2. Шваб К. Четвертая промышленная революция / Пер. с англ. М.: Эксмо-Пресс, 2020.
3. Вызов 2035 / И.Р. Агамирзян и [др.]. М.: Олимп-Бизнес, 2016.
4. Industry 4.0: Building the digital enterprise. 2016 Global Industry 4.0 Survey / PwC, 2016. [Электронный ресурс]: <https://www.pwc.com/gx/en/industries/industries-4.0/landing-page/industry-4.0-building-your-digital-enterprise-april-2016.pdf> (дата обращения 15.10.2020).
5. Программа «Цифровая экономика Российской Федерации». Утверждена распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р. [Электронный ресурс]: <http://static.government.ru/media/files/9gFM4FHj4PsB79I5v7yLVuPgu4bvR7M0.pdf> (дата обращения 15.10.2020).
6. Ананьин В.И., Зимин К.В., Лугачев М.И., Гимранов Р.Д., Скрипкин К.Г. Цифровое предприятие: трансформация в новую реальность // Бизнес-информатика. 2018. № 2 (44). С. 45–54. DOI: 10.17323/1998-0663.2018.2.45.54.
7. Ананьин В.И., Зимин К.В., Гимранов Р.Д., Лугачев М.И., Скрипкин К.Г. Реальное время управления предприятием в условиях цифровизации. 2019. Т. 13. № 1. С. 7–17. DOI: 10.17323/1998-0663.2019.1.7.17
8. Талей Н.Н. Черный лебедь. Под знаком непредсказуемости / Пер. с англ. М.: КоЛибри, 2020.
9. Коротков Э.М. Антикризисное управление. М.: ИНФРА-М, 2003.
10. Шеремет А.Д., Негашев Е.В. Методика финансового анализа деятельности коммерческих организаций. М.: ИНФРА-М, 2016.
11. Федеральный закон РФ от 21 декабря 1994 года № 68-ФЗ «О защите населения и территорий от чрезвычайных ситуаций природного и техногенного характера». [Электронный ресурс]: http://www.consultant.ru/document/cons_doc_LAW_5295/ (дата обращения: 15.12.2020).
12. Гапоненко А.Л., Савельева М.В. Теория управления. М.: Юрайт, 2020.
13. Пригожин И., Сенгерс И. Порядок из хаоса: Новый диалог человека с природой / Пер. с англ. М.: URSS, 2008.
14. Хакен Г. Информация и самоорганизация: Макроскопический подход к сложным системам / Пер. с англ. М.: URSS, 2014.
15. Курдюмов С.П., Малинецкий Г.Г., Потапов А.Б. Синергетика — новые направления. М.: Знание, 1989.
16. Малинецкий Г.Г. Математические основы синергетики: Хаос,

Statistical sustainability of a digital organization

Vladimir I. Ananyin^a

E-mail: v.ananiin@gmail.com

Konstantin V. Zimin^b

E-mail: konst.zimin@gmail.com

Mikhail I. Lugachev^c

E-mail: mlugachev@gmail.com

Rinat D. Gimranov^d

E-mail: gimranov_rd@mail.ru

^a Russian Presidential Academy of National Economy and Public Administration

Address: 82 build. 1, Prospect Vernadskogo, Moscow 119571, Russia

^b The Russian Union of CIO

Address: 34, Seleznevskaya Street, Moscow 123056, Russia

^c Lomonosov Moscow State University

Address: 1 build. 46, GSP-1, Leninskie Gory, Moscow 119991, Russia

^d PJSC Surgutneftegaz

Address: 1 block 1, Grigoriya Kukuevitskogo Street, Surgut 628415, Russia

Abstract

An important feature of a digital organization is its ability to change rapidly. For an organization to remain capable of rapid change, it must be on the brink of resilience, since a resilient organization always resists change. The article examines the borderline state of the organization, which is on the verge of its stability and instability. In this state, the organization begins to lose predictability in the details of behavior, but still retains predictability in general. The authors called this borderline state the statistical sustainability of the organization. The phenomenon of statistical sustainability of an organization is very similar to the property of stability of the frequency of mass events and average values described in mathematical statistics by a similar term. To analyze the nature of the statistical sustainability of the organization, the authors used the ideas of strange attractors and modes with sharpening from the theory of complex systems. A strange attractor is an area of the organization's behavior that, outside this area, is an area of stability for the organization, and inside it is an area of complete unpredictability. The theory of complex systems has shown that it is in the regions of strange attractors that the conditions for the variability of systems are created, and the theory of modes with aggravation shows the conditions under which this variability can lead to self-organization, that is, the spontaneous emergence of new structures. This article shows that systematic digitalization objectively leads to the formation of the statistical sustainability of the organization and creates the preconditions for maintaining the organization's ability to make rapid changes. In traditional management, the statistical sustainability of an organization is viewed as a threat and a source of risk. Therefore, in the context of systematic digitalization, traditional approaches to management should be significantly refined.

Key words: digital organization; variability; stability/instability of the organization; exponential organization; self-organization; strange attractor; diversity; scattering; digital transformation; statistical sustainability.

Citation: Ananyin V.I., Zimin K.V., Lugachev M.I., Gimranov R.D. (2021) Statistical sustainability of a digital organization. *Business Informatics*, vol

3. Agamirzyan I.R., et. al (2016) *Challenge 2035*. Moscow: Olymp Business (in Russian).
4. PwC (2016) *Industry 4.0: Building the digital enterprise. 2016 Global Industry 4.0 Survey*. Available at: <https://www.pwc.com/gx/en/industries/industries-4.0/landing-page/industry-4.0-building-your-digital-enterprise-april-2016.pdf> (accessed 15 October 2020).
5. The Government of the Russian Federation (2017) *The program "Digital economy of the Russian Federation."* Approved by Order of the Government of the Russian Federation No 1632-r of 28 July 2017. Available at: <http://static.government.ru/media/files/9gFM4FHj4PsB79I5v7yLVuPgu4bvR7M0.pdf> (accessed 15 December 2020) (in Russian).
6. Ananyin V.I., Zimin K.V., Lugachev M.I., Gimranov R.D., Skriprin K.G. (2018) Digital organization: Transformation into the new reality. *Business Informatics*, no. 2, pp. 45–54. DOI: 10.17323/1998-0663.2018.2.45.54.
7. Ananyin V.I., Zimin K.V., Gimranov R.D., Lugachev M.I., Skriprin K.G. (2019) Real time enterprise management in the digitalization era. *Business Informatics*, vol. 13, no 1, pp. 7–17. DOI: 10.17323/1998-0663.2019.1.7.17.
8. Taleb N.N. (2007) *The black swan: The impact of the highly improbable*. New York: Random House.
9. Korotkov E.M. (2003) *Anti-crisis management*. Moscow: INFRA-M (in Russian).
10. Sheremet A.D., Negashev E.V. (2016) *Methodology of financial analysis of commercial organizations' activities*. Moscow: INFRA-M (in Russian).
11. The Russian Federation (1994) *Federal Law of the Russian Federation No 68-FZ of 21 December 1994 "On the protection of the population and territories from natural and man-made emergencies."* Available at: http://www.consultant.ru/document/cons_doc_LAW_5295/ (accessed 15 December 2020) (in Russian).
12. Gaponenko A.L., Savelieva M.V. (2020) *Theory of management*. Moscow: Urait (in Russian).
13. Prigogine I., Stengers I. (1984) *Order out of chaos: Man's new dialogue with nature*. Bantam New Age Books.
14. Haken H. (2006) *Information and self-organization: A macroscopic approach to complex systems*. Berlin, Heidelberg: Springer.
15. Kurdyumov S.P., Malinetsky G.G., Potapov A.B. (1989) *Synergetics – new directions*. Moscow: Znanie (in Russian).
16. Malinetsky G.G. (2017) *Mathematical foundations of synergetics: Chaos, structures, computational experiment*. Moscow: URSS (in Russian).
17. Maturana H.R., Varela F.J. (1980) *Autopoiesis and cognition: The realization of the living*. Dordrecht: D. Reidel Publishing.
18. Vasilkova V.V. (1999) *Order and chaos in the development of social systems: Synergetics and theory of social self-organization*. St. Petersburg: Lan' (in Russian).
19. Kershengolts B.M., Chernobrovkina T.V., Shein A.A., Khlebny E.S., Anshakova V.V. (2009) *Nonlinear dynamics (synergetics) in chemical, biological, and biotechnological systems*. Yakutsk: Ammosov Yakut State University (in Russian).

About the authors

Имитационное моделирование развития отдельных отраслей тяжелой промышленности*

Е.В. Кислицын 

E-mail: kev@usue.ru

В.В. Городничев

E-mail: helltoaster@yandex.ru

Уральский государственный экономический университет

Адрес: 620144, г. Екатеринбург, ул. 8 Марта, д. 62

Аннотация

В настоящее время повышается актуальность задачи обеспечения устойчивого развития отраслей тяжелой промышленности. Однако для этого одних лишь теоретических и аналитических исследований недостаточно, и экономическая наука нуждается в разработке принципиально новых подходов к исследованию развития промышленных отраслей. Статья посвящена разработке и апробации имитационной модели развития отдельных отраслей экономики. В качестве объекта исследования выступает металлургическая отрасль, а также такие смежные отрасли, как добыча руд, машиностроение и производство готовых металлических изделий. Теоретическую основу исследования составляет системный подход, объединивший в себе теорию отраслевых рынков, экономического роста, экономики промышленности, системной динамики и математической экономики. В качестве основных методов исследования применены системный анализ, статистический анализ (для выявления тенденций изменения основных экономических показателей), эконометрическое моделирование (для построения производственных функций), а также математическое моделирование макроэкономических систем. В результате предложена имитационная модель, разработанная в нотации системной динамики с использованием программного продукта AnyLogic, позволяющая оценить развитие отдельных отраслей промышленности с учетом различных изменений. Данная модель построена на базе трехсекторной модели национальной экономики, где в качестве структурных элементов выделены отдельные отрасли, связанные динамическими петлями обратной связи. В работе подробно отражена структура имитационной модели, базирующейся на динамических уравнениях первого порядка, балансовых уравнениях и нелинейных производственных функциях. Имитационная модель позволила спрогнозировать ряд сценариев развития отраслей металлургии с учетом изменения трудовых ресурсов и инвестирования в основные производственные фонды. Результаты работы могут использоваться для формирования предложений по промышленной политике, а также мониторинга состояния и эффективности деятельности отдельных отраслей промышленности.

* Статья опубликована при поддержке Программы НИУ ВШЭ «Университетское партнерство»

Ключевые слова: имитационное моделирование; системная динамика; промышленность; металлургия; трехсекторная модель экономики; производственная функция; математическая модель.

Цитирование: Кислицын Е.В., Городничев В.В. Имитационное моделирование развития отдельных отраслей тяжелой промышленности // Бизнес-информатика. 2021. Т. 15. № 1. С. 59–77. DOI: 10.17323/2587-814X.2021.1.59.77

Введение

Устойчивое развитие отраслей тяжелой промышленности остается важнейшим приоритетом национальной экономики. Механизмы обеспечения такого развития ложатся в основу промышленной политики РФ, отдельных национальных проектов и программ [1]. Поэтому является актуальной задача поиска новых способов, методов и инструментов обеспечения и поддержки устойчивого развития отраслей тяжелой промышленности. В современных российских и зарубежных исследованиях в качестве таковых используются как классические экономические [2, 3], так и новые междисциплинарные подходы [4, 5]. Однако разработать такие инструменты без использования современного математического инструментария и вычислительных систем практически невозможно.

Аналитические исследования в данной области, в большинстве своем, сводятся к анализу текущей ситуации в отдельных отраслях [6, 7], регионах [8–10] и экономике в целом [11], а также к построению эконометрических моделей и интерпретации полученных зависимостей для дальнейшей корректировки промышленной политики. Однако такие модели практически не учитывают, во-первых, нелинейное поведение экономических систем и, во-вторых, наличие причинно-следственных зависимостей между отдельными показателями этих систем. Использование математического и имитационного моделирования в исследованиях отраслей промышленности встречается в современных исследованиях все чаще, однако не всегда предлагаемые авторами модели теоретически обоснованы и эмпирически апробированы.

В данной работе предпринята попытка использовать для исследования развития отраслевых экономических систем идеи математической экономики и, в частности, моделирования макроэкономических динамических систем, в сочетании с имитационным моделированием. Такой синтез возможен благодаря использованию причинно-следственных и потоковых нотаций имитационного моделирования. Особенностью предлагаемого подхода является использование методов системной динамики,

впервые предложенных в работах Дж. Форрестера [12, 13] и Д. Медоуз [14, 15], и впоследствии развитых в работах В.Н. Сидоренко [16], А.С. Аконова [17–19] и других исследователей. В основе динамических моделей макроэкономических систем, как правило, лежат дифференциальные уравнения первого порядка, которые можно представить в нотации системной динамики, используя потоковую стратификацию.

Цель данного исследования – разработать имитационную модель развития отдельных отраслей промышленности на примере металлургии и смежных с ней отраслей (добыча металлических руд, машиностроение, производство готовых металлических изделий), используя идеи трехсекторной модели экономики и системной динамики. В соответствии с целью исследования, были поставлены и решены следующие задачи:

- 1) провести анализ имеющихся исследований в области имитационного моделирования экономических систем;</

рый агрегированный продукт. Основой здесь является односекторная модель Солоу [20], в которой экономическая система рассматривается как некий неструктурированный аппарат, производящий единственный универсальный продукт. Данная модель отражает процесс воспроизводства и позволяет анализировать соотношение между потреблением и накоплением.

Развитием односекторной модели Солоу стали двухсекторная [21] и трехсекторная модели [22]. Вид таких моделей позволяет не только провести анализ роста экономики в целом, но и сфокусировать внимание на отдельных ее секторах. Однако развитие данных идей в основном заключается в исследованиях макроэкономических систем. Вместе с тем, не менее важной является проблема развития отдельных отраслей, в первую очередь, промышленности. Идеи, изложенные в трехсекторной модели [22, 23], по мнению автора, можно использовать для исследования экономического роста не секторов экономики в целом, а отдельных ее отраслей, наиболее тесно связанных друг с другом. Единственным допущением, которое необходимо соблюдать при модификации данной модели, является необходимость представить отрасли из трех секторов экономики — материального, фондосоздающего и потребительского.

Математический инструментарий, представленный в трехсекторной модели экономики, представляет собой совокупность дифференциальных уравнений, балансовых уравнений и нелинейных производственных функций. Аналитическое решение трехсекторной модели экономики для сбалансированного экономического роста подробно описано в работах [23, 24]. В рамках данного исследования предлагается использовать принципиально иной подход — имитационное моделирование.

Имитационное моделирование — это экспериментальный метод исследования реальной или проектируемой системы на основе ее имитационной модели, который сочетает в себе особенности экспериментального подхода со специфическими условиями применения вычислительной техники [25]. Согласно принятой классификации, существует три подхода к созданию имитационных моделей: дискретно-событийное моделирование, системная динамика и агентное моделирование [26].

Дискретно-событийное моделирование в основном используется для анализа бизнес-процессов и производственных систем на предприятиях, где

логика может быть описана в виде последовательности конкретных действий. К примеру, в работе [27] имитационное моделирование применяется для оценки рисков промышленных предприятий. Данная имитационная модель представляет собой процессную (дискретно-событийную) модель производственного процесса, на которой проводились статистические испытания и исследовалась вариация значения показателя рентабельности в зависимости от различных входных факторов.

числе уровень роста потребительских цен, уровень безработицы и т.д., а также четыре функциональных зависимости и несколько десятков вспомогательных переменных.

Е.И. Пискун [32] использовала системную динамику для построения модели взаимовлияния переменных оценки результативности стратегии инновационного развития производственно-экономических систем, с использованием для анализа нескольких сценариев развития системы. В модели рассчитываются такие показатели, как средний цепной темп роста, а также средний темп роста уровней инновационной интенсивности, эффективности инвестиционной деятельности, инновационного потенциала и развития производственно-экономических систем.

Особое внимание уделяется исследованиям взаимодействия экономического роста и ресурсопотребления, что также можно выразить с помощью дифференциальных математических моделей. Так, в работе [33] разработана модель в нотации системной динамики, в которой экономический рост и ресурсы рассматриваются с системной точки зрения, с целью выявления зависимости роста экономики от задерживаемых эффектов обратной связи от истощения ресурсов [34]. Модель состоит из двух подсистем — возобновляемых природных ресурсов и экономик, где в качестве накопителей используются уровень ресурсов и ВВП.

Г.Л. Бекларян [35] разработала имитационную модель экономического развития Дальневосточного федерального округа, в которой использованы методы системной динамики и агентного подхода. Основная идея модели заключается в исследовании влияния на экономику федерального округа ряда факторов, таких как темпы роста инвестиций в основные фонды, средней заработной платы, дотаций из федерального бюджета и цен на природные ресурсы. Макроэкономические показатели в модели описаны в нотации системной динамики, а регионы федерального округа представлены в виде отдельных агентов, находящихся в одном из трех состояний.

Моделирование процессов экономической динамики по отраслям представлено в работе [36], где в качестве основных модулей модели выделены государственное управление, производство, потребление, трудоустройство, демография, образование и финансовая система. В качестве основных методов в этой работе используются агентное моделирова-

ние, системная динамика, межотраслевые балансовые модели, элементы искусственного интеллекта и когнитивные технологии. В основе модели лежит определение потребности в различных ресурсах (товар, кадры и т.д.), валовое накопление основного капитала и инвестиционный спрос. В данной имитационной модели в качестве агентов представлены домохозяйства и юридические лица различных форм [37], поведение которых моделируется с помощью специальных

- ♦ материальный сектор: 07 – Добыча металлических руд; 24 – Производство металлургическое;
- ♦ фондосоздающий сектор: 28 – Производство машин и оборудования, не включенных в другие группировки (в том числе: 28.4 – Производство станков, машин и оборудования для обработки металлов, и прочих твердых металлов; 28.91 – Производство машин и оборудования для металлургии);
- ♦ потребительский сектор: 25 – Производство готовых металлических изделий (кроме машин и оборудования).

Другие отрасли в данной модели напрямую не рассматриваются, однако учитываются в качестве некой «другой» отрасли. Материальный сектор здесь представлен производством предметов труда (железные руды, руды цветных металлов, чугун, сталь, ферросплавы, стальные трубы, драгоценные и цветные металлы, литье), в качестве фондосоздающего сектора рассматриваются средства труда (машины и оборудование), а потребительского – предметы потребления (готовые металлические изделия). Здесь следует отметить, что потребителями могут являться как другие отрасли экономики, так и домохозяйства. В рамках настоящей модели рассматривается только сектор металлургии и наиболее тесные с ним отрасли.

Технологический уклад в данной модели считается постоянным, а выпуск в отраслях задается неоклассическими производственными функциями:

$$X_i = A_i \cdot K_i^{\alpha_i} \cdot L_i^{\beta_i}, \quad (1)$$

где X_i – объем выпуска в i -й отрасли;

A_i – коэффициент нейтрального технического процесса в i -й отрасли;

K_i – основные производственные фонды (далее – ОПФ) i -й отрасли;

L_i – число занятых в i -й отрасли;

α – коэффициент эластичности по фондам;

β – коэффициент эластичности по труду;

риальных затрат, начальное значение числа занятых, начальное распределение занятых по отраслям, начальные значения ОПФ отраслей и параметры производственных функций. Соответственно, управление в модели осуществляется путем распределения трудовых и инвестиционных ресурсов.

Основываясь на уравнениях (1)–(6) и основных постулатах системной динамики, автором спроектирована и разработана в программной среде AnyLogic имитационная модель развития металлургической и смежных с ней отраслей промышленности. Такая модель составлена в соответствии со следующими принципами:

1. Динамические элементы представлены в виде диаграмм потоков и накопителей, каждый динамический элемент представляет собой накопитель, а потоком зад

В качестве накопителей в модели представлены объем основных производственных фондов (капитала) каждой из исследуемых отраслей и число занятых в промышленности, которые рассчитываются по заданным дифференциальным уравнениям (2) и (4) с помощью специальных элементов — потоков. Потоки задают скорость изменения накопителей за единицу времени динамическим путем, т.е. непрерывно. С помощью динамических переменных описываются выпуск продукции отрасли, а также число занятых и объем инвестиций (*таблица 1*).

Параметрами в имитационной модели обозначаются значения, которые задаются и изменяются пользователем, а не функционально. Это данные, полученные статистическим и эмпирическим путем, рассчитанные математически и эконометрически и занесенные в модель. Параметры могут быть как стат

Таким образом, описана структура имитационной модели развития нескольких отраслей промышленности, включая металлургию, добычу металлических руд, машиностроение и производство готовых металлических изделий.

3. Результаты имитационного моделирования развития металлургической отрасли

Для апробации модели необходимо рассчитать значения экзогенных переменных (параметров) модели и задать уравнения (1)–(6) с соответствующими переменными. С целью вычисления коэффициентов A_i , α_i и β_i построены производственные функции для каждой отрасли. Для построения производственных функций использованы официальные данные Федеральной службы государственной статистики за период с 2002 по 2018 гг. К сожалению, использовать более ранние данные не удалось из-за их несопоставимости. Однако период протяженностью в 17 лет вполне позволяет построить регрессионную модель

с двумя объясняющими переменными на высоком уровне значимости. Стоимостные показатели пересчитаны в сопоставимые цены 2002 года.

В качестве показателей X_i были использованы показатели объема отгруженных товаров собственного производства по каждой из исследуемых отраслей: добыча металлических руд, производство металлургическое, производство готовых металлических изделий (кроме машин и оборудования), производство машин и оборудования, не включенных в другие группировки. Показатель K_i рассчитан как значение показателя основных фондов коммерческих организаций по каждой из отраслей. Показатель L_i рассчитан как значение показателя среднегодовой численности работников организаций (таблица 3)¹.

В результате расчетов с использованием языка программирования R и пакета RStudio были получены следующие производственные функции исследуемых отраслей промышленности:

Таблица 3.

Дескриптивная статистика исходных данных для построения производственных функций

Показатель ²	Среднее значение	Стандартная ошибка	Медиана	Минимум	Максимум	Дисперсия	Среднеквадратическое отклонение
X_07	407056,8	46944,1	988000,0	343633,8	484446,1	2034227172	45102,4

$$\begin{aligned} X_{07} &= K_{07}^{0,89} \cdot L_{07}^{0,28} \\ X_{24} &= 0,68 \cdot K_{24}^{0,74} \cdot L_{24}^{0,55} \\ X_{25} &= 5,9 \cdot K_{25}^{0,38} \cdot L_{25}^{0,49} \\ X_{28} &= K_{28}^{0,97} \cdot L_{28}^{0,22} \end{aligned} \quad (7)$$

Здесь следует отметить, что сумма коэффициентов эластичности по фондам и труду больше единицы, что говорит о наличии возрастающей отдачи. Данный тезис основан на результатах исследования С.В. Ореховой и Е.В. Кислицына [41], посвященного анализу совокупной производительности факторов, под которым понимается необъясненный остаток роста конечного продукта. В частности, в исследовании доказано наличие совокупной производительности факторов в отраслях тяжелой промышленности, в т.ч. металлургии, что дает основание предположить наличие возрастающей отдачи.

Все построенные функции и их коэффициенты статистически значимы на уровне 5%, коэффициенты детерминации колеблются от 0,4 для металлургии до 0,99 для машиностроения. Значения всех экзогенных переменных (параметров), описанных в таблице 2, рассчитаны и представлены в таблице 4.³

Основная идея имитационных моделей состоит в том, чтобы можно было моделировать различные ситуации, возникающие в системе, т.е. моделировать сценарии. В рамках данной работы было исследовано шесть различных сценариев развития отраслей металлургии и смежных с ней. Первый

сценарий – развитие экономической системы без существенных изменений в соответствии с заданными производственными функциями. При таком сценарии предполагается, что все параметры модели являются константами (т.е. сохраняются все темпы прироста основных показателей). Такой сценарий отражает существующую картину и позволяет спрогнозировать развитие экономик различных отраслей. В качестве анализируемых переменных выделены производительность труда, фондовооруженность (рисунок 3) и валовый выпуск продукции (рисунок 4).

Сразу следует отметить, что общие объемы выпуска в некоторых отраслях могут не отражать действительные значения на начало анализируемого периода, поскольку, во-первых, производственные функции строились по ценам 2002 года и, во-вторых, начальное значение выпуска в данной модели не задавалось. Основная цель здесь – проследить темпы роста отдельных отраслей.

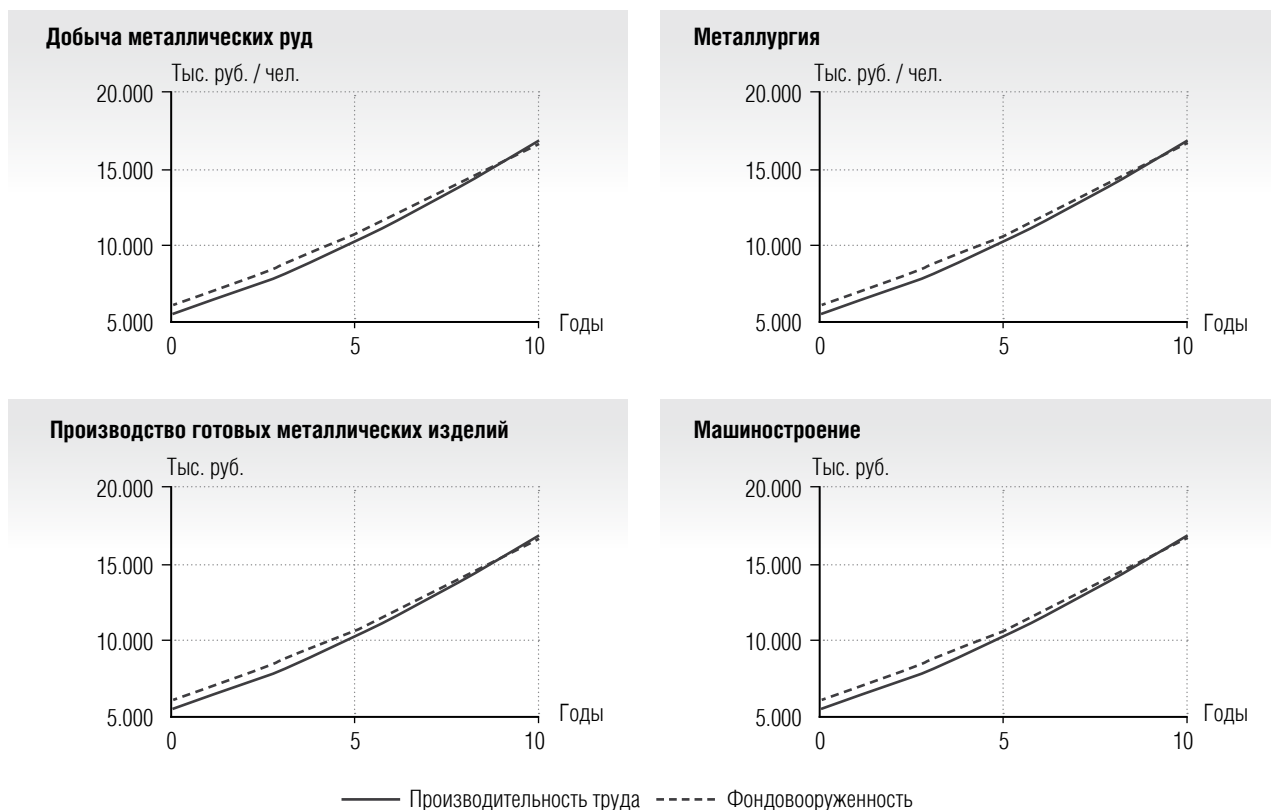


Рис. 3. Динамика изменения производительности труда и фондовооруженности по сценарию №1

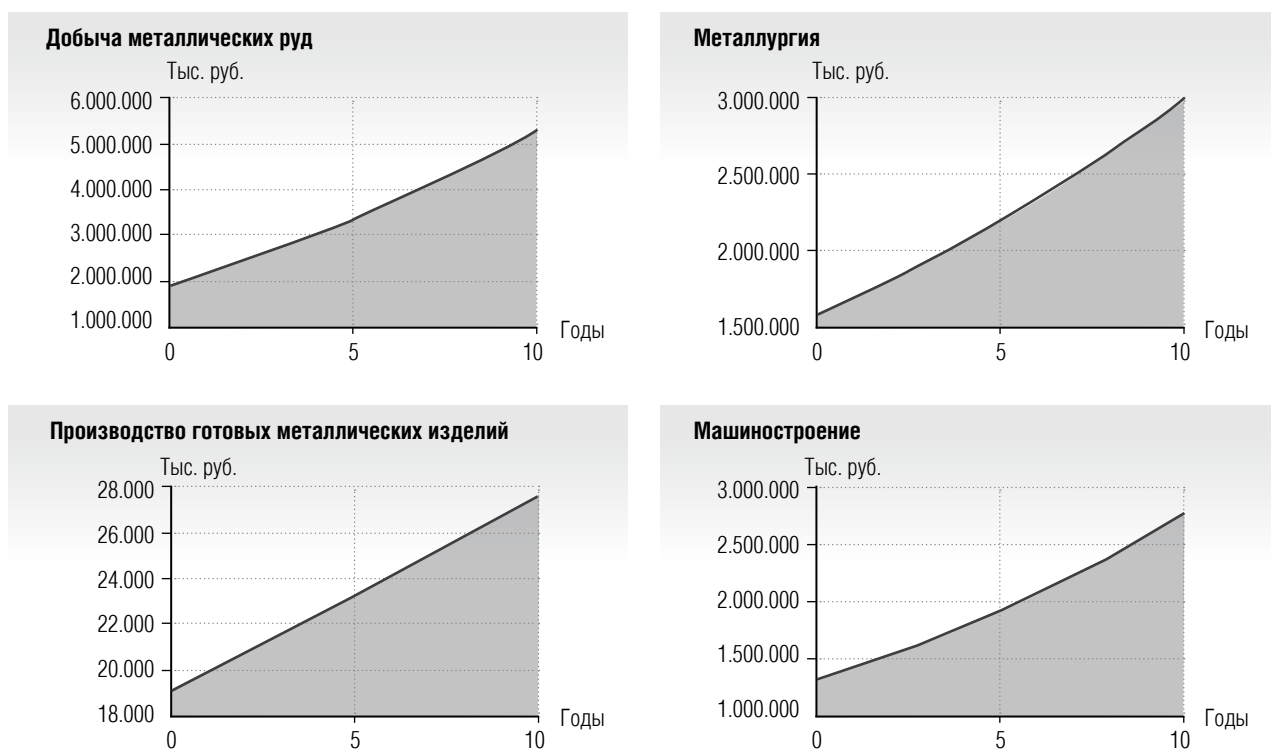


Рис. 4. Динамика изменения валового выпуска продукции по сценарию №1

Далее анализируются три пессимистических и два оптимистических сценария. Второй сценарий – это отражение уже существующей тенденции сокращения числа работников в промышленности. В имитационной модели задано сокращение темпов роста численности работников в промышленности на 1% в год (рисунки 5).

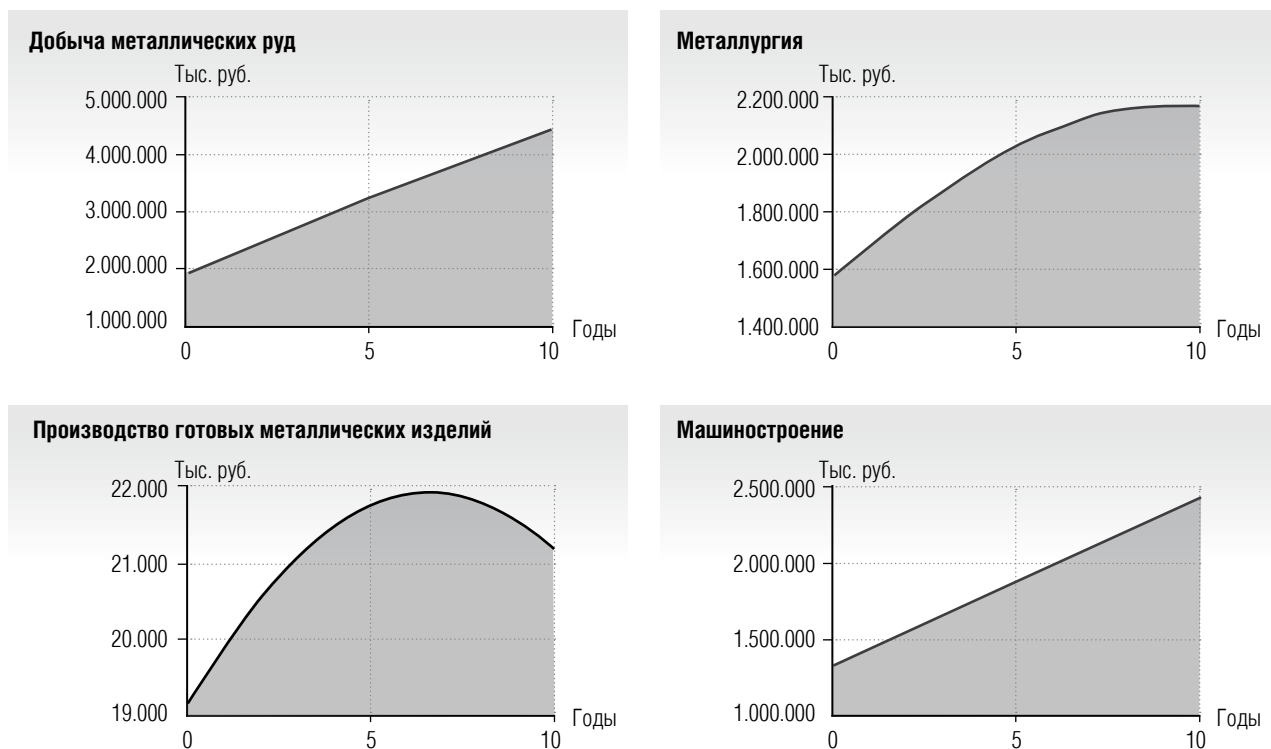
В первую очередь, сокращение притока новой рабочей силы в промышленность в целом коснется отраслей, в которых производительность труда влияет на выпуск в большей степени, чем фондовооруженность. Согласно рассчитанным значениям, выпуск в отрасли металлургии будет постепенно замедляться и на девятый год практически остановится. В отрасли производства готовых металлических изделий ситуация еще более негативная – уже на шестой год объемы выпуска начнут падать. Отрасли добычи металлических руд и машиностроения не так сильно пострадают, однако и у них темпы роста выпуска продукции сократятся.

Третий сценарий – сокращение доли работников отрасли машиностроения в общем числе работников промышленности на 25% в год. Данный сценарий является попыткой проследить причинно-следственную зависимость между выпуском отраслей и производством основных фондов (рисунки 6).

Нетрудно заметить, что темпы выпуска продукции отрасли машиностроения замедляются достаточно явно уже с третьего года, тогда как при общем снижении численности работников во всех отраслях такой тенденции не наблюдалось. Кроме того, аналогичные показатели в других отраслях также снижаются. Так, темп роста выпуска продукции в отрасли добычи металлических руд составит всего 2,28, металлургии – 1,27, а в производстве готовых металлических изделий – 1,34 за 10 лет, что существенно ниже, чем при сценарии №1.

Следующий сценарий – сокращение доли инвестиций в отрасль машиностроения на 25% в год (рисунки 7) в общем объеме инвестиций в промышленное производство. Как ожидается, здесь должна наблюдаться тенденция, аналогичная третьему сценарию. Тем не менее, при сохранении общих тенденций, количественная разница с предыдущим сценарием все же есть.

Выпуск отрасли машиностроения страдает сильнее: его темпы начинают падать уже на восьмой год прогноза. Однако, темпы выпуска других отраслей падают медленнее. Здесь сказывается балансовое уравнение (5), т.е. инвестиции в другие отрасли остаются на прежнем уровне. Даже когда темпы производства в отрасли машиностроения начинают падать, другие



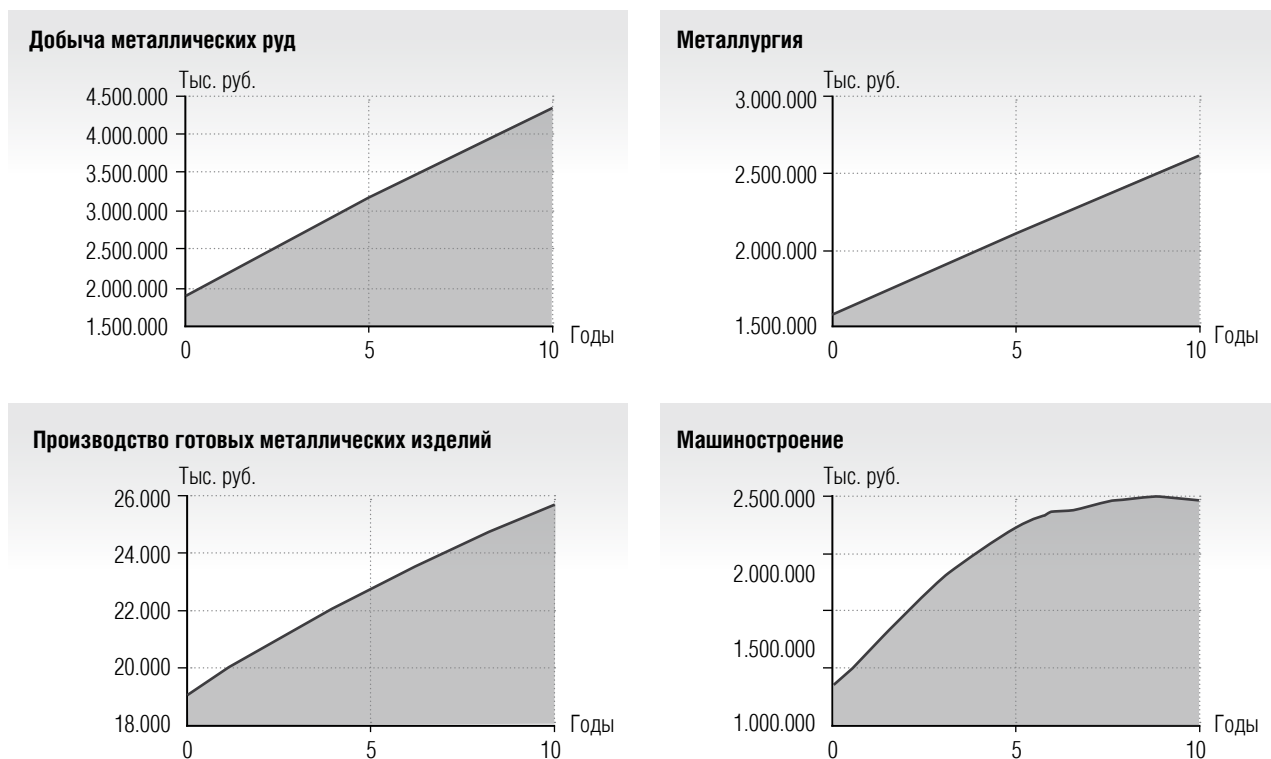
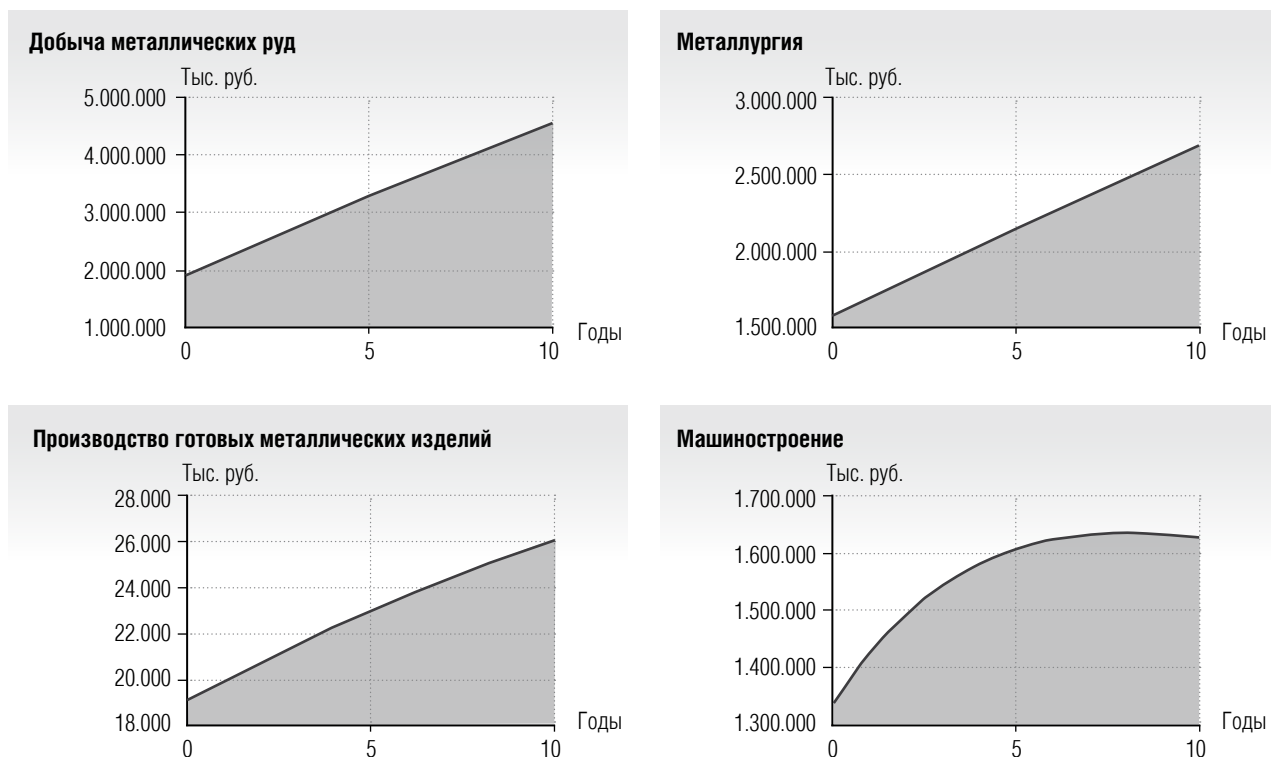


Рис. 6. Динамика изменения валового выпуска продукции по сценарию №3



отрасли ощущают это в меньшей степени, что является весьма необычным явлением. Получается, что исследуемые отрасли компенсируют приток инвестиций из других отраслей промышленности (например, строительства) в большей степени, чем отток работников из отрасли машиностроения.

Пятый сценарий является оптимистическим – увеличение прироста числа работников в промышленности с сохранением их распределения по отраслям на 1% в год (рисунк 8).

Прирост численности работников позитивно сказывается на темпах роста продукции по всем отраслям. Более того, активно начинает работать петля обратной связи, которая связывает выпуск фондообразующих отраслей (в частности, машиностроение) с приростом капитала в другие отрасли. Таким образом, чем дальше мы строим прогноз, тем большую долю в увеличении выпуска продукции занимает прирост капиталовложений в основные фонды. Такой эффект называется самовоспроизводящейся петлей обратной связи.

Последний сценарий в данном исследовании посвящен анализу развития отраслей при увеличении доли инвестиций в отрасль машиностроения на 10% в год (рисунк 9).

Естественно, в данном сценарии также наблюдается увеличение темпов роста выпуска продукции всех отраслей по сравнению со стандартным сценарием №1. Однако, следует отметить, что в первую очередь сильно увеличивается выпуск в отрасли машиностроения – в 3,7 раза за 10 лет. Однако, в металлургических отраслях ускорение темпов роста более медленное по сравнению со сценарием №5. Здесь следует отметить, что, во-первых, прирост трудовых ресурсов в целом увеличивает прирост в большей степени линейно и, во-вторых, прирост инвестиций в одну отрасль согласно балансового уравнения (5) предполагает их отток (хоть и небольшой) из других отраслей. Тем не менее, экспоненциальный рост отрасли машиностроения в более долгосрочной перспективе повлечет за собой аналогичный рост металлургических отраслей, который впоследствии превысит темпы роста в сценарии №5. Однако разработанная имитационная модель не может строить достаточно правдоподобные прогнозы на период, превышающий 10 лет.

Таким образом, продемонстрирована возможность использования разработанной имитационной модели для анализа и прогнозирования развития отраслей металлургии и смежных с ней. Естественно, данный набор сценариев является

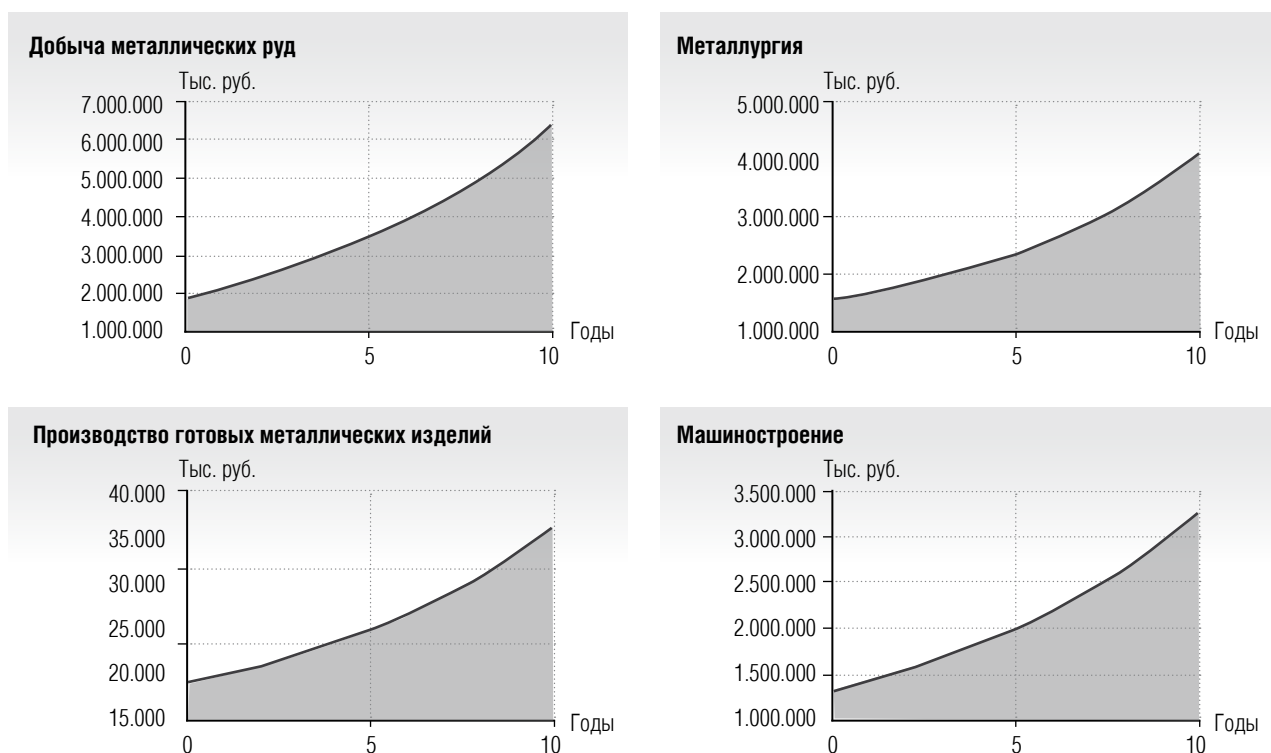


Рис. 8. Динамика изменения валового выпуска продукции по сценарию №5

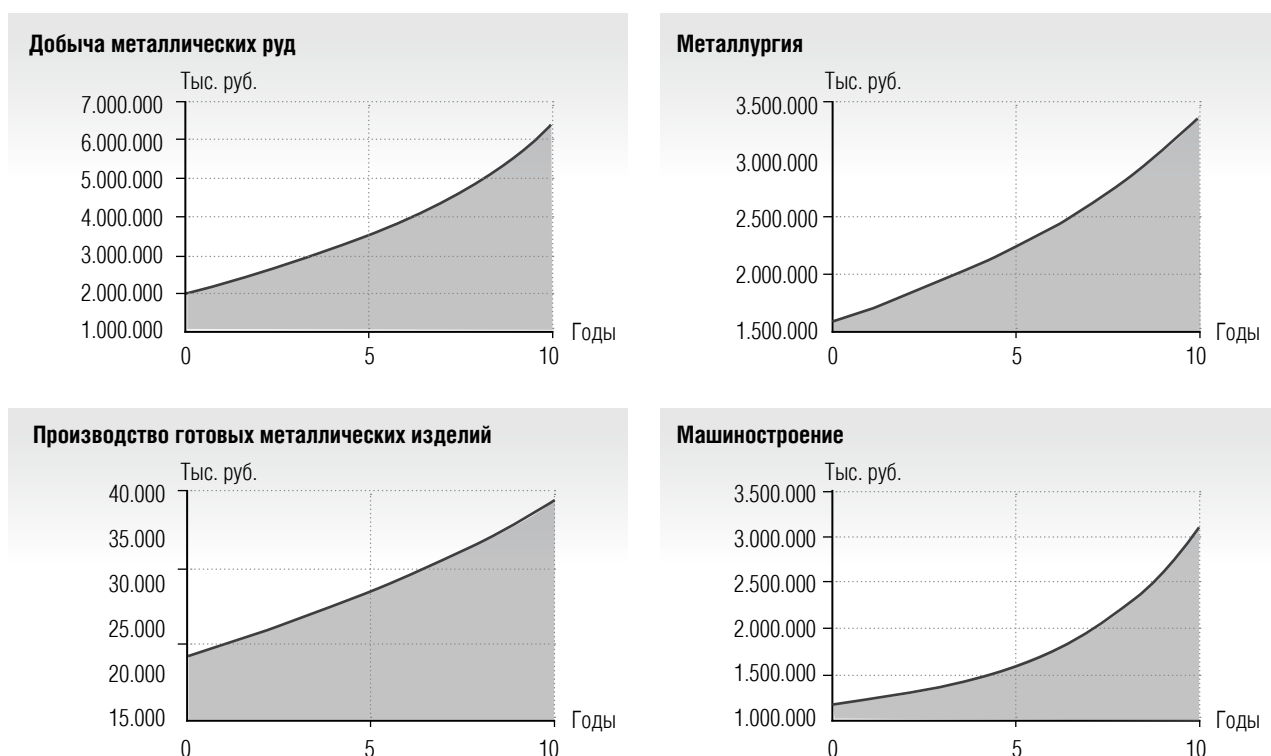


Рис. 9. Динамика изменения валового выпуска продукции по сценарию №6

условным. Тем не менее, имитационная модель позволяет добавлять новые сценарии в соответствии с промышленной политикой РФ.

Заключение

В ходе настоящего исследования, были получены следующие результаты.

Анализ научных работ, посвященных построению имитационных моделей экономических систем, показал, что исследования в этой области крайне актуальны, и их число увеличивается с каждым годом. Однако использование системной динамики в основном ограничено исследованиями либо национальной экономической системы в целом, либо отдельных регионов. Тем не менее, развитие отдельных отраслей также необходимо исследовать с помощью современных инструментальных средств.

Модифицирована трехсекторная модель экономики В.А. Колемаева, где в качестве секторов рассматриваются не глобальные сектора национальной экономики, а отдельные отрасли промышленности: в рамках данной работы это металлургия,

добыча железных руд, машиностроение и производство готовых металлических изделий. Построено пять динамических уравнений первого порядка, три балансовых уравнения и четыре нелинейных функции, которые в совокупности позволяют анализировать, прогнозировать и оптимизировать устойчивое развитие отдельных отраслей промышленности.

На основе аналитической модели металлургических отраслей разработана имитационная модель развития отдельных отраслей промышленности в нотации системной динамики. Определены накопители (основные производственные фонды каждой отрасли и численность работников в промышленности), изменяющие их потоки, а также динамические переменные и параметры модели. Построенная имитационная модель представляет собой комплекс потоковых причинно-следственных диаграмм, отражающих не только структуру исследуемой экономической системы, но и ее поведение в течение нескольких периодов времени.

Построенная имитационная модель апробирована с учетом имеющихся данных за 2002–2018 годы. Построены производственные функции и эмпи-

рически рассчитаны все экзогенные переменные модели. Для демонстрации работы имитационной модели было предложено шесть сценариев развития исследуемой экономической системы – стандартный, три пессимистических и два оптимистических. Сценарный анализ показал основные точки

управления моделью и, как следствие, всей исследуемой экономической системой. Результаты данного исследования могут быть применены, в первую очередь, органами государственной власти для корректировки промышленной политики. ■

Литература

- Орехова С.В. Ресурсы и устойчивый рост промышленного металлургического предприятия: эмпирическая оценка // Современная конкуренция. 2017. Т. 11. № 3 (63). С. 65–76.
- Tretyakova E.A. Evolution of research and evaluation methodology of sustainable development of social and economic systems // World Applied Sciences Journal. 2013. Vol. 25. No 5. P. 756–759. DOI: 10.5829/idosi.wasj.2013.25.05.13335.
- Орехова С.В. Оценка устойчивости экономического роста металлургического комплекса // Вестник НГУЭУ. 2017. № 2. С. 204–220.
- Baranenko S.P., Dudin M.N., Lyasnikov N.V., Busygin K.D. Use of environmental approach to innovation-oriented development of industrial enterprises // American Journal of Applied Sciences. 2014. Vol. 11. No 2. P. 189–194. DOI: 10.3844/ajassp.2014.189.194.
- Socio-economic role of service-sector small business in sustainable development of the Russian economy / L.G. Rudenko [et al.] // European Research Studies Journal. 2015. Vol. 18. No 3. P. 219–234. DOI: 10.35808/ersj/468.
- Dubrovsky V., Yaroshevich N., Kuzmin E. Transactional approach in assessment of operational performance of companies in transport infrastructure // Journal of Industrial Engineering and Management. 2016. Vol. 9. No 2. P. 389–412.
- Орехова С.В., Кислицын Е.В. Малый бизнес и структурные сдвиги в промышленности // Terra Economicus. 2019. Т. 17. № 4. С. 129–147. DOI: 10.23683/2073-6606-2019-17-4-129-147.
- Демидова О.А., Иванов Д.С. Модели экономического роста с неоднородными пространственными эффектами (на примере российских регионов) // Экономический журнал Высшей школы экономики. 2016. Т. 20. № 1. С. 52–75.
- Сурнина Н.М., Илюхин А.А., Илюхина С.В. Демографический ландшафт региона: факторы, динамика, тенденции, прогнозы // Известия Уральского государственного экономического университета. 2017. № 4 (72). С. 32–44. DOI: 10.29141/2073-1019-2017-16-4-3.
- Бархатов В.И., Бенц Д.С. Промышленные рынки Уральского региона: экономический рост в условиях «новой нормальности» // Управленец. 2019. Т. 10. № 3. С. 83–93. DOI: 10.29141/2218-5003-2019-10-3-8.
- Кульков В.М., Кайманаков С.В., Теняков И.М. Экономический рост в России: национальная модель, качество и безопасность // Национальные интересы: приоритеты и безопасность. 2014. Т. 10. № 38 (275). С. 9–19.
- Forrester J.W. Urban dynamics. Waltham, MA: Pegasus Communications, 1969.
- Forrester J.W. Industrial dynamics: A major breakthrough for decision makers // Harvard Business Review. 1958. Vol. 36. No 4. P. 37–66.
- Meadows D.H. Limits to growth: A report for the Club of Rome's project on the predicament of mankind. N.Y.: Universe Books, 1972.
- Meadows D.H., Randers J. Meadows D.L. Limits to growth: The 30 year update. London: Earthscan, 2005.
- Сидоренко В.Н. Системная динамика. М.: ТЕИС, 1998.
- Акопов А.С. Системно-динамический подход в управлении инвестиционной деятельностью нефтяной компании // Аудит и финансовый анализ. 2006. № 2. С. 165–200.
- Akopov A.S. Designing of integrated system-dynamics models for an oil company // International Journal of Computer Applications in Technology. 2012. Vol. 45. No 4. P. 220–230. DOI: 10.1504/IJCAT.2012.051122.
- Акопов А.С., Хачатрян Н.К. Системная динамика. М.: ЦЭМИ, 2014.
- Solow R.M. Contribution to the theory of economic growth // Quarterly Journal of Economics. 1956. Vol. 70. No 1. P. 65–94. DOI: 10.2307/1884513.
- Intriligator M. Mathematical optimization and economic theory. N.Y.: Prentice-Hall, 1971.
- Колемаев В.А. Математическая экономика. М.: ЮНИТИ-ДАТА, 2005.
- Колемаев В.А. Оптимальный сбалансированный рост открытой трехсекторной экономики // Прикладная эконометрика. 2008. Т. 11. № 3. С. 15–42.
- Трофимова В.Ш., Озерова К.А. Экономико-математическое моделирование макроэкономической динамики // Математическое и компьютерное моделирование в экономике, страховании и управлении рисками. 2017. № 2. С. 85–89.
- Кислицын Е.В. Принципы построения имитационной модели рынка с ограниченной конкуренцией (на примере рынка операторов сотовой связи Екатеринбурга) // Вестник Забайкальского государственного университета. 2017. Т. 23. № 10. С. 101–110. DOI: 10.21209/2227-9245-2017-23-10-101-110.

26. Borshchev A., Karpov Y., Kharitonov V. Distributed simulation of hybrid systems with AnyLogic and HLA // *Future Generation Computer Systems*. 2002. Vol. 18. No 6. P. 829–839. DOI: 10.1016/S0167-739X(02)00055-9.
27. Бочаров Е.П., Алексенцева О.Н., Ермошин Д.В. Оценка рисков промышленных предприятий на основе имитационного моделирования // *Прикладная информатика*. 2008. № 1 (13). С. 15–24.
28. Кислицын Е.В. Имитационное моделирование процесса кредитования физических лиц с использованием кредитного рейтинга // *Вестник Воронежского государственного университета*. Серия: Экономика и управление. 2018. № 3. С. 112–118.
29. Воронцовский А.В. Современные подходы к моделированию экономического роста // *Вестник Санкт-Петербургского университета*. Экономика. 2010. № 3. С. 105–119.
30. Шульц Д.Н., Якупова И.Н. Агентное моделирование влияния микроструктуры на свойства экономики // *Журнал экономической теории*. 2016. № 1. С. 70–81.
31. Яндыбаева Н.В., Кушников В.А. Математическая модель для прогнозирования показателей экономической безопасности Российской Федерации // *Вестник Астраханского государственного технического университета*. Серия: Управление, вычислительная техника и информатика. 2014. № 3. С. 93–101.
32. Пискун Е.И. Оценка результативности реализации стратегии инновационного развития производственно-экономических систем // *Вестник Волгоградского государственного университета*. Серия 3: Экономика. Экология. 2014. № 5 (28). С. 55–68. DOI: 10.15688/jvolsu3.2014.5.6.
33. Mathur M., Agarwal S. Sustainability dynamics of resource use and economic growth. A discussion on sustaining the dynamic linkages between renewable natural resources and the economic system / Discussion paper, 2015. [Электронный ресурс]: http://www.teriin.org/policybrief/files/aug15/files/downloads/Discussion_paper_Sustainability_Aug2015.pdf (дата обращения 15.12.2020).
34. Матур М., Агарвал С. Динамика устойчивого взаимодействия ресурсопотребления и экономического роста // *Карельский научный журнал*. 2015. № 4 (13). С. 48–59.
35. Бекларян Г.Л. Система поддержки принятия решений для устойчивого экономического развития Дальневосточного федерального округа // *Бизнес-информатика*. 2018. № 4 (46). С. 66–75. DOI: 10.17323/1998-0663.2018.4.66.75.
36. Машкова А.Л., Савина О.А., Маматов А.В., Новикова Е.В. Компьютерное моделирование процессов экономической динамики в отраслевом разрезе // *Известия Юго-Западного государственного университета*. 2018. Т. 22. № 5. С. 96–108. DOI: 10.21869/2223-1560-2018-22-5-96-108.
37. Машкова А.Л. Прогнозирование долгосрочного развития макроэкономических систем на базе агент-ориентированных моделей // *Государственное управление*. Электронный вестник. 2016. № 57. С. 49–68.
38. Новикова Е.В., Машкова А.Л. Создание первоначального поколения агентов в компьютерной модели отраслевого развития экономики России // *Сборник трудов VII Международной научно-технической конференции «Информационные технологии в науке, образовании и производстве»*. Белгород, 17–19 октября 2018 г. С. 313–318.
39. Молодецкая Е.Ю. Имитационное моделирование систем цепочек поставок на макроэкономическом уровне как инструмент управления экономическим развитием региона // *Экономика: вчера, сегодня, завтра*. 2017. Т. 7. № 9А. С. 180–191.
40. Lychkina N., Molodetskaya E., Morozova Yu. The simulation model of supply chains on the macroeconomic level is the tool to control the economic development of the region // *Strategic Innovative Marketing*. Springer, 2017. P. 357–362. DOI: 10.1007/978-3-319-56288-9_47.
41. Орехова С.В., Кислицын Е.В. Совокупная производительность факторов в промышленности России: малые vs крупные предприятия // *Journal of New Economy*. 2019. Т. 20. № 2. С. 127–144. DOI: 10.29141/2073-1019-2019-20-2-8.

Об авторах

Кислицын Евгений Витальевич

кандидат экономических наук;

и.о. заведующего кафедрой информационных технологий и статистики, Уральский государственный экономический университет, 620144, г. Екатеринбург; ул. 8 Марта, д. 62;

E-mail: kev@usue.ru

ORCID: 0000-0003-1518-0043

Городничев Виктор Владимирович

старший преподаватель кафедры информационных технологий и статистики, Уральский государственный экономический университет, 620144, г. Екатеринбург; ул. 8 Марта, д. 62;

E-mail: helltoaster@yandex.ru

Simulation of development of individual heavy industry sectors

Evgeniy V. Kislitsyn

E-mail: kev@usue.ru

Victor V. Gorodnichev

E-mail: helltoaster@yandex.ru

Ural State University of Economics

Address: 62, 8 Marta Street, Yekaterinburg 620144, Russia

Abstract

Nowadays, in the context of the coronavirus crisis, the issue of ensuring the sustainable development of heavy industries is acute. However, theoretical and analytical researches alone are not sufficient for this, and economic science needs to develop fundamentally new approaches to the study of the development of industrial sectors. This article is devoted to the creation and testing of a simulation model for the development of individual sectors of the economy. The object of research is the metallurgical industry, as well as related ore mining, mechanical engineering and production of finished metal products. The theoretical basis of the research is a systematic approach that combines the theory of industry markets, economic growth, industrial economics, system dynamics and mathematical economics. The main research methods used are system analysis, statistical analysis to identify trends in changes in the main economic indicators, econometric modeling to build production functions, as well as mathematical modeling of macroeconomic systems. As a result, a simulation model developed in system dynamics notation is proposed, which makes it possible to evaluate the development of individual industries taking into account various changes. This model is built on the basis of the three-sector model of the national economy, where separate adjacent industries connected by dynamic feedback loops are identified as structural elements. The paper details the structure of the simulation model based on first-order dynamic equations, balance equations and nonlinear production functions. The simulation model allowed us to predict a number of scenarios for the development of metallurgical industries, taking into account changes in the labor force and investment in fixed assets. The results of the work can be used for forming proposals on industrial policy, monitoring the condition and efficiency of individual industries.

Key words: simulation; system dynamics; industry; metallurgy; three-sector model of the economy; production function; mathematical model.

Citation: Kislitsyn E.V., Gorodnichev V.V. (2021) Simulation of development of individual heavy industry sectors. *Business Informatics*, vol. 15, no 1, pp. 59–77. DOI: 10.17323/2587-814X.2021.1.59.77

References

1. Orekhova S.V. (2017) Resources and sustainable growth of an industrial metallurgical enterprise: an empirical assessment. *Modern Competition*, vol. 11, no 3, pp. 65–76 (in Russian).
2. Tretyakova E.A. (2013) Evolution of research and evaluation methodology of sustainable development of social and economic systems. *World Applied Sciences Journal*, vol. 25, no 5, pp. 756–759. DOI: 10.5829/idosi.wasj.2013.25.05.13335.
3. Orekhova S.V. (2017) Assessment of the economic growth stability of the metallurgical complex. *Vestnik NGUEU*, no 2, pp. 204–220 (in Russian).
4. Baranenko S.P., Dudin M.N., Lyasnikov N.V., Busygin K.D. (2014) Use of environmental approach to innovation-oriented development of industrial enterprises. *American Journal of Applied Sciences*, vol. 11, no 2, pp. 189–194. DOI: 10.3844/ajassp.2014.189.194.
5. Rudenko L.G., Zaitseva N.A., Larionova A.A., Chudnovsky A.D., Vinogradova M.V. (2015) Socio-economic role of service-sector small business in sustainable development of the Russian economy. *European Research Studies Journal*, vol. 18, no 3, pp. 219–234. DOI: 10.35808/ersj/468.
6. Dubrovsky V., Yaroshevich N., Kuzmin E. (2016) Transactional approach in assessment of operational performance of companies in transport infrastructure. *Journal of Industrial Engineering and Management*, vol. 9, no 2, pp. 389–412.

7. Orekhova S.V., Kislitsyn E.V. (2019) Small business and structural changes in industry. *Terra Economicus*, vol. 17, no 4, pp. 129–147 (in Russian). DOI: 10.23683/2073-6606-2019-17-4-129-147.
8. Demidova O.A., Ivanov D.S. (2016) Models of economic growth with heterogeneous spatial effects (on the example of Russian regions). *HSE Economic Journal*, vol. 20, no 1, pp. 52–75.
9. Surnina N.M., Ilyuhin A.A., Ilyuhina S.V. (2017) Demographic landscape of the region: factors, dynamics, trends, forecasts. *Journal of the Ural State University of Economics*, no 4, pp. 32–44 (in Russian). DOI: 10.29141/2073-1019-2017-16-4-3.
10. Barkhatov V.I., Benz D.S. (2019) Industrial markets of the Ural region: Economic growth under “new normal”. *Upravlenets – The Manager*, vol. 10, no 3, pp. 83–93 (in Russian). DOI: 10.29141/2218-5003-2019-10-3-8.
11. Kulkov V.M., Kaimanakov S.V., Tenyakov I.M. (2014) The economic growth in Russia: national model, quality and security. *National Interests: Priorities and Security*, vol. 10, no 38, pp. 9–19 (in Russian).
12. Forrester J.W. (1969) *Urban dynamics*. Waltham, MA: Pegasus Communications.
13. Forrester J.W. (1958) Industrial dynamics: A major breakthrough for decision makers. *Harvard Business Review*, vol. 36, no 4, pp. 37–66.
14. Meadows D.H. (1972) *Limits to growth: A report for the Club of Rome's project on the predicament of mankind*. N.Y.: Universe Books.
15. Meadows D.H., Randers J., Meadows D.L. (2005) *Limits to growth: The 30 year update*. London: Earthscan.
16. Sidorenko V.N. (1998) *System dynamics*. Moscow: TEIS (in Russian).
17. Akopov A.S. (2006) System-dynamic approach to managing the investment activity of an oil company. *Audit and Financial Analysis*, no 2, pp. 165–200 (in Russian).
18. Akopov A.S. (2012) Designing of integrated system-dynamics models for an oil company. *International Journal of Computer Applications in Technology*, vol. 45, no 4, pp. 220–230. DOI: 10.1504/IJCAT.2012.051122.
19. Akopov A.S., Khachatryan N.K. (2014) *System dynamics*. Moscow: CEMI (in Russian).
20. Solow R.M. (1956) Contribution to the theory of economic growth. *Quarterly Journal of Economics*, vol. 70, no 1, pp. 65–94. DOI: 10.2307/1884513.
21. Intriligator M. (1971) *Mathematical optimization and economic theory*. N.Y.: Prentice-Hall.
22. Kolemaev V.A. (2005) *Mathematical economics*. Moscow: UNITY-DANA (in Russian).
23. Kolemaev V.A. (2008) Optimal balanced growth of the open three-sector economy. *Applied Econometrics*, vol. 11, no 3, pp. 15–42 (in Russian).
24. Trofimova V.Sh., Ozerova K.A. (2017) Economic and mathematical modeling of macroeconomic dynamics. *Mathematics and Computer Modeling in Economics, Insurance and Risk Management*, no 2, pp. 85–89 (in Russian).
25. Kislitsyn E.V. (2017) Principles of building a simulation model of a market with limited competition (for example, the market of mobile operators in Yekaterinburg). *Transbaikal State University Journal (Bulletin of ZabGU)*, vol. 23, no 10, pp. 101–110 (in Russian). DOI: 10.21209/2227-9245-2017-23-10-101-110.
26. Borshchev A., Karpov Y., Kharitonov V. (2002) Distributed simulation of hybrid systems with AnyLogic and HLA. *Future Generation Computer Systems*, vol. 18, no 6, pp. 829–839. DOI: 10.1016/S0167-739X(02)00055-9.
27. Bocharov E.P., Aleksenceva O.N., Ermoshin D.V. (2008) Assessment of risks of industrial enterprises on the basis of simulation modeling. *Applied Informatics*, no 1, pp. 15–24 (in Russian).
28. Kislitsyn E.V. (2018) Simulation of the process of crediting individuals using a credit rating. *Proceedings of Voronezh State University. Series: Economics and Management*, no 3, pp. 112–118 (in Russian).
29. Vorontsovskiy A.V. (2010) Modern approaches to modeling of the economic growth. *St. Petersburg University Journal of Economic Studies*, no 3, pp. 105–119 (in Russian).
30. Shults D.N., Yakupova I.N. (2016) Agent-based modeling of the influence of microstructure on economic properties. *Russian Journal of Economic Theory*, no 1, pp. 70–81 (in Russian).
31. Yandybaeva N.V., Kushnikov V.A. (2014) Mathematical model for forecasting of indicators of economic safety of the Russian Federation. *Vestnik of Astrakhan State Technical University. Series: Management, Computer Sciences and Informatics*, no 3, pp. 93–101 (in Russian).
32. Piskun E.I. (2014) The evaluation of effectiveness of the implementation of the strategy of innovative development of industrial and economic systems. *Science Journal of VolsU. Global Economic System*, no 5, pp. 55–68 (in Russian). DOI: 10.15688/jvolsu3.2014.5.6.
33. Mathur M., Agarwal S. (2015) *Sustainability dynamics of resource use and economic growth. A discussion on sustaining the dynamic linkages between renewable natural resources and the economic system. Discussion paper*. Available at: http://www.teriin.org/policybrief/files/aug15/files/downloads/Discussion_paper_Sustainability_Aug2015.pdf (accessed 15 December 2020).
34. Mathur M., Agarwal S. (2015) Dynamics of sustainable interaction between resource consumption and economic growth. *Karelian Scientific Journal*, no 4, pp. 48–59 (in Russian).
35. Beklaryan G.L. (2018) Decision support system for sustainable economic development of the Far Eastern Federal District. *Business Informatics*, no 1, pp. 66–75. DOI: 10.17323/1998-0663.2018.4.66.75.
36. Mashkova A.L., Savina O.A., Mamatov A.V., Novikova E.V. (2018) Computer modeling of sectoral economic dynamics. *Proceedings of the Southwest State University*, vol. 22, no 5, pp. 96–108 (in Russian). DOI: 10.21869/2223-1560-2018-22-5-96-108.
37. Mashkova A.L. (2016) Forecasting long-term development of macroeconomic systems based on agent modeling. *Public Administration. E-Journal*, no 57, pp. 49–68 (in Russian).

38. Novikova E.V., Mashkova A.L. (2018) Creation of the initial generation of agents in the computer model of industrial development of the Russian economy. Proceedings of the VII International Scientific and Technical Conference "Information Technologies in Science, Education and Manufacturing", Belgorod, Russia, 17–19 October 2018, pp. 313–318 (in Russian).
39. Molodetskaya E.Yu. (2017) Simulation of supply chain systems at the macroeconomic level as a tool for managing the economic development of the region. *Economics: Yesterday, Today and Tomorrow*, vol. 7, no 9A, pp. 180–191 (in Russian).
40. Lychkina N., Molodetsk

Big data analysis of IoT-based supply chain management considering FMCG industries

Hamed Nozari^a

Introduction

An intelligent supply chain is an innovative supply chain that utilizes information technology and other technology tools to improve efficiency, improve processes and increase service levels. In today's world, the modern business environment is dealing with data and this has created

1. Literature review

In order to illustrate the effects of big data and IoT concepts on supply chain management, a literature review of big data and IoT technologies is provided in this section.

works, middleware, cloud computing, IoT applications. On the other hand, it is recognized that IoT are faced with several challenges such as scalability, self-Organizing, data volumes, data interpretation, interoperability, automatic discovery, software complexity, security and privacy, fault tolerance, power supply,

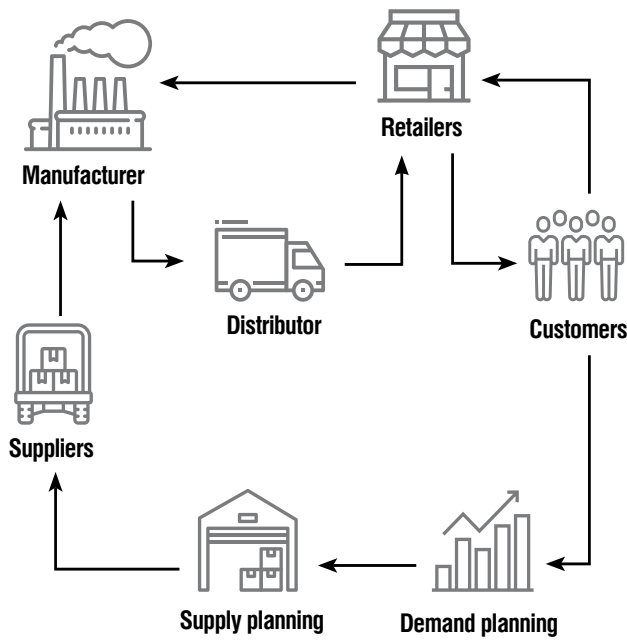
Chen [80] proposed the intelligent IoT-enabled system in green supply chain to simulate complex system by linked physical and digital objects with relationships. Yan et al. [81] illustrated an intelligent supply chain integration and management system to provide flexible and agile approaches to facilitate the resource sharing and participant collaboration in the whole life cycle of supply chain. Parry et al. [82] demonstrated how the IoT may be operationalized in the domestic setting to capture data on a consumer's use of products and the implications for reverse supply chains. Kang et al. [36] proposed a sensor-integrated RFID data repository-implementation model that can integrate and store a large amount of IoT-generated data collected from supply chain processes. Thoben et al. [83] indicated that smart manufacturing can include different technologies such as IoT, robotics/automation, big data analytics and Cloud computing.

Therefore, due to lack of academic publications that address big data analysis of a smart supply chain in an IoT environment, it is important to investigate the applications and interactions of

3. Smart supply chain

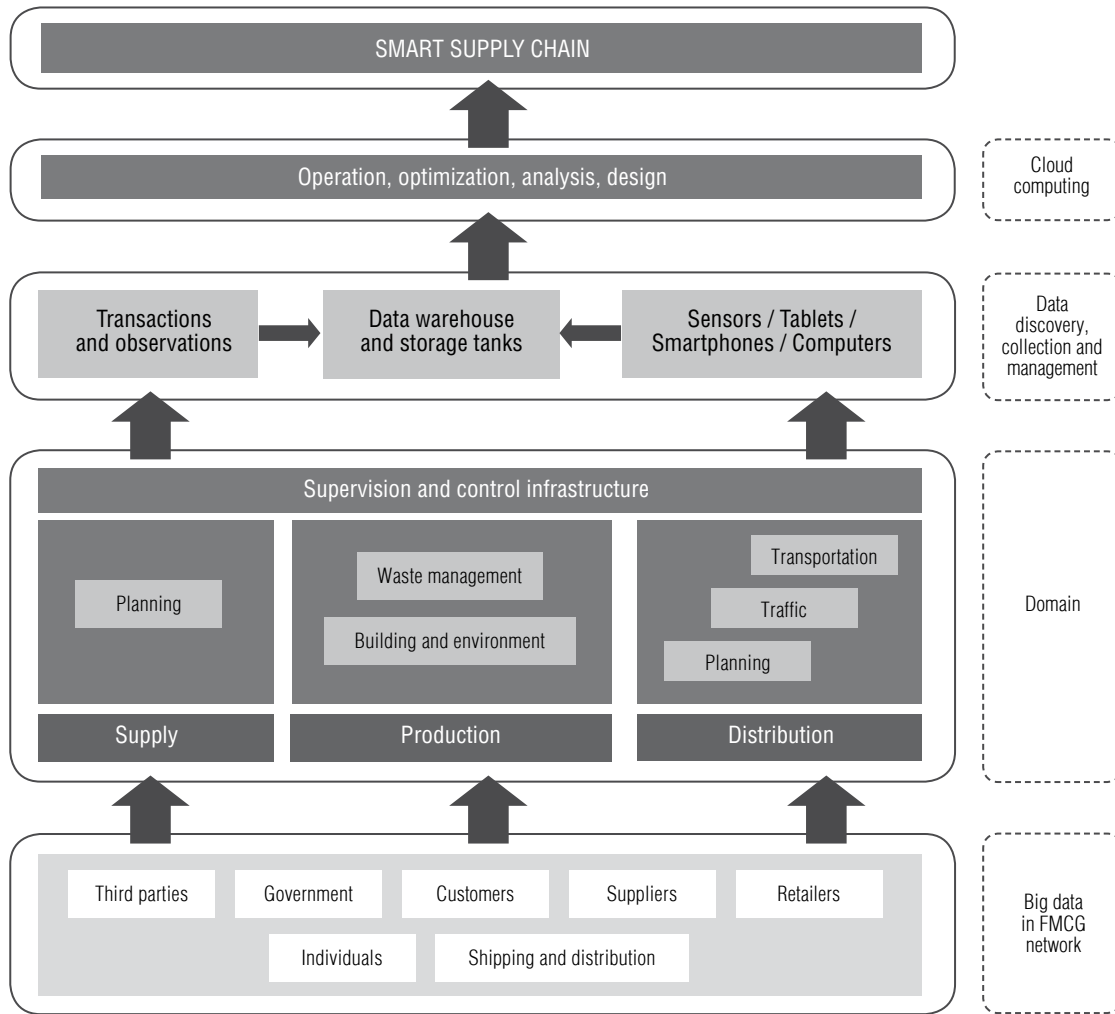
3.1. Supply chain and IoT

Supply chain management integrates supply and demand processes throughout the organization. Supply chain management claims to offer solutions to help industry professionals better manage the entire supply chain from suppliers to end customers. Supply chain systems have been empowered using the internet and this has led to increased efficiency and productivity in the supply chain. In traditional supply chain management systems, there exist several problems such as overstocking, delivery delays and stock out. These problems return to several factors such as complexity and uncertainty as exist in real supply chains [91]. IoT technology can help the industrial system to manage these uncertainties and improve supply chain efficiency [92]. The benefits of managing an electronic supply chain can include operational and strategic improvement in communications, collaboration and coordination to meet organizational barriers [



ment. By the generated data, customers can also make more precise decisions to buy products. It should be noted that every year hundreds of billions value are generated by big data in the FMCG supply chain and that illustrates the importance of big data in this industry.

Logistical data is provided by distribution activities and people involved in processes as well as government information. But one of the most important benefits of using big data for the FMCG industry is that it can be used to enhance the ability of distributors and the timely delivery of goods and products to customers. In addition, companies can optimize many of their activities using this data and consequently they gain a better understanding of the decision



Infrastructure monitoring: IoT devices can be used to improve disaster management, improve emergency response coordination, improve service quality, and reduce operating costs in all infrastructure-related areas. As an ancillary product, IoT infrastructure enables efficient maintenance of planning activities by coordinating tasks between service providers and users of these infrastructures.

With the growing demand for IoT as well as the analysis of big data in smart systems, the use of these technologies also faces challenges. Some of these challenges include:

- ◆ providing security for information;
- ◆ lack of understanding of many features and information;
- ◆ technical challenges;
- ◆ the cost of many technologies;
- ◆ quality assurance and data access;

So in addition to the positive and undeniable effects of using IoT and big data analytics, these challenges must also be addressed.

After the design, the framework was provided and approved by 24 supply chain specialists in FMCG companies. This study was tested using a five-point Likert scale questionnaire and the results showed the acceptability of this framework. It should be noted, however, that the selection of these experts was also challenging. Because these professionals had to be selected among those who had a thorough knowledge of IoT technologies and had a track record of using some of these technologies in their background.

Conclusion

IoT is a new form of pervasive computing and applications for big data that is increasingly being taken into account with operational performance and planning for sustainable development. Therefore, the use of IoT in the supply chain can greatly enhance its intelligence and thus improve its performance. On the other hand, analyzing big data and using it to realize key features of a smart and sustainable supply chain (such as operation and service efficiency, resource optimization, and intelligent infrastructure and facility management) has a huge impact. For this reason, the purpose of this paper was to review and integrate the related literature in order to identify and discuss IoT-based big data applications for supply chain sustainability. In this paper, the feasibility of developing smart supply chains using big data from IoT is explored to achieve the level of re-

quired intelligence. In this paper, the supply chain of FMCG companies was selected. Products in these industries have a special nature and as a result their supply chain has different characteristics. Timely distribution and delivery system is one of the main features of this supply chain. The most important data-driven applications in these supply chains that are enabled through IoT include transportation, dynamics, energy, environment, infrastructure monitoring and management and supply chain planning. Therefore, expanding the intelligence landscape of the smart supply chain using sensor-based big data has great potential to promote environmental sustainability. To this end, a framework has been outlined that provides a lot of information on the smart supply chain.

This framework defines the scope of decision-making in the core parts of the supply chain in the FMCG industry according to their nature. With this framework, the ways in which big data can be accessed through the internet of things (which can itself be a path to sustainability and in some cases being green) are identified in this industry. By placing appropriate analytical tools in the computational parts of the framework (such as statistical analysis and optimization), one can gain a good understanding of the audience's beliefs. Finally, analyzing this data can help to make the organization more intelligent. One of the most important advantages of using this framework can be the production of optimal products based on customer preferences.

The implementation of the framework will always have challenges. One of the problems is always the use of huge volumes of data and therefore the high complexity of analysis. Since data is produced using the IoT technology without any restrictions, maintenance and computation on this data is very time consuming and complex and therefore there is a need for high knowledge in this field. In addition, IoT devices are not always available and can be costly to obtain. Therefore, all of these must always be considered when implementing the framework.

This framework provides the basis for developing research opportunities to integrate this type of supply chain as well as providing analytical insights into future research. This framework can be expanded and operational solutions in analytics and computing can be customized to suit the different industries. Also, providing computational solutions and incorporating them precisely into the framework can provide a clear pathway for solving new challenges. ■

References

1. Evtodieva T.E., Chernova D.V., Ivanova N.V., Wirth J. (2020) The internet of things: Possibilities of application in intelligent supply chain management. *Digital transformation of the economy: Challenges, trends and new opportunities*. Series: Advances in Intelligent Systems and Computing, vol. 908 (S. Ashmarina, A. Mesquita, M. Vochozka, eds.). Cham: Springer, pp. 395–403. DOI: 10.1007/978-3-030-11367-4_38.
2. Tajfar A.H., Gheysari M. (2016) Analysis the effects of internet of things technology in managing supply chain. *International Journal of Information & Communication Technology Research*, vol. 8, no 3, pp. 15–25.
3. Grossmann I.E. (2018) Optimization and management in manufacturing engineering: resource collaborative optimization and management through the internet of things. *Optimization Methods and Software*, vol. 34, no 1, pp. 220–223. DOI: 10.1080/10556788.2018.1527332.
4. Wazid M., Das A.K., Hussain R., Succi G., Rodrigues J.J.P.C. (2019) Authentication in cloud-driven IoT-based big data environment: Survey and outlook. *Journal of Systems Architecture*, vol. 97, pp. 185–196. DOI: 10.1016/j.sysarc.2018.12.005.
5. Bibri S.E. (2015) *The shaping of ambient intelligence and the internet of things: Historicoepistemic, socio-cultural, politico-institutional and eco-environmental dimensions*. Atlantis Press. DOI: 10.2991/978-94-6239-142-0.
6. Bibri S.E., Krogstie J. (2016) On the social shaping dimensions of smart sustainable cities: A study in science, technology, and society. *Sustainable Cities and Society*, vol. 29, p. 219–246. DOI: 10.1016/j.scs.2016.11.004.
7. Muñuzuri J., Onieva L., Cortés P., Guadix J. (2020) Using IoT data and applications to improve port-based intermodal supply chains. *Computers & Industrial Engineering*, vol. 139, article no 105668. DOI: 10.1016/J.CIE.2019.01.042.
8. Borah M.D., Naik V.B., Patgiri R., Bhargav A., Phukan B., Basani S.G.M. (2019) Supply chain management in agriculture using blockchain and IoT. *Advanced applications of blockchain technology*, Studies in big data, vol. 60 (S. Kim, G. Deka, eds). Singapore: Springer, pp. 227–241. DOI: 10.1007/978-981-13-8775-3_11.
9. Dai H.-N., Wang H., Hu G., Wan J., Imran M. (2020) Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. *Enterprise Information Systems*, vol. 14, no 9–10, pp. 1279–1303. DOI: 10.1080/17517575.2019.1633689.
10. Cox M., Ellsworth D. (1997) Application-controlled demand paging for out-of-core visualization. Proceedings of the 8th IEEE Visualization Conference (IEEE Vis 1997), Phoenix, AZ, USA, 19–24 October 1997, pp. 235–244. DOI: 10.1109/VISUAL.1997.663888.
11. Addo-Tenkorang R., Helo P.T. (2016) Big data applications in operations/supply-chain management: A literature review. *Computers & Industrial Engineering*, vol. 101, pp. 528–543. DOI: 10.1016/j.cie.2016.09.023.
12. Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A.H. (2011) *Big data: The next frontier for innovation, competition, and productivity*. Available at: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#> (accessed 15 December 2020).
13. Laney D. (2001) *3D data management: Controlling data volume, velocity, and variety*. Technical report. META Group.
14. Russom P. (2011) *Big data analytics*. TDWI best practice report, 4th quarter. Renton, WA: TWDI.
15. Kwon O., Sim J.M. (2013) Effects of data set features on the performances of classification algorithms. *Expert Systems with Applications*, vol. 40, no 5, pp. 1847–1857. DOI: 10.1016/j.eswa.2012.09.017.
16. McAfee A., Brynjolfsson E. (2012) Big data: the management revolution. *Harvard Business Review*, vol. 90, no 10, pp. 61–67.
17. Oracle (2013) *Oracle: Big data for the enterprise*. White paper. Redwood Shores, CA: Oracle Corp.
18. Wamba S.F., Akter S., Edwards A., Chopin G., Gnanzou D. (2015) How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, vol. 165, pp. 234–246. DOI: 10.1016/j.ijpe.2014.12.031.
19. White M. (2012) Digital workplaces: Vision and reality. *Business Information Review*, vol. 29, no 4, pp. 205–214. DOI: 10.1177/0266382112470412.
20. IDC (2013) *Big data in 2020*. Available at: <https://www.emc.com/leadership/digital-universe/2012iview/big-data-2020.htm> (accessed 15 December 2020).
21. Boyd D., Crawford K. (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, vol. 15, no 5, pp. 662–679. DOI: 10.1080/1369118X.2012.678878.
22. Chen H., Chiang R., Storey V. (2012) Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, vol. 36, no 4, pp. 1165–1188. DOI: 10.2307/41703503.
23. Dremel C., Herterich M., Wulf J., Brocke J. (2020) Actualizing big data analytics affordances: A revelatory case study. *Information & Management*, vol. 57, no 1, article no 103121. DOI: 10.1016/j.im.2018.10.007.
24. Ghasemaghaei M. (2020) The role of positive and negative valence factors on the impact of bigness of data on big data analytics usage. *International Journal of Information Management*, vol. 50, pp. 395–404. DOI: 10.1016/J.IJINFORMGT.2018.12.011.
25. Mikalef P., Krogstie J., Pappas I., Pavlou P. (2020) Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, vol. 57, no 2, article no 103169. DOI: 10.1016/j.im.2019.05.004.
26. Strawn G.O. (2012) Scientific research: How many paradigms? *Educause Review*, vol. 47, no 3, pp. 26–34.
27. Golchha N. (2015) Big data – The information revolution. *International Journal of Applied Research*, vol. 1, no 12, pp. 791–794.
28. Wamba S.F., Akter S. (2015) Big data analytics for supply chain management: A literature review and research agenda. *Enterprise and Organizational Modeling and Simulation (EOMAS 2015)* (J. Barjis, R. Pergl, E. Babkin, eds.). *Lecture Notes in Business Information Processing*, vol. 231, pp. 61–72. DOI: 10.1007/978-3-319-24626-0_5.
29. Mital R., Coughlin J., Canaday M. (2014) Using big data technologies and analytics to predict sensor anomalies. Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS 2014), Wailea, Maui, Hawaii, 15–18 September 2014, pp. 84.
30. Chen M., Mao S., Zhang Y., Leung V.C. (2014) *Big data: Related technologies, challenges and future prospects*. Springer. DOI: 10.1007/978-3-319-06245-7.
31. Marjani M., Nasaruddin F., Gani A., Karim A., Hashem I.A.T., Siddiqua A., Yaqoob I. (2017) Big IoT data analytics: Architecture, opportunities, and open research challenges. *IEEE Access*, vol. 5, pp. 5247–5261.
32. Chen C.P., Zhang C.-Y. (2014) Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, vol. 275, pp. 314–347. DOI: 10.1016/j.ins.2014.01.015.
33. Pfaffl M.W. (2001) A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Research*, vol. 29, no 9, p. e45. DOI: 10.1093/nar/29.9.e45.

34. Jourdan Z., Rainer P.K., Marshall T.E. (2008) Business intelligence: An analysis of the literature. *Information Systems Management*, vol. 25, no 2, pp. 121–131. DOI: 10.1080/10580530801941512.
35. Bifet A., Holmes G., Kirkby R., Pfahringer B. (2010) MOA: Massive online analysis. *Journal of Machine Learning Research*, no 11, pp. 1601–1604.
36. Kang Y.-S., Park I.-H., Rhee J., Lee Y.-H. (2016) MongoDB-based repository design for IoT generated RFID/sensor big data. *IEEE Sensors Journal*, vol. 16, no 2, pp. 485–497. DOI: 10.1109/JSEN.2015.2483499.
37. Jiang L., Xu L.D., Cai H., Jiang Z., Bu F., Xu B. (2014) An IoT-oriented data storage framework in cloud computing platform. *IEEE Transactions on Industrial Informatics*, vol. 10, no 2, pp. 1443–1451. DOI: 10.1109/TII.2014.2306384.
38. O’Leary D.E. (2013) ‘Big data’, the ‘internet of things’ and the ‘internet of signs’. *Intelligent Systems in Accounting, Finance and Management*, vol. 20, no 1, pp. 53–65. DOI: 10.1002/isaf.1336.
39. Hagstrom M. (2012) High-performance analytics fuels innovation and inclusive growth: use big data, hyperconnectivity and speed to intelligence to get true value in the digital economy. *Journal of Advanced Analytics*, no 2, pp. 3–4.
40. Kenny J. (2014) *Big data can have big impact on supply chain management: The use of data analytics is underused in supply chain management to minimize risk exposure*. InsideCounsel.
41. Vasan S. (2014) Impact of big data and analytics in supply chain execution. *Supply Chain Digital*. Available at: <https://www.supplychaindigital.com/logistics-1/impact-big-data-and-analytics-supply-chain-execution> (accessed 15 December 2020).
42. Heudecker N., Buytendijk F., Kart L. (2013) *Survey analysis: Big data adoption in 2013 shows substance behind the hype*. Available at: <https://www.gartner.com/en/documents/2589121/survey-analysis-big-data-adoption-in-2013-shows-substance> (accessed 15 December 2020).
43. Kamble S., Gunasekaran A. (2018) Big data-driven supply chain performance measurement system: a review and framework for implementation. *International Journal of Production Research*, vol. 58, no 1, pp. 65–86. DOI: 10.1080/00207543.2019.1630770.
44. Dubey R., Gunasekaran A., Childe S.J. (2019) Big data analytics capability in supply chain agility: The moderating effect of organizational flexibility. *Management Decision*, vol. 57, no 8, pp. 2092–2112. DOI: 10.1108/MD-01-2018-0119.
45. Mayer-Schönberger V., Cukier K. (2013) *Big data: A revolution that will transform how we live, work, and think*. New York: Eamon Dolan/Mariner Book.
46. Rozados I.V., Tjahjono B. (2014) Big data analytics in supply chain management: Trends and related research. Proceedings of the 6th International Conference on Operations and Supply Chain Management (OSCM), Sanur, Bali, 10–12 December 2014, pp. 1096–1107. DOI: 10.13140/RG.2.1.4935.2563
47. Waller M.A., Fawcett S.E. (2013) Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, vol. 34, no 2, pp. 77–84. DOI: 10.1111/jbl.12010.
48. Bates D.W., Saria S., Ohno-Machado L., Shah A., Escobar G. (2014) Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, vol. 33, no 7, pp. 1123–1131. DOI: 10.1377/hlthaff.2014.0041.
49. Hopkins J. (2016) A comparative study examining academic cohorts with transnational migratory intentions towards Canada and Australia. *Higher Education Quarterly*, vol. 70, no 3, pp. 246–263. DOI: 10.1111/hequ.12091.
50. Kim G.-H., Trimi S., Chung J.-H. (2014) Big-data applications in the government sector. *Communications of the ACM*, vol. 57, no 3, pp. 78–85. DOI: 10.1145/2500873.
51. Sellitto C., Hawking P. (2015) Enterprise systems and data analytics: A fantasy football case study. *International Journal of Enterprise Information Systems*, vol. 11, no 3, pp. 1–12.
52. Hopkins J., Hawking P. (2017) Big data analytics and IoT in logistics: a case study. *International Journal of Logistics Management*, vol. 29, no 2, pp. 575–591. DOI: 10.1108/IJLM-05-2017-0109.
53. Sarma S., Brock D.L., Ashton K. (2000) *The networked physical world: Proposals for engineering the next generation of computing, commerce & automatic-identification*. White paper. Available at: https://cocoa.ethz.ch/downloads/2014/06/None_MIT-AUTOID-WH-001.pdf (accessed 15 December 2020).
54. Greengard S. (2015) *The internet of things*. Cambridge, MA: MIT Press.
55. Xu L.D., He W., Li S. (2014) Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, vol. 10, no 4, pp. 2233–2243. DOI: 10.1109/TII.2014.2300753.
56. Ben-Daya M., Hassini E., Bahroun Z. (2019) Internet of things and supply chain management: a literature review. *International Journal of Production Research*, vol. 57, no 15–16, pp. 4719–4742. DOI: 10.1080/00207543.2017.1402140.
57. Atzori L., Iera A., Morabito G. (2010) The internet of things: A survey. *Computer Networks*, vol. 54, no 15, pp. 2787–2805. DOI: 10.1016/j.comnet.2010.05.010.
58. Gubbi J., Buyya R., Marusic S., Palaniswami M. (2013) Internet of things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, vol. 29, no 7, pp. 1645–1660. DOI: 10.1016/j.future.2013.01.010.
59. Leminen S., Rajahonka M., Wendelin R., Westerlund M. (2020) Industrial internet of things business models in the machine-to-machine context. *Industrial Marketing Management*, vol. 84, pp. 298–311. DOI: 10.1016/j.indmarman.2019.08.008.
60. Zhou K., Liu T., Zhou L. (2015) Industry 4.0: Towards future industrial opportunities and challenges. Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2015), Zhangjiajie, China, 15–17 August 2015, pp. 2147–2152. DOI: 10.1109/FSKD.2015.7382284.
61. Li L. (2011) Application of the internet of thing in green agricultural products supply chain management. Proceedings of the 2011 Fourth International Conference on Intelligent Computation Technology and Automation (ICICTA), Shenzhen, China, 28–29 March 2011, vol. 1, pp. 1022–1025. DOI: 10.1109/ICICTA.2011.256.
62. Kopetz H. (2011) *Real-time systems. Design principles for distributed embedded applications*. Springer.
63. Xia F., Yang L., Wang L., Vinel A. (2012) Internet of things. *International Journal of Communication Systems*, vol. 25, no 9, pp. 1101–1102. DOI: 10.1002/dac.2417.
64. Wortmann F., Flüchter K. (2015) Internet of things. *Business & Information Systems Engineering*, vol. 57, no 3, pp. 221–224. DOI: 10.1007/s12599-015-0383-3.
65. Postcapes (2017) *IoT standards and protocols*. Available at: <https://www.postcapes.com/internet-of-thingsprotocols/> (accessed 15 September 2020).
66. Lee I., Lee K. (2015) The internet of things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, vol. 58, no 4, pp. 431–440. DOI: 10.1016/j.bushor.2015.03.008
67. Zeinab K.A.M., Elmustafa S.A.A. (2017) Internet of things applications, challenges and related future technologies. *World Scientific News*, vol. 67, no 2, pp. 126–148.

68. Davies R. (2015) *The internet of things opportunities and challenges*. Available at: <https://epthinktank.eu/2015/05/21/the-internet-of-things-opportunities-and-challenges/> (accessed 15 December 2020).
69. Porter M.E., Heppelmann J.E. (2014) How smart, connected products are transforming competition. *Harvard Business Review*, vol. 92, pp. 11–64.
70. Uckelmann D., Harrison M., Michahelles F. (2011) An architectural approach towards the future internet of things. *Architecting the internet of things*. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-19157-2_1.
71. Sun C. (2012) Application of RFID technology for logistics on internet of things. *AASRI Procedia*, vol. 1, pp. 106–111. DOI: 10.1016/j.aasri.2012.06.019.
72. Bandyopadhyay D., Sen J. (2011) Internet of things applications, challenges and related future technologies. *Wireless Personal Communications*, no 58, pp. 49–69. DOI: 10.1007/s11277-011-0288-5.
73. Langley D.L., van Doorn J., Ng I.C.L., Stieglitz S., Lazovik A., Boonstra A. (2020) The internet of everything: Smart things and their impact on business models. *Journal of Business Research*, vol. 122, pp. 853–863. DOI: 10.1016/j.jbusres.2019.12.035.
74. Yuvaraj S., Sangeetha M. (2016) Smart supply chain management using internet of things (IoT) and low power wireless communication systems. *Proceedings of the IEEE 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 23–25 March 2016*, pp. 555–558. DOI: 10.1109/WiSPNET.2016.7566196.
75. Tao F., Ying Z., Xu L.D., Zhang L. (2014) IoT-based intelligent perception and access of manufacturing resource toward cloud manufacturing. *IEEE Transactions on Industrial Informatics*, vol. 10, no 2, pp. 1547–1557. DOI: 10.1109/TII.2014.2306397.
76. Gnimpieba Z.D.R., Nait-Sidi-Moh A., Durand D., Fortin J. (2015) Using internet of things technologies for a collaborative supply chain: Application to tracking of pallets and containers. *Procedia Computer Science*, vol. 56, pp. 550–557. DOI: 10.1016/j.procs.2015.07.251.
77. Verdouw C.N., Robbemon R.M., Verwaart T., Wolfert J., Beulens A.J.M. (2018) A reference architecture for IoT-based logistic information systems in agri-food supply chains. *Enterprise Information Systems*, vol. 12, no 7, pp. 755–779. DOI: 10.1080/17517575.2015.1072643.
78. Decker C., et al. (2008) Cost-benefit model for smart items in the supply chain. *The internet of things* (C. Floerkemeier, M. Langheinrich, E. Fleisch, F. Mattern, S.E. Sarma, eds.). *Lecture Notes in Computer Science*, vol. 4952, pp. 155–172. DOI: 10.1007/978-3-540-78731-0_10.
79. Xu L.D. (2011) Information architecture for supply chain quality management. *International Journal of Production Research*, vol. 49, no 1, pp. 183–198. DOI: 10.1080/00207543.2010.508944.
80. Chen R.-Y. (2015) Intelligent IoT-enabled system in green supply chain using integrated FCM method. *International Journal of Business Analytics*, vol. 2, no 3, pp. 47–66. DOI: 10.4018/IJBAN.2015070104.
81. Yan J., Xin S., Liu Q., Xu W., Yang L., Fan L., Chen B., Wang Q. (2014) Intelligent supply chain integration and management based on cloud of things. *International Journal of Distributed Sensor Networks*, vol. 10, no 3, article no 624839. DOI: 10.1155/2014/624839.
82. Parry G.C., Brax S.A., Maull R.S., Ng I.C.L. (2016) Operationalising IoT for reverse supply: the development of use-visibility measures. *Supply Chain Management*, vol. 21, no 2, pp. 228–244. DOI: 10.1108/SCM-10-2015-0386.
83. Thoben K.D., Wiesner S., Wuest T. (2017) “Industrie 4.0” and smart manufacturing – A review of research issues and application examples. *International Journal of Automation Technology*, vol. 11, no 1, pp. 4–16.
84. Fleisch E. (2010) *What is the internet of things? An economic perspective, business process and applications*. White Paper WP-BIZAPP-053. Available at: https://cocoa.ethz.ch/downloads/2014/06/None_AUTOIDLABS-WP-BIZAPP-53.pdf (accessed 15 December 2020).
85. Abdel-Basset M., Nabeeh N.A., El-Ghareeb H.A., Aboelfetouh A. (2020) Utilising neutrosophic theory to solve transition difficulties of IoT-based enterprises. *Enterprise Information Systems*, vol. 14, no 9–10, pp. 1304–1324. DOI: 10.1080/17517575.2019.1633690.
86. Gonzalez J. (2011) *The impact of the internet of things on business and society*. Foundation de la Innovation Bankinter.
87. Poirier C.C., Bauer M.J. (2000) *E-supply chain: Using the internet to revolutionize your business*. San Francisco: Berrett-Koehler.
88. Vervest P., Preiss K., van Heck E., Pau L.-F. (2004) The emergence of smart business networks. *Journal of Information Technology*, vol. 19, no 4, pp. 228–233. DOI: 10.1057/palgrave.jit.2000024.
89. Wirtz B.W., Weyerer J.C., Schichtel F.T. (2019) An integrative public IoT framework for smart government. *Government Information Quarterly*, vol. 36, no 2, pp. 333–345. DOI: 10.1016/j.giq.2018.07.001.
90. Janssen M., Luthra S., Mangla S., Rana N.P., Dwivedi Y.K. (2019) Challenges for adopting and implementing IoT in smart cities: An integrated MICMAC-ISM approach. *Internet Research*, vol. 29, no 6, pp. 1589–1616. DOI: 10.1108/INTR-06-2018-0252.
91. Abdel-Basset M., Manogaran G., Mohamed M. (2018) Internet of things (IoT) and its impact on supply chain: A framework for building smart, secure and efficient systems. *Future Generation Computer Systems*, vol. 86, pp. 614–628. DOI: 10.1016/j.future.2018.04.051.
92. Rong K., Hu G., Lin Y., Shi Y., Guo L. (2015) Understanding business ecosystem using a 6C framework in Internet-of-Things-based sectors. *International Journal of Production Economics*, vol. 159, pp. 41–55. DOI: 10.1016/j.ijpe.2014.09.003.
93. Cegielski C.G., Jones Farmer L.A., Wu Y., Hazen B.T. (2012) Adoption of cloud computing technologies in supply chains. *International Journal of Logistics Management*, vol. 32, no 2, pp. 184–211. DOI: 10.1108/09574091211265350.
94. Suguna S.K., Kumar S.N. (2019) Application of cloud computing and internet of things to improve supply chain processes. *Edge computing: From hype to reality*. Springer, pp. 145–170.
95. Arlbjorn J.S., de Haas H.D., Munksgaard K.B. (2011) Exploring supply chain innovation. *Logistics Research*, vol. 3, no 1, pp. 3–18. DOI: 10.1007/s12159-010-0044-3.
96. Lee I. (2015) The Internet of Things (IoT) for supply chain innovation: a conceptual framework and analysis of Fortune 200 companies. *Asia Pacific Journal of Innovation and Entrepreneurship*, vol. 9, no 1, pp. 81–103.
97. Nozari H., Najafi E., Fallah M., Lotfi F.H. (2019) Quantitative analysis of key performance indicators of green supply chain in FMCG industries using non-linear fuzzy method. *Mathematics*, vol. 7, no 11, article no 1020. DOI: 10.3390/math7111020.
98. Kaur J., Sidhu R., Awasthi A., Srivastava S.K. (2019) A Pareto investigation on critical barriers in green supply chain management. *International Journal of Management Science and Engineering Management*, vol. 14, no 2, pp. 113–123. DOI: 10.1080/17509653.2018.1504237.
99. Rehman M.H., Yaqoob I., Salah K., Imran M., Jayaraman P.P., Perera C. (2019) The role of big data analytics in industrial Internet of Things. *Future Generation Computer Systems*, no 99, pp. 247–259. DOI: 10.1016/j.future.2019.04.020.
100. Yu M., Nagurney A. (2013) Competitive food supply chain networks with application to fresh produce. *European Journal of Operational Research*, vol. 224, no 2, pp. 273–282. DOI: 10.1016/j.ejor.2012.07.033.
101. Ji G., Hu L., Tan K.H. (2017) A study on decision-making of food supply chain based on big data. *Journal of Systems Science and Systems Engineering*, no 26, pp. 183–198. DOI: 10.1007/s11518-016-5320-6.
102. Osman A.M.S. (2019) A novel big data analytics framework for smart cities. *Future Generation Computer Systems*, no 91, pp. 620–633.

About the authors

Hamed Nozari

Ph.D.;

Lecturer, Department of Industrial Engineering, Islamic Azad University, Central Tehran Branch, Hamila Blvd., Poonak Sqr., Tehran 1469669191, Iran;

E-mail: Ham.nozari.eng@iauctb.ac.ir

ORCID: 0000-0002-6500-6708

Mohammad Fallah

Ph.D.;

Associated Professor, Department of Industrial Engineering, Islamic Azad University, Central Tehran Branch, Hamila Blvd., Poonak Sqr., Tehran 1469669191, Iran;

E-mail: Mohammad.fallah43@yahoo.com

ORCID: 0000-0001-5541-4284

Hamed Kazemipoor

Ph.D.;

Assistant Professor, Department of Industrial Engineering, Islamic Azad University, Central Tehran Branch, Hamila Blvd., Poonak Sqr., Tehran, 1469669191 Iran;

E-mail: Hkazemipoor@yahoo.com

ORCID: 0000-0002-1848-8927

Seyed Esmail Najafi

Ph.D.;

Assistant Professor, Department of Industrial Engineering, Islamic Azad University, Science and Research Branch, Daneshgah Blvd., Simon Bulivar Blvd., Tehran 1477893855, Iran;

E-mail: E.najafi@srbiau.ac.ir

ORCID: 0000-0002-3125-4439

Анализ больших данных в управлении цепями поставок на основе интернета вещей в отраслях FMCG

Х. Нозари^a

E-mail: Ham.nozari.eng@iauctb.ac.ir

М. Фаллах^a

E-mail: Mohammad.fallah43@yahoo.com

Х. Каземипур^a

E-mail: Hkazemipoor@yahoo.com

С.Э. Найяфи^b

E-mail: E.najafi@srbiau.ac.ir

^a Islamic Azad University, Central Tehran Branch

Адрес: Hamila Blvd., Poonak Sqr., Tehran 1469669191, Iran

^b Islamic Azad University, Tehran Branch, Science and Research

Адрес: Daneshgah Blvd., Simon Bulivar Blvd., Tehran 1477893855, Iran

Аннотация

Цепи поставок являются одним из наиболее существенных элементов производственных систем и промышленных компаний, и их «умные» характеристики помогают бизнесу стать интеллектуальным. Поэтому использование инновационных технологий, направленных на повышение бизнес-интеллекта, представляет собой актуальную задачу. «Умная» цепь поставок использует инновационные инструменты для повышения качества, увеличения производительности и облегчения процесса принятия решений. Интернет вещей (internet of things, IoT) является одним из ключевых элементов ИТ-инфраструктуры для развития интеллектуальных цепей поставок, обладающих высоким потенциалом для обеспечения устойчивости производственных систем. Кроме того, интернет вещей является одним из важнейших источников больших данных. Большие данные и стратегии анализа данных, предоставляющие широкие возможности для оптимизации решений и повышения производительности, в настоящее время быстро развиваются. В связи с этим в данной статье предпринята попытка рассмотреть информативные стратегии развития цепей поставок (на основе частного случая – цепей поставок в отраслях FMCG), а также разработать аналитический подход к построению устойчивых «умных» цепей поставок с использованием анализа больших данных на основе интернета вещей. Предлагаемый подход основан на методологии внедрения интернета вещей, с акцентом на использование исходных больших данных и экспертных оценок. Учитывая характер индустрии FMCG, это может привести к существенным улучшениям в области производственных решений.

Ключевые слова: большие данные; интернет вещей; управлении цепями поставок на основе интернета вещей; цепи поставок отраслей FMCG.

Цитирование: Nozari H., Fallah M., Kazemipoor H., Najafi S.E. Big data analysis of IoT-based supply chain management considering FMCG industries // Business Informatics. 2021. Vol. 15. No 1. P. 78–96.
DOI: 10.17323/2587-814X.2021.1.78.96

Литература

1. Evtodiev T.E., Chernova D.V., Ivanova N.V., Wirth J. The internet of things: Possibilities of application in intelligent supply chain management // Digital transformation of the economy: Challenges, trends and new opportunities. Series: Advances in Intelligent Systems and Computing. Vol. 908 (S. Ashmarina, A. Mesquita, M. Vochozka, eds.). Cham: Springer, 2020. P. 395–403. DOI: 10.1007/978-3-030-11367-4_38.
2. Tajfar A.H., Gheysari M. Analysis the effects of internet of things technology in managing supply chain // International Journal of Information & Communication Technology Research. 2016. Vol. 8. No 3. P. 15–25.
3. Grossmann I.E. Optimization and management in manufacturing engineering: resource collaborative optimization and management through the internet of things // Optimization Methods and Software. 2018. Vol. 34. No 1. P. 220–223. DOI: 10.1080/10556788.2018.1527332.
4. Authentication in cloud-driven IoT-based big data environment: Survey and outlook / M. Wazid [et al.] // Journal of Systems Architecture. 2019. Vol. 97. P. 185–196. DOI: 10.1016/j.sysarc.2018.12.005.
5. Bibri S.E. The shaping of ambient intelligence and the internet of things: Historicoepistemic, socio-cultural, politico-institutional and eco-environmental dimensions. Atlantis Press, 2015. DOI: 10.2991/978-94-6239-142-0.
6. Bibri S.E., Krogstie J. On the social shaping dimensions of smart sustainable cities: A study in science, technology, and society // Sustainable Cities and Society. 2016. Vol. 29. P. 219–246. DOI: 10.1016/j.scs.2016.11.004.
7. Muñuzuri J., Onieva L., Cortés P., Guadix J. Using IoT data and applications to improve port-based intermodal supply chains // Computers & Industrial Engineering. 2020. Vol. 139. Article no 105668. DOI: 10.1016/j.cie.2019.01.042.
8. Supply chain management in agriculture using blockchain and IoT / M.D. Borah [et al.] // Advanced applications of blockchain technology, Studies in big data. Vol. 60 (S. Kim, G. Deka, eds). Singapore: Springer, 2019. P. 227–241. DOI: 10.1007/978-981-13-8775-3_11.
9. Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies / H.-N. Dai [et al.] // Enterprise Information Systems. 2020. Vol. 14. No 9–10. P. 1279–1303. DOI: 10.1080/17517575.2019.1633689.
10. Cox M., Ellsworth D. Application-controlled demand paging for out-of-core visualization // Proceedings of the 8th IEEE Visualization Conference (IEEE Vis 1997). Phoenix, AZ, USA, 19–24 October 1997. P. 235–244. DOI: 10.1109/VISUAL.1997.663888.
11. Addo-Tenkorang R., Helo P.T. Big data applications in operations/supply-chain management: A literature review // Computers & Industrial Engineering. 2016. Vol. 101. P. 528–543. DOI: 10.1016/j.cie.2016.09.023.
12. Big data: The next frontier for innovation, competition, and productivity / J. Manyika [et al.], 2011. [Электронный ресурс]: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#> (дата обращения 15.12.2020).
13. Laney D. 3D data management: Controlling data volume, velocity, and variety. Technical report. META Group, 2001.
14. Russom P. Big data analytics. TDWI best practice report, 4th quarter. Renton, WA: TWDI, 2011.
15. Kwon O., Sim J.M. Effects of data set features on the performances of classification algorithms // Expert Systems with Applications. 2013. Vol. 40. No 5. P. 1847–1857. DOI: 10.1016/j.eswa.2012.09.017.
16. McAfee A., Brynjolfsson E. Big data: the management revolution // Harvard Business Review. 2012. Vol. 90. No 10. P. 61–67.
17. Oracle: Big data for the enterprise. White paper. Redwood Shores, CA: Oracle Corp., 2013.
18. How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study / S.F. Wamba [et al.] // International Journal of Production Economics. 2015. Vol. 165. P. 234–246. DOI: 10.1016/j.ijpe.2014.12.031.

19. White M. Digital workplaces: Vision and reality // *Business Information Review*. 2012. Vol. 29. No 4. P. 205–214. DOI: 10.1177/0266382112470412.
20. Big data in 2020 / IDC, (2013). [Электронный ресурс]: <https://www.emc.com/leadership/digital-universe/2012view/big-data-2020.htm> (дата обращения 15.12.2020).
21. Boyd D., Crawford K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon // *Information, Communication & Society*. 2012. Vol. 15. No 5. P. 662–679. DOI: 10.1080/1369118X.2012.678878.
22. Chen H., Chiang R., Storey V. Business intelligence and analytics: From big data to big impact // *MIS Quarterly*. 2012. Vol. 36. No 4. P. 1165–1188. DOI: 10.2307/41703503.
23. Dremel C., Herterich M., Wulf J., Brocke J. Actualizing big data analytics affordances: A revelatory case study // *Information & Management*. 2020. Vol. 57. No 1. Article no 103121. DOI: 10.1016/j.im.2018.10.007.
24. Ghasemaghaei M. The role of positive and negative valence factors on the impact of bigness of data on big data analytics usage // *International Journal of Information Management*. 2020. Vol. 50. P. 395–404. DOI: 10.1016/j.IJINFOMGT.2018.12.011.
25. Mikalef P., Krogstie J., Pappas I., Pavlou P. Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities // *Information & Management*. 2020. Vol. 57. No 2. Article no 103169. DOI: 10.1016/j.im.2019.05.004.
26. Strawn G.O. Scientific research: How many paradigms? // *Educause Review*. 2012. Vol. 47. No 3. P. 26–34.
27. Golchha N. Big data – The information revolution // *International Journal of Applied Research*. 2015. Vol. 1. No 12. P. 791–794.
28. Wamba S.F., Akter S. Big data analytics for supply chain management: A literature review and research agenda // *Enterprise and Organizational Modeling and Simulation (EOMAS 2015)* (J. Barjis, R. Pergl, E. Babkin, eds.). Lecture Notes in Business Information Processing. 2015. Vol. 231. P. 61–72. DOI: 10.1007/978-3-319-24626-0_5.
29. Mital R., Coughlin J., Canaday M. Using big data technologies and analytics to predict sensor anomalies // *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS 2014)*. Wailea, Maui, Hawaii, 15–18 September 2014. P. 84.
30. Chen M., Mao S., Zhang Y., Leung V.C. Big data: Related technologies, challenges and future prospects. Springer, 2014. DOI: 10.1007/978-3-319-06245-7.
31. Big IoT data analytics: Architecture, opportunities, and open research challenges / M. Marjani [et al.] // *IEEE Access*. 2017. Vol. 5. P. 5247–5261.
32. Chen C.P., Zhang C.-Y. Data-intensive applications, challenges, techniques and technologies: A survey on big data // *Information Sciences*. 2014. Vol. 275. P. 314–347. DOI: 10.1016/j.ins.2014.01.015.
33. Pfaffl M.W. A new mathematical model for relative quantification in real-time RT–PCR // *Nucleic Acids Research*. 2001. Vol. 29. No 9. P. e45. DOI: 10.1093/nar/29.9.e45.
34. Jourdan Z., Rainer P.K., Marshall T.E. Business intelligence: An analysis of the literature // *Information Systems Management*. 2008. Vol. 25. No 2. P. 121–131. DOI: 10.1080/10580530801941512.
35. Bifet A., Holmes G., Kirkby R., Pfahringer B. MOA: Massive online analysis // *Journal of Machine Learning Research*. 2010. No 11. P. 1601–1604.
36. Kang Y.-S., Park I.-H., Rhee J., Lee Y.-H. MongoDB-based repository design for IoT generated RFID/sensor big data // *IEEE Sensors Journal*. 2016. Vol. 16. No 2. P. 485–497. DOI: 10.1109/JSEN.2015.2483499.
37. An IoT-oriented data storage framework in cloud computing platform / L. Jiang [et al.] // *IEEE Transactions on Industrial Informatics*. 2014. Vol. 10. No 2. P. 1443–1451. DOI: 10.1109/TII.2014.2306384.
38. O’Leary D.E. ‘Big data’, the ‘internet of things’ and the ‘internet of signs’ // *Intelligent Systems in Accounting, Finance and Management*. 2013. Vol. 20. No 1. P. 53–65. DOI: 10.1002/isaf.1336.
39. Hagstrom M. High-performance analytics fuels innovation and inclusive growth: use big data, hyperconnectivity and speed to intelligence to get true value in the digital economy // *Journal of Advanced Analytics*. 2012. No 2. P. 3–4.
40. Kenny J. Big data can have big impact on supply chain management: The use of data analytics is underused in supply chain management to minimize risk exposure. InsideCounsel, 2014.
41. Vasan S. Impact of big data and analytics in supply chain execution. Supply Chain Digital / 2014. [Электронный ресурс]: <https://www.supplychaindigital.com/logistics-1/impact-big-data-and-analytics-supply-chain-execution> (дата обращения 15.12.2020).
42. Heudecker N., Buytendijk F., Kart L. Survey analysis: Big data adoption in 2013 shows substance behind the hype / 2013. [Электронный ресурс]: <https://www.gartner.com/en/documents/2589121/survey-analysis-big-data-adoption-in-2013-shows-substance> (дата обращения 15.12.2020).
43. Kamble S., Gunasekaran A. Big data-driven supply chain performance measurement system: a review and framework for implementation // *International Journal of Production Research*. 2018. Vol. 58. No 1. P. 65–86. DOI: 10.1080/00207543.2019.1630770.
44. Dubey R., Gunasekaran A., Childe S.J. Big data analytics capability in supply chain agility: The moderating effect of organizational flexibility // *Management Decision*. 2019. Vol. 57. No 8. P. 2092–2112. DOI: 10.1108/MD-01-2018-0119.
45. Mayer-Schönberger V., Cukier K. Big data: A revolution that will transform how we live, work, and think. New York: Eamon Dolan/Mariner Book, 2013.
46. Rozados I.V., Tjahjono B. Big data analytics in supply chain management: Trends and related research // *Proceedings of the 6th International Conference on Operations and Supply Chain Management (OSCM)*. Sanur, Bali, 10–12 December 2014. P. 1096–1107. DOI: 10.13140/RG.2.1.4935.2563
47. Waller M.A., Fawcett S.E. Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management // *Journal of Business Logistics*. 2013. Vol. 34. No 2. P. 77–84. DOI: 10.1111/jbl.12010.
48. Big data in health care: using analytics to identify and manage high-risk and high-cost patients / D.W. Bates [et al.] // *Health Affairs*. 2014. Vol. 33. No 7. P. 1123–1131. DOI: 10.1377/hlthaff.2014.0041.
49. Hopkins J. A comparative study examining academic cohorts with transnational migratory intentions towards Canada and Australia // *Higher Education Quarterly*. 2016. Vol. 70. No 3. P. 246–263. DOI: 10.1111/hequ.12091.
50. Kim G.-H., Trimi S., Chung J.-H. Big-data applications in the government sector // *Communications of the ACM*. 2014. Vol. 57. No 3. P. 78–85. DOI: 10.1145/2500873.
51. Sellitto C., Hawking P. Enterprise systems and data analytics: A fantasy football case study // *International Journal of Enterprise Information Systems*. 2015. Vol. 11. No 3. P. 1–12.
52. Hopkins J., Hawking P. Big data analytics and IoT in logistics: a case study // *International Journal of Logistics Management*. 2017. Vol. 29. No 2. P. 575–591. DOI: 10.1108/IJLM-05-2017-0109.

53. Sarma S., Brock D.L., Ashton K. The networked physical world: Proposals for engineering the next generation of computing, commerce & automatic-identification. White paper, 2000. [Электронный ресурс]: https://cocoa.ethz.ch/downloads/2014/06/none_MIT-AUTOID-WH-001.pdf (дата обращения 15.12.2020).
54. Greengard S. The internet of things. Cambridge, MA: MIT Press, 2015.
55. Xu L.D., He W., Li S. Internet of things in industries: A survey // *IEEE Transactions on Industrial Informatics*. 2014. Vol. 10. No 4. P. 2233–2243. DOI: 10.1109/TII.2014.2300753.
56. Ben-Daya M., Hassini E., Bahroun Z. Internet of things and supply chain management: a literature review // *International Journal of Production Research*. 2019. Vol. 57. No 15–16. P. 4719–4742. DOI: 10.1080/00207543.2017.1402140.
57. Atzori L., Iera A., Morabito G. The internet of things: A survey // *Computer Networks*. 2010. Vol. 54. No 15. P. 2787–2805. DOI: 10.1016/j.comnet.2010.05.010.
58. Gubbi J., Buyya R., Marusic S., Palaniswami M. Internet of things (IoT): A vision, architectural elements, and future directions // *Future Generation Computer Systems*. 2013. Vol. 29. No 7. P. 1645–1660. DOI: 10.1016/j.future.2013.01.010.
59. Leminen S., Rajahonka M., Wendelin R., Westerlund M. Industrial internet of things business models in the machine-to-machine context // *Industrial Marketing Management*. 2020. Vol. 84. P. 298–311. DOI: 10.1016/j.indmarman.2019.08.008.
60. Zhou K., Liu T., Zhou L. Industry 4.0: Towards future industrial opportunities and challenges // *Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2015)*. Zhangjiajie, China, 15–17 August 2015. P. 2147–2152. DOI: 10.1109/FSKD.2015.7382284.
61. Li L. Application of the internet of thing in green agricultural products supply chain management // *Proceedings of the 2011 Fourth International Conference on Intelligent Computation Technology and Automation (ICICTA)*. Shenzhen, China, 28–29 March 2011. Vol. 1. P. 1022–1025. DOI: 10.1109/ICICTA.2011.256.
62. Kopetz H. Real-time systems. Design principles for distributed embedded applications. Springer, 2011.
63. Xia F., Yang L., Wang L., Vinel A. Internet of things // *International Journal of Communication Systems*. 2012. Vol. 25. No 9. P. 1101–1102. DOI: 10.1002/dac.2417.
64. Wortmann F., Flüchter K. Internet of things // *Business & Information Systems Engineering*. 2015. Vol. 57. No 3. P. 221–224. DOI: 10.1007/s12599-015-0383-3.
65. IoT standards and protocols / Postcapes, 2017. [Электронный ресурс]: <https://www.postscapes.com/internet-of-thingsprotocols/> (дата обращения 15.09.2020).
66. Lee I., Lee K. The internet of things (IoT): Applications, investments, and challenges for enterprises // *Business Horizons*. 2015. Vol. 58. No 4. P. 431–440. DOI: 10.1016/j.bushor.2015.03.008.
67. Zeinab K.A.M., Elmustafa S.A.A. Internet of things applications, challenges and related future technologies // *World Scientific News*. 2017. Vol. 67. No 2. P. 126–148.
68. Davies R. The internet of things opportunities and challenges / 2015. [Электронный ресурс]: <https://epthinktank.eu/2015/05/21/the-internet-of-things-opportunities-and-challenges/> (дата обращения 15.12.2020).
69. Porter M.E., Heppelmann J.E. How smart, connected products are transforming competition // *Harvard Business Review*. 2014. Vol. 92. P. 11–64.
70. Uckelmann D., Harrison M., Michahelles F. An architectural approach towards the future internet of things. Architecting the internet of things. Berlin, Heidelberg: Springer, 2011. DOI: 10.1007/978-3-642-19157-2_1.
71. Sun C. Application of RFID technology for logistics on internet of things // *AASRI Procedia*. 2012. Vol. 1. P. 106–111. DOI: 10.1016/j.aasri.2012.06.019.
72. Bandyopadhyay D., Sen J. Internet of things applications, challenges and related future technologies // *Wireless Personal Communications*. 2011. No 58. P. 49–69. DOI: 10.1007/s11277-011-0288-5.
73. The internet of everything: Smart things and their impact on business models / D.L. Langley [et al.] // *Journal of Business Research*. 2020. Vol. 122. P. 853–863. DOI: 10.1016/j.jbusres.2019.12.035.
74. Yuvaraj S., Sangeetha M. Smart supply chain management using internet of things (IoT) and low power wireless communication systems // *Proceedings of the IEEE 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. Chennai, India, 23–25 March 2016. P. 555–558. DOI: 10.1109/WISPNET.2016.7566196.
75. Tao F., Ying Z., Xu L.D., Zhang L. IoT-based intelligent perception and access of manufacturing resource toward cloud manufacturing // *IEEE Transactions on Industrial Informatics*. 2014. Vol. 10. No 2. P. 1547–1557. DOI: 10.1109/TII.2014.2306397.
76. Gnimpieba Z.D.R., Nait-Sidi-Moh A., Durand D., Fortin J. Using internet of things technologies for a collaborative supply chain: Application to tracking of pallets and containers // *Procedia Computer Science*. 2015. Vol. 56. P. 550–557. DOI: 10.1016/j.procs.2015.07.251.
77. A reference architecture for IoT-based logistic information systems in agri-food supply chains / C.N. Verdouw [et al.] // *Enterprise Information Systems*. 2018. Vol. 12. No 7. P. 755–779. DOI: 10.1080/17517575.2015.1072643.
78. Cost-benefit model for smart items in the supply chain / C. Decker [et al.] // *The internet of things (C. Floerkemeier, M. Langheinrich, E. Fleisch, F. Mattern, S.E. Sarma, eds.)*. Lecture Notes in Computer Science. 2008. Vol. 4952. P. 155–172. DOI: 10.1007/978-3-540-78731-0_10.
79. Xu L.D. Information architecture for supply chain quality management // *International Journal of Production Research*. 2011. Vol. 49. No 1. P. 183–198. DOI: 10.1080/00207543.2010.508944.
80. Chen R.-Y. Intelligent IoT-enabled system in green supply chain using integrated FCM method // *International Journal of Business Analytics*. 2015. Vol. 2. No 3. P. 47–66. DOI: 10.4018/IJBAN.2015070104.
81. Intelligent supply chain integration and management based on cloud of things / J. Yan [et al.] // *International Journal of Distributed Sensor Networks*. 2014. Vol. 10. No 3. Article No 624839. DOI: 10.1155/2014/624839.
82. Parry G.C., Brax S.A., Maull R.S., Ng I.C.L. Operationalising IoT for reverse supply: the development of use-visibility measures // *Supply Chain Management*. 2016. Vol. 21. No 2. P. 228–244. DOI: 10.1108/SCM-10-2015-0386.
83. Thoben K.D., Wiesner S., Wuest T. “Industrie 4.0” and smart manufacturing – A review of research issues and application examples // *International Journal of Automation Technology*. 2017. Vol. 11. No 1. P. 4–16.
84. Fleisch E. What is the internet of things? An economic perspective, business process and applications. White Paper WP-BIZAPP-053, 2010. [Электронный ресурс]: https://cocoa.ethz.ch/downloads/2014/06/none_AUTOIDLABS-WP-BIZAPP-53.pdf (дата обращения 15.12.2020).
85. Abdel-Basset M., Nabeeh N.A., El-Ghareeb H.A., Aboelfetouh A. Utilising neutrosophic theory to solve transition difficulties of IoT-based enterprises // *Enterprise Information Systems*. 2020. Vol. 14. No 9–10. P. 1304–1324. DOI: 10.1080/17517575.2019.1633690.
86. Gonzalez J. The impact of the internet of things on business and society. Foundation de la Innovation Bankinter, 2011.

87. Poirier C.C., Bauer M.J. E-supply chain: Using the internet to revolutionize your business. San Francisco: Berrett-Koehler, 2000.
88. Vervest P., Preiss K., van Heck E., Pau L.-F. The emergence of smart business networks // *Journal of Information Technology*. 2004. Vol. 19. No 4. P. 228–233. DOI: 10.1057/palgrave.jit.2000024.
89. Wirtz B.W., Weyerer J.C., Schichtel F.T. An integrative public IoT framework for smart government // *Government Information Quarterly*. 2019. Vol. 36. No 2. P. 333–345. DOI: 10.1016/j.giq.2018.07.001.
90. Challenges for adopting and implementing IoT in smart cities: An integrated MICMAC-ISM approach / M. Janssen [et al.] // *Internet Research*. 2019. Vol. 29. No 6. P. 1589–1616. DOI: 10.1108/INTR-06-2018-0252.
91. Abdel-Basset M., Manogaran G., Mohamed M. Internet of things (IoT) and its impact on supply chain: A framework for building smart, secure and efficient systems // *Future Generation Computer Systems*. 2018. Vol. 86. P. 614–628. DOI: 10.1016/j.future.2018.04.051.
92. Hu G., Lin Y., Shi Y., Guo L. Understanding business ecosystem using a 6C framework in Internet-of-Things-based sectors / K. Rong [et al.] // *International Journal of Production Economics*. 2015. Vol. 159. P. 41–55. DOI: 10.1016/j.ijpe.2014.09.003.
93. Cegielski C.G., Jones Farmer L.A., Wu Y., Hazen B.T. Adoption of cloud computing technologies in supply chains // *International Journal of Logistics Management*. 2012. Vol. 32. No 2. P. 184–211. DOI: 10.1108/09574091211265350.
94. Suguna S.K., Kumar S.N. Application of cloud computing and internet of things to improve supply chain processes // *Edge computing: From hype to reality*. Springer, 2019. P. 145–170.
95. Arlbjorn J.S., de Haas H.D., Munksgaard K.B. Exploring supply chain innovation // *Logistics Research*. 2011. Vol. 3. No 1. P. 3–18. DOI: 10.1007/s12159-010-0044-3.
96. Lee I. The Internet of Things (IoT) for supply chain innovation: a conceptual framework and analysis of Fortune 200 companies // *Asia Pacific Journal of Innovation and Entrepreneurship*. 2015. Vol. 9. No 1. P. 81–103.
97. Nozari H., Najafi E., Fallah M., Lotfi F.H. Quantitative analysis of key performance indicators of green supply chain in FMCG industries using non-linear fuzzy method // *Mathematics*. 2019. Vol. 7. No 11. Article No 1020. DOI: 10.3390/math7111020.
98. Kaur J., Sidhu R., Awasthi A., Srivastava S.K. A Pareto investigation on critical barriers in green supply chain management // *International Journal of Management Science and Engineering Management*. 2019. Vol. 14. No 2. P. 113–123. DOI: 10.1080/17509653.2018.1504237.
99. The role of big data analytics in industrial Internet of Things / M.H. Rehman [et al.] // *Future Generation Computer Systems*. 2019. No 99. P. 247–259. DOI: 10.1016/j.future.2019.04.020.
100. Yu M., Nagurny A. Competitive food supply chain networks with application to fresh produce // *European Journal of Operational Research*. 2013. Vol. 224. No 2. P. 273–282. DOI: 10.1016/j.ejor.2012.07.033.
101. Ji G., Hu L., Tan K.H. A study on decision-making of food supply chain based on big data // *Journal of Systems Science and Systems Engineering*. 2017. No 26. P. 183–198. DOI: 10.1007/s11518-016-5320-6.
102. Osman A.M.S. A novel big data analytics framework for smart cities // *Future Generation Computer Systems*. 2019. No 91. P. 620–633.

Об авторах

Хамед Нозари

Ph.D.;

Lecturer, Department of Industrial Engineering, Islamic Azad University, Central Tehran Branch, Hamila Blvd., Poonak Sqr., Tehran 1469669191, Iran;

E-mail: Ham.nozari.eng@iauctb.ac.ir

ORCID: 0000-0002-6500-6708

Мохаммад Фаллах

Ph.D.;

Associated Professor, Department of Industrial Engineering, Islamic Azad University, Central Tehran Branch, Hamila Blvd., Poonak Sqr., Tehran 1469669191, Iran;

E-mail: Mohammad.fallah43@yahoo.com

ORCID: 0000-0001-5541-4284

Хамед Каземипур

Ph.D.;

Assistant Professor, Department of Industrial Engineering, Islamic Azad University, Central Tehran Branch, Hamila Blvd., Poonak Sqr., Tehran, 1469669191 Iran;

E-mail: Hkazemipoor@yahoo.com

ORCID: 0000-0002-1848-8927

Сейед Эсмаил Найяфи

Ph.D.;

Assistant Professor, Department of Industrial Engineering, Islamic Azad University, Science and Research Branch, Daneshgah Blvd., Simon Bulivar Blvd., Tehran 1477893855, Iran;

E-mail: E.najafi@srbiau.ac.ir

ORCID: 0000-0002-3125-4439