

БИЗНЕС- ИНФОРМАТИКА

НАУЧНЫЙ ЖУРНАЛ НИУ ВШЭ

СОДЕРЖАНИЕ

Ф.А. Белоусов, Н.К. Хачатрян, И.В. Неволин

Снижение размерности в задаче
оптимального управления парком
грузовых вагонов с использованием
беспилотных локомотивов.....7

Е.С. Морозевич, В.С. Коротких, Е.А. Кузнецова

Разработка модели формирования
индивидуальных образовательных
траекторий с использованием методов
машинного обучения21

Б.Б. Славин

Технологии коллективного интеллекта
в управлении бизнес-процессами
организации36

Р.Р. Сухов, М.Б. Амзараков, Е.А. Исаев

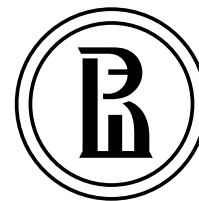
Новые метрики энергоэффективности
ИТ-отрасли.....49

С.М. Авдошин, Е.Ю. Песоцкая

Доверенный искусственный интеллект
как способ цифровой защиты.....62

С.А. Смоляк

О назначении сроков службы технических
систем в условиях инфляции74



Издатель:
Национальный
исследовательский университет
«Высшая школа экономики»

Подписной индекс
Объединенного каталога
«Пресса России» – E79128

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

И.о. главного редактора
Е.П. Зараменских

Компьютерная верстка
О.А. Богданович

Администратор веб-сайта
И.И. Хрусталева

Адрес редакции:
119049, г. Москва,
ул. Шаболовка, д. 26-28
Тел./факс: +7 (495) 772-9590 *28509
<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 – 100 экз.,
англоязычная версия – 100 экз.,
онлайн-версии на русском и английском –
свободный доступ

Отпечатано в типографии НИУ ВШЭ
г. Москва, Измайловское шоссе, д. 44, стр. 2

© Национальный
исследовательский университет
«Высшая школа экономики»

«Бизнес-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 дис-

сертаций 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: 2.3.4 — Управление в организационных системах (технические науки); 2.3.5 — Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (технические науки); 1.2.2 — 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 Scopus, 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).

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

И.О. 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 Андраник 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

Баракшин 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

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 Йоахим

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 Гедеванович

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

Штраус Кристина

Университет Вены, Вена, Австрия

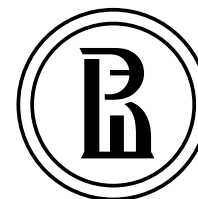
ISSN 1998-0663 (print), ISSN 2587-8166 (online)

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

Vol. 16 No. 2 – 2022

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

CONTENTS

F.A. Belousov, N.K. Khachatryan, I.V. Nevolin

Reduction of dimension in the problem
of optimal management of a freight cars fleet
using unmanned locomotives7

E.S. Morozevich, V.S. Korotkikh, Y.A. Kuznetsova

The development of a model
for a personalized learning path
using machine learning methods21

B.B. Slavin

Technologies of collective intelligence
in the management of business processes
of an organization36

R.R. Sukhov, M.B. Amzarakov, E.A. Isaev

New energy efficiency metrics
for the IT industry49

S.M. Avdoshin, E.Yu. Pesotskaya

Trusted artificial intelligence:
Strengthening digital protection62

S.A. Smolyak

On assigning service life for technical
systems under inflation74

Acting Editor-in-Chief

E. Zaramenskikh

Computer Making-up:

O. Bogdanovich

Website Administration:

I. Khrustaleva

Address:

26-28, build. 4, Shablovka Street
Moscow 119049, Russia

Tel./fax: +7 (495) 772-9590 *28509

<http://bijournal.hse.ru>

E-mail: bijournal@hse.ru

Circulation:

English version – 100 copies,

Russian version – 100 copies,

online versions in English and Russian –
open access

Printed in HSE Printing House

44, build. 2, Izmaylovskoye Shosse,
Moscow, Russia

© National Research University
Higher School of Economics

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 Scopus, 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).

EDITORIAL BOARD

ACTING 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. Gavrilova

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

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

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

Снижение размерности в задаче оптимального управления парком грузовых вагонов с использованием беспилотных локомотивов

Ф.А. Белоусов 

E-mail: sky_tt@list.ru

Н.К. Хачатрян 

E-mail: nerses-khachatryan@yandex.ru

И.В. Неволин 

E-mail: i.nevolin@cemi.rssi.ru

Центральный экономико-математический институт Российской академии наук
Адрес: Россия, 117418, г. Москва, Нахимовский проспект, д. 47

Аннотация

В работе рассматривается задача оптимального управления парком грузовых вагонов транспортным железнодорожным оператором. Решением данной задачи является оптимальный план, представляющий собой расписание движения грузовых и порожних железнодорожных вагонов, следуя которому транспортный оператор получит максимальную прибыль за расчетный период времени. Данная проблема сводится к задаче линейного программирования большой размерности. В отличие от работ других авторов по данной тематике, в которых в основном речь идет о методах численного решения соответствующих задач линейного программирования, в данной статье акцент делается на алгоритме, позволяющем снижать их размерность. Этого удастся достичь благодаря исключению из расчета тех маршрутов, которые заведомо не смогут быть задействованы в решении, либо вероятность участия которых в итоговом решении оценивается как крайне низкая. Эффективность предложенного модифицированного алгоритма была подтверждена как на модельном примере (несколько станций, короткий горизонт планирования), так и на реальном примере (более 1000 станций, длинный горизонт планирования). В первом случае наблюдалось снижение размерности задачи на 44%, тогда как во втором – в 30 раз.

Ключевые слова: железнодорожные грузоперевозки, оптимальный план, оптимальное управление парком вагонов, линейное программирование, теория расписаний, исследование операций, беспилотные локомотивы

Цитирование: Белоусов Ф.А., Хачатрян Н.К., Неволин И.В. Снижение размерности в задаче оптимального управления парком грузовых вагонов с использованием беспилотных локомотивов // Бизнес-информатика. 2022. Т. 16. № 2. С. 7–20. DOI: 10.17323/2587-814X.2022.2.7.20

Введение

Одним из наиболее востребованных для грузоперевозок видов транспорта в Российской Федерации является железнодорожный. Публикации, посвященные железнодорожной логистике, по типу исследуемых задач можно разбить на следующие основные группы:

- ♦ задачи проектирования инфраструктуры железнодорожной сети;
- ♦ задачи железнодорожного планирования;
- ♦ задачи управления парком локомотивов и вагонов.

В первой группе можно выделить работы [1–4]. Вторая группа, в частности, представлена задачами формирования расписания движения грузовых поездов, а также задачами формирования грузовых потоков [5, 6]. Один из подходов к формированию грузопотока представлен в работах Хачатряна Н.К. и Бекларяна Л.А. [7–16]. В них представлены макроскопические динамические модели, в которых процесс организации железнодорожных грузоперевозок представляет собой формирование грузопотока на основе взаимодействия соседних станций. Эти модели позволяют прогнозировать динамику загруженности станций и потоков, возникающих на железнодорожной сети, при заданной процедуре организации грузопотока. Рассматривается несколько конфигураций участка железнодорожной сети. Первая – представляет собой протяженный участок железнодорожной сети и характеризуется бесконечным числом станций, как в одну, так и в другую сторону и предполагает отсутствие узловых станций. Такая конфигурация транспортной сети подходит для описания транснациональных перевозок (например, перевозки по транссибирской железнодорожной магистрали протяженностью более 9000 км). Вторая конфигурация задает движение грузопотока по замкнутой цепочке станций. Третья – характеризуется конечным числом станций и определяет движение грузопотока между двумя узловыми станциями.

Представленная работа посвящена задаче оптимального управления парком железнодорожных грузовых вагонов. Перед железнодорожными транспортными операторами стоит задача оптимального с точки зрения максимизации прибыли управления парком грузовых вагонов. Такое управление осуществляется, с одной стороны, на основе дислокации вагонов, с другой – на основе заявок на транспортировку грузов. Заявки подаются клиентами, в каждой заявке указывается пункт отправле-

ния, пункт назначения, объем перевозимого груза, выраженный в вагонах, и ставка, которую заплатит заказчик за каждый вагон перевезенного груза. Кроме ставки, которую заказчик платит транспортному оператору за предоставление вагонов, он также оплачивает услуги РЖД за транспортировку груженых вагонов. Затраты на транспортировку порожних вагонов покрываются транспортным оператором. Из всего списка заявок транспортный оператор выбирает те, которые ему наиболее выгодно исполнять. Любая выбранная заявка может быть выполнена как полностью, так и частично. В соответствии с дислокацией вагонов и имеющимся списком заявок создание плана управления вагонами подразумевает составление расписания движения груженых и порожних вагонов с учетом известных временных нормативов. Таким образом, ставится задача поиска оптимального плана управления парком грузовых вагонов на определенный период времени (как правило, план составляется на месяц). В [17] она представлена в виде многопродуктовой потоковой модели, заданной на пространственно-временном графе, и предложен алгоритм, позволяющий свести ее к задаче линейного программирования большой размерности.

Задачи с близкой постановкой ранее рассматривались в работах [18, 19]. В указанных источниках авторы делают упор на способах численного решения получаемой задачи линейного программирования, в частности в них идет речь о применении метода генерации колонок [20, 21] и модификации этого метода [22, 23], основой которой является декомпозиция Данцига-Вульфа [24]. Похожая задача рассматривалась в более ранней статье [25]. Представленная в работе модель была разработана по запросу одного из крупнейших транспортных операторов Латинской Америки. Ее особенностью является то, что план грузовых перевозок составляется с учетом заранее известного расписания движений локомотивов, тогда как в данной статье и в работах [18, 19] время движения вагонов по каждому из маршрутов определяется исключительно исходя из нормативов РЖД (т.е. расписание движения локомотивов в нашем случае неизвестно и для нас оно неважно, это внутреннее дело РЖД).

В отличие от указанных работ авторы этой статьи не акцентируют своего внимания на методах численного решения задач линейного программирования, а предлагают способы конструирования пространственно-временных графов, служащих основой для задания как целевой функции, так и

ограничений задачи линейного программирования, имеющей меньшую размерность. При этом сам вид целевой функции и ограничений не меняется. Таким образом, в данной работе предложена модификация алгоритма, описанного в [17], позволяющая значительно снизить размерность модели. Прежде чем перейти к ее описанию, приведем постановку самой задачи линейного программирования. Для этого введем следующие обозначения:

N – количество станций, участвующих в планировании;

T – горизонт планирования, который измеряется в днях. Для упрощения за горизонт планирования принят один месяц (т.е. $T = 30$ или 31);

t – дискретный параметр, характеризующий время. Он измеряется в днях и принимает значения $t = 1, 2, \dots, T$;

$C = \{C_{ij}\}_{i,j=1}^N$ – $(N \times N)$ -матрица, элементы которой характеризуют тарифы, установленные РЖД за порожний перегон одного вагона от станции i до станции j ;

$\Theta 1 = \{\Theta 1_{ij}\}_{i,j=1}^N$ – $(N \times N)$ -матрица, элементы которой характеризуют время (в днях) движения груженого вагона от станции i до станции j в соответствии с нормативами РЖД (время округляется до большего целого числа);

$\Theta 2 = \{\Theta 2_{ij}\}_{i,j=1}^N$ – $(N \times N)$ -матрица, элементы которой характеризуют время (в днях) движения порожнего вагона от станции i до станции j в соответствии с нормативами РЖД (время округляется до большего целого числа);

$P = \{P_{ij}\}_{i,j=1}^N$ – $(N \times N)$ -матрица, элементы которой характеризуют ставку, указанную заказчиком в заявке на транспортировку одного вагона груза от станции i до станции j ;

$\bar{Q} = \{\bar{Q}_{ij}\}_{i,j=1}^N$ – $(N \times N)$ -матрица, элементы которой характеризуют количество грузовых вагонов, указанное в соответствующей заявке на транспортировку груза от станции i до станции j . Все элементы матрицы принимают неотрицательные целочисленные значения;

$\bar{S}^0(t) = \{\bar{S}_i^0(t)\}_{i=1}^N$ – вектор размерности N , характеризующий начальное расположение вагонов в день t ; i -й элемент данного вектора равен количеству вагонов, отправка которых была осуществлена в предыдущем месяце, прибывших на станцию i в момент времени $t \in \{1, \dots, T\}$. Все значения данного вектора принимают неотрицательные целочисленные значения.

Тогда указанная выше задача линейного программирования принимает вид

$$PC^T \cdot K \rightarrow \max_{K \geq 0}, \quad (1)$$

при ограничениях

$$(A_{out} - A_{in}) \cdot K = S_0, \quad (2)$$

$$A_Q \cdot K \leq Q, \quad (3)$$

где K – вектор, первая часть которого отвечает за груженные маршруты, вторая за порожние, по существу данный вектор является планом перевозок;

PC – вектор, первая часть которого отвечает за ставки на грузовые перевозки, вторая часть соответствует издержкам за порожние перегоны, произведение $PC^T \cdot K$ дает прибыль, которая получится за горизонт планирования;

A_{out} – матрица, учитывает исходящие из каждой станции маршруты;

A_{in} – матрица, учитывает входящие в каждую станцию маршруты;

S_0 – вектор начального распределения вагонов по станциям и по времени;

A_Q – матрица, произведение которой на вектор K даёт результат в виде количества вагонов, исполненных по каждой из заявок в соответствии с решением K ;

Q – вектор объема заказов (в вагонах), указанных в заявках.

Ограничение (2) является балансовым ограничением, то есть его выполнение гарантирует, что в каждый период количество въехавших в станцию вагонов будет совпадать с количеством уехавших вагонов. Выполнение ограничения (3) гарантирует, что объем исполненных грузовых маршрутов не будет превышать объемов, указанных в заявках.

Алгоритм, представленный в работе [17] в итоге давал размерность задачи (1)–(3) равную $2TN^2$, т.е. количество элементов в векторах K и PC составляет $2TN^2$. Размерность матриц A_{out} и A_{in} получается равной $TN \times 2TN^2$, размерность матрицы A_Q составляет $N^2 \times 2TN^2$, размерность вектора S_0 равна TN , размерность Q равна N^2 . Модификация алгоритма, приведенная в данной работе, заключается в снижении размерности указанных матриц и векторов.

1. Алгоритмы формирования матриц и векторов для задачи (1)–(3)

Наиболее заметного снижения размерности задачи (1)–(3) можно добиться путем учета только тех груженных маршрутов в векторе K , использование которых приведет к частичному или полному исполнению заявок. Другими словами, снижение размерности задачи можно организовать за счёт удаления из вектора K тех грузовых маршрутов, которые в любом случае задействованы не будут. Кроме вектора K соответствующие преобразования должны быть произведены над всеми остальными векторами и матрицами задачи (1)–(3) так, чтобы все они были согласованы.

В данном разделе описан алгоритм формирования всех матриц и векторов сокращенной размерности для задачи (1)–(3). Кроме удаления из рассмотрения ненужных грузовых маршрутов в алгоритме также заложена возможность исключения из расчета порожних маршрутов, отобранных по какому-либо признаку.

Векторы PC , Q , S_0 и динамические списки

Введем новые переменные $routes_from_station_cargo$, $routes_to_station_cargo$ и $routes_from_station_empty$, $routes_to_station_empty$, которые представляют собой динамические списки с элементами, принимающими целочисленные значения. Элементы этих переменных отвечают за номера тех маршрутов, которые будут учтены в расчете. Список $routes_from_station_cargo$ содержит номера исходящих станций для каждого из учитываемых грузовых маршрутов, $routes_to_station_cargo$ содержит соответствующие номера входящих станций для этих же грузовых маршрутов. Аналогично, $routes_from_station_empty$ содержит номера исходящих станций для порожних маршрутов, $routes_to_station_empty$ содержит номера входящих станций для соответствующих порожних маршрутов. Если учитывать всевозможные маршруты, как это сделано в работе [17], то количество элементов, содержащееся в новых переменных, будет равно N^2 . Однако введены эти переменные для того, чтобы можно было учитывать меньшее число маршрутов, снижая тем самым размерность задачи.

Заполним переменные $routes_from_station_cargo$ и $routes_to_station_cargo$. Будем учитывать только те маршруты, которые есть в заявках, поэтому, если существует заявка из станции i в станцию j , то есть $P_{ij} > 0$, то к переменной $routes_from_station_cargo$

справа добавляем элемент i , к переменной $routes_to_station_cargo$ также справа добавляем элемент j . Одновременно с каждым добавлением элементов в переменные $routes_from_station_cargo$ и $routes_to_station_cargo$ будем составлять вектор p последовательным добавлением снизу элементов P_{ij} , а также будем составлять вектор Q , последовательным добавлением снизу элементов $\bar{Q}_{ij}$, характеризующих объемы соответствующих заявок. Таким образом, на каждой итерации размерность векторов p и Q , а также переменных $routes_from_station_cargo$ и $routes_to_station_cargo$ увеличивается на единицу.

Таким образом, полученный вектор Q учитывает только грузовые маршруты, а не всевозможные маршруты, как это было в работе [17]. За счет этого размерность вектора Q снижается с N^2 до N_{cargo} .

Будем считать, что для каждой пары станций i и j из станции i в станцию j может быть не более одной заявки. Если для какой-то пары станций i и j из i в j существует две заявки, то в этом случае создается дубль станции i , назовем его $\hat{i}$, и в переменные $routes_from_station_cargo$ и $routes_to_station_cargo$ добавим не только i и j , но и $\hat{i}$ и j соответственно. В случае двух заявок вектор p заполняется соответствующими ставками в той же последовательности, что и переменные $routes_from_station_cargo$ и $routes_to_station_cargo$. Сообщение между станцией i и её дублем $\hat{i}$ осуществляется мгновенно и бесплатно. При этом обратного сообщения из станции $\hat{i}$ в станцию i нет. Все входящие маршруты направлены в станцию i , в станцию $\hat{i}$ можно попасть только через станцию i . Сделано это для того, чтобы в решениях не появлялись циклические потоки из станции i в $\hat{i}$ и обратно. Случай, когда из станции i в станцию j может быть больше двух заявок в рамках данной статьи рассматриваться не будет — это отдельная тема, раскрытие которой оставим на последующие работы в этом направлении.

Аналогичным путем переменные $routes_from_station_empty$ и $routes_to_station_empty$ заполняются номерами станций только тех маршрутов, которые были выбраны по каким-либо критериям. Одновременно с заполнением переменных $routes_from_station_empty$ и $routes_to_station_empty$, по аналогии с вектором p , заполняется новый вектор $\tilde{C}$ путем добавления снизу элементов C_{ij} (стоимость порожнего перегона из станции i в станцию j). Порядок добавления элементов в вектор $\tilde{C}$ соответствует порядку добавления элементов в переменные $routes_from_station_empty$ и $routes_to_station_empty$.

station_empty. Таким образом, на каждой итерации размерность вектора $\tilde{C}$, а также переменных *routes_from_station_empty* и *routes_to_station_empty* увеличивается на единицу.

Один из способов снизить размерность за счет порожних маршрутов состоит в удалении из рассмотрения тех порожних маршрутов, станции прибытия в которых не являются станциями отправления ни для одной из заявок на груженные рейсы. Идея в том, что нет надобности приезжать в такие станции порожними рейсами, поскольку уехать оттуда вагоны могут только другими порожними рейсами, что вряд ли окажется оптимальным. Исключением являются маршруты из этих станций в себя, такие маршруты в численном расчете лучше оставить, поскольку в эти станции либо приезжают вагоны из предыдущего периода, либо они являются конечным пунктом для некоторых груженных маршрутов. Тут необходимо сделать оговорку о том, что этот способ удаления порожних маршрутов из рассмотрения оправдан при условии, что временные и финансовые затраты порожних перегонов из произвольной станции А в произвольную станцию В всегда не больше, чем при осуществлении двух последовательных порожних перегонов из А в некоторую станцию С и из С в В. На практике это условие выполняется, поэтому исключение из рассмотрения подобных двойных порожних маршрутов не приводит к ухудшению целевых показателей полученных решений.

Через N_{cargo} обозначим размерность вектора p , которая совпадает с количеством элементов в переменных *routes_from_station_cargo* и *routes_to_station_cargo*, через N_{empty} обозначим размерность вектора $\tilde{C}$, которая также совпадает с количеством элементов в переменных *routes_from_station_empty* и *routes_to_station_empty*. Другими словами, N_{cargo} характеризует количество всех грузовых маршрутов, соответствующих списку заявок, N_{empty} характеризует количество всевозможных порожних маршрутов, которые берутся в расчет при поиске оптимального плана.

Построим вектор PC меньшей размерности по сравнению с аналогичным вектором в работе [17]. Для этого осуществим $T - 1$ раз последовательную конкатенацию вектора p так, чтобы получился вектор размерности $T \cdot N_{cargo}$. Далее к полученному вектору также последовательной конкатенацией $T - 1$ раз присоединим вектор $\tilde{C}$. Вектору PC присвоим значение полученного вектора, его размерность составит $T \cdot (N_{cargo} + N_{empty})$. Новая размерность

задачи (1)–(3) будет также равна $T \cdot (N_{cargo} + N_{empty})$. Очевидно, элементы вектора K соответствуют тем же маршрутам, которые соответствуют и ставкам, и тарифам вектора PC .

Алгоритм создания вектора S_0 , описанный в работе [17], останется неизменным. А именно, система векторов $\bar{S}^0(t)$, $t \in \{1, \dots, T\}$ преобразуется в вектор S_0 путем последовательной конкатенации векторов, соответствующих каждому моменту времени. При такой конкатенации первые N элементов нового вектора соответствуют вектору $\bar{S}^0(1)$, следующие N элементов соответствуют вектору $\bar{S}^0(2)$, и т.д. Таким образом, размерность вектора S_0 составит TN .

Матрицы A_{in} и A_{out}

Каждая из матриц A_{in} и A_{out} состоит из двух частей; первая часть отвечает за грузовые маршруты, вторая – за порожние. Данные матрицы являются разреженными матрицами, любой ненулевой элемент в них принимает единичное значение. Сконструируем матрицу A_{out} , которая отвечает за учет исходящих из каждой станции маршрутов. На нулевой итерации матрица A_{out} представляет собой нулевую матрицу размером $T \cdot N \times T \cdot (N_{cargo} + N_{empty})$.

Обозначим через *Index_cargo_out*[1] динамический список, в который записываются такие индексы $k \in \{1, \dots, N_{cargo}\}$ переменной *routes_from_station_cargo*[k] для которых *routes_from_station_cargo*[k] = 1. Другими словами, переменная *Index_cargo_out*[1] содержит номера тех грузовых маршрутов среди первых N_{cargo} элементов вектора K , точкой отправления для которых является станция 1. Для произвольной станции $i \in \{1, \dots, N\}$ интерпретация переменной *Index_cargo_out*[i] аналогична.

Для учета исходящих маршрутов из станции 1 в первый период времени необходимо для всех таких $k \in \text{Index_cargo_out}[1]$ элементам $A_{out}[1, k]$ присваивать значение 1. Для учета исходящих маршрутов из станции 1 во второй период времени необходимо уже элементам $A_{out}[1 + N, k + N_{cargo}]$, $k \in \text{Index_cargo_out}[1]$ присваивать единичные значения. Первой компоненте координат матрицы A_{out} прибавляется N , поскольку в векторе S_0 период равен N , другими словами первые N элементов в этом векторе отвечают за N станций в первый период времени, следующие N элементов отвечают за эти же N станций во второй период времени и т.д. Второй компоненте матрицы A_{out} прибавляется N_{cargo} элементов, поскольку период в первой части вектора PC , отвеча-

ющей за ставки груженых рейсов, равен N_{cargo} . Иначе говоря, первые N_{cargo} элементов в первой части вектора PC , отвечающих за маршруты, стартующие в первый период времени, следующие N_{cargo} элементов отвечают за эти же маршруты, но стартующие уже во второй период времени, и так далее. Продолжая эту логику далее получаем, что для учета исходящих маршрутов из станции 1 в произвольный период времени $t \in \{1, \dots, T\}$ присвоить единице необходимо все элементы матрицы A_{out} с координатами $[1 + (t - 1) \cdot N, k + (t - 1) \cdot N_{cargo}]$, $k \in Index_cargo_out[1]$. Таким образом, для получения матрицы A_{out} необходимо для каждой станции $i \in \{1, 2, \dots, N\}$ создать динамический список $Index_cargo_out[i]$ с такими номерами $k \in \{1, \dots, N_{cargo}\}$, для которых $routes_from_station_cargo[k] = i$. Далее, для всех $i \in \{1, \dots, N\}$ для которых $Index_cargo_out[i] \neq 0$, единице присваиваются элементы матрицы A_{out} с координатами $[i + (t - 1) \cdot N, k + (t - 1) \cdot N_{cargo}]$, $k \in Index_cargo_out[i]$, $t \in \{1, \dots, T\}$. Сформирована первая часть матрицы A_{out} , отвечающая за грузовые маршруты. Аналогично формируется вторая часть этой матрицы, отвечающая за порожние маршруты. Для этого для каждой станции $i \in \{1, \dots, N\}$ формируются другие списки $Index_empty_out[i]$ с такими номерами $k \in \{1, \dots, N_{empty}\}$, для которых $routes_from_station_empty[k] = i$. Другими словами, переменная $Index_cargo_out[i]$ содержит номера тех порожних маршрутов среди N_{empty} элементов вектора K , идущих после $T \cdot N_{cargo}$ элементов, точкой отправления для которых является станция i .

Далее, для всех $i \in \{1, \dots, N\}$, для которых $Index_empty_out[i] \neq 0$, единице присваиваются элементы матрицы A_{out} с координатами $[i + (t - 1) \cdot N, k + (t - 1) \cdot N_{empty} + T \cdot N_{cargo}]$, $k \in Index_empty_out[i]$, $t \in \{1, \dots, T\}$. Так как первые $T \cdot N_{cargo}$ элементов в вектора K отвечают за грузовые маршруты, а остальные ответственны за порожние, то в случае рассмотрения порожних маршрутов во второй компоненте координат матрицы A_{out} присутствует дополнительно слагаемое $T \cdot N_{cargo}$. После осуществления всех описанных операций формирование матрицы A_{out} оказывается завершенным.

На следующем этапе формируется матрица A_{in} . Эта матрица отвечает за учет входящих в каждую станцию маршрутов. На нулевой итерации матрица A_{in} представляет собой нулевую матрицу размером

$T \cdot N \times T \cdot (N_{cargo} + N_{empty})$. Для формирования этой матрицы нам дополнительно понадобится информация о времени движения по каждому из маршрутов, то есть значения матриц $\Theta 1$ и $\Theta 2$. Обозначим через $Index_cargo_in[1]$ динамический список тех индексов $k \in \{1, \dots, N_{cargo}\}$ переменной $routes_to_station_cargo[k]$, для которых $routes_to_station_cargo[k] = 1$. Другими словами, переменная $Index_cargo_in[1]$ содержит номера тех грузовых маршрутов среди первых N_{cargo} элементов вектора K , пунктом назначения для которых является станция 1. Для произвольной станции $i \in \{1, \dots, N\}$ интерпретация переменной $Index_cargo_in[i]$ аналогична. Тогда для учета входящих маршрутов в станцию 1, отправления по которым осуществляется в первый период времени, необходимо для всех $k \in Index_cargo_in[1]$ элементам матрицы A_{in} с координатами $[1 + \Theta 1[routes_from_station_cargo[k], 1] \cdot N, k]$ присвоить значение 1. Аналогично для учета входящих маршрутов в станцию 1, отправления по которым осуществляется в период времени $t \in \{1, \dots, T\}$ для всех $k \in Index_cargo_in[1]$ единице присваиваются элементы матрицы A_{in} с координатами $[1 + (\Theta 1[routes_from_station_cargo[k], 1] + t - 1) \cdot N, k + (t - 1) \cdot N_{cargo}]$.

Для произвольной станции $i \in \{1, \dots, N\}$ составляются динамические переменные $Index_cargo_in[i]$ из тех индексов $k \in \{1, \dots, N_{cargo}\}$ переменной $routes_to_station_cargo[k]$, для которых $routes_to_station_cargo[k] = i$. Для всех $i \in \{1, \dots, N\}$, для которых $Index_cargo_in[i] \neq 0$, единице присваиваются элементы матрицы A_{in} с координатами $[i + (\Theta 2[routes_from_station_cargo[k], i] + t - 1) \cdot N, k + (t - 1) \cdot N_{cargo}]$, $k \in Index_cargo_in[i]$, $t \in \{1, \dots, T\}$.

Первая часть матрицы A_{in} , относящаяся к груженым маршрутам, построена. Осталось построить вторую часть этой матрицы, относящуюся к порожним маршрутам. Обозначим через $Index_empty_in[i]$ динамический список тех индексов $k \in \{1, \dots, N_{empty}\}$ переменной $routes_to_station_empty[k]$, для которых $routes_to_station_empty[k] = i$. Для всех станций $i \in \{1, \dots, N\}$, для которых $Index_empty_in[i] \neq 0$, единице присваиваются элементы матрицы A_{in} с координатами $[i + (\Theta 2[routes_from_station_cargo[k], i] + t - 1) \cdot N, k + (t - 1) \cdot N_{empty} + T \cdot N_{cargo}]$, $k \in Index_empty_out[i]$, $t \in \{1, \dots, T\}$. Формирование матрицы A_{in} завершено.

Матрица A_0

Матрица A_0 нужна для расчета объема исполненных заявок, поэтому при вычислении этого показателя учитываются только груженные маршруты. Это означает, что в матрице A_0 , имеющей размерность $N_{cargo} \times T \cdot (N_{cargo} + N_{empty})$, последние $T \cdot N_{empty}$ столбцов состоят исключительно из нулевых элементов, ненулевые элементы находятся только в первых $T \cdot N_{cargo}$ столбцах. На нулевой итерации в качестве матрицы A_0 возьмем нулевую матрицу. Поскольку первые $T \cdot N_{cargo}$ элементов вектора K упорядочены с периодом N_{cargo} , т.е. первые N_{cargo} элементов отвечают за груженные маршруты, исходящие в первый период времени, следующие N_{cargo} элементов отвечают за эти же маршруты, исходящие во второй период времени и т.д., то в первой строке матрицы A_0 записывается T единиц, первая из которых ставится на первую позицию, следующая на позицию $N_{cargo} + 1$, следующая на позицию $2N_{cargo} + 1$ и т.д. Другими словами, в первой строке матрицы A_0 единице присваиваются T элементов начиная с первого элемента и далее с периодом N_{cargo} элементов. В следующей строке матрицы A_0 единице также присваивается T элементов с периодом N_{cargo} , но начиная уже со второго элемента второй строки. В третьей строке матрицы A_0 алгоритм повторяется, но единице присваиваются элементы, начиная с третьего элемента третьей строки. Так продолжается до последней строки N_{cargo} . В результате, если рассмотреть первые N_{cargo} строк и первые N_{cargo} столбцов, получится единичная матрица, если рассмотреть следующие N_{cargo} столбцов, то также получится единичная матрица, и т.д. Если рассмотреть первые $T \cdot N_{cargo}$ столбцов матрицы A_0 , то увидим T последовательно составленных единичных матриц размерности $N_{cargo} \times N_{cargo}$, остальные столбцы матрицы являются нулевыми.

В данном разделе приведены алгоритмы формирования всех составляющих задачи (1)–(3) – целевой функции и ограничений. Показано, что

размерность и вектора переменных целевой функции, и матриц ограничений заметно снижается. Продemonстрируем это как на модельном примере (несколько станций, короткий горизонт планирования), приведенном в [17], так и на реальном примере (более 1000 станций, длинный горизонт планирования).

2. Снижение размерности задачи на модельном и реальном примерах

Приведем постановку модельного примера из [17]. Число станций равно 4 ($N = 4$), горизонт планирования составляет 3 дня ($T = 3$). Список поступивших заявок состоит из пяти позиций, которые приведены в таблице 1.

На основе списка заявок необходимо составить две матрицы – матрицу ставок P , элементы которой записаны в условных единицах, и матрицу объемов заявок $\bar{Q}$:

$$P = \begin{pmatrix} 0 & 0 & 2,9 & 0 \\ 1,1 & 0 & 2,3 & 0 \\ 0 & 1,9 & 0 & 2,1 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \bar{Q} = \begin{pmatrix} 0 & 0 & 3 & 0 \\ 5 & 0 & 4 & 0 \\ 0 & 7 & 0 & 6 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Далее приведем время движения как груженных, так и порожних маршрутов в виде значений матриц $\Theta 1$ и $\Theta 2$:

$$\Theta 1 = \begin{pmatrix} 0 & 2 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 2 \\ 1 & 2 & 1 & 0 \end{pmatrix}; \Theta 2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 \end{pmatrix}.$$

Напомним, что диагональные элементы матрицы $\Theta 2$ берутся равными единице. Связано это с тем, что если вагоны необходимо оставить на станции до следующего дня, то это равносильно тому, что они как бы отправляются из данной станции на саму себя в рейс длительностью один день.

Таблица 1.

Список заявок на перевозку груза в модельном примере

№	Пункт отправления	Пункт назначения	Объем заявки (в вагонах)	Ставка (в условных единицах)
1	1	3	3	2,9
2	2	1	5	1,1
3	2	3	4	2,3
4	3	2	7	1,9
5	3	4	6	2,1

Значения тарифов РЖД на порожние перегоны, также как и ставки, выраженные в условных единицах, характеризуются значениями элементов матрицы C :

$$C = \begin{pmatrix} 0 & 1,9 & 1,3 & 1,9 \\ 1,2 & 0 & 1,8 & 0,9 \\ 1,1 & 1,2 & 0 & 1,6 \\ 1,3 & 1,5 & 1,2 & 0 \end{pmatrix}.$$

В рамках данной задачи предполагается, что вагоны могут оставаться на станциях до следующего дня бесплатно, поэтому диагональные элементы матрицы C нулевые.

Начальное распределение вагонов характеризуется следующими векторами:

$$\bar{S}^0(1) = \begin{pmatrix} 0 \\ 2 \\ 1 \\ 3 \end{pmatrix}; \quad \bar{S}^0(2) = \begin{pmatrix} 5 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

В период $t = 3$ вагоны не прибывают, что эквивалентно нулевому вектору $\bar{S}^0(3)$.

Размерность задачи

Размерность представленной в работе [17] задачи равна $2TN^2 = 96$. Рассчитаем размерность при решении этой же задачи с помощью представленного в данной работе алгоритма. Предложенный алгоритм даёт размерность $T \cdot (N_{cargo} + N_{empty})$, поэтому для того, что бы ее рассчитать необходимо знать значения параметров N_{cargo} и N_{empty} . Для определения N_{empty} необходимо понимать, какие порожние маршруты планируется включить в расчет, а какие — нет. Исключим из рассмотрения те порожние маршруты, станции прибытия в которых не являются станциями отправления ни для одной из заявок на груженные рейсы. Такая станция одна и это четвертая станция. Уберем из рассмотрения все порожние маршруты, в которых пунктом назначения является станция 4, оставим только маршрут из 4 в 4 (вагон остается на станции до следующего периода). Получается, из рассмотрения убираются порожние маршруты из 1 в 4, из 2 в 4 и из 3 в 4. Получаем, что $N_{empty} = N^2 - 3 = 13$. Что касается N_{cargo} , то его значение равно числу заявок, т.е. в нашем случае $N_{cargo} = 5$. Таким образом, размерность при использовании нового алгоритма получается равной $T \cdot (N_{cargo} + N_{empty}) = 54$. Получается, что конкретно для этого примера размерность сократилась при-

мерно на 44%. Отдельно отметим, что на практике из расчета можно исключать порожние маршруты и по другим признакам, например, можно не учитывать порожние маршруты, тариф по которым выше некоторого порогового значения. Поэтому в реальных задачах можно добиться еще большего сокращения размерности по сравнению с $2TN^2$.

В качестве примера можно привести задачу поиска оптимального плана, которая решалась на практике для $N = 1126$ станций, со сроком планирования $T = 30$ дней и количеством заявок на транспортировку грузов равным 1616. Размерность задачи при решении её алгоритмом из работы [17] составляет $2TN^2 = 76072560$. Для определения размерности задачи, которая получится с помощью представленного в данной работе алгоритма рассчитаем N_{cargo} и N_{empty} . Очевидно, $N_{cargo} = 1616$, что соответствует количеству заявок. Для расчета N_{empty} из рассмотрения необходимо убирать все порожние маршруты в направлении станций, которые не фигурируют в заявках в качестве станций отправления. Кроме этого, из рассмотрения убираются порожние маршруты, тарифы по которым превышают 50000 руб. В результате этого количество порожних маршрутов, которые будут учитываться при расчете, получается равным $N_{empty} = 82058$, т.е. в расчет попадает примерно 6,5% от всех возможных порожних маршрутов, количество которых равно $N^2 = 1267877$. В итоге размерность задачи получается равной $T \cdot (N_{cargo} + N_{empty}) = 2510220$, т.е. уменьшается примерно в 30 раз.

Задача линейного программирования

Выпишем задачу линейного программирования (1)–(3) для модельного примера. Для этого определим значения матриц A_{in} , A_{out} и A_Q , а также векторов PC , S_0 , Q . После этого решим данную задачу и сравним полученное решение с решением из [17].

Составим динамические списки *routes_from_station_cargo*, *routes_to_station_cargo* и *routes_from_station_empty*, *routes_to_station_empty*:

$$routes_from_station_cargo = \{1, 2, 2, 3, 3\},$$

$$routes_to_station_cargo = \{3, 1, 3, 2, 4\},$$

$$routes_from_station_empty = \{1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4\},$$

$$routes_to_station_empty = \{1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, 4\}.$$

Составим векторы p и c :

$$p = (2,9 \quad 1,1 \quad 2,3 \quad 1,9 \quad 2,1)^T$$

$$c = (0 \quad 1,9 \quad 1,3 \quad 1,2 \quad 0 \quad 1,8 \quad 1,1 \quad 1,2 \quad 0 \quad 1,3 \quad 1,5 \quad 1,2 \quad 0)^T.$$

Вектор PC получается последовательной конкатенацией полученных векторов:

$$PC = (p^T, p^T, p^T, c^T, c^T, c^T).$$

Вектор Q , отвечающий за объем заявок, принимает следующий вид:

$$Q = (3 \ 5 \ 4 \ 7 \ 6)^T.$$

Вектор S_0 , характеризующий начальное распределение вагонов по времени и по станциям, примет вид:

$$S_0 = (0 \ 2 \ 1 \ 3 \ 5 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0)^T.$$

Получим матрицы A_{in} , A_{out} и A_Q . Поскольку эти матрицы являются разреженными матрицами, и ненулевые элементы могут принимать только единичные значения, поэтому, как и в работе [17], запишем эти матрицы в разреженном формате, указывая координаты элементов, принимающих единичные значения. Выпишем единичные координаты матрицы A_Q , размерность которой равна $N_{cargo} \times T \cdot (N_{cargo} + N_{empty}) = 5 \times 54$ (здесь и далее нумерация строк и столбцов начинается с единицы):

(1, 1), (2, 2), (3, 3), (3, 4), (4, 5), (1, 6), (2, 7), (3, 8), (4, 9), (5, 10), (1, 11), (2, 12), (3, 13), (4, 14), (5, 15).

Координаты единичных элементов матрицы A_{in} примут следующий вид:

(7, 1), (5, 2), (11, 3), (10, 4), (12, 5), (11, 6), (9, 7), (5, 16), (6, 17), (7, 18), (5, 19), (6, 20), (7, 21), (5, 22), (6, 23), (7, 24), (5, 25), (10, 26), (7, 27), (8, 28), (9, 29), (10, 30), (11, 31), (9, 32), (10, 33), (11, 34), (9, 35), (10, 36), (10, 37), (9, 38), (11, 40), (12, 41).

Список координат единичных элементов матрицы A_{out} :

(1, 1), (2, 2), (2, 3), (3, 4), (3, 5), (5, 6), (6, 7), (6, 8), (7, 9), (7, 10), (9, 11), (10, 12), (10, 13), (11, 14), (11, 15), (1, 16), (1, 17), (1, 18), (2, 19), (2, 20), (2, 21), (3, 22), (3, 23), (3, 24), (4, 25), (4, 26), (4, 27), (4, 28), (5, 29), (5, 30), (5, 31), (6, 32), (6, 33), (6, 34), (7, 35), (7, 36), (7, 37), (8, 38), (8, 39), (8, 40), (8, 41), (9, 42), (9, 43), (9, 44), (10, 45), (10, 46), (10, 47), (11, 48), (11, 50), (12, 51), (12, 52), (12, 53), (12, 54).

Размерность матриц A_{in} и A_{out} равна $T \cdot N \times T \cdot (N_{cargo} + N_{empty}) = 12 \times 54$.

Выпишем решение, которое было получено с помощью программного продукта MatLab. Поскольку вектор K , состоящий из $T \cdot (N_{cargo} + N_{empty}) = 54$ элементов, также в основном состоит из нулевых элементов, выпишем значения только ненулевых элементов: $K_3 = 2$; $K_4 = 1$; $K_6 = 3$; $K_{13} = 2$; $K_{14} = 4$; $K_{15} = 6$; $K_{26} = 1$; $K_{28} = 2$; $K_{31} = 2$; $K_{40} = 3$.

Выпишем это же решение в более понятном формате матриц $K1(t)$ и $K2(t)$, представляющие из себя $(N \times N)$ -матрицы, элементы которой характеризуют количество отправляемых груженных ($K1(t)$) и порожних ($K2(t)$) вагонов от станции i до станции j на момент времени $t \in \{1, \dots, T\}$.

$$K1(1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad K1(2) = \begin{pmatrix} 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

$$K2(1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}; \quad K1(3) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 4 & 0 & 6 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

$$K2(2) = \begin{pmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{pmatrix}; \quad K2(3) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Графически данное решение представлено на рис. 1, ширина каждой полосы в котором характеризует количество вагонов, отправленное в заданном направлении.

Целевое значение итоговой прибыли, вычисляемое по правилу $PC^T \cdot K$, получается равным 32,3.

Если сравнивать полученное решение с решением из [17], то видно, что они отличаются, но значения целевого функционала, выражающего итоговую прибыль, при этом совпадают. Таким образом сравнение решений одной и той же задачи показывает, что поставленная задачи имеет как минимум два различных решения.

Заключение

Данная статья является продолжением работы [17], в ней представлен модифицированный алгоритм решения задачи оптимального управления парком грузовых вагонов. Суть предложенного подхода состоит в исключении из расчета тех груженных или порожних маршрутов, про которые заранее известно, что они либо не будут задействованы в итоговом решении, либо вероятность появления этих маршрутов в решении оценивается как очень низкая. На представленном в работе модельном примере видно, что применение модифицированного алгоритма приводит к сокращению размерности примерно на 44%. В задачах, которые встречаются на практике, как пра-

более тонкими методами. Например, при построении пространственно-временного графа для станций второго и третьего типов можно дополнительно учитывать то, с какого самого раннего момента времени в этих станциях могут начать появляться вагоны и до этого момента модельного времени для соответствующих станций граф не строить. В этом случае в борьбе за снижение размерности единственной платой является еще большее усложнение алгоритмов формирования матриц и векторов для задачи (1)–(3), что в свою очередь увеличивает вероятность ошибок при создании таких алгоритмов.

Помимо усилий по дальнейшей оптимизации алгоритмов формирования матриц и векторов другим направлением для развития данного типа задач является модернизация постановки задачи оптимального управления парком грузовых вагонов с целью учета большего числа ограничений. В текущей версии транспортная задача представляет ис-

ключительно научный интерес, но никак не практический. Для железнодорожных транспортных операторов, являющихся основными заказчиками подобных моделей, важна возможность учета достаточно большого числа факторов, среди которых можно отметить учет различных типов вагонов, запрет для некоторых из типов вагонов заезжать на определенные территории, учет станций отстоя, ограничения по минимальному или максимальному количеству вагонов, которое должно двигаться за расчетный период по заданным направлениям и т.д. Изучение описанных выше проблем может стать предметом будущих исследований. ■

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-29-06003.

Литература

1. Higgins A., Ferreira L., Kozan E. Modeling single-line train operations // *Transportation Research Record*. 1995. Vol. 1489. P. 9–16.
2. Kraay D., Barker P., Chen B. Optimal pacing of trains in freight railroads: model formulation and solution // *Operations Research*. 1991. Vol. 39. No. 1. P. 82–99. <https://doi.org/10.1287/opre.39.1.82>
3. Ferreira L., Murray M. Modelling rail track deterioration and maintenance: current practices and future needs // *Transport Reviews*. 1997. Vol. 17. No. 3. P. 207–221. <https://doi.org/10.1080/01441649708716982>
4. LeBlanc L. Global solutions for a nonconvex nonconcave rail network model // *Management Science*. 1976. Vol. 23. No. 2. P. 131–139. <https://doi.org/10.1287/mnsc.23.2.131>
5. Лазарев А.А., Мусатова Е.Г., Кварацхелия А.Г., Гафаров Е.Р. Теория расписаний. Задачи управления транспортными системами. М.: МГУ, 2012.
6. Лазарев А.А., Мусатова Е.Г., Гафаров Е.Р., Кварацхелия А.Г. Теория расписаний. Задачи железнодорожного планирования. М.: ИПУ РАН, 2012.
7. Beklaryan L., Khachatryan N. Traveling wave type solutions in dynamic transport models // *Functional Differential Equations*. 2006. Vol. 13. No. 2. P. 125–155.
8. Бекларян Л.А., Хачатрян Н.К. Об одном классе динамических моделей грузоперевозок // *Журнал вычислительной математики и математической физики*. 2013. Т. 53. № 10. С. 1649–1667. <https://doi.org/10.7868/S0044466913100037>
9. Хачатрян Н.К. Динамическая модель организации грузоперевозок при ограниченности емкостей перегонных путей // *Бизнес-информатика*. 2013. Т. 26. № 4. С. 62–68.
10. Хачатрян Н.К., Акопов А.С. Модель организации грузоперевозок с начальной станцией отправления и конечной станцией распределения грузов // *Бизнес-информатика*. 2017. Т. 39. № 1. С. 25–35. <https://doi.org/10.17323/1998-0663.2017.1.25.35>
11. Хачатрян Н.К., Акопов А.С., Белоусов Ф.А. О квазирешениях типа бегущей волны в моделях организации грузоперевозок // *Бизнес-информатика*. 2018. Т. 43. № 1. С. 61–70. <https://doi.org/10.17323/1998-0663.2018.1.61.70>
12. Хачатрян Н.К., Бекларян Г.Л., Борисова С.В., Белоусов Ф.А. Исследование динамики емкостей перегонов в модели организации грузоперевозок между двумя узловыми станциями // *Бизнес-информатика*. 2019. Т. 13. № 1. С. 59–70. <https://doi.org/10.17323/1998-0663.2019.1.59.70>
13. Beklaryan L., Khachatryan N., Akopov A. Model for organization cargo transportation at resource restrictions // *International Journal of Applied Mathematics*. 2019. Vol. 32. No. 4. P. 627–640. <https://doi.org/10.12732/ijam.v32i4.7>
14. Хачатрян Н.К., Бекларян Л.А. Исследование динамики потока в модели организации грузоперевозок по круговой цепочке станций // *Экономика и математические методы*. 2021. Т. 57. № 1. С. 83–91. <https://doi.org/10.31857/S042473880013024-5>
15. Khachatryan N. Study of flow dynamics in the model of cargo transportation organization between node stations // *International Journal of Applied Mathematics*. 2020. Vol. 33. No. 5. P. 937–949. <https://doi.org/10.12732/ijam.v33i5.14>
16. Khachatryan N. Modeling the process of cargo transportation between node stations // *International Journal of Applied Mathematics*. 2021. Vol. 34. No. 6. P. 1223–1235. <https://doi.org/10.12732/ijam.v34i6.12>
17. Белоусов Ф.А., Неволин И.В., Хачатрян Н.К. Моделирование и оптимизация планов грузовых железнодорожных перевозок, выполняемых транспортным оператором // *Бизнес-информатика*. 2020. Т. 14. № 2. С. 21–35. <https://doi.org/10.17323/2587-814X.2020.2.21.35>

18. Sadykov R., Lazarev A., Shiryayev V., Stratonnikov A. Solving a freight railcar flow problem arising in Russia // Proceedings of the 13th Workshop on Algorithmic Approach for Transportation Modelling, Optimization, and Systems (ATMOS'13), Leibniz, Germany, 5 September 2013. P. 55–67. <https://doi.org/10.4230/OASIS.ATMOS.2013.55>
19. Лазарев А.А., Садыков Р.Р. Задача управления парком грузовых железнодорожных вагонов // XII Всероссийское совещание по проблемам управления (ВСПУ 2014), Москва, ИПУ РАН, 16–19 июня 2014. С. 5083–5093.
20. Desaulniers G., Desrosiers J., Solomon M. Column generation. New York: Springer, 2005. <https://doi.org/10.1007/b135457>
21. Lübbecke M. Column generation. Wiley Encyclopedia of Operations Research and Management Science. John Wiley & Sons, 2011. <https://doi.org/10.1002/9780470400531.eorms0158>
22. Frangioni A., Gendron B. 0–1 reformulations of the multicommodity capacitated network design problem // Discrete Applied Mathematics. 2009. Vol. 157. No. 6. P. 1229–1241. <https://doi.org/10.1016/j.dam.2008.04.022>
23. Sadykov R., Vanderbeck F. Column generation for extended formulations // EURO Journal on Computational Optimization. 2013. Vol. 1. Nos. 1–2. P. 81–115. <https://doi.org/10.1007/s13675-013-0009-9>
24. Dantzig G., Wolfe P. Decomposition principle for linear programs // Operations Research. 1960. Vol. 8. No. 1. P. 101–111. <https://doi.org/10.1287/opre.8.1.101>
25. Fukasawa R., Aragão M.P., Porto O., Uchoa E. Solving the freight car flow problem to optimality // Electronic Notes in Theoretical Computer Science. 2002. Vol. 66. No. 6. P. 42–52. [https://doi.org/10.1016/S1571-0661\(04\)80528-0](https://doi.org/10.1016/S1571-0661(04)80528-0)

Об авторах

Белоусов Федор Анатольевич

кандидат экономических наук;

старший научный сотрудник, лаборатория динамических моделей экономики и оптимизации, Центральный экономико-математический институт, Российская академия наук, 117418, г. Москва, Нахимовский проспект, д. 47;

E-mail: sky_tt@list.ru

ORCID: 0000-0002-3040-3148

Хачатрян Нерсес Карленович

кандидат физико-математических наук;

заместитель директора ЦЭМИ РАН по научной работе, Центральный экономико-математический институт, Российская академия наук, 117418, г. Москва, Нахимовский проспект, д. 47;

E-mail: nerses-khachatryan@yandex.ru

ORCID: 0000-0003-2495-5736

Неволин Иван Викторович

кандидат экономических наук;

ведущий научный сотрудник, лаборатория экспериментальной экономики, Центральный экономико-математический институт, Российская академия наук, 117418, г. Москва, Нахимовский проспект, д. 47;

E-mail: i.nevolin@cemi.rssi.ru

ORCID: 0000-0002-8462-9011

Reduction of dimension in the problem of optimal management of a freight cars fleet using unmanned locomotives

Fedor A. Belousov

E-mail: sky_tt@list.ru

Nerses K. Khachatryan

E-mail: nerses-khachatryan@yandex.ru

Ivan V. Nevolin

E-mail: i.nevolin@cemi.rssi.ru

Central Economics and Mathematics Institute, Russian Academy of Science

Address: 47, Nakhimovsky Prospect, Moscow 117418, Russia

Abstract

This paper considers the problem of optimal management of a fleet of freight cars by a transport railway operator. The solution to this problem is an optimal plan, which is a timetable for the movement of freight and empty railway cars, following which the transport operator will receive the maximum profit for the estimated period of time. This problem is reduced to the problem of linear programming of large dimension. Unlike the works of other authors on this topic, which mainly deal with methods of numerical solution of the corresponding linear programming problems, this article focuses on an algorithm that allows one to reduce their dimensionality. This can be achieved by excluding from the calculation those routes that obviously cannot be involved in the solution, or whose probability of participation in the final solution is estimated as extremely low. The effectiveness of the proposed modified algorithm was confirmed both on a model example (several stations, a short planning horizon) and on a real example (more than 1 000 stations, a long planning horizon). In the first case, there was a decrease in the dimension of the problem by 44%, while in the second – by 30 times.

Keywords: railway freight transportation, optimal plan, optimal management of the fleet of cars, linear programming, theory of schedules, operations research, unmanned locomotives

Citation: Belousov F.A., Khachatryan N.K., Nevolin I.V. (2022) Reduction of dimension in the problem of optimal management of a freight cars fleet using unmanned locomotives. *Business Informatics*, vol. 16, no. 2, pp. 7–20. DOI: 10.17323/2587-814X.2022.2.7.20

References

1. Higgins A., Ferreira L., Kozan E. (1995) Modeling single-line train operations. *Transportation Research Record*, vol. 1489, pp. 9–16.
2. Kraay D., Barker P., Chen B. (1991) Optimal pacing of trains in freight railroads: model formulation and solution. *Operations Research*, vol. 39, no. 1, pp. 82–99. <https://doi.org/10.1287/opre.39.1.82>
3. Ferreira L., Murray M. (1997) Modelling rail track deterioration and maintenance: current practices and future needs. *Transport Reviews*, vol. 17, no. 3, pp. 207–221. <https://doi.org/10.1080/01441649708716982>
4. LeBlanc L. (1976) Global solutions for a nonconvex nonconcave rail network model. *Management Science*, vol. 23, no. 2, pp. 131–139. <https://doi.org/10.1287/mnsc.23.2.131>
5. Lazarev A., Musatova E., Kvaratskheliya A., Grafov E. (2012) *Schedule theory. Problems of transport system management*. Moscow: MSU (in Russian).
6. Lazarev A., Musatova E., Grafov E., Kvaratskheliya A. (2012) *Schedule theory. Problems of railway planning*. Moscow: ICS RAS (in Russian).
7. Beklaryan L., Khachatryan N. (2006) Traveling wave type solutions in dynamic transport models. *Functional Differential Equations*, vol. 13, no. 2, pp. 125–155.
8. Beklaryan, L., Khachatryan N. (2013) On one class of dynamic transport models. *Computational Mathematics and Mathematical Physics*, vol. 53, no. 10, pp. 1649–1667. <https://doi.org/10.7868/S0044466913100037>
9. Khachatryan N. (2013) Dynamic model of organization of cargo transportation with a limited capacity of the distillation ways. *Business Informatics*, vol. 26, no. 4, pp. 62–68.
10. Khachatryan N., Akopov A. (2017) Model for organizing cargo transportation with an initial station of departure and a final station of cargo distribution. *Business Informatics*, vol. 39, no. 1, pp. 25–35. <https://doi.org/10.17323/1998-0663.2017.1.25.35>
11. Khachatryan N., Akopov A., Belousov F. (2018) About quasi-solutions of traveling wave type in models for organizing cargo transportation. *Business Informatics*, vol. 43, no.1, pp. 61–70. <https://doi.org/10.17323/1998-0663.2018.1.61.70>
12. Khachatryan N.K., Beklaryan G.L., Borisova S.V., Belousov F.A. (2019) Research into the dynamics of railway track capacities in a model for organizing cargo transportation between two node stations. *Business Informatics*, vol. 13, no. 1, pp. 59–70. <https://doi.org/10.17323/1998-0663.2019.1.59.70>
13. Beklaryan L., Khachatryan N., Akopov A. (2019) Model for organization cargo transportation at resource restrictions. *International Journal of Applied Mathematics*, vol. 32, no. 4, pp. 627–640. <https://doi.org/10.12732/ijam.v32i4.7>
14. Khachatryan N., Beklaryan L. (2021) Study of flow dynamics in the model of cargo transportation organization along a circular chain of stations. *Economics and Mathematical Methods*, vol. 57, no. 1, pp. 83–91. <https://doi.org/10.31857/S042473880013024-5>
15. Khachatryan N. (2020) Study of flow dynamics in the model of cargo transportation organization between node stations. *International Journal of Applied Mathematics*, vol. 33, no. 5, pp. 937–949. <https://doi.org/10.12732/ijam.v33i5.14>
16. Khachatryan N. (2021) Modeling the process of cargo transportation between node stations. *International Journal of Applied Mathematics*, vol. 34, no. 6, pp. 1223–1235. <https://doi.org/10.12732/ijam.v34i6.12>

17. Belousov F.A., Nevolin I.V., Khachatryan N.K. (2020) Modeling and optimization of plans for railway freight transport performed by a transport operator. *Business Informatics*, vol. 14, no. 2, pp. 21–35. <https://doi.org/10.17323/2587-814X.2020.2.21.35>
18. Sadykov R., Lazarev A., Shiryaev V., Stratonnikov A. (2013) Solving a freight railcar flow problem arising in Russia. Proceedings of the 13th Workshop on Algorithmic Approach for Transportation Modelling, Optimization, and Systems (ATMOS'13), Leibniz, Germany, 5 September 2013, pp. 55–67. <https://doi.org/10.4230/OASICS.ATMOS.2013.55>
19. Lazarev A., Sadykov R. (2014) Management problem of railway cars fleet. Proceedings of the XII All-Russian Meeting on Management Issues (VSPU 2014), ICS RAS, Moscow, Russia, 16–19 June 2014, pp. 5083–5093 (in Russian).
20. Desaulniers G., Desrosiers J., Solomon M. (2005) *Column generation*. New York: Springer. <https://doi.org/10.1007/b135457>
21. Lübbecke M. (2011) *Column generation*. *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons. <https://doi.org/10.1002/9780470400531.eorms0158>
22. Frangioni A., Gendron B. (2009) 0–1 reformulations of the multicommodity capacitated network design problem. *Discrete Applied Mathematics*, vol. 157, no. 6, pp. 1229–1241. <https://doi.org/10.1016/j.dam.2008.04.022>
23. Sadykov R., Vanderbeck F. (2013) Column generation for extended formulations. *EURO Journal on Computational Optimization*, vol. 1, nos. 1–2, pp. 81–115. <https://doi.org/10.1007/s13675-013-0009-9>
24. Dantzig G., Wolfe P. (1960) Decomposition principle for linear programs. *Operations Research*, vol. 8, no. 1, pp. 101–111. <https://doi.org/10.1287/opre.8.1.101>
25. Fukasawa R., Aragão M.P., Porto O., Uchoa E. (2002) Solving the freight car flow problem to optimality. *Electronic Notes in Theoretical Computer Science*, vol. 66, no. 6, pp. 42–52. [https://doi.org/10.1016/S1571-0661\(04\)80528-0](https://doi.org/10.1016/S1571-0661(04)80528-0)

About the authors

Fedor A. Belousov

Cand. Sci. (Econ.);

Senior Researcher, Laboratory of Dynamic Models of Economy and Optimization, Central Economics and Mathematics Institute, Russian Academy of Science, 47, Nakhimovsky Prospect, Moscow 117418, Russia;

E-mail: sky_tt@list.ru

ORCID: 0000-0002-3040-3148

Nerses K. Khachatryan

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

Deputy Director of CEMI RAS for Scientific Work, Central Economics and Mathematics Institute, Russian Academy of Science, 47, Nakhimovsky Prospect, Moscow 117418, Russia;

E-mail: nerses-khachatryan@yandex.ru

ORCID: 0000-0003-2495-5736

Ivan V. Nevolin

Cand. Sci. (Econ.);

Leading Researcher, Laboratory of Experimental Economics, Central Economics and Mathematics Institute, Russian Academy of Science, 47, Nakhimovsky Prospect, Moscow 117418, Russia;

E-mail: i.nevolin@cemi.rssi.ru

ORCID: 0000-0002-8462-9011

Разработка модели формирования индивидуальных образовательных траекторий с использованием методов машинного обучения*

Е.С. Морозевич^a 

E-mail: katyamorozevich@mail.ru

В.С. Коротких^b 

E-mail: vskor@bk.ru

Е.А. Кузнецова^a 

E-mail: zhenya.kuz-1997@yandex.ru

^a Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева
Адрес: Россия, 660037, г. Красноярск, проспект им. газеты Красноярский рабочий, д. 31

^b Муниципальное бюджетное общеобразовательное учреждение средняя общеобразовательная школа №4
Адрес: Россия, 663091, г. Дивногорск, ул. Набережная, д. 9

Аннотация

Сегодня экономика переживает цифровую трансформацию, ключевыми барьерами которой являются нехватка квалифицированных кадров, компетенций и знаний, а также внутреннее сопротивление в организациях. Преодолеть их можно через качественное развитие и обучение персонала. Актуальной задачей является построение индивидуальных образовательных траекторий. Современные исследования направлены на имплементацию рекомендательных систем в целях подбора релевантного материала. Однако основой таких рекомендаций являются цифровые следы; полный личностный профиль обучающегося, а также организационные ценности не учитываются. Целью исследования является создание интеллектуального проводника, который бы сопровождал сотрудника на протяжении всей жизни в организации, вовлекая его в процесс обучения согласно персонифицированной траектории на основе данных о комплексном личностном профиле и реакциях на образовательный материал, тренируя soft и hard skills в соответствии с ценностями организации и сотрудника. В исследовании использовались методы системного анализа, системной инженерии, психодиагностического исследования (модель DISC, методика Алана Роу «Стиль принятия решений», методика определения стилей деятельности Хони и Мамфорда, тест на психотип), проектирования программного обеспечения и искусственного интеллекта (матричная факторизация и нейронные сети). Исследование проведено на уникальной базе данных, собранной в рамках его проведения и состоящей из образовательных задач для развития soft skills, данных по их выполнению пользователями с разными soft skills профилями. Разработана модель интеллектуального проводника и произведена ее реализация в виде программного компонента для системы управления предприятием, основу которого составили модули психодиагностики, организационного управления, обучения и рекомендаций. Интеллектуальность разработанной системы позволяет качественно сформировать индивидуальную образовательную траекторию, которая вовлечет

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

сотрудника не только в процесс обучения и развития, но и в достижение организационных целей. Вкладывая в развитие сотрудников, организация получает T-shaped специалистов, обладающих проактивной позицией и способных к самоорганизации. Результаты исследования могут использоваться предприятиями не только на организационном уровне, но и через трансляцию в системе образования для формирования экосистемы образования в соответствии с требованиями инновационного развития экономики региона.

Ключевые слова: обучение и развитие, надпрофессиональные компетенции, адаптивность, рекомендательная система, машинное обучение, матричная факторизация, нейронная сеть

Цитирование: Морозевич Е.С., Коротких В.С., Кузнецова Е.А. Разработка модели формирования индивидуальных образовательных траекторий с использованием методов машинного обучения // Бизнес-информатика. 2022. Т. 16. № 2. С. 21–35. DOI: 10.17323/2587-814X.2022.2.21.35

Введение

Сегодня экономика переживает цифровую трансформацию, которая представляет собой революционный процесс преобразования бизнес-модели организации не только с применением цифровых технологий, но и внесением коренных организационных изменений в технологии, культуру, операции и принципы создания новых продуктов.

Связующим звеном интеграции глубинных преобразований организаций и новых технологий выступают люди — сотрудники с их знаниями, навыками и ценностями. По данным ООО «Делойт Консалтинг» уже сегодня 53% организаций понимают, что в текущих условиях от 50 до 100% их сотрудников должны приобрести новые навыки и способности [1].

Организации создаются людьми с вложением своей системы ценностей в создаваемую структуру. Сотрудник, приходящий в организацию на определенном этапе жизненного цикла, не осознает, что происходит, потому что: он не знает — что было, он не догадывается — что будет. Ввиду этого актуальной задачей является развитие персонала с учетом точек соприкосновения ценностей сотрудника и организации.

Цифровизация деятельности порождает рост объема данных, которые требуют последующей обработки. Именно данные лежат в основе цифровой трансформации, являясь источником жизненной силы процесса перехода к цифровой экономике.

Уже в 2020 году по данным Data Age Report человечеством было сформировано около 51 зеттабайта информации, а к 2025 году объем этих данных увеличится почти в 3,5 раза и составит 175 зеттабайтов [2], что показывает тенденцию экспоненциального роста цифровых данных.

Активность пользователей, фиксируемая различными регистрирующими устройствами, представляет собой их цифровые следы. Сегодня цифровые следы занимают существенную часть облака больших данных и ключевое направление их использования — это извлечение информации о предпочтениях потенциальных клиентов и предложение продукции, которая им будет интересна, то есть повышение объема продаж [3].

Однако извлекаемую информацию из цифровых следов можно использовать не только во внешних взаимодействиях, но и во внутренней деятельности компании. Современные компании все чаще используют данные для повышения вовлеченности сотрудников в достижение и разделение целей, то есть для повышения продуктивности как сотрудников, так и процессов организации, тем самым создавая новые источники для своего конкурентного преимущества [4].

Таким образом, неотъемлемой частью крупных игроков рынка сегодня являются рекомендательные системы. Рекомендательная система представляет собой комплекс алгоритмов, программ и сервисов, предназначенных для формирования релевантных рекомендаций пользователям в отношении объекта информационного поиска [5].

Во многих исследованиях отмечается [6–8], что преодолеть ключевые барьеры цифровой трансформации в виде нехватки квалифицированных кадров, компетенций и знаний, а также внутреннего сопротивления в организациях [9–10] можно путем создания системы обучения и развития персонала, в основе которой лежит формирование индивидуальной образовательной траектории [11] на базе ценностей организации и сотрудника и релевантном образовательном материале, то есть на тех объ-

ектах, действиях, задачах, которые сотрудник будет выполнять с повышенной заинтересованностью и как следствие с максимальной эффективностью.

В сфере образования использование информационных систем, основанных на рекомендациях, призвано решать одну из главных проблем обучающегося – проблему выбора образовательного материала. Зачастую таким материалом являются программы обучения и образовательные курсы. При подборе образовательного материала в таких системах обычно не учитываются личностные характеристики обучающегося, а используются его предпочтения [12]. В последнее время исследователи начали активно изучать вопрос использования личностных характеристик в образовательных рекомендациях. Например, использование информации о стиле обучения, эмоциональных реакциях [13]. Кроме того, в исследованиях все чаще поднимается вопрос об использовании рекомендательных систем в онлайн обучении [14].

Исходя из вышесказанного, актуальной целью исследования является создание интеллектуального проводника, который бы сопровождал сотрудника на протяжении всей жизни в организации, вовлекая его в процесс обучения согласно персонализированной траектории на основе данных о личностном профиле и реакциях на образовательный материал, тренируя soft и hard skills в соответствии с ценностями организации и сотрудника.

1. Концепция интеллектуальной системы обучения и развития персонала

В основу разработки интеллектуального проводника для развития и обучения сотрудников легли принципы системной инженерии. Отрывной точкой в создании интеллектуального проводника

явился сбор требований заинтересованных сторон к разрабатываемой системе.

В результате анализа требований было выявлено, что модульная архитектура системы позволит не только учесть все требования заинтересованных сторон, но и разграничит доступ к различным функциям в зависимости от занимаемых позиций. Предлагаемая к реализации модульная архитектура в соответствии с выявленными требованиями представлена на *рис. 1*.

Для реализации интеллектуального проводника и поддержки работы предложенных модулей была создана база данных с помощью СУБД MySQL.

Основной задачей модуля диагностики является формирование цифрового профиля сотрудника, который позволит учесть все особенности личности и станет основой для ее эффективного обучения и развития.

Современные тенденции таковы, что организации все чаще делают упор на надпрофессиональные навыки или так называемые soft skills, которые гораздо труднее воспроизвести с помощью инструментов цифровой трансформации и которые лежат в основе происходящих изменений [15]. Soft skills являются основой для эффективного развития профессиональных навыков или так называемых hard skills.

Ввиду этого было предложено сформировать цифровой профиль сотрудника, состоящий не только из персональных данных (пол, возраст, образование, должность, опыт работы), но и раскрывающий уровни развития soft skills сотрудника, востребованных на рынке труда в условиях цифровой экономики.

Для определения уровня владения тем или иным soft skills была разработана комплексная автоматизированная

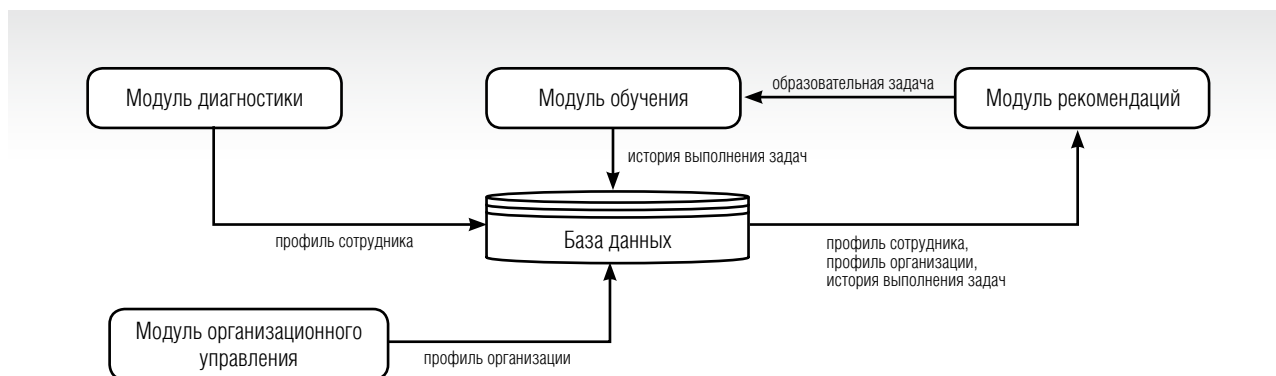


Рис. 1. Модель интеллектуального проводника.

зированной диагностика (рис. 2). В качестве диагностического материала использовались модель DISC, методика Алана Роу «Стиль принятия решений», методика определения стилей деятельности Хони и Мамфорда, тест на психотип. Данные методики раскрывают психоэмоциональные особенности личности в 16 проекциях с соответствующими им компетенциями.

Модуль организационного управления предназначен для фиксации организационных ценностей через формирование профилей должностей. В первую очередь через данный модуль производится создание организационной структуры. После этого для каждой рассматриваемой должности руководителем структурного подразделения определяются точки критического и желаемого профиля сотрудника в соответствии с трудовым функционалом и трудовыми действиями, а также обозначенными на основе организационных ценностей, надпрофессиональными навыками.

Таким образом, составив профиль должности для каждого элемента организационной структуры, формируется комплексный профиль организации, в котором отражены все особенности ее деятельности.

Основой модуля обучения является организация образовательного процесса. Пользователи в рамках

развития soft skills выполняют уникальные образовательные задачи (рис. 3). Для каждой задачи пользователь оценивает насколько понравилась ему задача, ее эффективность и сложность. Таким образом формируется набор данных, характеризующий выполнение пользователями образовательных задач и применяемый в дальнейшем для формирования релевантных рекомендаций.

Модуль рекомендаций призван формировать индивидуальную образовательную траекторию в соответствии с ценностями сотрудника и организации и подбирать максимально релевантный образовательный материал для сотрудника.

Имея профиль сотрудника и профиль должности можно определить направления развития soft skills сотрудника в рамках интересов организации. Однако, как было обозначено ранее, необходимо синхронизировать ценности сотрудника и организации, поэтому требуется также определить желаемые направления развития для сотрудника. Для этого в рамках модуля рекомендаций разработан опросник потребностей развития сотрудника в зависимости от его ценностных ориентиров.

После прохождения диагностики сотруднику предлагается ответить на перечень вопросов для формирования желаемого профиля сотрудника и

Вопрос: 5/64.

Когда я сталкиваюсь с проблемой, я:

Опираюсь на проверенные подходы	0	1	2	3	4
Провожу тщательный анализ	0	1	2	3	4
Ищу творческое решение	0	1	2	3	4
Опираюсь на мои чувства	0	1	2	3	4

выставьте каждому ответу оценку от 0 до 4, характеризующую на сколько утверждение близко вам. 4 - наиболее близко

[Вернуться в личный кабинет](#)

Рис.2. Модуль диагностики.

ДИАГНОСТИКА

ОБУЧЕНИЕ

ОРГАНИЗАЦИОННОЕ УПРАВЛЕНИЕ

ЛИЧНЫЕ ДАННЫЕ

Задачи

Активная задача

[компетенция: Доброжелательность]

Прочитайте стихотворение о дружелюбии «Путь Дружелюбия»

завершить задачу

Доступные задачи (подобраны случайно, переопределено администратором)

[компетенция: Объективность]

Читайте научно-популярную литературу: Кратчайшая история времени, Стивен Хокинг, Леонард Млодинов

взять задачу

[компетенция: Энергичность]

Во время занятия спортом включайте любимую музыку

взять задачу

[компетенция: Экспериментальность]

Почаще подвергайте себя новым испытаниям. Сходите в церковь не вашей веры.

взять задачу

[компетенция: Логическое мышление]

Старайтесь думать и делать выбор сам, иначе за вас будут делать другие.

взять задачу

[компетенция: Исполнительность]

Найдите способ тактично сказать своему сотруднику о увольнении.

взять задачу

[компетенция: Креативное мышление]

Прислушайтесь к звукам за окном, вдохните запах, доносящийся с кухни, осмотритесь вокруг, почувствуйте прикосновение теплого одеяла к ногам. Попытайтесь сохранить это состояние хотя бы на несколько минут. С каждым разом увеличивайте время.

взять задачу

Последние завершённые задачи (всего: 18)

[компетенция: Нетерпимость]

Как вы относитесь к взяточничеству в стране?

36.0 сек. 1 2

[компетенция: Абстрактность]

Сравните мопед и мотоцикл найдите общее и различное

55.0 сек. 1 3

Рис.3. Модуль обучения.

определения тех компетенций, которые он хотел бы развить.

Наложение полученных профилей друг на друга позволяет наглядно представить уровень выраженности компетенций и обозначить направления личностного развития сотрудника в соответствии с потребностями организации, что приводит к формированию их общей траектории развития.

Следующим шагом является непосредственный подбор релевантного образовательного материала с помощью рекомендательной системы. Для выбора алгоритма ее работы было проведено исследование таких технологий, как машинное обучение на основе матричной факторизации и нейронная сеть.

2. Вычислительный эксперимент

2.1. Постановка задачи

Есть множество пользователей $U = \{u_1, u_2, \dots, u_n\}$ и множество образовательных задач $T = \{t_1, t_2, \dots, t_m\}$.

Матрица $V_{n \times m}$ содержит оценки, выставленные пользователями образовательным задачам. На месте v_{ij} ($i \in 1, \dots, n; j \in 1, \dots, m$) будет стоять некоторое число, если пользователь u_i оценил задачу t_j и пусто в противном случае.

Требуется найти вектор $\hat{v}_i$, содержащий уже известные оценки пользователя v_{ij} , а также предполагаемые $\hat{v}_{ij}$. На основе полученного вектора произвести ранжирование списка образовательных задач T для пользователя u_i .

2.2. Машинное обучение на основе матричной факторизации

Матричная факторизация подразумевает декомпозицию исходной матрицы в произведение двух матриц малого ранга. Взаимодействие пользователя с объектом моделируется как скалярное произведение векторов представления пользователя и объекта в факторном пространстве. Факторизационные модели эффективно работают с сильно разреженными матрицами [16–18].

Представим матрицу оценок V в виде произведения двух матриц:

- ♦ матрицы $W_{n \times k}$, которая содержит в себе как латентные характеристики пользователей, так и явные (пол, образование, трудовой стаж и значения профилей пользователей, состоящих из 16 параметров);
- ♦ матрицы $H_{k \times m}$, характеризующей образовательные задачи.

Заполним латентные характеристики матриц W и H случайными величинами на основе закона равномерного распределения на интервале

$$[0; \sqrt{\max\{v_{ij}\} / k}].$$

Далее решим задачу минимизации:

$$\text{argmim} \|V - \hat{V}\| + \alpha \|H\| + \beta \|W\|, \quad (1)$$

где $\hat{V}$ – матрица, полученная аппроксимацией из W и H ;

α, β – параметры алгоритма.

Каждая итерация алгоритма поиска решения минимизации ошибки состоит из следующих шагов:

1. Фиксируем матрицу H .
2. Находим ошибку $\delta = |v_{ij} - \hat{v}_{ij}|, j \in 1, \dots, n$.
3. Находим новые значения

$$W_{ip} = W_{ip} - \epsilon (\delta H_{pj}^T + \lambda W_{ip}),$$

где $p \in 1, \dots, k$;

λ – регуляризационный параметр;

ϵ – скорость обучения.

4. Фиксируем матрицу W .
5. Находим ошибку $\delta = |v_{ij} - \hat{v}_{ij}|, j \in 1, \dots, m$.
6. Находим новые значения

$$H_{pj} = H_{pj} - \epsilon (\delta W_{ip}^T + \lambda H_{pj}),$$

где $p \in 1, \dots, k$;

λ – регуляризационный параметр;

ϵ – скорость обучения.

При решении поставленной задачи опытным путем был найден ранг матриц W и H $k = 40$. Данное значение является граничным для заданного случая, при котором среднеквадратичная и абсолютная ошибки решения задачи факторизации являются оптимальными. При значении $k > 40$ преимущество в результате не прослеживается, время расчетов увеличивается. И наоборот, при значении $k < 40$ получается слишком грубая аппроксимация.

На основе результатов тестирования (таблица 1) при увеличении скорости обучения ϵ со значения 0,005 до 0,03 наблюдается улучшение результатов (уменьшение погрешности, увеличение точности предсказания). При значении $\epsilon = 0,04$ результаты как правило эквивалентны результатам при $\epsilon = 0,03$ либо хуже. При количестве итераций 20 ($\epsilon = 0,03$) уже наблюдается высокая точность факторизации на заданных выборках. На основе данных результатов используем следующие параметры для алгоритма: $\epsilon = 0,03$, количество итераций 20.

При увеличении объемов полученных данных, следовательно, уменьшении разреженности данных, может потребоваться увеличение количества итераций. Контроль количества итераций был автоматизирован с помощью выполнения тестирования после окончания факторизации и сравнения точности предсказания с прошлым тестированием. При уменьшении точности предсказания, производится увеличение количества итераций, иначе количество итераций остается прежним. Еще одна альтернатива – задание допустимой точности факторизации, на которой необходимо останавливать алгоритм.

Одной из основных задач при построении рекомендательных систем является решение проблемы «холодного» старта [19]. Она возникает, когда в системе появляются новые элементы: будь то пользователь или объекты предпочтений.

В рассматриваемой задаче можно выделить два варианта холодного старта пользователей: пользователь не выполнял образовательные задачи; пользователь выполнял образовательные задачи, но задача факторизации решалась до его регистрации, то есть в системе отсутствуют факторы пользователя.

Рассмотрим первый случай. О предпочтениях пользователя ничего неизвестно. Только его профиль, являющийся некоторым вектором, который сформирован из значений пола, стажа, должности. Если пользователь прошел тестирование, то в векторе будет также заложен soft skills профиль из шестнадцати компонент.

Таблица 1.

Скорость и ошибки факторизации ($n = 14000$)

Количество итераций	ϵ	0,005	0,01	0,02	0,03	0,04
20	ср. абс.:	0,197	0,119	0,064	0,05	0,048
	ср. кв.:	0,24	0,09	0,013	0,007	0,0064
	время, с.	6	6	5	5	5
40	ср. абс.:	0,113	0,057	0,028	0,02	0,022
	ср. кв.:	0,082	0,013	0,002	0,001	0,001
	время, с.	11	10	10	11	11
60	ср. абс.:	0,075	0,03	0,017	0,014	0,014
	ср. кв.:	0,03	0,003	0,0009	0,0006	0,0006
	время, с.	16	15	16	15	15

Проблему холодного старта в данном случае предлагается решать с помощью косинусного сходства. Для этого находится косинус между новым пользователем и каждым существующим пользователем, ранжируется по убыванию, выбирается тот пользователь, которому соответствует максимальный косинус. В результате для нового пользователя формируются рекомендации уже на основании найденного через косинус-сходство пользователя.

$$\cos \varphi = \frac{(\overline{W_a}, \overline{W_b})}{\|\overline{W_a}\| \times \|\overline{W_b}\|}. \quad (2)$$

В рамках второго случая проблему предлагается разрешить путем решения задачи факторизации для конкретного пользователя. В этом случае матрицы W и V примут вид векторов-строк. Матрица факторов задач H будет зафиксирована, а поиск значений будет произведен только для вектора-строки W . Алгоритм поиска сводится к частному случаю: $i = 1$, шаги алгоритма матричной факторизации с четвертого по шестой исключаются.

Также стоит выделить проблему «холодного» старта для образовательных задач. На текущем этапе исследования задачи, для которых отсутствуют факторы, случайным образом предлагаются пользователям.

Для оценки качества предсказаний множество оценок V было разделено на выборки для обучения V_{train} и тестирования V_{test} . Тестирование модели проводилось при следующих количественных характеристиках:

- ◆ количество пользователей: 303;
- ◆ количество задач: 11726;
- ◆ количество выполненных пользователями задач: 17733;
- ◆ обучающая выборка: 75%;
- ◆ тестовая выборка: 25%;
- ◆ ранг матриц: 40;
- ◆ шаги факторизации: 20;
- ◆ диапазон значений исходной матрицы: 0–10 (нормализованные данные).

Матрица V является очень разреженной, большинство ячеек не заполнены (более 95% ячеек). При оценке работы модели проводилось три типа аппроксимации: для эффективности, сложности и предпочтений. Были рассмотрены четыре случая:

- ◆ использование только латентных факторов (на графиках обозначены как *FREE*);
- ◆ использование фиксированных факторов пользователей наряду с латентными (soft skills профиль, пол, должность, образование; обозначены как *UV*);
- ◆ использование фиксированных факторов задач наряду с латентными (средние значения показателей эффективности, сложности и предпочтений, выставленные пользователями, уровень сложности, ориентация задач на компетенции, обозначены как *TV*);
- ◆ использование фиксированных факторов и пользователей, и задач (обозначены как *UV+TV*).

Для каждого случая проводилось 50 факторизаций на случайных данных, далее эти данные усреднялись. При тестировании также учитывался холодный старт пользователей на основе косинус-сходства. Холодный старт задач не учитывался в тестировании.

Данные для ошибки предсказания предпочтений представлены на *рис. 4–7*, для ошибки предсказания эффективности – на *рис. 8–11*.

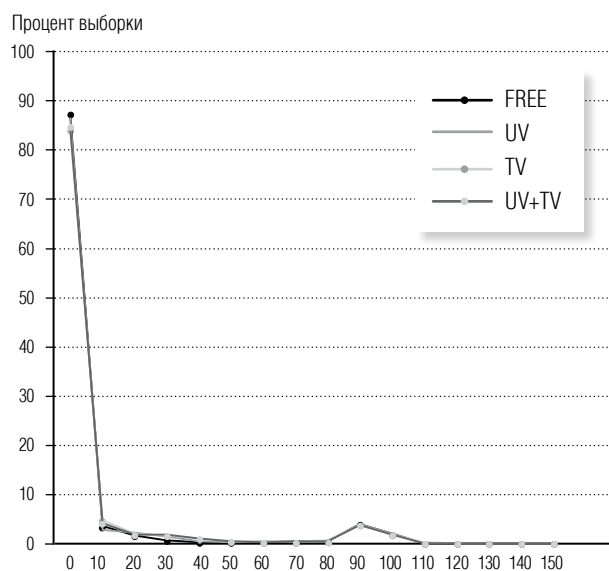


Рис. 4. Ошибка предсказания предпочтения с помощью факторизации на процентных интервалах.

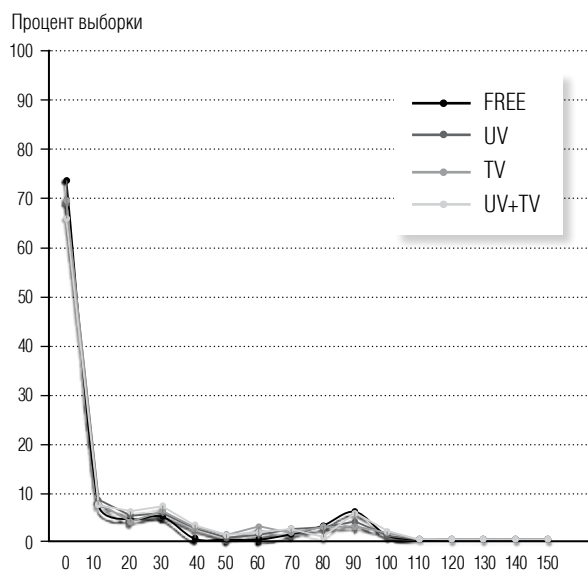


Рис. 6. Ошибка предсказания предпочтения с помощью косинус-сходства на процентных интервалах.

Как видно из графиков (*см. рис. 4–11*), использование фиксированных факторов в основном ухудшает результат. Это может быть следствием малого объема набора данных.

2.3. Нейронная сеть

Нейронная сеть представляет собой вычислительную структуру, собранную в сеть из вычислительных элементов, созданных на подобию биоло-

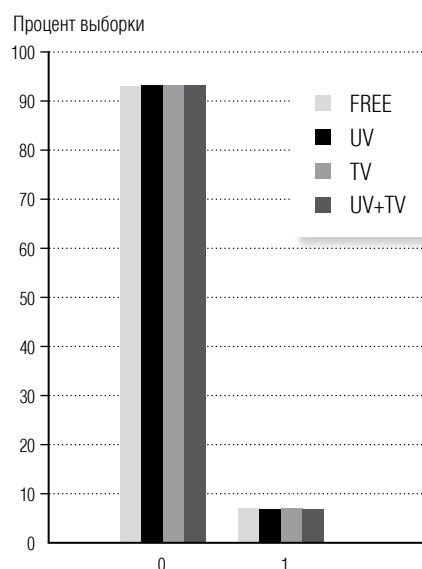


Рис. 5. Ошибка предсказания предпочтения с помощью факторизации на дискретных величинах [0; 1].

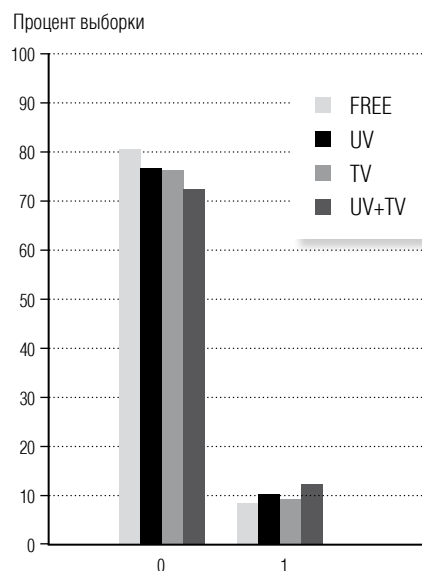


Рис. 7. Ошибка предсказания предпочтения с помощью косинус-сходства на дискретных величинах [0; 1].

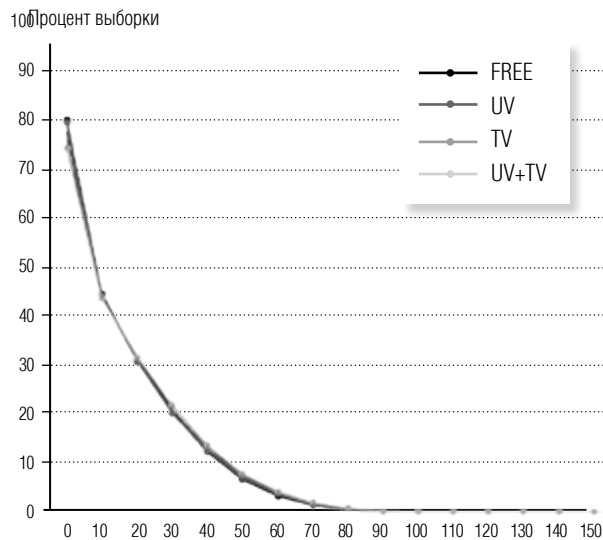


Рис. 8. Ошибка предсказания эффективности с помощью факторизации на процентных интервалах.

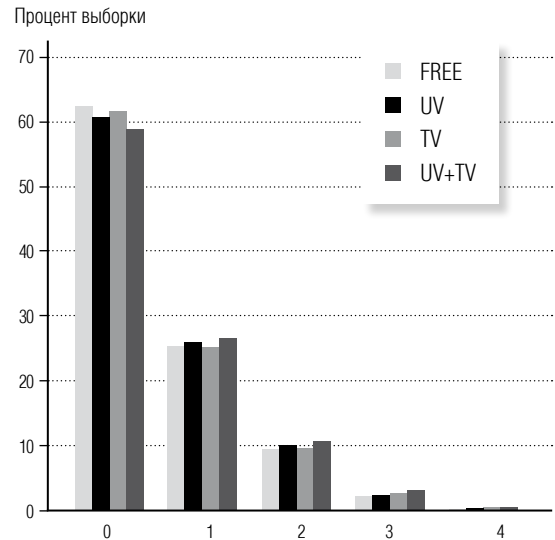


Рис. 9. Ошибка предсказания эффективности с помощью факторизации на дискретных величинах [1; 5].

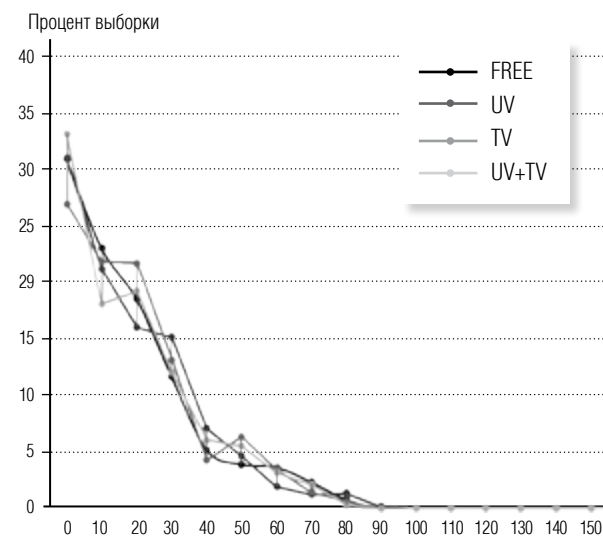


Рис. 10. Ошибка предсказания эффективности с помощью косинус-сходства на процентных интервалах.

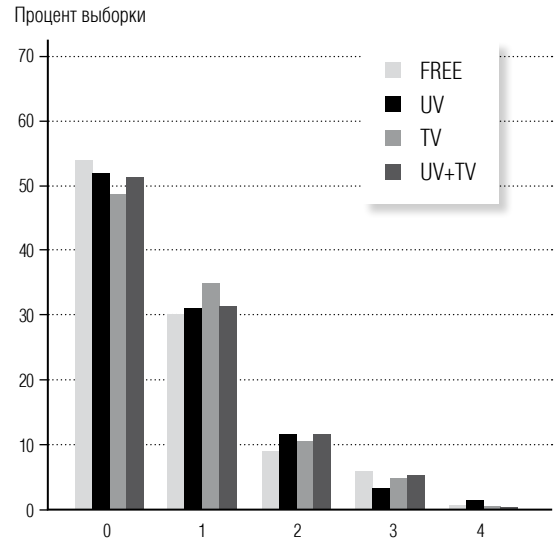


Рис. 11. Ошибка предсказания эффективности с помощью косинус-сходства на дискретных величинах [1; 5].

гического нейрона [20]. Нейронные сети способны решать широкий спектр задач в различных сферах деятельности [21].

Для решения поставленной задачи рассматривались такие архитектуры нейронных сетей как сверточная нейронная сеть [22] и многослойный персептрон по Румельхарту [23]. Наиболее эффективно показала себя последняя. К тому же, сверточ-

ная нейронная сеть требует большего количества данных и времени на обучение.

Для векторизации образовательной задачи было исследовано два способа: кодирование слов уникальным номером и вложение слов (Word2Vec) [24]. Кодирование слов уникальным номером является более быстрым, однако не позволяет фиксировать никакие отношения между словами, а также сходство слов.

Использование модели Word2Vec для вложения слов позволяет фиксировать тонкие взаимосвязи между словами, но требует для обучения большого объема данных и является длительным процессом. Ввиду отсутствия необходимости постоянной векторизации образовательных задач, а лишь при пополнении базы образовательных задач, была использована модель Word2Vec.

Входной слой нейронной сети имеет один нейрон, на который поступает вектор, содержащий данные о пользователе и задаче.

На скрытых слоях задействована функция активации «ReLu», на выходном слое – функция активации «Sigmoid», который позволяет усиливать слабые сигналы и не насыщаться от сильных сигналов [25].

Для обучения и тестирования нейронной сети был использован набор данных, состоящий из 16 862 векторов, содержащих данные о пользователях и задачах. Данные были разделены на обучающую и тестовую выборки в пропорции 80/20.

На исследуемом наборе данных наибольшую точность показала нейронная сеть, имеющая следующую топологию: два скрытых слоя (первый слой – 64 нейрона, второй слой – 128 нейронов), размер батча – 32, количество эпох – 20.

Результаты работы нейронной сети и точность предсказания представлены в *таблице 2*.

Для построения вектора $\hat{v}_i$ запуск сети производится столько раз, сколько значений содержится в базе образовательных задач. Таким образом, после всех итераций формируется вектор $\hat{v}_i$, на основе которого можно произвести выбор наиболее предпочтительных задач для пользователя.

3. Сравнение технологий искусственного интеллекта

В рамках исследования рассмотрены два подхода к формированию рекомендаций: нейронная сеть и машинное обучение на основе матричной факторизации. Оба подхода имеют как достоинства, так и недостатки.

В отличие от векторов, подаваемых на вход нейронной сети, матрица факторов состоит из уникальных векторов пользователей, которые содержат латентные признаки, не учитываемые нейронной сетью. Нейронная сеть, в свою очередь, опирается только на заданный известный вектор пользователя, состоящий из параметров, заложенных исследователем. Это позволяет судить о более персонализированном подходе при использовании матричной факторизации.

Нейронная сеть также требует обязательное наличие не только всех данных о пользователе, но и их корректности. При использовании машинного обучения на основе матричной факторизации можно выдавать рекомендации уже после нескольких оценок, не имея более никакой информации.

Недостатком матричной факторизации является холодный старт, когда пользователь еще не оценивал ни одной задачи, а ему требуется выдать рекомендации. Однако данная проблема решается «на лету» с помощью способа, описанного ранее.

Несмотря на то, что перед исследователями стояла задача ранжирования, необходимо было сравнить работу этих методов по трем критериям: предпочтения (бинарная величина), сложность (дискретная [1; 4]) и эффективность (дискретная [1; 5]). Результаты сравнения представлены в *таблице 3*.

Как видно из *таблицы 3*, на дискретных величинах точность предсказаний нейронной сети падает по сравнению с матричной факторизацией. Исходя из вышесказанного, исследователями было принято решение использовать в качестве ядра рекомендательного модуля машинное обучение на основе матричной факторизации.

Заключение

В статье рассмотрена проблема низкой адаптивности управления в условиях цифровой трансформации. В рамках исследования разработана модель интеллектуального проводника сотрудника, которая позволяет нивелировать данную проблему за счет

Таблица 2.

Результаты работы нейронной сети

	Предпочтение	Сложность	Эффективность
Среднеквадратичное отклонение	0,056	0,102	0,098
Точность, %	93,31	68,79	61,0

Таблица 3.

Точность предсказаний технологий искусственного интеллекта

	Предпочтение	Сложность	Эффективность
Нейронная сеть	93,3%	68,8	61%
Матричная факторизация	93,1%	76,4	62,4%

качественного развития компетентностного профиля сотрудника с учетом ценностных ориентиров самого сотрудника и организации, а также произведена ее реализация в виде программного компонента для системы управления предприятием.

Освоение новых компетенций приведет к расширению профиля сотрудника, возможности выполнения новых задач в интересующей области, тем самым получая качественный результат. Организация, в свою очередь, получит новые возможности для развития на рынке и быстрой адаптации к изменяющимся условиям. Таким образом, вкладывая в развитие сотрудников, организация получает T-shaped специалистов, обладающих проактивной позицией и способных к самоорганизации. Что непосредственно приводит к реализации потенциалов как сотрудника, так и организации.

Интеллектуальный проводник имеет модульную архитектуру, что обусловлено междисциплинарностью и новизной исследования, а также простотой внедрения полученных разработок в реальную управленческую практику. Такой подход позволил провести качественные исследования в двух проекциях: разработка комплексной автоматизированной диагностики soft skills профиля сотрудника и релевантного наполнения индивидуальной образовательной траектории образовательным материалом с помощью искусственного интеллекта, а затем объединить полученные результаты в единую систему для обучения и развития персонала.

Искусственный интеллект в предлагаемом подходе выражен с помощью машинного обучения с использованием матричной факторизации. Такая интеллектуальность позволяет качественно подобрать образовательный материал, который будет интересен сотруднику с его личностных позиций и максимально вовлечет его не только в процесс обучения и развития, но и в достижение организационных целей.

Предложенная структура базы данных позволяет собирать цифровые следы пользователей, описыва-

ющие биографические характеристики, навыки и уровень их владения на определенном этапе развития, историю выбора и выполнения образовательных задач, и в дальнейшем использовать собранные данные в качестве основы для работы модуля рекомендаций и формирования персонифицированных предложений.

Междисциплинарность и новизна исследования также обуславливают вариативность его развития. В рамках дальнейшего исследования планируется интеграция программного компонента в систему управления бизнес-процессами предприятия. Предполагается прогнозировать время отсутствия загрузки трудового ресурса и предлагать для его заполнения образовательные задачи. Это станет основной принципиально нового подхода к рассмотрению простоев не как потерь, а как возможности для развития и обучения сотрудников.

Разрабатываемый программный компонент может использоваться предприятиями не только на организационном уровне, но и через трансляцию в систему образования. С одной стороны, предприятия смогут сократить простои персонала, расширить личностно-компетентностный профиль сотрудников, увеличить рост производственного потенциала как самих сотрудников, так и предприятий на рынке. С другой стороны, образовательные учреждения смогут осуществлять подготовку высококвалифицированных кадров в рамках ориентации на ценности рынка и ценности студентов с формированием индивидуальных образовательных траекторий. Такая кооперация позволит сформировать экосистему образования в соответствии с требованиями инновационного развития экономики региона.■

Благодарности

Данное исследование выполнено при поддержке Сибирского государственного университета науки и технологий им. М.Ф. Решетнева в соответствии

с соглашением о выполнении научно-исследовательской работы от 01.12.2020 №04/20-ВГ по проекту «Исследование и разработка интеллектуально-го проводника для расширения компетентностного профиля сотрудника в рамках индивидуальных образовательных траекторий».

Литература

1. Результаты исследования «Международные тенденции в сфере управления персоналом – 2020» по России // Делойт Инсайт, 2020. [Электронный ресурс]: <https://www2.deloitte.com/content/dam/Deloitte/ru/Documents/human-capital/russian/hc-trends-2020-Russia.pdf> (дата обращения 22.11.2021).
2. Большие данные (Big Data) мировой рынок // TAdviser, 2020. [Электронный ресурс]: <https://www.tadviser.ru/a/129607> (дата обращения 22.11.2021).
3. Фатьянов А.А. Большие данные в цифровой экономике: ценность и правовые вызовы // Экономика. Право. Общество. 2018. № 4. С. 37–40.
4. Reinsel D., Gantz J., Rydning J. The digitization of the world – from edge to core // IDC White Paper, 2018. [Электронный ресурс]: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> (дата обращения 22.11.2021).
5. Ali N.M., Gadallah A.M., Hefny H.A., Novikov B.A. Online web navigation assistant // Вестник Удмуртского университета. Математика. Механика. Компьютерные науки. 2021. Т. 31. Вып. 1. С. 116–131. <https://doi.org/10.35634/vm210109>
6. Что такое цифровая экономика? Тренды, компетенции, измерение / Г.И. Абдрахманова и [др.]; науч. ред. Л. М. Гохберг; Нац. исслед. ун-т «Высшая школа экономики». М.: Издательский дом Высшей школы экономики, 2019 г.
7. Куликова М.Х., Куликова М.Х., Магоматов М.А. ИТ в образовании: сегодня, завтра и всегда // Материалы I студенческой научно-практической конференции «Информационные технологии в бизнесе и образовании», Грозный, 20 февраля 2020 года. Чеченский государственный университет им. А.А.Кадырова, 2020. С. 32–35. <https://doi.org/10.36684/21-2020-1-32-35>
8. Geissbauer R., Lübken E., Schrauf S., Pillsbury S. Digital champions. How industry leaders build integrated operations ecosystems to deliver end-to-end customer solutions // PwC Strategy & Global Digital Operations Study 2018 [Электронный ресурс]: <https://www.strategyand.pwc.com/gx/en/insights/industry4-0/global-digital-operations-study-digital-champions.pdf> (дата обращения 22.11.2021).
9. Harvard Business Review. High-performance sourcing and procurement driving value through collaboration // Harvard Business School Publishing, 2017. [Электронный ресурс]: <https://hbr.org/resources/pdfs/comm/scoutfrp/HighPerformanceSourcing.pdf> (дата обращения 22.11.2021).
10. Долганова О.И., Деева Е.А. Готовность компании к цифровым преобразованиям: проблемы и диагностика // Бизнес-информатика. 2019. Т. 13. № 2. С. 59–72. <http://doi.org/10.17323/1998-0663.2019.2.59.72>
11. Комаров В.А., Сарафанов А.В. Системы интернета вещей в процессе междисциплинарной подготовки кадров для цифровой экономики и их проектирование // Бизнес-информатика. 2021. Т. 15. № 2. С. 47–59. <http://doi.org/10.17323/2587-814X.2021.2.47.59>
12. Rivera A.C., Tapia-Leon M., Lujan-Mora S. Recommendation systems in education: A systematic mapping study // Rocha, Á., Guarda, T. (eds) Proceedings of the International Conference on Information Technology & Systems (ICITS 2018). Advances in Intelligent Systems and Computing. Vol. 721. P. 937–947. https://doi.org/10.1007/978-3-319-73450-7_89
13. Bustos López M., Alor-Hernández G., Sánchez-Cervantes J., Paredes-Valverde M., Salas-Zárate M.P. EduRecomSys: An educational resource recommender system based on collaborative filtering and emotion detection // Interacting with Computers. 2020. Vol. 32. No. 4. P. 407–432. <https://doi.org/10.1093/iwc/iwab001>
14. Urdaneta-Ponte M.C., Mendez-Zorrilla A., Oleagordia-Ruiz I. Recommendation systems for education: Systematic review // Electronics. 2021. Vol. 10. No. 14. Article ID 1611. <https://doi.org/10.3390/electronics10141611>
15. Bughin J., Hazan E., Lund S., Dahlström P., Wiesinger A., Subramaniam A. Skill shift: Automation and the future of the workforce // McKinsey & Company, 2018. [Электронный ресурс]: <https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce> (дата обращения 22.11.2021).
16. Рубцов В.Н. Матричная факторизация на основе глубокого обучения для коллаборативной фильтрации: дис. студента / НИУ ВШЭ, Москва, 2017. [Электронный ресурс]: <https://www.hse.ru/edu/vkr/206744221> (дата обращения 22.11.2021).
17. Strömquist Z. Matrix factorization in recommender systems: How sensitive are matrix factorization models to sparsity? // Uppsala University Publications, 2018. [Электронный ресурс]: <https://uu.diva-portal.org/smash/get/diva2:1214390/FULLTEXT01.pdf> (дата обращения 22.11.2021).
18. Мойсюк-Дранько П.А., Ревотюк М.П. Методы матричной факторизации для систем рекомендации // Информационные технологии и системы 2020 (ИТС 2020): материалы международной научной конференции. Минск: БГУИР, 2020. С. 193–194. [Электронный ресурс]: https://libdoc.bsuir.by/bitstream/123456789/41339/1/Moysuk_Dranko_Metody.pdf (дата обращения 22.11.2021).
19. Кузнецов И.А. Методы и алгоритмы машинного обучения для предобработки и классификации слабоструктурированных текстовых данных в научных рекомендательных системах: дис. ... канд. техн. наук / ФГАОУ ВО НИЯУ «МИФИ». Москва, 2019. [Электронный ресурс]: https://ds.mephi.ru/documents/90/Кузнецов_И_А_Текст_диссертации.pdf (дата обращения 22.11.2021).
20. Головкин В.А., Краснопрошин В.В. Нейросетевые технологии обработки данных // Минск: БГУ, 2017. [Электронный ресурс]: [https://ds.mephi.ru/documents https://elib.bsu.by/bitstream/123456789/193558/1/Golovko.pdf](https://ds.mephi.ru/documents/https://elib.bsu.by/bitstream/123456789/193558/1/Golovko.pdf) (дата обращения 22.11.2021).

21. Лисовский А.Л. Применение нейросетевых технологий для разработки систем управления // Стратегические решения и риск-менеджмент. 2020. Т. 11. № 4. С. 378–389. <https://doi.org/10.17747/2618-947X-923>
22. Bezdan T., Bacanin Džakula N. Convolutional neural network layers and architectures // Sinteza 2019 – International Scientific Conference on Information Technology and Data Related Research. Belgrade, Singidunum University, 2019. P. 445–451. <https://doi.org/10.15308/Sinteza-2019-445-451>
23. Постарнак Д.В. Критический анализ моделей нейронных сетей // Вестник Тюменского государственного университета. 2012. № 4. С. 162–167. [Электронный ресурс]: <https://elibrary.ru/item.asp?id=17758787> (дата обращения 22.11.2021).
24. Young T., Hazarika D., Poria S., Cambria E. Recent trends in deep learning based natural language processing // IEEE Computational Intelligence Magazine. 2018. Vol. 13. No. 3. P. 55–75. <https://doi.org/10.1109/MCI.2018.2840738>
25. Iqbal T., Qureshi S. The survey: Text generation models in deep learning // Journal of King Saud University – Computer and Information Sciences. 2020. <https://doi.org/10.1016/j.jksuci.2020.04.001>

Об авторах

Морозевич Екатерина Сергеевна

процессный архитектор, научно-исследовательское управление, Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, 660037, Сибирский федеральный округ, Красноярский край, г. Красноярск, проспект им. газеты Красноярский рабочий, д. 31;

E-mail: katyamorozevich@mail.ru

ORCID: 0000-0001-8796-3157

Коротких Владимир Сергеевич

системный администратор, Муниципальное бюджетное общеобразовательное учреждение средняя общеобразовательная школа №4, 663091, Сибирский федеральный округ, Красноярский край, г. Дивногорск, ул. Набережная, д. 9;

E-mail: vskor@bk.ru

ORCID: 0000-0001-9270-3376

Кузнецова Евгения Александровна

инженер, научно-исследовательское управление, Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, 660037, Сибирский федеральный округ, Красноярский край, г. Красноярск, проспект им. газеты Красноярский рабочий, д. 31;

E-mail: zhenya.kuz-1997@yandex.ru

ORCID: 0000-0002-3559-8394

The development of a model for a personalized learning path using machine learning methods

Ekaterina S. Morozevich^a

E-mail: katyamorozevich@mail.ru

Vladimir S. Korotkikh^b

E-mail: vskor@bk.ru

Yevgeniya A. Kuznetsova^a

E-mail: zhenya.kuz-1997@yandex.ru

^a Reshetnev Siberian State University of Science and Technology
Address: 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russia

^b Municipal budget general education institution secondary school No. 4
Address: 9, st. Naberezhnaya, Divnogorsk 663091, Russia

Abstract

Today, the economy is undergoing a digital transformation. Its key barriers are a lack of qualified personnel, competencies and knowledge, as well as internal resistance in organizations. It can be overcome through quality staff development and training. An urgent problem is to build a personalized learning path. Modern research is aimed at the implementation of recommendation systems in order to select relevant material. However, these recommendations are based on digital traces; the student's full personal profile, as well as organizational values are not considered. This study aims to create an intelligent guide that would accompany an employee throughout his life in the organization, involving him in the learning process according to a personalized path based on a complex personal profile and reactions to educational material, training soft and hard skills in accordance with the values of the organization and the employee. Methods of system analysis, system engineering, psychodiagnostic research (the DISC model, Rowe's "Decision-making Style" methodology, Honey and Mumford's method of determining activity styles, psychotype test), software design and artificial intelligence (matrix factorization and neural networks) were used in this study. The study was conducted on a unique database collected as part of its implementation and consisting of educational tasks for soft skills development, plus data on their implementation by users with different soft skills profiles. An intelligent guide model has been developed and implemented as a software component for an enterprise management system. The basis consists of psychodiagnostic modules, organizational management, training and recommendations. The intelligence of the system we developed allows you to qualitatively form a personalized learning path that will involve an employee not only in the learning and development process, but also in achieving organizational goals. The organization receives T-shaped specialists who have a proactive position and are capable of self-organization by investing in the development of employees. The results of this study can be used by enterprises not only at the organizational level, but also through broadcasting in the education system to form an education ecosystem in accordance with the requirements of innovative development of a given region's economy.

Keywords: training and development, supra-professional competencies, adaptability, recommendation system, machine learning, matrix factorization, neural network

Citation: Morozovich E.S., Korotkikh V.S., Kuznetsova Y.A. (2022) The development of a model for a personalized learning path using machine learning methods. *Business Informatics*, vol. 16, no. 2, pp. 21–35. DOI: 10.17323/2587-814X.2022.2.21.35

References

1. Deloitte Insights (2020) *Results of the study "International trends in human resource management – 2020" in Russia*. Available at: <https://www2.deloitte.com/content/dam/Deloitte/ru/Documents/human-capital/russian/hc-trends-2020-Russia.pdf> (accessed 22 November 2021) (in Russian).
2. TAdviser (2020) *Big Data global market*. Available at: <https://tadviser.com/a/e.php?id=129607> (accessed 22 November 2021).
3. Fatyanov A.A. (2018) Big Data in the digital economy: its value and legal challenges. *Economics. Law. Society*, no. 4, pp. 37–40 (in Russian).
4. Reinsel D., Gantz J., Rydning J. (2018) *The digitization of the world – from edge to core*. IDC White Paper. Available at: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> (accessed 22 November 2021).
5. Ali N.M., Gadallah A.M., Hefny H.A., Novikov B.A. (2021) Online web navigation assistant. *Vestnik Udmurtskogo Universiteta. Matematika. Mekhanika. Komp'yuternye Nauki*, vol. 31, no. 1, pp. 116–131. <https://doi.org/10.35634/vm210109>
6. Gokhberg L. (ed.) (2019) *What is the digital economy? Trends, competencies, measurement*. Moscow: HSE (in Russian).
7. Kulikova M.Kh., Kulikova M.Kh., Magomadov M.A. (2020) IT in education: today, tomorrow and always. Proceedings of the I Student Scientific and Practical Conference "Information technologies in business and education", Grozny, 20 February 2020, pp. 85–90 (in Russian). <https://doi.org/10.36684/21-2020-1-32-35>
8. Geissbauer R., Lübken E., Schrauf S., Pillsbury S. (2018) *Digital Champions. How industry leaders build integrated operations ecosystems to deliver end-to-end customer solutions*. PwC Strategy & Global Digital Operations Study. Available at: <https://www.strategyand.pwc.com/gx/en/insights/industry4-0/global-digital-operations-study-digital-champions.pdf> (accessed 22 November 2021).
9. Harvard Business Review (2017) *High-performance sourcing and procurement driving value through collaboration*. Harvard Business School Publishing. Available at: <https://hbr.org/resources/pdfs/comm/scoutfrp/HighPerformanceSourcing.pdf> (accessed 22 November 2021).
10. Dolganova O.I., Deeva E.A. (2019) Company readiness for digital transformations: problems and diagnosis. *Business Informatics*, vol. 13, no 2, pp. 59–72. <http://doi.org/10.17323/1998-0663.2019.2.59.72>
11. Komarov V.A., Sarafanov A.V. (2021) IoT systems in the process of multidisciplinary training of personnel for the digital economy and their design. *Business Informatics*, vol. 15, no 2, pp. 47–59. <http://doi.org/10.17323/2587-814X.2021.2.47.59>

12. Rivera A.C., Tapia-Leon M., Lujan-Mora S. (2018) Recommendation systems in education: A systematic mapping study. Proceedings of the *International Conference on Information Technology & Systems (ICITS 2018)*, Libertad City, Ecuador, 10–12 January 2018 (eds. A. Rocha, T. Guarda). Advances in Intelligent Systems and Computing, vol. 721, pp. 937–947. https://doi.org/10.1007/978-3-319-73450-7_89
13. Bustos López M., Alor-Hernández G., Sánchez-Cervantes J., Paredes-Valverde M., Salas-Zárate M.P. (2020) EduRecomSys: An educational resource recommender system based on collaborative filtering and emotion detection. *Interacting with Computers*, vol. 32, no. 4, pp. 407–432. <https://doi.org/10.1093/iwc/iwab001>
14. Urdaneta-Ponte M.C., Mendez-Zorrilla A., Oleagordia-Ruiz I. (2021) Recommendation systems for education: Systematic review. *Electronics*, vol. 10, no. 14, article ID 1611. <https://doi.org/10.3390/electronics10141611>
15. Bughin J., Hazan E., Lund S., Dahlström P., Wiesinger A., Subramaniam A. (2018) *Skill shift: automation and the future of the workforce*. McKinsey & Company. Available at: <https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce> (accessed 22 November 2021).
16. Rubtsov V.N. (2017) *Matrix factorization based on deep learning for collaborative filtering*. Student Theses. Moscow: HSE. Available at: <https://www.hse.ru/en/edu/vkr/206744221> (accessed 22 November 2021) (in Russian).
17. Strömquist Z. (2018) *Matrix factorization in recommender systems: How sensitive are matrix factorization models to sparsity*. Uppsala University Publications. Available at: <https://uu.diva-portal.org/smash/get/diva2:1214390/FULLTEXT01.pdf> (accessed 22 November 2021).
18. Mojsyuk-Dran'ko P.A., Revotyuk M.P. (2020) Matrix factorization methods for recommendation systems. Proceedings of the *International Conference Information Technologies and Systems 2020 (ITS 2020)*. Minsk: Belarusian State University of Informatics and Radioelectronics, pp. 193–194. Available at: https://libelloc.bsuir.by/bitstream/123456789/41339/1/Mojsyuk_Dranko_Metody.pdf (accessed 22 November 2021) (in Russian).
19. Kuznetsov I.A. (2019) *Machine learning methods and algorithms for preprocessing and classification of semi-structured text data in scientific recommendation systems*. Moscow: NRNU MEPhI. Available at: https://ds.mephi.ru/documents/90/Кузнецов_И_А_Текст_диссертации.pdf (accessed 22 November 2021) (in Russian).
20. Golovko V.A., Krasnoproshech V.V. (2017) *Neural network data processing technologies*. Minsk: Belarusian State University. Available at: <https://elib.bsu.by/bitstream/123456789/193558/1/Golovko.pdf> (accessed 22 November 2021) (in Russian).
21. Lisovsky A.L. (2020) Application of neural network technologies for management development of systems. *Strategic decisions and risk management*, vol. 11, no. 4, pp. 378–389 (in Russian). <https://doi.org/10.17747/2618-947X-923>
22. Bezdan T., Bacanin Džakula N. (2019) Convolutional neural network layers and architectures. Proceedings of the *Sinteza 2019: International Scientific Conference on Information Technology and Data Related Research, Belgrade, 20 April 2019* (ed. Milovan Stanišić), pp. 445–451. <https://doi.org/10.15308/Sinteza-2019-445-451>
23. Postarnak D.V. (2012) The critical analysis of models of neural networks. *Vestnik Tyumenskogo gosudarstvennogo universiteta*, no. 4, pp. 162–167. Available at: <https://elibrary.ru/item.asp?id=17758787&> (accessed 22 November 2021) (in Russian).
24. Young T., Hazarika D., Poria S., Cambria E. (2018) Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75. <https://doi.org/10.1109/MCI.2018.2840738>
25. Iqbal T., Qureshi S. (2020) The survey: Text generation models in deep learning. *Journal of King Saud University — Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2020.04.001>

About the authors

Ekaterina S. Morozovich

Process architect, Research Management Department, Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk 660037, Russia;

E-mail: katyamorozovich@mail.ru

ORCID: 0000-0001-8796-3157

Vladimir S. Korotkikh

System administrator, Municipal budget general education institution secondary school No. 4, 9, st. Naberezhnaya, Divnogorsk 663091, Russia;

E-mail: vskor@bk.ru

ORCID: 0000-0001-9270-3376

Yevgeniya A. Kuznetsova

Engineer, Research Management Department, Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk 660037, Russia;

E-mail: zhenya.kuz-1997@yandex.ru

ORCID: 0000-0002-3559-8394

Технологии коллективного интеллекта в управлении бизнес-процессами организации

Б.Б. Славин 

E-mail: bbslavin@gmail.com

Финансовый университет при Правительстве РФ
Адрес: Россия, 125993, г. Москва, Ленинградский проспект, д. 49

Аннотация

По мере цифровизации экономики возрастает креативная составляющая деятельности организации. Стандартные методы управления бизнес-процессами перестают работать ввиду увеличения неопределенности времени решения задач. В настоящее время отсутствуют эффективные технологии управления процессами интеллектуальной деятельности в организациях. В литературе уже давно обсуждается роль технологий коллективного интеллекта для управления знаниями в организациях, но конкретных предложений, как это организовать — до сих пор нет. Цель данной работы — показать, как технологии коллективного интеллекта могут решить проблемы управления бизнес-процессами интеллектуальной деятельности. Для демонстрации возможности технологий коллективного интеллекта в части увеличения производительности труда предложены модели распределения задач по компетенциям и синергии от совместной работы. В работе показано, что компетенции являются основной метрикой, которой можно измерять работу со знаниями в организации. Но также они должны учитываться при организации групповой деятельности. На простом модельном примере показано, что правильное распределение задач по компетенциям позволяет увеличить скорость решения задач группой в несколько раз. В реальных случаях необходимы расчеты с использованием вычислительных ресурсов. Также предложена модель, которая демонстрирует эффект увеличения скорости решения задач от совместной деятельности креативного сотрудника и аналитика. Показано, что управление бизнес-процессами должно быть дополнено маппингом модели компетенций и вариантов групповой работы на этапы бизнес-процессов. Это позволит управлять бизнес-процессами интеллектуальной деятельности.

Ключевые слова: коллективный интеллект, компетенции, системы управления знаниями, бизнес-процессы, интеллектуальная деятельность, синергия, брейнсторминг, групповая работа

Цитирование: Славин Б.Б. Технологии коллективного интеллекта в управлении бизнес-процессами организации // Бизнес-информатика. 2022. Т. 16. № 2. С. 36–48. DOI: 10.17323/2587-814X.2022.2.36.48

Введение

Организации, которые прошли или проходят этап цифровой трансформации, начинают конкурировать в основном на рынке инноваций. Для этого им необходимо создавать новые продукты и сервисы в существенно большем масштабе, чем раньше. В доцифровую эпоху подразделения, занимающиеся внедрением инноваций, были немногочисленны, но вполне справлялись со своей работой. Сегодня в создание инновационной продукции включается большая часть технологов, разработчиков и менеджеров. Именно поэтому популярными становятся технологии DevOps и BizOps, предполагающие создание непрерывного конвейера начиная с разработки новых продуктов (Dev – Development) до передачи их в эксплуатацию (Ops – Operations) и обратно, или даже начиная с идей, генерируемых бизнесом (Biz – Business). В этой связи доля сотрудников таких организаций, занимающихся креативной интеллектуальной деятельностью, возрастает кратно. Не случайно именно отрасли, находящиеся в центре «цифрового вихря», сегодня являются основным потребителем креативных кадров, и ощущают в них «голод».

Однако не только проблемы с рынком труда возникают при переходе к обществу знаний. Увеличение доли креативной деятельности требует радикального пересмотра подходов к управлению бизнес-процессами. Покажем это на простом примере. На *рис. 1* приведен простой бизнес-процесс, состоящий из четырех этапов. Для каждого этапа бизнес-процесса приведены графики вероятности

его исполнения во времени, где значение «1» означает исполнение. Условием того, что весь бизнес-процесс будет управляем, является то, что каждый этап должен заканчиваться в срок. Обычно на этап закладывают даже чуть больше времени, чтобы с высокой вероятностью он был исполнен. Более того, время исполнения нормируют, и по таким нормативам можно точно предсказать, когда бизнес-процесс будет исполнен. Например, в авто-сервисах так рассчитывают и трудоемкость, и оценивают время исполнения заказа. На выполнении этого, достаточно очевидного условия, построена вся система управления бизнес-процессами в организации.

Однако в случае, если бизнес-процесс касается креативной деятельности, время окончания этапов будет непредсказуемым. Этап может закончиться намного раньше отведенного срока, а может закончиться и много позже. На *рис. 2* показаны плотности вероятностей завершения этапов в случае креативной деятельности.

Фактически такой бизнес-процесс будет неуправляем, поскольку вероятности того, что этапы не завершатся в срок, будут перемножаться, и время окончания всего процесса станет непредсказуемым. Можно, конечно, существенно увеличить время на каждый из этапов, но тогда эффективность бизнес-процесса будет чрезмерно низкой, сотрудники будут большую часть времени простаивать. Когда подразделения, занимающиеся инновациями, были малочисленными, их просто выводили за рамки системы управления бизнес-процессами и ставили цели с неопределенными сроками. В

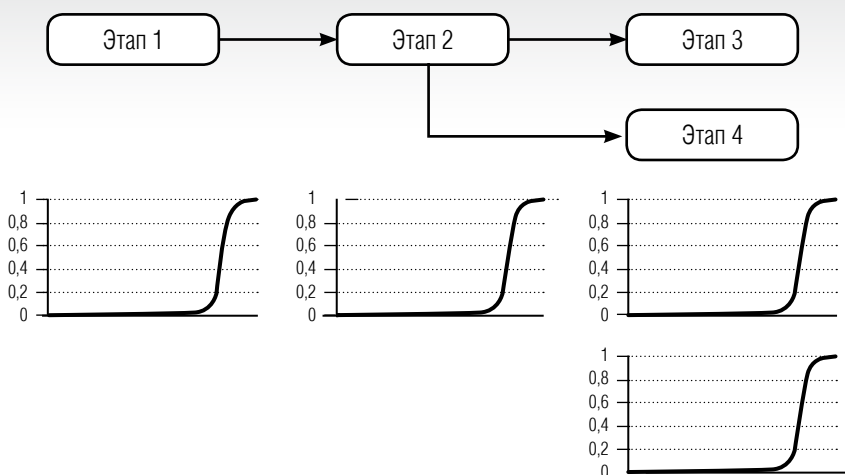


Рис. 1. Схема бизнес-процесса и вероятности исполнения этапов в срок в обычной деятельности.

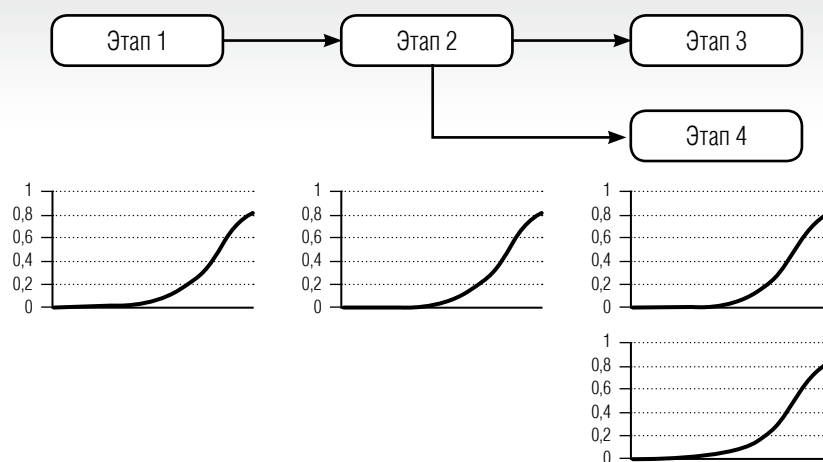


Рис. 2. Схема бизнес-процесса и вероятности исполнения этапов в креативной деятельности.

случае же, когда инновационные подразделения встроены в общие бизнес-процессы, как это сегодня происходит с компаниями, оказавшимися в центре цифровой трансформации, системы управления бизнес-процессами работать не будут. Решить эту проблему и должны помочь технологии коллективного интеллекта и компетентностный подход к управлению процессами.

1. От управления знаниями к управлению компетенциями

В работе [1] показано, что аналогом товара в экономике знаний будет не само знание, а способности людей оперировать со знанием, т.е. их компетенции [2]. В отличие от знаний стоимость компетенций пропорциональна себестоимости, т.е. расходам, необходимым для обучения человека. Причем, если со стороны затрат на преподавателей определенная тиражируемость все-равно присутствует (а значит такие затраты могут быть снижены), то временные ресурсы самого обучаемого не тиражируются, и в будущем будут, видимо, определять существенную долю себестоимости обучения. Более того, инвестиции в компетенции носят вполне рыночный характер, и дают такую же отдачу, как и инвестиции в производство товаров и услуг. Так, согласно оценкам экономистов¹ один год обучения в среднем повышает зарплату на 10%. Это в свою очередь означает, что инвестировать в компетенции человека

по мере перехода к экономике знаний будет становиться все выгоднее.

В силу особой роли компетенций в инновационной экономике становятся актуальными исследования возможности построения систем управления компетенциями, организации творческой деятельности, измерения и повышения стоимости человеческого капитала [3]. В настоящее время широкое распространение пока получили лишь информационные системы учета кадров, однако на рынке растет спрос уже и на такие субъектно-ориентированные информационные системы, как Управление талантами, Планирование развития карьеры, Управление компетенциями и т.п. Управление компетенциями как понятие впервые возникло в сфере образования еще 40 лет назад [4], но широкое распространение в бизнесе оно получило позже [5]. Д. Дарнтон сформулировал основные компоненты процесса управления компетенциями, к которым отнес: взаимосвязь компетенций сотрудников со стратегией и задачами предприятия, условия развития компетенций, их классификация, планирование развития и контроль компетенций [6].

Управление компетенциями является частью общей системы управления знаниями (СУЗ или Knowledge Management – КМ) [7], если под знаниями иметь в виду явные и неявные знания. Впервые выделить «молчаливые» или неявные знания (tacit knowledge) еще в 1958 году предложил Михаэль По-

¹ Такую оценку, в частности, озвучивал на лекции нобелевский лауреат по экономике Кристофер Антониу Писсаридес: <http://www.fa.ru/News/2016-03-28-open-lecture.aspx>

лани, отнеся к ним знания, который человек имеет сверх того, о чем может сказать² [8, с. 23]. Применительно к знаниям в организации термин «неявные знания» использовали Нонака и Такеучи в своей книге «Компания — создатель знания» [9], называя его чаще «неформализованным» знанием: «Неформализованное знание (или «неявное») — личное и зависящее от ситуации и поэтому с трудом поддающееся формализации и распространению» [9, с. 84]. Компетенции включают в себя неявные знания, и, в отличие от знаний, измеримы. Поскольку говорить об управлении чем-либо можно только, если объект управления измерим, понятие управление компетенциями имеет смысл в отличие от управления знаниями. Впрочем, можно по-прежнему использовать понятие управление знаниями, понимая под ним именно управление компетенциями людей по созданию и использованию знаний.

2. Технологии коллективного интеллекта

Особую роль компетентностный подход играет в технологиях коллективного интеллекта. Понятие коллективного интеллекта имеет широкое толкование, и в той или иной форме (мудрость толпы, коллективный разум и т.п.) его можно найти в научной литературе, датируемой многими сотнями лет назад [10]. Непосредственно сам термин коллективный интеллект (Collective Intelligence) по всей видимости впервые ввел Дэвид Векслер, создатель так называемых векслеровских шкал оценки интеллекта. Векслер утверждал, что коллективный интеллект возникает лишь тогда, когда члены группы используют общие интеллектуальные ресурсы в своей деятельности [11, с. 906]. Возможностям коллективной творческой деятельности в конце прошлого века было посвящено много работ. Отметим в качестве примера книгу супругов Фишер «Распределенные умы: достижение высокой производительности через коллективный интеллект рабочих групп» [12], в которой обсуждаются подходы к коллективизации знаний в организациях. И все-таки проблемы коллективного интеллекта наибольшее внимание получили лишь с развитием сети Интернет [13]. Именно эпоха Интернета ознаменовалась бурным интересом к проблемам коллективного интеллекта.

Канадский публицист Пьер Леви еще в конце прошлого века опубликовал книгу с названием «Коллективный интеллект: развивающийся мир человечества в киберпространстве» [14], в которой призвал создавать общество, где кибертехнологии оказывают гуманизирующее влияние и способствуют появлению «коллективного интеллекта». Френсис Хейлиген (автор книги «Мировой Суперорганизм: эволюционно-кибернетическая модель формирующегося сетевого общества» [15]) писал, что очень важно научиться использовать сетевые коммуникации для повышения «коллективного разума» таким образом, чтобы групповой интеллект превышал сумму интеллектов членов группы [16, с. 92]. Особое место исследователи коллективного интеллекта уделяют проекту Википедии. Так, например, американские ученые из Университета Карнеги-Меллона выявили взаимосвязь сложности контента Википедии и компетентности редакторов этого проекта [17]. М. Хорост, который вообще все сетевые ресурсы рассматривает как глобальный мозг, имеющий память, узлы и синапсы, писал о Википедии как о коллективной базе знаний: «Википедия отличается своей “интеллектуальностью”, которую она развивает благодаря коллективному сознанию и редактированию контента. И вновь мы видим общую сумму множества индивидуальных суждений о том, что важно, а что нет... Возникающее при этом знание отличается от PageRank, однако оба ресурса замечательно гармонично дополняют друг друга. В сочетании они как бы формируют зарождающиеся лобные доли, гиппокамп и своего рода долговременную память Сети» [18, с. 251].

О том, что технологии коллективного интеллекта являются одним из инструментов управления знаниями в сети Интернет, писали К. Зетцу и Я. Кийоки [19]. В работе [20] вообще было предложено все социальные сети рассматривать как инфраструктуру знаний (knoware) коллективного интеллекта. Авторы вводят такое понятие как «суперсеть знаний», которая включает в себя медиасети, сети пользователей и сети знаний. Большое число исследований по теме коллективного интеллекта проводится в Массачусетском Технологическом Институте США группой под руководством Т. Малоуна [21]. Ученые этой группы изучают различные способы применения технологий коллективного интеллекта, как для организации глобальных сетевых проектов, так и для повышения

² «know more than we can tell»

эффективности бизнеса. Использованию технологий коллективного интеллекта в качестве особых информационных систем предприятий посвящены такие работы как [22, 23]. В работе [24] технологии коллективного интеллекта рассматривались как технологии повышения эффективности деятельности человека по аналогии с использованием инструментов бизнес-аналитики.

В работе [25] Малоуном с коллегами была предложена классификация, которую они назвали «геномом» коллективного интеллекта. Однако, по сути, эта классификация не выявила особенностей технологий коллективного интеллекта, а просто позволила ранжировать все глобальные сетевые проекты. Многие исследователи вслед за Малоуном также не делают различий между краудсорсинговыми технологиями и технологиями коллективного интеллекта [26, 27]. Однако имеется и другая точка зрения. Так Т. Грубер, описывая краудсорсинговые технологии и социальные сети, пишет, что они могут претендовать лишь на то, чтобы называться «собранием интеллектов», но не являются единым коллективным интеллектом, поскольку не поддерживают группового мышления [28, с. 4].

В настоящее время нет консенсуса, что должны включать в себя технологии коллективного интеллекта. В данном исследовании поддерживается и развивается точка зрения, что технологии коллективного интеллекта представляют собой инструменты и системы, «которые объединяют в группы необходимое число людей, имеющих собственные индивидуальные цели, но организованных таким образом, что общий интеллект и эффективность группы возрастает»³ [29, с. 219]. В рамках такого подхода

можно дать определение технологиям коллективного интеллекта как особой формы «информационных технологий, способствующих коллективному решению интеллектуальных и творческих задач с использованием сетевых коммуникаций» [10].

3. Роль компетенций в технологиях коллективного интеллекта

Рассмотрим несколько примеров, которые показывают эффективность технологий коллективного интеллекта при организации творческой деятельности. Первое, что позволяют такие технологии – за счет правильного учета компетенций при распределении задач, которые решает группа, – существенно ускорить их решение. Предположим, что мы имеем группу сотрудников из четырех человек, имеющих различные компетенции (пусть их будет 6), которые обозначают вероятность решения задачи на данную компетенцию (так обычно измеряют индекс интеллектуальности человека, IQ). Пусть для простоты эти вероятности равны либо 0 (нет компетенции), либо 1 (компетенция позволяет решить задачу с вероятностью 1). Тогда спектр компетенций такой группы можно описать прямоугольной матрицей, показанной на *рис. 3а*. Первый сотрудник однозначно решает задачи первых двух и четвертой компетенции, второй – с третьей по пятую и т.д.

Пусть также эта группа решает задачи, собранные в четыре набора по шесть компетенций каждый. И предположим, что эти задачи распределены равномерно между участниками (*рис. 3б*). При таких допущениях задачи будут решены лишь в том случае, если у решающих будут аналогичные компетенции – результаты работы показаны в

a)	<table> <tr><td>1</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>1</td></tr> </table>	1	0	1	0	1	0	1	1	0	1	0	0	1	1	0	0	0	1	1	0	0	0	1	1	b)	<table> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> </table>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0	1	0																																																
1	0	1	1																																																
0	1	0	0																																																
1	1	0	0																																																
0	1	1	0																																																
0	0	1	1																																																
1	1	1	1																																																
1	1	1	1																																																
1	1	1	1																																																
1	1	1	1																																																
1	1	1	1																																																
1	1	1	1																																																
			3 3 4 2																																																

Рис. 3. Матрица компетенций (а), групповая матрица распределения (б).

³ “...which hosts an adequately large group of people, who act for their individual goals, but whose group actions aim and may result – through technology facilitation – in a higher-level intelligence and benefit of the community.”

строке под матрицей на *рис. 3б*. Видно, что в среднем участники такой группы решат по три задачи.

Однако, если мы распределим задачи по участникам другим образом (смотри групповую матрицу распределения на *рис. 4б*), сохранив нагрузку на каждого участника — по шесть задач, можно добиться того, что каждый из участников решит все шесть задач. То есть производительность группы будет в два раза выше, причем просто за счет правильного подбора компетенций. Именно нахождение групповой (или коллаборационной) матрицы распределения необходимо при организации работы в рамках технологии коллективного интеллекта.

В случае, когда в группе участвуют большее число сотрудников, а вероятности решения ими задач отличается от 0 или 1, необходимо проводить численные вычисления, при этом разница в производительности может быть еще выше. Алгоритм организации групповой работы, основанный на расчете коллаборационной матрицы, является аналогом разделения труда, но для интеллектуальной деятельности. Понятно, что точно измерить вероятность решения тех или иных задач сложно, но оценивать скорости их решения тем или иным специалистом можно. Правильное разделение общей задачи на подзадачи и правильный подбор персонала позволяет эффективно использовать интеллектуальные ресурсы. В текущей практике проведения сложных научно-исследовательских работ такое распределение руководители пока осуществляют, полагаясь лишь на интуицию.

4. Синергия в коллективном интеллекте

Разделение людей по техническим компетенциям — не единственное условие эффективности коллективной интеллектуальной деятельности. Важным является правильная организация совместной

работы над одной задачей, или синергия, которая учитывает креативные и аналитические компетенции. Разделение экспертов на аналитиков и «генераторов идей» является важной составляющей метода брейнсторминга, мозгового штурма. Альтшуллер Г.С. в своей книге «Алгоритм изобретения» так описывает этот метод, предложенный американским журналистом Алексом Осборном в конце 30-х годов прошлого века: «Есть люди, которые по складу ума хорошо «генерируют» идеи, но плохо справляются с их анализом. И наоборот: некоторые люди больше склонны к критическому анализу идей, чем к их «генерации». Осборн решил разделить эти процессы. Пусть одна группа, получив задачу, только выдвигает идеи, хотя бы и самые фантастические. Другая группа пусть только анализирует выдвинутые идеи» [30, с. 10]. Несмотря на то, что «генерацию» идей и их анализ можно считать разными компетенциями, учитывать их при организации интеллектуальной деятельности требуется особо, поскольку одну задачу невозможно разделить на фазу выработки идей и фазу их конкретизации, необходима совместная работа.

Чтобы понять, каким образом при взаимодействии креативного участника («генератора» идей) и аналитика достигается эффект синергии, можно воспользоваться модельными функциями плотности вероятности решения задачи. Если предположить, что время решения задачи у обоих специалистов одинаково (и равно 10), вероятности решения ими задачи будут выглядеть примерно так, как показано на *рис. 5*, где F_i — вероятность решения задачи «генератором» идей, а F_a — аналитиком. Эксперт аналитик маловероятно решит задачу раньше времени $t = 6$, и почти наверняка ее решит ко времени $t = 14$, в то время как эксперт, обладающий креативными компетенциями, точно решит задачу лишь ко времени $t = 20$, но вполне вероятно может решить задачу и при небольших значениях t .

а)

1	0	1	0
1	0	1	1
0	1	0	0
1	1	0	0
0	1	1	0
0	0	1	1

б)

3	0	1	0
0	0	1	3
0	4	0	0
3	1	0	0
0	1	3	0
0	0	1	3
6	6	6	6

Рис. 4. Матрица компетенций (а), неоднородная групповая матрица распределения (б).

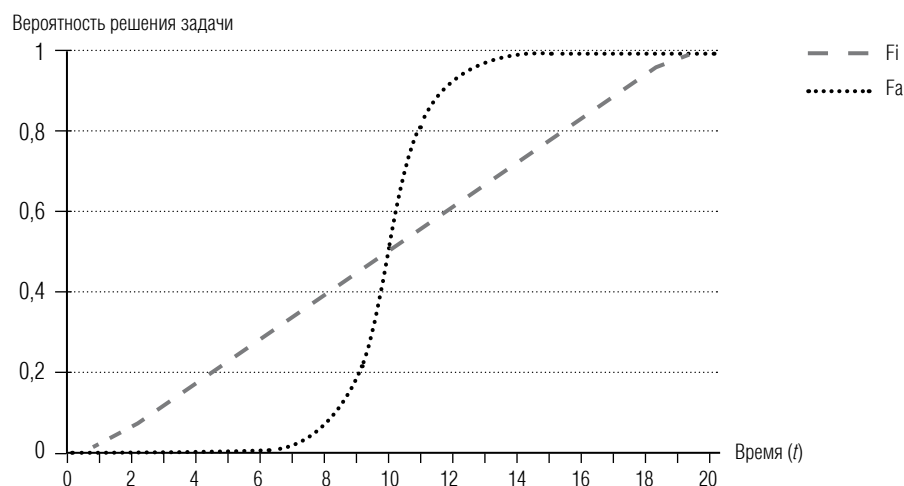


Рис. 5. Вероятности решения задачи «генератором» идей и аналитиком.

Функцию распределения можно интерпретировать не только как плотность вероятности решения задачи, но и как процент выполнения задачи. Конечно, отдельная задача может быть либо полностью решена, либо она решена не будет. Но в отдельных случаях частичное решение задачи имеет реальный смысл — например, при выполнении какого-нибудь исследования, когда один ученый может провести лишь часть исследования, а закончить его может другой. Такая интерпретация функции распределения позволяет промоделировать ситуацию, когда над задачей работают сразу два специалиста, причем один является аналитиком, а второй — «генератором» идей. При передаче задачи друг другу ее объем (или вероятность) не должен меняться. Математически это означает, что функция вероятности совместного решения задачи (в случае передачи задачи от одного к другому) должна быть непрерывной.

Непрерывность функции вероятности совместного решения задачи достаточно очевидна, но только это свойство не позволяет определить момент, когда можно передавать задачу другому участнику. Можно сформулировать гипотезу о том, что при передаче задачи от одного участника к другому необходимо равенство не только объема решенной задачи, но и динамики ее решения. Доказательств этой гипотезы пока нет, но есть эмпирические факты, частично подтверждающие ее верность. Так в работе [31] изучалась коллаборация студентов, которая осуществлялась дистанционно с использованием сетевых инструментов (блогов, wiki и др.), и было показано, что студенты с большим успехом участвуют в со-

вместной работе, когда стиль решения задач (навыки, знания, цели и планы) их напарников им ближе и понятнее. Данная гипотеза означает, что коллаборационная функция вероятности совместного решения должна быть не только непрерывной, но и гладкой (непрерывной в первой производной или непрерывной для функции плотности вероятности).

«Генератор» идей рассматривает возможные решения задачи быстрее аналитика, поскольку он их не проверяет сразу. В определенный момент времени (обозначим его τ_i) объем решенной им задачи и скорость решения могут оказаться равными тому, как это бы решал аналитик, но существенно позже, в момент времени τ_a . Если в этот момент передать задачу от «генератора» идей аналитику, общее решение задачи сократится на величину $(\tau_a - \tau_i)$. В некотором смысле такая передача решения от «генератора» идеи к аналитику моделирует «озарение» или «инсайт» при групповом решении задачи. Благодаря такому «озарению» распределение вероятности смещается по оси времени влево — изображено линией «Collab» на рис. 6. При выбранных параметрах распределений значение времен будет таким: $\tau_a \sim 7,8$, а $\tau_i \sim 1,8$, и, следовательно, время решения задачи может быть уменьшено на величину равную 6, т.е. среднее время решение задачи сокращается более чем вдвое ($t = 10$).

Заметим, что креативный специалист, «генерирующий» идеи, участвует в решении задачи меньше времени, чем аналитик (в приведенном случае более чем в 4 раза). Это говорит о том, что для эффективного использования коллаборации в творческой деятель-

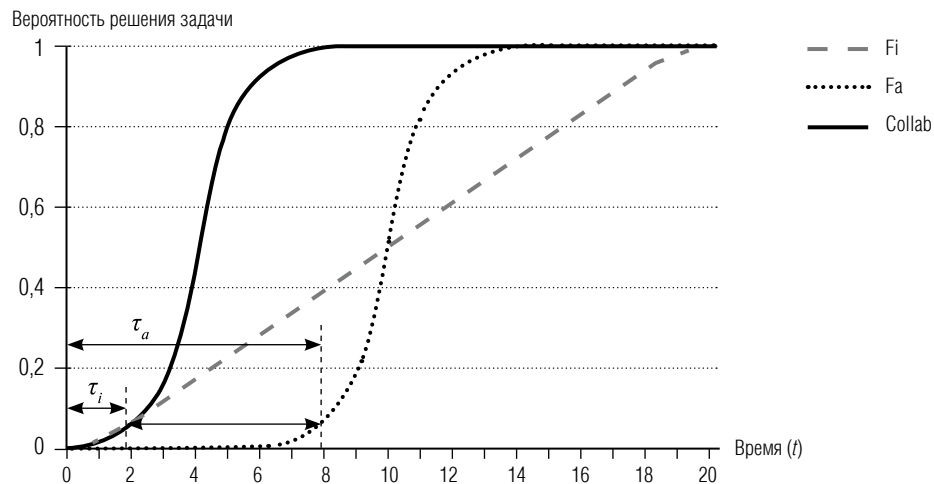


Рис. 6. Решение задачи в результате коллаборации.

ности целесообразно одного «генератора» идей использовать для работы с несколькими аналитиками. Практика управленческой деятельности, в которой роль креативного специалиста часто играет руководитель подразделения, подтверждает это — сотрудники, которых всегда несколько, доводят идеи, высказанные одним руководителем до завершённой формы.

Таким образом, технологии коллективного интеллекта помимо технических компетенций групповой деятельности должны учитывать способности быть аналитиком или «генератором» идей, причем такие способности могут у человека меняться местами в зависимости от области знаний. Человек, занимающийся интеллектуальной деятельностью в одиночку, вынужден играть обе роли, часто откладывая исследование, чтобы потом взглянуть на него с другой стороны. Нетрудно понять, что такой подход будет всегда проигрывать коллективной работе, если, конечно, при коллаборации учитываются способности и компетенции человека. При организации научной или исследовательской деятельности очень важно учитывать, как участник решает задачи — как «генератор» идей, или как аналитик, чтобы более эффективно встроить его в командную работу.

Можно показать (используя аналогичный вероятностный подход), что синергия проявляется не только на начальной стадии решения задачи, но и при его завершении. Например, при подготовке отчетов о научных исследованиях часто один и тот же текст читают разные участники исследования, рецензируя и внося свои исправления. Это происходит не потому, что компетенции пишущего текст

автора меньше компетенций рецензентов — взгляд со стороны позволяет лучше увидеть недостатки. Кроме того, трудозатраты на экспертизу, как правило, на порядок меньше трудозатрат на подготовку первоначального документа, что позволяет привлечь к работе сразу несколько человек, имеющих разные компетенции и опыт. Разделение участников групповой работы на тех, кто создает документ, и тех, кто его рецензирует, положено в основу метода эволюционного согласования [32], и может быть использовано в деятельности различных организаций, требующих интеллектуальной работы, включая поиск решений [33].

5. Учет компетенций и коллаборации в бизнес-процессах

Описанные выше технологии позволяют решить проблему бизнес-процессов, в которых основную роль играет творческая интеллектуальная деятельность. Необходимо, с одной стороны, учитывать компетенции участников процесса, а с другой — организовать совместную работу над решениями задач. Фактически речь идет о маппинге модели (классификатора) компетенций и вариантов групповой деятельности на этапы бизнес-процесса. На рис. 7 показан пример такого маппинга. Технологии коллективного интеллекта фактически являются связующим звеном между информационными системами, автоматизирующими бизнес-процессы организации, и сотрудниками организации, которые не только обладают определенными компетенциями, но и решают групповые задачи.

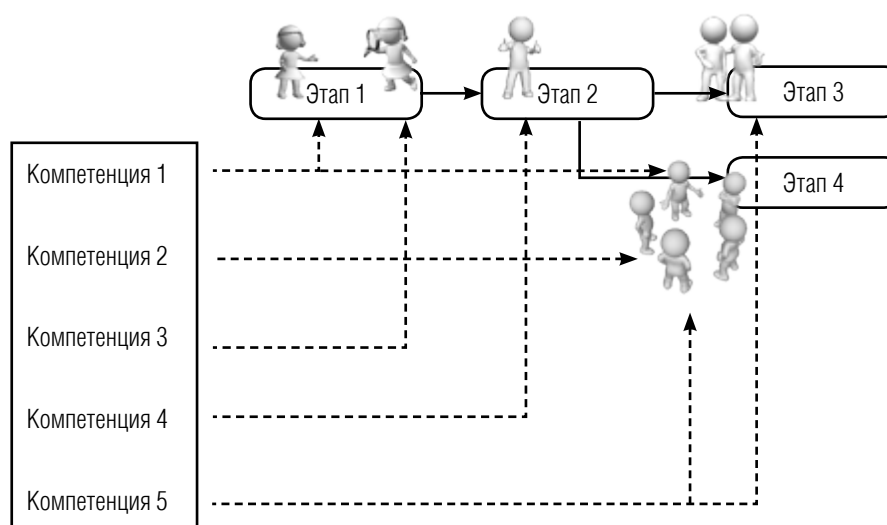


Рис. 7. Маппинг модели компетенций на бизнес-процесс с учетом групповой деятельности.

В обычной деятельности, когда для участия в бизнес-процессе необходимо небольшое число стандартных компетенций, сотрудник организации подбирается таким образом, чтобы его компетенции соответствовали бизнес-процессу, возможно после соответствующего обучения. Когда речь заходит об интеллектуальной деятельности, число необходимых компетенций существенно возрастает, причем они касаются не только профессиональной области, но и организационных, творческих способностей, которым нелегко (а в некоторых случаях и невозможно) обучить. При этом групповая работа становится важным элементом, и не учитывать ее в управлении бизнес-процессами нельзя.

Компетенции человека обязательно должны оцениваться в процессе реальной деятельности, причем оценка должна служить не для наказания или поощрения сотрудников, а для более точного распределения их в творческой работе и для обучения. По мере того, как реализуются бизнес-процессы, увеличивается и качество информации о компетенциях сотрудников. Взаимосвязь компетенций и бизнес-процессов, по сути дела, является взаимосвязью между неявным и явным знанием в организации. Именно в такой связи следует говорить об эффективном управлении знанием. Задача технологий коллективного интеллекта как раз и заключается в том, чтобы максимально использовать человеческий интеллектуальный капитал при работе

с явным знанием или с организационным капиталом компании.

Технологии коллективного интеллекта сегодня находят все больше и больше применений в различных сферах. Так, например, в работе [34] исследуется возможность использования технологий коллективного интеллекта в онлайн сообществах МООК⁴. Авторы показали, что в образовательных сообществах по мере их становления снижается роль фасилитаторов (старшекурсников, преподавателей) и возрастает роль взаимодействия со сверстниками. Работа [35] посвящена исследованию возможности использования технологий коллективного интеллекта в предиктивном анализе. Большое число исследований посвящено возможностям технологий коллективного интеллекта в организации научной [36] и экспертной [37] деятельности.

Роль компетенций и необходимость развития технологий управления ими в новой экономике еще недостаточно понята. Частично это связано с тем, что отсутствует теоретическая основа технологий управления интеллектуальной деятельностью человека. На бизнес-форумах сегодня все чаще обсуждается необходимость развития человеческого капитала в связи с тем, что инновации становятся одной из основных деятельности компании, но пока научные исследования касаются предмета в целом, а не конкретных технологий.

⁴ Массовые открытые онлайн курсы

Эдвард Деминг, который пропагандировал сотрудничество для эффективной организации корпоративной работы, приводил в качестве эталонного примера оркестр: «Музыканты не играют соло, а внимательно прислушиваются друг к другу. Они собираются, чтобы поддержать друг друга... Так, каждый из 140 музыкантов Королевской филармонии в Лондоне поддерживает остальных 139 коллег. Звучание оркестра оценивают слушатели; при этом роль играет не известность исполнителей, а то, что у них получается в результате» [38, с. 87]. В отличие от музыкантов, слушающих звук коллег, интеграции профессионалов в области интеллектуальной деятельности способствуют электронные коммуникации, что позволяет говорить о создании единого сетевого разума.

Заключение

Таким образом, можно сказать, что именно технологии коллективного интеллекта, опирающиеся на компетентностный подход и учитывающий синергию от групповой работы, позволят управлять бизнес-процессами в условиях творческой деятельности, которая все больше и больше востребована организациями. В южнокорейских школах уже давно обучают детей, рассаживая их вокруг

круглых столов. Это делают намеренно, чтобы приучить школьников к групповой работе с детства. По мере того, как деятельность компаний будет носить все более креативную деятельность именно групповая работа с учетом конкретных компетенций и организационных особенностей сотрудников сможет снизить неопределенность в завершении задач. Управление бизнес-процессами на основе технологий коллективного интеллекта потребует внедрение компетентностного подхода, причем измерение компетенций необходимо будет осуществлять непрерывно в рамках обратной связи. Измерение компетенций позволит подстраивать систему управления бизнес-процессами под меняющиеся условия, менять или доучивать сотрудников. Организации, которые смогут первыми наладить такие системы управления бизнес-процессами, получат конкурентные преимущества в области инновационного развития. ■

Благодарности

Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финансового университета при Правительстве РФ.

Литература

1. Славин Б.Б. От экономики товара к экономике человека // Экономика и управление: проблемы, решения. 2017. Т. 7. № 8. С. 79–84.
2. Male S.A., Bush M.B., Chapman E.S. Perceptions of competency deficiencies in engineering graduates // Australasian Journal of Engineering Education. 2010. Vol. 16. No. 1. P. 55–68. <https://doi.org/10.1080/22054952.2010.11464039>
3. Лосева О.В. Методика оценки состояния и анализа динамики развития человеческого интеллектуального капитала в организации // Известия Пензенского государственного педагогического университета. 2009. Т. 16. № 12. С. 75–71.
4. Spady W.G. The concept and implications of competency-based education // Education Leadership. 1978. P. 16–22.
5. Homer M. Skills and competency management // Industrial and Commercial Training. 2001. Vol. 33. No. 2. P. 59–62. <https://doi.org/10.1108/00197850110385624>
6. Darnton G. Modelling requirements and architecting large-scale on-line competence-based learning systems // Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT 2002), Kazan, Russia. 2002. P. 170–174.
7. Алтухова Н.Ф., Данилина О.М. К вопросу о компетенциях в контексте управления корпоративными знаниями // Вестник Университета. 2008. Т. 21. № 11. С. 9–16.
8. Polanyi M. The tacit dimension. Chicago: University of Chicago Press, 2009. <https://doi.org/10.1007/s11016-010-9328-0>
9. Нонака И., Takeuchi Х. Компания – создатель знания. М: Олимп-Бизнес, 2011.
10. Славин Б.Б. Технологии коллективного интеллекта // Проблемы управления. 2016. № 5. С. 2–9.
11. Wechsler D. Concept of collective intelligence // American Psychologist. 1971. Vol. 26. P. 904–907.
12. Fisher K., Fisher M.D. The distributed mind: Achieving high performance through the collective intelligence of knowledge work teams. New York: Amacom, 1997.
13. Weiss A. The power of collective intelligence // netWorker. 2005. Vol. 9. P. 16–23. <https://doi.org/10.1145/1086762.1086763>
14. Levy P. Collective intelligence: Mankind's emerging world in cyberspace. Cambridge: Perseus Books, 1997.
15. Heylighen F. The global superorganism: An evolutionary-cybernetic model of the emerging network society // Social Evolution & History. 2007. Vol. 5. No. 1. P. 57–117.

16. Хейлиген Ф. Концепция глобального мозга // Рождение коллективного разума: О новых законах сетевого социума и сетевой экономики и об их влиянии на поведение человека. М.: Ленанд, 2014.
17. Kittur A., Lee B., Kraut R.E. Coordination in collective intelligence: The role of team structure and task interdependence // Proceedings of the 27th International Conference on Human Factors in Computing Systems, Boston, MA, USA, April 4–9, 2009. P. 1495–1504. <https://doi.org/10.1145/1518701.1518928>
18. Хорост М. Всемирный разум. М.: Эксмо, 2011.
19. Zetssu K., Kiyoki Y. Towards knowledge management based on harnessing collective intelligence on the Web // Proceedings of the 15th International Conference on Managing Knowledge in a World of Networks (EKAW'06), Podebrady, Czech Republic, October 2006. P. 350–357. https://doi.org/10.1007/11891451_31
20. Luo S., Xia H., Yoshida T., Wang Z. Toward collective intelligence of online communities: a primitive conceptual model // Journal of Systems Science and Systems Engineering. 2008. Vol. 18. No. 2. P. 203–221. <https://doi.org/10.1007/s11518-009-5095-0>
21. Woolley A., Aggarwa I., Malone T. Collective intelligence and group performance // Current Directions in Psychological Science. 2015. Vol. 24. No. 6. P. 420–424. <https://doi.org/10.1177/0963721415599543>
22. Leimeister J.M. Collective intelligence // Business & Information Systems Engineering. 2010. No. 4. P. 245–248. <https://doi.org/10.1007/s12599-010-0114-8>
23. Gregg D.G. Designing for collective intelligence // Communications of ACM. 2010. Vol. 53. No. 4. P. 134–138. <https://doi.org/10.1145/1721654.1721691>
24. Alag S. Collective intelligence in action. Greenwich: Manning Publications Co., 2008.
25. Malone T.W., Laubacher R., Dellarocas C. Harnessing crowds: Mapping the genome of collective intelligence // MIT Sloan Research Paper No. 4732-09. 2009. <https://doi.org/10.2139/ssrn.1381502>
26. Buecheler T., Sieg J., Fuchslin R., Pfeifer R. Crowdsourcing, open innovation and collective intelligence in the scientific method: a research agenda and operational framework // Proceedings of the 12th International Conference on the Synthesis and Simulation of Living Systems (Artificial Life XII), Odense, Denmark, 19–23 August 2010. P. 679–686. <https://doi.org/10.21256/zhaw-4094>
27. Bothos E., Apostolou D., Mentzas G. Collective intelligence for idea management with Internet-based information aggregation markets // Internet Research. 2009. Vol. 19. No. 1. P. 26–41. <https://doi.org/10.1108/10662240910927803>
28. Gruber T. Collective knowledge systems: Where the social web meets the semantic web // Web Semantics: Science, Services and Agents on the World Wide Web. 2008. Vol. 6. P. 4–13. <https://doi.org/10.1016/j.websem.2007.11.011>
29. Lykourantzou I., Vergados D., Kapetanios E., Loumos V. Collective intelligence systems: Classification // Journal of Emerging Technologies in Web Intelligence. 2011. Vol. 3. No. 3. P. 217–226. <https://doi.org/10.4304/jetwi.3.3.217-226>
30. Альтшуллер Г.С. Алгоритм изобретения. М.: Московский рабочий, 1969.
31. Alterman R., Harsch K. A more reflective form of joint problem solving // International Journal of Computer-Supported Collaborative Learning. 2017. Vol. 12. P. 9–33. <https://doi.org/10.1007/s11412-017-9250-1>
32. Протасов В. Метод эволюционного согласования решений. Компьютерная и математическая модели // Горный информационно-аналитический бюллетень. 2011. Т. 1. № 12. С. 360–379.
33. Протасов В.И., Славин Б.Б. Совершенствование инструментов электронной демократии с использованием технологий коллективного интеллекта // Информационное общество. 2017. № 2. С. 37–44.
34. Garreta-Domingo M., Sloep P.B., Hernandez-Leo D., Mor Y. Design for collective intelligence: pop-up communities in MOOCs // AI & Society. 2018. Vol. 33. No. 4. P. 91–100. <https://doi.org/10.1007/s00146-017-0745-0>
35. Kenneth J., et al. The promise of prediction markets // Science. 2008. Vol. 320. No. 5878. P. 877–888. <https://doi.org/10.1126/science.1157679>
36. Yu C., Chai Y., Liu Y. Literature review on collective intelligence: a crowd science perspective // International Journal of Crowd Science. 2018. Vol. 2. No. 3. P. 64–73. <https://doi.org/10.1108/IJCS-08-2017-0013>
37. Славин Б. Современные экспертные сети // Открытые системы. 2014. № 7. С. 30–33.
38. Деминг Э. Новая экономика. М.: Эксмо, 2006.

Об авторе

Славин Борис Борисович

доктор экономических наук;

профессор Департамента бизнес-информатики, Финансовый университет при Правительстве РФ, 125993, Москва, Ленинградский проспект, д. 49;

E-mail: bbslavin@gmail.com

ORCID: 0000-0003-3465-0311

Technologies of collective intelligence in the management of business processes of an organization

Boris B. Slavin

E-mail: bbslavin@gmail.com

Financial University under the Government of the Russian Federation
Address: 49, Leningradsky Prospekt, Moscow 125993, Russia

Abstract

With the digitalization of the economy, the creative component of the organization's activities increases. Standard business process management methods stop working due to the rise in uncertainty of the task solution time. Currently, there are no effective technologies for managing intellectual activity processes in organizations. The role of collective intelligence technologies for knowledge management in organizations has long been discussed in the literature, but there are still no concrete proposals on implementation. This work aims to show how collective technologies can solve the problems of managing business processes of intellectual activity. The possibility of collective intelligence technologies for increasing labor productivity is demonstrated. Models for distributing tasks by competencies and synergy from the collaboration are proposed for this demonstration. The paper shows that competencies are the primary metric that can be used to measure work with knowledge in an organization. But they should also be considered when organizing group activities. A simple model example shows that the correct distribution of tasks by competencies allows you to increase the speed of solving tasks by a group by several times. In real cases, calculations using computing resources are necessary. A model is also proposed that demonstrates increasing the joint activity of a creative employee and an analyst. It is shown that business process management should be supplemented by mapping the competence model and group work options to the stages of business processes. This will allow you to manage the business processes of intellectual activity.

Keywords: collective intelligence, competencies, knowledge management systems, business processes, intellectual activity, synergy, brainstorming, group work

Citation: Slavin B.B. (2022) Technologies of collective intelligence in the management of business processes of an organization. *Business Informatics*, vol. 16, no. 2, pp. 36–48. DOI: 10.17323/2587-814X.2022.2.36.48

References

1. Slavin B.B. (2017) From commodity economy to human economy. *Economics and management: problems, solutions*, vol. 7, no. 8, pp. 79–84 (in Russian).
2. Male S.A., Bush M.B., Chapman E.S. (2010) Perceptions of competency deficiencies in engineering graduates. *Australasian Journal of Engineering Education*, vol. 16, no. 1, pp. 55–68. <https://doi.org/10.1080/22054952.2010.11464039>
3. Loseva O.V. (2009) Methodology for assessing the state and analysis of the dynamics of human intellectual capital development in the organization. *Izvestiya Penza State Pedagogical University*, vol. 16, no. 12, pp. 75–71 (in Russian).
4. Spady W.G. (1978) The concept and implications of competency-based education. *Education Leadership*, pp. 16–22.
5. Homer M. (2001) Skills and competency management. *Industrial and Commercial Training*, vol. 33, no. 2, pp. 59–62. <https://doi.org/10.1108/00197850110385624>
6. Darnton G. (2002) Modelling requirements and architecting large-scale on-line competence-based learning systems. Proceedings of the *IEEE International Conference on Advanced Learning Technologies (ICALT 2002)*, Kazan, Russia, pp. 170–174.
7. Altukhova N.F., Danilina O.M. (2008) On the issue of competencies in the context of corporate knowledge management. *Bulletin of the University*, vol. 21, no. 11, pp. 9–16 (in Russian).
8. Polanyi M. (2009) *The tacit dimension*. Chicago: University of Chicago Pres. <https://doi.org/10.1007/s11016-010-9328-0>
9. Nonaka I., Takeuchi H. (2011) *The company is the creator of knowledge*. Moscow: Olymp-Business (in Russian).
10. Slavin B.B. (2016) Technologies of collective intelligence. *Problems of management*, no. 5, pp. 2–9 (in Russian).
11. Wechsler D. (1971) Concept of collective intelligence. *American Psychologist*, vol. 26, pp. 904–907.
12. Fisher K., Fisher M.D. (1997) *The distributed mind: Achieving high performance through the collective intelligence of knowledge work teams*. New York: Amacom.

13. Weiss A. (2005) The power of collective intelligence. *netWorker*, vol. 9, pp. 16–23. <https://doi.org/10.1145/1086762.1086763>
14. Levy P. (1997) *Collective intelligence: Mankind's emerging world in cyberspace*. Cambridge: Perseus Books.
15. Heylighen F. (2007) The global superorganism: An evolutionary-cybernetic model of the emerging network society. *Social Evolution & History*, vol. 5, no. 1, pp. 57–117.
16. Heylighen F. (2014) The concept of the global brain. *The Birth of the collective mind: On the new laws of the network society and the network economy and their impact on human behavior*. Moscow: Lenand (in Russian).
17. Kittur A., Lee B., Kraut R.E. (2009) Coordination in collective intelligence: The role of team structure and task interdependence. Proceedings of the 27th International Conference on Human Factors in Computing Systems, Boston, MA, USA, April 4–9, 2009, pp. 1495–1504. <https://doi.org/10.1145/1518701.1518928>
18. Chorost M. (2011) *World Mind*. Moscow: Eksmo (in Russian).
19. Zetssu K., Kiyoki Y. (2006) Towards knowledge management based on harnessing collective intelligence on the Web. Proceedings of the 15th International Conference on Managing Knowledge in a World of Networks (EKAW'06), Podebrady, Czech Republic, October 2006, pp. 350–357. https://doi.org/10.1007/11891451_31
20. Luo S., Xia H., Yoshida T., Wang Z. (2008) Toward collective intelligence of online communities: a primitive conceptual model. *Journal of Systems Science and Systems Engineering*, vol. 18, no. 2, pp. 203–221. <https://doi.org/10.1007/s11518-009-5095-0>
21. Woolley A., Aggarwa I., Malone T. (2015) Collective intelligence and group performance. *Current Directions in Psychological Science*, vol. 24, no. 6, pp. 420–424. <https://doi.org/10.1177/09637214155599543>
22. Leimeister J.M. (2010) Collective intelligence. *Business & Information Systems Engineering*, no. 4, pp. 245–248. <https://doi.org/10.1007/s12599-010-0114-8>
23. Gregg D.G. (2010) Designing for collective intelligence. *Communications of ACM*, vol. 53, no. 4, pp. 134–138. <https://doi.org/10.1145/1721654.1721691>
24. Alag S. (2008) *Collective intelligence in action*. Greenwich: Manning Publications Co.
25. Malone T.W., Laubacher R., Dellarocas C. (2009) *Harnessing crowds: Mapping the genome of collective intelligence*. MIT Sloan Research Paper No. 4732-09. <https://doi.org/10.2139/ssrn.1381502>
26. Buecheler T., Sieg J., Fuchslin R., Pfeifer R. (2010) Crowdsourcing, open innovation and collective intelligence in the scientific method: a research agenda and operational framework. Proceedings of the 12th International Conference on the Synthesis and Simulation of Living Systems (Artificial Life XII), Odense, Denmark, 19–23 August 2010, pp. 679–686. <https://doi.org/10.21256/zhaw-4094>
27. Bothos E., Apostolou D., Mentzas G. (2009) Collective intelligence for idea management with Internet-based information aggregation markets. *Internet Research*, vol. 19, no. 1, pp. 26–41. <https://doi.org/10.1108/10662240910927803>
28. Gruber T. (2008) Collective knowledge systems: Where the social web meets the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, pp. 4–13. <https://doi.org/10.1016/j.websem.2007.11.011>
29. Lykourantzou I., Vergados D., Kapetanios E., Loumos V. (2011) Collective intelligence systems: Classification. *Journal of Emerging Technologies in Web Intelligence*, vol. 3, no. 3, pp. 217–226. <https://doi.org/10.4304/jetwi.3.3.217-226>
30. Altshuller G.S. (1969) *Algorithm of invention*. Moscow: Moskovsky rabochy (in Russian).
31. Alterman R., Hirsch K. (2017) A more reflective form of joint problem solving. *International Journal of Computer-Supported Collaborative Learning*, vol. 12, pp. 9–33. <https://doi.org/10.1007/s11412-017-9250-1>
32. Protasov V. (2011) Method of evolutionary coordination of solutions. Computer and mathematical models. *Mining information and analytical bulletin*, vol. 1, no. 12, pp. 360–379 (in Russian).
33. Protasov V.I., Slavin B.B. (2017) Improving the tools of electronic democracy using technologies of collective intelligence. *Information Society*, no. 2, pp. 37–44 (in Russian).
34. Garreta-Domingo M., Sloep P.B., Hernandez-Leo D., Mor Y. (2018) Design for collective intelligence: pop-up communities in MOOCs. *AI & Society*, vol. 33, no. 4, pp. 91–100. <https://doi.org/10.1007/s00146-017-0745-0>
35. Kenneth J., et al. (2008) The promise of prediction markets. *Science*, vol. 320, pp. 877–888. <https://doi.org/10.1126/science.1157679>
36. Yu C., Chai Y., Liu Y. (2018) Literature review on collective intelligence: a crowd science perspective. *International Journal of Crowd Science*, vol. 2, no. 3, pp. 64–73. <https://doi.org/10.1108/IJCS-08-2017-0013>
37. Slavin B. (2014) Modern expert networks. *Open systems*, no. 7, pp. 30–33 (in Russian).
38. Deming E. (2006) *New Economy*. Moscow: Eksmo (in Russian).

About the author

Boris B. Slavin

Dr. Sci. (Econ.);

Professor of the Department of Business Informatics, Financial University under the Government of the Russian Federation, 49, Leningradsky Prospekt, Moscow 125993, Russia;

E-mail: bbslavin@gmail.com

ORCID: 0000-0003-3465-0311

Новые метрики энергоэффективности ИТ-отрасли

Р.Р. Сухов^a 

E-mail: r.sukhov@uptimetechnology.ru

М.Б. Амзараков^a 

E-mail: m.amzarakov@uptimetechnology.ru

Е.А. Исаев^b 

E-mail: is@itaec.ru

^a АНО «Институт «Аптайм»

Адрес: Россия, 123060, г. Москва, ул. Маршала Рыбалко, д. 7

^b Институт математических проблем биологии РАН – филиал Института прикладной математики им. М.В. Келдыша РАН
Адрес: Россия, 142290, Московская область, г. Пушкино, ул. профессора Виткевича, д. 1

Аннотация

Снижение техногенного влияния деятельности человека на экологию планеты – задача, которая из разряда теоретической в настоящее время все больше переходит в практическую плоскость. Экологическая ситуация серьезна, и требует пристального внимания к себе. Одним из существенных факторов отрицательного влияния человека на окружающую его среду являются выбросы вредных веществ, возникающие при производстве электроэнергии. Техническое развитие человечества и повсеместное внедрение ИТ-технологий, характеризуется взрывным ростом количества электронных устройств, объемов данных, передаваемых по информационным сетям, и способствует увеличению потребностей в вычислительных ресурсах для хранения и обработки этих данных. Как следствие – потребность в электроэнергии также значительно увеличивается. Вычислительное оборудование за последние 15 – 20 лет многократно нарастило вычислительную мощность, количество эксплуатируемых серверов в настоящее время исчисляется многими миллионами единиц, а суммарное энергопотребление серверного парка становится весьма значимым в структуре энергозатрат во всех развитых странах. В данной статье мы проанализируем способ снижения энергозатрат при работе серверов и центров обработки данных, применение которого обладает высоким потенциалом к экономии электроэнергии, а также приведем пример нового способа оценки эффективности работы ИТ оборудования с помощью нового коэффициента – коэффициента простоя сервера (SIC).

Ключевые слова: энергоэффективность, энергозатраты, сервер, центр обработки данных

Цитирование: Сухов Р.Р., Амзараков М.Б., Исаев Е.А. Новые метрики энергоэффективности ИТ-отрасли // Бизнес-информатика. 2022. Т. 16. № 2. С. 49–61. DOI: [10.17323/2587-814X.2022.2.49.61](https://doi.org/10.17323/2587-814X.2022.2.49.61)

Введение

Развитие глобальных (мировых) и региональных (на уровне отдельных стран) социальных сетей, интернет-сервисов, повсеместное внедрение информационных технологий во всех отраслях экономики приводит к необходимости повышения эффективности использования вычислительных ресурсов. В настоящее время это выражается в консолидации серверного оборудования в специализированных местах эксплуатации — центрах обработки данных, что позволяет снизить издержки за счет глубокой оптимизации энергоснабжения и охлаждения серверного оборудования, а также в разработке и производстве серверов с улучшенными характеристиками по энергопотреблению и вычислительной мощности.

Современные центры обработки данных, использующие инновационные способы распределения энергии и охлаждения уже приблизились к теоретическим пределам энергоэффективности. Дальнейшие технологические разработки в инженерных системах центров обработки данных позволят незначительно повышать энергоэффективность [1], при этом существенно удорожая их стоимость.

С одной стороны, на основании имеющихся исследований [2] можно сделать вывод, что, например, в Великобритании примерно 10% от общего производства электроэнергии потребляется коммерческими и государственными дата центрами и размещенными в них ИТ-системами.

С другой стороны, современные серверы вплотную приблизились к пределам компактности и энергоэффективности. Текущий «кремниевый» технологический базис не даст в ближайшем будущем

существенного снижения энергопотребления при сопоставимой вычислительной мощности.

И, несмотря на активное развитие в последние десятилетия так называемой «зеленой» энергетики — то есть энергетики, использующей источники энергии альтернативные традиционным, функционирующим на нефти, добываемом природном газе и угле, традиционные источники энергии при сгорании выделяют в атмосферу углекислый газ, способствующий росту парникового эффекта и глобальному потеплению.

В настоящее время более двух третей источников энергии в мировом производстве приходится именно на традиционные, причиняющие существенный вред окружающей среде (рис. 1) [3].

Консолидация ИТ ресурсов позволяет повысить эффективность как за счет оптимизации расходов на обслуживание, так и за счет снижения издержек при энергоснабжении и последующих дополнительных расходах на охлаждение. Появление новой отрасли — центры обработки данных — это логичное развитие ИТ отрасли в целом.

Современный центр обработки данных (ЦОД) — это высокотехнологичное предприятие, обеспечивающее непрерывное и надежное энергоснабжение серверов. Основным ресурсом, которым управляет, распределяет и снабжает центр обработки данных — это электроэнергия, эффективность использования которой определяет общую эффективность ЦОД в частности и ИТ отрасли в целом.

Понимание имеющихся ограничений подталкивает исследователей во всем мире к поиску новых способов снижения энергозатрат в ИТ отрасли.

В работе ЦОД можно выделить два самых значимых потребителя электроэнергии: серверное оборудо-

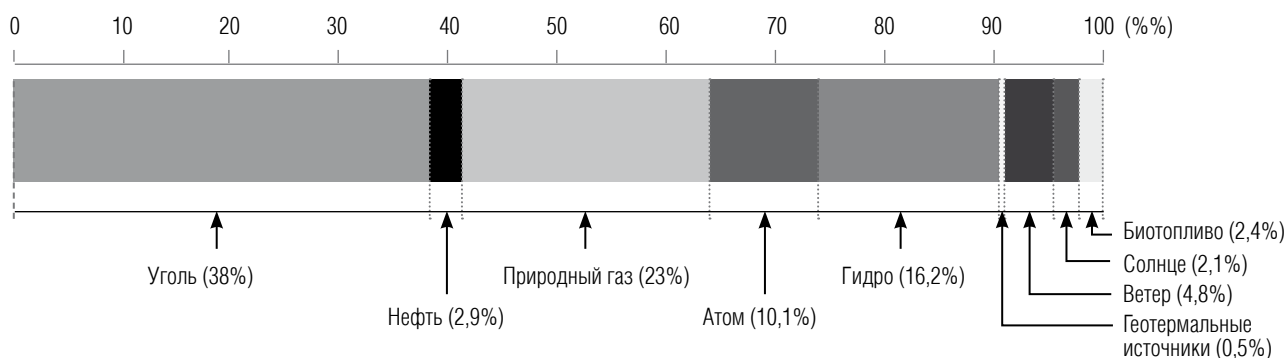


Рис. 1. Доля производства электроэнергии по источникам энергии в мире.

дование и вспомогательные инженерные системы (кондиционирование, электрораспределение и бесперебойное питание и т.п.). По каждому виду потребителей существуют различные метрики энергоэффективности, в той или иной степени позволяющие дать качественную, либо количественную оценку каждого из потребителей. Однако единая метрика, которая бы позволила комплексно объединить в себе обе оценки и дать возможность оценить влияние на итоговую энергоэффективность работы ЦОД авторам данной статьи не известна.

Именно поэтому становится важным создание единой метрики, которая позволила бы оценивать эффективность использования электроэнергии в ЦОД при проведении вычислений вне зависимости от того, какой процессор используют сервер, или какие технологии охлаждения использованы в центре обработки данных [1, 2].

Но задача измерения энергоэффективности сервера не так проста, как кажется на первый взгляд.

1. Актуальные показатели энергоэффективности

Эффективность ЦОД, с точки зрения энергозатрат на поддержание работы серверного оборудования, оценивается с помощью коэффициента энергоэффективности PUE [4]. Данный коэффициент появился в 2007 г., прочно вошел в обиход специалистов и позволяет мгновенно оценить энергоэффективность ЦОД как объекта инженерной инфраструктуры.

PUE вычисляется как отношение общего энергопотребления ЦОД (включая все энергозатраты, как затраты на ИТ, так и вспомогательные затраты) к затратам на энергоснабжение серверного оборудования ЦОД. Т.е. PUE показывает сколько электроэнергии расходует ЦОД на то, чтобы серверное оборудование работало должным образом.

$$PUE = \frac{P_{total}}{P_{IT}}, \text{ где} \quad (1)$$

P_{total} — общее количество энергии, затраченные дата-центром;

P_{IT} — количество энергии, затраченное всем ИТ оборудованием дата-центра за то же время.

Согласно данным Uptime Institute коэффициент PUE резко снижался с 2006 по 2013 годы [5], однако после 2013 года коэффициент PUE остается примерно на одном и том же уровне (рис. 2) и колеблется на уровне 1,5–1,7.

коэффициент PUE

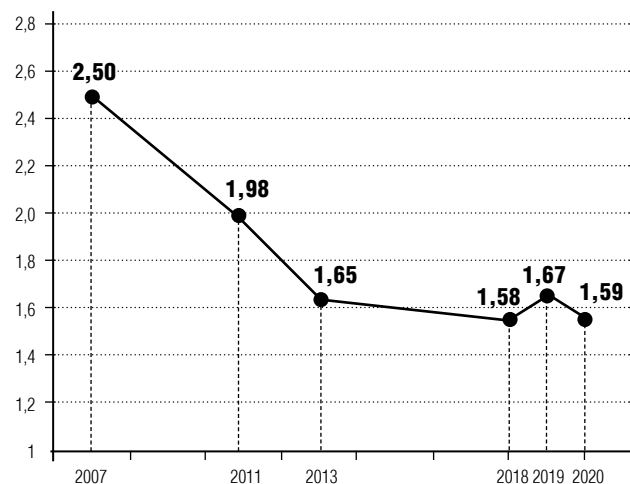


Рис. 2. Снижение динамики улучшения энергетической эффективности дата центров [5].

Каждый ватт электроэнергии, потребленный сервером, связан с энергетическими затратами на его «доставку»: передача, преобразование, охлаждение, освещение и т.п. В настоящее время эти дополнительные затраты, необходимые для обеспечения бесперебойной работы сервера, составляют 0,5–0,8 Вт на каждый ватт, потребляемый сервером.

В профессиональной среде такой эффект называется каскадным эффектом (рис. 3).

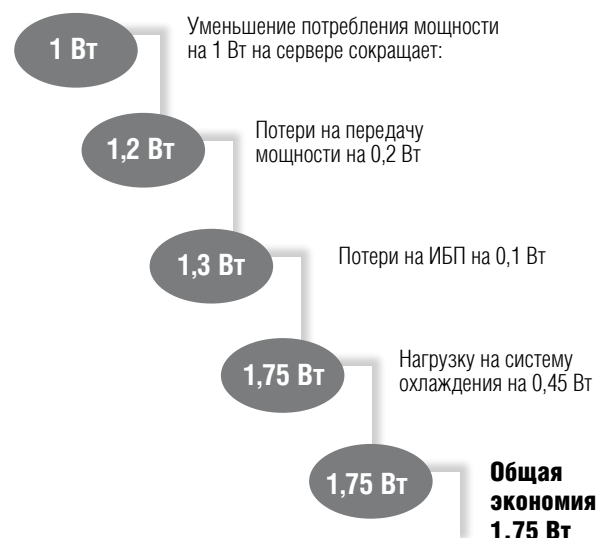


Рис. 3. Каскадный эффект.

На текущий момент не приходится ожидать, что энергоэффективность ЦОД может быть значительно улучшена. Отдельные проекты ЦОД показывают феноменально низкое значение PUE, равное 1,06–1,1. Однако стоит отметить, что такие значения достигаются при очень ограниченных условиях, при применении жестких, комплексно зависящих друг от друга условий эксплуатации инженерных систем [6] и ИТ оборудования, и в большинстве случаев сложно достижимы или практически нереализуемы [7].

Другая составляющая энергозатрат, имеющая высокий вес в общем энергопотреблении, кроется непосредственно в ИТ оборудовании.

Практически со дня появления первых компьютеров шла неустанная борьба за снижение габаритов и энергопотребления компьютеров. И в этой области достигнуты потрясающие результаты. Для примера можно посмотреть на результаты, достигнутые компанией AMD [8]. За последние 6 лет AMD повысила энергоэффективность вычислений, производимых своими процессорами, предназначенными для мобильных устройств, в 31,7 раза (рис. 4).

При этом мы наблюдаем интересный эффект. В последние 15–20 лет среднее энергопотребление непосредственно процессора и компьютера (сервера) в целом снижается незначительно, а производительность процессора, т.е. компьютера (сервера) растет значительно. Формально говоря, процессор



Рис. 4. Повышение энергоэффективности мобильных процессоров [8].

компьютера становится значительно энергетически эффективней, ведь на выполнение 1 операции требуется существенно меньше энергии, чем это требовалось раньше.

И, рассуждая в таком ключе, можно было бы предположить, что в таком случае общее энергопотребление ИТ отрасли и отрасли ЦОД должно снижаться, что является благоприятным фактором для снижения потребления ресурсов.

Однако этого не происходит. Основная причина на наш взгляд кроется в том, что все повышение производительности процессоров «расходуется» на то, чтобы в лучшем виде удовлетворять потребности общества, например, сделать более качественными видеоконтент, более быстро доставить данный контент до пользователя, сгенерировать новый контент, привлечь новых пользователей, создать новые ИТ сервисы, которые все более широко используют нейронные сети, построенные в свою очередь также на серверных кластерах и т.п.

Пока мы не видим, что в этой «гонке вооружений» есть какой-то предел, но можем ли мы при этом найти способы снижения энергопотребления, которые в меньшей степени будут зависеть от вышеописанных драйверов, порождающих спрос на ИТ услуги?

2. Управление энергоснабжением

Основываясь на данных, приведенных в статье [1], можно принять за основу распределение энергии в среднестатистическом сервере, приведенное в таблице 1.

Таблица 1.

Распределение затрат энергии в сервере

Компонент	Потребление, Вт	Потребление, %
Процессор	115	63,89%
Оперативная память	15	8,33%
Жесткие диски	2	13,33%
Передача данных по сети	5	2,78%
Охлаждение	8,87	4,93%
Блок питания	11,13	6,18%
Прочие энергозатраты	1	0,56%

Из таблицы видно, что основным потребителем является процессор сервера — это почти 64% общего энергопотребления сервера. При этом также известно, что в среднем, в зависимости от характера производимых вычислений, серверные процессоры заняты полезной работой от 2 до 30 процентов времени [9, 10]. Из этого можно сделать вывод, что при том, что процессор потребляет значительную часть энергии в сервере, по факту процессор большую часть времени не выполняет полезную работу.

Для производителей серверов и других участников ИТ отрасли данный факт не является тайной, и они для минимизации потерь энергии при простоях сервера, а также для снижения риска перегрева процессоров и возникновения отрицательных эффектов в полупроводниках (туннельный эффект) уже достаточно давно встраивают в системы управления серверами специальные алгоритмы и способы для управления энергосбережением, которые позволяют достаточно гибко настроить работу всех элементов сервера (процессор, оперативная память, постоянная память, видеокарта и т.п.) и тем самым снизить непродуктивные потери сервера в целом. Исследования в направлении, как сделать системы управления энергосбережением сервера эффективнее ведутся непрерывно.

В качестве примеров таких исследований можно отметить:

- ♦ исследование лаборатории Беркли: Comparing server energy use and efficiency using small sample sizes (сравнение энергопотребления и эффективности серверов на примере серверов типовых размеров малого формата) [11];
- ♦ доклад на международной конференции по высоко производительным вычислениям, передаче данных, хранению и анализу: Energy-aware data transfer algorithms (энергосберегающие алгоритмы передачи данных) [12];
- ♦ доклад на 9-й международной конференции по прикладной энергетике: Development of a simple power consumption model of information technology (IT) equipment for building simulation (разработка простой модели энергопотребления оборудования информационных технологий (ИТ) для моделирования строительства) [13].

И в будущем мы увидим новые поколения серверов и телекоммуникационного оборудования, которые благодаря технологиям, созданным на базе таких исследований, будут обладать лучшими характеристиками, чем современные.

Как основные направления, которые бы позволили повысить энергоэффективность, можно выделить следующие:

- ♦ измерение энергопотребления;
- ♦ контроль энергопотребления;
- ♦ управление энергопотреблением.

3. Новые тенденции в повышении эффективности

Как мы отметили ранее, энергоэффективность сервера зависит от того, на сколько он загружен полезной работой, выполняя вычисления.

Существует множество различных метрик по измерению энергоэффективности вычислений. Самая простая метрика — измерение количества энергии, затраченной для производства вычисления с плавающей запятой [14].

Для персональных компьютеров распространена метрика SPEC [15], вычисляющая количество энергии, затраченной на выполнение типовых действия на компьютере под управлением разных операционных систем.

Energy Star (организация при Агентстве по охране окружающей среды EPA, USA) разработало целую программу по измерению эффективности — SERT [16], включающую измерение электропотребления при производстве различных операций.

SUN Microsystems предложила метрику SWAP (space, wattage and performance) [17].

Однако у всех упомянутых метрик есть один общий недостаток. А именно, необходимость определить производительность сервера. Для метрики SWAP полезная работа определяется непосредственно указанием выполняемых комплексных действий. Для метрики SPEC производительность определяется производительностью работы известных программ. Для SERT — это производительность выполнения специализированного ПО.

И эти ограничения существенно сужают возможности использования таких метрик, что в свою очередь не дает возможности распространить ту или иную методику на все типы и виды серверов и вычислений.

Но со стороны общества, представляемого государством, есть серьезный запрос на поиск и внедрение новых механизмов, технологий, позволяющих получить дополнительную экономию электроэнергии при использовании серверов.

И в этом смысле государство выполняет очень важную функцию, а именно задает правила и стандарты, устанавливает метрики, либо граничные значения, достижение которых становится обязательным, тем самым стимулируя вовлеченные отрасли в поиск новых решений и их внедрения в технологии и оборудование.

Широко распространенным в мире примером повышения энергоэффективности является введение классов энергоэффективности в промышленном и бытовом электрооборудовании, а также установление контрольных дат, когда начинают действовать ограничения, либо полный запрет на оборот техники с низкими показателями энергоэффективности, как это введено в России в виде государственного отраслевого стандарта (ГОСТ) [18], или на примере положений комиссии ЕС [2, 19].

Другим показательным примером реализации программы повышения энергоэффективности является программа LEAP («Lower Energy Acceleration Program») — Программа развития пониженного энергопотребления (ПРПЭ) [20]. Данная программа является инициативой, реализуемая правительством Нидерландов, и является частью такой программы Евросоюза. В рамках этой программы производятся исследования, направленные на поиск тех направлений, где возможны значимые улучшения с точки зрения повышения энергоэффективности.

В частности, очень интересным представляется отчет, подготовленный компаниями Certios и WCoolIT по заказу Netherlands Enterprise Agency (Нидерландское промышленное агентство). Отчет называется «LEAP Track 1 “Powermanagement” Pilot analysis» — ПРПЭ Этап 1 Пилотный анализ энергосбережения [9].

Прежде чем перейти непосредственно к метрике SIC будет полезно описать критерии, применяемые при вычислении метрики.

Авторы метрики SIC не претендуют на определение производительности, либо полезной работы, однако для функционирования любого вычислительного устройства характерно выполнение «паразитных вычислений» (для процессора — инструкция NOOP). С точки зрения авторов, любое вычислительное действие является полезным, вне зависимости от того, какое именно вычисление производится, и насколько оно «полезно» потребителю. Однако, большую часть времени процессор не производит вычислительных действий, а находится в режиме ожидания, при котором выполня-

ется инструкция NOOP. Энергия, затраченная на выполнение данной операции, считается прямой потерей.

Идея исследования состояла в том, чтобы попытаться понять, насколько эффективно используется процессорное время, насколько влияют на энергосбережение, внедренные в аппаратную часть серверов, либо операционных систем, возможности по управлению энергосбережением, а также, есть ли возможность представить фактическую энергоэффективность сервера в одинаково понятном для всех виде.

Для исследования была взята статистика пула серверов, работающих с реальной нагрузкой и разными профилями самой нагрузки. Были проверены все основные режимы энергосбережения, встроенные в серверы и некоторые операционные системы в различных режимах нагрузок.

Полученные данные были проанализированы и представлены в очень наглядном виде, демонстрирующем реальную ситуацию с энергопотреблением данных серверов. На *рис. 5* показана зависимость энергопотребления в зависимости от нагрузки на сервер. Приведен пример работы сервера с включенным режимом управления энергопотреблением. Хорошо видна прямая зависимость энергопотребления от загрузки процессора. Также мы видим, что данный сервер имеет явно выраженные максимумы и минимумы полезной нагрузки на процессор, связанные со спецификой приложений, работающих на данном сервере.

Оценка энергоэффективности проводилась с применением методики измерения времени полезной работы процессора сервера и времени простоя, которые соотносились с данными контроля энергопотребления в эти моменты времени.

В качестве итогового коэффициента полезной работы сервера в данном исследовании был предложен коэффициент SIC (Server Idle Coefficient) — коэффициент простоя сервера.

$$SIC = \frac{E_{total}}{E_{total} - E_{idle}}, \quad (2)$$

$$SIC\% = 100\% \cdot \frac{E_{total}}{E_{idle}}, \text{ где} \quad (3)$$

E_{total} — полная энергия, затраченная сервером;

E_{idle} — энергия, затраченная сервером во время простоя.

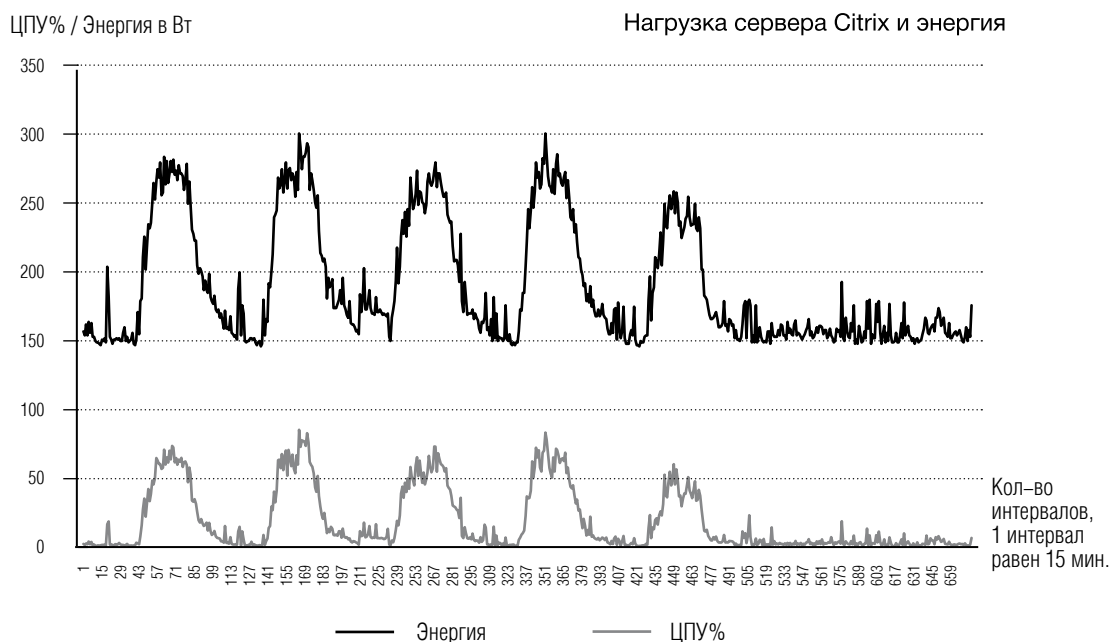


Рис. 5. Графики зависимости потребляемой энергии процессора от нагрузки [9].

Показатель (2) интерпретируется по аналогии с коэффициентом PUE, т.е. чем ближе значение к единице, тем большее количество энергии из общего количества потраченной энергии сервер тратит на полезную работу.

Показатель (3) представляет собой процентное отношение расхода энергии, когда сервер был в простое, к общему количеству потраченной энергии. Т.е., чем ближе данный показатель к 100%, тем меньше времени данный сервер производит полезную работу.

В исследовании отмечено, что:

- ♦ показатели SIC% в разных группах серверов варьировались от 34% до 91%;
- ♦ у технического персонала, отвечающего за эксплуатацию серверов, существует пробел в знаниях о роли виртуализации в управлении энергоснабжением;
- ♦ до сих пор сильны предрассудки в отношении существенного снижения производительности систем, настроенных на использование динамических режимов энергосбережения;
- ♦ у большинства эксплуатантов серверных кластеров отсутствуют какие-либо четкие правила и политики в отношении управления питанием серверов. А там, где эти политики существуют, то они чаще всего отдают предпочтение в пользу максимальной производительности сервера.

Полученные исследовательские данные демонстрируют очень высокий потенциал по снижению неэффективного расходования электроэнергии.

В качестве практического примера можно проанализировать данные, представленные на рис. 6.

На графиках рис. 6 представлены данные о загрузке процессора сервера и его энергопотреблении в двух режимах работы: верхний график — режим высокой производительности, нижний график — режим энергосбережения под управлением операционной системы сервера.

Полученные данные были обработаны и представлены в итоговых значениях потребленной энергии в разных режимах и был вычислен SIC для режима включенного энергосбережения:

- ♦ общая потребленная энергия за период: 24,5 кВт;
- ♦ общая энергия, потраченная на периоды простоя: 8,43 кВт;
- ♦ среднее значение времени простоя ЦПУ: 60,4%.

Используя формулы (2, 3), вычислен коэффициент SIC в процентном соотношении и в виде отношения количества общей потраченной энергии, к энергии, затраченной на вычислительные нужды:

$$\text{SIC}\% = 8,43/24,4 = 34,4\%;$$

$$\text{SIC} = 1,5;$$

и в режиме отключенного энергосбережения:

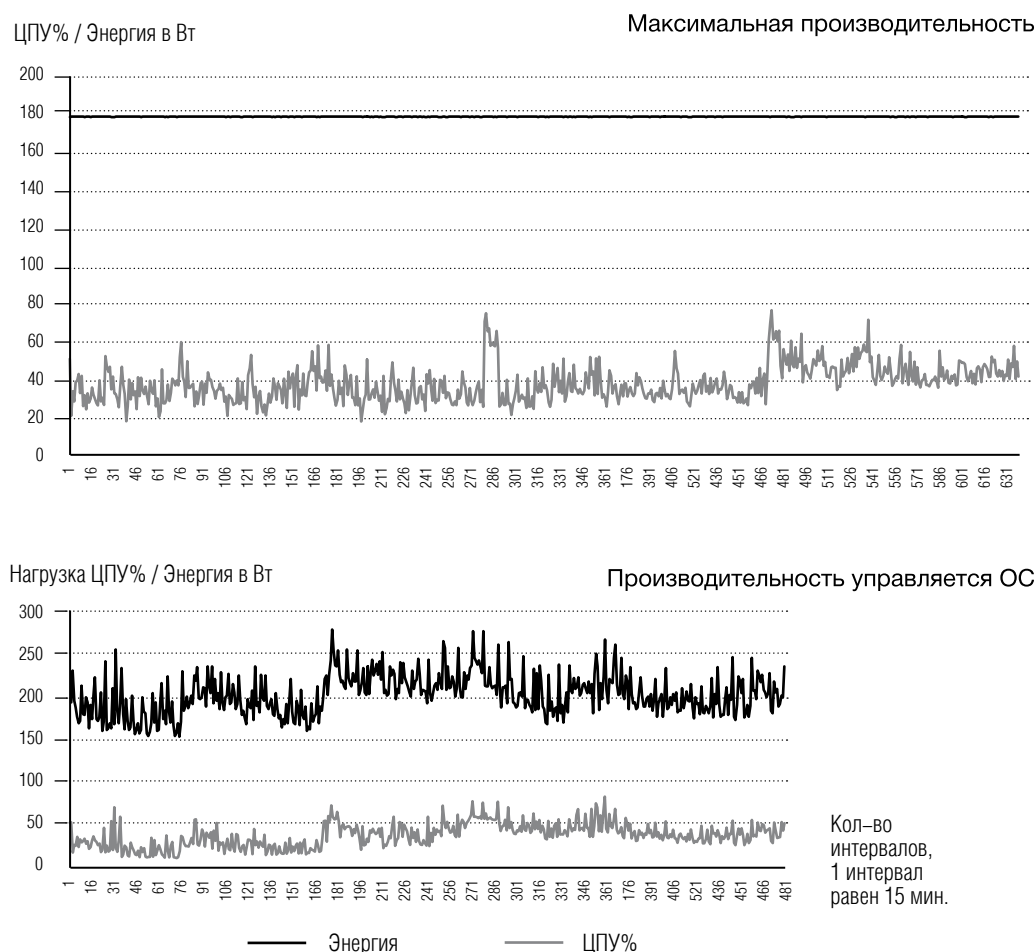


Рис. 6. Нагрузка сервера и разные режимы энергосбережения [9].

$SIC\% = \text{время простоя ЦПУ} = 60,4\%$

$SIC = 2,53$.

Таким образом, коэффициент SIC дает наглядное представление об эффективности работы сервера.

Данный подход представляется интересным и перспективным, т.к. позволяет оценивать эффективность работы серверного оборудования без привязки к его реальным параметрам энергопотребления, что дает возможность оценивать эффективность комплекса: сервер – программное обеспечение – кластеризация – динамическое управление нагрузками.

Развитие данного подхода безусловно потребует более глубокого изучения, разработки единой методологии и механизмов извлечения данных для оценки и последующей интерпретации полученных данных.

4. Оценка практической пользы применения коэффициента простоя сервера

Улучшение значения коэффициента простоя возможно двумя путями. Первое направление – это улучшение планирования исполнения инструкций операционной системой и программным обеспечением. Что возможно, например, за счёт лучшего распределения вычислений между серверами. Задача повышения времени полезного использования вычислительных мощностей – это задача разработчиков информационных систем и ПО.

Второе направление – снижение электропотребления серверным оборудованием при отсутствии вычислений – это задача для разработчиков процессоров и серверного оборудования.

Можно утверждать, что повсеместное внедрение метрики SIC позволяет создать единую базисную среду для сравнимой и достоверной оценки эффективности работы современных серверов, что в комбинации с государственным регулированием даст большой стимул разработчикам программного обеспечения в повышении полезной загрузки процессоров вычислительной техники, а разработчикам оборудования в повышении энергоэффективности их устройств.

Коэффициент SIC также позволяет применить новые методы практической реализации программ повышения энергоэффективности в дата центрах с минимальными инвестиционными затратами на внедрение таких практик.

Переход на контроль энергопотребления серверами с помощью коэффициента SIC позволяет выполнять эти процедуры в режиме реального времени, не затрачивая ресурсы на дорогостоящие и длительные специальные измерения коэффициента PUE или его производных.

Автоматизированное фоновое вычисление SIC на каждом сервере и агрегация в единую систему анализа может позволить практически мгновенно выявлять серверное оборудование, используемое неэффективно.

Ориентируясь на осторожные оценки авторов исследования [9] в возможностях энергосбережения в размере не менее 10% для высоконагруженных серверов при переходе к динамическому управлению энергопотреблением серверов, оценим размер потенциальной экономии электроэнергии для небольшого типового дата центра, обладающего следующими характеристиками:

- ♦ общее количество эксплуатируемых серверов: 2 000 шт.;
- ♦ среднее энергопотребление одного сервера: 200 Вт;
- ♦ влияние каскадного эффекта (коэффициент PUE) дата центра: 1,7.

Предположим, что в результате контроля коэффициента SIC нам в среднем удалось снизить потребляемую мощность одного сервера на 10%, мы получим:

$$P_{\text{экономии}} = 200 \text{ (Вт)} \cdot 10\% = 20 \text{ (Вт)} -$$

для одного сервера;

Принимая во внимание общее количество серверов, равно 2000 шт., мы получим:

$$P_{\text{общ. экономии ИТ}} = 20 \text{ (Вт)} \cdot 2000 = 40 \text{ (кВт)} -$$

для всей ИТ нагрузки.

Таким образом мы видим, что незначительные на первый взгляд 20 Вт экономии потребляемой мощности на одном сервере на масштабе нашего примера, в количестве 2000 серверов, позволяют получить общее снижение потребляемой электрической мощности в размере 40 кВт.

Приняв во внимание коэффициент PUE, получим:

$$P_{\text{общ. экономии}} = 40 \text{ (кВт)} \cdot 1,7 = 68 \text{ (кВт)} -$$

для дата центра,

Снижение потребляемой мощности на вспомогательных инженерных системах, необходимых для нормального функционирования серверов добавляет в общую экономию еще 28 кВт электрической энергии.

Перейдем от потребляемой электроэнергии, измеряемой в кВт к потребленной, учитываемой в кВт·ч и вычислим объем экономии электрической энергии в месяц:

$$P_{\text{общ. экономии}} = 68 \text{ (кВт)} \cdot 720 \text{ (часов)} = 48\,960 \text{ (кВт·ч)},$$

где 720 часов — это усредненное количество часов в месяц.

Далее вычислим объем экономии электрической энергии в год:

$$P_{\text{общ. экономии}} = 68 \text{ (кВт)} \cdot 8640 \text{ (часов)} = 587\,250 \text{ (кВт·ч)},$$

где 8 640 часов — это усредненное количество часов в год.

Если принять, что в Москве стоимость одного кВт·ч электрической энергии в 2021 году составляет 7,11 руб., то за год мы получим общую экономию в денежной форме:

$$S_{\text{общ. экономии}} = 587\,250 \text{ (кВт·ч)} \cdot 7,11 \text{ (руб.)} = 4\,175 \text{ (млн руб.)}.$$

Приведенный пример дата центра, в сравнении с гигантами отрасли, обладает небольшой мощностью, около 700 кВт, однако этот пример наглядно показывает, какой существует потенциал в снижении общего энергопотребления дата центрами и ИТ отраслью.

Заключение

В настоящей статье рассмотрен вопрос основных тенденций в области повышения эффективности использования электроэнергии при эксплуатации серверных кластеров и систем.

Энергоэффективность производимых вычислений на уровне отдельных устройств из расчета количества операций на единицу мощности, и снижение

эксплуатационных потерь инженерных систем дата-центров [21] – являются основным драйвером повышения энергоэффективности ИТ отрасли в целом.

Показано, что снижение энергопотребления на 10% на серверах снижает необходимую мощность с 700 до 632 кВт и дает существенную экономию в затратах на оплату потребленной электроэнергии.

Необходимо обратить пристальное внимание на новые способы повышения энергоэффективности ИТ отрасли, а именно на изменение подходов в управлении серверными системами. Объединение единичных устройств – серверов в кластеры и системы с одновременным управле-

нием загрузкой процессорных мощностей, а также динамическое управление энергоснабжением элементов сервера позволяет в дополнение к основным способам повышения энергоэффективности дать еще один весомый инструмент управления и контроля.

Введение в практику новых способов оценки эффективности работы серверного оборудования, таких как коэффициент простоя сервера (SIC), могут дать качественно новую оценку эффективности производимых вычислений, вне зависимости от того, насколько сам по себе процессор сервера энергоэффективен. ■

Литература

1. Амзараков М.Б., Сухов Р.Р., Исаев Е.А., Амзаракова А.М. Энергоэффективность центра обработки данных как совокупности инженерной и ИТ инфраструктуры // Приборы и системы. Управление, контроль, диагностика. 2019. № 12. С. 47–52.
2. A new European roadmap to cleaner, greener data centres // BCS, The Chartered Institute for IT, 2021. [Электронный ресурс]: <https://www.bcs.org/content-hub/a-new-european-roadmap-to-cleaner-greener-data-centres/> (дата обращения 28.02.2022).
3. World gross electricity production by source, 2019 // IEA, 2021. [Электронный ресурс]: <https://www.iea.org/data-and-statistics/charts/world-gross-electricity-production-by-source-2019> (дата обращения 28.02.2022 г.)
4. Power usage effectiveness // Wikipedia, the free encyclopedia, 2022. [Электронный ресурс]: https://en.wikipedia.org/wiki/Power_usage_effectiveness (дата обращения 28.02.2022).
5. Lawrence A. Data center PUEs have been flat since 2013 // Uptime Institute, 2020. [Электронный ресурс]: <https://www.datacenterdynamics.com/en/opinions/data-center-pues-have-been-flat-2013/> (дата обращения 28.02.2022).
6. Обзор «зеленых» энергетических стратегий и методик для современных центров обработки данных // Delta Power Solution, 2014. [Электронный ресурс]: https://www.deltapowersolutions.com/media/download/White-Paper_Overview-of-Green-Energy-Strategies-and-Techniques-for-Modern-Data-Centers_WP0014_ru-ru.pdf (дата обращения 28.02.2022).
7. Supermicro second annual green data center report finds opportunity for saving millions in energy costs, and reductions in E-Waste // Super Micro Computer, Inc., 2019. [Электронный ресурс]: <https://www.supermicro.com/en/pressreleases/supermicro-second-annual-green-data-center-report-finds-opportunity-saving-millions> (дата обращения 28.02.2022).
8. AMD 25x20 energy efficiency initiative // Advanced Micro Devices, Inc, 2022. [Электронный ресурс]: <https://www.amd.com/en/technologies/25x20> (дата обращения 28.02.2022).
9. Harryvan D., Verzijl M., Amzarakov M. LEAP Track 1 'Powermanagement' Pilot analysis // Netherlands Enterprise Agency, 2020. [Электронный ресурс]: <https://amsterdameconomicboard.com/app/uploads/2020/10/LEAP-Track-1-'Powermanagement'-Pilot-analysis.pdf> (дата обращения 28.02.2022).
10. Meisner D., Gold B.T., Wenisch T.F. PowerNap: eliminating server idle power // ACM SIGARCH Computer Architecture News. 2009. Vol. 37. No. 1. P. 205–216. <https://doi.org/10.1145/2528521.1508269> (дата обращения 28.02.2022).
11. Coles H.C., Qin Y., Price P.N. Comparing server energy use and efficiency using small sample sizes // Lawrence Berkeley National Laboratory, 2014. [Электронный ресурс]: <https://buildings.lbl.gov/publications/comparing-server-energy-use-and> (дата обращения 28.02.2022).
12. Alan I., Arslan E., Kosar T. Energy-aware data transfer algorithms // Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC'15). 2015. Article No.: 44. P. 1–12. <https://doi.org/10.1145/2807591.2807628>
13. Cheung H., Wang S., Zhuang C. Development of a simple power consumption model of information technology (IT) equipment for building simulation // Energy Procedia. 2017. Vol. 142. P. 1787–1792. <https://doi.org/10.1016/j.egypro.2017.12.564>
14. Green500 results // TOP500.org. [Электронный ресурс]: <https://www.top500.org/lists/green500/> (дата обращения 28.02.2022).
15. Standard Performance Evaluation Corporation. [Электронный ресурс]: <http://www.spec.org/index.html> (дата обращения 28.02.2022).
16. Fogle R. How to measure server efficiency with SERT // Energy Star. [Электронный ресурс]: <https://www.energystar.gov/products/ask-the-experts/how-to-measure-server-efficiency-with-sert-> (дата обращения 28.02.2022).

17. Sun introduces new metric for server efficiency // Phys.org, 2005. [Электронный ресурс]: <https://phys.org/news/2005-12-sun-metric-server-efficiency.html> (дата обращения 28.02.2022).
18. ГОСТ Р 51749-2001. Энергосбережение. Энергопотребляющее оборудование общепромышленного применения. 2002. [Электронный ресурс]: <http://docs.cntd.ru/document/1200012993> (дата обращения 28.02.2022).
19. Commission Regulation (EU) 2019/424 of 15 March 2019 // Official Journal of the European Union. 2019. No. L 74. P. 46–66. [Электронный ресурс]: <https://eur-lex.europa.eu/eli/reg/2019/424/oj> (дата обращения 28.02.2022).
20. LEAP – Lower Energy Acceleration Program // The Amsterdam Economic Board, 2020. [Электронный ресурс]: <https://amsterdameconomicboard.com/en/results/lancering-lower-energy-acceleration-programme-leap/> (дата обращения 28.02.2022).
21. Амзараков М.Б., Сухов Р.Р., Исаев Е.А. Модульный центр обработки данных: Целостный взгляд // Бизнес-информатика. 2014. Т. 29. № 3. С. 7–14.

Об авторах

Сухов Рафаэль Ряхимович

финансовый управляющий, АНО «Институт «Аптайм», 123060, г. Москва, ул. Маршала Рыбалко, д. 7;

E-mail: r.sukhov@uptimetechnology.ru

ORCID: 0000-0002-8124-137X

Амзараков Максим Борисович

директор АНО «Институт «Аптайм», 123060, г. Москва, ул. Маршала Рыбалко, д. 7;

E-mail: m.amzarakov@uptimetechnology.ru

ORCID: 0000-0001-6229-8592

Исаев Евгений Анатольевич

кандидат технических наук;

старший научный сотрудник, Институт математических проблем биологии РАН – 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

New energy efficiency metrics for the IT industry

Rafael R. Sukhov^a

E-mail: r.sukhov@uptimetechnology.ru

Maxim B. Amzarakov^a

E-mail: m.amzarakov@uptimetechnology.ru

Evgeny A. Isaev^b

E-mail: is@itaec.ru

^a INO Uptime Technology

Address: 7, Marshala Rybalko st., Moscow 123060, Russia

^b Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences

Address: 1, Professor Vitkevich st., Pushchino 142290, Russia

Abstract

Reducing the technogenic impact of human activity on the ecology of the planet is a problem that is increasingly moving from a theoretical category into a practical one. The environmental situation is serious and requires more attention. One of the significant factors of the negative impact of humans on their environment is the emissions of harmful substances that occur during the production of electricity. The technical development of humanity and the widespread introduction of information technologies are characterized by an explosive growth in the number of electronic devices and the amount of data transmitted over information networks. This contributes to an increase in the need for computing resources for storing and processing this data, and as a result, the need for electricity is also increasing. Over the past 15–20 years, computing equipment has increased its computing power many times. The number of servers in operation is currently estimated at many millions of units, and the total energy consumption of the server park is becoming very significant in the structure of energy costs in all developed countries. In this article, we will analyze a way to reduce energy costs in the operation of servers and data centers, the application of which has a high potential for saving energy. We will give an example of a new way to evaluate the efficiency of IT equipment using a new factor – the server idle coefficient (SIC).

Keywords: energy efficiency, energy consumption, server, data center

Citation: Sukhov R.R., Amzarakov M.B., Isaev E.A. (2022) New energy efficiency metrics for the IT industry. *Business Informatics*, vol. 16, no. 2, pp. 49–61. DOI: 10.17323/2587-814X.2022.2.49.61

References

1. Amzarakov M.B., Sukhov R.R., Isaev E.A., Amzarakova A.M. (2019) Energy efficiency of the Data Processing Center as a combination of engineering and IT infrastructure. *Instruments and Systems: Monitoring, Control, and Diagnostics*, no. 12, pp. 47–52 (in Russian).
2. BCS, the Chartered Institute for IT (2021) *A new European roadmap to cleaner, greener data centres*. Available at: <https://www.bcs.org/content-hub/a-new-european-roadmap-to-cleaner-greener-data-centres/> (accessed 28 February 2022).
3. IEA (2021) *World gross electricity production by source, 2019*. Available at: <https://www.iea.org/data-and-statistics/charts/world-gross-electricity-production-by-source-2019> (accessed 28 February 2022).
4. Wikipedia, the free encyclopedia (2022) *Power usage effectiveness*. Available at: https://en.wikipedia.org/wiki/Power_usage_effectiveness (accessed 28 February 2022).
5. Lawrence A. (2020) *Data center PUEs have been flat since 2013*. Uptime Institute. Available at: <https://www.datacenterdynamics.com/en/opinions/data-center-pues-have-been-flat-2013/> (accessed 28 February 2022).
6. Delta Power Solution (2014) *Overview of green energy strategies and methodologies for modern data centers*. Available at: <https://www.deltapowersolutions.com/en-us/mcis/white-paper-overview-of-green-energy-strategies-and-techniques-for-modern-data-centers.php> (accessed 28 February 2022).
7. Super Micro Computer, Inc. (2019) *Supermicro second annual green data center report finds opportunity for saving millions in energy costs, and reductions in E-Waste*. Available at: <https://www.supermicro.com/en/pressreleases/supermicro-second-annual-green-data-center-report-finds-opportunity-saving-millions> (accessed 28 February 2022).
8. Advanced Micro Devices, Inc. (2022) *AMD 25x20 energy efficiency initiative*. Available at: <https://www.amd.com/en/technologies/25x20> (accessed 28 February 2022).
9. Harryvan D., Verzijl M., Amzarakov M. (2020) *LEAP Track 1 'Powermanagement' Pilot analysis*. Netherlands Enterprise Agency. Available at: <https://amsterdameconomicboard.com/app/uploads/2020/10/LEAP-Track-1-'Powermanagement'-Pilot-analysis.pdf> (accessed 28 February 2022).
10. Meisner D., Gold B.T., Wenisch T.F. (2009) PowerNap: eliminating server idle power. *ACM SIGARCH Computer Architecture News*, vol. 37, no. 1, pp. 205–216. <https://doi.org/10.1145/2528521.1508269>
11. Coles H.C., Qin Y., Price P.N. (2014) *Comparing server energy use and efficiency using small sample sizes*. Lawrence Berkeley National Laboratory. Available at: <https://buildings.lbl.gov/publications/comparing-server-energy-use-and> (accessed 28 February 2022).
12. Alan I., Arslan E., Kosar T. (2015) Energy-aware data transfer algorithms. Proceedings of the *International Conference for High Performance Computing, Networking, Storage and Analysis (SC'15)*. Article No.: 44, pp. 1–12. <https://doi.org/10.1145/2807591.2807628>
13. Cheung H., Wang S., Zhuang C. (2017) Development of a simple power consumption model of information technology (IT) equipment for building simulation. *Energy Procedia*, vol. 142, pp. 1787–1792. <https://doi.org/10.1016/j.egypro.2017.12.564>

14. TOP500.org (2021) *Green500 results*. Available at: <https://www.top500.org/lists/green500/> (accessed 28 February 2022).
15. *Standard Performance Evaluation Corporation*. Available at: <http://www.spec.org/index.html> (accessed 28 February 2022).
16. Fogle R. *How to measure server efficiency with SERT*. Energy Star. Available at: <https://www.energystar.gov/products/ask-the-experts/how-to-measure-server-efficiency-with-sert-> (accessed 28 February 2022).
17. Phys.org (2005) *Sun introduces new metric for server efficiency*. Available at: <https://phys.org/news/2005-12-sun-metric-server-efficiency.html> (accessed 28 February 2022).
18. GOST R 51749-2001 (2002) *Energy saving. Energy-consuming equipment for general industrial use*. Available at: <http://docs.cntd.ru/document/1200012993> (accessed 28 February 2022).
19. Commission Regulation (EU) 2019/424 of 15 March 2019 (2019) *Official Journal of the European Union*, no. L 74, pp. 46–66. Available at: <https://eur-lex.europa.eu/eli/reg/2019/424/oj> (accessed 28 February 2022).
20. The Amsterdam Economic Board (2020) *LEAP – Lower Energy Acceleration Program*. Available at: <https://amsterdameconomicboard.com/en/results/lancering-lower-energy-acceleration-programme-leap/> (accessed 28 February 2022).
21. Amzarakov M.B., Sukhov R.R., Isaev E.A. (2014) The modular data center: a holistic view. *Business Informatics*, vol. 29, no. 3, pp. 7–14.

About authors

Rafael R. Sukhov

Financial manager, INO Uptime Technology, 7, Marshala Rybalko st., Moscow 123060, Russia;

E-mail: r.sukhov@uptimetechnology.ru

ORCID: 0000-0002-8124-137X

Maxim B. Amzarakov

Director, INO Uptime Technology, 7, Marshala Rybalko st., Moscow 123060, Russia;

E-mail: m.amzarakov@uptimetechnology.ru

ORCID: 0000-0001-6229-8592

Evgeny A. Isaev

Cand. Sci. (Tech.);

Senior Research Fellow, Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, 1, Professor Vitkevich st., Pushchino 142290, Russia;

E-mail: is@itaec.ru

ORCID: 0000-0002-3703-447X

Доверенный искусственный интеллект как способ цифровой защиты

С.М. Авдошин 

E-mail: savdoshin@hse.ru

Е.Ю. Песоцкая 

E-mail: epesotskaya@hse.ru

Национальный исследовательский университет «Высшая школа экономики»

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

Аннотация

Статья посвящена аспектам, связанным с феноменом цифровой защиты и массового внедрения нейронных сетей в современном обществе. С одной стороны, проникновение цифровых технологий расширит возможности бизнеса и социальных институтов, а также общества, которым поручено выполнять многочисленные задачи с большей эффективностью. Однако есть и серьезные угрозы, связанные с использованием искусственного интеллекта в части работы с данными пользователей. Авторы исследуют вызовы и формулируют новую концепцию надежного искусственного интеллекта для противодействия некорректному цифровому проникновению. Целью статьи является изучение потенциальных опасностей злоупотребления нейронными сетями со стороны владельцев технологий и сопротивления им с опорой на доверенный искусственный интеллект. Изучая различные аспекты цифровой трансформации и использования технологий искусственного интеллекта, авторы формализуют опасности и предлагают подход к использованию технологий цифровой защиты, которым можно доверять.

Ключевые слова: нейронные сети, цифровая защита, доверие, искусственный интеллект, общество

Цитирование: Авдошин С.М., Песоцкая Е.Ю. Доверенный искусственный интеллект как способ цифровой защиты // Бизнес-информатика. 2022. Т. 16. № 2. С. 62–73. DOI: [10.17323/2587-814X.2022.2.62.73](https://doi.org/10.17323/2587-814X.2022.2.62.73)

Введение

Искусственный интеллект (ИИ) занимает все большее место в современном обществе, становясь основой для инноваций и прогресса [1–5]. Не являются исключением сферы, активно внедряющих цифровые техноло-

гии с целью модернизации процессов и ускорения экономического развития.

Концепция искусственного интеллекта (ИИ) была введена Джоном Маккарти в 1955 году [6, 7]. Понятие ИИ подразумевает способность системы интерпретировать данные, извлекать из них уро-

ки и использовать полученные сведения для достижения конкретных целей и задач посредством гибкой адаптации [6].

Стоит отметить, что обсуждение определения ИИ к настоящему времени не привело к результату, удовлетворяющему всех заинтересованных сторон в технологиях ИИ. С 2016 года индустрия искусственного интеллекта развивается благодаря поддержке облачных вычислений и больших данных. Сегодня искусственные нейронные сети, развернутые с использованием облачных технологий, а также глубокое машинное обучение формируют основу большинства приложений, которые нам известны под названием ИИ.

Использование таких сетей наряду с компьютерными мощностями даёт возможность решать различные проблемы, поскольку искусственные нейронные сети состоят из абстрактной модели взаимосвязанных нейронов, специальное расположение и связь которых могут быть использованы для решения прикладных задач в различных областях, таких как статистика, технологии или экономика [8].

Аналитические и когнитивные инструменты ИИ позволяют владельцем технологий анализировать колоссальные по объёму данные, а значит своевременно выявлять и эффективно реагировать на изменения в ситуации в бизнесе и обществе. Именно нейронные сети могут стать важным шагом на пути к реализации принципов открытого общества, повысить уровень прозрачности, оптимизировать внутренние процессы межведомственного взаимодействия и стимулировать инновационную деятельность, что в итоге приведет к высокому уровню доверия [9].

На повестке дня для многих компаний и организаций стоит вопрос создания и последующего внедрения централизованной цифровой экосистемы, направленной на улучшение взаимосвязи и взаимодействия всех заинтересованных лиц (государства, бизнеса, объединений и частных лиц) [9–11]. Главная роль в этой трансформации будет возложена именно на ИИ.

Однако ситуация неоднозначна и ставит под вопрос своевременность широкого использования ИИ [12]. Помимо уже всесторонне определенных рисков нарушения приватности, хакерских атак, технологической сингулярности и некорректного цифрового проникновения, имеют место конкретные предпосылки другой серьезной опаснос-

сти [5, 13–15]. Речь идет об аккумулировании и последовательном использовании возможностей цифровых технологий в интересах узкой группы людей. Однозначно можно утверждать, что набор современных алгоритмов открывает практически безграничные возможности. Подробный сценарий отражает концепцию появления в современном мире цифровой диктатуры (digital dictatorship) и необходимости цифровой защиты для регулирования недобросовестного использования технологий.

Выходом может стать повсеместное установление принципов доверенного ИИ (Trusted Artificial Intelligence) [6, 12]. В рамках этой концепции предполагается найти золотую середину между прогрессивностью стратегий применения ИИ и защитой этических и моральных аспектов человеческой жизнедеятельности. При имеющейся незначительной законодательной базе и в отсутствие международного консенсуса применение нейронных сетей на данный момент – это не только вызов традиционной модели работы властей, но и угроза основным человеческим ценностям [6, 7, 12]. Более того, имеется огромный разрыв между основами международного права и реальностью, где эти принципы постоянно нарушаются. Крайне важно выстроить продуманную нормативно-правовую систему, отвечающую запросам научно-технического прогресса и регулирующую все многообразие алгоритмов нейросетей. Важно также разработать конкретные технологии, которые позволят противодействовать возможным угрозам.

Таким образом, целью данной статьи является изучение рисков злоупотребления возможностями ИИ со стороны владельцев технологий и способов сопротивления им с опорой на концепцию доверия. Чтобы избежать проблем, которые могут нанести вред человеку путем искажения, хищения или утечки данных, необходимо убедиться, что результатам работы ИИ можно доверять. В данной работе исследуются потенциальные проблемы, связанные с вопросами доверия, конфиденциальности и надежности, рассматривается понятие «доверенного искусственного интеллекта» а также феномен цифровой защиты. В разделе 1 подробно разбирается природа нарушений прав граждан с точки зрения четырех различных сфер жизни: политической, социальной, культурной и экономической. Мы указываем на необратимость процесса внедрения нейронных сетей в контексте цифровой трансформации и описываем идею доверенного ИИ.

Раздел 2 посвящен обсуждению возможных стратегий реагирования и предложения по установлению цифровой защиты; завершается наше исследование кратким изложением ключевых выводов.

1. Перспективы и методы искусственного интеллекта

1.1. Новые возможности искусственного интеллекта

Научно-технический прогресс не остановить. Под давлением стремительного развития IT-сектора мировой экономики государства оказываются вынуждены отвечать меняющимся запросам бизнеса, в противном случае рискуя столкнуться с настоящей цифровой пропастью в управлении [9]. Искусственный интеллект плавно проник в нашу жизнь, умные системы применяются во множестве сфер, от клиентской аналитики и поисковых систем до голосовых помощников и медицинских исследований. В медицинской области разрабатываются системы, распознающие патологии по видеозаписям эндоскопических исследований; в транспортной — автопилоты и системы управления транспортным потоком; в финансовой — системы, опознающие клиентов или выявляющие подозрительные операции, которые могут указывать на уклонение от налогов или отмывание денег. Можно с уверенностью сказать, что дальнейшее мировое экономическое развитие и прогресс напрямую зависят от того, насколько эффективно различные индустрии научатся пользоваться искусственным интеллектом. Однако с развитием технологии более остро встает проблема доверенного искусственного интеллекта, поскольку любые неполадки могут иметь тяжелые последствия. Пользователи хотят быть уверены, что модель обладает высокой точностью, а ее результаты справедливы, легко интерпретируемы. Например, неверно откалиброванные датчики автомобиля, оборудованного автопилотом, могут стать причиной аварии. Ошибки в контроле над инфраструктурой, на которую полагается ИИ, могут привести к утечке персональных данных пациентов или краже личности. Низкий уровень программного обеспечения (ПО) ИИ может привести к постановке неверного диагноза или задержке в транспортировке, разрушающей дальнейшую цепочку поставок.

С одной стороны, вычислительные и аналитические мощности нейронных сетей, превосходящие возможности человеческого мозга, открывают

новые горизонты для бизнеса и социальных институтов (например, в предиктивной аналитике). Помимо этого, ИИ не имеет ограниченного запаса выносливости и доступен для использования в любой момент. Благодаря этим характеристикам возможны колоссальные преобразования во многих секторах экономики: системах дорожного движения, здравоохранении, поддержании общественного порядка и персонализации государственных услуг, в том числе образования [15]. Внушительные успехи уже достигнуты в предоставлении услуг гражданам и юридическим лицам. При этом, конечно, важно отметить ограниченность выполняемых ИИ на данный момент задач. Среди основных категорий — проработка запросов (о социальных выплатах, миграции, по вопросам граждан и т. д.), заполнение и поиск документов, перевод текстов [4].

Значительные успехи нейронным сетям пророчат и в области управления экономикой. Планируется качественное повышение эффективности в распределении ресурсов и логистики через переустройство и оптимизацию каналов поставок, систем складов и утилизации [15]. Также нельзя не упомянуть многообещающую концепцию «Умный город» (smart city), опирающуюся именно на нейронные сети в вопросах контроля камер видеонаблюдения, электрических сетей, водоснабжения, транспортных систем и т. д. [7, 16].

Нужно отметить, что значительные изменения произошли под влиянием пандемии COVID-19 — в частности в области преподавания и обучения. Академические учреждения переходят на цифровые технологии, чтобы предоставить своим студентам больше ресурсов. Благодаря технологиям у студентов теперь есть больше возможностей учиться и улучшать навыки в своем собственном темпе и по индивидуальной траектории, проходить этапы контроля при помощи онлайн-тестов. Все больше популярности набирают сервисы онлайн-прокторинга, при которых лицо испытуемого идентифицируется и анализируется, чтобы предсказать его эмоции. Кроме того, обнаруживаются такие аспекты, как телефон, книга или присутствие другого человека. Эта комбинация моделей создает интеллектуальную систему логического вывода, основанную на правилах, которая способна определить, имело ли место какое-либо злоупотребление во время прохождения экзамена или теста.

Здесь может вставать вопрос о корректности работы системы и адекватности оценивания поведе-

ния. Любые сбои и злоупотребления чреваты негативными последствиями в части академической честности, дискредитации идеи как офлайн, так и онлайн-обучения. Среди ключевых рисов стоит отметить нарушение конфиденциальности, скомпрометированная доступность или скомпрометированные учетные записи, утечка персональных данных или искажение результатов.

Для успешной разработки моделей машинного обучения необходимы большие данные. Количество, качественно и доступность больших данных влияют на эффективность и точность обучаемых моделей, поэтому многие компании заинтересованы в организации постоянного сбора данных о своих потребителях. Многие системы собирают информацию, не подлежащую разглашению: видео и фотографии с видеокамер, записи речи, финансовые транзакции. Неразумное использование этой информации, ошибки в модели или хищение данных могут стать причиной угрозы безопасности отдельных личностей или даже предприятий и государственных организаций.

1.2. Потенциальные проблемы и угрозы цифрового проникновения

Утечка данных для искусственного интеллекта особенно опасна из-за того, что большие данные обычно несут в себе очень много конфиденциальной информации, из которой можно получить много сведений об объекте, на который осуществлена атака. При этом утечка данных может произойти на любом этапе разработки: обучении или использовании готовой модели.

Нарушение конфиденциальности является другой важной деталью, поскольку именно личная информация выступает катализатором цифровой трансформации, становясь основой для моделей обучения. Процесс разработки многих нейронных сетей практически неотделим от опоры на собираемые данные включая речь, интернет-активность, изображения, финансовые потоки, медицинские показатели и т. д. Таким образом, вопрос доступа к большим данным оказывается одной из самых существенных проблем, связанных с интеграцией нейронных сетей в общество [13]. Любой третьей стороне будут очевидны потенциальные риски использования собираемой владельцами технологий информации касательно информации пользователей, их права на частную жизнь и защиты их персональных данных [17, 18]. Этот аспект широко об-

суждается обществом, так как за ним стоит прямая угроза здоровью и жизни человека. Актуальность проблемы многократно озвучивалась видными научными и государственными деятелями, а соответствующие отчеты и исследования не раз были представлены в рамках международной дискуссии (IEEE, специализированных комитетов ЕС, ОЭСР и т. д.) [6].

Публично гарантируя прозрачность, полный аудит и объективность благодаря внедрению нейросетей, владельцы технологий в эпоху цифровизации способны осуществлять любые манипуляции. Данная возможность вытекает напрямую из непонимания обществом и пользователями цифровых технологий структуры и принципов работы алгоритмов. Сложные модели искусственного интеллекта выполняют колоссальные вычисления, которые не поддаются всестороннему анализу даже разработчиков [5, 7]. Таким образом, для большинства пользователей процесс работы нейронных сетей будет непрозрачным [15]. Ученые называют это проблемой «черного ящика». Пользуясь этим феноменом, недобросовестные разработчики могут использовать нейронные сети по своему усмотрению.

Еще одна проблема, связанная с данными для обучения, это их низкая доступность. Зачастую небольшие объемы данных принадлежат разным лицам, не имеющим оснований доверять друг другу или разработчику, и составить один датасет — набор данных — достаточного размера оказывается невозможным. Если же датасет необходимого объема существует, он все равно может быть недоступен, если данные в нем конфиденциальны. Даже если подходящие данные удастся найти, необходимо удостовериться, что они отражают реальное положение вещей. В частности, в них не должно содержаться скрытых предвзятостей, как это, например, произошло с некоторыми системами распознавания лиц: из-за несбалансированности данных они, например, значительно лучше справлялись с распознаванием светлокожих мужчин, чем темнокожих женщин.

Данные, с которыми работает готовая система ИИ, также могут быть низкого качества: они могут исходить из ненадежных источников или содержать информацию с высокой степенью неопределенности. Кроме того, базы данных, с которыми взаимодействует система, могут оказаться под угрозой, если сама система взломана. Например, система проверки биометрических данных может быть

подвержена нескольким типам атак с целью заставить ее принять злоумышленника за владельца. В обрабатываемые данные добавляется шум так, чтобы уже обученная модель определила объект на фотографии в неправильный класс. Такие атаки могут применяться в компьютерном зрении, например, заставляя модель неправильно определять дорожные знаки.

Предвзятость модели может крайне негативно повлиять на результат использования цифровых технологий. Среди возможных причин предвзятости – неравномерное распределение данных в обучающей выборке, алгоритмически заложенные предпочтения и предвзятое отношение к отдельным группам лиц. Даже классический спектр потенциальных кризисных кейсов огромен, начиная с эйджизма и сексизма при приеме на работу и заканчивая расизмом при выявлении потенциальных преступников [7].

Более того, недопустимо исключать ситуации, когда предвзятость модели является осмысленной политикой ее создателей, продиктованной их интересами, что несет еще большую опасность. Без должного контроля разработчики могут получить серьезную власть над обществом. При этом степень разрушительности последствий дискриминации трудно переоценить: систематические нарушения прав определенных социальных групп приведут к определению ИИ как бесчеловечного механизма.

Чтобы избежать проблем, которые могут нанести вред человеку путем искажения, хищения или утечки данных, необходимо убедиться, что результатам работы ИИ можно доверять. Так возникает потребность в понятии «доверенный искусственный интеллект»: система ИИ, в отношении которой пользователь может быть уверен, что она способна качественно выполнять возложенные на нее задачи [19].

2. Результаты и обсуждение: внедрение доверенного ИИ

Стоит признать, что внедрение искусственного интеллекта представляется весьма выгодным процессом. Однако именно доверие является ключевым фактором в вопросе применения ИИ, так как отторжение данной технологии массами исключит всю потенциальную пользу [5, 8, 13]. Следовательно, установление рациональных доверительных отношений, исключающих чрезмерное доверие или

его отсутствие, позволит добиться выгоды для всего общества. Сущность этой идеи отражена в концепции доверенного ИИ.

Доверие к ИИ на физическом уровне означает уверенность в корректной работе всех его физических компонент, таких как датчики, и в качестве получаемых системой данных. Доверие к инфраструктуре, окружающей ИИ, означает уверенность в безопасности данных, с которыми ИИ взаимодействует, и в контроле над доступом к самой системе. Доверие на прикладном уровне означает уверенность в корректной работе ПО.

Если система ИИ считается доверенной, то есть доверие к ней проявляется на всех её уровнях, то такую систему можно допускать к решению задач с масштабными последствиями, так как пользователь может быть уверен в результатах ее работы.

Вопрос доверия достаточно комплексный и напрямую зависит от целого спектра особенностей человеческой психики [13, 20, 21]. Исследователи подчеркивают, что доверие есть желание и готовность отдельного лица зависеть от поступков другой стороны с целью извлечения некоторой выгоды, несмотря на потенциальные риски уязвимого положения. В основе этого феномена – совпадение нравственных ценностей, позволяющее предугадывать дальнейшие действия стороны, облеченной доверием, и уверенность в ее достаточной компетенции [20]. Также необходимо отметить, что кумулятивный характер доверия ставит дальнейшие отношения в зависимость от первого опыта. Значит, уже на изначальном этапе массового внедрения нейросетей должны быть четко выработаны стратегии выстраивания доверенного ИИ, отвечающего критериям этики и компетенции.

Далее представим несколько ключевых подходов, которые кажутся наиболее важными на пути внедрения доверенного искусственного интеллекта, для каждой из заинтересованных сторон.

А. Международное сообщество

Первый этап заключается в разработке законодательной базы, регулирующей работу ИИ в каждой из индустрий. Уже сейчас различные международные комитеты и организации заняты составлением общих рекомендаций и руководств к действию [21]. Несмотря на то, что отдельные попытки контролировать технологическую задачу предпринимались и ранее [22], отправной точкой в дискуссии вокруг этики и компетентности ИИ считается до-

кумент ОЭСР Principles on Artificial Intelligence [23, 24]. Представив первичные наработки и ключевые принципы, авторы создали базу для будущих более серьезных документов [23, 25, 26]. Объединяя в себе наиболее значимые критерии доверенного ИИ, они нацелены на способствование выработке собственной политики странами — лидерами в области цифровых технологий [20].

Ключевые принципы включают постоянный человеческий контроль, надежность, подотчетность, конфиденциальность и прозрачность, но не ограничиваются ими. В будущих конвенциях могут также содержаться понятия непредвзятости и справедливости, сохранения социального и экологического благополучия, гибкости алгоритмов и введения так называемого права на объяснение [23, 25–27]. Важно подчеркнуть, что, пока органы власти будут обязаны не только постоянно адаптировать правовую базу, но и расширять ее, международные конвенции должны обеспечивать базисность объявленных принципов. В итоге, опираясь на данные принципы, власти смогут выработать сбалансированную систему норм, разграничив области регулирования между диспозитивными (soft laws) и императивными актами (hard laws) [7]. Важно осознавать созидательный характер процесса, когда вопрос контроля

ИИ в госуправлении решается на основе уже имеющейся законодательной базы, а не вопреки ей. Таким образом, демократические государства смогут противодействовать риску нарушения прав человека, стоящих во главе угла и оберегаемых независимыми инстанциями и законом. Избегая даже потенциальной возможности предвзятого отношения как на этапе разработки, так и в процессе применения, государство предложит этические основания доверять ИИ. Однако, несмотря на краеугольное значение будущих норм, они лишь указывают на важность соблюдения закона в вопросах безопасности и качества, при этом оставаясь абстрактными правилами.

Концепция доверенного ИИ представлена на рис. 1 и затрагивает множество аспектов.

В. Научное сообщество

Первостепенное значение имеет способность ученых и разработчиков защищать нейронные сети. В противном случае основные надежды на доверие нейросетям будут подорваны, и ситуация потенциально будет развиваться по сценарию смещения в сторону цифровой зависимости граждан от третьих сторон. Однако эволюция методов защиты происходит почти синхронно с аналогич-

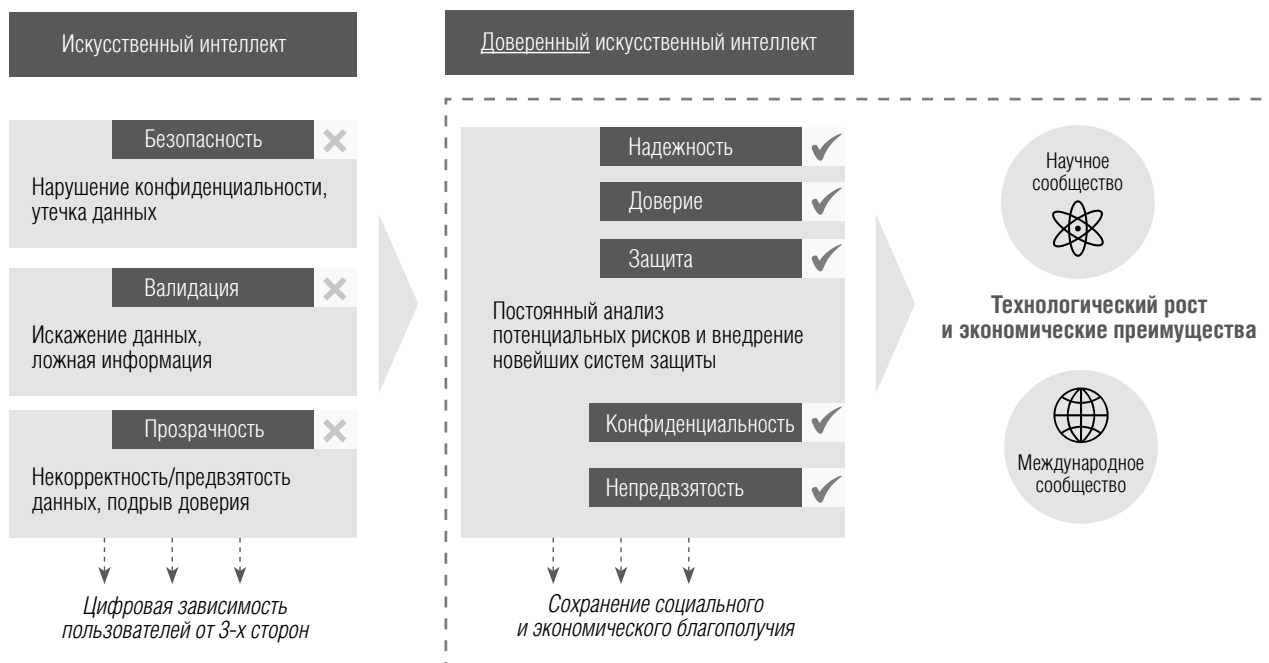


Рис. 1. Концепция использования доверенного ИИ.

ным процессом для атак на искусственный интеллект. Следовательно, в данном случае наиболее важно создание стандартов, которые гарантировали бы не только постоянный анализ потенциальных рисков, но и внедрение новейших систем защиты. Например:

1. Нарушение конфиденциальности (privacy breaches) – одна из наиболее вероятных проблем в будущем. Утечки личной информации могут возникнуть практически в любой момент работы нейронной сети – от обучения до вывода результатов. Очевидно, что данные частных лиц крайне выгодны определенным третьим лицам во многих сферах. В настоящий момент фокус разработчиков для защиты от потенциальных атак смещается с псевдонимизации и деидентификации к целому ряду новых технологий повышения конфиденциальности (PET). Одна из самых заметных – OPAL (open algorithms project) – заключается в предоставлении алгоритмам удаленного и контролируемого доступа к данным вместо пересылки анонимизированных данных. Таким образом, благодаря целому ряду специальных механизмов гарантируется возврат разработчикам модели только агрегированного результата, а запись всех операций обеспечивает возможность полной проверки процесса обучения [28, 29].

2. Другой сценарий, считающийся одним из наиболее опасных, – это отравление данных (*data poisoning*). Его идея состоит во включении в обучающую выборку ложной информации, либо обесценивающей результаты всей системы, либо подталкивающей нейросеть к принятию неверных решений, выгодных третьей стороне. Подобные технологии могут как потенциально нанести значительный урон репутации отдельной личности, так и повлиять на поведение масс (к примеру, на выборах) [20]. Конкретного решения пока не предложено. Однако, несмотря на весь свой угрожающий потенциал, риски в этой области пока минимальны: данные технологии все еще находятся на раннем этапе развития, не подразумевающим особых угроз. Тем не менее в литературе уже описано несколько моделей для борьбы с отравлением данных, почти каждая из которых акцентирует внимание на важности человеческого контроля и защиты.

3. Серьезное беспокойство вызывает и так называемая атака уклонения (evasion attack). В отличие от отравления данных, используемого во время фазы обучения, здесь угроза возникает на этапе применения ИИ. Незаметно видоизменяя входные значения, можно получить кардинально новый от-

вет сети. Даже малейшие преобразования могут привести к серьезным последствиям [7]. Имеется несколько механизмов противодействия атаке уклонения, однако самый широко описанный – это состязательная тренировка (adversarial training), подразумевающая включение разработчиками на этапе обучения заведомо неверных данных, позволяющих модели в дальнейшем не обращать внимания на потенциальные шумы [17, 30].

4. Очевидна интеллектуальная ценность уже обученной нейронной сети, особенно при использовании собираемых государством больших данных. Инверсия модели (model extraction) может потенциально означать колоссальные утечки персональных данных. Подобное нарушение конфиденциальности способно привести к крайне тяжелым последствиям, например, при утечке медицинских карт [7]. Так как сама технология существует пока только в научных статьях и абстрактных моделях, методы борьбы с ней тоже представлены лишь в теории. Среди наиболее признанных можно выделить приватное агрегирование обучающих моделей (PETA). Идея данной концепции состоит в разъединении данных на несколько наборов – «учителей», каждый из которых обучает отдельную нейросеть. Далее эти независимые модели объединяются для обучения нейросети «ученик» путем голосования, в итоге, не давая последней доступа к изначальным данным [18, 31].

5. В научной литературе представлен целый ряд метрик, позволяющих с определенной эффективностью обнаружить предвзятости модели (difference in means, difference in residuals, equal opportunity, equal mis-opportunity, disparate impact, normalized mutual information). Опираясь на данные математические подходы, разработчики смогут выявлять несовершенства своих нейросетей и вносить соответствующие изменения. Смягчение предвзятости может быть достигнуто за счет одной из трех категорий алгоритмов: pre-processing algorithms, in-processing algorithms и post-processing algorithms [32].

Наконец, важным аспектом политики объединенного научного сообщества должна оказаться своевременность представления конкретных технологий, которые позволят более эффективно реализовывать заявленные принципы. Даже сейчас разрыв между используемыми алгоритмами и абстрактными идеалистическими теориями очевиден. В то же время уже давно проводятся многочисленные углубленные исследования, изучающие возможные сценарии и инструменты.

Несколько подходов к решению вышеуказанных проблем уже были представлены. Один из них побуждает нас комплексно разобрать технологию блокчейн — непрерывную цепочку блоков, связанных в обратном порядке посредством хеш-сумм. Каждый блок, помимо собственного хеша и хеша предыдущего блока, содержит некоторую информацию. Таким образом, блокчейн представляет собой распределенный реестр общего пользования, позволяющий хранить и передавать данные с особо высокой надежностью и без возможности вмешательства. Интеллектуальные контракты (smart contracts), то есть программный код, обеспечивающий выполнение всех установленных условий, в связке с машинным обучением гарантирует надежность итоговых результатов.

Следовательно, объединение технологии блокчейн с ИИ позволит создать децентрализованную систему, хранящую информацию любой ценности и предоставляющую ее для обучения нейросети. В итоге не только гарантируется безопасность данных, но и решается проблема черного ящика благодаря полной и при этом децентрализованной истории операций, исключающей внешнее вмешательство или определенную предвзятость [24, 33].

Также появляются инструменты для отслеживания прозрачности и справедливости моделей. Компания IBM представила несколько инструментов для разработки доверенного искусственного интеллекта: библиотеку на Python, помогающую проверять модели на устойчивость к различным атакам — Adversarial Robustness Toolbox (ART); алгоритмы, анализирующие процесс работы моделей ИИ для создания прозрачности — AI Explainability 360; библиотеку с метриками, показывающими справедливость модели — AI Fairness 360 [34].

Выводы

Повсеместная интеграция цифровых технологий и создание децентрализованных экосистем способны открыть новые горизонты и устранить ряд текущих проблем. Ключевая роль нейронных сетей в этом процессе заставляет уделять особое внимание каждой потенциальной угрозе. Особого внимания заслуживает вопрос возникновения цифровой защиты и использование доверенного искусственного интеллекта. Совместный контроль за соблюдением принципов доверенного ИИ позволит как научному сообществу, так и международным организациям создать и регулярно обновлять механизмы противодействия рискам. Более того, у человечества в целом появится шанс грамотно распорядиться своим политическим опытом, накопленным за столетия.

Несмотря на то, что почти каждая проблема имеет конкретное решение, представленные технологические стратегии — не более чем теоретизация на тему планирующихся в ближайшем будущем процессов. Другими словами, идеалистические тезисы, основывающиеся на биоэтике, будут неизменно сталкиваться с реальностью, обусловленной ситуацией в каждой отдельной стране. Именно поэтому важно постоянно адаптировать абстрактные принципы под изменяющуюся реальность, расширять законодательную базу и находить новые решения. В то же время крайне необходима твердая коллективная приверженность сохранению демократических институтов.

Научный прогресс уже сейчас позволяет делать первые шаги в сторону массового внедрения ИИ. Хотя нынешние масштабы относительно скромны, перспективы вдохновляют. Если обществу удастся достойно справиться с предстоящими вызовами, перечисленные выше изменения поднимут человечество на совершенно новый уровень существования. ■

Литература

1. Vogt T., Winter P., Nessler B., Doms T. Trusted artificial intelligence: towards certification of machine learning applications. Vienna: TÜV Austria Holding AG, 2021.
2. Kuleshov A., Ignatiev A., Abramova A., Marshalko G. Addressing AI ethics through codification // 2020 International Conference Engineering Technologies and Computer Science (EnT), 2020. P. 24–30. <https://doi.org/10.1109/EnT48576.2020.00011>
3. Harrison T., Luna-Reyes L. Cultivating trustworthy artificial intelligence in digital government // Social Science Computer Review. 2021. Vol. 40. No. 2. P. 494–511. <https://doi.org/10.1177/0894439320980122>
4. Recommendation of the council on digital government strategies. Paris: OECD Publishing, 2014. [Электронный ресурс]: <https://www.oecd.org/gov/digital-government/Recommendation-digital-government-strategies.pdf> (дата обращения: 22.09.2021).
5. Chakraborty A., Alam M., Dey V., Chattopadhyay A., Mukhopadhyay D. Adversarial attacks and defences: A survey. Working paper arXiv: 1810.00069, 2018. <https://doi.org/10.48550/arXiv.1810.00069>

6. Haenlein M., Kaplan A. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence // *California Management Review*. 2019. Vol. 61. No. 4. P. 5–14. <https://doi.org/10.1177/0008125619864925>
7. Wei J. Research progress and application of computer artificial intelligence technology // *MATEC Web of Conferences*. 2018. Vol. 176. Article Number 01043. <https://doi.org/10.1051/mateconf/201817601043>
8. Mijwil M., Esen A., Alsaadi A. Overview of neural networks. 2019. [Электронный ресурс]: https://www.researchgate.net/publication/332655457_Overview_of_Neural_Networks (дата обращения: 22.09.2021).
9. Sibai F.N. AI crimes: A classification // 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security). 15–19 June 2020. P. 1–8. <https://doi.org/10.1109/CyberSecurity49315.2020.9138891>
10. Jastroch N. Trusted artificial intelligence: On the use of private data // *Product Lifecycle Management Enabling Smart X. PLM 2020* (eds. F. Nyffenegger, J. Ríos, L. Rivest, A. Bouras). IFIP Advances in Information and Communication Technology. 2020. Vol. 594. https://doi.org/10.1007/978-3-030-62807-9_52
11. Nemitz P. Constitutional democracy and technology in the age of artificial intelligence // *Phil. Trans. R. Soc. A*. 2018. Vol. 376. <http://doi.org/10.1098/rsta.2018.0089>
12. Misra S.K., Das S., Gupta S., Sharma S.K. Public policy and regulatory challenges of artificial intelligence (AI) // *Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation. TDIT 2020* (eds. S.K. Sharma, Y.K. Dwivedi, B. Metri, N.P. Rana). IFIP Advances in Information and Communication Technology. 2020. Vol. 617. https://doi.org/10.1007/978-3-030-64849-7_10
13. Павлютенкова М.Ю. Электронное правительство vs цифровое правительство в контексте цифровой трансформации // *Мониторинг общественного мнения: Экономические и социальные перемены*. 2019. № 5. С. 120–135. <https://doi.org/10.14515/monitoring.2019.5.07>
14. Концепция цифрового правительства как политический проект для России: перспективы реализации в условиях вызовов и рисков цифровизации общества / А.П. Кочетков и [др.] // *Власть*. 2021. Т. 29. № 1. С. 317–331. <https://doi.org/10.31171/vlast.v29i1.7963>
15. Williams M., Valayer C. Digital government benchmark – study on digital government transformation. European Union, 2018. [Электронный ресурс]: <https://joinup.ec.europa.eu/collection/elise-european-location-interoperability-solutions-e-government/document/report-digital-government-benchmark-study-digital-government-transformation> (дата обращения: 22.09.2021).
16. Carter D. Regulation and ethics in artificial intelligence and machine learning technologies: Where are we now? Who is responsible? Can the information professional play a role? // *Business Information Review*. 2020. Vol. 37. No. 2. P. 60–68. <https://doi.org/10.1177/0266382120923962>
17. Kamat G. Algorithms for private data analysis. Lecture 14 – Private ML and stats: Modern ML. [Электронный ресурс]: <http://www.gautamkamath.com/CS860notes/lec14.pdf> (дата обращения: 02.09.2021).
18. Hinnefeld J., Cooman P., Mammo N., Deese R. Evaluating fairness metrics in the presence of dataset bias. Working paper arXiv: 1809.09245, 2018. <https://doi.org/10.48550/arXiv.1809.09245>
19. Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения. ГОСТ Р 59276-2020 // Национальный стандарт Российской Федерации, 2021.
20. Tinholt D., Carrara W., Linden N. Unleashing the potential of Artificial Intelligence in the Public Sector. Capgemini, 2017. [Электронный ресурс]: <https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf> (дата обращения: 22.09.2021).
21. Sharma G.D., Yadav A., Chopra R. Artificial intelligence and effective governance: A review, critique and research agenda // *Sustainable Futures*. 2020. Vol. 2. Article ID 100004. <https://doi.org/10.1016/j.sfr.2019.100004>
22. McCormick T.R., Min D. Principles of Bioethics. University of Washington, 2020. [Электронный ресурс]: <https://depts.washington.edu/bhdept/ethics-medicine/bioethics-topics/articles/principles-bioethics> (дата обращения: 22.09.2021).
23. Lindgren I., Veenstra A.F. Digital government transformation: a case illustrating public e-service development as part of public sector transformation // *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, Delft, The Netherlands, 2018. P. 1–6.
24. Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, Version 2. The IEEE global initiative on ethics of autonomous and intelligent systems. IEEE, 2021. [Электронный ресурс]: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (дата обращения: 22.09.2021).
25. Брадул Н.В., Лебезова Э.М. Концептуализация понятия «Smart Government»: наукометрический подход // *Управленец*. 2020. Т. 11. № 3. С. 33–45. <https://doi.org/10.29141/2218-5003-2020-11-3-3>
26. Thiebes S., Lins S., Sunyaev A. Trustworthy artificial intelligence // *Electron Markets*. 2021. Vol. 31. P. 447–464. <https://doi.org/10.1007/s12525-020-00441-4>
27. Falco G., Viswanathan A., Caldera C., Shrobe H. A master attack methodology for an AI-based automated attack planner for smart cities // *IEEE Access*. 2018. Vol. 6. P. 48360–48373. <https://doi.org/10.1109/ACCESS.2018.2867556>
28. AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias / R. Bellamy [et al.] Working paper arXiv: 1810.01943, 2018. <https://doi.org/10.48550/arXiv.1810.01943>
29. Towards enabling trusted artificial intelligence via Blockchain / K. Sarpatwar [et al.] // Calo, S., Bertino, E., Verma, D. (eds) *Policy-Based Autonomic Data Governance. Lecture Notes in Computer Science*. Vol. 11550. P. 137–153. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-17277-0_8

30. Montjoye Y.D., Farzanehfar A., Hendrickx J., Rocher L. Solving artificial intelligence's privacy problem // Field Actions Science Reports. 2017. Special Issue 17. P. 80–83. [Электронный ресурс]: <https://journals.openedition.org/factsreports/pdf/4494> (дата обращения: 22.09.2021).
31. Open Algorithms (OPAL) Project. DataCollaboratives.org, 2021. [Электронный ресурс]: <https://datacollaboratives.org/cases/open-algorithms-opal-project.html> (дата обращения: 22.09.2021).
32. Salah K., Rehman M.H.U., Nizamuddin N., Al-Fuqaha A. Blockchain for AI: Review and open research challenges // IEEE Access. 2019. Vol. 7. P. 10127–10149. <https://doi.org/10.1109/ACCESS.2018.2890507>
33. Baker-Brunnbauer J. Management perspective of ethics in artificial intelligence // AI and Ethics. 2021. Vol. 1. P. 173–181. <https://doi.org/10.1007/s43681-020-00022-3>
34. IBM. [Электронный ресурс]: <https://www.ibm.com/us-en/> (дата обращения: 22.09.2021).

Об авторах

Авдошин Сергей Михайлович

кандидат технических наук;

профессор департамента компьютерной инженерии, Московский институт электроники и математики им. А.Н. Тихонова, Национальный исследовательский университет «Высшая школа экономики», 101000, г. Москва, ул. Мясницкая, д. 20;

E-mail: savdoshin@hse.ru

ORCID: 0000-0001-8473-8077

Песоцкая Елена Юрьевна

кандидат экономических наук;

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

E-mail: epesotskaya@hse.ru

ORCID: 0000-0003-2129-4645

Trusted artificial intelligence: Strengthening digital protection

Sergey M. Avdoshin

E-mail: savdoshin@hse.ru

Elena Yu. Pesotskaya

E-mail: epesotskaya@hse.ru

HSE University

Address: 20, Myasnitskaya Street, Moscow 101000, Russia

Abstract

This article is devoted to aspects associated with the up-coming need for mass implementation of neural networks in the modern society. On the one hand, the latter will fully expand the capabilities of state institutions and society delegated to perform numerous tasks with higher efficiency. However, a significant threat to democratic institutions obliges society to set out the concept of reliable artificial intelligence (AI). The authors explore a new concept of a trusted AI necessary for the scientific and international community to counter improper future digital penetration. Explaining to what extent digital transformation is mandatory, the authors emphasize the numerous dangers

associated with the applications of artificial intelligence. The purpose of the article is to study the potential hazards of neural networks' abuse by the authorities and the resistance to them with reliance on the trusted AI. Studying various aspects of digital transformation and the use of artificial intelligence technologies, the authors formalize the dangers associated with the emergence and propose an approach to the use of digital protection technologies that can be trusted.

Keywords: neural networks, digital protection, trust, artificial intelligence, society

Citation: Avdoshin S.M., Pesotskaya E.Yu. (2022) Trusted artificial intelligence: Strengthening digital protection. *Business Informatics*, vol. 16, no. 2, pp. 62–73. DOI: 10.17323/2587-814X.2022.2.62.73

References

1. Vogt T., Winter P., Nessler B., Doms T. (2021) *Trusted Artificial Intelligence: Towards Certification of Machine Learning Applications*. Vienna: TÜV Austria Holding AG.
2. Kuleshov A., Ignatiev A., Abramova A., Marshalko G. (2020) Addressing AI ethics through codification. *2020 International Conference Engineering Technologies and Computer Science (EnT)*, pp. 24–30. <https://doi.org/10.1109/EnT48576.2020.00011>
3. Harrison T., Luna-Reyes L. (2021) Cultivating trustworthy artificial intelligence in digital government. *Social Science Computer Review*, vol. 40, no. 2, pp. 494–511. <https://doi.org/10.1177/0894439320980122>
4. OECD (2014) *Recommendation of the council on digital government strategies*. Paris: OECD Publishing. Available at: <https://www.oecd.org/gov/digital-government/Recommendation-digital-government-strategies.pdf> (accessed 22 September 2021).
5. Chakraborty A., Alam M., Dey V., Chattopadhyay A., Mukhopadhyay D. (2018) *Adversarial attacks and defences: A survey*. Working paper arXiv: 1810.00069. <https://doi.org/10.48550/arXiv.1810.00069>
6. Haenlein M., Kaplan A. (2019) A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, vol. 61, no. 4, pp. 5–14. <https://doi.org/10.1177/0008125619864925>
7. Wei J. (2018) Research progress and application of computer artificial intelligence technology. *MATEC Web of Conferences*, vol. 176, Article Number 01043. <https://doi.org/10.1051/mateconf/201817601043>
8. Mijwil M., Esen A., Alsaadi A. (2019) *Overview of neural networks*. Available at: https://www.researchgate.net/publication/332655457_Overview_of_Neural_Networks (accessed 22 September 2021).
9. Sibai F.N. (2020) AI crimes: A classification. *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pp. 1–8. <https://doi.org/10.1109/CyberSecurity49315.2020.9138891>
10. Jastroch N. (2020) Trusted artificial intelligence: On the use of private data. In: *Product Lifecycle Management Enabling Smart X. PLM 2020* (eds. F. Nyffenegger, J. Ríos, L. Rivest, A. Bouras). IFIP Advances in Information and Communication Technology, vol. 594. https://doi.org/10.1007/978-3-030-62807-9_52
11. Nemitz P. (2018) Constitutional democracy and technology in the age of artificial intelligence. *Phil. Trans. R. Soc. A.*, vol. 376. <https://doi.org/10.1098/rsta.2018.0089>
12. Misra S.K., Das S., Gupta S., Sharma S.K. (2020) Public Policy and Regulatory Challenges of Artificial Intelligence (AI). In: *Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation. TDIT 2020* (eds. S.K. Sharma, Y.K. Dwivedi, B. Metri, N.P. Rana). IFIP Advances in Information and Communication Technology, vol. 617. https://doi.org/10.1007/978-3-030-64849-7_10
13. Pavlutenkova M. (2019) Electronic government vs digital government in context of digital transformation. *Monitoring of Public Opinion: Economic and Social Changes Journal*, no. 5, pp. 120–135 (in Russian). <https://doi.org/10.14515/monitoring.2019.5.07>
14. Kochetkov A.P., Vasilenko I.A., Volodenkov S.V., Gadzhiev K.S., Kovalenko V.I., Soloviev A.I., Kirsanova E.G. (2021) Political Project for Russia: Prospects for implementation in the context of challenges and risks of digitalization of society. *Vlast' (The Authority)*, vol. 29, no. 1, pp. 317–331 (in Russian). <https://doi.org/10.31171/vlast.v29i1.7963>
15. Williams M., Valayer C. (2018) *Digital government benchmark – study on digital government transformation*. European Union. Available at: <https://joinup.ec.europa.eu/collection/elise-european-location-interoperability-solutions-e-government/document/report-digital-government-benchmark-study-digital-government-transformation> (accessed 22 September 2021).
16. Carter D. (2020) Regulation and ethics in artificial intelligence and machine learning technologies: Where are we now? Who is responsible? Can the information professional play a role? *Business Information Review*, vol. 37, no. 2, pp. 60–68. <https://doi.org/10.1177/0266382120923962>
17. Kamat G. (2014) *Algorithms for private data analysis. Lecture 14 – Private ML and stats: Modern ML*. Available at: <http://www.gautamkamath.com/CS860notes/lec14.pdf> (accessed 22 September 2021).
18. Hinnefeld J., Cooman P., Mammo N., Deese R. (2018) *Evaluating fairness metrics in the presence of dataset bias*. Working paper arXiv: 1809.09245. <https://doi.org/10.48550/arXiv.1809.09245>

19. National Standard of the Russian Federation (2021) *Artificial intelligence systems. Methods for ensuring trust. General. GOST R 59276-2020* (in Russian).
20. Tinholt D., Carrara W., Linden N. (2017) *Unleashing the potential of Artificial Intelligence in the Public Sector*. Capgemini. Available at: <https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf> (accessed 22 September 2021).
21. Sharma G.D., Yadav A., Chopra R. (2020) Artificial intelligence and effective governance: A review, critique and research agenda. *Sustainable Futures*, vol. 2, Article ID 100004. <https://doi.org/10.1016/j.sfr.2019.100004>
22. McCormick T.R., Min D. (2020) *Principles of Bioethics*. University of Washington. Available at: <https://depts.washington.edu/bhdept/ethics-medicine/bioethics-topics/articles/principles-bioethics> (accessed 22 September 2021).
23. Lindgren I., Veenstra A.F. (2018) Digital government transformation: a case illustrating public e-service development as part of public sector transformation. *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, Delft, The Netherlands*, pp. 1–6.
24. IEEE (2021) *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2*. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Available at: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (accessed 22 September 2021).
25. Bradul N.V., Lebezova E.M. (2020) Conceptualization of Smart Government: A scientometric approach. *Upravlenets (The Manager)*, vol. 11, no. 3, pp. 33–45. <https://doi.org/10.29141/2218-5003-2020-11-3-3> (in Russian).
26. Thiebes S., Lins S., Sunyaev A. (2021) Trustworthy artificial intelligence. *Electron Markets*, vol. 31, pp. 447–464. <https://doi.org/10.1007/s12525-020-00441-4>
27. Falco G., Viswanathan A., Caldera C., Shrobe H. (2018) A master attack methodology for an AI-based automated attack planner for smart cities. *IEEE Access*, vol. 6, pp. 48360–48373. <https://doi.org/10.1109/ACCESS.2018.2867556>.
28. Bellamy R. et al. (2018) *AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias*. Working paper arXiv: 1810.01943. <https://doi.org/10.48550/arXiv.1810.01943>
29. Sarpatwar K. et al. (2019) Towards enabling trusted artificial intelligence via Blockchain. In: Calo, S., Bertino, E., Verma, D. (eds) *Policy-Based Autonomic Data Governance. Lecture Notes in Computer Science*, vol. 11550, pp. 137–153. Springer, Cham. https://doi.org/10.1007/978-3-030-17277-0_8
30. Montjoye Y.D., Farzanehfard A., Hendrickx J., Rocher L. (2017) Solving artificial intelligence’s privacy problem. *Field Actions Science Reports*, Special Issue 17, pp. 80–83. Available at: <https://journals.openedition.org/factsreports/pdf/4494> (accessed 22 September 2021).
31. DataCollaboratives.org (2021) *Open Algorithms (OPAL) Project*. Available at: <https://datacollaboratives.org/cases/open-algorithms-opal-project.html> (accessed 22 September 2021).
32. Salah K., Rehman M.H.U., Nizamuddin N., Al-Fuqaha A. (2019) Blockchain for AI: Review and open research challenges. *IEEE Access*, 2019, vol. 7, pp. 10127–10149. <https://doi.org/10.1109/ACCESS.2018.2890507>
33. Baker-Brunnbauer J. (2021) Management perspective of ethics in artificial intelligence. *AI and Ethics*, vol. 1, pp. 173–181. <https://doi.org/10.1007/s43681-020-00022-3>

About the authors

Sergey M. Avdoshin

Cand. Sci. (Tech.);

Professor, School of Computer Engineering, HSE Tikhonov Moscow Institute of Electronics and Mathematics (MIEM HSE), National Research University Higher School of Economics, 20, Myasnitskaya Street, Moscow 101000, Russia;

E-mail: savdoshin@hse.ru

ORCID: 0000-0001-8473-8077

Elena Yu. Pesotskaya

Cand. Sci. (Econ.);

Associate Professor, School of Software Engineering, Faculty of Computer Science, National Research University Higher School of Economics, 20, Myasnitskaya Street, Moscow 101000, Russia;

E-mail: epesotskaya@hse.ru

ORCID: 0000-0003-2129-4645

О назначении сроков службы технических систем в условиях инфляции

С.А. Смоляк 

E-mail: smolyak1@yandex.ru

Центральный экономико-математический институт Российской академии наук

Адрес: Россия, 117418, г. Москва, Нахимовский проспект, д. 47

Аннотация

Техническая система используется предприятием-участником рынка для выполнения определенных работ. В процессе работы операционные характеристики системы ухудшаются. Возможный отказ системы наносит ущерб предприятию, при этом она выводится из эксплуатации и утилизируется. Оказывается выгодным назначить такой технической системе срок службы, по истечении которого (если не произошло отказа) она подлежит утилизации. Решается задача оптимизации такого срока. Обычно при ее решении не учитывается инфляция, а критериями оптимальности выступают средние затраты в единицу времени и иные показатели, не в полной мере отражающие коммерческие интересы предприятия, владеющего системой. Использование принципов и методов стоимостной оценки позволяет построить экономико-математическую модель и предложить формулы, позволяющие с учетом инфляции рассчитать оптимальный назначенный срок службы системы и одновременно оценить рыночную стоимость работ, выполняемых технической системой, и изменение рыночной стоимости системы с возрастом. При этом в данной задаче критерием оптимальности оказывается отношение ожидаемых дисконтированных затрат к ожидаемому дисконтированному объему работ, выполняемых системой. Показано, что такой критерий позволяет максимизировать стоимость владеющего системой предприятия. Приводятся примеры применения построенной модели. Полученные результаты могут быть использованы как для решения других оптимизационных задач теории надежности, так и для практической стоимостной оценки некоторых видов машин и оборудования.

Ключевые слова: надежность технической системы, отказ, назначенный срок службы, критерий оптимальности, стоимостная оценка, рыночная стоимость

Цитирование: Смоляк С.А. О назначении сроков службы технических систем в условиях инфляции. 2022. Т. 16. № 2. С. 74–88. DOI: [10.17323/2587-814X.2022.2.74.88](https://doi.org/10.17323/2587-814X.2022.2.74.88)

Введение

Установление сроков службы машин и оборудования являлось, по крайней мере, с 20-го века, актуальной задачей как для экономистов, так и для технических специалистов. Обзор соответствующей литературы потребовал бы слишком много места, и мы его не приводим. Упомянем лишь, что еще в 1920–30-х годах был опубликован ряд статей по экономическим аспектам оптимизации сроков службы машин, не потерявших своей актуальности и сегодня. Характерно, однако, что в своих исследованиях экономисты обычно игнорировали вероятностный характер процесса эксплуатации машин, тогда как технические специалисты основное влияние уделяли вопросам их надежности и ремонта, игнорируя влияние физического износа на их производительность и (порой) на операционные затраты. При этом экономисты рассматривали приобретение и эксплуатацию машин как инвестиционный проект и опирались соответственно на теорию оценки эффективности инвестиционных проектов, тогда как технические специалисты при установлении назначенных сроков службы технических систем выбирали критерий оптимальности, по сути, произвольно, без должных обоснований (подробнее см. следующий раздел). Существенно, что почти во всех известных работах по данной проблеме влияние инфляции не учитывалось. Между тем, при практической стоимостной оценке реальных активов (зданий, сооружений, машин и оборудования) оценщики учитывают не только инфляцию, но и физический износ и надежность оцениваемых активов, а соответствующие стандарты оценки включают общие принципы такого учета.

В связи с этим представляется необходимым совмещение технических и экономических аспектов в единой оптимизационной модели, учитывающей как надежность, так и физический износ технических систем, и опирающейся, в отличие от известных экономико-математических моделей, на теорию стоимостной оценки. Целью настоящей статьи является построение подобной модели применительно к задачам назначения срока службы техническим системам, подвергающимся случайным отказам.

1. Постановка задачи. Основные понятия

Объектами рассмотрения в данной статье являются технические системы (ТС), обычно — это машины и оборудование, которые:

- ♦ используются предприятиями-участниками рынка;
- ♦ подвергаются случайным отказам;
- ♦ являются неремонтируемыми, т.е. не подлежат капитальному или среднему ремонту, а в случае отказа — выводятся из эксплуатации и утилизируются.

Предметами нашего рассмотрения будут назначенные сроки службы таких ТС и стоимости ТС разного возраста. Необходимые сведения о стоимостной оценке ТС и их надежности мы излагаем в соответствии со стандартами оценки [1] и книгой [2]. Основными операционными характеристиками ТС, зависящими от ее технического состояния, являются:

- ♦ (эксплуатационная) производительность (объем работ, производимых ТС в малую единицу времени);
- ♦ интенсивность операционных затрат (операционные затраты, необходимые для использования ТС по назначению в течение малой единицы времени);
- ♦ опасность отказа (вероятность отказа ТС в течение малой единицы времени).

В теории надежности рассматривают задачу установления назначенного срока службы неремонтируемой ТС. Мы займемся обоснованием критерия оптимальности такого срока. По этой причине мы ограничиваемся рассмотрением следующей достаточно простой ситуации.

Предприятие приобретает на рынке и использует неремонтируемую ТС определенного вида (например, машину определенной модели). В процессе работы ТС подвергается деградации, ее операционные характеристики ухудшаются, несмотря на проводимые техническое обслуживание и текущий ремонт (затраты на эти операции мы включаем в состав операционных затрат). Это положение нуждается в комментариях. Многие специалисты, как и в годы плановой экономики, считают, что такие характеристики машин и оборудования, как их производительность и годовые операционные затраты, должны в течение всего срока службы поддерживаться на неизменном (проектном) уровне. Неизменная производительность ТС принимается и в известных публикациях по проблемам оптимизации назначенных сроков службы и сроков плановых ремонтов ТС. Между тем, анализ работы машин и оборудования различного назначения показывает, что с возрастом

их характеристики имеют тенденцию к ухудшению, причем капитальный ремонт компенсирует такое ухудшение лишь частично (подробнее см. [3]). Часто это пытаются объяснить тем, что с возрастом растут только стоимость и продолжительность технического обслуживания и ремонтов. Однако имеющиеся данные свидетельствуют о том, что с возрастом снижается часовая производительность строительных машин, растут затраты топлива на 1 км пробега транспортных средств, у станков увеличивается длительность цикла обработки деталей.

По этой причине считать операционные характеристики ТС неизменными во времени представляется нам неправомерным. Далее мы предполагаем, что техническое состояние ТС определяется ее возрастом (наработкой), так что ее операционные характеристики являются известными функциями возраста.

Поскольку ТС свободно обращаются на рынке и обладают полезностью для предприятий-участников рынка, они являются объектами стоимостной оценки. Имеется много видов стоимости, но основным является рыночная стоимость (РС). Мы не будем давать подробного определения этому понятию, отсылая читателей к стандартам стоимостной оценки [1]. Отметим лишь, что РС объекта оценки на определенную дату (дату оценки) отражает цену этого объекта в совершаемой на эту дату сделке (реальной или гипотетической) между независимыми и ведущими себя экономически рационально типичными участниками рынка. При этом РС объекта отражает одновременно его полезность для типичных участников рынка и вклад объекта в РС владеющего объектом предприятия – типичного участника рынка.

Для стоимостной оценки объектов используются (по отдельности или в различных сочетаниях) три подхода:

- ♦ при *сравнительном* подходе стоимость объекта находят, исходя из цен сделок с аналогичными объектами, скорректированных на различия в характеристиках объектов, дате и условиях сделки;
- ♦ при *затратном* подходе стоимость объекта оценивают по затратам, необходимым для его создания (получения);
- ♦ при *доходном* подходе стоимость объекта находят, учитывая поток выгод, которые получит владелец от его использования.

Поскольку ТС является движимым имуществом, ее РС может быть различной в зависимости от того, оценивается ли она «на месте» или «для перемещения». Новые ТС обычно продаются производителями или дилерами на первичном рынке и являются точными аналогами друг друга. Разброс цен при этом небольшой, поэтому РС новой ТС «для перемещения к месту эксплуатации» находят, используя сравнительный подход, что не представляет особых сложностей. РС этой же ТС «на месте ее эксплуатации» выше, поскольку включает дополнительные затраты на транспортировку приобретенной ТС и ее монтаж. РС новой ТС «на месте» мы считаем известной величиной K .

Подержанные ТС продаются на вторичном рынке. Такие ТС не имеют точных аналогов, покупатели не имеют возможности оценить их техническое состояние, а продавцы не сообщают покупателям (или даже не знают) историю их эксплуатации. В результате цены машин/оборудования одной марки и одного возраста имеют очень большой разброс, а оценка подержанных ТС представляет для оценщиков значительные трудности.

В качестве аналога подержанной ТС удобно принять такую же новую ТС – мы будем называть ее новым аналогом. РС нового аналога ТС называют стоимостью воспроизводства оцениваемой ТС. Уменьшение РС подержанной ТС по сравнению со стоимостью ее воспроизводства называется обесценением (depreciation) оцениваемой ТС, а отношение этих величин, выраженное в долях единицы или в процентах, – коэффициентом/процентом годности или относительной стоимости (Percent Good Factor, PGF). В оценочной литературе предлагается ряд формул и таблиц, отражающих (не всегда корректно) зависимость коэффициента/процента годности от возраста ТС [3].

ТС, использовать которую по назначению далее невозможно или неэффективно, выводится из эксплуатации и утилизируется. Рыночная стоимость такой ТС называется утилизационной (УС). Обычно ее находят по данным о ценах машин, продаваемых «в лом» или рассчитывают как стоимость годных к дальнейшему использованию элементов ТС (включая и металлолом) за вычетом расходов на демонтаж и доставку элементов ТС к месту их дальнейшего использования. УС машин и оборудования невелика по сравнению со стоимостью их воспроизводства – отношение этих величин (относительная УС) для машин и оборудования обычно составляет от 0,03 до 0,09.

Методам стоимостной оценки машин и оборудования посвящены небольшие разделы в стандартах оценки и учебниках. Как правило, там не предусматривается учета ни ухудшения операционных характеристик ТС с возрастом, ни вероятностного характера срока их службы (исключение составляют методы, применяемые в системах национальных счетов [4], но они рассчитаны на оценку больших групп активов и неадекватно учитывают деградацию машин и оборудования [5]). В то же время достаточно адекватно учесть указанные факторы можно, опираясь на основные принципы стоимостной оценки и оценочную практику, используя при этом модели и методы теории надежности.

Одной из основных характеристик надежности ТС является опасность отказа. Мы будем считать ее зависящей от возраста ТС (t) и обозначим через $p(t)$. Эта величина имеет простой «физический смысл»: если ТС работала безотказно t лет, то вероятность ее отказа в интервале времени $(t, t + dt)$ составляет $p(t)dt$. При этом (случайный) момент наступления отказа имеет распределение с плотностью $p(t)e^{-P(t)}$, где $P(t) = \int_0^t p(x)dx$. Как известно (см., напр., [2, 6]), вероятность безотказной работы ТС в течение срока t составляет $e^{-P(t)}$, а среднее время безотказной работы (средняя наработка до отказа), если только оно специально не ограничивается, $-\int_0^\infty e^{-P(t)} dt$.

Существенно, что отказ ТС приводит к выходу из строя не только самой этой ТС, но и другого имущества предприятия. При этом предприятию нанесен ущерб, вообще говоря, случайный. Методы определения таких ущербов (потерь) изложены в соответствующей литературе, например [7]. Мы будем считать, что среднее значение ущерба известно, и в его величине учтена стоимость годных к дальнейшему использованию элементов отказавшей ТС и другого имущества предприятия.

Чтобы по возможности сократить потери от отказа ТС, ей назначают определенный срок службы (наработку) S , по достижении которого ТС должна выводиться из эксплуатации независимо от своего технического состояния. Мы хотим найти оптимальный такой срок S . От выбора S зависят и средние остаточные сроки службы (среднее время безотказной работы) ТС разного возраста.

Рассмотрим ТС, дожившую до возраста t . Пусть $T(t)$ — средний остаточный срок ее службы. Чтобы найти его, заметим, что остаточный срок службы ТС составит $x - t$, если она откажет в возрасте

$x < S$, и будет равен $S - t$, если за назначенный срок отказа не произойдет. Учитывая, что вероятность безотказной работы ТС, дожившей до возраста t , в течение всего назначенного срока равна $e^{P(S)-P(t)}$, а отказ в интервале времени $(t, t + dt)$, $t < S$, возможен с вероятностью $p(t)e^{P(S)-P(t)}dt$, мы получаем:

$$T(t) = \int_t^S (x-t)p(x)e^{P(t)-P(x)}dx + (S-t)e^{P(t)-P(S)} = \int_t^S e^{P(t)-P(x)}dx. \quad (1)$$

Для многих видов ТС принимается, что время безотказной работы имеет распределение Вейбулла, при котором опасность отказов $p(t)$ растет пропорционально некоторой степени возраста. Значения показателя степени для различных видов ТС рассчитывались рядом авторов, в том числе для целей национального счетоводства, а рекомендуемые значения для ряда объектов машиностроения приведены в [6]. Отметим, что для некоторых видов машин (морозильники, холодильники, пылесосы, микроволновые печи, видеоманитофоны, стиральные машины, отопительные электронагревательные приборы, небольшие легковые автомобили, оборудование для ремонта и технического обслуживания автомобилей, железнодорожные вагоны) этот показатель оказывается близким к единице [8–10]. Позднее в качестве примера мы рассмотрим именно такие ТС. У них время безотказной работы имеет распределение Рэлея, $p(t) = t/\omega^2$, где ω — параметр масштаба. При этом средний остаточный срок службы ТС возраста t составит

$$T(t) = e^{t^2/2\omega^2} \int_t^S e^{-x^2/2\omega^2} dx = \omega\sqrt{2\pi} \left[\Phi\left(\frac{S}{\omega}\right) - \Phi\left(\frac{t}{\omega}\right) \right] e^{t^2/2\omega^2}.$$

При $t = 0$ отсюда вытекает формула для среднего полного срока службы ТС:

$$T_m = T(0) = \omega\sqrt{2\pi} \left[\Phi\left(\frac{S}{\omega}\right) - \frac{1}{2} \right]. \quad (2)$$

При $S = \infty$, когда срок службы не назначается, среднее время безотказной работы ТС составит $\omega\sqrt{2\pi}$.

В теории надежности для решения оптимизационных задач обычно используются различные критерии, например, средние за срок службы количество или стоимость ремонтов, средние затраты на ремонт в единицу времени [11], средние затраты в

единицу времени за межремонтный цикл [12], отношение средних за срок службы затрат к среднему сроку службы [13–15], суммарные дисконтированные затраты или эквивалентный им аннуитет [16, 17]. Указанным критериям свойственен ряд общих недостатков.

1. Выбор критерия носит формально-математический характер и ориентирован на использование измеряемых характеристик системы (прежде всего, затратных и временных). Интересы конкретного бизнеса, использующего ТС, при этом не учитываются [12].

2. Сравнивать варианты использования ТС по затратам корректно, только если они дают тождественные результаты [18]. Если же производительность ТС меняется с возрастом, то варианты с разными назначенными сроками службы будут различаться по величине и распределению во времени получаемых результатов — объемов выполняемых работ.

3. При исчислении таких показателей, как суммарные или средние за срок службы затраты, не учитывается влияние инфляции. Это можно было бы оправдать, если бы речь шла о ТС с коротким (до 1–2 лет) сроком службы, однако эти сроки в строительном, машиностроительном и транспортном секторах экономики превышают 5–15 лет, и для них учет инфляции необходим.

Учитывая изложенное, для решения поставленной задачи мы считаем более адекватным использовать общие принципы стоимостной оценки, ориентированные на максимизацию рыночной стоимости предприятия.

При этом, в отличие от работ по теории надежности, оказывается возможным учесть не только инфляцию, но и влияние деградации ТС на ее производительность¹ и операционные затраты, а также утилизационную стоимость ТС. Особого рассмотрения при этом заслуживает учет инфляции, чему будет посвящен следующий раздел.

2. Групповая инфляция

Выбор экономически рациональных решений по управлению техническими системами в условиях инфляции связан со значительными трудностями. Дело в том, что для этого необходим прогноз экономических характеристик системы на весь срок

ее службы, что предполагает, как минимум, прогнозирование цен на производимые системами продукты (товары, работы или услуги), а также на различные потребляемые при этом ресурсы. К сожалению, заниматься разработкой таких прогнозов специалисты по управлению техническими системами не в состоянии. Поэтому в математических моделях оптимизации технических решений, ориентированных на практическое применение, инфляция обычно не учитывается. Это относится и к рассматриваемой нами задаче назначения сроков службы ТС. Между тем, выяснить, как влияет инфляция на решение этой задачи, можно, опираясь на теорию стоимостной оценки и практику оценочной деятельности. Предлагается учитывать только наиболее важные для стоимостной оценки и поддающиеся измерению характеристики инфляции, пренебрегая всеми остальными.

Анализ цен первичного и вторичного рынка и опыт оценки подержанных машин и оборудования показывают, что цены подержанных машин обычно меняются синхронно с ценами аналогичных новых. Это обусловлено тем, что ведущий себя экономически рационально покупатель подержанной машины на вторичном рынке сопоставляет намечаемую покупку с альтернативой приобретения ее нового аналога на первичном рынке.

В связи с этим естественным представляется допущение о том, что в условиях, когда растут рыночные стоимости новых ТС некоторого вида, РС подержанных ТС этого вида растут в той же пропорции. Такое явление названо групповой инфляцией [3]. Однако дать ему строгое определение не просто, поскольку все подержанные ТС одного вида, в отличие от новых, не имеют аналогов, находятся в разном техническом состоянии, их количество и состав на вторичном рынке все время меняется, так что само понятие о росте стоимостей ТС на вторичном рынке становится неопределенным.

Тем не менее, суть групповой инфляции можно пояснить условным примером, относящимся к группе машин одной модели, техническое состояние которых со временем изменяется. Представим себе рынок, на котором в некоторый момент времени T_1 представлена совокупность C_1 из M новых и подержанных машин этой группы, находящихся в разных состояниях, причем стоимости всех этих машин известны. Допустим теперь, что в более

¹ В работах по проблеме оптимизации назначенных сроков службы и сроков плановых ремонтов ТС их производительность обычно считается постоянной.

поздний момент T_2 на рынке также оказалась совокупность C_2 из M машин группы, причем каждая i -я из них находится точно в таком же техническом состоянии, в каком была i -я машина из совокупности C_1 . Иными словами, в этой ситуации все машины группы, представленные на рынке в момент T_1 , как бы перенесли на более позднюю дату без изменения их технического состояния. При этом за счет инфляции изменятся рыночные стоимости всех машин. Однако при групповой инфляции они изменятся пропорционально, и окажется, что отношение стоимости подержанной машины к стоимости такой же новой (коэффициент годности) будет зависеть только от состояния этой машины, но не от того, оценена ли она в момент T_1 или в момент T_2 . Это условие мы и положим в основу определения самого понятия групповой инфляции.

Будем говорить, что для ТС определенного вида в некотором интервале времени имеет место групповая инфляция, если коэффициенты годности этих ТС зависят только от состояния ТС, но не от даты оценки этих ТС (в указанном интервале времени). Темп групповой инфляции при этом определяется как темп роста РС новых ТС. Поэтому для анализа и краткосрочного прогноза этого темпа достаточно использовать открытую и доступную информацию о ценах производителей или дилеров в периоде, близком к дате оценки.

Допущение о наличии групповой инфляции существенно упрощает решение задачи оптимизации назначенного срока службы ТС, что мы увидим в следующем разделе. Однако правомерность этого допущения подтверждается и другими доводами.

Обычно техническое состояние продаваемых или проданных подержанных машин оценщикам неизвестно², они знают только их возраст. Поэтому они вынуждены (как и в этой статье) характеризовать состояние машины ее возрастом. Тогда для приближенного расчета коэффициентов годности они формируют выборку сложившихся в некотором базовом периоде (например, в текущем году) рыночных цен машин разного возраста и на ее основе строят регрессионную зависимость $F(t)$ цен машин от их возраста (t). При этом значение $F(t)$ отразит

среднюю РС машин возраста t , которые были (или могли бы быть³) представлены на рынке в базовом периоде, а $F(0)$ – РС новых машин в этом периоде. Коэффициенты годности при этом находятся по формуле: $k(t) = F(t)/F(0)$. При отсутствии групповой инфляции подобные зависимости, построенные для машин той же марки по данным разных лет, могли бы существенно различаться (это означало бы, например, что под влиянием инфляции цены более старых по возрасту машин снижаются или растут медленнее, чем цены более молодых). Однако такого явления обычно не наблюдается.

Наоборот, в условиях групповой инфляции построенная функция $k(t)$ не зависит от того, к какому периоду относятся цены машин. Да и российские оценщики, по существу, принимают допущение о групповой инфляции, используя зависимости коэффициента годности от возраста, построенные для групп машин разного назначения по рыночным данным за давно прошедшие годы. Примерно такая же ситуация сложилась в некоторых штатах США, где машины и оборудование являются объектами налогообложения. Там ежегодно строятся регрессионные зависимости коэффициентов годности от возраста, правда, применительно к довольно широким группам машин (например, сельскохозяйственные или строительные машины и оборудование). Анализ соответствующих публикаций (например, [19–21]) показывает, что эти зависимости из года в год меняются незначительно, что так же свидетельствует о групповой инфляции.

Важно отметить, что допущение о групповой инфляции позволяет при построении зависимости $k(t)$ включать в выборку цены машин, сложившиеся в разные годы (одновременно рассчитывая при этом и РС новых машин в эти годы), что дает возможность существенно увеличить объем выборки и повысить точность рассчитанной зависимости $k(t)$.

Разумеется, в иных ситуациях групповую инфляцию можно понимать и иначе, используя вместо возраста наработку или иные поддающиеся объективному измерению характеристики. В таких случаях в приводимую далее модель необходимо будет внести соответствующие корректировки.

² Казалось бы, количественные оценки технического состояния машин и оборудования дают автоматизированные системы диагностики. Однако эти системы не оценивают ни опасности отказа машины, ни других важных для участников рынка ее характеристик (например, состояния рамы грузового автомобиля или кузова легкового автомобиля, существенно влияющих на их рыночную стоимость).

³ Такая оговорка связана с тем, что построенная зависимость позволяет вычислять $F(t)$ для машин любого возраста t , тогда как в базовом периоде машины некоторых возрастов на рынке могли не выставляться.

3. Оптимизационная модель

Для оптимизации назначенного срока службы ТС рассматриваемого вида необходимо задать определенный критерий оптимальности. С этой целью мы будем решать другую задачу, на первый взгляд, достаточно далекую от теории надежности, а именно, будем оценивать рыночные стоимости ТС разного возраста «на месте». В качестве даты оценки мы примем момент принятия решения о назначении срока службы ТС. Кроме того, примем, что для ТС данного вида в течение небольшого периода, близкого к дате оценки, имеет место групповая инфляция с известным темпом i .

Будем считать известными также следующие характеристики ТС: РС новой ТС K , утилизационную стоимость U , средний (ожидаемый) ущерб от отказа ТС L , опасность $p(t)$ отказа исправной ТС возраста t , ее эксплуатационную производительность $Q(t)$ и интенсивность $C(t)$ ее операционных затрат⁴.

Назначенный срок службы ТС, подлежащий оптимизации, обозначим через S , а РС исправной ТС возраста t «на месте» — через $V(t)$.

Заметим, что работа, выполняемая ТС, обладает полезностью для участников рынка и измеряется в определенных физических единицах (кубометрах перемещенного грунта, декалитрах разлитой жидкости, количестве условных банок консервов и т.п.). Раз так, то она, согласно стандартам оценки [1], имеет и свою рыночную стоимость. Другое дело, что владельцы машин ее обычно не знают, да и профессиональные оценщики редко оценивают РС работ (за исключением, пожалуй, строительно-монтажных и ремонтных). Неизвестную РС единицы выполняемых ТС работ на дату оценки обозначим через B .

В общем случае выгоды от использования объекта за период определяются как РС выполняемой ею за этот период работы за вычетом понесенных при этом затрат. В зависимости от способа измерения стоимостей и затрат выделяют разные виды выгод:

- ♦ если стоимости и затраты измеряют в ценах соответствующего периода (т.е. с учетом инфляции), выгоды именуют *номинальными*, если же для

этого используют цены на некоторую фиксированную дату — *реальными*;

- ♦ если в составе затрат учитывают налог на прибыль, выгоды именуют *посленалоговыми*, в противном случае — *доналоговыми*.

В данной статье выгоды от использования ТС считаются номинальными и доналоговыми. В таком случае интенсивность выгод от использования исправной ТС возраста t (т.е. размер выгод, приносимых ею за малую единицу времени) составит на дату оценки $BQ(t) - C(t)$.

Для оценки ТС воспользуемся принципом ожидания выгод, лежащим в основе доходного подхода к оценке имущества и упоминаемом (но не раскрываемом) в стандартах оценки. Приведем его наиболее подробную формулировку, следуя [3].

РС объекта на дату оценки равна ожидаемой сумме дисконтированных к этой дате выгод от его использования в последующем прогнозном периоде (включая и РС объекта в конце периода), если объект используется наиболее эффективно, и не меньше ее в противном случае.

К этой формулировке необходимо сделать ряд комментариев.

1. Длительность прогнозного периода может быть выбрана произвольно, поскольку РС объекта от этого выбора не зависит.

2. Включение стоимости объекта в конце периода в состав суммарных за прогнозный период выгод может пониматься и как выгода от продажи объекта по рыночной стоимости. Тем самым, продажа объекта рассматривается как один из способов его использования.

3. В условиях вероятностной неопределенности термин «ожидаемый» понимается как «математическое ожидание» (в [1] — «взвешенное по вероятностям»), а риски, связанные с неопределенностью выгод, в ставке дисконтирования не учитываются — такую ставку назовем «безрисковой».

4. Вид ставки дисконтирования определяется видом выгод: номинальные (реальные) выгоды должны дисконтироваться по номинальной (ре-

⁴ Строго говоря, производительность и операционные затраты ТС зависят от колебания спроса на продукцию предприятий, использующих ТС, и, следовательно, от потребности в выполняемых ТС работах. В нашей модели под $Q(t)$ и $C(t)$ понимаются средние (по типичным предприятиям — владельцам ТС) значения соответствующих характеристик, отвечающие среднему режиму использования исправных ТС возраста t .

альной) ставке, а посленалоговые (доналоговые) выгоды — по после-(до-)налоговой ставке⁵. Поэтому в данной статье при оценке ТС мы используем номинальную доналоговую безрисковую ставку дисконтирования, обозначаемую через r .

5. Стандарты оценки [1] требуют при оценке любого вида стоимости объекта указывать и предпосылку о способе его использования. Здесь и далее способ использования объекта предполагается наиболее эффективным, так что соответствующая предпосылка опускается.

Переходя к определению неизвестных стоимостей $V(t)$, вначале заметим, что РС новой ТС — известная величина $V(0) = K$. Кроме того, ТС, достигшая назначенного срока службы S , утилизируется, так что ее РС $V(S) = U$.

Рассмотрим теперь ТС, достигшую возраста $t < S$. Чтобы оценить ее РС $V(t)$ на дату оценки, применим к этой ТС принцип ожидания выгод, выбрав в качестве прогнозного малый период времени длительности dt . При этом учтем, что в нашей модели все возможные способы использования ТС отличаются только назначенным сроком службы S .

Если за время dt произойдет отказ, что возможно с вероятностью $p(t)dt$, то предприятию наносится ущерб, составляющий в среднем L , а ТС будет выведена из эксплуатации. В противном же случае ТС принесет выгоды $[BQ(t) - C(t)]dt$, а в конце периода ее возраст станет равным $t + dt$. ТС такого возраста на дату оценки имеет РС $V(t + dt)$, однако при групповой инфляции с темпом i за время dt она вырастет в $1 + i dt$ раз и станет равной $(1 + i dt) V(t + dt)$. Учитывая вероятности обоих случаев, отсюда получаем:

$$\begin{aligned} V(t) \geq & p(t)dt \cdot (-L) + \\ & + [1 - p(t)dt] \{ [BQ(t) - C(t)]dt + \\ & + (1 - rdt)(1 + idt)V(t + dt) \}, \end{aligned} \quad (3)$$

причем равенство здесь достигается при наиболее эффективном использовании ТС, т.е. при оптимальном значении S .

Введем теперь в рассмотрение «реальную»⁶ ставку дисконтирования:

$$\rho = r - i. \quad (4)$$

Используя ее, после простых преобразований неравенства (3) найдем:

$$\begin{aligned} BQ(t) - C(t) - \rho U - (L + U)p(t) \leq \\ \leq [\rho + p(t)][V(t) - U] - [V(t) - U]'. \end{aligned}$$

Умножим это неравенство на $e^{-\rho t - P(t)}$ и проинтегрируем по t от $t = s$ до $t = S$. Учитывая, что $V(S) = U$, мы получим:

$$\begin{aligned} \int_s^S [BQ(t) - C(t) - \rho U - (L + U)p(t)] e^{-\rho t - P(t)} ds \leq \\ \leq [V(s) - U] e^{-\rho s - P(s)}, \end{aligned} \quad (5)$$

или, что тоже самое:

$$V(s) \geq U + BQ_\Sigma(s, S) - C_\Sigma(s, S), \quad (6)$$

где обозначено:

$$\begin{cases} Q_\Sigma(s, S) = \int_s^S Q(t) e^{-\rho(t-s) - P(t) + P(s)} dt, \\ C_\Sigma(s, S) = \int_s^S [C(t) + \rho U + (L + U)p(t)] e^{-\rho(t-s) - P(t) + P(s)} dt. \end{cases} \quad (7)$$

Подставив $s = 0$, $V(0) = K$ в формулу (6) и выразив из нее B , мы получим:

$$B \leq \frac{K - U + C_\Sigma(0, S)}{Q_\Sigma(0, S)} = Z. \quad (8)$$

При этом равенства в (5), (6) и (8) достигаются при оптимальном S .

Отсюда следует, что при оптимальном назначенном сроке службы достигается минимум показателя Z , имеющего вид дроби. Выясним экономический смысл ее числителя и знаменателя, используя формулы (7).

На первый взгляд, он очевиден. В знаменателе дроби стоит ожидаемый дисконтированный объем выполняемых ТС работ за весь назначенный срок службы, а в числителе — ожидаемые дисконтированные затраты на ее приобретение и использова-

⁵ Ставки дисконтирования оценщики определяют по данным о наблюдаемых на рынке номинальных доналоговых доходностях финансовых инструментов.

⁶ Термин «реальная» взят в кавычки, поскольку в формуле (4) номинальная ставка уменьшается на темп групповой инфляции, а не на темп общей инфляции в стране, как это требуется при исчислении «настоящей» реальной ставки.

ние (за вычетом утилизационной стоимости ТС)⁷, включая и ущерб от отказов. Поэтому дробь Z отражает ожидаемые удельные (на единицу работ) дисконтированные затраты (ОУДЗ).

На самом деле выделенный термин не вполне корректен, поскольку входящие в формулы (7) величины $Q(t)$, $C(t)$, $p(t)$ отражают не динамику характеристик исправной ТС на протяжении срока ее службы, а характеристики исправных ТС разных возрастов на дату оценки. И лишь тогда, когда на протяжении всего срока службы ТС все цены в стране растут с одним и тем же темпом i , числитель и знаменатель дроби Z могут трактоваться как ожидаемые дисконтированные затраты и результаты, связанные с приобретением и использованием ТС до конца срока ее службы. Тем не менее, введенный термин и его сокращение ОУДЗ представляются удобными и наглядными, и мы будем их использовать.

Как видно из формулы (8), оптимальному назначенному сроку службы должно отвечать минимальное значение ОУДЗ. Более того, это значение будет равно рыночной стоимости единицы выполняемых объектом работ B . Такой метод оценки РС работ согласуется с затратным подходом к оценке, хотя оценщики в таком виде его не применяют.

Мы видим также, что применение критерия ОУДЗ ориентирует предприятие на наиболее эффективное использование ТС и максимизацию своей рыночной стоимости⁸. Аналогичный критерий удельных дисконтированных затрат для установления сроков службы машин в детерминированной ситуации ранее обосновывался с помощью оптимизационных моделей (см., например, [18, 19]). Близкие по содержанию критерии использовались и при решении некоторых оптимизационных задач теории надежности (см., например, [23, 24]), однако при этом производительность и интенсивность операционных затрат ТС предполагались постоянными, а ее утилизационная стоимость не учитывалась. В отличие от критерия отношения средних затрат к среднему сроку службы, часто используемого в работах по теории надежности, в ОУДЗ учитывается разновременность затрат и результатов работы ТС, но не учитывается время на ликвидацию последствий отказа.

Вычисление оптимального значения S и стоимости единицы работ B можно упростить, воспользовавшись неравенством (5). Поскольку с ростом t подынтегральное выражение в (5) убывает, нетрудно убедиться, что максимум в левой части (5) достигается, когда S является единственным корнем уравнения: $BQ(S) - C(S) - \rho U - (L + U)p(S) = 0$. В этом случае неравенство (8) становится равенством, а S и B будут решением системы уравнений:

$$B = \frac{K - U + C_\Sigma(S)}{Q_\Sigma(S)} = \frac{C(S) + \rho U + (L + U)p(S)}{Q(S)}. \quad (9)$$

Зависимость стоимости ТС от возраста теперь вытекает из неравенства (6), которое при оптимальном S становится равенством:

$$V(s) = U + BQ_\Sigma(s, S) - C_\Sigma(s, S). \quad (10)$$

Подставив в (10) B из первого равенства (9), мы получим:

$$V(s) = U + (K - U) \frac{Q_\Sigma(s, S)}{Q_\Sigma(S)} + \left[C_\Sigma(S) \frac{Q_\Sigma(s, S)}{Q_\Sigma(S)} - C_\Sigma(s, S) \right]. \quad (11)$$

Это равенство, по сути, является одной из модификаций известной формулы Львова [25], применявшейся в СССР для оценки эффективности новой техники и установления цен на новую технику, а в настоящее время применяемой при оценке машин и оборудования (об истории формулы Львова и ее модификациях см. [3, 26, 27]). В отличие от других модификаций формулы Львова, формула (11) учитывает утилизационную стоимость ТС, риск ее отказа и применяется к оценке подержанной ТС.

К числу характеристик машин и оборудования, используемых при их оценке, относятся средние полный и остаточный сроки службы. Их можно рассчитать и для рассматриваемых ТС. Казалось бы, в этих целях можно использовать формулы (1) и (2), связывающие указанные сроки с назначенным сроком службы (S). Однако это будет не вполне корректно, поскольку при выводе формулы (1) предполагалось, что до момента утилизации ТС назначенный срок ее службы не меняется. Такое до-

⁷ Поскольку дисконтируемые производительность и затраты относятся к исправной (не отказавшей) ТС, для дисконтирования здесь, по существу, используется «реальная» и учитывающая зависящий от возраста ТС риск отказа ставка $\rho + p(s)$.

⁸ Рыночная стоимость предприятия при предположении, что оно назначит ТС иной срок службы, уменьшится.

пушение было бы оправдано, если бы, начиная с даты оценки, для ТС имела бы место групповая инфляция с неизменным темпом. Однако сроки службы ТС могут составлять десятки лет, при этом темпы роста цен на машины могут сильно колебаться. Но тогда будут меняться и оптимальные значения назначенных сроков службы ТС. В таких условиях рассчитанные по формулам (1) и (2) величины $T(t)$ и T_m уже нельзя назвать средними остаточным и полным сроками службы ТС. Скорее, они отражают такие сроки, исчисленные при допущении о групповом характере и неизменном темпе инфляции на протяжении всего назначенного срока службы ТС. Впрочем, подобную оговорку необходимо делать и в отношении других технико-экономических норм (например, нормативов периодичности ремонтов или коэффициентов запаса прочности), поскольку все они разрабатывались применительно к конкретным экономическим условиям и потому будут экономическими обоснованными, пока эти условия существенно не изменятся.

4. Экспериментальные расчеты

По построенной модели были проведены экспериментальные расчеты в нескольких вариантах. Время измерялось в годах и долях года. Во всех вариантах принималось $V(0) = K = 100$, $Q(0) = 1$. Рыночные стоимости ТС разного возраста при этом численно совпадают с процентами их годности (PGF). Принималось, что зависимости $C(t)$ и $Q(t)$ – линейные, а время безотказной работы ТС имеет распределение Рэлея:

$$Q(t) = 1 - \alpha t, \quad C(t) = C_0(1 + \beta t), \quad p(t) = t/\omega^2,$$

где α и β – (базисные) темпы соответственно снижения производительности и роста операционных затрат, 1/год;

ω – параметр распределения Рэлея, годы.

Оптимальные назначенные сроки службы ТС (S) и РС единицы выполняемых ТС работ на дату оценки (B) находились путем численного решения системы уравнений (9).

Базовым был принят вариант со следующими значениями параметров:

$$U = 7; C_0 = 100; L = 100; \rho = 0,1; \\ \alpha = 0,01; \beta = 0,02; \omega = 8.$$

Варьируя параметрами этого варианта, можно выявить их влияние на оптимальное значение S назначенного срока службы ТС (рис. 1–4).

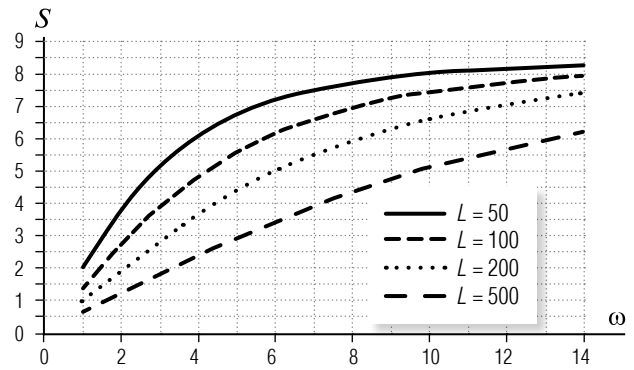


Рис. 1. Влияние ω на оптимальный назначенный срок службы при разных L .

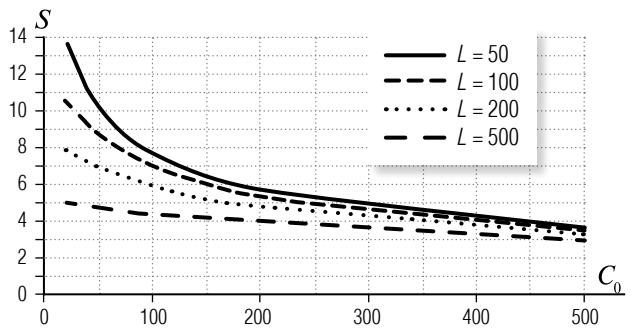


Рис. 2. Влияние C_0 на оптимальный назначенный срок службы при разных L .

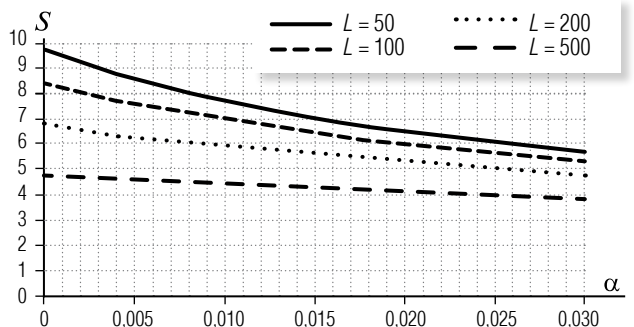


Рис. 3. Влияние α на оптимальный назначенный срок службы при разных L .

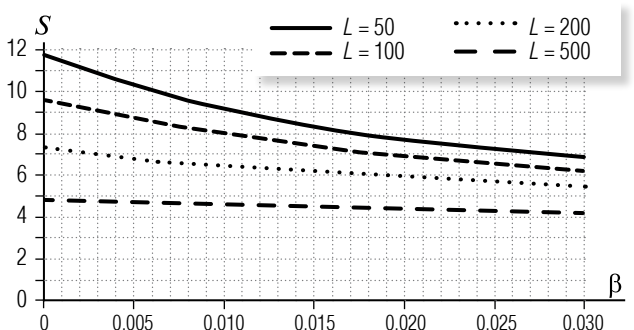


Рис. 4. Влияние β на оптимальный назначенный срок службы при разных L .

Отметим также, что оптимальное S немного уменьшается с ростом УС U и растет с увеличением ставки дисконтирования (особенно сильно – при малых значениях ущерба L).

Рассчитанные по формуле (2) зависимости среднего срока службы от интенсивности операционных затрат в начале эксплуатации (C_0) и ω при разных значениях ущерба представлены на рис. 5–6.

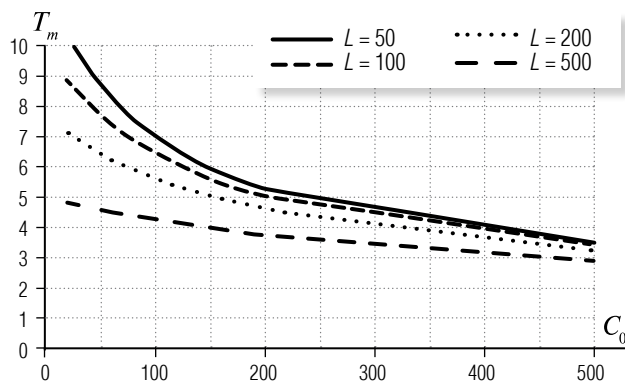


Рис. 5. Влияние C_0 на средний срок службы ТС при разных L .

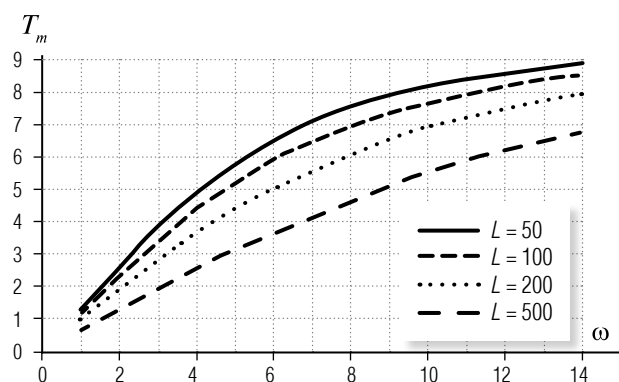


Рис. 6. Влияние ω на средний срок службы ТС при разных L .

Как видим, у новых ТС, требующих относительно больших (по сравнению со своей стоимостью) операционных затрат, средние сроки службы оказываются меньше. С уменьшением интенсивности отказов (ростом ω) средние сроки службы ТС растут, но только до определенного предела («экономического» срока службы аналогичных ТС, не подвергающихся отказам).

Средние значения РС подержанных ТС как функции от возраста, можно рассчитать по формуле (10). Правда, соответствующие графики трудно соп-

ставлять, поскольку средние и назначенные сроки службы по вариантам существенно различаются. Однако положение меняется, если на базе этих расчетов построить зависимости средних процентов годности от относительного возраста (τ) – отношения возраста ТС (t) к среднему сроку их службы (T_m).

Четыре варианта таких зависимостей представлены на рис. 7. Во всех этих вариантах принималось $K = 100$; $U = 7$; $Q_0 = 1$; $\rho = 0,1$; $\alpha = \beta = 0,01$. Значения других параметров ТС, а также назначенных и средних сроков службы для этих вариантов сведены в таблице 1.

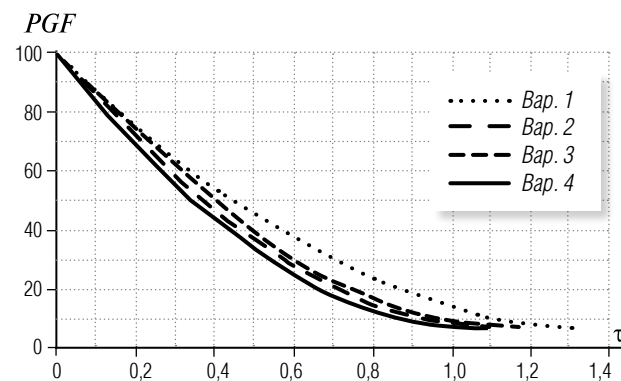


Рис. 7. Зависимости среднего процента годности ТС (PGF) от относительного возраста (τ) по вариантам расчета.

Таблица 1.

Основные параметры ТС по вариантам расчета

Вариант	1	2	3	4
C_0	20	100	40	300
L	100	200	200	500
ω	10	10	5	5
S	13,36	7,44	4,94	2,78
T_m	10,26	6,80	4,24	2,64

Обратим внимание, что, несмотря на существенные различия в основных параметрах ТС, соответствующие графики оказываются достаточно близкими. В то же время они дают меньшие значения процентов годности, чем при использовании гиперболической модели, принятой в российской системе национальных счетов для оценки машин и оборудования.

Заключение

Традиционный подход к установлению назначенных сроков службы объектов не в полной мере отвечает интересам участников рынка. Для решения этой задачи предложено ориентироваться на максимизацию рыночной стоимости предприятия — владельца объекта и базироваться на теории стоимостной оценки. Оказывается, что при таком подходе появляется возможность достаточно простого учета инфляции, а назначенный срок службы объекта должен обеспечивать минимальные ожидаемые удельные дисконтированные затраты — отношение ожидаемых дисконтированных затрат на

приобретение и использование ТС и устранение последствий ее возможных отказов к ожидаемому дисконтированному объему работ, выполненных ТС. Одновременно построенная экономико-математическая модель позволяет оценить рыночную стоимость выполняемых ТС работ и установить зависимость рыночной стоимости ТС от ее возраста. Предложенный подход может быть использован и в ситуации, когда деградация исправной ТС описывается случайными процессами, а также при решении других оптимизационных задач теории надежности, например, для оптимизации графика капитальных ремонтов. ■

Литература

1. Международные стандарты оценки / пер. с англ. под ред. И.Л. Артеменкова, С.А. Табаковой. М.: Саморегулируемая общероссийская общественная организация «Российское общество оценщиков», 2020.
2. Гнеденко Б.В., Беляев Ю.К., Соловьев А.Д. Математические методы в теории надежности. М.: Наука, 1965.
3. Смоляк С.А. Стоимостная оценка машин и оборудования (секреты метода ДДП). М.: Опцион, 2016.
4. System of National Account 2008 / European Commission, International Monetary Fund, Organization for Economic Co-operation and Development, United Nations, World Bank. New York, 2009.
5. Смоляк С.А. Оценка обесценения машин с учетом положений систем национальных счетов // Труды ИСА РАН. 2021. Т. 71. № 1. С. 44–54. <https://doi.org/10.14357/20790279210105>
6. Острейковский В.А. Теория надежности: Учебник для вузов. М.: Высшая школа, 2003.
7. Методические рекомендации по оценке ущерба от аварий на опасных производственных объектах (РД 03-496—02). Серия 03. Выпуск 19. М.: Закрытое акционерное общество «Научно-технический центр исследований проблем промышленной безопасности», 2010.
8. Lutz J.D., Hopkins A., Letschert V., Franco V.H., Sturges A. Using national survey data to estimate lifetimes of residential appliances // HVAC&R Research. 2011. Vol. 17. No. 5. P. 726–736. <https://doi.org/10.1080/10789669.2011.558166>
9. Erumban A.A. Lifetimes of machinery and equipment evidence from Dutch manufacturing // Review of Income and Wealth. 2008. Vol. 54. No. 2. P. 237–268. <https://doi.org/10.1111/j.1475-4991.2008.00272.x>
10. Nomura K., Suga Y. Measurement of depreciation rates using microdata from disposal survey of Japan // The 35th IARIW General Conference, Copenhagen, Denmark, 2018.
11. Алексеев В.В., Хоменко И.В., Прохорский Р.А. Модели планирования ремонтов и замен элементов в процессе жизненного цикла сложных технических систем // Вестник Воронежского института МВД России. 2011. № 3. С. 94–102.
12. Van Horenbeek A., Pintelon L., Muchiri P. Maintenance optimization models and criteria // International Journal of System Assurance Engineering and Management. 2010. Vol. 1. No. 3. P. 189–200.
13. Barlow R., Hunter L. Optimum preventive maintenance policies // Operations Research. 1960. Vol. 8. No. 1. P. 90–100. <https://doi.org/10.1287/opre.8.1.90>
14. Jiang R. Performance evaluation of seven optimization models of age replacement policy // Reliability Engineering & System Safety. 2018. Vol. 180. pp. 302–311. <https://doi.org/10.1016/j.res.2018.07.030>
15. Smith D.J. Reliability, maintainability and risk: Practical methods for engineers. Elsevier Ltd, Eighth edition, 2011.
16. Aven T. Optimal replacement under a minimal repair strategy: A general failure model // Advances in Applied Probability. 1983. Vol. 15. No. 1. P. 198–211. <https://doi.org/10.2307/1426990>
17. Weersink A., Stauber S. Optimal replacement interval and depreciation method for a grain combine // Western Journal of Agricultural Economics. 1988. Vol. 13. No. 1. P. 18–28. <https://doi.org/10.22004/ag.econ.32156>
18. Виленский П.Л., Лившиц В.Н., Смоляк С.А. Оценка эффективности инвестиционных проектов: Теория и практика. М.: ПолиПринтСервис, 2015.
19. 2021 Personal Property Factors & Tables / State of Colorado. The Department of Local Affairs. [Электронный ресурс]: <https://drive.google.com/file/d/1FZI9Da3QebSCDSVTPpZEX0H-6t0c9MNd/view> (дата обращения 22.05.2022).
20. 2021 Recommended Personal Property Valuation Schedule / Utah State Tax Commission. Property Tax Division. [Электронный ресурс]: https://propertytax.utah.gov/personal-property/val_schedule_2021.pdf (дата обращения 22.05.2022).

21. 2022 Personal Property Depreciation Schedules and Trend Tables / Montana Department of Revenue. [Электронный ресурс]: <https://mtrevenue.gov/?mdocs-file=59493> (дата обращения 22.05.2022).
22. Лившиц В.Н., Смоляк С.А. Модели динамики экономического износа оборудования // Экономика и математические методы. 1990. Т. 26. Вып. 5. С. 871–882.
23. Christer A.H., Goodbody W. Equipment replacement is unsteady economy // Journal of the Operational Research Society. 1980. Vol. 31. P. 497–506.
24. Dohi T., Ashioka A., Osaki S., Kaio N. Optimizing the repair-time limit replacement schedule with discounting and imperfect repair // Journal of Quality in Maintenance Engineering. 2001. Vol. 7. No. 1. P. 71–84. <https://doi.org/10.1108/13552510110386973>
25. Львов Д.С. Экономические проблемы повышения качества промышленной продукции. М.: Наука, 1969.
26. Смоляк С.А. О формуле Д.С. Львова для оценки машин // Имущественные отношения в Российской Федерации. 2018. Т. 205. № 10. С. 19–25. <https://doi.org/10.24411/2072-4098-2018-10101>
27. Смоляк С.А. Обобщенная формула Д.С. Львова для машин, подвергающихся деградации // Труды ИСА РАН. 2020. Т. 3. С. 3–12. <https://doi.org/10.14357/20790279200301>

Об авторе

Смоляк Сергей Абрамович

доктор экономических наук;

кандидат физико-математических наук;

главный научный сотрудник, Центральный экономико-математический институт, Российская академия наук, 117418, г. Москва, Нахимовский проспект, д. 47;

E-mail: smolyak1@yandex.ru

ORCID: 0000-0001-5287-4285

On assigning service life for technical systems under inflation

Sergey A. Smolyak

E-mail: smolyak1@yandex.ru

Central Economics and Mathematics Institute, Russian Academy of Science

Address: 47, Nakhimovsky Prospect, Moscow 117418, Russia

Abstract

A technical system is used by an enterprise that is a typical market participant to perform specific work. During operation, the operating characteristics of the system deteriorate. In case of a possible failure of the system, it is decommissioned and this causes losses for the enterprise. It turns out that it is beneficial to assign a certain service life to the system after which (if no failure has occurred) it is subject to decommissioning. We are solving the problem of optimizing this assigned service life. Usually, when solving it, inflation is not taken into account, and the optimality criteria are the average costs per unit of time and other indicators that do not fully reflect the commercial interests of the enterprise owning the system. Using the principles and methods of valuation, we build a mathematical model and propose formulas that allow us, taking into account inflation, to find the optimal assigned service life of the system and

at the same time estimate the market value of the work performed by the system and calculate the change in the market value of the system with age. Moreover, in this problem, the optimality criterion is the ratio of the expected discounted costs to the expected discounted volume of work performed by the system. We show that such a criterion maximizes the market value of the enterprise owning the system. We give examples of using the constructed model. The results obtained can be used both for solving other optimization problems of the reliability theory and for practical valuation of some types of machinery and equipment.

Keywords: technical system reliability, failure, assigned service life, optimization criterion, valuation theory, market value

Citation: Smolyak S.A. (2022) On assigning service life for technical systems under inflation. *Business Informatics*, vol. 16, no. 2, pp. 74–88. DOI: 10.17323/2587-814X.2022.2.74.88

References

1. *International Valuation Standards* (2019) London: International Valuation Standards Council.
2. Gnedenko B.V., Belyaev Yu.K., Soloviev A.D. (1965) *Mathematical methods in the reliability theory*. Moscow: Nauka (in Russian).
3. Smolyak S.A. (2016) *Machinery and Equipment Valuation (secrets of the DCF method)*. Moscow: Option (in Russian).
4. System of National Account 2008 (2009) New York: European Commission, International Monetary Fund, Organization for Economic Co-operation and Development, United Nations, World Bank.
5. Smolyak S.A. (2021) Assessment of machinery depreciation taking into account the provisions of the systems of national accounts. *Proceedings of the Institute for System Analysis of the Russian Academy of Sciences*, vol. 71, no. 1, pp. 44–54 (in Russian). <https://doi.org/10.14357/20790279210105>
6. Ostreykovsky V.A. (2003) *Reliability theory: Textbook for universities*. Moscow: Vysshaya shkola (in Russian).
7. *Methodical recommendations for assessing damage from accidents on hazardous production facilities (Regulatory Document 03-496-02)* (2010) Moscow: Closed Joint Stock Company «Scientific and Technical Center for Research on Industrial Safety Problems» (in Russian).
8. Lutz J., Hopkins A., Letschert V.E., Franco V., Sturges A. (2011) Using national survey data to estimate lifetimes of residential appliances. *HVAC&R Research*, vol. 17, no. 5, pp. 726–736. <https://doi.org/10.1080/10789669.2011.558166>
9. Erumban A.A. (2008) Lifetimes of machinery and equipment evidence from Dutch manufacturing. *Review of Income and Wealth*, vol. 54, no. 2, pp. 237–268. <https://doi.org/10.1111/j.1475-4991.2008.00272.x>
10. Nomura K., Suga Y. (2018) Measurement of depreciation rates using microdata from disposal survey of Japan. In: *The 35th IARIW General Conference, Copenhagen, Denmark*.
11. Alekseev V.V., Khomenko I.V., Prokhorsky R.A. (2011) Models of planning repairs and replacement of elements in the life cycle of complex technical systems. *Bulletin of the Voronezh Institute of the Ministry of Internal Affairs of Russia*, no. 3, pp. 94–102 (in Russian).
12. Van Horenbeek A., Pintelon L., Muchiri P. (2010) Maintenance optimization models and criteria. *International Journal of System Assurance Engineering and Management*, vol. 1, no. 3, pp. 189–200.
13. Barlow R. and Hunter L. (1960) Optimum preventive maintenance policies. *Operations Research*, vol. 8, no. 1, pp. 90–100. <https://doi.org/10.1287/opre.8.1.90>
14. Jiang R. (2018) Performance evaluation of seven optimization models of age replacement policy. *Reliability Engineering & System Safety*, vol. 180, pp. 302–311. <https://doi.org/10.1016/j.ress.2018.07.030>
15. Smith D.J. (2011) *Reliability, maintainability and risk: Practical methods for engineers*. Elsevier Ltd, Eighth edition.
16. Aven T. (1983) Optimal replacement under a minimal repair strategy. *Advances in Applied Probability*, vol. 15, no. 1, pp. 198–211. <https://doi.org/10.2307/1426990>
17. Weersink A., Stauber S. (1988) Optimal replacement interval and depreciation method for a grain combine. *Western Journal of Agricultural Economics*, vol. 13, no. 1, pp. 18–28. <https://doi.org/10.22004/ag.econ.32156>
18. Vilensky P.L., Livchits V.N., Smolyak S.A. (2015) *Evaluation of the investment projects efficiency: theory and practice*. Moscow: PolyPrintService (in Russian).
19. State of Colorado. The Department of Local Affairs (2021) *2021 Personal Property Factors & Tables*. Available at: <https://drive.google.com/file/d/1FZi9Da3QebSCDSVTPpZEX0H-6t0c9MNd/view> (accessed 22 May 2022).
20. Utah State Tax Commission. Property Tax Division (2021) *2021 Recommended Personal Property Valuation Schedule*. Available at: https://propertytax.utah.gov/personal-property/val_schedule_2021.pdf (accessed 22 May 2022).
21. Montana Department of Revenue (2022) *2022 Personal Property Depreciation Schedules and Trend Tables*. Available at: <https://mtrevenue.gov/?mdocs-file=59493> (accessed 22 May 2022).

22. Livshits V.N., Smolyak S.A. (1990) Models of the dynamics of economic wear of equipment. *Economics and Mathematical Methods*, vol. 26, no. 5, pp. 871–882 (in Russian).
23. Christer A.H., Goodbody W. (1980) Equipment replacement in unsteady economy. *Journal of the Operational Research Society*, vol. 31, pp. 497–506.
24. Dohi T., Ashioka A., Osaki S., Kaio N. (2001) Optimizing the repair-time limit replacement schedule with discounting and imperfect repair. *Journal of Quality in Maintenance Engineering*, vol. 7, no. 1, pp. 71–84. <https://doi.org/10.1108/13552510110386973>
25. Lvov D.S. (1969) *Economic problems of improving the quality of industrial products*. Moscow: Nauka (in Russian).
26. Smolyak S.A. (2018) On D.S. Lvov's formula for the valuation of machines. *Property relations in the Russian Federation*, vol. 205, no. 10, pp. 19–25 (in Russian). <https://doi.org/10.24411/2072-4098-2018-10101>
27. Smolyak S.A. (2020) Generalized D.S. Lvov's formula for machines subject to degradation. *Proceedings of the Institute for System Analysis of the Russian Academy of Sciences*, vol. 3, pp. 3–12 (in Russian). <https://doi.org/10.14357/20790279200301>

About the author

Sergey A. Smolyak

Dr. Sci. (Econ.);

Principal Science Researcher, Central Economics and Mathematics Institute, Russian Academy of Science, 47, Nakhimovsky Prospect, Moscow 117418, Russia;

E-mail: smolyak1@yandex.ru

ORCID: 0000-0001-5287-4285