

БИЗНЕС- ИНФОРМАТИКА

МЕЖДИСЦИПЛИНАРНЫЙ НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ

СОДЕРЖАНИЕ

Информационные системы и технологии в бизнесе

M. Amzarakov, R. Sukhov, E. Isaev

Modular data center: The holistic view7

Анализ данных и интеллектуальные системы

*A. Mitsyuk, A. Kalenkova, S. Shershakov,
W. van der Aalst*

Using process mining for the analysis
of an e-trade system: A case study 15

G. Kukharev, Y. Matveev, N. Shchegoleva

Barcode generation for face images28

Ю.Г. Сметанин, М.В. Ульянов

Мера символьного разнообразия:
подход комбинаторики слов к определению
обобщенных характеристик временных рядов40

Математические методы и алгоритмы решения задач бизнес-информатики

Л.Н. Ясницкий, Д.В. Иванов, Е.В. Липатова

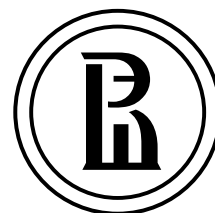
Нейросетевая система оценки
вероятности банкротства банков.....49

Н.Л. Коровкина, М.С. Фай

Обоснование инвестиций в информационные
технологии на основе дерева бизнес-драйверов57

Ю.П. Ехлаков, Н.В. Пермякова

Нечеткая модель оценки рисков продвижения
программных продуктов69



Издатель:

Национальный
исследовательский университет
«Высшая школа экономики»

**Подписной индекс
в каталоге агентства
«Роспечать» –72315**

Выпускается ежеквартально

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

*Главный редактор
А.О. Голосов*

*Заместители главного редактора
А.Р. Горбунов, Д.В. Исаев*

*Научный редактор
Н.Н. Лычкина*

*Технический редактор
В.И. Осипов*

*Дизайн обложки
С.Н. Борисова*

*Компьютерная верстка
О.А. Богданович*

*Администратор веб-сайта
Д.С. Проценко*

Журнал рекомендован ВАК
для научных публикаций

Адрес редакции:
105187, г. Москва, ул. Кирпичная, д. 33
Тел./факс: +7 (495) 771-32-38
<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 500 экз.

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

© Национальный
исследовательский университет
«Высшая школа экономики»

О ЖУРНАЛЕ

«**Б**изнес-информатика» – рецензируемый междисциплинарный научный журнал, выпускаемый с 2007 года Национальным исследовательским университетом «Высшая школа экономики» (НИУ ВШЭ). Администрирование журнала осуществляется факультетом бизнес-информатики НИУ ВШЭ.

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:

- ◆ корпоративные информационные системы;
- ◆ информационные технологии в бизнесе;
- ◆ организационные и управленческие проблемы создания и внедрения информационных систем;
- ◆ математическое моделирование социально-экономических процессов;
- ◆ методы анализа информации;
- ◆ интеллектуальные системы и управление знаниями в бизнесе;
- ◆ информационные сети и телекоммуникации;
- ◆ программная инженерия;
- ◆ информационная безопасность;
- ◆ электронный бизнес;
- ◆ инновации и бизнес в сфере информационных технологий;
- ◆ стандартизация, сертификация и качество;
- ◆ правовые вопросы бизнес-информатики;
- ◆ обучение в сфере бизнес-информатики.

В соответствии с решением президиума Высшей аттестационной комиссии Российской Федерации с 2010 года журнал включен в Перечень российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней доктора и кандидата наук.

Журнал выпускается ежеквартально и распространяется как в печатном виде, так и в электронной форме.

Журнал «Бизнес-информатика» зарегистрирован в «Роскомнадзоре».
Свидетельство ПИ № ФС 7752404 от 28 декабря 2012 г.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

ГЛАВНЫЙ РЕДАКТОР

ГОЛОСОВ Алексей Олегович —
кандидат технических наук, Президент компании
«ФОРС — Центр разработки»

ЗАМЕСТИТЕЛИ ГЛАВНОГО РЕДАКТОРА

ГОРБУНОВ Алексей Рэмович —
кандидат экономических наук, старший научный сотрудник,
Института США и Канады РАН

ИСАЕВ Дмитрий Валентинович —
кандидат экономических наук, доцент кафедры бизнес-
анализа, факультет бизнес-информатики, Национальный
исследовательский университет «Высшая школа экономики»

ЧЛЕНЫ РЕДКОЛЛЕГИИ

АБДУЛЬРАБ Абиб —
PhD, профессор департамента математики и программной
инженерии, Национальный институт прикладных наук, Руан,
Франция

АВДОШИН Сергей Михайлович —
кандидат технических наук, заведующий кафедрой управления
разработкой программного обеспечения, заведующий
отделением программной инженерии, Национальный
исследовательский университет «Высшая школа экономики»

АЛЕСКЕРОВ Фуад Тагиевич —
доктор технических наук, профессор, руководитель
департамента математики, факультет экономики, Национальный
исследовательский университет «Высшая школа экономики»

БАБКИН Эдуард Александрович —
кандидат технических наук, PhD, заведующий кафедрой
информационных систем и технологий, факультет бизнес-
информатики и прикладной математики (Нижний Новгород),
Национальный исследовательский университет
«Высшая школа экономики»

БАЙЕР Алекс —
PhD, Директор KAFAN FX Information Services, Нью-Йорк, США

БАРАНОВ Александр Павлович —
доктор физико-математических наук, профессор,
заведующий кафедрой информационной безопасности, факуль-
тет бизнес-информатики, Национальный исследовательский
университет «Высшая школа экономики»

БЕККЕР Йорг —
PhD, проректор, профессор, директор Европейского
исследовательского центра в области информационных систем
(ERCIS) Мюнстерского университета, Мюнстер, Германия

БЕЛОВ Владимир Викторович —
доктор технических наук, профессор кафедры вычислительной
и прикладной математики, факультет вычислительной техники,
Рязанский государственный радиотехнический университет.

ГРИБОВ Андрей Юрьевич —
кандидат экономических наук, Генеральный директор
компании «КиберПлат»

ГРОМОВ Александр Игоревич —
кандидат химических наук, профессор, заведующий кафедрой
моделирования и оптимизации бизнес-процессов, факультет
бизнес-информатики, Национальный исследовательский
университет «Высшая школа экономики»

ГУРВИЧ Владимир Александрович —
PhD, приглашенный профессор и исследователь,
Центр исследования операций, Ратгерский университет
(Университет Нью-Джерси), США

ДЖЕЙКОБС Лоренц —
PhD, профессор медицинского факультета, Университет
Цюриха, Швейцария

ЗАНДКУЛЬ Курт —
PhD, заведующий кафедрой информационных систем
для бизнеса, институт информатики, факультет информатики
и электротехники, Университет Ростока, Германия

ИЛЬИН Николай Иванович —
доктор технических наук, профессор, заместитель начальника
Управления специальной связи, Федеральная служба охраны
Российской Федерации (ФСО России)

КАЛЯГИН Валерий Александрович —
доктор физико-математических наук, профессор,
заведующий кафедрой прикладной математики и информатики,
факультет бизнес-информатики и прикладной математики
(Нижний Новгород), Национальный исследовательский
университет «Высшая школа экономики»

КАМЕННОВА Мария Сергеевна —
кандидат технических наук, директор компании «ИДС Шер»,
Россия и страны СНГ

КУЗНЕЦОВ Сергей Олегович —
доктор физико-математических наук, профессор, заведующий
кафедрой анализа данных и искусственного интеллекта,
заведующий отделением прикладной математики, Национальный
исследовательский университет «Высшая школа экономики»

МАЛЬЦЕВА Светлана Валентиновна —
доктор технических наук, профессор, заведующий кафедрой
инноваций и бизнеса в сфере информационных технологий,
декан факультета бизнес-информатики, Национальный
исследовательский университет «Высшая школа экономики»

МЕЙОР Питер —
PhD, заместитель директора консультативной группы
по радиокommunikациям, Международный телекоммуникаци-
онный союз (ITU), заместитель руководителя Комиссии ООН
по науке и технологиям, Женева, Швейцария

МИРКИН Борис Григорьевич —
доктор технических наук, профессор кафедры анализа данных
и искусственного интеллекта, отделение прикладной мате-
матики и информатики, Национальный исследовательский
университет «Высшая школа экономики»

МОТТЛЬ Вадим Вячеславович —
доктор технических наук, профессор кафедры
информационной безопасности, факультет кибернетики,
Тульский государственный университет

ПАЛЬЧУНОВ Дмитрий Евгеньевич —
доктор физико-математических наук, заведующий кафедрой
общей информатики, факультет информационных технологий,
Новосибирский государственный университет

ПАРДАЛОС Панайот (Панос) —
PhD, почетный профессор, директор центра прикладной
оптимизации, департамент промышленной и системной
инженерии, Университет Флориды, США

СИЛАНТЬЕВ Альберт Юрьевич —
доктор технических наук, профессор кафедры
информационных бизнес систем, Институт информационных
бизнес-систем, Национальный исследовательский
технологический университет «МИСиС»

ТАРАТУХИН Виктор Владимирович —
кандидат технических наук, PhD, руководитель научной группы
Европейского исследовательского центра в области
информационных систем (ERCIS) Мюнстерского
университета, Мюнстер, Германия

УЛЬЯНОВ Михаил Васильевич —
доктор технических наук, профессор кафедры управления
разработкой программного обеспечения, отделение
программной инженерии, Национальный исследовательский
университет «Высшая школа экономики»

ШАЛКОВСКИЙ Алексей Геннадьевич —
кандидат технических наук, проректор, директор Института
информационных технологий, Национальный
исследовательский университет «Высшая школа экономики»

ISSN 1998-0663

BUSINESS INFORMATICS

INTERDISCIPLINARY ACADEMIC JOURNAL

CONTENTS

Information systems and technologies in business

M. Amzarakov, R. Sukhov, E. Isaev
Modular data center: The holistic view7

Data analysis and intellectual systems

*A. Mitsyuk, A. Kalenkova, S. Shershakov,
W. van der Aalst*
Using process mining for the analysis
of an e-trade system: A case study 15

G. Kukharev, Y. Matveev, N. Shchegoleva
Barcode generation for face images28

Y. Smetanin , M. Ulyanov
Measure of symbolical diversity:
Combinatorics on words as an approach
to identify generalized characteristics
of time series.....40

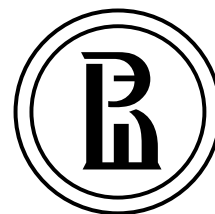
Mathematical methods and algorithms for business informatics problems

L. Yasnitsky, D. Ivanov, E. Lipatova
Neural network designed to estimate
probability of bank bankruptcies49

N. Korovkina, M. Fay
Justification of IT investments
based on business drivers tree57

Y. Ekhlakov, N. Permyakova
A fuzzy risk assessment model
for software promotion risks69

№3(29)-2014



Publisher:
National Research University –
Higher School of Economics

**Subscription index
in the «Rospechat» catalogue –
72315**

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 Ministry
of Education and Science
of the Russian Federation*

Editor-in-Chief:
A. Golosov

Deputies Editor-in-Chief:
A. Gorbunov, D. Isaev

Scientific Editor:
N. Lychkina

Technical Editor:
V. Osipov

Design:
S. Borisova

Computer Making-up:
O. Bogdanovich

Website Administration:
D. Protsenko

Address:
33, Kirpichnaya str., Moscow,
105187, Russian Federation

Tel./fax: +7 (495) 771-32-38
<http://bijournal.hse.ru>
E-mail: bijournal@hse.ru

Circulation – 500 copies

Printed in HSE Printing House
3, Kochnovskiy proezd, Moscow,
Russian Federation

© 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 Faculty of Business Informatics.

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:

- ◆ Corporate information systems
- ◆ Information technologies in business
- ◆ Organizational and managerial problems of information systems development and implementation
- ◆ Mathematical modeling of economic and social processes
- ◆ Methods of information analysis
- ◆ Intellectual systems and knowledge management in business
- ◆ Information networks and telecommunications
- ◆ Software engineering
- ◆ Information security
- ◆ Electronic business
- ◆ Innovations and business in the sphere of information technologies
- ◆ Standardization, certification and quality
- ◆ Legislation in the field of business informatics
- ◆ Education in the field of business informatics.

Since 2010 the journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Ministry of Education and Science of the Russian Federation.

The journal is published quarterly and distributed both in printed and electronic forms.

EDITORIAL BOARD

EDITOR-IN-CHIEF

Dr. Alexey GOLOSOV –

President of FORS Development Center, Russian Federation

DEPUTY EDITOR-IN-CHIEF

Dr. Alexey GORBUNOV –

Senior Researcher, Institute of US and Canada Studies

Dr. Dmitry ISAEV –

Associate Professor, Department of Business Analytics, Faculty of Business Informatics, National Research University Higher School of Economics, Russian Federation

EDITORIAL BOARD

Prof. Habib ABDULRAB –

Professor, Mathematical and Software Engineering Department, National Institute of Applied Sciences – Institut national des sciences appliquées de Rouen (INSA de Rouen), Rouen, France

Dr. Sergey AVDOSHIN –

Head of Software Management Department, Head of School of Software Engineering, National Research University Higher School of Economics, Russian Federation

Prof. Fuad ALESKEROV –

Head of Department of Mathematics, Faculty of Economics, National Research University Higher School of Economics, Russian Federation

Dr. Eduard BABKIN –

Head of Department of Information Systems and Technologies, Faculty of Business Informatics and Applied Mathematics (Nizhny Novgorod), National Research University Higher School of Economics, Russian Federation

Dr. Alex BAYER –

Head of KAFAN FX Information Services, New York, USA

Prof. Alexander BARANOV –

Head of Department of Information Security Management, Faculty of Business Informatics, National Research University Higher School of Economics, Russian Federation

Prof. Jorg BECKER –

Vice Rector, Professor, Director of European Research Center for Information Systems (ERCIS), University of Munster, Germany

Prof. Vladimir BELOV –

Professor, Department of Computational and Applied Mathematics, Faculty of Computer Engineering, Ryazan State Radio Engineering University, Russian Federation

Dr. Andrey GRIBOV –

Director General, CyberPlat Company, Russian Federation

Prof. Alexander GROMOV –

Head of Department of Modeling and Business Process Optimization, Faculty of Business Informatics, National Research University Higher School of Economics, Russian Federation

Prof. Vladimir GURVICH –

Invited Professor and Researcher, Rutgers Center for Operations Research, Rutgers, The State University of New Jersey, USA

Prof. Laurence JACOBS –

Professor, Medical School, University of Zurich, Switzerland

Prof. Kurt SANDKUHL –

Head of Department of Business Information Systems, Institute of Computer Science, Faculty of Computer Science and Electrical Engineering, University of Rostock, Germany

Dr. Nikolay ILYIN –

Deputy Head, Administration of Special Communication, Federal Security Guard, Russian Federation

Prof. Valery KALYAGIN –

Head of Department of Applied Mathematics and Informatics, Faculty of Business Informatics and Applied Mathematics (Nizhny Novgorod), National Research University Higher School of Economics, Russian Federation

Dr. Maria KAMENNOVA –

director, IDS Scheer Russia and CIS Countries, Russian Federation

Prof. Sergey KUZNETSOV –

Head of Department of Data Analysis and Artificial Intelligence, Head of School of Applied Mathematics and Information Science, National Research University Higher School of Economics, Russian Federation

Prof. Svetlana MALTSEVA –

Head of Department of Innovation and Business in Information Technologies, Acting Dean of Faculty of Business Informatics, National Research University Higher School of Economics, Russian Federation

Dr. Peter MAJOR –

Vice-chairman, Radiocommunication Advisory Group of International Telecommunication Union (ITU), vice-chairman of the UN Commission on Science and Technology for Development (CSTD), Geneva, Switzerland

Prof. Boris MIRKIN –

Professor of Department of Data Analysis and Artificial Intelligence, School of Applied Mathematics and Information Science, National Research University Higher School of Economics, Russian Federation

Prof. Vadim MOTIL –

Professor, Department of Information Security Management, Faculty of Cybernetics, Tula State University, Russian Federation

Prof. Dmitry PALCHUNOV –

Head of Department of General Informatics, Faculty of Information Technologies, Novosibirsk State University, Russian Federation

Prof. Panagote (Panos) PARDALOS –

Distinguished Professor and University of Florida Research Foundation Professor, Director of Center for Applied Optimization, Department of Industrial and Systems Engineering, University of Florida, USA

Dr. Albert SILANTYEV –

Professor, Department of Information Business Systems, Institute of Information Business Systems, National University of Science and Technology «MISIS», Russian Federation

Dr. Victor TARATOUKHIN –

Managing Director European Research Center for Information Systems (ERCIS) Competence Center ERP, Head of ERCIS Lab. Russia, University of Munster, Germany

Prof. Mikhail ULYANOV –

Professor of Software Management Department, School of Software Engineering, National Research University Higher School of Economics, Russian Federation

Dr. Alexey SHALCOVSKY –

Vice Rector, Director of Institute of Information Technologies, National Research University Higher School of Economics, Russian Federation

MODULAR DATA CENTER: THE HOLISTIC VIEW

Maxim AMZARAKOV

Director, INO Uptime Technology

Address: 9, Chasovaya str., Moscow, 125315, Russian Federation

E-mail: m.amzarakov@uptimetechnology.ru

Rafael SUKHOV

Finance Manager, INO Uptime Technology

Address: 9, Chasovaya str., Moscow, 125315, Russian Federation

E-mail: r.sukhov@uptimetechnology.ru

Eugene ISAEV

Professor, Head of Department for Education Stack Group,

National Research University Higher School of Economics; Head of the Laboratory,

P.N.Lebedev Physical Institute of the Russian Academy of Sciences

Address: 20, Myasnitskaya str., Moscow, 101000, Russian Federation

E-mail: eisaev@hse.ru

Datacenter modularity is a new term in data processing. The given article aims at making distinction between the modular, mobile, modern and traditional datacenter, and reviews solutions available in the market.

The present research provides a systematic view of the currently existing data center technology solutions, as well as the major factors influencing the cost and possible engineering pitfalls and determining basic rules of calculating and evaluating their cost and further maintenance. The concept of energy efficiency is studied here, as well as its influence on the primary cost of a data center, its maintenance cost, and thus its final cost. The conditions for the modular solutions for data processing centers emergence have also been studied here. Classifying and identifying key features allows precise positioning of the applicability of existing technologies. For this purpose, the paper provides major features of the applicability limits of available technologies, while technological solutions from different vendors are evaluated for containing engineering systems. The received estimations are presented in a convenient and comparable tabular form and.

The research results are provided in the form of a summarizing table allowing comparing the features of each solution in several aspects: form factor, complete solution, modularity, flexibility, further development in several key engineering solutions.

Key words: datacenter, modular datacenter, IT industry, information technologies.

Introduction

The «Datacenter world» today is a numerous set of products and solutions providing end-user with the computational and storage resources. Some people treat the datacenter as «High level» services such as Cloud term SaaS, which is quite different in fact. A datacenter is defined as a facility used to host

computer systems and associated components, such as telecommunications and storage systems. It generally includes redundant or backup power supplies, redundant data communications connections, environmental controls (e.g., air conditioning, fire suppression) and security devices [1]. The diversity of meaning of the term «datacenter» confuses the understanding of the topic.

Datacenters do not provide any other services except IT and telecommunication equipment hosting. But the most common view of a datacenter is «something I can connect to» or «something which does the data processing for me». The truth is somewhere in between. Both approaches to datacenters can be applied if we say that a datacenter is both an infrastructure to host computers and computers with the network it self. In order to clarify the view of the topic we define a datacenter as follows:

A data center is a highly technological facility to host computer technologies and data storage to ensure the continuity and reliability of data processing.

In other words, a datacenter is the physical layer supplying computer equipment with power and network wired connections. This notion is used in consideration for the purposes of this article.

1. What a datacenter needs

In general, the aim of a datacenter facility is to host the computing equipment, data storing and telecommunication equipment.

Basically, all requirements for datacenters come from the business. These demands are directly or indirectly linked to requirements of the IT equipment/systems.

These requirements can be ranged by their reliability, sustainability, maintainability, power consumption, data floor space etc.

In the early days of the IT the datacenter design was fixed and depended on the IT equipment for which a datacenter was designed. «Computer room» was the word to describe infrastructure supplying power and cooling for specific IT hardware. As computer industry grew, server rooms became more and more univesal and could host different types of IT hardware. The diversity of IT equipment today causes headaches for datacenter designers. And datacenters are becoming more and more flexible.

The source of the requirements (inputs) changes the datacenter design and the design process itself. Two following examples will show you how datacenter design can vary and be sophisticated depending on the source of the requirements.

Example 1: indirect requirements

One big international company is planning to deploy the Enterprise Resource Planning (ERP) system. The company has 200K employees accessing the system. The company also intends to improve the documenting and the internal processes by implementing remote desktop environment based on the IaaS services.

This requires a number of IT equipment, data storage and telecommunication infrastructure installed in the datacenter. Designing these systems requires a certain amount of computing resources and data storage, as does designing the telecommunication (network) part the system. This kind of the requirements can be easily transformed to ones for datacenter facility infrastructure. The designer of a datacenter selects infrastructure components on the basis of a fixed set of «consumers» of datacenter capacities.

The datacenter, designed and set up in this example, is specially developed for the specific needs (inputs) of the equipment hosted in it. This is an «ideal» example of datacenter design from the designer's point of view.

This is an example of how indirect business should define a datacenter. The company's direct requirements are reflected in its IT system.

Example 2: direct requirements

The second example comes from direct business needs – a commercial data center.

The company is planning to build a datacenter to provide basic services like location of goods.

The requirements in this example can only be described though the number of crates (e.g. floor space) and the planned power capacity, otherwise the designer will not be able to choose appropriate cables, sockets, network cabling and etc. Also no information was provided about the required reliability of the infrastructure.

Details

The specifics of the data center allow seeing the kinds of criteria which are simpler/harder to reach.

- ◆ The data center today (based on the IT needs) can be described by the folloThe floor space footage;
- ◆ The power density per square feet;
- ◆ The power efficiency;
- ◆ The reliability (or expected uptime);
- ◆ The lifetime.

The floor space and power capacity are defined straightforward in both examples. Power efficiency is more valuable for commercial DC, we assume. Reliability can be defined simply in the first example and becomes a complicated question in the second one.

Most of the customers coming into the datacenter can say: I want to have a Tier 3 [2] level of uptime for my servers and PUE less than 1.2 [3, 4, 5]! In the examples above the reliability of the last datacenter can be higher than that of the first one and much higher than that of

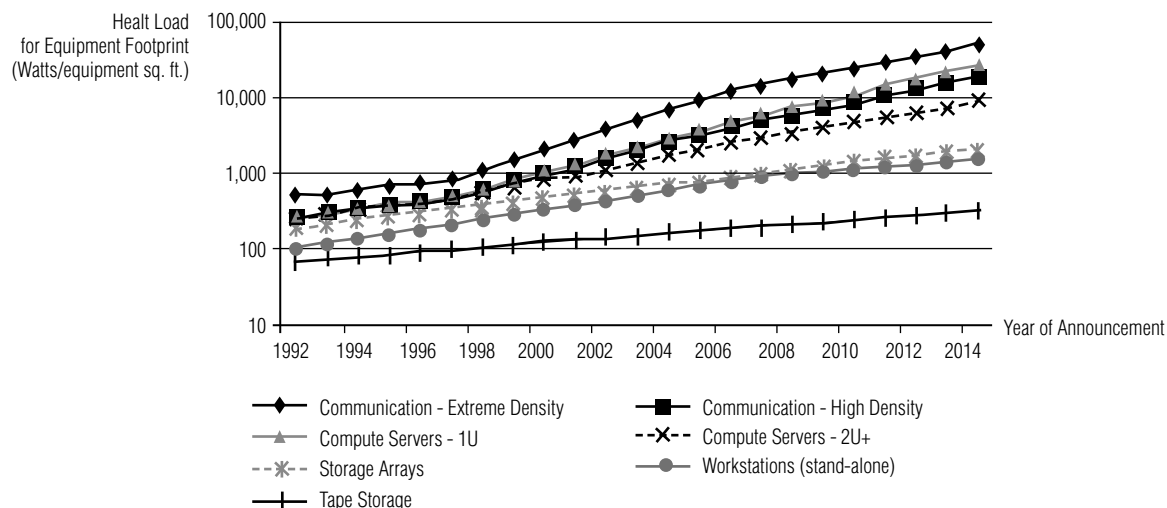


Fig. 1. Historic and projected head load by equipment type.

Source: D. Floyer, «Network Go GrEEN», Wikibon, May 07, 2011 (http://wikibon.org/wiki/v/Networks_Go_GrEEN)

the third (XaaS seems to be more redundant). What is the reliability standard you should choose for your application? What is reliability standard the designer has to take into account in the absence of the customer whom he can ask about it in the last example?

Lifetime

Lifetime of the IT equipment is nearly three years [6]. The requirements of the new IT equipment do not change in the form-factor. But these parameters will be completely different in terms of network ports number, power consumption per unit (see fig. 1), etc. So after 3 years the data center equipment shall be renovated.

The lifetime of the facility technologies, like cooling or power distribution, is close to 10 years after the installation.

We are well aware that in this case it is practically hard to justify the investment, if the equipment has served for three years only [2].

More details. Flexibility

As you already can see, the first example becomes less and less simple in the design as we start thinking farther than three years.

Design is much more detailed in the first example. All plugs, network cords, places, spaces are predesigned, hence we have to provide IT system and IT hardware design.

In the case of unpredictable IT hardware demand, the cabling infrastructure, power distribution infrastructure of the datacenter should be flexible enough to easily adapt to new IT equipment.

Keeping in mind the «life time issue», the flexibility in capacities, in network connections, in space layout,

there tends to be an ultimate solution for the datacenter.

This can be true. Is the datacenter design flexible enough to make the final step of the datacenter growth completely different than designed initially? If it is possible, can the datacenter be modified by the capacity reduction or changed after several years of operations without the critical load interruption?

Power

The necessity to modify the datacenter design also comes from the technical characteristics of IT equipment to be placed in it. As mentioned above, the lifetimes of the IT equipment and datacenter facility infrastructure differ significantly. The latter one is three times more than the former one. At this point we expect the datacenter requirements will have to be substantially changed every three years because of the IT equipment. The regular server in the early 2Ks consumed 50W; modern server consumes at least 300W [7, 8]. The regular servers seven years ago had one or two Ethernet interfaces and most of them were connected to one port, today servers have three interfaces at least (one is for management) and most of them are connected to the datacenter network [9]. So we can say once every three or five years we face a significant change[2].

Timeframe

We spoke before about power demand, space capacities and possible change in three or five years. What if the change comes earlier? Is there a reason to have this change? The growth of IT is very rapid all over the world. This enhances the computer and automation system

growth and change, IT growth and changes, datacenter facilities grow and change, one after another. There are a lot of examples when a data center becomes outdated before it starts working, because IT requirements of the company changed too fast.

The usual time frame of the datacenter development is 6 months for design and one year to buy and build it. Can one imagine how many changes IT world can face with in the one year and a half? A new iPhone is released every year with half its version new. A new Microsoft Windows appears approximately every 1,5 years [10]. This can be the planned growth you can predict. But can you predict the future?

Power efficiency

One of the data center characteristics is caused not by IT, but by the money limit. The power consumption of the IT equipment (*see fig. 1*) and the price per kilowatt are growing rapidly. Both facts lead to the basic answer: we have to save money!

The efficiency of the data center is described by the proportion between the power consumed by the data center and the power supplied to the IT equipment. These characteristics are called PUE and calculated by the formula:

$$PUE = \frac{P_{total}}{P_{IT}},$$

where PIT is the amount of power consumed by IT hardware.

The number theoretically can be as low as 1. The data center power efficiency is better as PUE number is lower.

Drivers

There is always a solution when the requirements are identified. Modular datacenters come to cover the above-mentioned needs [4]. There is also a specific use of such solutions, but this will be covered later in this article.

To recap above mentioned, we should say the demands which drive the datacenter technology development are the following:

- ◆ reliability;
- ◆ flexibility;
- ◆ deployment time;
- ◆ mobility;
- ◆ power efficiency.

2. Terms

Keeping in mind examples and requirements discussed above, we must define mobile, modular, modern and traditional types of data center.

We could say mobile and modular types are modern, while modern is not only mobile and modular.

Traditional

The traditional approach to building a data center is to construct concrete/brick solid building and bring all necessary infrastructure in. Everything is fixed and cannot (or is very expensive to) be changed.

Modern

The modern data center reflects the modern needs of flexibility, deployment time, mobility, reliability and power efficiency. Which cannot be achieved in a «concrete» building. And of course, a modern data center can be a «concrete» solid building, but very efficient as it must be today.

Mobile

The data center can be disassembled and moved to another place. Such kind of solution is highly prefabricated. And most of the times it is reliable as the consequences of «factory made product». Mobile solutions address the deployment time issue. The TCO of containerized solution can be as little as the lifetime of the IT hardware filled in the container. Should it be lifetime issue solution? Mobile can also be modular.

Modular

If necessary, flexibility can be the main driver for the modular approach of the datacenter design.

Flexibility of all the components in the data center and the flexibility of the investments are the major characteristics of the modular solutions.

A modular data center (as the word suggests) consists of modules. The modules are the infrastructure components of the datacenter. Modules give the possibility of «on-demand» growth of the datacenter. These modules can also be highly prefabricated, reducing the delivery and installation time.

3. Modular containers

The above-mentioned demands have their solution. Containerized datacenters are coming in to the market. This chapter describes five world known and real operation-proven solutions of this kind.

In this and the following chapters we intend to describe the existing modular solutions briefly. The tables describing the solution show which component of the data center is included, which is not included, and which is partly included (checked in both rows).

Microsoft

Microsoft modular datacenter's key feature is the virtual computing environment. The solution is based on integration of IT and networking equipment in one container. The container is produced and shipped to the customer from the factory where all the systems inside it are prefabricated and tested [10].

The centralized power supply with autonomous power generating and cooling supplies have to be ready on site to allow container installation.

The centralized systems can be modular in units, but not easily modular in distribution. The backbone distribution system of power and cooling has to be prebuilt on site and has the «end situation» capacity. This requires more investments at the first growth stage and restricts the change of the end state capacity and sometimes the stage of growth.

Table 1.

Microsoft MDC

	UPS	Fire	Security	Network	Cooling	Power	Office
Included	✓	✓	✓		✓	✓	
Not included				✓	✓	✓	✓

SUN MD

The first well-known modular datacenter was SUN Project Black box. The modularity of this solution meant the ability to bring the computing resources to the place where they were necessary. Sun Project Black box included up to 280 servers built into a ISO-standardized shipping container. Sun Microsystems was purchased by Oracle in 2010. Now the project is named Sun MD s20 [11]. It unifies infrastructure and IT hardware monitoring and controls. The power and cooling for the IT equipment has to be supplied by site infrastructure. The solution is applicable for fast deployment where IT demands are not vast and can fit into one or a couple of containers.

Table 2.

SUN MD

	UPS	Fire	Security	Network	Cooling	Power	Office
Included	✓	✓	✓	✓	✓		
Not included				✓	✓	✓	✓

AST Modular

AST started with the containerized datacenter and then came to the modular datacenter infrastructure. AST today has a set of modular products and solutions [12]. IT container has integrated uninterruptible power supply. Availability of water-cooled block container, direct and indirect cooling block container allows choosing the appropriate cooling solution for certain site and weather conditions. Power container and vestibule container complete the containerized modular data center solution.

Table 3.

AST Modular

	UPS	Fire	Security	Network	Cooling	Power	Office
Included	✓	✓	✓		✓	✓	
Not included							

Huawei CDC

20 feet or 40 feet sea containers [13]. The key feature, beside low price, is the indirect free-cooling incorporated into the container. The solution also has the «cooling container», «tambour» and power container which can be stacked to the «data container» providing the modular data center solution.

Table 4.

Huawei CDC

	UPS	Fire	Security	Network	Cooling	Power	Office
Included	✓	✓	✓	✓	✓	✓	
Not included				✓			✓

HP POD

HP Performance Optimized Datacenter (POD) 20c is based on the superstructure of special design [14]. It does not use the «traditional» ISO shipping container, which restricts entrance openings, inside interior etc. The block size is the same because of transportable dimensions. High temperature chilled water close-coupled heat removal provides higher energy efficiency but requires the cooling plant on site.

Table 5.

HP POD

	UPS	Fire	Security	Network	Cooling	Power	Office
Included	✓	✓	✓	✓	✓	✓	
Not included				✓			✓

4. Modular solutions

All of above-mentioned solutions use containers of fixed size. Most of these solutions require the site infrastructure for power and cooling to connect to.

The container solution is the best choice in case of accidental or partial IT demand.

The free choice in power and cooling infrastructure, the necessity of interconnection between the «building block» using the containerized design, the fixed data-floor space and server rack dimensions restrictions are the drawbacks of containerized solutions and limit their usability as the basis of large datacenters.

Examples of that are the AST NON ISO Modular Data center and Colt Modular Datacenter.

AST Modular

AST NON ISO Modular datacenter is based on the same set of cooling, power, IT room building blocks as its containerized solution. But the outside is different. The cooling is provided by modules of different capacities and sizes. Availability of IT room block of two sizes (25' and 45') allows to assembling the data room according to the customer's needs.

Table 6.

AST Modular NON ISO

	UPS	Fire	Security	Network	Cooling	Power	Office
Included	✓	✓	✓	✓	✓	✓	✓
Not included							

Colt

The Colt modular data center is one of the containerized solutions. Colt has the growth «block» of data-

center space and power. [15] Most of the construction blocks are produced at the factory. The customer selects the floor space block (server rooms) of his datacenter. Then Colt is able to provide the block-to-block growth of the datacenter.

Table 7.

Colt Modular DC

	UPS	Fire	Security	Network	Cooling	Power	Office
Included	✓	✓	✓	✓	✓	✓	✓
Not included							

Conclusion

There is no «general» description of modular data-center today. You can see from examples above many directions of modularity in technologies. The means of modularity, noted in the article of J. Rath «DCK Guide to the Modular Datacenter» [16], are the following:

- ◆ deployment;
- ◆ consumption;
- ◆ financing.

The modular datacenter differs from traditional at least by flexibility and modularity in the deployment process. At the same time modular datacenter does not require to be mobile, but can also be constructed from mobile blocks.

The author's definitions of pure modular datacenter are:

- ◆ step by step growth of basic capacities: space, power, cooling;
- ◆ changeability of the capacities during the DC operations;
- ◆ flexibility in easy IT demand adaptation.

We include all the described above features of mobile and modular solutions into one table to make the comparison easier. Table 8 allows comparing the advantages and disadvantages of each solution. Does the solution include Power and Cooling or is this required on site? Is it possible to modular growth of the power inside the container? Further development availability and the flexibility are also shown in the table. ■

Table 8.

Modular datacenter comparison

Vendor	Solution	Solution			Modular			Growth			Flexibility			
		Space	Power	Cooling	Space	Power	Cooling	Space	Power	Cooling	Cabinets	kW/Rack	Cooling	Power
Oracle	Sun MD	●	○	○	◐	○	○	●	○	○	○	○	○	○
Microsoft		●	◐	○	◐	◐	○	●	○	○	○	○	○	○
AST	ISO	●	●	●	●	●	●	●	●	●	○	●	●	●
Huawei	CDC	●	●	●	●	◐	●	●	●	●	○	●	●	●
HP	POD	●	◐	○	◐	◐	○	●	○	○	○	○	○	○
AST	NON ISO	●	●	●	●	●	●	●	●	●	●	●	●	●
Colt		●	●	●	●	●	●	●	●	●	●	●	●	●

References

1. Kelley C., Cooley J. (2011) *Deploying and Using Containerized / Modular Data Center Facilities*. The Green Grid White Papers.
2. The Uptime Institute: <http://www.uptimeinstitute.com/>(accessed 30.07.2013).
3. Sukhov R.R., Amzarakov M.B., Isaev E.A. (2013) Advanced Data Center Economy. *Business Informatics*, no. 2 (24), pp.13-18.
4. The Green Grid Consortium: <http://www.thegreengrid.org/Global/Content/white-papers/The-Green-Grid-Data-Center-Power-Efficiency-Metrics-PUE-and-DCiE> (accessed 05.07.2013).
5. Uptime Technology Russia: http://www.uptimetechnology.ru/evaluation_pue.html (accessed 30.07.2013).
6. Intel: <http://www.intel.com/>(accessed30.07.2013)
7. Turner IV P.E., Brill K.G.(2010) *Cost Model: Dollars per kW plus Dollars per Square Foot of Computer Floor*. The Uptime Institute White paper.
8. Koomey J. (2007) *Estimation Total Power Consumption by Servers in the U.S. and the World*. Jonathan Koomey Analytics Press.
9. Minas L., Ellison B. (2009) *The Problem of Power Consumption in Servers*. Intel.
10. Microsoft Company: <http://www.microsoft.com/>(accessed3 0.07.2013)
11. Oracle: <http://www.oracle.com/>(accessed 30.07.2013)
12. AST Modular: <http://www.astmodular.com/>(accessed 30.07.2013)
13. Huawei: <http://www.huawei.com/>(accessed 30.07.2013)
14. Hewlett Packard Company: <http://www.hp.com/>(accessed 30.07.2013)
15. Colt Technology Services Group: <http://www.colt.net/> (accessed 30.07.2013)
16. Rath J. (2013) DCK Guide To Modular Data Centers: Second Edition. *DataCenterKnowledge Journal*.

МОДУЛЬНЫЙ ЦЕНТР ОБРАБОТКИ ДАННЫХ: ЦЕЛОСТНЫЙ ВЗГЛЯД

М.Б. АМЗАРАКОВ

Директор АНО «Институт «Аптайм»
Адрес: 125315, г. Москва, ул. Часовая, д. 9
E-mail: m.amzarakov@uptimetechnology.ru

Р.Р. СУХОВ

Финансовый управляющий АНО «Институт «Аптайм»
Адрес: 125315, г. Москва, ул. Часовая, д. 9
E-mail: r.sukhov@uptimetechnology.ru

Е.А. ИСАЕВ

кандидат технических наук, профессор, заведующий базовой кафедрой
Группы компаний «Стек», Национальный исследовательский университет
«Высшая школа экономики»; заведующий лабораторией, Физический институт
имени П.Н. Лебедева Российской академии наук
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: eisaev@hse.ru

Настоящая работа ставит перед собой цель систематизировать существующие в мире технологические решения в области центров обработки данных. В статье рассмотрены основные факторы, влияющие на стоимость и инженерную сложность дата-центров и определяющие базовые правила формирования и оценки стоимости их создания и последующей эксплуатации. Рассмотрено понятие энергоэффективности и ее влияние на первичную стоимость дата-центра, стоимость его эксплуатации и, как следствие, совокупную стоимость владения. Также рассмотрены предпосылки появления модульных решений для центров обработки данных. Систематизация и определение ключевых отличий и особенностей позволяют точно позиционировать применимость существующих технологий. Для этого в работе изложены ключевые особенности и определены границы применимости существующих технологий, а технологические решения разных производителей оценены на предмет укомплектованности инженерными системами. Полученные оценки приведены в удобном для сопоставления табличном виде.

Итогом работы является обобщенная таблица, позволяющая сравнить возможности каждого решения в нескольких аспектах, включая форм-фактор, законченность решения, модульность, гибкость, возможность развития по нескольким ключевым инженерным решениям.

Ключевые слова: центр обработки данных, модульный центр обработки данных, ИТ-индустрия, информационные технологии.

Литература

1. Kelley C., Cooley J. Deploying and Using Containerized / Modular Data Center Facilities. The Green Grid White Papers, 2011.
2. The Uptime Institute: <http://www.uptimeinstitute.com/> (дата обращения 30.07.2013).
3. Sukhov R.R., Amzarakov M.B., Isaev E.A. Advanced Data Center Economy // Бизнес-информатика. 2013. № 2 (24). С.13-18.
4. The Green Grid Consortium: <http://www.thegreengrid.org/Global/Content/white-papers/The-Green-Grid-Data-Center-Power-Efficiency-Metrics-PUE-and-DCiE> (дата обращения 05.07.2013).
5. Uptime Technology Russia: http://www.uptimetechnology.ru/evaluation_pue.html (дата обращения 30.07.2013).
6. Intel: <http://www.intel.com/> (дата обращения 30.07.2013).
7. Turner IV P.E., Brill K.G. Cost Model: Dollars per kW plus Dollars per Square Foot of Computer Floor. The Uptime Institute White paper, 2010.
8. Koomey J. Estimation Total Power Consumption by Servers in the U.S. and the World. Jonathan Koomey Analytics Press, 2007.
9. Minas L., Ellison B. The Problem of Power Consumption in Servers. Intel, 2009.
10. Microsoft Company: <http://www.microsoft.com/> (дата обращения 30.07.2013).
11. Oracle: <http://www.oracle.com/> (дата обращения 30.07.2013).
12. ASTModular: <http://www.astmodular.com/> (дата обращения 30.07.2013).
13. Huawei: <http://www.huawei.com/> (дата обращения 30.07.2013).
14. Hewlett Packard Company: <http://www.hp.com/> (дата обращения 30.07.2013).
15. Colt Technology Services Group: <http://www.colt.net/> (дата обращения 30.07.2013).
16. Rath J. DCK Guide To Modular Data Centers: Second Edition. Data Center Knowledge Journal, 2013.

USING PROCESS MINING FOR THE ANALYSIS OF AN E-TRADE SYSTEM: A CASE STUDY

Alexey MITSYUK

Analyst, International Laboratory of Process-Aware information Systems (PAIS Lab.), National Research University Higher School of Economics

Address: 20, Myasnitskaya str., Moscow, 101000, Russian Federation

E-mail: amitsyuk@hse.ru

Anna KALENKOVA

Research Fellow, International Laboratory of Process-Aware information Systems (PAIS Lab.), National Research University Higher School of Economics

Address: 20, Myasnitskaya str., Moscow, 101000, Russian Federation

E-mail: akalenkova@hse.ru

Sergey SHERSHAKOV

Research Fellow, International Laboratory of Process-Aware information Systems (PAIS Lab.), National Research University Higher School of Economics

Address: 20, Myasnitskaya str., Moscow, 101000, Russian Federation

E-mail: sshershakov@hse.ru

Wil van der AALST

Academic Supervisor, International Laboratory of Process-Aware information Systems (PAIS Lab.), National Research University Higher School of Economics;

Full Professor, Department of Mathematics & Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

Address: P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands

E-mail: w.m.p.v.d.aalst@tue.nl

E-trade systems are widely used to automate sales processes. Inefficiencies and bottlenecks in the sales processes lead to business losses. Conventional approaches to identifying problems require much time and result in subjective conclusions. This paper proposes an approach for the analysis of e-trade system processes based on the application of process mining techniques. Process mining aims to discover, analyze, repair and improve real business processes on the basis of behavior of an information system recorded in an event log. Using process mining techniques, we have analyzed process running in an online ticket booking information system. This work has shown that process mining can give insight into the e-trade processes and can produce information for their improvement. The case study carried out allows formulating appropriate recommendations. The article also presents the real outcome of using process mining techniques. We have generalized the applied approach and showed how it could be used to the investigation of a wide spectrum of e-trade information systems. During the case study we mostly used a software framework named ProM, which includes a substantial number of plugins implementing process mining methods. Using software for automatic process analysis and discovery, one should be careful with the interpretation of particular methods' output. Pitfalls and difficulties of applying process mining techniques to the logs of e-trade systems have also been shown.

Key words: process mining, process analysis, data analysis, e-trade system.

1. Introduction

Process mining is a new and fast-growing research area in the field of Business Process Management. The idea of process mining is to discover, analyze and improve processes by extracting knowledge from real-life event logs of an information system [1, 2]. Such event logs are usually produced by most modern information systems. There are only two requirements for process mining: (1) there is a notion of a process, and (2) there is an event log that keeps recorded behavior of a process in a structured form. The event log has to contain information about process steps (events) together with timestamps and, perhaps, additional information (actors, resources). If both of these requirements are met, it is possible to apply a wide range of process mining techniques, including those implemented in ProM Framework [3]. Process mining includes (1) process discovery, (2) conformance checking, and (3) process enhancement [1]. Discovery aims to learn a process model from an event log, i.e. to derive a process model from observed behavior recorded in event log. Conformance checking answers the question whether the modeled behavior matches the observed behavior. Model enhancement comprises model improvement, extension, and optimization based on information obtained from event logs.

This paper describes an application of process mining to the analysis of e-trade system processes. This analysis is crucial for finding process bottlenecks and improving an information system. E-trade systems are widespread. Typically, a today's e-trade system consists of a server that processes the requests and a set of client software applications or a web-based client interface generating requests. When one wants to buy something (goods or services), they use a web site (in the open system case) or a client application (in the case of internal corporate system) to browse the list of available offers, then they form a request and send it to the server. An application at the server site receives this request and processes it in a number of ways using a particular process scheme. Eventually, a staff member should be involved in approving the request or preparing a ready supply.

The analysis of business process models, like the ones considered here, is far from trivial. In most cases, information systems have a rather complex structure, and involve a lot of services and people. Frequently, there is no explicit process model describing the system behavior. Developers and analysts often rely on an implicit model of the process, which is not well correlated with reality, i.e., real-life behavior is very different. When something goes wrong in such a process, it is a sophisticated task to

get insight into the problem and solve it. Since e-trade information systems generate event logs, process mining techniques can be used for analysis and improvement of such processes. Moreover, the recording of all trade operations is typically regulated by law. Using process mining methods, one can investigate the functioning of an information system, obtain models of real processes, analyze these models, locate inefficiencies, and propose improvements.

The paper presents a real case study involving an online e-trade system that was analyzed using process mining techniques.

Process mining has been applied in many other domains. For example, several papers have been published on process mining of healthcare processes, cf. the papers by Mans, van der Aalst et al. [5], Kirchner, Herzberg, et al. [6], and other works [7, 8, 9]. Another interesting application for process mining techniques is business process auditing [10, 11, 12, 13, 14]. There are also papers that consider using process mining in insurance [15]. Even maritime vessel behavior has been analyzed using process mining [16]. Process mining is a new rapidly developing area, thus applying process mining in real-life situations is of particular interest both for practice and further research.

Process mining uses many heuristics, and the direct application of process mining methods without any pre-processing usually is not helpful. The results of applying process mining strongly depend on the problem definition and questions asked. One has to be very precise with conditions and software settings to obtain relevant outcome (see [18]). Selection of appropriate techniques according to the subject area is an important preliminary step of analysis. Note that while dealing with a specific problem, one has not only to play with the parameters but also to extend existing methods.

The rest of the paper is organized as follows. Section 2 contains a general description of the problem. Section 3 presents analysis of the studied online e-trade information system. Finally, section 4 gives some conclusions and further research directions.

2. Online ticket booking information system

In this paper we consider a case study aimed at finding inefficiencies in a typical e-trade information system process that deals with booking travel tickets, and at proposing changes that would possibly lead to higher turnovers. To achieve these goals, various data analysis and process mining techniques were used.

The system is a portal designed to provide ticket booking services. It is a website that allows the users to search tickets according to a number of criteria (destination city, date, carrier, class of service, etc.). The resulting tickets are offered to the user. After booking, the user can purchase the reserved ticket by paying with a credit card or in cash. There is also an additional service when purchasing tickets: the user is advised to buy travel insurance. The server processes the requests and stores all the data, including event logs of the system behavior. Thus, we can apply process mining techniques.

Usually the average number of purchases per unique site visitor is used to evaluate the effectiveness of this kind of portal. The metric value for the portal is lower than the average value for similar projects in Russia, according to the information received from experts of the portal owner company. Thus, there are problems or bottlenecks in portal functioning. The portal owner had the feeling that potential clients left the travel portal after starting browsing and filling the forms without completing purchase of a ticket. The goal was to confirm or to refute this idea, and, in the latter case, to answer the question why this happens.

Event data gathered by the portal were used as input for this study. Initially, a period of one month was analyzed. Two tables provided by the portal and containing information about its functioning were used as input for creating an event log. The main fields of these tables are listed below (*Tab. 1, Tab. 2*). Each event in the log relates to an activity (a step in a process) and belongs to a process instance (a case). Table 1 contains cases, and Table 2 is filled with types of events recorded by the server.

Table 1.

Cases

ID	record serial number (page ID)
SESSION_ID	client session ID
ACTION_COUNT	number of actions on a page
ORDER_STATUS	status of an order for which the user entered data

3. Analysis of the system behavior

At the start of this research, the owner of the portal had no strictly formalized process model for the system, only a general description and a vague scheme of how it should function. Therefore, it was necessary to design a model. One preliminary step was needed before: to obtain and preprocess the event log.

Table 2.

Events

ID	entry serial number
PAGE_ID	ID of a page on which specific actions were performed; ID field from Table 1
OBJECT	page structure object that the client submitted to an action
WINDOW	window
PAYMETHOD	payment type
CONFIRM_SUBMIT	<-book> button
ACCEPT	acceptance of the fare conditions
SURNAME	surname
NAME	name
DOCNUMBER	document number
BIRTHDAY	date of birth
EXIST_DOCEXP	expiration
DOCEXP	valid until
FARE_DETAIL	link to information about the fare
FF_CARD_NUMBER_ADD	link for adding a frequent flyer card number
FF_CARD_NUMBER	frequent flyer card number
C_PHONE_NUMBER	cell phone
C_EMAIL	e-mail
INSURED_PERSON	adding insurance
ACTION	action on an object; possible options: LOAD, UNLOAD, CLICK, CHECK, UNCHECK, FILL, SELECT, CLEAR

The two tables containing information about the portal functioning were considered as an event log. In order to apply process mining techniques, it was necessary to have a single log file in a specific strictly formalized format [19]. Thus, the tables were merged to a single file by using the unique field identifiers and «PAGE_ID» field.

The preprocessing of the event log was performed using MySQL RDBMS [20], as well as ProM framework with additional software tools [19]. First of all, it was necessary to identify those fields which constitute events (i.e., event class identifiers in the information system). The combination of fields «OBJECT» + «ACTION» was chosen, as it identifies all the unique user actions. Taken separately, these fields do not completely describe an event in the portal information system. The user may perform different actions on the same object («click», «clear» and «fill»), at the same time the same action can be performed with regard to different objects (e.g., «pressing the left mouse button»). However, the pair of these fields uniquely characterizes an event (for example, «pressing the left mouse button on «submit» button»).

The event log was filtered in various ways before being analyzed. The significant and insignificant parts were identified. The timestamps of the log events were analyzed. It was important to filter out all the actions of the portal administration team, which was done using the selection based on user IP addresses. In the next chapter we will show statistical characteristics of the booking process.

3.1. Preliminary analysis

We analyzed an event log containing the records of the portal operation over a short period of time. As an event classifier, the pair of primary keys «ACTION» and «OBJECT» was chosen. «SESSION ID» field was selected as a trace classifier. The total number of events in the log was 84760 (50 different classes of events), and the total number of unique traces was 16818.

The ten types of events that are the most frequently represented in the log are shown in Fig. 1. It can be seen, that about 40% of all events available in the log are events of page loading and unloading. Importantly, the number of unloading events does not match that of page loadings. This effect is caused by cutting off the events that are outside the considered timeframe.

One can see that for 7564 traces (i.e., about a half), users attempted to select a payment method. Only in 4909

traces out of all launched (16818) users tried to submit a filled form to the server. Other traces can be considered unfortunate for the seller. Several traces without completion are the traces with a cut-off, but not all of them. This means there are problems with stability of the web site. Users have problems during filling and submitting forms.

The five most common classes of events in the log after removing «WINDOW LOAD» and «WINDOW UNLOAD» events are shown in Fig. 2.

The distribution of final events in the user traces is noteworthy. Fig. 3 shows the statistics for the five most frequent final trace events. One can see that only half of the sessions (49.85 %) end with attempts to submit data to the server. Approximately 17 % of customers finalize browsing the site after pressing «select a payment method» button («PAYMETHOD CLICK» action), which indicates the inadequacy of the payment options provided.

Another common event occurring prior to unloading the page is the event of displaying the fare conditions («FARE_DETAIL CLICK» action). In 367 cases, the visitors left the portal after viewing the fare. This value is not too large (it is obvious that some users will not be satisfied with the proposed fares).

Total number of classes: 50		
Class	Occurrences (absolute)	Occurrences (relative)
LOAD+WINDOW	16818	19,843%
UNLOAD+WINDOW	14962	17,653%
CLICK+PAYMETHOD	7564	8,925%
CLICK+CONFIRM_SUBMIT	4909	5,792%
FILL+SURNAME	3798	4,481%
FILL+DOCNUMBER	3729	4,4%
FILL+NAME	3688	4,351%
CHECK+ACCEPT	3681	4,343%
FILL+BIRTHDAY	3524	4,158%
UNCHECK+INSURED_PERSON	2897	3,418%

Fig. 1. The most frequent events in the log

CLICK+PAYMETHOD	7564	14,278%
CLICK+CONFIRM_SUBMIT	4909	9,267%
FILL+SURNAME	3798	7,169%
FILL+DOCNUMBER	3729	7,039%
FILL+NAME	3688	6,962%

Fig. 2. The five most frequently occurring events after removing the page loading and unloading events

Class	Occurrences (absolute)	Occurrences (relative)
CLICK+CONFIRM_SUBMIT	3242	49,854%
CLICK+PAYMETHOD	1124	17,284%
CHECK+INSURED_PERSON	544	8,365%
CLICK+FARE_DETAIL	367	5,644%
UNCHECK+INSURED_PERSON	332	5,105%

Fig. 3. The final events



Fig. 4. Characteristics of the event log after removing the loading and unloading events

The most of traces contain two exact events. These are traces consisting of «WINDOW LOAD» and «WINDOW UNLOAD» events. It takes from 30 seconds to 1 hour between the two events. Such traces must be associated with the users who only browse various offers, as well as with the web crawlers, which, of course, have no effect on booking.

Fig. 4 shows characteristics of the event log after removing the page loading and unloading events (and correspondingly the traces consisting only of opening and closing the portal page). Thus, the real average number of events in a trace is 8 (6 plus the two events for opening and closing of the page). Below we consider the filtered event log consisting of 52000 rather than 84000 events.

By using process mining it is possible to identify factors affecting the user’s desire to use portal’s services and buy a ticket on it.

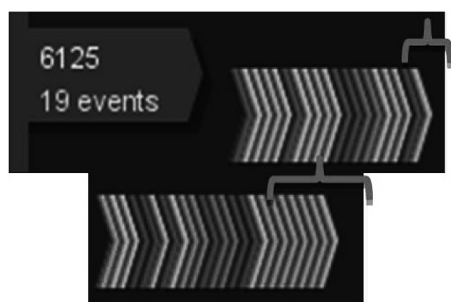


Fig. 5. Typical traces (sequences of activities)

One of the potentially problematic areas of the website is its reliability. When working with the portal event log, the following fact was identified: many users repeatedly (up to 9 times, Fig. 5) produce the action of submitting a completed form to the server, which is designated by «CONFIRM_SUBMIT_CLICK» event (such behavior was observed in more than a half of the cases). This behavior indicates a problem with bandwidth and connection efficiency of the channel between the user interface and the portal server/database. As a result of such purely technological problems, many users may leave the attempt to submit data to the server and therefore refuse to buy tickets using the portal.

3.2. Fuzzy model of the ticket booking process

The general scheme of users’ access to the portal can be represented by a fuzzy model. The fuzzy model is a directed graph, its vertices corresponding to the events (i.e., user actions). The arcs denote the time dependencies. If some user action is preceded by (not necessarily immediately) another action, this dependence is denoted in the graph by an arc from the preceding action to the following one. To derive a fuzzy model Fuzzy Miner Plugin for ProM framework was used [3].

The model contains information about the frequency of events occurrence and other characteristics. Fig. 6 shows an example of diagram fragment where «SUR-

NAME FILL» and «WINDOW LOAD» vertices correspond to the actions of completing «Name» field and loading the page, respectively. For each node a relative frequency of occurrence of an event in the log is shown. For the arcs a relative frequency of existence of a temporal relationship between two events in the log was derived. The indicated «correlation» (see Fig. 6) is calculated on the basis of event name similarity and matching of common attribute values.



Fig. 6. A fragment of the fuzzy model of the complete event log

The fuzzy model contains only the elements with numerical characteristics above a certain threshold value. This makes the model more compact and allows considering only significant elements and connections which define patterns in the analyzed process.

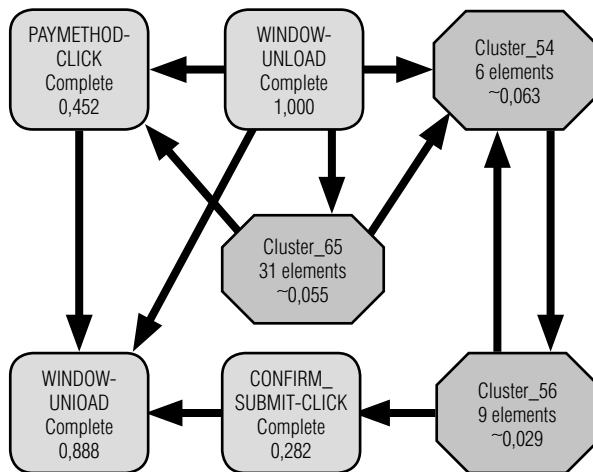


Fig. 7. The simplified fuzzy model of the complete event log

The fuzzy model (Fig. 7), as supported by a ProM plug-in, helps to group the sets of events into clusters and to hide excessive details.

On the basis to the generated models, we can conclude that among the most common user actions, that precede (but not necessarily immediately) the closing of the portal page, are the actions of opening the portal page, selecting a method of payment and confirming the booking of tickets.

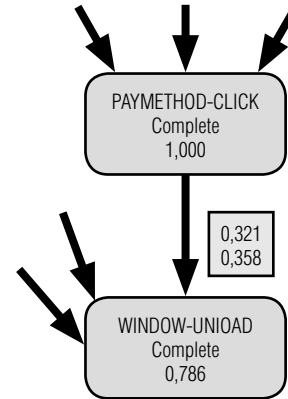


Fig. 8. Dependence of closing the portal page on viewing information about payment methods

By filtering out the traces containing accomplished orders from the log (i.e., «ORDER_STATUS» attribute value is set to «finalized»), we can make assumptions about the reasons for users to leave the portal. On a fragment from the detailed fuzzy model (Fig. 8) we can see that the relative frequency of the identified relation between the actions of closing the portal page and choosing a payment method is calculated as 0.321.

However, to obtain this and other dependencies more explicitly it is necessary to filter out (sanitize) the log by removing all traces containing only two events of opening and closing the portal page.

A fragment of the fuzzy model built for the traces that contain more than two events is presented in Fig. 9.

This fuzzy model allows us to conclude that for the given event log in 36.7 % of cases the closing of the portal page (not necessarily immediately) is preceded by a reservation confirmation, in 32 % of cases – by viewing the payment methods, in 12.9 % of cases – by the addition of insured persons and in 8.1 % – by reading the terms of payment.

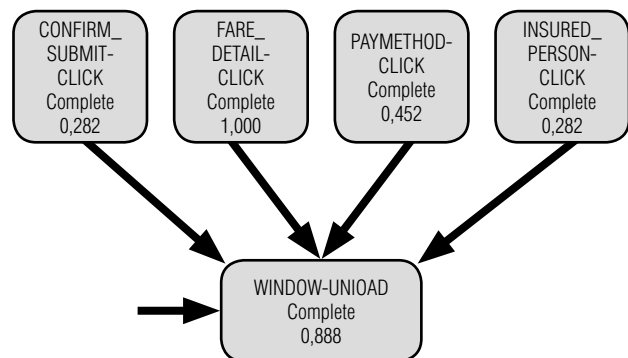


Fig. 9. A fragment of the fuzzy model built for the traces with more than two events

This suggests that for the case of unfinished ticket acquisition, the most frequent activities before leaving the portal are (1) reservation, (2) viewing payment methods, (3) adding insured persons, (4) reading the terms of payment.

3.3. Heuristic model of the ticket booking process

Frequency characteristics of the log can be represented using a heuristic model. The heuristic model is a directed graph whose vertices correspond to the events. For each vertex (event) the number of traces that contain this event is indicated. Two graph vertices are connected by an arc if the corresponding two events in the event log follow one another directly. For each arc its frequency parameter (a number of traces containing the corresponding dependency) is given. The heuristic model contains the arcs with the frequency characteristics exceeding a certain threshold value. Heuristic miner was used to obtain a model of this type [17].

We can see that after a user has opened the portal page, in 58.37 % its closing is performed (Fig. 10).

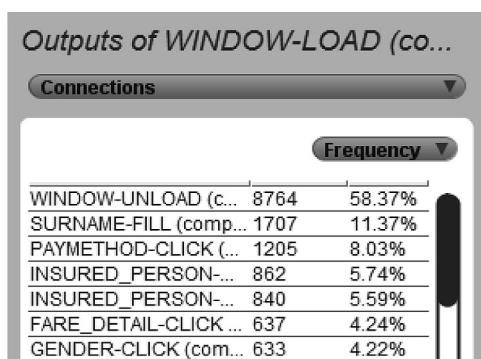


Fig. 10. The frequency characteristics of outgoing links of the page opening action

This model can determine a user's actions preceding closing the portal page. The easiest way to do this is to derive a model from an event log containing neither the traces with «finalized» order status, nor the traces that are formed only by two events: opening and closing the portal page. A fragment of the heuristic net filtered according to these constraints on the event log is presented in Fig. 11.

Note that the most frequent events immediately preceding the user's leaving the portal page (Fig. 12) are the following:

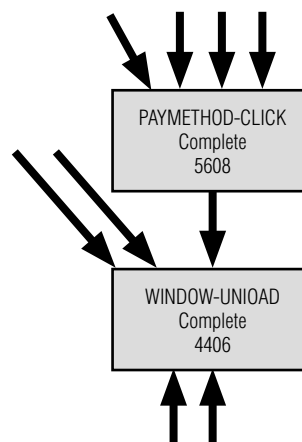


Fig. 11. A fragment of the heuristic net for the filtered event log

- ◆ verification of payment method («PAYMETHOD-CLICK») – 43,14 %,
- ◆ confirmation of booking («CONFIRM_SUBMIT-CLICK») – 38,29 %,
- ◆ acceptance of the fare conditions («ACCEPT-CHECK») – 33,88 %,
- ◆ removal of an insurance policy («INSURED_PERSON-UNCHECK») – 20,47 %,
- ◆ filling in the e-mail field («C_EMAIL-FILL») – 16,97 %,
- ◆ adding an insurance policy («INSURED_PERSON-CHECK») – 12,43 %,
- ◆ reading the information about the terms of payment («FAIR_DETAIL-CLICK») – 11,47 %.

Each case needs to be considered individually. Checking the method of payment occurs in most cases immediately after the user opens the ticketing page (Fig. 13).

Furthermore, the user usually presses the button to select a payment method more than once, and after selecting a payment method leaves the portal page.

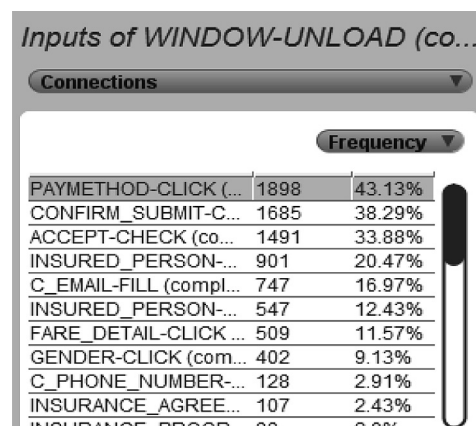


Fig. 12. Events immediately preceding the closing of the portal page

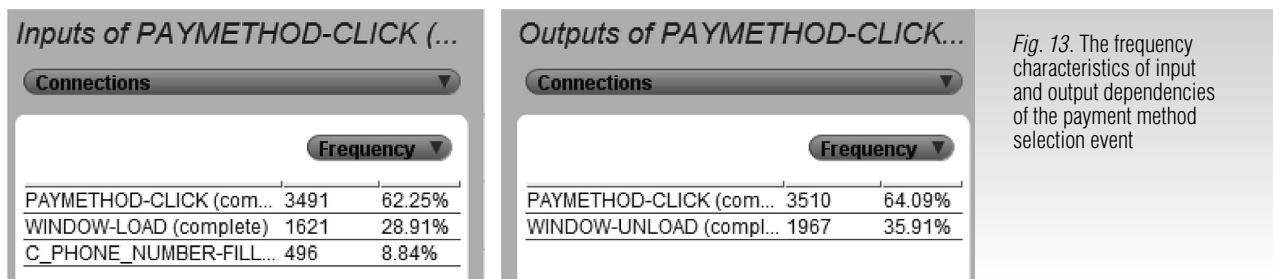


Fig. 13. The frequency characteristics of input and output dependencies of the payment method selection event

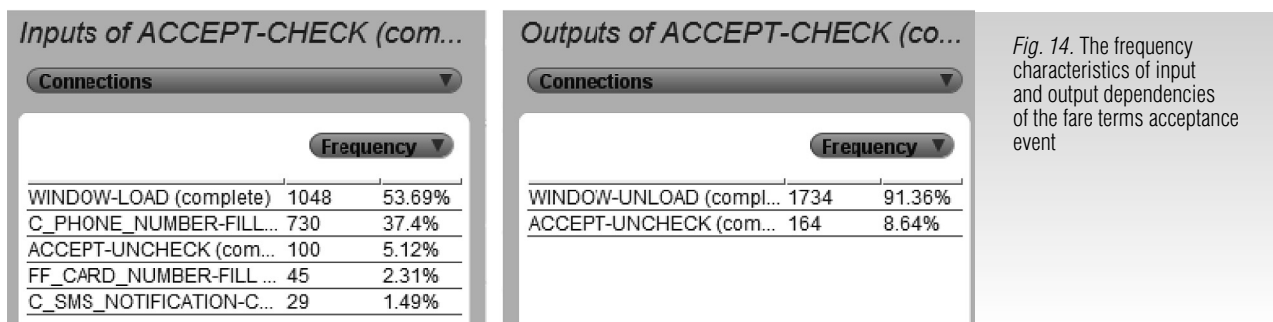


Fig. 14. The frequency characteristics of input and output dependencies of the fare terms acceptance event

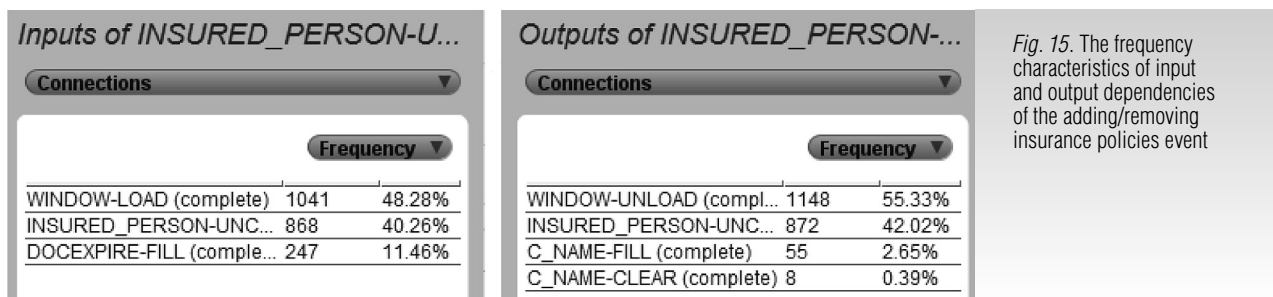


Fig. 15. The frequency characteristics of input and output dependencies of the adding/removing insurance policies event

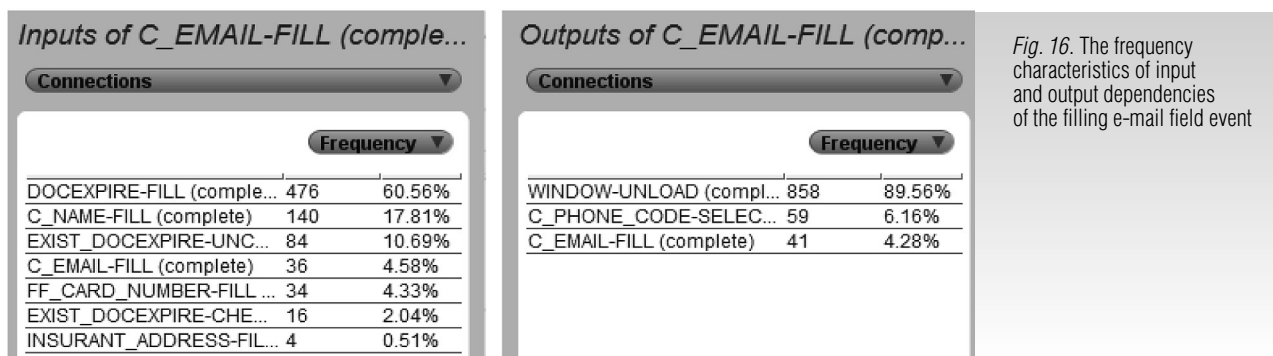


Fig. 16. The frequency characteristics of input and output dependencies of the filling e-mail field event

It was verified that for all the traces with the «reservation made» order status a reservation confirmation event precedes an event of closing the portal page. For other traces occurrence of a reservation confirmation event did not lead to an accomplished order or was accidental (i.e., it immediately followed the portal page loading event).

The event of accepting the fare conditions («ticking») leads to the user leaving the portal page. There is another variant: the events of rejection/acceptance of the

fare terms occur in cycle, which also leads to the user leaving the portal page (Fig. 14). The acceptance of the fare conditions is preceded by loading the portal page or a standard set of events of filling the form.

The events of adding/removing insurance policies for passengers in most cases lead to the leaving the portal event (Fig. 15).

In most cases, after filling the e-mail field user leaves the portal without specifying the phone number (Fig. 16).



Fig. 17. The frequency characteristics of input and output dependencies of the reading information about payment terms event

Reading the terms of payment in many cases immediately precedes closing the portal page (Fig. 17).

3.4. Workflow model of the ticket booking process

On the basis of the fuzzy and heuristic models, as well as from the results of a detailed study of the event log and the portal web site, a formal model of ticket booking process was developed.

The formal workflow model provides insight into the structure and complexity level of the process. This model reflects possible ways of a user's interaction with the portal. For example, entering first name, surname and date of birth can occur in any order (surname-name-date, date-name-surname, name-surname-date). Checking this model against the event log [21, 22] showed that the process of browsing the site has a highly linear structure. Overall consistency of the developed model with the event log was only about 20 % (whereas ideal models have consistency of 90-93 %).

The event log was filtered so that 40 % of the most common events remained. It is possible to build a much more accurate (but not ideal) model corresponding to the actions of portal users reproduced in the most sessions of the portal.

The most common user strategy, according to the log, is: to enter name, date of birth and document details, to enter frequent flyer program data, to uncheck ordering an additional insurance policy, to enter contact information (phone number, e-mail address), to accept the terms of service, to select the payment method and to submit data to the server. Conformance checking shows that this scheme fully corresponds to about 46 % of the traces in the system event log.

This means that rather a substantial part of the system traces fall out of the scheme. For example, as shown earlier, there are quite a large number of traces that contain duplicate user actions, multiple clicks on the same element, etc. On the other hand, there are traces containing only page loading and unloading events. Quite a large percentage of the traces consist

of three events: page loading, clicking on the button of payment method selection, and page unloading. On the whole, the button of payment method selection was used in the vast (over two thirds) majority of the valid log traces. The log shows that a large number of traces (about a third of approximately 46 % corresponding to the given scheme) contain the event sequence «clicking on the payment method selection button» and «unloading the page». This indicates that attention should be paid to the provided payment methods.

When checking the conformance of the model and the event log, it was revealed that about a third of traces (46 % of valid traces) contain an insurance cancellation event. Indeed, when the page is loaded «order additional insurance» is checked. Many people refused taking this insurance and unchecked the box.

3.5. Analysis results

The analysis of the general model shows that *portal users most often change their mind about ordering tickets when selecting a payment method*, which is mainly related to the different ticket prices of portal acquiring for bank cardholders and the other users. In addition, *many users leave the portal after viewing the fare conditions and payment terms* that may not be suitable for them. *Users may not decide whether to add or remove insurance for passengers* and eventually also leave the portal. Another reason for users to leave the portal should be noted: after specifying their e-mail address in the contact information section *people are hesitant to specify their phone and close the portal page*.

A *reassessment* and, perhaps, a change (or an addition) *to payment methods have been recommended* to the portal owner. *The analysis also shows the need to revise the fares policy* in order to increase the number of tickets purchased.

As a result of the research, a hypothesis has been generated as to why the portal is inefficient and recommendations on how to change the ticket e-trade information system to reduce the amount of unfinished cases (see Tab. 3).

Table 3.

The recommendations

1.	To update the content of the portal.
2.	To change the scheme for offers displayed.
3.	To improve the hardware performance.
4.	To increase the bandwidth of the channel between the user interface and the portal server/database.
5.	To better the functionality (i.e., to improve the purchase scheme).
6.	To reassess the payment methods.
7.	To revise the fares policy.

All the recommendations have been presented to the portal owner. After making the proposed changes, a new accumulated data log has to be submitted to further study. Both logs, the old and the new ones, can be the basis for the use of additional process mining methods, more detailed analysis and identification of hidden patterns of the services acquisition process.

4. Conclusions

In this work we have analyzed a ticket e-trade system on the basis the observed behavior. The analysis approaches are quite general and can be reused. Moreover, we think that the revealed problems are typical for the modern e-trade systems. The crucial question for e-trade systems is: how to increase the number of purchases? An answer to this (not the only one) is the following: to exclude rea-

sons for users to leave the e-shop without any purchase. Process mining, as a very powerful approach for analyzing processes, could help to obtain new insights into real business processes information system involved.

When analyzing processes, it is essential to enlist the services of a domain expert. Evaluation of the situation requires interpretation of the results. Without adaptation and a wise selection of methods for a specific case, the results derived by software are meaningless. Some methods (fuzzy modeling, for example) can give only vague and rough estimates of a process. One should treat them with care and turn to more precise models if necessary. Another key step is interpretation. One has to be very careful in order to interpret the result in the right way. For instance, accuracy of the evaluation results strongly depends on software, data, and settings used.

The assumptions made during the study can be used for changing processes in the e-trade system to achieve better performance or to optimize costs. Thus, using process mining techniques, one can substantially support business process improvement, whereas conventional approaches are more time consuming and more subjective.

5. Acknowledgements

This work is output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE). ■

References

1. van der Aalst W.M. (2011) *Process mining: discovery, conformance and enhancement of business processes*. Springer.
2. IEEE (2012) Task force on process mining. Process mining manifesto. *Business process management workshops*, Berlin, Heidelberg: Springer, pp. 169-194.
3. Verbeek H.M.W., Buijs J.C.A.M., van Dongen B.F., van der Aalst W.M. (2010) Prom 6: The process mining toolkit. *Proceedings of the BPM Demonstration Track*, 615, pp. 34-39.
4. Günther C.W., van der Aalst W.M. (2007) Fuzzy mining—adaptive process simplification based on multi-perspective metrics. *Business Process Management*, Berlin, Heidelberg: Springer, pp. 328-343.
5. Mans R.S., van der Aalst W.M., Vanwersch R.J., Moleman A.J. (2013) Process mining in healthcare: Data challenges when answering frequently posed questions. *Process Support and Knowledge Representation in Health Care*, Berlin, Heidelberg: Springer, pp. 140-153.
6. Kirchner K., Herzberg N., Rogge-Solti A., Weske M. (2013) Embedding conformance checking in a process intelligence system in hospital environments. *Process Support and Knowledge Representation in Health Care*, Berlin, Heidelberg: Springer, pp. 126-139.
7. Mans R.S., Schonenberg M.H., Song M., van der Aalst W.M., Bakker P.J. (2009) Application of process mining in healthcare—a case study in a Dutch hospital. *Biomedical Engineering Systems and Technologies*, Berlin, Heidelberg: Springer, pp. 425-438.

8. Yang W.S., Hwang S.Y. (2006) A process-mining framework for the detection of healthcare fraud and abuse. *Expert Systems with Applications*, no. 31 (1), pp. 56-68.
9. Rebuge Á., Ferreira D.R. (2012) Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, no. 37 (2), pp. 99-116.
10. Jans M., Alles M., Vasarhelyi M. (2011) *Process mining of event logs in internal auditing: a case study*. Paper presented at the 2nd International Symposium on Accounting Information Systems, Rome, 2011.
11. Jans M., Alles M., Vasarhelyi M. (2010) *Process mining of event logs in auditing: Opportunities and challenges*. Paper presented at the International Symposium on Accounting Information Systems, Orlando, 2010.
12. Jans M., Alles M., Vasarhelyi M. (2013) The case for process mining in auditing: Sources of value added and areas of application. *International Journal of Accounting Information Systems*, no. 14, pp. 1-20.
13. Jans M., Alles M., Vasarhelyi M. (2012) Process Mining of Event Logs in Auditing: A Field Study of Procurement at a Global Bank. Proceedings of the *9th International Conference on Enterprise Systems, Accounting and Logistics (ICESAL 2012)*, June 3-5, Chania, Crete, Greece, pp. 7-31.
14. Huang Z.M., Cong Q.S., Hu J.B. (2012) Information system risk auditing model based on process mining. Proceedings of the *IEEE International Conference on Management Science and Engineering (ICMSE)*, September 20-22, Dallas, TX, USA, pp. 39-45.
15. Suriadi S., Wynn M.T., Ouyang C., ter Hofstede A.H., van Dijk N.J. (2013) Understanding process behaviours in a large insurance company in Australia: A case study. *Advanced Information Systems Engineering*, Berlin, Heidelberg: Springer, pp. 449-464.
16. Maggi F.M., Mooij A.J., van der Aalst W.M. (2013) Analyzing vessel behavior using process mining. *Situation Awareness with Systems of Systems*, New York: Springer, pp. 133-148.
17. Weijters A.J.M.M., van der Aalst W.M., De Medeiros A.A. (2006) *Process mining with the heuristics miner-algorithm*. Technische Universiteit Eindhoven, Tech. Rep. WP, no. 166.
18. van der Aalst W.M., Reijers H.A., Weijters A.J., van Dongen B.F., Alves de Medeiros A.K., Song M., Verbeek H.M.W. (2007) Business process mining: An industrial application. *Information Systems*, no. 32 (5), pp. 713-732.
19. Verbeek H.M.W., Buijs J.C., Van Dongen B.F., van der Aalst W.M. (2011) XES, xESame, and proM 6. *Information Systems Evolution*, Berlin, Heidelberg: Springer, pp. 60-75.
20. Vaswani V., Smith P. (2004) *MySQL: The complete reference*. McGraw-Hill/Osborne.
21. Rozinat A., van der Aalst W.M. (2006) Conformance testing: Measuring the fit and appropriateness of event logs and process models. *Business Process Management Workshops*, Berlin, Heidelberg: Springer, pp. 163-176.
22. Rozinat A., van der Aalst W.M. (2008) Conformance checking of processes based on monitoring real behavior. *Information Systems*, no. 33 (1), pp. 64-95.

ПРИМЕНЕНИЕ МЕТОДОВ PROCESS MINING В АНАЛИЗЕ СИСТЕМ ЭЛЕКТРОННОЙ ТОРГОВЛИ: ПРАКТИЧЕСКОЕ ИССЛЕДОВАНИЕ

А.А. МИЦЮК

аналитик Международной научно-учебной лаборатории процессно-ориентированных информационных систем, Национальный исследовательский университет «Высшая школа экономики»

*Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: amitsyuk@hse.ru*

А.А. КАЛЕНКОВА

кандидат физико-математических наук, научный сотрудник Международной научно-учебной лаборатории процессно-ориентированных информационных систем, Национальный исследовательский университет «Высшая школа экономики»

*Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: akalenkova@hse.ru*

С.А. ШЕРШАКОВ

научный сотрудник Международной научно-учебной лаборатории процессно-ориентированных информационных систем, Национальный исследовательский университет «Высшая школа экономики»

*Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: sshershakov@hse.ru*

В. ван дер ААЛСТ

Prof.dr.ir., научный руководитель Международной научно-учебной лаборатории процессно-ориентированных информационных систем, Национальный исследовательский университет «Высшая школа экономики»; профессор Технического университета города Эйндховен (Нидерланды)

*Адрес: P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands
E-mail: w.m.p.v.d.aalst@tue.nl*

Системы электронной торговли применяются повсеместно в автоматизации торговли. Неэффективность и узкие места в процессах электронных продаж ведут к коммерческим потерям. Традиционные подходы, применяемые для выявления проблем при выполнении процессов, требуют большого количества времени и сильно зависят от субъективных оценок. В статье предложен новый подход, основанный на применении методов process mining. Методы process mining предназначены для извлечения, анализа, исправления и усовершенствования бизнес-процессов. При этом используется информация о реальном поведении информационной системы, записанная в так называемом журнале событий. В данной работе анализируется бизнес-процесс, исполняемый информационной системой онлайн-бронирования и продажи билетов. Разобран конкретный пример практического применения предложенного подхода. Показаны результаты применения методик Process mining для анализа информационной системы электронной торговли. Основываясь на этих результатах, выдвинуты гипотезы и предложены пути для усовершенствования бизнес-процессов, обеспечивающие улучшение экономических показателей функционирования информационной системы бронирования и продажи билетов. Рекомендации, сформулированные по итогам анализа логов событий системы, приводятся в этой работе для иллюстрации реальных возможностей, выгод и недостатков применения process mining. Предложенный подход обобщается для применения к широкому спектру

информационных систем электронной торговли. В работе использовалась программная среда ProM, состоящая из множества подсистем, реализующих различные методы process mining. Инструменты автоматического анализа логов событий необходимы для решения поставленных задач, однако необходимо избегать ошибок, связанных, прежде всего, с неправильной или неточной интерпретацией результатов работы методов. В статье показаны возможные трудности и подводные камни, возникающие при решении практических задач с использованием process mining.

Ключевые слова: process mining, извлечение и анализ процессов, анализ процессов, анализ данных, системы электронной торговли.

Литература

1. van der Aalst W.M. Process mining: discovery, conformance and enhancement of business processes. Springer, 2011.
2. IEEE Task force on process mining. Process mining manifesto // Business process management workshops. Berlin, Heidelberg: Springer, 2012. P. 169-194.
3. Verbeek H.M.W., Buijs J.C.A.M., van Dongen B.F., van der Aalst W.M. Prom 6: The process mining toolkit // Proceedings of the BPM Demonstration Track, 615, 2010. P. 34-39.
4. Günther C.W., van der Aalst W.M. Fuzzy mining—adaptive process simplification based on multi-perspective metrics // Business Process Management. Berlin, Heidelberg: Springer, 2007. P. 328-343.
5. Mans R.S., van der Aalst W.M., Vanwersch R.J., Moleman A.J. Process mining in healthcare: Data challenges when answering frequently posed questions // Process Support and Knowledge Representation in Health Care. Berlin, Heidelberg: Springer, 2013. P. 140-153.
6. Kirchner K., Herzberg N., Rogge-Solti A., Weske M. Embedding conformance checking in a process intelligence system in hospital environments // Process Support and Knowledge Representation in Health Care. Berlin, Heidelberg: Springer, 2013. P. 126-139.
7. Mans R.S., Schonenberg M.H., Song M., van der Aalst W.M., Bakker P.J. Application of process mining in healthcare—a case study in a Dutch hospital // Biomedical Engineering Systems and Technologies. Berlin, Heidelberg: Springer, 2009. P. 425-438.
8. Yang W.S., Hwang S.Y. A process-mining framework for the detection of healthcare fraud and abuse // Expert Systems with Applications. 2006. No. 31 (1). P. 56-68.
9. Rebuge Á., Ferreira D.R. Business process analysis in healthcare environments: A methodology based on process mining // Information Systems. 2012. No. 37 (2). P. 99-116.
10. Jans M., Alles M., Vasarhelyi M. Process mining of event logs in internal auditing: a case study // Paper presented at the 2nd International Symposium on Accounting Information Systems, Rome, 2011.
11. Jans M., Alles M., Vasarhelyi M. Process mining of event logs in auditing: Opportunities and challenges // Paper presented at the International Symposium on Accounting Information Systems, Orlando, 2010.
12. Jans M., Alles M., Vasarhelyi M. The case for process mining in auditing: Sources of value added and areas of application // International Journal of Accounting Information Systems. 2013. No. 14. P. 1-20.
13. Jans M., Alles M., Vasarhelyi M. (2012) Process Mining of Event Logs in Auditing: A Field Study of Procurement at a Global Bank // Proceedings of the 9th International Conference on Enterprise Systems, Accounting and Logistics (ICESAL 2012), June 3-5, Chania, Crete, Greece. P. 7-31.
14. Huang Z.M., Cong Q.S., Hu J.B. (2012) Information system risk auditing model based on process mining // Proceedings of the IEEE International Conference on Management Science and Engineering (ICMSE), September 20-22, Dallas, TX, USA. P. 39-45.
15. Suriadi S., Wynn M.T., Ouyang C., ter Hofstede A.H., van Dijk N.J. Understanding process behaviours in a large insurance company in Australia: A case study // Advanced Information Systems Engineering. Berlin, Heidelberg: Springer, 2013. P. 449-464.
16. Maggi F.M., Mooij A.J., van der Aalst W.M. Analyzing vessel behavior using process mining // Situation Awareness with Systems of Systems. New York: Springer, 2013. P. 133-148.
17. Weijters A.J.M.M., van der Aalst W.M., De Medeiros A.A. Process mining with the heuristics miner-algorithm. Technische Universiteit Eindhoven. Tech. Rep. WP. No 166. 2006.
18. van der Aalst W.M., Reijers H.A., Weijters A.J., van Dongen B.F., Alves de Medeiros A.K., Song M., Verbeek H.M.W. Business process mining: An industrial application // Information Systems. 2007. No. 32 (5). P. 713-732.
19. Verbeek H.M.W., Buijs J.C., Van Dongen B.F., van der Aalst W.M. XES, xESame, and proM 6 // Information Systems Evolution. Berlin, Heidelberg: Springer, 2011. P. 60-75.
20. Vaswani V., Smith P. MySQL: The complete reference. McGraw-Hill/Osborne, 2004.
21. Rozinat A., van der Aalst W.M. Conformance testing: Measuring the fit and appropriateness of event logs and process models // Business Process Management Workshops. Berlin, Heidelberg: Springer, 2006. P. 163-176.
22. Rozinat A., van der Aalst W.M. Conformance checking of processes based on monitoring real behavior // Information Systems. 2008. No. 33 (1). P. 64-95.

BARCODE GENERATION FOR FACE IMAGES¹

Georgy KUKHAREV

Full professor, Department of Multimedia Systems, West Pomeranian University of Technology
Address: 49, Żołnierska, Szczecin, 70-310, Poland
E-mail: gkukharev@wi.zut.edu.pl

Yuri MATVEEV

Professor of the Department of Speech Information Systems,
Saint Petersburg National Research University of Information Technologies,
Mechanics and Optics (ITMO University); Chief Researcher, STC Innovation Ltd.
Address: 49, Kronverkskiy av., St. Petersburg, 197101, Russian Federation
E-mail: matveev@mail.ifmo.ru

Nadezhda SHCHEGOLEVA

Associate Professor, Department of Software Engineering and Computer Applications,
Faculty of Computer Technologies and Informatics,
Saint Petersburg Electrotechnical University (LETI University)
Address: 5, Professora Popova str., St. Petersburg, 197376, Russian Federation
E-mail: nlschegoleva@etu.ru

The paper addresses the problem of linear barcode generation for face images. The history of the problem and possible approaches to its solution in mobile oriented systems are discussed.

Two methods are presented: the first one is based on intensity histograms, and second one is based on intensity gradients, calculated over images using their original features. Then these features are averaged over limited number of intervals, quantized in the range of decimal digits from 0 to 9 and converted into standard barcode. Structure of barcode generation system is proposed, and description of its blocks is presented.

The methods have been tested by using «Face94», «Face Sketch FERET Database» databases, as well as a database composed of different age faces. The tests have demonstrated invariance of barcode in respect to changes in local sizes of face images, in tilt in the XY plane, to changes in the view, mirror rotation about vertical axis, as well as changes in facial expressions and age-related face changes.

Therefore, the presented methods constitute novel solutions to practical applications in real conditions of dynamical parameters changes in face images. Moreover, both methods require neither big computational resources, nor using of special software packages for image processing, allowing generation of linear barcode in real time systems. A generated standard barcode contains information about a person's face and can be used for indexing, identification, recognition, and searching for people.

Key words: face images, barcode, real time systems.

Introduction

Barcode encoding techniques are used for various goods identification, in finance documents, payments, advertising materials and services,

and have established wide and everyday applications. Reading of barcodes is fast and simple by using specialized tools, and comprehension of their content poses no problem, because decoding is solved in the frame of list search.

¹ This work was partially financially supported by the Government of the Russian Federation, Grant 074-U01.

Standard barcode for people identification was proposed for the first time ever in the patent [1] only in 1999. It was assumed that identification took place when making electronic payment – i.e. in real time, by using the unique barcode printed on his hand and read by special head.

But this technique of people identification with barcode has not found wide application, despite implementation of biometrical standards for anthropometric identification, and biometrical methods of people identification have better ability than people themselves. Barcode printed on skin, however, are now advertised as fashionable tattoo [2]. Barcodes printed on skin does not contain any biometrical characteristics of a person [2]. But we can expect with certainty that if a barcode contains some biometrical characteristics, it will be accepted not only as «fashionable token» but also as usable way of people identification. It is hard to imagine the way of identification using barcode hidden under the clothing, or on the body in hardly accessible place and uncomfortable for reading, as presented in *Fig. 1*.



Fig. 1. Examples of tattoos [2]

An ideal solution will be to generate a barcode only when it is needed, and in real time – e.g. based on face or voice. In such case identification can be made «from distance» – and not observable by an individual, not requiring any activity and storing generated code on his/her skin.

Assuming that such task can be solved in principle, obtained solution can be used to encode a face image or voice in the form of barcode. Such codes can be used in mobile biometrical systems of access control AC, portable interface systems «man – computer» for user identification, for people recognition in video monitoring systems recording face or/and voice.

This paper will focus on one of possible information sources: face or voice, i.e. face images, presenting a problem leading to solution, it implies encoding face image with a barcode. In practice of using face image

as the source of biometrical information in the form of barcode a problem is associated with dynamical change of image parameters in real conditions (e.g. tilt and direction of face region in the image, image intensity and facial expression). This means that basically it is impossible to represent a face image using invariant features, which don't depend on dynamic changes in mentioned parameters. As a result, the lack of invariance leads to generation of unstable barcodes, independent of its form. In AC systems variability of input face images is not big, because of using stable image exposition conditions (face illumination, controlled rotation and face expression, etc.), also in this application obtaining stable barcodes is problematic.

For these reasons the problem of stable representation of face images with barcodes has remained relevant, practically since the first application of a computer system for people recognition based on their face images, also for intensively developing mobile systems, using new technology base – tablets and smartphones.

1. Short review of existing approaches

10 years after submission of the patent [1], psychologists S.C. Dakin and R.J. Watt noticed in their paper [3], that all information about human face is contained in horizontal lines, like eyebrows line, eyes line and lips line. In addition, they have shown that this information can be successfully represented as a set of white and black lines (i.e. as some kind of barcode, named «biological barcode» by the authors), paying attention to some invariant characteristics of «biological barcode of human face».

Example of this representation of face region is shown in *Fig. 2*, where: *a* – original face image; *b* – variant of extracted face characteristics; *c* – «biological barcode» of human face; *d* – variant of standard barcode, representing required solution.

However, in the paper [3] and further publications of these authors (e.g. [4]), they have not presented any algo-

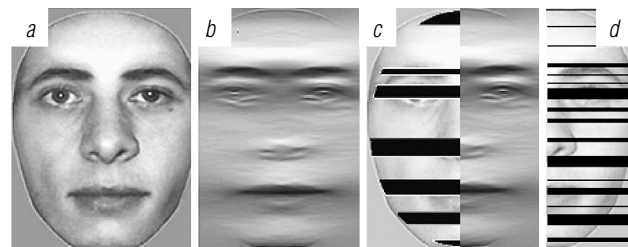


Fig. 2. Ways of representing face region [3, 4]

rithm for generation of similar barcodes, and concluded that representation of face region only using thick white and black bars cannot be the exact model of a specific person face. Such model constitutes precise «biological barcode of face», at the same time it represents «very rough approximation» of specific person face, what is visible in Fig. 2, comparing its «c» – «d» parts.

Approximate character of this representation is associated with the morphology of human face – it is very difficult to detect border hair/forehead, border lines of eyebrows and their specific features, borders of eyes lines, border of nose line and lips and even more difficult to describe these by using thick straight lines. More precise representation requires thinner and more complicated set of lines. It is possible that human face and corresponding barcode could look like presented in Fig. 2d, as was suggested in [5].

Within five years the most important experimental investigation concerning face representation with barcodes has appeared [6]. The authors have presented an algorithm to generate face barcode HCC2D, encompassing advantages of two-dimensional code HCCB (High Capacity Color Barcode) and two-dimensional code QR, distinguishing high density of information concentration. In the framework of real time and mobile systems, using tablets and smartphones, the algorithm proposed in [5] has little chance to be implemented in the near future. Such conclusion can be supported by two serious flaws of the algorithm for generation of two-dimensional color barcode HCC2D:

- ◆ the algorithm [6] implements a rather complicated process of singular (key) points searching in face image and the algorithm for describing its neighborhood by using SIFT (Scale Invariant Feature Transform) and SURF (Speeded Up Robust Features) procedures;

- ◆ the result of singular (key) point search with SURF algorithm is not invariant in scale, requiring in consequence decomposition of an original face image into pyramid of multiscale regions and multiple run of SURF method for each region of the pyramid (making the whole algorithm of key points searching even more complex).

Consequently, the algorithm presented in [6] for generation of two-dimensional color barcode HCC2D requires high throughput computer system, supported with dedicated image processing software packages.

Finally, representation of face image by two-dimensional code (i.e. HCC2D) is a departure from the concept of linear «biological barcode» [3] and will not be considered further in this paper.

The closest to the concept of linear «biological barcode» is the algorithm presented in monograph [7, pp. 170 and 214]. Fig. 3 illustrates the concept of this algorithm.

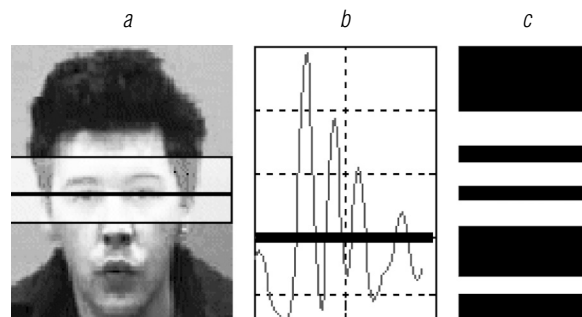


Fig. 3. Diagram of forming binary codes [7]

At the first phase the intensity gradient for symmetrical (mirror) located windows slides over a face image in top – down direction, as shown in segment *a*. Then the difference between current values of gradients and their mean values is calculated, and obtained differences are coded. Differences of gradients, which are less than mean value, are coded as «0», and those, which are greater than mean value, as «1», resulting in binary like code, representing each original image. The form of resulting code is similar to «biological barcode» [3]. These results are shown in Fig. 3 segments 3: *b* – current gradient values and its mean value; *c* – «biological barcode» for original image.

This approach [7] has the drawback, namely we cannot obtain the same binary code for face images of the same person, but taken in different exposition conditions. Such changes include small, but «eye» perceptible changes in face region size, its tilt (in plane XY), face expression and difference in intensities of some face areas caused by local shadows. This flaw makes application of the mentioned approach impossible in situations with dynamical changes in face image expositions described above, and consequently in transforming face images into the corresponding code.

General appraisal of presented methods of transforming face image into binary code suggests, that they can be used only in case of classification of generalized class of «human face images» among other classes of images, not containing faces, as was stressed by the authors of [3, 4].

This paper proposes two possible approaches for representation of a human face in the form of linear barcode of type EAN-8, EAN-13 (or UPS) in conditions of perceptible dynamical changes of face image param-

eters. Both approaches require neither big computational resources, nor specialized software packages for image processing, making possible their applications for generating linear barcodes in real time within mobile systems.

The paper is organized as follows: section 2 presents the structure and function of the barcode generation system, section 3 presents structure and functions of the barcode generation system corresponding to input face image, section 4 describes the algorithms used to generate a barcode based on face image intensity histogram, section 5 is devoted to the barcode generation algorithm based on face image difference gradients. Results of testing of the proposed algorithms and examples of generated barcodes are presented in section 6. The paper ends with short commentaries and bibliography.

2. Barcode generation system

The structure of the barcode generation system is shown in Fig. 4; it is composed of of four basic blocks: 1 – input image preprocessing; 2 – feature extractor 3 – feature encoding; 4 – barcode generator; system input and output.

System input receives an image and produces a standardized barcode corresponding to an input image. Taking as an example the AS system, we formulate the input image requirements.

The reason for selecting this application is that in AC systems a person standing in front of the camera is asked to fulfill recommendations concerning stable exposition conditions – i.e. taking the front view with neutral face expression. These requirements are necessary in order to obtain a stable barcode. In practice, however, it is hard to meet these conditions, and we are forced to allow some variations in face images input to the AC system. An example of allowed variations is presented in Fig. 5a (taken from source [8]).

These variations concern: face area size, tilt in XY plane, face expression, as well as intensities in some regions of face (caused by local shadows or spectacles).

As mentioned, block 1 (input image preprocessing) executes two basic tasks. The first one is input image analysis with its parameters estimation – size, range of colors and deflection of eye line from horizontal line. The second one implies rotation in XY plane, and correction of image size and intensity equalization, if necessary as a result of the analysis. Such problems are discussed, e.g. in monograph [6].

Often in bimodal AC systems input data include image and voice. Then face expression changes while uttering some control word, sentence or during conversation. An example of images in mentioned situations is presented in Fig. 5b. Ensuring stable image encoding in cases presented in Fig. 5 and 6, depends on selected features for face representation. This task is solved in block 2 (feature extractor). At least, two approaches are possible.

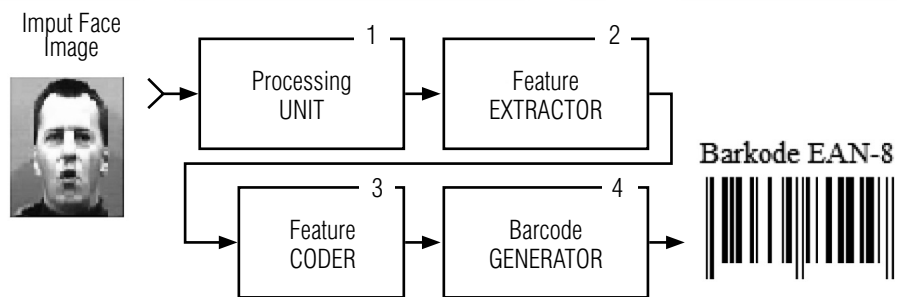


Fig. 4. Structure of system



Fig. 5. Possible changes of face image

The first approach involves representation of the whole input image using a set of features that ensures imperceptible influence of face expression. The implementation may be based, for example, on intensity histogram of input face image. As shown in [7, 9], the histogram gives high efficiency of image retrieval in big databases and frequently is the only facility in CBIR applications (Content-based image retrieval).

The second approach is feature extraction using only 3/4 of the top part of a face image, excluding the part below the middle of the nose/mouth segment (shown in Fig. 5b). The top part of face is not influenced by utterance, ensuring stability of representing its features.

In both approaches block 3 addresses the encoding task using by the necessary number of decimal digits. This is achieved by averaging the features to a limited number of quantization intervals and quantizing these in the range of decimal digits from 0 to 9. This stage eliminates the influence of dynamic change of input image parameters, which have not been taken into account at the earlier stages, on encoding result.

In this case the task of block 4 is trivial – to change the encoding result obtained in block 3 based on algorithm of a standard barcode generation [10]. Generation of a standard barcode includes evaluation of code checksum produced in block 3 and transformation of this code into a binary matrix, representing input image in graphic form of a standard barcode.

The approaches presented in this paper enable to generate linear barcodes based on face images in EAN-8 format and can be used to obtain other linear barcodes, e.g. EAN-13 and UPS [10].

3. Barcode generation based on histogram

We assume that an input image with GRAY scale has got size $M \times N$. Our task is to represent this image in the format of EAN-8 barcode.

To generate EAN-8 barcode we have to calculate the intensity histogram of an original image and to convert it into a code consisting of 7 decimal digits. Therefore, for the histogram calculation we have chosen the initial value of the parameter $\text{bins} \geq 7$. Let $\text{bins} = 8$, then the histogram of the input image evaluated originally for 256 intensity levels (from 0 to 255) will be averaged exactly in 32 intervals. As a result, we obtain a histogram H with the values of $H(b)$, for $b = 1, 2, \dots, 8$.

The next stage implements the preliminary encoding of histogram values:

$$\overline{H}(b) = f[\text{scale}(H(b) / \max(H))], b = 1, \dots, 8,$$

where: $f[\cdot]$ – the nearest integer value (removing the remainder); scale – scale factor, $9 < \text{scale} < 10$, is not an integer.

In this case values of histogram are quantized in the interval decimal digits from 0 to 9. As a result we have got the decimal number encoding with 8 digits. Subsequently, the formation of EAN-8 barcode involves only seven digits, for example, from the 1st to the 7th, or from 2nd to 8th.

Then the result is converted into the final EAN-8 barcode, consisting of 8 digits, and the 8th digit identifies the checksum for the first 7 digits. For this purpose we apply the standard algorithm, as in [10].

In conclusion, please note the following: evaluation of input image histogram requires neither big computational resources, nor special software packages for image processing. This enables to generate a barcode with «histogram method» by using the mobile systems. Moreover, most of them have built-in face area detection function, as well as barcode reading programs. Such program can be easily adopted for synthesis of barcode based on given decimal number.

4. Barcode generation based on difference gradients

This method is based on intensity difference gradients of an input image, averaging of gradients on the limited number of intervals, quantizing the result in the range of decimal digits from 0 to 9 and transformation into standard barcode according to a table.

Let an input image with GRAY scale have the size $M \times N$. Original feature extraction for the given input image is based on the calculation procedure of intensity difference gradient of two symmetrically located windows with height $H \geq 1$ pixels and length equaling the width of an input image. Windows are shifted over a face image with step $S \geq 1$. Windows move only in vertical direction from top to bottom; and for each step the distance (e.g. euclidean distance) between corresponding fragments of image is covered by the windows. Shifting of windows starts in practice on the border «hair/forehead» and finishes on lower border of nose, or just beneath nose. Difference gradients in windows, transformed into distance, enhance jumps in intensity on the border hair/forehead, borders of eyebrow line, eye line and the line nose/lips – i.e. along the lines of «biological face code». Calculated distance values represent integral

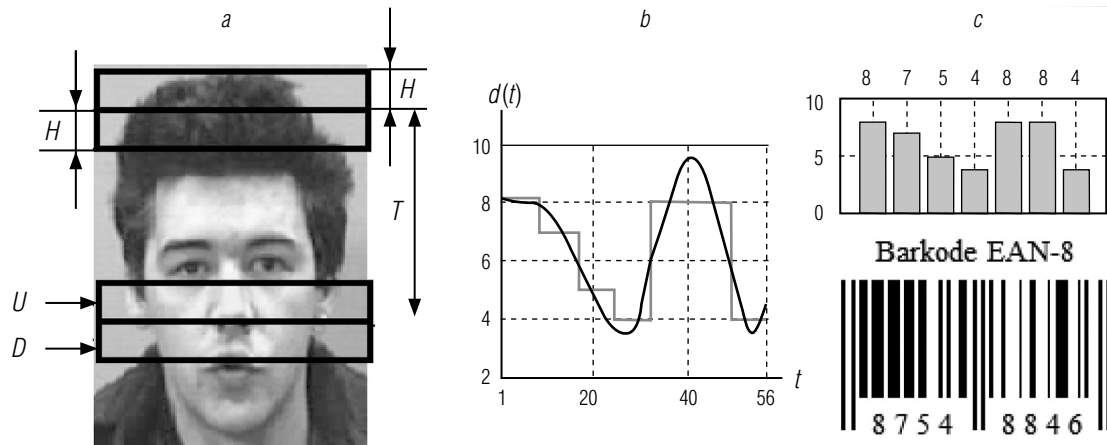


Fig. 6. Clarification of gradient method

characteristics of the least variable part of a face under the condition of discernible dynamical changes of face image parameters.

The idea of forming difference gradients and their encoding by using the windows is clarified in Fig. 6, where we use the following descriptions:

The total number of shift steps is $T = LA$, where L – code length, and A – smoothing interval.

Generation of the barcode in EAN-8 format, $L = 7$, and the parameter $A \geq 2$, generally chosen from the condition of border intersection

$$T = LA \leq (M - H). \quad (1)$$

The value of T should, for instance, touch the lower border of nose or be between nose and mouth, in order to exclude the lower part of a face, thus eliminating the influence of face expression on stability of generated barcode. From the other side, if the value of $T > (M - H)$, then it will be necessary to increase the size of input image in order to meet the condition (1). Increasing is executed in block 1 at the stage of image size correction, as described above.

Now we can determine the distance $d(t)$ between these windows:

$$d(t) = \|U_t - D_t\|, \forall t = 1, 2, \dots, T, \quad (2)$$

and normalization result:

$$\bar{d}(t) = d(t) / \max(d), t = 1, \dots, T. \quad (3)$$

The results obtained in (3) are averaged within interval A and quantized in the range of decimal digits from 0 to 9 using scaling factor scale:

$$\bar{d}(l) = f\left[\frac{scale}{A} \sum_{j=1}^A d(A(l-1) + j)\right], l = 1, 2, \dots, L, \quad (4)$$

where: $f[\cdot]$ – the nearest integer value (removing the remainder); scale – scale factor, $9 < scale < 10$, is not an integer.

The result calculated according to (4) from output of block 3 is fed at input of block 4, where the final barcode is formed, consisting of 8 digits. To this end we have used the standard algorithm described in [10].

The result calculated in (2)-(4) is shown in Fig. 6: b – normalized values of distances and their thresholding in averaging intervals A (broken curve); c (upper index) – representation of thresholding result in the form of 7 digits barcode; c (lower index) – resulting barcode.

5. Experiments

5.1. Tests of «histogram method»

The test is composed of 4 computer experiments: two with «Face 94» face database [8] and two with CUFSF database (CUHK Face Sketch FERET, [12]); and extended modification presented in [13].

The experiments with «Face 94» database have used 100 classes each containing 20 images, and EAN-8 barcodes have been generated. Input images of size 200×180 pixels have been transformed into GRAY format. No other image analysis and rotation operations have been carried out to obtain horizontal orientation of eyes line, although, as mentioned above, parameters of images were not very stable, as required, e.g. in case of AC systems. The barcode generation process has been controlled by selecting the following parameters: $bins = 8$; $scale = 9$.

A portrait with number 1 was selected from each class, and barcode was generated for the first seven values of the histogram. Then barcodes were generated also for remaining nineteen portraits of each class (from number 2

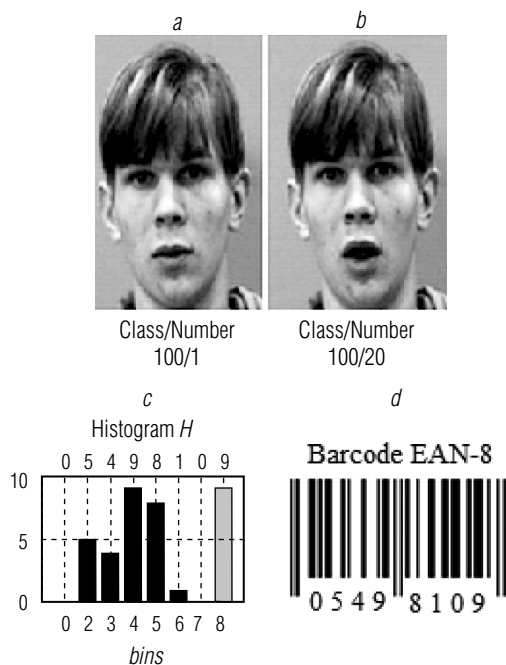


Fig. 7. Test results using «Face94» database

to 20) by using the same encoding control parameters. The test results for all 100 classes gave 670 related image pairs, having the same code. These related pairs belong exclusively to their class, giving 6 to 7 images per class. It is important to note that face expressions related to speech presented no hindrance to generate the stable barcode.

These test results are shown in Fig. 7 and present: *a* and *b* – input images (1st and 20th in class 100); *b* – histogram, consisting of first 7 values (dark strips) and the value of checksum (light strip); *c* – barcode.

Further 6 tests – checking the stability of generated barcodes obtained for reduced size images – reduction by 10%. For example, if the size of an input image is 200 × 180 pixels, the size of smallest one is – 80 × 72 (40% of input image). The total number of generated barcodes for 6 tests amounts to {661, 671, 665, 673, 675, 620} also with complete lack of doubling between classes. These results have shown that barcodes generated using intensity histograms in all conducted tests have no doubling

errors. The pair number of identical barcodes changes slightly in each test, confirming the stability of this method. We can note that the barcodes have not been changed with mirror rotation of test images around the vertical axis as well as rotation by 90° (right/left). This «invariance property» of intensity histogram makes this method applicable in real life with different variants of people images (we can call these conditions as «mobile photography»).

Experiments involving sketches were conducted with 100 classes of database [13]. Each class contained 10 sketches, the first one is the original sketch from CUFSF base [12], remaining 9 are its modifications. The modified sketches differ from the original – these have different size of the face area, shifted up or down of the central part of face, disturbed symmetry of the face area. With mentioned modifications the phase correlation between the original sketch and its modifications is on average close to 0.2 that confirms the importance of these modifications [13]. Control parameters for barcode generation had the values *bins* = 8; *scale* = 9. Control barcode was generated from original sketches, and their 9 modifications were used for the test. The tests results for total 100 classes gave 524 corresponding pairs of sketches (from 900 tested), which had the same barcodes. These corresponding pairs belonged exclusively to their respective classes, amounting to 5 or 6 sketches per class. In the experiment four pairs of classes had identical barcodes, and this can be attributed to «non-specific texture» of images, showing the sketches. However, selecting histograms with *bins* > 8 and using barcodes EAN-13, probability of doubling is practically excluded.

Fig. 8 illustrates the result of the second experiment for sketches.

Here the following parameters are shown: *a* – the original sketch from CUFSF database [12]; *b* – a modification of the first image (changes in the size of the face area with effect similar to «fish eye»); *c* – the intensity histogram is shown in the form of 0012211111234598 consisting of 16 decimal digits (this histogram is identical for two presented sketches); *d* – variants of EAN-8

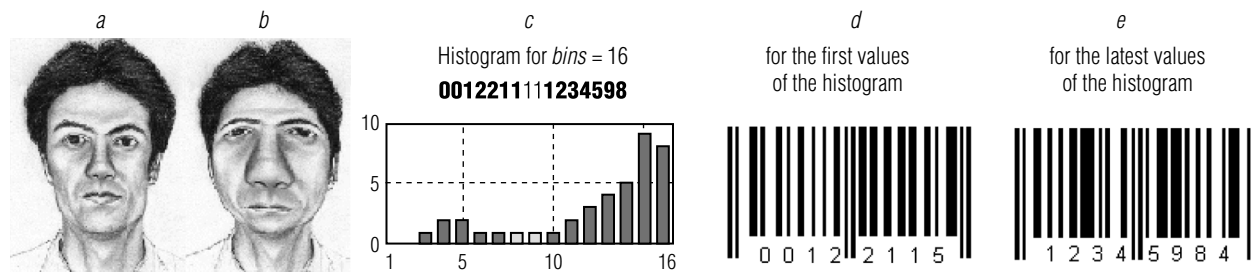


Fig. 8. Result of experiment with CUFSF database

barcode evaluated on the first seven values of the histogram; *e* – variants of EAN-8 barcode evaluated on the last seven values of the histogram. These results show that changes in the sketch parameters had no influence on the stability of generated barcode.

But in real conditions we can expect some problems with application of the histogram method. In case of AC systems, which use separate lighting of face, and in case of certain face database, the histogram method provides very stable generation of a barcode in real time.

5.2. Tests for «gradient method»

Two experiments have been carried out for these tests: one – by using «Face 94» face image database [8] and the other – by using database of composed faces, representing face changes with human age [14]. «Face94» database consists of 100 classes, 11 images in each class, for which EAN-8 barcode was generated. Input images were transformed into GRAY format and scaled to the size 112×92 pixels. Images were not preprocessed in order to obtain the standard position of eye line or scale of anthropometric face characteristics.

In test 1 portrait number 1 was selected from each class, and its barcode was generated by using the following control parameters of encoding process:

$$\{H = 23; S = 1; L = 7; A = 8; T = 56; scale = 9.5\}. \quad (5)$$

Then the barcode was generated for the remaining ten portraits of each class (from number 2 to number 11) by using the same parameters of encoding process. As a result of testing all 1000 pair images were obtained, more than 700 related image pairs having the identical barcodes. These related image pairs belong exclusively to own classes. The example result for conducted test is illustrated in Fig. 9.

It is shown: *a* and *b* – the first and last image used in class 100; *c* – phase correlation between corresponding distance vectors (confirming nearly 100% of their similarity); *d* – normalized and quantized values of the distances presented in barcodes graphics; *e* – barcodes for two images. Correlation, as additional verification means, has been introduced into the test on purpose. We can note that in spite of variation in face expression, the same barcodes were obtained.

Fig. 10 illustrates the result of barcode generation for images with different face expression.

Here we have 2 pairs of images and corresponding barcodes generated in standard EAN-8.

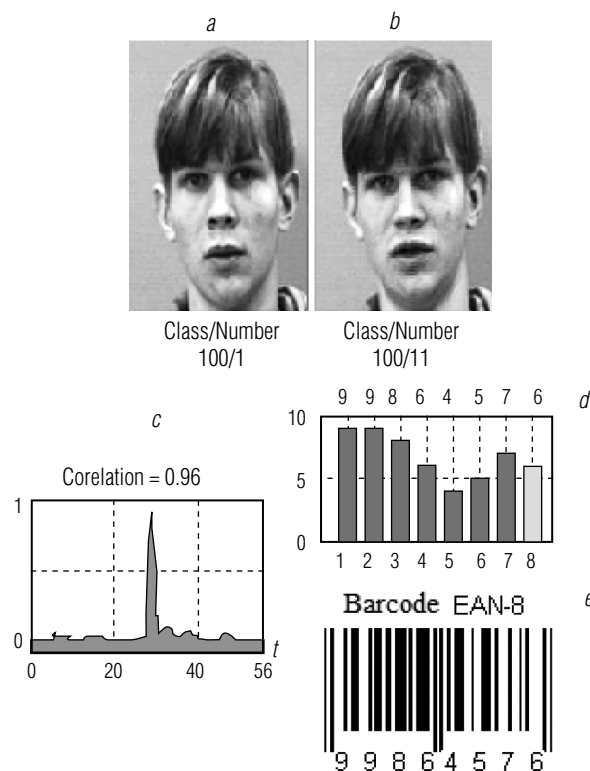


Fig. 9. Result for gradient method

We can observe the following: changes in face expression, changes of face area size, and changes in eyes area (open and closed eyes) have no influence on the generated barcode. The same results are obtained in extended test 1.

The next test aimed at checking the stability of the generated barcode after mirror rotation of input image, its preservation after additional dynamic changes of image parameters – local face size, tilt in XY plane, image rotation, changes in face expression and presence of local shadows caused by local lighting. The obtained results are illustrated in Fig. 11.



Fig. 10. Example results for images with different face expression



Fig. 11. Example results for images with mirror symmetry

Face images for each class (from number 2 to number 11) were mirror rotated about vertical axis, and the barcodes were also generated for such images. Here we can see that mirror rotation has not changed the barcodes.

The goal of the last test is to check the stability of the generated barcode with face «aging». The experiment was conducted with 38 composite faces, representing changes with human age [14] from 20-24 up to 61-64 years. The following encoding parameters were used:

$$\{H = 22; L = 7; A = 7; T = 49; scale = 9.7 \div 9.9\}, \quad (6)$$

and the parameter *scale* could have any value in the indicated interval. We can note that independent of face aging, for all of them we have get the same barcode.

The final results are illustrated in Fig. 12. Fig. description: in column *a* and *b* – faces in age of 20-24 и 60-64 years; in column *c* – phase correlation change and similarity index for the whole range of image changes (face

ages), and *d* – normalized and quantized values of the distances, presented in barcodes graphic and standard barcode for input images.

All 38 images have the same barcode. This experiment has shown that in the framework of difference gradient and with parameters (6) the barcode remains unchanged with face aging modeling.

Conclusion

The paper has discussed the problem of generating linear EAN-8 barcodes from facial images. The history of the problem has been presented, and possible approaches for mobile systems have been formulated. Two methods have been proposed: the first one is based on the intensity histogram as invariant features, which represent the original image of a face, and the second one relies on intensity gradients calculated over images using their original features also representing the characteristics of face images. Testing has been performed at the stage of the computer simulation. The tests have been conducted by involving «Face94», «CUHK Face Sketch FERET» bases and the base of composite faces simulating aging. The test results have shown that the proposed methods ensure the stability of the generated barcodes in case of mirror reflection of the original image, when there are scale, pose and facial expression variations and when there are shadows on facial images from local lighting. Furthermore, the method generates the standard barcodes directly from the facial images, and thus contains information about a person’s face. In this case the generated barcode can be used for indexing, identification, recognition, and searching for people, and proposed by the methods determined by the direction of further research and applications of barcodes in practice. ■

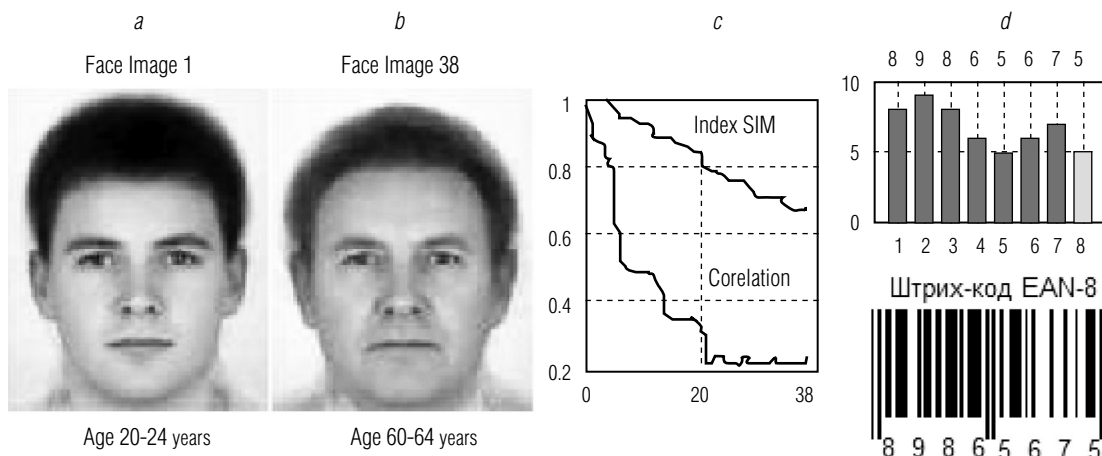


Fig. 12. Barcodes for face aging

References

1. Method for verifying human identity during electronic sale transactions – Patent US 5878155A (1999). Available at: <http://www.google.com/patents/US5878155> (accessed 15 March 2014).
2. Barcode/Tattoos. Available at: <http://www.barcodeart.com/store/wearable/tattoos/> (accessed 15 March 2014).
3. Dakin S.C., Watt R.J. (2009) Biological «bar codes» in human faces. *Journal of Vision*, vol. 9, no. 4, pp. 1–10.
4. Goffaux V., Dakin S.C. (2010) Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, vol. 1, no. 43, pp. 1–14.
5. Facial Barcodes Help Us Identify People. Available at: <http://www.barcodesinc.com/news/?p=92> (accessed 15 March 2014).
6. Querinini M., Italiano G.F. (2013) Facial recognition with 2D color barcodes. *International Journal of Computer Science and Application*, vol. 10, no. 1, pp. 78–97.
7. Kukharev G.A., Kamenskaya E.I., Matveev Yu.N., Shchegoleva N.L. (2013) *Metody obrabotki i raspoznavanija izobrazhenij lic v zadachah biometrii* [Methods of facial images processing and recognition in biometrics]. Edited by M.V.Hitrov. SPb.: Politehnika. (in Russian)
8. Face94 data base. Available at: <http://cswww.essex.ac.uk/mv/allfaces/faces94.html> (accessed 15 March 2014).
9. Forczmanski P., Kukharev G. (2007) Comparative analysis of simple facial features. *Journal Real-Time Image Processing*, no.1, pp. 239–255.
10. Barcode generation algorithm for EAN-8 и EAN-13. Available at: http://www.cherry-notes.spb.ru/barcode_ean8.htm (accessed 15 March 2014).
11. Kukharev G.A., Matveev Yu.N., Shchegoleva N.L. (2014) Ekspress-metod formirovanija shtrih-koda po izobrazhenijam lic [Express method of barcode generation from facial images]. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, no. 2 (90), pp. 99–106.
12. CUHK Face Sketch FERET Database. Available at: <http://mmlab.ie.cuhk.edu.hk/cufsf> (accessed 15 March 2014).
13. Kukharev G.A., Buda K., Shchegoleva N.L. (2014) Methods of face photo-sketch comparison. *Pattern Recognition and Image Analysis*, vol. 24, no. 1, pp. 82–93.
14. Burt D.M., Perrent D.I. (1995) Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour. *Proceedings of the Royal Society London*, vol. 259, no. 1355, pp. 137–143.
15. Kukharev G.A., Matveev Yu.N., Shchegoleva N.L. (2014) Formirovanie shtrih-koda po izobrazhenijam lic na osnove gradientov jarkosti [Creating barcodes for facial images based on intensity gradients]. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, no. 3 (91), pp. 88–95.

МЕТОДЫ ФОРМИРОВАНИЯ ШТРИХ-КОДА ПО ИЗОБРАЖЕНИЯМ ЛИЦ

Г.А. КУХАРЕВ

доктор технических наук, профессор кафедры мультимедийных систем,
факультет информатики, Западно-поморский технологический университет

Адрес: Польша, 70-310, г. Щецин, ул. Жолниерска, д. 17

E-mail: gkukharev@wi.zut.edu.pl

Ю.Н. МАТВЕЕВ

доктор технических наук, профессор кафедры речевых информационных систем,
Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики (НИУ ИТМО);

главный научный сотрудник ООО «ЦРТ-инновации»

Адрес: 197101, г. Санкт-Петербург, Кронверкский проспект, д. 49

E-mail: matveev@mail.ifmo.ru

Н.Л. ЩЕГОЛЕВА

кандидат технических наук, доцент кафедры математического обеспечения
и применения ЭВМ, факультет компьютерных технологий и информатики,
Санкт-Петербургский государственный электротехнический университет (ЛЭТИ)

Адрес: 197376, г. Санкт-Петербург, ул. Проф. Попова, д. 5

E-mail: nlschegoleva@etu.ru

В статье обсуждается задача формирования линейного штрих-кода по изображениям лиц. Представлена история задачи и возможные подходы к ее решению с ориентацией на мобильные системы.

Для решения задачи предложены два метода: первый метод основан на гистограммах яркости, второй – на разностных градиентах яркости, представляющих изображения лиц в форме исходных признаков. Далее в каждом методе эти признаки усредняются на ограниченном числе интервалов, квантуются в диапазоне десятичных цифр от 0 до 9 и преобразуются в стандартный штрих-код. Предложена структура системы формирования штрих-кода и приводится описание основных блоков.

Тестирование выполнено на базах «Face94», «Face Sketch FERET Database» и базе композитных лиц различных возрастов. В рамках тестов показано, что сформированный штрих-код не изменяется при изменении локальных размеров лиц, наклона в плоскости XY, изменения ракурса и зеркального поворота вокруг вертикальной оси, а также при изменениях речевой мимики лиц и возрастных изменениях лиц.

Таким образом, предложенные методы предлагают новое решение для практики использования в реальных условиях – динамики изменения параметров изображений лиц. При этом оба подхода не требуют больших вычислительных затрат, а также использования специализированных пакетов программ по обработке изображений, что создает условия для формирования линейных штрих-кодов в рамках систем реального времени. Сформированный штрих-код содержит информацию о лице конкретного человека и может быть использован для индексирования, идентификации, распознавания и поиска людей.

Ключевые слова: изображения лиц, штрих-код, системы реального времени.

Литература

1. Method for verifying human identity during electronic sale transactions – Patent US 5878155A (1999). [Электронный ресурс]: <http://www.google.com/patents/US5878155> (дата обращения 15.03.2014).
2. Barcode/Tattoos. [Электронный ресурс]: <http://www.barcodeart.com/store/wearable/tattoos/> (дата обращения 15.03.2014).
3. Dakin S.C., Watt R.J. Biological «bar codes» in human faces // Journal of Vision. 2009. Vol. 9, No. 4. P. 1-10.
4. Goffaux V., Dakin S.C. Horizontal information drives the behavioral signatures of face processing // Frontiers in Psychology. 2010. Vol. 1, No. 43. P. 1-14.

5. Facial Barcodes Help Us Identify People. [Электронный ресурс]: <http://www.barcodesinc.com/news/?p=92> (дата обращения 15.03.2014).
6. Querinini M., Italiano G.F. Facial recognition with 2D color barcodes // International Journal of Computer Science and Application. 2013. Vol. 10, No. 1. P. 78-97.
7. Методы обработки и распознавания изображений лиц в задачах биометрии / Кухарев Г.А. и [др.]. Монография. СПб: Политехника, 2013. 388 с.
8. Face94 data base. [Электронный ресурс]: <http://cswww.essex.ac.uk/mv/allfaces/faces94.html> (дата обращения 15.03.2014).
9. Forczmanski P., Kukharev G. Comparative analysis of simple facial features // Journal Real-Time Image Processing. 2007. No. 1. P. 239-255.
10. Barcode generation algorithm for EAN-8 и EAN-13. [Электронный ресурс]: http://www.cherry-notes.spb.ru/barcode_ean8.htm (дата обращения 15.03.2014).
11. Кухарев Г. А., Матвеев Ю. Н., Щеголева Н. Л. Экспресс-метод формирования штрих-кода по изображениям лиц // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 2(90). С. 99-106.
12. CUHK Face Sketch FERET Database. [Электронный ресурс]: <http://mmlab.ie.cuhk.edu.hk/cufsf> (дата обращения 15.03.2014).
13. Kukharev G.A., Buda K., Shchegoleva N.L. Methods of face photo-sketch comparison // Pattern Recognition and Image Analysis. 2014. Vol. 24, No. 1. P. 82-93.
14. Burt D. M., Perrett D. I. Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour // Proceedings of the Royal Society London. 1995. Vol. 259, No. 1355. P. 137-143.
15. Кухарев Г.А., Матвеев Ю.Н., Щеголева Н.Л. Формирование штрих-кода по изображениям лиц на основе градиентов яркости // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 3(91). С. 88-95.

МЕРА СИМВОЛЬНОГО РАЗНООБРАЗИЯ: ПОДХОД КОМБИНАТОРИКИ СЛОВ К ОПРЕДЕЛЕНИЮ ОБОБЩЕННЫХ ХАРАКТЕРИСТИК ВРЕМЕННЫХ РЯДОВ¹

Ю.Г. СМЕТАНИН

доктор физико-математических наук, главный научный сотрудник,
Вычислительный центр им. А.А. Дородницына, Российская академия наук
Адрес: 119333, г. Москва, ул. Вавилова, д. 40
E-mail: smetanin.iury2011@yandex.ru

М.В. УЛЬЯНОВ

доктор технических наук, профессор кафедры прикладной математики
и моделирования систем, Институт коммуникаций и медиабизнеса,
Московский государственный университет печати им. Ивана Федорова;
профессор кафедры управления разработкой программного обеспечения,
департамент программной инженерии, факультет компьютерных наук,
Национальный исследовательский университет «Высшая школа экономики»
Адрес: 101000, г. Москва, ул. Мясницкая, д. 20
E-mail: muljanov@mail.ru

В настоящее время рассматриваются разнообразные подходы к исследованию временных рядов в аспекте их прогнозирования. По мнению авторов, интерес представляет подход кластерного анализа, в котором объектом исследования является множество временных рядов, порожденных различными источниками. Пространство кластеризации строится на основе обобщенных универсальных характеристик временных рядов, каждая из которых является координатой этого пространства. Одному временному ряду в таком пространстве соответствует точка в координатах универсальных характеристик. Применение методов кластерного анализа позволяет выделить временные ряды, близкие по метрике пространства, а для полученных кластеров возможно решение задачи о выборе рационального метода прогнозирования. Построение специального метрического пространства для анализа временных рядов является объектом исследования данной статьи. Предметом исследования являются координаты этого пространства – обобщенные характеристики временных рядов. Авторами в ряде предыдущих работ уже были введены две координаты такого пространства – сложность временного ряда по Колмогорову и гармоническая сложность временного ряда. Настоящая статья посвящена разработке новой обобщенной характеристики временного ряда с использованием аппарата комбинаторики слов – мере символического разнообразия. Применение подхода символического кодирования позволяет получить представление временных рядов в пространстве слов некоторого выбранного алфавита. Исследование полученного представления методами комбинаторики слов позволяет получить оценку энтропии сдвигов как функцию длины скользящего окна. На основе исследования особенностей первой конечной разности этой функции предлагается мера символического разнообразия временного ряда. Предложенная обобщенная характеристика может быть использована для последующего выявления характерных особенностей временных рядов, в частности, как одна из осей пространства кластеризации.

Ключевые слова: временные ряды, обобщенные характеристики, символические описания, оценка энтропии слов, мера символического разнообразия.

¹ Работа выполнена при поддержке гранта РФФИ 13-07-00516.

1. Введение

Одним из перспективных подходов к исследованию временных рядов в аспекте их прогнозирования является подход кластерного анализа [1, 2], приводящий к выявлению групп временных рядов, обладающих близкими свойствами. Кластерный анализ базируется на построении специального метрического пространства, оси координат которого соответствуют обобщенным универсальным характеристикам временных рядов. В таком пространстве конкретному временному ряду соответствует кортеж – точка данного пространства. Последующий кластерный анализ приводит к выделению кластеров, элементами которых являются временные ряды, близкие по выбранной метрике. Для полученных кластеров может быть решена задача о назначениях методов прогнозирования: такой подход может способствовать повышению точности прогнозов за счет выбора метода, учитывающего специфику временных рядов, принадлежащих данному кластеру.

С точки зрения авторов наибольший интерес представляет именно построение специального метрического пространства, координатами которого являются обобщенные универсальные характеристики временных рядов. В аспекте построения пространства кластеризации авторами в предыдущих работах и докладах [2, 3, 4] были введены некоторые координаты такого пространства – сложность временного ряда по Колмогорову и гармоническая сложность временного ряда. Настоящая статья посвящена новой обобщенной характеристике временного ряда – мере символьного разнообразия.

Содержательно предлагаемая обобщенная характеристика отражает границу наблюдаемого разнообразия подслов над фиксированным алфавитом в слове, представляющим собой символьный код рассматриваемого временного ряда. Используемый при этом аппарат основан на методах символьческой динамики и символьском кодировании значений временного ряда.

2. Символьческая динамика и комбинаторика слов

В данной статье для исследования временных рядов предложено использовать подход, основанный на методах комбинаторики слов и анализе энтропии, в анализе информации, представленной в виде слов над конечным алфавитом.

Комбинаторика слов – термин, введенный в широкое обращение группой математиков, публикующих результаты своих исследований под псевдонимом M.Lothaire [5]. Этим термином объединяются направления исследований, связанные общими подходами, которые ранее оставались разбросанными по различным областям математики и информатики, от теории чисел и теории динамических систем до анализа естественных языков или биологических последовательностей. В комбинаторике слов исследуется внутренняя структура слов. Плодотворность этого подхода проявилась в эффективном применении методов из одной предметной области в других областях. Типичными примерами являются применения в теории формальных языков и автоматов [6], теории групп [7], теории хаоса [8], фрактальном анализе [9], символической динамике [10] и анализе временных рядов, биоинформатике [11, 12], лингвистике и некоторых других областях.

Анализ временных рядов или очень длинных последовательностей (например, ДНК [11, 13]) также тесно связан с задачами символической динамики [14].

Символической динамикой называют раздел теории динамических систем, в котором для описания последовательностей измерений состояния системы используются символы из некоторого алфавита. Последовательности соответствуют траекториям изучаемой системы. Наиболее эффективными методами символической динамики оказываются в исследовании детерминированных систем, в которых из-за ограниченных возможностей измерения возникает сходство со случайными процессами. Описание динамики получается в терминах топологических аналогов марковских процессов – матриц возможных переходов между состояниями системы. Для построения такого описания необходимо выбрать алфавит, наиболее подходящий для представления разбиения пространства состояний системы на области, соответствующие измеряемым значениям. Сложность систем естественно оценивать числом различных конечных слов, входящих в допустимые бесконечные последовательности. Для количественной оценки целесообразно применять метрическую энтропию по Колмогорову или топологическую энтропию. Для оценки сложности индивидуальных траекторий системы можно строить оценки на основе сложности по Колмогорову. С помощью теста Колмогорова-Мартина-Лефа можно принимать решения о том, является ли индивидуальная траектория случайной.

В качестве примера можно привести задачу распознавания вторичной структуры белков [15], которая заключается в следующем. Белок можно представлять как одномерную последовательность аминокислот или как одномерную последовательность характерных локальных конфигураций. В настоящее время общепринятым является допущение, что первичная структура однозначно определяет вторичную. При этом задача определения вторичной структуры (структуры локальных конфигураций) формулируется как задача преобразования слов в алфавите имен аминокислот в слова над алфавитом локальных конфигураций с помощью кодов скользящего блока.

В данной работе рассмотрена задача анализа бесконечных последовательностей (временных рядов) на основе их достаточно длинных конечных отрезков. Предложен метод оценки энтропийных характеристик, полезный для выделения признаков в целях последующей классификации временных рядов.

3. Постановка задачи

Объектом исследования является временной ряд (произвольной природы)

$$V = \{ (f_i, t_i), f_i \in R^1, i = 1, \dots, n \}, \quad (1)$$

где f_i — значение характеристики наблюдаемого процесса в момент t_i , n — число наблюдений (отсчетов).

Предметом исследования является построение обобщенной характеристики ряда, отражающей разнообразие наблюдаемых значений. В этой постановке мы формулируем следующие задачи относительно ряда V :

- ◆ задачу символьного кодирования значений временного ряда;
- ◆ задачу построения функции оценки энтропии сдвигов;
- ◆ задачу определения меры символьного разнообразия временного ряда.

Дальнейшее изложение посвящено описанию решений сформулированных задач, предложенных авторами.

4. Символьное кодирование временного ряда по значениям

Требование универсальности пространства кластеризации налагает, очевидно, и требования к обобщенным характеристикам временных рядов,

конкретные значения которых интерпретируются как координаты точки, представляющей данный временной ряд в осях этого пространства. Проблема универсализации связана с тем, что различные временные ряды имеют различную точность измерений (число значащих цифр в значении f_i) и различный масштаб по значениям. Решение проблемы авторы видят в едином масштабировании значений наблюдаемой функции процесса и построении на этой основе строки символов, отражающей динамику числовых значений исследуемого ряда.

В целях такого масштабирования мы вводим разбиение $x_i, i = 1, \dots, m$ диапазона размаха варьирования значений f_i от $x_1 = \min_{i=1, n} f_i$ до $x_m = \max_{i=1, n} f_i$. Поскольку значения временного ряда могут попасть в точки разбиения, мы рассматриваем множества $[x_i, x_{i+1}) = \{x | x_i \leq x < x_{i+1}\}$, которые далее будем называть полусегментами. Определение числа и границ полусегментов доставляется бикритериальным методом построения гистограмм [16]. Подробное изложение этого метода по отношению к символьному кодированию приведено в [3]. Число полученных этим методом полусегментов определяет мощность алфавита кодирования. Заметим, что последний элемент разбиения является, очевидно, сегментом. Выбор символов алфавита, по сути, не принципиален, но мы в дальнейшем будем использовать прописные символы латинского алфавита. Далее каждый полусегмент кодируется соответствующим символом алфавита, и мы получаем представление временного ряда в виде строки символов, например (для алфавита $\Sigma = \{A, B, C, D\}$): «*BABCDEEEDDCCCBBAABB....*» При этом числовое значение ряда кодируется символом полусегмента, в котором оно находится. Для временного ряда, содержащего n наблюдений, мы получаем его представление в виде строки из n символов над алфавитом Σ . Полученная строка символьного кода значений может быть основой и для решения задачи символьного кодирования временного ряда по тенденциям, более подробно см. в [3].

Отметим еще одно преимущество предлагаемого подхода символьного кодирования. За редким исключением значения в отсчетах временных рядов не являются точными. Одними из таких исключений являются, например, ряды курсов валют. Для значений, имеющих погрешность измерений, в математической статистике принято строить доверительные интервалы. Используемый авторами бикритериальный метод построения гистограмм как раз и определяет ширину полусегмента гисто-

граммы, а, следовательно, и «ширину» значений для кодирующего этот полусегмент символа, на основе доверительной вероятности для среднего значения [16]. Таким образом, подход символического кодирования более достоверно отражает исследуемый процесс с точки зрения математической статистики.

5. Функция оценки энтропии сдвигов

С целью построения предлагаемой меры разнообразия полученная символическим кодированием временного ряда строка подвергается обработке, первым этапом которой является оценка энтропии слов. Эта оценка используется как в символической динамике [10], так и в биоинформатике для оценки сложности нуклеотидных геномных последовательностей [17].

Оценка энтропии слов строится следующим образом [18]. Фиксируется длина подслова m и алфавит Σ , и далее рассматриваются все подслова длины m над алфавитом Σ . Множество различных подслов есть Σ^m , мощность этого множества $M = |\Sigma^m|$ есть общее число подслов, очевидно, что $M = k^m$. Для рассматриваемого значения m вводится произвольная нумерация подслов $i = \overline{1, M}$ и вводятся счетчики числа подслов c_i , которые изначально обнуляются. Изначально позиционированное в начале анализируемого слова длины n , окно ширины m сдвигается каждый раз на один символ. Таким образом, мы имеем $n - m + 1$ позиций окна, и для каждого его положения распознается подслово, полученное в окне. Если мы наблюдаем в текущей позиции окна ширины m подслово, которое имеет номер i в принятой нумерации, то значение счетчика c_i увеличивается на единицу. По полученным значениям c_i , $i = \overline{1, M}$ и рассчитывается оценка энтропии слов C_m по следующей формуле:

$$C_m = -\sum_{i=1}^M \left(\frac{c_i}{n-m+1} \right) \log_M \left(\frac{c_i}{n-m+1} \right). \quad (2)$$

Заметим, что применение основания M у логарифма приводит автоматически к нормировке значений $C(m)$ – значение 0 означает, что все подслова длины m одинаковы и состоит из одного и того же подслова (фундаментальное отсутствие разнообразия). Просто показать, что значение $C(m) = 1$ соответствует равночастотности всех возможных подслов в исходном слове. На основании оценки энтропии слов мы строим функцию $C(m) = C_m$, аргументом которой является длина подслова m , с

областью определения: $1 \leq m \leq n$. Функция $C(m)$ вычисляется при фиксированном m сдвигом окна ширины m по исходному слову по формуле (2) и увеличением на единицу ширины окна при изменении m от 1 до n . В соответствии с принятой в символической динамике терминологией [10] будем называть $C(m)$ функцией оценки энтропии сдвигов.

6. Монотонность функции оценки энтропии сдвигов

Для дальнейшего построения предлагаемой меры временного ряда рассмотрим поведение функции оценки энтропии сдвигов $C(m)$ как функции длины подслова m . Очевидно, что при $m = n$ мы наблюдаем всего одно подслово, совпадающее с исходным словом, и, в соответствии с (2) $C(n) = 0$. При $m = 1$ максимум $C(m)$ будет равен единице в случае, если частота символов алфавита в символическом представлении временного ряда одинакова.

Покажем, что при малых значениях m (длины подслова) функция $C(m)$ является монотонно убывающей. Далее мы сформулируем условие на «малость» m . Пусть алфавит $\Sigma = \{s_1, s_2, \dots, s_k\}$, где k – мощность алфавита, c_i – кратность вхождения подслова $w_i = (\alpha_1 \alpha_2 \dots \alpha_m)$, где $\alpha_i \in \Sigma$, в рассматриваемое слово. При длине подслова m функция оценки энтропии сдвигов в соответствии с (2) может быть записана в виде

$$H_i(m) = \frac{c_i}{n-m+1} \log_M \left(\frac{n-m+1}{c_i} \right), \quad C(m) = \sum_{i=1}^M H_i(m). \quad (3)$$

При увеличении длины подслова на 1, $m \rightarrow m+1$, из каждого подслова длины $m - w_i = (\alpha_1 \alpha_2 \dots \alpha_m)$ образуются подслова длины $m+1 - w_{i0} = (\alpha_1 \alpha_2 \dots \alpha_m s_1)$, $w_{i1} = (\alpha_1 \alpha_2 \dots \alpha_m s_2)$, ..., $w_{i,k-1} = (\alpha_1 \alpha_2 \dots \alpha_m s_k)$, при этом, возможно, некоторые из них будут иметь нулевую кратность, поэтому в функцию $C(m)$ при длине подслова $m+1$ вместо i -го слагаемого в (3)

$$H_i(m) = \frac{c_i}{n-m+1} \log_M \left(\frac{n-m+1}{c_i} \right)$$

войдет слагаемое

$$H_i(m+1) = \sum_{j=0}^{k-1} \frac{c_{ij}}{n-m} \log_{(Mk)} \left(\frac{n-m}{c_{ij}} \right),$$

где через c_{ij} мы обозначили кратность вхождения подслова w_{ij} .

Это слагаемое достигает максимума, когда кратности всех образовавшихся подслов w_{ij} одинаковы, $c_{ij} = c_i / k$, $j = 0, 1, \dots, (k-1)$, то есть его значение не превышает величины

$$H_i(m+1)_{max} = \frac{c_i}{n-m} \log_{(Mk)} \left(\frac{k(n-m)}{c_i} \right).$$

При $M \cdot k = k^{m+1} < \frac{n-m}{c_i}$ выполняется неравенство

$$\log_{(Mk)} k < \log_{(Mk)} \left(\frac{n-m}{c_i} \right)^{1/(m+1)},$$

следовательно,

$$\begin{aligned} H_i(m+1)_{max} &< \frac{c_i}{n-m} \log_{(Mk)} \left(\frac{n-m}{c_i} \right)^{1+1/(m+1)} = \\ &= \frac{c_i}{n-m} \frac{(1 + \frac{1}{m+1}) \log_M \left(\frac{n-m}{c_i} \right)}{\log_M(Mk)} = \\ &= \frac{\left(1 + \frac{1}{m+1}\right) c_i}{1 + \frac{1}{m}} \log_M \left(\frac{n-m}{c_i} \right) = \\ &= \frac{m(m+2)}{(m+1)^2} \cdot \frac{n-m+1}{n-m} \left(\frac{c_i}{n-m+1} \log_M \left(\frac{n-m+1}{c_i} \right) \right). \end{aligned}$$

Можно показать, что при $n > m^2 + 3m$

$$\frac{m(m+2)}{(m+1)^2} \cdot \frac{n-m+1}{n-m} < 1,$$

следовательно,

$$H_i(m+1)_{max} < H_i(m).$$

Таким образом, при достаточно малой длине подслова m по отношению к длине слова n (т.е. при $m^2 + 3m < n$) значение функции оценки энтропии сдвигов уменьшается при переходе от m к $m+1$, следовательно, $C(m)$ как функция целочисленного аргумента является функцией, спадающей от начального значения $C(1)$, которое при больших n и близких частотах символов алфавита, как правило, близка к единице, до $C(n) = 0$.

7. Мера символьного разнообразия

Интерес представляет изучение характера убывания значений $C(m)$ с ростом аргумента. Поскольку функция $C(m)$ — «убывающая по совокупности», рассмотрим инверсную конечную разность функции $C(m)$:

$$\Delta C(m) = C(m) - C(m-1), m = \overline{2, n}. \quad (4)$$

По определению $C(m)$ значения $\Delta C(m)$ ограничены, и $0 \leq \Delta C(m) \leq 1$, но поведение $\Delta C(m)$ может быть достаточно сложным. Предположим, что начальное

значение $C(1) \approx 1$, т.е. символы алфавита кодирования временного ряда имеют слабо отличающуюся частотную встречаемость. Тогда близкие к нулю начальные значения $\Delta C(m)$, характеризуют нашу символьную последовательность как последовательность, обладающую достаточно богатым разнообразием подслов. Однако функция $C(m)$ не может долго «держаться единицу». Определим пороговое значение \hat{m} , при котором теоретически функция оценки энтропии сдвигов еще может быть равной единице. Поскольку в сдвигающемся окне ширины m при мощности алфавита кодирования $k = |\Sigma|$ может наблюдаться максимально $M = |\Sigma^m| = k^m$ различных подслов, а всего в слове длины n мы имеем $n - m + 1$ позиций окна, то максимально возможная длина подслова при котором еще можно наблюдать полное разнообразие подслов, определяется из уравнения $M = k^{\hat{m}} = n - \hat{m} + 1$, что с учетом целочисленности \hat{m} приводит к уравнению $\hat{m} = \lfloor \log_k(n - \hat{m} + 1) \rfloor$. В предположении, что $m \ll n$, значение $\hat{m} \approx \lfloor \log_k n \rfloor$. В окне ширины $\hat{m} + 1$ максимально наблюдаемое разнообразие слов в k раз меньше полного разнообразия в алфавите мощности k . Поэтому мы ожидаем резкого падения значения функции $C(m)$ при $m = \hat{m} + 1$, и, следовательно, резкого скачка $\Delta C(m)$ даже для псевдослучайной последовательности символов исходного слова, обладающего конечной длиной. Таким образом, наличие ярко выраженного максимума у функции $\Delta C(m)$ при $m < \hat{m}$ означает, что начиная с данного значения m разнообразие подслов резко уменьшилось, и исходное слово обладает определенной регулярностью или периодичностью.

На основе этих рассуждений авторы и вводят меру символьного разнообразия временного ряда $\mu_s(V)$ в виде отношения значения аргумента функции $\Delta C(m)$, доставляющего ее максимум к максимально возможной ширине окна, сохраняющей полное разнообразие подслов. В этих целях определим максимум функции $\Delta C(m)$, и обозначим через m^* аргумент этого максимума

$$m^* = \arg \max_{1 \leq m \leq n} \Delta C(m).$$

Тем самым значение m^* определяет положение скачка конечной разности. Рассмотрим отношение $\mu_s(V) = m^*/\hat{m}$. Оно нормировано в $[0, 1]$, и малые значения свидетельствуют о раннем наступлении потери разнообразия и большей «простоте» исследуемого слова. Учитывая принцип построения меры, мы окончательно получаем меру символьного разнообразия временного ряда $\mu_s(V)$ в виде

$$\mu_s(V) = \frac{m^*}{\hat{m}} = \frac{\arg \max_{1 \leq m \leq n} \Delta C(m)}{\lfloor \log_k n \rfloor} \quad (5)$$

Здесь авторы понимают меру символьного разнообразия как функционал, т.е. как индивидуальную меру на одноэлементных подмножествах. Заметим, что в соответствии с (4) для периодических слов с малым периодом с ростом длины слова (числа отсчетов исходного временного ряда) значение $\mu_s(V)$ будет уменьшаться, что соответствует логике введенной меры — для длинного периодического слова с малым периодом символьное разнообразие очевидно мало.

8. Модельный пример

Приведем модельный пример вычисления предложенной меры символьного разнообразия для бесконечной периодической строки $(ABBAAB)_n$ в алфавите $\Sigma = \{A, B\}$. Вычисленные по формулам (3) и (4) значения функций $C(m)$ и $\Delta C(m)$ приведены в табл. 1.

Таблица 1.

Значения функции оценки энтропии сдвигов и ее конечной разности для модельной строки

m	$C(m)$	$\Delta C(m)$
1	1,000	
2	0,959	0,041
3	0,862	0,097
4	0,646	0,215
5	0,517	0,129
6	0,431	0,086
7	0,369	0,062
8	0,323	0,046

Соответствующие графики приведены на рис. 1 и 2. Обе функции являются функциями целочисленного аргумента, но мы показываем их как кусочно-линейные для наглядности отображения тенденций.

Заметим, что поскольку модельное слово имеет период 6, то, начиная с окна ширины 3, мы наблюдаем всего 6 различных подслов, а поскольку $\Sigma = \{A, B\}$, то полное разнообразие подслов увеличивается вдвое при увеличении ширины окна на единицу. Таким образом, при $m = 3$ мы наблюдаем 6 подслов из 8 возможных, а при $m = 4$ — всего 6

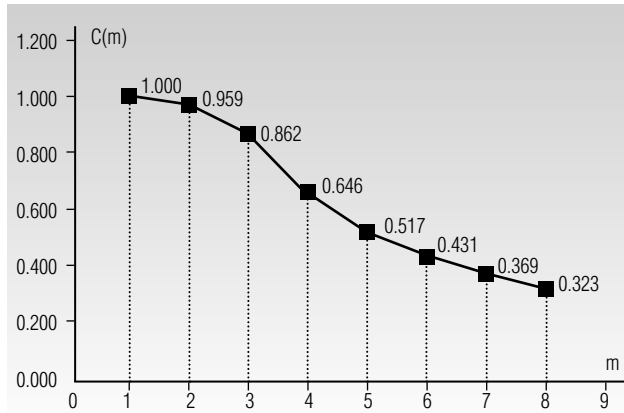


Рис. 1. График функции оценки энтропии сдвигов $C(m)$ для модельной строки

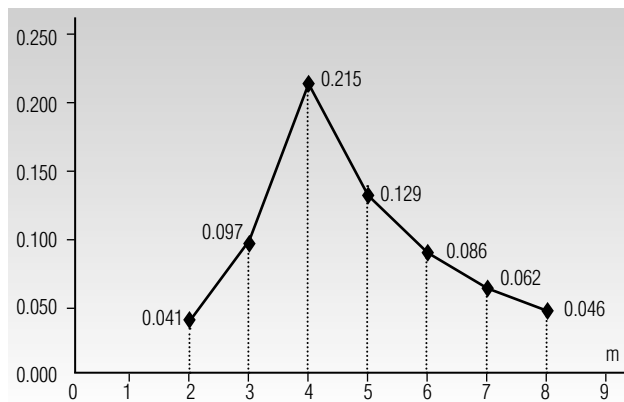


Рис. 2. График функции конечной разности $\Delta C(m)$ для модельной строки

подслов из 16 возможных, и максимум функции $\Delta C(m)$ фиксирует потерю разнообразия в окне ширины 4. Значение предложенной меры зависит от длины строки в соответствии с (5). Значения в табл. 1 рассчитаны для бесконечного слова, однако при больших длинах, например, при $n = 1033$ эти значения изменятся незначительно. Для такой строки $m^* = 4$, а $\hat{m} = 10$ и значение меры $\mu_s(V) = 0,40$.

9. Заключение

Представление временного ряда, полученное на основе символьческого кодирования полусегментов с использованием бикритериального метода построения гистограмм, является основой для построения функции оценки энтропии сдвигов $C(m)$, аргументом которой является ширина окна. Построение конечной разности $\Delta C(m)$ позволяет изучить особенности разнообразия подслов в исследуемом слове, а максимум этой разности сви-

детельствует о падении разнообразия как в смысле отклонения от равномерности частот подслов, так и в смысле собственно наблюдаемого разнообразия подслов. На основе исследования поведения функции $\Delta C(m)$ авторы вводят меру символического разнообразия временного ряда $\mu_s(V)$ как отношение значения аргумента функции $\Delta C(m)$, доставляюще-

го ее максимум, к максимально возможной ширине окна, сохраняющей полное разнообразие подслов. По принципам построения, малые значения $\mu_s(V)$ соответствуют «простым» временным рядам с вероятной хорошей возможностью их прогнозирования, а большие, близкие к единице, значения – рядам с выраженной хаотичностью. ■

Литература

1. Грабуст П. Способы оценок сходства временных рядов // Научные труды Международной конференции «Теория вероятностей, случайные процессы, математическая статистика и приложения», Минск, БГУ, 15-19 сентября 2008 г. Минск: Белорусский государственный университет, 2008. С. 23–24.
2. Ульянов М.В., Сметанин Ю.Г. Об одном подходе к построению кластерного пространства временных рядов: колмогоровская и гармоническая сложность // Proceedings of the International scientific-practical conference «Information Control Systems and Technologies» (ICST 2013). Odessa, 2013. С. 30-36.
3. Ульянов М.В., Сметанин Ю.Г. Подход к определению характеристик колмогоровской сложности временных рядов на основе символических описаний // Бизнес-информатика. 2013. №2 (24). С. 49-54.
4. Сметанин Ю.Г., Ульянов М.В. Пространство обобщенных характеристик для классификации временных рядов: характеристика гармонической сложности // Проблемы автоматизации и управления в технических системах: Сборник статей Международной научно-технической конференции / Под ред. д.т.н., проф. М.А.Щербакова. Пенза: Изд. ПГУ, 2013. С. 125-128.
5. Lothaire M. Algebraic combinatorics on words. Cambridge, UK: Cambridge University Press, 2002. 455 pp.
6. Хопкрофт Д., Мотвани Р., Ульман Дж. Введение в теорию автоматов, языков и вычислений. М.: Издательский дом «Вильямс», 2008. 528 с.
7. Morse M., Hedlund G. Unending chess, symbolic dynamics and a problem in semigroups // Duke Mathematical Journal. 1944. No.11. P. 1-7.
8. Симиу Э. Хаотические переходы в детерминированных и стохастических системах. М.: Физматлит, 2007. 208 с.
9. Афраймович В., Угальде Э., Уриас Х. Фрактальные размерности для времен возвращения Пуанкаре. Москва, Ижевск: Институт компьютерных исследований, R&C Dynamics, 2011. 292 с.
10. Lind D., Marcus B. An introduction to symbolic dynamics and coding. Cambridge, UK: Cambridge University Press, 1995. 495 pp.
11. Математические методы для анализа последовательностей ДНК. М.: Мир, 1999. 349 с.
12. Гамов Г. Комбинаторные принципы в генетике // Прикладная комбинаторная математика / Под ред. Э.Беккенбаха. М.: Мир. 1968. С. 288-308.
13. Гасфилд Д. Строки, деревья и последовательности в алгоритмах: Информатика и вычислительная биология / Пер с англ. СПб.: Невский диалект; БХВ-Петербург, 2003. 654 с.
14. Боуэн Р. Методы символической динамики. М.: Мир, 1979. 245 с.
15. Рудаков К.В., Торшин И.Ю. Об отборе информативных значений признаков на базе критериев разрешимости в задаче распознавания вторичной структуры белка // ДАН. 2011. Т. 441, № 1. С. 1–5.
16. Петрушин В.Н., Ульянов М.В. Бикритериальный метод построения гистограмм // Информационные технологии и вычислительные системы. 2012. № 4. С. 22–31.
17. Орлов Ю.Л. Анализ регуляторных геномных последовательностей с помощью компьютерных методов оценок сложности генетических текстов // Дисс. на соискание уч. ст. канд. биол. наук. Новосибирск, 2004. 148 с.
18. Орлов Ю.Л. Компьютерная реализация оценок сложности текстов // Материалы Российской конференции «Дискретный анализ и исследование операций» (ДАОР), Новосибирск, Институт математики СО РАН, 28 июня – 2 июля 2004 г. Новосибирск: Изд-во Института математики СО РАН, 2004. С. 225.

MEASURE OF SYMBOLICAL DIVERSITY: COMBINATORICS ON WORDS AS AN APPROACH TO IDENTIFY GENERALIZED CHARACTERISTICS OF TIME SERIES

Yuri SMETANIN

Chief Researcher, Dorodnitsyn Computing Centre, Russian Academy of Sciences

Address: 40, Vavilova street, Moscow, 119333, Russian Federation

E-mail: smetanin.iury2011@yandex.ru

Mikhail ULYANOV

Professor, Department of Applied Mathematics and Systems Modeling,

Institute of Communications and Media Business,

Moscow State University of Printing Arts;

Professor, Software Management Department, School of Software Engineering,

Faculty of Computer Science, National Research University Higher School of Economics

Address: 20, Myasnitskaya street, Moscow, 101000, Russian Federation

E-mail: muljanov@mail.ru

Currently various approaches to time series analysis are being investigated in terms of their forecasting. In the authors' opinion, an approach to cluster analysis, which research object constitutes sets of time series generated by various sources, is of particular interest. The clusterization space is constructed by using generalized universal characteristics of time series each of which is a coordinate in this space. In such space for each time series there is a corresponding point in the coordinates of universal characteristics. Application of cluster analysis methods enables to identify time series that are space metric, and for the obtained clusters it is possible to solve the problem of choosing an efficient method of forecasting.

Construction of a special metric space to analyze time series constitutes the research object of this article. The research subject is this space coordinates – generalized characteristics of time series. In their previous articles, the authors have already defined two coordinates of such space: the Kolmogorov complexity of the time series and its harmonic complexity. This paper focuses on elaboration of a new generalized characteristic of time series by using combinatorics on words technique: a measure of symbolic diversity. The application of the symbolic coding approach enables to represent time series in a space of words in a selected alphabet. Investigation of the representation generated by combinatorics on words methods enables to estimate the entropy of shifts as a function of the length of the sliding window. A measure of symbolic diversity of time series has been proposed based on investigation of specifics of the first finite difference of this function. The proposed generalized characteristic may be applied for further identification of specific features of time series; in particular as one of the axes in the clusterization space.

Key words: time series, generalized characteristics, symbolic descriptions, words entropy assessment, measure of symbolic diversity.

References

1. Grabust P. (2008) Sposoby otsenok skhodstva vremennykh riadov [Methods of time series estimation], Nauchnye trudy Mezhdunarodnoi konferentsii «Teoria veroiatnostei, sluchainye protsessy, matematicheskaya statistika i prilozhenia» [Proceedings of the *Probability Theory, Random Processes, Mathematical Statistics and Applications: International conference* (Minsk, Belarus, September 15-19, 2008), Minsk, Belarus State University, pp. 23–24. (in Russian)
2. Ulyanov M., Smetanin Y. (2013) Ob odnom podkhode k postroeniui klaster'nogo prostranstva vremennykh riadov: kolmogorovskaia i garmonicheskaya slozhnost' [On an approach to constructing a cluster space of time series: Kolmogorov and harmonic complexity] Proceedings

- of the *Information Control Systems and Technologies ICST 2013: International scientific-practical conference*, Odessa, Ukraine, pp. 30–36. (in Russian)
3. Ulyanov M., Smetanin Y. Podkhod k opredeleniiu kharakteristik kolmogorovskoi slozhnosti vremennykh riadov na osnove simvol'nykh opisani [Approach to determining of characteristics of Kolmogorov complexity of time series: An approach based on symbolical descriptions]. *Business Informatics*, no. 2 (24), pp. 49–54. (in Russian)
 4. Smetanin Y., Ulyanov M. Prostranstvo obobshchennykh kharakteristik dlia klassifikatsii vremennykh riadov: kharakteristika garmonicheskoi slozhnosti [A Space of generalized characteristics for the classification of time Series: A characteristic of harmonic complexity. Proceedings of the *International Scientific and Technical Conference (Penza, Russia, 2013)* (Ed. M.Shcherbakov), Penza: Penza State University, pp. 125–128. (in Russian)
 5. Lothaire M. (2002) *Algebraic combinatorics on words*, Cambridge, UK: Cambridge University Press.
 6. Hopcroft J.E., Motwani R., Ullman J.D. (2000) *Introduction to Automata theory, languages, and computation*, Pearson Publication.
 7. Morse M., Hedlund G. (1944) Unending chess, symbolic dynamics and a problem in semigroups. *Duke Mathematical Journal*, no. 11, pp. 1–7.
 8. Simiu E. (2009) *Chaotic transitions in deterministic and stochastic dynamical systems: Applications of Melnikov processes in engineering, physics, and neuroscience*, Princeton University Press.
 9. Afraimovich V., Ugalde E., Urias J. (2006) *Fractal dimensions for Poincare recurrences*, Elsevier.
 10. Lind D., Marcus B. (1995) *An introduction to symbolic dynamics and coding*, Cambridge University Press.
 11. Mir (1999) *Matematicheskie metody dlia analiza DNK* [Mathematical Methods of DNA Analysis]. Moscow: Mir. (in Russian)
 12. Gamov G. (1968) Kombinatornye metody v genetike [Combinatorial Methods in Genetics]. *Prikladnaia kombinatornaia matematika* [Applied Combinatorial Mathematics] (Ed. E.Beckenbach), Moscow: Mir, pp. 288–308. (in Russian)
 13. Gusfield D. (1997) *Algorithms on strings, trees, and sequences*, Cambridge, UK: Cambridge University Press.
 14. Bowen R. (1979) *Metody simvolicheskoi dinamiki* [Methods of symbolical dynamics], Moscow: Mir. (in Russian)
 15. Rudakov K., Torshin I. (2011) Ob otbore informativnykh znachenii priznakov na baze kriteriev razreshimosti v zadache raspoznavaniia vtorichnoi struktury belka [On the selection of informative features using the solvability criteria in the problem of recognition of secondary structure of the protein]. *Doklady Mathematics (Doklady Akademii Nauk)*, vol. 441, no. 1, pp. 1–5. (in Russian)
 16. Petrushin V., Ulyanov M. (2012) Bikriterial'nyi metod postroeniia gistogramm [Bicriterial method of histogram construction]. *Information Technologies and Computer Systems*, no. 4, pp. 22–31. (in Russian)
 17. Orlov Y. (2004) *Analiz regulatornykh genomnykh posledovatel'nostei s pomoshchiu komp'yuternykh metodov otsenok slozhnosti geneticheskikh tekstov* [Analysis of regulatory genome sequences using computer methods of genetic texts complexity estimation] (PhD Thesis), Novosibirsk: Institute of Mathematics. (in Russian)
 18. Orlov Y. Komp'yuternaia realizatsiia otsenok slozhnosti tekstov [Computer implementation of text complexity estimations]. Proceedings of the *Discrete Analysis and Operation Research: Russian Conference (Novosibirsk, Russia, June 28 – July 2, 2004)*, Novosibirsk: Institute of Mathematics, p. 225. (in Russian)

НЕЙРОСЕТЕВАЯ СИСТЕМА ОЦЕНКИ ВЕРОЯТНОСТИ БАНКРОТСТВА БАНКОВ

Л.Н. ЯСНИЦКИЙ

доктор технических наук, профессор кафедры информационных технологий в бизнесе, факультет бизнес-информатики, Национальный исследовательский университет «Высшая школа экономики» (Пермский филиал); Председатель Пермского отделения Научного Совета РАН по методологии искусственного интеллекта
Адрес: 614046, г. Пермь, ул. Студенческая, д. 38
E-mail: yasn@psu.ru

Д.В. ИВАНОВ

аспирант кафедры информационных систем и математических методов в экономике, экономический факультет, Пермский государственный национальный исследовательский университет
Адрес: 614990, г. Пермь, ул. Букирева, д. 15
E-mail: idv_1988@mail.ru

Е.В. ЛИПАТОВА

студент факультета экономики, Национальный исследовательский университет «Высшая школа экономики» (Пермский филиал)
Адрес: 614046, г. Пермь, ул. Студенческая, д. 38
E-mail: lipatova_katya@mail.ru

Предметом исследования является банковская система России. Цель работы – создание математической модели, предназначенной для оценки вероятности банкротств банков по причине отзыва лицензии. Инструмент создания модели – аппарат нейронных сетей, обучаемых на материалах финансовой отчетности Центрального банка Российской Федерации. Погрешность тестирования (обобщения) обученной и оптимизированной нейронной сети составила 6,3%. Исследования моделируемой области, – банковской сферы РФ, – выполнены путем проведения виртуальных компьютерных экспериментов, в ходе которых вычисления с помощью нейронной сети производились при изменении одного из пятнадцати входных параметров, характеризующих банки, в то время как остальные параметры сохранялись неизменными. В частности, исследовалось влияние на вероятность банкротства банка коэффициента долгосрочной ликвидности, вида организационно-правовой формы, показателя размера крупных кредитных рисков, места регистрации банка. В результате сделан вывод о том, что повышение коэффициента долгосрочной ликвидности снижает вероятность банкротства банка, однако, начиная с определенного уровня, зависящего от других параметров конкретного банка, повышение данного показателя увеличивает вероятность его банкротства. Существенное влияние на успешность функционирования банка оказывает организационно-правовая форма банка, а также место его регистрации. Однако это влияние неоднозначно и в каждом конкретном случае может проявляться по-разному, в зависимости от множества других параметров банка и его деятельности. Приведен пример применения математической модели для разработки рекомендаций по снижению вероятности банкротства одного из банков.

Ключевые слова: банк, лицензия, банкротство, ликвидность, активы, капитал, прогноз, модель, нейронная сеть.

1. Введение

Актуальность проблемы прогнозирования банкротства предприятий, в частности, банкротства банков привела к тому, что сегодня в мире существует большое количество различных методик. В то же время, по мнению специалистов, не существует единого общепринятого метода. Кроме того, как отмечают многие российские авторы, многочисленные попытки применения иностранных моделей прогнозирования банкротств в отечественных условиях не позволили получить достаточно точные результаты.

Целью настоящей работы является создание математической модели, предназначенной для оценки вероятности банкротства банков по причине отзыва лицензии. Какие закономерности существуют между вероятностью отзыва лицензии и параметрами банка, чем руководствуется Центральный банк Российской Федерации (ЦБ РФ), отзывая лицензии банков – можно узнать, если проанализировать его финансовую отчетность. Перспективным инструментом для выявления скрытых в статистической информации закономерностей и построения на их основе математических моделей являются нейросетевые технологии.

Нейрокомпьютерные и нейросетевые технологии – это одна из наиболее эффективных стратегий искусственного интеллекта. Они наследуют от своего прототипа – мозга, его полезные свойства: способность извлечения знаний из статистических данных, способность обобщения их в виде закономерностей моделируемых предметных областей, свойство интуиции [1], как способность делать правильные прогнозы и принимать верные решения в тех случаях, когда обычная логика оказывается бессильной. Как убедительно показывает наш собственный опыт ([2, 3], www.PermAi.ru), хорошо спроектированные и правильно обученные нейронные сети являются мощным инструментом, позволяющим выявлять закономерности практически любых предметных областей и строить адекватные математические модели в промышленности, в экономике и бизнесе, в политологии, в криминалистике, в медицине, в экологии, в исторических науках, спорте и др.

Возможности применения нейросетевых технологий в прогнозировании банкротства предприятий исследовались в работах ряда зарубежных авторов. Так, в работе [4] рассмотрены различные подходы к построению моделей прогнозирования банкротства предприятий в Тайване. В результате исследования

был сделан вывод о том, что нейронная сеть обладает такой же эффективностью в оценке вероятности финансового краха, как CRIS (composite rule induction system), и большей эффективностью, чем логит-модель. Результаты работы [5] демонстрируют более высокую, чем модель на основе MDA (множественный дискриминантный анализ) эффективность нейронных сетей. Среди их преимуществ указывается способность прогнозировать банкротство на более ранних стадиях. В работе [6], посвященной прогнозированию платежеспособности российских предприятий, подчеркивается преимущество нейросетевого метода в случае нелинейного характера связей между показателями.

2. Математическая постановка задачи

При постановке задачи была введена гипотеза о том, что между финансово-юридическими параметрами банка и вероятностью отзыва его лицензии существуют некие закономерности, которые, впрочем, могут нарушаться, например, в случаях коррупции, рейдерского захвата и пр. Было введено предположение, что таких случаев в России относительно мало, и что они должны отображаться в статистической информации в виде отдельных выбросов. В качестве входных параметров нейросетевой математической модели были выбраны критерии, которые являются наиболее широко распространенными, включенными в официальные нормативы ЦБ РФ для банков и, в то же время, доступными. Всего было использовано 15 входных параметров:

- x_1 – достаточность собственных средств;
- x_2 – коэффициент мгновенной ликвидности;
- x_3 – коэффициент текущей ликвидности;
- x_4 – коэффициент долгосрочной ликвидности;
- x_5 – показатель максимального размера риска на одного заемщика или группу связанных заемщиков;
- x_6 – показатель максимального размера крупных кредитных рисков;
- x_7 – показатель максимального размера кредитов, банковских гарантий и поручительств, предоставленных банком своим участникам (акционерам);
- x_8 – показатель совокупной величины риска по инсайдерам банка;
- x_9 – показатель использования собственных средств (капитала) банка для приобретения акций (долей) других юридических лиц;
- x_{10} – рентабельность активов;
- x_{11} – год внесения в книгу государственной регистрации;
- x_{12} – место регистрации банка;

x_{13} – правовая форма;
 x_{14} – размер уставного капитала;
 x_{15} – размер активов.

Среди входных параметров модели имеются качественные факторы, характеризующие место регистрации банка (x_{12}) и его организационно-правовую форму (x_{13}). Эти показатели были закодированы следующим образом. Если банк зарегистрирован в Москве или Санкт-Петербурге, то параметру x_{12} присваивается значение «1», если же он зарегистрирован в других городах, то «0». Если банк зарегистрирован как ОАО, то параметру x_{13} присваивается значение «1», ООО – «2» и ЗАО – «3».

Выходная переменная y принимает значение, равное единице, если у банка была отозвана лицензия, и значение ноль, если банк продолжает свою деятельность.

В качестве источника информации для обучения нейронных сетей были использованы данные финансовой отчетности с сайта ЦБ РФ. Причинами банкротств банков, данные которых брались с сайта, были: неспособность удовлетворить требования кредиторов по денежным обязательствам и исполнить обязанность по уплате обязательных платежей, потеря ликвидности, неисполнение обязательств перед кредиторами, бездействие по восстановлению финансового положения, отсутствие адекватных резервов на возможные потери по ссудам, предоставление недостоверной финансовой отчетности, проведение операций, связанных с переводами денежных средств в пользу нерезидентов по различным сомнительным сделкам, несоблюдение сроков направления в Росфинмониторинг сообщений об операциях, подлежащих обязательному контролю, а также противодействие проведению инспекционной проверки и совершению надзорных действий.

На основании этой информации было сформировано множество, содержащее 111 примеров, из которых 52 примера являются данными обанкротившихся банков, 59 примеров относятся к работающим банкам. Все множество примеров разбито на обучающее, использованное для обучения сети, и тестирующее, предназначенное для проверки ее прогностических свойств. Естественно, что примеры тестирующего множества при обучении сети не использовались.

Проектирование, оптимизация, обучение, тестирование нейронной сети и эксперименты над нейросетевой математической моделью выполнялись с помощью нейропакета [2]. Оптимальная структура нейронной сети представляла собой персептрон

[3], имеющий пятнадцать входных нейронов, один скрытый слой с пятью нейронами и один выходной нейрон. В качестве активационных функций нейронов скрытого слоя и выходного нейрона использовались сигмоидные функции [3].

После обучения, прогностические свойства сети проверялись на тестирующих примерах. Изначально среднеквадратичная ошибка тестирования нейронной сети составила 13,5 %.

В процессе обучения и тестирования нейросети были обнаружены выбросы, выявленные с помощью методики [7]. Суть этой методики состоит в поочередном исключении примеров из обучающего множества и наблюдении за погрешностью нейросети, обученной на этих урезанных множествах. Если пример обучающего множества является выбросом и выпадает из закономерности, характерной для исследуемой предметной области, то его удаление из обучающего множества приводит к падению погрешности обучения сети и повышению ее обобщающих свойств, тогда как исключение обычных примеров существенного влияния на качество сети, как правило, не оказывает.

В число выбросов, обнаруженных с помощью этой методики, попали данные о пяти банках, среди которых: «Сембанк» и «Анкор-банк». Данные о первом банке были квалифицированы нейронной сетью как выброс, по-видимому, в связи с тем, что у него была отозвана лицензия, несмотря на то, что банк имел хорошие показатели достаточности капитала и коэффициенты ликвидности. Второй банк, наоборот, не подвергся санкциям ЦБ РФ несмотря на то, что имел довольно низкие показатели достаточности капитала и самый маленький размер уставного капитала из всей выборки.

После удаления обнаруженных нейросетью выбросов среднеквадратичная ошибка тестирования (обобщения) снизилась до 2,6%. Причем дополнительное тестирование нейронных сетей по методу multi-fold cross-validation [3] не показали сколько-нибудь заметного увеличения погрешности тестирования.

Обучение, оптимизация и тестирование нейронных сетей проводилось согласно методике, принятой в Пермской научной школе искусственного интеллекта ([2-3], www.PermAi.ru).

3. Исследование предметной области

После проверки разработанной нейросетевой математической модели на адекватность и точность, ее можно использовать для исследования предмет-

Таблица 1.

Параметры исследуемых банков

Банк	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	Y
1. Банкрот	10	19	92	30	24	306	0	1	10	0	1992	1	1	150000000	1695047	0,99
2. Не банкрот	10	19	92	30	24	306	0	1	0,1	0,3	1992	1	1	550000000	18991000	0,09

ной области. Эта задача решалась путем проведения виртуальных компьютерных экспериментов, в ходе которых вычисления с помощью нейронной сети производились при изменении одного из входных параметров, в то время как остальные параметры «замораживались», т.е. сохранялись неизменными.

3.1. Влияние коэффициента долгосрочной ликвидности

Рассматривались характеристики двух банков, первый из которых изначально классифицировался как банкрот, а второй банк – как не банкрот. Значения параметров рассматриваемых банков приведены в табл. 1. Как видно из таблицы, банки существенно отличаются показателем использования собственных средств (капитала) для приобретения акций (долей) других юридических лиц (x_9), рентабельностью активов (x_{10}), размером уставного капитала (x_{14}) и размером активов (x_{15}).

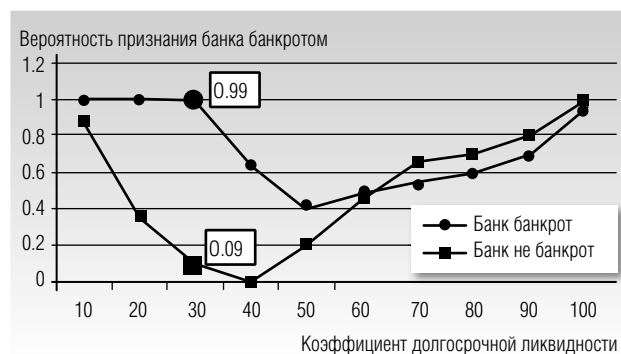


Рис. 1. Влияние коэффициента долгосрочной ликвидности на вероятность признания банка банкротом. Маркером увеличенного размера отмечено реальное состояние исследуемых банков

Анализируя результаты исследований, представленные на рис. 1, можно заключить, что значение коэффициента долгосрочной ликвидности оказывает существенное влияние на вероятность признания банка банкротом, причем кривые имеют U-образную форму. Повышение коэффициента долгосрочной ликвидности до определенного уровня снижает вероятность банкротства. Однако с определенного момента дальнейшее повышение данного показателя увеличивает вероятность банкротства. Оптимальным показателем долгосрочной ликвидности для изначально надежного банка является значение 40, а для банка банкрота – 50.

Нужно отметить, что вывод об U-образной форме зависимости подтверждается и теоретическими исследованиями. Слишком высокая ликвидность снижает прибыльность операций банка, поскольку большая часть денежных средств банка не инвестируется с целью получения доходов. Это влечет недополучение банком потенциальной прибыли и, как следствие, менее устойчивое финансовое состояние.

С другой стороны, слишком низкая ликвидность связана с риском дефолта банка вследствие невозможности удовлетворить в необходимые сроки финансовые требования вкладчиков, кредиторов и других клиентов. Например, невозможность банка вернуть депозит и/или межбанковский кредит, а также проводить операции по расчетным счетам клиентов в необходимые сроки является одной из наиболее частых причин отзыва лицензий у банков.

Таблица 2.

Параметры исследуемых банков

Банк	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	Y
1. Банкрот	25	50	60	28	14	200	0	1	0	0,2	1994	0	2	200000000	1220000	0,997
2. Не банкрот	14	50	60	60	14	200	0	0	0	0,2	1991	0	2	300000000	21991000	0,0

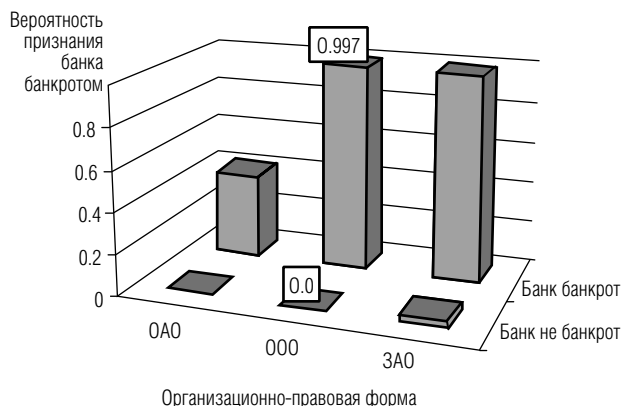


Рис. 2. Влияние организационно-правовой формы на вероятность признания банка банкротом

3.2. Влияние организационно-правовой формы

Рассматривались два банка. Первый банк изначально классифицировался как банкрот, второй банк – как не банкрот. Организационно-правовая форма обоих банков – ООО. Параметры рассматриваемых банков приведены в *табл. 2*.

Как видно из *рис. 2*, где приведены результаты исследований, для изначально нормально функционирующего банка (для него $y=0$) изменение организационно-правовой формы не оказывает существенного влияния на признание его банкротом. Для рассматриваемого банка банкрота (для него $y=0,997$) существует определенная зависимость, а именно: банк, существующий в форме ООО, имел бы меньшую вероятность банкротства, если бы был зарегистрирован в форме ОАО.

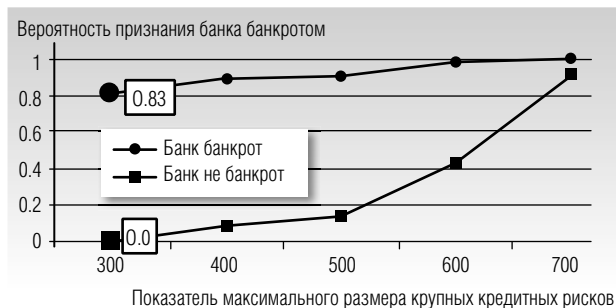


Рис. 3. Влияние показателя максимального размера крупных кредитных рисков на вероятность признания банка банкротом

3.3. Влияние показателя максимального размера крупных кредитных рисков

Рассмотрены два банка с одинаковым показателем максимального размера крупных кредитных рисков, отличающиеся по остальным входным факторам. Параметры банков приведены в *табл. 3*. По результатам исследований нейросетевой математической модели, приведенным на *рис. 3*, можно сделать вывод о прямой зависимости между показателем максимального размера крупных кредитных рисков и вероятностью признания банкротства как для изначально нормально функционирующего банка, так и для банка, признанного банкротом.

3.4. Влияние места регистрации банка

Рассмотрены характеристики трех банков, значения параметров которых представлены в *табл. 4*, а результаты исследований – на *рис. 4*.

Банк №1 изначально классифицировался как нормально функционирующий, и смена места регистрации банка не повлияла на вероятность признания его банкротом.

Банк №2 изначально рассматривался, как зарегистрированный в регионах и был признан банкротом ($y=0,87$). Смена регистрации банка на Москву и Санкт-Петербург изменила категорию данного банка на нормально функционирующий.

Банк №3 обладает теми же характеристиками, что и банк №2, кроме параметра x_4 (коэффициент долгосрочной ликвидности). Этот параметр банка №3 ниже на 43%. На *рис. 4* можно видеть, что данный банк является банкротом в Москве и Санкт-Петербурге, а в регионах он банкротом не является.

Таким образом, в результате выполненных исследований можно сделать заключение о существовании связи между местом регистрации банка и вероятностью признания его банкротом. Однако при этом стоит учитывать влияние других факторов. Если у банка хорошие показатели финансовой устойчивости (Банк №1), то он остается нормально

Таблица 3.

Параметры исследуемых банков

Банк	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	y
1. Банкрот	10	24	60	20	10	300	19	1	0	0	1991	1	2	200000000	1220000	0,83
2. Не банкрот	12	19	63	30	14	300	14	0	0	0,1	1989	1	1	3400000000	21991000	0,0

Таблица 4.

Входные параметры банков и результаты исследования

Банк	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	y
Банк №1	16	45	76	90	14	72	0	1	0	0,5	1989	0	1	40000000000	25000000	0,0
	16	45	76	90	14	72	0	1	0	0,5	1989	1	1	40000000000	25000000	0,0
Банк №2	13	55	64	70	9	21	18	0	0	0	1991	0	1	160000000000	10000000	0,87
	13	55	64	70	9	21	18	0	0	0	1991	1	1	160000000000	10000000	0,11
Банк №3	13	55	64	40	9	21	18	0	0	0	1991	0	1	160000000000	15000000	0,34
	13	55	64	40	9	21	18	0	0	0	1991	1	1	160000000000	15000000	0,99

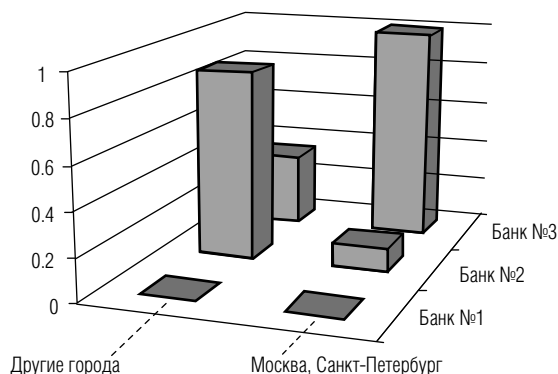


Рис. 4. Влияние места регистрации банка на вероятность признания его банкротом

функционирующим при смене Москвы (Санкт-Петербурга) на другой город. Если региональный банк признан банкротом (Банк №2), то по результатам исследования нейронной сети, он может улучшить свое положение, если позиционируется в Москве или в Санкт-Петербурге. Однако может быть и наоборот. Например, в ситуации, если банк характеризуется низким коэффициентом долгосрочной ликвидности и успешно работает в регионе (Банк №3), то в Москве и в Санкт-Петербурге такой банк может стать банкротом.

3.5. Пример разработки рекомендаций по снижению вероятности банкротства банка «Акционерный коммерческий банк «Кодекс»

Характеристика параметров рассматриваемого банка приведена в табл. 5, из которой видно, что банк является банкротом.

Как следует из рис. 5, уменьшение коэффициента долгосрочной ликвидности (x_4) с 82,5 до 55 приводит к снижению вероятности признания банкротства с 1 до 0,65. Из этого же рисунка видно, что уменьшение показателя совокупной величины риска по инсайдерам (x_9) с 2 до 0 приводит к снижению вероятности

признания банка банкротом с 1 до 0,62. Аналогично, из рис. 5 следует, что уменьшение показателя использования собственных средств (капитала) банка для приобретения акций (долей) других юридических лиц (x_9) с 18 до 9 приводит к снижению вероятности банкротства банка с 1 до 0,66. И наконец, как видно из таблицы 6, одновременное изменение всех трех параметров (x_4 , x_8 , x_9) в указанных выше пределах позволяет снизить вероятность банкротства банка «Кодекс» до нуля.

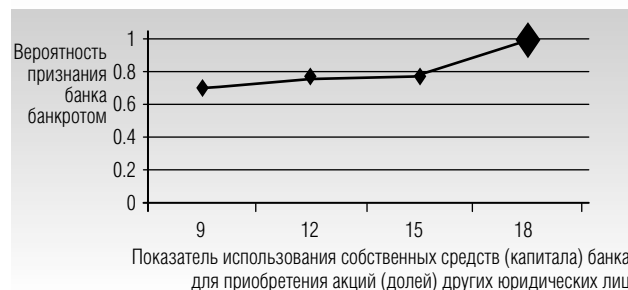
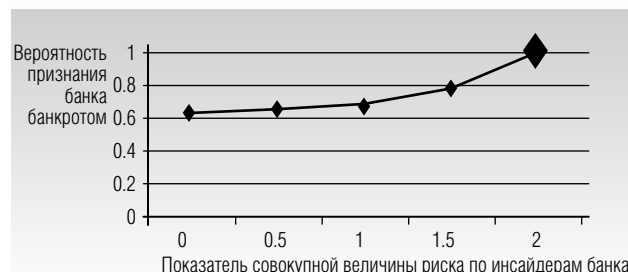
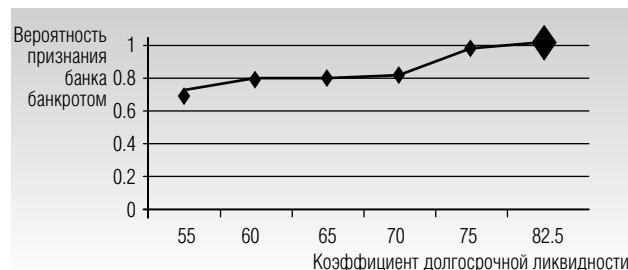


Рис. 5. Влияние параметров банка «Кодекс» на вероятность признания его банкротом. Маркером увеличенного размера отмечены реальные значения параметров банка

Таблица 5.

Параметры банка «Кодекс»

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	y
27,8	55,3	67,3	82,5	23	218,6	0	2	18	0	1994	1	1	80000000	733755	1

Таблица 6.

Параметры банка «Кодекс» после корректировки

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	y
27,8	55,3	67,3	55	23	218,6	0	0	9	0	1994	1	1	80000000	733755	0

4. Заключение

Применение метода нейросетевого моделирования позволило создать систему, с высокой точностью оценивающую вероятность банкротства банков по пятнадцати входным параметрам. С помощью вычислительных экспериментов над разработанной математической моделью выявлен ряд закономерностей исследуемой предметной области – банковской системы России. В частности, сделан вывод о том, что повышение коэффициента долгосрочной ликвидности до определенного момента снижает вероятность банкротства банка, однако с определенного уровня повышение данного

показателя увеличивает вероятность банкротства банка. Оптимальное значение коэффициента долгосрочной ликвидности для каждого конкретного банка может быть определено путем вычислений с помощью разработанной нейросетевой системы.

Вычислительные эксперименты позволили установить, что организационно-правовая форма банка, а также его место регистрации оказывают существенное влияние на состояние банка, однако это влияние неоднозначно и в каждом конкретном случае может проявляться по-разному, в зависимости от множества других параметров, характеризующих банк. ■

Литература

1. Ясницкий Л.Н., Данилевич Т.В. Современные проблемы науки. М.: БИНОМ. Лаборатория знаний, 2008. 295 с.
2. Ясницкий Л.Н., Богданов К.В., Черепанов Ф.М. Технология нейросетевого моделирования и обзор работ Пермской научной школы искусственного интеллекта // *Фундаментальные исследования*. 2013. № 1–3. С. 736–740.
3. Ясницкий Л.Н. Интеллектуальные информационные технологии и системы. Пермь: Пермский университет, 2007. 271 с.
4. Ko L.-J., Blocher E.J., Lin P.P. Prediction of corporate financial distress: An application of the composite rule induction system // *The International Journal of Digital Accounting Research*. 2009. Vol. 1, No. 1. P. 69–85.
5. Coats P.K., Fant L.F. Recognizing financial distress patterns using a neural network tool // *Financial Management*. 1993. Vol. 22. P. 142–155.
6. Богданова Т.К., Шевгунов Т.Я., Уварова О.М. Применение нейронных сетей для прогнозирования платежеспособности российских предприятий обрабатывающих отраслей // *Бизнес-информатика*. 2013. №2 (24). С. 40–48.
7. Черепанов Ф.М., Ясницкий Л.Н. Нейросетевой фильтр для исключения выбросов в статистической информации // *Вестник Пермского университета. Серия: Математика. Механика. Информатика*. 2008. № 4. С. 151–155.

NEURAL NETWORK DESIGNED TO ESTIMATE PROBABILITY OF BANK BANKRUPTCIES

Leonid YASNITSKY

Professor, Department of Information Technologies in Business, Faculty of Business Informatics, National Research University Higher School of Economics (Perm branch); The Chairman of the Perm office of Scientific Council of the Russian Academy of Sciences on methodology of artificial intelligence
Address: 38, Studencheskaya street, Perm, 614046, Russian Federation
E-mail: yasn@psu.ru

Dmitry IVANOV

Post-graduate student, Department of Information Systems and Mathematical Methods in Economics, Faculty of Economics, Perm State University
Address: 15, Bukireva street, Perm, 614990, Russian Federation
E-mail: idv_1988@mail.ru

Ekaterina LIPATOVA

Student, Faculty of Economics, National Research University Higher School of Economics (Perm branch)
Address: 38, Studencheskaya street, Perm, 614046, Russian Federation
E-mail: Lipatova_katya@mail.ru

The object of research is the banking system of Russia. The study purpose is to build a mathematical model to estimate probability of bank bankruptcies due to license revocation. An instrument to build the model is neural networks to be trained on financial statements of the Central Bank of the Russian Federation. The testing error of the trained and optimized neural network has constituted 6.3%. The studies of the modeled area – the banking system of the Russian Federation – have been carried out through virtual computer experiments. The neural network calculations have been made by changing one of fifteen bank-related input parameters with other parameters remaining constant.

In particular, the impact of long-term liquidity ratio, the type of business legal status, the exposure to large credit risks and bank place of registration on bank bankruptcy probability has been investigated. As a result the conclusion has been formulated that the increase of long-term liquidity ratio reduces the bank bankruptcy probability. However, starting with a certain level, depending on other parameters of a specific bank, the increase of this indicator increases the probability of its bankruptcy. Essential impact on successful bank performance is exerted by bank's business legal status, as well as the place of its registration. However, this impact is ambiguous and may manifest itself differently in each individual case, depending on many other bank parameters and its operations. A case study involving the mathematical model application to formulate recommendations to reduce bankruptcy probability of a bank is given.

Key words: bank, license, bankruptcy, liquidity, assets, capital, forecast, model, neural network.

References

1. Yasnitsky L.N., Danilevich T.V. (2008) *Sovremennye problemy nauki* [Modern problems of science]. Moscow: BINOM. (in Russian)
2. Yasnitsky L.N., Bogdanov K.V., Cherepanov F.M. (2013) Tehnologija nejrosetevogo modelirovanija i obzor rabot Permskoj nauchnoj shkoly iskusstvennogo intellekta [Neural network technology and an overview of the Perm scientific school of artificial intelligence]. *Fundamental research*, no. 1–3, pp. 736–740. (in Russian)
3. Yasnitsky L.N. (2007) *Intellektual'nye informacionnye tehnologii i sistemy* [Intelligent information technologies and systems]. Perm: Perm University. (in Russian)
4. Ko L.-J., Blocher E.J., Lin P.P. (2009) Prediction of corporate financial distress: An application of the composite rule induction system. *The International Journal of Digital Accounting Research*, vol. 1, no. 1, pp. 69–85.
5. Coats P.K., Fant L.F. (1993) Recognizing financial distress patterns using a neural network tool. *Financial Management*, vol. 22, pp. 142–155.
6. Bogdanova T.K., Shevgunov T.Y., Uvarova O.M. (2013) Primenenie nejronnyh setej dlja prognozirovanija platjezhnesposobnosti rossijskih predpriyatij obrabatyvajushih otraslej [Application of neural networks to predict the solvency of Russian enterprises manufacturing industries]. *Business Informatics*, no. 2 (24), pp. 40–48. (in Russian)
7. Cherepanov F.M., Yasnitsky L.N. (2008) Nejrosetevoj fil'tr dlja iskljuchenija vybrosov v statisticheskoj informacii [Neural network filter for eliminating outliers in the statistical information]. *Bulletin of the University of Perm. Series: Mathematics. Mechanic. Informatics*, no. 4, pp. 151–155. (in Russian)

ОБОСНОВАНИЕ ИНВЕСТИЦИЙ В ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ НА ОСНОВЕ ДЕРЕВА БИЗНЕС-ДРАЙВЕРОВ

Н.Л. КОРОВКИНА

доцент кафедры корпоративных информационных систем,
факультет бизнес-информатики, Национальный исследовательский
университет «Высшая школа экономики»

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

E-mail: nkorovkina@hse.ru

М.С. ФАЙ

студент, Институт им. Хассо Платтнера, School of Design Thinking

Адрес: Germany, D-14482, Potsdam, Prof.-Dr.-Helmert-Str., 2-3

E-mail: msfay@edu.hse.ru

В статье предложен подход к разработке методике обоснования инвестиций в ИТ-проекты за счет установления их взаимосвязи с бизнес-драйверами компании. Представляя собой факторы повышения ценности компании, бизнес-драйверы позволяют определить, каким образом стратегические преимущества от реализации ИТ-проектов влияют на показатели предприятия. Дерево бизнес-драйверов базируется на модели Val IT, отображая иерархию нескольких финансовых факторов повышения добавочной экономической стоимости, и дополняется несколькими индивидуальными для индустрий и компаний нефинансовыми факторами (такими, как внутренняя оптимизация, способность к инновациям). Таким путем обеспечивается основа для определения стратегической целесообразности инвестирования в проект.

Подобный метод формулирования потенциальных выгод от информационных технологий (ИТ) в терминах бизнеса становится одним из критериев принятия решений об инвестициях и дополняет идею концепции Value-Based Management. В работе обосновывается, что достаточный уровень покрытия проектом бизнес-драйверов компании способен в ряде случаев компенсировать негативное значение ожидаемых финансовых выгод. Проведенный анализ возможностей обоснования инвестиций в ИТ делает возможным построение матрицы принятия решений на основе оценки риска, количественных показателей, а также соответствия приоритетам компании. Реализация формируемых в работе предложений призвана обеспечить большую прозрачность инвестиций, а также гармонизацию бизнеса и ИТ. Обеспечивается учет особенностей бизнеса и более комплексная оценка эффекта от вложений в ИТ, что нивелирует риски неверной оценки роста денежного потока. Предложенный подход успешно апробирован для оценки портфеля потенциальных ИТ-проектов в крупной производственной компании в России.

Ключевые слова: информационные технологии (ИТ), инвестиции, обоснование проекта, ценность ИТ, портфель ИТ-проектов, бизнес-драйвер, эффективность ИТ.

1. Введение

В 2011 году в журнале Harvard Business Review [18] были опубликованы результаты опроса руководителей ведущих мировых компаний. Согласно результатам опроса, 72% руководителей

предприятий подтвердили, что «широкое использование технологий способствовало созданию дополнительной бизнес-ценности», а также отметили, что ИТ-подразделения стали полноценным фактором влияния на конкурентные преимущества и прибыль компании, конкурируя за инвестиции

даже с проектами в сфере основного производства. Все эти свидетельства конкурентной природы инвестиционного процесса подтверждают, что рассмотрения отдачи от выполнения ИТ-проекта и выбора самого проекта для реализации вне привязки к достигаемым бизнес-выгодам и бизнес-целям недостаточно. И уже сейчас (по результатам проведенного изданиями Information Week и CIO Insight масштабного опроса ИТ-директоров [10]), четыре из пяти компаний отмечают частую невозможность представления выгод от ИТ-проектов в количественном выражении, а три четверти ИТ-директоров говорят о неполном охвате бизнес-выгод от ИТ-проектов текущими методами оценки инвестиций.

Согласно оценке практик оценки и управления выгодами от ИТ предприятий малого и среднего бизнеса, опубликованном профессорами Edith Cowan University и Brunel University [9], отсутствие единой модели управления ИТ-инвестициями¹ приводит к нескольким последствиям:

- ◆ отказ от вложений в ИТ-инфраструктуру, снижающих прибыльность компании в долгосрочном периоде;

- ◆ «интуитивное» и спонтанное инвестирование;

- ◆ непринятие во внимание особенностей бизнеса и отраслевой специфики, а значит, и потеря конкурентных преимуществ.

Исходя из сформулированной проблематики, необходима методика обоснования инвестиций в ИТ, учитывающая стратегию и факторы повышения ценности компании при определении бизнес-выгод от ИТ-проектов, что и обуславливает актуальность исследования.

Цель исследования — разработка комплексной методики обоснования ИТ-инвестиций для обеспечения учета бизнес-приоритетов при формировании портфеля ИТ-проектов.

Исследование было выполнено с учетом двух основных ограничений:

- ◆ отказ от рассмотрения методов экономической

оценки эффективности в пользу ее оценки с точки зрения стратегии бизнеса;

- ◆ ограничение анализа результатов практического применения предложенной методики качественными параметрами на основе проведенных интервью с руководством компаний.

В качестве предпосылки работы выступает утверждение о том, что целесообразность инвестирования в конкретные ИТ-проекты может быть определена при их рассмотрении с точки зрения факторов развития бизнеса, различающихся для различных компаний и индустрий.

2. Учет стратегических приоритетов компании при разработке методики обоснования ИТ-инвестиций

Любая компания, вне зависимости от отрасли, имеет целью своей деятельности получение определенных выгод, являющихся по сути функцией различных факторов: они могут быть представлены в виде роста выручки, роста маржинальности бизнеса, либо повышения эффективности основной деятельности (при допущении, что в рассматриваемые примеры не будут включены некоммерческие организации и предприятия, не имеющие акционерного капитала). Далее для обозначения подобных факторов повышения ценности компании будет применяться термин «бизнес-драйвер» (business driver). По определению одного из самых известных словарей в области технологий, Technopedia², бизнес-драйвер понимается как «ресурс, процесс или условие, необходимые для обеспечения роста и развития бизнеса» [15]. Этот термин также широко используется крупными международными компаниями, в частности, Accenture³, одним из лидеров в отрасли ИТ-консалтинга [14]. В литературе также встречается синонимичное понятие «ценностный драйвер» (value driver), разработанное консалтинговой компанией McKinsey⁴ для описания факторов повышения ценности компании (ее доходности и роста) [8].

¹ Здесь и далее в работе под термином «ИТ-инвестиции» будем понимать «инвестирование денежных средств/капитала в ИТ-технологии с целью получения прибыли/возврата от инвестиций в виде процента, дохода или повышения ценности», данное известным в мире ресурсом dictionary.reference.com. В свою очередь, ИТ-инвестиции осуществляются в проекты, реализуемые ИТ-департаментами компаний

² <http://www.techopedia.com/>. Technopedia — один из крупнейших в мире репозиториев информации в области программного и аппаратного обеспечения, ИТ-процессов и проектов

³ <http://www.accenture.com/>. Accenture — международная компания, оказывающая услуги в области управленческого консалтинга, технологий и аутсорсинга.

⁴ <http://www.mckinsey.com/>. McKinsey & Company — ведущая международная компания, предоставляющая услуги в области управленческого консалтинга

По результатам анализа результатов нескольких исследований [7, 14, 17] представляется возможным выделить две основные категории бизнес-драйверов: финансовые и нефинансовые (определяемые стратегией и рыночной динамикой). Финансовые драйверы – факторы экономического успеха предприятия, напрямую (математически) связанные с ключевыми статьями финансовой отчетности. Нефинансовые бизнес-драйверы – факторы успеха в конкурентной среде, важные для владельцев и руководства организации, опосредованно (логически) связанные с ценностью и успешностью компании. Именно в том, чтобы реализуемые проекты в совокупности вносили максимальный вклад в развитие бизнес-драйверов (и повышение стоимости компании), и заинтересовано руководство компании.

На основании исследования, проведенного Oana Velcu [12] представляется возможным выделить наиболее распространенные выгоды от ИТ-инвестиций, установив их связь с бизнес-результатами. Это позволит реализовать сформулированное для методики требование о формировании понимания стратегии развития бизнеса и определения факторов, посредством которых информационные системы влияют на показатели компании.

Далее в работе предлагаемый подход рассмотрен более детально.

3. Требования к методике обоснования ИТ-инвестиций и критериям ее эффективности

В первую очередь, определим основные проблемы обоснования ИТ-инвестиций, с которыми сталкиваются российские компании. При этом в качестве исходной базы будут использоваться результаты исследования в области эффективности ИТ, проведенного рабочей группой проекта IT-value.ru⁵ [1]. В частности, рассмотрим список наиболее часто используемых методов оценки эффективности инвестиций (выделенных на основе анализа данных более чем 300 российских компаний) для каждой группы заинтересованных лиц предприятия (например, инвесторов и менеджеров). Среди упомянутых в исследовании IT-value.ru проблем по конкретным используемым компаниями методикам оценки выгод от ИТ можно отметить следующие:

♦ методика определения совокупной стоимости владения не позволяет оценить как информационные технологии влияют на развитие бизнеса [1];

♦ показатели оценки денежного потока (ROI, NPV, IRR и другие) имеют высокую вероятность ошибки в оценках;

♦ оценка выгод через сравнение с аналогичными проектами (бенчмаркинг) не позволяет понять, как особенности предлагаемого проекта влияют на показатели компании, и является субъективной по своей сути.

Основываясь на вышеприведенных проблемах различных методик оценки выгод, сформулируем требования к разрабатываемой методике обоснования ИТ-инвестиций и критерии ее проверки на соответствие данным требованиям.

Таблица 1.

Определение требований к методике обоснования ИТ-инвестиций

Выделенные проблемы	Требования к методике обоснования ИТ-инвестиций	Критерий проверки соответствия методики заявленным требованиям
Отсутствие учета особенностей бизнеса	Построение методики на основе специфики стратегии и рисков конкретного предприятия	Использование в основе методики подхода, учитывающего специфику бизнес-модели и рисков компании
Высокая ошибка оценки роста денежного потока, вызванного инвестициями в ИТ. Стандартные показатели оценки денежного потока применяются для всех типов инвестиций и не учитывают специфику ИТ (в частности, наличие системных/комплексных эффектов)	Формирование потенциальных выгод от ИТ в терминах бизнеса и стратегии	Наличие возможностей определения нефинансовых потенциальных выгод для бизнеса
Оценка выгод через сравнение с аналогами представляет собой уход от особенностей конкретного оцениваемого проекта и является субъективной	Формирование критериев включения проекта в портфель ИТ-инвестиций	Наличие возможностей определения целесообразности включения нового проекта в портфель ИТ-инвестиций

⁵ IT-value.ru – научно-исследовательский проект, участниками которого на основе анализа множества работ мирового уровня в области эффективности ИТ сформулировали несколько фундаментальных гипотез для проверки на практике в условиях российской специфики.

Исходя из данных требований возможно моделирование и разработка методики, позволяющей обеспечить учет бизнес-приоритетов при формировании портфеля ИТ-проектов на основе разработки комплексной методики обоснования ИТ-инвестиций. По информации об используемых компаниями методик оценки инвестиций [3, 16, 19] для дальнейшего исследования и доработки была выделена методика Val IT, разработанная Международной ассоциацией – Институтом корпоративного управления ИТ ISACA/ITGI. Она рекомендована к применению организациями в ряде случаев, в частности, когда инвестиции в ИТ не соответствуют бизнес-стратегии или не дают ожидаемого эффекта, либо если отсутствуют механизмы пересмотра портфеля проектов и их «замораживания» в случае необходимости [2].

Краеугольным камнем методики является учет четырех критериев при выборе проекта для реализации:

- ◆ соответствие проекта стратегии компании;
- ◆ значения ожидаемых финансовых показателей;
- ◆ перечень ожидаемых нефинансовых выгод (изменения на уровне функций/ операций/ модели бизнеса, ассоциированные с реализацией проекта);
- ◆ приемлемость для компании уровня риска проекта.

Весь процесс обоснования ИТ-инвестиций в данном случае может быть представлен в виде единой логической цепочки. По результатам обследования предприятия и сбора данных составляется матрица с пятью столбцами, которая призвана дать ответ на вопрос, какие из проектов принимать к реализации. Она отображает все возможные сочетания комбинаций «да – нет» по четырем критериям (приемлемость уровня риска проекта и т.д.) и для каждого из них дает в пятом столбце решение, включать ли

проект в общую программу проектов. Фрагмент подобной матрицы, в оригинале содержащей 16 строк (по числу всех возможных комбинаций ответов) и основанной на результатах ее применения в сотнях компаний, приведен в табл. 2.

Непосредственно для принятия решения по конкретному проекту возможно составление подобной матрицы с ответами по проекту (например, «да – да – да – да») и сопоставление этой комбинации с табл. 2, чтобы из предлагаемого решения в результирующем столбце узнать, следует ли реализовывать ли проект. Однако данный подход не предлагает вариантов формулирования нефинансовых выгод (в особенности, в терминах бизнеса) и явным образом не определяет критерии соответствия проекта стратегии. Именно поэтому в качестве совершенствования приведенной в Val IT матрицы принятия решения об ИТ-проектах предполагается использование дерева бизнес-драйверов как способа определения нефинансовых выгод и степени соответствия стратегии. Крайне важно отметить, что основой для включения в методику системы учета приоритетов бизнеса (дерева бизнес-драйверов) может быть выбрана другая (отличная от Val IT) базовая методология, используемая компанией, если при этом она будет соответствовать сформулированным критериям адекватности методики. Далее последовательно сформулированы и описаны компоненты методики для обеспечения учета требований, приведенных в табл. 1.

4. Описание потенциальных выгод от ИТ в терминах бизнеса и стратегии

В соответствии с выделенными требованиями, опишем процесс построения модели, основываясь на ранее представленной иерархической структу-

Таблица 2.

Адаптированная матрица принятия решения об ИТ-инвестициях (на основе ValIT, фрагмент)

Результат анализа данных ИТ-проекта				Решение об инвестировании
Приемлем ли для компании уровень риска при реализации проекта?	Приемлем ли уровень ожидаемых финансовых показателей?	Четко ли определены нефинансовые выгоды от проекта?	Соответствует ли проект стратегии бизнеса?	
Да	Да	Да	Да	Включить проект в портфель проектов, если рассчитанная величина риска ниже величины рисков реализуемых проектов
Да	Нет	Да	Нет	Отказ от включения проекта в портфель проектов

ре и практиках формулирования самих операционных факторов успешности компании, приведенных в публикации консалтинговой компании McKinsey [7]. Именно в ней представлена разработанная McKinsey концепция управления, основанного на ценностях (Value-Based Management, VBM). Выбор этого подхода был осуществлен на основе анализа нескольких исследований (преимущественно консалтинговых компаний), наиболее часто по сравнению с другими оперирующими терминами «бизнес» и «ценностных» драйверов и принципами их выделения. Отметим, что основным критерием для принятия решения является ранее упомянутое требование к наличию возможностей определения нефинансовых потенциальных выгод для бизнеса.

Переходя к рассмотренным в процессе анализа исследованиям, отметим, что Stewart&Co[11] концентрируется преимущественно на рассмотрении показателя экономической добавленной стоимости (economic value added, EVA) в качестве основного элемента определения ценности компании. BCG⁶, еще один глобальный игрок рынка бизнес-консалтинга, использует основанную на четырех финансовых компонентах метрику CFROI⁷, исходящую из гипотезы о том, что на рынке наиболее важна способность компании генерировать денежные потоки [5].

Наиболее интересными и представляющими возможность для развития идеи выгод для бизнеса, выраженных в нефинансовых терминах являются подходы L.E.K. Consulting⁸ и McKinsey. Первая концепция [17] выделяет три вида драйверов: драйверы роста, эффективности и финансовой состоятельности, связывая их с определенными заинтересованными сторонами в организации (маркетинг и продажи, операции и продукция, финансы и бухгалтерия соответственно). Следующим шагом L.E.K. предлагает разработку «карты драйверов», определяющих для различных «контролируемых компанией факторов» (стоимость материалов, контракты на техобслуживание и ремонт) соответствующие «драйверы ценности» (расходы на ТОиР (техническое обслуживание и ремонт)). Однако данная модель преимущественно концентрируется на затратах, не рассматривая стратегические аспекты.

Наконец, в работе McKinsey [7] предлагается идея двух уровней факторов повышения ценностей компании: общих для всех организаций и индустрий операционно-финансовых факторов доходности и роста (1-й уровень), а также специфических для разных компаний приоритетов более высокого уровня детализации (2-й уровень). Соответственно, согласно данному подходу соответствие ИТ-инициатив стратегическому плану развития компании достигается через выделение критических факторов успеха и связи конкретных инициатив с этими факторами. Именно данный иерархический подход представляет широкие возможности для дальнейшего развития в сторону выделения на одном из уровней стратегических, нефинансовых бизнес-драйверов, повышающих ценность компании.

Выделим основные источники информации для определения того, какие именно бизнес-драйверы способствуют развитию компании (выраженной в подходе VBM в виде экономической добавленной стоимости). В первую очередь, в ходе предшествующего моделированию бизнес-анализа должны быть рассмотрены:

- ◆ годовой отчет (и/или утвержденная стратегия), содержащая актуальные стратегические цели компании;
- ◆ утвержденные показатели деятельности компании в целом и подразделений компании;
- ◆ интервью с руководителями компании, подтверждающие соответствие деятельности компании декларируемым целям и/или актуализирующие их;
- ◆ интервью с руководством бизнес-подразделений и ИТ-департамента.

Как уже было отмечено, построение дерева бизнес-драйверов производится на основе модели ценностных драйверов VBM [16], поэтому первым уровнем всегда будут служить «Доходность» (под элементами которого являются «Выручка» и «Затраты»), а также «Рост/расширение компании» (выражающееся драйверами «Инвестированный капитал» и «Средневзвешенная стоимость капитала, WACC»). Отметим, что концепция VBM предполагает и дальнейшую детализацию этих факто-

⁶ <http://www.bcg.com/>. BCG (Boston Consulting Group) – одна из ведущих международных компаний в области стратегического бизнес-консалтинга, основана в США в 1963 году.

⁷ CFROI, доходность инвестиций на основе денежного потока (cash flow return on investment) – интегральный показатель, включающий такие компоненты, как Валовые инвестиции, Валовый денежный поток, Неамортизируемые активы, Экономический срок жизни активов.

⁸ <http://www.lek.com/>. L.E.K. consulting – одна из ведущих международных компаний в области стратегического бизнес-консалтинга, основана в Лондоне в 1983 году.

ров. Компании могут самостоятельно определять, какие именно элементы они считают более важными для своего бизнеса. К примеру, металлургическая компания для драйвера «Затраты» выделит «Производственные издержки» и «Новые технологии» как наиболее значительные статьи расходов и, соответственно, важные для оптимизации при помощи информационных технологий. А компания-мобильный оператор для драйвера «Выручка» может определить «Маркетинговые активности» и «Привлечение новых абонентов» как наиболее важные факторы, влияющие на выручку и общую доходность бизнеса. Что касается бизнес-драйверов категории «Инвестированный капитал» и «WACC», то они представляются наиболее важными для крупных акционерных компаний, для которых взаимоотношения с инвесторами играют критическую роль.

Как уже было отмечено, для обеспечения корректности модели важно наличие возможностей определения потенциальных выгод для бизнеса в нефинансовых терминах. Поэтому представляется возможным расширить предложенную модель, выделив отдельно группу факторов «Оптимизация бизнес-стратегии». На основе анализа ранее приведенного списка источников информации для выбранной для тестирования модели производственной компании были выделены такие драйверы, как «Эффективность взаимодействия с контрагентами», «Эффективность аналитики» и «Совершенствование системы информирования акционеров». Подобный выбор объясняется стремлением анализируемой компании в ближайшей перспективе организовать публичное размещение акций, а значит, на первый план выходят корректность и своевременность предоставления всей отчетной информации, а также взаимоотношения с потенциальными инвесторами, поставщиками и потребителями.

На основе выделенных приоритетов компании в качестве следующего этапа работы предлагается перейти к другим активностям методики Val IT, а именно – анализу специфики стратегии и рисков компании.

5. Построение методики с учетом специфики стратегии и рисков конкретного предприятия

Ранее в исследовании уже были приведены основные критерии выбора проектов для реализации по методике Val IT. Отметим, что авторами Val IT пред-

лагается перевод всех критериев в шкалу «да – нет» для достижения максимальной однозначности их понимания. Соответственно, итоговый список вопросов, на основе которых представляется возможным определить целесообразность реализации проекта, будет выглядеть следующим образом [16]:

◆ Приемлем ли для компании уровень риска (при реализации проекта)?

◆ Приемлем ли уровень ожидаемых финансовых показателей?

◆ Четко ли определены нефинансовые выгоды от проекта?

◆ Соответствует ли проект стратегии бизнеса?

Однако разработчики методики Val IT на основании проведенного анализа крупнейших компаний в мире отмечают, что при определении ожидаемых денежных потоков, необходимых для достижения целевых показателей прибыли, компании проводят суммирование по отдельным проектам для оценки совокупной прибыли реализации программы. В случае, если потенциальные проекты не достигают критического значения показателей (например, положительного NPV), разницу должны компенсировать нефинансовые выгоды. В случае, если потенциальные проекты не достигают критического значения показателей, принятие решения может (и должно) базироваться на «степени соответствия стратегии компании», с учетом веса этого показателя. Однако в силу того, что методика Val IT является верхнеуровневой, дальнейшей детализации принципа определения данного критерия не приводится. В результате каждая компания использует свои наработки, регулярно их изменяя и корректируя.

Именно на систематизацию и определение оптимального способа обоснования соответствия стратегическим приоритетам направлена предложенная ранее концепция дерева бизнес-драйверов на основе модели VBM. Теперь, чтобы ответить на вопрос «Соответствует ли проект стратегии?» необходимо определить, к каким бизнес-драйверам компании относится потенциальный проект и какова степень приоритетности именно этих бизнес-драйверов.

Рассмотрим данные рассуждения на примере. Как в некоммерческих организациях, так и в других компаниях особое внимание уделяется интеллектуальному капиталу, отношениям с клиентами и поставщиками, и даже просто абстрактной «инновационности» предприятия [12]. Это значит, что конкурентное преимущество, отличающее успешные компании, может быть реализовано за

счет поддерживающих их ИТ-проектов. При этом ИТ-инвестиции, позволяющие повысить удовлетворенность клиентов или снижение себестоимости товара через понижение затрат на транзакцию, представляют собой вполне оценимые выгоды, соотносимые с бизнес-драйверами.

Расстановка приоритетов для данных драйверов и определение наиболее важных среди них в конкретной ситуации – важный элемент выбора между несколькими проектами для реализации, который определенно является отдельной темой для продолжения исследования. В качестве базового метода определения приоритетов проектов предлагается использование системы весовых коэффициентов Фишберна [4]. Данная система присваивает каждому проекту уровень значимости на основе выявления отношений между проектами (строгое/нестрогое предпочтение либо безразличие) и использования простейшей арифметической прогрессии для получения итогового взвешенного коэффициента. Так как оценка по Фишберну отвечает максимуму энтропии информационной неопределенности в отношении объекта исследования (проектов) и является непротиворечивой, она может использоваться для ранжирования по любому из факторов оценки.

Для всех комбинаций возможных вариантов Val IT предлагается определенное решение в области инвестиций. Так, в случае, если уровни риска и финансовых показателей являются приемлемыми, но проект не соответствует общей стратегии бизнеса, рекомендуется «отклонить проект, если только выгоды не смогут быть достигнуты в краткосрочном периоде без оказания негативного влияния на бизнес» [16].

6. Определение единых критериев включения проекта в существующий портфель ИТ-инвестиций

Сформируем итоговую матрицу принятия решения и оценки проекта на основе анализа источников данных о бизнес-требованиях к ИТ, а также приоритетах компании. Основой ее формирования

будут служить четыре выделенных в предыдущем разделе фактора для принятия решений об ИТ-проектах на основе методики Val IT [16].

Рассмотрим пример производственной компании, в которой была впервые апробирована данная методика, в виде трех потенциальных проектов для реализации: «Единый центр обслуживания», «Радиочастотная идентификация (RFID)¹⁰ для повышения точности обработки данных и контроля продукции», «Виртуальное присутствие на расстоянии (телепрезенс)¹¹» (табл. 3).

Таблица 3.

Матрица соответствия ИТ-проектов, бизнес-выгод и драйверов

Проект	Бизнес-выгоды	Бизнес-драйверы
Единый центр обслуживания	Обеспечение высококачественного клиентского сервиса, снижение затрат на обработку транзакций	Эффективность взаимодействия с контрагентами, операционные затраты
Радиочастотная идентификация (RFID) для повышения точности обработки данных и контроля продукции	Точность учета и предотвращение потерь	Автоматизация процессов, повышение производительности
Виртуальное присутствие на расстоянии (телепрезенс)	Снижение компонента SG&A-затрат (в виде оплаты командировок и приемов)	Торговые, общие и административные затраты (SG&A)

В соответствии с ранее описанным принципом, для выбранной производственной компании, выходящей на рынок акций, бизнес-драйверы были сформулированы и соотнесены с предлагаемыми к реализации ИТ-инвестициями. Проект «Единый центр обслуживания» необходим для обеспечения более низкой стоимости транзакции (за счет усовершенствованных механизмов обработки заявок и нескольких линий поддержки), что снижает операционные затраты. В свою очередь, более оперативный и качественный клиентский сервис относится к эффективности взаимодействия с

⁹ Радиочастотная идентификация (RFID, radiofrequency identification) – бесконтактная технология, позволяющая наносить метки на изделия для последующего распознавания. Отличие RFID от традиционных технологий в возможности считывания сотен меток одновременно даже без прямой видимости, высокой оперативности, долговечности нанесенных меток и сложности их копирования.

¹⁰ Телепрезенс (telepresence) – коммуникационная технология, обеспечивающая эффект «виртуального присутствия» за счет элементов интерактивного взаимодействия и передачи высококачественного изображения/звука.

**Анализ ИТ-проектов
на основе матрицы принятия решений, фрагмент**

Проект	Результат анализа данных ИТ-проекта			
	Приемлем ли для компании уровень риска при реализации проекта?	Приемлем ли уровень ожидаемых финансовых показателей?	Четко ли определены нефинансовые выгоды от проекта?	Соответствует ли проект стратегии бизнеса? (На основе дерева бизнес-драйверов)
Единый центр обслуживания	Да	Да	-	Да
Радиочастотная идентификация (RFID) для повышения точности обработки данных и контроля продукции	Да	Нет	Да	Нет
Виртуальное присутствие на расстоянии (телепрезентс)	Да	Нет	Да	Да

контрагентами. Инвестиции в проект «Радиочастотная идентификация» позволят увеличить точность и оперативность обработки данных, достигнуть повышения производительности и автоматизации процессов. Наконец, третий потенциальный проект, – «Виртуальное присутствие на расстоянии (телепрезентс)», – снижает крайне высокие в период подготовки к публичному размещению акций компании общие и административные затраты.

Соответственно, в зависимости от того, какие драйверы будут отмечены наиболее приоритетными на ближайший период (например, один год), можно выделить те проекты, для которых даже недостаточно приемлемый уровень рассчитанных финансовых показателей не будет являться препятствием для финансирования.

Результирующая матрица принятия решения о целесообразности реализации ранее выделенных проектов приведена в *табл. 4*. Ее построению предшествует последовательная реализация выделенных компонентов, расширяющих методику: определение стратегических приоритетов компании в виде дерева бизнес-драйверов и сопоставление с ними конкретных проектов. Именно за счет этих двух активностей на основе дерева бизнес-драйверов возможно построение пятого столбца матрицы – соответствия проекта стратегии.

Именно после построения итоговой матрицы для всех потенциальных инвестиций становится возможным формирование портфеля оцениваемых ИТ-проектов. Так, проект «Виртуальное присутствие на расстоянии (телепрезентс)», несмотря на высокое соотношение за-

трат и возможных финансовых выгод, принят к реализации в силу приоритетности для компании бизнес-драйверов, к которым он относится. В данной ситуации принимающий решение ИТ-комитет может посчитать для этих инвестиций достаточным покрытие разницы с критическими финансовыми показателями ($NPV < 0$) и одобрить проекты в силу их соответствия приоритетам бизнеса компании. Ранее от данных проектов предполагалось отказаться. Проект Единого центра обслуживания принят к реализации в силу своего полного соответствия необходимым критериям, а проект по радиочастотной идентификации, напротив, отклонен (по крайней мере, временно).

7. Анализ результативности методики

В итоге разработанная на основе дерева бизнес-драйверов методика позволяет решить выделенные ранее проблемы (*табл. 5*) и, что важно, без привязки к тому или иному методу количественной оценки инвестиций (а значит, и без значительных ограничений в дальнейшем применении компаниями).

Говоря о достоинствах и ограничениях методики, крайне важно упомянуть, что, как замечает один из российских исследователей проблемы обоснования инвестиций в ИТ П. Солопов [3], чтобы гарантировать адекватные результаты проводимого анализа, требуется выполнение двух важных условий – наличия стратегии развития предприятия, а также заинтересованности руководства в проведе-

Таблица 5.

Покрытие проблем разработанной методикой

Критерий проверки соответствия разработанной методики заявленным требованиям	Покрытие проблем компонентами разработанной методики
Использование в основе методики подхода, учитывающего специфику бизнес-модели и рисков компании	Использование в качестве базы для доработки методики Val IT, содержащей критерии стратегического соответствия, финансовых/ нефинансовых выгод и уровня рисков проекта
Наличие возможностей определения нефинансовых потенциальных выгод для бизнеса	Потенциальные выгоды от ИТ в терминах бизнеса формулируются при построении иерархии бизнес-драйверов на основе модели VBM, учитывающей стратегические приоритеты компании
Наличие возможностей определения целесообразности включения нового проекта в портфель ИТ-инвестиций	Адаптированная матрица принятия решения об ИТ-инвестициях

нии оценки и готовности воспринимать ее результаты. Это может гарантировать как достоверность поступающих для анализа исходных данных, так и возможность дальнейшей реализации подготовленных рекомендаций и их применение на практике.

8. Заключение

Предлагаемая методика принятия решений об инвестициях в ИТ-проекты для определения условий ее оптимального использования была применена на практике в одной из крупнейших российских производственных компаний, выходящих на международный Рынок и пересматривающих в связи с этим событием свою стратегию. Ранее компания отмечала, что ранее основной технологической причиной реализации ИТ-проектов были замена устаревших систем и потребность в апгрейде. По первым результатам проекта внедрения предложенной методики обоснования ИТ-инвестиций на первый план выходит соответствие текущим приоритетам стратегии, также легкость дальнейшей подстройки ландшафта систем под требования бизнеса. Также отмечено, что такие проекты, как развертывание на производстве сети Wi-Fi, были бы невозможны в ближайшей перспективе без использования основанной на дереве бизнес-драйверов методики обоснования.

Таким образом, предложенная методика по своей сути дополняет существующие методики оценки и выбора ИТ-проектов для реализации за счет построения дерева бизнес-драйверов как способа учета бизнес-приоритетов компании. Она может применяться компаниями как самостоятельный инструмент обоснования нефинансовых выгод от ИТ-инвестиций, а также совмещаться с обычными количественными методами оценки финансовых показателей проектов. Однако отметим, что подобное сочетание требует корректного определения соотношения веса финансовых и нефинансовых выгод – и методика на основе дерева бизнес-драйверов не призвана заменять расчет количественных параметров проектов, позволяющих планировать финансовую деятельность компании и проводить мониторинг достижения плановых параметров.

В качестве перспектив работы возможно дальнейшее развитие математической модели приоритизации бизнес-драйверов для принятия решений об очередности и срочности инвестиций, а также выделение уровня индустрий в процессе детализации дерева бизнес-драйверов как промежуточного звена между моделью VBM и низкоуровневыми драйверами компании. Реализация методики в других компаниях, а также ее оценка и совершенствование на основе получаемых результатов позволят повысить эффективность деятельности компании в целом. ■

Литература

1. Ананьин В., Зимин К. Ценность ИТ. Теория и практика// СIO congress Kirov. 28.04.2012. [Электронный ресурс]: <http://bit.ly/139QkWJ> (дата обращения 18.02.2013).
2. Группа компаний 4x4. Повышение эффективности инвестиций в ИТ (Val IT). [Электронный ресурс]: http://4by4.ru/ru/services/val_it (дата обращения 02.05.2013).
3. Солопов П. Оценка целесообразности инвестиций в ИТ. Презентация KPMG. [Электронный ресурс]: <http://bit.ly/WexW7U> (дата обращения 15.04.2013).

4. Фишберн П. Теория полезности для принятия решений. М.: Наука, 1978.352 с.
5. Ameels A., Bruggeman W., Scheipers G. (2002) Value-based management. Control processes to create value through integration. A literature review. Vlerick Leuven Gent Management School, p.27-29. [Электронный ресурс]: http://www.valuebasedmanagement.net/articles_ameels_valuebased_full.pdf (дата обращения 19.04.2014).
6. Hubbell W.W. Combining Economic Value Added and Activities-Based Management // Journal of Cost Management. Spring 1996. P.18-29
7. Koller T., McKinsey & Company Inc. et al. Valuation: Measuring and managing the value of companies. Hoboken, NJ: John Wiley & Sons Inc., 2010.
8. Koller T. What is value-based management? // The McKinsey Quarterly. 1994. No. 3. P.87-101.
9. Love P.E.D., Irani Z. An exploratory study of information technology evaluation and benefits management practices of SMEs in the construction industry // Information & Management. 2004. No. 42. P.227-242.
10. Pisello T. IT Value Chain Management: Maximizing the ROI from IT investments. Performance metrics and management methodologies every IT stakeholder should know. P.17-18. [Электронный ресурс]: <http://www.p2080.co.il/go/p2080h/files/1214612662.pdf> (дата обращения 20.02.2014).
11. Thorp J.A. The Challenge of Value Governing IT Investments. ISACA Victoria, [Электронный ресурс]:<http://www.slideshare.net/khanyasmin/value-of-it-investment.ppt> (дата обращения 12.03.2013).
12. Velcu O. Drivers of ERP systems' business value. Helsinki: Swedish School of Economics and Business Administration, 2009.
13. Waldo R.F. A method to estimate incomes from IT investments. Master thesis XR-n.EE-ICS 2008:016. Sweden: KTH, 2008.
14. Business in Development Study 2012, p.11-13 [Электронный ресурс]: <http://www.enterprise-development.org/download.ashx?id=2136> (дата обращения 19.04.2014).
15. «Business driver». Technopedia dictionary. [Электронный ресурс]: <http://www.techopedia.com/definition/28013/business-driver> (дата обращения 20.04.2013).
16. Enterprise Value. Governance of IT Investments. Business case. Sea Quation Investment Research. IT Governance Institute (ITGI), 2006.
17. Identifying and managing key value drivers // L.E.K. Consulting Executive Insights. 1998. Vol. 1, issue 1. P.2-4.
18. Innovation from efficiency: Unlocking the value in IT systems. Harvard Business School Publishing, 2012.
19. Tops ВI. Методы оценки инвестиций в ИТ: блеск и нищета. [Электронный ресурс]: <http://www.topsbi.ru/?artID=49> (дата обращения 18.01.2013).

JUSTIFICATION OF IT INVESTMENTS BASED ON BUSINESS DRIVERS TREE

Nina KOROVKINA

Associate Professor, Department of Corporate Information Systems,
Faculty of Business Informatics, National Research University
Higher School of Economics

Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation.

E-mail: nkorovkina@hse.ru

Maria FAY

Student, Hasso-Plattner-Institut für Softwaresystemtechnik,
School of design thinking

Address: 2-3, Prof.-Dr.-Helmert-Str., Potsdam, D-14482, Germany

E-mail: msfay@edu.hse.ru

This article presents an approach to developing a methodology for IT project investment feasibility estimation through establishing a link to company business drivers. Representing the factors enhancing the company value, business drivers enable to identify how strategic benefits from IT project implementation influence enterprise performance indicators. A business driver tree is based on the Val IT model, displacing the hierarchy of several financial factors that increase economic value added and being complemented by several industry- and company-specific nonfinancial factors, such as «internal optimization» and «ability to innovate». Thus an underlying basis is provided to evaluate strategic feasibility of project investment.

Such method of formulating potential benefits from information technologies (for IT) in business terms is regarded as a criterion for investment decision-making and complements the Value-Based Management concept. The paper argues that in a number of cases the adequate level of company business drivers coverage by a project can compensate for the negative value of expected financial benefits. The analysis of IT investment feasibility opportunities enables to build a decision-making matrix based on risk assessment, quantitative indicators and alignment with company priorities. Implementation of the recommendations formulated in the paper is intended to ensure a higher IT investments transparency and to achieve harmonization of business and IT. It also enables to take into account business specifics and to carry out a more comprehensive appraisal of IT investment-related effects that mitigates the risk of inaccurate cash flow estimation. The proposed approach has been successfully tested in assessment of a potential IT projects portfolio at a large Russian manufacturing company.

Key words: information technologies (IT), investments, project feasibility, IT value, IT projects portfolio, business driver, IT efficiency.

References

1. Ananyin V., Zimin K. (2012) *Tsennost IT. Teoriya i praktika*. CIO congress Kirov. [Value of IT. Theory and practice]. CIO congress Kirov. Paper presented at CIO congress Kirov 28.04.2012, Kirov, Russian Federation. (in Russian)
2. Gruppa kompaniy 4x4. *Povyshenie effektivnosti investitsiy v IT (Val IT)* [Improving IT investment effectiveness]. Available at: http://4by4.ru/services/val_it (accessed 02.05.2013). (in Russian)
3. Solopov P. (2005) *Otsenka tselesoobraznosti investitsiy v IT* [Assessment of IT investment reasonability]. CIO (electronic journal). Available at: <http://www.osp.ru/cio/2005/09/379612/> (accessed 15.04.2013). (in Russian)
4. Fishburn P. (1978) *Teoriya poleznosti dlya prinyatiya resheniy*. [Utility theory of decision making]. Moscow: Nauka. (in Russian)
5. Ameels A., Bruggeman W., Scheipers G. (2002) Value-based management. Control processes to create value through integration. A literature review. Vlerick Leuven Gent Management School, p. 27-29. Available at: http://www.valuebasedmanagement.net/articles_ameels_valuebased_full.pdf (accessed 19.04.2014).
6. Hubbell W.W. (1996) Combining Economic Value Added and Activities-Based Management. *Journal of Cost Management*, Spring 1996, pp. 18-29.
7. Koller T., McKinsey & Company Inc. et al. (2010) *Valuation: Measuring and Managing the Value of Companies*. Hoboken, New Jersey: John Wiley & Sons Inc.

8. Koller T. (1994) What is value-based management? *The McKinsey quarterly*, no. 3, pp. 87-101.
9. Love P.E.D., Irani Z. (2004) An exploratory study of information technology evaluation and benefits management practices of SMEs in the construction industry. *Information & Management*, no. 42, pp. 227–242.
10. Pisello T. (2003) *IT Value Chain Management: Maximizing the ROI from IT investments. Performance metrics and management methodologies every IT stakeholder should know*. p. 17-18. Available at: <http://www.p2080.co.il/go/p2080h/files/1214612662.pdf> (accessed 20.02.2014).
11. Thorp J.A. The Challenge of Value Governing IT Investments. ISACA Victoria. Available at: <http://www.slideshare.net/khanyasmin/value-of-it-investment.ppt> (accessed 12.03.2013).
12. Velcu O. (2009) *Drivers of ERP systems' business value*. Helsinki: Swedish School of Economics and Business Administration.
13. Waldo R.F. (2008) *A method to estimate incomes from IT investments*. Master thesis XR-n.EE-ICS 2008:016. Sweden: KTH.
14. Accenture (2012) *Business in Development Study 2012*, p.11-13 Available at: <http://www.enterprise-development.org/download.ashx?id=2136> (accessed 19.04.2014).
15. «Business driver». Technopedia dictionary. Available at: <http://www.techopedia.com/definition/28013/business-driver> (accessed 20.04.2013).
16. IT Governance Institute ITGI (2006) *Enterprise Value. Governance of IT Investments. Business case*. p. 22. USA: SeaQuation Investment Research.
17. L.E.K. consulting (1998) Identifying and managing key value drivers. *L.E.K. Consulting Executive Insights*, volume 1, issue 1, pp. 2-4.
18. Harvard Business Review Analytic (2012) *Innovation from Efficiency: Unlocking the Value in IT Systems*. Harvard Business School Publishing.
19. Tops BI. *Metody otsenki investitsiy v IT: blesk i nisheta* [Methods for IT investment evaluation] Available at: <http://www.topsbi.ru/?artID=49> (accessed 18.01.2013). (in Russian)

НЕЧЕТКАЯ МОДЕЛЬ ОЦЕНКИ РИСКОВ ПРОДВИЖЕНИЯ ПРОГРАММНЫХ ПРОДУКТОВ

Ю.П. ЕХЛАКОВ

*доктор технических наук, профессор,
заведующий кафедрой автоматизации обработки информации,
факультет систем управления, Томский государственный университет
систем управления и радиоэлектроники*

Адрес: 634034, г. Томск, ул. Вершинина, д. 74

E-mail: ure@tusur.ru

Н.В. ПЕРМЯКОВА

*аспирант кафедры автоматизации обработки информации,
факультет систем управления, Томский государственный университет
систем управления и радиоэлектроники*

Адрес: 634034, г. Томск, ул. Вершинина, д. 74

E-mail: pnv@mima.tusur.ru

В статье рассматривается задача оценки и анализа рисков при разработке и реализации программы продвижения прикладных программных продуктов, показана целесообразность использования для решения задачи математического аппарата нечетких систем. Сформулирована маркетинговая цель программы как «достижение планового объема продаж в определенном интервале времени при ограничениях на бюджет программы». Учитывая явную логическую взаимосвязь между целями программы и возможными рисками, определены три вида рисков: срыв плановых сроков реализации программы, срыв плановых объемов продаж, превышение бюджета программы. На основе анализа литературы выявлены причины (рискообразующие факторы), способствующие возникновению рисков, предложена классификация рискообразующих факторов, их качественные и количественные характеристики. Рассмотрен реальный пример построения нечеткой модели оценки и анализа рисков при выводе на рынок веб-ориентированной геоинформационной технологии генерального плана предприятия. Выделены одиннадцать входных лингвистических переменных (риск образующих факторов), влияющих на риски программы и три выходных переменных (степень влияния факторов на итоговый риск проекта, степень превышения бюджета проекта и степень достижения плана по объему продаж). Построены две базы правил: правила первой базы используются для определения степени влияния факторов на итоговый риск, правила второй базы используются для определения степени рискованности основных целей программы. На основе алгоритма нечеткого вывода Мамдани получены численные характеристики каждого из рисков и предложены сценарии по реагированию на них. Полученные результаты в практическом плане полезны руководителям малых ИТ-компаний, а также маркетологам, при решении задач продвижения новых товаров на промышленные рынки.

Ключевые слова: классификация рисков, оценка рисков, управление рисками программных проектов, рискообразующий фактор, качественное описание риска, количественное описание риска, категории рисков, метод нечеткой логики, алгоритм Мамдани.

1. Введение

Каждая ИТ-компания, разрабатывая рыночный программный продукт (ПП) стремится обеспечить его дальнейшую коммерциализацию и тиражирование, тем самым закладывая

фундамент для своего устойчивого финансового положения не только в краткосрочной, но и в долгосрочной перспективах. Как правило, эти вопросы рассматриваются в рамках разработки и реализации программы продвижения ПП на рынок. С точ-

ки зрения маркетингового подхода цель программы продвижения можно определить как «достижение определенного объема продаж ПП в определенном интервале времени при ограничениях на бюджет рекламной компании (программы)» [1]. Процесс продвижения программы предполагает изучение рынка потенциальных потребителей и выбор его целевых сегментов, определение наиболее приоритетных бизнес-моделей поставок ПП отдельным группам (сегментам) потребителей, выбор бизнес-моделей организации взаимодействия с потребителями через наиболее эффективные каналы и инструменты коммуникаций. Сложность реализации программы заключается в высокой конкуренции на рынке ПП, отсутствии либо низком уровне компетенций разработчиков по вопросам продвижения и продаж, низкой результативности рекламной и выставочно-ярмарочной деятельности и т.д. Перечисленные особенности позволяют отнести процесс продвижения ПП к высокорискованному бизнесу, поэтому при разработке программы вопросы идентификации и оценки рисков вывода ПП на рынок, а также принятия одного из возможных сценариев реагирования на риски (принятие, уклонение, передача, снижение) являются ключевыми.

2. Идентификация рисков и рискообразующих факторов

В литературе понятие риска предлагается рассматривать с трех точек зрения [2]: риск как возможность угрозы бизнесу; риск как негативное событие, не позволяющее достичь в полной мере цели проекта; риск как неопределенность между возникающими неблагоприятными ситуациями и возможными действиями по их устранению. Учитывая в явную логическую взаимосвязь между целями и возможными рисками, содержащуюся в этих определениях, можно предположить, что при реализации программы продвижения ПП могут возникнуть три типа (категории) рисков:

- ◆ срыв плановых показателей по объему продаж;
- ◆ срыв плановых сроков реализации программы;
- ◆ превышение бюджета программы.

Появление каждого из рисков возможно при наличии причин (процессов или явлений), способствующих его возникновению и поясняющих, почему наступление риска неизбежно. Такие явления принято называть *рискообразующими факторами* [2]. Для классификации рискообразующих факторов будем использовать иерархический метод классифика-

ции. На первом уровне в качестве основания классификации можно выделить внешние и внутренние факторы. Проявление внешних факторов обуславливается как политикой государства в отношении бизнеса малых ИТ-компаний, так и различными ситуациями на финансовом и продуктовом рынках, а также на рынке трудовых ресурсов. В качестве оснований для классификации внутренних факторов можно использовать элементы процесса разработки и реализации программы: программа как конечный продукт, персонал, технологии реализации продукта, технологии управления продуктом.

Проведенный анализ литературы позволил выявить следующее множество внешних и внутренних факторов риска программы продвижения ПП [2-9].

К внутренним первичным факторам риска программы продвижения относятся следующие.

1. *Продукт*: недостаток финансирования мероприятий по продвижению программы; нестабильное финансирование мероприятий по продвижению программы; нереальные сроки выхода на планируемые объемы продаж; ошибки в расчетах трудоемкости и финансовых затрат на разработку и продвижение; недостатки в планировании при разработке и реализации программы, появление «забытых» работ.

2. *Персонал*: дефицит трудовых ресурсов; высокая текучесть кадров; отсутствие опыта, необходимого для разработки и реализации программы; разрыв в квалификации специалистов разных областей знаний; недостаточная поддержка программы членами команды; саботаж отдельных членов команды; недостатки во внутренней организации работ; неумение работать в реальном времени.

3. *Технологии реализации продукта*: неполные или нечеткие требования к информационным технологиям продвижения; недостаточная зрелость технологий, применяемых при продвижении; высокая скорость устаревания применяемых технологий; ошибки при выборе программно-аппаратной платформы и средств продвижения; недостаточные навыки владения исполнителями информационными технологиями продвижения; ошибки выбора каналов и инструментов коммуникаций; недостаточная проработка коммуникационных сообщений; отсутствие эффективного взаимодействия с потенциальными пользователями.

К внешним первичным факторам риска программы продвижения относятся следующие.

1. *Государство*: изменение нормативно-правовых механизмов ведения бизнеса в ИТ-отрасли; из-

менение нормативного регулирования бизнес-процессов у потенциальных потребителей; отсутствие устоявшейся законодательской практики по защите авторских и имущественных прав ПП; изменение экономической ситуации в государстве, отрасли, регионе.

2. *Финансовый рынок*: колебания курсов валют; изменение ставок по кредитам.

3. *Рынок труда*: отсутствие на рынке необходимых специалистов.

4. *Продуктовый рынок (потребители)*: неполнота и неточность оценки потребностей потенциального рынка; несоответствие функциональных характеристик ПП потребностям потребителей; слабое влияние внедрения ПП на совершенствование бизнес-процессов компаний потребителей; несовместимость предлагаемого продукта с ПП компаний-потребителей; несоответствие общесистемных характеристик ПП, имеющимся у потребителей, программно-аппаратным средствам и коммуникациям; ошибки при выборе потребительских предпочтений пользователей; ошибочный выбор целевого сегмента; ошибочные прогнозы объема продаж; несоответствие рыночной цены ПП возможностям потенциальных потребителей; ухудшение финансовой ситуации компаний, являющихся потенциальными потребителями; невостребованность ПП рынком; скрытое противостояние специалистов-потребителей внедрению ПП; низкий уровень подготовки пользователей у потенциальных потребителей ПП.

5. *Продуктовый рынок (партнеры, конкуренты)*: появление на рынке новых аналогичных продук-

тов; непредсказуемое поведение конкурентов; дискредитация ПП со стороны конкурентов; пиратское распространение копий ПП; ненадежная работа аутсорсинговых компаний; изменение цен на услуги связи; изменение цен на размещение рекламы.

Очевидно, что приведенный перечень рисков образующих факторов не претендует на абсолютную корректность и полноту. Вместе с тем эти сведения будут полезны менеджерам проектов при первичном отборе факторов, влияющих на результаты конкретного проекта.

Содержательное описание каждого фактора проводится по схеме: «условия возникновения → последствия проявления → влияние на результат» (табл. 1).

3. Качественное и количественное описание рисков образующих факторов

Для проведения количественной оценки и анализа влияния фактора на результаты реализации программы продвижения необходимо определить вероятность проявления фактора и уровень негативных последствий от него (ущерба, убытков, потерь). Учитывая высокую степень неопределенности, описание рисков образующих факторов целесообразно проводить в качественных шкалах, а для их количественной интерпретации использовать интервальные оценки (табл. 2). Оценки вероятности и воздействий в виде процентов отклонения от целевых показателей программы продвижения должны быть проведены по каждому из результатов программы.

Таблица 1.

Описание рисков образующих факторов

	Факторы	Условие	Последствия	Воздействие на цели
1	Изменение нормативного регулирования бизнес-процессов у потенциальных потребителей	Принятие законов по регулированию бизнес-процессов в области применения ПП	Необходимость доработок функционала ПП	Увеличение бюджета проекта
2	Появление новых аналогичных продуктов	Выход на рынок новых аналогичных продуктов	Усиление конкуренции	Сокращение объемов продаж
3	Отсутствие у команды проекта необходимого опыта по продвижению ПП	Ошибки при реализации проекта	Необходимость доработок проекта	Срыв сроков разработки. Увеличение бюджета проекта Сокращение объемов продаж
4	Ошибки выбора каналов и инструментов коммуникаций	Снижение необходимого уровня информирования целевой аудитории	Несоответствие плановых и фактических показателей результативности программы продвижения	Сокращение объемов продаж

Таблица 2.

Количественное оценивание рискообразующих факторов

Показатели	Шкалы				
	Маловероятно	Возможно	Вероятно	Очень вероятно	
Вероятность наступления фактора	< 0,3	0,2–0,6	0,4–0,8	> 0,7	
Возможные потери в объемах продаж, увеличение бюджета проекта	Незначительные	Умеренные	Критичные	Катастрофические	
	Потери менее 15 %	Потери от 10 до 40 %	Потери от 35 % и менее 70 %	Потери более 65 %	
Степень воздействия фактора	Низкая	Ниже среднего	Средняя	Умеренно-высокая	Высокая
	0–3	1–5	3–7	5–9	7–10

Пороговые значения вероятности наступления фактора и возможные потери от его проявления определялись с учетом рыночной конкуренции программного продукта и рекомендаций по выбору шкал влияния риска на цели проекта, приведенных в [2; 10].

Отсутствие точных статистических количественных оценок первичных рискообразующих факторов, наличие факторов, оценка которых существует только в качественной, либо в интервальной категории, привлечение для оценки факторов экспертов [11], объективно требуют использование для этих целей математического аппарата теории нечеткой логики [12].

4. Итоговая оценка риска программы продвижения и выбор стратегии реагирования

Рассмотрим нечеткую модель принятия решений по управлению рисками программы продвижения на рынок веб-ориентированной геоинформационной технологии формирования и мониторинга электронного генерального плана инженерной инфраструктуры [13]. В качестве целевых сегментов рынка выбраны предприятия металлургии и нефтехимии, где разработчики уже имели опыт внедрения подобных ПП. Стоимость ПП определяется исходя из рыночного окружения, и не может превышать стоимости аналогичных программных продуктов, представленных на рынке. Согласно маркетинговой цели программы продвижения заданы объемы продаж и бюджет программы. Требуется оценить риски реализации программы продвижения программных продуктов в выделенных сегментах рынка и выбрать одну из возможных стратегий по управлению рисками:

принятие риска, уклонение от риска, передача риска, снижение рисков.

Этап 1. Идентификация рискообразующих факторов.

В качестве входных лингвистических переменных модели экспертами были выбраны и описаны следующие первичные рискообразующие факторы, влияющие на процессы продвижения ПП:

- изменение нормативного регулирования бизнес-процессов у потенциальных потребителей (x_1);
- колебания курса валют (x_2);
- появление на рынке новых аналогичных продуктов (x_3);
- пиратское распространение копий ПП (x_4);
- несоответствие функциональных характеристик ПП потребностям потребителей (x_5);
- недостаточные навыки владения исполнителями информационными технологиями продвижения (x_6);
- ошибочный выбор целевого сегмента (x_7);
- ошибки в расчетах трудоемкости и финансовых затрат на разработку и продвижение (x_8);
- ошибки при выборе потребительских предпочтений (x_9);
- ошибки выбора каналов и инструментов коммуникаций (x_{10});
- недостаточная проработка коммуникационных сообщений (x_{11}).

Фрагмент описания рискообразующих факторов представлен в *табл. 1*.

Этап 2. Качественная и количественная оценка рискообразующих факторов.

В соответствии с *табл. 2*, при задании лингвисти-

Таблица 3.

Оценка первичных рискообразующих факторов

Фактор	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
Вероятность наступления	1	3	1	1	1	3	1	3	2	2	3
Сила воздействия фактора	2	3	3	3	2	3	4	2	1	2	2

Таблица 4.

Матрица нечетких правил

Сила воздействия фактора	Катастрофическая	Средняя	Умеренно-высокая	Умеренно-высокая	Высокая
	Критичная	Ниже среднего	Средняя	Средняя	Высокая
	Умеренная	Ниже среднего	Ниже среднего	Средняя	Умеренно-высокая
	Незначительная	Низкая	Низкая	Ниже среднего	Средняя
Степень влияния фактора на итоговый риск		Маловероятная	Возможная	Вероятная	Очень вероятная
Возможность реализации фактора					

ческих переменных, характеризующих факторы риска, использовались следующие терм-множества:

— для характеристики вероятности наступления: $T1 = \{ \text{Маловероятно (1), Возможно (2), Вероятно (3), Очень вероятно (4)} \}$;

— для характеристики силы воздействия фактора на итоговый риск: $T2 = \{ \text{Незначительная (1), Умеренная (2), Критичная (3), Катастрофическая (4)} \}$.

В табл. 3 представлены экспертные оценки вероятности наступления первичных факторов и силы их воздействия на итоговый риск. Исходными данными для оценки экспертами вероятности наступления рискообразующих факторов и силы их воздействия на цели проекта является описание причинно-следственных связей проявления факторов, представленных в табл. 1.

Этап 3. Итоговая оценка риска программы продвижения и выбор стратегии реагирования.

Для вычисления будем использовать алгоритм нечеткого вывода Мамдани [14].

Для определения степени воздействия факторов на итоговый риск продвижения продукта при задании лингвистической переменной «Степень воздействия фактора на итоговый риск» будем использовать терм-множество $T3 = \{ \text{Низкая (1), Ниже$

$\text{среднего (2), Средняя (3), Умеренно-высокая (4), Высокая (5)} \}$.

В табл. 4 представлена матрица соответствий возможности проявления первичного фактора и силы его воздействия на итоговый риск.

Интерпретация последствий наступления первичных факторов и возможные убытки (снижение объема продаж и увеличение бюджета программы), выполненная на основании табл. 1 и оценок экспертов, представлена в табл. 5¹.

Таблица 5.

Последствия наступления факторов риска

Фактор	Степень влияния фактора риска	Снижение объема продаж, %	Увеличение бюджета, %
x_1	Низкая	Незначительное	Незначительное
	Средняя	Незначительное	Незначительное
	Высокая	Умеренное	Незначительное
...			
x_2	Низкая	Незначительное	Незначительное
	Средняя	Незначительное	Незначительное
	Высокая	Умеренное	Незначительное

¹ Здесь и далее представление неполных данных не снижает наглядности приведенного примера. Авторы статьи по требованию могут представить все расчетные данные.

Выбор способа реагирования на риск

		Увеличение бюджета			
		1	2	3	4
Снижение объема продаж	1	Принятие	Принятие	Снижение	Уклонение
	2	Снижение	Снижение	Уклонение	Уклонение
	3	Снижение	Уклонение	Уклонение	Передача
	4	Уклонение	Уклонение	Передача	Передача

Для описания выходных переменных «Снижение объема продаж» и «Увеличение бюджета программы» будем использовать терм-множество $T4 = \{\text{Незначительное (1), Умеренное (2), Критичное (3), Катастрофическое (4)}\}$.

В табл. 6 отражена предложенная экспертами зависимость способов реагирования на риск от степени влияния риска на объем продаж и бюджет программы.

Для оценки степени влияния рискообразующих факторов на итоговый риск, возможного снижения объема продаж и увеличения бюджета программы на основании данных табл. 4 и 5 сформированы две базы правил нечеткого вывода в виде:

если посылка a_i [и/или посылка a_j] то заключение b_k .

Фрагменты базы правил представлены в табл. 7.

Таблица 7.

Нечеткие правила модели (фрагмент)

Посылка	Заключение
«возможность реализации фактора x_i » <i>маловероятна</i> и «сила воздействия фактора x_i » <i>критичная</i>	степень влияния фактора на итоговый риск – <i>ниже среднего</i>
«возможность реализации фактора x_i » <i>очень вероятно</i> и «сила воздействия фактора x_i » <i>катастрофическая</i>	степень влияния фактора на итоговый риск – <i>высокая</i>
«степень влияния x_1 » <i>низкая</i>	увеличение бюджета программы <i>незначительное</i>
«степень влияния x_3 » <i>высокая</i>	снижение объема продаж <i>критичное</i>

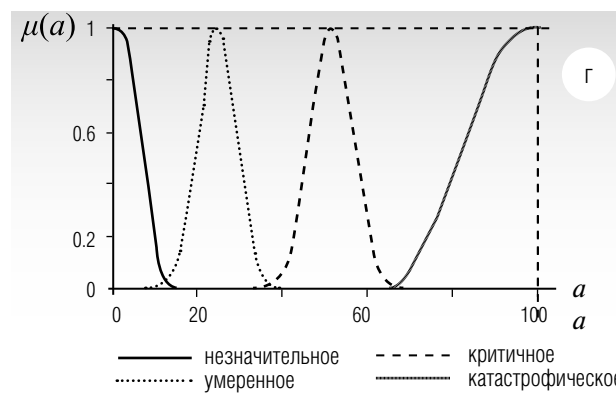
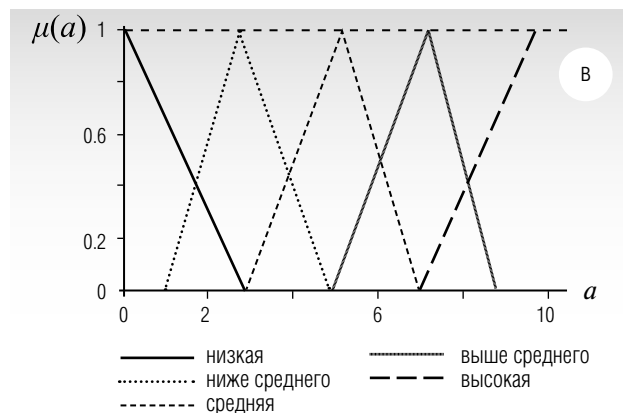
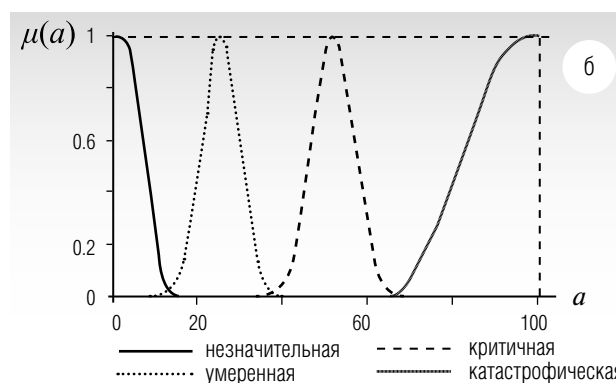
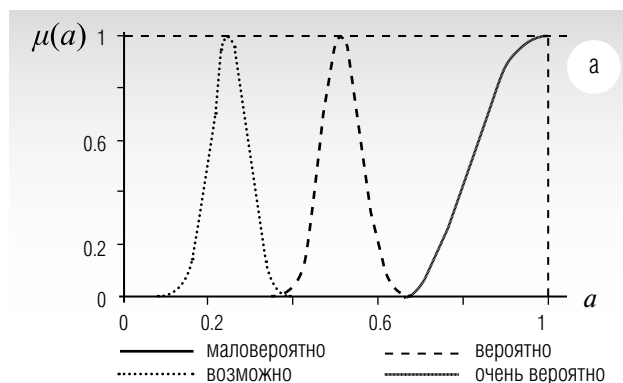


Рис. 1. Функции принадлежности терм-множеств модели: а – T1, б – T2, в – T3, г – T4

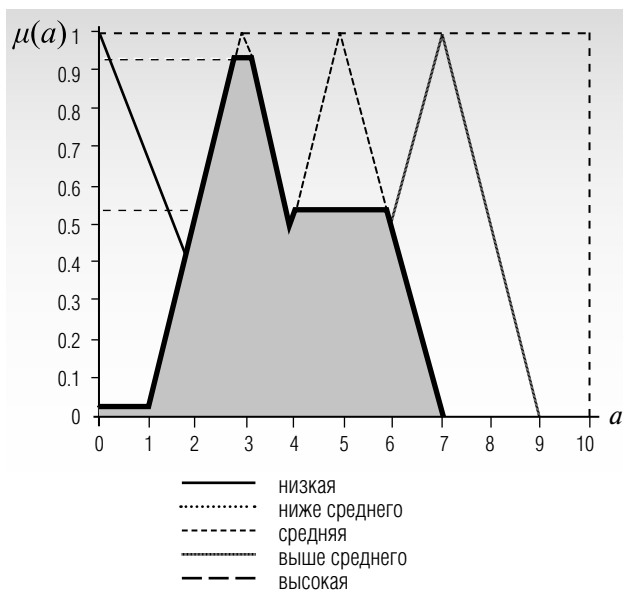


Рис. 2. Функция принадлежности нечеткой переменной «Степень воздействия рискообразующих факторов на итоговый риск»

На этапе *фазификации* были получены функции принадлежности для входных и выходных переменных нечеткой модели. Выбор функций принадлежности проводился экспертами в соответствии с рекомендациями, приведенными в [15]. По виду дискретных рядов, полученных после обработки экспертных данных, аппроксимирующими функциями для описания терм-множеств T_1 , T_2 , T_4 выбраны Z -гармонический сплайн, Гауссова функция, S -гармонический сплайн (рис. 1), для терм-множества T_3 — треугольные функции (рис. 1).

На этапе *агрегации* для всех правил модели (табл. 7) были вычислены истинностные значения предпосылок. В данном примере это минимум из всех возможных значений соответствующих функций принадлежности (*min-конъюнкция*) [12]. Для полу-

чения значения выходной нечеткой переменной в алгоритме использовался минимизирующий логический вывод (*min-активизация*).

На этапе *аккумуляции* по проведенным на предыдущем шаге вычислениям получены: нечеткие множества для выходных переменных — «степень воздействия рискообразующих факторов на итоговый риск» (рис. 2), «снижение объема продаж», «увеличение бюджета программы» (рис. 3).

Выполнение этапа *дефаззификации* предполагает использование метода центра тяжести для получения численных оценок выходных переменных [14]. Для перехода от качественных категорий описания и оценки к количественным будем использовать десятибалльную шкалу.

Результаты расчетов показали следующее:

♦ «Степень воздействия рискообразующих факторов на итоговый риск» программы продвижения близка к средней и равна 3,817. Пренебрегать зависимостью такой силы воздействия не следует и в дальнейшем рекомендуется исследовать степень влияния выделенных факторов на отдельные мероприятия программы.

♦ «Снижение объема продаж» может быть оценено как («умеренное») и равно 0,246. Следует ожидать, что при наступлении риска объем продаж может уменьшиться на 25%.

♦ «Увеличение бюджета программы» может быть оценено как («незначительное») и равно 0,045. Бюджет программы продвижения может потребовать увеличения всего на 5%.

В целом по результатам расчетов, следуя рекомендациям экспертов по управлению рисками (табл. 6), можно порекомендовать руководителям программного проекта принять стратегию снижения риска.

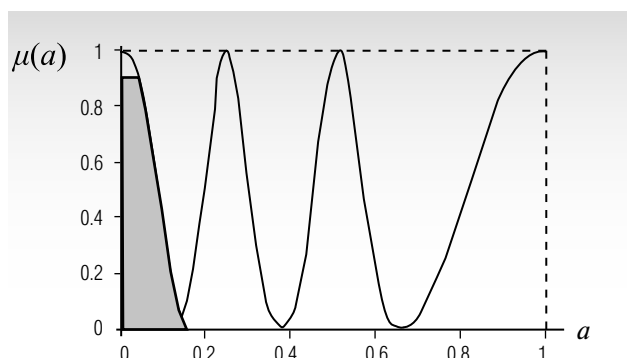
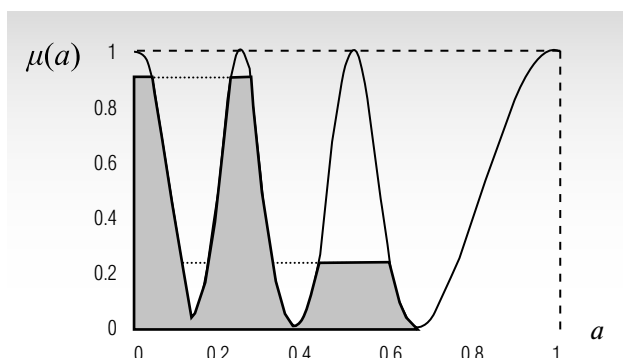


Рис. 3. Функции принадлежности выходных переменных. а — «Снижение объема продаж» б — «Увеличение бюджета»

5. Заключение

Полученные результаты в практическом плане полезны руководителям малых ИТ-компаний, а также маркетологам при решении задач продвижения на промышленные рынки новых товаров. Предложенные в статье множество рискообразующих факторов и модель оценки рисков могут использоваться командой проекта при самостоятельном продвижении программных продуктов, не прибегая к аутсорсингу. Иерархическая структура классификации делает ее открытой. Менеджеры могут добавлять в нее новые элементы, присущие конкретному проекту, при этом элементы могут добавляться на любом уровне иерархии.

Использование математического аппарата нечеткой логики позволяет экспертам работать с переменными, описания которых существуют только в интервальном виде (в качественных категориях) без перехода к средним значениям или к рангам. Построенная модель не чувствительна к количеству входных данных: при увеличении или уменьшении количества рискообразующих факторов соответственно увеличивается или уменьшается количество правил вывода, логика модели при этом не меняется. Это позволяет использовать модель для оценки рисков не только программ продвижения на рынок программных продуктов, но и рисков программ продвижения новых товаров. ■

Литература

1. Ехлаков Ю.П., Бараксанов Д.Н. Основные положения по разработке программы продвижения программных продуктов в сети Интернет // Бизнес-информатика. 2012. №4 (22). С. 26–32.
2. Авдошин С.М., Песоцкая Е.Ю. Информатизация бизнеса. Управление рисками. М.: ДМК Пресс, 2011. 176 с.
3. Архипенков С. Лекции по управлению программными проектами. 2009. [Электронный ресурс]: http://citforum.ru/SE/project/arkhi-penkov_lectures (дата обращения 10.10.2013).
4. Фатрелл Р.Т., Шафер Д.Ф., Шафер Л.И. Управление программными проектами. Достижение оптимального качества при минимуме затрат. М.: Изд. дом «Вильямс», 2004. 1136 с.
5. Липаев В.В. Анализ и сокращение рисков проектов сложных программных средств. М.: СИНТЕГ, 2005. 224 с.
6. Шапкин А.С., Шапкин В.А. Экономические и финансовые риски. М.: Дашков и К, 2008. 543 с.
7. Фатхутдинов Р.А. Инновационный менеджмент: учебник для вузов. 6-е изд. СПб.: Питер, 2008. 448 с.
8. Авдошин С.М., Песоцкая Е.Ю. Информационные технологии для управления финансовыми рисками // Бизнес-информатика. 2011. №1 (15). С. 42-49.
9. Пырлина И.В. Классификация операционных рисков при сервисно-ориентированном подходе к созданию информационной системы // Бизнес-информатика. 2011. №4 (18). С. 54-61.
10. Руководство к своду знаний по управлению проектами (РМВОК). 4-е изд. М.: Project Management Institute, 2010. 496 с.
11. Заде Л. Понятие лингвистической переменной и ее роль в принятии приближенных решений. М.: Мир, 1976. 168 с.
12. Глушенко С.А. Применение системы MATLAB для оценки рисков информационной безопасности организации // Бизнес-информатика. 2013. №4 (26). С. 35-42.
13. Ехлаков Ю.П., Гриценко Ю.Б., Жуковский О.И. Автоматизация технологий формирования и мониторинга электронного генерального плана инженерной инфраструктуры. Доклады ТУСУРа. Томск: В-Спектр, 2013. № 1. С. 100-106.
14. Алгоритм Мамдани в системах нечеткого вывода [Электронный ресурс]: <http://habrahabr.ru/post/113020/> (дата обращения 01.12.2013).
15. Борисов А.Н., Крумберг О.А., Федоров И.П. Принятие решения на основе нечетких моделей: примеры использования. Рига: Знание, 1990. 184 с.

A FUZZY RISK ASSESSMENT MODEL FOR SOFTWARE PROMOTION RISKS

Yuri EKHLAKOV

Professor, Head of Department of Data Processing Automation, Faculty of Control Systems,
Tomsk State University of Control Systems and Radioelectronics

Address: 74, Verшинina str., Tomsk, 634034, Russian Federation

E-mail: upe@tusun.ru

Natalya PERMYAKOVA

Post-graduate student, Department of Data Processing Automation, Faculty of Control Systems,
Tomsk State University of Control Systems and Radioelectronics

Address: 74, Verшинina str., Tomsk, 634034, Russian Federation

E-mail: pnv@muma.tusun.ru

This paper discusses the issue of risk assessment and analysis in development and implementation of a software promotion scheme, and demonstrates feasibility of fuzzy analysis as a mathematical tool for this purpose. It formulates the marketing goal of the scheme as being «to achieve the target sales within the specified period of time under a limited budget». Given the clear logical connection between the goals of the scheme and the associated risks, the paper identifies three types of risks: failure to meet scheme implementation schedule, failure to meet target sales, failure to stay within the budget. Based on analysis of available publications, risk-contributing factors have been identified, a classification of such factors and their qualitative and quantitative characteristics have been offered. A real case study of building of a fuzzy risk assessment and analysis model for market launch of a Web-oriented geo-information technology for an enterprise master plan has been considered. The analysis has identified eleven input linguistic variables (risk-contributing factors) having impact on scheme risks and three output variables (degree of factors' impact on the total risk of the project, budget overrun rate and target sales achievement rate). Two databases of rules have been built: the rules of the first database are used to determine the degree of factors' impact on the total risk. The rules of the second database are applied to determine the risk exposure of the main goals of the scheme. The authors have used Mamdani's fuzzy inference algorithm to calculate values of each of the risks and to offer their risk response scenarios. In practical terms the results are useful for heads of small IT companies and marketing experts in promotion of new products in industrial markets.

Key words: risk classification, risk assessment, software projects risk management, risk-contributing factor, qualitative description of risk, risk quantification, risk categories, fuzzy logic method, Mamdani's algorithm.

References

1. Ekhlakov Yu.P., Baraksanov D.N. (2012) Osnovnye polozheniia po razrabotke programmy prodvizheniia programnykh produktov v seti Internet [The main provisions of the development of software promotion in the Internet]. *Business Informatics*, no. 4 (22), pp. 26-32.
2. Avdoshin S.M., Pesotskaia E.Yu. (2011) *Informatizatsiia biznesa. Upravlenie riskami* [Informatization of business. Risk management.] Moscow: DMK Press. (in Russian)
3. Arkhipenkov S. (2009) *Lektsii po upravleniiu programmnymi proektami* [Lectures on software project management]. Available at: http://citforum.ru/SE/project/arkhi-penkov_lectures (accessed 10.10.2013). (in Russian)
4. Fatrell R.T., Shafer D.F., Shafer L.I. (2004) *Upravlenie programmnymi proektami. Dostizhenie optimal'nogo kachestva pri minimume zatrat* [The management of software projects. The achievement of optimum quality at minimum expenses]. Moscow: Publishing house «Williams». (in Russian)
5. Lipaev V.V. (2005) *Analiz i sokrashchenie riskov proektov slozhnykh programnykh sredstv* [Analysis and reduction of risks of complex software projects]. Moscow: SINTEG. (in Russian)
6. Shapkin A.S., Shapkin V.A. (2008) *Ekonomicheskie i finansovye riski* [Economic and financial risks]. Moscow: Dashkov and Co. (in Russian)
7. Fatkhutdinov R.A. (2008) *Innovatsionnyi menedzhment: uchebnik dlia vuzov* [Innovation management: textbook for universities]. 6-th ed. Saint-Petersburg: Piter. (in Russian)
8. Avdoshin S.M., Pesotskaia E.Yu. (2011) Informatsionnye tekhnologii dlia upravleniia finansovymi riskami [Information technologies for managing financial risks]. *Business Informatics*, no. 1 (15), pp. 42-49.
9. Pyrlina I.V. (2011) Klassifikatsiia operatsionnykh riskov pri servisno-orientirovannom podkhode k sozdaniiu informatsionnoi sistemy [Classification of operational risks under a service-oriented approach to the creation of an information system]. *Business Informatics*, no. 4 (18), pp. 54-61.
10. PMBOK (2010) *Rukovodstvo k svodu znanij po upravleniiu proektami* [Guide to the Project Management Body of Knowledge]. 4-th ed. Moscow: Project Management Institute. (in Russian)

Представляемая для публикации статья должна быть актуальной, обладать новизной, отражать постановку задачи (проблемы), описание основных результатов исследования, выводы, а также соответствовать указанным ниже правилам оформления.

Текст должен быть тщательно вычитан автором, который несет ответственность за научно-теоретический уровень публикуемого материала.

Материалы представляются в электронном виде по адресу:
bijournal@hse.ru.

ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ

ТЕКСТ СТАТЬИ представляется в редакцию в электронном виде (в формате MS Word, версия 2003 или выше).

ОБЪЕМ. Ориентировочный объем статьи составляет 20-25 тысяч знаков (с пробелами).

ШРИФТ, ФОРМАТИРОВАНИЕ, НУМЕРАЦИЯ СТРАНИЦ

ШРИФТ – Times New Roman, кегль набора – 12 пунктов, полуторный интервал, форматирование по ширине. Нумерация страниц – сверху по центру, поля: левое – 2,5 см, верхнее, нижнее и правое – по 1,5 см.

НАЗВАНИЕ СТАТЬИ приводится на русском и английском языках. Название статьи должно быть информативным и раскрывать содержание статьи.

СВЕДЕНИЯ ОБ АВТОРАХ приводятся на русском и английском языках и включают следующие элементы:

- ◆ фамилия, имя, отчество всех авторов полностью
- ◆ должность, звание, ученая степень каждого автора
- ◆ полное название организации – места работы каждого автора в именительном падеже, полный почтовый адрес каждой организации (включая почтовый индекс)
- ◆ адрес электронной почты каждого автора.

АННОТАЦИЯ К СТАТЬЕ представляется на русском и английском языках.

- ◆ Объем – 200-300 слов.
- ◆ Аннотация должна быть информативной (не содержать общих слов).
- ◆ Аннотация должна отражать основное содержание статьи и быть структурированной (следовать логике описания результатов в статье).
- ◆ Структура аннотации: предмет, цель, метод или методологию проведения исследования, результаты исследований, область их применения, выводы.
- ◆ Метод или методологию проведения исследований целесообразно описывать в том случае, если они отличаются новизной или представляют интерес с точки зрения данной работы. В аннотациях статей, описывающих экспериментальные работы, указывают источники данных и характер их обработки.
- ◆ Результаты работы описывают предельно точно и информативно. Приводятся основные теоретические и экспериментальные результаты, фактические данные, обнаруженные взаимосвязи и закономерности. При этом отдается предпочтение новым результатам и дан-

ным долгосрочного значения, важным открытиям, выводам, которые опровергают существующие теории, а также информации, которая, по мнению автора, имеет практическое значение.

- ◆ Выводы могут сопровождаться рекомендациями, оценками, предложениями, гипотезами, описанными в статье.
- ◆ Сведения, содержащиеся в названии статьи, не должны повторяться в тексте аннотации. Следует избегать лишних вводных фраз (например, «автор статьи рассматривает...»).
- ◆ Исторические справки, если они не составляют основное содержание документа, описание ранее опубликованных работ и общеизвестные положения, в аннотации не приводятся.
- ◆ В тексте аннотации следует употреблять синтаксические конструкции, свойственные языку научных и технических документов, избегать сложных грамматических конструкций.
- ◆ В тексте аннотации следует применять значимые слова из текста статьи.

КЛЮЧЕВЫЕ СЛОВА приводятся на русском и английском языках. Количество ключевых слов (словосочетаний) – 6-10. Ключевые слова или словосочетания отделяются друг от друга точкой с запятой.

ФОРМУЛЫ. При наборе формул, как выключных, так и строчных, должен быть использован редактор формул MS Equation. В формульных и символических записях греческие (русские) символы, а также математические функции записываются прямыми шрифтами, а переменные аргументы функций в виде английских (латинских) букв – наклонным курсивом (пример «cos a», «sin b», «min», «max»). Нумерация формул – сквозная (по желанию авторов допускается двойная нумерация формул с указанием структурного номера раздела статьи и, через точку, номера формулы в разделе).

РИСУНКИ (графики, диаграммы и т.п.) могут быть оформлены средствами MS Word или MS Excel. Ссылки на рисунки в тексте обязательны и должны предшествовать позиции размещения рисунка. Допускается использование графического векторного файла в формате wmf/emf или cdr v. 10. Фотографические материалы предоставляются в формате TIF или JPEG, с разрешением изображения не менее 300 точек на дюйм. Нумерация рисунков – сквозная.

ТАБЛИЦЫ оформляются средствами MS Word или MS Excel. Нумерация таблиц – сквозная.

СПИСОК ЛИТЕРАТУРЫ составляется в соответствии с требованиями ГОСТ 7.0.5-2008. Библиографическая ссылка (примеры оформления размещены на сайте журнала <http://bi.hse.ru/>). Нумерация библиографических источников – в порядке цитирования. Ссылки на иностранную литературу – на языке оригинала без сокращений.

СПИСОК ЛИТЕРАТУРЫ ДЛЯ АНГЛОЯЗЫЧНОГО БЛОКА оформляется в соответствии с требованиями SCOPUS (примеры оформления размещены на сайте журнала <http://bi.hse.ru/>). Для транслитерации русскоязычных наименований можно воспользоваться сервисом <http://translit.ru/>.

ЛИЦЕНЗИОННЫЙ ДОГОВОР

Для размещения полнотекстовых версий статей на сайте журнала с авторами заключается лицензионный договор о передаче авторских прав.

Плата с авторов за публикацию рукописей не взимается.

AUTHORS GUIDELINES

Articles should be topical and original, should outline tasks (issues), describe key results of the author's research and appropriate conclusions.

Manuscripts are submitted via e-mail: bijournal@hse.ru.

MANUSCRIPT REQUIREMENTS

TEXT FILES should be submitted in electronic form, as a MS Word document (version 2003 or higher).

LENGTH. Articles should be between 20 and 25 thousand characters (incl. spaces).

FONT, SPACING, MARGINS. The text should be in Times New Roman 12 pt, 1.5 spaced, fit to the width, margins: left – 25 mm, all other – 15 mm.

TITLE of the article should be submitted in native language and English.

AUTHORS' DETAILS are presented in native language and English. The details include:

- ◆ Full name of each author
- ◆ Position, rank, academic degree of each author
- ◆ Affiliation of each author, at the time the research was completed
- ◆ Full postal address of each affiliation (incl. postcode / ZIP)
- ◆ E-mail address of each author.

ABSTRACT are presented in native language and English.

- ◆ The abstract should be between 200 and 300 words.
- ◆ The abstract should be informative (no general words), original, relevant (reflects your paper's key content and research findings); structured (follows the logics of results' presentation in the paper)
 - ◆ The recommended structure: purpose (mandatory), design / methodology / approach (mandatory), findings (mandatory), research limitations / implications (if applicable), practical implications (if applicable), originality / value (mandatory).
- ◆ It is appropriate to describe the research methods/methodology

if they are original or of interest for this particular research. For papers concerned with experimental work the data sources and data procession technique should be described.

- ◆ The results should be described as precisely and informatively as possible. Include your key theoretical and experimental results, factual information, revealed interconnections and patterns. Give special priority in the abstract to new results and long-term impact data, important discoveries and verified findings that contradict previous theories as well as data that you think have practical value.

- ◆ Conclusions may be associated with recommendations, estimates, suggestions, hypotheses described in the paper.

- ◆ Information contained in the title should not be duplicated in the abstract. Authors should try to avoid unnecessary introductory phrases (e.g. «the author of the paper considers...»).

- ◆ Authors should use the language typical of research and technical documents to compile your abstract and avoid complex grammatical constructions.

- ◆ The text of the abstract should include key words of the paper.

KEYWORDS are presented in native language and English. The number of key words / words combinations are from 6 to 10 (separated by semicolons).

FORMULAE should be prepared using Math Type or MS Equation tool.

FIG.S should be of high quality, black and white, legible and numbered consecutively with Arabic numerals. All Fig.s (charts, diagrams, etc.) should be submitted in electronic form (photo images – in TIF, PSD or JPEG formats, minimum resolution 300 dpi). Appropriate references in the text are required.

REFERENCES should be presented in Harvard style and carefully checked for completeness, accuracy and consistency.

The publication is free of charge.