



Мин МИНСК

Часть 1

# BIG DATA

and Advanced Analytics

VI МЕЖДУНАРОДНАЯ  
НАУЧНО-ПРАКТИЧЕСКАЯ  
КОНФЕРЕНЦИЯ

2020

Министерство образования Республики Беларусь  
Учреждение образования «Белорусский государственный университет  
информатики и радиоэлектроники»  
Национальная академия наук Беларуси  
Объединенный институт проблем информатики  
DHTechnologies & Data Nubes (Austin, USA)  
University of Technology (Kielce, Poland)  
ActiveCloud (Москва, Россия)  
BEZNext (Chicago, USA)  
IBM (NY, USA)  
Ташкентский университет информационных технологий  
(Ташкент, Узбекистан)  
ИООО «ЭПАМ Системз»  
ИЧПУП «Годел Текнолоджис Юроп»

## **BIG DATA AND ADVANCED ANALYTICS**

## **BIG DATA И АНАЛИЗ ВЫСОКОГО УРОВНЯ**

СБОРНИК МАТЕРИАЛОВ  
VI МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ

(Республика Беларусь, Минск, 20-21 мая 2020 года)

В трех частях

Часть 1

Минск  
«Бестпринт»  
2020

УДК 004.6(082)

ББК 32.97.3

Б59

Редакционная коллегия:

**В.А. Богуш**, доктор физико-математических наук, профессор, ректор учреждения образования «Белорусский государственный университет информатики и радиоэлектроники» Республики Беларусь;

**Boris Zibitsker**, MS, PhD and Honorable Doctor at BSUIR, President and CEO BEZNext, Professor DePaul University in Chicago, USA;

**С.К. Дик**, кандидат физико-математических наук, доцент, депутат Палаты представителей Национального собрания Республики Беларусь седьмого созыва;

**Д.В. Лихачевский**, кандидат технических наук, доцент, декан факультета компьютерного проектирования учреждения образования «Белорусский государственный университет информатики и радиоэлектроники» Республики Беларусь;

**К.Д. Яшин**, кандидат технических наук, доцент, заведующий кафедрой инженерной психологии и эргономики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники» Республики Беларусь

Рецензенты:

**Boris Zibitsker**, MS, PhD and Honorable Doctor at BSUIR, President and CEO BEZNext, Professor DePaul University in Chicago, USA;

**А.В. Тузиков**, доктор физико-математических наук, профессор, член-корреспондент Национальной академии наук Беларуси, генеральный директор государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларуси» Республики Беларусь;

**Domonique A. Heger**, PhD, President and CEO  
DHTechnologies & Data Nubes, Austin, USA

**Alain Biem**, PhD, Opera Systems, formerly at IBM Watson  
Research – Big Data, NY, USA;

**Н. Успенский**, руководитель направления Big Data & Analytics (Беларусь, Украина, Чехия) в корпорации Oracle;

Б59 **BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня: сб. материалов VI Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 20-21 мая 2020 года): в 3 ч. Ч. 2 / редкол. : В.А. Богуш [и др.]. – Минск : Бестпринт, 2020. – 428 с.**

ISBN 978-985-90533-7-5.

В сборнике опубликованы результаты научных исследований и разработок в области BIG DATA and Advanced Analytics для оптимизации IT-решений и бизнес-решений, а также тематических исследований в области медицины, образования и экологии.

УДК 004.6(082)

ББК 32.97.3

ISBN 978-985-90533-7-5 (ч. 1)

ISBN 978-985-90533-2-0

© УО «Белорусский государственный университет информатики и радиоэлектроники», 2020  
© Оформление. УП «Бестпринт», 2020

## **ОРГАНИЗАЦИОННЫЙ КОМИТЕТ КОНФЕРЕНЦИИ**



### **Председатель, Богущ В.А.**

Ректор Белорусского государственного университета информатики и радиоэлектроники, доктор физико-математических наук, профессор



### **Сопредседатель, Boris Zibitsker**

MS, PhD and Honorable Doctor at BSUIR, President and CEO BEZNext, Adjunct Associate Professor, DePaul University in Chicago



### **Сопредседатель, Dominique A. Heger**

PhD, Founder of DHTechnologies, Data Analytica, Hotshot Analytics and AI/ML firm, Austin, USA



### **Заместитель председателя, Дик С.К.**

Депутат Палаты Представителей Национального собрания Республики Беларусь седьмого созыва, кандидат физико-математических наук, доцент, Республика Беларусь

### **Члены организационного комитета**



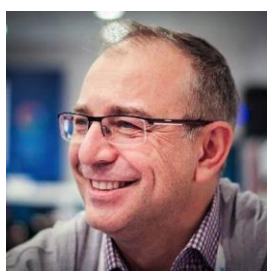
**Appolo Tankeh**, PhD, IBM Treasury, NY, USA



**Alain Biem**, PhD, VP of Analytics at Opera Solutions, formerly at ABM Watson Research, NY, USA



**Dirk Marc Guy Stroo**, PhD, Owner of Invisi, Netherlands, Owner of Act On Insight, Belarus, Information Innovation Leader, Business Intelligence Consultant: Royal Agio Cigars, City of Rotterdam, Nuon



**Leon Katsnelson**, Director and CTO at IBM Analytics Platform Emerging Technologies, USA



**Батура М.П.**, Научный руководитель НИЛ 8.1 БГУИР, доктор технических наук, профессор, академик «Международной академии наук высшей школы», заслуженный работник образования Республики Беларусь, Республика Беларусь



**Лихачевский Д.В.**, декан факультета компьютерного проектирования Белорусского государственного университета информатики и радиоэлектроники, кандидат технических наук, доцент, Республика Беларусь



**Лукашевич М.М.**, декан факультета компьютерных систем и сетей БГУИР, кандидат технических наук, доцент, Республика Беларусь



**Прытков В.А.**, проректор по учебной работе БГУИР, кандидат технических наук, доцент, Республика Беларусь



**Тузиков А. В.**, генеральный директор Объединенного института проблем информатики Национальной академии наук Беларуси, доктор физико-математических наук, профессор, член-корреспондент, Национальной академии наук Беларуси, Республика Беларусь



**Пархименко В.А.**, заведующий кафедрой экономики БГУИР, кандидат экономических наук, доцент, Республика Беларусь



**Яшин К.Д.**, заведующий кафедрой инженерной психологии и эргономики Белорусского государственного университета информатики и радиоэлектроники, кандидат технических наук, доцент, Республика Беларусь

## ОРГАНИЗАТОРЫ КОНФЕРЕНЦИИ

Министерство образования Республики Беларусь



Учреждение образования  
«Белорусский государственный университет  
информатики и радиоэлектроники»



Национальная академия наук Беларуси  
Объединенный институт проблем  
информатики



ООО «АктивХост РУ»



BEZNext (Chicago, USA)



DHTechnologies & Data Nubes (Austin, USA)



«Politechnika Świętokrzyska» Kielce University of  
Technology (Kielce, Poland)



IBM (NY, USA)



Ташкентский университет информационных  
технологий



ИОО «ЭПАМ Системз»

## СПОНСОРЫ КОНФЕРЕНЦИИ





## ОГЛАВЛЕНИЕ

<b>B. Zibitsker, A. Lupersolsky</b> How to apply modeling and optimization to select the appropriate cloud platform .....	11
<b>B. Zibitsker, A. Podelko</b> Performance testing and modeling for new analytic applications .....	19
<b>A. Adamov, A.I. Buranbaeva, S.I.Gindin, M. Glesner, I. Horton, A. Keevallik, A.D. Khomonenko, Y.G. Klyuyeva, I. Skliarova, V. Sklyarov, V.V. Yavorskiy</b> Data processing in high-performance computing systems .....	33
<b>I. A. Trubin</b> Cloud servers usage efficiency and cost optimization.....	53
<b>М.Н. Калимолдаев, В.В. Яворский, М.А. Сонькин, В. Войчик, И.Т. Утепбергенов, А.Т. Ахмедиярова, Д.Т. Касымова, Е.Г. Ключева, Н.В. Байдикова, М.М. Есмагамбетова</b> Формирование хранилища и анализ больших данных передвижений в городе .....	59
<b>D.A. Paulenka, V.A. Kovalev, E.V. Snezhko, V.A. Liauchuk, E.I. Pechkovsky</b> Recognition of underlying surface using a convolutional neural network on a single-board computer .....	71
<b>R.M. Asimov, S.V. Valevich, I. Kruse, V.S. Asipovich</b> Digital twin for pv plant's power generation analysis .....	78
<b>A.V. Kuchynski, U.N. Hutkouski, I.I. Piletski</b> Nosql databases. technology for protecting data from unauthorized access .....	90
<b>I.N. Tsyrelchuk, N. M. Mamatova, M. Y. Abdul-Azalova</b> Optimization of business processes via big data .....	96
<b>M. Vaskouski</b> Applications of second order ornstein unlenbeck stochastic processes to credit risk modeling.....	105
<b>N.V. Guletskii, Ya. M. Radyna</b> Estimating the inefficiency of machinery and its operators using hierarchical bayesian models .....	112
<b>S.K. Kurbanov, S.S. Beknazarova</b> Deep learning approaches to biomedical image segmentation .....	118
<b>V.G. Gorovoy, A.S. Sidorovich</b> Improving the efficiency and use of information resources .....	123
<b>K.M. Karaneuski, M.A. Konovalova, K.N. Meziyanaya, K.D. Yashin</b> Basic principles and analysis of high-level techniques for designing computer programs for virtual world .....	128
<b>M. W. Alhamd, T. M. Salman, M. Melqonyan</b> Reducing of radiation dose for x-ray using contourlet transform and block thresholding technique .....	137
<b>A.K.Nishanov, G.P.Jurayev, N.B.Narziyev</b> Adaptable genetic algorithm for primary processing of medical data based on random search .....	147

<b>S.S. Beknazarova, M.K. Jaumitbayeva</b> Method for preprocessing a distorted mage based on the deconvolution procedure .....	156
<b>G. Polat, A. Aydin, F.F. Altinoglu</b> Interpretation seismic and gravity data of hazar lake and surroundings, eastern anatolia .....	162
<b>I.N. Tsyrelchuk, D.S. Yakhshibaev, Kh.R.Jiyanbekov, P.I. Soo, Yu. V. Pisetsky</b> Perspectives for the using of big data to ensure the security of critical information systems .....	165
<b>C.S. Dzik, I.I. Pilecki</b> Anomaly detection using autoencoder for data quality monitoring in cloud.....	170
<b>L.P. Varlamova, A.G. Davydovsky, K.D. Yashin</b> Processing images of technological objects.....	176
<b>Z. A. Alqaisi, M. W. Alhamd, N. A. Abbood</b> The radiation doses diagnostic x-rays and quality assurance tests for devices.....	183
<b>S.N. Ibodullayev, A.A. Bahromov</b> A variety of virtual reality implementations forcreative learning and 5 ways to implement virtual reality in the learning process.....	195
<b>O.N. Karasik, A.A. Prihozhy</b> Cooperative multi-thread scheduler for solving large-size tasks on multi-core systems .....	202
<b>F.F. Saidov, D.R. Mamasoatov</b> Use of information and communication technologiesfor the export of economic subjects of samarkand region.....	213
<b>М.П. Батура, И.И. Пилецкий, В.А. Прытков, Н.А. Волорова</b> Интеллектуальная система комплексного анализа данных интернет- источников. ....	220
<b>Д.И. Черемисинов, Л.Д. Черемисинова</b> Глубокий анализ и декомпиляция кмп-схемы на транзисторном уровне в иерархическую схему из логических элементов .....	242
<b>Л.Ю. Шилин, А.А. Навроцкий, Р. В. Козарь</b> Модификация процессов распознавания медицинских изображений в задачах компьютерной диагностики .....	252
<b>А.Г. Зенкевич, Т.В. Казак</b> Глубокий анализ системы подбора кадров в учреждениях высшего образования.....	257
<b>Ц. Ма, С.Х. Жэнь, В.Ю. Цветков, В.К. Конопелько</b> Математическая модель и алгоритм скелетизации изображений на основе орта и zhang-suen.....	265
<b>А.С. Прокопеня, И.С. Азаров</b> Сверточные нейронные сети для распознавания изображений.....	271
<b>А.С. Чочиева, И. И. Пилецкий</b> Выбор алгоритмов кластеризации .....	281
<b>А.А. Каракулько, О.С. Медведев, П.А. Мороз, О.Н. Дудич, В.Л. Красильникова, В.С. Осипович, К.Д. Яшин</b> Использование нейронных сетей в обработке медицинских изображений.....	294

<b>Пархименко В.А., Савчик О.А., Верняховская В.В., Лапко М.В., Князева Л.П., Дершень А.Г., Дершень В.В.</b> Цифровая трансформация в маркетинговой деятельности: от автоматизации к алгоритмическому маркетингу.....	303
<b>В.В. Шаталова, Т.В. Казак, Лихачевский Д.В.</b> Анализ и особенности технологии bigdata в управлении человеческими ресурсами .....	319
<b>М.П. Ревотюк, Н.В. Хаджинова</b> Ускорение встречного поиска кратчайших путей на больших динамических графах методом бутстрэппинга .....	324
<b>Р.А. Лащётко, М.М. Меженная, Т.В. Калилец, С.К. Дик, И.В. Кишкевич, Е.Н. Рункевич</b> Компьютерное моделирование поверхностных биотканей человека с микроциркуляторным руслом под воздействием лазерного излучения.....	332
<b>А.Д.Тусупов, Н.И.Листопад, А.Т. Тохметов</b> Гигабитные пассивные оптические сети с расширенной зоной досягаемости.....	340
<b>А.Н. Никитин</b> Динамический вычислительный граф для обобщения разрозненных научных данных в универсальную модель на примере поведения цезия в системе «почва-растение» .....	351
<b>Д.С. Сенькович, А.В. Жвакина</b> Нейросетевая система поддержки принятия банковских решений при выдаче кредитов.....	358
<b>А.И. Слаута, О.Б. Цехан</b> Моделирование динамики макроэкономических показателей для прогнозирования действий заемщиков коммерческого банка.....	367
<b>В.С. Маркевич, Л.Н. Марченко</b> Фреймворки языка программирования python для распознавания объектов в режиме фотографии.....	375
<b>С.М. Боровиков, С.С. Дик, В.Т. Лэ, К.И. Клинов</b> Анализ и оценка надёжности прикладных компьютерных программ.....	382
<b>Е.А. Пац, Е.В. Насуро</b> Система распознавания объектов в видеопотоке на основе технологий виртуализации и облачных вычислений.....	391
<b>А.В. Суша, М.И. Вашкевич</b> Технологии глубокой нейронной сети многомасштабного детектирования лиц .....	399
<b>В.В. Старовойтов</b> Проблемы статистических оценок при анализе несбалансированных классов больших данных .....	409

UDK 004.75

## HOW TO APPLY MODELING AND OPTIMIZATION TO SELECT THE APPROPRIATE CLOUD PLATFORM



**B. Zibitsker†**  
BEZNext



**A. Lupersolsky**  
BEZNext

BEZNext, Chicago, IL

E-mail: bzibitsker@beznex.com, alupersolsky@beznex.com

**Abstract.** Organizations want to take advantage of the flexibility and scalability of Cloud platforms. By migrating to the Cloud they hope to develop and implement new applications faster with lower cost. Amazon AWS, Microsoft Azure, Google, IBM, Oracle and other Cloud providers support different DBMS like Snowflake, Redshift, Teradata Vantage, etc. These platforms' different architecture, mechanism of allocation and management of resources, and sophistication of DBMS optimizers affect performance, scalability and cost. As a result, the response time, CPU service time and the number of I/Os for the same query accessing the similar table in the Cloud could be significantly different than On Prem.

In order to select the appropriate Cloud platform, we use modeling and optimization. First, we perform a Workload Characterization for On Prem Data Warehouse. Each Data Warehouse workload represents a specific line of business and includes activity of many users generating concurrently simple and complex queries accessing data from different tables. Each workload has different demand for resources and different response time and throughput Service Level Goals (SLG).

In this paper we will review results of the workload characterization for On Prem Data Warehouse environment.

Secondly, we collect measurement data for standard TPC-DS benchmark tests performed in AWS Vantage, Redshift and Snowflake Cloud platforms for different sizes of the data sets and different number of concurrent users.

During third step we use the results of the workload characterization and measurement data collected during the benchmark to modify BEZNext On Prem closed queueing network model to model individual Clouds.

And finally, during the fourth step we use the model to consider differences in concurrency, priorities and resource allocation to different workloads. BEZNext Capacity Planning optimization algorithms incorporate gradient search mechanism to find the AWS instance type and minimum number of instances which will be required to meet SLGs for each of the workloads. Publicly available information about the cost of the different AWS instances, storage and DBMS software is used to predict the cost of supporting workloads in the Cloud month by month during next 12 months.

**Keywords:** Cloud Platform, Service Level Goals, Workload Characterization, Workload Forecasting, Seasonality Determination, Benchmarking, Modeling, Optimization.

### **Introduction.**

Organizations planning to move On Prem workloads to Cloud are looking for Cloud platform that will be able to continuously satisfy SLGs of the individual workloads (response time and throughput) with the lowest cost. Perpetual change in demand for resources caused by growth in the number of users, volume of data and implementation of new applications increase the contention for resources and cause unstable performance. Cloud elasticity addresses these problems but requires proactive actions to control performance and cost.

Every Cloud platform has a different architecture, elasticity implementation, workload management and DBMS options affecting the workloads' scalability and performance. For example, Teradata Vantage architecture [11] shown on Figure 1 uses sophisticated optimizer and DBMS management options. Database and query tuning can provide significant benefits. Change of Vantage's rules setting workloads' priorities and concurrency is used to control resource allocation

and management. On the other hand, Vantage autoscaling is limited and it is not as simple to use as in Snowflake.



Figure 1. – Teradata Vantage Architecture

Snowflake architecture [9,10] shown on Figure 2 has limited use of Materialized Views for complex queries, but it provides Virtual Warehouses to isolate workloads and reduce contention for resources. Snowflake automatically scales out by adding a cluster of EC2 instances when the number of queries to be processed exceeds a predefined level (8-32). It also scales up by changing its Virtual Warehouse size from X-Small to 4X-Large.

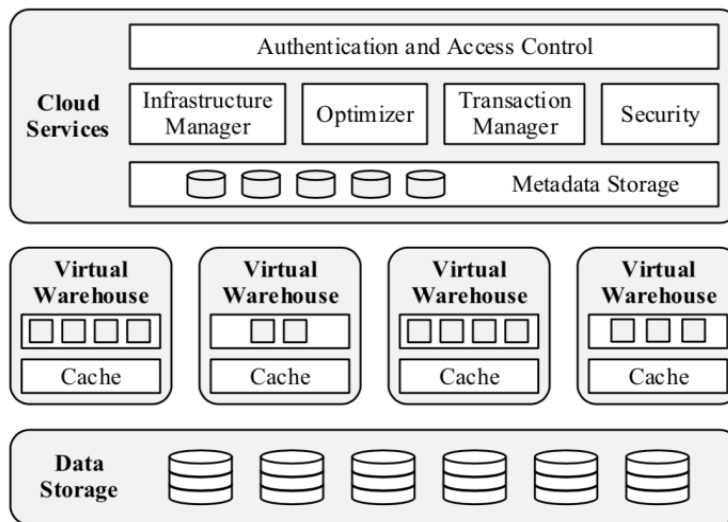


Figure 2. – Overview of Snowflake Architecture

AWS Redshift [5,6] architecture shown on Figure 3 allows customers to select from variety of instance types but has limited concurrency management options, and there is a contention for resources between all workloads.

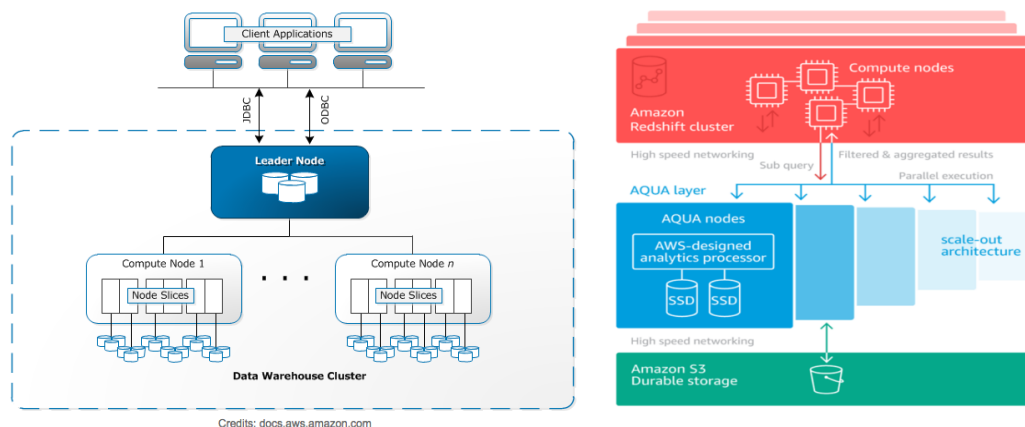


Figure 3. – Redshift Architecture

New release of Redshift based on ra3 instance types incorporates AQUA layer that accelerates queries execution by running data intensive tasks such as filtering and aggregation, compression and encryption closer to the storage layer.

Preparation to migration to the Cloud requires significant efforts to modify applications/SQL and database design.

Another challenge is to estimate how many instances will be required to meet SLGs (for example, if ETL workloads will be able to finish data load on time).

Many organizations run Proof of Concepts projects and execute representative queries against a subset of data because it is too costly to move all production data and execute production workloads in different Clouds. Typically, it is difficult to extrapolate results of the benchmark tests and compare scalability, performance and cost of configurations which will be required to host constantly changing and growing On Prem production workloads. Each workload includes a mix of DML queries generated by different numbers of concurrent users accessing databases with the various level of parallelism depending on their size and structure.

The “trial and error” approach during selection of the Cloud is too expensive. The Cloud cost over time could be very different from what was expected. In this paper we will review a methodology and use case illustrating how BEZNext modeling and performance optimization software is used to determine the minimum configuration required in each Cloud to meet SLGs for each of On Prem Data Warehouse workloads every shift of the day during next 12 months and to predict the corresponding cost.

**Methodology.** The presented methodology is based on use of modeling and optimization of On Prem environment and each Cloud option.

Measurement data continuously collected On Prem in the production environment are aggregated into business workloads. Workload Characterization is used to generate Performance, Resource Utilization and Data Usage profiles for each production workload, and to determine the seasonality of service demands of each workload.

Measurement data collected during standard TPC-DS benchmark tests against different data sets sizes with different numbers of concurrent users are used to determine how different Cloud Architectures change the CPU service time, and KB per query. We also analyze the differences in Cloud scalability during processing of the benchmark queries.

Results of the benchmark tests are used then to modify CPU Service time and #I/O operations for each business workload in BEZNext Queueing Network Models to model their performance in different Clouds. BEZNext optimization algorithms are used to find the minimum configuration required to meet SLGs for each workload on each of the optional Cloud platforms. We predict the impact of the workload and volume of data growth and periodical changes of workload parameters on response time and throughput for each workload On Prem and on each Cloud platform. Instance

types available for each Cloud and results of the workload forecasting are used as input to the models and optimization algorithms.

One of the unique solutions presented in this paper is how optimization algorithms are applied and used to determine the minimum configuration which will be required to meet SLGs for each workload for every shift of the day during next 12 months. Predicting the minimum configuration makes it possible to estimate the cost of running production Data Warehouse workloads on different Cloud platforms and justify the final Cloud selection.

#### *Use case*

The organization is evaluating moving Teradata Data Warehouse workloads to Vantage, Redshift or Snowflake Cloud platforms.

#### *Current Hardware Configuration On Prem*

Data Warehouse is based on Teradata server with the following configuration:

Massively Parallel Processing Architecture (MPP) with 49 processing nodes:

6800 nodes, 56 CPUs per node, 512GB memory

DiskArray1: with RAID1, 90, SSD, disk capacity 1,600GB, transfer rate 450 MBps, average seek time 0.1 ms, 1017 disks.

DiskArray2: RAID1, HDD, 600GB disks with transfer rate 130 MBps, average seek time 3.5 ms, rotational speed 10,000 rpm, 3159 disks

OS SLES 11, DBMS Teradata

Interconnect Bandwidth 11200 MBps

#### *Hardware Configurations used for Benchmark Tests*

Teradata: 6700 56 CPU 512GB RAM 48 SSD 84 HDD

Vantage: AWS m4.4xlarge 16 vCPU 64GB RAM 25 SSD

Redshift: AWS dc2.large 2 vCPU 16GB RAM 1 SSD

Snowflake: AWS c5d.2xlarge 8 vCPU 16GB RAM 1 SSD

#### *Data Collection*

During this project we performed TPC-DS benchmark tests and collected performance measurement data On Prem and on each of the Cloud platforms as it shown on Figure 4 including the average values of CPU Service Time and KB I/O per query, elapsed time and execution time grouped by dataset size (1GB, 10GB and 100GB) and by the number of concurrent users (10, 20, 30, 40 and 50).

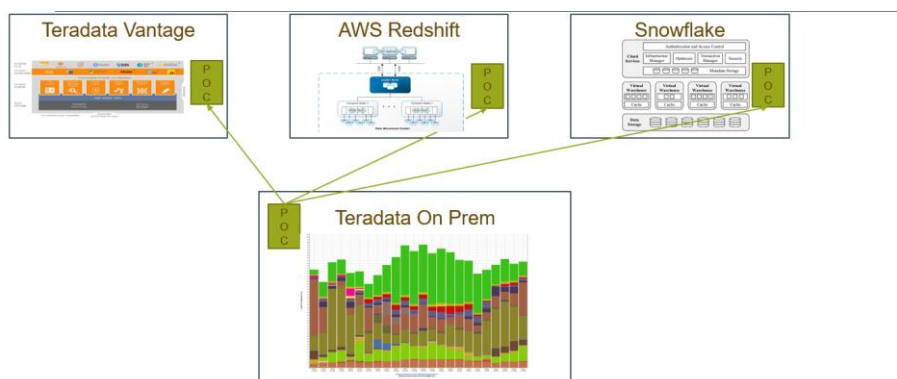


Figure 4. – Data Collection and Workload Characterization On Prem and each of the Cloud platform

We used BEZNext data collection agents to collect measurement data On Prem from ResUsage, DBQL, TDWM tables every hour.

On each Cloud our agents extracted measurement data stored in system tables. A sample of measurement data collected during TPC-DS Benchmarks is shown in Table 1.

Table 1. – Measurement data collected during the benchmark tests On Prem and on each Cloud platform includes the average values of CPU Service Time and KB I/O per Query Type, Elapsed Time and Execution Time grouped by Data set size by the number (#) of concurrent users

Data size # sessions Query type	Teradata Lab				Redshift				Vertica				Snowflake					
	Elapsed Time	Execution Time	CPU seconds	ID KB Phys	Elapsed Time	Execution Time	CPU seconds	ID KB	Elapsed Time	Execution Time	CPU seconds	ID KB Phys	Elapsed Time	Execution Time	CPU seconds	ID KB		
30 TPCDS_Query1	0.478	0.473	3.315	5012607.510	457.340	1.018	1.021	0.178	40183.487	1.234	1.209	0.410	710841.617	1099.587	283.13	275.96	17.87	6.243.576.175.00
TPCDS_Query2	0.795	0.789	4.134	3479687.813	456.267	0.745	0.740	0.071	14713.020	2.080	1.834	5.499	880930.295	2962.817	341.96	337.03	13.89	3.879.023.625.00
TPCDS_Query3	0.843	0.827	2.900	1261046.709	241.172	2.440	2.441	0.018	59542.194	2.009	1.390	4.395	840661.424	1513.106	41.65	43.20	4.89	1.025.992.625.00
TPCDS_Query4	0.269	0.265	0.654	501240.410	97.321	0.239	0.236	0.004	1877.031	0.403	0.409	0.011	45860.325	123.759	76.57	12.92	8.21	104.960.917.50
TPCDS_Query5	0.537	0.494	2.745	3207007.000	49.894	0.585	0.580	0.000	794.031	2.055	1.156	4.383	848102.973	88.791	4.89	4.16	0.74	10.045.281.23
TPCDS_Query6	0.265	0.257	0.473	403531.893	350.683	0.532	0.537	0.012	6721.260	0.689	0.561	0.646	443483.184	549.388	187.93	183.99	28.02	2.285.156.230.00
TPCDS_Query7	0.492	0.397	1.807	984213.107	131.941	0.481	0.477	0.001	31935.497	1.093	0.962	3.210	498945.424	1573.420	1.49	1.22	0.41	327.551.16
TPCDS_Query8	0.469	0.464	4.062	807113.714	451.671	0.284	0.279	0.013	6895.320	1.269	1.100	4.544	544464.528	818.157	157.96	154.86	122.80	2.314.634.125.00
TPCDS_Query9	0.387	0.381	2.565	811010.903	81.340	0.501	0.497	0.003	11842.404	1.238	1.081	3.267	494604.856	1778.734	21.12	17.50	1.06	386.625.000.00
TPCDS_Query10	0.389	0.378	3.329	910508.425	46.997	0.471	0.466	0.006	15371.220	1.283	1.138	3.124	649564.585	121.647	251.63	247.49	50.74	2.215.796.875.00
TPCDS_Query11	1.291	1.275	42.516	6680574.562	4227.277	2.420	2.426	0.100	96302.206	1.426	7.804	30.387	6286417.862	480769.134	25.00	9.71	2.94	1.967.671.675.00
TPCDS_Query12	0.649	0.643	44.864	4681208.876	5138.589	0.529	0.525	0.006	4254.269	35.983	35.735	41.945	4206224.717	54015.604	11.74	58.54	0.91	308.184.062.50
TPCDS_Query13	0.243	0.240	1.239	104418.199	41.393	0.284	0.280	0.000	10242.821	0.929	0.811	2.705	412860.952	39.381	13.45	18.18	1.48	385.742.387.50
TPCDS_Query14	0.232	0.228	1.184	603626.393	34.443	0.238	0.234	0.015	14304.421	0.929	0.819	2.714	402161.226	80.481	17.50	13.93	7.95	586.156.230.00
TPCDS_Query15	0.512	0.506	1.432	861973.598	90.396	0.281	0.276	0.007	5039.199	1.114	0.940	1.712	640734.303	371.040				
TPCDS_Query16	0.226	0.220	2.410	3311051.042	502.320	1.517	1.520	0.008	41057.728	2.346	2.346	4.605	719118.887	1706.467				
TPCDS_Query17	1.964	1.961	4.840	1227480.713	349.838	1.453	1.450	0.000	15204.892	1.834	1.528	5.718	884046.550	2879.712				
TPCDS_Query18	1.675	1.568	3.030	1218177.292	100.361	1.437	1.570	0.122	60480.254	1.588	2.930	4.441	818369.688	2565.447				
TPCDS_Query19	0.817	0.813	0.662	520808.817	52.440	0.802	0.797	0.003	1184.729	1.405	1.404	0.942	410613.825	190.049				
TPCDS_Query20	1.340	1.321	2.788	5382402.305	81.894	0.763	0.756	0.000	557.624	1.329	2.352	4.405	602766.640	69.676				
TPCDS_Query21	0.873	0.871	0.525	648448.892	44.915	0.911	0.911	0.011	5836.131	1.565	1.426	0.682	442941.425	127.486				
TPCDS_Query22	1.035	1.032	1.874	919198.822	75.794	1.563	1.125	0.095	37480.954	2.060	1.924	3.247	501150.791	1590.838				
TPCDS_Query23	1.142	1.147	4.122	608015.486	46.933	0.809	0.809	0.011	6102.148	1.305	2.136	4.574	546293.743	1348.474				
TPCDS_Query24	0.982	0.994	2.476	817348.297	46.767	1.111	0.948	0.048	12704.005	2.240	2.072	3.114	504730.322	1674.844				
TPCDS_Query25	1.081	1.083	2.430	932405.208	73.825	1.101	0.813	0.044	14211.663	2.448	2.283	3.430	812429.734	111.612				
TPCDS_Query26	1.769	1.763	44.053	6946801.563	4160.794	1.767	1.112	0.584	30376.799	11.772	13.376	10.772	6286217.822	653063.297				
TPCDS_Query27	13.858	14.054	12.050	4711386.049	5495.268	0.955	0.408	0.006	3938.968	25.045	24.869	44.612	4210913.209	17569.288				
TPCDS_Query28	0.896	0.888	1.284	340712.032	46.385	0.807	0.370	0.019	9191.490	1.752	1.612	2.727	413103.282	88.149				
TPCDS_Query29	0.896	0.768	1.212	609789.886	46.820	0.717	0.417	0.013	8922.110	1.711	1.586	2.739	410781.863	81.796				
TPCDS_Query30	1.293	1.289	1.559	895203.099	49.739	0.841	0.878	0.006	4794.224	1.181	1.909	1.704	642881.720	235.579				

### Workload Characterization On Prem

On Prem measurement data are automatically aggregated into business workloads.

Results of the workload characterization are used for performance analysis, Anomaly and Root cause detection and seasonality determination.

An example of CPU consumption by different workloads for 24 hours is shown on Figure 5. Each shift has a different pattern of resource usage that will be used in our use case to decide when to allocate and deallocate resources to take advantage of Clouds' elasticity.

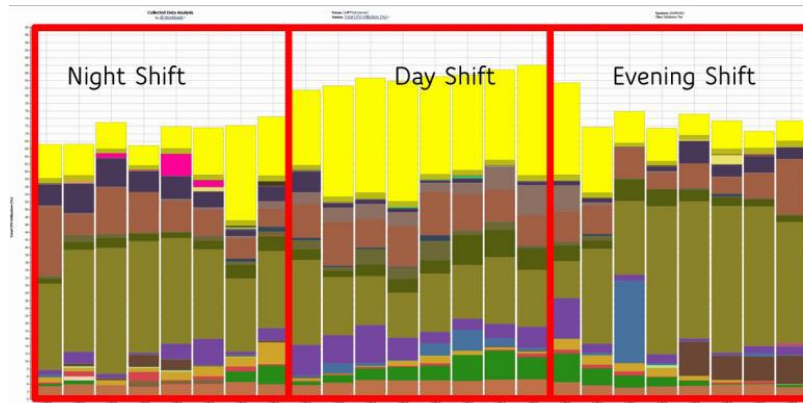


Figure 5. – Example of hourly CPU utilization for On Prem Data Warehouse workloads

Results of the workload characterization are also used to determine the Seasonal Peaks for each workload. Each Seasonal Peak is characterized by start time, duration and amplitude of peak. This information is used to ensure that sufficient resources are proactively allocated exactly when needed and deallocated after then. Automatic determination of seasonal peaks (Figure 6) is used to proactively modify rules of resource allocation and deallocation to continuously meet SLGs and lower the cost [2].

Figure 6 below shows an example of the seasonal peaks for each workload On Prem, where three workloads have peaks in resource utilization daily.



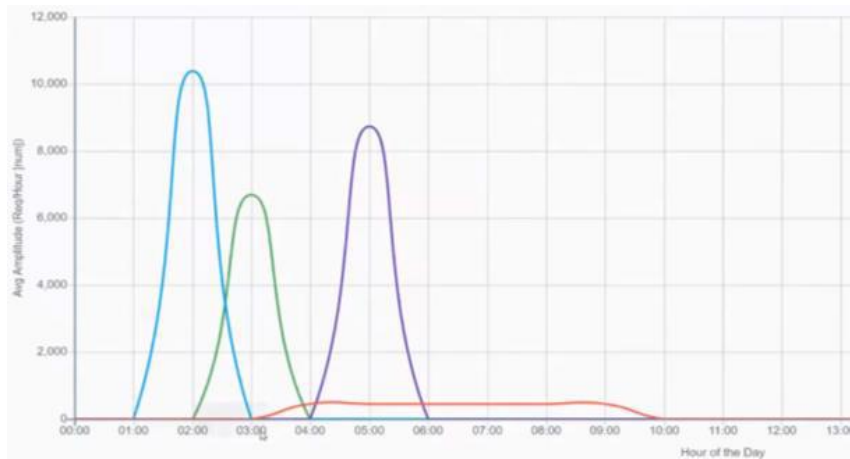


Figure 6. – Example of determining Daily Peaks for Production Workloads

*Workload Characterization Change for the Clouds*

The main goal of running benchmarks on each Cloud is to compare the difference in CPU Service time and KB per query for each of the business workloads On Prem and in each of the Clouds.

For each workload we compared the TPC-DS CPU service time On Prem with CPU Service time consumed by TPC-DS queries on each of the Cloud platform and calculated the average CPU time ratio. The same for KB per query. We applied the ratios to modify service time of production workloads in the queueing network model of each Cloud.

**Modeling.** The goal of modeling is to predict the impact of the expected workload and volume of data growth and impact of new application implementation on performance of each workload for On Prem and for each Cloud platform. We apply Queueing Network models, ML and AI algorithms (Figure 7) to determine the minimum configuration required to meet SLGs for each workload for every hour of the Day during the next 12 months and calculate cost.

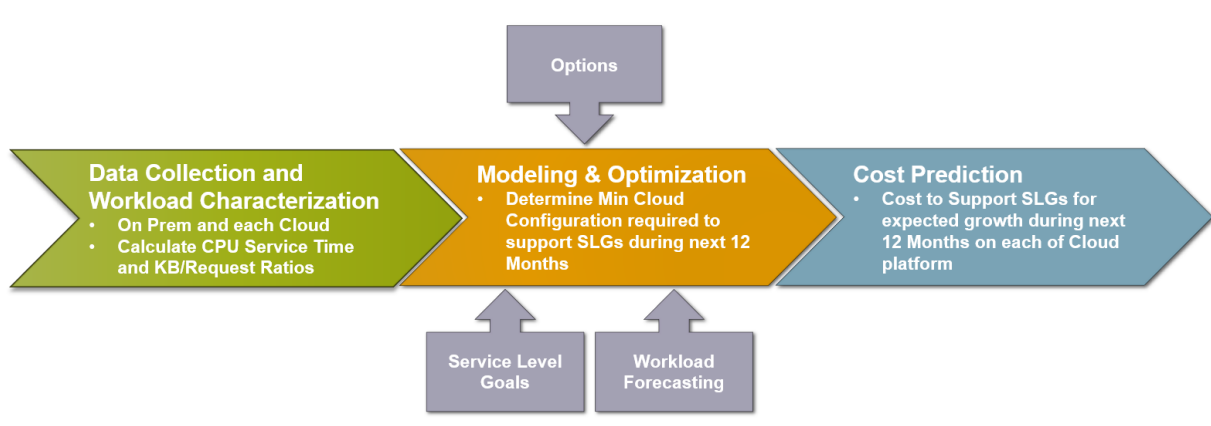


Figure 7. – Modeling and Optimization is used to determine the minimum configuration required to meet SLGs for each workload for every hour of the Day during the next 12 months and calculate cost.

BEZNext incorporates Closed Queueing Network Models modified to be able to accurately predict not only the impact of the workload and volume of data growth, impact of changing hardware and software configuration in multi-tier, distributed, parallel processing environment, but also to predict the impact of changing Priorities, Level of concurrency and resource allocation on performance of each workload and many other “what if” questions [3]. A BEZNext Capacity Planning

Adviser based on Gradient Search optimization technology is used to find the minimum number of instances and instance types which will be required to meet SLGs for each workload every hour of the day during the next 12 months for each of the possible Cloud platforms (Figure 8).

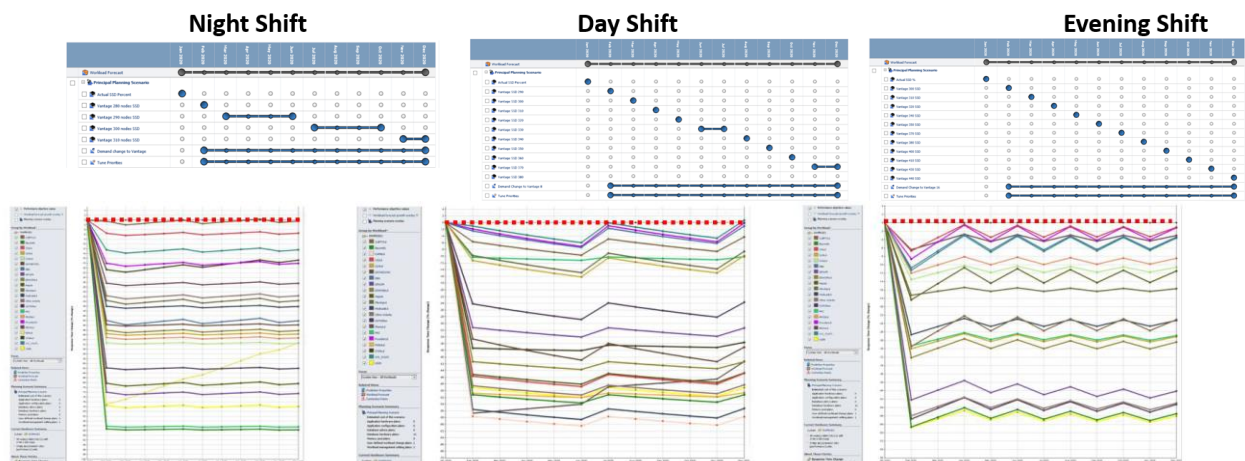


Figure 8. – Example of predicting the minimum configuration required for Vantage to provide Response Time better than current level On Prem during next 12 months for night, day and evening shifts

Minimum configuration determined during modeling and publicly available information about cost of instances is used to predict monthly and yearly cost of running On Prem workloads on each of the optional Cloud platform.

Table 2. – Pricing information for different Instance types

Cloud Name	Instance Type	Cost per Instance / Hour	Software Advanced	Storage
Vantage	m4.16XL	\$3.20	\$18.86	\$0.71 per 5TB per hour
Redshift	dc2.8XL	\$4.80	NA	2.6TB/instance included
Redshift	ra3.16XL	\$13.04	NA	\$0.024 GB/month
Snowflake	2XL with 32 c5d.2Xlarge	\$192/hour	NA	\$23 per TB / Month
Snowflake	3XL with 64 c5d.2Xlarge	\$384/hour	NA	\$23 per TB / Month
Snowflake	4XL with 128 c5d.2Xlarge	\$768/hour	NA	\$23 per TB / Month

**Conclusion.** The methodology and use case illustrate how BEZNext Performance Assurance modeling and optimization software is used to evaluate AWS Redshift, Vantage and Snowflake platforms to select the best Cloud platform for Data Warehouse workloads.

We used measurement data collected during TPC-DS benchmark against different data set sizes with different number of concurrent users to estimate changes of CPU service time and KB per query for each workload in each Cloud comparing with the On Prem Data Warehouse environment.

We illustrated results of applying ML and AI algorithms for workload characterization and seasonality determination.

Queueing Network Models were used to predict the impact of the expected workload and volume of data growth on performance and resource utilization for each workload. Optimization technology was used to predict the minimum configuration of each Cloud which will be required to meet SLGs. These results were used to compare cost and select the Cloud platform with minimum cost.

### ***References***

- [1] B. Zibitsker, «How to select appropriate IT infrastructure to support digital transformation», CMG Impact 2020.
- [2] B. Zibitsker, «Performance Assurance for Big Data and Cloud Computing», Big Data Advanced Analytics, Minsk, 2019.
- [3] J. Buzen, B. Zibitsker, «Challenges of Performance Prediction in Multi-tier Parallel Processing Environments», CMG 2006.
- [4] B. Zibitsker, A. Lupersolsky, “How to Apply Modeling to Compare Options and Select the Appropriate Cloud Platform”, ICPE WOSP-C 2020, Edmonton, Alberta.
- [5] <https://hevodata.com/blog/redshift-architecture/>.
- [6] <https://hevodata.com/blog/amazon-redshift-pros-and-cons/>.
- [7] <https://aws.amazon.com/redshift/>.
- [8] <https://hevodata.com/blog/snowflake-architecture-cloud-data-warehouse/>
- [9] <https://docs.snowflake.net/manuals/user-guide/intro-key-concepts.html>.
- [10] <https://hevodata.com/blog/snowflake-data-warehouse/>.
- [11] <https://www.teradata.com/Products/Software/Vantage>.

UDK 004.75

## PERFORMANCE TESTING AND MODELING FOR NEW ANALYTIC APPLICATIONS



**B.Zibitsker**  
*PHD, CEO BEZNext*



**A.Podelko**  
*PHD, Consulting of Technical Staff at Oracle*

**Abstract.** Traditional load testing (optimized for the waterfall software development process) was mostly focused on pre-production realistic tests. Drastic changes in the industry in recent years - agile development and cloud computing probably the most - opened new opportunities for performance testing. Instead of the single way of doing performance testing, we now have a full spectrum of different tests which can be done at different moments - so deciding what and when to test has become a very non-trivial task heavily depending on the context.

Due to increased sophistication and scale of systems, in most situations full-scale realistic performance testing is not viable anymore. In many cases we may have different partial performance test results. So, results analysis and interpretation have become more challenging - and may require modeling to make meaningful conclusions about performance of the whole system.

Performance testing provides response times and resource utilization for specific workloads. Together with knowledge about architecture and environments, it allows creation of a model to predict a system's performance (to be verified by larger-scale performance tests if necessary). This is a proactive approach to mitigating performance risks - but it requires significant skills and investments to be implemented properly. So, for existing systems it is often complemented (or even completely replaced) by reactive approaches of observing the production system. However, this does not work for new systems. If you are creating a new system, proactive methods such as early performance testing and modeling are needed to make sure that the system will perform as expected.

Modeling becomes important during the design stage because we need to investigate performance and cost consequences of different design decisions. In this case, production data are not available and waiting until the system is fully developed and deployed is too risky for any non-trivial system.

Big data systems are one of the best examples of performance risk mitigated by a combination of performance testing and modeling. The enormous size of the system makes creating a full-scale prototype almost impossible. However associated performance risks are very high - implementing a wrong design may be not fixable and can lead to a complete re-design from scratch. So, building a model to predict the system's cost and performance based on early/partial prototype performance test results and knowledge about architectures and environments becomes the main way to mitigate associated risks. A few examples of such models will be discussed in this paper.

Early load testing provides valuable information, but does not give any insight as to how the new application will perform in a production environment with large number of concurrent users accessing large volumes of data. It does not answer how implementing the new application will affect the performance of existing applications and how to change the workload management parameters affecting priorities, concurrency and resource allocation to meet business Service Level Goals. It does not answer whether the production environment has enough capacity to support expected workload growth and increase in volume of data. Should new application be part of Data Warehouse or Big Data environment? Should new application use Cloud platform? What is the best Cloud platform for new applications?

In this paper, we will review the value and limitations of available Load Testing tools and discuss how modeling and optimization technology can expand results of Load Testing. We will review a use case based on using BEZNext Performance Assurance software. We will cover data collection and workload characterization in small test and large production environments. We will review results of anomaly and root cause detection and seasonality determination. We will demonstrate how modeling and optimization are used to predict the impact of new application implementation, find the appropriate platform, develop proactive recommendations and set realistic expectations. This approach reduces the risk of performance surprises and enables automatic results verification.

**Keywords:** Load Testing, Performance Testing, Application Performance, Cloud Platform, Big Data, Data Warehouse, Service Level Goals, Workload Characterization, DevOps, Modeling, Optimization.

### Traditional Load Testing

Traditional performance testing, optimized for the waterfall development process, was mostly focused on the ability of systems to handle peak load. Usually traditional performance testing was conducted as:

- last moment before deployment;
- last step in the waterfall process;
- extensive tools requiring special skills;
- protocol level record-and-playback;
- lab environment;
- scale-down environment;
- checking against given requirements/SLAs.

There is always the risk of system crash or experiencing performance issues under heavy load - and the only way to mitigate it is to actually test the system. Even stellar performance in production and a highly scalable architecture don't guarantee that the system won't crash under a slightly higher load.

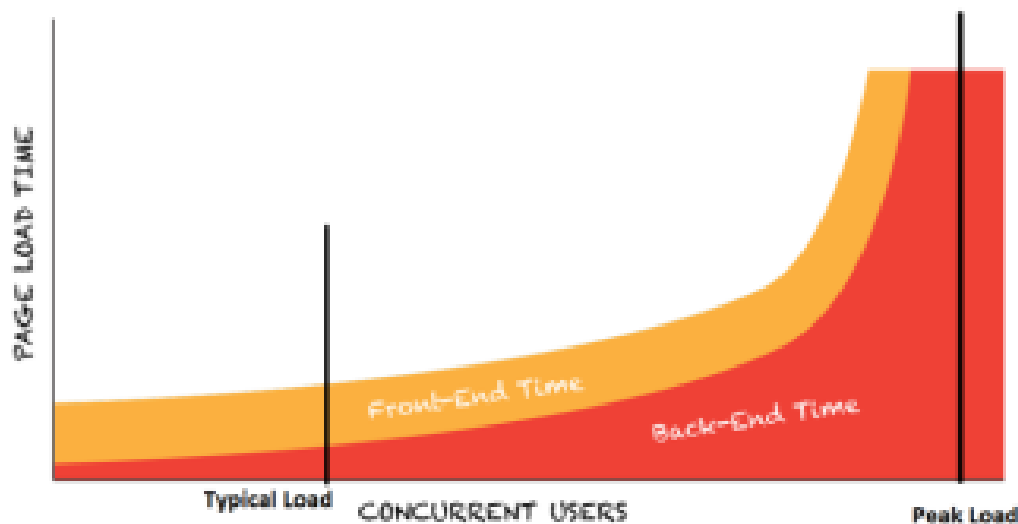


Figure 1. – Typical response time curve

A typical response time curve is shown on figure 1, adapted from Andy Hawkes' post discussing the topic [HAWK13]. As it can be seen, a relatively small increase in load near the curve knee may kill the system - so the system would be unresponsive (or crash) under the peak load.

Traditional performance testing as a part of waterfall development process did mitigate that particular risk if was done properly - however the feedback was usually provided in the very end, when the system was ready to be released. If serious issues were discovered, the cost of fixing them and deployment delays were very high.

Meanwhile the cost of fixing performance and scalability issues early may be many times lower. The idea that we need to test early, as the cost of fixing defects skyrocket later in the development lifecycle, may be traced at least to [BOEH76] and further developed in his book [BOEH81]. See figure 2 from [HICK18] as an attempt to quantify the dependency.

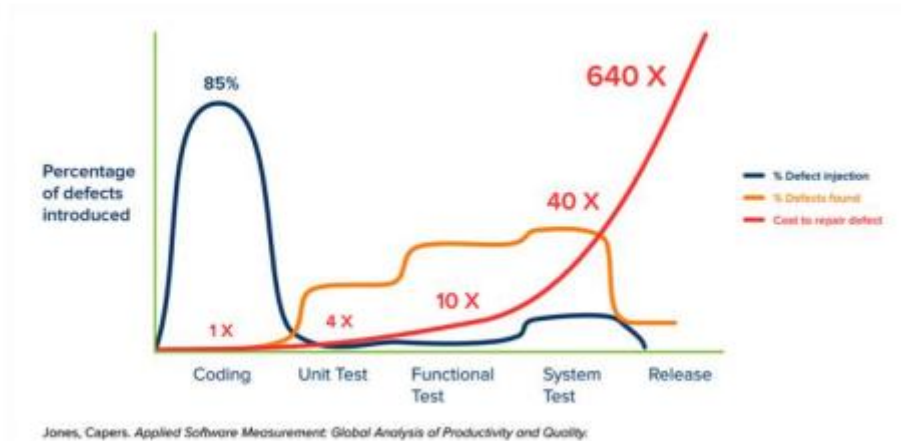


Figure 2. – Cost of fixing defects

However, although it was a common wisdom, not much could be done early as not much was available to test during waterfall development until the very end. Of course, we could define performance requirements and do architecture analysis (fully developed, for example, in [Smith90]). But, without testing, we had very limited information input from the system - until the very late moment when the cost of fixing found defects was very high.

### Today's Performance Testing

Agile / iterative development and DevOps provided an opportunity to start performance work early as we are supposed to get a working system (or at least some of its components) on each iteration [PODE16, PODE19]. So, finally, it is possible to get performance feedback from the system from the first development iterations - so architecture could be verified early and defects found as soon as they appear. That alone drastically increases the value of performance testing. Now performance testing becomes an integral part of DevOps cycle as shown on figure 3.

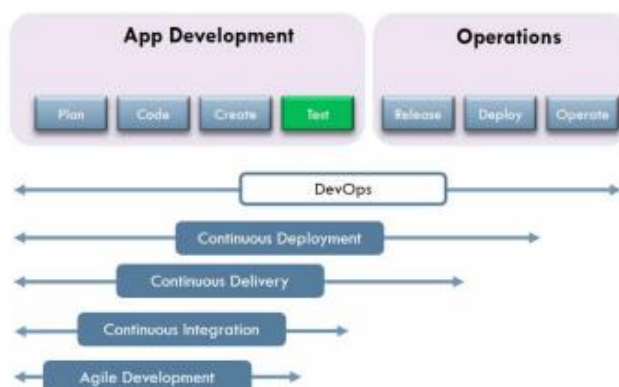


Figure 3. – Performance Testing is an Integral Part of DevOps Process

However, it is not viable in most cases to run full-scale performance tests inside Continuous Integration. There are no generic guidelines anymore - what and how should be tested depends on specific context. Instead of the well-defined traditional way of performance testing, we have a continuum of options along multiple dimensions.

**Scope.** While a full-scale system test still remains an option, performance tests could be run on any level - unit, component, service, subsystem - and with any intensity and kind of load. In many

cases load should be selected explicitly to load a specific part of the system deployed in a particular way.

**Environment.** options nowadays include traditional internal (and external) labs; cloud as 'Infrastructure as a Service' (IaaS), when some parts of the system or everything are deployed there; and service, cloud as 'Software as a Service (SaaS)', when vendors provide load testing service. There are some advantages and disadvantage of each model. Depending on specific goals and systems to test, one deployment model may be preferred over another.

For example, to check the effect of performance improvement (performance optimization), using an isolated lab environment may be a better option to see even small variations introduced by a change. To test the whole production environment end-to-end to make sure that the system will handle load without any major issue, testing from the cloud or a service may be more appropriate. To create a production-like test environment without going bankrupt, moving everything to the cloud for periodical performance testing may be a solution. For comprehensive performance testing, you probably need to use several approaches - for example, lab testing (for continuous performance testing and performance optimization to get reproducible results) and distributed, realistic outside testing (to check real-life issues that can't simulate in the lab). Limiting yourself to one approach limits the risks you will mitigate.

**Testing Approach.** Instead of a traditional testing approach of full-scale realistic workload to simulate the production system, we have a whole dimension of options from early exploratory / agile performance testing (somewhat corresponding to new systems which we don't know much about) to automated / regression testing (somewhat corresponding to well-known systems where only small enhancements get implemented) - and the traditional approach may be depicted as a dot on that dimension (see fig.4 to illustrate that idea).

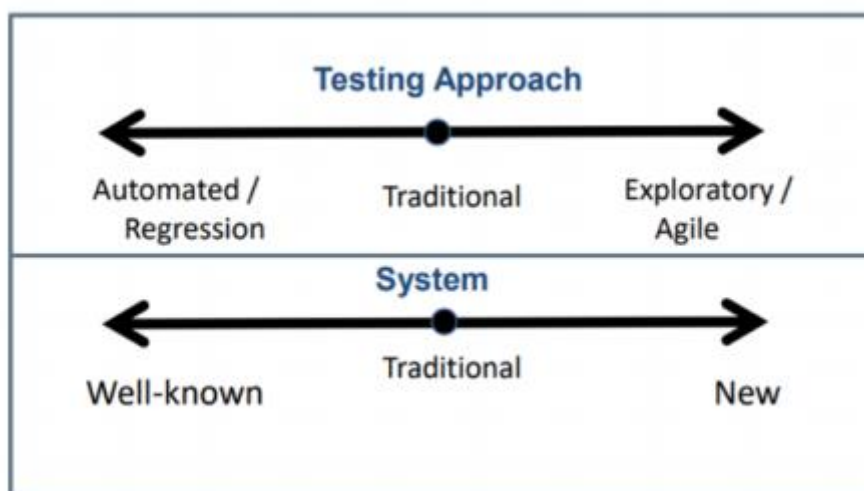


Figure 4. – Testing approach continuum of options

**Load Generation.** Quite often the whole area of load testing is reduced to pre-production testing using protocol-level recording/playback. While protocol-level recording/playback was (and still is) the mainstream approach to testing applications, it is definitely just one type of load testing using only one type of load generation.

The time when all communication between client and server was using simple HTTP is in the past and the trend is to provide more and more sophisticated interfaces and protocols. While load generation is a rather technical issue, it is the basis for load testing - you can't proceed until you figure out a way to generate load. As a technical issue, it depends heavily on the tools and functionality supported.

There are three main approaches to workload generation [PODE14] and every tool may be evaluated on which of them it supports and how.

**Protocol-level recording/playback.** This is the mainstream approach to load testing: recording communication between two tiers of the system and playing back the automatically created script (usually, of course, after proper correlation and parameterization). As far as no client-side activities are involved, it allows the simulation of a large number of users. Such a tool can only be used if it supports the specific protocol used for communication between two tiers of the system.

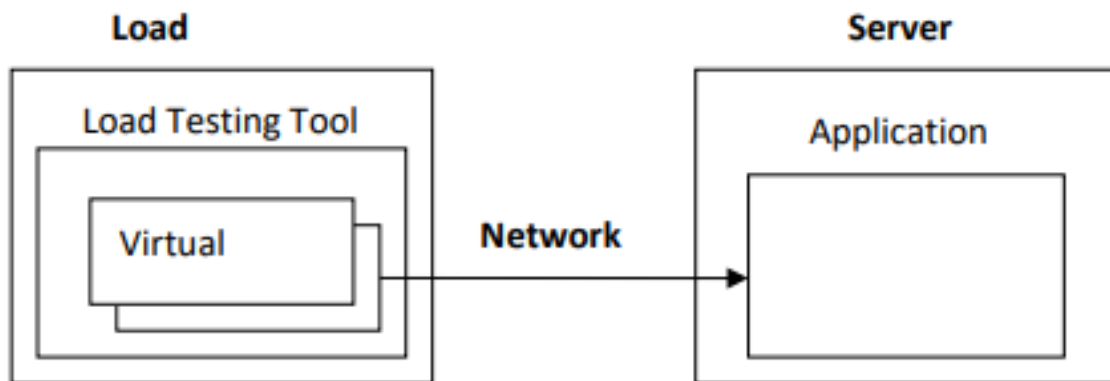


Figure 5. – Record and playback approach, protocol level

**UI-level recording/playback.** This option has been available for a long time, but it is much more viable now. In the past, a separate machine was needed for each virtual user (or at least a separate terminal session). This drastically limited the load level that could be achieved.

New UI-level tools for browsers, such as Selenium, have extended the possibilities of the UI-level approach, allowing running of multiple browsers per machine (limiting scalability only to the resources available to run browsers). Moreover, UI-less browsers, such as HtmlUnit or PhantomJS, require significantly fewer resources than real browsers.

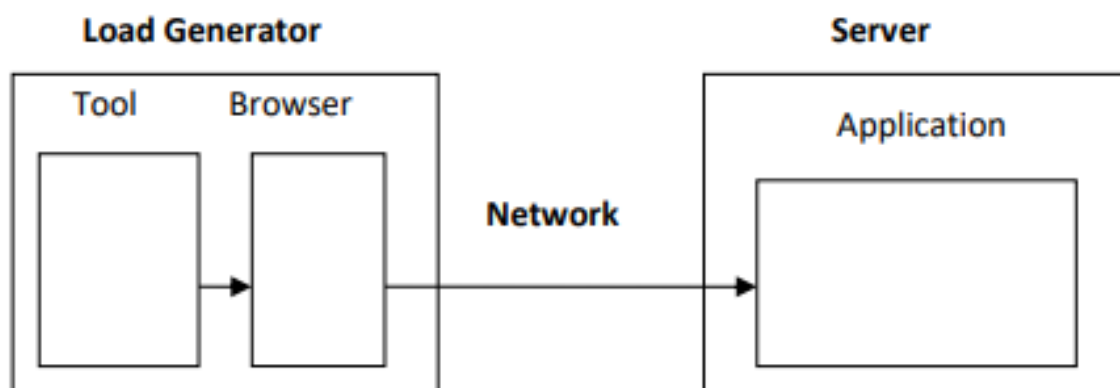


Figure 6. – Record and playback approach, browser user

**Programming.** There are cases when recording can't be used at all, or when it can, but with great difficulty. In such cases, API calls from the script may be an option. Often it is the only option for component performance testing. Other variations of this approach are web services scripting or use of unit testing scripts for load testing. And, of course, there is a need to sequence and parameterize your API calls to represent a meaningful workload. The script is created in whatever way is



appropriate and then either a test harness is created or a load testing tool is used to execute scripts, coordinate their executions, and report and analyze results.

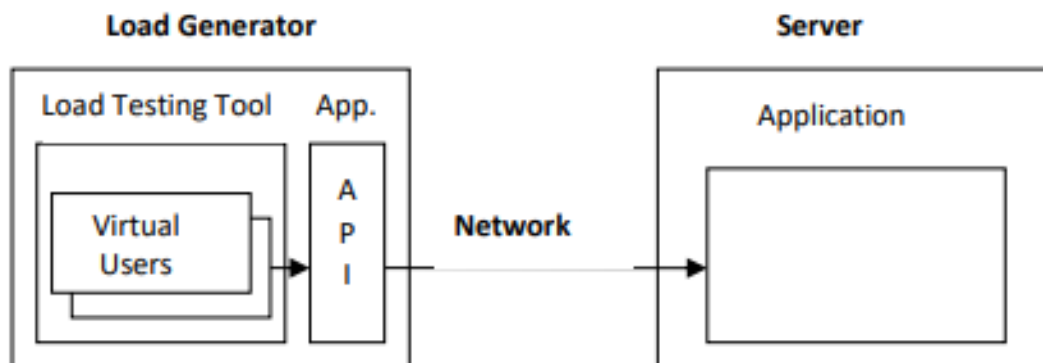


Figure 7. – Programming API using a Load Testing Tool.

To do this, the tool should have the ability to add code to (or invoke code from) your script. And, of course, if the tool's language is different from the language of your API, you would need to figure out a way to plumb them. The importance of API programming increases in agile / DevOps environments as tests are run often during the development process. In many cases APIs are more stable than GUI or protocol communication – and even if something changed, the changes usually can be localized and fixed – while GUI- or protocol-based scripts often need to be re-created. Load Testing Tool Virtual Users Load Server Generator Application Network Tool Browser Load Generator Server Application Network Load Testing Tool App. Virtual Users Load Generator Server Application Network A P I All Rights Reserved 6 Load Testing Tools. There are quite a few load testing tools [PODE14, LONN20]. While most load testing tools look similar at first glance, they are actually quite different in supporting different performance testing options discussed above. And, unfortunately, generic descriptions (for example, from the vendor website) are usually useless in understanding the differences. Each situation is different. A tool may be very good in one situation and completely useless in another. The value of the tool is not absolute; rather it is relative to a specific situation. To list some examples, Microfocus LoadRunner family, Microfocus Silk Performer, Neotys NeoLoad, IBM Rational Performance Tester, RadView WebLoad, and SmartBear LoadNinja may be mentioned among commercial tools. Apache JMeter, Gatling, k6, and Locust may be mentioned among open source tools. Broadcom BlazeMeter, Tricentis Flood.io, RedLine13, and Octoperf are examples of commercial extensions of open source tools.

**Testing Strategy.** Seeing that continuum of performance testing options along different dimensions, it is obvious the testing strategies became very non-trivial - as a specific set of tests and their timing is defined by specific context. “Automation” is only one part of it - continuous performance testing is very important for iterative development, but it is just part of performance testing strategy addressing a specific risk - performance regression between builds. Moreover, performance testing should be considered as part of a larger performance engineering strategy [PODE18].

Asking the right questions may help to formulate a proper testing strategy. For example, such questions could be:

- What are performance risks we want to mitigate?
- What part of these risks should be mitigated by performance testing?
- Which performance tests will mitigate the risk?
- When we should run them?
- What process/environment/approach/tools will we need in our context to implement them?

### **Value and Limitations of Performance Testing**

Performance testing provide an immense value as a proactive way to mitigate performance risk. Early problem detection prevents costly redesigns and delays. Considering the flexibility of today's performance testing the strategy may be optimized for specific context to provide the best return on investments.

Moreover, early / continuous performance testing provides a constant stream of real performance-related information. Even for existing systems, it provides important input to see a possible performance impact on production systems. But it becomes really invaluable for new systems as only early performance feedback enables developers to identify and fix performance issues before new application deployment.

However, it is important to understand the limitation of performance testing. It is expensive on a high-scale level, so the number of large-scale tests that can be run is limited. Smaller-scale tests provide very important, but partial information - which by themselves don't provide a holistic view.

Modeling complements performance testing here, enabling a big picture view and answering what-if questions from disjointed performance testing results. It is invaluable for evaluating options and developing proactive recommendations for the system's architecture and design.

### **Role of Modeling During DevOps Process**

#### **Value of modeling results for Application Developers**

##### **– Predict new applications implementation impact**

- predict how new application will perform in production environment;
- identify anomalies and their root causes during testing of new applications;
- develop recommendations to application developers/

##### **– Predict how new application will affect existing production applications**

- predict how implementation of new applications will affect Response Time and Throughput of existing applications;
- develop capacity planning recommendations;
- set up realistic expectations.

### **Value of modeling for Operations**

#### **1 Develop Proactive Performance Management and Workload Management**

##### **Recommendations**

- compare performance measurement results after implementation of the new application with expected;
- develop proactive performance tuning recommendations;
- develop proactive workload management recommendations;
- reevaluate Capacity Planning recommendations.

Modeling plays the central role in Performance Assurance

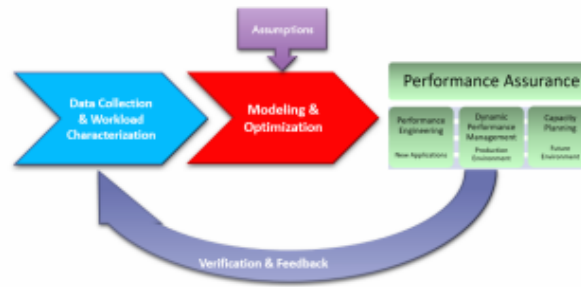


Figure 8. – Modeling and Optimization are a foundation of Performance Assurance, which includes Performance Engineering for new applications, Dynamic Performance Management after deployment of the new applications in production and Capacity Planning supporting expected growth and selection of the appropriate platform for new application

Modeling results predict the impact of the different measures on performance for each workload.

Workload and volume of data growth affect workloads' queueing and software delay time/ Response time of new application in production has:

- different Response Time, Service Time, Queueing time and Delay Time for new application;
- response time of production workloads and it's Queueing Time and Delay Time are changed;
- move workloads to the Cloud can affect Service Time, Queueing Time and Delay Time for all workloads ;
- change of the Workload Management Rules (Priorities, Concurrency and Resource Allocation) affect the Queueing time and Delay Time of each workload.



Figure 9. – Modeling results predict how different changes will affect the major components of the Response Time, Throughput and resource utilization by each workload

We use an optimization engine to run models iteratively to find the optimum workload management parameters and additional resources which will be required to meet Service Level Goals (SLGs ) for each of the growing and changing workloads.

Let's review 10 steps of using measurement results of performance testing during DevOps to build models and predict new application impact and develop proactive recommendations about what should be done to meet SLGs for new and existing production applications with lowest cost. All Rights Reserved 8.

The first step in using modeling is data collection during performance tests in typically small test environments and data collection for production workloads in a large production environment. We use OS agents to collect data about each process and DBMS agents to collect data from systems

tables. Measurement data are aggregated and transformed into common format, which includes the following data types used to build models:

- hardware and Software Configuration;
- information by each node/server by user and application, including;
- response Time;
- throughput;
- CPU Utilization and CPU Service Time per request;
- disk Utilization, I/O rate, #I/O operations per request and KB/Request, Channel Utilization;
- memory utilization;
- network utilization;
- level of concurrency.



Figure 10. – Data Collection in Test and Production Environments

The Second Step is workload characterization of test and production environments. Each workload represents the activity of a group of users using a set of applications supporting a specific Line of Business. Workload Aggregation process aggregates measurement data by Module of the New Application or by Line of Business / Workload, Aggregation use rules describing user names and program names which belong to the specific module or line of business.

Workload characterization is an automated process performed hourly. Each workload is characterized by performance, usage of resources and patterns of accessing data. Example of the workload characterization results showing CPU utilization and number of I/O operations by each workload during 24 hours is shown on figure 11.

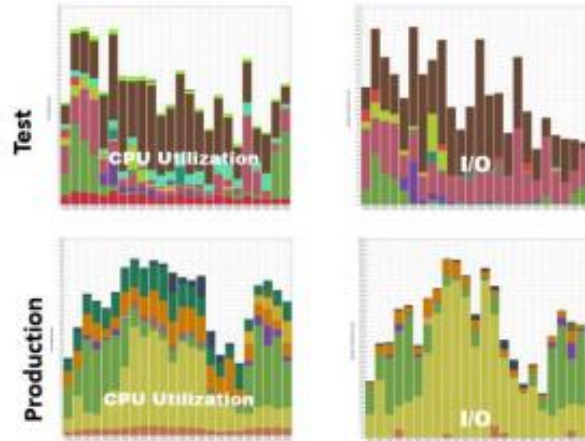


Figure 11. – Workload Characterization in Test and Production Environments

Results of the workload characterizations are used for Anomaly and Root cause determination and as input for a model.

The third step is anomaly detection. Measurement data collected during performance tests after each build are used to detect the anomalies and their root causes. Information about the most severe anomalies and root causes determined in test and production environments are passed to application developers and operations (figure 12.) We assume that the most severe anomalies will be fixed prior to deployment of the new application in production.



Figure 12. – Detection of the workloads with highest frequency of Anomalies and Programs and Users causing the most severe problems during DevOps provide immediate information to Application developers and narrows down the scope of tuning efforts

Fourth Step is Workload Forecasting for New and existing Production Applications. Workload Forecasting for a new application is typically based on Business Plan and for existing Production workloads based on analysis of the historical data.

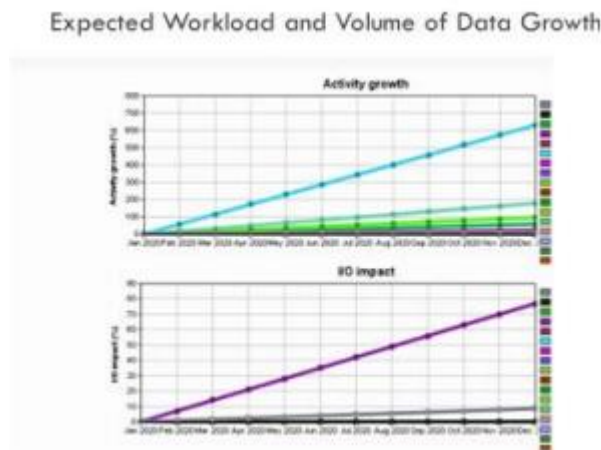


Figure 13. – Workload Forecasting shows expected increase in number of transactions and volume of data accessed by workload

Fifth Step is predicting impact of the expected increase in number of users and volume of data on performance of existing workloads in Production environment. Predicted results show when existing production workloads will not meet SLGs on current production environment even without the new application impact. [ZIB1].

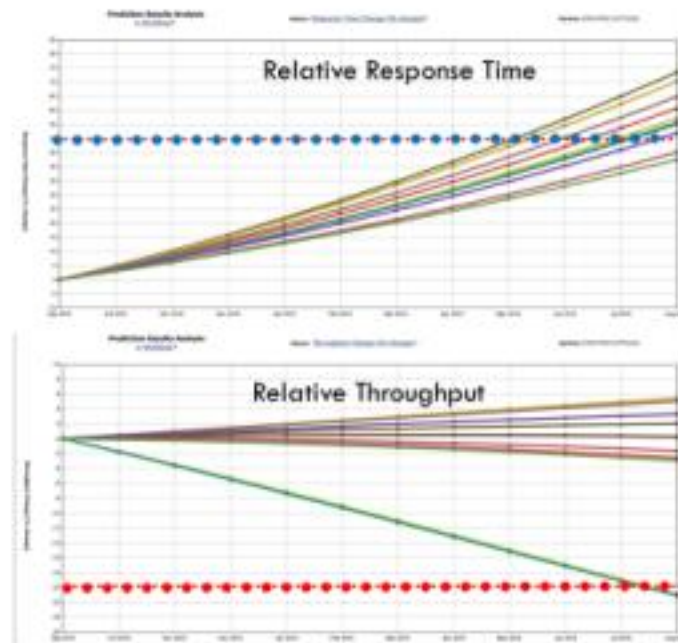


Figure 14. – As a result of the workload growth and increase in Volume of Data SLGs will not be met

Sixth Step is Predicting the New Application Implementation Impact Performance prediction results showing expected Performance (response time, throughput), Resource utilization (CPU utilization, Disk utilization, Network utilization, Memory utilization) after deployment of new application.



Figure 15. – After deployment of new application, existing production workload will not be able to meet SLGs and new application will have performance problems starting in November. So, can tuning and workload management optimization avoid an expensive upgrade

Seventh Step is Predicting Impact of the Workload Management Optimization Workload management rules of changing priorities, limitation of concurrency and allocation of resources for critical workloads allow operations to influence reallocation of resources between workloads to meet SLGs without adding new instances. BEZNext modeling and optimization technology [ZIB2] automatically evaluates different options and recommends rules for different time of day, and sets realistic expectations of response time, throughput and resource utilization for each workload. Figure 16 shows that changes of workload management rules will not be sufficient to meet SLGs and additional resources will be required in 6 months.



Figure 16. – Change of the workload management rules affecting priorities between all workloads will not be sufficient to meet SLGs.

Eighth Step is Predicting the Minimum On Prem Upgrade Required to meet SLGs after Deployment of new Application According to the performance prediction results [ZIB3] as we can see from Figure 17 at least 4 additional nodes will be required in 6 months to meet SLGs through the end of the year. It is pretty expensive and an alternative platform can be evaluated.

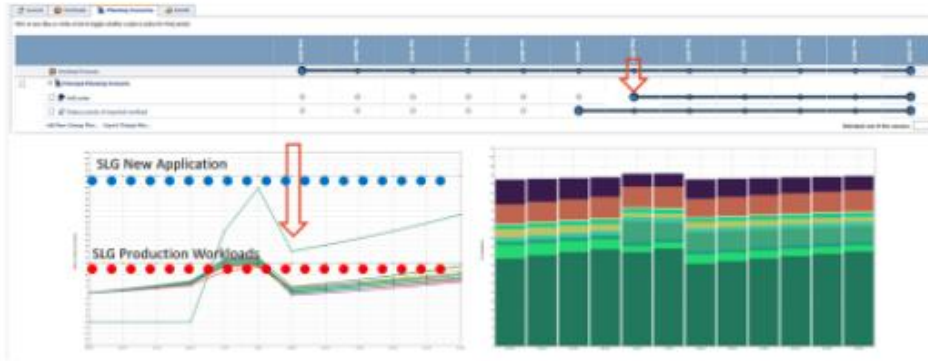


Figure 17. – Predicting when and how much additional resources will be required to meet SLGs for existing and new workloads. According to the model, an additional 4 nodes will be required in 6 months On Prem to meet SLGs through the end of the year.

**Ninth Step is Determining Appropriate Cloud Platform for New Application.**

BEZNext Modeling and Optimization use measurement data characterizing new application performance, resource utilization and hardware and software configuration of the test environment to predict how new application will perform on different Cloud platforms.

It takes into consideration the expected workload and volume of data growth and predicts the minimum number of instances and instance types which will be required to meet SLGs for new workload. It also predicts the impact of moving a new workload to one of the existing Clouds [ZIB1].

Predicted information about the number and type of instances which will be required during different hours of the day and different months of the year is used to predict the cost of supporting a new workload in different Cloud environments.

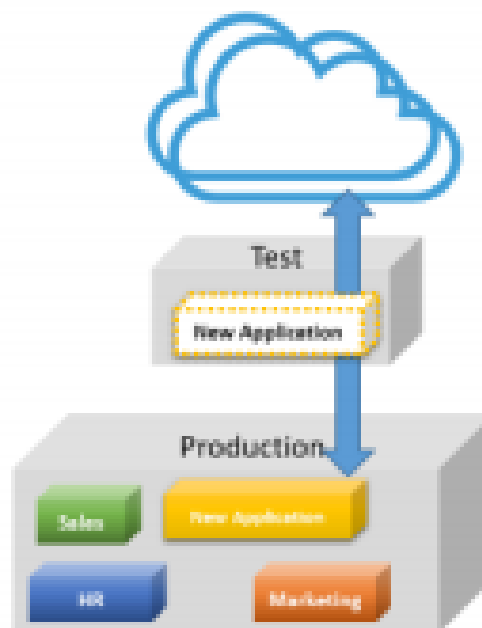


Figure 18. – Applying Modeling and Optimization to select appropriate Cloud Platform

Tenth Step is Automatic Results Verification Automatic comparison of the actual measurement data with expected / predicted results identifies new anomalies and enables creation of the continuous Performance Assurance Process.



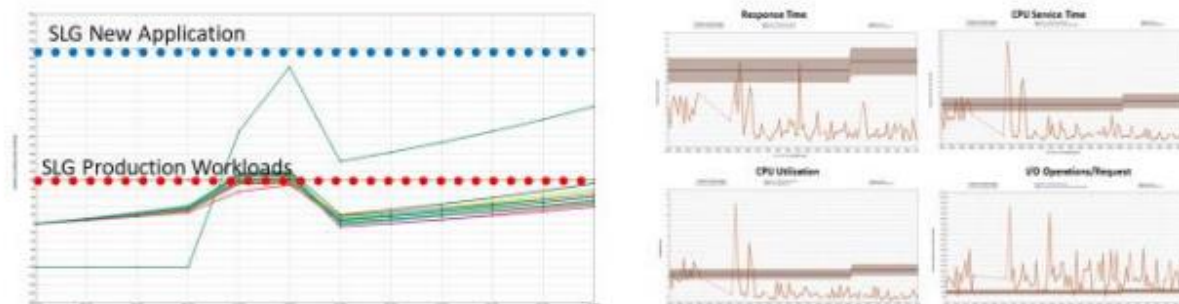


Figure 19. – Verification by comparing actual results with expected

### Summary:

1. Performance testing is the main source of performance measurement data during development process
2. Performance measurement data is needed to create and validate models predicting new applications performance
3. Modeling complements performance testing allows fast and inexpensive analysis of what-if scenarios
4. Modeling results provide value to Application Developers and Operations during DevOps process
5. Combination of Performance Testing and Modeling is a way to mitigate performance risks early and avoid performance surprises.

### References

- [1] Lönn, R. Open source load testing tool review 2020. <https://k6.io/blog/comparing-best-open-source-load-testing-tools>
- [2] Andy Hawkes, A. When 80/20 Becomes 20/80. <http://www.speedawarenessmonth.com/when-8020-becomes-2080/>
- [3] Load Testing at the Speed of Agile. Neotys White Paper, 2014. [http://www.neotys.com/documents/whitepapers/whitepaper\\_agile\\_load\\_testing\\_en.pdf](http://www.neotys.com/documents/whitepapers/whitepaper_agile_load_testing_en.pdf)
- [4] Podelko, A. Context-Driven Performance Testing, CMG imPACT, , 2019.
- [5] Podelko, A. Context-Driven Performance Engineering, Performance Calendar, , 2018. [PODE16] Podelko, A. Reinventing Performance Testing, CMG imPACT, , 2016.
- [6] Podelko, A. Performance and Capacity Conference by CMG, 2014
- [7] B. Zibitsker, A. Lupersolsky, “How to Apply Modeling to Compare Options and Select the Appropriate Cloud Platform”, ICPE 2020, Canada
- [8] B. Zibitsker, IEEE Conference, Delft, Netherlands, March 2016, Big Data Performance Assurance
- [9] B. Zibitsker, Proactive Performance Management for Data Warehouses with Mixed Workload, Teradata Partners, 2008, 2009
- [10] Hicken, A. The Shift-Left Approach to Software Testing, 2018 <https://www.stickyminds.com/article/shift-left-approach-softwaretesting>
- [11] Smith, C. Performance Engineering of Software Testing, 1990.
- [12] Boehm, B. Software Engineering, IEEE Trans. Computers, 1976.
- [13] Boehm, B. Software Engineering Economics, 1981.

UDK 004.75

## **DATA PROCESSING IN HIGH-PERFORMANCE COMPUTING SYSTEMS**



**A. Adamov**  
*Director, Center for Data Analytics Research (CeDAR)*



**A.I. Buranbaeva**  
*Associate Professor*



**S.I. Gindin**  
*Associate Professor of Information and Computer Systems Department*



**M. Glesner**  
*Professor, IEEE Fellow, head of Microelectronic System Research Group*



**I. Horton**  
*Author of books dedicated to programming languages C/C++/Java and software development*



**A. Keevallik**  
*Professor of Tallinn University of Technology*



**A.D. Khomonenko**  
*Professor, the head of Information and Computer Systems Department*



**Y.G. Klyuyeva**  
*Senior lecturer of Information Technology and Security Department*



**I. Skliarova**  
*Auxiliary Professor of Aveiro University*



**V. Sklyarov**  
*Professor, honorary doctor*



**V.V. Yavorskiy**  
*Professor of Information Technology and Security Department*

*Center for Data Analytics Research (CeDAR), Baku, Azerbaijan  
Kazakh Academy of Transport and Communications (Institute of Information and Computer Technologies, Ministry of Education and Science of Kazakhstan), Almaty, Kazakhstan  
Information and Computer Systems Department of Emperor Alexander I St. Petersburg State Transport University, St. Petersburg, Russia  
Darmstadt Technical University, Darmstadt, Germany  
Stratford-upon-Avon, England  
Tallinn University of Technology, Tallinn, Estonia  
Karaganda State Technical University, Karaganda, Kazakhstan*

*Aveiro University, Aveiro, Portugal  
E-mail: skl@ua.pt*

**A. Adamov**

*Director, Center for Data Analytics Research (CeDAR), School of IT & Engineering at ADA University, General chair of the 14th IEEE International conference on Application of Information and Communication Technologies (details about upcoming conference are available at: [www.aict.info](http://www.aict.info)), Baku, Azerbaijan. The main research area is data science and analytics.*

**A.I. Buranbaeva**

*Associate Professor of Kazakh Academy of Transport and Communications (Institute of Information and Computer Technologies, Ministry of Education and Science of Kazakhstan), Almaty, Kazakhstan. The main research area is application of big data.*

**S.I. Gindin**

*Associate Professor of Information and Computer Systems Department of Emperor Alexander I St. Petersburg State Transport University, St. Petersburg, Russia. The main research area is big data technologies, databases, and modeling of IT systems.*

**M. Glesner**

*Professor, IEEE Fellow, head of Microelectronic System Research Group at Darmstadt Technical University, honorary doctor of Tallinn Technical University (Estonia), Bucharest Polytechnic University (Romania), Mongolian Technical University (Mongolia) and the University of Liepaja (Latvia), Darmstadt, Germany. The main research area is in the scope of embedded systems design, high-level synthesis and physical design.*

**I. Horton**

*He worked for IBM for more than 30 years and was responsible for implementing large systems in automotive, aerospace, and ship building. From 1972 to 1976 he was project manager for the design and installation of the IBM systems at Kamaz (Naberezhnye Chelny, Russia). After retirement from IBM he started writing books about programming.*

**A. Keevallik**

*Professor of Tallinn University of Technology, honorary doctor of Helsinki University of Technology (Finland) and Caucasus University (Georgia), Tallinn, Estonia. He is author and editor of 10 books. The main research area is digital systems decomposition based on divide-and-conquer algorithms.*

**A.D. Khomonenko**

*Professor, the head of Information and Computer Systems Department of Emperor Alexander I St. Petersburg State Transport University, St. Petersburg, Russia. He is author of 25 books. The main research area is big data technologies, databases, and modeling of IT systems.*

**Y.G. Klyuyeva**

*Senior lecturer of Information Technology and Security Department of Karaganda State Technical University, Karaganda, Kazakhstan. She is author of 3 books. The main research area is big data technologies.*

**I. Skliarova**

*Auxiliary Professor of Aveiro University, Aveiro, Portugal. She is author of 6 books written in English and Chinese. The main research area is data processing, FPGA-based designs, hardware accelerators and combinatorial optimization.*

**V. Sklyarov**

*Professor, honorary doctor of Tallinn University of Technology (Estonia), Aveiro, Portugal. He is author of 23 books written in Russian, English and Chinese. The main research area is reconfigurable computing, data processing, and hardware accelerators.*

**V.V. Yavorskiy**

*Professor of Information Technology and Security Department of Karaganda State Technical University, Academician of RANS.. He is author of 14 books. The main research area is big data technologies.*

**Abstract.** The paper integrates the results of a large group of authors working in different areas that are important in the scope of big data, including but not limited to: overview of the basic solutions for the development of data centers; storage and processing; decomposition of a problem into sub-problems of lower complexity (such as applying divide and conquer algorithms); models and methods allowing broad parallelism to be realized; alternative techniques for potential acceleration; programming languages; and practical applications.

**Keywords:** big data, data processing, data centers, MapReduce concept, hardware accelerators.

*Introduction.* Engineering and scientific applications have become more data intensive [1] demanding emerging high-performance interactive frameworks. Existing processing techniques may be applied at different stages from upper levels aimed at developing data processing centers with the primary focus on software for high performance accelerators that involve additional hardware components. Hence, a range of diverse types of expertise are required, which resulted in the large group of authors being involved. This allowed the scope of big data and acceleration at lower levels to be combined and reviewed. The importance and complexity of big data [2] forces computer engineers to seek new solutions for processing them. Processing speed still continues to rise exponentially while memory bandwidth increases at a much slower rate [3]. Working with big data requires huge storage, extensive processing capability, and high-performance communications for data exchange [3]. It is explicitly stated in [2] that based on previous knowledge we need to introduce new techniques for data manipulation, analysis, and mining. The experience of the authors of the paper falls into three different areas: big data management; high-performance computing systems and hardware accelerators; practical applications necessitating working with big data. It is difficult or perhaps impossible to discuss all the potential features in the areas selected above. Thus, particular topics (given sections of the paper) have been chosen in accordance with the authors' expertise and they are dedicated to: basic solutions for the development of data centers; a review of potential hardware accelerators and how they can be applied to the selected processing technique, namely MapReduce; big data storage and processing involving emerging technologies and platforms; operations over data that may be accelerated significantly, and potential practical applications. We will also briefly discuss useful mathematical models and methods that can efficiently be applied at different stages of big data processing, from the top level to hardware acceleration relying on divide-and-conquer procedures and different types of system decomposition.

For acceleration purposes we will talk mainly about the MapReduce processing technique and the relevant model for distributed computations that was based originally on Java programming language and contains two important steps. *Map* takes and divides data into subsets of individual elements that are of lower complexity, and are represented by tuples (key/values pairs). A simple example can be found in tutorial [4]. *Reduce* at the second step takes the outputs from the first step and builds a smaller set of tuples. We will consider the *MapReduce* processing technique from slightly different point of view and apply hardware acceleration at specific sub-steps.

As a simplified example, let us consider a group of students with the main objective being to find out the number of occurrences of each student's name. Similar problems appear in many practical applications requiring acquisition, analysis and filtering of large data sets [5]. In [6] a data mining problem is explained, with an analogy to a shopping card. A basket is the set of items purchased at one time. A frequent item is an item that often occurs in a database. A frequent set of items often occur together in the same basket. A researcher can request a particular support value and find the items which occur together in a basket either a maximum or minimum number of times within the database [6]. Similar problems arise to determine frequent queries on the Internet, customer transactions, credit card purchases, etc. These require very large volumes of data to be processed within a day [6]. As an example, the following list depicts the first names of students from part of a group in Aveiro University (Portugal): Tiago, Maria, Maria, Maria, Tiago, Ana, Beatriz, Ana, Tiago, Carolina, Maria, Beatriz. The name Maria is the most frequent (4 times), and the name Carolina is the least frequent (once). Figure 1 shows the operations that are executed in the MapReduce procedure.

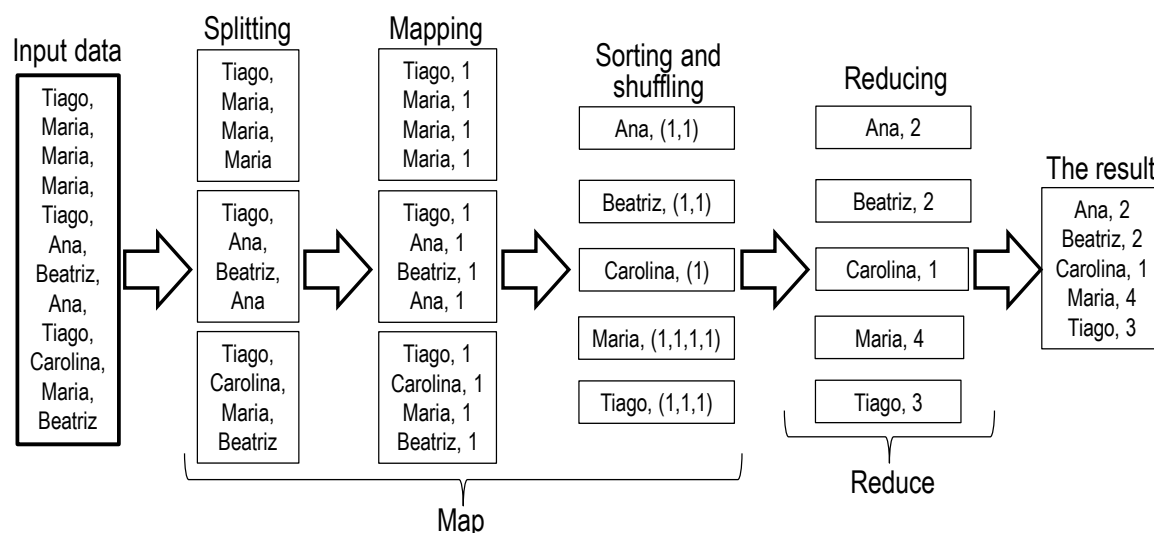


Figure 1. – An example of the MapReduce procedure

Initially, the data set containing the names is divided into three subsets with four names in each. The number of objects (names) and the number of subsets have been chosen somehow. Initially, the number of appearances for each name is assigned to 1. The key is the name and the value is 1, for example «Tiago, 1». Then sorting and shuffling are done and all the tuples with the same name (key) are sent to the corresponding block. For example, the name Tiago from the first and the third mapping blocks are mapped to the fifth block created after sorting and shuffling: «Tiago, (1,1,1)». Counting the Hamming weight (the number of ones) gives the final output shown on the right-hand side of Fig. 1. As we can see, the operations applied are: extraction of non-repeated names; sorting; and counting the number of ones for each such name.

There are many hardware accelerators for executing the operations identified above that are proposed in [5] with all necessary details including hardware specifications in VHDL (Very high speed integrated circuits Hardware Description Language) and software models described in Java. In [7-9] many other useful and easily understandable software models are described in C/C++ and Java languages.

The remainder of this paper is organized in 6 sections. Section 2 reviews and discusses basic solutions for the development of data centers that provide support for managing big data. Section 3 presents a brief review of FPGA-based hardware accelerators for some selected important problems; these are sorting, the extraction of non-repeated names/values, discovering the most frequently occurring items, and Hamming weight computations. The importance of divide-and-conquer algorithms and system decomposition is emphasized. Section 4 is dedicated to emerging technologies and platforms for big data storage and processing. Section 5 overviews various operations over data. Section 6 demonstrates some feasible practical applications from the information system for electoral processes in Kazakhstan. The conclusion is given in section 7.

The material of the paper has been distributed between the authors as it is shown below:

- Sections 1, 3 and 5 have been prepared by M. Glesner (Germany), I. Horton (England), A. Keevallik (Estonia), I. Skliarova, and V. Sklyarov (Portugal) collaborating for more than 30 years. They have extensive experience in data processing based on divide-and-conquer algorithms, system decomposition at various levels, optimization applying discreet mathematical models, accelerations with FPGAs (Field-Programmable Gate Arrays) / PSoC (Programmable Systems-on-Chip), embedded systems, and computer-aided design techniques.

- Section 2 has been written by Y.G. Klyuyeva and V.V. Yavorskiy from Kazakhstan with long experience in big data management.

– Section 4 has been prepared by A. Adamov from Azerbaijan (sub-sections 4.1, 4.2, 4.4, 4.5, 4.6) and S.I. Gindin and A.D. Khomonenko from Russia (sub-section 4.3) with extensive experience in big data management. It should be noted that A. Adamov is the organizer of fourteen International IEEE conferences closely related to big data and analytics.

– Section 6 has been written by A.I. Buranbaeva (Kazakhstan) based on her practical experience.

### ***Basic solutions for the development of data centers***

One of the main goals of developing data processing centers (DPCs) is to provide information support for computer-based decisions for all or main activities of the relevant organization. The nature and scale of the data analysis tasks that need to be solved also determines the approaches to choosing architecture and designing a data center.

Several large companies provide solutions for data centers and leading positions are held by Oracle, Microsoft, IBM, and Teradata [10]. Oracle is a leader in the field of database management system (DBMS) software and supports all possible architecture options, including clusters, symmetrical multiprocessor systems, and over 80 variants of the operating environment, including IBM mainframes, DEC VAX, UNIX, Windows, and many other platforms. Offered solutions ranging from entry-level, to high-performance systems are scalable and secure, which is also critical for business applications [11]. Oracle also provides support for compatibility with older solutions (three quarters of Oracle customers have been working with Oracle Database for more than 10 years). The disadvantages of Oracle include the high cost and complexity of licensing, as well as problems associated with the release of updates. When designing a data center, Oracle offers a wide range of products, from using a certified configuration to a device that is ready to configure the data center. Exadata proprietary solutions are also available: Oracle Exadata X2–2 for data centers and mixed workloads, Oracle Exadata X2–8 for cloud solutions, and Oracle Exadata Storage Expansion Rack X2–2 for increasing data center capacity. The competitive capabilities of Oracle are determined by the following factors:

–there is a set of ready-made applications for the development of data centers, providing a complete life cycle;

–the company is one of the leaders in sales in the field of data analysis;

–achieving compatibility with products made by other companies.

Microsoft has also a strong position in the database market. The target is creating an instrumental and technological environment that would minimize the cost of developing data centers and making this process accessible to the mass user. The solutions offered by the company in the field of data warehousing concentrate on the development of OnLine Analytical Processing (OLAP) tools [12]. The offered solutions are SQL Server DBMS and the Azure SQL DATABASE cloud service. The company received the highest satisfaction from customers for meeting their needs, value for money, service, support, and overall experience. Microsoft has also increased competitiveness by launching the free Developer Edition of SQL Server and Database Migration Service tools for migrating SQL Server and Oracle databases to Azure SQL Database. SQL Server is used to work with small and medium-sized databases, as well as for large enterprise-wide databases. But, despite the strengths, many corporate customers still do not consider this DBMS suitable for mission-critical applications. According to the experience if the number of users exceeds 2000, then you need to switch to a higher-level DBMS, for example, from Oracle. In the data center market, Microsoft offers its Microsoft Data Warehouse and Fast Track Data Warehouse solutions to provide data centers for customers who do not need a massively parallel database management system. Microsoft released its own mass-parallel architecture data center device, SQL Server 2008 R2 Parallel Data Warehouse, in November 2010. This architecture is implemented in future versions of products [13].

IBM Corporation offers both standalone DBMS solutions and data center devices. The IBM Smart Analytics System (ISAS) family is currently available on the market, with IBM Data Center

software that is InfoSphere Warehouse for Unix, Linux, Windows, and z / OS. There are thousands of database clients worldwide. It is easy to rich functionality of the solutions, including the cloud and hybrid capabilities that the company's products have, as well as the active use of popular open source solutions (Hadoop, Kafka, Parquet, Spark, etc.) and backup and restore functions in or from Swift and AWS S3. The core for data center is the DB2 family of object-relational DBMSs. The data manipulation language is SQL. The advantage of IBM solutions is manifested when both the operational data processing systems and the data center are managed by IBM software, that is, when the so-called closed standard solution is offered. However, IBM's revenue and market share in the operational database systems has been declining for several years. The DB2 database system loses to most competitors in terms of transaction processing speed and data loading. There are also difficulties with pricing and licensing.

Teradata works for more than 30 years in the data center market in combination with the established equipment and specialized database analytics software. The products include data mining tools, departmental data mining and enterprise solutions, as well as cloud and big data products. Aster Data has added new features to the Teradata product line (such as MapReduce, unstructured data, and graphical analysis).

Hence, there are several options for the implementation of data centers within the framework of a typical architecture. Let us consider the features of technological solutions for some of them.

Virtual data warehouse. The architecture provides a real-time access to live data through the middleware software. This solution is based on a metadata repository that describes data sources, procedures for their preliminary processing, and formats for presenting information to the end user. The disadvantages of such solution are intensive network traffic, a decrease in the performance of a non-existent system, and the threat of data integrity violation in the event of unsuccessful data center users.

Data Kiosks. The architecture is a lightweight version of the data center thematic focus. There are data kiosks associated with an integrated or unrelated (stand-alone) data center.

Global data warehouse. The architecture is a single source of integrated organization data.

Data warehouses with a multilevel (mainly three-tier) architecture, or corporate data centers. The architecture is a kind of global data center in which three levels are technologically implemented. At the first level, the organization's corporate data center is located. At the second level, related topic-specific data kiosks based on multidimensional DBMS are supported. At the third level there are user client applications with data analysis tools installed.

Built-in (combined) data storage. The architecture is a data center that is organically integrated into a virtual enterprise (Enterprise Information Factory, EIF) or used as a component of analytical support of business functions.

Corporate Information Factory (CIF). It is a development of enterprise data warehouse (EDW) architecture involving coordinated data extraction from sources, loading them into a relational database with a structure in the third normal form, using the constructed data center to populate additional repositories of presentation data.

Data Warehouse Bus architecture. In this architecture, the data center is not a single physical repository (unlike CIF). This is a "virtual" component, representing a collection of data items, each of which has a star architecture.

United (federated) data center. In this architecture, there are a number of data center instances that operate on a semi-autonomous basis and, as a rule, they are organizationally or geographically dispersed, but can be considered and managed as one large entity.

Significant differences in software from different manufacturers are determined by the following factors:

- the used data model;
- the degree of coverage of the life cycle;

- built-in support for various architectures;
- data processing language capabilities.

Attention may be paid to the following main trends. Manufacturers offer complete solutions for creating data warehouses. Leading software companies in the field of design and development of information systems with databases launch their own programs for data storage systems and provide a full life cycle for the development and maintenance. They offer off-the-shelf embedded storage solutions significantly reducing the time for the design and development of data centers.

From the point of view of using software and hardware platforms, solutions in the field of creating decision support system (DSS) based on data warehouses can be divided into three classes:

- combination of finished products (solutions) of different companies without direct programming;
- using a complete closed chain of products (solutions) of one supplier company;
- using the contour of products (solutions) of one supplier's company with the addition to a closed chain of compatible products of third companies.

An example of a simple scalable solution can be based on the use of Crystal Enterprise and Crystal Reports (Business Objects) as end-user tools. The data center is implemented on DBMS Oracle, DB2, MS SQL Server or other similar components with open database connectivity (ODBC) interface or a direct access interface with Crystal Enterprise. Typically, classic data center architecture without data kiosks is used. For this decision, careful design of the data center structure and queries is of great importance. It is necessary to develop and create applications for data cleaning (or use the tools available from suppliers). The advantages of this solution are [14]:

- the amount of programming is minimized, since all stages are covered with ready-made box products;
  - time is reduced for the development and creation of data centers (by eliminating the labor-intensive process of writing programs);
  - the development time for a typical request is from 2 to 6 hours and for a typical report: 1-2 days;
  - such a solution is good for creating prototypes of data centers, because almost all the necessary requests and reports are processed;
  - an excellent environment for using non-core queries is created;
  - this solution is also perfect for creating virtual data centers.
- The disadvantages of this solution are:
- developing complex cross-querying system may require long time;
  - this solution is not suitable for complex analytical data processing, which needs the development of special applications for analysis.

A closed standard solution can be based on the use of a closed chain of products of one supplying company, for example Microsoft, Oracle, IBM. The advantages of this solution are:

- as a rule, all business areas are supported by ready-made services;
- the time of development and creation of a data center can be rigorously described and reasonably accurate;
- such a solution is good for creating data centers that are supposed to be used for a long time;
- such solutions are suitable for complex analytical data processing, which requires the development of special applications for analysis.

The disadvantages of closed standard solution are:

- high level cost of development and creation;
- requiring high-qualified staff to work with a set of products of the selected company.



### **Brief review of FPGA-based hardware accelerators**

This section emphasizes FPGAs because this technology serves as a core of many modern systems-on-chip incorporating high-speed processing blocks, advanced interfacing capabilities (Peripheral Component Interconnect – PCI, Advanced eXtensible Interface – AXI, etc.), and GPUs (Graphics Processing Units) [15-18]. Besides, FPGAs themselves can be seen as very powerful architectural solutions for high-performance accelerators [5]. The related system design challenges in ubiquitous computing environments and open-source hardware initiatives have also been discussed in invited and keynote talks [19, 20].

A detailed review of FPGA-based hardware accelerators for the *MapReduce* model is given in [21] with many details and examples of practical applications (see also section 4 below). FPGAs were invented by Xilinx in 1985 and they are the first truly programmable logic devices that could be configured after manufacturing. Nowadays they have quite impressive 35-years old history, showing evolution from relatively simple systems with a few tens of modest configurable logic blocks and programmable interconnects up to extremely complex chips combining the features of traditional FPGA with embedded multi-core processors and related peripherals, leading in this way to hardware Programmable Systems-on-Chip (PSoC). One of the most-recently unveiled reconfigurable accelerator platforms from Xilinx, Versal integrates FPGA technology with ARM CPU cores, Digital Signal Processing (DSP) and Artificial Intelligence (AI) processing engines and it is intended to be employed to process data-intensive workloads running datacenters [16-18]. Thus, they are appropriate for data processing. All necessary details about FPGAs are presented in [5, 22] and the book [22] has also been translated to Chinese in 2018.

Sections 1 and 4 list frequently used operations in the *MapReduce* model, namely sort, search, count, extraction of non-repeated names/values, discovering the most frequent items, Hamming weight computations and some others. One of the fastest sorters can be implemented in sorting networks, the majority of which use Batcher even-odd and bitonic mergers [23, 24]. A sorting network is a set of vertical lines composed of comparators that can swap data to change their positions in the input multi-item vector. The data propagate through the lines from left (from inputs) to right (to outputs) to produce the sorted multi-item vector on the outputs of the rightmost vertical line. Thus many-items data may be completely sorted in combinational circuits in such a way that input items are converted to sorted output items without applying clock cycles. There are also very efficient iterative solutions [25]. The book [5] describes effective hardware circuits including communication-time architectures allowing data to be completely sorted almost immediately as soon as a new item is received and, hence, minimizing communication overhead that is frequently pointed out as the main bottleneck in system performance. Thus, sorting illustrated in Fig. 1 may be implemented very fast.

The results of sorting may be further processed by the circuits proposed in [5]. Suppose there is a set of  $N$  sorted data items, which eventually includes repeated objects and we need the most frequently repeated item to be found. One possible solution for this problem is given in [5, pp. 98-102] (see Fig. 2 from [5]) where  $N-1$  parallel comparators form a binary vector  $V$ . The most frequently repeated item can be discovered if we find the maximum number of consecutive ones in the vector and take the item from any input of the comparators that forms the sub-vector with the maximum number of consecutive ones.

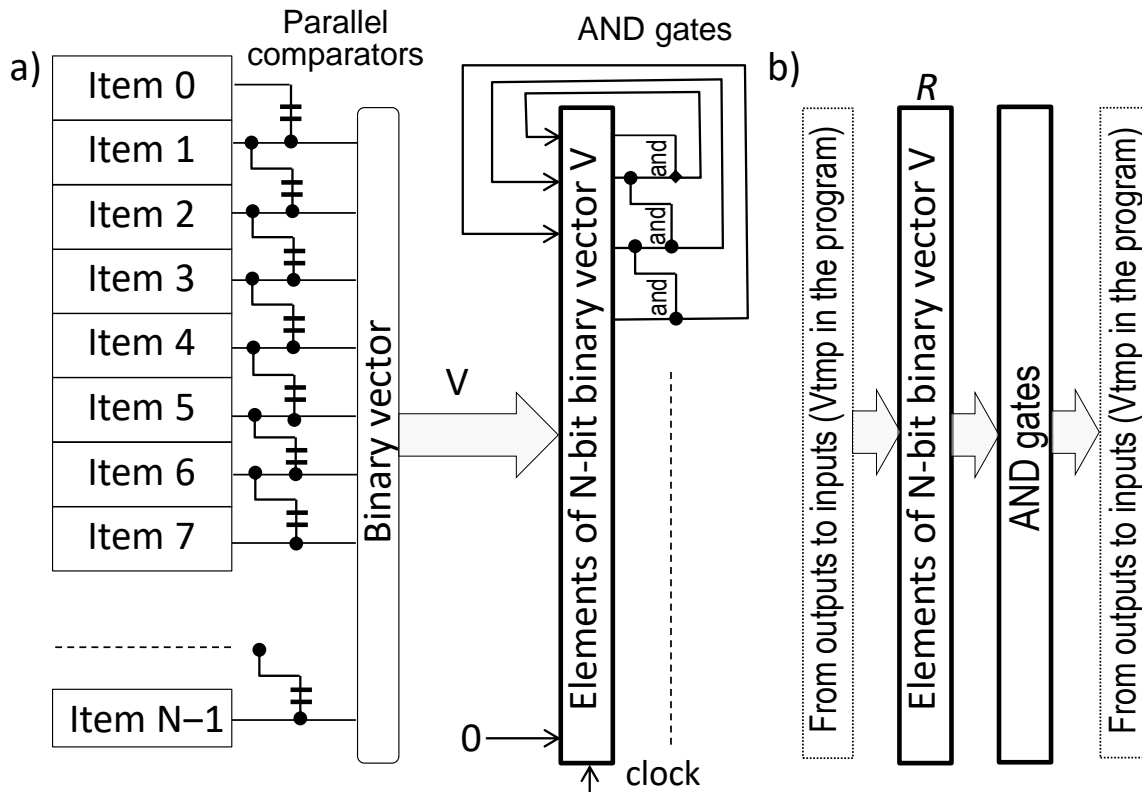


Figure 2. – Architecture for the most frequent item computation in a given sorted set of data items (a), copying the vector from the outputs of AND gates to the register R (b)

Much like it was done in [5], the considered circuits may be slightly modified, which permits to discover the number of repetitions for all the items or to find only such items that are repeated more than  $\kappa$  times where  $\kappa$  is a given threshold. Thus, sorting and shuffling may completely be implemented in hardware. It is shown in [5, pp. 34-36] that the number of consecutive ones can be found in a very fast FPGA LUT (Look-Up Table)-based combinational circuits. To understand the functionality of this circuit a Java program modeling the respective behavior is given. Note that computing frequencies of integers in large data sets may be effectively done with the address-based technique [5], proposed in [26].

Other useful hardware designs permit the Hamming weight of a binary vector to be computed that is helpful for the reducing procedure (see Fig. 1). There are a number of very effective solutions that have been proposed for solving such problems [27-30]. Comparisons of hardware and software implementations are done in the shown above publications. Communication mechanisms between different modules described in software programs may be formally transformed to the relevant hardware circuits using the methods [31] relying on the model of hierarchical finite state machine [32].

Note that the described above methods are applicable for data sets with limited number of items. The operations over such sets are executed in hardware very fast significantly improving the relevant throughput comparing to software implementations. Clearly the proposed methods cannot be applied directly to large sets that are common for big data. The main idea is working with blocks of data with limited complexity in such a way that large data sets are decomposed on blocks and completing the applied operations (within the blocks) makes possible to further handle large blocks of data instead of individual items. A good example is merge sorting in which the preliminary sorted blocks are merged in concurrently executed computational units. Thus, items in the blocks are sorted in very-high speed hardware accelerators with subsequent merging of the sorted blocks in parallel software. Interfacing circuits undoubtedly reduce the performance. However, deep parallelization and

optimization can be applied. Similar idea may be used for other types of data processing. Many examples can be found in [5, 22], which are supported by relevant hardware-level specifications and Java models. It is shown in [5] that such operations as Hamming weight computations may be completely implemented in on-chip (on-FPGA) solutions for hundreds of thousands bits with minimal delays from inputs to the results. Particular circuits and systems (used for the MapReduce model) have been discussed in [21] with examples. Note that open-source hardware may support operations over blocks [20], which is important for the future. The Gathering for Open Science Hardware (GOSH) is a diverse, global community working to enhance the sharing of publicly-available scientific technologies.

It should be noted that the techniques reviewed above can be used for complete on-chip solutions in embedded applications that (although are outside of the scope of big data) can be seen as very important for numerous practical needs [19]. Complete tasks (including software and hardware) may be solved on a single and relatively cheap microchips, such as [33], which provide very efficient interfaces for data exchange between processing units and programmable logic that have been evaluated in [34,35]. Other practical applications can be found in the scope of combinatorial optimization [36-38] based on mathematical models from [39, 40], such as transformations and operations over Boolean and ternary matrices. The major work in this area has been done by a corresponding member of Byelorussian Academy of science A.D. Zakrevskij. Three authors responsible for this section (A. Keevallik, I. Skliarova, V. Sklyarov) have been collaborating with A.D. Zakrevskij during many years, which is indicated in particular in the book [41]. An excellent paper [42] about scientific heritage of A.D. Zakrevskij was published by his colleagues in 2017. Many other practical applications [e.g. 3,43] may benefit from the proposed highly parallel hardware accelerators. The methods [5] permit LUTs contents to be generated from Java programs based on specification of the desired functionality in software. Thus, new fast hardware blocks can be designed easier because they may be carefully tested in software and then automatically retrieved and converted to hardware description language specifications. Information about network-based circuits in Russian can be found in [44].

The growing importance of hardware modules in different areas is emphasized in the keynote talk [20] where the following ideas are highlighted. In the contrary of open source software that is a well-known topic; open source hardware is a fairly new one. Indeed, as we approach the limits of available electronic systems in silicon, many specific designs in the past moved to generic standard chips. Best examples are electronic components like Arduino and Raspberry-Pi. In this domain we see an increased offer of high performance boards from companies like Intel. According to the executive summaries of the GOSH-community hardware forms a vital part of the scientific experimental process and the current supply chain limits access and impedes creativity and customization through high mark-ups and proprietary designs. Open source hardware enables designs to be shared, which often take advantage of modern digital fabrication techniques. Expanding the reach of this approach within academic research, citizen science and education has potential to increase access to experimental tools and ease their customization and reuse while lowering costs. In the DIY (Do It Yourself)-communities a growing number of people around the world are developing and using open source hardware in the context of the wider movement for Open Science that is a trend referred in proposals of the world-wide acting Open Science Hardware (OScH) community. However, in Europe such a movement is only slowly making progress. Thus, the status of the OScH approach has to be widely discussed to develop an open science hardware ecosystem which will play an important role for future smart system designs in different areas of the engineering community, including the basic directions of this conference. The referenced above reusable modules from [5, 22] can be seen as certain efforts done in such a field and targeted to high-performance data processing.

### ***Big Data Storage and Processing: Emerging Technologies and Platforms***

*Preliminary remarks.* The extremely fast grow of Internet Services, Web and Mobile Applications and advancements in the Pervasive, Ubiquity and Cloud Computing concepts have stimulated production of tremendous amounts of data partially available online (call metadata, texts, emails, social media updates, photos, videos, location, etc.). Even with the power of today's modern computers it is still big challenge for business and government organizations to manage, search, analyze, and visualize this vast amount of data turning it into actionable information and knowledge.

Just small part of this huge amount is structured (Databases, XML, logs) or semi-structured (web pages, email), over 90% of this information is unstructured, what means data does not have predefined structure and model. Generally, unstructured data is useless unless applying data mining and analysis techniques. At the same time, just in case if organizations are able to process and understand operational data, their efforts worth anything, otherwise they becomes useless. Data-Intensive computing, which is intended to address these problems, attracts great attention of research community over recent years yielding strong results. Data intensive computing framework is a complex system which includes hardware, software, communications, and Distributed File System (DFS) architecture.

Two key components of any data-intensive system are: Data Storage and Data Processing. So, which technologies, techniques, platforms, and tools are the best for Big Data storing and processing? How Big Data Era effects technological landscape? Why particular concepts and paradigms are better for Big Data? These and many other questions will be addressed in this research.

*Big Data Problem.* Moving 1 TB Data from one physical machine to another through network may take from 3 to several hours depending on specifications of platform and devices. Processing of 1 TB data may take from hours to forever depending on the Data format and how sophisticated the processing task is. It becomes clear that traditional platforms can't cope with continuously increasing amounts of data. A single machine, even exceptionally powerful (supercomputer), cannot process or even store all the data.

The only solution is to use specially designed platform/architecture that distributes data across large cluster of physical machines. But, when we deal with huge amounts of data and many computers, the following questions can't be ignored:

- How to split Data across many machines?
- How to minimize or avoid Data movement? This is because moving Data over network is expensive.
- Having many nodes/machines in any cluster increases failures. How to deal with failures?

Hadoop is open source common platform that combines two main tasks of any operating system: storing and processing data. Unlike to traditional systems, Hadoop accomplishes these tasks towards Big Data. The popularity of Hadoop increases day by day, because of simplicity, scalability and affordability that it provides for, thanks to its distributed architecture. Although, the Hadoop Core consists of two main components (Hadoop Distributed File System – HDFS and MapReduce) and has a limited functionality, thanks to many other components available in the Hadoop Ecosystem under Apache License, this platform can cover almost any demand to manage and process data regardless to its size and format [45].

*Big Data Storage with Hadoop Distributed File System – HDFS.* Data storage domain is in the center of the big data scientific studies and technological practices focused on machine learning, data mining and business intelligence. Storing large amounts of data allows many engineering problems to be efficiently solved such as: search and analyze the patterns, optimization for fastest access in business processes, etc.

HDFS is one of the most well-known examples of an open source distributed processing framework that manages data storage for big data applications. HDFS's main advantage is the cost-

efficient storage capacity for extremely large amounts of data and possibility to transfer data at high speed to user applications.

Typically organized in large clusters, user applications run on thousands of spread servers directly connected to the storage devices. Distributing data storage and computing resources between multiple servers allows scaling based on low-end hardware (if necessary) and creating an economically profitable system of any size. Hadoop clusters enable computing resources, storage capacity and input/output channel bandwidth to be adjusted by modifying the number of workstations in distributed network. Hadoop systems are very widely used to handle various forms of structured and unstructured data, known as Data Lakes [46], giving users more flexibility for collecting, processing, analyzing and managing data than relational databases and data warehouses provide.

Hadoop is the fundamental Big Data storage and processing technology. It is one of the most popular projects of the Apache Software Foundation. It is supported by a full stack variety of related technologies, such as:

- YARN (Yet Another Resource Negotiator) – the system for task planning and cluster management providing sharing, scaling, and reliability of distributed applications.
- Hadoop *MapReduce* – the platform for programming and performing distributed *MapReduce* [47,48] computing using a large number of nodes organized in a cluster.
- Apache Hive – the data warehouse software project built on top of Apache Hadoop for providing data query and analysis.
- Apache Spark – the open-source distributed general-purpose cluster-computing framework.
- Sqoop – the tool designed to efficiently transfer large amounts of data between Hadoop and structured DBMSs (for example, relational) in both directions.

The example of integration of multiple Big Data Processing Tools for Image Classification is considered in [48,49]. DataStore functionality built into MATLAB is used to develop and debug algorithms with a piece of data and then apply them on a distributed Hadoop cluster within the Hadoop MapReduce environment on the complete set of data.

The HDFS file system stores metadata and application data separately on a metadata server (NameNode) and application data servers (DataNode). All servers communicate with each other through protocols based on the transmission control protocol (TCP). Unlike the Lustre and parallel virtual file system (PVFS), the HDFS application data servers (DataNode) do not rely on such data protection mechanisms as redundant array of independent disks (RAID) to increase the reliability of data storage. On the contrary, similarly to the Google file system (GFS), the contents of a file is distributed and duplicated among many application data servers for its consistent storage. Ensuring the reliability of data storage, the HDFS strategy allows to take advantage of multiplying the bandwidth of the data transfer channel, and so when required, it becomes possible to carry out calculations in parallel with the delivery of the necessary data.

HDFS uses the classic UNIX tree structure of directories, user concept with read-write-execute operations of rights granting, and even a similar set of console commands. The data is divided into blocks (usually 64 MB or 128 MB), for each file the NameNode stores its path, a list of blocks with file data and their replicas.

The NameNode ensures that each block always has a given number of copies. It checks the fact that there are not enough copies or excessive copies of a block at the time of delivery of the block report from the DataNodes. In case the block has extra copies, the NameNode selects which copy to delete. The HDFS logic is designed to maintain the number of physical racks on which copies are stored, and to remove the copy from the DataNode on which the least amount of available disk space is available. The purpose of this policy is to balance the use of storage devices of data servers of application data without reducing the availability of blocks.

When the number of copies of a block becomes insufficient, the block enters the priority replication queue. A block with only one copy has the highest priority, and a block with several copies that make up more than two-thirds of the required replication volume has the lowest priority. A background program thread periodically scans the beginning of the replication queue to decide where to place the new block copies. The block replication process uses a simple policy for placing new copies of blocks. If one copy of the block is available, HDFS places the next copy on the server from another rack. In the case when the block has two available copies, if two existing copies are on the servers of one rack, the third copy is created on the server from the other rack; otherwise, the third copy is hosted on another server of the same rack used by the existing copy. In this case, the goal of the policy is to reduce the cost of resources for creating new copies.

User applications access the file system using the HDFS client, which is a dedicated library that exports the HDFS file system interface. Much like the most traditional file systems, HDFS supports reading, writing and deleting files, as well as creating and deleting directories. The user describes files and directories supplying paths in the namespace. The user application does not need to worry about the fact that metadata and file data from the file system are stored on different servers or that the blocks have many copies.

When the application reads from a file, the HDFS client first queries the metadata server for a list of application data servers that store copies of the data blocks of the required file. The list is sorted based on the distance to the client within the network topology. The client connects directly to the application data server and requests the transfer of the necessary block. When a client writes, it first requires the metadata server to select application data servers to store copies of the first block of the file. A client organizes a channel between several servers and sends data. When the first block is transmitted, the client requests the selection of the following application data servers for storing copies of the next block. A new channel is organized and the client sends data to the next block. The choice of application data servers for each block is likely to be different.

An important area of current research is the evaluation of operational efficiency of solutions and algorithms for processing big data. Given the complexity of the interactions architecture in the distributed systems, increasing the speed of data exchange in the file system is one of the main incentive targets to boost the overall Big Data implementation.

Examples of modeling research for assessing the efficiency of access to big data storage systems are considered in [50-53].

*Big Data Parallel Processing with MapReduce.* MapReduce is a distributed, large-scale data processing paradigm (also known as parallel programming concept) that was initially developed by Google [54]. The idea of MapReduce is very much similar to high order functions in functional programming languages. For example, Map and Fold (Reduce) functions in Haskell programming language.

Leveraging Big Data technologies, hundreds of SMEs and tech giants like Google, Amazon, Facebook, Twitter, etc. use Hadoop Ecosystem as a testing and production environment building their critical services on the top of this framework [55].

MapReduce is a programming concept designed to build algorithms for distributed computing (Figure 3).

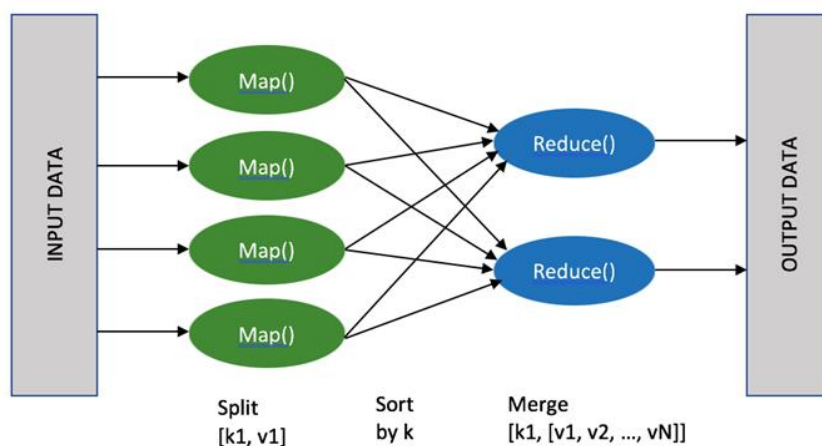


Figure 3. – MapReduce Architecture

When someone develops an algorithm for distributed system that consists of many connected nodes, there appears a need to deal with many sophisticated issues not having direct relations to the problem intended to be solved. These issues are the followings:

- How to split a job to tasks to be distributed over computing nodes?
- What happens if any of nodes that was assigned with a task, fails at the mid of computation?
- How to distribute data/messages between nodes in the most effective way?
- How to minimize data flow between nodes of the distributed system?

And many others...

It is known that *MapReduce* is not equally well for any kind of distributed algorithms. Developers of *MapReduce* had a specific list of tasks (search, sort, count, etc.) for those this concept is tuned. The framework that implements *MapReduce* concept is optimized to accomplish these tasks in the most effective way, at the same time taking care about all low-level common issues we mentioned earlier. The problem with this concept is: in order to run specific task in the framework, it should be ensured that the task fits into frames of *MapReduce*. In other words, task must follow all prerequisites of *MapReduce* [56]. As it has been mentioned *MapReduce* is far away of being panacea for parallel programming tasks. It is designed for batch processing and is not efficient for the certain types of massive processing like:

- stream and Real-time data processing;
- iterative data processing;
- low-latency data processing.

*Understanding MapReduce Lifecycle in detail.* MapReduce is suitable for large problems that can be decomposed into sub-tasks not having dependency to each other. The introduced constraint enables almost infinite scalability. Depending on the scale, the problem (data) can be parallelized involving from several to thousands computing nodes. Besides, Hadoop provides fault-tolerance not just at the level of HDFS, but also at the level of MapReduce taking care of possible failure of nodes in each cluster. From programmer's point of view, it is quite easy to implement parallel programming on the top of MapReduce framework. Generally, a programmer deals with Map- and Reduce-functions, and framework itself takes care about all possible system-level operations, including fault-tolerance, data distribution, etc. Any task that is intended to be run using MapReduce framework has to be divided into three essential functions [57] (see Fig. 4): 1) Map; 2) Sort/Shuffle/Merge; 3) Reduce.

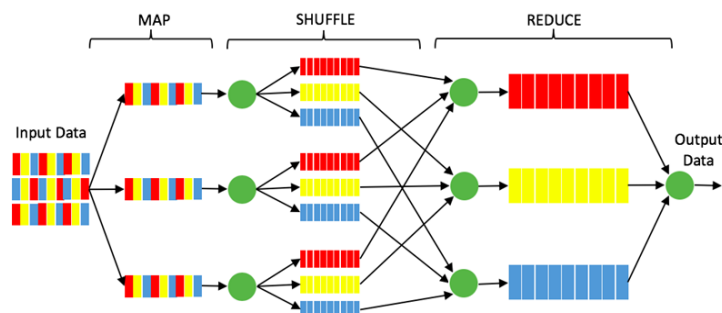


Figure 4. – Phases of MapReduce Lifecycle

The *Map*-function turns the input data that usually comes as blocks from HDFS into key-value pairs and applies specific functions to each element of input data. The *Map* function can be compared to mapping in functional programming languages (for example Haskell) `map func [e1, e2, e3, ..., eN]`, where function `func` is applied to each element of the list generating new list. The second task (Sort/Shuffle/Merge) sorts (or shuffles according to predefined conditions) pairs from all Mappers preparing data for the next task. The Reduce-function extracts each unique key and associates it with all values (similar to group by key operation) prepared by Map-function. It is important to notice that, while MapReduce runs each of Map-Shuffle-Reduce functions, it follows to requirement of Data Locality. Doing so, *Map*-function is executed on particular node where one of 3 data-block replicas (default replication factor in HDFS) resides avoiding any need for data transfer. In the same way, Shuffle- and Reduce-functions are executed on the node where Map-function stores its output. As a result, computations (in the form of software code) that implement those functions are transferred to where data is, while data itself remain immovable [58].

There are three categories of data locality with different level:

- Node-level locality - processing takes place on the same node where data resides;
- Rack-level locality - processing takes place on the node different from that where data resides, but on the node that belongs to the same rack;
- Cluster-level locality - processing takes place on the node located on different rack.

The fundamental features of MapReduce that made this paradigm de-facto standard for large-scale data processing are the followings:

- Move computation to the data rather than transferring data. Transferring data, especially huge amounts of data, over network is timely and costly process. For example, to move 1 TB file over 1 GB Ethernet network will take between 3 to 6 hours depending on many parameters of computers and network (performance of hard disks, performance of computers, network traffic usage, etc.).

- MapReduce deals with data stored in the Hadoop's native file system HDFS. HDFS with unique features of high availability, fault-tolerance, distributed storage and replication is available for all phases of MapReduce processing. The input, intermediate and output data are stored on distributed HDFS.

- MapReduce manages all resources available in a cluster. It is particularly true taking into account MapReduce transformation into YARN (Yet Another Resource Negotiator). YARN does not just manage resources (memory and processing power) among applications and jobs, but also responsible for provision of fault-tolerance, creating new containers, re-assigning jobs to other machines in case of failure.

*Logical Model of MapReduce.* To understand the logical model of MapReduce framework, let's look to the following example, which can be considered as starting point for implementation of Natural Language Processing (Text Analytics) application on the top of MapReduce. Let  $W$  be large textual dataset of length  $n$ . The  $i^{th}$  word in the sequence  $W$  is depicted as  $\omega_i$ . Since MapReduce expects



the input to be a list of key-value pairs, the  $W$  is presented as the list of  $n$  pairs  $\langle i, \omega_i \rangle$ . The MapReduce lifecycle can be introduced as following process:

- Each word in textual dataset is depicted as the key, and the position of word as the value. So, the *Map*-function can be defined as  $\mu_i(\langle i; \omega_i \rangle) = \langle \omega_i; i \rangle$ .
- After grouping by key (particular word) as an output of *Map*-function, *Reduce*-function will receive the unique key (word) and the list of all positions of this word appears at. So, the *Reduce*-function can be defined as  $r_i(\langle \omega_i; \{\rho_1, \rho_2, \dots, \rho_m\} \rangle)$ .
- All pairs with the same key will be submitted to the same reducer (see Fig. 3) [59]:

$$r(\langle \omega_i; \{\rho_1, \rho_2, \dots, \rho_m\} \rangle) = \langle \omega_i; \sum_i \rho_i \rangle$$

### ***Operations over Data***

There are many frequently executed operations over data and just a part of them from [5] is listed below:

- Retrieving from a given set items with the maximum and/or minimum values.
  - Retrieving from a given set items with the maximum and/or minimum values of given parameters, such as the Hamming weight (HW) of lines/column of binary/ternary matrices, the maximum and/or minimum difference in the HWs in a given set of binary/ternary vectors, etc.
  - Finding the most repeated items.
  - Testing if in a given data set there are no repeated items, i.e. all values of the items are unique.
  - Ordering repeated items that satisfy some predefined conditions, for example they must be repeated not less than a given threshold.
  - Decomposing a given set in intervals  $0, \dots, N_{sets}-1$  and solving the listed above tasks in each separate interval.
  - Finding extreme values in the set of intervals, for example, retrieving all minimum values in each interval and finding the maximum value among the retrieved minimum values (finding the greatest minimum); similarly the smallest maximum can be requested to be found.
  - Extracting sorted maximum/minimum subsets from a given set.
  - Filtering data, i.e. extracting subsets with values that are within the defined limits or satisfy some characteristics.
  - Dividing data items into intervals or subsets  $\Theta_0, \dots, \Theta_{E-1}$  and finding the minimum/maximum/average values in each subset, sorting each subset, or solving some other problems partially described in the next point.
  - Operations with the subsets  $\Theta_0, \dots, \Theta_{E-1}$  such as union  $\Theta_i \cup \Theta_j$  ( $x : x \in \Theta_i$  or  $x \in \Theta_j$ ), intersection  $\Theta_i \cap \Theta_j$  ( $x : x \in \Theta_i$  and  $x \in \Theta_j$ ), difference  $\Theta_i - \Theta_j$  ( $x : x \in \Theta_i$  and  $x \notin \Theta_j$ ). Some other operations may also be considered, for example, removing from a given subset all repeated or indicated items, etc.
  - Locating and ordering repeating relationships between different objects, for example ordering the most/less frequent items or items satisfying some other criteria.
  - Removing all duplicated items from a given set.
  - Multi-targeted sorting, i.e. ordering the same set by different criteria.
  - Solving the problems indicated above for matrices (for rows/columns of matrices).
- The above operations may be needed when big data are processed.

### ***Practical Application***

The described below practical application is taken from an experience in Kazakhstan. When designing an information system for electoral processes, one must take into account the need to

process a large amount of streaming data at all stages of voting [60]. In this regard, it is important to consider scalable, flexible and fault-tolerant technologies. Thus, the use of micro-service architecture for electronic voting systems has become popular. The concept of «micro-service architecture of information systems» has gained popularity over the past few years, as a description of the method of designing an information system in the form of a set of independently deployable services. While there is no exact description of this architectural style, there is a certain common set of characteristics: organization of services around business needs, automatic deployment, transfer of logic from the message bus to receivers (endpoints) and decentralized control over languages and data.

The main reason for using services instead of libraries is independent deployment. If you are developing an application consisting of several libraries working in one process, any change in these libraries leads to redesign of the entire application. But if the application is divided into several services, then the changes affecting any of them will require a reconstruction of only the changed service.

If the micro-service of one of the modules stops responding as a result of an accident, its clients should be immediately redirected to the backup one. So-called message queues are often used to control the flow of requests. The use of a message queuing system ensures reliability and the necessary performance in data collection and processing. A message queue is a form of an asynchronous communication between modules (services) of a system. Messages are queued until they are processed and deleted. Each message is processed only once and only by one consumer. Message queues can be used for the separation complex processes, to buffer or organize batch handling, and to smooth out peak loads. A queue can be used by multiple sources and recipients, but each message is processed by one recipient only once. For this reason, this type of messaging is often referred to as one-to-one or direct.

A micro-service with message queues for the electronic voting system can use the data collection module. Figure 5 illustrates the operation algorithm of such a data acquisition micro-service module.

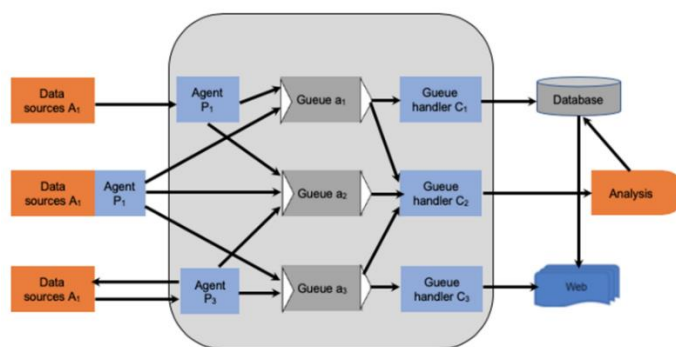


Figure 5. – The algorithm of the micro-service data acquisition module: A - data sources; P - agents to set the data in the queue; a - message queues; C - queue handlers

The module is a set of distributed agents that is responsible for the direct collection of data, the initial processing of data, and bringing the data to the standard form of the internal ecosystem of the electronic voting platform. The source of information for the data collection module, depending on the functionality and type of user, can be client personal computers, tablets or smartphones of voters (A1, A3), voting terminals and electronic ballot boxes (A2). Input agents (P1, P3) receive appropriate data streams or take portion from data sources (P2). Data queues and output queue processing agents are formed using special algorithms that reflect the specifics of information processes ensuring election confidentiality. The received big data is transferred to the next level of the system to the corresponding modules. The presence of many agents solves the problems of various

data formats, because it allows for initial data processing the computational load on the analysis unit to be reduced. The considered module of the electronic voting system can use Apache Kafka platform developed by LinkedIn Company subsequently transferred to Apache Foundation [61]. Apache Kafka is a disseminated streaming platform, as well as a very fast distributed message broker that allows the required functionality to be implemented. Kafka agents (producer and consumer) can be implemented in various modern programming languages, such as Python, Java, Scala. The Kafka architecture is based on the Publisher-Subscriber concept.

Let us compare the functionality of Kafka with its analogues. The functionality is limited to only one pattern, thereby being inferior in functionality to many other platforms, such as the RabbitMQ platform [62]. However, other platforms delete messages after successful delivery, while Kafka stores them for any required period, by default a week, thus increasing message delivery guarantees in case of failure of subsequent dependent modules, which is very important for the electronic voting system. Also, the requirements for the network bandwidth and equipment resources on which Kafka will be launched are lower than those of the popular ActiveMQ and RabbitMQ message brokers, as well as the bandwidth figures: 20 Kbit / s messages per second for Kafka versus 100 Kbit / s for RabbitMQ. In terms of clustering, Kafka has an advantage over analogues ease of customization and scalability. Since one Kafka broker is a cluster, to add a second instance of Kafka is enough to inform the first instance about the presence of the second. Kafka saves messages to disk, which allows synchronizing end layouts between cluster nodes instead of data replication, while maintaining the logic of processing topics. Also, the output throughput grows exponentially, since a single topic can be processed by several brokers [61].

The joint use of this solution based on the micro-service architecture on the Kafka platform and hardware accelerators based on FPGAs with the MapReduce model [20] (see also section 3 above) can be a breakthrough option in the automation of elective processes.

### ***Conclusion***

The paper combines expertise of a large group of authors actively working in different areas that are important for big data processing. Various sections of this paper briefly describe their results. Distribution of the material between the authors is shown at the end of section 1. Additional information may be taken from publications of the authors shown in the list of references given below.

### ***References***

- [1]. High Performance Computing for Big Data. Methodologies and Applications. Edited by Chao Wang. CLR Press by Taylor & Francis Group, 2018.
- [2]. C.L.P. Chen and C.Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: a survey on big data", Information Sciences, vol. 275, 2014, Pp. 314-347.
- [3]. B. Parhami, "Computer architecture for big data", in Encyclopedia of Big Data Technologies, S. Sakr, A. Zomaya (eds.), Springer, 2018.
- [4]. MapReduce Tutorial – Fundamentals of MapReduce with MapReduce Example. Available at <https://www.edureka.co/blog/mapreduce-tutorial>.
- [5]. I.Sklyarova, V.Sklyarov. FPGA-based Hardware Accelerators. Springer, Switzerland, 2019, 245 p.
- [6]. Baker Z.K., Prasanna V.K. An Architecture for Efficient Hardware Data Mining using Reconfigurable Computing Systems. In: Proceedings of the 14th Annual IEEE Symposium on Field-Programmable Custom Computing Machines – FCCM'06, Napa, USA, April 2006, Pp. 67-75.
- [7]. Ivor Horton, Beginning C - From Novice to Professional Apress, 2018, 614 p.
- [8]. Ivor Horton, Using the C++ Standard Template Libraries, Apress, 2015, 489 p.
- [9]. Ivor Horton, Beginning Java, Wiley Publishing, 2011, 1112 p.
- [10]. Калимолдаев М.Н., Утепбергенов И.Т., Яворский В.В., Ахмедиярова А.Т., Ключева Е.Г. Организация хранилищ данных для Smart систем городского общественного транспорта. Алматы: Институт информационных и вычислительных технологий КН МОН РК, 2019, 119 с.
- [11]. Bruce Nelson, Saurabh Gupta and others. Oracle Big Data Handbook. Oracle Press, 2013, 464 p.
- [12]. Brian Mitchell, Christopher Price, Dan Clark, John Welch, James Rowland-Jones, Adam Jorgensen. Microsoft Big Data Solutions. Wiley, 2014, 408 p.

- [13]. Деян Сарка, Матия Лах, Грега Йеркич, Microsoft SQL Server 2012. Реализация хранилищ данных. Учебный курс Microsoft Русская редакция, 2012, 816 с.
- [14]. Су Кеннет, Анналин Ын. Теоретический минимум по Big Data. Всё что нужно знать о больших данных. Питер, 2019, 208 с.
- [15]. M. Santarini, "Xilinx 16nm UltraScale+ devices yield 2-5X performance/watt advantage", XCell Journal, issue 90, 2015, Pp. 8-15.
- [16]. Xilinx Press Releases (2018) Xilinx Unveils Versal: The First in a New Category of Platforms Delivering Rapid Innovation with Software Programmability and Scalable AI Inference. <https://www.xilinx.com/news/press/2018/xilinx-unveils-versal-the-first-in-a-new-category-of-platforms-delivering-rapid-innovation-with-software-programmability-and-scalable-ai-inference.html>.
- [17]. Xilinx Inc. (2018) Versal: The First Adaptive Compute Acceleration Platform (ACAP). White Paper: Versal ACAPs. [https://www.xilinx.com/support/documentation/white\\_papers/wp505-versal-acap.pdf](https://www.xilinx.com/support/documentation/white_papers/wp505-versal-acap.pdf). Xilinx Inc. (2018) Xilinx AI Engines and Their Applications. White Paper: AI Engines. [https://www.xilinx.com/support/documentation/white\\_papers/wp506-ai-engine.pdf](https://www.xilinx.com/support/documentation/white_papers/wp506-ai-engine.pdf).
- [18]. Manfred Glesner, Thomas Hollstein, Tudor Murgan: System Design Challenges in Ubiquitous Computing Environments, in: Proc. of the Intl. Conf. on Microelectronics (ICM), S.11-14, 2004. Invited Talk.
- [19]. Manfred Glesner, "The rebirth of hardware: Open-source hardware initiatives for IoT- and MINT-based developments". Proceeding of the 2nd International Symposium on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC 2017), July 12-14, 2017, Madrid, Spain (<http://www.cei.upm.es/reco soc17/keynotes.html>).
- [20]. V. Sklyarov, I. Skliarova, I. Uterbergenov, A. Akhmediyarova, "High-Performance Information Processing in Distributed Computing Systems", International Journal of Innovative Computing, Information and Control, vol. 15, no. 1. 2019, Pp. 321-335.
- [21]. Sklyarov V, Skliarova I, Barkalov A, Titarenko L. Synthesis and Optimization of FPGA-Based Systems, Springer, Switzerland, 2014, 432 p.
- [22]. Batcher K.E. Sorting networks and their applications. In: Proceedings of AFIPS Spring Joint Computer Conf., USA, 1968.
- [23]. Aj-Haj Baddar S.W., Batcher K.E. Designing Sorting Networks. A New Paradigm. Springer, 2011, 132 p.
- [24]. Sklyarov V., Skliarova I., "High-performance implementation of regular and easily scalable sorting networks on an FPGA", Microprocessors and Microsystems, vol. 38, no. 5, July 2014, Pp. 470-484.
- [25]. Sklyarov V, Skliarova I, Mihhailov D, Sudnitson A. Implementation in FPGA of Address-based Data Sorting. In: Proceedings of the 21st International Conference on Field Programmable Logic and Applications, Crete, Greece, 2011, Pp. 405-410.
- [26]. Sklyarov V, Skliarova I. Design and implementation of counting networks. Computing, 2015, 97(6):557-577.
- [27]. Sklyarov V, Skliarova I. Digital Hamming weight and distance analyzers for binary vectors and matrices. International Journal of Innovative Computing, Information and Control, 2013, 9(12):4825-4849.
- [28]. Sklyarov V, Skliarova I, Silva J. On-Chip Reconfigurable Hardware Accelerators for Popcount Computations. International Journal of Reconfigurable Computing, 2016:8972065.
- [29]. Sklyarov V, Skliarova I. Multi-core DSP-based Vector Set Bits Counters/Comparators. Journal of Signal Processing Systems, 2015, 80(3):309-322.
- [30]. Sklyarov V, Skliarova I. Hardware Implementations of Software Programs Based on HFSM Models. Computers & Electrical Engineering, 2013, 39(7):2145-2160.
- [31]. V.Sklyarov. Hierarchical Finite-State Machines and Their Use for Digital Control. IEEE Transactions on VLSI Systems, 1999, Vol. 7, No 2, Pp. 222-228.
- [32]. Xilinx Inc. (2018) Zynq-7000 All Programmable SoC Technical Reference Manual. [https://www.xilinx.com/support/documentation/user\\_guides/ug585-Zynq-7000-TRM.pdf](https://www.xilinx.com/support/documentation/user_guides/ug585-Zynq-7000-TRM.pdf).
- [33]. Silva J, Sklyarov V, Skliarova I. Comparison of On-chip Communications in Zynq-7000 All Programmable Systems-on-Chip. IEEE Embedded Systems Letters, 2015, 7(1):31-34.
- [34]. Sklyarov V, Skliarova I, Silva J, Sudnitson A. Analysis and Comparison of Attainable Hardware Acceleration in All Programmable Systems-on-Chip. In: Proceedings of the Euromicro Conference on Digital System Design - Euromicro DSD'2015, Madeira, Portugal, August, 2015, Pp. 345-352.
- [35]. Skliarova I, Ferrari A.B. Reconfigurable Hardware SAT Solvers: A Survey of Systems. IEEE Transactions on Computers, 2004, 53(11):1449-1461.
- [36]. Skliarova I, Ferrari A.B. A Software/Reconfigurable Hardware SAT Solver. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2004, 12(4):408-419.
- [37]. Skliarova I, Ferrari A.B. The Design and Implementation of a Reconfigurable Processor for Problems of Combinatorial Computation. Journal of Systems Architecture. Special Issue on Reconfigurable Systems, 2003, 49(4-6):211-226.
- [38]. A. Zakrevskij, Y. Pottosin, L. Cheremisinova. Combinatorial Algorithms of Discrete Mathematics. Edited by Andres Keevallik. Tallinn, TUT Press, 2008, 193 p.

- [39]. A.Zakrevskij, Y.Pottosin, L.Cheremisinova. Optimization in Boolean Space. Edited by Andres Keevallik. Tallinn, TUT Press, 2009, 241 p.
- [40]. А.Д. Закревский. В Томском университете, Минск, 2014, 224 с.
- [41]. П.Н.Бибило, Ю.В.Поттосин, Л.Д.Черемисинова. О научном наследии члена-корреспондента А.Д.Закревского. Информатика, N 1 (январь-март), 2017, стр. 112-124.
- [42]. Chee C.H., Jaafar J, Aziz I.A., Hasan M.H., Yeoh W. Algorithms for frequent itemset mining: a literature review. Artificial Intelligence Review, 2018.
- [43]. Скляров В.А., Склярова Ю.В., "Обработка данных в программно-аппаратных системах логического управления на основе поисковых сетей", Автоматика и Телемеханика, no. 1, 2017, Pp. 121-136.
- [44]. Abzetdin Adamov, Large-scale Data Modelling in Hive and Distributed Query Processing using MapReduce and Tez, DiVAI 2018 - Distance Learning in Applied Informatics, 02 - 04 October, 2018, Štúrovo, Slovakia.
- [45]. Walker, Coral; Alrehamy, Hassan. "Personal Data Lake with Data Gravity Pull". IEEE Fifth International Conference on Big Data and Cloud Computing, 2015) Pp. 160–167.
- [46]. Ghazi M.R., Gangodkar D. Hadoop, MapReduce and HDFS: A Developers Perspective. Procedia Computer Science. 2015 Jan 1;48:45-50.
- [47]. Paramonov I.Yu., Smagin V.A., Kosykh N.E., Khomonenko A.D. Methods and models for the study of complex systems and big data processing. Ed. V.A.Smagin and A.D. Khomonenko. St. Petersburg Publishing House "Lan", 2020. 236 p.
- [48]. Nikita E. Kosykh, Anatoly D. Khomonenko, Alexander P. Bochkov, Anatoly V. Kikot. Integration of Big Data Processing Tools and Neural Networks for Image Classification. Proceedings of Models and Methods of Information Systems Research Workshop in the frame of the Betancourt International Engineering Forum. St. Petersburg, Russian Federation, Dec. 4-5, 2019. CEUR Workshop Proceedings. Vol-2556, Pp. 52-58.
- [49]. A.D. Khomonenko, A.G. Basyrov, V.P. Bubnov [et al.]. Models and methods of research of information systems. Edited by A. D. Khomonenko. St. Petersburg: Publishing House "Doe". Russia. 2019. - 204 p.
- [50]. A.D.Khomonenko; S.I.Gindin; Khalil Maad Modher. A cloud computing model using multi-channel queuing system with cooling. XIX IEEE International Conference on Soft Computing and Measurements (SCM). 2016, Pp 103 - 106.
- [51]. V.A.Lokhvitskii, A.D.Khomonenko and M.A.Bol'shakov, On the Construction of a Cybervisor for the Intelligent Monitoring and Control of Data Centers Automatic Control and Computer Sciences, 2019, Vol. 53, No. 8, pp. 870–873. © Allerton Press, Inc., 2019. Russian Text © The Author(s), 2019, published in Problemy Informatsionnoi Bezopasnosti, Komp'yuternye Sistemy.
- [52]. A.D.Khomonenko, S.I.Gindin. Performance evaluation of cloud computing accounting for expenses on information security. 18th Conference of Open Innovations Association and Seminar on Information Security and Protection of Information Technology (FRUCT-ISPIT). 18-22 April 2016, Pp. 100-105.
- [53]. Jeffrey Dean and Sanjay Ghemawat, MapReduce: Simplified Data Processing on Large Clusters, Google Inc.
- [54]. Apache Wiki, Powered by Apache Hadoop, <https://cwiki.apache.org/confluence/display/HADOOP2/PoweredBy>.
- [55]. Bala, M., Boussaid, O., & Alimazighi, Z. A Fine-Grained Distribution Approach for ETL Processes in Big Data Environments. Data & Knowledge Engineering, 2017, 111, Pp. 114–136.
- [56]. Sterling, T., Anderson, M., & Brodowicz, M. MapReduce. High Performance Computing, 2018, Pp. 579–589.
- [57]. Lee, S., Jo, J.-Y., & Kim, Y. Hadoop Performance Analysis Model with Deep Data Locality. Information, 2019, 10(7), 222 p.
- [58]. Howard Jeffrey Karloff, Siddharth Suri, Sergei Vassilvitskii, A Model of Computation for MapReduce, SODA '10: Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete algorithms, January 2010, Pp. 938–948.
- [59]. I.T. Utepbergenov, A.I. Buranbaeva. Анализ международного опыта применения информационно-коммуникационной технологоо в выборных процессах. Almaty, 2018, 78 p.
- [60]. Humphrey P. Understanding When to use RabbitMQ or Apache Kafka: <https://content.pivotal.io/blog/understanding-when-to-use-rabbitmq-or-apache-kafka.pdf>.
- [61]. RabbitMQ vs. Kafka vs. ActiveMQ Technoblog StackShare.io. <https://stackshare.io/stackups/activemq-vs-kafka-vs-rabbitmq.pdf>.

UDK 004.75

## CLOUD SERVERS USAGE EFFICIENCY AND COST OPTIMIZATION



**I. A. Trubin**

*PhD, IT Manager, Capital One Bank, USA  
Director, Computer Measurement Group, USA*

*Computer Measurement Group, Capital One Bank, USA  
E-mail: Igor@Trub.in*

### **I.A. Trubin**

*Igor Trubin has started in 1979 as an IBM/370 system engineer. In 1986 he got his PhD. in Robotics at St. Petersburg Technical University (Russia) and then worked as a professor teaching CAD/CAM, Robotics for about 12 years. He published 30 papers and made several presentations for conferences related to the Robotics and Artificial Intelligent fields. In 1999 he moved to the US and worked at Capital One bank as a Capacity Planner. His first CMG.org paper was written and presented in 2001. The next one, "Exception Detection System Based on MASF Technique," won a Best Paper award at CMG 2002 and was presented at UKCMG 2003 in Oxford, England. He made other tech. presentations at IBM z/Series Expo, Southern and Central Europe CMG and ran several workshops covering his original method of Anomaly and Change Point Detection ([www.Performalist.com](http://www.Performalist.com)). He is an author of the online class "Performance Anomaly Detection" (<https://cmg1.teachable.com>). After working more than 2 years as the Capacity team lead for IBM, he had worked for SunTrust Bank for 3 years and then at IBM for 2+ years as Sr. IT Architect. Now he works for Capital One bank as IT Manager at the Cloud Engineering department and since 2015 he is a member of CMG.org Board of Directors. He runs his tech blog at [www.Trub.in](http://www.Trub.in)*

**Abstract.** The public cloud has unlimited capacity if you have an unlimited budget to buy it. But the reality is that budgets are never truly unlimited, and one needs to do rightsizing of the cloud objects to stop wasting money on unused or unneeded cloud capacity.

During this session, Igor Trubin will discuss methods of tracking and reporting on cloud usage. He will discuss how cloud cost optimization can be done to show the efficiency of individual applications, lines of businesses (LOB), or the entire infrastructure. Lastly, participants will learn how to access and plan for organic cloud growth and cost increases.

**Keywords:** cloud, servers.

### ***Problem statement***

–The public cloud has unlimited capacity if you have unlimited budget to buy it. But if the budget is tight, one needs to do Rightsizing of the cloud objects to stop wasting money on capacity that are not used.

–TOOLS to analyze cost and get Rightsizing recommendations are:

–AWS (*Amazon Cloud*)

–Amazon EC2 resource optimization recommendations/\_AWS Cost Management

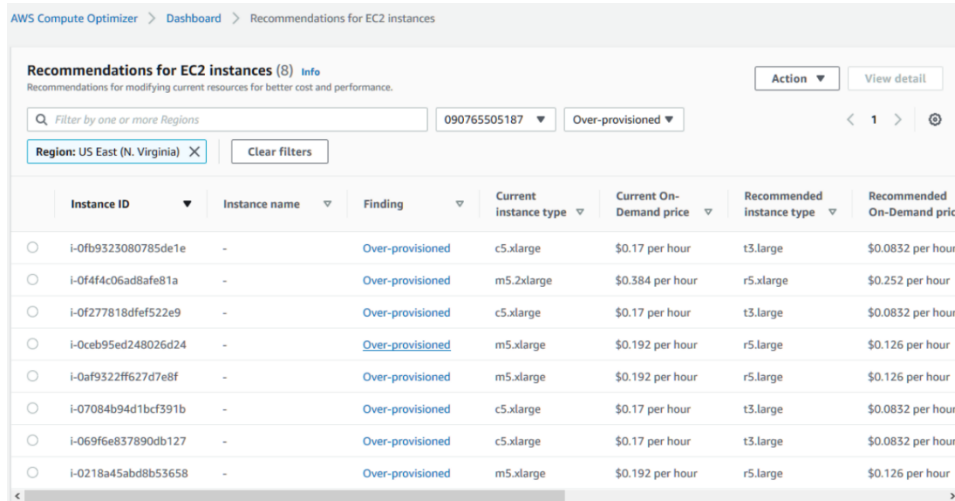


Figure 1. – Recommendations for EC2 instances

EC2 sizing recommendation example from *AWS Compute Optimizer*:

- *Trusted Advisor*;
- *VMWARE CloudHealth*;
- *Rightsizing*;
- *Cost management*;

### ***Solution to solve the problem***

For growing business cost management tools most of the time show growing expenses regardless of rightsizing efforts. The typical trend is below:

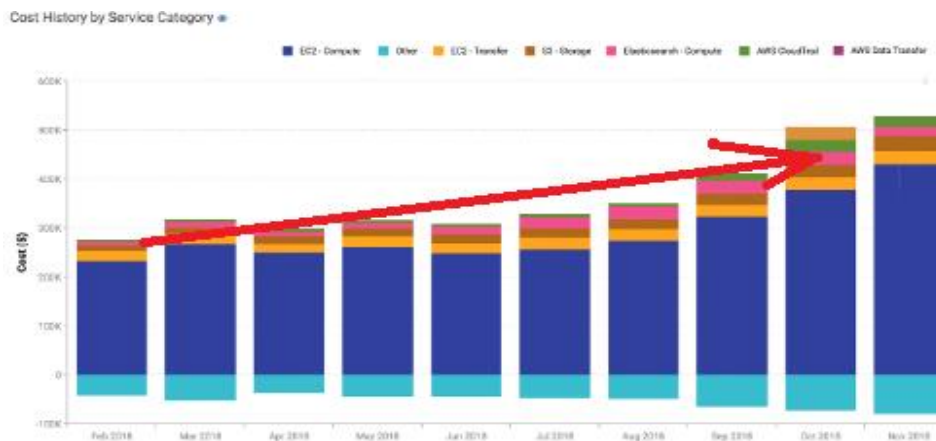


Figure 2. – The typical trend

To show how effective the cloud is used another following normalized approach is suggested. Multidimensional capacity utilization report should cover:

1. Compute capacity utilization
2. Memory (RAM) capacity utilization
3. Disk I/O bandwidth utilization
4. Network bandwidth utilization

To show how effective the Compute capacity is used there is a need to do normalization and aggregation of all different virtual servers (EC2) types (sizes).

One of the approach is to use AWS Elastic Compute Units (ECUs) That is the comparable “horse power” metric and can be get from AWS EC2 price list (the fragment of the list is below):

server type	vCPU	ECU	Memory (GiB)	Instance Storage (GB)	Linux/UNIX Usage
m5.large	2	8	8 GiB	EBS Only	\$0.096 per Hour
m5.xlarge	4	16	16 GiB	EBS Only	\$0.192 per Hour
m5.2xlarge	8	31	32 GiB	EBS Only	\$0.384 per Hour
m5.4xlarge	16	60	64 GiB	EBS Only	\$0.768 per Hour
c5.large	2	9	4 GiB	EBS Only	\$0.085 per Hour
c5.xlarge	4	17	8 GiB	EBS Only	\$0.17 per Hour
c5.2xlarge	8	34	16 GiB	EBS Only	\$0.34 per Hour
c5.4xlarge	16	68	32 GiB	EBS Only	\$0.68 per Hour
c5.9xlarge	36	141	72 GiB	EBS Only	\$1.53 per Hour
c5.18xlarge	72	281	144 GiB	EBS Only	\$3.06 per Hour
c5d.large	2	9	4 GiB	1 x 50 NVMe SSD	\$0.096 per Hour

Figure 3. – The fragment of the price list

It could be aggregated by Applications or/and by LOBs into Compute Capacity Utilization (CCU) that has a natural (as %) way to check a progress: closer to 100% is better. For example, combined compute power of *m5.xlarge* and *c5.4xlarge* would be 16+68=84 ECUs

CCUt - Compute Capacity utilization

CCA (Compute Capacity Available) is overall (sum) of all “i” ECUs for particular application or LOB. That gives the capacity amount purchased and available:

$$CCA = \sum ECU_i$$

CCU (Compute Capacity Used) is how much compute capacity used:

$$CCU = \sum (ECU_i * CPU_i\% / 100)$$

Where  $CPU_i\%$  is CPU utilization of “i” server (EC2), that could be get from the AWS CloudWatch tool or any other performance tools like DataDog. Finally Compute Capacity Utilization should be calculated as

$$CCUt \% = (CCU / CCA) * 100\%$$

CCA vs. CCU can be used to compare size and efficiency of cloud usage of two (or more) applications (or LOBs). Below is the example of two applications efficiency comparison, which shows that application APP\_1 has much more opportunity to be downsized and respectively to get bigger cost savings.

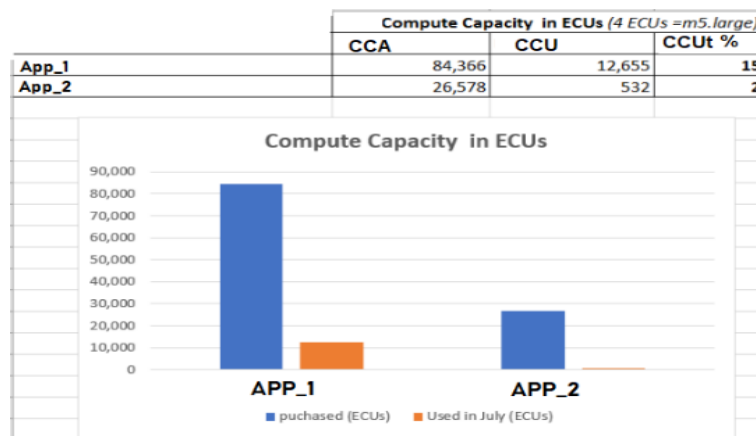


Figure 4. – Compute Capacity in ECU

RAM, Disk I/Os and Network Capacity Utilization



For workloads that are memory or I/Os intensive the downsizing only based of Compute capacity optimization could not be done correctly. Respectively similar calculation needs to be done to get the following:

- Memory (RAM) capacity utilization as a sum of all (total) RAM sizes in Gb available vs. RAM used in Gb.
- Disk I/O bandwidth utilization as a sum of IOs (per sec.) \*IO\_size (KB) vs. sum of IO\_bandwidths (KB per sec)
- Network bandwidth utilization as a sum of actual bandwidth (Gbit per sec) vs. Gbit used (per sec)

Considering all 4 dimensions of capacity usage allows to see how rightsizing works (current capacity usage vs. case when all recommendations are implemented), the example below shows how the capacity usage of all 4 dimensions are improved for two applications:

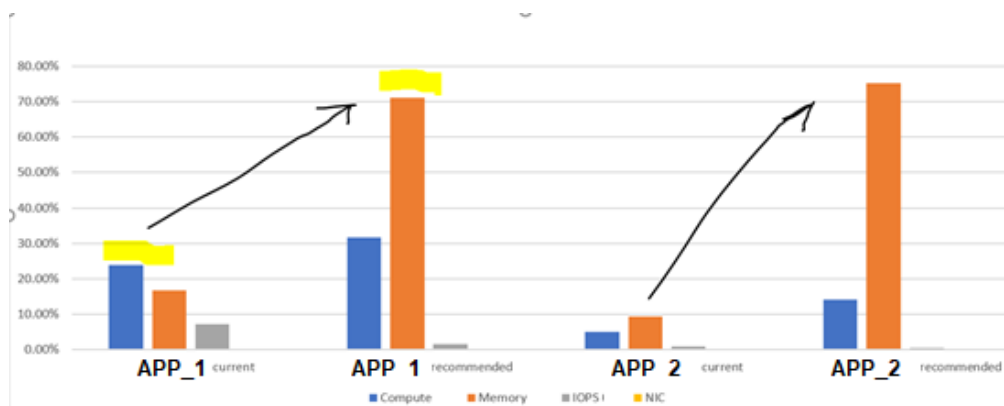


Figure 5. – The capacity usage of all 4 dimensions

Note, the culprit (bottleneck – least optimized dimension) could be changed or stay the same in the recommended right size.

*Typical patterns (Current vs. Recommended)*

And finally by keeping history of those 4 capacity usage metrics the trend could be built and regardless of adding additional workload that trend should be flat or going up to saturation level of 70% for good optimized cloud , while the actual cost still might keep growing, but that growth would be not excessive and just reflecting the growth of the business!

On the next few tend charts the typical pattern and antipatterns are shown:

- Far from Optimum – a lot of savings opportunities:

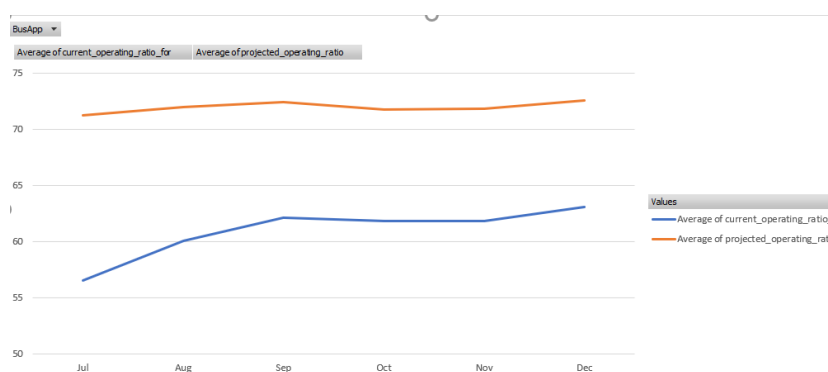


Figure 6. – A lot of savings opportunities

- Optimum is achieved!

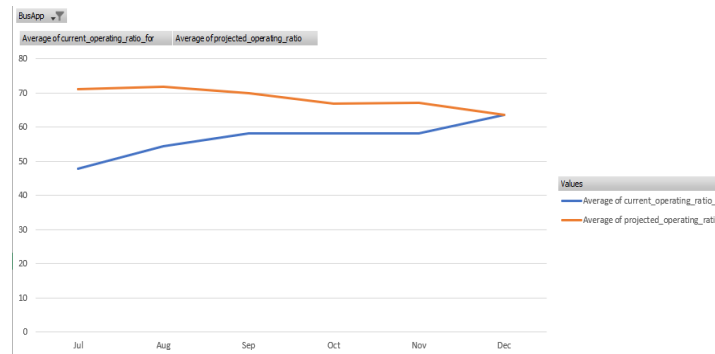


Figure 7. – Optimum

– After optimum there is a bit of overutilized:

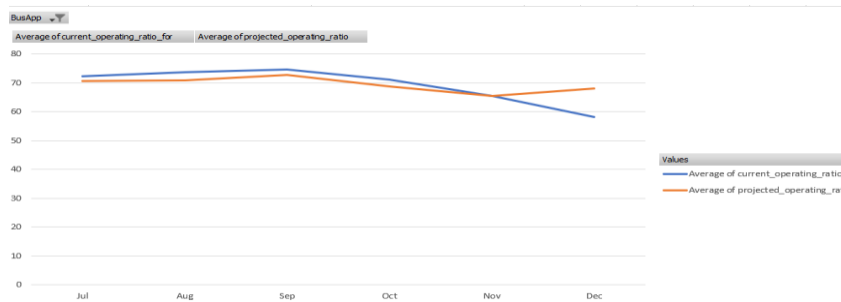


Figure 8. – After optimum

– BAD trend, efficiency is declining:

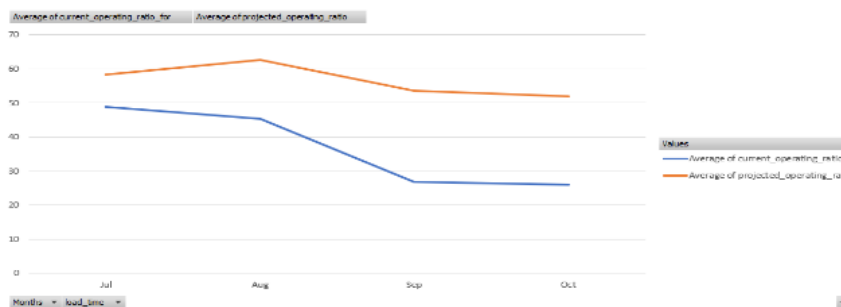


Figure 9. – BAD trend

– DANGER! Overutilization. Could be a capacity issue. Buy more capacity!

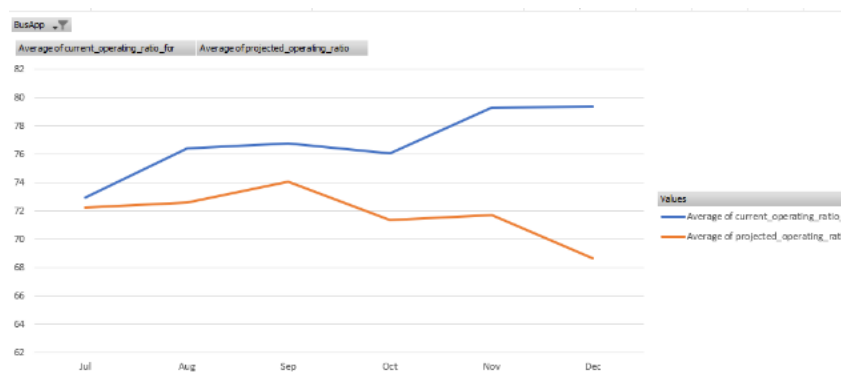


Figure 10. – Overutilization

### EC2 Overall Fleet Cost efficiency Treemap

To identify applications that have largest cost savings opportunities or opposite (– need some investing), the treemap (heat-chart) could be used as shown on the picture below:

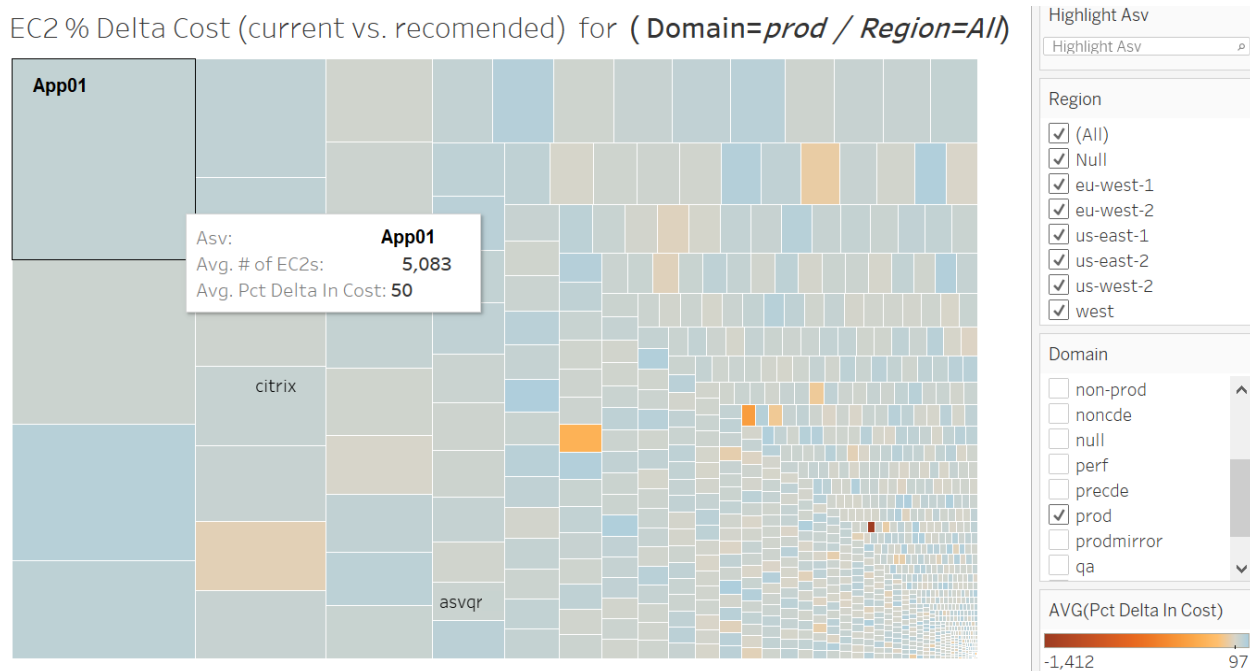


Figure 11. –EC2 % Delta Cost

UDK 004.04

## ФОРМИРОВАНИЕ ХРАНИЛИЩА И АНАЛИЗ БОЛЬШИХ ДАННЫХ ПЕРЕДВИЖЕНИЙ В ГОРОДЕ



**М.Н. Калимолдаев**  
Доктор технических наук, профессор, академик НАН РК



**В.В. Яворский**  
Доктор технических наук, профессор



**М.А. Сонькин**  
Доктор технических наук, профессор



**В. Войчик**  
Доктор технических наук, профессор



**И.Т. Утепберенов**  
Доктор технических наук, профессор



**А.Т. Ахмедиярова**  
соискатель научной степени



**Д.Т. Касымова**  
магистр техники и технологий



**Е.Г. Ключева**  
магистр



**Н.В. Байдикова**  
магистр технологий



**М.М. Есмагамбетова**  
магистр техники и технологий, докторант

Институт информационных и вычислительных технологий КН МОН РК, г. Алматы, Республика Казахстан,

Карагандинский государственный технический университет, г. Караганда, Республика Казахстан

Национальный исследовательский Томский политехнический университет, г. Томск, Российская Федерация

Института электронных и информационных технологий, Люблин, Республика Польша

Евразийский национальный университет им. Л.Н. Гумилева, г. Нур-Султан, Республика Казахстан

Карагандинский индустриальный университет, г. Темиртау, Республика Казахстан

E-mail: ms.bnatalya@mail.ru, mnk@ipic.kz, i.utepbergenov@gmail.com, dika.cat@mail.ru, aat.78@mail.ru, yavorskiy-v-v@mail.ru, e.klyueva@kstu.kz, sonkin@tpu.ru, wojcik@pollub.pl, marzhan1983@mail.ru

**М.Н. Калимолдаев**

*Генеральный директор Института информационных и вычислительных технологий КН МОН РК, обладатель «Государственной научной стипендии для ученых и специалистов, внесших выдающийся вклад в развитие науки и техники» (2019г.), нагрудного знака «За заслуги в развитии науки в Республике Казахстан» (2015г.). Сфера научных интересов: устойчивость, стабилизация, управляемость и оптимальность фазовых (электроэнергетических), технических и экономических систем, управление робототехническими системами.*

**В.В. Яворский**

*Томский политехнический университет в 1973 году, зав. отделом Территориальных АСУ НИИ АМ г. Томск до 1978 г., зав. каф. Автоматизированные информационные системы Карагандинского государственного технического университета (КарГТУ) 1998 – 2008гг, обладатель нагрудного знака «За заслуги в развитии науки в Республике Казахстан» (2016г.), обладатель гранта и звания «Лучший преподаватель» Республики Казахстан (2006, 2007, 2012гг.). Сфера научных интересов: интеллектуальные информационные системы, компьютерное моделирование. д.т.н., профессор.*

**М.А. Сонькин**

*Директор института кибернетики ТПУ, проректор ТПУ по научной работе, зам губернатора Томской области по научно-технической политике 2007-18гг., председатель совета директоров ГК ИНКОМ, д.т.н., профессор. Сфера научных интересов: интеллектуальные информационные системы и технические средства обеспечения безопасности в условиях чрезвычайного положения.*

**В. Войчик**

*Директор Института электроники и информационных технологий Люблинского Технического Университета (2012-по н/в), обладатель: медали комиссии народного образования Польши (2002г.), Рыцарского креста ордена Полонии Ресцута (2011г.), Ордена II-степени святого откровения апостолам Марии Магдалины (2018г.), член международных комитетов: IEEE, современных проблем радиотехники, телекоммуникаций и информатики, физики и технологий тонких пленок и наносистем, д.т.н., профессор.*

**И.Т. Утепбергенов**

*Заведующий лабораторией инновационных и смарт-технологий Института информационных и вычислительных технологий КН МОН РК (2012-по н/в), обладатель гранта и звания «Лучший преподаватель» Республики Казахстан (2009г. и 2014г.), Ветеран труда МОН Республики Казахстан (2019г.). Сфера научных интересов: информационные системы, информационные технологии, автоматизированные системы управления, системный анализ, д.т.н., профессор.*

**А.Т. Ахмедиярова**

*PhD по специальности «Информационные системы», с.н.с., Института информационных и вычислительных технологий КН МОН РК. Сфера научных интересов: моделирование и управление движением транспорта, интеллектуальные транспортные системы.*

**Д.Т. Касымова**

*Магистр техники и технологии по специальности «Вычислительная техника и программное обеспечение», н.с., Института информационных и вычислительных технологий КН МОН РК. Сфера научных интересов: интеллектуальные транспортные системы, большие данные.*

**Е.Г. Ключева**

*Магистр техники и технологии по специальности «Вычислительная техника и программное обеспечение». Старший преподавателем кафедры «Информационные технологии и безопасность» Карагандинского государственного технического университета. Сфера научных интересов: параллельные вычисления, базы данных, компьютерные сети, разработка программного обеспечения.*

**Н.В. Байдикова**

*Магистр техники и технологии, преподаватель кафедры вычислительной техники Карагандинского промышленного университета. Сфера научных интересов: интеллектуальные информационные системы, компьютерное моделирование.*

**М.М. Есмагамбетова**

*Старший преподаватель Карагандинского экономического университета. Магистр техники и технологии, докторант PhD Евразийского национального университета имени Л.Н. Гумилева. Сфера научных интересов: компьютерные модели массового поведения и информационное обеспечение безопасности в условиях*

*чрезвычайного положения.*

**Аннотация.** В статье рассмотрена концепция построения хранилища данных путей передвижения для интеллектуальной транспортной системы города. Приведен математический аппарат формализованного описания транспортной системы на основе гиперсети, дуги которой являются путями передвижения. Рассмотрен рекурсивный алгоритм получения данных о путях передвижения, а также механизм использования матриц смежности по маршрутным связям. Исследования проводятся в рамках проекта AP05133699 «Исследование и разработка инновационно-телекоммуникационных технологий с использованием современных кибер-технических средств для интеллектуальной транспортной системы города».

**Ключевые слова:** хранилище данных, интеллектуальные транспортные системы, матрицы корреспонденций.

### *Введение*

Современные городские транспортные системы (ГТС) – сложные образования, объединяющие инфраструктуру, функциональную структуру и среду потребления транспортных услуг [1].

Важнейшей задачей цифровизации управления городским транспортом является создание современных хранилищ данных и центров обработки, что наряду со средствами телекоммуникаций позволит эффективно использовать информацию, а также модели анализа процессов передвижения населения в городе, процессов обслуживания пассажиров и движения транспортных единиц.

SMART управление перевозками пассажиров может осуществляться как на макроуровне (рисунок 1), так и на микроуровне [1]. Микросистемы предполагают использование логистических принципов при организации транспортного обслуживания работников предприятия как одного из аспектов производственной деятельности.

К макро SMART относятся системы, участвующие в организации транспортного обслуживания. Управление на макроуровне предусматривает решение следующих задач:

- разработку общей концепции построения маршрутной сети;
- выбор рациональных направлений перевозок;
- отбор операторов и определение их объема работ;
- оптимизацию распределения объектов инфраструктуры по территории региона.

Центральной проблемой решения этих задач является анализ потоков на маршрутных сетях взаимодействующих видов транспорта. Для решения этой задачи необходимо создать хранилище данных о базовой транспортной сети города, состоящей из множества пунктов и проложенных между ними коммуникаций (связей).

Хранилища данных (ХД) – это современные информационные технологии, методы и средства проектирования и построения автоматизированных систем, ориентированных на анализ данных.

На транспортной сети формируется маршрутная сеть, представленная совокупностью транспортных маршрутов, по которым организовано движение транспорта. Здесь на маршрутах осуществляется производство конечного продукта транспортной системы - перевозок. Между пунктами транспортной сети осуществляются передвижения (корреспонденции), которые обобщенно можно реально представлять интенсивностью потока. С использованием данной информации необходимо оценивать и анализировать работу транспортных маршрутов.

*Целью настоящей работы* является разработка методики и алгоритмов формирования хранилища данных о передвижениях с целью дальнейшего использования для оптимизации транспортной системы.

Качественная характеристика системы

Социальная характеристика системы

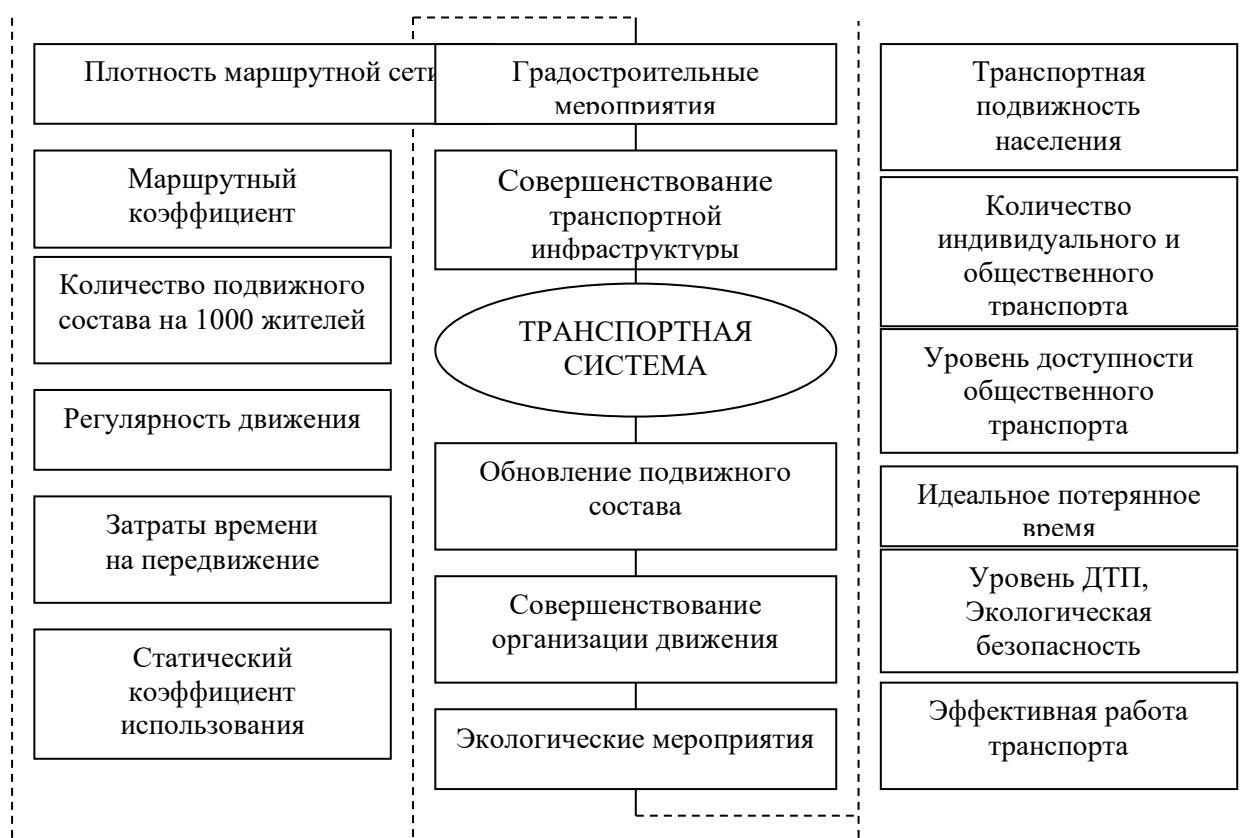


Рисунок 1. – Параметры качества транспортного обслуживания на макроуровне

### Материалы и методы

Наличие разных подходов к решению задачи распределения потоков на маршрутных сетях транспортных систем городов связано, на наш взгляд, с различным пониманием вопроса о детальности представления путей передвижения в городе.

Наиболее распространенный подход обычно предусматривает распределение корреспонденций по кратчайшим путям и направлениям, которые представляют собой объединение нескольких видов транспорта, обладающих определенным общим свойством, например, для системы городского общественного транспорта (ГОТ) такими направлениями являются: наземный уличный транспорт, внеуличный транспорт, метро, железнодорожный пассажирский транспорт и т.д.

Предлагаемый подход опирается на возможность формирования хранилища данных о возможных путях передвижения для крупных транспортных систем, а также получение обобщенной информации для планирования и управления их развитием. Наиболее сложным и важным в этом подходе является вопрос формирования путей передвижения, адекватных существующим способам передвижения потоков на транспортных сетях. Модели массового поведения и обслуживания, при явном определении путей передвижения и их параметров, позволяют получить детальные характеристики процессов передвижения в транспортной системе.

Ввиду многогранности и сложности учета множества взаимосвязанных элементов в качестве основного инструмента для изучения транспортных систем городов используется методология моделирования сложных систем массового поведения. Такая имитационная

модель позволяет с определенной степенью адекватности показать динамику взаимодействия транспортных систем, представить поведение исследуемого объекта в зависимости от изменения входных параметров, обеспечивает возможность многовариантного проектирования.

Предлагаемый комплекс моделей проектирования базируются на разработанном формализованном описании транспортной системы и имитационном моделировании процессов обслуживания на маршрутных сетях. Использование этого комплекса моделей включает следующие основные этапы:

- построение хранилища данных путей передвижения (ХДПП) между всеми корреспондирующими пунктами города;
- моделирование процессов обслуживания потоков на маршрутной сети;
- формирование реального потокораспределения на маршрутной сети с использованием моделирования процессов обслуживания на маршрутах.

Концепция ХД предполагает не просто единый логический взгляд на данные, а реализацию единого интегрированного источника данных.

ХДПП строятся на основе многомерной модели данных, которая подразумевает выделение отдельных измерений и фактов, анализируемых по выбранным измерениям. Прежде всего, это данные о маршрутах, в частности координаты остановок и интенсивности потоков входящих и выходящих пассажиров, данные о дорожной сети. Многомерная модель данных физически может быть реализована как в многомерных СУБД, так и в реляционных.

Основные требования к ХДПП:

- предметная ориентированность;
- интегрированность;
- неизменчивость;
- поддержка хронологии.

Математический аппарат формализованного описания транспортной системы рассмотрен в работе [2]. В рамках исследований по проекту AP05133699 «Исследование и разработка инновационно-телекоммуникационных технологий с использованием современных кибер-технических средств для интеллектуальной транспортной системы города» института ИВТ КН МОН РК были предложены собственные модели [3, 4]. Транспортная сеть, маршрутная сеть, корреспонденции между пунктами на транспортной сети, потоки на маршрутной сети, пути передвижения между пунктами транспортной системы предлагается представлять гиперсетью, дуги которой являются путями передвижения.

Путь  $h^{p_{kij}}$  между пунктами  $i$  и  $j$  на МС можно описать следующим образом:

$$h^{p_{kij}} = \{i, j; t(h^{p_{kij}}); p; k_1, \xi_1, \eta_1; k_2, \xi_2, \eta_2; \dots; k_l, \xi_l, \eta_l, \dots, k_p, \xi_p, \eta_p\}, \quad (1)$$

где  $t(h^{p_{kij}})$  – затраты времени на передвижение по пути  $h^{p_{kij}}$ ;  $p$  – порядок пути;  $k_l$  – индекс ТМ, осуществляющего  $l$ -ю маршрутную связь  $(\xi_l, \eta_l)$ ;  $\xi_l$  и  $\eta_l$  – начальный и конечный пункты в маршрутной связи  $(\xi_l, \eta_l)$ ,  $l = \overline{1, p}$ .

Для получения совокупности дуг  $\{h_{ij}^{q+1}\}$  берутся совокупности  $\{h_{i\xi}^q\}$  и к ним присоединяются маршрутно-пешеходные связи из совокупности  $\{h_{\xi j}^{MP}\}$ .

Операцию склеивания  $\otimes$  можно определить следующим образом:



$$\begin{aligned}
 & (i, \xi; t_s; q; k_1, \dots, k_q; \xi_1, \eta_1, \dots, \xi_q, \eta_q) \otimes (\xi, j; t_V; k_V; \eta_V) = \\
 & = \begin{cases} (i, j; t_s + t_V; q + 1; k_1, \dots, k_q, k_V; \xi_1, \eta_1, \dots, \xi_q, \eta_q, \xi, \eta_V), & \text{если} \\ q + 1 \leq q^M \wedge t_s + t_V \leq \alpha t_{ij}^M \wedge (k_1, \dots, k_q, k_V) - \text{элементарная после-} & (2) \\ \text{довательность} \wedge (i, \xi_1, \eta_1, \dots, \xi_q, \eta_q, \xi, \eta_V, j) - \text{элементарный путь.} \\ \emptyset - \text{в противном случае.} \end{cases}
 \end{aligned}$$

Стратегия выбора пассажирами  $\xi$ -го пути следования из  $i$ -го района в  $j$ -й из всей совокупности возможных путей может быть задана вероятностной функцией, зависящей от параметров пути передвижения.

Распределение корреспонденций по имеющимся путям передвижения в соответствии со стратегией выбора, очевидно, является «идеальной». Идеальное потокораспределение соответствует интересам пассажиров, осуществляющих поездки, однако оно не учитывает процессы, которые возникают при их обслуживании. Для получения реального потокораспределения разработана имитационная модель. С использованием разработанной модели может быть описан процесс обслуживания пассажиров. Так как пассажиры на основе своего опыта имеют информацию о процессе обслуживания, который их ожидает, то будет происходить перераспределение идеальных потоков и приспособление передвижения к реальной пропускной способности маршрутов гиперсети. Для того, чтобы учесть это, используется процедура пересчета временных параметров гиперграфа  $L(Z, W)$ .

Рассмотрим разработанный алгоритм построения хранилища данных путей передвижения.

1. Выбирается маршрут на текущей остановке, до тех пор, пока не будут рассмотрены все маршруты, проходящие через текущую остановку.

2. Если маршрут пешеходный: 2.1. Выбирается остановка до тех пор, пока не будут рассмотрены все пешеходные передвижения от текущей остановки до всех существующих остановок. 2.2. Если выбранная остановка присутствует в списке найденных ветвей текущего пути, то возврат в п. 2.1 2.3. Если ветвь от текущей остановки до выбранной является текущим решением, то эта ветвь вносится в текущий путь. Вызывается рекурсивно данную процедуру, содержащую параметры текущего пути. После окончания рекурсии последняя найденная ветвь удаляется из текущего пути. 2.4 Возврат в п. 2.1

3. Если маршрут не пешеходный: 3.1. Если данный маршрут уже был использован передвижениями по найденным ветвям текущего пути, то возврат в п. 1 3.2. Если количество использованных непешеходных маршрутов на 1 больше числа возможных пересадок, то выбранный маршрут должен быть последним. Следовательно, если ветвь от текущей остановки реку до пункта назначения с использованием выбранный маршрут является текущим решением, то эта ветвь вносится в текущий путь, который заносится в результирующий набор. Последняя найденная ветвь удаляется из текущего пути. 3.3. Возврат в п.1. 3.4. Если выбранный маршрут - не последний, то выбираем остановку на выбранном маршруте до тех пор, пока не будут рассмотрены все передвижения от текущей остановки до всех остановок по данному маршруту. 3.5. Если выбранная остановка присутствует в списке найденных ветвей текущего пути, то возврат в п. 3.4. 3.6. Если ветвь от текущей остановки до выбранной является текущим решением, то эта ветвь вносится в текущий путь. Если выбранная остановка совпадает с пунктом назначения, то текущий путь заносится в результирующий набор, иначе вызывается рекурсивно описываемая процедура, содержащая параметры текущего пути. После окончания рекурсии последняя найденная ветвь удаляется из текущего пути. 3.7. Возврат в п. 3.4. 3.8. Возврат в п. 1.

Проверка на достоверность решения включает в себя:

- проверку на совпадение текущей остановки с выбранной остановкой;
- проверку на существование пути из текущей остановки в выбранную остановку;
- проверку на отклонение по времени.

Рассмотрим пример работы алгоритма поиска на простейшей транспортной системе. Дана транспортная система:

–остановки: [1, 2, 3, 4, 5].

–список маршрутов [0, А, В].

–маршруты:

0 – все пешеходные передвижения [4-5];

А – [1, 2, 3, 4];

В – [1, 5, 3].

Схема дорог с маршрутами:

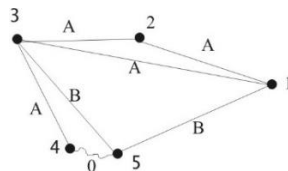


Рисунок 1. – Схема дорог с маршрутами

Рассмотрим структуру поиска всех возможных путей к передвижениям из 1 в 3.

Условия поиска:

–ограничение по времени не учитывается (только для данного демонстрационного случая);

–возможны 3 пересадки;

–каждый маршрут может быть использован только 1 раз.

*Шаг 1.* {текущая остановка 1, маршрут 0}.

Результат поиска: не найдено пешеходных передвижений из 1.

*Шаг 2.* {текущая остановка 1, маршрут А}.

Результат поиска: найдено передвижение из 1 в 2 по маршруту А.



Рисунок 2. – Результат поиска: найдено передвижение

*Шаг 3.* {текущая остановка 2, маршрут А}.

Результат поиска: не найдено передвижений из 2. Возврат.



Рисунок 3. – Результат поиска: найдено передвижение

*Шаг 4.* {текущая остановка 1, маршрут А}.

Результат поиска: найдено передвижение из 1 в 3 по маршруту А. Текущий путь внесен в результирующий набор (дерево решений). Возврат.

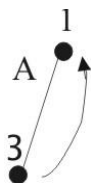


Рисунок 4. – Результат поиска: найдено передвижение из 1 в 3 по маршруту А

Шаг 5. {текущая остановка 1, маршрут В}

Результат поиска: найдено передвижение из 1 в 5 по маршруту В.

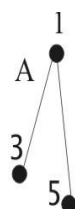


Рисунок 5. – Результат поиска: найдено передвижение из 1 в 5 по маршруту В

Шаг 6. {текущая остановка 5, маршрут В}

Результат поиска: найдено пешеходное передвижение из 5 в 4.

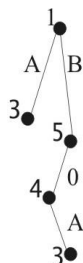


Рисунок 6. – Результат поиска: найдено пешеходное передвижение из 5 в 4.

Шаг 7. {текущая остановка 4, маршрут А}

Результат поиска: найдено передвижение из 4 в 3 по маршруту А. Текущий путь внесен в результирующий набор (дерево решений). Возврат.

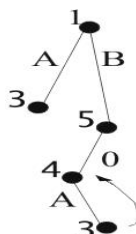


Рисунок 7. – Результат поиска: найдено передвижение из 4 в 3 по маршруту А

Шаг 8. {текущая остановка 4, маршрут А}

Результат поиска: не найдены передвижения из 4 по маршруту А.

Шаг 9. {текущая остановка 1, маршрут В}

Результат поиска: найдено передвижения из 1 в 3 по маршруту В. Текущий путь внесен в результирующий набор (дерево решений). Возврат.

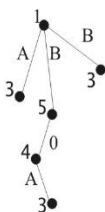


Рисунок 8. – Результат поиска: найдено передвижения из 1 в 3 по маршруту В

Шаг 10. {текущая остановка 1, маршрут 0}

Результат поиска: не найдены передвижения из 1.

Шаг 11. {текущая остановка 1, маршрут А}

Результат поиска: не найдены передвижения из 1.

Шаг 12. {текущая остановка 1, маршрут В}

Результат поиска: не найдены передвижения из 1.

Шаг 13.

Результирующий набор (дерево решений):

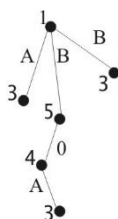


Рисунок 9. – Дерево решений

### Результаты исследований

Полученное дерево решений может быть записано в ХДПП.

После того, как ХДПП создана и все данные внесены в систему начинается главная процедура анализа. Она осуществляется в два этапа:

1. Получение первого приближения матрицы потенциальных корреспонденций (МПК);
2. Составление системы уравнений и получение МПК с заданной точностью.

Первое приближение МПК находится приближенным методом энтропийного моделирования с использованием данных об обследовании, средних данных по городу и различных коэффициентов, полученных во время проведенных обследований. Этот метод заключается в распределении пассажиропотоков, отправляющихся из транспортных районов города (остановок) в другие районы в соответствии с функцией, отражающей трудность передвижения между районами. Для реализации обычно выбирают экспоненциальную модель функции тяготения вида  $e^{-l/l_{cp}}$ , где  $l$  – расстояние между транспортными районами, а  $l_{cp}$  – средняя дальность поездки по городу. Для выравнивания матрицы обычно применяют метод балансировки корреспонденций.

Система уравнений, отражающая взаимосвязь пассажиропотоков на маршрутах с корреспонденциями между транспортными районами города, сформированная с использованием ХДПП позволяет рассчитать реальные корреспонденции.

Важный вопрос урбанистики - достижимость между районами города. На основе данных, полученных в ходе построения ХДПП, формируем матрицы смежности по маршрутным связям и более сложных передвижений.

Задача анализа достижимости фактически сводится к формированию и анализу экстремальных путей на гиперсети передвижений в городе. Она является вычислительно

сложной ввиду большой размерности исходных матриц смежности. Поэтому предлагается использовать концепцию параллельных вычислений умножения с использованием нескольких гетерогенных узлов.

Основной задачей при этом является нахождение оптимального разбиения данных, которое определяет способы распределения частей матриц среди доступных вычислительных элементов при выполнении операции умножения матриц. Разбиение данных производится с целью оптимизировать показатели по определенным критериям работы компьютерных приложений.

*Цель разделения данных* - оптимально распределить вычислительную нагрузку при выполнении матричных вычислений среди доступных процессоров. Скорость, с которой отдельный процессор может выполнять основные операции, такие как сложение и умножение, определяет долю общей задачи, на которую он будет назначен. Таким образом, при параллельных вычислениях все процессоры будут одновременно выполнять вычисления, и ни один из процессоров не будет простаивать.

Оценка оптимальности формы разделения должна быть произведена на основе объективных факторов, в качестве которых будет выступать время вычисления искомой матричной операции.

Для проведения вычислений в рамках исследования используется методика разбиения элементов матрицы на трех абстрактных гетерогенных процессорах, объединенных полносвязной топологией с различными пропускными способностями сети передачи данных, представленная в работах 5-7. В зависимости от исходной вычислительной мощности кластера выбирается оптимальная форма разбиения данных для конкретной задачи.

#### *Заключение*

В ходе проведенного исследования рассмотрены модели передвижения пассажиров городского пассажирского транспорта на основе гиперсетей. Показано, что равновесное состояние соответствует минимальным затратам времени на передвижение пассажиров и определяет распределение пассажиров по путям следования в соответствии со стратегией выбора пути. В предлагаемой модели равновесное состояние определяет распределение пассажиров по путям следования в соответствии со стратегией выбора пути. Рассмотрены возможности моделирования передвижения пассажиров городского пассажирского транспорта на основе гиперсетей.

Таким образом распределение потоков пассажиров по дугам графа  $L(Z, W)$  в равновесном состоянии  $\hat{G}(h, e)$ ,  $(h, e) \in W$  учитывает стратегию выбора пассажирами путей следования и реальные процессы обслуживания на гиперсети ГОТ.

В целом для предлагаемой модели справедлив вывод, который заключается в том, что сходимость имеет место, если функции  $\{P_{\xi}^{ij}\}$  и  $\{T_{\xi}^{ij}\}$  являются непрерывными.

На основании полученной математической модели может быть использован алгоритм добычи данных для ХДПП, основанный на методе поиска с возвратом. Результирующий набор представляет собой многомерное дерево решений, данные которого вносятся в хранилище.

В системе данные хранятся в виде многомерной базы данных, созданной в СУБД MS SQL Server 2017. Часть данных непосредственно вводится в БД с использованием интерфейса разработанного программного обеспечения. Другая часть генерируется подпрограммами системы [8] и также вносится в БД.

Приведенные алгоритмы позволяют строить хранилище данных, удовлетворяющее основным требованиям к ХДПП.

#### *Список литературы*

[1.] Организация хранилищ данных для Смарт систем городского общественного транспорта: монография для магистрантов специальностей 7М06101 "Информационные системы", 7М06102

"Вычислительная техника и программное обеспечение" / М. Н. Калимолдаев [и др.]. - Алматы: Институт информационных и вычислительных технологий КН МОН РК, 2019. - 119/1 с.

[2.] Яворский В.В., Утепбергенов И.Т. Структурные методы совершенствования управления транспортными системами городов. Караганда: КарГТУ, 2006. – 227с.

[3.] Яворский В.В., Утепбергенов И.Т., Ахмедиярова А.Т. Models of analysis of distribution of passenger traffics in routed transport systems // Материалы международной конференции XIII Balkan Conference on Operational Research (BALCOR 2018). — Belgrade, Serbia, 2018.

[4.] Яворский В.В., Ключева Е.Г., Мутовина Н.В., Касымова Д.Т. Расчет транспортной потребности на основе анализа линейных многообразий // Материалы научной конференции ИИВТ КН МОН РК «Инновационные IT и Smart-технологии», — Алматы : ИИВТ МОН РК, 2019. — С. 278-286.

[5.] Ключева Е.Г. Оптимальная форма разбиения данных при реализации алгоритма последовательной коммуникации с наложением на трех гетерогенных процессорах, объединенных полносвязной топологией с различными пропускными способностями сети передачи данных// Вестник государственного университета имени Шакарима города Семей. – 2018. - № 2(82) - С.46-51.

[6.] Ключева Е.Г., Адамов А.А., Оспанова А.Е., Сницарь Л.Р., Кулбаева Л.Н. Исследование оптимальной формы разбиения данных для умножения матриц на трех гетерогенных процессорах с полносвязной топологией и различными пропускными способностями // Современные наукоемкие технологии. – 2019. – № 2 – С. 83-88.

[7.] Ключева Е.Г., Яворский В.В., Адамов А.А. Определение оптимального разбиения элементов матриц при параллельном умножении на гетерогенных процессорах // Материалы научной конференции ИИВТ КН МОН РК «Инновационные IT и Smart-технологии», посвященной 70-летию юбилею профессора Утепбергенова И.Т. — Алматы : ИИВТ МОН РК, 2019. — С. 176-181.

[8.] Свидетельство о внесении сведений в государственный реестр прав на объекты, охраняемые авторским правом. Яворский В.В., Ключева Е.Г., Ахмедиярова А.Т., Касымова Д.Т. Программа для ЭВМ, Модуль «RouteSearch», 21.01.2019.

## **FORMATION OF STORAGE AND ANALYSIS OF BIG MOVEMENT DATA IN THE CITY**

**M.N. Kalimoldaev**  
*Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences*

**V.V. Yavorsky**  
*Doctor of Technical Sciences, Professor of Karaganda State Technical University*

**M.A. Sonkin**  
*Doctor of Technical Sciences, Professor of the National Research Tomsk Polytechnic University*

**V. Voychik**  
*Doctor of Technical Sciences, Professor (Technology), Director of the Institute of Electronics and Informatics*

**I.T. Utepbergenov**  
*Doctor of Technical Sciences, Professor, Head of the Laboratory of Innovative and Smart Technologies*

**A.T. Akhmediyarova**  
*Senior Researcher, Institute of Information and Computing Technologies*

**D.T. Kasymova**  
*Master of Engineering and Technology, Senior Researcher, Institute of Information and Computing Technologies*

**E.G. Klyueva**  
*Master of Karaganda State Technical University*

**N.V. Baidikova**  
*Master of Technology, Karaganda State Industrial University*

**M.M. Esmagambetova**  
*Master of Engineering and Technology, doctoral student at Eurasian National University named after L.N. Gumileva*

*Institute of Information and Computing Technologies of the National Academy of Science of the Republic of Kazakhstan, Almaty, Republic of Kazakhstan,  
Karaganda State Technical University, Karaganda, Republic of Kazakhstan  
National Research Tomsk Polytechnic University, Tomsk, Russian Federation  
Institute of Electronic and Information Technology, Lublin, Republic of Poland  
Eurasian National University named after L.N. Gumileva, Nur-Sultan, Republic of Kazakhstan  
Karaganda Industrial University, Temirtau, Republic of Kazakhstan  
E-mail: ms.bnatalya@mail.ru, marzhan1983@mail.ru, wojcik@pollub.pl, sonkin@tpu.ru, yavorskiy-v-v@mail.ru, e.klyueva@kstu.kz, mnk@ipic.kz, i.utepbergenov@gmail.com, dika.cat@mail.ru, aat.78@mail.ru*

**Abstract.** The article discusses the concept of constructing a repository of data on travel routes for the intellectual transport system of the city. The mathematical apparatus of a formalized description of a transport system based on a hypernet, whose arcs are paths of movement, is given. A recursive algorithm for obtaining data on paths of movement, as well as a mechanism for using adjacency matrices for route connections, is considered. Research is carried out as part of the project AR05133699 “Research and development of innovative telecommunication technologies using modern cyber-technical means for the intellectual transport system of the city”.

**Keywords:** data warehouse, intelligent transport systems, correspondence matrices.

UDK 004.93:004.032.26

## **RECOGNITION OF UNDERLYING SURFACE USING A CONVOLUTIONAL NEURAL NETWORK ON A SINGLE-BOARD COMPUTER**



**D.A. Paulenka**  
*Software Engineer,  
UIIP NASB,  
dmitri.pavlenko@gmail.com*



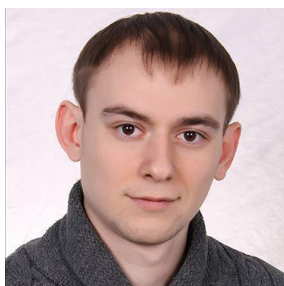
**V.A. Kovalev**  
*Head of Biomedical Image  
Analysis Laboratory, Ph.D.,  
UIIP NASB,  
vassili.kovalev@gmail.com*



**E.V. Snezhko**  
*Leading Researcher, Ph.D.,  
UIIP NASB,  
eduard.snezhko@gmail.com*



**V.A. Liauchuk**  
*Researcher, Ph.D.,  
UIIP NASB,  
vitali.liauchuk@gmail.com*



**E.I. Pechkovsky**  
*Leading Software Engineer,  
UIIP NASB,  
zver927@gmail.com*

*The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, (UIIP NASB, 6 Surganova Str., 220012, Minsk, Belarus)*

**Abstract.** The article presents development results for hardware and software system (micromodule), which detects and classifies underlying surface images of Earth. Given device has size 5.2×7.4×3.1 cm, mass 52 g and uses convolutional neural network based on MobileNetV2 architecture for image classification. The micromodule can be installed on board of a small spacecraft or a light unmanned aerial vehicle (drone). The information provided in this paper could be useful for engineers and researchers who are developing compact budget mobile systems for processing, analyzing and recognition of images.

**Keywords:** image recognition, convolutional neural network, deep learning, single-board computer, edge device, mobile system

**Introduction.** Automatic classification of underlying surface images, as well as search and recognition of objects in video and photo images are important tasks, the solution of which ensures the appropriateness of their application. Automatic terrain recognition using images of the underlying surface eliminates the long and tedious viewing of video materials by ground control operators, reduces time required for data preparation, increases reliability of choice, decreases power consumption from on-board sources due to the automatic regions of interest selection and diminish delay time between received signal and response to it.



Modern single-board microcomputers and smartphones have sufficient computing power to install software for automatic search and classification of images. This allows data preprocessing directly on the receiving device (edge device) without transferring a large amount of information to the server. Thus, the data preprocessing can be done aboard an unmanned aerial vehicle (UAV) or a small spacecraft. The authors developed a micromodule which is a hardware and software system and consists of a single-board microcomputer with necessary auxiliary equipment and a convolutional neural network (CNN) based on MobileNetV2 architecture for data processing.

#### **Main micromodule functions**

1. The micromodule is an autonomous system for automatic images recognition of underlying surface of various types: different kinds of forest, industrial structures, reservoirs, roads, fields, shrubs, country houses, farmland of various types with vegetation at different stages of growth, etc.

2. Target classes are set in the flight task by means of CNN which is prepared using Deep Learning technology.

3. Through the built-in video camera, the micromodule receives surface images and then classifies them in accordance with the flight task. If necessary, frames of the video sequence can be stored on the micromodule's external media for subsequent analysis.

4. The interaction between micromodule and UAV is minimal. And reduced to real-time transferring from micromodule to UAV the identified class number of a current underlying surface and the probability of its reliable determination. Class "0" means that the current image does not belong to any of the target classes specified in the flight task. The decision to perform high-quality shooting, recording and image transmission to the ground control is made outside of the micromodule.

5. The micromodule is a computing system with all necessary software development tools, operating system and interface equipment. Implementation with circuit solutions from separate specialized components such as microcontrollers, digital signal processors (DSP), chipsets, systems on a chip, etc. is not considered by the authors in this article.

#### **Micromodule's hardware**

Micromodule's hardware (Figure 1) consists of protective case, single-board Raspberry Pi Zero Wireless microcomputer, microSDHC memory card, Raspberry Pi Zero V1.3 Mini Camera, 720 mAh ROBITON Li-Po battery, battery charging microchip and connecting wires.

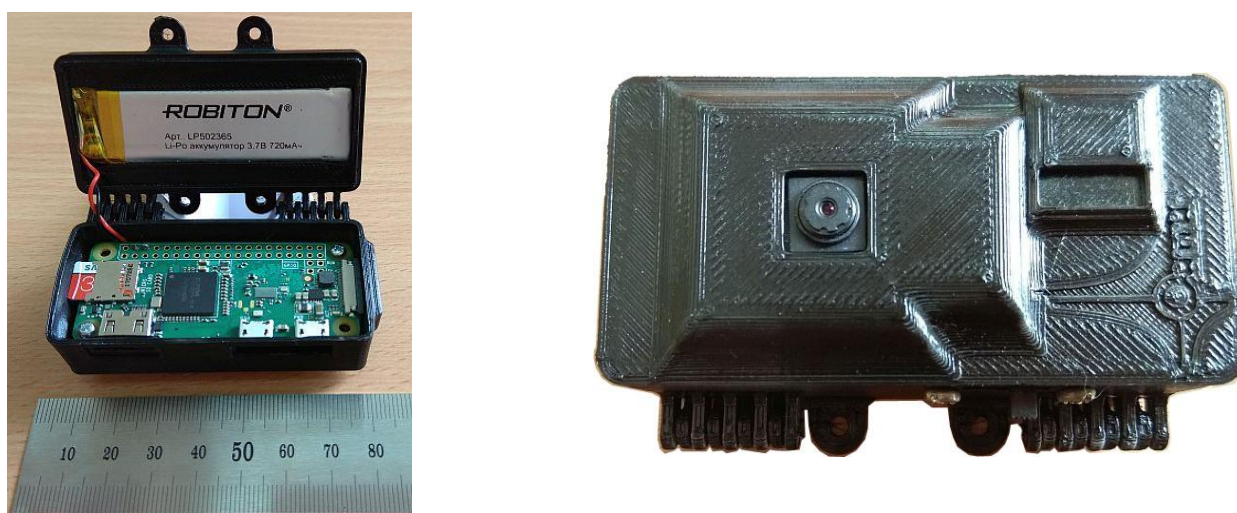


Figure 1. – Micromodule's hardware

To reduce production time, decrease manufacturing complexity and product cost the protective case was made on a 3D printer by the layer-by-layer extrusion of molten yarn from polycarbonate material.

The advisability of equipment selection, comparative analysis of computing microplatforms and technical requirements to hardware and software system are carefully described in [1, 2].

Micromodule size is  $5.2 \times 7.4 \times 3.1$  см. Micromodule weight is 52 г. Maximum power consumption at payload is  $1,75 \pm 0,25$  Вт.

### Micromodule's software

Micromodule's software consists of flying and ground parts (Figure 2).

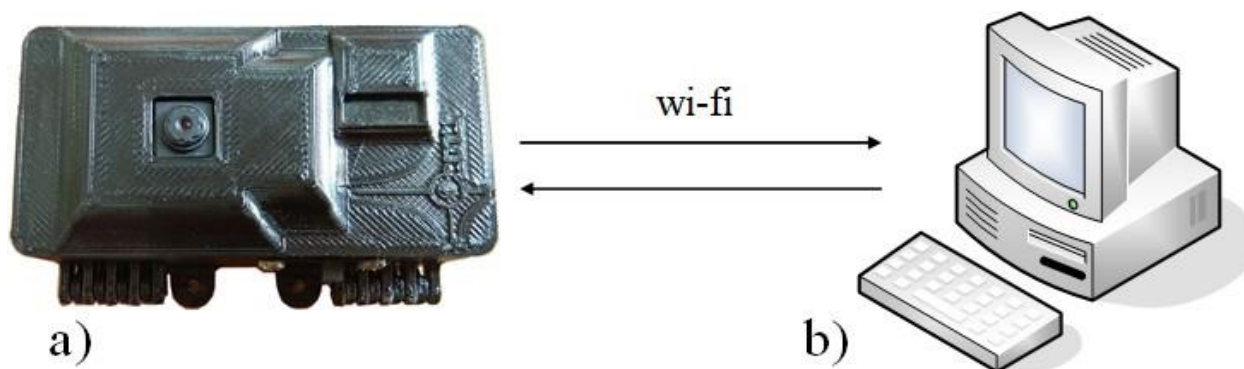


Figure 2. – Flying (a) and ground (b) parts of the micromodule

The flying part processes images and video stream frames, picks and compresses them. The ground part creates image datasets and constructs flight tasks. A wireless Wi-Fi connection is established between ground and flying parts to launch flight task, exchange data and monitor software functioning. In its turn, the flying part operates in two modes: the image processing mode and the video stream processing mode.

In the image processing mode, the flying part of the micromodule gets input *image files*, performs real-time data recognition, selection and compression and returns images with required classes of underlying surface according to the flight task.

In the video stream processing mode, the flying part of the micromodule gets input *video stream frames*, performs real-time data recognition, selection and compression and returns frames with required classes of underlying surface according to the flight task.

Micromodule's software is written in the Python and C/C++ programming languages. Operating system of the flying part is Raspbian Jessie. Operating system of the ground part is Ubuntu Linux.

The human operator of the ground part forms a flight task and copies it to the flying part. The flying part is mounted on a quadcopter or a UAV. Micromodule is attached to the aircraft but is not interlinked with its software or hardware. During the flight, micromodule in real-time mode processes video stream frames from the camera or scans input satellite images received via Wi-Fi. At the output, it has indexes and images of the found target classes in accordance with the flight task. For image processing, micromodule uses CNN which is based on the MobileNetV2 architecture [3].

Micromodule's flying part processes frames at a speed of more than three frames per second (see the next section).

### The architecture of developed convolutional neural network

The main function of the micromodule is the real-time images recognition of the underlying surface, that is, the analysis of a current video sequence frame or an image from the scanning window and assigning it to one of the classes indicated in the flight task.

Micromodule has low computational speed of a one-core processor with ARMv6 architecture (1 GHz) and small size of RAM (512 MB). This does not allow the use of any well-known neural

networks (AlexNet, VGG16, GoogLeNet, etc.). Therefore, a specialized CNN architecture based on MobileNetV2 was implemented [3].

Developed CNN has eleven blocks (106 layers in total), each block is the same as in MobileNetV2, as well as an output block adapted for image classification of Earth's underlying surface. In addition to reducing the number of main blocks from thirteen to eleven, a series of optimization techniques to speed up calculations were used. These techniques are taken into account the micromodule's hardware architecture. In general, optimization techniques reduced the time to get the output CNN vector by about five times on the same data compared with the original MobileNetV2 architecture.

The process of solving an image recognition task has two stages:

- training the neural network on sample images of underlying surface target classes, which is performed on the ground part of the micromodule's software;
- real-time image recognition on the flying part of the micromodule's software.

Software for flight task construction on the micromodule's ground part uses Keras library and trains a neural network on a powerful video card (Graphical Processing Unit, GPU). Flight task construction result is a trained neural network. This trained neural network, or rather a file with weights of the trained neural network is then transferred to the micromodule's flying part.

For real-time image classification on the micromodule's flying part, a special functions library was developed using C programming language. This function library imports CNN's weights to the neural network model and then calculates output vector values from the input images.

With this implementation, the classification time for a single three-channel (RGB) image 96×96 pixels in size on a single-core processor with ARMv6 architecture and a frequency of 1 GHz is on average 280 ms per image.

### Image recognition quality assessment

Two image datasets with various underlying surface classes were prepared (Figure 3).



Figure 3. – Examples of underlying surface images

Maps with underlying surface were downloaded using free of charge SASPlanet software (URL: <http://www.sasgis.org>). Map projection is “Mercator / Google Maps”. Map scale is Z18.

Maps markup and creation of image dataset with underlying surface classes were made by “Manual image annotation with polygons” software tool (URL:

[https://github.com/foobar167/junkyard/tree/master/manual\\_image\\_annotation1](https://github.com/foobar167/junkyard/tree/master/manual_image_annotation1)). This tool was developed by the authors of the article and is a component of the micromodule's ground part.

The first image dataset consists of eight classes: lake, city, field, forest, highway, road, river, village. To create the flight task, 40 000 images, each with size 256×256 pixels were used. To avoid overfitting effect, satellite maps for training were taken from Minsk city neighborhood, and for testing from Brest city and Lake Naroch neighborhood. Resulting confusion matrix for eight classes of developed CNN is in the Table 1, which gives the relative number of images that are assigned to one of the classes.

Table 1. – Confusion matrix for eight classes of developed CNN

class	lake	city	field	forest	highway	road	river	village
lake	<b>0.81</b>	0.00	0.05	0.13	0.00	0,00	0,01	0,00
city	0.03	<b>0.60</b>	0.04	0.01	0.07	0,01	0,01	<b>0,23</b>
field	0.04	0.00	<b>0.95</b>	0.00	0.00	0,01	0,00	0,00
forest	<b>0.60</b>	0.00	0.02	<b>0.32</b>	0.00	0,00	0,05	0,00
highway	0.00	0.01	0.04	0.01	<b>0.73</b>	0,11	0,10	0,01
road	0.00	0.00	0.01	0.00	0.05	<b>0.91</b>	0,04	0,00
river	0.03	0.00	0.09	0.02	0.05	<b>0,29</b>	<b>0,47</b>	0,04
village	0.00	0.00	0.01	0.00	0.00	0,06	0,00	<b>0,92</b>

Image percentage with minimum confidence coefficient 0.9 is equal to 0.86. Average classification accuracy is equal to 0.71.

The table shows that the class “city” is often confused with the class “village”, which is obvious, because “city” and “village” are similar by the texture. Class “forest” frequently falsely defined as class “lake”, possibly because the “lake” stands out along with the forested shores. Class “river” is confused with the “road”, possibly due to the fact that country roads often go along the rivers, so the rivers were marked along with the roads in the images. I.e. there are several classes on the image: lake and forest, river and road, etc.

For comparison, Table 2 shows the confusion matrix for the popular neural network GoogLeNet, from which it can be seen that the recognition quality is only slightly better than in Table 1.

Table 2. – Confusion matrix for eight classes of GoogLeNet

class	lake	city	field	forest	highway	road	river	village
lake	<b>0.91</b>	0,00	0,01	0,05	0,01	0,00	0,02	0,00
city	0,10	<b>0.60</b>	0,00	0,00	0,15	0,01	0,01	<b>0,13</b>
field	0,04	0,00	<b>0.94</b>	0,00	0,00	0,01	0,01	0,00
forest	<b>0.55</b>	0,00	0,00	<b>0.44</b>	0,00	0,00	0,01	0,00
highway	0,01	0,00	0,05	0,00	<b>0.73</b>	0,07	0,13	0,00
road	0,01	0,00	0,00	0,00	0,01	<b>0.95</b>	0,03	0,00
river	0,01	0,00	0,00	0,00	0,01	<b>0,11</b>	<b>0.87</b>	0,00
village	0,00	0,00	0,00	0,00	0,00	0,08	0,00	<b>0.91</b>

Image percentage with minimum confidence coefficient 0.9 is equal to 0.88. Average classification accuracy is equal to 0.79.

The second image dataset consists of eleven classes: canyon, city, desert1, desert2 (with a different texture), field, forest, lake, mountains, savannah, sea, swamp. In total 282 700 images, each with size 256×256 pixels were used.

Table 3. – Confusion matrix for eleven classes of developed CNN

class	canyon	city	desert1	desert2	field	forest	lake	mountains	savannah	sea	swamp
canyon	<b>9991</b>	0	0	0	0	0	8	1	0	0	0
city	0	<b>8000</b>	0	0	0	0	0	0	0	0	0
desert1	0	0	<b>5967</b>	33	0	0	0	0	0	0	0
desert2	0	0	0	<b>5000</b>	0	0	0	0	0	0	0
field	0	0	0	0	<b>7500</b>	0	0	0	0	0	0
forest	0	0	0	0	0	<b>6000</b>	0	0	0	0	0
lake	0	0	0	0	0	0	<b>6000</b>	0	0	0	0
mountains	16	0	0	0	17	0	0	<b>6959</b>	0	0	4
savannah	3	1	0	0	0	0	0	0	<b>7984</b>	0	0
sea	0	0	0	0	2	0	0	0	0	<b>6998</b>	0
swamp	0	0	0	0	0	0	0	0	0	0	<b>5500</b>

High recognition quality for the second image dataset is explained by the fact that it was *intentionally* made up of very similar textures of Earth, which, unlike the first image dataset, are completely different from each other.

A comparative analysis of the recognition quality between the developed CNN and the popular neural network architectures: LeNet, AlexNet and GoogLeNet was carried out. The results of the comparative analysis are given in [4] and show that the CNN architecture developed at UIIP NASB is comparable in quality to the popular architectures of deep convolution networks.

### Conclusion

A hardware and software system (micromodule) for image classification of the underlying surface of the Earth was developed at UIIP NASB.

Micromodule's hardware has low power consumption, small size and weight. Cost of micromodule components is less than hundred Belarusian rubbles.

Developed convolutional neural network is reliable and is comparable by work quality to the popular deep convolutional neural network architectures. Small image processing time allows to install this software on mobile platforms with a small computing power. However, recognition accuracy depends on the image dataset, which is used to train developed neural network.

In a further development of the hardware and software system authors plan to move from the *image classification* task to the *object detection* task.

**Gratitude.** This work was carried out with the financial support by the project No. 3.2.4.1 of the Union State programme “Technology-SG”.

### References

[1] Kovalev V.A., Paulenka D.A., Snezhko E.V., Liauchuk V.A. Comparative analysis of budget computing platforms for a portable micromodule of on-board image classification // BIG DATA and Advanced Analytics: collection of materials of the fourth international scientific and practical conference. (Minsk, Belarus, May 3 – 4, 2018) / editorial board: M. Batura [etc.]. – Minsk, BSUIR, 2018. – pp. 31–42.

[2] Сравнительный анализ вычислительных платформ для бортового микро модуля предварительного распознавания изображений / В.А. Ковалев [и др.] // Информатика. – ISSN 1816-0301 (print). ISSN 2617-6963 (online). 2018. – Т. 15, № 3. – С. 7–21.

[3] MobileNetV2: Inverted Residuals and Linear Bottlenecks, Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L.C. arXiv preprint. arXiv:1801.04381, 2018.

[4] Интеллектуальная технология распознавания подстилающей поверхности Земли / С.В. Кругликов, В.А. Ковалев, Д.А. Павленко, Э.В. Снежко, В.А. Левчук // Радиоэлектронные технологии. – 2019. – № 1. – С. 90–94.

## **РАСПОЗНАВАНИЕ ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ ЗЕМЛИ С ПОМОЩЬЮ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ НА ОДНОПЛАТНОМ МИКРОКОМПЬЮТЕРЕ**

<b>Д.А. Павленко</b> <i>Инженер-программист 1 кат. ОИПИ НАН РБ</i>	<b>В.А. Ковалев</b> <i>Кандидат технических наук, зав. лаб. анализа биомедицинских изображений ОИПИ НАН РБ</i>	<b>Э.В. Снежко</b> <i>Кандидат технических наук, ведущий научный сотрудник ОИПИ НАН РБ</i>
---	---	---

<b>В.А. Левчук</b> <i>Кандидат технических наук, научный сотрудник</i>	<b>Е.И. Печковский</b> <i>Ведущий инженер- программист</i>
---	---

*Объединенный институт проблем информатики (ОИПИ) Национальной академии наук Беларуси (ОИПИ НАН РБ, ул. Сурганова 6, 220012, Минск, Республика Беларусь)*

**Аннотация.** Приводятся результаты разработки программно-аппаратного комплекса (микромодуля) по обнаружению и классификации изображений подстилающей поверхности Земли. Полученное устройство имеет размеры 5,2×7,4×3,1 см, массу 52 г и использует сверточную нейронную сеть на основе архитектуры MobileNetV2 для классификации изображений. Микромодуль может использоваться на борту малых космических аппаратов либо легких беспилотных летательных аппаратов (дронов). Приведенные в статье сведения могут быть полезны инженерам и научным работникам, разрабатывающим компактные бюджетные мобильные системы обработки, анализа и распознавания изображений.

**Ключевые слова:** распознавание изображений, сверточная нейронная сеть, глубокое обучение, одноплатный микрокомпьютер, edge device, мобильная система

UDK 004.42

## DIGITAL TWIN FOR PV PLANT'S POWER GENERATION ANALYSIS



**R.M. Asimov<sup>1</sup>**,  
PhD, CEO of  
Sensotronica Ltd



**S.V. Valevich<sup>3</sup>**  
Graduate student BSUIR



**I. Kruse<sup>2</sup>**,  
CEO of SunSniffer GmbH



**V.S. Asipovich<sup>3</sup>**  
PhD, associate professor BSUIR

<sup>1</sup>Sensotronica Ltd, Kulman, 9, 373, Minsk, 220010, Republic of Belarus. E-mail: roustam.asimov@sensotronica.com

<sup>2</sup>SunSniffer GmbH & Co. KG, Ludwig-Feuerbach-Str. 69, Nuremberg 90489, Germany. E-mail: ingmar.kruse@sunsniffer.de

<sup>3</sup>Belarussian State University of Informatics and Radioelectronics, P. Brovki, 6, Minsk, 220013, Republic of Belarus. E-mail: v.osipovich@bsuir.by

### **R.M. Asimov**

В 1990 г. Окончил Таджикский Государственный Университет, факультет физики. Занимался исследовательской работой в институте Физики НАН Беларуси и университете Paris-Nord, Франция. В 2012 г. Защитил диссертацию по теме «лазерно-индуцированной фотодиссоциации комплексов гемоглобина». Является автором 134 научных работ, 14 патентов, 7 учебно-методических пособий. В настоящее время возглавляет компанию, резидент Белорусского Парка Высоких Технологий ООО «Сенсотроника». Направления научной деятельности: разработка алгоритмов и технологий обработки больших данных, математическое моделирование фотофизических процессов.

### **S.V. Valevich**

In 2017 graduated from the Belarusian State University of Informatics and Radioelectronics with degree in Engineering. In 2018 got Master's degree in Engineering.

### **I. Kruse**

Since the early 80s he develops innovative technology. In 1983, a year before his graduation, he founded his first company. He studied Business Administration in Nuremberg and Computer Science in Atlanta, USA. In 1996 he received a rare license from Apple for the production of Apple computers. Since 2002 he is in photovoltaic industry. He holds several patents in different fields, including photovoltaics.

### **V.S. Asipovich**

In 2004, he graduated from the Belarusian State University of Informatics and Radioelectronics with a degree in Microelectronics, and in 2005, he got a master's degree in the same specialty. In 2010, he defended the thesis for the degree of a candidate of technical sciences in Devices, systems and medical items. He is the author of 115 publications,

*2 patents, 12 teaching aids. Annually he orally reports at international scientific forums. Main areas of research activity: development of algorithms and technologies for big data processing, research and development in the software processing of medical images.*

**Abstract.** Summary. A method of PV array-based calculation has been proposed, implemented and tested. The results showed the following. Array-based calculation is able to provide quite accurate results in Pmpp values for array, but defective modules with some electrical issues in particular array could be identified only with some additional module-based analysis.

**Key words:** Digital twin, Photovoltaic, Photovoltaic Array Calculation, Defective Photovoltaic Modules.

### ***Introduction***

Global solar energy market growth resides around 30% per year. Under optimal conditions, the world's solar generation plant capacity could reach up to 1,270.5 GW by the end of 2022 [1].

For solar energy cost-effectiveness and predictable power generation play a key role in any PV installations. On the other hand, the same module will behave differently in another location, with various weather conditions, climate, dust and shadow conditions of a particular cell, and so on. It's quite hard to estimate how all these factors will affect the cell's lifetime and its efficiency. Also, some electrical issues may appear and it's almost impossible to predict them.

All of the above issues may be handled using monitoring via multiple sensors on PV module arrays, on cells, some external sensors, and systems that aggregate all the data. Monitoring helps to understand in a timely manner when anything on the plant requires maintenance, which will result in reduced operating costs.

Typical PV plant consists of multiple PV module arrays, and each module requires its own sensors, devices, and additional handling like regular cleaning procedures, especially for dusty areas [2]. Usually, PV equipment allows predicting plant's power generation using some maximum power point tracking (MPPT) methods combined with the raw data (temperature, irradiation, output params like the voltage, current, and power) from sensors located on each module.

The issue occurs when there's a lack of various equipment and sensors which may be suitable for remote monitoring and timely maintenance. It's true for most of the current active PV stations. Many existing papers and researches suggest methods which are suitable for single parameter or effect, so each individual plant needs to search and combine suitable devices, sensors and implement particular methods, while it would be great to have some platform which aggregates all params and effects together and provides extensive monitoring data. Digital Twin (DT) concept was proposed to fill this gap [3, 4]. It is a laboratory that accumulates data from sensors and allows us to monitor, predict, and fix various issues as soon as possible. DT consists of multiple modules which analyze all existing effects and factors around

PV module actual state. For instance, publication [3] demonstrates the ability to diagnose module states using DT.

However, the following issue occurs: there are plants without required voltage and temperature sensors on each module. In this case DT analysis for each module becomes impossible. Nevertheless, most of the PV plants are equipped with sensors that allow measuring temperature and irradiation for the whole plant and sensors for voltage and current parameters of module arrays.

This paper is aimed at the DT system's ability to detect defective modules or any other issues during PV plant operation based on telemetric data from module arrays. Currently DT calculates params by each module in order to analyze the technical plant state, but it requires temperature and voltage sensors located on each module to gather and analyze all the data which increases the plant's overall cost, especially for large PV plants.

### ***Average module calculation***

PV plant in Nürnberg, Germany, named Südstadt-Forum is used for data aggregation and calculations in this paper. Plant includes three inverters (SUN2000-20KTL, Sinvert PVM17, and Sinvert PVM20 models) with multiple strings (PV module arrays). Most of the strings consist of 18



PV monocrystalline modules. Each string and module provide various raw data from their sensors. Also ten additional devices for the whole plant are presented including SR05 pyranometer for temperature and irradiation.

In order to aggregate and prepare all the data software application written in Node.JS was used.

Digital Twin platform prepared API for module-by-module calculations based on input data. Input data includes the following parameters: voltage, current, temperature, irradiation from devices, temperature from devices, timestamp.

The output contains the following params: maximum power point (MPP), voltage and current at MPP, series and parallel resistance, short circuit current and open-circuit voltage params.

During the experiment, all the data from August 2018 was used; this month includes a lot of sunny days and provides more accurate results in the case of MPPT. Data was collected using Sunsniffer API.

The idea is to compare the results of calculations by each module and by each string (which is faster).

Input params for module-level calculation were: module voltage, string current, module temperature, and irradiation. The time alignment between the module readings and the SR05 pyranometer, made using the timestamps of the individual data points.

Input params for average module calculation (based on string-level measurement) included average module voltage  $U_{avg}$ , string current, and average module temperature  $T_{avg}$ , temperature, and irradiation from SR05 pyranometer.

$$U_{avg} = \frac{U_s}{n},$$

where  $U_s$  – string voltage,  $n$  – count of PV-modules in this string.

$$T_{avg} = \frac{\sum_i^n T_n}{n},$$

where  $T_n$  – temperature of module,  $n$  – count of modules in this string.

Based on output params additional fields were calculated:

$$P_{mpp\ diff} = \frac{\sum_i^n P_{mpp\ module\ i}}{n} - P_{mpp\ string},$$

where  $P_{mpp\ module\ i}$  – maximum power of module,  $P_{mpp\ string}$  – maximum power of average string.

$$P_{mpp\ diff\ percentage} = \frac{|P_{mpp\ diff}|}{P_{mpp\ string}} \%$$

$$P_{Pmpp\ avg} = P_{month} \frac{n}{\sum_i^n P_{mpp\ module\ i}} \%,$$

where  $P_{month}$  – power, produced by string during the month.

$$P_{Pmpp\ string} = \frac{P_{month}}{P_{mpp\ string}} \%,$$

$$P_{percentage\ diff} = P_{Pmpp\ avg} - P_{Pmpp\ string}$$

All parameters were calculated once per month for each string and each module. Data dynamics were analyzed based on time periods and params changes between the strings.

For the next analysis and hypothesis verification some criteria were required in order to determine defective modules. The following parameters were introduced:

$$P_{mpp\ delta} = P_{mpp\ max} - P_{mpp\ min},$$

where  $P_{mpp\ max}$  – maximum power, produced by module across the string during the month,  $P_{mpp\ min}$  – minimum power, produced by module across the string during the month.

$$P_{mpp\ delta\ \%} = \frac{P_{mpp\ delta}}{P_{mpp\ max}} \%,$$

$P_{mpp\ delta}$  allows to find out if there're some defective modules in strings more precisely using max and min module power parameters.

Some of the devices are synchronized and provide data with exactly the same timestamps, and others may vary a bit, so mapping between different sources includes finding the nearest data points.

On average, one new data point is acquired every 7 minutes. For example, Module 1.1\_1 has 3943 points during August after all the filtering on the app side. Most of the filtering stays inside DT calculation except some simple app-side preprocessing like removing points with invalid temperatures, points with missing parts of the data, and so on.

Aggregated results for individual modules, and average modules are passed to DT calculation API.

#### **Calculation analysis**

Diagrams of  $P_{mpp}$  distribution by modules during August 2018 are presented on figures 1, 2 (String 1.6 and String 2.2 for example).

On figure 2 it's clearly visible that String 2.2 distribution looks abnormal, we compared some valid string (String 1.6) with this one which has defective module (String 2.2).

For String 1.6 on string level average  $P_{mpp} = 173W$ , on module level average  $P_{mpp}$  across modules = 178W, min  $P_{mpp} = 174W$ . Difference between string and average module level  $P_{mpp} = 2.8\%$ .

For String 2.2 on string level average  $P_{mpp} = 175W$ , on module level average  $P_{mpp}$  across modules = 181W, min  $P_{mpp} = 135W$ . Difference between string and average module-level  $P_{mpp} = 3.6\%$ .

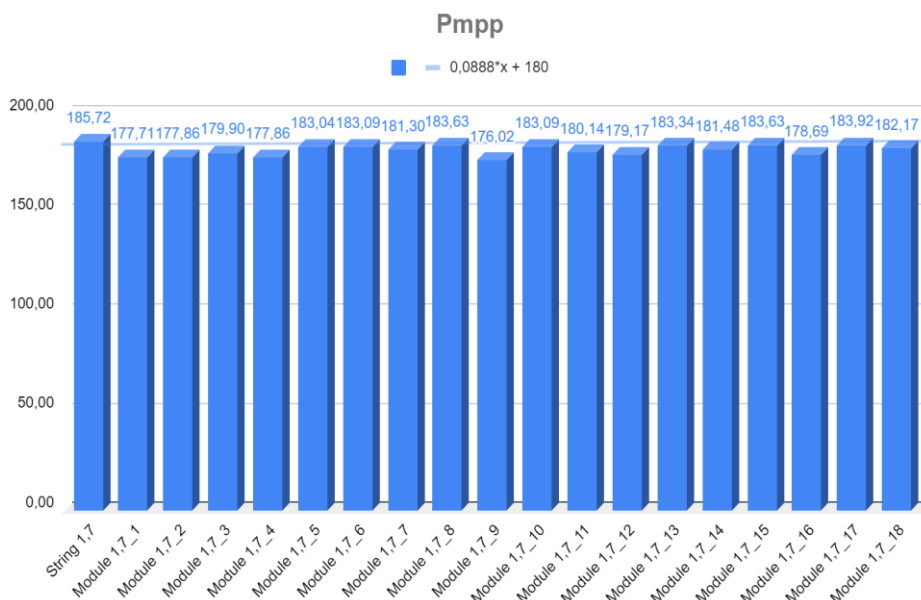


Figure 1. – Distribution of String 1.6 module power in August, 2018

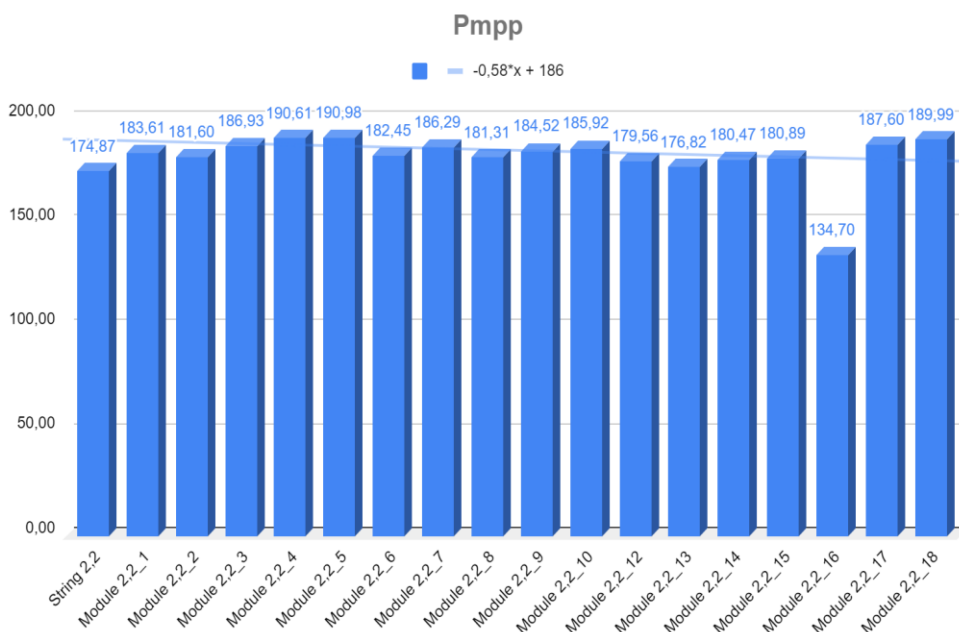


Figure 2. – Distribution of String 2.2 module power in August, 2018

Statistical distribution is presented in figures 3 and 4 for String 1.6 and 2.2. For String 2.2 distribution appears more uniform and overall  $P_{mpp}$  is higher despite defective module, results of calculation on average string level correlate with it. Therefore the average string method allows estimating the strings's state adequately.

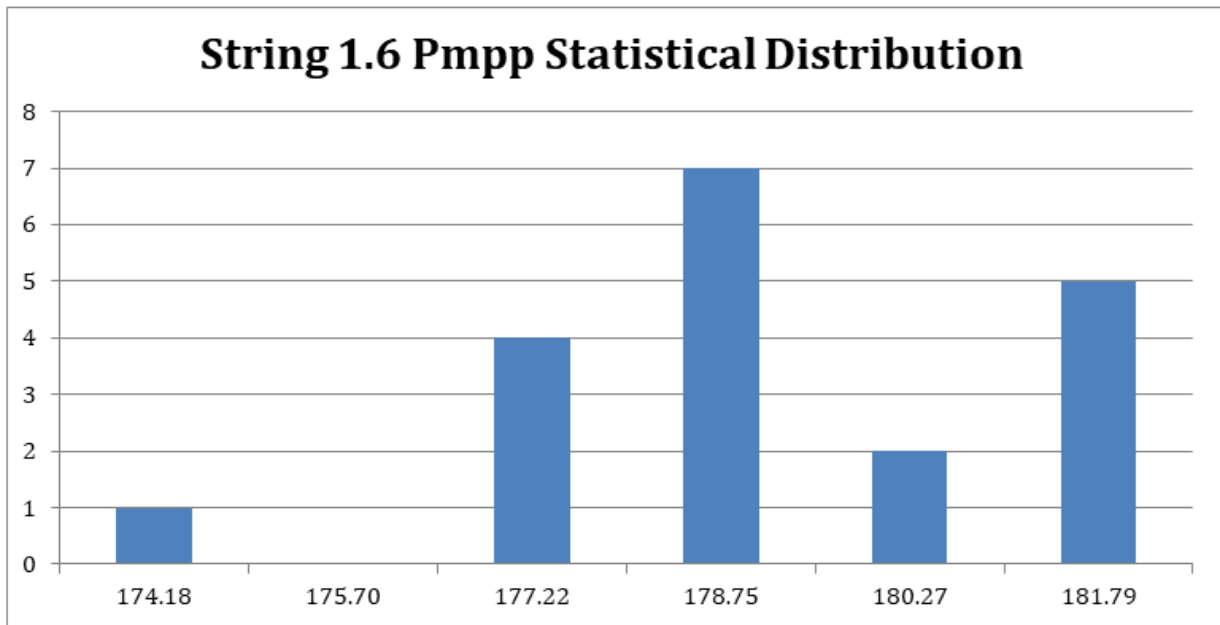


Figure 3. – String 1.6 power statistical distribution in August 2018

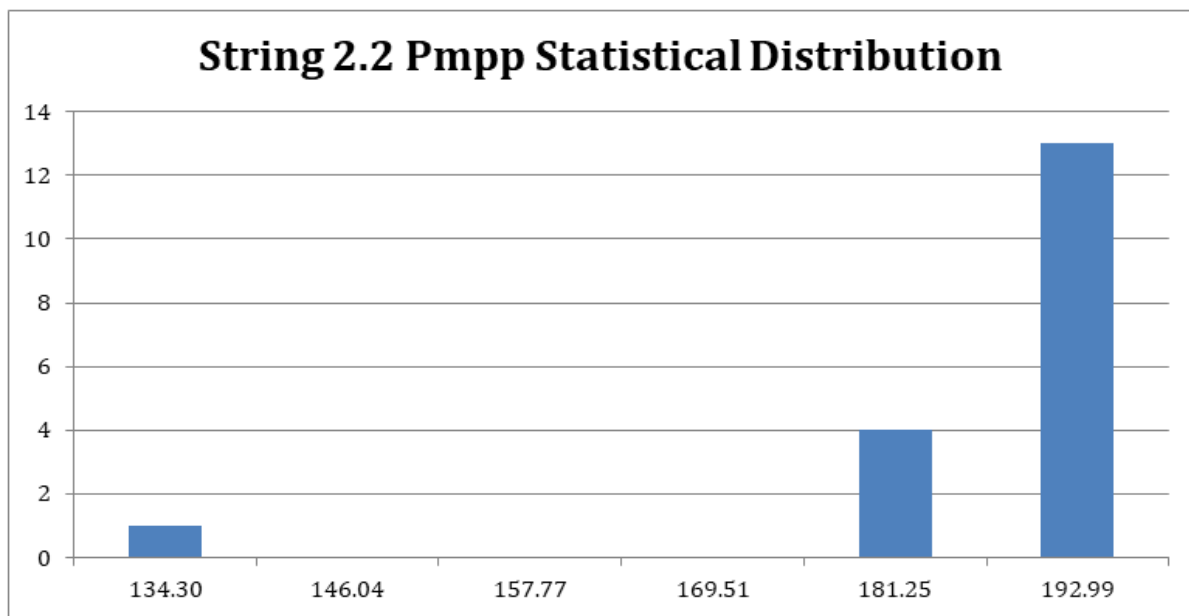


Figure 4. – String 2.2 power statistical distribution in August 2018

More detailed analysis of Module 2.2\_16 which contains voltage dynamics for this module is presented on figure 5.

It is clearly visible that the voltage on Module 2.2\_16 falls to zero. That is an indication that module underperforms, and the bypass diodes activated under high current conditions. Either shadowing or defective module can cause such behavior.

Additional analysis showed that for Module 2.2\_16  $R_p = 50$ ,  $R_s = 1.7$  while valid modules have values around  $R_p = 400$ ,  $R_s = 0.5$ .

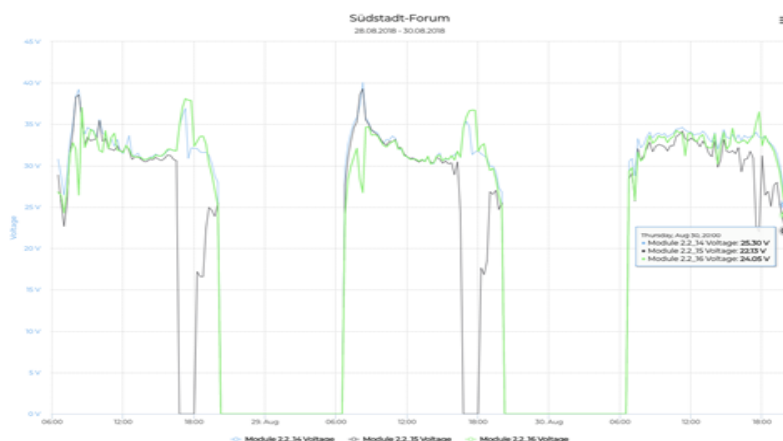


Figure 5. – String 2.2 Module voltages at DT calculation day (29.08.2018)

DT calculation results show that Module 2.2\_16 has significantly higher series and lower parallel (shunt) resistance. That is an indication of an electrical defect of the module.

Also, it is clearly visible that the parameter of the virtual module that describes string level data, does not allow us to detect the electrical failure within the string.

Further analysis of average string parameters dynamics during the half-year period from June till November was carried out.

The following parameters were gathered during the calculations for each separate month:  $P_{mpp\ diff}$ ,  $P_{mpp\ diff\ percentage}$ ,  $P_{Pmpp\ avg}$ ,  $P_{Pmpp\ string}$ ,  $P_{percentage\ diff}$ ,  $P_{mpp\ delta}$ ,  $P_{mpp\ delta\ \%}$ ,  $P_{mpp\ diff\ percentage}$  and  $P_{percentage\ diff}$  values are presented in Tables 1, 2. N/A values (String 1.4) used when data isn't available for this period. November results look less accurate due to low temperatures during this month.

Table 1. –  $P_{mpp\ diff\ percentage}$  values

	June	July	August	September	October	November
String 1.1	1,08%	2,35%	1,89%	6,70%	0,95%	4,47%
String 1.2	2,43%	2,75%	2,41%	0,06%	1,73%	11,19%
String 1.3	4,45%	2,76%	2,57%	3,88%	3,11%	27,30%
String 1.4	0,15%	1,39%	N/A	13,81%	N/A	N/A
String 1.5	3,08%	4,03%	3,69%	7,22%	4,12%	23,21%
String 1.6	3,14%	1,51%	2,57%	2,62%	0,42%	0,75%
String 1.7	5,35%	1,88%	3,16%	2,06%	3,74%	4,03%
String 1.8	2,74%	1,16%	2,66%	2,12%	3,57%	6,6%
String 1.9	1,73%	2,76%	0,22%	0,35%	0,07%	4,01%
String 1.10	3,10%	2,45%	1,67%	3,36%	4,78%	5,40%
String 1.11	1,27%	3,25%	2,74%	2,51%	3,60%	1,59%
String 1.12	3,19%	2,17%	3,55%	2,96%	2,93%	4,41%
String 2.1	0,88%	1,20%	2,53%	1,37%	1,48%	2,88%
String 2.2	3,00%	2,25%	3,64%	1,47%	1,44%	1,30%
String 2.3	1,73%	1,90%	2,81%	2,27%	5,93%	1,59%
String 2.4	3,67%	2,44%	3,18%	5,15%	0,19%	6,55%

Table 2. –  $P_{percentage\ diff}$  values

	June	July	August	September	October	November
String 1.1	-0,51%	-0,79%	-0,72%	-3,77%	-0,69%	-8,49%
String 1.2	-1,16%	-0,92%	-0,92%	-0,03%	-1,14%	18,37%
String 1.3	1,95%	0,86%	0,90%	1,91%	2,06%	45,57%
String 1.4	0,21%	0,45%	N/A	-7,33%	N/A	N/A
String 1.5	1,35%	1,24%	1,27%	3,36%	2,39%	36,69%
String 1.6	1,45%	0,49%	0,94%	1,31%	-0,27%	-19,49%
String 1.7	-2,72%	-0,64%	-1,25%	-1,10%	-2,53%	-7,42%
String 1.8	-1,36%	-0,40%	-1,04%	-1,13%	-2,41%	-21,36%
String 1.9	0,82%	-0,95%	-0,08%	-0,18%	0,04%	-6,97%
String 1.10	-1,51%	-0,83%	0,63%	-1,78%	-3,14%	-9,44%
String 1.11	0,59%	1,02%	0,97%	1,22%	2,10%	2,44%
String 1.12	-1,55%	-0,73%	-1,35%	-1,53%	-1,83%	-7,01%
String 2.1	-0,68%	-0,41%	-0,96%	-0,71%	-0,90%	-4,48%
String 2.2	1,41%	0,74%	1,35%	0,75%	0,87%	1,97%
String 2.3	0,83%	0,63%	1,04%	1,14%	3,44%	2,34%
String 2.4	-1,83%	-0,84%	-1,21%	-2,93%	0,12%	-11,56%

Table 3 includes additional data about specific month (August for example), e.g. produced power for each String, all  $P_{mpp}$  values. All result params are combined together for better view. Figure 6 presents  $P_{percentage\ diff}$  values from the table.

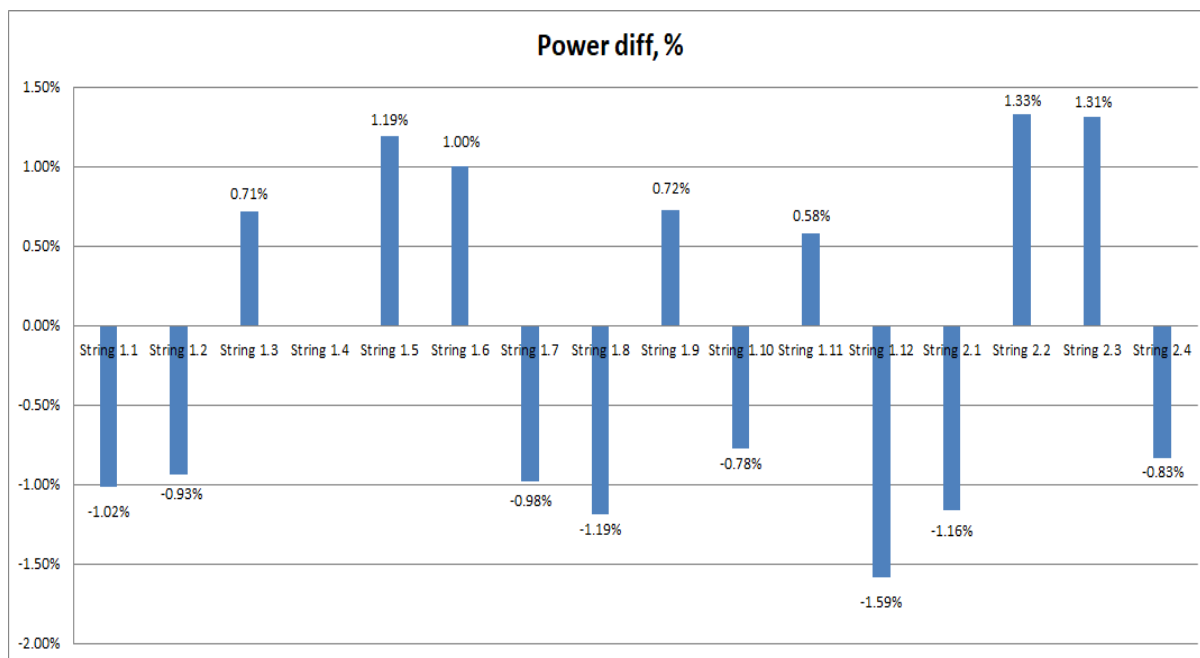


Figure 6. – Diagram of the difference between power’s produced by String for August division by two  $P_{mpp}$  types (by modules and average by strings)

For Figure 6, there are top 3 strings - 1.5, 2.2, and 2.3. The two highest values belong to 2.2 and 2.3, which have defective modules 2.2\_16 and 2.3\_10 (clearly visible by  $P_{mpp}$  diagrams for each String above).

Table 3. – Calculation results compared together with power produced by each String during the August

String name	String power, kWh	$P_{mpp}$ avg by modules	$P_{mpp}$ string	$P_{mpp}$ diff	$P_{mpp}$ diff percentage	$P_{Pmpp}$ avg	$P_{Pmpp}$ string	$P_{percentage}$ diff
String 1.1	491.61	176.3	181.3	-5	-2.76%	35.86%	36.88%	-1.02%
String 1.2	503.17	180.2	184.9	-4.7	-2.54%	35.81%	36.75%	-0.93%
String 1.3	519.13	180.8	177.1	3.7	2.09%	34.83%	34.11%	0.71%
String 1.4	504.16	179.6	N/A	N/A	N/A	35.62%	N/A	N/A
String 1.5	520.69	179.6	173.4	6.2	3.58%	34.49%	33.30%	1.19%
String 1.6	490.81	178	173.1	4.9	2.83%	36.27%	35.27%	1.00%
String 1.7	490.71	180.9	185.7	-4.8	-2.58%	36.86%	37.84%	-0.98%
String 1.8	480.34	177.6	183.3	-5.7	-3.11%	36.97%	38.16%	-1.19%
String 1.9	483.16	177.6	174.1	3.5	2.01%	36.76%	36.03%	0.72%
String 1.10	489.91	178	181.8	-3.8	-2.09%	36.33%	37.11%	-0.78%
String 1.11	516.682	182.6	179.6	3	1.67%	35.34%	34.76%	0.58%
String 1.12	516.617	182.8	191	-8.2	-4.29%	35.38%	36.97%	-1.59%
String 2.1	515.51	184.7	190.7	-6	-3.15%	35.83%	36.99%	-1.16%
String 2.2	488.85	181.4	174.9	6.5	3.72%	37.11%	35.78%	1.33%
String 2.3	481.531	176.2	169.9	6.3	3.71%	36.59%	35.28%	1.31%
String 2.4	517.342	185.1	189.4	-4.3	-2.27%	35.78%	36.61%	-0.83%

For more accurate detection of defective modules additional params were used:  $P_{mpp}$  delta and  $P_{mpp}$  delta % (tables 4, 5).

Table 4. –  $P_{mpp\ delta}$  values

	June	July	August	September	October	November
String 1.1	7,55	7,66	8,77	25,67	12,71	20,06
String 1.2	15,08	18,89	9,11	9,9	14,87	16,86
String 1.3	8,94	17,94	7,17	7,08	14,52	27,69
String 1.4	12,81	13,41	12,1	9,57	N/A	N/A
String 1.5	9,62	15,98	8,69	8,62	9,7	19,35
String 1.6	12,02	6,31	6,06	7,26	9,6	14,98
String 1.7	15,71	10,48	6,8	5,36	9,22	11,5
String 1.8	9,9	12,73	11,95	8,36	12,39	11,7
String 1.9	8,79	5,07	6,04	5,91	13,85	20,41
String 1.10	8,51	12,04	4,72	5,77	13,84	11,86
String 1.11	13,74	10,94	9,64	7,83	7,86	101,41
String 1.12	6,98	11,58	8,05	5,08	9,68	95,64
String 2.1	44,12	11,38	13,05	10,85	13,17	11,67
String 2.2	12,72	61,67	58,7	9,73	9,25	22,91
String 2.3	54,32	51,2	25,55	40,01	55,3	65,8
String 2.4	3,99	11,95	8,74	9,82	9,95	9,38

Table 5. –  $P_{mpp\ delta\ \%}$  values

	June	July	August	September	October	November
String 1.1	4,24%	4,31%	4,87%	14,05%	6,87%	10,76%
String 1.2	8,26%	10,09%	4,96%	5,29%	7,96%	9,11%
String 1.3	4,91%	9,61%	3,88%	3,80%	7,64%	13,69%
String 1.4	7,11%	7,26%	6,51%	5,13%	N/A	N/A
String 1.5	5,31%	8,62%	4,77%	4,68%	5,22%	10,87%
String 1.6	6,64%	3,55%	3,33%	3,91%	5,12%	7,80%
String 1.7	8,48%	5,74%	3,69%	2,85%	4,91%	6,09%
String 1.8	5,48%	6,97%	6,53%	4,52%	6,78%	6,27%
String 1.9	4,87%	2,88%	3,35%	3,24%	7,27%	11,08%
String 1.10	4,78%	6,59%	2,61%	3,16%	7,50%	6,39%
String 1.11	7,33%	5,86%	5,14%	4,23%	4,25%	55,42%
String 1.12	3,82%	6,21%	4,29%	2,72%	5,18%	51,63%
String 2.1	20,02%	5,99%	6,77%	5,67%	6,83%	6,09%
String 2.2	6,76%	32,04%	30,43%	5,16%	4,87%	11,92%
String 2.3	30,42%	28,64%	14,19%	21,83%	29,91%	36,37%
String 2.4	2,20%	6,27%	4,63%	5,23%	5,23%	4,95%

Based on the analysis of tables 4, 5 defective strings were identified. Additional verification of DT calculation using internal electrical parameters for each module confirmed existing of defective modules and allowed to create the following list.

List of known strings with defective modules during June - November period:

1) June - String 2.1 (Module 2.1\_11,  $P_{mpp} = 220,38W$ ,  $P_{mpp\ diff\ percentage} = 0,88\%$ ,  $R_p = 1000$ ,  $R_s = 0,9$ );



- 2) June - String 2.3 (Module 2.3\_10,  $P_{mpp} = 124,26$  W,  $P_{mpp}$  diff percentage = 1,73%,  $R_p = 95$ ,  $R_s = 2,49$ );
- 3) July - String 2.2 (Module 2.2\_16,  $P_{mpp} = 130,8$  W,  $P_{mpp}$  diff percentage = 2,25%,  $R_p = 95$ ,  $R_s = 1,89$ );
- 4) July - String 2.3 (Module 2.3\_10,  $P_{mpp} = 127,6$  W,  $P_{mpp}$  diff percentage = 1,9%,  $R_p = 185$ ,  $R_s = 2,09$ );
- 5) August - String 2.2 (Module 2.2\_16,  $P_{mpp} = 134,2$  W,  $P_{mpp}$  diff percentage = 3,64 %,  $R_p = 95$ ,  $R_s = 1,69$ );
- 6) August - String 2.3 (Module 2.3\_10,  $P_{mpp} = 154,56$  W,  $P_{mpp}$  diff percentage = 2,81 %,  $R_p = 95$ ,  $R_s = 1$ );
- 7) September - String 2.3 (Module 2.3\_10,  $P_{mpp} = 143,25$  W,  $P_{mpp}$  diff percentage = 2,27 %,  $R_p = 500$ ,  $R_s = 1,69$ );
- 8) October - String 2.3 (Module 2.3\_10,  $P_{mpp} = 129,6$  W,  $P_{mpp}$  diff percentage = 5,93 %,  $R_p = 140$ ,  $R_s = 2,68$ );
- 9) November - String 1.11 (Module 1.11\_15,  $P_{mpp} = 71,59$  W,  $P_{mpp}$  diff percentage = 1,59 %,  $R_p = 50$ ,  $R_s = 4,56$ );
- 10) November - String 1.12 (Module 1.12\_11,  $P_{mpp} = 89,61$  W,  $P_{mpp}$  diff percentage = 4,41 %,  $R_p = 95$ ,  $R_s = 4,17$ );
- 11) November - String 2.3 (Module 2.3\_10,  $P_{mpp} = 115,14$  W,  $P_{mpp}$  diff percentage = 1,59%,  $R_p = 95$ ,  $R_s = 3,57$ ).

Values from Table 1 for these defective strings ( $P_{mpp}$  diff percentage) look like average values across all strings (especially for String 2.3 which was defective during all 6 months) and cannot definitely identify string with a defective module.

However, with additional analysis for each module or with  $P_{mpp}$  delta % parameters defective modules could be easily identified (yellow values from table 5).

Dynamics analysis for  $P_{percentage}$  diff and  $P_{mpp}$  diff percentage values (tables 1, 2) in comparison with  $P_{mpp}$  delta and  $P_{mpp}$  delta % (tables 4, 5) shows that the relationship between these parameters and defective modules is not clear.

### Conclusions

Digital Twin estimates the string's state with 0,6 – 3,5% variation. It could be used during analysis of the actual PV plant's state by comparing real parameters with those calculated from DT.

However, the parameter of the virtual module that describes String level data does not allow us to detect the electrical failure within the string.

### References

- [1.] M. Schmela, A. Beauvals // Global Market Outlook For Solar Power 2018-2022 // SolarPower Europe, Brussels, 2018
- [2.] S.A. Sharaf, M.S. Abd-Elhady, H.A. Kandil // Feasibility of solar tracking systems for PV panels in hot and cold regions // Renewable Energy, Volume 85, 2016, pp. 228-233.
- [3.] Asimov R.M., Valevich S.V., Kruse I., Asipovich V.S. Virtual laboratory for testing of solar power plants in big data analysis // Collection of materials of the V International Scientific and Practical Conference «BIG DATA and ADVANCED ANALYTICS», March 13–14, 2019, Minsk, BSUIR, pp. 61–65.
- [4.] Osipovich V.S., Asimov R.M., Chernoshey S.V. Digital twin in the Analysis of a Big Data // Collection of materials of the IV International Scientific and Practical Conference «BIG DATA and ADVANCED ANALYTICS», May 3–4, 2018, Minsk, BSUIR, pp. 69–78.

## ИСПОЛЬЗОВАНИЕ ЦИФРОВОГО ДВОЙНИКА СОЛНЕЧНОЙ ПАНЕЛИ ДЛЯ АНАЛИЗА ВЫРАБОТКИ ЭЛЕКТРОЭНЕРГИИ СОЛНЕЧНОЙ ЭЛЕКТРОСТАНЦИЕЙ

**Р.М. Азимов**  
Директор  
Sensotronica Ltd

**С.В. Валевиц**  
Магистрант  
кафедры ИПиЭ

**И. Круз**  
Директор  
SunSniffer

**В.С. Осипович**  
Доцент кафедры  
ИПиЭ

**Аннотация.** Предложен, реализован и протестирован метод расчета на уровне массива панелей. Результаты тестирования показали следующее. Расчет на уровне массива панелей позволяет получить достаточно точные результаты по значениям  $P_{mpp}$  для массива, но для определения дефектных панелей с какими-то электрическими неисправностями в конкретном массиве необходим дополнительный анализ на уровне конкретных панелей

**Ключевые слова:** Цифровой двойник, фотоэлектрический, расчет для фотоэлектрического массива, дефектные фотоэлектрические панели.

UDK 004.6:004.632

## NOSQL DATABASES. TECHNOLOGY FOR PROTECTING DATA FROM UNAUTHORIZED ACCESS



**A.V. Kuchynski**  
*Student of Belarusian State University of Informatics and Radioelectronics. Software engineer IBA-Group.*



**U.N. Hutkouski**  
*Student of Belarusian State University of Informatics and Radioelectronics. Software engineer IBA-Group.*



**I.I. Piletski**  
*PhD, Associate Professor of Informatics Department of the BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus  
IBA-Group, Republic of Belarus  
E-mail: alexkuchinskydev@gmail.com*

### **A.V. Kuchynski**

*Pre-graduate student at Belarusian State University of Informatics and Radioelectronics. Certified AWS Architect, works as a Big Data engineer at IBA Group. Focuses on cloud computing, distributed frameworks, security of systems.*

### **U.N. Hutkouski**

*Pre-graduate student at Belarusian State University of Informatics and Radioelectronics. Works as a project manager at IBA Group. Interested in full stack programming, high-availability and elastic architectures.*

### **I.I. Piletski**

*PhD of Maths, associate professor of BSUIR. In the field of IT for more than 47 years. Participation in the development of several dozen major projects: the chief designer of the project, the chief architect of software and information support, the project manager, the head of the department, the head of the laboratory (Scientific Research Institute of Computer Science, Academy of Sciences of Belarus, IBA, BSUIR). The author of dozens of research and publications.*

**Abstract.** Companies such as banks, medical facilities, government units does not want their data was available and used by third parties (It can be result of hackers' attacks, data leaks, and employee fraud). Just write encrypted data to your data storage is not enough. This data storage has to allow business users to work with its content performing CRUD (create, read, update, delete) operations and search queries. During encryption, we are using some function  $F(x)$  to change content of our data, getting cyphertext as a result, after encryption takes place we lose this ability of performing search and basic operations. We are proposing approach, which allows managing encrypted data by executing queries on it without decryption, results returned in encrypted form to client where he/she can decrypt it using self-managed keys.

**Keywords:** NoSQL, graph database, protection from unauthorized access, security.

**Introduction.** Basically there are two states of data when it's vulnerable for potential thieves:

1. Data at rest – is data that is not actively moving from device to device or network to network such as data stored on a hard drive, laptop, flash drive, or archived/stored in some other way. Data protection at rest aims to secure inactive data stored on any device or network, so in case if somebody physically takes disks from the computer they will not be able to reveal the data stored on it.

2. Data in transit - or data in motion, is data actively moving from one location to another such as across the internet or through a private network. Data protection in transit is the protection of this data while it's traveling from network to network or being transferred from a local storage device to a cloud storage device

There are two places where data can be encrypted:

1. On the client side (client-side encryption) - information is encrypted before it's send to database. Client itself chooses encryption algorithm and manages encryption keys. Data is transferred and stored already encrypted.

2. On the server side (server-side encryption) - server performs encryption of the data when it's written and decryption when it's read. In this case encryption keys can be stored by client or by server.

Choosing the place where data is encrypted defines security level:

1. Level 0. No encryption
2. Level 1. Server side encryption, server keeps keys.
3. Level 2. Server side encryption, keys are kept by client
4. Level 3. Client-side encryption.

*Challenges of search in encrypted NoSQL storages.* Firstly, when encryption takes place and we write encrypted data into database, we lose ability of standard search over this data (using of indexes). Such operations as: greater than, less than, equals, in range, etc. become not supportable as columns contains ciphertext [1].

Secondly, specification of NoSQL databases have to be considered. In this types of storages search is performed based on the key (almost all NoSQL databases are key-value storages). Value which is mapped to the key can be unstructured data (or I would prefer to name it multistructured), so there is no definite database scheme as in relational databases. So there is a question: over which fields exactly we are searching over? Because most of the times all business crucial information is contained in the key (several attributes can be concatenated) [1, 2].

It's also worth mentioning that special tools do exist for performing column search. This services are built on the top of the database, maintain their own indexes, define their own data schemas and mappings, provide separate interface for querying data (Example: Solr, Phoenix).

*Basic principles.* During construction process of the system which supports search over encrypted data it would be preferable to keep up in mind the following principles:

– Transparency – implementation of encryption/decryption mechanism and mechanism for search over encrypted data shouldn't require changes of existing components in the system or if changes takes place it shouldn't be huge (shown on figure 1).

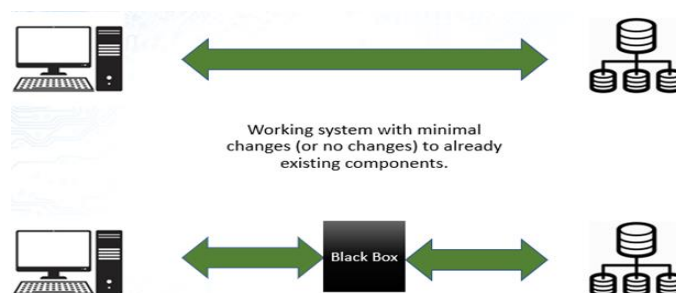


Figure 1. – Illustration of transparency concept

– SQL-like queries. We discussed previously search constraints of NoSQL (key-value) databases. In this types of storages we can search specific entities only by key or part of the key. Examples of get methods for reading from NoSQL database:

- `table.get("row_idificator")`
- `table.scan()`
- `table.scan("A*")`

Otherwise, SQL-like query allows to search over all columns:

- `SELECT * FROM People Where name like 'A*'`
- `SELECT * FROM People WHERE age < 25 and 0.14*year_salary > 5000`

– Proper type of algorithm. There are different classes of encryption algorithms which allows to perform specific operations on encrypted data:

- Homomorphic encryption algorithms - allows execute arithmetic operations on two encrypted values without decryption (without key)

- Order revealing encryption - allows execute comparison operations on two encrypted values without decryption (without key)

– Rationalism between security level of algorithm and space efficiency/performance.

*Proposed framework.* The main problem of encrypted data in database is: we cannot use default-indexing tool, which is based on simple values comparison. Theoretically, ORE (order revealing encryption) algorithms have special function `Compare(enc1, enc2)` which allows us to know which value is greater:

**Order Revealing Encryption (ORE):** Three algorithms:

$(sk, pk) \leftarrow \text{Keygen}$  outputs a secret key and a public “comparison” key

$c \leftarrow E_{sk}(x)$  outputs ciphertext

$b \leftarrow \text{Compare}(pk, c_1, c_2)$  outputs a bit

**Correctness:**  $x_1 \leq x_2 \Leftrightarrow \text{Compare}(E_{sk}(x_1), E_{sk}(x_2)) = 1$  (**w.h.p.**)

Our framework is aimed to make it possible to build indexes for encrypted data using graph database and overriding comparison function (`Compare(enc1, enc2)`) so this indexes can be balanced and searched. Graph database also stores data entities itself. Framework components (shown on Figure 2):

–Client – user application performing queries and CRUD operation.

–Proxy – proxy-server which handles incoming requests and responses. Performs encryption and decryption of query and response.

–Query Engine – parses query and as a result producing execution plan in form of tree where leaf - it's one condition (unit, condition like  $a < 5$ ); and node - logical primitives (and, or).

–Query Unit – library containing search functions (find less than, greater than, equals, etc.), functions for maintaining indexes in graph database (insert node, balance tree, etc.).

–Database – graph database.

Technical overview of the components:

–**Client Application** interacts with database via http protocol, currently available functionality:

- Write data to the database
- Send queries to the database

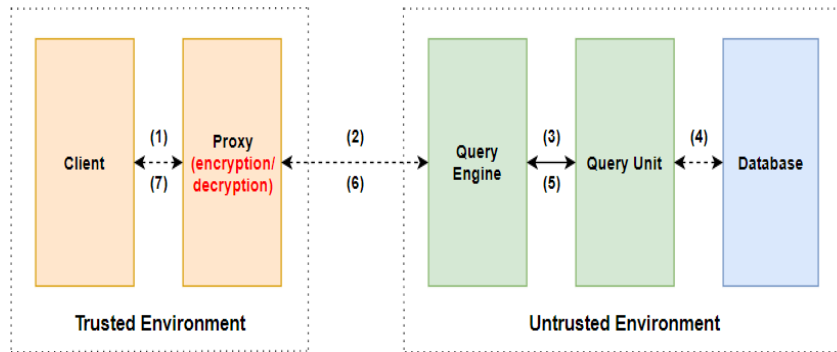


Figure 2. – High-level framework architecture

–**Proxy** checks incoming and outgoing responses. If there is a special symbols in url, it changes the content of the request (encrypts data to be inserted, encrypts query, decrypts the result of the query) and passes it further to destination or back to client (when it's response). Proxy takes care of encryption/decryption and key is now stored inside the proxy. Proxy also supports caching of encryption/decryption values, so encryption/decryption function will not be used if this value already was encrypted/decrypted in this session. If value is supposed to be indexed it's encrypted with use of OPE algorithm, else 3DES is used [3, 4].

–**Query Engine** builds tree and performs tree traversal executing each leaf, then each node which defines logic operation is applied on the results of each child. Leafs are runned in parallel. Query example: *select \* from db.table where (c < 10 and r > 90) or (a = 5 and (m in (9, 50) or k <= 0))*. Execution plan of this query in form of the tree shown on Figure 3.

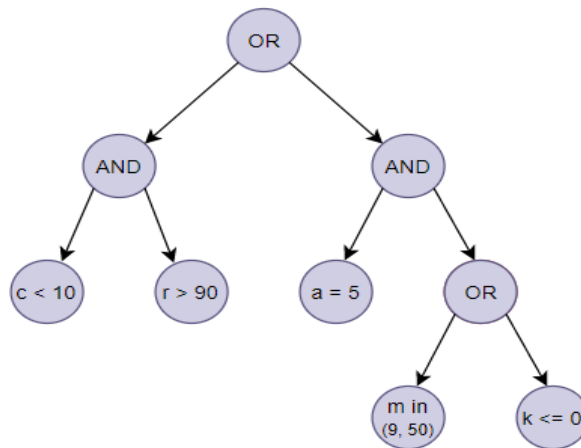


Figure 3. – Execution plan for the query.

–**Query Unit** is just a library of functions which are used by Query Engine and when insertion of new data takes place.

–**Database**. Indexes are built using AVL Tree data structure. It allows perform search for complexity  $\log(n)$  (shown on Figure 4).

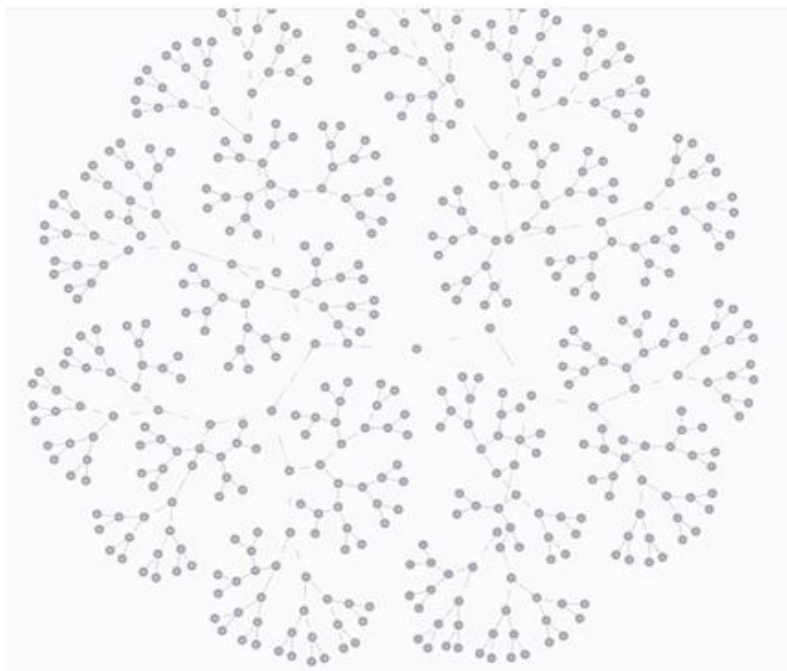


Figure 4. –View of AVL index tree

For each index column there is an index tree. Tree is updated on each insert operation (update, delete in future). Each node of the tree represents encrypted value. Nodes which are not in the tree - data entities. Inserted entity gets reference(s) to node in the tree (shown on Figure 5).

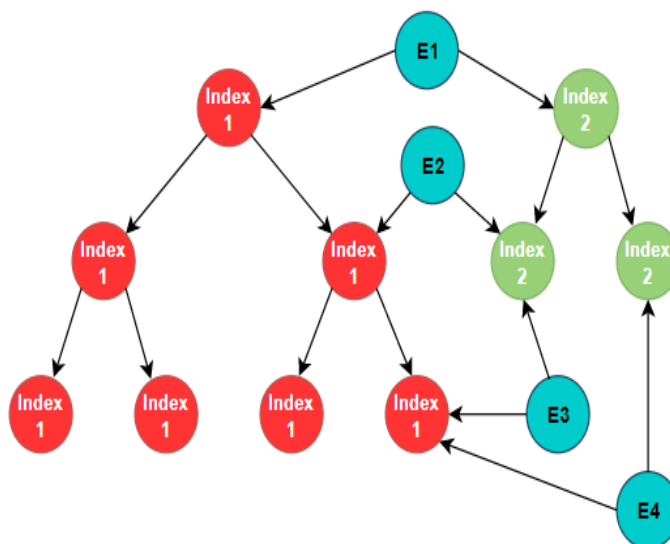


Figure 5. –View of index and entities nodes

**Conclusion.** Proposed framework allows performing search queries on encrypted data without need to decrypt it on query stage. Data is stored securely, thus cannot be used by unauthorized users. Architecture uses NoSQL graph database as a storage for primarily data and index structures. Order revealing encryption used for fields which have to be searched.

### **References**

- [1.]NoSQL Data Architecture & Data Governance: Everything You Need to Know [Электронный ресурс] / Режим доступа: <https://www.dataversity.net/nosql-data-architecture-data-governance-everything-need-know/> Дата доступа: 24.02.2020.
- [2.]The NoSQL Ecosystem [Электронный ресурс] / Режим доступа: <https://www.aosabook.org/en/nosql.html> Дата доступа: 24.02.2020.
- [3.]Secure Parallel Processing of Big Data Using Order-Preserving Encryption on Google BigQuery [Электронный ресурс] / Режим доступа: <https://arxiv.org/pdf/1608.07981.pdf> Дата доступа: 24.02.2020.
- [4.]What is 3DES encryption and how does DES work? [Электронный ресурс] / Режим доступа: <https://www.comparitech.com/blog/information-security/3des-encryption/> Дата доступа: 24.02.2020.

## **NOSQL DATABASES. TECHNOLOGY FOR PROTECTING DATA FROM UNAUTHORIZED ACCESS**

**А.В. Кучинский**  
Студент БГУИР.  
Инженер-программист  
IBA Group.

**В.Н. Гутковский**  
Студент БГУИР.  
Инженер-программист  
IBA Group.

**И.И. Пилецкий**  
Доцент кафедры  
информатики БГУИР,  
кандидат физико-  
математических наук,  
доцент, старший научный  
сотрудник

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь  
IBA-Group, Республика Беларусь  
E-mail: alexkuchinskydev@gmail.com*

**Аннотация.** Компании (банки, мед. учреждения, государственные компании) не хотят, чтобы их информация (в случае хакер-атаки, утечки данных, неосторожности обслуживающего персонала и т.д.) была доступна злоумышленниками и использовалась 3-й стороной. Просто сложить зашифрованные данные в хранилище недостаточно. Это хранилище должно использоваться и давать возможность работать со своим содержимым посредством обработки поисковых запросов. При шифровании мы меняем реальные значения на результат работы некой функции, одновременно с этим мы теряем возможность поиска по этим данным. Мы предлагаем подход, при котором можно бы было осуществлять поиск в зашифрованных данных без выполнения операции дешифрования при непосредственном поиске.

**Ключевые слова:** NoSQL, графовая база данных, защита от несанкционированного доступа, безопасность.



UDK 519.257

## OPTIMIZATION OF BUSINESS PROCESSES VIA BIG DATA



**I.N. Tsyrelchuk**  
Advisor to Rector, TUIT



**N. M. Mamatova**  
Deputy Dean of the joint  
faculty of the JFIT TUIT-  
BSUIR, Postgraduate student  
of the TUIT



**M. Y. Abdul-Azalova**  
Senior lecturer of the Information  
technologies Department of the  
TUIT, Postgraduate student of  
the TUIT

Joint faculty of JFIT TUIT-BSUIR, Republic of Uzbekistan  
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Republic of  
Uzbekistan  
E-mail: bonu444@mail.ru

**I.N. Tsyrelchuk**  
Advisor to Rector, TUIT. PhD.

**N. M. Mamatova**  
She graduated from the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi. Deputy Dean of the joint faculty of SFIT TUIT-BSUIR. Postgraduate student of the TUIT. Conducts research on algorithms and models of integrated, intelligent information exchange systems.

**M. Y. Abdul-Azalova**  
She graduated from the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi. Senior lecturer of the Department of "Information technologies" of TUIT. Postgraduate student of the TUIT. Conducts research on algorithms and adaptive models of an extensive multiparametric e-business management system.

**Abstract.** Big data is a collection of huge useful information that cannot be read using standard computing structures. Big data is not just data, it has already become a whole area that includes a set of tools, contexts, and structures. It uses complex data sets to select direction, course, and direct management within organizations. By improving and performing calculations, important structures can be obtained that are necessary for an accurate and deep understanding of the results obtained through the study of the organization's data. In this study, we examined different types of data and their use for e-Commerce, and specific methods for securing and developing data for use in complex organizations. In addition, we looked at big data issues related to the Internet and how to use it in web business.

**Keywords:** web analytics, predictive analytics, personalization, dynamic analysis, sentiment analysis, Big Data, e-commerce, e-business.

**Introduction.** Big Data is a continually progressing term. It is a great deal of sort out amorphous data that can be excavated for information. These educational accumulations are immense and complex that standard data getting ready isn't fit to process them. Enormous Data is being used in various sectors. We will see the effect of Big Data Analytics in changing the E-Commerce business, with the objective that the company surveyed as these E-exchange can benefit the most customers in the relationship from using Big Data because there will be information of the data accumulated on regular bases.

Various gigantic retailers regard this present data's information and cause them for predicting the customer interests and give their customers relative and charming looks when they shop on their

site, with the objective that they pull in the customer by providing the required and relevant journeys of things or things. These tendencies are inside and out-delivered from the Big Data examination. Huge Data contains two sorts of data one are composed, and the other one is unstructured.

*Materials and methods.* Starting late, China’s cross-edge web business has been creating rapidly. In the year 2017, the gross volume of China conveys online business accomplished 6.3 trillion Chinese Yuan with a yearly advancement rate of 14.5%. In China exchange the web-based business, B2B speaks to 80.9% while B2C and C2C speak to simply 19.1%. B2B is so far a standard exhibit anyway B2C is required to increase speedier. The principal products of China conveys online business 3C electronic products(20.8%), clothes(9.5%), house and home items(6.5%), outdoors products(5.4%). In the year 2017, the essential objective countries of China’s cross border web business are the USA(15%), France(11.4%),England(8.7%)and Brazil(6.5%) which exhibits that the USA and some made countries in Europe are up ’til now the objective rule countries, while the as of late creating business division in America, Middle Europe is growing fast [1].

Big Data, similarly as dispersed registering, have been associated in electronic business for a period, which has helped web-based business stages to recommend things even more correctly and rapidly, improve customer web shopping information, streamline collaboration structure and distortion security measure, and so forth. Starting late, a square chain begins to be associated in the web-based business, brings lower trade costs and progressively active portion. Likewise, non-modifying features diminish business distortion and assurance buyer astounding organization. Later on, with the more significant and progressively broad application, enormous information will pass on new a motivating force to cross-edge internet business.

Straightforwardly, there is no headed together a definition for the articulation “Big Data”, regardless, the most, for the most part, recognized the significance of Big Data is similar to 3 characteristics, volume, speed, and combination moreover implied as 3 V’s – Variety insinuates the heterogeneous nature, Velocity outlines the rate at which data is gotten, and Volume suggests the proportion of data. Due to these qualities, it is hard to direct and examination gigantic data using regular databases effectively. Nevertheless, using modern gadgets and progressions, Big data feasibly regulated. Also, when different data mining estimation, (for instance, machine learning and gathering count) are familiar with the extensive data insightful framework, one can get learning from the data (Figure 1.).

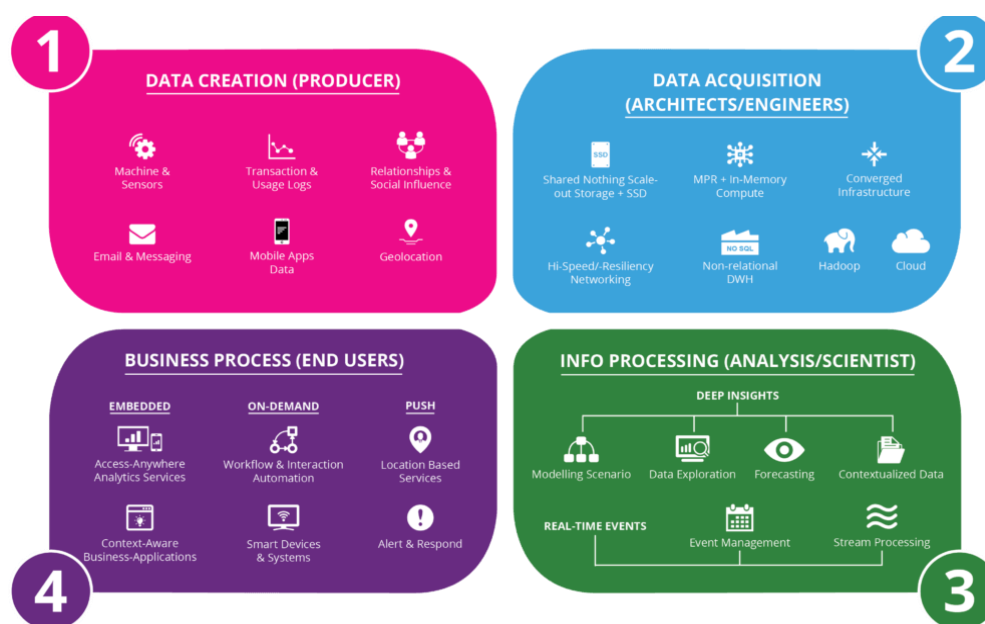


Figure 1. – Big data ecosystem

With the real objective of this examination, we will limit the investigation of the significant data examination to three classes as seeks after:

1. Web-Based Analytics: Refers to a review of a large volume of data made from internet organizing applications/areas.
2. Farsighted Analytics: Refers to the use of evident data to figure on buyer direct and designs.
3. Flexible Analytics: This implies the examination of an enormous volume of data made from mobiles, tablets and convenient contraptions.

A possible instance of such E-exchange business is Amazon.com – by utilizing exceptional programming to separate treats and click stream on customer programs, Company can perceive plans in buyers’ shopping penchant and therefore can give revamp/democratized offers, advancements, and points of confinement to such client [2].

Internet business implies the online trades: moving stock and adventures on the web, either in one trade (e.g., Amazon, Zappos, eBay, Expedia) or through a constant trade (e.g., Netflix, Match.com, LinkedIn, etc.). Web-based business firms going from Amazon to Netflix get distinctive sorts of data (e.g., orders, containers, visits, customers, suggesting joins, catchphrases, inventories examining), which can be broadly portrayed into four orders:

1. Exchange action information analytics.
2. Click-Stream information.
3. Video Information.
4. Voice Information.

In E-Commerce, information is the best approach to pursue buyer purchasing behavior to tweak provide, which are accumulated after some time using the customer examining and esteem based core interests. This fragment discusses different sorts of Big Data close by their proposals for web business (Figure 2.).

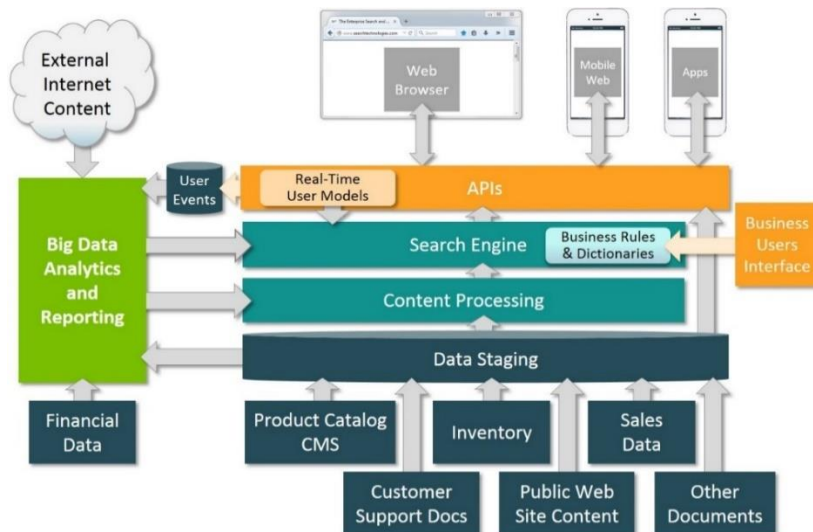


Figure 2. – Big Data architecture for e-commerce search platform

The Internet-based life Analytics (SMA) incorporates the social affair of data from electronic life goals/applications, (for instance, Wikipedia, Twitter, Facebook, GooglePlus, online diaries, etc.) and surveying such data to get encounters/learning. Web-based life data can be named big data as it bears the 3V properties. (For instance, every day there is around 35 million notification and more than 100,000 tweets for every minute on Twitter). Online life goals are frameworks of catenated people, yet virtual system, where people team up, exchange information and offer suppositions. These pursuits is prepared for affecting the buyer’s acknowledgment of a particular brand [3].

Fundamentally there are two basic methods for investigating the internet based life information; they include Text Mining and Sentimental Analysis.

Text Mining is exceedingly subject to the usage of substance based substance from sites and electronic life regions to make the judgment on the significance of an issue. The E-commerce sponsor makes once-over of watchwords identifying with the thing being checked. These watchwords can be used to perceive suspicions around an idea.

Predictive analysis alludes to the distinguishing proof of occasions before they happen using big data. The use of predictive analysis relies on strong information mining. In this unique situation, CEO of Caesar's Entertainment, expressed that: "[t]he most ideal approach to participate in data-driven publicizing is to amass progressively increasingly express information about customer tendencies, run preliminaries and examinations on new data, and choose techniques for connecting with [casino game] players' interests. We understood that the data in our database, combined with choice science apparatus that empowered us to foresee Singular client's possible incentive to us, would enable us to make advertising mediation that gainfully tended to players' interesting inclinations." Therefore, cautious examination encourages firms to set up their income spending plans. The readiness of these financial plans helps e-commerce firms to perceive future deals design from past deals information (e.g., annually or quarterly). It, thus, causes firms to all the more likely conjecture and decide stock prerequisites, along these lines prompting the shirking of item stock out and lost clients.

Right when customers put in demand on an online stage, it is sound for them to foresee that associations would give the organization of following the demand while the stock are in movement. Kopp (2013) cleared up that customers envision key information, for instance, the cautious openness, current status, and zone of the solicitations. E-Commerce organizations normally face an inconvenience in watching out for these wants from customers as different pariahs, for instance, warehousing and transportation are related with the store arrange a process. Big Data Analysis (BDA) expect a key occupation in this setting by the social affair diverse information from various get togethers on various things, and thusly precisely prompts the ordinary transport date to customers.

Another key zone in which web business firms can use Big Data is a customer organization. Customer grumblings bestowed by strategies for contact shapes in online stores together with tweeting enable internet business firms to make customers feel regarded when they call the organization center realizing a short organization movement. So additionally, Miller cleared up that, by offering proactive upkeep (i.e., taking preventive measures beforehand a failure happens or is even distinguished) using tremendous information obtained from sensors built up in things, online business firms can offer innovations after arranging organization.

A Clustering Algorithm system toils by recognizing social occasions of customers that have near tendencies. These customers are then packed into a single assembling and are given a unique identifier. New customer cluster are anticipated by finding out the typical similarities of the individual people in that gathering. Customer is mostly an individual from more than one gathering depending of the largeness of the customer typical appraisal in this case.

The vital utilization of web-based information for e-commerce firms is the course of action of a tweaked organization or changed things. Studies have fought that customers consistently like to buy with a comparable retailer using distinctive channels and that large data from these different coordinates can be modified ceaselessly. Persistent data examination enables firms to offer tweaked organizations including one of a kind substance and headway to customers. Similarly, these altered organizations help firms to detach reliable customer from a new customer and to make limited time offers as necessities are. As shown by Liebowitz, personalization can manufacture bargains by 10% or high and offer five numerous occasions the ROI on advancing utilization. Bloom spot, in such way, explored customer charge card data to pursue the spending records of the most immovable customers and to offer them rewards [4].

Take eBay, the B2C goliath, for instance. eBay is the biggest internet exchanging site on the planet. Purchasers conveyed more than 190 nations around the globe, more than 25 million dynamic vendors, 157 million dynamic purchasers, and 800 million dynamic items. In such a substantial number of clients and exchanges, information turns into the best need of eBay. eBay's Big Data the stage comprises of three layers: (1) Information mix layer: which is in charge of information ETL including information, obtaining, handling and cleaning, involving the group and continuous preparing abilities, related business items, and open origin items; (2) Information stage layer: which is primarily made out of the conventional undertaking information stockroom (EDW) with all-out limit surpassing 10PB, the Singularity putting away semi-organized and profound organized information with all-out limit 36PB and Hadoop bunches with an all-out limit exceeding 100PB; (3) Information get to layer: which can get to and break down information for business clients and examiners through different apparatus and stages, for example, MapReduce, Spark, Hive, HBase, which can give wealthy data getting ready and progression capacities.

The productivity of information use decay with time, the higher usage rate, the more up to date the information, the lower the entrance recurrence, the more established the info. In eBay's Hadoop, HDFS underpins various leveled stockpiling of different freshness information. HOT information is put away on a quick plate; WARM information is placed elsewhere in a quick circle and chronicled stockpiling. COLD and Frozen information is set away in a documenting, holding available to the above application. Putting away information with various freshness by stratification guarantees the pace of information preparing, and that the data which is at present in low esteem, however, may create new an incentive, later on, won't be erased. Of course, as the data scale creates with the extension of the customers' social occasion, to guarantee that the customer can get to and explore the tremendous scale enlightening accumulation set away on Hadoop with the most insignificant deferral and that the data obtainment, taking care of and examination in the Hadoop gathering can aggregate meanwhile, the eBay China Research and Development Department center started the OLAP around Hadoop adventure. The endeavor made metadata by the modeler by portraying the related estimations, and fabricated the metadata-based engine to normally create related Hive questions, MapReduce errands, and HBase exercises, so the data is examined out and pre-decided from HIVE, and the results are secured in HBase to give a natural request capacity of PB or even TB level, enlightening lists for front-end business customers and agents with only second measurement or even sub-second measurements delay.

A few shoppers are ending up progressively mindful of value segregation in Amazon.com. For example, CNN announced that a few clients of the Amazon are bothered over value separation on the cost of a specific DVD. One the purchasers revealed that the cost of a DVD in the wake of erasing treats on his PC, varied by \$2.50 edge. Another occasion, CNN detailed that the Amazon made utilization of dynamic estimating calculation while moving an item whoop "Jewel Rio MP3 Player" for \$51 not as much as its unique value.

Starting late, China cross-periphery online business has been creating fast. In 2017, the gross volume of China conveys electronic business accomplished 6.3 trillion Chinese Yuan with a yearly advancement rate of 14.5%. In China exchange web-based business, B2B speaks to 80.9% while B2C and C2C speak to simply 19.1%. B2B is up 'til now a direction show, yet B2C is depended upon to increase faster. The central products of China exchanges internet business are 3C electronic products(20.8%), house and home items(6.5%). outside products(5.4%). clothes(9.5%). In the year 2017, the standard objective countries of China cross online border business are the USA(15%),Russia(12.5%), France(11.4%),England(8.7%)and Brazil(6.5%) which demonstrates that the USA and some made countries in Europe are so far the essential objective countries while the as of late creating business part in countries like Latin America, Middle, and East Europe are growing quick [5].

A Positive element of implementing the Big Data examination request incorporates offering data look, a suggestion framework, a dynamic evaluating and client administration to collaborate with

the network part. By gathering characteristic information in the Big Data period, for example, geographic circulation, enthusiastic propensities, client conduct on shopping just as the social association, side interests, organizations can accomplish request introduction, biased introduction, a relationship introduction, and different approaches to fulfill clients.

Informative search shows that data standard and looking management quality. Data quality is a proportion of significant worth seen by yield given by a site. Data properties, for example, refresh, valuable, nitty-gritty, exact, and finish has been recognized as essential segments of data quality.

Suggestion System incorporates an association amongst e-dealers and buyers whereby the buyers give their information, for instance, relaxation exercises and tendencies, while the seller offers a proposition fitting their prerequisites, like this benefiting both. Nuances are given on key measures behind proposal systems: a customer based significant strain which used likeness in customer rankings to envision their interests and thing based network strain as centers in the space of words.

Giving a unique customer organization is the primary key to keep customers happy. Big Data engages you in improving your organizations. Using significant data analytics, you can overhaul your customer organization achieving progressively upbeat customers. A couple of customers may not simply protest of things or organizations through the official channels offered by the website, but may moreover go social about their get-together. You need data of such customers and exercise other alarm with the objective that grumblings of such customers are watched out for twofold quick. Enormous Data is secondhand to improve business shapes. Retailers can update their stock reliant on desires from web look designs, customer direct and atmosphere measures. One different application for the business procedure is the examination underway system or movement course. In light of scenery position and radio repeat recognition, the stimulus is used to pursue items or moving vehicles. This system enables customers to continue their solicitations. From that, customer organizations can be improved and increase shopper devotion [6].

Shopping habit is continuous and under perceived social dependence. Conduct compulsion is people's inability to see the quality of post-fixation longings and a failure to control want. For shopping addicts, shopping ends up uncontrolled, and they did not just purchase things they need, or they like, yet also genuinely spend their cash and are on edge to pass up on a decent chance to buy something. These items may is not utilized after buy. Utilizing the uses of Big Data investigation, the site can prescribe clients different things as a substitute or complementary pieces. This application is precious for clients with questions they need to purchase yet this is likewise unsafe for clients. They should invest more energy to audit more things to settle on a choice. It additionally suggests another correlative piece which the client feels they have to buy to build the obtained details. For instance, a client has purchased an exceptionally great pink dress, and the site prescribes her applicable sacks or shoes that are appropriate with the dress. They are wanted to consolidate together to give consumer loyalty. The client needs to invest energy and cash to purchase these corresponding items due to a decent chance to get them, even with less money. Shopping addictions are found to appear under two necessary measurements: propensity to spend and post-buy feeling.

The security of Big Data is another tremendous concern and one that increments with regards to Big Data. Because of the unmistakable attributes of Big Data in thee-trade environment, it can identify with protection and security concern. The high volume and convergence of information makes an all the more engaging focus for programmers. Moreover, higher information volume builds the likelihood that the information files and reports may contain characteristically important and touchy data. Information with the end goal of Big Data examination are along these lines a potential goldmine for digital culprits. As of late, ponders demonstrated that there is an expanding shopper worry over protection with regards to constant social publicizing and attaching advances, for example, treats. The Internet publicizing firms Double Click and Avenue A, product firm Intuit and others have confronted claims for utilizing treats to target promoting. A high assorted variety of Big data lead to associations coming up short on the capacity to oversee and understand these information, and

outsiders have chances to get to information. They may not conform to information insurance directions [7].

*Results.* In recent years, big data has become more popular as businesses of all sizes have discovered how to tap into it when it comes to making crucial decisions. Big data has assisted in finding and fixing problems, tracking business progress, and justifying causes for action. Although there is still much more that we can do when it comes to fully understanding big data, learning more means that we are able to utilize it to an even larger extent. And, many professionals believe that big data is the future of ecommerce, opening several doors for online businesses including analyzing sales, efficient transaction tracking, forecasting demand and supply, planning out expansion, and much more. So, how can big data help your ecommerce site?

#### 1. Provide a Personalized User Experience

The ecommerce experience was once severely lacking in any personal touch, which prevented shoppers from building a connection with the business. Buyers had to go through all the products with no option for highlighting the ones that they were interested in. Thanks to big data analytics, today's ecommerce owners are able to recommend items that their users are actually looking for. You can use big data to track the products that shoppers are interested in, allowing you to send personalized recommendations to the user that really mean something for them.

#### 2. Price More Competitively

Your ecommerce store's pricing strategy is one of the key determinants as to whether a shopper chooses your store or one of your competitors and it should be a vital part of your marketing strategy plan. It's easy for your users to check other ecommerce sites in order to find the best deal, which is why offering competitive prices is so important. To customers, it makes no sense to pay more for an item that they can get for less on another site. However, manually checking your competitor's sites on a regular basis would be a tedious task. Thanks to big data analytics, you have the ability to check how your competitors have priced their products, allowing you to respond accordingly.

#### 3. Turn Visitors into Buyers

When it comes to putting time, effort and resources into understanding big data, your ultimate goal should be attaining a higher conversion rate or turning visitors into actual buyers. Big data allows you to get a clearer picture of why visitors are leaving a site without making a purchase, whether it's down to on-site issues such as load speed or shopper concerns such as lack of site security, high prices or inadequate logistics solutions. This allows you to make more informed improvement decisions, such as switching e-commerce platforms or providing better shipping options. If you want to convert customers into buyers it helps to treat your website like a shopfloor, build trust with your visitors and ensure your website is responsive. You can then evaluate the data before and after implementing these changes to measure what works best for your brand.

#### 4. Better Manage Your Inventory

In addition to the above, you can also use big data to better manage your inventory. It allows you to automatically detect products that have been through the sales process successfully, in addition to tracking products that are returned. Big data analysis can also be used to track demand, allowing you to ensure that you get the right amount of inventory. Using this, you can determine the right number of products that are needed for certain periods, allowing you to stop wasting money on having too much or losing out on sales due to being out of stock.

#### 5. Predict Trends

Every online retailer wants to know about the next bestselling product before their competitors. Utilizing big data is a strategy that you must not miss out on when it comes to predicting trends in the market. Trend forecasting algorithms allow retailers to eliminate the guesswork when it comes to the next big products, combing data from social media posts and user web browsing habits to determine which products are causing a buzz. Sentiment analysis is another big data strategy that retailers can use to determine the context in which products are being talked about online. Are

conversations positive or negative? The data collected can be used to accurately predict the next popular products in various categories.

#### 6. Optimize Customer Service

Last but not least, the customer service is a vital component of your ecommerce store and without it, you're going to lose out to your competition very quickly. In addition to competitive prices, popular products, and a personalized customer experience, your customers want to feel that they are truly valued and understood by your brand. Exceptional customer service is crucial for both customer satisfaction and retention in the retail industry, both online and offline. Big data allows you to better determine the service strategies that work well for your customers, in addition to those that don't have a great effect. As a result, you will be able to optimize your customer service strategies and provide an enhanced customer experience, allowing you to gain an edge over your competition.

There are always beneficial ways to use big data to improve your ecommerce business, no matter what point you are in your company's growth. It's an especially exciting time to be a business owner. Do you use big data to improve your ecommerce store? We'd love to hear from you in the comments.

*Conclusion.* As the Big Data is used in various parts, it extensively smashes E-Commerce benefits and expects a vital activity in business choosing. The use of colossal data has significantly created in E-Commerce. Different gigantic retailer regards this present data's information and roots them for envisioning customer interests and give their customers similar and captivated looks for when they shop on their site. The objective that they attract the customer by providing the necessary and critical endeavors of things or things. Using the related information from this paper, the examiners can come up with vital and testing systems to expanding the upsides of Big data apply toward online business for both the customers similarly as the retailers. In this engaged and brisk condition customers generally, keep running with the online notification or through web crawlers by decreasing the inefficiency of the continuous markets. Our paper helps to perceive the various use of extensive data into web business so we can know the criticalness of astronomical data, it improves understanding of Usage of significant data and its fragments. We similarly talk about by the survey made on the issues related to E-Commerce if huge data isn't compared to that, so pros can tackle the problems associated to Big data and extend their work on that. An authoritative trial of Big Data examination is to make business regard from their impact of significant data. We also have analyzed the genuine troubles related to reliable data, so we expand our investigation in finding a response to one of the challenges identified with that. Researchers can get information about the issues concerning large data and critical troubles identified with that. So they can get concise information about colossal data which supports them in expanding their examination tackle extensive data related to online business.

#### *References*

- [1.] H. Liu and X. Wang, Shanghai, 2018, "Study on the application of big data in accurate marketing of cross-border e-commerce in China," 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA), pp. 24-27.
- [2.] X. Zhao, "A Study on the Applications of Big Data in Cross-Border E-Commerce," 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE), Xi'an, 2018, pp. 280-284.
- [3.] E. Makki and L. Chang, "Leveraging social big data for performance evaluation of E-commerce websites," 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, 2016, pp. 2525-2534.
- [4.] Shahriar Akter & Samuel Fosso Wamba, 2016, "Big data analytics in E-commerce: a systematic review and agenda for future research", IIM University of St. Gallen.
- [5.] S. Suguna, M. Vithya and J. I. C. Eunaicy, "Big data analysis in e-commerce system using HadoopMapReduce," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, 2016, pp. 1-6
- [6.] S. Pradeep and J. S. Kallimani, "A survey on various challenges and aspects in handling big data," 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICECCOT), Mysuru, 2017, pp. 1-5.



## **ОПТИМИЗАЦИЯ БИЗНЕС ПРОЦЕССОВ ПОСРЕДСТВОМ BIG DATA**

**И.Н. Цырельчук**  
Советник ректора,  
ТУИТ

**Н.М. Маматова**  
Заместитель декана  
совместного факультета  
СФИТ ТУИТ-БГУИР,  
соискатель ТУИТ

**М. Я. Абдул-Азалова**  
Старший преподаватель  
кафедры «Информационные  
технологии» ТУИТ, соискатель  
ТУИТ

*Совместный факультет СФИТ ТУИТ-БГУИР, Республика Узбекистан  
Ташкентский университет информационных технологий имени Мухаммада ал-Хорезми,  
Республика Узбекистан  
E-mail: bonu444@mail.ru*

**Аннотация.** Большие данные – набор огромной полезной информации, которая не может быть прочитана с помощью стандартных вычислительных структур. Большие данные – это не только данные, они уже стали целой областью, которая включает набор инструментов, контекстов и структур. Она использует сложные наборы данных для выбора направления, курса и проведения непосредственного управления внутри организаций. С помощью усовершенствования и выполнения вычислений могут быть получены важные структуры, необходимые для точного и глубокого понимания полученных результатов через исследование данных организации. В этом исследовании мы изучили разные типы данных и их использование для электронной торговли, и особые методы для обеспечения безопасности и развития данных для использования в сложных организациях. Кроме того, мы рассмотрели вопросы больших данных, касающихся интернета, и пути их использования в веб-бизнесе.

**Ключевые слова:** веб-аналитика, прогнозируемая аналитика, персонализация, динамический анализ, анализ тональности, большие данные, электронная торговля, электронный бизнес.

UDK 519.234

## APPLICATIONS OF SECOND ORDER ORNSTEIN UNLENBECK STOCHASTIC PROCESSES TO CREDIT RISK MODELING



**M. Vaskouski**

*Associate professor of  
Belarusian State University,  
Ph.D. in mathematics; senior  
statistical analyst at HiQo  
Solutions Ltd*

*Department of Higher Mathematics, Belarusian State University, Republic of Belarus  
HiQo Solutions, Ltd, USA  
E-mail: vaskovskii@bsu.by, maxim.vaskouski@hiqo-solutions.com*

### **M. Vaskouski**

*Associate professor at Department of Higher Mathematics of Belarusian State University, Ph.D. in mathematics, senior statistical analyst at HiQo Solutions Ltd. I am an expert in theory of stochastic differential equations.*

**Abstract.** We consider applications of second order stochastic processes for analysis and forecasting credit loss. In contrast to the Vasicek model based on the one-dimensional Ornstein-Uhlenbeck stochastic differential equation driven by the Wiener process, we study two-dimensional analogues of Ornstein-Uhlenbeck processes driven by fractional Brownian motions. These processes are applied to extrapolation of macroeconomic factors for modeling account loss probability. Second order Ornstein-Uhlenbeck stochastic processes capture local behavior of economic factors providing more realistic tools in comparison with the first order Ornstein-Uhlenbeck processes. The obtained results are applied to different types of account loss rate models in frame of FASB's Current Expected Credit Loss (CECL) and IASB's International Financial Reporting Standards 9 (IFRS 9) rules.

**Keywords:** Ornstein-Uhlenbeck processes, mean reverting, macroeconomic factors, rough path integration theory.

**Introduction.** Accounts loss rate models are core ingredients for credit loss estimation in frame of FASB's Current Expected Credit Loss (CECL) and IASB's International Financial Reporting Standards 9 (IFRS 9) rules [1]. Simple account loss models (time series, roll rate models) are based on fitting of loss rate time series by proper combinations of macroeconomic factors (e.g. unemployment rate, house price index, gross domestic product). More complicated models (age-period-cohort models, discrete time multihorizon survival models) extract exogenous components of historical accounts loss rate. These exogenous components, called environment functions, are modeled by macroeconomic factors.

Economic scenarios are applied to estimate future losses. According to the CECL rules, it is possible to use annual macroeconomic scenarios provided by the US Federal Reserve System for a short-term (24 months) economic prediction. Long-term economic prediction is based on using conception of mean reverting [1]. In contrast to the Vasicek model based on the one-dimensional Ornstein-Uhlenbeck stochastic differential equation driven by the Wiener process, we study two-dimensional analogues of Ornstein-Uhlenbeck processes driven by fractional Brownian motions.

Second order Ornstein-Uhlenbeck processes capture local behavior of economic factors and improve credit risk models in comparison with the first order Ornstein-Uhlenbeck processes.

This paper is organized in the following way. In section 2 we introduce necessary tools from rough path integration theory [2, 3]. In section 3 we deduce exact formulas for solutions of second order Ornstein-Uhlenbeck equations driven by fractional Brownian motions. In section 4 we derive new extrapolation method of macroeconomic factors by second order Ornstein-Uhlenbeck processes. Finally, Section 5 is devoted to application of the obtained results to CECL estimation in frame of the Age-Period-Cohort model [1, 4].

### Elements of rough path integration theory

A stochastic process  $B^H(t)$ ,  $t \geq 0$ , defined on a probability space  $(\Omega, F, P)$  is called a fractional Brownian motion with Hurst index  $H \in (0, 1)$  if  $B^H(t)$  is a Gaussian process with properties: 1)  $B^H(t) = 0$  a.s., 2)  $E(B^H(t)) = 0$  for any  $t \geq 0$ , 3)  $E(B^H(t)B^H(s)) = \frac{1}{2}(t^{2H} + s^{2H} + |t-s|^{2H})$  for any  $t, s \geq 0$  [5].

We consider a stochastic differential equation

$$dX(t) = b(X(t))dt + h(X(t))dW(t) + \sigma(X(t))dB^H(t), \quad t \in [0, T], \quad (1)$$

driven by independent Wiener process  $W(t)$  and fractional Brownian motion with Hurst index  $H \in \left(\frac{1}{3}, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right)$ , where functions  $b, h, \sigma: \check{Y} \rightarrow \check{Y}$  are deterministic.

Let us define the process  $B(t) = (t, W(t), B^H(t))^T$  and function  $B^H(t)$   $\alpha \in \left(\frac{1}{3}, \min\left(\frac{1}{2}, H\right)\right)$  and consider  $\alpha$ -Holder continuous rough path  $\mathbf{B} = (B, \mathbb{B})$ , where the two parametric random process  $\mathbf{B}: [0, T]^2 \rightarrow \check{Y}^{3 \times 3}$  is defined as follows:

$$\begin{aligned} B^{(1,1)}(s, t) &= \frac{1}{2}(t-s)^2, & B^{(2,1)}(s, t) &= \int_s^t (W(\tau) - W(s))d\tau, & B^{(3,1)}(s, t) &= \int_s^t (B^H(\tau) - B^H(s))d\tau, \\ B^{(1,2)}(s, t) &= -B^{(2,1)}(s, t) + (t-s)(W(t) - W(s)), & B^{(1,3)}(s, t) &= -B^{(3,1)}(s, t) + (t-s)(B^H(t) - B^H(s)), \\ B^{(2,2)}(s, t) &= \frac{1}{2}(W(t) - W(s))^2 - \frac{1}{2}(t-s), & B^{(3,3)}(s, t) &= \frac{1}{2}(B^H(t) - B^H(s))^2, \\ B^{(2,3)}(s, t) &= \lim_{|\Pi| \rightarrow 0} \sum_{k \in \Pi} (W(t_k) - W(s))(B^H(t_{k+1}) - B^H(t_k)), \\ B^{(3,2)}(s, t) &= -B^{(2,3)}(s, t) + (W(t) - W(s))(B^H(t) - B^H(s)), \end{aligned}$$

where  $\Pi = \{s = t_0 < t_1 < \dots < t_l = t\}$  is an arbitrary partition of segment  $[s, t]$ ,  $|\Pi| = \max_k (t_{k+1} - t_k)$ .

**Definition 1.** A random process  $X(t)$ ,  $t \in [0, T]$ , is called a solution of Eq. (1) if the process  $X(t)$  is an  $F_t$ -adapted process (here  $F_t$  is the filtration generated by the processes  $W(t)$ ,  $B^H(t)$ ), admits  $\alpha$ -Holder continuous paths a.s., and the function  $t \rightarrow X(t, \omega)$  is a solution of the rough path differential equation  $dz(t) = f(z(t))\mathbf{B}(t, \omega)$  for almost all  $\omega \in \Omega$ . Solution  $X(t)$  of Eq. (1) with the

initial condition  $X(0) = x \in \check{Y}$  is called unique if  $P(Y(t) = Z(t) \forall t \in [0, T]) = 1$  for any two solutions  $Y(t), Z(t)$  of Eq. (1) with the initial condition  $X(0) = x \in \check{Y}$ .

### Second order Ornstein-Uhlenbeck processes

Let us consider the following system of stochastic differential equations

$$\begin{cases} dx(t) = (\theta(\mu - x(t)) + v(t))dt, \\ dv(t) = -\theta_1 v(t)dt + \sigma_1 dW(t) + \sigma_2 dB^H(t), t \in [0, T], \end{cases} \quad (2)$$

where  $W(t)$  and  $B^H(t)$  are independent Wiener process and fractional Brownian motion with Hurst index  $H \in \left(\frac{1}{3}, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right)$ ;  $\mu$  is a long term mean,  $\sigma_1$  and  $\sigma_2$  are non-negative volatility coefficients,  $\theta$  and  $\theta_1$  are positive constants,  $\theta \neq \theta_1$ .

We consider the second equation of System (2) as a partial case of Eq. (1) with  $B(t) = (t, W(t), B^H(t))^T$ ,  $f(X) = (-\theta_1 X, \sigma_1, \sigma_2)$ . Let us show that the process

$$v(t) = v_0 e^{-\theta_1 t} + \int_0^t e^{-\theta_1(t-s)} dM(s) \quad (3)$$

is a solution of the second equation of System (2) with the initial condition  $v(0) = v_0$ , where integral in the right-hand side is the Young integral with respect to  $M(t) = \sigma_1 W(t) + \sigma_2 B^H(t)$ .

Since

$$v(t) = v(s) - \int_s^t \theta_1 v(\tau) d\tau + \sigma_1 (W(t) - W(s)) + \sigma_2 (B^H(t) - B^H(s))$$

for any  $t, s \in [0, T]$ , the remainder terms can be estimated as follows:

$$R^{f(v)}(s, t) := f(v(t)) - f(v(s)) - Df(v(s))f(v(s))(B(t) - B(s)) = O(|t - s|^{2\alpha}),$$

where  $\alpha \in \left(\frac{1}{3}, \min\left(\frac{1}{2}, H\right)\right)$ .

Taking into account the relation

$$\begin{aligned} & \lim_{|\Pi| \rightarrow 0} \sum_{t_i, t_{i+1} \in \Pi} Df(v(t_i))f(v(t_i))B(t_i, t_{i+1}) = \\ & = \lim_{|\Pi| \rightarrow 0} \sum_{t_i, t_{i+1} \in \Pi} (\theta_1^2 v(t_i) \Delta_i t - \theta_1 \sigma_1 \Delta_i W - \theta_1 \sigma_2 \Delta_i B^H) \Delta_i t = 0, \end{aligned}$$

where  $\Delta_i g = g(t_{i+1}) - g(t_i)$ , we conclude that the rough path integral  $\int_0^t f(v(s)) d\mathbf{B}(s)$  coincides

with the Young integral  $\int_0^t f(v(s)) dB(s)$ . Therefore, relation (3) gives a solution of the second equation of System (2). Substituting this relation into System (2), we get

$$x(t) = \mu + C_1 e^{-\theta t} + C_2 e^{-\theta_1 t} + \frac{1}{\theta - \theta_1} \int_0^t (e^{-\theta_1(t-s)} - e^{-\theta(t-s)}) dM(s), \quad (4)$$

where  $C_1$  and  $C_2$  are arbitrary constants. This family of processes is called second order Ornstein-Uhlenbeck processes.

### Economic scenarios based on second order Ornstein-Uhlenbeck processes

Let  $y(t)$ ,  $t \in [t_{\min}, t_{\max}]$ , be a time series of some macroeconomic factor. We are going to create extrapolation of time series  $y(t)$  for  $t > t_{\max}$  basing on the family of processes given by (4). It's clear that the best estimate of the process  $x(t)$  is the conditional expectation  $\bar{x}(t) = E(x(t) | F_0)$  with respect to the trivial  $\sigma$ -algebra  $F_0$ . Thus,

$$\bar{x}(t) = \mu + C_1 e^{-\theta t} + C_2 e^{-\theta_1 t}.$$

For convenience we suppose that the processes  $y(t)$  and  $\bar{x}(t)$  have the same discrete time scaling (usually we use one month as the minimal step that corresponds to  $\Delta t = \frac{1}{12}$ ). Here  $\mu$  is the long term mean (usually  $\mu$  is selected to be equal to the average value across values of  $y(t)$ ), parameters  $C_1, C_2, \theta, \theta_1$  are free. Denote  $t_n = t_{\max}$ ,  $t_{n-1} = t_{\max} - \Delta t$ ,  $y_n = y(t_n)$ ,  $y_{n-1} = y(t_{n-1})$ ,  $\bar{x}_n = \bar{x}(t_n)$ ,  $\bar{x}_{n-1} = \bar{x}(t_{n-1})$ . Setting up  $y_n = \bar{x}_n$ ,  $y_n - y_{n-1} = \bar{x}_n - \bar{x}_{n-1}$ , we get values of  $C_1, C_2$  which provide smooth continuation of the process  $y(t)$  for  $t > t_{\max}$ :

$$C_2 = \frac{y_{n-1} - \mu - (y_n - \mu)e^{\Delta t}}{e^{-\theta_1 t_n} (e^{\theta_1} - e^{\theta})}, \quad C_1 = (y_n - \mu)e^{\theta t_n} - C_2 e^{(\theta - \theta_1)t_n}.$$

Without loss of generality we assume that  $0 < \theta < \theta_1$ . Then parameter  $\theta$  corresponds to rate of reverting to mean value  $\mu$  depending on the length of the economic cycle. Parameter  $\theta_1$  can be estimated from 24-months Consensus economic scenarios or FRB economic scenarios.

The confidence intervals corresponding to 95% probability can be estimated in the following way  $x_{lower}(t) = \bar{x}(t) - 1.96\Xi(t)$ ,  $x_{upper}(t) = \bar{x}(t) + 1.96\Xi(t)$ , where

$$\Xi^2(t) = \frac{1}{(\theta - \theta_1)^2} E \left( \int_0^t (e^{-\theta_1(t-s)} - e^{-\theta(t-s)}) dM(s) \right)^2.$$

### Application to credit risk modeling

Let us consider discrete time Age-Period-Cohort model for estimation of Current Expected Credit Loss (CECL) on US mortgage loans, see [1, 4].

Denote by  $aa$ ,  $tt$ ,  $vv$  age of credit (in months), calendar date and origination date, it's clear that  $t = a + v$ . We consider historical account loss rate (or Probability of Default) and attrition rate (or Probability of Attrition) defined as

$$PD(a, t) = \frac{N_{def}(a, t)}{N_{act}(a, t-1)}, \quad PA(a, t) = \frac{N_{attr}(a, t)}{N_{act}(a, t-1)},$$

where  $N_{act}$ ,  $N_{def}$ ,  $N_{attr}$  are the numbers of active, charge-off, attrition accounts respectively.

Also we define exposure at default

$$EAD(a, t) = \frac{B_{def}(a, t)}{N_{def}(a, t)},$$

where  $B_{def}$  is charge-off amount.

The first stage of analysis is to find Age-Period-Cohort decompositions for variables PD, PA, EAD:

$$\text{logit}(PD(a, t)) = F_{PD}(a) + H_{PD}(t) + G_{PD}(v), \quad (5)$$

$$\text{logit}(PA(a, t)) = F_{PA}(a) + H_{PA}(t) + G_{PA}(v), \quad (6)$$

$$\log(EAD(a, t)) = F_{EAD}(a) + H_{EAD}(t) + G_{EAD}(v), \quad (7)$$

here  $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$  is the inverse function to the logistic

function  $g(z) = \frac{1}{1+e^{-z}}$   $z \in \mathbb{R}$ .

The following graphs show the Age-Period-Cohort decomposition of account loss rate for US 30-years fixed rate mortgage loans with high risk (FICO Score is less than 660) [6]. This decomposition was obtained by method described in [7].

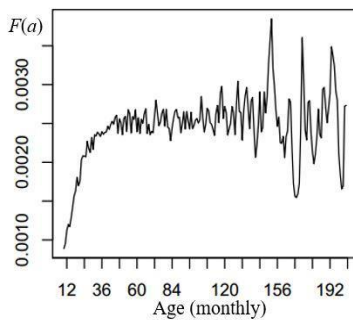


Figure 1a. Lifecycle, PD, probability scale

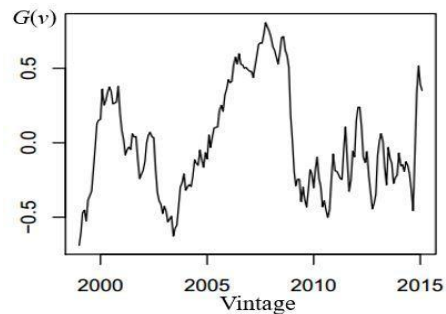


Figure 1b. Credit Quality Function, PD

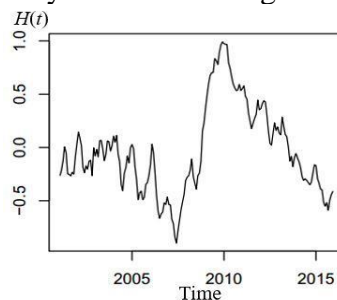


Figure 1c. Environment Function, PD

The environment function  $H_{PD}(t) H(t)$  captures macroeconomic impact on account loss rate. We create a linear regression model (which is called economic model) for  $H_{PD}(t)$  :

$$H_{PD} \sim \beta_0 + \sum_{i=1}^n \beta_i f_i,$$

where  $f_i(t) f_i$  ( $i = 1, \dots, n$ ) are macroeconomic factors,  $\beta_i$  ( $i = 1, \dots, n$ ) are regression coefficients. Usually a seasonality component is included as a part of the Age-Period-Cohort decomposition or economic model.

Table 1. – Parameters of the economic model for  $H_{PD}(t) H_{PD}(t)$  .

Economic variable	Transformation	Optimal value of lag	Optimal value of win	Regression coefficient, $\beta_i$
Real DPI	LogRatio	2	23	-1.734
Unemployment rate	Diff	5	24	0.078
House Price Index	LogRatio	12	17	-2.678
10-year Treasury yield	Diff	7	24	-0.106

Here the transformations LogRatio and Diff are defined as follows:

$$\text{LogRatio}(x, t, \text{lag}, \text{win}) = \log \frac{x(t - \text{lag})}{x(t - \text{lag} - \text{win})},$$

$$\text{Diff}(x, t, \text{lag}, \text{win}) = x(t - \text{lag}) - x(t - \text{lag} - \text{win}).$$

Anyone, who wants to predict future loss, needs to know extrapolation of the environment function for forthcoming dates. Basing on economics history, we are able to calculate values of the economic fit  $\hat{H}_{PD}$  up to present moment  $T_0$  . According to CECL rules, we get extrapolation of  $\hat{H}_{PD}$  for the first two years by using FRB economic scenarios. Starting from  $t = T_0 + 2$  we apply the economic model coefficients to the extrapolated macroeconomic factors, where mean reverting extrapolation based on second order Ornstein-Uhlenbeck processes is applied to the transformed economic factors (see Section 4).

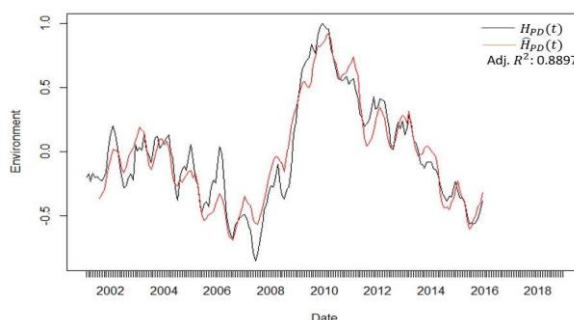


Figure 2.– Economic fit  $\hat{H}_{PD}$  for the Environment Function  $H_{PD}(t) H_{PD}$

The same procedure is applied for the environment functions of the variables PA, EAD.

The lifecycle functions  $F_{PD}(a)$  and  $F_{PA}(a)$  have to be truncated after 10-years age and extrapolated with a constant. The lifecycle  $F_{EAD}(a)$  should be truncated after 10-years age and extrapolated to be logarithmically decaying to zero.

Finally, CECL can be evaluated from relations (5) – (7), where the environment functions are replaced by extrapolated economic fits and original lifecycles are replaced by truncated and extrapolated lifecycles.

*Remark 1.* The same approach to get extrapolations of time functions can be applied to other credit risk models such as time series, roll rate, state transition, discrete time multihorizon survival models described in [1, 4].

*Acknowledgment.* The author thanks Dr J. Breeden and Dr A. Yablonski for their attention and valuable remarks.

### References

- [1.]Breeden, J.L. Living with CECL: Mortgage modeling alternatives / J.L. Breeden. – Middletown, 2018.
- [2.]Levakov, A. Stochastic differential equations and inclusions / A. Levakov, M. Vaskouski. – Minsk: BSU, 2019.
- [3.]Vaskouski, M. Asymptotic expansions of solutions of stochastic differential equations driven by multivariate fractional Brownian motions having Hurst indices greater than 1/3 / M. Vaskouski, I. Kachan // Stochastic Analysis and Applications. 2018. – Vol. 36, № 6. – P. 909–931.
- [4.]Breeden, J.L. Current expected credit loss procyclicality: it depends on the model / J.L. Breeden, M. Vaskouski // Journal of Credit Risk. 2020. – Vol. 16, № 1. – P. 1 – 22. DOI: 10.21314/JCR.2020.258.
- [5.]Mandelbrot, B.B. Fractional Brownian Motions, fractional noises and applications / B.B. Mandelbrot, J.W. van Ness // SIAM Review. 1968. – Vol. 10, № 4. – P. 422 – 437.
- [6.]Vaskouski, M. Analysis and forecasting of credit loss with help of discrete time survival models / M. Vaskouski, A. Zadorozhnyuk // Proceedings of the International conference “Big Data and Advanced Analytics”. 2019. – Vol. 1. – P. 285 – 293.
- [7.]Schmid, V. Bayesian age-period-cohort modeling and prediction – BAMP/ V.J. Schmid, L. Held // Journal of Statistical Software. 2007. – Vol. 21, №8 – P. 1–15.



UDK 519.245

## ESTIMATING THE INEFFICIENCY OF MACHINERY AND ITS OPERATORS USING HIERARCHICAL BAYESIAN MODELS



**N.V. Guletskii**  
*Systems analyst at HiQo Solutions,  
master's student at BSU*



**Ya. M. Radyna**  
*Statistical analyst at HiQo Solutions,  
associate professor at BSU*

*HiQo Solutions, Ltd., Republic of Belarus  
E-mail: nikolay.guletskiy@hiqo-solutions.com*

### **N. Guletskii**

*Graduated from the Belarusian State University. Master's student at BSU. Works as a systems analyst at HiQo Solutions, Ltd, Republic of Belarus, Minsk. Fields of research include functional analysis, abstract harmonic analysis, optimization, artificial neural networks, computer vision, and Bayesian modeling.*

### **Ya. Radyna**

*PhD in Mathematics. Associate professor at Belarusian State University. Works as a statistical analyst at HiQo Solutions, Ltd, Republic of Belarus, Minsk. Fields of research include Functional Analysis, Statistics and Machine Learning.*

**Abstract.** Inefficient operation and deterioration of machinery leads to increased costs and energy consumption. Here we describe a method for disentangling operator inefficiency from the inefficiency of their equipment by building a hierarchical Bayesian model to model the fuel consumption of each operation.

**Keywords:** Hierarchical Bayesian models, Bayesian Analysis, Markov Chain Monte Carlo, Statistical Modeling.

**Introduction.** The goal of this paper is to detect and attribute increased fuel consumption in an industrial process involving a set of machines, grouped by model, and a set of operators, which operate the machinery. Each operation involves a single piece of machinery and a single operator but may consist of multiple sub-operations which have different fuel consumption dynamics. The fuel consumption is measured on a per-operation basis, meaning that we do not have fuel consumption measurements for individual sub-operations. We assume that each set of machines used by a given operator and vice versa is sufficiently diverse to be able to draw conclusions about the overall (marginal) efficiency of that operator/machine.

As a baseline one might consider linear and log-linear models that predict the fuel consumption from operation duration, operator id, machine id and model id (encoded using dummy variables). The parameters corresponding to operators and machines can then be reinterpreted as absolute inefficiencies. Unfortunately, linear models only allow additive inefficiency, while log-linear models require a log-transform of fuel consumption. Both restrictions are sub-optimal since they do not accurately reflect the dynamics of fuel consumption: operation inefficiency should be multiplicative, but the fuel consumption should be almost linear in operation duration. Moreover, such approaches do not allow us to incorporate clustering by model and are unlikely to produce interpretable results for quantities such as the ideal fuel consumption for a given operation.

We propose using non-linear hierarchical Bayesian models to model the dynamics of fuel consumption. Unlike basic frequentist methods, like lasso regression [1], Bayesian hierarchical modeling allows us to express arbitrary physical processes involved in the industrial process and infer their latent parameters, which can then be used to perform comparisons with the machinery's reference documentation.

*The dataset.* Our proprietary dataset consists of approximately 700000 rows with 128 operators and 21 machines, grouped into 3 machine models. The dataset has many outliers, exhibits multi-modality, and likely suffers from dataset shift. Moreover, we can't distinguish between multiple sub-operation subtypes within the dataset: there are multiple tasks involving the same sub-operations which result in slightly different fuel consumption.

*Our model.* For each machine model and each sub-operation type, we introduce an ideal fuel consumption coefficient which determines the minimal fuel consumption per unit of time that can be achieved with a perfectly maintained machine and a perfect operator. For each machine and operator, we introduce an inefficiency coefficient, which, when added to 1, acts as a multiplier for a sub-operation's ideal fuel usage. The inefficiency is either taken into account or ignored depending on the nature of the sub-operation, but we expect at least one sub-operation type to include both machine and operator inefficiency. The expected fuel consumption for an operation is declared to be the sum of expected fuel consumption for sub-operations, which may be modeled as non-linear functions of both latent and observed parameters. We then assume that the actual fuel consumption is sampled from a log-normal distribution located at the expected fuel consumption. The scale of the log-normal distribution is controlled by a global scale parameter  $\sigma$ . The general formula for the fuel consumption  $\varphi_i$  during operation  $i$  for  $m$  subtasks given vectors of measurements  $x_{i,l}$  and physical models  $f_s(x_{i,s})$  for all subtasks  $s$ , as well as the operator's and vehicle's inefficiency coefficients  $\theta_{op}$  and  $\theta_{mc}$ , may be described as follows:

$$\tilde{\varphi}_i = \sum_{s=1}^m (1 + \theta_{op})^{q_{op,s}} (1 + \theta_{mc})^{q_{mc,s}} f_s(x_{i,s}),$$

$$q_{op,s}, q_{mc,s} \in \{0,1\},$$

$$\varphi_i \sim \text{Lognormal} \left( \log(\tilde{\varphi}_i) - \frac{\sigma^2}{2}, \sigma^2 \right).$$

*Kernel trick.* To reduce the heteroskedasticity caused by the presence of multiple operation sub-types which can't be distinguished directly, we introduce a fuel consumption adjustment coefficient  $b_m(x_{i,s,dur})$ , which depends on the duration of a specific sub-task and acts as an additional non-linearity inside the physical model. To apply the kernel trick, we associate fixed-length weight, center and scale vectors  $w_m$ ,  $y_m$ , and  $\gamma_m$  with each machine model  $m$ , and calculate the coefficient as a dot product of the weight vector with a vector of RBF kernels with the corresponding centers and scales. To ensure that different MCMC chains converge to the same posterior distribution, we restrict the vector of centers to be a monotonically increasing sequence of numbers, which is acceptable in our case because the kernel trick is applied to a 1-dimensional space (representing a sub-operation's duration). The monotonicity is achieved by reparametrizing the vector of centers to be a cumulative sum of positive offsets  $\beta_m$ , which define how far each following center is from the previous one:

$$b_m(x_{i,s,dur}) = \sum_{k=1}^n w_{m,k} \exp \left( -\gamma_{m,k} \left| x_{i,s,dur} - \sum_{j=1}^k \beta_{m,j} \right|^2 \right).$$

*Priors.* For ideal fuel consumption, we use a fixed truncated normal prior with fixed parameters determined from domain experience. For other parameters of the physical model, we also use normal and truncated normal distributions with weakly-informative hyperpriors. For inefficiency coefficients, we use half-normal priors with a weakly-informative (Half-Cauchy(5)) hyperprior for each group (i.e. two hyperpriors, one for operators and one for machines). For the observation noise scale parameter, we also use a weakly-informative (Half-Cauchy(5)) hyperprior. For the kernel trick’s weights and center offsets, we use Half-Normal(1) priors. For the kernel scales, we use a truncated normal distribution with a unit mean and scale.

*Parametrization.* Despite the relatively large size of the dataset, we use non-centered parametrizations for all truncated normal and half-normal distributions. To sample from a truncated normal distribution  $\mathcal{TN}(\mu, \sigma^2, l)$ , where  $l$  is the minimum value for numbers sampled from the distribution, we multiply the samples from a truncated normal distribution  $\mathcal{TN}(0, 1, l')$  by the desired scale  $\sigma$  and add the desired mean  $\mu$ . The value of  $l'$  corresponding to the desired minimum value  $l$  is given by

$$l' = \sigma^{-1} \left[ l - \left( \mu - \sigma \frac{-p_{\mathcal{N}(0,1)}((l-\mu)/\sigma)}{1 - \text{cdf}_{\mathcal{N}(0,1)}((l-\mu)/\sigma)} \right) \right].$$

*Inference.* Our implementation is based on numpyro[2]. We fit the model using the No-U-Turn Sampler (NUTS) [3], an adaptive sampler based on Hamiltonian Monte Carlo. We set the target acceptance probability to 0.99 to make sampling more robust to the high curvature introduced by the non-linear physical model. We achieve convergence despite the complex geometry of the typical set (Gelman-Rubin statistic =  $1 \pm 1e-2$  for all trained models).

*Preprocessing.* To speed up convergence, we normalize the operation durations, fuel consumption and other positive observations to have a mean of 1.

*Evaluation.* To validate whether our model is capable of predicting the fuel usage for individual operations, we perform a 1%-to-99% stratified train-test split and compare the mean absolute error of our model to four 5-fold cross-validated lasso regression models with 3rd degree polynomial features: two linear models, two log-linear models, two models with categorical variables and two models without categorical variables. We also compare our hierarchical model to a model obtained by removing the inefficiency coefficients from the main model. The small size of the training set is due to the computational complexity of training the Bayesian model.

To validate whether the model is capable of determining the inefficiencies of various parties, we build an auxiliary dataset by randomly selecting 10 operators and 5 machines and artificially adding inefficiencies to their operations by multiplying fuel usage by a per-operator/per-machine random number which we call jitter. We then train two models: one on a 5% subsample of the original dataset, and one on the auxiliary dataset constructed from the subsample. The relative increase in fuel consumption between datasets can then be derived from the physical model’s definition and compared to the random jitter applied during the creation of the auxiliary dataset.

*Results on the prediction task.* Our model has the largest coefficient of determination ( $R^2$ ) and the lowest mean average error (MAE) on the hold-out set. Table 1 holds the results of an ablation study performed on a 1%-to-99% split of the dataset. For all Bayesian models, we ran NUTS for 1000 iterations, the first 500 of which were used for adaptation.

Table 1: – Prediction error statistics on the hold-out set.

Model	MAE	$R^2$
Linear without categorical	0.128	0.641
Linear with categorical	0.128	0.641
Log-linear without categorical	0.129	0.624
Log-linear with categorical	0.129	0.625
Bayesian without inefficiencies	0.119	0.643
Bayesian	0.119	0.655
<b>Bayesian with a kernel trick</b>	<b>0.116</b>	<b>0.661</b>

This demonstrates that linear models do not benefit from operator/machine id information, while introducing inefficiency coefficients into the Bayesian model noticeably reduces the unexplained variance.

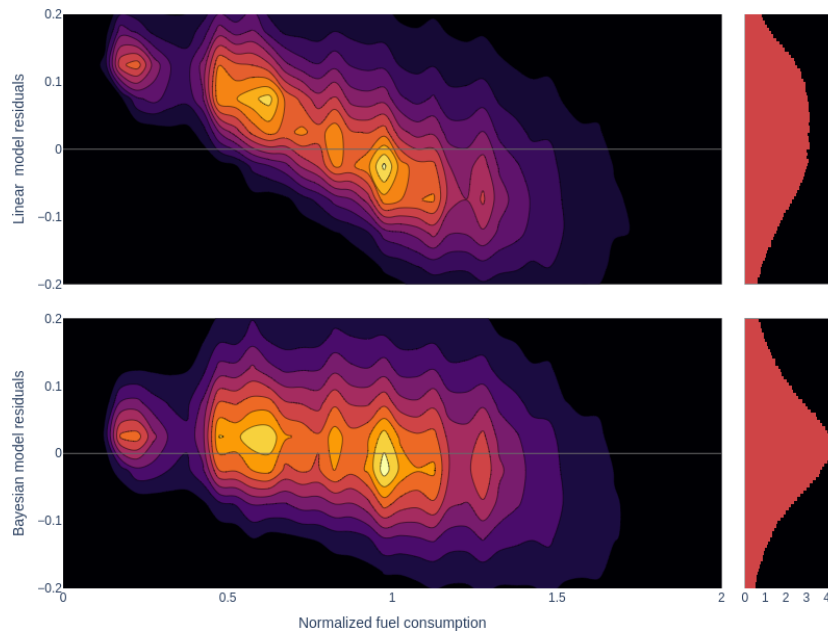


Figure 1: – Residual plots for the best linear model and the best Bayesian model

Linear models and basic Bayesian models exhibit major heteroskedasticity, while the best Bayesian model is relatively homoscedastic.

*Results on the synthetic inefficiency detection task.* To fit the Bayesian model to both the original and the auxiliary datasets, we ran NUTS for 2000 iterations, the first 1500 of which were used for adaptation. To recover the jitter multiplier  $\gamma$  from the original predicted inefficiency  $\theta_{orig}$  and the predicted inefficiency on the jittered dataset  $\theta_{jitt}$ , we use the following relationships:

$$(1 + \theta_{orig})(1 + \gamma) \approx (1 + \theta_{jitt}),$$

$$\gamma \approx \frac{\theta_{jitt} - \theta_{orig}}{1 + \theta_{orig}}.$$

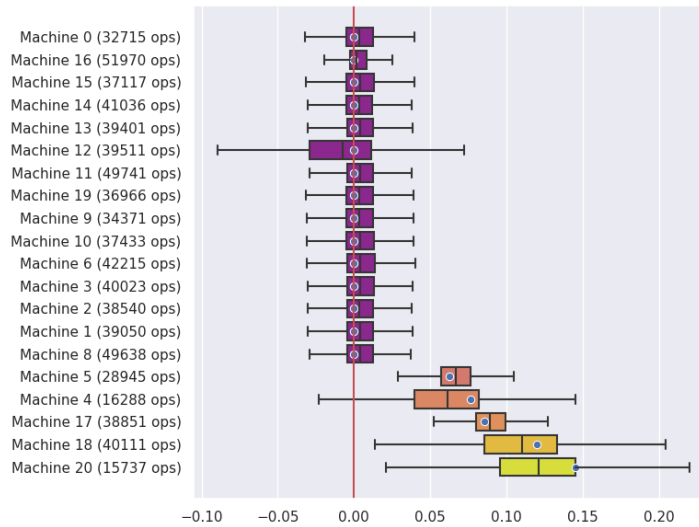


Figure 2: – A comparison of the predicted relative machine inefficiency jitter with the actual jitter used to construct the auxiliary dataset. The box plot summarizes the posterior distribution of the jitter multiplier, while the dots represent the actual fuel usage jitter multipliers. The closer the median line is to the blue dot, the better.

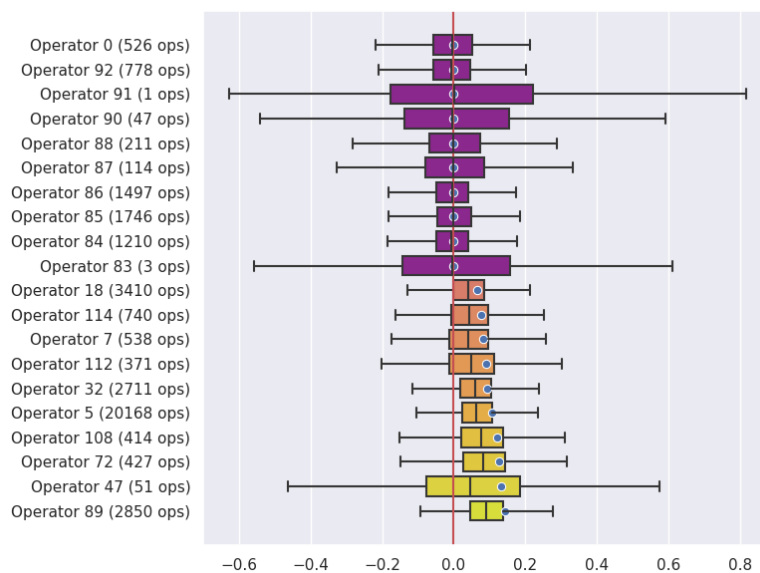


Figure 3: – A comparison of the predicted relative operator inefficiency jitter with the actual jitter used to construct the auxiliary dataset. The box plot summarizes the posterior distribution of the jitter multiplier, while the dots represent the actual fuel usage jitter multipliers. The closer the median line is to the blue dot, the better.

Our model successfully identifies the inefficiencies caused by the machines. For operators, the estimates are significantly more conservative, but this may be explained by the model’s low confidence in its predictions for operators who are underrepresented in the sample. Nevertheless, the

estimates for operators which were not affected during the creation of the modified dataset are near zero, while the estimates for operators which were affected are almost always significantly larger, and for most operators, the actual jitter is within the inter-quartile range. This indicates that the model gives conservative estimates, which is ethically desirable.

*Conclusion.* The proposed Bayesian model is superior to the proposed baselines in multiple ways. Firstly, it provides credible intervals for parameters describing the inefficiencies of operators and machines, allowing us to judge whether it is reasonable and fair to make judgements from the inferred parameters. Secondly, it allows us to integrate non-linear physical relationships into our model, giving us interpretable parameters, which can be inspected by a domain expert to determine whether the obtained estimates are reasonable. Thirdly, it outperforms the baselines on challenging predictive power benchmarks and is capable of dealing with various problems present within the dataset, such as multimodality and the presence of outliers. Our experiments demonstrate that the proposed model detects artificially added inefficiencies and that it is not overconfident about its predictions.

### **References**

- [1] Tibshirani R. Regression Shrinkage and Selection via the Lasso // Journal of the Royal Statistical Society. Series B (Methodological). 1996. Vol. 58, № 1. P. 267–288.
- [2] Phan D., Pradhan N., Jankowiak M. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro // arXiv preprint arXiv:1912.11554. 2019.
- [3] Hoffman M.D., Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. // Journal of Machine Learning Research. 2014. Vol. 15, № 1. P. 1593–1623.

## **ОЦЕНКА НЕЭФФЕКТИВНОСТИ ОБОРУДОВАНИЯ И ЕГО ОПЕРАТОРОВ С ПОМОЩЬЮ ИЕРАРХИЧЕСКИХ БАЙЕСОВСКИХ МОДЕЛЕЙ**

***Н.В. Гулецкий***

*Системный аналитик, СООО “ХайКво  
Солюшенс”, магистрант БГУ*

***Е.М. Радыно***

*Системный аналитик в СООО “ХайКво  
Солюшенс”, доцент Белорусского  
государственного университета, канд.  
физ.-мат. наук*

*СООО “ХайКво Солюшенс”, Республика Беларусь, Минск  
E-mail: nikolay.guletskiy@hiko-solutions.com*

**Аннотация.** Ненадлежащее состояние оборудования и его неэффективное использование приводит к завышенному потреблению энергоресурсов и повышению затрат. В данной статье описывается метод распутывания неэффективностей операторов оборудования от неэффективностей самого оборудования, основывающийся на иерархической Байесовской модели, моделирующей использование топлива.

**Ключевые слова:** Иерархические Байесовские модели, Байесовское моделирование, Методы Монте-Карло, Статистическое Моделирование.

UDK 57.089:004.932.7

## DEEP LEARNING APPROACHES TO BIOMEDICAL IMAGE SEGMENTATION



**S.K. Kurbanov**

*Assistant-professor of Tashkent university of information technologies named after Muhammad al-Khwarizmi*



**S.S. Beknazarova**

*Professor of Tashkent university of information technologies named after Muhammad al-Khwarizmi, Doctor of technical science*

*Audiovisual technologies department, Television technologies faculty  
Tashkent university of information technologies named after Muhammad al-Khwarizmi, Uzbekistan  
E-mail: abstrakt88@gmail.com*

### **S.K. Kurbanov**

*He has graduated Tashkent university of information technologies named after Muhammad al-Khwarizmi. He is working at the department of Audiovisual technologies Tashkent university of information technologies named after Muhammad al-Khwarizmi.*

### **S.S. Beknazarova**

*Professor of the department of Audiovisual technologies Tashkent university of information technologies named after Muhammad al-Khwarizmi. Doctor of technical science.*

**Abstract.** The review considers automatic image segmentation using deep learning methods in the field of medical imaging. Current developments in the field of machine learning, especially those related to deep learning, are useful for identifying and quantifying models in medical images. The main point of these achievements is the significant ability of deep learning approaches to obtain hierarchical representations of characteristics directly from images, which, in turn, eliminates the need for manually created functions. Deep learning is rapidly transforming into state-of-the-art medical imaging technology and has led to improved productivity in various clinical applications. This review discusses the basics of deep learning methods, as well as an overview of successful implementations that include image segmentation for various medical applications. Finally, some research questions are highlighted, and the need for further improvements is indicated in the future.

**Keywords:** machine learning, deep learning-based classifier (DLC), convolutional neural network (CNN)

**Introduction.** Medical imaging is an integral part of a modern healthcare system for performing non-invasive diagnostic procedures. This includes the creation of visual and functional representations of the inside of the human body and organs for clinical analysis. Its various types include: x-ray methods such as conventional radiography, computed tomography (CT), and mammography; molecular imaging; magnetic resonance imaging (MRI) and ultrasound (USA). In addition to these medical imaging techniques, clinical images are increasingly being used to diagnose various conditions, especially skin-related [1].

There are two components of medical imaging: 1) image formation and reconstruction, and 2) image processing and analysis [2]. Image formation includes a set of processes by which two-dimensional (2D) images of three-dimensional (3D) objects are generated, while reconstruction is based on a set of iterative algorithms for generating 2D and 3D images, as a rule, from projection data of an object. Image processing, on the other hand, involves the use of algorithms to improve image

properties, such as noise removal, while image analysis extracts quantitative information or a set of characteristics from an image to identify or classify an object.

#### Machine learning

A typical application of the machine learning-based image segmentation approach is the ROI classification, for example, a sick region or a healthy region. The development steps for such an application begin with a preprocessing step, which may include using a filter to remove any noise or to increase contrast. After the pre-processing step, the image is segmented using a segmentation technique such as threshold processing, clustering based approach and edge based segmentation. After segmentation, elements are extracted based on color, texture, contrast, and size information from the area of interest. Dominant features are then determined using feature selection methods, such as basic component analysis (PCA) or statistical analysis. Subsequently, the selected functions are used as input to the ML classifier, such as SVM or NN. The ML classifier uses the vector of input objects together with the labels of the target class to determine the optimal boundary separating each class [3]. After training the ML classifier, it can be used to classify new unknown data to determine its class. Typical problems include determining the appropriate pre-processing requirements based on the raw image properties, determining the appropriate features and the length of the feature vector, and the type of classifier among others.

#### Deep learning-based classifier (DLC)

The DLC can process the raw image directly, which means that no preprocessing, segmentation, or feature extraction is required. However, most deep learning approaches require image resizing due to input limitation. While some methods require normalization of intensity and increased contrast, which can be avoided if during training methods of increasing data are used, which will be discussed later. As a result, the DLC has a higher classification accuracy, since it can avoid errors associated with an erroneous feature vector or inaccurate segmentation [3]. A comparison of the ML and DLC approaches is shown in Fig. 1 below. DLC-based approaches have shifted the focus of research from traditional image processing techniques to designing functions to designing a network architecture for optimal results. DLC networks typically have several hidden layers, which means that more mathematical operations are performed than ML-based approaches, and thus models require more computational resources.

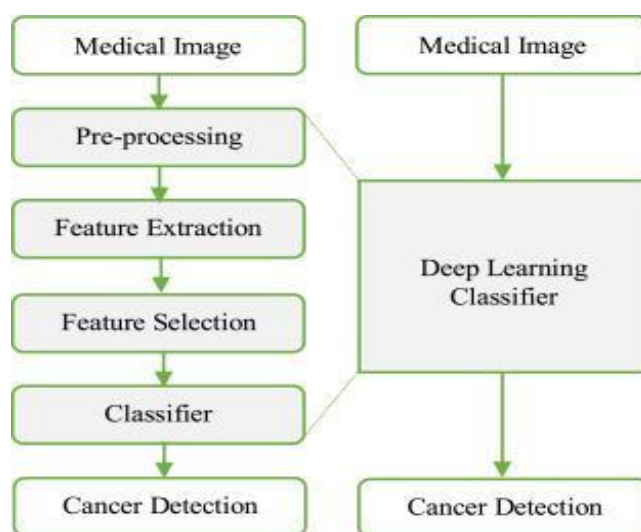


Figure 1. – Change in classifier approach using typical machine learning algorithm and deep learning.



### **Deep learning architecture – convolutional neural network (CNN)**

Among the many deep-learning architectures, CNN is the most widely used because it is very similar to regular NN. Unlike typical NN shown in fig. 2 (a), CNN accepts the image as input and has a three-dimensional arrangement of neurons that connects to a small area of the previous layer instead of the entire layer, as shown in Fig. 2 (b) below. CNN contains levels that include a convolutional level, a non-linear activation level, such as a straightened linear unit (ReLU) level, a union level, or a fully related level. The convolutional layer applies the convolution operation between the pixels of the input image and the filter to obtain the volumes of maps of objects containing objects extracted by the filter. ReLU is a non-linear activation level that applies the function  $f(x) = \max(0, x)$  to input values to increase non-linearity and increase learning speed. The pool layer lowers the sample of input values to reduce the spatial dimension of the image, to improve computational costs and prevent overlapping, and are invariant for translation, since the calculations are based on neighboring pixels [4]. A fully connected layer is usually the last CNN layer and is similar to the hidden layers of a traditional NN in the sense that all the neurons in this layer are connected to the neurons in the previous layer. As mentioned earlier, CNN is commonly used for classification tasks. To use CNN for semantic segmentation, the input image is subdivided into small sections of the same size. CNN classifies the center pixel of a patch. The patch then slides forward to classify the next center pixel. However, this approach is ineffective, since the overlapping elements of the sliding sections are not reused, which leads to the loss of spatial image information when the objects are moved to the final fully connected network layers. To overcome this problem, the use of a fully convolutional network (FCN) was proposed in which the final fully connected CNN layers were replaced by transposed convolutional layers, as shown in Fig. 2 (c), which applies up sampling on low resolution maps of objects to restore the original spatial dimensions while performing semantic segmentation [5].

Typically, deep neural networks are trained using the back propagation algorithm in combination with an optimization algorithm such as gradient descent. The process includes determining the gradient of the loss function, which is used by the optimization algorithm to update the network weights in order to minimize the value of the loss function.

#### **Types of Biomedical Images**

There are various types of biomedical images that depend on the imaging method. Some of the commonly used methods of biomedical imaging are presented below. The list below is not exhaustive, because with the development of technology, new imaging methods are being introduced to ensure better and timely diagnosis.

##### **Clinical images**

Clinical images are digital images of the patient's body and are often used to document injuries, burns, or skin lesions. The automatic analysis of these images can be used to track the effectiveness of treatment over time. These images are widely used for dermatological and cosmetic procedures to track before and after the presentation of the skin or anatomical structure. The most widely used application of clinical imaging for the detection of skin cancer known as melanoma.

##### **X-ray imaging**

Radiography is the most widely used imaging technique for detecting bone fractures and dislocations. The generated image is two-dimensional. The National Institute of Health (NIH) provided open access to 100,000 chest x-rays with relevant data and diagnoses to improve image analysis methods [6]. Similarly, the Massachusetts Institute of Technology (MIT) has published a dataset containing a collection of more than 350,000 chest x-rays for developing machine learning models to automatically detect 14 common diseases, such as pneumonia or an punctured lung, etc. [7].

### Computed tomography (CT)

CT refers to a computer imaging procedure in which X-rays are directed at the patient 360 degrees to obtain detailed images of cross sections of internal organs, bones, soft tissues and blood vessels in the body. Images are traditionally taken in the axial or transverse plane and perpendicular to the long axis of the body. However, these images, which are also known as slices, can be reformatted into several planes and can generate a three-dimensional image. It is widely used to detect cancer by localizing the presence of tumors and their size and is one of the most widely solved problems of biomedical imaging. The National Institute of Health (NIH) provided open access to 32,000 CT scans with associated data and diagnoses to improve lesion recognition accuracy [8].

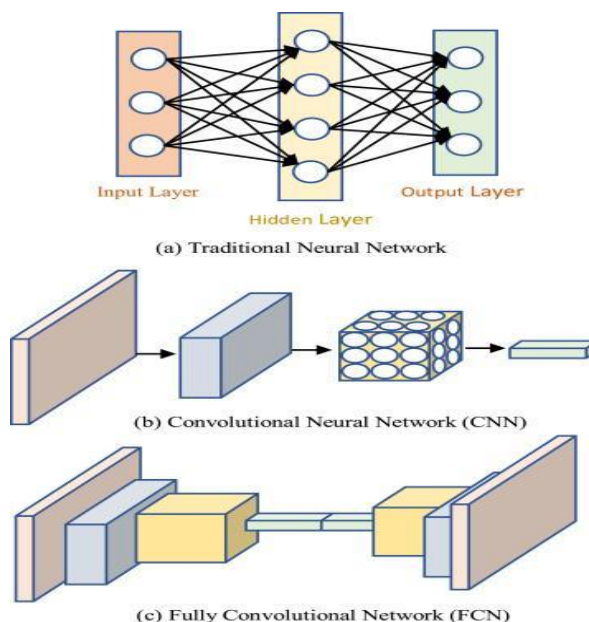


Figure 2. – (a) A 2-layer Neural Network (one hidden layer of 4 neurons and one output layer with 3 neurons), and three inputs, (b). Convolutional Neural Network (CNN) and (c) Fully Convolutional Network (FCN).

### Conclusion

A review of deep learning approaches for segmenting biomedical images highlighted some key points. All of these studies were based on empirical results that demonstrated the effectiveness of the proposed approach for this application with limited data sets. The question that remains is why deep learning approaches work for this problem. Understanding the answer to this question is an open area of research. Many researchers are working on creating new visual approaches that will help to intuitively understand the characteristics maps obtained from hidden layers [9,10,11]. In addition, many researchers do not address the problem of generalizing the network response if the source of data changes. This is what will be affected by the change in the data acquisition device, as this can lead to changes in image characteristics, such as light levels or color intensities. Lack of generalization can adversely affect network performance.

Deep learning methods have made unprecedented effects in a variety of biomedical applications - from automated imaging analysis of CT to segmentation of skin lesions. However, if more tagged images are open to all, there is more to be done. Manually setting image data by experts is a serious challenge for discovering fundamental truths. In the absence of basic truths, more emphasis should be placed on the study of uncontrolled learning methods.

### **References**

- [1.] Severity grading of psoriatic plaques using deep CNN based multi-task learning / Pal, A. Chaturvedi, U. Garain, A. Chandra, R. Chatterjee // 23rd international conference on pattern recognition, ICPR (2016) / 2016 - P. 1478-1483.
- [2.] A perspective on deep imaging / G. Wang // IEEE Access / 4 (2016) - P. 8914-8924.
- [3.] Overview of deep learning in medical imaging / K. Suzuki // Radiol. Phys. Technol. / 10 (3) (2017) - P. 257-273.
- [4.] Deep learning for visual understanding: a review / Y. Guo, Y. Liu, A. Oerlemans,
- [5.] S. Lao, S. Wu, M.S. Lew // Neurocomputing / 187 (2016) - P. 27-48.
- [6.] Deep learning and its application to medical image segmentation / H.R. Roth, et al. // (2018) - P. 1-6.
- [7.] Chest X-ray NIHCC / N. I. of H.-C. Center // (2017)[Online]. Available <https://nihcc.app.box.com/v/ChestXray-NIHCC>.
- [8.] MIMIC-chest X-ray database (MIMIC-CXR) / T. M. I. of T. (MIT)'s L. for C. Physiology // [Online]. Available. <https://physionet.org/content/mimic-cxr/2.0.0/>.
- [9.] DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning / K. Yan, X. Wang, L. Lu, R.M. Summers // J Med Imaging / 5 (Jul. 2018), - P. 103.
- [10.] Deep learning for visual understanding: a review / Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew // Neurocomputing / 187 (2016) - P. 27-48
- [11.] Visualizing and understanding convolutional networks / M.D. Zeiler, R. Fergus // European conference on computer vision (ECCV) / 2014 - P. 818-833.
- [12.] Deep inside convolutional Networks : visualising image classification models and saliency maps / arXiv : 1312 . 6034v2 [ cs . CV ] 19 Apr 2014 / 2013 - P. 1-8.

## **ПРИМЕНЕНИЕ МЕТОДОЛОГИИ DEEP LEARNING В БИМЕДИЦИНСКОЙ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЯ**

**С.К.Курбанов**

*Ассистент-профессор  
Ташкентского университета  
информационных технологий имени  
Мухаммеда аль-Хорезми*

**С.С.Бекназарова**

*Профессор Ташкентского  
университета информационных  
технологий имени Мухаммеда аль-  
Хорезми, Доктор наук*

*Кафедра аудиовизуальных технологий, факультет телевизионных технологий  
Ташкентский университет информационных технологий имени Мухаммеда аль-Хорезми, Узбекистан  
Ташкентский университет информационных технологий имени Мухаммеда аль-Хорезми, Узбекистан  
E-mail: abstrakt88@gmail.com*

**Аннотация.** В статье проводится обзор автоматической сегментации изображений с использованием методов глубокого обучения в области медицинской визуализации. Современные разработки в области машинного обучения, особенно связанные с глубоким обучением, полезны для идентификации и количественной оценки моделей в медицинских изображениях. Главным моментом этих достижений является значительная способность подходов обучения получать иерархические представления характеристик непосредственно из изображений, что, в свою очередь, устраняет необходимость создания особых функций. Обучение быстро превращается в современную технологию медицинской визуализации и привело к повышению производительности в различных клинических приложениях. В этом обзоре рассматриваются основы методов обучения, а также обзор успешных реализаций, включающих сегментацию изображений для различных медицинских приложений. К тому же, освещаются некоторые исследовательские вопросы, указывается на необходимость дальнейших улучшений в будущем.

**Ключевые слова:** машинное обучение, классификатор на основе глубокого обучения (DLC), сверточная нейронная сеть (CNN)

UDK 338.242

## **IMPROVING THE EFFICIENCY AND USE OF INFORMATION RESOURCES**



**V.G. Gorovoy**

*Subdean of Faculty of Computer Systems and Networks, BSUIR*



**A.S. Sidorovich**

*Senior teacher of Electronic Computing Machines department, BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus,  
E-mail: gorovoi@bsuir.by, sidorovich@bsuir.by*

### **V.G. Gorovoy**

*Graduated from the Faculty of Economics of BSUIR. Senior teacher of the Department of Economics.*

### **A.S. Sidorovich**

*Graduated from BSUIR. Senior teacher of Electronic Computing Machines department*

**Abstract.** The article is about improving the efficiency of the use of information resources. The problem is solved by the example of the largest bank in the Republic of Belarus - the Belarusbank.

In the current environment, attention must be paid directly to the use of information resources. One of the directions of this application should be centralization based on modern computer technologies of accounting and management accounting in the head units of domestic large economic entities, which should lead to the elimination of intermediate links in their organizational structure and a significant reduction in the number of their administrative and managerial personnel.

**Keywords:** information resources, centralization, management system, economy, bank system, organizational structure, structural divisions, tax system.

One of the directions of organizational changes in the management of the economy at various levels, which allows to increase the efficiency of the use of information resources, is the centralization of accounting and management. This centralization will allow to optimize the organizational structure of the management of large entity, reducing them to the intermediate levels, and as a result, to achieve absolute and (or) related downsizing of administrative staff .

The possibility of achieving significant positive results from such centralization can be considered on the example of the activities of the largest domestic banking institution of Joint Stock Company «Savings Bank «Belarusbank».

Belarusbank has a developed regional network, for the management of which a four-level management system has been used for many years : the central office, regional administrations, branches and exchange offices. Each unit of level 1 - 3 (central office, regional administrations, branches) has its own balance, accounting, and databases.

Since 2005, Belarusbank has been actively working to centralize the management system and all types of accounting. Filial agency began to be reorganized in stages either into branches with inclusion in the structure of other branches, or into banking service centers - structural divisions that are at the third level of management, but do not have their own balance and accounting. Their balance and accounting is transferred either to regional departments or to the remaining branches.

Based on the use of modern computer technology, such a reorganization gives positive results. What can be evidenced by the data on the number of staff and structural divisions of Belarusbank for the period from 2002 to 2017, presented in table 3.1.

The table shows that over the first five years of the study period, the number of bank employees grew. The increase in the number was 1,500 people, or 6.4%. Moreover, this happened against the background of a reduction in the number of structural divisions of the bank (filial agencies, branches). During this period, their total number decreased by 83 units (4%), and the number of branches by 53 units (34.6%). Filial agencies, as noted above, were transformed into filial agencies or centers of banking services, sometimes even completely closed. Filial agencies were liquidated. There was a reduction in the bank's branch network and, accordingly, the elimination of a significant number of jobs. But the number of staff was growing.

*Table 1. – The Number of employees and structural divisions of Belarusbank (without foreign representative offices and subsidiaries)*

Indicator	Years of the period under review			
	2002	2007	2012	2017
1. The number of staff	23 500	25 000	21,813	21,777
2. The number of structural units (filial agencies, banking services centers, branches), units including filial agencies:	2 0 59 153	1 976 100	2 087 29th	1 619 23

Note – source: [ 2 , 3 , 4 , 5 ]

From 2007 to 2012, we are already seeing the opposite situation. Now the total number of structural divisions has grown (by 100 units or 5%), while the number of filial agencies has continued to decrease (by 71 units or 71%). The number of staff at the same time decreased by 3178 people or 12.7 %. In total, from 2002 to 2012, the total number of structural divisions of Belarusbank increased by 28 units (1.4%), the number of filial agencies decreased by 124 units (81%), the number of employees decreased by 1687 people (7.2%) .

The positive effect of the centralization of management functions and accounting carried out at Belarusbank, based on the use of modern computer technologies, obtained during this period, is obvious: with a smaller number of its employees, the bank was able to achieve a wider and denser level of coverage of the country, making thus, banking services are more affordable without compromising on their quality. Reducing the number of filial agencies and, as a result, the administrative and managerial staff, Belarusbank increased the number of structural units of the fourth level (branches and exchange offices) directly engaged in the provision of banking services and having a small number of staff. Computer technology has allowed even in such units as exchange offices, in which only 1 person works, to provide a fairly wide range of banking services.

The development of the centralization of management and accounting allows Belarusbank to gradually switch to a two-level management system: 1) the central office; 2) various types of units and points of sale. However, in fairness, it should be noted that this process is not always deliberate and smooth. As evidenced by a lot of questions, the dynamics of the considered indicators in the period from 2012 to 2017: the number of units decreased by 468 units (22.4%), but the number of employees decreased by 36 people ( 0.17%)! Planed (even in 2012) transition to a two-tier management system has not yet taken place.

Undoubtedly, the reduction in the number of staff in general, and administrative and managerial staff in particular, is not, and should not be, the goal of using modern information resources in managing economic processes. The purpose of this application is a significant increase

in labor productivity of the category of workers in question and, as a consequence, the release of their working time. The time freed from the performance of highly demanding routine operations should be redirected to the solution of other important tasks that require several great mental and creative efforts. In modern conditions of rapid scientific and technological progress, changes in the socio-economic sphere, emphasis on the “knowledge economy” and the development of “human capital”, a significant part of the released working time should be directed to improving the qualification level of workers and their retraining.

The effective use of information resources in the managerial sphere does not mean the automatic reduction of administrative staff. The reduction in the number of staff, as you know, is absolute and relative. The use of modern computer technology should lead, first of all, to a relative reduction in the number of employees.

Another reason that obstructs the increase in the efficiency of the use of information resources both in the banking sector and in the economy of the country as a whole is our legislation, its often ill-conceived and frequent changes, as well as the existing bureaucratic system. A large number of a wide variety of legal acts prescribes to perform many different actions, to create a huge number of all kinds of documents (reports, requests, acts, etc.). Moreover, it would be possible to do without many of the actions prescribed by law and without many documents. But they exist and hinder the increase in the effectiveness of the use of information resources.

The activity of any domestic business entity is associated with so many different regulatory acts (instructions, decrees, orders, etc.) that it is very difficult to understand them and conduct business activities in strict accordance with them. A huge number of these documents, published by various ministries and departments, significantly complicates their coordination. As a result, in a vast sea of regulatory acts, there is also a significant percentage of conflicting accounts. A business entity operating in strict accordance with one regulatory document may not intentionally break another. Just because the provisions of one document will be ambivalent to the provisions of another.

The current situation is getting worse by the huge number of changes introduced into all legal acts. Thus, to understand the already complicated system of documents, which are regulating the activities of business entities, becomes even more difficult. It is very problematic to carry out economic activities strictly within the framework of the existing legal field. Frequent, rapid, as in a kaleidoscope, changes in the rules of the game naturally lead to their violation. What yesterday was completely permissible and legal, today is becoming unacceptable, going beyond the bounds of legality. But it's impossible to immediately adapt to new business conditions. And often, enterprises, not having time to reorganize in time, willingly offend a particular regulatory act.

This state of affairs in domestic law forces even relatively small enterprises to maintain a large staff of accountants, lawyers and other categories of administrative staff. Otherwise, to cope with many instructions and the same many changes and additions to them is simply impossible. Computer technologies in such cases are ineffective, since they are not yet able to independently monitor, analyze these changes and "tune" in accordance with them. All this is done by man. Automated information systems of business entities operate according to a specific, predetermined algorithm, designed to operate in the current conditions of the enterprise's life. As soon as these conditions change, a need arises to change this algorithm. It is clear that comprehending the changes and the new conditions for the functioning of the organization, the development of the corresponding principles for the operation of an automated information system requires a lot of labor, time and money.

It should also be noted that in the republic the system of tax reporting in electronic form is still poorly developed, which does not allow to fully switch to its automated processing. Processing paper reports is a rather time-consuming process, requiring a significant expenditure of time and effort, and, consequently, a significant number of tax workers. Also, a large number of tax specialists are required to conduct checks on the correct calculation and payment of taxes by business entities.

To improve the situation, it is necessary to ensure, first of all, the stability of the tax system. Indeed, as many researchers rightly point out, it is better to have a complex tax system, but stable than a simple but constantly changing one. When you do not know by what rules you will have to pay taxes in a month. Therefore, the Ministry of Taxes and Duties of the Republic of Belarus should concentrate its efforts not on the development of numerous and poorly thought out "improvements" of the tax system, but on its stabilization and removal of contradictions.

Moreover, this should be done with increasing automation operations level, both in calculating tax payments and in monitoring strict compliance with tax laws. For this, it is necessary not only to develop an electronic document management system between commercial organizations and tax authorities, but also to develop and implement a single electronic system for calculating and paying taxes at all business entities of the republic, which will greatly simplify the procedure for verifying the correctness of calculating and paying taxes and reduce its complexity.

Such steps should first of all lead to an absolute (relative) reduction in the number of both administrative and managerial staff of enterprises and employees of tax control bodies, and, therefore, lead to absolute (relative) savings in the wage fund in business entities, as well as budget labor costs of employees of tax authorities.

Legislation also plays a significant role in improving the efficiency of the use of information resources. First of all, its stability. If the regulatory framework of the business entity is stable, then it has much more ability to automate operations carried out by its employees. Indeed, in this case, the technology of the operation does not change for a long time, which means that the information system of an economic entity can be easily configured to perform this operation in automatic mode. The role of a person is reduced only to insignificant control of the process of performing this operation, which leads to a decrease in the workload on him and an absolute and (or) relative reduction in the number of staff.

If the regulatory framework is constantly changing, then there is a need for careful monitoring of these changes, and the introduction of appropriate adjustments to the work of the information system. On the one hand, this increases the workload of the employee, on the other hand, it does not allow making a number of operations fully automated.

It is also necessary to pay attention directly to the use of information resources. One of the directions of this application should be centralization based on modern computer technologies of accounting and management accounting in the head units of domestic large economic entities, which should lead to the elimination of intermediate links in their organizational structure and a significant reduction in the number of their administrative and management staff.

### ***References***

- [1.] "An automated mess does not become order." Does Belarus have a digital future ? [Electronic resource].
- [2.] Annual report to the general meeting of shareholders for 2002 [Electronic resource]. - Internet portal Belarusbank. - Access mode: <http://belarusbank.by/ru/deyatelnost/1648/1651>. - Date of access: 04/29/2013.
- [3.] Annual report to the general meeting of shareholders for 2007 [Electronic resource]. - Internet portal Belarusbank. - Access mode: <http://belarusbank.by/ru/deyatelnost/1648/16775>. - Date of access: 04/29/2013.
- [4.] Annual report on the general meeting of shareholders for 2012 [Electronic resource]. - Internet portal Belarusbank. - Access mode: <http://belarusbank.by/ru/deyatelnost/1648/16775>. - Date of access: 04/29/2013.
- [5.] Annual report for 2017 [Electronic resource]. - Internet portal Belarusbank. - Access mode: <http://belarusbank.by/ru/deyatelnost/1648/16775>. - Date of access: 04.25.2019 .
- [6.] Message from President A.G. Lukashenko to the Belarusian people and the National Assembly in 2006 [Electronic resource]. - The Internet portal of the Security Council of Belarus today. - Access mode: <http://law.sb.by/451/>. - Access date 04/29/2013.

## **ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ И ИСПОЛЬЗОВАНИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ**

***В.Г. Горовой***

*Заместитель декана факультета  
компьютерных систем и сетей БГУИР*

***А.С. Сидорович***

*Старший преподаватель кафедры  
электронных вычислительных машин  
БГУИР*

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь,*

*E-mail: gorovoy@bsuir.by, sidorovich@bsuir.by*

**Аннотация.** Статья посвящена повышению эффективности использования информационных ресурсов. Задача решается на примере крупнейшего банка в Республике Беларусь - Беларусбанк. В нынешних условиях необходимо уделять внимание непосредственно применению информационных ресурсов. Одним из направлений данного процесса должна стать централизация на базе современных компьютерных технологий бухгалтерского и управленческого учета в головных подразделениях отечественных крупных хозяйствующих субъектов, что должно привести к ликвидации промежуточных звеньев в их организационной структуре и существенному сокращению численности их административно-управленческого персонала.

**Ключевые слова.** Информационные ресурсы, централизация, система управления, экономика, банковская система, организационная структура, структурные подразделения, налоговая система.



UDK 159.9.016.4

## BASIC PRINCIPLES AND ANALYSIS OF HIGH-LEVEL TECHNIQUES FOR DESIGNING COMPUTER PROGRAMS FOR VIRTUAL WORLD



**K.M. Karaneuski**  
Neurologist and postgraduate student of Human Engineering and Ergonomics of the BSUIR



**M.A. Konovalova**  
Master student of the Department of Human Engineering and Ergonomics of the BSUIR



**K.N. Mezhianaya**  
Psychiatrist and has been a Research Associate of Human Engineering and Ergonomics of the BSUIR



**K.D. Yashin, PhD**  
Head of the Department of Human Engineering and Ergonomics of the BSUIR

Belarussian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus  
E-mail: yashin@bsuir.by

### **K.M. Karaneuski**

Neurologist and postgraduate student of Human Engineering and Ergonomics of the Belarussian State University of Informatics and Radioelectronics. He has published 13 articles.

### **M.A. Konovalova**

Master student of the Department of Human Engineering and Ergonomics of the Belarussian State University of Informatics and Radioelectronics.

### **K.N. Mezhianaya**

Psychiatrist and has been a Research Associate of Human Engineering and Ergonomics of the Belarussian State University of Informatics and Radioelectronics since 2014. She has published 16 articles in the field of psychiatry since 2001.

### **K.D. Yashin**

Head of the Department of Human Engineering and, Associate Professor, Candidate of Technical Sciences. Inventor of the USSR, he has 55 copyright certificates for inventions and patents. He graduated from the Belarussian State University and specialized in physical chemistry of semiconductors. Research interests: human engineering. The author and main executor of the project of creation of the Department of Human Engineering and Ergonomics at the Belarussian State University of Informatics and Radioelectronics, which provides training and graduation of specialists with higher education in the specialty "Engineering-Psychological Maintenance of Information Technologies", qualification of system engineer (software engineer).

**Abstract.** At the beginning of 2020, the world-famous Cambridge Scholars Publishing (Great Britain) published for the first time a scientific monograph called "Virtual World: Creators, Residents, and Tourists". The book contains the results of scientific research that was conducted for four years. The research was carried out at the Belarussian State University of Informatics and Radioelectronics (Minsk, Belarus). Employees of the Department of Human Engineering and Ergonomics, as well as employees of the research group "Systems and devices for environmental monitoring in the management of human life safety" took part in the research.

Part of the research material was not included in the book. In this paper, we present another part of the research that was prepared a little later. It was not published in the book. This part deals with the issues of seriality and simulacra. In our opinion, seriality and simulacrum are currently the basic principles for designing computer entertainment programs in the virtual space.

**Keywords:** high-level analysis, simulator, seriality, virtual space, epic.

### *Introduction*

At the beginning of 2020, the world-famous Cambridge Scholars Publishing (UK) first published a scientific monograph called "Virtual World: Creators, Residents, and Tourists". [1] The research was conducted at the Belarusian State University of Informatics and Radioelectronics. Here is an abstract of this monograph in the language of Cambridge Scholars Publishing.

The book acquaints the reader with the virtual world, which has become a part of the contemporary entertainment industry, an environment for both professional activity and communication. It discusses the role of computer games in people lives and the results of computer addiction studies. A separate section is devoted to the theme of Scandinavian mythology, the subjects of which are often present in the virtual world, allowing a comparison to be made between images from ancient myths and legends and their reflection in computer games. The book will appeal to a wide audience interested in interaction in the virtual world.

The scientific monograph "Virtual World: Creators, Residents, and Tourists" consists of four chapters [1]. The first chapter is "Virtualization Process Drivers". It deals with the following three issues. 1) The Scientific and Technological Revolution and its Impact on People's Lives. 2) Virtualization Drivers. 3) Information and Transformation.

The second chapter of the scientific monograph [1] is "Gamer's Dreams and Anticipation of Happiness". The chapter is devoted to the following three problems. 1) Oda to Games. 2) The Problem of moral Choice and moral Character. 3) Simulacra of Social media.

The third chapter of the scientific monograph [1] is "Battle with Virtual Life". It deals with the following three issues. 1) Computer Addiction as a multitude of States. 2) The Flow Experience and the Psychological Dominant. 3) The Virtual World in the Life of Students.

The fourth chapter of the scientific monograph [1] is "The Realm of Virtual World Life". It is devoted to the following three problems. 1) Pagan Motifs in the Realm of Virtual World. 2) Who is who in the Virtual World? 3) Vampires and their Incarnations in Various Games.

Scientific research was conducted by the authors [1] for four years. During this time, several dozen scientific articles were published in Russian and English in various scientific journals, in collections of scientific and practical conferences, as well as in the media. However, some of the research material was not included in the book. Therefore, in this paper we present another part of scientific research. It was not published in the book [1].

*The aim of this work* is to analyze the appearance of such principles as seriality, simulacra, and pagan motifs in the design technologies of computer entertainment (video films and video games).

### ***Results of research and analysis of technologies***

In this paper, we attempt to analyze video films and video games from the point of view of two basic principles of their creation: seriality and simulacrum. The rapid development of BIG DATA and cloud technologies has created a number of new phenomena in the development of people and society as a whole.

Figure 1 shows the way (algorithm diagram) to consider the analysis of two phenomena (seriality and simulacrum) that appeared during the development of technologies for creating virtual space (computer videos and computer video games). Let's have a look at these questions in order.

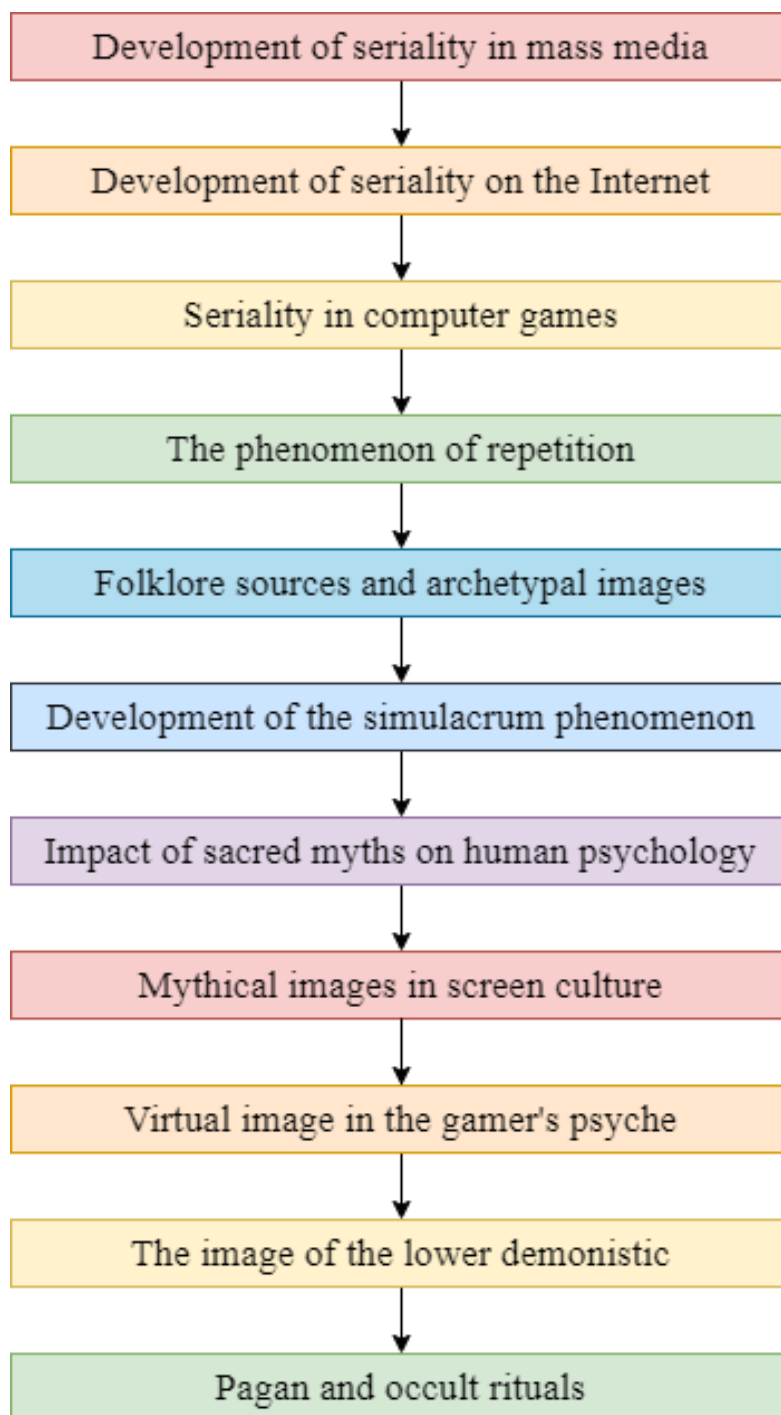


Figure 1. – Algorithm for consideration and analysis of two phenomena – seriality and simulacrum

*Development of seriality in mass media.* Mass media products placed in the virtual space have their own prehistory of development. Thus, the principle of seriality, embodied in videos and computer games, was actively developed and implemented on television in the middle of the XX century. The American researcher J. Monaco in his book "How to understand a film", wrote that since the 50s, "a series" has been understood as "having developing characters and a continuing narrative", and an episode "presents the same characters in different situations" [2].

Similarly to movies, computer games, as a type of screen art, are also created on the principle of seriality. Scientist A. McCartney considers that source of the series is a form of commercialization of fiction that has reached its height in the mass novel and in magazines published in the nineteenth

century. [3] An important role in the distribution of serial works in European countries was played by the formation of the detective genre, and then fiction.

The implementation of the principle of seriality in products that filled the Internet attracted a mass audience. Over the past decade, a new type of computer addiction has developed – from videos. A game TV movie in digital format is created in the same way, both for TV broadcasting and for information and computer technologies (ICT). The advance release is only due to the commercial capabilities of the broadcasting company.

Italian philosopher and writer Umberto Eco believed that repetition is an integral component of world art [4]. In his book "Innovation and repetition", he provided a classification of repetitions in screen culture. Retake, remake, cara, and episode are listed here. An episode, according to the writer, is an artistic form that can satisfy a new aesthetic sensibility [4].

The particularity of the relationship between technology and culture was analyzed by the classics of media theory V. Benjamin [5], M. McLuhan [6] and R. Williams [7]. In the study entitled "Television: technology and cultural form", R. Williams analyzed the technological, social and media aspects of TV. In the subgenre model, he also highlighted a series and an episode.

*Development of seriality on the Internet.* In social terms, seriality is implemented at the level of both television and Internet audiences. Russian philosopher V. A. Emelin in his dissertation research "Information technologies in the context of postmodern philosophy" (1999) identified seriality as one of the manifestations of a key characteristic of the postmodern worldview. He called it fragmentary. He also distinguished intertextuality, simulation, and pluralism as other characteristics [8].

The postmodern approach to the problem was used by American researchers A. Croker and D. Cook in the work "Television and the triumph of culture", in which television is called serial culture [9].

*Seriality in computer games.* The principle of seriality can be traced in all game products created for entertainment in the virtual space: videos, online games, as well as thematic meetings and discussions in social networks. Serial organization of production makes the viewer want to meet their favorite characters, heroes and TV hosts. Gamers are waiting for the next session of the online game and the release of a new version [10].

Similar to television, another type of screen art includes products placed in the virtual space for entertainment. It is no less naturally included in the private space of each user and successfully uses such mechanisms as the effect of presence and trust, which helps to erase the psychological border between the world on the screen and the environment of our perception [11]. At the same time, if watching a TV movie, the viewer to some extent compares himself with the characters, then in a computer game, he identifies himself with the hero of this game.

*The phenomenon of repetition.* It is known that the appearance of the highest technical possibilities of mass-production and the widest distribution of mass media significantly changed the nature of art. The characteristic features of mass media art are repetition, copying, subordination to a preset scheme, and abundance. The phenomenon of repetition has become particularly relevant. This is due to its widespread use in screen art forms (visual aesthetic space). With the development of ICT, the aesthetic needs of large masses of people have become impossible to meet through the creation of new and original products in art.

Postmodern aesthetics is based on the doubt about the possibility of permanent creation of a large number of works of art by the old methods, as an original product of creativity. Today, sophisticated people have a strong feeling that "everything has already happened". According to the French philosopher J. Baudrillard, modern art has entered the stage of simulation: "We live among countless reproductions of ideals, fantasies, images, dreams, the originals of which are left behind us" [12].

Various techniques of repetition are widely used in the media. In production, seriality is stamping, i.e. identical (literal) repetition of a sample. However, even mass production of material

products in China and other countries is not a perfect repetition of the sample. It is not the same. It is usually of lower quality. These are the so-called "generics". The serial principle in art also generates the non-identical. In this regard, J. Baudrillard states: "There is no more basic rules, no criteria of judgement or of pleasure" [13].

*Folklore sources and archetypal images.* The popularity of mass culture products is largely due to the use of folklore sources and the reproduction of archetypal images. According to the Russian researcher L. P. Saenkova, "all series and films that have become an example of mass adoration are created on the principle of restoring mythological plots and archetypal schemes" [14]. Some of them, created many years ago and received great popularity among the audience, have left a significant mark on the art of cinema and the memory of people.

The use of these stories led to the appearance of an industry for creating a variety of series. All this contributed to the development of specific terminology. Thus, the continuation, chronologically following the primary text, began to be called a sequel (from lat. sequo – I continue, I follow). Today, sequels are extremely common. These are movies and computer games.

Prequels (from lat. pre- and sequo – preceding to) show events that precede events in the primary text. There was also an extremely extensive network of sequels: the triquel – sequel of the sequel, i.e. the third in a series of chronologically developing sequels. The quadriquel is the fourth continuation, and finally the midquel, that is a continuation created later. The events depicted in it are chronologically related to the period of the original story. The series share a common theme, genre, and characters.

Other on-screen sequels include a retake (a re-shoot with characters who had success in a different narrative), a remake (a re-told story), and a spin-off where minor characters become the main actors. The division of films by the criterion of completeness/incompleteness (prolonged/non-prolonged) such as "closed end" / "open end" was also approved.

The principle of seriality is used both when creating videos and computer games. They are also characterized by potential incompleteness and endlessness. The concept of "product" as a complete and closed system is replaced by the concepts of "process" ("accumulating processes", "additive process"). Seriality has become one of the fundamental principles of virtual space design.

*Development of the simulacrum phenomenon.* Another design principle characteristic of the virtual space is simulacrum. A simulacrum is a compiled artistic image that is perceived by the human senses: visually, audibly, and kinestatically [14]. It is created by combining certain physical traits or other characteristic qualities that are idealized in a given culture (Chinese, African, European, etc.), taken from different sources. This can be a person, an image of nature, or other products, for example, that appeared in musical culture. Relatively speaking, this is a copy that is then mass produced. Plato defined a simulacrum as "a copy of a copy which original never existed."

In postmodern aesthetics, the simulacrum takes the place of the classic "artistic image". The French philosopher Jean Deleuze stated: "A simulacrum is not a degraded copy. It contains a positive charge that denies both the original and the reproduction" [15]. Other researchers, on the contrary, see the spread of simulacra as negative.

J. Baudrillard believed: "The relationship between the simulacrum and mass culture is obvious. Any connection with reality has been lost, meaning has been completely lost, and art has been turned into design. The result is an absolutely empty world of simulacra, in other words, the world of advertising" [13]. In his book "Simulacra and simulation", he identified three orders of simulacra: 1) simulacra natural, naturalistic, based on the image, imitation and fake; 2) simulacra productive, aimed at increasing productivity, based on energy, force, its material embodiment in the machine or in the entire production system; 3) simulacra of simulations based on information, modeling, and cybernetic play [13].

The imagination of science fiction writers anticipated the appearance of new mechanisms and machines. Scientific and technical thought of the twentieth century created the possibility for the appearance of a large number of simulacra of the second order. The computer is one of them.

Currently, the computer attracts people's attention more than others. With the development of information and computer technologies, "simulacra of simulations" with their hyperreality began to be created. [13]. And now, according to J. Deleuze: "Any model is lost in the dizzying abyss of simulacrum" [16].

The inclusion in the virtual space of computer games created exclusively for entertainment with a greater share of fantasy stories will help people to believe in the existence of other worlds, the main images and objects of which are simulacra. Users of the virtual world are put in the world of simulacra.

*The impact of sacred myths on human psychology.* The possibility of the impact of sacred myths on human psychology has been recognized by many prominent scientists: Paul Janet, Sigmund Freud, Carl Gustav Jung, Johan Huizinga, Mircho Eliade, and others. A number of studies and publications are devoted to this important topic of the cultural life of society. Work on the search and study of the heritage of humanity does not stop at the present time. This is not just honour to the past or an abstract interest.

In the process of the formation of Christianity, some of the mythological beliefs were included in its religious and cultural paradigm in a modified state. Thus, the Scandinavian "Saga of the Niflungs" turned into a medieval chivalric epic "The death of Nibelungs" [17]. Other stories were banned, but later some of them became fairy tales, which lost the ritual part of sacred myths.

The world of fabulous wealth and other norms of behavior that differ from Christian morals, full of mysticism and rituals, perfectly removes the psychological stress of gamers and brings them unexpected satisfaction. Carl Gustav Jung wrote about this effect of rituals on people in his book "The soul and myth: six archetypes of man" [18]. This proved to be a successful counterbalance to the settled understanding of the sublime essence of God in the Christian society. In the minds of most people, the idea has been formed that God, who stays in the transcendental sphere, cannot be known. This led to an increase in skepticism and disbelief. This helped to increase the distance between a person and his spiritual ideal.

*Mythological images in screen culture.* Social upheavals and rapid development of science in Europe (during the XIX – XX centuries) contributed to scientific and technological progress. This raised the cultural level of people in a short time. This has improved the material life of people. This significantly weakened the position of religious teachings. The increased cognitive and creative activity of people turned to science their expectations in explaining all the secrets of the universe.

The gap in knowledge was filled by writers and myth-makers. They widely used folklore sources in their works. Skilfully using people's curiosity and thirst for new experiences, they quickly created works that became bestsellers. The world they created was recreated in movies and video games. This marked the appearance of new trends in screen culture. According to the Belarusian scientist L. P. Saenkova: "All series and films that have become an example of mass adoration are created on the principle of restoring mythological plots and archetypal schemes" [19].

An arbitrary interpretation of mythological images and unconscious myth-making gave rise to archaic patterns of character behavior in movie and video game scenarios. Such scenarios described a dark world of mythological entities. In the countries of Eastern Europe, in a society raised in the spirit of atheism, an incredible number of psychics, fortune tellers and other representatives of the occultism appeared. Most of them have been considered witches and sorcerers by Christianity for centuries [20].

Ethnologists say that antiquity and the wild state are in our modern life. Their existence is proved by studies of sociologists [21, 22]. Scientists are studying the effect of presence experienced by a person in the virtual space. In the process of perception of events unfolding on the screen, the line between the fictional world and reality is blurred [23]. In some cases, the gamer may feel like not only an actor, but even a person with the story of the hero. Often, the gamer associates himself in games with the hero and his actions, which, as it seems to the gamer, ensure the omnipotence and death of others at his desire.

*Virtual image in the gamer's psyche.* We conducted research among students studying technical specialties at a university in Minsk (151 people). The results showed that 60% of respondents construct new worlds with a fairy-tale-mystical plot and characters like "orc" or "vampire" [24, 25]. The desire to gain all-power over the situation for some gamers becomes dominant. The role of a God or a demon in a computer game is preferred by a fifth of respondents (32 people). In the role-playing game (RPG), playing from the first person, they manage global processes with actions corresponding to their character.

It is significant that in half of the cases of respondents-gamers (17 people) who manage such entities, after exiting the game, a vision appears that is an autonomous virtual image, AVI [25]. AVI takes place in a real environment after exiting the game and/or shutting down the computer. The AVI generated by the gamer's internal representations appears as a visual image. Sometimes the AVI appears in the auditory sphere of perception: "the image from the game is sitting, walking, talking." The AVI of four students at the age of 18 acquired a higher hierarchy. Respondents said: "The image from the game controls me."

*The images of the lower demonistic.* It is known that primitive man spiritualized the objects surrounding him, the natural power, and his actions. He correlated them with the possibility of interacting with the invisible power.

Elves (Alves in Scandinavian mythology) and fairies were created by representations of stone age people. Therefore, elves and fairies are afraid of one type of metal, and immediately disappear from a metal clink, lose their magical power. According to Irish legends, they have empty bodies. These creatures are small, and their children have frightening appearance and monstrous appetite.

Elves and fairies are no strangers to theft. These creatures need human help for procreation. However, for such a person, who got into that world and came into contact with them, there is a threat to stay forever in their otherworldly dark world [26].

In computer games, dark mythological entities are portrayed as powerful associates of gamers. These associates are ready to oblige the master's will like fairy-tale servants. Fortunately, Microsoft canceled the release of the game *Scalebound* using a similar image. The developers said on Twitter that canceling the game would be useful primarily for gamers.

*Pagan and occult rituals.* In folk sagas and treatises of travelers, telling about pagan rituals of human sacrifice, terrible scenes of murder, bloodlessness of bodies, and sprinkling of sacrificial blood with brooms on a banquet table are described. People who already professed a monotheistic religion (Christianity or Islam) at that time were shocked by such feasts [27]. At the same time, the bearers of rites in the tribal system firmly believed in the need of such actions.

The plots of computer games and movies (along with images of archaic entities) have the similarity of ancient occult rituals. In Scandinavian myths, there is the use of organs from the bodies of people sacrificed to the god of war. These organs were used to predict the future. This ritual of mutilation was most widely carried out on prisoners.

The extant mythology of pagan beliefs was written in the last period of the tribal system. It preserves descriptions of traditions that have remained in the memory of generations. These traditions were passed on and followed. This allowed many peoples to survive and maintain their integrity and community in that barbaric era.

### ***Conclusion***

It becomes obvious that in the artistic culture of mass media, represented by the means of information and computer technologies, such general patterns that are characteristic of postmodern art as the principle of seriality and the principle of simulation are being updated. One of the key characteristics of the principle of seriality is repetition. Repeatable structures of mass art have naturally gained a place in the virtual space since social and humanitarian knowledge has been successfully processed on computers and transmitted through a network.

Just as technical capacities are already having a devastating effect on natural processes, the impact of humanitarian information transformed with the help of technical means will have a similar effect on their mental processes, with the aim of sophisticated influence on the human imagination.

The humanistic ideals of Christianity gave everyone, the weakest person and even the sinner, a chance for life and a better future. The revival of forgotten beliefs in the minds of people, the imitation of bloody scenes in computer games, can have a significant impact on the psyche and worldview of people. Science has no evidence that moral norms and the obligation to follow them are inherited. The degree of human civilization is determined by the social context: education and environment [28]. The virtual world has also become a habitat for many people. A significant number of users spend 40 or more hours a week in it for entertainment purposes.

Manipulation of the gamer's emotional experiences in a game with a mythological plot may cause him to experience a state of ecstasy similar to that experienced by people of the pre-Christian era and the early middle ages when performing certain rites. Devotion to bright, dynamic images from myths inevitably activates users' interest in the subject of ritual actions. Erich Fromm defined the danger of such devotion as "... readiness to accept any ideology and any culture as a promise of an exciting life" [29].

### **References**

- [1]. K.N. Mezianaya, K.M. Karaneuski, K.D. Yashin. Virtual world: Creators, Resolents, and Tourists // Cambridge Scholars Publishing, UK, 2019. – 242 p.
- [2]. Monaco J. How to read a film / J. Monaco. – 3rd ed. – New York: Harbor Electronic Publ., 2000. – 672 p.
- [3]. McCarthy A. The television genre book / ed.: G. Creeber, T. Miller, J. Tulloch. – London, 2001. – P. 47-46.
- [4]. Eco U. Innovation and repetition. Between the aesthetics of modernity and postmodernity / U. Eco // Philosophy of the postmodern era: a collection of translations and abstracts / edited by A.R. Usmanov. – Minsk, 1996. – P. 49-73.
- [5]. Benjamin W. Work of art in the age of mechanical reproduction / W. Benjamin. – Moscow: Medium, 1996. – 240 p.
- [6]. McLuhan M. Understanding of media: external extensions of a person [Electronic resource] / M. McLuhan; translated from English. By V. G. Nikolaev. Moscow: KANON-press-C, Zhukovsky: Kuchkovo pole, 2003. – Mode of access: <http://lib.rus.ec/b/154221/> – Data of access: 01.08.2015.
- [7]. Williams R. Television: technology and cultural form [Electronic resource] / R. Williams/ – London: Routledge, 2003. – Mode of access: [http://classes.dma.ucla.edu/Winter13/8/Williams\\_Television\\_Flow.pdf](http://classes.dma.ucla.edu/Winter13/8/Williams_Television_Flow.pdf). – Data of access: 09.08.2014.
- [8]. Emelin V.A. Information technologies in the context of postmodern philosophy / V.A. Emelin, Moscow: Moscow state University, 1999. – 26 p.
- [9]. Crowker A. Television and the triumph of culture / A. Crowker, D. Cook // Comments, 1997. – No. 11. – P. 159-168.
- [10]. Agafonova N.A. General theory of cinema and fundamentals of film analysis / N.A. Agafonova. – Minsk: Theseus, 2008. – 388 p.
- [11]. Bagirov E.G. Essays on the theory of television / E.G. Bagirov. – Moscow: Iskusstvo, 1978. – 151 p.
- [12]. Baudrillard J. The Transparency of Evil / J. Baudrillard. – Moscow: Dobrosvet: KDU, 2009. – 257 p.
- [13]. Baudrillard J. Simulacra and simulation (translated by A. Kachalov). – Moscow: Postum, 2016. – 238 p.
- [14]. Saenkova L.P. Mass culture: evolution of spectacular forms / L.P. Saenkova. – Minsk: Belarusian State University, 2003. – 123 p.
- [15]. Deleuze G. The logic of meaning / G. Deleuze. – Moscow: Academic project, 2011. – 472 p.
- [16]. Deleuze, G. Difference and repetition / G. Deleuze; translated by N.B. Mankovskaya, E.P. Yurovskaya. – Saint Petersburg: Petropolis, 1998. – 384 p.
- [17]. Cardini F. The origins of medieval chivalry (translated by V.P. Gaiduk). – Moscow: Progress, 1987. – 360 p.
- [18]. Jung K.G. Soul and myth: six archetypes. (translated by A. Spector). – Moscow: AST, Minsk: Harvest, 2005. – 400 p.
- [19]. Saenkova L.P. Mass culture: evolution of spectacular forms / L.P. Saenkova. – Minsk: Belarusian State University, 2003. – 123 p.
- [20]. Goethe I.V. Collection "Erlkönig" / I.V. Goethe / Translation by V.A. Zhukovsky. – Moscow: Komsomolskaya Pravda, St. Petersburg: Amphora, 2012. – 238 p.
- [21]. Tylor E.B. Primitive culture. Moscow: Politizdat, 1989. – 572 p.



- [22]. Benhabib S. The Claims of Culture: Equality and Diversity in the Global Era / S. Benhabib. The Claims of Culture. / translated and edited by V.L. Inozemtsev. – Moscow: Logos, 2005 – LX. 289 p.
- [23]. Hoffman D.L., Novak T.P. Marketing in hypermedia computer-mediated environments: Conceptual foundations // Journal of Marketing, 1996. – P. 50-68.
- [24]. Lee S., Kim G.J., Rizzo A., Park H. Formation of spatial presence: By form or content? // Proc. 7th Annual International Workshop on Presence / M.A. Raya, B.R. Solaz (eds.). –Valencia: Universidad Politecnica de Valencia, 2004. – P. 20-27.
- [25]. Yashin K.D., Mezryanaya K.N., Zalivako S.S., Korenevsky K.M. Influence of the virtual world on the personality of the student. // Information Technologies, 2013, T. 206, T. No. 10. – P. 50–55.
- [26]. Nosov N.A. Virtual psychology. – Moscow: Agraf, 2000. – 430 p.
- [27]. Petrukhin V.Ya. Myths of Ancient Scandinavia. – Moscow: Astrel, 2002. – 464 p.
- [28]. Doroshevich E.K. Use of folklore in the formation of the worldview of the subject of culture / E.K. Doroshevich // Minsk: Art and culture. – 2011, vol. 4 (4). – P. 114-120.
- [29]. Fromm E. Escape from freedom. – Moscow: Progress, 1990. – 269 p.

UDK 614.875:004

## REDUCING OF RADIATION DOSE FOR X-RAY USING CONTOURLET TRANSFORM AND BLOCK THRESHOLDING TECHNIQUE



**M. W. Alhamd**

*PhD, Ministry of Science and Technology Radiological and Nuclear Safety Directorate*



**T. M. Salman**

*PhD, Mustansiriyah University Engineering Faculty- Electrical Engineering Department*



**M. Melqonyan**

*PhD, Lecturer Yerevan University after Movses Khorenats*

*Ministry of Science and Technology Radiological and Nuclear Safety Directorate, Iraq*

*Mustansiriyah University, Iraq*

*Yerevan University after Movses Khorenats, Armenia*

*E-mail: melqonyan-1981@mail.ru, yaseen\_mohsin@yahoo.com, tjibori@gmail.com*

### **Dr. M. W. Alhamd**

*He was born in Baghdad, Iraq in 1975 He received a Master's degree in Physics in 2005. He received a master's diploma degree in Physics in 2014. And PhD in Nanotechnology in 2019 from Belarussian State University of Informatics and Radio Electronics, Republic of Belarus. He got a PhD in Nanotechnology. Inventions and registered patents in Belarus and Iraq on nanotechnology, Patent in Nuclear Physics, Patent in medical physics. He has several works in fields of magnetic nuclear resonance imaging, radioactive waste, UV rays with oxide compounds. aluminum, cadmium sulfide and zinc oxide, Radiation from X-ray machines and safety uses of nuclear energy and electromagnetic energy.*

### **Tariq M. Salman**

*He was born in Baghdad, Iraq in 1972. He obtained his B.Sc. in Electrical Engineering in 1995, M.Sc. in Communication Engineering in 2003 at University of Technology, Iraq and PhD in Telecommunication and Network devices in 2012 at Belarussian State University of Informatics and Radio Electronics, Belarus. From 2006 to 2012 he worked as a lecturer in the Electrical Engineering Faculty, at Al-Mustansiriyah University, Iraq. Since the beginning of 2018, he works as an assistant professor in the same Faculty. He is a consultant member of the Iraqi Engineering union since 2013. He is interested in the subject of wireless and network devices, video and image processing systems.*

### **Manushak Melqonyan**

*She was born in Yerevan, Armenia in 1981. She obtained her B.Sc. in applied mathematics, information and computer sciences in 1995, M.Sc. in computer science and applied mathematics in 2015 at Yerevan University Movsesa Xorenaci and PhD in Applied Mathematics at 2017. She worked as Lecturer on Yerevan University after Movses Khorenats in Republic of Armenia. She is interested on Image processing techniques and machine learning; she had several publications in the fields of applied mathematics and informatics.*

**Abstract.** Medical Image Denoising represents one of the fundamental challenges in the field of biological image processing and computer vision. X-Ray imaging is one of the widest used image acquisition technique in hospitals. Image denoising goal is to enhance the original X-Ray image by suppressing noise from a noise-contaminated version of the image. In this paper, a comprehensive survey of the types of noise added in X-Ray examination images was conducted, and to develop a database of five known types of X-Ray examinations in the hospital, three devices were randomly selected and 100 patients included in this survey. The survey results have been compared with the results of the International Atomic Energy Agency (IAEA) and the data have been useful to serve as a national guide or reference.

Secondly, a denoising algorithm using Contourlet transform with blocking method was proposed to improve the quality of X-Ray images, which reduces the radiation dose resulting from repeated radiographic attempts to patients and workers.

**Keywords:** X-ray, Blocking, Block thresholding, Contourlet, MRI images and Standard Deviation of Blocks

*Introduction.* To ensure the safety of workers in the radiation fields, as well as to reduce the dose received by the patient, an image improvement program should be applied by image processing of diagnostic radiology departments in medical institutions, as it leads to accurate diagnosis without exposure to unjustified radiation doses by reducing the noise generated by the devices. It takes a set of measures related to the performance, equipment, and X-Ray devices used in medical diagnosis, including digital image processing. Quality assurance programs have been established and approved in the rules and laws of most countries in the world and in the safety standards of international organizations [1]. The image improvement system aims to:

- 1 - Improve the image of the various body tissue structures on the X-Ray film or screen.
- 2 - Reducing radiation exposure to patients and workers.
- 3 - Reducing costs (film consumption, radiant tube life).

The quality assurance program is defined as the organized effort by workers in the diagnostic medical institution in order to maintain that there is no change in the technical specifications of devices over time.

A database of radiation doses has been created and has helped in comparison. For the first time, we did with radiopathic physicians working in X-Ray units performed 500 examinations. The result was that the annual dose 0.05 mSv and chest exams accounted for 63% of the total, that the radiation doses decreased After conducting quality assurance tests for the devices in this study. The dose was measured by (RMI Model 240) device and is intended for X-Ray with quality assurance tests. Analysis of test results showed that the causes of added noise and distortion was as below:

- 92% have tube voltage within the permissible limit.
- 90% have exposure time within the permissible limit.
- 86% to 98% straightness of its X-Ray beam is acceptable.
- 81% has a good radiation focus.
- 88% have *mAs* within the permissible limit.

The medical image denoising target lies primarily in the preservation category of image data quality in which the image signal transfer process affects the signal using many types of communication channels [2]. Denoising images is an important technique for image processing analysis. In this work, two noise models, Additive Noise Model (AWGN), and Impulsive Noise were used (Salt & Pepper) [3], for experimental requirements, noise signal gets added to the original signal (Image) to produce a corrupted noisy signal [4, 5]. The additive noise model follows the following rule:

$$W(x, y) = s(x, y) + n(x, y) \quad \dots (1)$$

Where,  $s(x, y)$  is the original image intensity and  $n(x, y)$  denotes the noise introduced to produce the corrupted signal  $W(x, y)$  at  $(x, y)$  pixel boundary, the two types of noise mentioned before (AWGN, Salt & Pepper) are evenly distributed over the signal [6]. This means that each pixel at the noise image, is the sum of true pixel value and random gaussian noise distributed value. Figure (1) shows the normal noise distribution over the transmission channel [6].

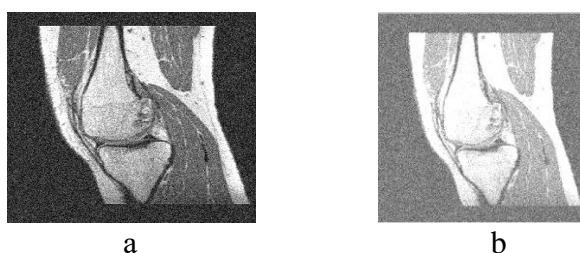


Figure 1. a – Noisy image of ( AWGN ) with, Mean ( $\mu$ )=0.5, Variance ( $\sigma^2$ ) = 0.05,  
 b – Noisy image of ( AWGN ) with Mean ( $\mu$ )=1.5, Variance ( $\sigma^2$ )= 0.5  
 The probability density of the normal distribution is:

$$PG(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-\mu)^2/2\sigma^2} \dots (2)$$

Where:  $\mu$  is the mean or expectation of the distribution,  $\sigma$  is the standard deviation,  $\sigma^2$  is the variance [7]. The Impulsive noise (Salt & Pepper) effect is shown in Figure (2). impulse noises are classified into two major types, salt and pepper noise (equal height impulses) impulse values are represented as 0 and 255 pixel value and random-valued impulse noise (unequal height impulses) impulse values are between 0 and 255.

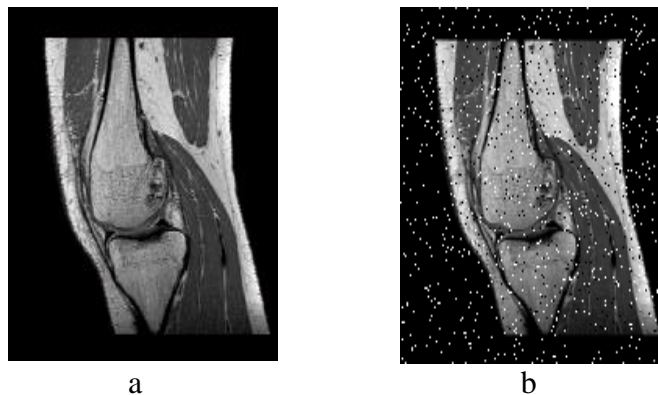


Figure 2. – The effect of impulsive noise: a – Original (X-Ray) image, b – noisy image with salt & pepper noise

*The Proposed Method (CTB).* In this paper, the focusing will be done with specific type of digital image (X-RAY). The multi-frame X-RAY scanner, that received from a digital photo-light scanner, has complain from many types of additive noise [8,9,10]. Denoising (noise removal) by comparing the variance  $\sigma^2$  of intensity occurred, with the mean value of normal standard image, such as Barbara. Compartment between the standard deviation variation of Barbara, figure 3a with the same variation of figure 3b of X-RAY -knee image, shows the difference of intensity shots in X-RAY much more than a slandered image such as Barbara, thus, the analysis and the next preprocessing techniques, such as 2D wavelet transform and contourlet transform, will concentrated for the field of X-RAY images [11]. Figure 3 shows the standard deviation variation differences between the slandered and X-RAY image [12, 13].

The recent ideas has been proposed by the researchers, for how to select the good threshold value for image denoise (or compression) [8], although they are mostly used the wavelet transform bands, they suggest other types of transform such as (DCT) [9]. In this paper, the focusing will be done on the contourlet transform with blocking operation [14].

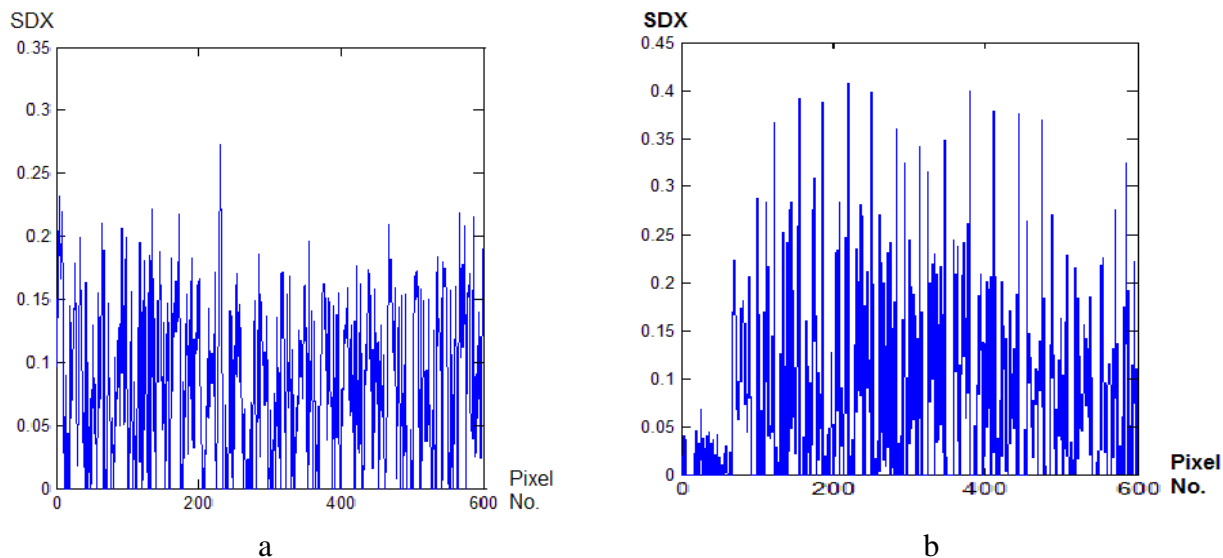


Figure 3. – The standard deviation variations: a – standard image [barbara], b – X-RAY image

*Blocking operation.* Blocking Algorithm has been used in the proposed system, instead of using the hole image, the sub-band of contourlet transform (with multi dimension arrays) instead of the sub-band wavelet transform (with two-dimension arrays). It is a new technique for maximizing signal to noise ratio.

For overall image size, (512x512) pixels, it can be use the block size of (32x32) pixel size, it means that the hole image will divide into (16x16) or 256 sub-image (block), as shown in figure 4.

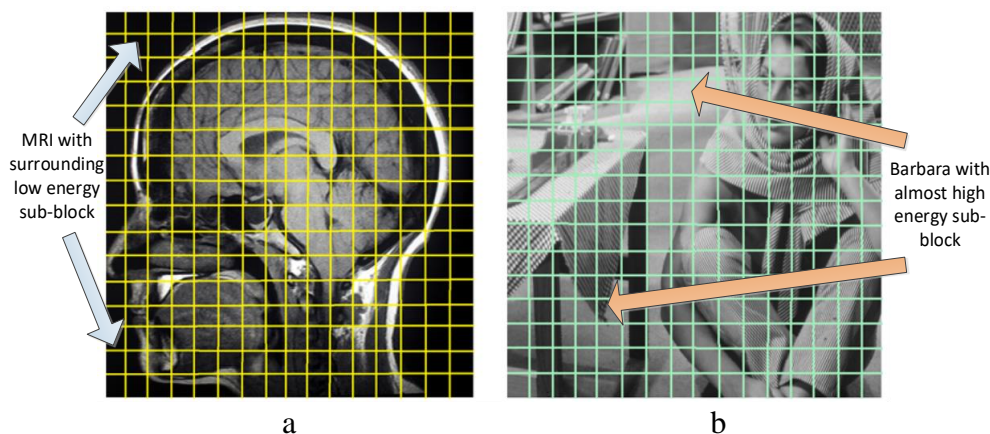


Figure 4. a – X-RAY image blocking, b – Barbara image blocking

The analysis of preprocessing by using the mentioned techniques (wavelet, contourlet- with and without blocking) are achieved by applying the system shown in the figure 5.

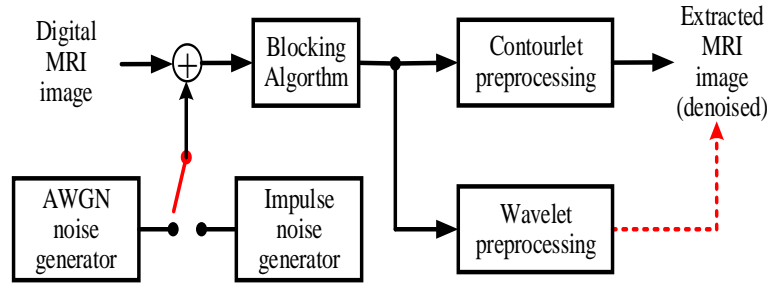


Figure 5. –The proposed system (CTB)

From the figure 4. It can be observe that the surrounding blocks, in figure 4a has lower energy (more darkness) as compared with the blocks inside the X-RAY object (skull), while there is a little difference in energy concentration of the blocks everywhere at the image of standard image (such as Barbara), figure 4b, as well seen in the figure 3(a,b) the standard deviation variation values of both, observe the more randomness of variation with the X-RAY than standard image. Whereas Do and Veterreli had proposed new multi-dimensional method for image denoising called contourlet transform this method is effectively able to find the contours and edges, its constructed of Laplacian pyramid (cluster) and filter banks [6,7,8]. The contourlet coefficients are generated by passing the bandpass output, then directed them to directional filter bank, the Laplacian pyramid is again received the output of lowpass in clustered manner till obtaining the image details like LL (Low-Low), LH (Low-High), HL (High-Low), and HH (High-Low). the schematic diagram of contourlet transform explained briefly in the Figure 6.

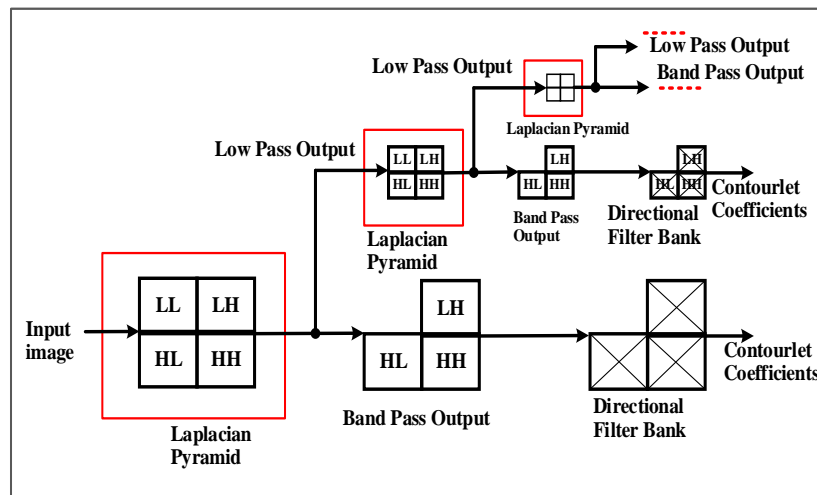


Figure 6. – Block diagram of Contourlet transform

The threshold value choosing for contourlet represents a problem in which solved in this work in automated choosing of the value of  $\sigma_x$ :

It means that, it should estimate the variance of noise value, and to find the optimal threshold value to be used in Contourlet transform with blocking technique. The variance of noise is calculated by:

$$\sigma_x^2 = \left\{ \frac{\text{med}(K_{i,j})}{c} \right\}^2 \dots\dots (2)$$

Where  $K$ , is the contourlet coefficients of the image,  $C$ , constant calculated by normalized mean value of individual block, the threshold value used for extracting contourlet coefficients is calculated by [6].

$$Th = \frac{3}{4} M \left( \frac{\sigma_x^2}{\sigma_d} \right) \dots \dots \quad (3)$$

Where  $M$ = the number of pixels in each individual block ( $x*y$ ),  $\sigma_d$  is the standered deviation of noisy source image.

*The experimental results.* 1- The results of manual methods of measuring and controlling image quality of Radiation Doses.

Radiation doses were evaluated in this paper using an RMI 240 device designed for X-Ray devices. The reading rate for the three tablets that represent the Entrance Surface Dose (ESD pp) is taken in the presence of the patient, and this dose is the dose that the patient receives and the dose resulting from X-Ray scattering from the patient's body, and these doses can be compared to the global dose limits for each type of examination And for every hometown. As in the figure (7). Table (2) represents the surface potions before and after the factory reset, it is noted that the resulting images differ in their accuracy due to human errors in reading, therefore for the mentioned reasons a method was proposed in the image processing that greatly reduces reading errors Image.



Figure 7. – X-ray measuring device RM 240

Table 1. – Represents surface doses of several parts of the body (A)PA: Posterior (B) LAT: Lateral AP: Projection Projection (C)

Projection Types						
Hospital	Chest	Skull		Lumber Spine		Pelvis
	PA	PA	LAT	AP	LAT	AP
before Noise removal						
A	0.27	3.81	3.53	12.36	14.88	11.70
B	0.28	5.05	4.02	7.00	15.70	2.6
C	0.38	4.32	5.14	6.54	13.50	7.16
After Noise removal						
A	0.29	3.58	2.84	8.39	8.03	9.74
B	0.09	4.7	3.74	3.9	8.4	1.2
C	0.24	5.12	4.37	5.78	11.26	6.42

The experiment has been done, mainly at X-RAY images, with the use of contourlet pre-processing and blocking algorithm, wavelet (2D) transform used for comparison of results for both, the standard image and blocking image. It is observed that the SNR has been improved when apply contourlet transform [10,11,12]. Figure (8, 9) represent Knee (X-RAY) with wavelet and AWGN [11] and Head (X-RAY) images.

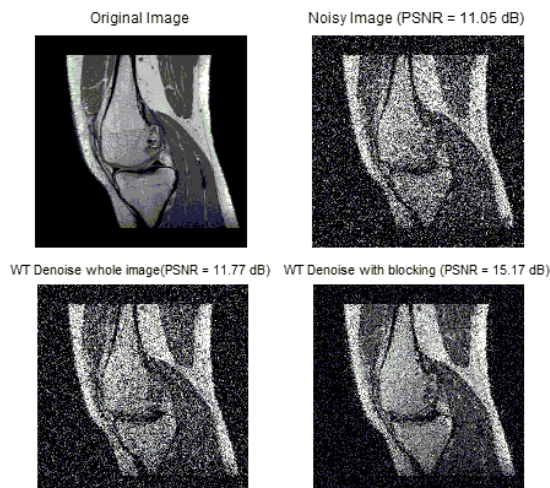


Figure 8. – Contourlet denoising for AWGN filtering (denoising) - with Barbara

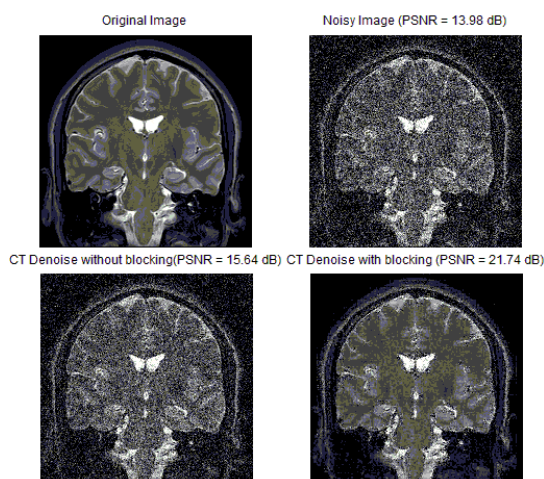


Figure 9. – Contourlet denoising for AWGN filtering (denoising)- with Head(X-RAY)

From these figures, it's very clear visually that the contourlet transform has more PSNR gain comparing the blocking method used, and hole image without blocking. For the figure 8, the noisy Barbara image with 13.98 dB, the contourlet denoising results of 19.84 dB, while it was without blocking results only a 15.22 dB, but with X-RAY images, got more visual clear effect and higher PSNR with blocking algorithm for 21.74 (for Head X-RAY). The noise added for the recent three examples was only AWGN. As compared with the wavelet transform, it is observed that got only a 15.17 dB PSNR with Knee (X-RAY). It's clear that the best image media for the best PSNR using for the proposed method are X-RAY images, with contourlet transform and blocking method (CTB).

For ensure the robustness of the CTB, in addition of adding AWGN, the impulsive noise also used, the salt& pepper (a type of impulse noise) has been added to test the system. It is more difficult in remove than the white noise, in the figure 10(a), it's clear that when comparing the results by using blocking algorithm, with the hole image denoising (without blocking), got advantage of 3 dB PSNR difference with wavelet, and about 10 dB with contourlet preprocessing for AWGN removal, Figure 10(b), thus, it is so clear that the wavelet is weak in this application and there is no sense to continue with it.



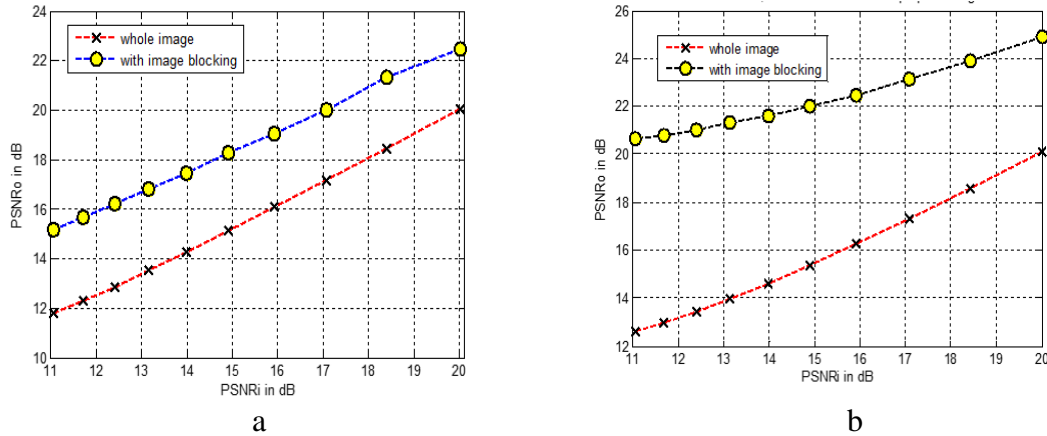


Figure 10. a – wavelet denoising for AWGN, b – contourlet denoising (both with blocking method)

Figure 11(a) shows good performance of removing salt and pepper noise, about 5 dB difference in PSNR by average, figure 11(b) shows the advantage of PSNR for MRI- Knee image.

$$APSNR = PSNR_o - PSNR_i \dots\dots (4)$$

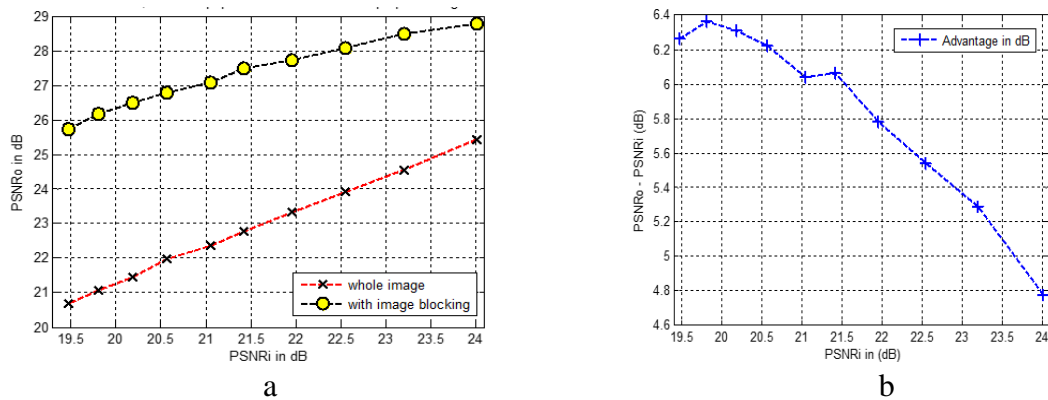


Figure 11. a – salt & pepper with contourlet preprocessing, b – PSNR advantage with salt and pepper with contourlet preprocessing (both for MRI- Knee image)

From the above figures clarify that the new method is more adequate for denoising to MRI image than other type of images such as (portrait, texture, and landscape images), because of high gain PSNR resultant of MRI compared with other types. Figure 12 (a) shows the  $PSNR_o$  with  $PSNR_i$  for Barbara, and figure 12 (b) shows the advantage of  $PSNR \approx 3$  dB between Barbara (an example for slandered image) and Knee (an example for MRI image) by applying CTB.

Table 2 shows the effect of changing the block size of MRI image to the PSNR improved in dB.

Table 2. – The effect of Block size for (MRI) image by using contourlet preprocessing

Block size (x*y)	32*32	64*64	128*128	256*256
PSNR improvement in dB	9.19	5.87	5.79	5.15

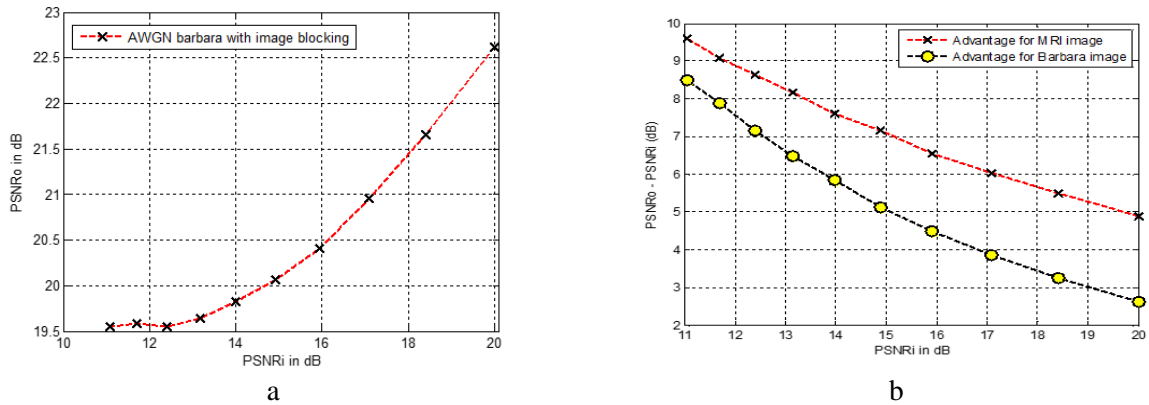


Figure 12. a – AWGN for Barbara with contourlet preprocessing, b – Advantage of PSNRo of MRI image with Barbara image using CTB

From the table shown, it can declare that as much as block size be smaller, the improvement of PSNR will be better, although there is a limitation below:

$$Bs \geq \sqrt[2]{2d} \dots \dots \dots (6)$$

While Bs is the minimum block size, d is the x dimension of original image, therefore, for a (512x512) image, the minimum block size will be (32x32) pixels. This limitation cause by the 3<sup>rd</sup> level cluster of contourlet transform.

*Discussion.* The average dose measured in this research for all devices in this study compared to the reference dose of the International Atomic Energy Agency (IAEA), but it is generally greater than the reference doses with a range of (1.7 - 1.2) as in Table (1), although this increase does not it poses a great risk to patients, but it is preferable that the dose be less than that according to levels in radiation protection, because the increase in the dose is not justified due to human errors and this increases are:

1. The age some devices used in this study.
2. The heavy operation load on the device during working hours.
3. Some radiographers, has low experience or insufficient training.
4. The psychological conditions of the medical staff will negatively affect quality.

As mentioned in Table (1), the quality assurance tests show that the devices (a, b) have succeeded in these tests and they work within the international specifications, while the device (C) has failed in most of the tests, therefore the proposed method, with image processing with contourlet transform and image blocking technique has been proposed. It shows that dividing the image into blocks improves the denoising process using Wavelet or Contourlet transforms. However, Contourlet transform denoising behavior outperforms the wavelet behavior. Simulation results on both ordinary and medical X-Ray images showed that the proposed blocking method is more powerful in medical X-Ray and MRI images than in ordinary images. Also, the results showed that the denoising improvement depends on the block size and the less block size is, the more PSNR advantage is. However, due the limited size of the LL band in Contourlet transform, there is a minimum block size to be used. This minimum block size depends of the size of the original image and number of levels in Contourlet transform.

### References

- [1]. IAEA. 2000. Technical Reports series No. 398 Absorbed dose determination in external beam radiotherapy.
- [2]. Wang, Zhou, and David Zhang. "Progressive switching median filter for the removal of impulse noise from highly corrupted images." IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing 46.1 (1999): 78-80.
- [3]. Delon, J., Desolneux, A., & Viano, A. (2016). Mixing Non-Local and TV-L1 methods to remove impulse noise from images.
- [4]. Yan, Ming. "Restoration of images corrupted by impulse noise and mixed Gaussian impulse noise using blind inpainting." SIAM Journal on Imaging Sciences 6.3 (2013): 1227-1245.
- [5]. Chang, S. Grace, Bin Yu, and Martin Vetterli. "Adaptive wavelet thresholding for image denoising and compression." IEEE transactions on image processing 9.9 (2000): 1532-1546.
- [6]. Wink, AlleMeije, and Jos BTM Roerdink. "Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing." IEEE transactions on medical imaging 23.3 (2004): 374-387.
- [7]. Luisier, Florian, et al. "Fast interscale wavelet denoising of Poisson-corrupted images." Signal Processing 90.2 (2010): 415-427.
- [8]. Rehani, M.M. 995 .Diagnostic imaging Quality Assurance.
- [9]. Thomas, R.H, Mendez, V.P. 2005. Advance in Radiation protection and Dosimetry Medicine, London.
- [10]. World Health Organization 2008. Quality Assurance in diagnostic Radiology.
- [11]. Basic quality Control Diagnostic Radiology.2008. American Association of Physists in Medicine, New York, AAPM report No. 4.
- [12]. Sadreazami, Hamidreza, M. Omair Ahmad, and MN Shanmukha Swamy. "Contourlet domain image denoising using normal inverse Gaussian distribution." Electrical and Computer Engineering (CCECE), 2014 IEEE 27th Canadian Conference on. IEEE, 2014.
- [13]. Wang, He, and Yunhong Xin. "Wavelet-Based Contourlet Transform and Kurtosis Map for Infrared Small Target Detection in Complex Background." Sensors 20.3 (2020): 755.
- [14]. Satheesh, S., and K. V. S. V. R. Prasad. "Medical image denoising using adaptive threshold based on contourlet transform." arXiv preprint arXiv: 1103. 4907 (2011).

## REDUCING OF RADIATION DOSE FOR X-RAY USING CONTOURLET TRANSFORM AND BLOCK THRESHOLDING TECHNIQUE

**Dr. M. W. Alhamd**  
PhD, Ministry of Science  
and Technology  
Radiological and Nuclear  
Safety Directorate

**Dr. Tariq M. Salman**  
PhD, Mustansiriyah University  
Engineering Faculty-  
Electrical Engineering  
Department

**Dr. Manushak Melqonyan**  
PhD, Lecturer  
Yerevan University after  
Moses Khorenats

Ministry of Science and Technology Radiological and Nuclear Safety Directorate, Iraq

E-mail: yaseen\_mohsin@yahoo.com

Mustansiriyah University, Iraq

E-mail: tjibori@gmail.com

Yerevan University after Moses Khorenats, Armenia

E-mail: melqonyan-1981@mail.ru

**Abstract.** Medical Image Denoising represents one of the fundamental challenges in the field of biological image processing and computer vision. X-Ray imaging is one of the widest used image acquisition technique in hospitals. Image denoising goal is to enhance the original X-Ray image by suppressing noise from a noise-contaminated version of the image. In this paper, a comprehensive survey of the types of noise added in X-Ray examination images was conducted, and to develop a database of five known types of X-Ray examinations in the hospital, three devices were randomly selected and 100 patients included in this survey. The survey results have been compared with the results of the International Atomic Energy Agency (IAEA) and the data have been useful to serve as a national guide or reference.

Secondly, a denoising algorithm using Contourlet transform with blocking method was proposed to improve the quality of X-Ray images, which reduces the radiation dose resulting from repeated radiographic attempts to patients and workers.

**Keywords:** X-ray, Blocking, Block thresholding, Contourlet, MRI images and Standard Deviation of Blocks

UDK [575:61]:004.6

## ADAPTABLE GENETIC ALGORITHM FOR PRIMARY PROCESSING OF MEDICAL DATA BASED ON RANDOM SEARCH



**A.K.Nishanov**  
*Professor, Information Technologies Software Department, Tashkent University of Information Technologies named after Muhammad al-Khorazmi, Uzbekistan*  
nishanov.a@tuit.uz



**G.P.Jurayev**  
*Independent researcher, Center for Information and Communication Technologies at the Tashkent University of Information Technologies named*



**N.B.Narziyev**  
*Senior teacher, Information Technologies Software Department, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan*  
n.b.narziyev@gmail.com

**Abstract.** In the article the topical software tool "Algorithmic and computer software for the primary processing of medical data" has been developed, and based on it several practical issues related to cardiology have been solved in visualized way.

The software is based on methods and algorithms for solving such issues as preprocessing of data generated by medical professionals, re-classifying study facilities and determining the importance of classes, creating table charts, and selecting a set of informative symbols that differentiate between class objects.

**Keywords:** Fishers Criterion, Feature Selection, Classification, Algorithms for an estimate calculation, Preprocessing of medical data, algorithm "A", flexible genetic algorithm based on random search.

### I. Introduction

With the ever-increasing amount of information about objects, processes and events in various areas of human activity around the world, great attention is being paid to the development and improvement of intellectual analysis systems.

One of the topical issues of modern, information-based health care systems is the transition from large-scale character space to small-scale medical data processing, namely the development of algorithms for identifying the most important symbols, and the classification of issues using them.

The software developed by the medical experts includes data processing, pre-classification of chosen objects and determination of classification levels [1, 4, 9,10], benchmarking, and selection of informative symbols that differentiate between class objects and finding a solution. The problems and issues most commonly encountered in most of the literature reviewed [5-8, 11-13] are based on the Fisher-type criteria for finding solutions. In addition, a tool for calculating the complexity of the problem under consideration was used in the calculation algorithms. [1-13].

The interface of the software "Algorithmic and computer software for the primary processing of medical data" consists of 6 modules called "Primary Data Reading", "Algorithms for Primary processing of Medical Information", "Classification of Medical Information", adaptive genetic algorithm based on random search ", " Classification based on informational character ", " General information about software ".

**I. “Preliminary Data Readings”.** The mechanism of this module reads the initial data prepared in the Exel program, which is displayed in the tutorial program window, indicating the need for the next step.

The activities of the software are one of the most important first steps in forming the learning sample based on primary data. Therefore, it is necessary to clarify the issue.

**Considering the issue.** Let's assume that the learning sample formed on the basis of primary data is divided into the elective classes and given as follows:

$$K_1 = \begin{bmatrix} x_{11}^1 & x_{11}^2 & \dots & x_{11}^N \\ x_{12}^1 & x_{12}^2 & \dots & x_{12}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{1m_1}^1 & x_{1m_1}^2 & \dots & x_{1m_1}^N \end{bmatrix} \dots K_r = \begin{bmatrix} x_{r1}^1 & x_{r1}^2 & \dots & x_{r1}^N \\ x_{r2}^1 & x_{r2}^2 & \dots & x_{r2}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{rm_r}^1 & x_{rm_r}^2 & \dots & x_{rm_r}^N \end{bmatrix}.$$

This can be summarized as follows:

$$K_p = \begin{bmatrix} x_{p1}^1 & x_{p1}^2 & \dots & x_{p1}^N \\ x_{p2}^1 & x_{p2}^2 & \dots & x_{p2}^N \\ \vdots & \vdots & \vdots & \vdots \\ x_{pm_p}^1 & x_{pm_p}^2 & \dots & x_{pm_p}^N \end{bmatrix}$$

Here  $p = \overline{1, r}$ ; as well the learning sample expressed as  $K = \bigcup_{p=1}^r K_p$ , they should be composed of non-intersecting classes, that is  $K_p \cap K_q = \emptyset, (p \neq q, p = \overline{1, r}; q = \overline{1, r};)$  terms are given.

Likewise, the components of object  $x_{pi}$  – consist of real numbers  $x_{pi}^j$ , and it is read as follows: p – related to class i – patient j - symbol. Here  $p = \overline{1, r}; i = \overline{1, m_p}; j = \overline{1, N}$ ; as well as, r is the total number of classes given,  $m_p$  - p – total number of patients in the class and N – denotes the total number of characters.

In the issues we are looking at, each class is treated as one type of disease, that is,  $K_1$ -class is “Severe angina” (number of objects - 140),  $K_2$ -class is “Acute myocardial infarction” (number of objects is 120), Class  $K^3$  "Arithmetic form" (number of objects - 40),  $K^4$  class is "Cardiosclerosis after infarction" (number of objects - 35). At the same time, the character space that characterizes each class (type of disease) is formed by experts in the field, consisting of 62 signs that characterize each class.

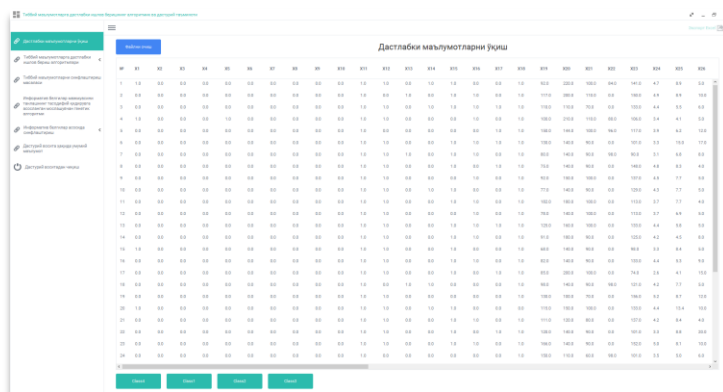


Figure 1. – View of the "Initial Data Reading" Module window

**II. "Algorithms of Medical Data Initial Processing"** is implemented in two steps: 1) determining the level of suitability of the data; 2) Converting the value of the characteristic of objects to 0 or 1.

**Stage 1:** In the data validation phase, the characterization criteria for each  $K_p$  class are evaluated in the following order:

a). Let's make the following notes :  $\bar{x}_p = (\bar{x}_p^1, \bar{x}_p^2, \dots, \bar{x}_p^N)$  vector,  $X_p$  middle class objects,  $p = \overline{1, r}$ . Calculate its components by the following formula:

$$\bar{x}_p^j = \frac{1}{m_p} \sum_{i=1}^{m_p} x_{pi}^j, p = \overline{1,3}; j = \overline{1,62}; i = \overline{1, m_p}. \quad (1)$$

b) The distance between  $x_{pi}$  and  $\bar{x}_p$  objects of  $X_p$  class is calculated by the following formula:

$$|x_{pi} - \bar{x}_p| = \sqrt{\sum_{j=1}^N (\bar{x}_p^j - x_{pi}^j)^2}, p = \overline{1,4}; j = \overline{1,62}; i = \overline{1, m_p}.$$

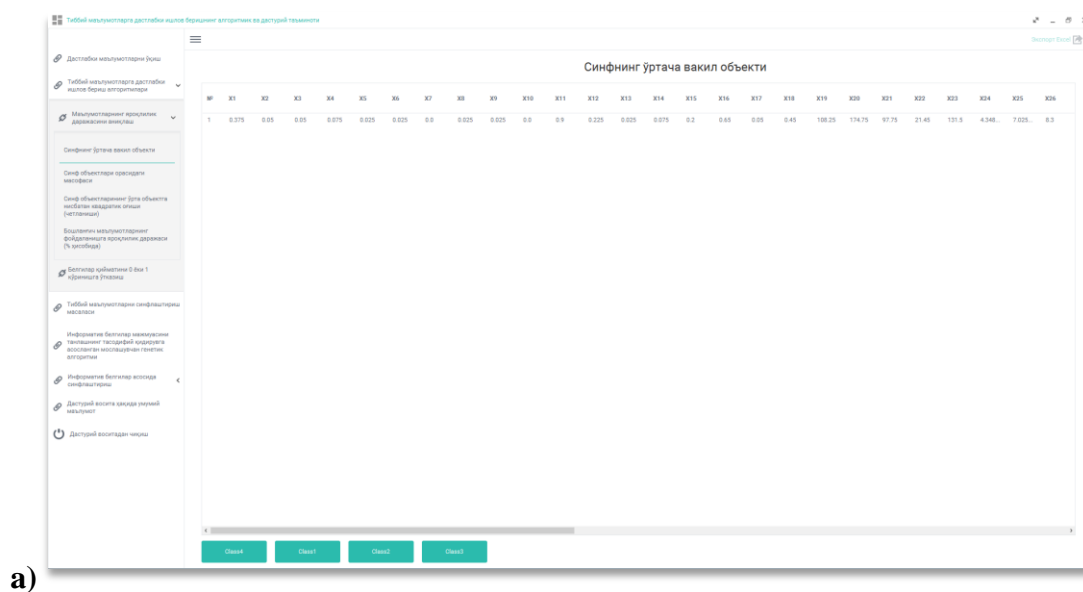
c). Quadratic deviation (deviation) of the  $X_p$ -class objects relative to the middle object  $D(\bar{x}_p)$  is calculated by the following formula:

$$D(\bar{x}_p) = \sqrt{\frac{1}{m_p} \sum_{j=1}^{m_p} |x_{pi} - \bar{x}_p|^2} = \sqrt{\frac{1}{m_p} \sum_{i=1}^{m_p} \sum_{j=1}^N (\bar{x}_p^j - x_{pi}^j)^2}, p = \overline{1,4}; j = \overline{1,62}; i = \overline{1, m_p}.$$

d). the following inequality is met and its performance is calculated as a percentage of class objects:

$$|x_{pi} - \bar{x}_p| \leq D(\bar{x}_p), p = \overline{1, r}; i = \overline{1, m_p}.$$

At the end of the process, the usability of each class is calculated as a percentage.



Шестая Международная научно-практическая конференция «BIG DATA and Advanced Analytics. BIG DATA и анализ высокого уровня», Минск, Республика Беларусь, 20-21 мая 2020 года

Синф объеклари орасидаги масофаси

№	Масофа	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25
1	437,895	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	495,0	27,9425	60,0425	5479	690,25	0,724	7,143
2	162,942	0,396	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,2025	2038	27,9425	0,0425	600,1	690,25	0,964	6,273
3	76,932	0,396	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	217,0	1207	60,0425	600,1	30,25	0,002	1,050
4	152,010	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,2025	340,0	2002	379,0	600,1	2,23	0,022	7,900
5	152,164	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,2025	2038	607,0	1,0425	520,0	6,25	0,002	3,900
6	148,242	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	68,0625	27,9425	0,0425	600,1	182,25	0,200	6,600
7	148,342	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	68,0625	27,9425	0,0425	600,1	342,25	0,400	6,225
8	159,630	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,302	280,0	4207	499,0	600,1	182,25	0,200	6,600
9	162,910	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	2078	27,9425	130,0	600,1	30,25	0,040	9,450
10	152,809	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,302	6201	22,9425	60,0425	620,0	110,25	0,121	0,140
11	127,824	0,396	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	284,0	627,0	1,0425	600,1	210,25	0,201	9,100
12	135,408	0,396	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	68,0625	1207	60,0425	607,0	182,25	0,200	3,300
13	151,078	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,302	1740	217,0	60,0425	600,1	602,25	0,700	0,000
14	170,763	0,396	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	280,0	4207	499,0	600,1	110,25	0,121	2,170
15	170,923	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	1207	60,0425	1,0425	600,1	30,25	0,040	11,28
16	307,173	0,396	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	1400	27,9425	0,0425	600,1	58,25	0,040	0,600
17	222,828	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,2025	130,0	627,0	1,0425	600,1	42,25	0,010	1,130
18	144,288	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,302	68,0625	1207	60,0425	607,0	182,25	0,200	4,100
19	239,242	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	1400	2047	130,0	600,1	25,25	0,020	0,600
20	189,330	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,302	1170	27,9425	0,0425	600,1	272,25	0,300	13,90
21	360,778	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	68,0625	1207	60,0425	607,0	182,25	0,200	6,615
22	187,988	0,396	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,422	0,002	0,2025	788,0	627,0	1,0425	600,1	2,25	0,020	7,700
23	89,3784	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	1940	217,0	60,0425	600,1	240,25	0,300	4,100
24	117,336	0,145	0,002	0,002	0,005	6,230	6,230	0,0	6,230	6,230	0,0	0,009	0,000	6,230	0,005	0,040	0,122	0,002	0,302	68,0625	27,9425	0,0425	600,1	1506,25	1,100	0,300

b)

Синф объекларнинг ўрта объектга нисбатан квадратик огши (четланлици)

Class4:	180,98
Class1:	159,63
Class2:	177,47
Class3:	184,48

c)

Бошланғич маълумотларнинг фойдаланишга яроқлилик даражаси (% ҳисобида)

Class4:	68,57%
Class1:	82,14%
Class2:	71,67%
Class3:	77,50%

d)

Figure 2. – View the Data Validation Dialog Box(a,b,c,d)

**Stage 2:** The initial data is presented in a continuous quantitative manner, during which the process of converting the character values of the class objects to values of 0 and 1 vector.

The process of converting the values of zero or one-character symbols to each of the above-mentioned  $K_p$  class objects in vector form is made by typing the following symbols in each class and all character sections:

a).  $\bar{x}_p = (\bar{x}_p^1, \bar{x}_p^2, \dots, \bar{x}_p^N)$  vector,  $K_p$  class middle objects,  $p = \overline{1, r}$ . Calculate its components by the following formula:

$$\bar{x}_p^j = \frac{1}{m_p} \sum_{i=1}^{m_p} x_{pi}^j, p = \overline{1, r}, j = \overline{1, N}, i = \overline{1, m_p};$$

b). Vectors  $a_p = (a_p^1, a_p^2, \dots, a_p^N)$  ва  $b_p = (b_p^1, b_p^2, \dots, b_p^N)$ , Let us define the following and calculate its components by this formula:

$$a_p^j = \frac{1}{m_p} \sum_{i=1}^{m_p} (\bar{x}_p^j - x_{pi}^j)^2, p = \overline{1, r}, j = \overline{1, N}.$$

$$b_{pi}^j = (\bar{x}_p^j - x_{pi}^j)^2, p = \overline{1, r}, j = \overline{1, N}.$$

c). The components of the  $K_p$  elements of the learning sample are converted from the actual numeric form to the view using the following procedure.

$$x_{pi}^j = \begin{cases} 1 \text{ equal, if } \frac{b_{pi}^j}{a_p^j} \leq 1, \\ \text{equal to 0, otherwise.} \end{cases}$$

At the end of this phase, the character values of the 4 class objects given are converted to 0 and 1 vector values.

Figure 3. – View stage conversion of character values to 0 or 1

III. Module "Classification of Medical Information" solves the problem of classification of  $K_p$  class objects, that is, whether each object in a class belongs to a class or a different one. At the



same time, each object belonging to the class  $X_p$  can be compared with objects in its class and other classes by introducing  $\rho_i(x_{p1}, x_{p2})$  in the informative character space:

$$\rho_i(x_{p1}, x_{p2}) = \begin{cases} 1 & \text{агар} & (x_{p1}^i - x_{p2}^i) = 0, i = \overline{1, N}, \\ 0 & & \text{otherwise.} \end{cases}$$

The first condition denotes the degree of similarity between the two objects, and the second condition indicates that they are different. The total of comparative evaluation is based on the following formula:

$$\Gamma_j(x_{pj}, x_{pk}) = \sum_{k=1}^{m_p} \sum_{i=1}^N \rho_i(x_{pj}, x_{pk}), j = \overline{1, m_p}; k = \overline{1, m_p}; j \neq k.$$

Comparative evaluation is calculated for each class, and the largest of the mean values obtained is the attribution of the object to that class.

In this section, there is a process of reclassifying classification objects to determine if a class object belongs to its own class or another class. This process is done step by step, with each step excluding objects that are not related to its class, and objects in the classes are complete until they reach their full class, ie 100%.

Figure 4. – Module "Problems of Classification of Medical Information"

IV. An informative character set is selected using the reference table formulated in the module "**Random search based adaptive genetic algorithm for selecting a set of informative characters**". Let us suppose that the resultant error in classification of the sample objects based on a reference table is  $\theta$ . That is, the proportion of the number of objects missing from class  $\eta$  to the total number of electives is set to  $\theta$ .

Let the mechanism of algorithm "A" be defined as follows. Using the convergence function proposed by the authors [1, 12], the degree of importance of each sign is determined. According to it, the column will be omitted from the table with a randomly selected icon. If at the end of the process all objects in the computation find their class at least  $\theta$ , the column removed from the table will not be redirected to the table otherwise the column will be returned to its original location by random selection and the process will be returned. The proposed process lasts up to  $\ell$ . If objects have found a different class in their class (switch to another class), the randomly selected symbol will be returned. This process takes place between N characters and is separated from the remaining symbols as informative characters at the end of the process.

Hence, the essence of the work performed at this stage is to select the most useful  $\ell$  from the set of symbols that characterize the objects being investigated, that is, the choice of an informative character tool.

The followings should be given:  $\ell \ll N$  characters to be selected are the number of characters in the tool ;  $p = (p^1, p^2, \dots, p^N)$  probability that the vector components must remove one character from the character tool. Initially,  $p^j = \frac{1}{N}, j = \overline{1, N}$  is taken, that is, the probability of removing an arbitrary character from an array is equal.

**1-step.** Algorithm "A" is used in the cross section of all marks of the given training sample. As a result, the error factor  $\theta(N)$  coefficient is calculated. Typically, to determine the error rate  $\theta(N)$ , the number of objects for which the error is detected is divided by the total number of objects.

**2-step.** Accidentally, with  $p^j = \frac{1}{N}, j = \overline{1, N}$  probability, one character is selected from  $\square$  characters, and it is omitted from the array. Then the algorithm "A" is run at the intersection of symbols – 1 characters. As a result, the error coefficient  $\square (\square - 1)$  for symbols – 1 is calculated.

**3-step.** If  $\theta(N) > \theta(N - 1)$ , in this case 2-step begins. This process is repeated  $h (h \leq \ell)$  time. If the process is repeated, even if  $h = \ell$  equality is met, the next step is to proceed. Likewise, if equality is not met, the process will continue until all the characters are considered one by one. After all the characters are reviewed, go to the next step.

**4-step.** Let us assume  $h = \ell$ , in this case with  $p^j = \frac{1}{2N-h}, j = \overline{1, h}; p^j = \frac{2}{2N-h}, j = \overline{h+1, N}$ ; probability out of  $N$  symbols one symbol is chosen and it is excluded from the set. As in the second step, the process is reversed, except that the probability that the previously selected and non-existent characters that are to be randomly selected differ from each other. Then the algorithm "A" is run at the intersection of  $N-1$  characters. As a result, the error coefficient  $\theta(N-1)$  for  $N-1$  is calculated.

**5-step.** If  $\theta(N) > \theta(N - 1)$ , then 4<sup>th</sup> step begins. This process  $h (h \leq \ell)$  is repeated one time and so on.

The proposed algorithm consists of two important parts: In the first part, genetics is a tool that identifies the generating characteristics of mutated objects, and the second part is a random search algorithm for selecting a new set.

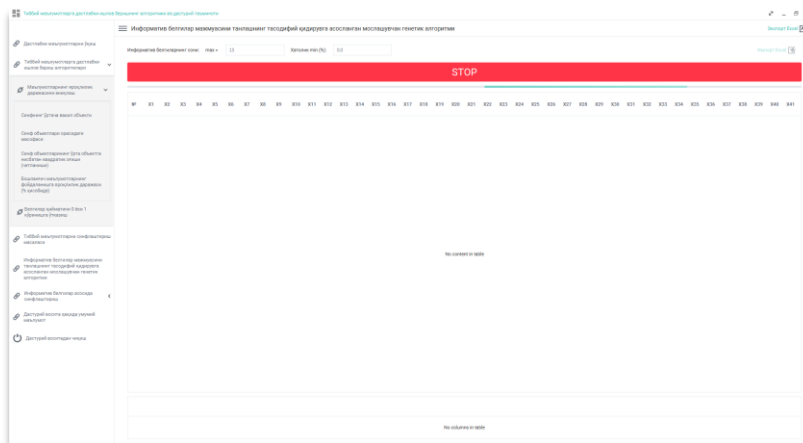


Figure 5. – Module window of "Random genetic search algorithm for selecting a set of informational symbols"

**V. "Classification based on informative symbols"** in the module, the classification process is carried out with the use of informative symbols, that is, in the matrix view, the columns containing informatively found symbols are retained and the remaining columns are omitted. The function of

inter-object proximity in the space of the stored informative symbols is done by entering  $r_i(x_{p1}, x_{p2})$ :

$$\rho_i(x_{p1}, x_{p2}) = \begin{cases} 1 & \text{if } (x_{p1}^i - x_{p2}^i) = 0, i = \overline{1, N}, \\ 0 & \text{otherwise.} \end{cases}$$

The first condition denotes the degree of similarity between the two objects, and the second condition indicates that they are different. The total of comparative evaluation is based on the following formula:

$$\Gamma_j(x_{pj}, x_{pk}) = \sum_{k=1}^{m_p} \sum_{i=1}^N \rho_i(x_{pj}, x_{pk}), j = \overline{1, m_p}; k = \overline{1, m_p}; j \neq k.$$

Comparative evaluation is calculated for each class, and the largest of the mean values obtained is the attribution of the object to that class.

№	Class1	Class2	Class3	Class4	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	0.941176470588236	6.91091818181818	5.68	6.075	0	0	0	1	1	1	1	1	0	1	
2	0.2088232941176	4.0962942962962	4.104	3.925	0	0	0	0	1	1	0	0	1	0	
3	0.32029411764707	6.02222222222222	4.696	5.175	0	0	0	1	0	1	1	1	1	0	
4	0.67647058823293	0.8	5.272	5.725	0	0	0	1	1	1	0	1	1	0	
5	0.32029411764707	6.02222222222222	4.696	5.175	0	0	0	1	0	1	1	1	1	0	
6	0.09882329411764	4.9027037027027	4.288	4.625	0	0	0	1	0	1	0	1	1	0	
7	7.411764705882363	8.21481481481481	3.904	2.975	0	0	0	0	1	1	0	0	1	0	
8	0.32029411764707	6.02222222222222	4.696	5.175	0	0	0	1	0	1	1	1	1	0	
9	0.97088232941176	5.54814814814814	4.696	5.275	0	0	0	1	1	1	1	0	0	1	
10	0.941176470588236	6.91091818181818	5.68	6.075	0	0	0	1	1	1	1	1	1	0	
11	0.941176470588236	6.91091818181818	5.68	6.075	0	0	0	1	1	1	1	1	1	0	
12	0.941176470588236	6.91091818181818	5.68	6.075	0	0	0	1	1	1	1	1	1	0	
13	0.32029411764707	6.02222222222222	4.696	5.175	0	0	0	1	0	1	1	1	1	0	
14	0.941176470588236	6.91091818181818	5.68	6.075	0	0	0	1	1	1	1	1	1	0	
15	0.67647058823293	0.8	5.272	5.725	0	0	0	1	1	1	0	1	1	0	
16	0.941176470588236	6.91091818181818	5.68	6.075	0	0	0	1	1	1	1	1	1	0	
17	0.09882329411764	4.9027037027027	4.288	4.625	0	0	0	1	0	1	0	1	1	0	
18	0.67647058823293	0.82222222222222	4.112	3.325	0	0	0	1	0	1	0	1	0	1	
19	0.2088232941176	4.17027027027027	4.888	3.925	0	0	0	1	1	1	0	0	1	0	
20	7.5	4.30000000000000	3.98	4.175	0	0	0	1	1	0	1	0	1	0	
21	0.941176470588236	6.91091818181818	5.68	6.075	0	0	0	1	1	1	1	1	1	0	
22	0.30292941176471	4.10000000000000	4.104	3.975	0	0	0	1	0	1	0	0	1	0	
23	0.67647058823293	0.8	5.272	5.725	0	0	0	1	1	1	0	1	1	0	
24	0.941176470588236	6.91091818181818	5.68	6.075	0	0	0	1	1	1	1	1	1	0	
25	0.67647058823293	0.8	5.272	5.725	0	0	0	1	1	1	0	1	1	0	
26	0.67647058823293	0.8	5.272	5.725	0	0	0	1	1	1	0	1	1	0	

Figure 6. – "Classification based on informational symbols" window view

VI. "General information about the software". This module gives an overview of the software.

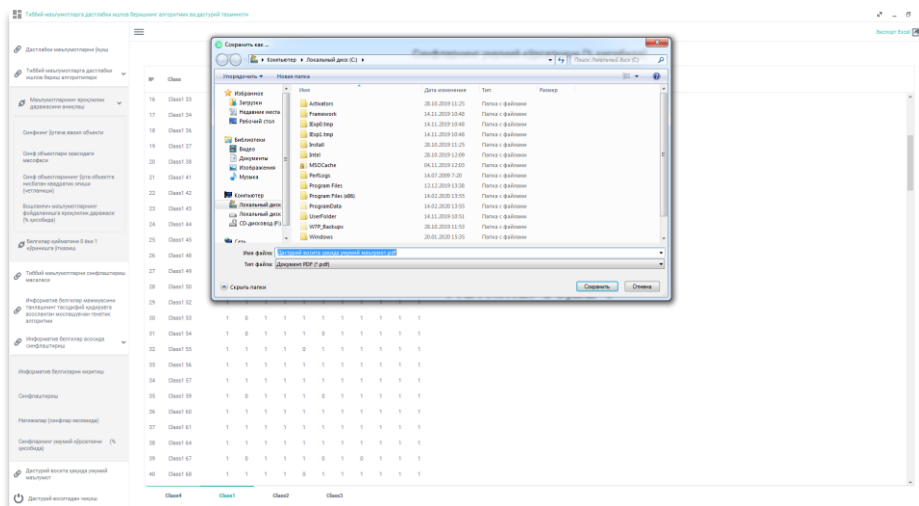


Figure 7. – «General information about the software» window view

### References

- [1]. Zhuravlev Yu.I. Selected scientific works. – M: ‘Magistr’ Publishing house, 1998. - 420 p.
- [2]. Yu, L., H. Liu. Efficient Feature Selection via Analysis of Relevance and Redundancy. – J. Mach. Learn. Res., Vol. 5, 2004, No Oct, pp. 1205-1224.
- [3]. Yu, L.: Toward Integrating Feature Selection Algorithm for Classification and Clustering. IEEE Transaction on Knowledge and Data Engineering 17(4), 491-502
- [4]. Yan, K., D. Zhang. Feature Selection and Analysis on Correlated Gas Sensor Data with Recursive Feature Elimination. – Sensors Actuators, B Chem., Vol. 212, Jun 2015, pp. 353-363.
- [5]. Nishanov A.Kh., Turakulov Kh.A., Turakhanov Kh.V. A decision rule for identification of eye pathologies // Biomedical Engineering (1999) 33(4) 178-179.
- [6]. Nishanov A. Kh., Turakulov Kh.A., Turakhanov Kh.V. A decisive rule in classifying diseases of the visual system // Meditsinskaia tekhnika (1999) (4) 16-18.
- [7]. Xiang Fang, Lina Wang Feature Selection Based on Fisher Criterion and Sequential Forward Selection for Intrusion Detection // Revista de la Facultad de Ingeniería U.C.V., Vol. 32, N° 1, pp.498-503, 2017.
- [8]. Linhui Sun, Sheng Fu and Fu Wang. Decision tree SVM model with Fisher feature selection for speech emotion recognition // Sun et al. EURASIP Journal on Audio, Speech, and Music Processing (2019) 2019:2 <https://doi.org/10.1186/s13636-018-0145-5>
- [9]. Yan, X., Tan, M., Yan, Y., Lü, M. (2012). “Research on Hidden Markov Model-Based And Neural Network-Based Intrusion Detections”. Computer Applications and Software. 29(2), 294–297.
- [10]. JING, X., WANG, H., NIE, K., LUO, Z. (2012). “Feature Selection Algorithm Based on IMGa and MKSVM to Intrusion Detection”. Computer Science, 39(7), 96–100.
- [11]. Koushal Kumar, Jaspreet Singh Bath. Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms // International Journal of Computer Applications (0975 – 8887) Volume 150 – No.12, September 2016, 1-13.
- [12]. Nishanov A.Kh., Djurayev G.P., Kasanova M.Kh. Improved algorithms for calculating evaluations in processing medical data // National Institute of Science Communication and Information Resources (NISCAIR)-India, 2019,-3158-3165.
- [13]. Kamilov M., Nishanov A., Beglerbekov R. Modified stages of algorithms for computing estimates in the space of informative features // International Journal of Innovative Technology and Exploring Engineering (2019) 8(6).

UDK 718.66.32

## METHOD FOR PREPROCESSING A DISTORTED IMAGE BASED ON THE DECONVOLUTION PROCEDURE



**S.S. Beknazarova**

*Professor of Tashkent university of information technologies named after Muhammad al-Khwarizmi, Doctor of technical science, associate of professor*



**M.K. Jaumitbayeva**

*Master of Tashkent university of information technologies named after Muhammad al-Khwarizmi*

*Department of Audiovisual technologies, Tashkent University of information technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan*

E-mail: [saida.beknazarova@gmail.com](mailto:saida.beknazarova@gmail.com)

**S.S. Beknazarova**

*Author of 189 research papers on audio, video, multimedia resources, and media education technologies.*

**M.K. Jaumitbayeva**

*Master of the Department of Audiovisual technologies, faculty of Television technologies. Conducts research in the field of video technologies.*

**Abstract.** In the article conducted the review of solution to the deconvolution problem is to reverse a two-dimensional convolution. The term "deconvolution" covers the most important and widely used image processing methods. The need for such an operation arises in all areas of science related to measurements. There are a large number of papers on deconvolution methods. The deconvolution problem can be solved in several ways. The choice of the most suitable method for solving this problem depends on a number of factors, including the shape and extent of the point propagation function (FRT), the nature of the original image, and the degree of truncation by the frame window of the recording device.

**Keywords:** method for preprocessing, a distorted image, based on the deconvolution procedure.

**Introduction.** Whatever method is used, it is almost always necessary to pre-process the specified distorted image to transform it into a form that is convenient for performing the deconvolution procedure. Preprocessing can be divided into five categories: smoothing, partitioning, Apodization (weighing the signal segment to be processed with a weight function), boundary expansion, and super-resolution. Image smoothing here refers to noise reduction. Partitioning involves splitting an image with a spatially dependent FRT into fragments, in each of which the FRT can be approximated as spatially invariant. Apodization is a method to reduce the impact of a frame window (recording device) that truncates an image. However, this method may be less effective than the border extension method, which we have successfully applied in a number of cases. There are two modifications of the method of expansion of borders is a simple extension of the with overlapping. The second method is generally preferred because it takes advantage of the consistency condition of periodic convolutions. This is another example of how the efficiency of the numerical method increases when the features of the problem under study are more fully taken into account in terms of

mathematical physics. Over-resolution is considered a pre-processing procedure because it ultimately reduces noise.

There is also multiplicative deconvolution, which is the most common method for restoring an image that is represented as a consistent convolution. A distorted image that is not one should be converted to a consistent convolution view.

The method of subtractive deconvolution is particularly useful when the defects present in the recorded image are not associated with loss of resolution, but with distortion of small details, for example, when the FRT has the same narrow main lobe as the resolved part, but has a wide tail of significant amplitude or is characterized by high side lobes. The subtractive deconvolution method can be easily modified to include space-dependent distortions, although the computational implementation of these methods then becomes very complex.

There are different approaches to the deconvolution problem. These approaches include non-recursive and recursive filtering methods in the image plane, direct matrix methods, and maximum entropy and maximum likelihood methods.

The coordinated deconvolution method, which originated from studies of complex zeros in the frequency plane, is basically a one-dimensional method that can be as widely applied as the multiplicative deconvolution method.

*Materials and methods.* One of the most important practical methods of deconvolution is the method of blind deconvolution. Note that all methods of processing speckle images can be considered as special cases of blind deconvolution.

In addition to the well-known traditional applications of deconvolution, there are also various exotic applications of deconvolution. One of the most remarkable applications is the restoration of recordings of famous singers voices on old gramophone records by blind deconvolution.

The convolution integral is represented by the expression

$$b(x) = f(x) * h(x) \quad (1)$$

where  $h(x)$  is the function that sets the distortion;  $f(x)$  is the function that needs to be restored.

According to the Fourier convolution theorem, the image of a quantity (1) is equal to

$$B(u) = F(u)H(u) \quad (2)$$

where  $F(u)$  is the function associated with the function  $f(x)$  by a two - dimensional Fourier transform;  $H(u)$  is the Fourier image of the optical transfer function.

The idealized problem of finite deconvolution is as follows: the functions  $b(x)$  and  $h(x)$  are given, and the function  $f(x)$  must be restored, provided that all three quantities have a finite extension.

It follows from the relation (2) that this problem can be solved as follows

$$b(x) = f(x) * h(x) \quad (3)$$

The division operation inside curly brackets in expression (3) is called simple inverse filtering. The term "filtering" is used here by analogy with the classical theory of circuits and the modern theory of signal processing. A classic filter is a device that changes the spectrum of time frequencies of a signal. The spectrum  $B(u)$  is a function of the spatial frequency.

The optical transfer function  $H(u)$  changes the spatial frequency spectrum  $B(u)$  as a result of applying the above division operation.

Since processed images are usually stored in computer memory as quantized values, image processing techniques typically use digital rather than classical analog filters. A digital filter is defined by a discrete array, generally speaking, of complex numbers, which changes the spectrum of spatial

frequencies during some processing operation. Therefore, both functions,  $h(x)$  in formula (1) and  $H(u)$  in formula (2), can be considered as filters (and in most applications they are implemented digitally). The generally accepted classification of digital filters originated in the theory of signal processing as functions of time, and this classification can be used in the theory of processing one-dimensional images, i.e. signals as functions of a (single) spatial variable. We will transfer the corresponding terminology to the two-dimensional case. The concept of "reference" in the theory of signal processing passes into the concept of "image element" in the theory of image processing. Both samples and image elements must be quantized in amplitude before they are digitally processed. The image to which the filtering operation is to be applied is called the specified image, and it is referred to as consisting of the specified image elements. Elements of a filtered image are called output image elements. In the case of a non-recursive digital filter, each output image element is a weighted sum of the specified image elements. In the case of a recursive digital filter, each image output element is a weighted sum of the specified image elements and the previously calculated image output elements. All practically implemented digital filters, of course, are described by arrays of finite sizes (in the one-dimensional case, the final filter is often called a short one). A digital filter is called a direct filter if it is applied in the image plane, and a spectral filter if it is applied in the frequency plane. The causal filter is one-way in the sense that its response always lags behind the input action (this is somewhat artificial in the two-dimensional case, but of course is very important for one-dimensional filtering operations that are the basis of signal processing as functions of time). Causal filters are almost always implemented as direct filters. A multiplicative digital filter is a spectral filter in which each output sample is obtained as the product of a given input signal element by one element of the filter array.

If all the essential aspects of the practical problems of deconvolution were reduced to the formula (3), then the entire content of this publication could easily be contained in a small article. However, there are many practical difficulties in the deconvolution problem. This is because the data being processed is always distorted in practice.

Before setting the practical problem of deconvolution, we examine some properties of convolution consistency.

In the one-dimensional case, the ratio (2) is represented as

$$B(w) = F(w)H(w) \tag{4}$$

where the real variable  $u$  is replaced by the complex variable  $B(w)$ . If the functions  $f(x)$  and  $h(x)$  have a finite extension, so that the extension of the function  $b(x)$  is also finite, then their spectra are characterized by sets of zeros in the complex  $w$ -plane. If a given set  $z_g$  is represented as a set of real zeros  $z_{gr}$  and zeros that can be complex  $z_{gc}$ , then we can write

$$z_b = z_f \cup z_k \tag{5}$$

This means that the one-dimensional deconvolution problem is consistent only if all zeros of the function  $H(w)$  are also zeros of the function  $B(w)$ . Therefore, the values  $b(x)$  and  $h(x)$  cannot be set independently; it must be known beforehand that they satisfy the relation (1). The same applies to two-dimensional convolutions.

*Results.* Now let's go back to the periodically extended (overlapping) ideal distorted  $imb(x)$  image and its  $IMB(x)$  spectrum. The latter can be written as

$$M_b(x) = \sum_l^\infty F_l H_{l,m} \delta(L_1 u - l) \delta(L_2 \gamma - m) \tag{6}$$

where  $\delta(*)$  is the Delta function;  $F_i, m$  are the Fourier coefficients of the true image  $f(x)$ , which are also counts of the function  $F(u)$  which are considered in the counting theorem. These samples are

taken at raster points  $(l/L_1, m/L_2)$  in the frequency plane. The values  $H_{l,m}$  included in expression (6) are samples of the optical transfer function  $H(u)$ , at the same points in the raster:

$$H_{l,m} = H\left(\frac{l}{L_1}, \frac{m}{L_2}\right) \quad (7)$$

where  $l$  and  $m$  are arbitrary integers.

Now we can set an idealized problem of periodic deconvolution: given the functions  $imb(x)$ , and  $h(x)$ , we need to find the function  $f(x)$ , [knowing that  $f(x)$ , and  $h(x)$ , are functions of finite extension, and  $imb(x)$  is a periodic function].

For a given function  $b(x)$ , you can calculate the function  $B(u)$  and immediately find that

$$H_{p,l,m} = B\left(\frac{l}{L_1}, \frac{m}{L_2}\right) \quad (8)$$

Similarly, the counts of the optical transfer function  $H_{l,m}$  are calculated. The expression (6) shows that each value of  $F_{l,m}$  is given by a division  $\frac{B_{p,l,m}}{H_{p,m}}$  operation, which can always be performed if the values of  $H_{l,m}$  are different from zero. This simple approach is adequate in the case of functions  $b(x)$  and  $f(x)$ , which are selected independently, since the function  $imb(u)$  in accordance with expression (6) actually exists only at the above points in the raster. But this approach is not acceptable in an idealized problem in the case of finite convolution, since then  $B(u)$ , is a continuous function of the variable  $u$ .

It is therefore surprising that the only consistency condition for periodic convolutions is the requirement that the values  $H_{l,m}$  can be zero only for those values  $l$  and  $m$  for which  $H_{l,m}=0$ . This condition is called the consistency condition for periodic convolutions. We emphasize that there is no value.  $H_{l,m}$  cannot be exactly zero in the real measurement of the function  $h(u)$ , or, equivalently, the function  $H(u)$ , so that periodic convolutions are always consistent in practice (they are, of course, very noisy when a large number of values  $H_{l,m}$  are "small" at values  $l$  and  $m$  that correspond to values significantly different from zero  $B_{p,l,m}$ ).

The practical problem of deconvolution is set as follows: the functions  $b(x)$ , and  $h(x)$  are given, you need to find the function  $f(x)$ , knowing that it is a truncated version of the function of the recorded image  $r(x)$ .

One of the "Golden rules" in image reconstruction is to avoid processing data that contains any discontinuities, of which clipping and truncation are the most undesirable, since they almost always produce false details (often called artifacts, especially in medical applications). Thus, as a rule, it is desirable to pre-process the image in order to fully compensate for all existing gaps and other removable defects.

*Conclusion.* Any type of preprocessing can, of course, contribute noise in addition to the image distortion  $f(x)$  already present in the recorded image  $r(x)$ . But if the gaps are not fixed, then the corresponding artifacts usually prevail over any additional noise introduced by preprocessing. The "aligned" shape of the image is denoted here by  $a(x)$ . and will be called a pre-processed recorded image. Although all three values must change as a result of preprocessing, there is rarely any way to estimate how much, and therefore it usually makes no sense to talk about the difference between images  $a(x)$  and  $r(x)$ . Next, we will treat these two images as identical, at least in the frame  $a$  (i.e., in the area of the image plane) where the pre-processed version of the meaning image fits  $a(x)$ . Therefore, we assume that

$$a(x) = f(x) * h(x) + c(x) \quad (9)$$



This assumption does not affect the generality of reasoning, since noise  $c(x)$  includes the effects of arbitrary additive distortion associated with preprocessing.

Now we will introduce the concept of "recoverable true image  $\hat{f}(x)$ ". This is an estimate of the image  $f(x)$  that can be obtained from the image  $h(x)$ .

In any rational approach to solving a practical deconvolution problem, the pre-processed image  $a(x)$  is first obtained from the specified image  $a(x)$ . Then, a suitable deconvolution procedure is selected to obtain  $\hat{f}(x)$  based on  $h(x)$  and  $a(x)$ . Some of these procedures can be seen as the process of obtaining  $\hat{h}(x)$  a modified point propagation function that is associated with a pre-processed recorded image and a recoverable true image ratio

$$a(x) = \hat{f}(x) * h(x) + c(x) \quad (10)$$

It is convenient to denote the Fourier coefficients  $\hat{f}(x)$  of a function by, and to denote the spectra  $\widehat{F}_{l,m}$  of functions  $a(x)$ ,  $c(x)$ , and use  $\hat{f}(x)$   $\hat{h}(x)$  the corresponding capital letters with or without a "hat".

If there is a concern that the differences between  $\hat{f}(x)$  and  $f(x)$  will greatly increase due to the lack of consistency between the functions  $a(x)$  and  $h(x)$ , taken explicitly finite, then you can refer to the formula

$$im_b(x) = \sum_{l,m \rightarrow \infty}^{\infty} b(x - lL_1, y - mL_2)$$

for a periodic  $imb(x)$  image with  $b$  replaced by  $a$ . Then the  $imb_u$  spectrum of the convolution is given by expression (6), but with the replacement of the values  $\widehat{F}_{l,m}$  and  $\widehat{H}_{l,m}$  the values  $F_l$ ,  $m$  and  $H_l$ ,  $m$ , respectively. Recall that periodic convolutions are not affected by inconsistency, which, as already mentioned, can distort convolutions of quantities that have finite extensions.

### References

- [1.] Bates R., McDonnell M. image Restoration and reconstruction: TRANS. - Moscow: Mir, 1989. - 336 p.
- [2.] Deacons Century MATLAB. Signal and image processing. Special reference book. - SPb.: Peter, 2002. - 608 p.
- [3.] Fu K., Gonzalez R., Lee K. Robotics: TRANS. / Under the editorship of V. G. Gradetsky. - Moscow: Mir, 1989. - 624 p.
- [4.] J. P. Gambotto et T. S. Huang, "Motion analysis of isolated targets in infrared image sequences", 7th Int. Conf. on Pattern Recognition, Montreal, 1984. P. 534-538.
- [5.] J. P. Gambotto, "Correspondence analysis for target tracking in infrared images", 7th Int. Conf. on Pattern Recognition, Montreal, 1984. P. 526-530.
- [6.] L. Gerardin, "La bionique", Hachette Paris, 1968.
- [7.] Gerardin L. Bionics. Per. FR. M. N. Kovaleva. Ed. and with a Preface. Professor V. I. Gusel'nikova. Moscow: Mir, 1971. 231 PP.
- [8.] J. F. Gilmore et W. B. Pemberton "Asurvey of aircraft classification algorithms", 7th Int. Conf. on Pattern Recognition, Montreal 1984. P. 559-562.
- [9.] R. C. Gonzales et M. G. Thomason, "Syntactic pattern recognition", Edit. Addison Wesley, Londres, 1978. P. 118-120.
- [10.] Perception and pattern recognition / Ed. With FR. A. V. Seredinsky; edited by G. P. Katys. - M.: Mashinostroenie, 1989, 272 p.: Il.
- [11.] N. Sedova, V. Sedov, R. Bazhenov, A. Karavka, S. Beknazarova. Automated Stationary Obstacle Avoidance When Navigating a Marine Craft //2019 International Multi-Conference on Engineering, Computer and Information Sciences, SIBIRCON 2019; Novosibirsk; Russian Federation; 21 October 2019.

## **СПОСОБ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ИСКАЖЕННОГО ИЗОБРАЖЕНИЯ НА ОСНОВЕ ПРОЦЕДУРЫ ДЕКОНВОЛЮЦИИ**

***С. С. Бекназарова***

*в.и.о.профессора Ташкентского университета информационных технологий имени Мухаммеда Аль-Хорезми, доктор технических наук, доцент*

***М.К. Жаумьтбиева***

*Магистр Ташкентского университета информационных технологий имени Мухаммеда Аль-Хорезми*

*Кафедра аудиовизуальных технологий, Ташкентский университет информационных технологий имени Мухаммеда Аль-Хорезми, Ташкент, Узбекистан*

*E-mail: c*

**Аннотация.** В статье проводится обзор решения задачи деконволюции, которое заключается в обратном обращении двумерной свертки. Термин "деконволюция" охватывает наиболее важные и широко используемые методы обработки изображений. Необходимость в такой операции возникает во всех областях науки, связанных с измерениями. Существует большое количество работ, посвященных методам деконволюции. Проблему деконволюции можно решить несколькими способами. Выбор наиболее подходящего метода решения этой задачи зависит от ряда факторов, в том числе от формы и протяженности функции распространения точки (FRT), характера исходного изображения и степени усечения кадровым окном регистрирующего устройства.

**Ключевые слова:** метод предварительной обработки, искаженное изображение, процедура деконволюции.

UDK159.9.016.4

## INTERPRETATION SEISMIC AND GRAVITY DATA OF HAZAR LAKE AND SURROUNDINGS, EASTERN ANATOLIA



**G. Polat**  
PhD, Yeditepe University



**A. Aydın**  
Professor  
Pamukkale University



**F.F. Altinoglu**  
PhD, Pamukkale  
University

Yeditepe University, Istanbul, Turkey  
Pamukkale University, Denizli, Turkey  
Email: aaydin@pau.edu.tr  
Pamukkale University, Denizli, Turkey

### **Gülten Polat**

She is a lecturer at Civil Engineering Department, Yeditepe University. Graduated from Boğaziçi University. Master of science PhD in School of Geological sciences, University College Dublin (UCD). She mainly studies on seismic anisotropy and tomography.

### **Ali Aydın**

He is lecturer in the Department of Applied Geophysics at Pamukkale University, graduated from Karadeniz Technical University and Doctor of Philosophy (Ph.D.) in Geophysics. His research interest include potential field, tectonic, magnetic susceptibility.

### **Fatma Figen Altinoglu**

She is a Research Fellow in the Department of Applied Geophysics at Pamukkale University, Ph.D. in Pamukkale University and focuses on tectonics, seismicity and hazard analysis of a region and geophysical imaging of fault.

**Abstract.** The frequency-magnitude distribution of the earthquakes occurred in the city of Elazığ and surrounding area was analysed with Bouguer gravity anomaly map. This region is located in Eastern Anatolia and tectonically very active and has a high seismic hazard risk. A complete set of 1550 earthquakes of  $M_d \geq 1$  from 1 January 2010 to 28 February 2019 was used. The spatial mapping of frequency-magnitude distribution was produced. Outstanding variations in the b-value were detected with b ranging from  $b \approx 0.7$  to 1.3. The gravity anomaly were showed the lineaments and especially important geologic units. The results show that the accumulated stresses are high at the Pütürge segment in contrast to the fault, drawn in the south, represents the main thrust belt that runs through Hani-Lice and stretches to Çüngüş. The relationships between two data, showing very high relation between them. The last activity on this area, the results of this study will give very clearly information to earthsciences.

**Keywords:** seismic, gravity, data

**Introduction.** The tectonic structure of the Anatolian plate is very complex because of the many plates such as the Eurasian, the Arabian and African plates significantly influenced its tectonic and geodynamic structure (Şengör and Yılmaz 1981). An interplay of this movement and also the geological process of the Mediterranean layer to a lower place the Anatolian plate causes a N-S expansion and E-W shortening in western Turkey. As a consequence of this, the area is underneath expansion in a NNE-SSW direction. The interactions cause a large variety of complex tectonic processes like collision, subduction, back-arc extension, strike-slip faulting and rotation of different blocks and microplates at intervals a comparatively little region (Ketin 1966). Elazığ located in Eastern Anatolia, where has a complex and active tectonic structure (Hempton

1983). The aim of the study is to investigate the seismic activity of the studied region by using a-value and b-value and its relation with Bouguer anomaly map.

*Data and Methods.* To calculate seismic parameters consisting of a-value, b-value and  $M_c$  (Magnitude of Completeness) of the study area, a complete set of 1550 earthquakes of  $M_d \geq 1$  from 1 January 2010 to 28 February 2019 taken from Bogazici University, Kandilli Observatory and Earthquake Research Institute, Regional Earthquake-Tsunami Monitoring Center (KOERI) was used. The spatial mapping of frequency-magnitude distribution was produced for the study region.  $\log_{10}N = a - bM$  is used to define the frequency of occurrence of earthquakes as a function of magnitude where  $N$  denotes the cumulative number of earthquakes with magnitude greater than  $M$  (Gutenberg and Richter, 1944). Bouguer gravity anomaly map taken by joint study of MTA and TP was used to determine the tectonic lineaments of the study area shown in Fig. 1. The region gives negative gravity anomalies range between negative 116 to 56 mGal. Regional anomalies with negative high amplitude dominate in Bouguer gravity anomaly map of Eastern Anatolia region due to the thick crust (Pamukcu et al 2007). High negative gravity anomalies (maximum anomalies) indicates topographic highs.

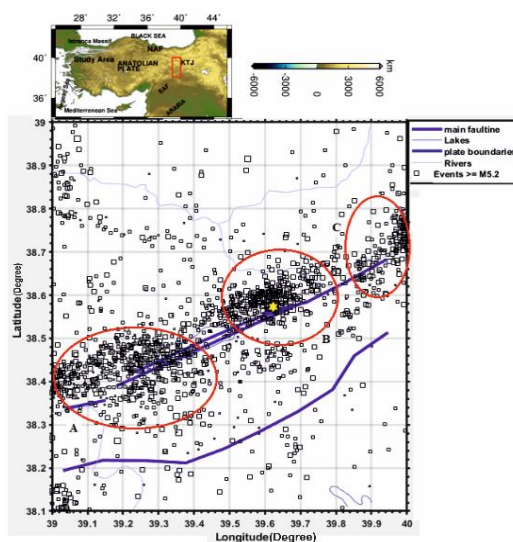


Figure 1. – The location of the study area within Turkey and the epicentres of the earthquakes

*Results.* Seismotectonic b- and a-values of the study area are investigated by the analysing of region-time characteristics of earthquakes in Elazig, Turkey. The b-value is computed as  $1.33 \pm 0.06$  with a completeness level of 2.6, and this result indicates that frequency-magnitude distribution of seismicity in Elazig is well represented with a b-value slightly greater than 1.0. The spatial b-value map for the region (Fig.2, 3) indicated that the b-values of seismicity for the whole region is not homogeneous. The observed variations in b value are probably related to changes in locally and regionally stresses accumulated in the region. The regional variation in b-value may be associated with changes in stress of the active fault systems. An inverse relation between gravity anomaly and b-value is found. Negative high gravity anomalies and intermediate to high b-values are associated with a comparatively thicker crust. On the other hand, high gravity anomaly and low b-values are all consistent with a thinner crust.

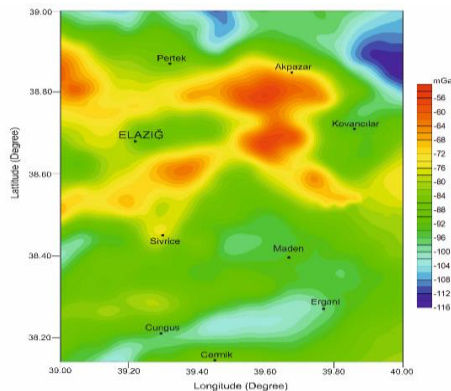


Figure 2. – Bouguer gravity anomaly map of the region.

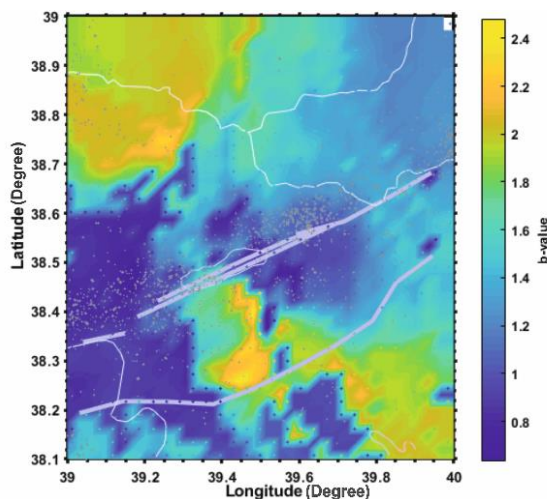


Figure 3. – Regional variations of b-value changes between 1 January 2010 - 28 February 2019

*Conclusion.* Negative high gravity anomalies and intermediate to high b-values are associated with a comparatively thicker crust. On the other hand, high gravity anomaly and low b-values are all consistent with a thinner crust.

### References

- [1]. Çetin H., Güneyli H., Mayer L. Paleoseismology of the Palu-Lake Hazar segment of the East Anatolian Fault Zone, Turkey: *Tectonophysics*, 374, 2003. – Pp. 163-197.
- [2]. Gutenberg R., Richter C.F. Earthquake magnitude, intensity, energy and acceleration. *Bull Seismol Soc Am* 32, 1944. – Pp. 163-191.
- [3]. Hempton M.R., Dunne L.A., Dewey J.F. Sedimentation in an active strike- slip basin, south-eastern Turkey. *J. Geol.*, 91, 1983. – Pp. 401-412.
- [4]. Ketin I. Tectonic units of Anatolia. *Maden Tetkik ve Arama Bulletin*, 1966. – Pp. 23-34.
- [5]. Pamukçu O.A., Akçığ Z., Demirbas S., Zor E. Investigation of crustal thickness in Eastern Anotolia using Gravity, Magnetic and Topographic data. *Pure and Applied Geophysics*, 164 (11), 2007.
- [6]. Şengör A.M.C., Yılmaz Y. Tethyan evolution of Turkey: a plate tectonic approach. *Tectonophysics* 75, 1981. – Pp. 181-241.

UDK 519.257

## PERSPECTIVES FOR THE USING OF BIG DATA TO ENSURE THE SECURITY OF CRITICAL INFORMATION SYSTEMS



**I.N. Tsyrelchuk**  
Advisor to Rector,  
TUIT



**D.S. Yakhshibaev**  
Dean of the  
Computer  
Engineering faculty  
of TUIT, PhD.



**Kh.R. Jiyanbekov**  
Deputy Dean of  
the Computer  
Engineering  
faculty of TUIT



**P.I. Soo**  
Professor of the  
Video Editing  
department of  
KUMA, South  
Korea



**Yu. V. Pisetsky**  
Dean of the Joint  
faculty of  
Information  
Technology of  
TUIT, DtS.

Computer Engineering faculty of TUIT, Republic of Uzbekistan  
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi,  
Republic of Uzbekistan  
E-mail: hurshidj@gmail.com

**I.N. Tsyrelchuk**  
Advisor to Rector, TUIT. PhD.

**D.S. Yakhshibaev**  
He graduated from the National University of Uzbekistan. He is working as a Dean of the Computer Engineering faculty of TUIT. Also, he is an Associate Professor of the Department "Higher Mathematics". Conducts research on algorithms and models of integrated, applied mathematics.

**Kh.R. Jiyanbekov**  
He graduated from the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi. He is a Deputy Dean of the Computer Engineering faculty, TUIT. In addition, he is an Associate Professor of the Department of "Information technologies" of TUIT. Conducts research on algorithms and models of big data analysis, data mining and cloud computing.

**Yu. V. Pisetsky**  
Doctor of technical Sciences, associate Professor, Dean of the joint faculty of information technologies of the Tashkent University of information technologies and the Belarusian State University of Informatics and Radioelectronics

**Park In Soo**  
He graduated from the Michigan State University. He is Professor of the Video Editing department of KUMA (Korean university of Media arts), South Korea.

**Abstract.** The Big Data plays a great role in the daily life of society. There are many types to ensure the security of systems. In this article were discussed new research method to ensure the security of critical information systems. Given new approach to support a defense to all kind of cyberattacks. In this article shown a simplified segment of the state graph of a modular-cluster network and its decomposition into core FLS subgraphs.

**Keywords:** big data, system, information system, critical information system, graph, subgraph, multigraph.

**Introduction.** There is a class of information systems, the safety of which is one of the basic requirements. These include, for example, systems for managing hazardous technological industries,

transport, infrastructure management facilities, and others. Such information systems are called critical. The security of critical systems largely depends on the security of communication networks, which form the basis of the critical information infrastructure of a country or company that has a critical system. At different stages of the process of ensuring the security of a critical information infrastructure or information system, it is necessary to have a reliable assessment of the security of the information architecture.

The main stages of the information security process, the implementation of which requires an assessment:

- risk assessment of information security breaches;
- analysis of information flows and architecture;
- analysis of the threat model and information security violator;
- development of an information security policy;
- building an information security system;
- assessment of the effectiveness of the information security system;
- management of information protection processes and incidents;
- information security control.

To obtain a reliable assessment of the security of a complex information system, it is necessary to classify and analyze various physical and logical elements and the relationships between them, taking into account the dynamics of state changes, which involves the use of various modeling methods, as well as reducing the state graph to simplify the model in order to ensure the solvability of the problem.

However, any model is not accurate. With increasing complexity of the system, types and number of analyzed objects, the accuracy of the model decreases. For critical information systems, this situation is not acceptable, since there is a risk of information security incidents with significant damage.

The combination of modeling methods with Big Data technologies makes it possible to increase the reliability of the results of dynamic modeling of the states of complex information systems when solving problems of ensuring information security.

Consider the possibility of using Big Data technology to build an information system model and assess its compliance with security policy using the example of the modular-cluster network (ISS).

*Materials and methods.* In the ISS, the architecture of the critical information system is represented in the form of a FLS-multigraph, the vertices of which are physical and virtual (logical) modules, and arcs are the information relations between the modules. Typing of modules is based on the analysis of their cluster properties. Each module property is an interface that determines the module's ability to interact with other modules that have a similar (paired) interface. Interfaces are distributed according to hierarchical levels of information interaction: F-physical, L-syntactic and S-semantic. A set of FLS-interfaces of the module determines its ability to implement information primitives for information processing (transfer, storage, copying, modification, etc.).

The figure 1 shows a simplified segment of the state graph of a modular-cluster network and its decomposition into core FLS subgraphs. If the modules have the same type of interfaces at three levels, then the system switches to a new state, in which new modules are created or existing ones (graph vertices) and interfaces (graph arcs) are deleted. The presence of interaction levels of the modules allows you to simulate the dynamics of the information processing process, formalize the security policy and take into account the impact on the information process of information protection tools. For example, to take into account in the model the influence on the process of differentiating physical access to equipment, the differentiation at the level of syntactic access to information by cryptographic means, semantic access by means of delimiting the powers of users and programs. Modeling the state transition allows you to find all the possible trajectories of the information process.

Comparison of the declared security policy with all system states makes it possible to identify dangerous process paths in which unauthorized access to protection objects is possible. To ensure the practical applicability of the method to complex systems, the state graph is reduced by typing the modules and their interfaces, which deduces the analysis task from the class of NP-complete problems. This reduces the accuracy of the model and the reliability of the analysis results.

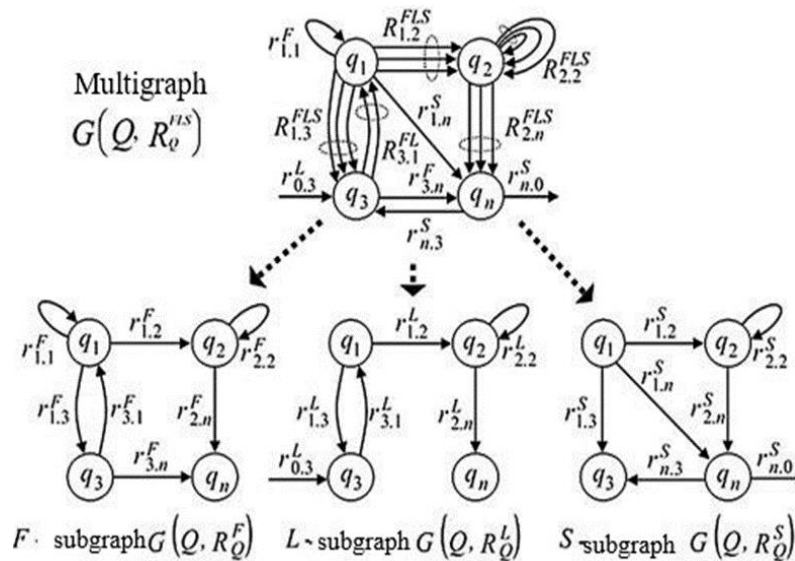


Figure 1. – A simplified segment of the state graph of a modular-cluster network and its decomposition into core FLS subgraphs.

Effective solutions are often found in a combination of different methods. Some areas of application of Big Data technology for solving information security problems already have practical application, others are at the stage of development of a methodology for application or implementation.

The use of Big Data technology is a promising direction for improving the results of modeling and evaluating information security in combination with the methods of constructing and analyzing a modular-cluster network, which will increase the completeness of the information system model, the reliability and applicability of the analysis results due to:

- reducing the time to build a modular-cluster model of CIS (accelerating the identification and typing of modules and their interfaces);
- representations in the model of a larger number and types of elements and relationships, which are reduced under standard conditions;
- ensuring state monitoring, incident management and decision-making on information security online (speeding up state search procedures, identifying security event anomalies, ensuring the completeness of the knowledge base for forming decision options, etc.);
- visualization of the state graph of the system indicating the places of violation of security policy, etc.

*Results.* In addition to methodological, there are legal aspects of using Big Data to solve information security problems. In many applications, Big Data technologies are used to analyze personal data, including cross-border processing, which may be in conflict with national law.

In order to ensure legal legitimacy, it is advisable to develop and adopt the Big Data international information security standard, which defines the basic principles for their use, cross-border transfer, ensuring the rights of individuals and the application of protection measures.



Such an international standard can be adopted, for example, by the International Telecommunication Union, in which the 17th Security Committee operates.

The basic principles for ensuring the security of personal data when using Big Data technology in accordance with the Moorish resolution:

–Openness in information on the composition of the information collected its processing, the purposes of use and transfer to third parties.

–Determining the purpose of collecting information directly during its collection and limiting the use of data to an exclusively established purpose.

–Obtaining consent to the use of data.

–Collection and storage of only that amount of data that is necessary for the implementation of the intended legal goals.

–Providing personal access to individuals to the data that was collected about them, providing information about data sources and any algorithms used for the further development of their profile.

–Providing individuals with the opportunity to correct and manage their information.

–Analyzing how the collection of information affects the user's privacy.

–Anonymization (depersonalization) of data.

–Restriction and control of access to personal data.

Conducting a regular analysis to confirm that the profiling results are “reliable, fair and ethical, and also meet the purpose for which the profiles are used.”

*Conclusion.* The use of Big Data technology in order to solve the problems of ensuring information security in critical information systems, including by telecom operators, is a promising direction for increasing the level of information protection. In particular, the technology improves the reliability of information security assessments in critical applications.

It is advisable to standardize the issues of safe use of Big Data technology for use in the field of information security at the international level, to harmonize legislation under the new realities of the information society.

### ***References***

- [1]. Informatica and Capgemini. The Big Data Payoff: Turning Big Data into Business Value, 2016.
- [2]. V. Kayser, B. Nehrke, D. Zubovic. Data Science as an Innovation Challenge: From Big Data to Value Proposition, Technology Innovation Management Review, 2018.
- [3]. M.S. Hopkins, R. Shockley. Big Data, Analytics and the Path from Insights to Value, MITSloan Management Review, 2011.
- [4]. Harvard Business Review. The Enterprise Lacks a Big Data Strategy for IoT Transformation, 2017, pp.1-12.
- [5]. S. Lavalle, M.S. Hopkins, E.Lesser, R. Schockley, N. Krushchwitz. Analytics: The New Path to Value, MITSloan Management Rev., 2010.
- [6]. S. Viaene, A. Van den Bunder. The secrets to managing business analytics projects, MITSloan Manag. Rev., 2011, pp. 65–69.
- [7]. A. Chebotko, A. Kashlev, S. Lu. A Big Data Modeling Methodology for Apache Cassandra. Proc. of IEEE Int. Congress on Big Data, 2015, pp.238- p.245.
- [8]. A. Fink, R. Guzzo. S. Roberts. Big Data at Work: Lessons from the Field, Society for Industrial and Organizational Psychology, 2017.
- [9]. S. Nalchigar, E. Yu. Business-driven data analytics: a conceptual modeling framework, Data & Knowledge Engineering, 2018, pp. 1-14.
- [10]. M.A. Berry, G.S. Linoff. Mastering data mining: the art and science of customer relationship management, Industrial management data system, 2000.

## **PERSPECTIVES FOR THE USING OF BIG DATA TO ENSURE THE SECURITY OF CRITICAL INFORMATION SYSTEMS**

***I.N. Tsyrelchuk***  
*Advisor to Rector,  
TUIT*

***D.S. Yakhshibaev***  
*Dean of the Computer Engineering  
faculty of TUIT, PhD.*

***Kh.R. Jiyانبekov***  
*Deputy Dean of the Computer  
Engineering faculty of TUIT*

***Park In Soo***  
*Professor of the Video  
Editing department of  
KUMA, South Korea*

***Yu. V. Pisetsky***  
*Dean of the Joint faculty of  
Information Technology of  
TUIT, DtS.*

*Computer Engineering faculty of TUIT, Republic of Uzbekistan  
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi,  
Republic of Uzbekistan  
E-mail: hurshidj@gmail.com*

**Abstract.** The Big Data plays a great role in the daily life of society. There are many types to ensure the security of systems. In this article were discussed new research method to ensure the security of critical information systems. Given new approach to support a defense to all kind of cyberattacks. In this article shown a simplified segment of the state graph of a modular-cluster network and its decomposition into core FLS subgraphs.

**Keywords:** big data, system, information system, critical information system, graph, subgraph, multigraph.

UDK 004.75

## ANOMALY DETECTION USING AUTOENCODER FOR DATA QUALITY MONITORING IN CLOUD



**C.S. Dzik**

*Utech Solutions, software and data engineer*



**I.I. Pilecki**

*Candidate of physico-mathematical sciences, Senior scientific researcher, Associate Professor of Informatics Department of the BSUIR*

*Utech Solutions LLC, USA*

*Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus*

*E-mail: constantine.dzik@gmail.com*

**Abstract.** Many works are dedicated to solving the data quality problem, a number of standards have been developed, but the problem has not been solved for decades. Moreover, the problem now requires a more complex solution due to processing large amounts of unstructured data in the cloud. This work presents the original project Autoencoder that focuses on the technology of analysis, detection and forecasting poor-quality data transmission based on machine learning and the use of neural networks.

**Keywords:** anomaly detection, autoencoder, artificial neural network, MLP, deep learning, AWS, S3, unsupervised learning.

**Introduction.** In today's world, cloud computing is a rapidly evolving technology that many organizations are adopting to enable their digital transformation. Cloud technology is opening up new competitive opportunities for companies globally and re-defining how they do business. The cloud makes resources, applications, platforms, and data available anytime, anywhere. Transferring data from outdated systems to the cloud enables you to scale your business, make your data productive, and make it more accessible.

Today most of the company's operations and strategic solutions rely heavily on the cloud data, so data quality is becoming an increasingly important characteristic. Data quality issues that arise when data and data applications are transferred to the cloud have a particular position among the challenges companies face. Cloud computing assumes new types and resources for potential data quality errors. In general, poor quality data can affect productivity, total and overall return on investment.

**Materials and Methods.** Before discussing data quality errors in the cloud, let's define what the term "data quality" means.

According to data quality experts, data is of high quality when it satisfies the requirements of its intended use. In other words, companies know that they have good quality data when they are able to use it to communicate effectively with their constituents, determine clients' needs, and find effective ways to serve their client base [5-8].

This data quality definition is broad enough to help companies with varying products, markets, and missions to understand if their data is up to standards.

Data quality is not good or bad, high or low. It is a range or an indicator of operability of the data that passes through a company. Data quality management ensures the context-dependent process

of improvement of suitability of the data, which is used for analysis and decision-making. The goal is to provide the vision of the “health” of the data by applying different processes and technologies to the increasingly complex data sets [5-7].

Quality of Data for business:

1. Consistency,
2. Completeness,
3. Reliability,
4. Accuracy,
5. Actuality

Data Quality as a process at the company level:

1. Analysis
2. Data profiling
3. Establishing data quality target indicators
4. Design and development of the data quality rules
5. Monitoring the correspondence of the actual data quality indicators to the target indicators
6. Improvement
7. Implementation of the rules in the data integration platform

First, good data quality management provides the framework for all the business initiatives, establishes the structure for all the units of the company for ensuring compliance with the data quality rules.

Second, accurate and up-to-date data provide a clear idea of the day-to-day operations of your company; therefore, you can be sure of the top and bottom indicators that all the data use. Data quality management also reduces unnecessary expenses.

Third, data quality management for compliance with the requirements and risk objectives. Good data management requires clear procedures and communication as well as good basic data

In our data quality study we focus on the data quality availability attribute.

The term “data availability” refers to the ability to ensure that the required data are always available at a particular location and time in an organization’s IT infrastructure, even if an error occurs. In practice, this means that the data that are not available if necessary parameters are useless. In fact, the situation described is even worse as data become more than just useless: if systems are configured on the basis of data availability, a catastrophic chain response of a system failure can occur; the data that users are counting on are missing, or worse, the data are in the system, but they are out of date or otherwise compromised [5-8].

Availability is defined as the degree of convenience for users to obtain data and related information, which is divided into the three elements of accessibility, authorization, and timeliness.

Availability indicators:

1. Within a given time, whether the data arrive on time
2. Whether data are regularly updated
3. Whether the time interval from data collection and processing to release meets requirements

Instead, we want to be sure that when we take advantage of the cloud to help data managing, we define data quality parameters at the same time.

The most obvious and compelling way to achieve the goal is to make sure we perform automatic data quality checks for all our data, wherever they are - in the cloud or elsewhere. We must always perform an on-site data quality check.

Virtually every company that works with data has a certain data quality (DQ) monitoring system. Some companies even hire an entire department that deals with the issue. This option is very expensive. In addition, most data quality checks are hard-coded and rule-based. In the event of a failure, the system notifies you of the risk indicator. Such rules are often critical to business continuity. For example, we cannot have a missing customer ID or a “risk profile” variable with an

incorrect value. As the amount of data grows, you cannot specify a rule for work with each attribute; not to mention the difficulty of working with hard-coded multidimensional control checks.

The best option is automated DQ (data quality) checks using Machine Learning to detect anomalies that we don't even need to explicitly program.

Anomaly detection (or outlier detection) is the identification of rare items, events or observations which raise suspicions by differing significantly from the majority of the data. Typically, abnormal data may be associated with a problem or rare event such as data quality, bank fraud, health problems, structural defects, faulty equipment, etc. This relationship is interesting in terms of the possibility to identify data points that can be considered anomalies, since the detection of such events is interesting from the point of view of sustainable business development [1].

This brings us to one of the key goals of this study: how to determine whether data points are normal or abnormal? In some simple cases, this can be determined at once (see figure below):

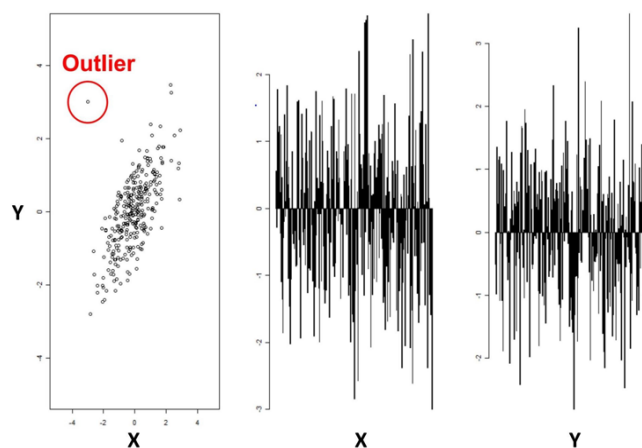


Figure 1. – Anomaly detection for two variables

Let us consider the case of two-dimensional data (X and Y): it is quite easy to visually identify anomalies through data points located outside the typical distribution. However, looking at the numbers on the right, it is not possible to identify the outlier directly from investigating one variable at the time.

An autoencoder is a type of artificial neural network used to learn efficient data codings in an unsupervised manner. The aim of an autoencoder is to learn a representation (encoding) for a set of data, typically for dimensionality reduction. Along with the reduction side, a reconstructing side is learnt, where the autoencoder tries to generate from the reduced encoding a representation as close as possible to its original input [3].

In terms of architecture, the simplest form of an autoencoder is a feedforward, non-recurrent neural network very similar to the many single layer perceptrons which makes a multilayer perceptron (MLP) – having an input layer, an output layer and one or more hidden layers connecting them – but with the output layer having the same number of nodes as the input layer, and with the purpose of reconstructing its own inputs.

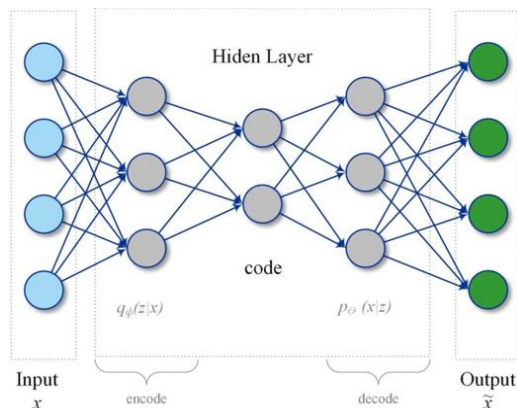


Figure 2. – Autoencoder network

Autoencoder anomaly detection – solve data quality issues without specifying them.

Autoencoders are neural networks that model their own input. In other words, autoencoders try to learn the identity function. They consist of two parts. Firstly, the encoder reduces the dimensionality of the input data. After that, the decoder tries to reconstruct the input from these reduced dimensions. Autoencoders are relevant for DQ because they can model whole data tables. They map all relevant fields of a data table to the input and the output layer of the network. Hidden layers between input and output layers learn the regular behaviour of the data. Of course, we need to have “good” data to train the model [2].

Imagine something unusual happened to the data, and we didn’t expect that. Even if the system conforms to all the classic hard-coded rules, an autoencoder trained to handle regular data will be completely malfunctioning and predicting incorrect data outputs. We will observe a large reconstruction error – a significant difference between predicted and actual values – and detect a data anomaly. The data doesn’t behave in the way that the autoencoder learned.

Such function is very useful when something new is happening in a dataset. For example, there are suddenly much more missing values (NA) than usual, or levels of categorical variables change or shift. The autoencoder allows you to omit the encoding of these rules. This feature also helps you to identify the source of the problem. The described properties of autoencoders allow the detection of anomalies at an early stage of their occurrence and reduce the cost of manual research [4].

**Approach and Results.** Neural Network Architecture [3]:

1. Trained with Keras/TensorFlow using Adam - an optimized version of backpropagation
2. Trained to minimize mean squared error between input vector and the output one:

$$MSE = \frac{1}{n} \sum_{i=0}^n (X_i - \hat{X}_i)^2$$

3. Three hidden layers, with number of units.
4. Activation functions: Tanh and ReLu

S3 object metadata:

```
{
  "timestamp": "2017-07-06T04:00:05.173Z",
  "bucket": "handler-data-lake",
  "key": "bin2/root/predictions/97e05fcc-61ff-11e7-b192-80e6501b37de.docx",
  "principalId": "ACQ34MYG55IWH",
  "filename": "97e05fcc-61ff-11e7-b192-80e6501b37de.docx",
  "extension": ".docx",
  "eventName": "ObjectCreated:Put",
  "eTag": "6797bde3578a82e62a4564de50f18cc0",
  "awsRegion": "us-west-2",
  "size": 242,
  "metadata": {
    "department": "021",
    "last": "Johnston",
    "first": "Amy"
  }
}
```

Figure 3. – Programm cod

Data preparation takes the input metadata and output the same records with additional features extracted from the input attributes. Features example: unique id for each file, file size, file size percentage change, timestamp etc.

The general idea behind unsupervised anomaly detection approaches is to find an approximate model that can capture the normal behavior of complex systems. The approximate model can then be used to flag anomalies if the deviation of the predicted behaviors of the trained model from the actual observation exceeds some certain threshold. Training on the normal data, the autoencoder is expected to produce higher reconstruction error for the abnormal inputs than the normal ones, which is adopted as a criterion for identifying anomalies (Figure 3). On Figure 4 we have two columns – time range difference and file size difference. Orange line represent the training curve, while blue line represents new input curve. Green dots indicate normal data points, while red dots indicates potential anomalies. As we can see on Figure 4, unsupervised anomaly detection approaches Autoencoder allows us to identify unusual patterns and behaviors.

**Conclusion.** The main purpose of unsupervised deep learning suggested approach is ability to detect anomalies based on load patterns for data files in S3. By using metadata from Objects in S3 and performs transformations and formatting on top of the metadata so it can be consumed by Autoencoder Artificial neural network. In the future our plan to apply the proposed technique for various applications. Also the plan to conduct a more in-depth theoretical analysis of the proposed technique.

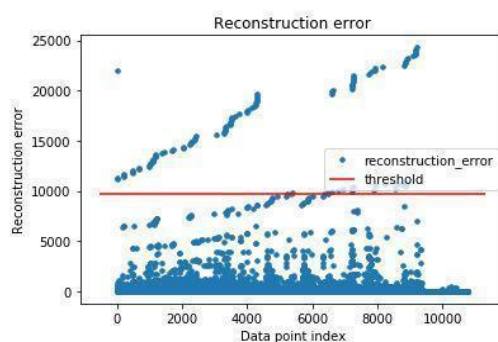


Figure 4. – Model allows to detect anomalies

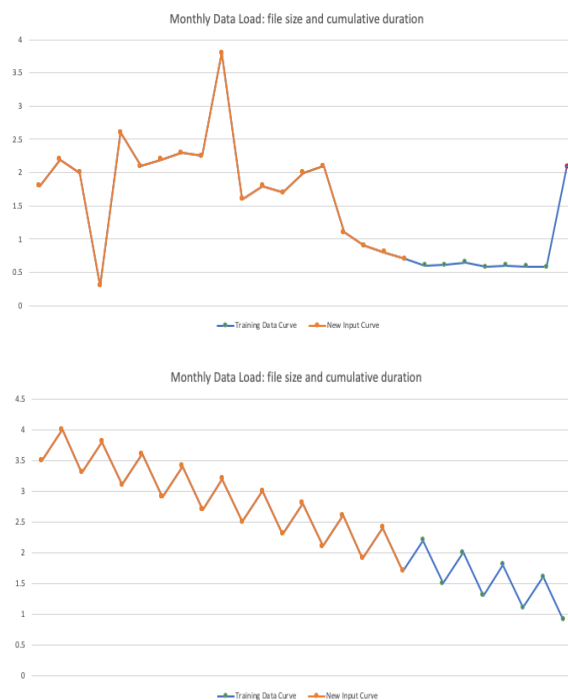


Figure 5. – Identify unusual patterns and behaviors

Benefits of the offered model:

1. Unsupervised learning: no need to know the labels of anomalies before learning. We are using an artificial neural network - Autoencoder algorithm. That allows us to identify unusual patterns or behaviours and label output with the results.
2. Able to take all factors that may cause anomalies into count at the same time.
3. Able to take multiple files together to run at the same time.

### References

- [1]. Jason Brownlee, Deep Learning for Natural Language Processing, 2017.
- [2]. Jason Brownlee, Machine Learning Mastery With Python Understand Your Data, Create Accurate Models and Work Projects End-To-End, 2016.
- [3.] Giancarlo Zaccone, Md. Rezaul Karim, Ahmed Menshawy, Deep Learning with TensorFlow, 2017.
- [4]. Chun Chat Tan, AUTOENCODER NEURAL NETWORKS: A Performance Study Based on Image Reconstruction, 2009.
- [5]. ISO 8000-2.
- [6]. ISO/TC 8000-110.
- [7]. ISO/TC 8000-120.
- [8]. [https://en.wikipedia.org/wiki/Data\\_quality](https://en.wikipedia.org/wiki/Data_quality).



UDK 62-787.1

## PROCESSING IMAGES OF TECHNOLOGICAL OBJECTS



**L.P. Varlamova, PhD**  
Candidate of Technical Sciences,  
Associate Professor  
Associate Professor of Tashkent  
University of  
Information Technologies  
named after al-Khwarizmi



**A.G. Davydovsky**  
Associate Professor of  
the Department of  
Engineering Psychology  
and Ergonomic,  
Candidate of Biological



**K.D. Yashin, PhD**  
Head of the Department of  
Human Engineering and  
Ergonomics of the Belarussian  
State University of Informatics  
and Radioelectronics

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus,  
Tashkent University of information technologies named after al-Khwarizmi, Tashkent, Republic of Uzbekistan  
E-mail: vlp@bk.ru, yashin@bsuir.by

### **L.P. Varlamova**

Associate Professor, Tashkent University of Information Technology named after al-Khwarizmi. Field of science my interests are research on the development directions of robotics and unmanned aerial systems, as well as their applications.

### **K.D. Yashin**

Head of the Department of Human Engineering and Ergonomics of the Belarussian State University of Informatics and Radioelectronics, Associate Professor, Candidate of Technical Sciences. Inventor of the USSR, he has 55 copyright certificates for inventions and patents. He graduated from the Belarussian State University The author and main executor of the project of creation of the Department of Human Engineering and Ergonomics at the Belarussian State University of Informatics and Radioelectronics, which provides training and graduation of specialists with higher education in the specialty "Engineering-Psychological Maintenance of Information Technologies", qualification of system engineer (software engineer).

### **A.G. Davydovsky**

Associate Professor, Department of Engineering Psychology and Ergonomics, Faculty of Computer Engineering BSUIR, Candidate of Biological Sciences, Associate Professor. Specialty "Biology". He graduated from the doctoral program of BSUIR with a degree in "System analysis, management and information processing." Conducted research in the field of social informatics, mathematical modeling of biological and biosocial systems, methodology of preventive risk management in socio-technical and innovative production systems. Member of a number of international scientific societies. The author of curricula and manuals for students and undergraduates.

**Annotation.** The paper considers the problem of creating a smoke detection and fire detection system for technological objects. Fire detection is carried out using a vision system. Monitoring of technological objects is carried out using an unmanned aerial vehicle. For further actions, processing and classification of images is necessary.

**Keywords:** image processing, system, unmanned aerial vehicle, fire, smoke, technological object.

### **Introduction**

To assess the state of technological objects and their parameters, technical vision systems built on the basis of artificial intelligence are currently widely used. One of the tasks to be solved by vision systems is the recognition of objects in static images, which include the task of recognizing smoke and fire [1].

An important problem of ensuring industrial and environmental safety is the fire protection of various objects and territories – providing automated operational detection of fires using unmanned aerial vehicle technologies [2-4].

The current trend is the expanding use of video technology in fire safety tasks, characterized as “video analytics”. Used video detectors can detect a fire in a room and in open areas automatically by specific signs in the image, allowing you to assess the situation at the production facility.

### **Results of research**

*Formulation of the problem.* Most automated video analytics systems are based on computer image processing and analysis of their changes. In this case, video detectors can be used in case conventional fire alarms are not applicable. The literature cites data according to which the probability of false alarms is <1%, and the recognition range is 10 km for a smoke area 10x10 m in size [5]. To monitor and identify fires and fires in the video image, an adaptive background model of the area observed by the video camera and a generalized color model of the fire can be proposed based on statistical analysis of a sample of images containing fire pixels, followed by transmission via radio or optoelectronic channels in digital data processing.

In enterprises of the chemical, petrochemical industry, the main requirement for the detection of fire or smoke is considered the need for early detection of an emergency. A good alternative to traditional chemical sensors is a smoke control video control system, which allows, in addition to the fact of smoke generation, to determine the degree of smoke, the number of smoke areas, the contours and sizes of these areas, as well as the direction of smoke propagation.

Thus, the problem of constructing an intelligent system for assessing the state of technological equipment based on an unmanned aerial vehicle is considered. It is proposed to use expert systems of a new generation [6].

*Monitoring technological equipment.* For integrated monitoring systems for technological objects, the detection of the effective surface of dispersion, reflection, or emission is of utmost importance. At the same time, in the microprocessor system of the unmanned aerial vehicle, a continuously adjusted reference map of the intensity of the reflected (absorbed) signals or radiation is formed based on the integration of the effective scattering, reflection and absorption surface in the scanning parameters and resolution elements of the reference map generated using the measuring complex. In one resolution element of the measuring complex, the reflectance (emissivity)  $S_{отр}$  of the observed object is found as the total value over the area (1):

$$S_{отр} = \int_S^{\infty} k_i ds, \text{ или } S_{отр} = \sum_{i=1}^n s_i k_i, \quad (1)$$

where:  $n$  - is the number of resolution elements of the map with area  $S_i$ , reflection coefficient  $k_i$  in the resolution element of the meter.

The most important parameter of a fire detector is the maximum detection range of a small-sized fire source by a subsystem with an automatic detection method based on the excess of the video signal generated by the sensor from the object over the threshold signal. In the process of automatic detection of an object (fire), the signal from the output of the photodetector after preliminary amplification is fed to a threshold device that detects the excess of the signal from the object above the threshold. The probability of detection in the presence of a noise signal clearly depends on the signal-to-noise ratio. The object (fire) is always placed on the background. The useful signal at the output of the radiation receiver  $U$  is the difference between the signals from the object with the fire source ( $\pi$ ) and background ( $f$ ) in the spectral range of the sensor (2):

$$U = U_{\pi} - U_{\phi} = \frac{A_{\pi} A_{06}}{R^2} \int_{\lambda_1}^{\lambda_2} \tau_{\alpha}(\lambda) \Delta L(\lambda) \tau_{06}(\lambda) S_D(\lambda) d\lambda, \quad (2)$$

where:  $\lambda$  - is the radiation wavelength;  $A_{\Pi}$  - the area of the fire;  $A_{об}$  - the area of the entrance pupil of the lens;  $R$  - distance to the object;  $\Delta L(\lambda) = L_{\Pi}(\lambda) - L_{\Phi}(\lambda)$  - the absolute contrast of the brightness of the fire source and background;  $\tau_a(\lambda)$  - spectral transmission of the atmosphere, which depends on the following parameters: range to the object, meteorological visibility range at relative humidity and temperature;  $\tau_{об}(\lambda)$  - spectral transmission of the lens.

The relationship between the spectral sensitivity of the receiver  $S_D(\lambda)$  and the specific detectivity  $D(\lambda)$  is described by the expression (3):

$$S_D(\lambda) = D(\lambda)U_{\Pi}/\sqrt{A_D\Delta F}, \quad (3)$$

where:  $\Delta f$  - is the effective noise band of the electronic path of the sensor;

$A_D$  is the area of the sensitive area of the receiver (pixel, if a matrix receiver is used);

$U_{\Pi}$  - noise signal generated during monitoring, due to the influence of secondary natural and man-made sources.

The detection range of a small fire with a single-element detector in the infrared range is up to 7–8 km for a clean atmosphere and up to 2-3 km in conditions of poor visibility. The brightness of the background radiation in the visible range of the spectrum (for a video camera) is determined by the formula (4):

$$B_{\phi\lambda} = \frac{\omega_{\lambda}S_0(\lambda)\mu_0\tau_a(\lambda)}{\pi} + B_{\alpha}(\lambda) + (1 - \omega_{\lambda})M_{\lambda}(\lambda, T_{\Phi}), \quad (4)$$

Where  $S_0(\lambda)$  is the spectral solar constant (illumination) on the surface of the Earth;

$\mu_0$  - is the cosine of the angle of the Sun with respect to the normal of the observed area;

$\omega_{\lambda}$  - is the albedo of the reflecting surface of the background;

$\tau_a(\lambda)$  - transmission of the atmosphere between the object and the device;

$B_{\alpha}(\lambda)$  - is the brightness of the atmospheric haze between the object and the device;

$M_{\lambda}(\lambda, T_{\Phi})$  - is the Planck radiation function for the background temperature.

The brightness of the flame radiation in the visible range is determined by the Planck function with the flame temperature  $T_{\Pi}$ . The expression for light illumination in the plane of the object has the form (5):

$$E_{CB} = 683 \int_0^{\infty} E(\lambda)V(\lambda)d\lambda, \quad (5)$$

Here:  $E_{CB}$  - light illumination of the object, lux;

$E(\lambda)$  - is the spectral density of the energy illumination of the object;

$V(\lambda)$  - is the relative spectral sensitivity of the eye;

683 - conversion factor of energy values into light, lm / W.

For spectral illumination in the plane of the entrance pupil in the case of solar radiation reflected from background objects, when the intrinsic background radiation at a temperature of 300 K corresponds to (6):

$$E_{\phi}(\lambda) = \frac{\omega_{\lambda}S_0(\lambda)\mu_0\tau_a(\lambda)A_{\Pi}}{\pi R^2} + E_{\alpha}(\lambda). \quad (6)$$

Here  $E_{\alpha}(\lambda)$  is the atmospheric illumination corresponding to the brightness of the atmospheric haze on the observation beam. Spectral illumination from radiation from a fire source (7):

$$E_{\Pi}(\lambda) = \frac{M_{\lambda}(\lambda, T_{\Pi})A_{\Pi}\tau_a(\lambda)}{R^2} + E_{\alpha}(\lambda). \quad (7)$$

For light illumination of the background at the entrance pupil in two cases (8) and (9):

$$E_{\phi \text{ CB}} = \frac{683\omega\mu_0 A_{\text{п}}\tau_{\alpha}}{\pi R^2} \int_0^{\infty} S_0(\lambda) V(\lambda) d\lambda + E_{\alpha \text{ CB}} \equiv E_{\phi \text{ отр}} + E_{\phi \text{ CB}}, \quad (8)$$

$$E_{\phi \text{ CB}} = \frac{683A_{\text{п}}\tau_{\alpha}}{R^2} \int_0^{\infty} S_0(\lambda) V(\lambda) d\lambda + E_{\alpha \text{ CB}} \equiv E_{\phi \text{ соб}} + E_{\phi \text{ CB}}. \quad (9)$$

Moreover, at a distance of 1 km, the probability of detecting a fire focus of 1 m in size with a video camera will be significantly less than 0.5. Fire places with a temperature of 1500 K<sup>0</sup> with a linear size of 1 m<sup>2</sup> can be detected using a video camera at a distance of up to 220 m or less. To increase the detection range of a fire or a fire at a distance of more than 1 km, it is advisable to use either traditional video cameras operating in the light range, but with a much higher sensitivity, or infrared sensors.

Information technology video analytics using an unmanned aerial vehicle can be useful for monitoring fire safety and remote monitoring the integrity of the structures of oil, gas and product pipelines and other hazardous production facilities with significant linear dimensions.

At the stage of image processing, the color scheme is balanced, the average value of each R, G, B color component of the image is calculated in order to obtain a real level of gray color.

At the next stage, smoke regions are distinguished, for which purpose the color characteristics of the smoke regions are used. Since smoke has a color from light to dark gray, this property is used to highlight potential areas of smoke in images in which the intensities of the color components are in the following ratio:

$$\begin{cases} |R - G| < T, \\ |G - B| < T, \\ |R - B| < T, \end{cases}$$

where T is a threshold that can be adjusted according to the training set of video files [7].

Naturally, the use of color characteristics to localize areas of smoke is not enough. It is known that the smoke areas are not stationary, but constantly move and change their shape. Therefore, the next step in detecting smoke areas will be the detection of frames on the video stream of objects.

The processing result is presented in binary form. Figure 1 shows the stages of image processing. An example of smoke generation is given.



Figure 1. – Image Processing Steps

When observing an object against a relatively uniform background, the problem arises of estimating the parameters of geometric transformations with the required accuracy. Thus, the

development of effective algorithms for the isolation, detection, and estimation of parameters of airborne objects remains a very urgent task to date [8].

*Hazards identification and detection algorithm.* As a model of image formation obtained using a video sensor, we consider the space-time model [8], where we consider only spatial information, then, omitting the frame number, this model can be written in the following form:

$$l(i, j) = h(i, j) r(i, j) + g(i, j)(1 - r(i, j)) + \xi(i, j), \quad i = \overline{0, N_x - 1}, \quad j = \overline{0, N_y - 1}, \quad (10)$$

where -  $N_x, N_y$  height and width of the frame;  $l(i, j)$  is the observed image;  $g(i, j)$  and  $h(i, j)$  are unknown functions whose values are the brightnesses of the background and object points, respectively;  $\xi(i, j) \approx N(0, \sigma_\xi^2)$  is the noise of the video sensor. The function  $r(i, j)$ , which determines the location of an object in the image, is defined by the rule:

$$r(i, j) = \begin{cases} 1, & \text{if at the point } (i, j) \text{ of the frame there is an object,} \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

The task of distinguishing objects is to estimate  $\hat{r}(i, j)$  from the observed image  $l(i, j)$ . The stage of object isolation, as a rule, precedes the stage of object detection - the decision on the presence or absence of an object in the image  $l(i, j)$ . In this case, the appearance of smoke. In [9], when observing air objects, a cloudy sky with smooth transitions of brightness is considered as a background component. In this case, it is proposed to use an autoregressive model of the form to describe such a background:

$$g(i, j) = \sum_{\substack{n=0 \\ (n,m)=(0,0)}}^{L-1} \sum_{m=0}^{L-1} a(n, m) g(i - n, j - m) + \mu(i, j)$$

where  $a(n, m)$  - are the autoregression coefficients;  $\mu(i, j)$  is an unobservable white generating noise with zero mathematical expectation, the quantity  $L$  determines the order of the model. To make a decision about the presence of objects in the frame, it is necessary to compensate for the background, which in turn involves an assessment of the parameters  $a(n, m)$ . In situations characterized by a priori uncertainty and spatial variability of the observed images, it is advisable to use adaptive methods of information processing in which automatic optimization of the parameters and structure of the algorithm relative to the current background characteristics is carried out.

Next, image processing is performed using the Laplace, Sobel, Prewitt, Gaussian [10-12] transformation methods, histogram processing.

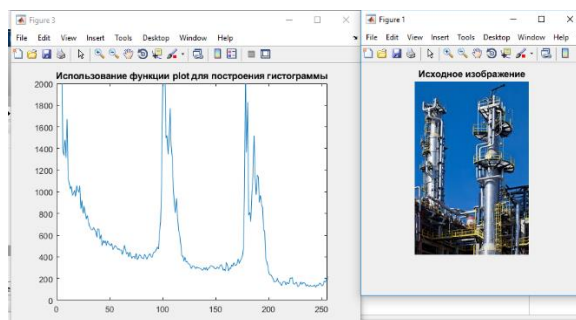


Figure 2. – Histogram of the source image

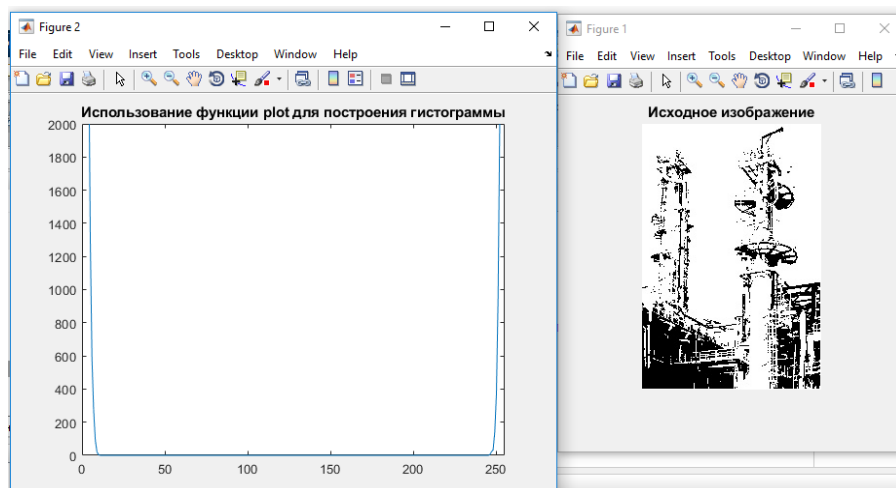


Figure 3. – Image histogram after Gaussian transformation

In [10], a function is considered that corresponds to Gaussian blur:

$$h(r) = -e^{\frac{r^2}{2\sigma^2}}, \quad (12)$$

where  $r^2 = x^2 + y^2$ ,  $\sigma^2$  - standard deviation. After convolution of the image with a similar function, we get the same image, but slightly blurred. The degree of blur is determined by the value of  $\sigma$ . If we calculate the Laplacian for this function, we get the following expression:

$$\nabla^2 h(r) = -\left(\frac{r^2 - \sigma^2}{\sigma^4}\right) e^{\frac{r^2}{2\sigma^2}}. \quad (13)$$

Since the characteristic feature of smoke is smoothing of the faces of objects, smoke areas are then estimated based on the combination of moving image areas and areas whose color components correspond to gray, various methods have been proposed for [11-14] forming contours at different resolution levels (Fig. 2.3) With the help of binarization and conversion, the coefficient of difference between the energy of the spectra and the gray opaque object present is determined - smoke is detected.

### Conclusion

When observing the technological apparatuses of the chemical and petrochemical industry, problems arise associated with the large size, distribution and extent of objects [15]. There may be situations in which there are violations of the technological regulations and, as a result, the occurrence of fire, fire and smoke. Using the technology of unmanned piloting, monitoring and monitoring of large-sized objects becomes more accessible, and the regulated flight of objects allows you to identify and prevent dangerous situations.

The approach proposed in the work of applying the spatio-temporal model of objects of observation and highlighting smoke areas allows us to solve the problem of detecting moving objects (smoke, flames) in the absence of a priori information, combining further moving areas of the images of objects to obtain hazard information.

### **References**

- [1]. Davydovsky A.G. and other problems of the use of unmanned aerial vehicles in ensuring public, industrial and environmental safety / A.G. Davydovsky, D.V. Likhachevsky, S.K. Dick, C.D. Yashin, L.P. Varlamova, J.A. Tazhiev // Big Data and advanced Analytics. 2019 = BIG DATA and high-level analysis: Sat materials of the V Intern. scientific-practical conf. Republic of Belarus, Minsk, 2019. At 2 hours, Part 2 – Pp. 305-320.
- [2]. Binenko, V.I. Some results and prospects of using unmanned aerial vehicles for environmental monitoring / V.I. Binenko, V.K. Donchenko, V.L. Andreev, R.V. Ivanov // Ecological chemistry. - 2001. - N1. - Pp. 21-31.
- [3]. Hayroyan, Z.A. Monitoring of oil trunk pipelines using unmanned aerial vehicles / Z.A. Hayroyan, O.A. Korkishko, G.V. Sukharev // Electronic scientific journal "Engineering Journal of the Don." - 2016. - N4. - [Electronic resource]. - Access mode: <http://www.ivdon.ru/ru/maga-zine/archive/n4y2016/3898>. - Access date: 12.17.2018.
- [4]. Popov NI, Emelyanova OV Dynamic features of monitoring overhead power lines using a quadcopter // Modern problems of science and education. - 2014. - No. 2. URL: <http://www.science-education.ru/ru/article/view?id=12773> (accessed date: 08.15.2018).
- [5]. Katkovsky, L.V. Metrological characteristics of video thermal fire detection systems / L.V. Katkovsky, S. Yu. Vorobyov // Journal of Applied Spectroscopy. - 2012. - T.79, N1. – P. 168-176.
- [6]. Varlamov O.O. and etc. Mivar expert systems for supporting production processes in transport. O.O. Varlamov, Adamova L.E., Nazarov K.V. Saraev D.V., Jha Punam, Varlamova I.A. // T-Comm. Vol. 11. # 5-2017. – Pp. 53-59. Access Mode: <https://cyberleninka.ru/article/n/mivarnye-ekspertnye-sistemy-dlya-soprovozhdeniya-proizvodstvennyh-protsessov-na-transporte>. - Date of access: 09.17.2019.
- [7]. Vogel, Wolfgang. A non-zero-divisor characterization of Buchsbaum modules. Michigan Math. J. 1981, no.2, Pp. 147-152. doi: 10.1307 / mmj / 1029002505. <https://projecteuclid.org/euclid.mmj/1029002505>.
- [8]. Alpatov B. A., Blokhin A. N., Muravyov V. S. Algorithm for image processing for systems of automatic tracking of air objects // Digital signal processing, 2010. No. 4. Pp. 12–17.
- [9]. Muravyov V. S., Muravyev S. I. Adaptive algorithm for the detection and detection of airborne objects in images // Information and control systems. 2011. No 5. Pp. 8–14.
- [10]. N. Senthilkumarn, R. Rajesh. "Edge Detection Techniques for Image Segmentation- A Survey of Soft Computing Approaches", IJRTE, voll, No 2, 2009. Pp. 250-254.
- [11]. Salem Saleh Al-amri, N.V. Kalyankar, Khamitkar S.D: Image Segmentation by Using Edge Detectors, International Journal on Computer Science and Engineering, Vol. 02, No. 03, 2010, Pp. 804-807.
- [12]. Y.Ramadevi, T.Sridevi, B.Poornima, B.Kalyani Segmentation And Object Recognition Using Edge Detection Techniques// International Journal of Computer Science & Information Technology (IJCSIT), Vol 2, No 6, 2010, Pp.153-161.
- [13]. Muthukrishnan.R1 and M.Radha Edge Detection Techniques for Image Segmentation // International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, 2011. Pp.259-267.
- [14]. Michael A. Wirth, Image Segmentation: Edge-based/University of Guelph Computing and Information Science, Image Processing Group, 2004.
- [15]. Korkishko, A.N. Location of leaks of oil, oil products and unstable hydrocarbon liquids on trunk pipelines / A.N. Korkishko, Sh.I. Rakhmatullin, V.G. Karamyshev // Journal "Problems of the collection, preparation and transport of oil and oil products." - 2011. - N2. - Pp. 142–147.

UDK 004

## THE RADIATION DOSES DIAGNOSTIC X-RAYS AND QUALITY ASSURANCE TESTS FOR DEVICES



**N. A. Abbood**  
Ministry of Science and  
Technology Treatment and  
Management of Radioactive  
Waste Directorate-



**M. W. Alhamd**  
Ministry of Science and  
Technology  
Radiological and Nuclear  
Safety Directorate



**Z. A. Alqaisi**  
Ministry of Science and  
Technology  
Radiological and Nuclear  
Safety Directorate

Ministry of Science and Technology Treatment and Management of Radioactive Waste Directorate, Iraq  
Ministry of Science and Technology Radiological and Nuclear Safety Directorate, Iraq  
E-mail: Yaseen\_mohsin@yahoo.com, zeki\_alqaisy@yahoo.com, engnazar777@yahoo.com

### **Nazar Ali Abbood**

He was born in Baghdad, Iraq in 1973. He obtained his B.Sc. in Electrical Engineering in 1998, M.Sc. in Communication Engineering in 2017 at University of Technology, he worked as a lecturer in the Nano Technology Engineering Faculty, Ministry of Science and Technology, Treatment and Management of Radioactive Waste Directorate-Iraq, Iraq. Since the beginning of 2018, he works as an assistant professor in the same Faculty. He is a consultant member of the Iraqi Engineering union since 2018. He is interested in the subject of Nano Technology Engineering, video and image processing systems.

### **Dr. M. W. Alhamd**

He was born in Baghdad, Iraq in 1975 He received a Master's degree in Physics in 2005. He received a master's diploma degree in Physics in 2014. And Ph.D. in Nanotechnology in 2019 from Belarussian State University of Informatics and Radio Electronics, Republic of Belarus. He got a PhD in Nanotechnology. Inventions and registered patents in Belarus and Iraq on nanotechnology, Patent in Nuclear Physics, Patent in medical physics. He has several works in fields of magnetic nuclear resonance imaging, radioactive waste, UV rays with oxide compounds. aluminum, cadmium sulfide and zinc oxide, Radiation from X-ray machines and safety uses of nuclear energy and electromagnetic energy.

### **Zaki Abduljabbar Alqaisi**

He was born in Baghdad, Iraq, in 1971. She obtained her B.Sc. in applied mathematics, information and computer sciences in 1999, M.Sc. in computer science and applied mathematics in 2009 at Baghdad University, He works in the field of radiation protection and the official of the radiation protection department in the Nuclear and Radiation Safety Directorate, and he has posts in the International Atomic Energy Agency.

**Abstract.** Right now estimated the passage surface portions of patients which result from ordinary radiography (chest, head, belly, appendages) by picking instruments in AL-sader. A gathering of patients (10 patient) are chosen for every assessment and the mean portions are determined. The surface dosages with understanding squeezed (ESDpp) were estimated utilizing thermoluminescence dosimeter (TLD) and passage surface dosages with nonattendance of patient (ESDpa) by utilizing versatile ionizing chamber Sweden made. We found that the perusing of thermoluminescence dosimetry (TLD) is more than normal estimation of ionization chamber which are utilized and the diverse originated from the dispersing of radiation in understanding body, the dissipating coefficient to all tests are determined and we found that coefficient extend between(1.05 – 1.1) and this worth is near global qualities which equivalent to (1.09). The patient dosages right now contrasted and worldwide qualities, it is discovered that our qualities are more noteworthy by a range between (1.2-1.7) in many instruments, and the range is shut to (1) at whatever point the instrument is new and the experience of radiographer is acceptable, The estimations of this examination are talked about in detail. For the



significance of (Quality Assurance) we measure seven tests for three instruments in particular, and these tests are (– voltage Test – Test of current Linearity – Test of Linearity of Time – Half Value Layer Test – Beam Alignment Test – Optical and Radiation Field Congruence Test – Focal Spot Test ), The tests demonstrated that the gadgets (F, C) breezed through quality confirmation assessments are unfit for work right now, yet the (A) didn't finish most assessments of value affirmation, accordingly prescribed not to utilize this machine.

**Introduction.** Radiological presentation coming about because of X-beam assessment is one of the most significant modern wellsprings of introduction for mankind. It was discovered that by figuring that about 21% of the complete portion of physical impacts (Total Somatic Dose) and about 10% of the absolute viable portion result from radiography. Hence, it was important to focus on this sort of assessment and arrange all the variables that influence the radiation portion of the patient while keeping up a decent beam of highlights and difference [1].

From the investigation in various emergency clinics and facilities, which use X-beams in the analysis, it was discovered that there are contrasts in presentation components to one assessment, which may prompt an expansion in the radiation portion assimilated in the patient, and among these elements is what is explicit to the idea of the gadget, for example, voltage, quality, current force and sort of combination produced using it Target, center territory, type and thickness of the channel, factors that rely upon down to earth understanding on the gadget, the nature of the assessment, for example, the utilization of the system (Grid) and power (boards), the nature of movies and the chronicle of cases and distinctive fermentation factors, for example, time and temperature. In this manner, it was important to be comfortable with the idea of the various elements, how they influence the patient's radiation portion, and how to get a very much separated, clear-beam picture with the least radiation portion for the patient [2]. Because of the across the board utilization of X-beams in clinical analysis, an expansion in radiological presentation to laborers, patients, and the overall population rose, which required the use of value affirmation programs for radiological gadgets so as to stay away from unexplained radiological introduction. Studies have demonstrated that numerous patients and laborers are presented to over the top and unjustified radiation portions, and a few examinations show The distinction in the measurements that patients get during radiography starting with one organization then onto the next or even inside one establishment because of the absence of congruity of the specialized details of the gadgets utilized. It likewise indicated that ineffectively working gadgets can prompt great radiographs, however patients get radiation dosages in overabundance of as far as possible [3].

**The Proposed Method.** The reagents are gadgets or frameworks having a place with radiation identification and estimation innovation with both molecule and wave types, and the discovery system depends on numerous common impacts of particles or photons with the material, and the vitality consumed by the radiation that is recognized in the locator is moved to another body, for example, electric flow. To guarantee the wellbeing of laborers in the radiation fields, just as to lessen the portion got by the patient, a quality confirmation program must be applied to indicative radiology divisions in clinical establishments, as it prompts an exact finding without presentation to unjustified radiation dosages. The quality affirmation program is characterized as the composed exertion by laborers in the indicative clinical establishment so as to keep up that there is no adjustment in the specialized details of gadgets after some time. [4] It incorporates:

A thermometers streak.

B radiation level estimating gadgets.

C quality assurance..

**Thermo luminescence Dosemetry (TLD).** The premise of the activity of blaze materials is the development of aggravating iotas, on the grounds that the falling particles or photons associate with the glimmer substance, shaping various electrons with bothered vitality, and upon after coming back to steadiness they produce light photons in flashes in the bright area or in the obvious light district, so the materials which can create this marvel can be utilized to identify ionizing radiation [4], yet with two conditions:

– The way toward making flashes should happen when radiation interfaces with the precious stone, and be joined with inciting the gem in another manner all together for the light to transmit.

– The way toward returning electrons to their steady state must be joined by the emanation of light and not by changing over the abundance vitality into heat.

Free electrons and holes are created by ionization, and an (electron-hole) pair frames because of disturbance, and as it comes back to a steady state it produces bright photons. The nearness of enacted polluting influences prompts the arrangement of another degree of bad tempered and stable state inside the illegal band that contains overabundance heat as warmth, which cools because of the bondage procedure, so the tainted gems don't have to work at low temperatures [5].

This light can be palpated or blazing with a thermometer framework and utilizing a Pitman Toledo mode - 654TLD-peruser. These meters require an optical speaker, which is corresponding to the power of the ionizing radiation consumed by the blazing material, and afterward you arrive at the meter to get the sum The tally that speaks to the radiation portion subsequent to adjusting the framework.

(Lithium Fluoride) is put as tablets with a distance across of 30mm. LiF-6 is utilized to quantify neutrons and Lif-7 to gauge X-beams and kama. A cycle begins before the estimation at 100C ° and the estimation cycle at 240C ° and the cleaning and cooling cycle at 300C °, so this gadget should be cooled, and nitrogen gas is utilized for this reason which has an immaculateness of 99.99% and is feeling the squeeze 400 mul/min as There is a channel between the jug and the gadget that makes the gas 100% unadulterated that enters the gadget [5]. Warm glimmer (TLD) is utilized to quantify the inward surface portion (ESD) of all pieces of the body, and Figure (1) outlines the warm blaze gadget (TOLEDO). These tablets are described by:

- Small size
- It has high strength
- It has extraordinary affectability and exactness
- Do not rely upon vitality or portion rate
- It has a wide scope of radiation portion (10<sup>-5</sup> - 10<sup>4</sup>) mGy
- Atomic numbers proportional to the body tissues.

**Standard Thermoluminescence Calibration.** Prior to beginning the way toward estimating any arrangement of tablets for individual presentation, the way toward adjusting the Toledo gadget, and getting ready to gauge the got portions, is done as follows:

1. Find the affectability of the TLD peruser

A gathering of (10) Lif-7 glimmer tablets is taken, which are warmed by a unique stove and the procedure is called Annealing, where the tablets are warmed to a specific temperature (80 °) and for five hours expel any residual dosages of introduction Previously, the radiation foundation of these circles (BG) is estimated by a TLD peruser, after which the plates are presented to a particular portion by the irradiator which contains the source (Sr90/Y90) and the producer of beta particles, as in Figure 1-1 The portion is controlled by deciding the quantity of turns of the gadget and the measure of portion got for every meeting is (0.03 mGy).

–dose received (mGy) = number of cycles x measure of portion got per meeting (mGY)

The (TLD peruser) gadget is introduced on the adjustment affectability of the TLD reaction estimation (99.99), at that point the greater part of these tablets are estimated by the gadget, and the rest of it is to guarantee the exactness of the new affectability and its appropriateness for estimation.

Measurements:

The new sensitivity is calculated from the following relationship:

$$\text{New sensitivity} = \text{calibration sensitivity} \times \frac{\text{Known radiation dose}}{\text{Readout rate - radiative background}}$$

(B.G) = 0.04 mGy

Dose received  $1.20 \text{ mGy} = 40 \times 0.03 \text{ (mGy)} = \text{Dose (mGy)}$

Table 1. – TLD reader readings

Number of readings	1	2	3	4	5	6	7	8
Read the device	336	353	337	358	350	349	333	343

The average of these readings is 344.9

$$\text{New sensitivity} = 0.35197 = \frac{1.20}{0.04 - 344.9} \times 99.99$$

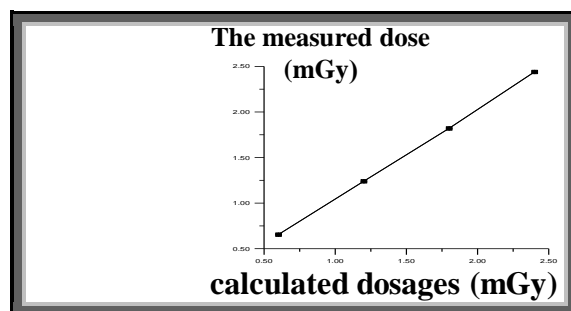


Figure 1. – The relationship between the known doses and the measured doses



Figure 2. – Image of a thermal flash device



Figure 3. – Thermal flash disk irradiation device

**Finding the Calibration Curve.** A particular arrangement of circles is taken and presented to various courses by the light gadget and the portion is determined in principle additionally, at that point it is estimated by the (TLD peruser) gadget, and the realistic relationship is drawn between the

hypothetically determined radiation dosages and the for all intents and purposes estimated portion by the TLD peruser, and Table 2 Figure 3 shows the measure of this portion.

Table 2. – Measured and known radiation doses of TLD device

calculated dosages(1) (mGy)	The average dose (mGy)	The measured dose (1) (mGy)	The measured dose(2) (mGy)
0.6	0.655	0.650	0.66
1.2	1.24	1.22	1.26
1.8	1.82	1.87	1.77
2.4	2.44	2.45	2.43

**Instruments of Radiation Level Measurement. Unfors Instrument.** This Swedish-made gadget comprises of a locator that has an ionization chamber associated with a covered wire associated with an advanced gas-filled counter. The gadget is intended for quality affirmation tests as it gauges the portion, time, portion rate and voltage straightforwardly. The most significant element of this meter is its little size, precision in estimation, and the figure (4) shows an image of the Unfors mobile device.



Figure 4. – Image of the Unfors mobile device

**Quality Assurance.** Quality Assurance Program: - It is a lot of methodology identified with the exhibition, hardware, and X-beam gadgets utilized in clinical diagnostics. It has built up quality confirmation programs and has been endorsed in the principles and laws of most nations on the planet and in the wellbeing models of global associations. The quality confirmation program means to: [6]

- 1 - Improve the picture of the different body tissue structures on the x-beam film or screen.
- 2 - Reducing radiation presentation to patients and laborers.
- 3 - Reducing costs (film utilization, brilliant cylinder life).

This program incorporates two primary lines:

First: Automated tests

Second: - Radiological tests

The primary line incorporates:

X-beam machine sturdiness, non-loss of gadget cutters, just as crafted by pointers, measures, working or cautioning lights, and heading of the X-beam shaft.

With respect to the subsequent line, it incorporates the accompanying tests, which will be clarified in detail:

**Check of Voltage kVp.** In this test, the difference in voltage between the two ends of the X-ray tube, which represents the anode voltage, is measured, because this voltage has an important relationship with the intensity of the X-ray coming out of the tube and the degree of image clarity.

The amount of voltage recorded on the control panel must be matched with the voltage measured by the devices; because the intensity of the rays The X-ray emitted from the device tube is proportional to the square of the peak voltages [6].

The voltages are measured using the digital scale directly, and the work of these devices depends on the use of two optical devices that convert the X-ray windows from two slices of copper of different thicknesses into a voltage difference that is proportional to the peak of the tube voltage after calibrating the device on such a measurement, and the device is usually placed at a distance of 100cm from The goal of the x-ray machine is exposed to x-rays, so the reading will be correct if the amount of change in voltage ranges between (10-25)%, and we need in the voltage test to two features:

- Precision
- Evidence.

**Linearity test mA.** The objective of the test is to set up tube current estimation. There are three strategies for estimation, the most significant of which are the accompanying:

Right now, top voltage is fixed (kVp = 70), the presentation time is fixed (0.4) seconds, the present changes unfailingly, and the introduction or portion is estimated a ways off of 100 cm. It ascertains the time pace of current (mAs) for each situation and afterward computes the coefficient of linearity, which ought not surpass 0.1 [6].

**Exposure Time Test.** Introduction time can be estimated straightforwardly utilizing the gadgets used to investigate the light emission, and such estimations rely upon the state of the radiation wave as the presentation time is set by the makers in a solitary structure, or joined with the current to gauge the sum (mAs). A turntable is utilized as a choice to time estimation for gadgets that are assessed by a full or half wave. Right now, number of spots on the x-beam movie gives time legitimately [6].

**Timer Linearity Test.** The point of the test is to gauge the exactness of the introduction time, and this is finished by utilizing a meter to quantify the portion or presentation and is set 100 cm away from the objective of the x-ray. Fixing both the pinnacle tube voltage and the normal time of the current and presenting to x-ray for introduction timethe average time takes the rate of each reading and plots Between exposure time on the x-axis and exposure (dose) on the y-axis [7].

**Optical and Radiation Field Congruence Test.** In this test the beam of light emitted from the lamp is identical to the X-ray beam and the reason for the mismatch is due to the deviation in the lamp location, the mirror and the anode focus point, and this test is done by taking a cassette inside the X-ray film and placed 100 cm from the target of the ray device X. The area of light falling on the film is organized by small L-shaped metal pieces placed at the edges of an area of 20 x 25 cm when the peak voltage (45kVp) and the current amount of current (6 mA.s) [7,6]. This type of test takes place every year or four months for each device, and there is another method for testing in which several beam boresight tests are used, which consists of a flat plate made of copper, which has length measures in the direction of x, y, positive and negative as in Figure (3). The test itself is repeated by placing this plate on top of the film, and then it tests radiation and light match.

**Beam Alignment Test.** The goal of this test is to know the integrity of the X-ray beam and its fall perpendicular to the film, when the radiation is not perpendicular, this leads to a blurring of the image, and for this the straightening test kit consisting of a 15-cm-long plastic cylinder is used to place the cylinder over the pointing kit on a table The patient is 100 cm away from the x-ray target as in Figure (5) and makes sure that it is leveled by a leveling balance and also by matching the light field with the dimensions of the film. The number and film were exposed to a x-ray beam with a voltage of (8 mAs, 55 kVp). After the film acidification process, the amount of deviation is known [6,7].

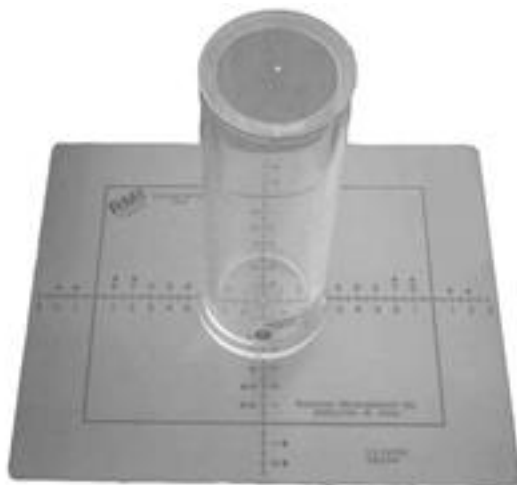


Figure 5. – Repayment Test Kit [6]

**Focal Spot Test.** The goal of the test is to measure the area of the X-ray focus, because the small area of the focus leads to the formation of an inconspicuous image, so the periodic test is very important to obtain a picture of good quality, and the measurement is done using several star, which consists of a disk that contains tapes that deviate from each other by an amount ( $1^\circ - 2^\circ$ ) and consists of lead 0.05 mm thick and arranged in a circular way. This plate is placed on the lancet and the ray film on the patient's bed and visualized by exposure to x-rays, and there is a lack of clarity in the image in the middle, including the focus area, shown in Figure 6 ).

**Filters Tes.** Filters are placed in front of the X-ray beam because it absorbs low energies that do not contribute to the medical diagnosis process and increases the patient's dose, because X-rays with low energies absorb into the skin according to the photoelectric phenomenon, and all their energies are deposited in it so the dose increases.

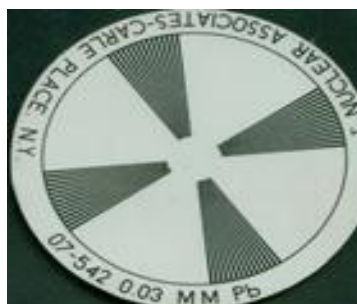


Figure 6. – Star Kit [6]

To test the thickness of the filter, it is necessary to measure the HVL (Half Value Layer), which is the thickness of the filter that absorbs half the intensity of the radiation falling on it. To carry out this test, a group of aluminum plates shall be taken with thicknesses (0, 0.1, 0.5, 1, 2) mm. By taking a portable meter to measure the radiation dose, it is placed 100 cm from the target of the x-ray, and the radiation field is paid to the x-ray to fall to the sensitive size of the counter, and the aluminum plates are placed one after the other in the middle of the distance between the meter and the target of the x-ray or by placing it on the diver after fixing the tube voltage on A certain effort, and taking the time amount of the current 20 millimeters. Secondly, from these measurements, a relationship is drawn between the thickness of the filter (linear scale) and the amount of the radiation dose (logarithmic scale). From the graph, we find the thickness in which the radiation dose decreases

to half its value without using the filter. The experiment is repeated for different values of peak tube voltage, and Table (3) shows the lowest thickness of half of the usual X-ray devices and those used in dentistry.

*Table 3. – The lowest half thickness of the aluminum filter in dental rays and other types of rays the lowest half thickness (mm) [7]*

<b>Other types of x-rays</b>	<b>Teeth rays</b>	<b>Peak tube voltage</b>
<b>0.4</b>	<b>1.5</b>	<b>40</b>
<b>0.5</b>	<b>1.5</b>	<b>50</b>
<b>1.3</b>	<b>1.5</b>	<b>60</b>
<b>2.1</b>	<b>2.1</b>	<b>70</b>
<b>2.3</b>	<b>2.3</b>	<b>80</b>
<b>2.5</b>	<b>2.5</b>	<b>90</b>
<b>2.7</b>	<b>2.7</b>	<b>100</b>
<b>3.0</b>	<b>3.0</b>	<b>110</b>
<b>3.2</b>	<b>3.2</b>	<b>120</b>
<b>3.5</b>	<b>3.5</b>	<b>130</b>
<b>3.8</b>	<b>3.8</b>	<b>140</b>

**Conclusions.** *Features of X-Ray Instruments/* Work was done on X-ray devices in the hospitals (Ali Ibn Abi Talib, Al-Sadr, Ibn Al-Baladi) located in Sadr City, and there are (6) devices. Table (1) shows the type of devices, the date of manufacture, and other information about the X-ray devices, and by means of a radiation dose measuring device (unfors).

*Table 4. – Specifications of the X-ray Devices*

<b>The distance between the film and the SID ray tube</b>	<b>The manufacturer and its history</b>	<b>mAs</b>	<b>Voltages kVp</b>	<b>Type of test</b>	<b>the hospital</b>
<b>130</b> <b>130</b>	Shimadzu 1980	25 35	<b>70</b> <b>85</b>	chest	A
<b>85</b> <b>85</b>	Toshiba 1980	30 35	80 85	Lumbar spine	B
<b>80</b>	Sedecal 2004	25	90	Abdomin	C
<b>60</b> <b>60</b>	Shimadzu 1989	25 <b>25</b>	75 90	Lumbar spine	D

**Radiation doses resulting from diagnostic tests.** In this study, new devices with different origins were used. The radiation dose was measured for several patients and different parts of the body, and it is clear from Table (5) that:

A- The ESDpp values for the patient's presence and measured using thermal flash (TLD) are greater than the values measured by the Swedish or German ionization chamber, which measures the dose in the absence of the patient (ESDpa) because the scattered rays in the patient's body lead to increased Radiation dose In Table (5), the dispersion factor resulting from the patient's body was

calculated, which is the ratio between the dose measured by thermal flash tablets (TLD) to the average dose measured by the two ionization chambers, and it was found that this value ranges between (1.05 - 1.1) and is a good approximation of the values The calculated world value of (1.09).

B- The surface entry dosage values for the Spanish-origin X-ray apparatus with the symbol (E) are close to the internationally approved values, because the device is newly imported and is still under maintenance by the company that regularly performs its work performance tests [1, 2, 3].

C- As for the measured dose values for the X-ray machine with the symbol (A), it is about twice the dose recommended globally, because the device is a quarter of a century old and did not undergo quality assurance tests previously, and these measurements show that the device is not fit for measurement, so the hospital administration uses it in a way Little and in the case of great momentum for patients.

D- The surface entry dose values for the X-ray machine with the symbol (F) are close to the reference values even though it is 10 years old, and this is due to the fact that it is a private sector device that maintains it well, in addition to that the radiographer is experienced in this field.

E- The surface entry dose values measured by the TLD are closer to the reference values, because the sensitive size of this meter is a substance equivalent to the live tissue, and it is exposed to X-rays, so the reading value of the dose in the air is not the same as in the case of the ionization room (Unfors ) Swedish, but rather the dose of live tissue, which is close to the dose of patients.

F-The average dose measured in this research for all devices in this study is an approximation to the reference dose of the International Atomic Energy Agency (IAEA), but it is generally greater than the reference doses with a range of those from (1.2 - 1.7) times, as in Table (6), although this The increase does not constitute a line for patients, but it is preferable that the dose be less than that according to the ALARA Law in radiation protection, because the increase in the dose is not justified and this increase is due to the following:

- \* Introduced some devices to work in this study, and one of them has been used for 25 years continuously.

- \* Most radiographers of these devices lack expertise and good training on these devices.

- \* The absence of a program for quality assurance measurements for these devices in Iraq, and there is no great interest in the radiation risks of patients, and the only important thing is to obtain a picture of good quality that the doctor can diagnose easily, because the radiographer uses a few kVp, and high exposure (mAs) to The image is good, formed to be the opposite by using high kVp and few (mAs) to have a low dose and good image quality.

G-Most of the films used in the measurement are of poor quality and require a high exposure time to obtain a good image.



Table 5. – Surface entry dosages measured by thermal flash disks and portable devices, and their comparison with the International Atomic Energy Agency (IAEA) values and the dispersion coefficient indication (8)

The device	Type of test	Dose mGy		Dispersion coefficient	Values IAEA mGy
		TLD	Unfors		
A	chest	± 1.95 ± 0.1	1.85 ± 0.1	1.05 1.06	A
B	Lumbar spine	± 1 14.8 1.32±41	± 1.1 13.65 2.1±38	1.08 1.07	B
C	Abdomin	± 1.2 18.1 ± 1.3 48.8 0.6±8.2	± 1.45 16.9 ± 1.8 45.4 1.2±7.85	1.07 1.07 1.05	C
D	Lumbar spine	59 ± 2 0.8±23	± 3.5 54 1.5±21	1.09 1.09	D
E	Skull	± 0.2 0.73 ± 0.65 5.4 ± 1.2 46.3 0.63±14.6	± 0.3 0.68 ± 0.5 5.1 ± 1.2 43 0.6±13.5	1.07 1.06 1.08 1.08	E

Table 6. – Radiological doses measured by radiography of a group of devices in a hospital Iraq in current research compared to global potions [9]

Type of test	معدل الجرعة mGy	Values IAEA mGy	The measured dose IAEA
chest	0.63 1.85	0.4 1.5	1.6 1.2
Lumbar spine	17.3 45.1	10 30	1.7 1.5
Skul	7.17 5.1	5 3	1.4 1.7
Abdomin	14.1	10	1.4

**Quality assurance tests. Voltage Stability Test.** The test of voltages is acceptable if the standard deviation in the measurement is less than ± 5% or if the fluctuation in its measured and recorded values is acceptable if the percentage of fluctuation ranges between (3 - 10)% for devices with three phases and is used in this research, and the oscillation constant is measured From the following law [8]:

$$\text{Volatility\%} = \frac{\text{measured (kVp)} - \text{recorded (kVp)}}{\text{measured (kVp)}} \times 100\%$$

Table (7) shows the amount of fluctuation in the measurement of voltages when the current is fixed (200) mA for device (A) is 18.4%, which is higher than the specified values, and also the

standard deviation in the measurement is greater than 5%, that is, the device does not meet warranty specifications Quality (QA), as for the two devices (F, C), the volatility value in the voltages is 3.58% and 5.56%, respectively, under the same conditions, which is less than the limits set for this fluctuation, which is 10%, and the standard deviation values are less than 5%.

**Current linearity test.** It is noted from Tables (3), (4) and (5) that the dose values measured by TLD disks and the two portable devices are very close within experimental and statistical errors, and that the Swedish device values are closer to them than the global values for the measurement accuracy in this device, and that the relationship is linear for the devices (E, C) and non-linearity of the device (A). For the purpose of finding the linear coefficient, the amount of linear coefficient was calculated from the law [9]

$$\text{linearity coefficient} = \frac{(\text{Maximum dose / mAs})_{\text{super}} - (\text{Maximum dose / mAs})_{\text{lower}}}{(\text{Dose / mAs})_{\text{super}} + (\text{dose / mAs})_{\text{lower}}}$$

It was found that the linear coefficient of the device (A) equals 0.2 which is greater than the largest specified value which is 0.1, and for the other two devices (E, C) they are within the specified values (0.07, 0.01) because it is less than 0.1 as shown in Table (7).

Table 7. – they are within the specified values (0.07, 0.01) because it is less than 0.1 as shown in Table (7).

kVp = 90		distance = 100cm			
Linear coefficient	average mAs	The average Max dose	average mAs	average Min dose	The device
0.2	160	14.47	9.8	1.345	A
0.01	140	20	8	1.163	C
0.07	28.8	10	10.3	1.433	F

**Time stability test.** It is noted from Table (8) that the percentage of time stability does not exceed 10% for the two devices (E, C) by applying the law of linear coefficient while it reaches 45% for the device (A), meaning that the time stability of the device is very bad.

**Test half for the filter.** From Table (8) it is clear that the thickness of the filter ranges between (0.62 - 2.5) mm and that the approved values for the half thickness should not be less than 1.8 mm when the kVp = 70)) For the device (A), the half thickness of the filter is less than the limits set [2, 3, 4]. Therefore, this device does not apply to quality assurance contexts. As for the two devices (E, C), the values are within the limits set, especially for the device (C).

**Package integrity test.** We see from Table (8) that the alignment is not good, because the X-ray beam is about 3 ° away from the vertical axis of the device (A), while we notice from Table (8) that the straightness is good, because the image of the beam beam is symmetrical around the point of intersection of the two axes and that the beam of rays The x-ray is about 1.5 ° from the horizontal vertical intersection of the device (C).

**Examination of the light and radiation fields.** It is noted from Table (8) that the congruence between the optical and radiation fields of apparatus A) is not good, but it is within the limits set [2, 3, 4] which is 2% of the distance between the source and the film, i.e. 2 cm in this case, but the convergence is good for the device (C).

**Radiation focus area test.** Note from Table (8), the focus area of the X-ray machine (A) is (2.26 mm \* 2.44mm), and it is slightly larger than the established limits, as can be seen from Table (8), the focus area of the X-ray machine (C) is (1.90 mm \* 1.98 mm), which is within the limits set [2, 3, 4]

Table 8. – Quality Assurance Tests for the Three X-ray Devices and their Comparison

with International Values [10]

Values IAEA	device (E)	device (C)	device (A)	Tests
10 %	5.56 %	3.58 %	14.8 %	kV <sub>p</sub>
0.1	0.07	0.01	0.2	Linear current
0.1	0.049	0.01	0.41	Linear time
mm 2.1 ≤	mm1.8	mm2.15	mm0.62	Half thickness
° 1.5 ≥	°1.5	°1.5	°3	Straightening the package
cm2 ≥	cm0.6	cm0.3	cm0.8	Congruence
1.75mm 2.25 mm	————	1.90 mm 1.98 mm×	2.26 mm 2.44 mm×	Focus area

In general, as shown in Table (8), the quality assurance tests indicate that the devices (E, C) have succeeded in these tests and they work within the international specifications, while the device (A) failed in most of the tests, therefore it should not be used for examination because it Patients are exposed to high doses, and images are not good, so radiographs will repeat imaging, exposing patients to additional unexplained doses. It is noted that:

1- The doses measured in this study are higher than the global doses due to:

A - Focusing on one device and not others in many tests.

B - Lack of experience for radiographers.

C - Not implementing quality assurance programs.

D - Change (kVp and mAs) to get a good image regardless of overdose.

E - Not taking into account the thickness of the patient by radiographers to choose the appropriate kVp and mAs workers, and the process is done in an estimated manner.

2- The use of poor films, which calls the doctor to restore the x-ray again.

3 -The X-ray machine (A) does not meet the required technical specifications because it did not pass the quality assurance tests.

### *References*

- [1] Rehani, M.M. Diagnostic imaging Quality Assurance. 1999.
- [2] ICRP. Publication 61. 2004.
- [3] Coggle. J.E , 2005. Biological Effect of Radiation New York.
- [4] IAEA and WHO. Calibration of dosimeters used in Radiotherapy. Technical Report No. 3742008.
- [5] IAEA. 2000. Technical Reports series No. 398 Absorbed dose determination in external beam radiotherapy.
- [6] Shrimpton, P.C, Wall, B.F and Hart,. Diagnostic Medical Exposures in U.K. Applied Radiation and Isotopes 50: 261-269. 2011.
- [7] Lavole, C, Don, C. Measurement method of radiation dose during barium enema examination. The British Journal of Radiology 70: 291- 295. 2014
- [8] Shrimpton, P.C, wall, B.F and Harrison , R.M. Dose to patients from routine diagnostic x-ray examination in England. British Journal of radiology 57:749- 58. 2016
- [9] Stanizeweka, M.A. .x-ray diagnostics as the main source of ionizing radiation for the polish population . Journal Radiological protection vol.3 No.4: 275- 278. 2017

UDK 718.66.58

## A VARIETY OF VIRTUAL REALITY IMPLEMENTATIONS FOR CREATIVE LEARNING AND 5 WAYS TO IMPLEMENT VIRTUAL REALITY IN THE LEARNING PROCESS



**S.N. Ibodullayev**

*Magistr of Tashkent university  
of information technologies named  
after Muhammad al-Khwarizmi*



**A.A. Bahromov A.A.**

*Teacher of Tashkent university of  
information technologies named after  
Muhammad al-Khwarizmi.*

*Department of Audiovisual technologies, Tashkent University of information technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan  
E-mail: Sniper\_2112@mail.ru, Pure.life.jojo@gmail.com*

### **S.N. Ibodullayev**

*Magistr of the Department of Audiovisual technologies, Tashkent University of information technologies named after Muhammad al-Khwarizmi. Author of 39 research papers on audio, video, multimedia resources, VR and AR and media education technologies.*

### **A.A. Bahromov**

*Teacher of the Department of Audiovisual technologies, Tashkent University of information technologies named after Muhammad al-Khwarizmi. Author of 10 research papers on audio, video, multimedia resources, VR and AR and media education technologies.*

**Abstract.** Prior studies on the use of digital prototyping and virtual reality (VR) in designing as well as evaluating new products have shown that VR reduces both development time and costs whilst augmenting student motivation and creativity. The current study demonstrates that VR and 3D prototyping in the context of project-based learning (PBL) promote effective communication, increase problem solving skills, and enhance learning outcomes. VR and digital prototyping have been extensively used in industries for the purpose of product design and usability evaluation. In the context of engineering education, many research studies have attempted to explore the effect of VR on teamwork, engagement, retention, and motivation. In this paper, VR is used in conjunction with PBL in self-directed approach to design and implement a product using 3D software whilst also using virtual reality immersive CAVE display to evaluate their design. The hypothesis is that the use of VR with a project-based-learning approach to facilitate the attainment of desirable goals in the engineering design project, improved achievement of course learning outcomes and promoted effective communication. According to the research findings, VR approach significantly affected the distribution of cumulative project grades. Students' project grades improved, particularly the implementation component. In addition, the course outcomes related to project design were better achieved in VR approach. The communication and problem-solving skills were improved in the VR approach as compared to traditional approach.

**Keywords:** modeling types, 3D model, VR and AR, education, Science, engineering, distance learning.

**Introduction.** Virtual reality (BP) (lat. Virtus - potential, possible; lat. Realis - real, existing) - the world created by technical means on any substrate and transmitted to the subject through his usual sensations of the material world: sight, hearing, smell, and others. Also abnormal reality, imperfect, not meeting certain criteria of truth. The concept of artificial reality was first introduced by Myron Krueger in the late 1960s. In 1989, Yaron Lanier introduced the more popular now virtual reality. In the fantastic subgenre literature, cyberpunk, virtual reality is a way for a person to communicate with "cyberspace" — a kind of medium of interaction between people and machines created in computer networks.

The first virtual reality system was the Aspen Movie Map, created at the Massachusetts Institute of Technology in 1977. This computer program simulated a walk through the city of Aspen, Colorado, giving the opportunity to choose between different ways to display the terrain. Summer and winter options were based on real photos. Currently, virtual reality technologies are widely used in various areas of human activity: design and design, mining, military technology, construction, simulators and simulators, marketing and advertising, the entertainment industry, etc. The market volume of virtual reality technologies is estimated at 15 billion dollars a year.[1]

Moreover, the possibilities provided by Virtual Reality (VR) Technology, combined with the complex intrinsic properties of architectural pedagogy, place this technology under architecture researchers' constant watch. With its experiential nature, VR technology can improve architectural students' learning. Although the study of VR applications for educational purposes is not new, this is rarely studied in the light of emerging learning theories in architectural education. In response, an educational application called LADUVR ("Learning Architectural Details Using Virtual Reality Technology") has been designed by the authors to show how VR would address the current shortcomings of architecture learning systems.

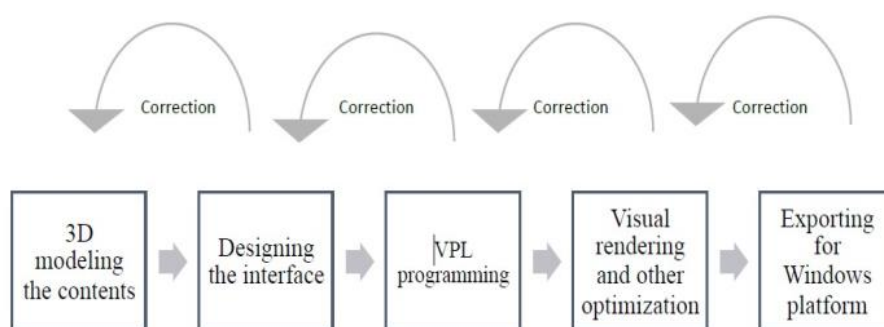


Figure 1. –The process of making LADUVR.[2] Main part. Applications for VR and AR in Education Medical Education

Numerous surgery trainers and simulators employ virtual reality, including laparoscopy (Huber et al., 2015), temporal bone surgery (Fang et al., 2014), and even dental training (Steinberg et al., 2007). Some of these VR applications give haptic (i.e. tactile) feedback and they all allow students to practice their skills in a safe environment and without the expense of practicing on human or animal cadavers. Furthermore, virtual reality has been used to help medical students visualize anatomy in 3D, providing a much greater sense of context and scale than the cutaway diagrams and pictures common to anatomy textbooks (Satava, 1995; Falah et al., 2014).

Science.

Early uses of virtual reality in science education focused on visualizing chemical reactions (Bell and Fogler, 1998) or learning about molecules by assembling them in a virtual environment (Byrne, 1996). More recent uses include marker-based augmented reality to visualize the process of respiration and human meiosis (Weng, 2016) and an astronomy application using a headmounted display to explore the solar system and give students a grasp for its scale (Hussein and Nätterdal, 2015). Virtual reality and augmented reality make it possible to visualize concepts that are abstract or difficult to relate to real-world experiences, such as a marker-based augmented reality application that helps teach electromagnetism and the interaction between different circuit elements (Ibáñez, 2014).

Engineering.

A variety augmented reality apps have been developed and tested in introductory electrical engineering courses (Martín-Gutiérrez et al., 2014). ElectARManual displays animations and instructions over electrical machines used in the lab to help students learn how to use the machines safely. ELECT3D is a markerless system that reads and interprets electrical diagrams. A third application is called ElectAR\_notes, which is a study assistant that recognizes markers located on the course study notes and illustrates the concepts with video, animations, and more detailed information.

Another study developed a virtual reality application to teach microcontrollers and Arduino boards with Google Cardboard headsets (Ray and Deb, 2016).

#### Distance Learning

The Internet has made distance learning far more accessible and rich in content than ever before, but in many cases the only forum for discussion and interaction with classmates is through online message boards or e-mail. Virtual reality can improve distance learning by allowing easier and more natural class discussions in the distance learning setting. The simplest examples are giving lectures in a virtual classroom, such as in Second Life (Jarmon, 2009). [3] Another misconception is related to the opinion that BP causes the effect of "motion sickness" among users. This effect is due to lack of synchronization between the image that transmits helmet, and the data that the user receives from his body. This happens when the brain expects to receive one image (according to corresponding to the normal order of things, following its cognitive model), and gets another. Such flaws were the result of the imperfections of the first helmets and poor-quality content. At present, there is no such problem. It is widely known how dangerous radiation is from modern media devices (phones, computers, televisions).[4]

Adjacent to the last misconception is that BP kills live communication and alienates people from each other. But the same was said about mobile phones, SMS-messages, social networks. However, those who consider the use of new technologies to escape from reality forget about their advantages: mobile communication makes it possible to connect with a person on the other side of the planet, and social networks can enrich remote communication with various types of content and activities. Virtual reality technology can be applied in the field of social interaction. Using VR helmets will soon be possible to communicate with friends in one virtual space, for example, on the river bank or in the forest. In this case, the picture will be clear and believable, and the movements of the head and the sound of the voice will create an effective illusion of live communication.[5]

#### Educational

The effect was low, only 20% of students were able to recall the information provided for memorization in vAcademia. Therefore, the use of vAcademia for conducting such forms of study as a lecture and a seminar (most common in the educational process) seems to be low productive, unlike a webinar.

Preparing a teacher for a lesson in virtual reality requires much more time than to the webinar. In addition to designing a new location or studying the characteristics of a typical location, the teacher has to think through alternative scenarios for the development of a lesson due to the possible unplanned behavior of students. "Pedagogical situations" in a virtual 3D environment.

There is much more happening than in a real audience or webinar. The teacher's avatar is very often not associated by students with the personality of the teacher, which creates an atmosphere of excessive informality and freedom of expression during the lesson. In conclusion, there are several points that should be paid attention to when designing VRcontent for education. The technology brings positive results when used in short sessions or in the form of simulators and simulators. It is inappropriate to use BP for lectures and seminars. When developing software solutions, you should focus on the latest models of BP equipment, the most environmentally friendly and ergonomic for users.[6]

**The calculation of generalized evaluations.** The main disadvantage of the virtual caves is the high cost, which makes the use of this type of immersive VR limited. On the other hand, the use of an HMD can often cause some level of cybersickness. In addition to, I will show working scheme of VR in education system.

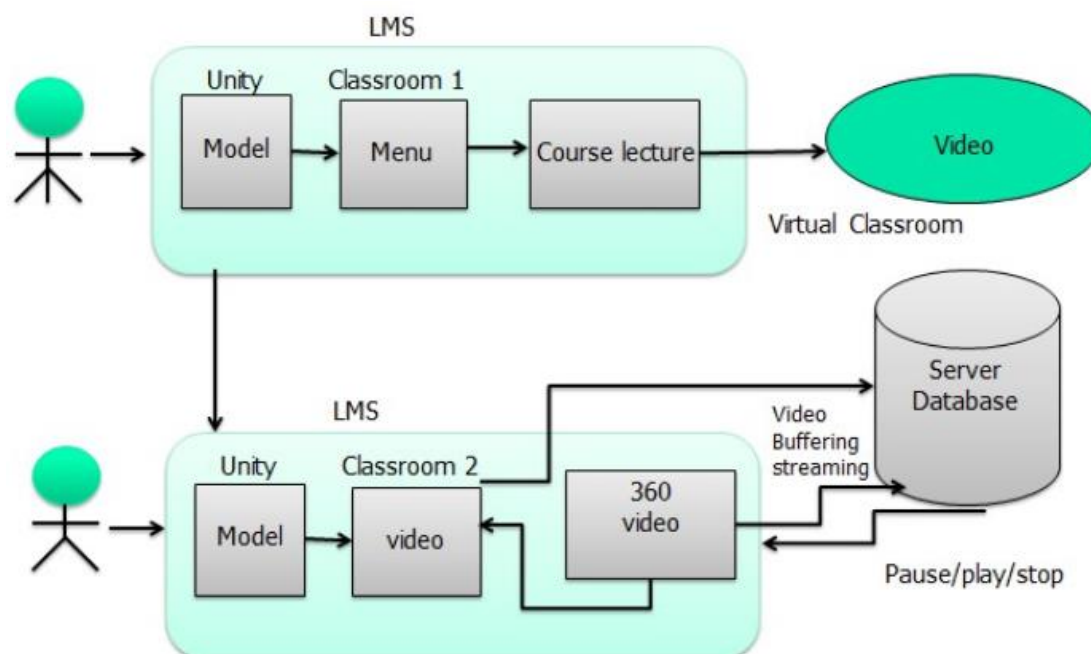


Figure 2. – Viewing of Rough idea integration of applications VR with most common LMS in authorization part and transfer of user activity data from VR to LMS.

My workshop took place in Second Life on an island owned Tashkent University of Information Technology in Uzbekistan. Students at the university made virtual buildings with virtual classrooms for e-learning activities. The authors have already carried out lessons and projects with students in these classrooms. These teachers were not familiar with the virtual world, so they needed to learn about it and how to create, register, and use an avatar prior to the official workshop (which was a great success). During a test run, some of the teachers experienced technical problems with their computers, microphones, etc. The instructor met the teachers in a virtual classroom that contained red chairs, tables, a podium, and a poster attached to a whiteboard. She presented five creative teaching techniques and discussed them in detail with the group.

Students can participate in virtual world activities with their teacher. They can take virtual tours of museums, different countries, etc. For young student, a teacher can project the virtual world on a large screen for the entire class to use and enjoy. In Second Life, students are able to experience sunrise, midday, sunset, and midnight.

As universities continue to increase their online presence, they are challenged to reevaluate their ways of teaching to best utilize the new opportunities that digital technologies offer. While these technologies present a variety of methods for online collaboration, virtual worlds offer the added ability to interact within a realistic graphical environment, situating learners together in the same space and time. These features give participants a highly social and immersive environment within which they can represent their role-play characters through avatars.

### ***How To Implement Virtual Reality In The Learning Process***

While VR adoption has been a bit slow across the board, educators are certainly finding some nifty ways to use this technology within the classroom. Used correctly, this technology can be employed by teachers to create some innovative learning opportunities for students. Anyone thinking of using VR in the classroom should consider one of these 5 creative applications.

#### **1. Immersive Language Learning**

Language learners have long touted immersion as an ideal way to help students master languages. Rather than simply learning through typical classroom experiences and instruments like lectures and worksheets, teachers often prefer to use experiential learning so that students ‘live’ the language rather than simply studying it. One example of this is opting to only speak the language

being learned in the classroom. Other teachers use cultural experiences. In some cases, language learning classes even culminate with exchange trips to other countries. As with any other educational experiences, the ability for teachers to engage in immersive language teaching is limited by resources and geography. Fortunately, there are several language learning apps developed to combine language with VR to create these experiences. One example of this is Mondly VR . Using this app, students are able to practice conversational language skills from within a series of realistic simulations. Students can also use VR to interact with students from other countries and go on virtual sightseeing tours.

#### 1. Creating Better Learning Opportunities For Underserved Populations

Some educators have the fortune to be able to create unique and enriching learning experiences by taking students on field trips. For others, that's a bit more difficult. They may be serving students who are:

1. Disabled to the extent that field trips are not optional
2. Homebound, institutionalized, or hospitalized
3. Living in remote areas

In addition to this, many educators simply work in school districts that don't have the resources that allow them to reserve school buses, cover admissions fees, or any other costs relating to traditional field trips.

This is unfortunate because in many instances the students who could most benefit from enriching experiences don't have access to them. Fortunately, teachers can use VR to give these student populations access to places they wouldn't ever have otherwise. Here are some examples of VR 'field trips' students can take:

1. Visiting and exploring national parks
  2. Touring museums
  3. Exploring historical locations and monuments
  4. Discovering remote regions such as Antarctica or Sub Saharan Africa
- #### 3. Giving Students The Opportunity To Create With VR

Students don't need to simply consume VR content. They can also spend time using VR technology to create. For example, Google's Tilt Brush tool allows users to create three-dimensional art in VR. The tool provides students with a variety of brush types and palettes. Students can then wear VR headsets and create beautiful art including characters, landscapes, skylines, and more.

That's not all. Coding in the classroom isn't new, but VR coding is. Just like students have learned coding skills to build games, apps, even robotics programming, now they can be taught to apply those skills to VR. Students can create VR apps and then test them out in the classroom.

#### 4. Modeling And Simulations

Imagine a drivers ed student driving a route over and over again without spending any money on gas. Now, picture an architectural engineering student testing different building materials to create a walkway that spans a city block. What about a squeamish student dissecting without complaint?

With Virtual Reality, students can use modeling to understand and test a variety of concepts and theories in engineering, math, and science. VR simulations can provide students with important experiences as well as providing needed information for ongoing assessments. In addition to this, once the original cost of hardware and software is covered, other expenses are minimal.

#### 5. Developing Empathy Within Students

A great number of students have never experienced war where they live. They haven't lived in abject poverty or with little to no resources. Many have never experienced a devastating natural event. Putting major tragedies aside, many students simply have not had experience of exposure to different cultures, traditions, or belief systems.

Current events curriculum, pen pal programs, and other efforts can help. Still, their impact is limited. Fortunately, VR can be used to put students in places they have never been and in contact with people they may not otherwise meet. Imagine students 'sitting' in a classroom in Kenya,



navigating the daily search for clean water in Flint Michigan, or being taken on tour in a country where there is currently political unrest. These experiences can increase understanding, and develop a sense of empathy.

### **Conclusion.**

We will review in this example model of University and there are students choose their subjects where they learning his timetable. Students can now experience the topics they are learning about. Use of virtual reality technology has been shown to increase student engagement and focus, while the immersive and interactive environment encourages the students to become active learners.

Using 3D models for educational environments is one of the most actual topics not only in our country, in abroad also. These are just a few ways that VR can be used in the learning process. The possibilities seem endless. Different combinations of classrooms and educational goals can lead teachers to conceive their own uses for this technology. In any case, students are guaranteed to come out ahead as long as educators are committed to using VR to its full potential.

### **Reference**

- [1]. Таев В.А. Контроль, оцифровка и обработка изображений сложных пространственных деталей в машиностроении. Методические указания. Владимир, 2015.
- [2]. Sayyed Amir Hossain Maghool, Seyed Hossein (Iradj) Moeini, Yasaman Arefazar. An educational application based on virtual reality technology for learning architectural details: challenges and benefits Archnet-IJAR: International Journal of Architectural Research DOI: <http://dx.doi.org/10.26687/archnet-ijar.v12i3.1719>.
- [3]. Brian Boyles. Virtual Reality and Augmented Reality in Education a 2- year faculty professional development program conducted by the Center for Teaching Excellence, United States Military Academy, West Point, NY, 2017.
- [4]. Селиванов В.В., Селиванова Л.Н. Виртуальная реальность как метод и средство обучения // Образовательные технологии и общество. 2014. Т. 17, № 3. С. 378–391.
- [5]. Севастьянова Ю. О дивный новый мир: пять заблуждений о виртуальной реальности [Электронный ресурс] // Понедельник – ежедневный интернет-журнал об образовании, карьере и бизнесе. [Б. м.], 2016. URL: <http://ponedelnikmag.com/post/o-divnyy-novyy-mir-pyat-zabluzhdeniy-o-virtualnoy-realnosti>.
- [6]. Nuraliyev F.M., Giyosov U.E., On the design of virtual reality environments in education, VIII International Scientific, Technical, Scientific and Methodological Conference "Actual Problems of Information Telecommunications in Science and Education", February 27- 28, 2019, Russia.
- [7]. Ibodullayev Sardor Nasriddinovich., Giyosov Ulugbek Eshpulatovich., A variety of virtual reality implementations for creative learning, International Scientific Conference «Global Science And Innovations 2019: Central Asia» Nur-Sultan, Kazakhstan, May 2019.

## **РАЗНООБРАЗИЕ РЕАЛИЗАЦИИ ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ ДЛЯ ТВОРЧЕСКОЕ ОБУЧЕНИЕ И 5 СПОСОБОВ ВНЕДРЕНИЯ ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ В ПРОЦЕСС ОБУЧЕНИЯ**

***Ибодуллаев Сардор Насриддинович***  
*Магистр ТУИТ*

***Бахромов Асрор Ахрор ўгли***  
*Преподаватель ТУИТ*

*Кафедра аудиовизуальных технологий, Ташкентский университет информационных технологий имени Мухаммеда аль-Хорезми, Ташкент, Узбекистан.*  
*E-mail: Pure.life.jojo@gmail.com, Sniper\_2112@mail.ru*

**Аннотация.** Предыдущие исследования по использованию цифрового прототипирования и виртуальной реальности (VR) при проектировании, а также при оценке новых продуктов показали, что VR сокращает как время разработки, так и затраты, одновременно повышая мотивацию и творческий потенциал учащихся. Текущее исследование демонстрирует, что виртуальное и трехмерное прототипирование в контексте проектного обучения (PBL) способствует эффективному общению, развитию навыков решения проблем и улучшению результатов обучения. VR и цифровое прототипирование широко используются в промышленности с целью разработки дизайна продукта и оценки его удобства. В контексте инженерного образования многие научные исследования пытались исследовать влияние виртуальной реальности на командную работу, вовлеченность, удержание и мотивацию. В этой статье VR используется совместно с PBL в самостоятельном подходе к проектированию и реализации продукта с использованием программного обеспечения 3D, а также с использованием иммерсивного дисплея CAVE в виртуальной реальности для оценки их дизайна. Гипотеза заключается в том, что использование виртуальной реальности с подходом, основанным на обучении на основе проектов, облегчает достижение желаемых целей в проекте технического проектирования, улучшает достижение результатов обучения по курсу и способствует эффективной коммуникации. Согласно результатам исследования, подход VR существенно повлиял на распределение совокупных оценок проекта. Улучшены оценки студентов, особенно компонент реализации. Кроме того, результаты курса, связанные с дизайном проекта, были лучше достигнуты в подходе VR. Навыки общения и решения проблем были улучшены в подходе VR по сравнению с традиционным подходом.

**Ключевые слова:** типы моделирования, 3D модель, VR и AR, образование, наука, инженерия, дистанционное обучение.

UDK [611.018.51+615.47]:612.086.2

## COOPERATIVE MULTI-THREAD SCHEDULER FOR SOLVING LARGE-SIZE TASKS ON MULTI-CORE SYSTEMS



**O.N. Karasik**

*Tech Lead at ISsoft Solutions  
(part of Coherent Solutions) in  
Minsk, Belarus, PhD in  
Technical Science*



**A.A. Prihozhy**

*Professor at the Computer and System  
Software Department,  
Doctor of Technical Sciences,  
Full Professor  
Belarusian National Technical University*

*ISsoft Solutions (part of Coherent Solutions), Belarus  
Belarusian National Technical University, Belarus  
E-mail: karasik.oleg.nikolaevich@gmail.com, prihozhy@yahoo.com*

### **O.N. Karasik**

*Tech Lead at ISsoft Solutions (part of Coherent Solutions) in Minsk, Belarus; PhD in Technical Science (2019).  
Interested in parallel computing on multi-core and multi-processor systems.*

### **A.A. Prihozhy**

*Full professor at the Computer and system software department of Belarusian national technical university, doctor of science (1999) and full professor (2001). His research interests include programming and hardware description languages, parallelizing compilers, and computer aided design techniques and tools for software and hardware at logic, high and system levels, and for incompletely specified logical systems. He has over 300 publications in Eastern and Western Europe, USA and Canada. Such worldwide publishers as IEEE, Springer, Kluwer Academic Publishers, World Scientific and others have published his works.*

**Abstract.** The architecture of a cooperative multi-thread scheduler for thread execution on multi-core systems that run Windows is proposed. The architecture is implemented using the User Mode Scheduling (UMS) mechanism, which allows the user application to organize cooperative thread execution. The architecture under development includes the necessary set of components: a user thread for executing user code; new synchronization primitive for organizing the interaction of user threads running on different cores of a multicore system; a control transfer mechanism between user threads running on the same core. The architecture allows the programmer to implement cooperative multi-threaded algorithms to accelerate the solution of large-scale problems on multi-core systems.

**Keywords:** multi-threaded application, multi-core system, cooperative multi-tasking, scheduler, large-size tasks.

*Introduction.* Solving large-scale problems in reasonable time is impossible without exploring parallelism of modern parallel systems. Effectiveness of the parallelization depends on the existence of a parallel solution for the problem, on the possibility to develop an effective parallel algorithm for the problem, and on the availability of hardware and software tools, which meet the requirements of the parallel algorithm under implementation [2]. Nowadays, a wide range of tools for the development of parallel applications exist: Windows API, OpenMP, Cilk Plus, Portable Operating System Interface Threads (POSIX Threads or PThreads), Threading Building Blocks (TBB), Open Computing Language (OpenCL), different implementations of the Message Passing Interface (MPI), and others [2-8]. Choosing the appropriate tool depends on various aspects, including the operating system compatibility, the availability (licensing, freeware or shareware), and the specifics of the

parallel problem. In this paper, we investigate the effectiveness of the cooperative multi-thread scheduler implementations [9] on two different parallel applications: the Gaussian Elimination algorithms [10, 11] for solving algebraic equations, and the parallel Floyd-Warshall algorithms for all-pairs shortest paths problem [11 – 14].

*Cooperative multi-tasking in Windows.* All modern versions of the Windows operating system use a preemptive scheduler as the default thread scheduler [15]. However, starting from Windows 7 (for workstations) and Windows Server 2008 R2 (for servers), Windows provides a User-Mode Scheduling (UMS) interface, which allows the execution of an application by operating system scheduler in such a way as to construct a custom thread scheduler [16]. UMS interface consists of three core entities: a worker thread – *UmsWorkerThread* (UMSWT), a scheduler thread – *UmsSchedulerThread* (UMSST), a signature of the scheduling procedure, which is invoked by operating system in certain circumstances (described below) – *UmsSchedulerProcedure* (USP) and a completion list of worker threads – *UmsCompletionList* (UMSCL).

*UmsSchedulerThread* is responsible for execution of UMSWT. It is implemented using standard operating system thread (executed by operating system scheduler), which is switched to scheduling mode by the invocation of an *EnterUmsSchedulingMode* procedure. The parameters of this procedure are a pointer to UMSCL and a pointer to an implementation of USP. The operating system invokes the provided implementation of USP in the cases as follows: at the initialization of UMSST; immediately after the call to *EnterUmsSchedulingMode* (using the *UmsSchedulerStartup* event); after blocking UMSWT at the system call (*UmsSchedulerThreadBlocked* event); when UMSWT passes the control to UMSST (*UmsSchedulerThreadYield* event). USP is also responsible for handling operating system callbacks, and for maintaining a list of all UMSWT created by the application. UMSST executes UMSWT using the *ExecuteUmsThread* procedure.

*UmsWorkerThread* is responsible for the execution of a user code. It is implemented with the standard operating system thread, which is transformed into UMSWT by setting up a set of attributes at the moment of creating. These attributes include pre-allocated *UmsContext* (used by the operating system and created using the *CreateUmsThreadContext* procedure), and a pointer to UMSCL. At the initialization, operating system pushes UMSWT to UMSCL previously specified as attribute. Starting from this moment UMSWT is under the control of UMSST bound to UMSCL (cooperative multi-threading). UMSWT can pass the control to UMSST using the *UmsThreadYield* procedure.

Figure 1 illustrates the execution of two operating system threads  $T_1$  and  $T_2$  under the control of operating system scheduler, the execution of one user-mode scheduling thread  $UMSST_1$ , and the execution of two user-mode scheduling worker threads  $UMSWT_1$  и  $UMSWT_2$ .

*Architecture of cooperative multi-thread scheduler.* The cooperative multi-thread scheduler consists of three core components: scheduler thread (CST), user thread (CUT) and synchronization primitive (CSP).

The scheduler implements:

- the memory management, in particular, aligned memory allocation, NUMA aware memory allocation, and memory buffering;
- procedures for creating / terminating CST, CUT and CSP as well as procedures for interaction between these core components.

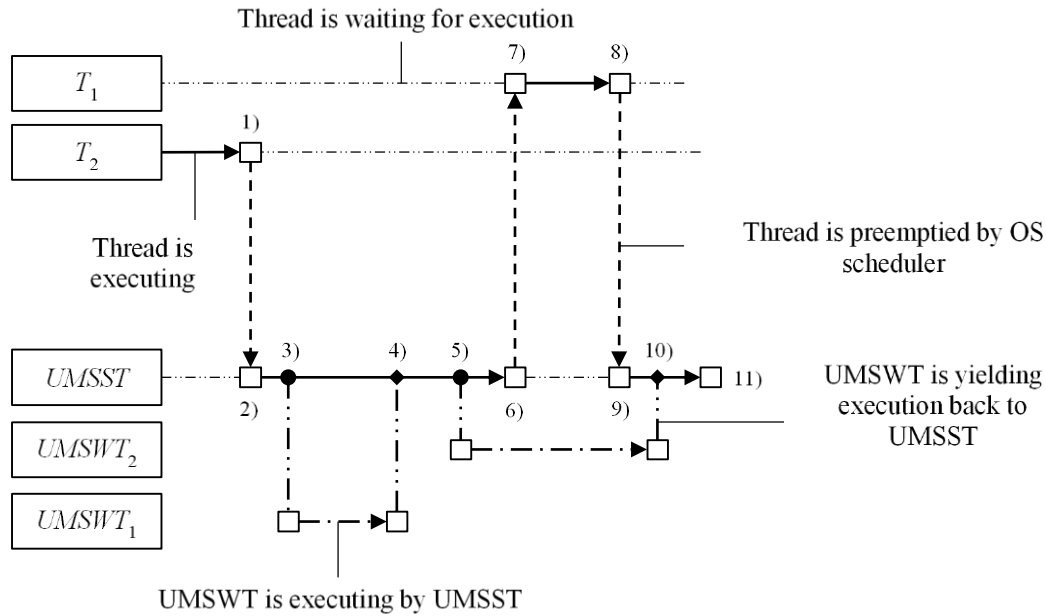


Figure 1. – Illustration of the execution of two operating system threads  $T_1$  and  $T_2$  (under control of the operating system scheduler), one user-mode scheduler thread  $UMSST_1$ , and two user-mode scheduling worker threads  $UMSWT_1$  и  $UMSWT_2$ : solid arrows represent the execution of operating system threads; long dash dot arrows represent the execution of user-mode scheduler worker threads; dash arrows represent the thread preemption done by the preemptive scheduler; square and diamond glyphs represent the begin and end of the user-mode scheduler thread execution

During the initialization, the scheduler allocates the following resource on each logical processor of the multi-core system:

- one  $UMSCL$ ;
- one ready user thread queue ( $RUTQ$ );
- one  $CST$  which is bound to previously allocated  $UMSCL$  and  $RUTQ$ .

Besides the mentioned above, the scheduler also includes arrays of all created  $CST$ ,  $CUT$  and  $CSP$ . A high-level view of cooperative multi-threaded scheduler is illustrated on Figure 2.

*Cooperative scheduler thread.*  $CST$  is responsible for handling the operating system callbacks and the user thread requests. It consists of  $UMSST$ , an implementation of  $UmsSchedulerProc$ , a pointer to  $RUTQ$  and a field representing its  $CST$  state.  $CST$  is created during scheduler initialization. During the initialization,  $CST$  switches from *Created* to *Initializing*, to *Initialized* and then to *Executing* state (executes an implementation of  $UmsSchedulerProc$ ). Starting from this point,  $CST$  can switch between the four states as follows: *Executing*, *WaitingWrldle* (if no requests to handle, scheduler thread is idling), *WaitingTaskExecuting* ( $CST$  is executing user thread), and *Terminated*.  $CST$  supports two type of  $CUT$  requests: direct control transfer between two  $CUT$ , and blocking-unblocking  $CUT$  by using  $CSP$ .

A high-level view of the control direct transfer implementation is presented on Figure 3. It is implemented as the following sequence of events:

1.  $CUT_1$  transfers execution back to  $CST$  using the  $UmsThreadYield$  procedure that passes information about  $CUT_2$  over the  $SchedulerParam$  argument (Figure 3, arc 1);
2.  $CST$  uses information from  $SchedulerParam$  and finds requested  $CUT_2$ ;
3.  $CST$  uses the  $ExecuteUmsThread$  procedure to execute  $CUT_2$  (Figure 3, arc 2).

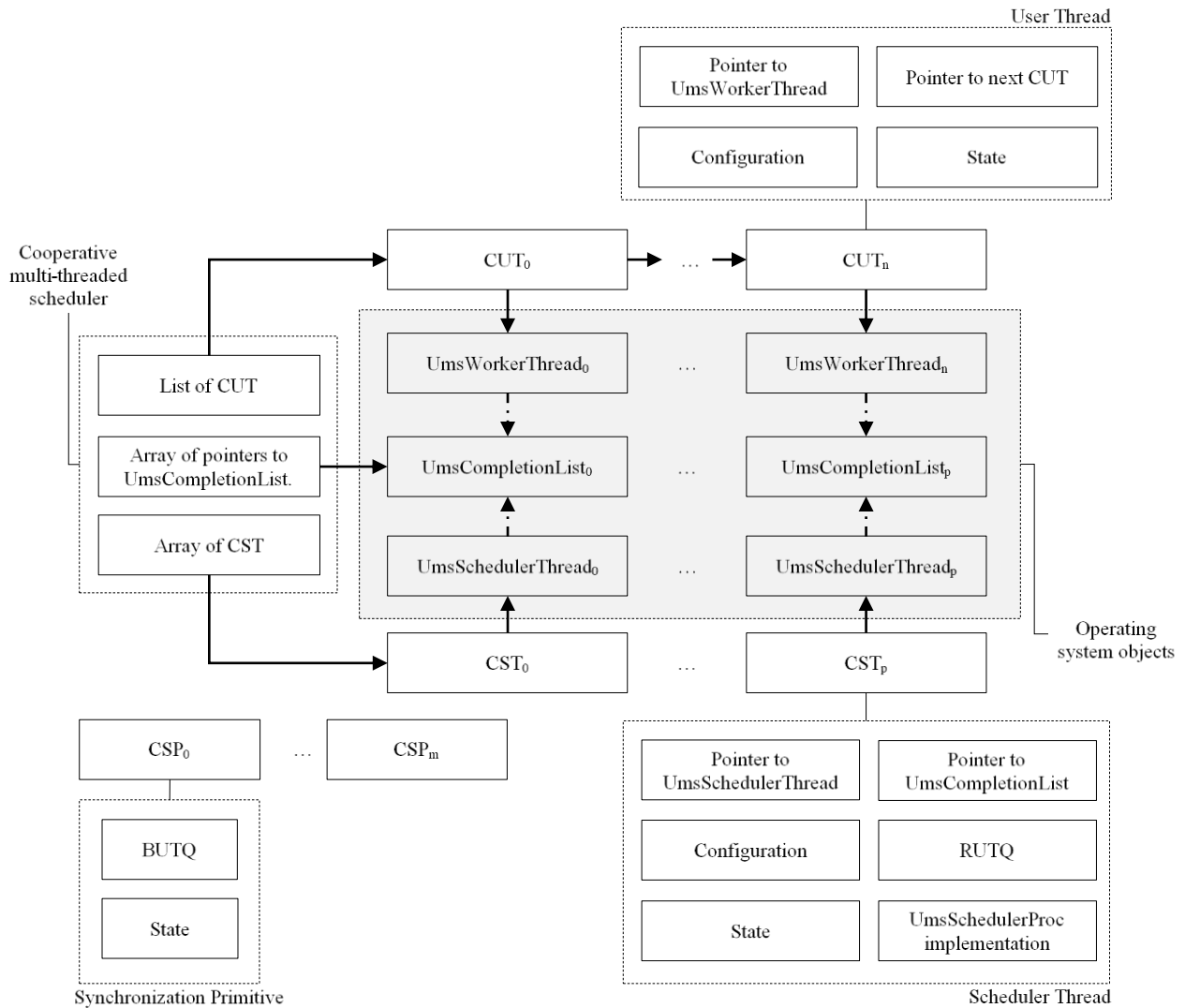


Figure 2. – High-level view of cooperative multi-thread scheduler architecture

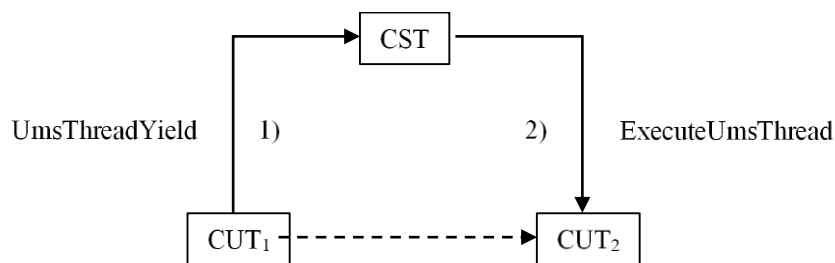


Figure 3. – High-level view of control direct transfer between CUT<sub>1</sub> and CUT<sub>2</sub>

The implementation of CUT on CSP synchronization request handler has been described in [9].

*Cooperative user thread.* CUT is responsible for the user thread execution. It consists of UMSWT and a field representing the CUT state. CUT is created by a user request which includes a pointer to the user defined procedure and an index of the logical core which CUT will execute on. During the initialization, CUT switches from the *Created* state to *Initializing*, then to *Initialized* and then to *Ready* states (CUT is ready for execution and currently resides in RUTQ). Starting from this moment, CUT could switch between these five states: *Ready*, *Executing*, *WaitingWrStopped* (CUT is blocked after performing the control transfer), *WaitingWrBlocked* (CUT is synchronized using CSP),

*WaitingWrSystemTrap* or *WaitingWrSystemCall* (CUT is blocked on a system call) and *Terminated*. When CUT is switching between the above states it sometimes holds one of the intermediate states: *StandBy* (when CUT has been selected from RUTQ but has not been executed yet, i.e. between *Ready* and *Executing*), and *WaitingWrYielded* (when CUT has transferred execution, but the request has not been decoded yet by CST, i.e. between *Executing* and *WaitingWrStopped*, or between *Executing* and *WaitingWrBlocked*).

*Cooperative synchronization primitive*. CSP represents an event. It is used to synchronize two or more CUTs executing on different cores. CSP allows to block CUT waiting for the event, and it resumes CUT's execution without support from the operating system. CSP is created by a user request and is not tied to any logical core, CST or CUT. CSP consists of a blocked user thread queue (BUTQ), which is used to hold blocked CUTs, and includes a field for representing CSP state.

At any moment of time, CSP can be in one of the following states: *Reset*, *Signaled*, *Signaling*, *SignalingActSignal*, *SignalingActReset*, *Joining*, *JoiningActSignal* and *JoiningActReset*. States *Reset* and *Signaled* are terminal states. Other states are transitional ones, which are required to support concurrent calls to CSP's procedures.

In order to better understand the role of each state, let us consider five examples.

*Example 1*. CSP is in state *Reset*. It indicates that the event represented by CSP has not happened yet. CUT<sub>1</sub> wants to notify about the event all the CUTs blocked on CSP. The notification should either unblock one of the CUTs or switch CSP into the *Signaled* state. Here is the sequence of events:

*Step 1*. CUT<sub>1</sub> switches CSP from state *Reset* to state *Signaling*. This is to indicate that there is a CUT which is executing the notify procedure.

*Step 2*. CUT<sub>1</sub> checks BUTQ for the blocked CUT. If BUTQ does not contain the blocked CUT then the behavior of CUT<sub>1</sub> is described by step 3a, otherwise by step 3b.

*Step 3a*. Because BUTQ does not contain a blocked CUT, CUT<sub>1</sub> switches CSP from state *Signaling* to state *Signaled*. It indicates that the event represented by CSP has happened.

*Step 3b*. Because BUTQ contains one or more blocked CUTs, CUT<sub>1</sub> pops one blocked CUT and moves it from BUTQ to RUTQ (simultaneously switching CUT to state *Ready*). Then CSP switches to state *Reset*.

*Example 2*. Imagine, that at the same time as CUT<sub>1</sub> executes the "signal" action, one more CUT<sub>2</sub> also want to execute the "signal" action to notify about the event. Here is how the sequence of events will look like:

*Step 1*. CUT<sub>2</sub> fails to switch CSP from state *Reset* to state *Signaling*. This is because CSP is already in state *Signaling*.

*Step 2*. Because CUT<sub>2</sub> cannot be sure whether CUT<sub>1</sub> has already checked for BUTQ, it switches CSP from state *Signaling* to state *SignalingActSignal*. This state indicates that if CUT<sub>1</sub> have not unblocked any of CUTs, then it should switch CSP back to state *Signaling* and repeat the check of BUTQ.

*Example 3*. Imagine, CUT<sub>2</sub> from the previous example executes the "reset" action rather than the "signal" one. In this case, the "reset" action will have the same sequence of events as the "signal" action has (see the previous example), with one exception, CUT<sub>2</sub> will switch CSP to state *SignalingActReset*. This will force CUT<sub>1</sub> to switch CSP to state *Reset* right after completion of the "signal" action.

The blocking of CUT is performed in a similar way, with the only exception that the blocking operation is performed by CST. The following two examples illustrate it.

*Example 4*. CUT<sub>3</sub> transfers control to CST<sub>1</sub> in order to synchronize itself on CSP, which is in state *Reset*. Here is how the execution of the "synchronize" action will look like:

*Step 1*. CST<sub>1</sub> switches CSP from *Reset* to *Joining* state. It indicates that CSP is now used by CST to synchronize CUT.

*Step 2*. CST<sub>1</sub> puts CUT<sub>3</sub> in BUTQ.

Step 3.  $CST_1$  switches CSP back to state *Reset*.

Example 5. Imagine, that at the same time as  $CUT_3$  is being synchronized on CSP, other  $CUT_1$  executes the “signal” action. In this case, the “signal” action execution is as follows:

Step 1.  $CUT_1$  fails to switch CSP from state *Reset* to state *Signaling*, as CSP is already in state *Joining*.

Step 2.  $CUT_1$  switches CSP from state *Joining* to state *JoiningActSignal*. This state instructs  $CST_1$  to switch CSP to state *Signaling* and to re-execute the “signal” action after the completion of the “synchronize” action.

Figure 4 shows the state transfer matrix of CSP.

	R	S	SG	SR	SS	JG	JR	JS
R			$f_s$			$f_j$		
S	$f_r$		$f_s$					
SG		$f_s$		$f_s, f_{cr}$	$f_s, f_{cs}$			
SR	-							
SS			-					
JG	-						$f_j, f_{cr}$	$f_j, f_{cs}$
JR	-							
JS			-					

**State abbreviations:**

- S – Signaled
- R – Reset
- SG – Signaling
- SR – SignalingActReset
- SS – SignalingActSignal
- J – Joining
- JR – JoiningActReset
- JS – JoiningActSignal

**Conditional flags:**

- $f_r$  – flag, indicating a reset operation was taken
- $f_s$  – flag, indicating a signal operation was taken
- $f_j$  – flag, indicating a join operation was taken
- $f_{cr}$  – flag, indicating a concurrent reset operation was taken
- $f_{cs}$  – flag, indicating a concurrent signal operation was taken
- $f_{cj}$  – flag, indicating a concurrent join operation was taken

Figure 4. – State transfer matrix of cooperative synchronization primitive

*Experimental environment.* All experiments were done on two multi-core systems. The first multi-core system was equipped with two Intel Xeon E5520 processors. Each processor has 4 cores. Each core runs on 2.26 GHz frequency and has high-speed hierarchical cache memory (L1 – 64 KB, L2 – 256 KB). Besides that, each core has the Intel Hyper-Threading technology built in, which allows the execution of two hardware threads on the single core. Each processor has access to shared L3 cache of 8 MB size, and access to local and remote memory with NUMA memory organization. The system is equipped with 16 GB of RAM and is controlled by Windows Server 2012 R2 (64 bits). The second multi-core system is equipped by one Intel Core i5-3450 processor (4 cores). Each core runs on 3.10 GHz frequency and has access to the high-speed local hierarchical cache memory (L1 – 256 KB and L2 - 1 MB), and the shared L3 cache memory with capacity of 6 MB. The system is equipped with 16 GB RAM and is controlled by Windows 10 Professional 1809.

*Cooperative scheduler and algorithms implementation.* Cooperative multi-thread scheduler is implemented in C/C++ language as dynamically linked library (.dll) using Visual C++ 14.1 compiler.

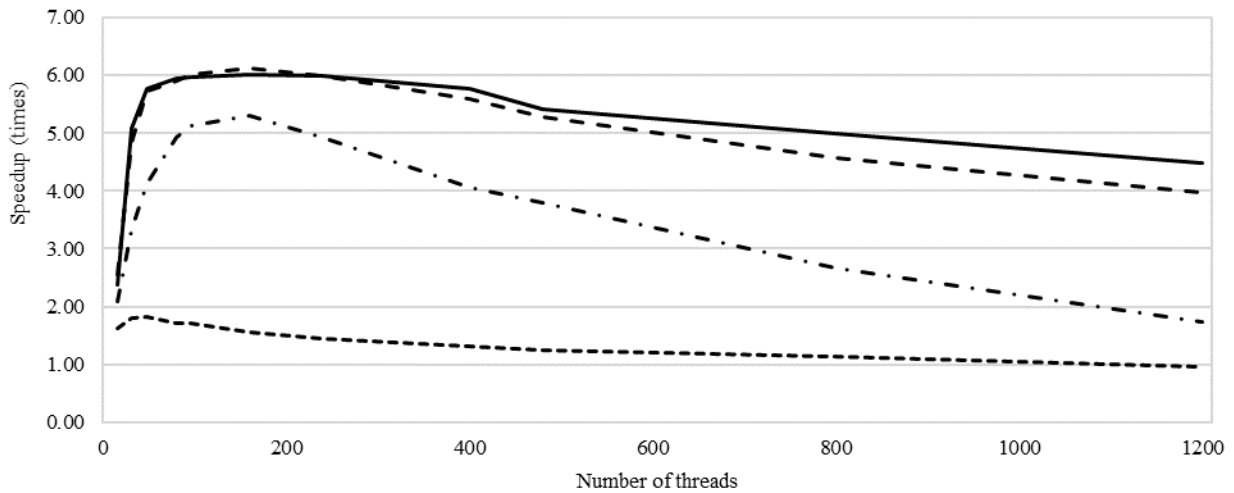


Parallel Gaussian Elimination algorithms ( $\mu_1$ ,  $\mu_2$ ) [13, 17] are implemented using the native Windows threads and the *AutoResetEvent* synchronization primitive. Cooperative algorithms ( $\mu_{1k}$  и  $\mu_{2k}$ ) [10, 11] are implemented using the developed scheduler library. The source code of  $\mu_1$ ,  $\mu_2$ ,  $\mu_{1k}$  and  $\mu_{2k}$  algorithms are written in C/C++ language and compiled into executions (.exe) using Visual C++ 14.1 compiler. Block-parallel all pairs shortest path Floyd-Warshall algorithm (BFW) is implemented using OpenMP directives for the task-based parallelism. The cooperative threaded block-parallel algorithm CTBPA [12] is implemented using the scheduler library. The source code of both algorithms is written in C/C++ language and is compiled using Intel Compiler 18 that is configured on maximum optimization (OV3) with additional options to do maximum optimization for underlying multi-core system (IvyBridge specific optimization and vectorization using Intel AVX).

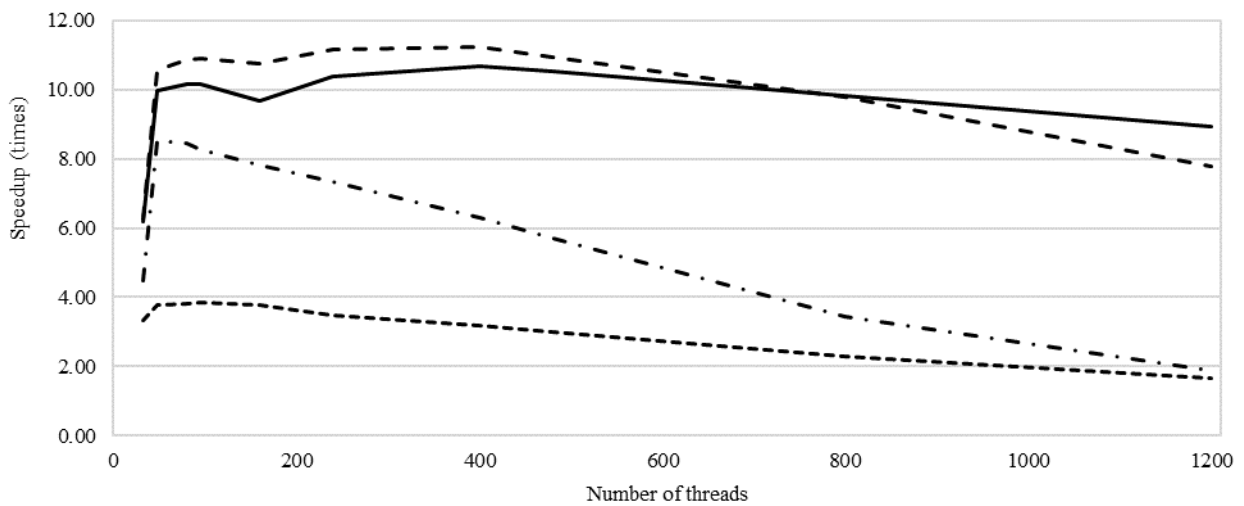
*Experimental results.* In order to demonstrate the effectiveness of the cooperative multi-thread scheduler and the developed cooperative algorithms, we conducted experiments using two different parallel implementations of the Gaussian Elimination and the block-parallel implementation of all-pairs shortest path Floyd-Warshall algorithm. Figures 5, *a* and 5, *b* report experimental results obtained in series of 1000 runs for each of four  $\mu_1$ ,  $\mu_2$ ,  $\mu_{1k}$  and  $\mu_{2k}$  algorithms depending on the number of threads. All experiments use linear algebraic equations of 2400 variables. The obtained results are compared against the results obtained for single-threaded implementations on both multi-core systems. The best execution time of the single-threaded implementation is 10.43 sec on the first multi-core system and is 4.91 sec on the second multi-core system. On first multi-core system, the cooperative algorithms  $\mu_{1k}$  and  $\mu_{2k}$  have shown a maximum speed up of 6.00 and 6.12 times, which exceeds the maximum speedup shown by the  $\mu_1$  and  $\mu_2$  algorithms (5.30 and 1.82 times respectively). On the second multi-core system, cooperative algorithms  $\mu_{1k}$  and  $\mu_{2k}$  have shown the maximum speed up of 10.68 and 11.23 times, which significantly overcomes the speed up obtained by the  $\mu_1$  and  $\mu_2$  algorithms (8.47 and 3.83 times respectively).

In addition to the execution time of four algorithms  $\mu_1$ ,  $\mu_2$ ,  $\mu_{1k}$  and  $\mu_{2k}$ , we have analyzed the execution time distribution in series of 1000 runs of each algorithm. Figure 6 presents histograms of the execution time and time intervals (in sec) for both multi-core systems. Table 1 reports the execution time intervals of best runs of the algorithms.

Figure 7 presents the execution speedup of algorithm CTBPA comparing to BFW algorithm depending on block size for graph of 2400 vertexes in series of 1000 runs done on Intel Core i5-3450 multi-core system. CTBPA demonstrated speedup for all block sizes with maximum speedup in 26.8% (0.842 vs. 0.664 seconds) for block size of 120x120. Histograms of execution intervals of both algorithms are presented on picture 8. Table 2 demonstrates significant advantage of developed (using developed cooperative multi-threading scheduler) CTBPA algorithm against existing parallel implementation of BFW, done using operating system scheduler.



a)



b)

a) – Intel Core i5-3450; b) – Intel Xeon E5520

Figure 5. – Speedup in times between execution time of  $\mu 1\kappa$  (solid),  $\mu 2\kappa$  (dash),  $\mu 1$  (dash dot),  $\mu 2$  (dot) algorithms and best execution time of single-threaded algorithm in series of 1000 runs depending on number of threads on SLAE of 2400 variables size

Table 1. – Execution time intervals of  $\mu 1$ ,  $\mu 2$ ,  $\mu 1\kappa$  и  $\mu 2\kappa$  algorithms for both experimental multi-core systems

Algorithm	Intel Core i5-3450, 4 cores			2 x Intel Xeon E5520, 8 cores		
	from (seconds)	to (seconds)	width (seconds)	from (seconds)	to (seconds)	length (seconds)
$\mu 1$ (160 threads)	0.89	1.05	0.16	1.09	1.85	0.76
$\mu 2$ (48 threads)	2.49	3.19	0.70	2.61	2.95	0.34
$\mu 1\kappa$ (160 threads)	0.80	1.35	0.55	0.969	1.005	0.044
$\mu 2\kappa$ (160 threads)	0.79	0.99	0.20	0.912	0.952	0.040

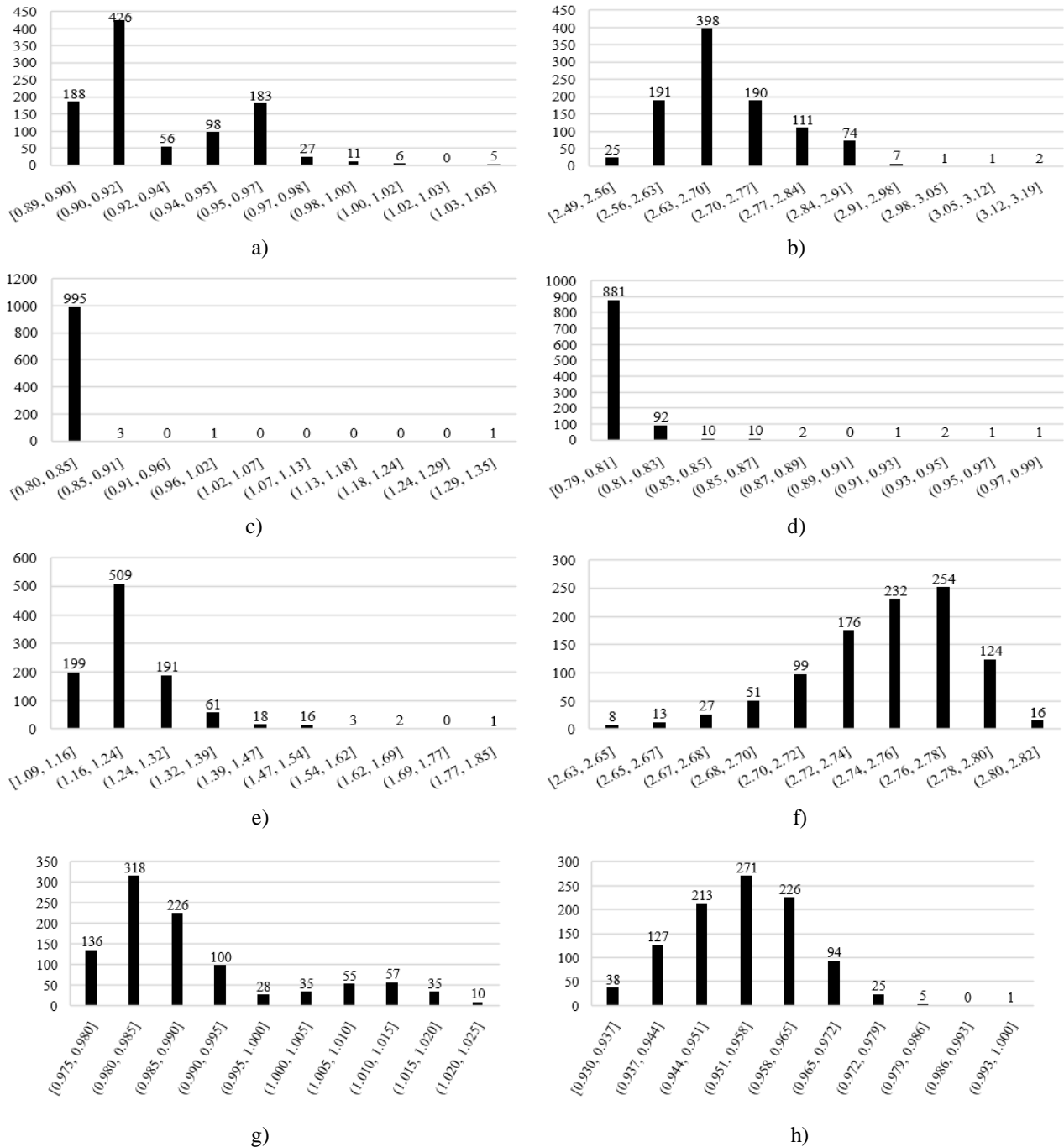


Figure 6. – Histograms of execution time intervals (in seconds) of  $\mu_1$ ,  $\mu_2$ ,  $\mu_1\kappa$  and  $\mu_2\kappa$  algorithms in series of 1000 runs for SLAE of 2400 variables size on Intel Core i5-3450 (4 cores) a), b), c), d)

and Intel Xeon E5520 (8 cores) e), f), g), h)

a) –  $\mu_1$  (160 threads); b) –  $\mu_2$  (48 threads); c) –  $\mu_1\kappa$  (160 threads); d) –  $\mu_2\kappa$  (160 threads);  
 e) –  $\mu_1$  (48 threads); f) –  $\mu_2$  (96 threads); g) –  $\mu_1\kappa$  (400 threads); h) –  $\mu_2\kappa$  (400 threads)

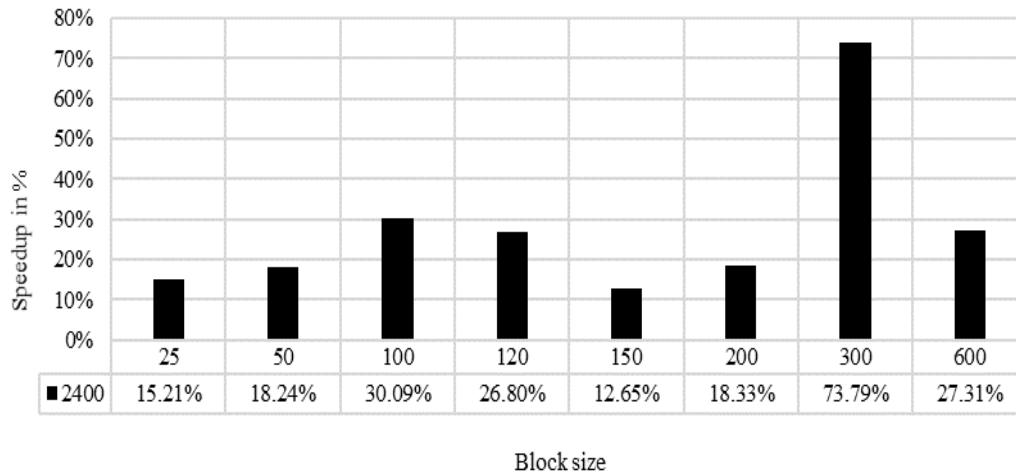


Figure 7. – Speedup in % given by algorithm CTBPA comparing to algorithm BFW vs. block size on graphs of 2400 vertices in series of 1000 runs done on Intel Core i5-3450

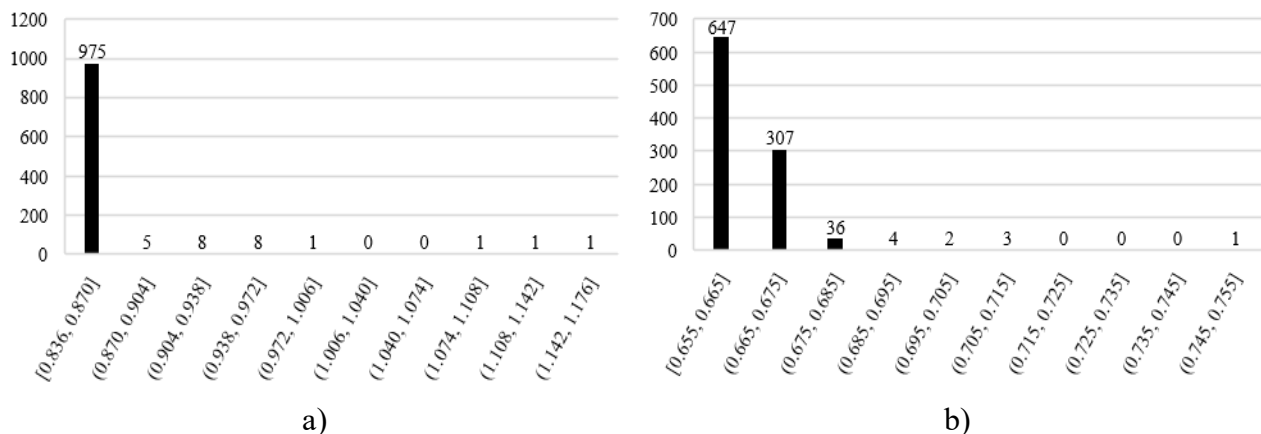


Figure 8. – Histograms of the execution time (sec) of a) BFW and b) CTBPA algorithms in series of 1000 runs on graphs of 2400 vertices and block size of 120×120

Table 2. – Execution time intervals given by algorithms BFW and CTBPA on block size 120x120

Intel Core i5-3450, 4 core			
Algorithm	from (sec)	to (sec)	width (sec)
BFW	0.836	1.176	0.340
CTBPA	0.655	0.755	0.100

*Conclusion* When solving a large task, the performance of a multi-core system depends heavily on the operating system and the system software built into it, on the one hand, and depends on the method of constructing parallel multi-threaded algorithms that solve the application problems and perform calculations on a large amount of data, on the other hand. In the paper, a cooperative multi-thread scheduler is proposed and experimentally investigated, which reduces the execution time compared to the preemptive scheduler of the operating system, and is used to create multi-threaded algorithms in such a way that the threads running on one core directly transfer control to each other, and the threads running on different cores interact with each other through the proposed synchronization primitive, that speeds up the processes of thread lock-unlock.

### **References**

- [1.] Prihozhy, A.A. Analysis, transformation and optimization for high performance parallel computing / A.A. Prihozhy / Minsk: BNTU, 2019. – 229 p.
- [2.] Richter, J. M. Windows via C/C++ / J. M. Richter, C. Nasarre. – 5th ed. – Microsoft Press, 2007. – 848 p.
- [3.] OpenMP. OpenMP [Электронный ресурс]. – Режим доступа: <https://www.openmp.org>. – Дата доступа: 24.02.2020.
- [4.] ISO/IEC/IEEE 9945:2009 Information technology — Portable Operating System Interface (POSIX®) Base Specifications, Issue 7 [Электронный ресурс]. – Режим доступа: <https://www.iso.org/standard/50516.html>. – Дата доступа: 24.02.2020
- [5.] MPI Forum. MPI Forum [Электронный ресурс]. – Режим доступа: [www.mpi-forum.org](http://www.mpi-forum.org). – Дата доступа: 24.02.2020.
- [6.] Google. The Go Programming Language [Электронный ресурс]. – Режим доступа: <https://golang.org>. – Дата доступа: 24.02.2020.
- [7.] The Khronos Group Inc. OpenCL [Электронный ресурс]. – Режим доступа: <https://www.khronos.org/opencl/>. – Дата доступа: 24.02.2020.
- [8.] Berkeley Lab. Berkeley UPC - Unified Parallel C [Электронный ресурс]. – Режим доступа: <http://upc.lbl.gov/>. – Дата доступа: 24.02.2020.
- [9.] Карасик, О. Н. Усовершенствованный планировщик кооперативного выполнения потоков на многоядерной системе / О. Н. Карасик, А. А. Прихожий // Системный анализ и прикладная математика. – 2017. – № 1. – С. 4–11.
- [10.] Прыхожы, А. А. Кааператыўныя блочна-паралельныя алгарытмы рашэння задач на шмат'ядравых сістэмах / А. А. Прыхожы, А. М. Карасік // Сістэмны аналіз і прыкладная інфарматыка. – 2015. – № 2. – С. 10–18.
- [11.] Прихожий, А.А. Кооперативная модель оптимизации выполнения потоков на многоядерной системе / А.А. Прихожий, О.Н. Карасик // Системный анализ и прикладная информатика, 2014, № 4, с. 13-20.
- [12.] Карасик, О. Н. Поточковый блочно-параллельный алгоритм поиска кратчайших путей на графе / О. Н. Карасик, А. А. Прихожий // Доклады БГУИР. – 2018. – № 2. – С. 77–84.
- [13.] Прихожий, А.А. Исследование методов реализации многопоточных приложений на многоядерных системах / А.А. Прихожий, О.Н. Карасик // Информатизация образования, 2014, № 1, с. 43-62.
- [14.] Прихожий, А. А. Кооперативная потоковая модель решения задач большой размерности на многоядерных системах / А. А. Прихожий, О. Н. Карасик // Big Data and Advanced Analytics. – 2018. – №. 4. – С. 381–386.
- [15.] Probert, D. Dave Probert: Inside Windows 7 - User Mode Scheduler (UMS) [Электронный ресурс]. / D. Probert. – Режим доступа: <https://channel9.msdn.com/shows/Going+Deep/Dave-Probert-Inside-Windows-7-User-Mode-Scheduler-UMS/>. – Дата доступа: 01.11.2019.
- [16.] Microsoft. User-Mode Scheduling [Электронный ресурс]. – Режим доступа: <https://docs.microsoft.com/en-us/windows/desktop/ProcThread/user-mode-scheduling>. – Дата доступа: 24.02.2020.
- [17.] Howe, J. Parallel Gaussian Elimination [Электронный ресурс]. / J. Howe, S. Bratcher. – Режим доступа: <http://www.cse.ucsd.edu/classes/fa98/cse164b/Projects/PastProjects/LU>. – Дата доступа: 01.11.2018
- [18.] Venkataraman, G. Blocked All-Pairs Shortest Paths Algorithm / G. Venkataraman, S. Sahni, S. Mukhopadhyaya // Journal of Experimental Algorithmics (JEA). – 2003. – Vol. 8. – P. 857–874
- [19.] Tang, P. Rapid development of parallel blocked all-pairs shortest paths code for multi-core computers / P. Tang // IEEE SOUTHEASTCON 2014. – Lexington, KY, USA: IEEE, 2014. – P. 1–7
- [20.] Park, J. Optimizing graph algorithms for improved cache performance / J. Park, M. Penner, V. K. Prasanna // IEEE Transactions on Parallel and Distributed Systems. – 2004. – Vol. 15, №. 9. – P. 769–782

UDK 6Т12.15.7:334.75(575.1)

## **USE OF INFORMATION AND COMMUNICATION TECHNOLOGIES FOR THE EXPORT OF ECONOMIC SUBJECTS OF SAMARKAND REGION**



**F.F. Saidov**

*Assistant-professor of Tashkent university of information technologies named after Muhammad al-Khwarizmi*



**D.R. Mamasoatov**

*Scientific Research Center "Scientific Research of the Economy of Uzbekistan Principles and Problems" Junior Researcher*

*Department of Audiovisual technologies, Tashkent University of information technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan*

*Department of Information system and technologies in economy, Tashkent State University of Economics, Tashkent, Uzbekistan*

*E-mail: farmaf@mail.ru, dilshodbek11@mail.ru*

### **F.F. Saidov**

*Teacher of the Department of Audiovisual technologies, faculty of Television technologies, Tashkent University of information technologies named after Muhammad al-Khwarizmi. Author of 22 research papers on information technologies, economics and corporate governance.*

### **D.R. Mamasoatov**

*Teacher of the Information system and technologies in economy, Faculty of Information Systems in Economics, Tashkent State University of Economics. Author of 35 research papers on information technologies, economics and Econometrics.*

**Abstract.** This article is aimed at improving the use of information and communication technologies to support the export of entrepreneurship, covering the processes of further expansion of business activity and creating conditions for domestic producers, including their export to foreign markets and exports.

**Keywords:** information and communication technologies, export, interactive services, modern technologies, action strategy, integrated electronic information system of integrated divisions, electronic catalogs. Introduction

### *Introduction*

The rating of developed countries in the global economy is determined by at least 80% of their GDP due to the creation of new products and services based on new knowledge, 60-70% of the share of small business and private entrepreneurship, and 30-35% of their share in exports. Therefore, small business, private entrepreneurship and its export potential are one of the priority areas of economic policy of each of the developed and developing countries.

According to the World Bank, the annual volume of goods and services produced by small businesses and private entrepreneurship is 6.4 trillion. US dollars, of which 2.1 trillion. US dollars in exports, world GDP grew by 16% [9]. In the conditions of fierce competition and trade in international markets, business entities need innovative technologies and ICT in order to quickly receive information about consumers, effectively organize logistics processes and capture new markets.

Particular attention was paid to economic measures in the last decade in our country for wide access to small business, private entrepreneurship, their support and promotion. Small business and private entrepreneurship are the leading sectors of the economy, accounting for 57% of the country's GDP, 27% of exports and 76% of the total number of employees [8].

However, the regions lack sufficient opportunities for the development of business entities, as well as their limited access to international markets, lack of access to modern technologies in their export orientation, customs, transport and logistics, cargo tracking, quality control, storage and shipment. Thus, to support the export activities of enterprises, access to international markets and the need to improve the use of ICTs for the delivery of competitive products are necessary.

The Development Strategy of the Republic of Uzbekistan for 2017-2021 defines the tasks of "creating a favorable business environment for the large-scale development of small business and private entrepreneurship, improving and simplifying the procedures for access of business entities to engineering networks and the gradual acquisition of new products and technologies, according to which competitiveness national goods in the domestic and foreign markets, liberalization of export operations, diversification of the structure and geography of exports, expansion and obilizatsiya export potential of industries and regions "[1].

Our first priority should be to provide all-round support to entrepreneurs, in particular, in creating favorable conditions for the import and introduction of high technologies and the latest scientific achievements into production"[2].

The effective implementation of these tasks requires the development of small business and private entrepreneurship in the regions of the republic and the availability of ICT in their export operations. The introduction of methods to improve the use of modern ICTs, which provide competitive products and exports that meet the requirements of world standards, quick access to international markets and efficient logistics, requires research to ensure the country's economic development.

In this regard, conducting research to improve the use of ICT for the export of business entities is one of the pressing problems of our time. Therefore, it is necessary to pay special attention to the development of economic entities and their export in our country. At the same time, today, research on improving the use of ICT is one of the urgent tasks in the regions when exporting business entities.

Based on the above, you should pay attention to the development of business entities and their exports in our country. At the same time, the use of ICT to further improve the quality of export development of business entities is one of the important tasks in the regions. In this aspect, we need to study the scientific work of a number of domestic and foreign scientists.

#### **Literature review**

At the same time, the use of information and communication technologies is one of the country's most important tasks for further improving the export of business entities. In this regard, we must study the scientific work of a number of foreign and domestic scientists. Including scientific and practical issues of the implementation of information and communication technologies in various sectors of the economy were studied by foreign scientists M. M. Bezkorovain, A. I. Kostogryzov, V. M. Lvov, A. V. Goncharov, S. L. Goldstein [3 ].

G. Kalyanov's work presents a method for automatic verification of the quality of information systems in the process of designing an information system using CASE-tools, which is built in graphical form [4].

A. V. Goncharov, S. L. Goldstein and others in their research work paid special attention to issues of assessing the quality of modeling business processes [5]. Scientific research in the field of the formation of the national information and communication market and the national information system in entrepreneurship during the liberalization of the economy of the republic was also carried out by our country's scientists B. B. Begalov [6] and A. Kenzhabaev [7].

The above scientists and experts have studied the scientific basis for the implementation of ICT in various sectors of the national economy, but it can be seen that issues such as the effective use

of ICT in the export of business entities are not sufficiently addressed. Therefore, this situation requires a theoretical and methodological justification for the effective use of information and communication technologies in the export of business entities in accordance with the development of the economy.

### Analysis and Results

A statistical analysis of the sector will be necessary to increase the export potential of business entities and study the state of ICT use. It is necessary to develop entrepreneurial qualifications by improving the use of information and communication technologies in the export of business entities, the widespread introduction of international best practices in entrepreneurial activity, and establishing close cooperation with leading foreign organizations.

First of all, if we pay attention to the share of small business and private entrepreneurship in the main types of economic activity in 2017, then today its share is different in the types of economic activity (Figure 1).

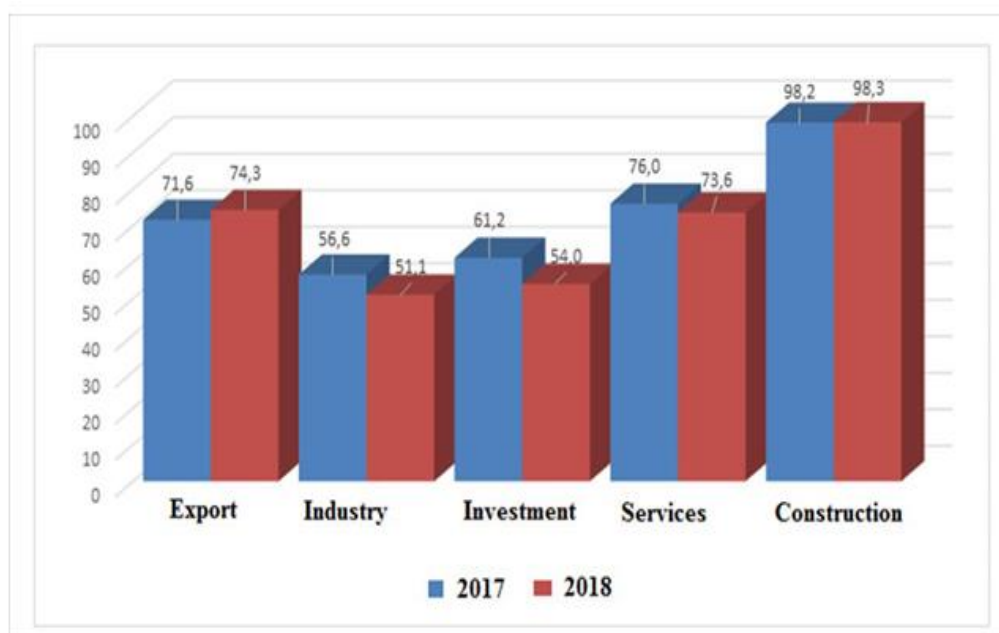


Figure 1. –The share of small business and private entrepreneurship in economic entities of Samarkand region%) [10]

**Source:** Information from Samarkand Regional Statistical Office.

Given the share of small business and private entrepreneurship in major economic activities, the share of construction (98.3%), services (73.6%) and industry (51.1%). The following indicators comprise export (74.3%) and investments (54.0%). In this regard, taking into account the foreign trade of small businesses and private entrepreneurship, this allows you to increase the share of small businesses in export due to state support of small businesses producing import-substituting and export-oriented products.

In 2018, the share of small business and private entrepreneurship in the total volume of exports in the republic amounted to 26.5%, in 2000 - 10.2%, in 2017 - 27%, i.e. increased by 16.8% unit (Figure 2).



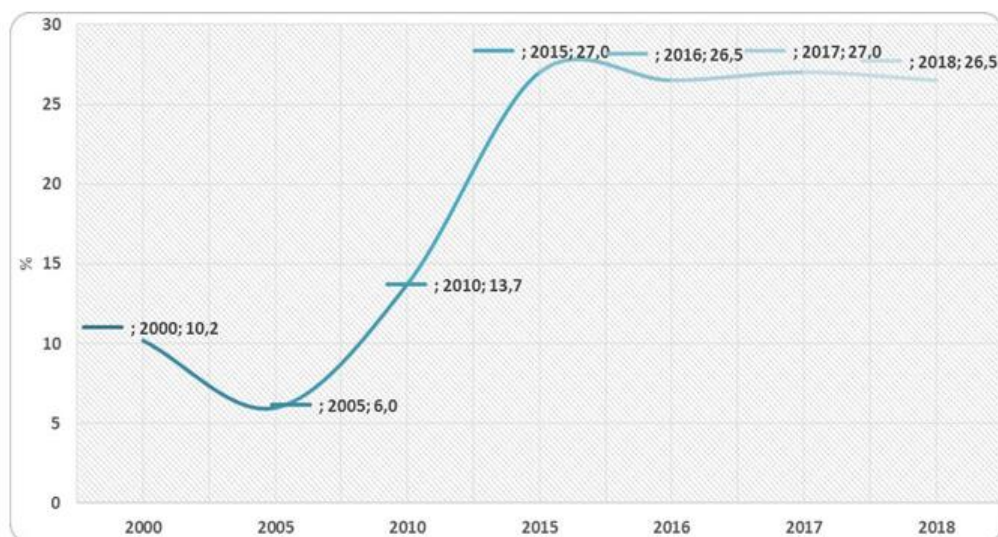


Figure 2. – The share of small business and private entrepreneurship in export (in% to the total volume)

**Source:** authoring based on data from the State Statistics Committee of the Republic of Uzbekistan.

As a result of a serious study of these processes, we must be sure that the share of exports over the past three years has amounted to about 27%. This is caused by a number of problems.

Along with the advantages of using information and communication technologies when exporting enterprises, a number of problems and contradictions arise. They should be divided into six main groups:

- technological;
- organizational;
- personal;
- financial;
- transaction costs and other problems (table 1).

Table 1. – Problems in the use of information and communication technologies in the export of business entities

Problems	Feature Description
<b>Technological</b>	<ul style="list-style-type: none"> <li>- restrictions in computer programs;</li> <li>- the complexity of installing software;</li> <li>- lack of Internet, the speed of which does not meet the requirements;</li> <li>- the inability to install fiber optic cables;</li> <li>- insufficient formation of export management processes at enterprises;</li> <li>- the need to modernize the technologies used;</li> <li>- Lack of the website of the enterprise;</li> </ul>
<b>Organizational</b>	<ul style="list-style-type: none"> <li>- not the readiness of managers to implement ICTs, lack of awareness about them</li> <li>- misconception about the organizational culture of ICT implementation;</li> <li>- the need for reorganization of the enterprise related to ICT;</li> <li>- the need to form a highly qualified team on ICT by managers;</li> <li>- non-use of remote interactive services;</li> <li>- lack of innovative consulting services;</li> <li>- insufficient monitoring of online surveillance in logistics;</li> </ul>

Continuation of *table 1*

<b>Private</b>	<ul style="list-style-type: none"> <li>- insufficient knowledge and skills of qualified ICT personnel;</li> <li>- lack of involvement of qualified ICT consultants;</li> <li>- unwillingness to perform additional tasks related to the use of ICT in export processes;</li> </ul>
<b>Financial</b>	<ul style="list-style-type: none"> <li>- the high cost of computer software;</li> <li>- the cost of installing computer software;</li> <li>- Internet costs;</li> <li>- marketing expenses;</li> <li>- advertising expenses</li> </ul>
<b>Transaction costs</b>	<ul style="list-style-type: none"> <li>- transaction costs due to market uncertainty;</li> <li>- automation of business processes;</li> <li>- search for information;</li> <li>- negotiation and conclusion of contracts;</li> <li>- certification, assessment</li> </ul>
<b>Other</b>	<ul style="list-style-type: none"> <li>- ignorance of the use of ICT;</li> <li>- operational changes in the marketing environment of the foreign market;</li> <li>- imperfection of the electronic payment system:</li> <li>- non-use of electronic retail space;</li> <li>- refusal to participate in local and international fairs and exhibitions, so the wishes of people are not taken into account when developing and improving the enterprise.</li> </ul>

**Source:** Compiled by the author based on the results of the study.

According to the study, the creation of an information system in an enterprise during the export process provides the following advantages: collecting organizational information about the export, storing important data, coordinating the export plan, analyzing costs and profits, making timely decisions in foreign markets and strategic planning.

As a result, the process of creating an information system for studying foreign markets by business entities is currently being formed. Based on the experience of foreign countries, it should be emphasized that for the development of exports at the enterprise, first of all, the creation and effective use of the ICT system is required.

Based on the results of the analysis, all the problems arising in connection with the use of ICTs not only affect the export process, but also lead to difficulties in analyzing and monitoring information about the foreign market.

As practice shows, when using ICT in the export of business entities, the following issues remain relevant:

The exchange of documents between business entities and organizations is often carried out in paper form and slows down many decision-making processes. Moreover, the costs of securities used in these processes are increasing day by day.

- introduction of an electronic document management system for ICT services when exporting business entities taking into account the requirements of the world market;

- in the event of an accident during the export process, the widespread use of ICT functionality and assessment from a distance of this provision

- difficulties in receiving orders from customers abroad;

- a lot of mistakes are found, and corruption in paperwork over a long period of time;

- problems on the participation of business entities in foreign fairs;

- business entities partially use ICTs, mainly for the automation of settlement operations and accounting reports;

- the presence of vague goals in the export activities of business entities;
- making unreasonable decisions, taking into account the low efficiency of exports, real estate and risk factors.

Based on the foregoing, to improve the use of ICT in the export of enterprises, the following is proposed:

- widespread use of ICT functionality to improve the exchange of documents between business entities and partner organizations in these processes;
- Improving the electronic system of registration of insurance services using ICT for export; - acceptance of orders from customers abroad in electronic form;
- financial support for the participation of entrepreneurs at foreign fairs;
- development of consulting services for the use of electronic trading platforms;
- development of an optimal system of electronic declaration of electronic procurement for all types of business entities;
- implementation in practice of a single integrated interdepartmental electronic information system of foreign trade operations. This system will help reduce and accelerate the process of drawing up foreign trade contracts, as well as ensure currency control over cash flows from domestic goods.
- Improving online interactive services for the export of business entities.
- Creation of a unified automated control system for transit goods and vehicles, which allows you to automate the processing and control of transit goods in real time.

Based on the above proposals, it is possible to achieve an increase in exports through a positive solution to the problems. In addition, the banks of our country need to introduce a wide range of remote banking services for small businesses and private entrepreneurship, such as Bank-Client, Internet banking, SMS banking. For example, entrepreneurs will be able to manage their accounts using mobile phones in real time.

In turn, the use of ICT tools for exporting goods is very important for business entities. Taking into account the level of ICT support for the export of business entities, it is necessary to improve the quality of ICT use in the export process, and the large-scale introduction of business entities in the export process (Figure 3).

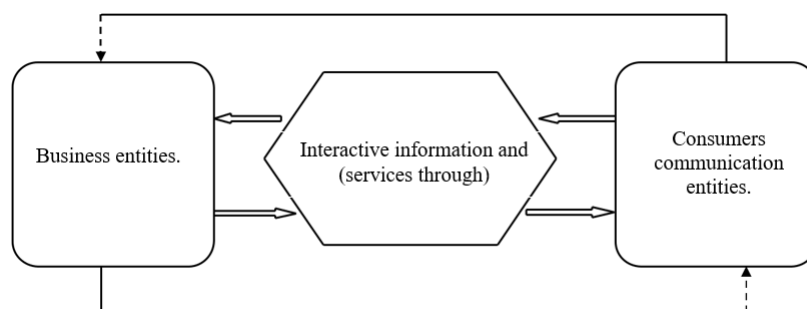


Figure 3. – Widespread adoption of information and communication technologies in the export of business entities

**Source:** Authoring

Through the use of interactive services through ICT, the export potential of enterprises is accelerating. This process is an innovation, a change in the business activity of the export of business entities, and its implementation provides high efficiency.

In conclusion, it can be noted that as a result of further improvement of the process of using ICT in the export of business entities, high indicators of export quality can be achieved. This, in turn, indicates the need for the use of ICT in the export of business entities.

Improving the use of information and communication technologies when entering foreign markets can solve the problems facing entrepreneurs. And this leads to a more effective solution to the problems of achieving high growth rates of finished goods exports by increasing the output of goods that are in demand on foreign markets.

### **Reference**

- [1.]Mirziyoyev Sh.M. Address by the President to the “Senate of the Oliy Majlis and members of the Legislative Chamber”. December 22, 2017.
- [2.]Decree of the President of the Republic of Uzbekistan “On the development strategy of the
- [3.]Republic of Uzbekistan for 2017-2021” dated February 7, 2017 No. PP-4947.
- [4.]Bezkorovayny M.M., Kostogryzov A.I., Lvov V.M. Instrument-modeling complex for assessing the quality of functioning of information systems. - M.: Arms, politics, conversion, 2001 - 303 p.
- [5.]Kalyanov G.N. CASE technology. Consulting in the automation of business processes. - M.: Hot line - Telecom, 2002. -- 320 p.
- [6.]Goncharov A.V., Goldstein S.L. Problems of assessing the quality of modeling business processes: literary review, analogues, prototype // Bulletin of Ural State Technical University-UPI. Modern technologies: problems and solutions. - Yekaterinburg: GOU VPOUGTU-UPI, 2004. - No. 5 (35). - S. 150-156.
- [7.]Begalov B.A. Doctoral dissertation on the topic “Ahborot - communication lar Bosorning shacklanish v rivoglanish trend larini econometrics modeillashtirish”, 2001.
- [8.]Kenzhabaev A.T. Doctoral dissertation on the topic “Tadbirkorlik faoliyatida ahborotlashtirish milliy tizimini shakllantirish muammolari”, 2005
- [9.]www.stat.uz - The data of the State Statistics Committee of the Republic of Uzbekistan are used.
- [10.] www.worldbank.org- The World Bank
- [11.] www.samstat.uz- Samarkand regional statistical office.

## **ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ ДЛЯ ЭКСПОРТА ЭКОНОМИЧЕСКИХ СУБЪЕКТОВ САМАРКАНДСКАЯ ОБЛАСТЬ**

**Ф.Ф. Саидов**

*Ассистент Ташкентского  
университета информационных  
технологий им.  
Мухаммед аль-Хорезми*

**Д.Р.Мамасоатов**

*Научно-исследовательский центр  
«Научные исследования принципов и  
проблем экономики Узбекистана», младший  
научный сотрудник*

*Ташкентский университет информационных технологий имени Мухаммеда Аль-Хорезми, Ташкент, Узбекистан*

*Ташкентский государственный экономический университет, Ташкент, Узбекистан*

*E-mail:farmaf@mail.ru, dilshodbek11@mail.ru*

**Аннотация:** Цель данной статьи - улучшить использование информационных и коммуникационных технологий для поддержки экспорта предпринимательства, охватывая процессы дальнейшего расширения деловой активности и создания условий для отечественных производителей, включая их экспорт на зарубежные рынки и экспорт.

**Ключевые слова:** информационно-коммуникационные технологии, экспорт, интерактивные услуги, современные технологии, стратегия действий, интегрированная электронная информационная система интегрированных подразделений, электронные каталоги. Вступление

УДК 004.6-024.11:004.738.5

## ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА КОМПЛЕКСНОГО АНАЛИЗА ДАННЫХ ИНТЕРНЕТ- ИСТОЧНИКОВ



**М.П. Батура**  
Заведующий  
лабораторией НИЛ  
8.1 «Новые  
обучающие  
технологии»  
БГУИР, Доктор  
технических наук,  
профессор,  
академик  
«Международной  
академии  
наук высшей  
школы»

**И.И. Пилецкий**  
Доцент кафедры  
информатики  
БГУИР,  
кандидат физико-  
математических  
наук, доцент,  
старший научный  
сотрудник

**В.А. Прытков**  
Проректор по  
учебной работе  
БГУИР, кандидат  
технических наук,  
доцент

**Н.А. Волорова**  
Заведующая  
кафедрой информатики  
БГУИР, кандидат  
технических  
наук, доцент

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь  
E-mail: [bmpbel@bsuir.by](mailto:bmpbel@bsuir.by), [ianmenski@gmail.com](mailto:ianmenski@gmail.com), [prytkov@bsuir.by](mailto:prytkov@bsuir.by), [volorova@bsuir.by](mailto:volorova@bsuir.by)

### **М.П. Батура**

Заведующий лабораторией НИЛ 8.1 «Новые обучающие технологии» БГУИР, Доктор технических наук, профессор, академик «Международной академии наук высшей школы», заслуженный работник образования Республики Беларусь. Область научных исследований: Системный анализ, управление и обработка информации в технических и организационных системах. Опубликовано более 150 научных работ, в том числе 4 монографии, учебник, выдержавший три издания, 4 учебных пособия, имеет авторские свидетельства на изобретения.

### **И.И. Пилецкий**

Кандидат физико-математических наук, доцент БГУИР. В сфере IT более 47 лет. Участие в разработке нескольких десятков крупных проектов: главный конструктор проекта, главный архитектор программно-информационного обеспечения, руководитель проекта, начальник отдела, заведующий лабораторией (НИИ информатики, Академия наук Беларуси, МБА, БГУИР). Автор десятков исследований, имеет более 95 публикаций.

### **В.А. Прытков**

Проректор по учебной работе БГУИР, кандидат технических наук, доцент. Автор более 40 научных работ. Область научных интересов: обработка изображений и текстурный анализ, синтаксические методы обработки информации, анализ слабоструктурированных данных.

### **Н.А. Волорова**

Заведующая кафедрой информатики БГУИР, кандидат технических наук, доцент. В сфере IT более 40 лет. Имеет более 140 публикаций, сфера научных интересов – модели сложных систем

**Аннотация.** В статье приводится описание инструментов мониторинга открытых интернет-источников с целью выявления экспертов в некоторой предметной области, определения тематик публикаций, оценки

популярности публикаций. Описываются принятые решения при разработке различных вариантов построения аналитического комплекса, полученные результаты его работы и применения методов искусственного интеллекта для проведения глубокого анализа данных.

**Ключевые слова:** интернет-источники, Big Data, мониторинг, анализ, Machine Learning, машинное обучение, Neo4j, Cassandra, Natural Language Processing, обработка естественного языка,

## **1. Введение**

В настоящее время уже привыкли получать данные и информацию из интернет-источников. В результате анализа данных интернет-источников полученная информация является базовой для принятия решений различными организациями.

Как правило, это неструктурированные текстовые данные, различные мультимедийные данные. Данные могут быть получены как из социальных сетей, так и тематических сайтов (газет, журналов, библиотек, компаний и т.д.), содержащих различные публикации (некоторые наиболее популярные платформы приведены на рисунке 1). Пользователи интернет-ресурсов и социальных сетей могут самостоятельно выбирать интересные им направления и читать публикации интересных им людей, которым они симпатизируют. В свою очередь создатели контента могут быть со своим контентом и своими подписчиками. Такие связи, как правило, бывают достаточно сложными и представляют собой многоуровневую циклическую сеть.

Несмотря на наличие различных систем для анализа данных из интернет-источников, данная тематика не только не утрачивает актуальности, а, напротив, становится все более востребованной. Поисковые системы помогают найти разнообразную информацию практически о любых явлениях в мире, деятельности организаций и людей, но основной проблемой является не то, как найти данные, а, *проблемой является то, как разобраться в этом многообразии источников данных и самих данных как превратить данные в информацию, а информацию – в знания для принятия разумных решений.*

Интерес к интеллектуальным системам анализа данных постоянно усиливается, поскольку создание новой техники и технологий становится все более наукоемким, причем эта тенденция характерна и для продукции массового рынка. Данный проект «Интеллектуальная система комплексного анализа данных интернет-источников» (ИС Анализа Данных) предполагает развитие системы «Система комплексного анализа данных интернет-источников» [2, 3, 4] расширении ее области применения и наделяет компонент новой системы *функциональностью интеллектуальных систем, позволяющей выполнять глубокий интеллектуальный анализ исходных данных с применением методов искусственного интеллекта.*

Соответственно, компании, внедряющие новые технологии, а также придающие своей продукции интеллектуальные функции, обладают определенным конкурентным преимуществом. На глобальном уровне это проявляется в определении своей ниши в мировой экономике, ее расширении.

## **2. Назначение и цели**

### **Назначение:**

1. «ИС Анализа Данных» - предназначена для поддержки принятия обоснованных решений, на основе *мониторинга и анализа данных* из открытых Интернет источников, в том числе и научных публикаций.

2. Выявления важных публикаций, ведущих специалистов и *поиске экспертов* определенных предметных областей.

3. Создание многоцелевого, модифицируемого кластера в Университете для подготовки специалистов Data Scientist.



Рисунок 1. – Разнообразие источников данных (источник: [1])

### Цели и задачи

1. Сбор и анализ данных некоторой (некоторых) предметной области с целью:
  - определения тематики и тональности публикаций, поиска экспертов (авторитетов) в предметной области, тематик их публикаций, а также просмотра их профилей;
  - поиска новых областей исследований, тематик публикаций, определение областей к которым интерес уже пропал;
  - просмотр различных индивидуальных параметров некоторой предметной области (публикации, снижение/увеличение интереса);
2. Разработке тематической масштабируемой системы, которая при минимальных модификациях может быть применена и в других областях, как корпоративных, так и правительственных, например, для *анализа СМИ или мониторинга настроений* в обществе и др.

Большой интерес представляют Задачи нахождения экспертов предметной области, которые имеют большую аудиторию подписчиков и на которую они могут влиять, их публикации и их связи.

Примеры задач: выявление лидера социального мнения, *группы лиц, связанных в соцсетях по некоторой тематике*, задачи *рекламы (маркетинга) определенного товара или группы товаров*, поиска экспертов и документов в некоторых научно-технических областях.

*Социальные отношения могут быть явными или неявными*, а социальные сети помогают идентифицировать прямые и косвенные отношения между людьми, группами людей, и характер их взаимодействий.

Перспективность проекта «ИС Анализа Данных» в том, что результаты могут быть использованы как для прикладных целей, например, определения наиболее перспективных и актуальных научных направлений с учетом конкретной специфики (для конкретной организации или страны, с учетом имеющегося задела и т. д.), определения экспертов в заданных предметных областях (для приглашения специалистов или формирования команды исследователей), так и для научно-образовательной цели – подготовки специалистов по анализу больших объемов неструктурированных данных (Data Scientist) на основе разработки и оптимизации аналитических ML-алгоритмов, применяемых в системе.

Модульность «ИС Анализа Данных» позволит при минимальной адаптации использовать данный программный комплекс для анализа и мониторинга различных явлений по заказу конкретных пользователей, например, анализа публикаций в «Твиттере», изучения популярности «бренда», поиска блогеров в социальных сетях, изучения конкурентов, изучения рынка сбыта, поведения групп людей и т.д.

«ИС Анализа Данных» позволит не только помочь с проведением анализа и принятия решения, но и позволит сэкономить весьма дефицитный в условиях конкуренции ресурс –

время. В современной экономике важно не только дать качественный продукт с новыми свойствами, но и сделать это в числе первых.

Ценность работы заключается не только в создании «Интеллектуальной системы комплексного анализа данных интернет-источников», но и в создании в Университете многоцелевого, модифицируемого кластера для анализа данных интернет-источников и глубокой математической подготовки специалистов по анализу больших объемов неструктурированных данных, специалистов Data Scientist, внедрении результатов исследований в научно-образовательный процесс при подготовке магистрантов в области обработки больших объемов информации.

### **3. Технология построения «Интеллектуальной системы комплексного анализа данных интернет источников»**

Основу технологии разработки системы должны составлять методы и алгоритмы построения и обслуживания графовой модели различных социальных сетей авторов (блогеров) и их публикаций (в том числе и в СМИ), ссылок на их публикации и определение рейтинга конкретного автора (блогера) публикаций, определение тематик публикаций и классификация их по областям знаний. И для глубокого интеллектуального анализа применение методов искусственного интеллекта на основе преобразований семантических данных из графовой модели в различные матричные представления для их анализа в ML.

Анализируя данные из социальных сетей (СМИ) можно выявить как прямые, так и скрытые отношения между людьми, группами людей, а также характер их взаимодействий. На основе определенных связей между субъектами группы можно сделать выводы об индивидуальных предпочтениях субъектов и прогнозировать выбор объекта предпочтения.

«ИС Анализа Данных» должна состоять из следующих компонент: сбора данных, фильтрации данных и составления «мешка слов» из N-грамм (векторизации), библиотеки аналитических модулей, хранилища данных, графовой базы данных и графа знаний, аналитического компонента, обеспечивающего, взаимодействие с пользователем и подготовки выдачи результата, клиентского модуля и универсальной интеграционной шины (управляющего компонента).

При необходимости набор модулей и компонент может быть расширен, а некоторые модули заменены новыми. Общая технология построения многофункционального комплекса по обработке данных и работы компонент, а именно чтения данных интернет-источников, фильтрации данных и векторизации, библиотеки ML модулей, хранилища и БД «граф знаний», приведена в более ранних публикациях [2, 3, 4]. В «ИС Анализа Данных» должна быть применена новая архитектура построения многофункциональных комплексов как набор постоянно работающих компонент в виде отдельных серверов, изменена предметная область и система дополнена многофункциональным компонентом БД «граф знаний» и аналитическим компонентом подготовки и выдачи результата.

Все взаимосвязи и взаимодействия компонент должны быть организованы на основе специально разрабатываемого управляющего компонента (универсальная шина) «ИС Анализа Данных». Данный компонент должен обладать функциональностью интеграции данных и приложений, реализовывать функции брокера, синхронного и асинхронного выполнения приложений. Управляющий компонент должен реализовать средства логирования, сбора статистики и мониторинга работы компонент системы.

Компонент скачивания публикаций и компонент извлечения текста из материалов должны использовать технологию Docker, инструменты мониторинга и предупреждения ошибок, развертывание и хранилище логов. Применять новую технологию извлечение медиа-данных из документов и обработки новых форматов документов (PostScript, заархивированные документы).

Компонент графовая база данных и граф знаний должен быть дополнен графовой моделью для поиска постов в социальных сетях на определённую тематику, и другими



графовыми моделями, которые могут быть максимально полезны и интересны пользователю, методами интеллектуального анализа графовой модели на основе применения ML.

Компонент хранилища данных должен быть дополнен новыми БД, соответствующими новым областям применения системы.

Библиотеки аналитических модулей новыми ML-модулями интеллектуального анализа данных.

### 3.1. Ретроспективный анализ семейства СКА

#### 3.1.1. «Аналитический комплекс анализа данных из открытых интернет источников»

«Аналитический комплекс анализа данных из открытых интернет источников» (СКА) [2], позволяет получить информацию популярности некоторого бренда (компании, объекта, товара, события, явления и возможно некоторого человека) на основе анализа существующих данных из электронных газет, журналов, статей, социальных сетей с помощью различных аналитических методов и оценить отношение в Интернет-СМИ к этому явлению. Значимость проекта состоит в создании наращиваемого многомодульного комплекса, позволяющего анализировать большие данные из интернет источников различными аналитическими методами и выборе для применения наиболее подходящих для данного явления.

Аналитический комплекс из состоит компонент (см. рисунок 2): сбора данных, фильтрации данных и составление «мешка слов» из N-грамм (векторизации), хранилища данных, библиотеки аналитических модулей, подготовки выдачи результата, клиентского модуля.

При необходимости набор модулей компонент может быть расширен, а некоторые модули заменены новыми.

Сами компоненты состоят из функциональных модулей, на рисунке 2 приведена логическая схема взаимодействия компонент и модулей.

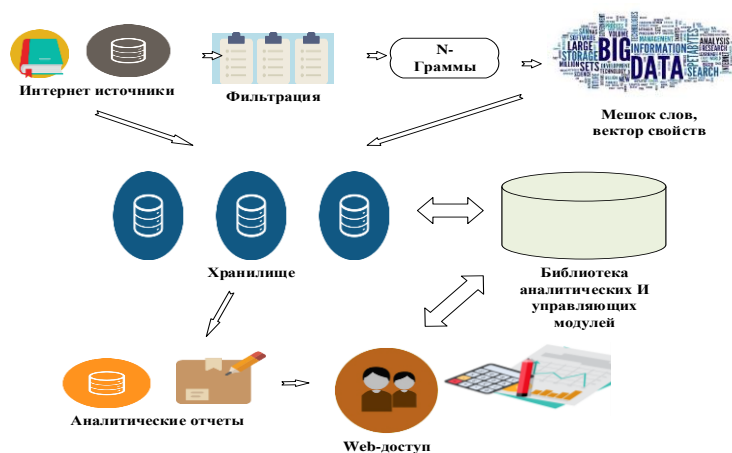


Рисунок 2. – Логическая схема взаимодействия компонент

Аналитический комплекс выполняет свои функции в несколько этапов:

– Компонент сбора данных – выполняет целевое сканирование выбранных социальных сетей, новостных порталов, сайтов и помещает полученные документы в специальном виде в базу данных документов (в хранилище);

– Компонент обработки данных выполняет фильтрацию исходных данных (токенизацию, лемматизацию, стемминг, удаление стоп слов, перевод к нижнему регистру), построения «мешка слов» и векторизацию исследуемых наборов документов («мешок слов» в специальном виде хранится в хранилище);

– Компонент библиотека аналитических модулей, содержит набор модулей, которые осуществляют обработку данных, полученных из интернет источников с целью поиска упоминаний о брендах, определения их тональности и формирования аналитических данных для передачи клиентскому модулю, а также содержит управляющие и служебные модули. В реализованной системе компонент библиотека аналитических состоит из ML модулей: SVM и LDA, PLSA [4, 5]. Данные векторные и вероятностные модули апробированы. Применение комбинированного подхода к оценке некоторого явления позволяет наиболее достоверно его оценить, выполнить ретроспективный анализ прошедшего события, идентифицировать новое события и принять правильные решения;

– Компонент подготовки выдачи результата – подготавливает информацию в виде отчетов для пользователей аналитического комплекса;

– Компонент хранилища данных – содержит данные из социальных сетей, предварительно обработанные и размеченные данные, необходимые для построения классификатора, «мешок слов», информацию о брендах, а также служебную информацию, необходимую для работы других модулей системы.

Основной недостаток в данном решении состоит в том, что он разработан по общепринятой архитектуре для аналогичных комплексов анализа данных:

– нет общего архитектурного решения для всей системы в целом, каждый компонент представляет собой отдельную функциональность;

– все компоненты работают последовательно;

– анализ данных выполняется в отложенном режиме.

К достоинству данного архитектурного решения следует отнести удобство при модернизации и функциональному расширению компонент комплекса, т.к. компоненты состоят из отдельных модулей, которые реализуют его функциональность.

### **3.1.2. Система комплексного анализа данных интернет-источников**

«Система комплексного анализа данных интернет-источников» (СКАД ИИ) [3], позволяет анализировать большие объемы данных из интернет-источников в области научных исследований и предназначена для сбора информации о научных публикациях, построения графа знаний, что дает возможность определять экспертов предметной области, тематики их работ, их взаимосвязи, а также определять передовые научные направления.

Система позволяет находить экспертов (авторитетов) в предметной области и выдавать оценку их рейтинга влияния. Например, лучше прочитать три книги признанных экспертов в определенной области, чем десять книг дилетантов.

Перспективность проекта СКАД ИИ в том, что результаты могут быть использованы как для прикладных целей, например, определения наиболее перспективных и актуальных научных направлений с учетом конкретной специфики (для конкретной организации или страны, с учетом имеющегося задела и т. д.), определения экспертов в заданных предметных областях (для приглашения специалистов или формирования команды исследователей), так и для научно-образовательной цели – подготовки специалистов по анализу больших объемов неструктурированных данных (Data Scientist) на основе разработки и оптимизации аналитических ML-алгоритмов, применяемых в системе.

Модульность СКАД ИИ позволяет при небольших изменениях использовать данный программный комплекс для анализа и мониторинга различных явлений по заказу конкретных пользователей, например, как для СКА [2] изучения популярности «бренда», поиска блогеров в социальных сетях, изучения конкурентов, изучения рынка сбыта, поведения групп людей и т.д.

СКАД ИИ позволит не только помочь с проведением анализа и принятия решения, но и позволит сэкономить весьма дефицитный в условиях конкуренции ресурс – время. В современной экономике важно не только дать качественный продукт с новыми свойствами, но и сделать это в числе первых.

Ценность работы заключается не только в создании «Системы комплексного анализа данных интернет-источников», но и в создании в Университете многоцелевого, модифицируемого кластера для анализа данных интернет-источников и глубокой математической подготовки специалистов по анализу больших объемов неструктурированных данных, специалистов Data Scientist, внедрении результатов исследований в научно-образовательный процесс при подготовке магистрантов в области обработки больших объемов информации.

Facebook, Google, другие популярные сервисы построены на использовании графовых моделей данных. Facebook, например, использует не только информацию о людях, их именах, профессиях и т. д., но также сведения о взаимосвязях между людьми, которые представляют ещё большую ценность. Социальные отношения могут быть явными или неявными, а социальные сети помогают идентифицировать как прямые, так и косвенные отношения между людьми, группами людей, и характер их взаимодействий. Gartner утверждает, что способность использовать эти графы обеспечивает «устойчивое преимущество в конкурентной среде».

В отличие от известных поисковых систем (например, Facebook, Google) СКАД ИИ позволяет найти наиболее перспективные и актуальные научные направления и определить экспертов в заданных предметных областях.

Основу технологии разработки системы составляют методы и алгоритмы построения и обслуживания графовой модели социальной сети авторов и их публикаций, ссылок на их публикации и определение рейтинга конкретного автора публикаций, определение тематик публикаций и классификация их по областям знаний.

Анализируя данные из социальных сетей можно выявить как прямые, так и скрытые отношения между людьми, группами людей, а также характер их взаимодействий. На основе определенных связей между субъектами группы можно сделать выводы об индивидуальных предпочтениях субъектов и прогнозировать выбор объекта предпочтения.

Графовая модель социальной сети может быть построена на применении классической графовой модели, которая включает узлы и взаимосвязи, а также их свойства и метки.

Основным назначением графовой базы данных является применение графовых алгоритмов для обработки полученных данных, выстраивание логических взаимосвязей и подготовка и выдача информации для пользователя. Также ключевой особенностью таких БД является формирование очень гибких запросов, наподобие следующих: ***Какие авторы наиболее часто сотрудничали с автором данной популярной статьи, является ли данный автор писателем только в области биологии, или же он пишет еще и на темы математического анализа, существуют ли математические публикации, которые по какой-то причине перекликаются с темой философии и т. д.***

СКАД ИИ состоит из следующих компонент (см. рисунок 3): сбора данных, фильтрации данных и составление «мешка слов» из N-грамм (векторизации), хранилища данных, библиотеки аналитических модулей, подготовки выдачи результата, клиентского модуля.

При необходимости набор модулей и компонент может быть расширен, а некоторые модули заменены новыми. Общая технология построения многофункционального комплекса по обработке данных из интернет источников и работы компонент, а именно чтения данных интернет-источников, фильтрации данных и векторизации, библиотеки и хранилища приведена в более ранних публикациях [2].

В СКАД ИИ апробирована технология построения многофункциональных комплексов как набор постоянно работающих компонент в виде отдельных серверов, изменена предметная область и система дополнена компонентом БД «граф знаний» и компонентом подготовки и выдачи результата.

Сами компоненты СКА состоят из набора функциональных модулей, на рисунке 3 приведена первоначальная логическая схема взаимодействия компонент и модулей. Данный

проект опирается на решения NLP & ML (см. рисунок 3, компоненты (2) и (3)), хранилище данных (компонент (4)), которое взаимодействует с компонентом граф – знаний (компонент (5)) и компонент взаимодействия с пользователем (компонент (6)). Для чтения данных их интернет источников используется компонент (1).

Достоинство данного решения состоит в том, что все сырые, обработанные данные и служебная информация собираются и хранятся в хранилище Hbase, которое используется для насыщения и построения графа – знаний. Хранилище может использоваться не только для накопления данных по текущей тематике, но для других областей применения, т.е хранилище может использоваться как озеро данных. Это значит, другие не графовые ML алгоритмы могут пополнять компонент (3) и использовать данные из хранилища.

Взаимодействие хранилища с графовой базой данных позволяет перестраивать и перезагружать граф - знаний, иметь множественное представление графа – знаний для различных областей применения.

Сама графовая база данных обладает рядом преимуществ и достоинств перед другими БД, она обладает свойствами OLTP & OLAP, поддерживает транзакции ACID (atomic, consistent, isolated и durable), чего не обеспечивает ни одна NoSQL БД.

Графовые технологии являются основой для построения интеллектуальных приложений, для применения алгоритмов искусственного интеллекта.

Но, как и для СКА (см. п. 3.1.1) данное решения уязвимо при возникновении сбоя при функционировании некоторых компонент. Обработка данных так же выполняется последовательно. И оказалось, что сама БД NoSQL – Hbase не обладает хорошими эксплуатационными характеристиками.

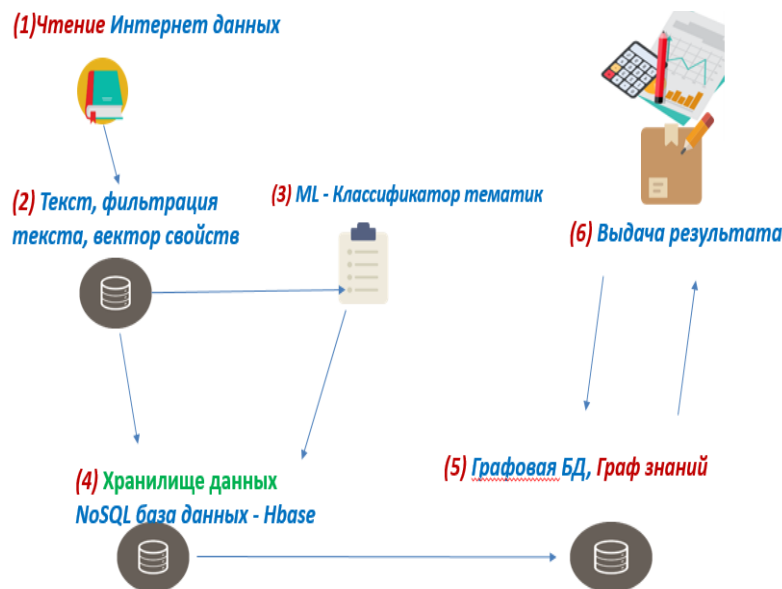


Рисунок 3. – Первоначальная логическая схема взаимодействия компонент СКАД ИИ

### 3.1.3. Модернизация система комплексного анализа данных интернет-источников

В процессе реализации проекта СКАД ИИ были приняты важные дополнительные архитектурные решения (см. рисунок 4, новая архитектура СКАД ИИ):

- все компоненты функционируют как постоянно работающие самостоятельные сервера;

- для создания компонента хранилища БД Hbase была заменена на лидера хранилища типа "семейство столбцов" Cassandra;

- для обеспечения взаимодействия компонент как по данным, так и по управлению, разработан управляющий компонент, который выполняет роль интеграционной шины (аналог USB - enterprise service bus, для работы в среде Big Data);
- остановка работы одного из компонент не приводит к остановке работы всего комплекса;
- обеспечен многопользовательский режим доступа;
- существенно модернизированы компоненты чтения данных из интернет источников и компонент анализа и выдачи отчетов.

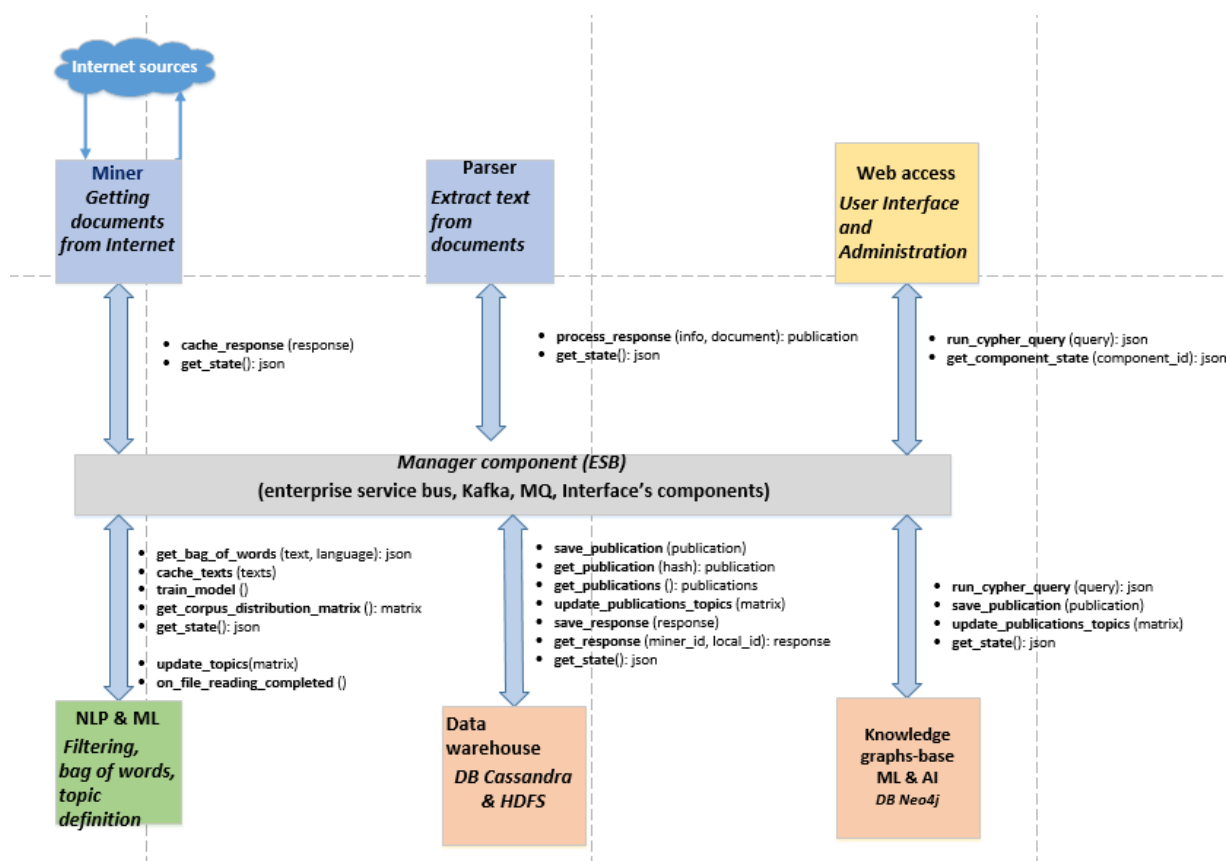


Рисунок 4. – Новая архитектура СКАД ИИ

- Miner Getting documents from Internet** Компонент сбора данных – выполняет целевое сканирование выбранных социальных сетей, новостных порталов, сайтов и помещает полученные документы в исходном виде и специально обработанном виде в хранилище (в качестве хранилища используются HDFS и база данных документов Cassandra);
- Parser Extract text from documents** Компонент обработки данных выполняет фильтрацию исходных данных построение «мешка слов», извлечение из необработанных файлов публикаций текста, авторов, тематики, названия, литературу и др. Данные в специальном виде помещаются в хранилище;
- NLP & ML Filtering, bag of words, topic definition** Компонент библиотека аналитических модулей содержит набор модулей, которые осуществляют обработку данных, полученных из интернет-источников, с целью определения тематик публикаций, и формирования аналитических данных для передачи клиентскому модулю, кроме того, содержит управляющие и служебные модули. Для тематического анализа текста применяется модифицированный алгоритм PLSA [5]. Тематический анализ с использованием EM-алгоритма позволяет выявить N самых важных тем во всём тексте, первоначально на основе

мешка слов, а в дальнейшем при пополнении корпуса документов, возможна корректировка классификатора документов. Здесь важно отметить, что для каждой темы определяется вероятность ее появления документе, что учитывается в графе знаний. Разработаны и апробированы модули построения и анализа графовых моделей, позволяющих получить знания для принятия решений [6, 7]. Область применения СКАД ИИ может быть расширена за счет пополнения компонента другими модулями, а применение комбинированного подхода к оценке некоторого явления позволит наиболее достоверно идентифицировать новые события и принять правильные решения;

Web access  
User Interface  
and  
Administration

Графический интерфейс обеспечивает взаимодействие с пользователем системы. Веб-интерфейс отправляет запросы к графу знаний через управляющий компонент, граф – знаний возвращает результат, который в виде различных отчетов выдается пользователю. Существуют различные алгоритмы определения наиболее важных статей, публикаций, блогеров [4, 7], в СКАД ИИ используется алгоритм `page_rank` [8].

Компонент хранилища данных – содержит данные из интернет-источников, предварительно обработанные и размеченные данные, необходимые для построения классификатора, «мешок слов», а также служебную информацию, необходимую для работы других модулей системы. В хранилище хранятся сырые данные с сайта, текст, фильтрованный текст, исходные документы, «мешок слов», тематика документов и служебная информация;

Knowledge  
graphs-base  
ML & AI  
DB Neo4j

Компонент графовая база данных и граф знаний [9] состоит из графовой БД, моделирующей предметную область, и программных модулей, позволяющих пополнять графовую БД данными из хранилища и извлекать из нее знания о запрашиваемых объектах, графовая модель является развитием графовой модели приведенной в [6]. «Граф знаний», динамически связывает в графической форме: названия, авторов, публикаций, тематики публикаций, ссылки на публикации и авторов, готовит и выдает различные отчеты.

Manager component (ESB)

(enterprise service bus, Kafka, MQ, Interface's components)

Управляющий компонент построен на базе Kafka, выполняет роль интеграционной шины, обеспечивает взаимодействие всех компонент, как по данным, так и по управлению. Является брокером

СКАД ИИ и обеспечивает гарантированную доставку сообщений (MQ).

#### 4. Графовые алгоритмы как база интеллектуального анализа информации

В настоящее время резко возрос интерес к применению графовых алгоритмов и аналитики для различных областей человеческой деятельности. Гибридная транзакционная и аналитическая обработка может потенциально переопределить способ выполнения некоторых бизнес-процессов, поскольку расширенная аналитика в реальном времени (например, планирование, прогнозирование и анализ «что, если») становится неотъемлемой частью самого процесса, а не отдельной выполняемой операцией после факта (Gartner research).

Здесь приведены некоторые примеры областей применения графовых алгоритмов, гибридной транзакционной и аналитической обработки данных:

- транзакции, маршрутизация, логистика, IoT, социальные сети, кибер-безопасность и аналитика;
- обнаружение мошенничества в реальном времени, графовая аналитика позволяет находить шаблоны построения связей и обнаружить мошенничество (например, в банковской сфере);
- выдавать рекомендации в реальном времени (например, в сфере услуг, соотносить данные о товаре, клиенте, предпочтениях, запасах, поставщике, логистике);
- мониторинг и управление различными физическими сложными сетями в реальном времени;

- идентификация и управление доступом клиентов, сотрудников к различным ресурсам в локальных и глобальных сетях в соответствии с их ролями;
- обеспечение работы различных приложений и функций социальные сети, предсказание поведения социальных групп.

Графовые технологии – это основа для создания интеллектуальных приложений, позволяющая делать более точные прогнозы и быстрее принимать решения. Графы лежат в основе широкого спектра вариантов использования искусственного интеллекта (ИИ).

Так, граф знаний – одна из основных областей ИИ, который позволяет понимать предписывающую аналитику и приложения ИИ (например, обработка и понимание естественного языка (NLP, NLU), PageRank).

В общем случае, огромное количество графовых алгоритмов классифицированы на алгоритмы: Pathfinnding, Centrality и Community Detection (см. примеры [10]).

Pathfinnding. Класс алгоритмов поиска кратчайших путей, с учетом различных весовых критериев (например, расстояния или скорости), например, найти самый быстрый маршрут для поездки, минимизировать трафик телефонных звонков.

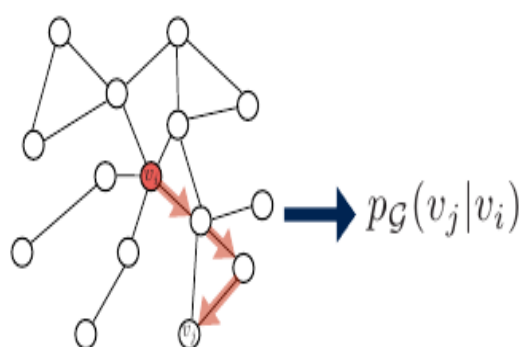


Рисунок 5. – Пример графа поиска путей

Centrality. Данный класс алгоритмов (центральности) заключается в понимании того, какие узлы наиболее важные в сети. Эти алгоритмы позволяют определить, как быстро можно распространять информацию в различных группах и между группами сущностей, предсказать появления новых тенденций в этих группах, выявлять уязвимости и возможные цели атаки в сетях связи и транспорта.

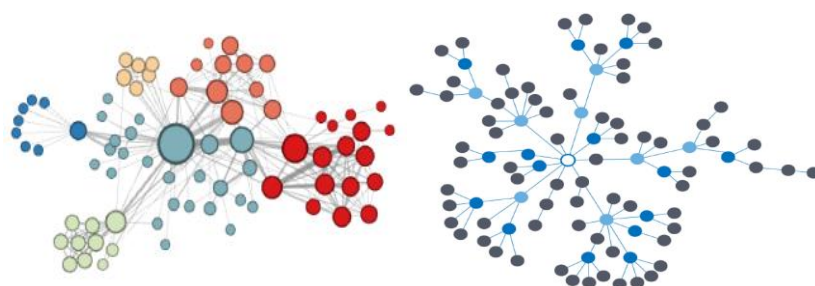


Рисунок 6. – Примеры графов групп сущностей и их связей

Community Detection (обнаружение сообщества). Класс алгоритмов, позволяющий изучать различные социальные сети, выявлять лидеров этих сетей, определять количественные характеристик различных групп. Оценивать иерархии, предсказывать тенденции поведения к видоизменению в этих группах, выявлять спамеров и потенциальных участников мошенничества.

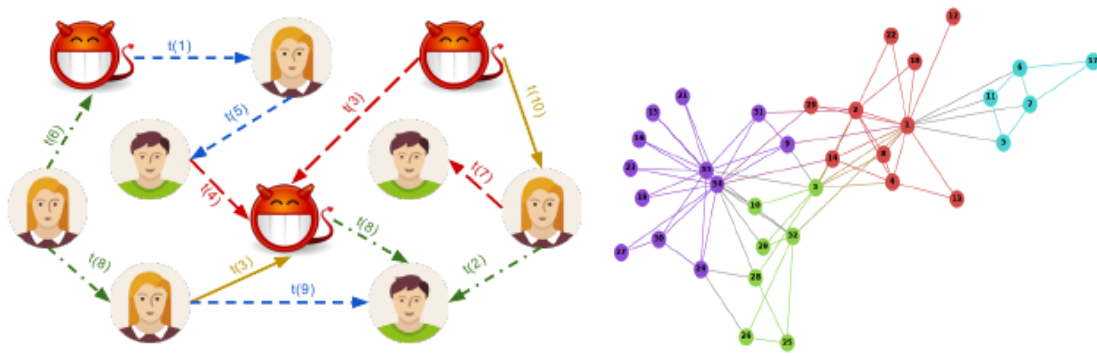


Рисунок 7. – Пример графов социальных сетей, выявления социальных групп, спамеров

Совместное использование информации графовых моделей и ML, позволяют получать скрытые зависимости и выполнять предиктивный анализ информации, получать ответы в режиме реального времени, реализовывать алгоритмы искусственного интеллекта, отслеживать решения ИИ. В настоящее время алгоритмы ИИ широко распространены для решения конкретной задачи, машина обучается выполнять какую-то задачу, например, автономное вождение автомобилей, управление различными дронами, автоматический поиск фото друзей на фотографиях и т.д. Для решения большинства из этих задач применяется графовая аналитика.

Графовая аналитика позволяет выявить закономерности в данных, например, в социальных данных. обнаружить сообщества или группу лиц, предсказать их поведение. Графическая визуализация помогает понять невидимые процессы ML алгоритмов, которые позволяют компьютерам учиться на примерах без явного программирования. Процесс глубокого обучения использует глубокие искусственные нейронные сети и ML в качестве моделей.

Так, на рисунке 8 и 9 приведены представления преобразования графовых данных для моделирования различных процессов в ML [11]. Для моделирования и предсказания из графовых моделей могут выполняться вырезки данных, а из отношений выбираются атрибуты, по которым нужно выполнять анализ данных.

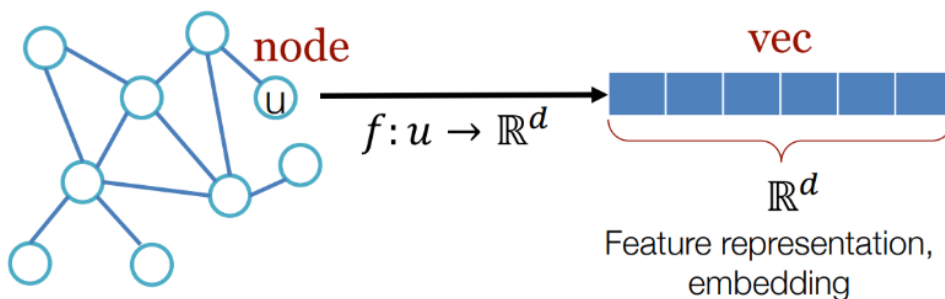


Рисунок 8. – Схема преобразования атрибутов графовой модели в векторное представление



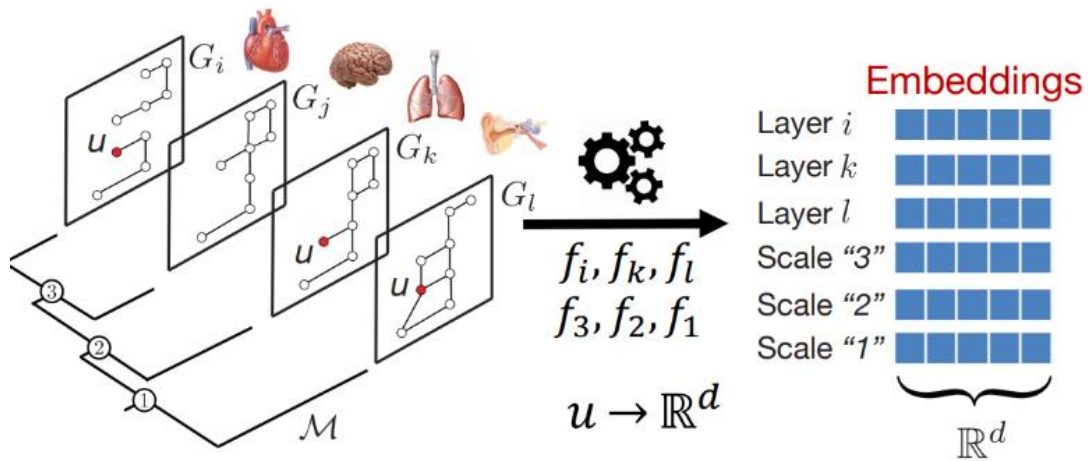


Рисунок 9. – Многоуровневая схема преобразования атрибутов графовой модели в векторное представление

Искусственные нейронные сети и процесс глубокого обучения также используют графовые модели (см. рисунок 10).

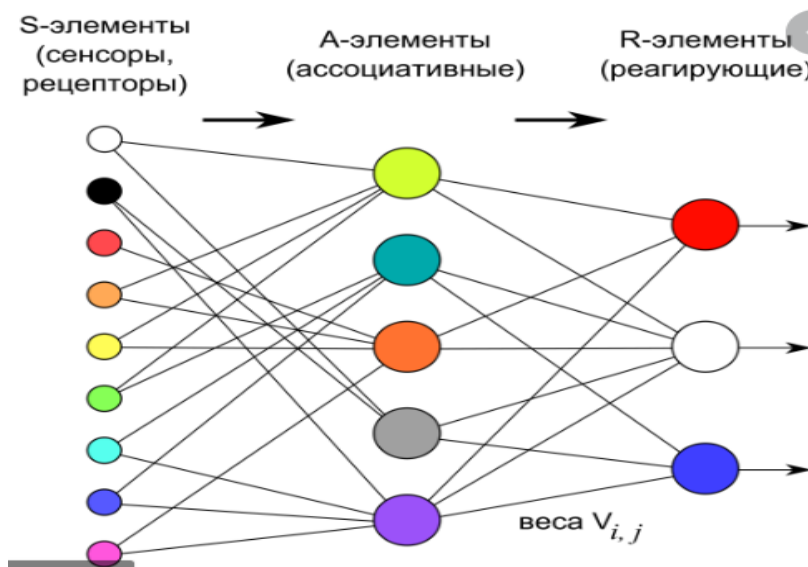


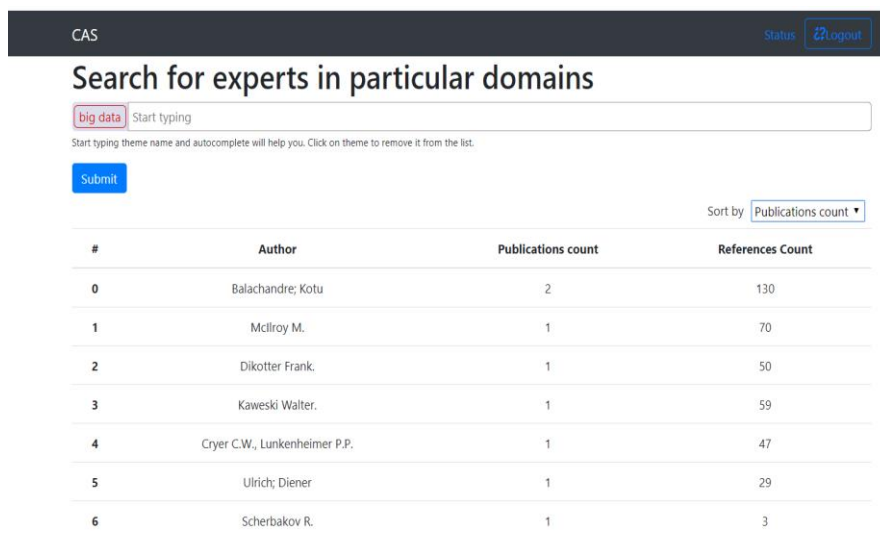
Рисунок 10. – Пример графовой модели схемы искусственной нейронной сети

## 5. Примеры применения ИС Анализа Данных

В настоящее время (написание статьи) СКАД ИИ находится в опытной эксплуатации на ЦОД БГУИР на этой базе создается «ИС Анализа Данных».

В данном разделе приведены результаты работы СКАД ИИ с учетом важности публикаций и их тематик. Графовая модель позволяет получать знания о публикациях в различных аспектах, например, связанных с тематикой «биология». В таких запросах важно указывать порог вероятности тематики в статьях больше некоторой величины. Ниже приведены примеры получения информации из БД в виде графа знаний.

Данные получены с сайтов, который используется для публикаций научных работ: <http://libgen.io>, <http://gen.lib.rus.ec/>, <https://arxiv.org>



The screenshot shows a search interface with a search bar containing 'big data'. Below the search bar is a table of results. The table has columns for '#', 'Author', 'Publications count', and 'References Count'. The results are sorted by 'Publications count' in descending order.

#	Author	Publications count	References Count
0	Balachandre; Kotu	2	130
1	McIlroy M.	1	70
2	Dikotter Frank.	1	50
3	Kaweski Walter.	1	59
4	Cryer C.W., Lunkenheimer P.P.	1	47
5	Ulrich; Diener	1	29
6	Scherbakov R.	1	3

Рисунок 11. – Эксперты в областях знаний (область – big data)

Показывается список экспертов в одной или нескольких областях знаний.

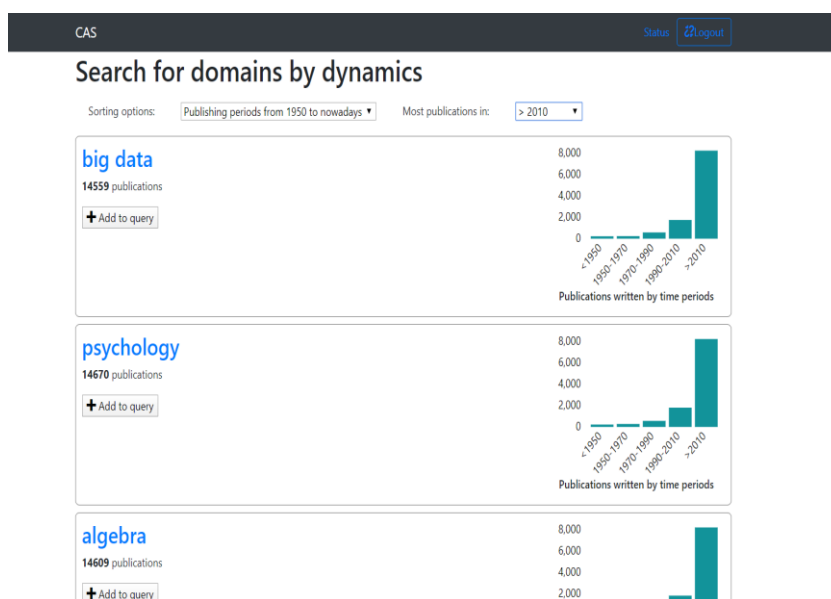


Рисунок 12. – Гистограмма «областей знаний» с 1950 года

Приведены области знаний с количеством публикаций в каждой из них и столбчатой диаграммой распределения публикаций по периодам времени.

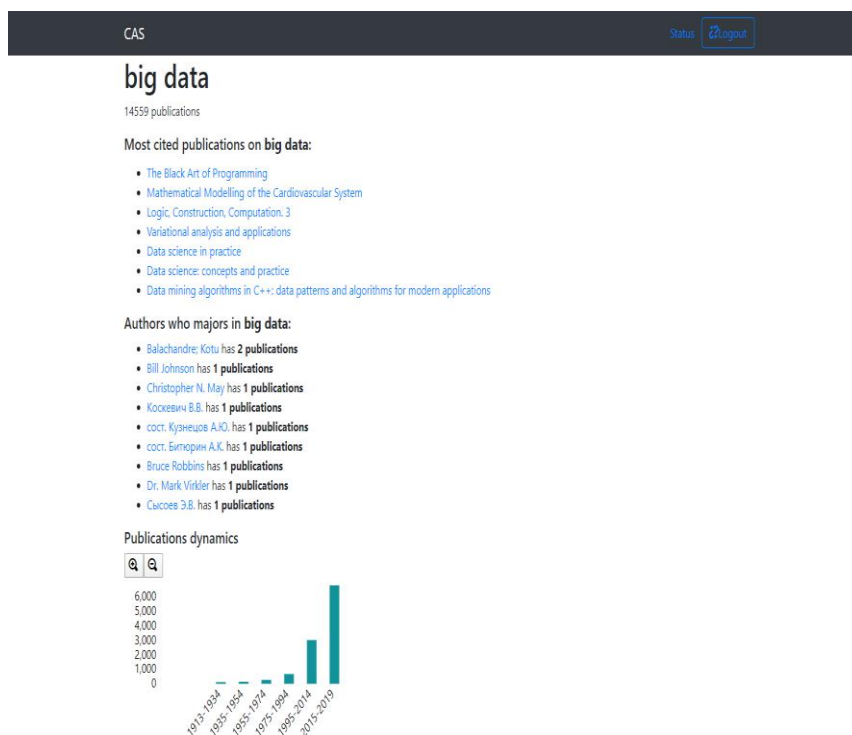


Рисунок 13. – Подробная информация о области знаний

Страница со списком наиболее цитируемых работ, списком авторов с наибольшим числом публикаций в данной области, масштабируемой диаграммой распределения публикаций по периодам времени.

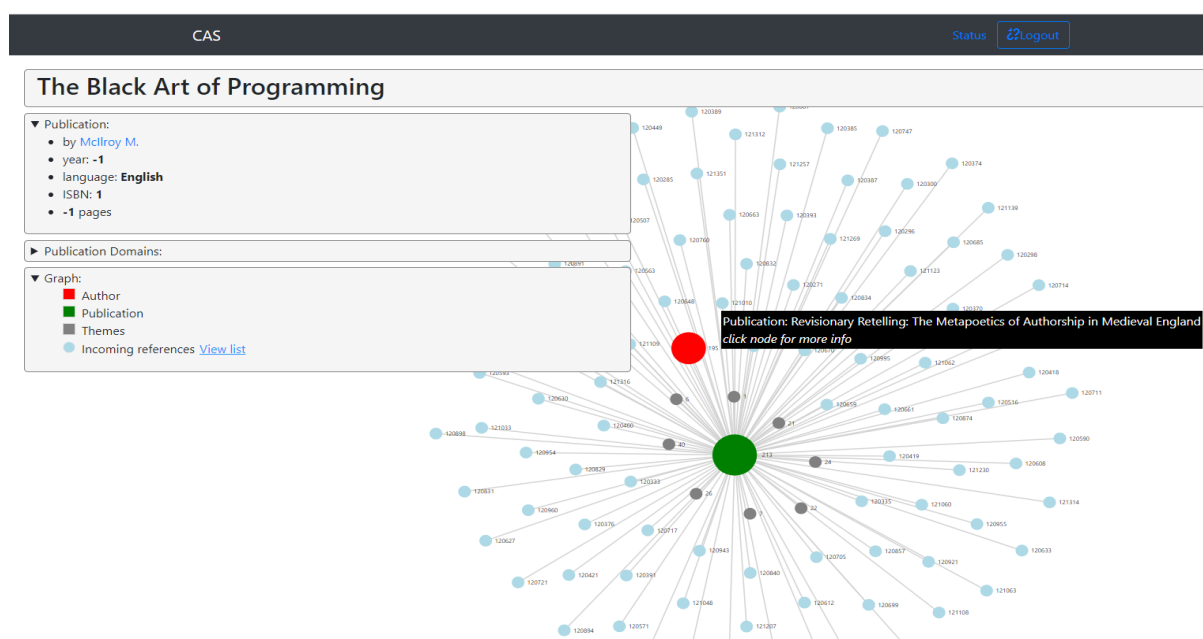


Рисунок 14. – Подробная информация о конкретной публикации

Страница со списком тем, к которым относится публикация, со списком входящих (не более 50) и исходящих (не более 50) ссылок. Дополнительно – список ФИО и ID авторов. Из этой информации строится визуальный граф.

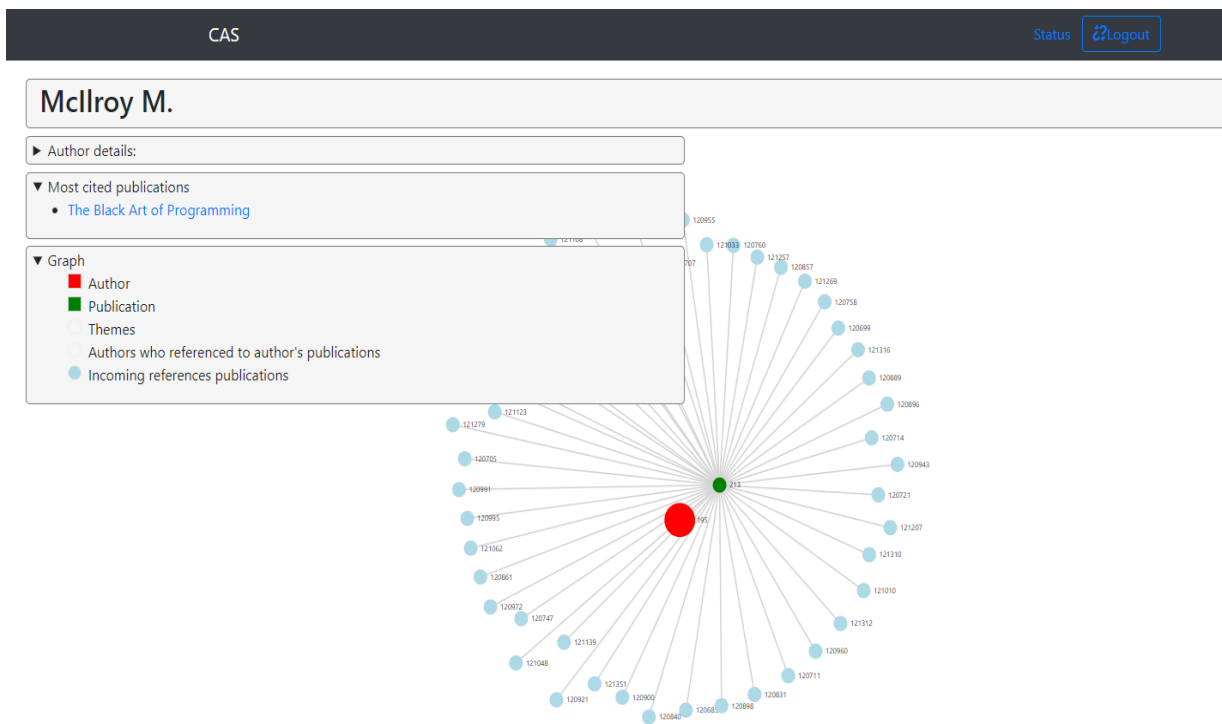


Рисунок 15. – Подробная информация о авторе

Страница с информацией об авторе, со списком его наиболее цитируемых публикаций (не более 10) и исходящих (не более 50) ссылок для каждой из них. Для каждой публикации-«ссылки» дополнительно отправляется ее автор. Примечание. При большом количестве ссылок запросы 14 и 15 являются наиболее затратными по времени.

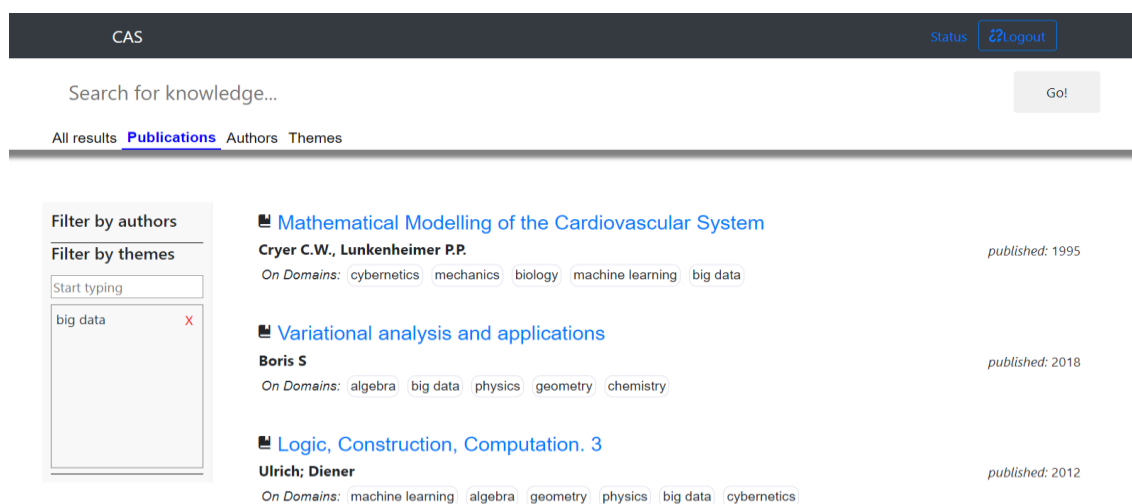


Рисунок 16. –«Универсальный поиск» – поиск публикаций, связанных с темой «big data»

Страница со списком из 10 ответов-ссылок на подробности, удовлетворяющих запросу. К каждому из ответов, в зависимости от его типа (публикация/автор/тема) подгружаются дополнительные характеристики (количество публикаций, топ-5 тем, к которым относится публикация и т.д.). После получения первых 10 результатов возможен запрос на следующие 10, т.е. постраничная загрузка.

CAS Status Logout

computer Go!

[All results](#) [Publications](#) [Authors](#) [Themes](#)

**Computer science: the hardware, software and heart of it**  
Alfred V. Blum published: 2011  
On Domains: linguistics cybernetics big data

**Computer vision. Principles, algorithms, applications, learning**  
Davies E.R. published: 2018  
On Domains: mechanics geometry machine learning physics

**Computer algebra in scientific computing: 20th International Workshop, CASC 2018, Lille, France, September 17-21, 2018, Proceedings**  
V. P. Koepf published: 2018  
On Domains: biotechnology physics psychology geology linguistics algebra cybernetics machine learning big data

**Computer Games and Technical Communication: Critical Methods and Applications at the Intersection**  
Jennifer deWinter published: 2014

Рисунок 17. – Поиск по запросу «Computer»

Пример, получения более сложной интеллектуальной информации о данных из предметной области, на который скорее всего невозможно получить ответ из не графовой БД, приведен ниже.

Вопрос: *Какие авторы наиболее часто сотрудничали с автором данной популярной статьи, является ли данный автор писателем только в области биологии или же он пишет еще и на темы математического анализа, существуют ли математические публикации, которые по какой-то причине перекликаются с темой философии.*

Запрос для известной публикации (ее ID=10) возвращает 10 авторов, которые имеют наибольшее число статей, ссылающихся на данную статью, выстроенных по убыванию числа ссылающихся статей. Плюс, возвращается сама публикация с ID=10 и ее автор (см. рисунок 18).

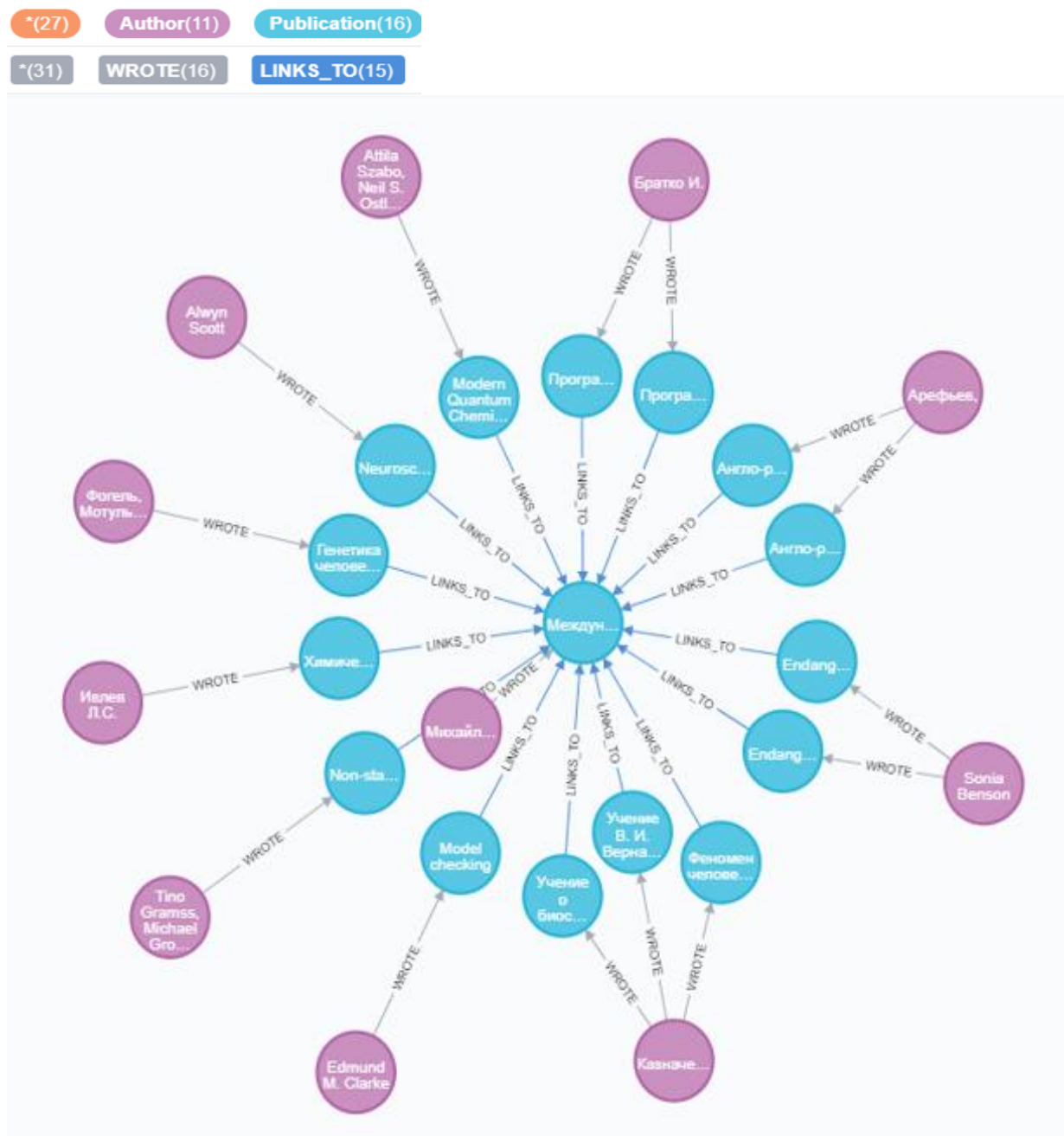


Рисунок 18. – Какие авторы наиболее часто сотрудничали с автором данной популярной статьи

Запрос для известного автора (ID=8220) возвращает его публикации, которые не имеют связей с темой под названием “biology”. Из всех связей, обозначающих отношение тематики (THEME\_RELATION) берутся лишь те, вероятность правильности отношения статьи к конкретной тематике у которых более 0.7 (см. рисунок 19).

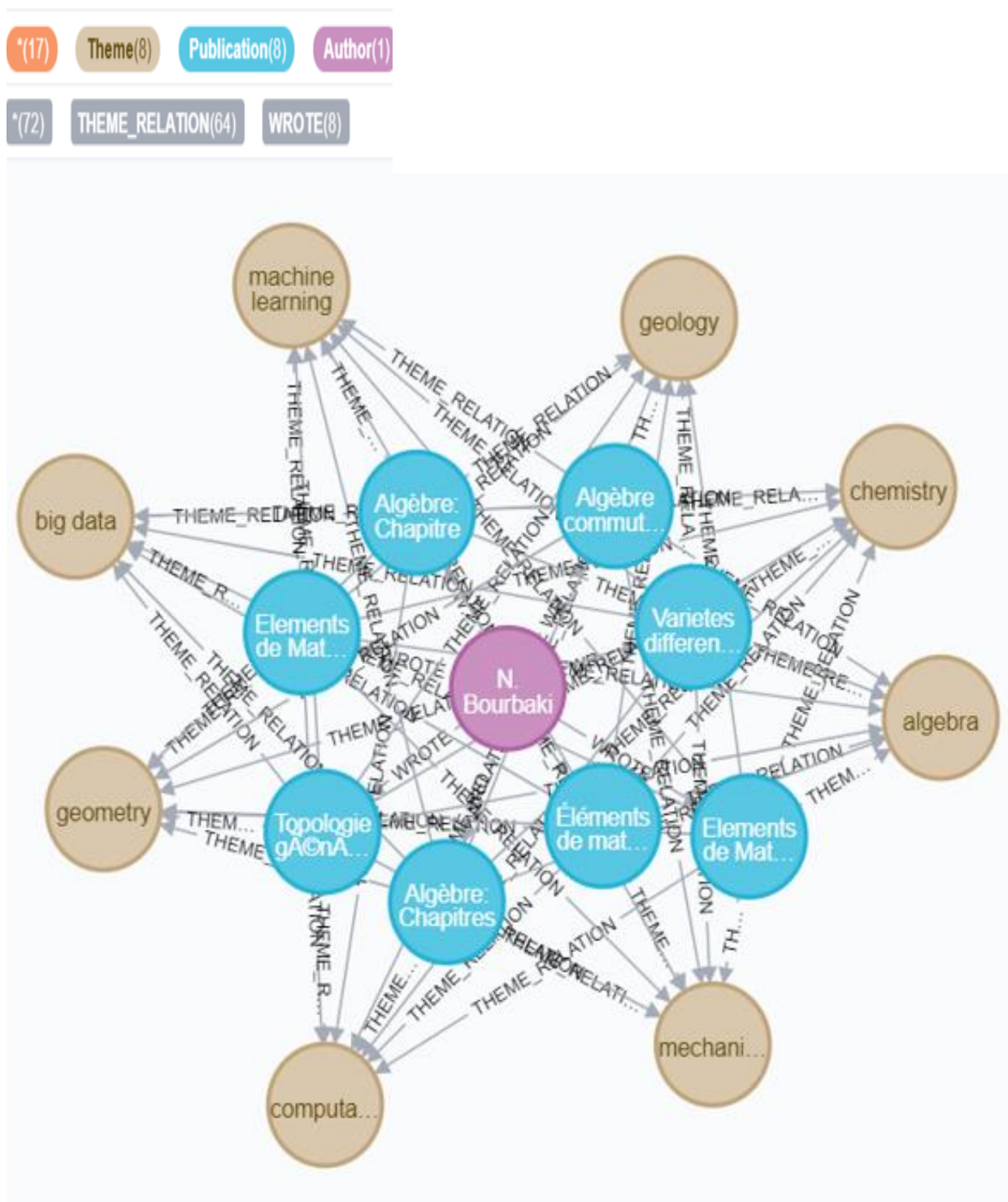


Рисунок 19. – Является ли данный автор писателем только в области биологии или же он пишет еще и на темы математического анализа

Запрос выбирает публикации, относящиеся к теме “algebra” с вероятностью, большей 0.7, имеющие также отношение к другим тематикам (с вероятностью большей 0.7). И, для построения графа, запрос возвращает тему “algebra” и, для каждой публикации, список других тем, к которым она относится (см. рисунок 20).

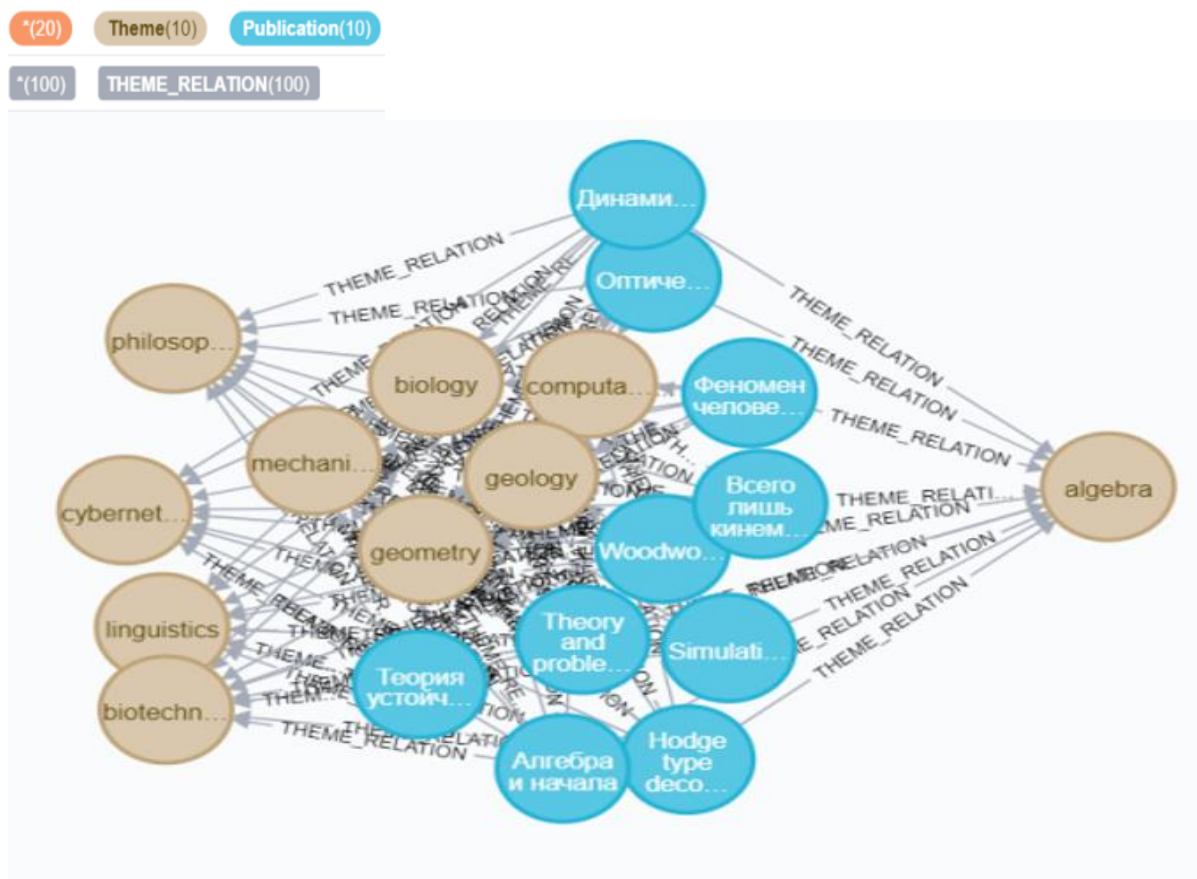


Рисунок 20. – Существуют ли математические публикации, которые по какой-то причине перекликаются с темой философии

## 6. Заключение

«Интеллектуальная система комплексного анализа данных интернет-источников» – это инновационный научно-образовательный проект БГУИР. Результаты выполнения проекта используются при обучении магистрантов по тематике «Обработка больших объемов информации», подготовке специалистов «Data Scientist», а также для получения экспертных данных при проведении исследовательских работ в университете. В настоящее время над проектом работает коллектив студентов, которые уже получили навыки работы в команде над большим проектом и знания по извлечению данных из интернет-источников, их обработке и анализу с помощью NLP и ML алгоритмов.

Основными результатами СКАД ИИ являются:

- интеллектуальное взаимодействие графовой базы данных с ML, расширение системы алгоритмами ИИ;
- разработка ESB СКАД ИИ, которая обеспечивает интеграцию данных и приложений (серверов). Созданная архитектура позволяет модернизировать и функционально наращивать СКАД ИИ в процессе эксплуатации. Данная архитектура построения многофункциональных комплексов позволяет работать СКАД ИИ как набор постоянно работающих компонент в виде отдельных серверов;
- компоненты и комплексы могут быть добавлены или заменены новыми;
- сами компоненты можно нарастить функционально, дополнить компоненты «хранилище», граф-знаний, «библиотека модулей» векторными алгоритмами и алгоритмами нейровычислений.



Компонент графовая база данных и граф знаний может быть дополнен графовой моделью для поиска постов в социальных сетях на определённую тематику, и другими графовыми моделями, которые могут быть максимально полезны и интересны пользователю, методами интеллектуального анализа графовой модели на основе применения ML.

Компонент «хранилище» данных может быть дополнен новыми БД, соответствующими новым областям применения системы.

Библиотеки аналитических модулей новыми ML-модулями интеллектуального анализа данных

В момент подготовки данной статьи в реализации проекта СКАД ИИ и его модернизации принимали участие студенты БГУИР кафедры информатики: Гутковский В.Н., Сернацкий В. И., Судникович К. И., Черныш Н. Н., Шпаков Н. И.

### Список литературы

[1.]Data Never Sleeps 7.0 [Электронный ресурс] / Режим доступа: <https://www.domo.com/learn/data-never-sleeps-7> Дата доступа: 14.02.2020.

[2.]Пилецкий, И. И. Аналитический комплекс анализа данных из открытых интернет источников / И. И. Пилецкий, В. А. Прытков, Н. А. Волорова // BIG DATA Advanced Analytics: collection of materials of the fourth international scientific and practical conference, Minsk, Belarus, May 3 – 4, 2018 – Minsk, BSUIR, 2018. – P. 193 – 199.

[3.]Батура М.П., Пилецкий И.И., Прытков В.А., Волорова Н.А., Козуб В.Н. Система комплексного анализа данных интернет источников // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. материалов V Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 13–14 марта 2019 года). В 2 ч. Ч. 2 / редкол. : В. А. Богуш [и др.]. – Минск : БГУИР, 2019. – 379 с. ISBN 978-985-543-484-0 (ч. 2).p/172-187

[4.]Романов А.А., Пилецкий И. И. Классификация тональности текстовых документов с помощью метода опорных векторов. Компьютерные системы и сети: материалы 53-й научной конференции аспирантов, магистрантов и студентов. – Минск: БГУИР, 2017 -06 мая 2017.

[5.]Чугаинов К. В., Пилецкий И. И. Методы тематической кластеризации новостных статей. Научно-практические исследования №2 (ISSN 2541-9528) – Омск: Дельта, – 2017 с. 295 – 298.

[6.]Прытков В.А. Нормализация словоформ при анализе репозитория университета с использованием графовой базы данных / В.А. Прытков, И.И. Пилецкий, Н.А. Волорова // Сб. мат. V межд. Науч.-практ. конф. “Big Data and Advanced Analytics (Big Data и анализ высокого уровня)”. - Минск, 2019. - с.209-220

[7.]В. Н. Козуб, И. И. Пилецкий. Использование алгоритмов обработки естественного языка и графовых баз данных для построения рекомендательной системы // BIG DATA Advanced Analytics: collection of materials of the fourth international scientific and practical conference, Minsk, Belarus, May 3 – 4, 2018 / editorial board: M. Batura [etc.]. – Minsk, BSUIR, 2018. – P. 274 – 277.

[8.]PageRank algo neo4j. [Электронный ресурс] – Режим доступа: <https://neo4j.com/docs/graph-algorithms/current/algorithms/page-rank/> Дата доступа: 22.01.2019

[9.]Шпаков Н.Н., Черныш Н.Н., Пилецкий И.И. ГРАФ ЗНАНИЙ КАК СРЕДСТВО АНАЛИЗА В СИСТЕМЕ КОМПЛЕКСНОГО АНАЛИЗА ДАННЫХ ИНТЕРНЕТ ИСТОЧНИКОВ//«GLOBAL SCIENCE AND INNOVATIONS 2019: CENTRAL ASIA» атты VI Халықар. ғыл.-тәж. конф. материалдары (X ТОМ)/ Құраст.: Е. Ешим, Е. Абиев т.б.– Нур-Султан, Мау 9-13, 2019 – 353 б. ISBN 978-601-341-186-6. С.120-123

[10.]Representation Learning on Graphs: Methods and Applications. [Электронный ресурс] / Режим доступа: <https://www.semanticscholar.org/paper/Representation-Learning-on-Graphs%3A-Methods-and-Hamilton-ing/ecf6c42d84351f34e1625aba2e4cc6526da45c74> Дата доступа: 24.02.2020

[11.]Где и как врубиться в эмбединги графов. [Электронный ресурс] / Режим доступа: <https://habr.com/ru/company/ods/blog/418727/> Дата доступа: 24.02.2020

## **SYSTEM FOR COMPLEX ANALYSIS OF DATA FROM INTERNET SOURCES**

**M. P. BATURA**

*Head of the  
Research  
Laboratory 8.1  
“New Learning  
Technologies”  
BSUIR, Doctor of  
Technical  
sciences,*

**I.I. PILETSKI**

*PhD  
Associate Professor of  
Informatics  
Department  
of the BSUIR*

**V.A. PRYTKOV**

*PhD  
Vice-rector for  
education BSUIR,  
Associate Professor*

**N.A. VOLARAVA**

*PhD  
Head of the  
Informatics  
Department of the  
BSUIR,  
Associate Professor*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus*

E-mail: [bmpbel@bsuir.by](mailto:bmpbel@bsuir.by), [ianmenski@gmail.com](mailto:ianmenski@gmail.com), [prytkov@bsuir.by](mailto:prytkov@bsuir.by), [volorova@bsuir.by](mailto:volorova@bsuir.by)

**Abstract.** The article describes the monitoring tools for open Internet sources in order to identify experts in a certain subject area, determine the topics of publications, and assess the popularity of publications. Describes the decisions taken in the development of different variants of the construction of analytical complex. The obtained results of the complex and the application of artificial intelligence methods for conducting in-depth data analysis are presented.

**Keywords:** Internet sources, Big Data, monitoring, analysis, Machine Learning, Natural Language Processing, Neo4j, Hbase, Cassandra.

УДК 681.32

## ГЛУБОКИЙ АНАЛИЗ И ДЕКОМПИЛИЦИЯ КМОП-СХЕМЫ НА ТРАНЗИСТОРНОМ УРОВНЕ В ИЕРАРХИЧЕСКУЮ СХЕМУ ИЗ ЛОГИЧЕСКИХ ЭЛЕМЕНТОВ



**Д.И. Черемисинов**

*Ведущий научный сотрудник ОИПИ НАНБ  
кандидат технических наук, доцент  
cher@newman.bas-net.by*



**Л.Д. Черемисинова**

*Главный научный сотрудник ОИПИ НАНБ  
доктор технических наук, профессор  
cld@newman.bas-net.by*

### **Д.И. Черемисинов**

*Окончил Томский государственный университет, кандидат технических наук, доцент. Круг научных интересов: программирование, логическое проектирование и тестирование дискретных систем управления, реализация параллельных алгоритмов управления.*

### **Л.Д. Черемисинова**

*Окончила Томский государственный университет, доктор технических наук, профессор. Круг научных интересов: дискретная математика, логическое проектирование и тестирование дискретных систем управления, реализация параллельных алгоритмов управления.*

**Аннотация.** Рассматривается проблема декомпиляции плоской КМОП схемы из транзисторов, заданной в формате SPICE, в иерархическую схему из логических элементов. Задача возникает при верификации лейаута СБИС, а также при перепроектировании схем. Рассматривается наиболее общий и сложный в теоретическом плане случай, когда библиотека исходных логических элементов не известна. Такая ситуация имеет место при перепроектировании схем транзисторного уровня, восстановленных из топологии, на новую технологическую библиотеку. Результатом работы процедуры декомпиляции является иерархическое SPICE-описание, состоящее из описаний идентифицированных логических элементов, соответствующих подсхемам взаимосвязанных транзисторов. Предлагается метод и программа распознавания подсхем, являющихся КМОП-вентильями.

**Ключевые слова:** КМОП схемы, верификация, перепроектирование СБИС, формат SPICE, распознавание подсхем, канонизация графов.

**Введение.** Для представления цифровых интегральных схем существует несколько уровней абстракций. Описание верхнего функционального уровня состоит из функциональных блоков, таких как защелки, триггеры, сумматоры, счетчики, каждый из них может состоять из более простых логических элементов. Логический уровень использует в качестве строительных блоков для описания схем логические элементы. Транзисторный уровень описывает схему на физическом уровне с точки зрения транзисторов и их взаимосвязей. Следующий уровень, на котором изготавливается набор фотошаблонов и собственно интегральная схема, является практически необратимым, так как процесс производства современной СБИС стоит очень дорого: только изготовление набора фотошаблонов требует затрат в несколько миллионов долларов. Поэтому прежде чем отдать проект на производственную линию для изготовления интегральной схемы разработчик должен спроектировать надежную электрическую схему и убедиться в этом, детально проверив ее в разных режимах работы и при воздействии всех возможных дестабилизирующих факторов. Перед изготовлением фотошаблонов обязательно выполняется

верификация топологии СБИС (LVS – layout versus schematic verification) проектируемой схемы.

Этап тестирования результатов проектирования становится все более ответственным этапом проектирования, на котором выявляются ошибки проектирования или устанавливается, что описание проекта на уровне транзисторов, которое будет реализовано «в кремнии», полностью соответствует спецификации проектируемого устройства. Известно, что в большая часть (до 70%) времени проектирования сложной системы используется именно на проведение верификации проектов. Недостатки верификации могут перечеркнуть все усилия разработчиков, нарушить сроки проектирования и, что наиболее важно, иметь результатом ненадежное устройство. При этом наибольшей трудоемкостью характеризуются задачи верификации схем транзисторного уровня в силу громадной сложности этих схем (сотни миллионов транзисторов) и низкого быстродействия средств их анализа. Проведение исследований в этом направлении остается актуальным в связи с отсутствием практически эффективных методов решения таких задач. Традиционный метод проверки, такой как моделирование схем переключательного уровня, является очень дорогим с точки зрения необходимых вычислительных ресурсов, так как симуляторы схем на уровне транзисторов, такие как SPICE (Simulation Program with Integrated Circuit Emphasis), работают крайне медленно для переключательных схем даже относительно небольших схем. А современные интегральные схемы могут содержать более сотни миллионов транзисторов, и примерно такое же количество соединений между ними.

Дополнить традиционные методы моделирования могут методы верификации, позволяющие установить функциональную эквивалентность двух описаний проекта на языках более высокого уровня. Схема уровня логических элементов является функциональным представлением и может использоваться для доказательства функциональной эквивалентности двух описаний схемы. Однако схема переключательного уровня является чисто структурным представлением. В связи с этим для того чтобы сделать возможной проверку функциональной эквивалентности, необходимо сначала получить функциональность, реализуемую транзисторной схемой. Задача решается путем восстановления (экстракции) из плоского структурного описания схемы транзисторного уровня иерархического структурного описания, блоки которого представляют логические элементы.

Верификация топологии интегральных схем выполняется в два этапа. На первом этапе восстанавливается описание электрической схемы СБИС из описания топологии. Электрическая схема цифровой СБИС, изготавливаемой по КМОП технологии, содержит только униполярные транзисторы, т.е. является плоским структурным описанием. Затем, во время второго этапа, электрическая схема проверяется с помощью статической верификации или с помощью моделирования [1]. Замена хотя бы части транзисторной схемы схемой более высокого уровня (уровня логических элементов значительно меньшей сложности) приводит к значительному увеличению скорости моделирования, а значит, позволяет анализировать транзисторные схемы большой размерности. К настоящему времени разработаны средства совместного моделирования схем разного уровня представления. Например, Electronic Workbench MultiSim [2] допускает совместное моделирование блоков описаний на языках VHDL, Verilog и SPICE, имитируя логическое и транзисторное описания частей одной и той же схемы.

Результаты моделирования и верификации используются проектировщиком, чтобы изменить описание топологии для устранения найденных ошибок. Для выполнения процедуры анализа за приемлемое время требуется построить по электрической схеме иерархическое структурное описание.

Средства распознавания высокоуровневых структур [3] в сетях транзисторов представляют собой ценный инструмент автоматизированного проектирования СБИС.

Проблема извлечения подсхем из сети транзисторов изучается уже в течение долгого времени. Первоначально автоматические экстракторы подсхем главным образом использовались для функциональной проверки результата физического проектирования относительно результата логического проектирования. Позже, экстракторы, извлекая структуры более высокого уровня, позволили ускорить моделирование транзисторной схемы. Извлечение иерархии также является основой логического перепроектирования на новый технологический базис [4].

Структурный анализ цифровых схем в прошлом широко исследовался [5]. Для распознавания подсхем в литературе описаны два класса подходов: библиотечные подходы и алгоритмические подходы. Подходы, основанные на библиотеке, используют схемы из библиотеки как шаблоны для распознавания подсхем, и их способность распознавания ограничена теми подсхемами, которые содержатся в библиотеке [6]. В алгоритмических подходах фрагмент схемы распознается как подсхема, если имеется возможность вычислить его логическую функцию [5]. В дальнейшем задача декомпиляции рассматривается для случая, когда библиотека подсхем не известна.

В настоящей работе разработаны и комплексированы в единый алгоритм извлечения логических подсхем из сети транзисторов методы и программные средства решения следующих задач: 1) поиск подсхем транзисторов, связанных по постоянному току; 2) распознавание подсхем, реализующих элементы, на основе набора правил, когда библиотека логических элементов не известна; 3) поиск часто встречающихся транзисторных подсхем и выделение их в качестве элементов. Предлагаемые методы для решения задачи декомпиляции сводятся к решению задач на помеченных графах большой размерности: нахождению компонент связности графов, проверки изоморфизма графов и изоморфных вложений.

Предлагаемый метод декомпиляции КМОП схемы из транзисторов реализован в виде программы на языке C++. Разработанный декомпилятор прошел апробацию на практических примерах [7].

*Постановка задачи.* Исходными данными для программы декомпиляции служит плоский нетлист КМОП схемы в формате SPICE, имя головной схемы и имена цепей питания. Результатом является иерархическое SPICE-описание, в которое включены модели всех идентифицированных КМОП-вентилей. Все шаги предлагаемого алгоритма декомпиляции КМОП схемы из транзисторов выполняются за линейное время от размерности исходных данных, поэтому программа имеет достаточное быстродействие, чтобы обрабатывать схемы из 100 тыс. транзисторов за несколько минут на персональной ЭВМ.

Предложенный метод распознавания подсхем извлекает структуру функционального уровня из схемы транзисторного уровня, собирая транзисторы в подсхемы (рисунок 1), когда библиотека логических элементов не известна. Метод реализует трехэтапный процесс. Сначала, используется структурный подход, при котором распознаются группы транзисторов, связанных по постоянному току. Затем из этих групп выделяются правильные, которые представляют собой некоторые КМОП-вентили. И, наконец, на множестве оставшихся транзисторов ищутся и выделяются в качестве псевдо элементов часто встречающиеся транзисторные подсхемы, разбивая схему на минимальное количество классов идентичных функциональных псевдо элементов.

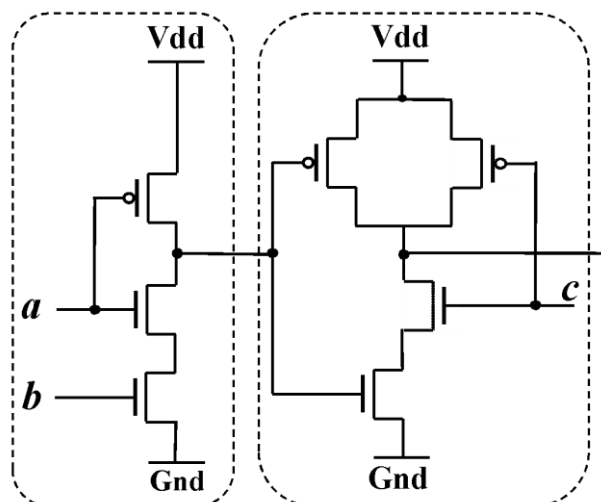


Рисунок 1. – Выделение в МОП схеме двух подсхем из транзисторов, связанных по постоянному току

Декомпиляция выполняется следующей последовательностью шагов.

1. Анализ исходного SPICE-описания плоской транзисторной схемы и построение графовой модели SPICE-описания – двудольного помеченного графа и иерархических хеш-таблиц для хранения синтаксических элементов анализируемой схемы.

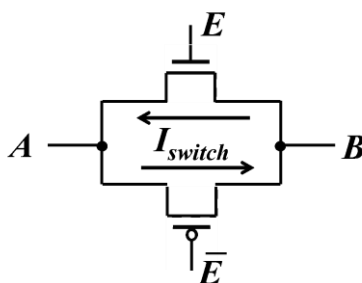


Рисунок 2. – Передаточный элемент

2. Поиск одиночных МОП-транзисторов, групп последовательно соединенных идентичных МОП-транзисторов (одного типа и с одним и тем же сигналом на затворе) и передаточных элементов (рисунок 2), представляющих собой пары связанных n-МОП и p-МОП [8].

3. Факторизация двудольного помеченного графа на группы связанных по постоянному току транзисторов (рисунок 1) .

4. Распознавание правильных групп транзисторов, которые являются КМОП-вентилем, и нахождение реализующей их булевой функции в виде алгебраической формулы (рисунок 3), а также классификация экземпляров вентиля, описываемых неизоморфными графами вследствие взаимозаменяемости стока и истока и несимметричности входов.

5. Выделение изоморфных транзисторных подсхем на множестве нераспознанных групп транзисторов.

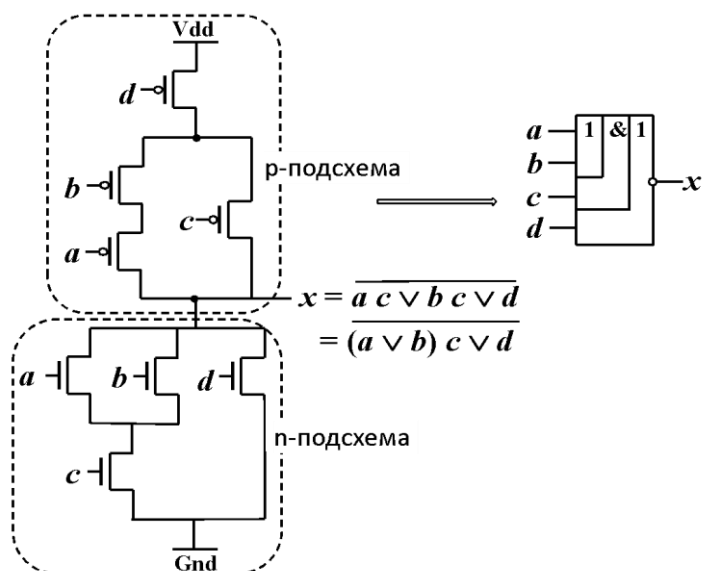


Рисунок 3. – Транзисторная схема, реализуемая им функция в виде факторизованной ДНФ и КМОП элемент

*Модель схемы на транзисторном уровне.* Одним из основных форматов для обмена электрическими схемами является SPICE (Simulation Program with Integrated Circuit Emphasis – программа моделирования радиоэлектронных компонентов). Этот формат используется в разработанном декомпиляторе для представления исходных транзисторных схем и полученных иерархических схемах из логических элементов.

Главной частью описания схемы в формате SPICE является список транзисторов, в котором для каждого вывода транзистора указано имя цепи, соединяющей его с остальными частями схемы [9]. Каждый транзистор имеет четыре вывода: сток, исток, затвор и подложка.

Общая форма описания униполярного транзистора в формате SPICE (рис. 4) имеет следующий вид:

M<name> <nd> <ng> <ns> <nb> <model-name> [L=value] [W=value],

где сток, затвор, исток и подложка соединены с цепями nd, ng, ns, и nb соответственно, model-name – тип транзистора, а L длина и W ширина. nd, ng, ns и nb – метки сети, связанные с выводами стока, затвора, истока и подложки соответствующего транзистора.

Например, список соединений (без каких-либо технических подробностей) КМОП инвертора, показанного графически на рисунок 4, выглядит следующим образом:

```
* subcircuit example
.subckt inverter in out 1
mp out in 1 1 mupmos
mn out in 0 0 mynmos
.model mupmos pmos( ... )
.model mynmos nmos( ... )
.ends
```

Список транзисторов в этом описании интерпретируется как задание бинарного отношения в виде списка пар. Например, «mp out in 1 1» является сокращенной записью пар

связей (mp.nd, out), (mp.ng, in), (mp.ns, 1), (mp.nb, 1), в которых имя «mp» транзистора вынесено.

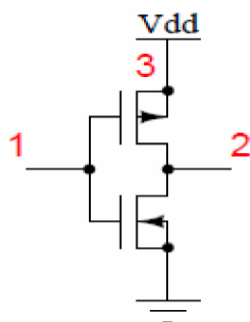


Рисунок 4. – Описание инвертора в формате SPICE и его электрическая схема

Электронные схемы на транзисторном уровне могут быть представлены в виде помеченных графов. Простой моделью для электрических схем является гиперграф, в котором вершины соответствуют транзисторам, а ребра – соединениям между ними. Однако для целей декомпиляции описания схемы на транзисторном уровне более удобной моделью является двудольный граф. Вершины двудольного графа разделены на два множества. Одно множество вершин составляют выходы транзисторов и выходы всей электрической схемы (входные и выходные порты схемы), а другое множество – соединения между выводами, т.е. цепи. Цепь является оптимальным способом представления связи между произвольным количеством транзисторов. Примером являются цепи питания и земли. К этим цепям подключено большинство элементов схемы.

Использование представления схемы в виде гиперграфа потребует значительно большего объема памяти при программной реализации, чем представление в виде двудольного графа. Можно сказать, что по требованию к объему памяти сложность первой структуры оценивается как  $O(n^2)$ , в то время как сложность второй – как  $O(n)$ , где  $n$  – число элементов схемы. Кроме того, двудольный граф является естественной формальной моделью описания схемы, заданной в формате SPICE, в котором электрические схемы состоят из элементов, которые соединены друг с другом соединениями – цепями (рисунок 1).

*Разбиение множества транзисторов на группы связанных по постоянному току.* КМОП технология позволяет реализовать функциональные элементы разными способами (стилями). Простейшим элементом является передаточный вентиль, состоящий из одного МОП-транзистора, он, однако, не обеспечивает усиление входного сигнала. Усиление сигналов обеспечивается схемой, в которой потенциал выхода элемента «подтягивается» к потенциалу шины «земли» или к потенциалу шины «питания». Элемент таким образом представляет собой группу транзисторов, связанных по току [10].

Группой транзисторов, соединённых по постоянному току, является произвольная схема из МОП-транзисторов с тремя типами внешних соединений:

- 1) Входы группы подаются только на затворы транзисторов группы;
- 2) Выходы группы подаются только на затворы транзисторов других групп;
- 3) Имеются соединения с цепями шины «питания» VDD и шины «земли» GND.

В МОП схеме группа транзисторов, соединённых по постоянному току, соответствует компоненте связности двудольного графа исходной транзисторной схемы, в которой предварительно удаляются выводы затворов (и соответствующие цепи), а также цепей питания VDD и GND, а выводы стока и истока каждого транзистора соединяются ребром. Алгоритм поиска компонент связности в таком графе построен на основе алгоритма «поиск сначала в глубину» (DFS). На рисунке 1 приведены две группы транзисторов, связанных по постоянному току.



*Распознавание подсхем, реализующих КМОП-вентили.* КМОП-вентиль состоит из  $n$ - и  $p$ -подсхем, состоящих соответственно из  $n$ -МОП и  $p$ -МОП транзисторов, включенных последовательно между цепями питания и разделенных выходной цепью (рисунок 3).

Логическая функция КМОП-вентили определяется отрицанием функции проводимости транзисторов  $n$ -МОП блока (или функции проводимости  $p$ -МОП блока при инвертировании входных переменных). Функция проводимости представляется в виде дизъюнктивной нормальной формы (ДНФ), которая задает список путей от цепи выхода до цепи питания (рисунок 3). Каждый путь представляется конъюнкцией входных переменных, управляющих проводимостью транзисторов (цепь, соединённая с затвором), входящих в этот путь.

Группа транзисторов, соединённых по току, представляет собой схему КМОП-вентиль, если

- 1) Выход является единственной цепью, соединяющей  $p$ - и  $n$ - подсхемы;
- 2) Группа содержит обе цепи питания VDD и GND;
- 3) Все пути из цепи выхода доходят до цепей питания (и наоборот);
- 4) Внутренние цепи группы не содержат выводов транзисторов, ей не принадлежащих;
- 5) Одинаковое число транзисторов в  $n$ - и  $p$ -подсхемах и одни и те же входы;
- 6) Функции проводимости  $n$ - и  $p$ -подсхем, представляемые в виде ДНФ взаимно инверсны (рисунок 3).

Для классификации схемы строится скобочная форма ДНФ функции проводимости в каноническом виде путем ее алгебраической факторизации [11]. В результате получается алгебраическое представление функции КМОП-вентили элемента в скобочной форме (рисунок 3).

При классификации КМОП-вентилей не требуется решать задачу изоморфизма соответствующих графов, но приходится учитывать не только реализуемые ими функции, но и некоторые особенности топологической реализации вентилей, которые требуют различать некоторые группы КМОП-вентилей как топологически не изоморфные на уровне транзисторов или одинаковые [7]. Это обеспечивает изоморфность исходной и декомпилированной схем. Например, разница между КМОП-вентилем, хотя и описываемыми одинаковыми логическими функциями, вызывается, например, отсутствием симметрии входов (рисунок 5) у реализующих их транзисторных подсхем (хотя сами КМОП-вентили и реализуют симметрические функции); взаимозаменяемостью стока и истока МОП-транзисторов и др. И наоборот, транзисторные подсхемы, реализующие КМОП-вентили и имеющие закороченные входы (рисунок 6), являются изоморфными аналогичным подсхемам с незакороченными входами.

*Выделение изоморфных транзисторных подсхем сводится к задаче установления изоморфизма графов, соответствующих сравниваемым подсхемам.* Как уже говорилось выше, подсхемы в формате SPICE эквивалентны помеченным двудольным графам. Причем МОП схемы имеют два типа транзисторов ( $n$ -МОП и  $p$ -МОП), каждый из них имеет по четыре вывода, соответствующие стоку, затвору, истоку и подложке. Выводы транзисторов и порты схемы (входы, выходы всей схемы) соответствуют вершинам первой доли двудольного графа. Соответственно каждая из вершин первой доли может иметь одну из восьми возможных пометок, если она соответствует выводу транзистора или пометку, большую 8, если соответствует портам схемы.

На этапе поиска изоморфных транзисторных подсхем имеется множество помеченных графов, причем они являются разреженными [12]. Для упрощения задачи канонизации [13] графы, представляющие подсхемы, дополнены ребрами, связывающими все выходы для каждого транзистора.

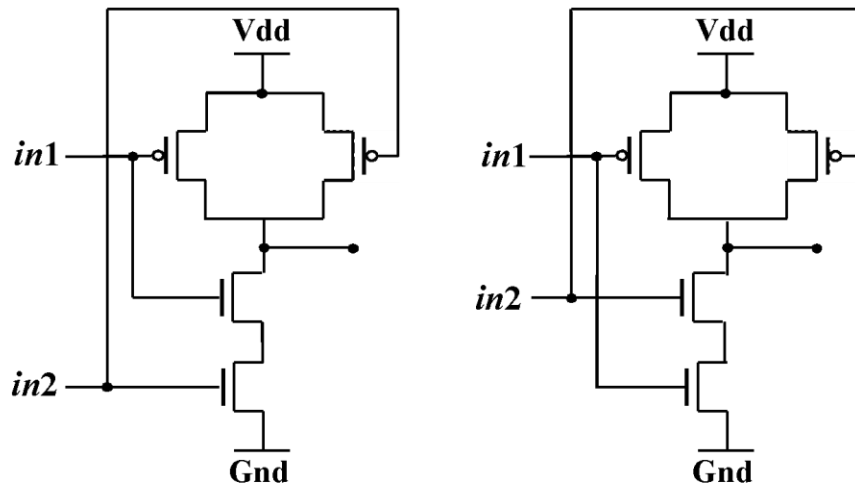


Рисунок 5. – Неизоморфные КМОП реализации элемента И-НЕ

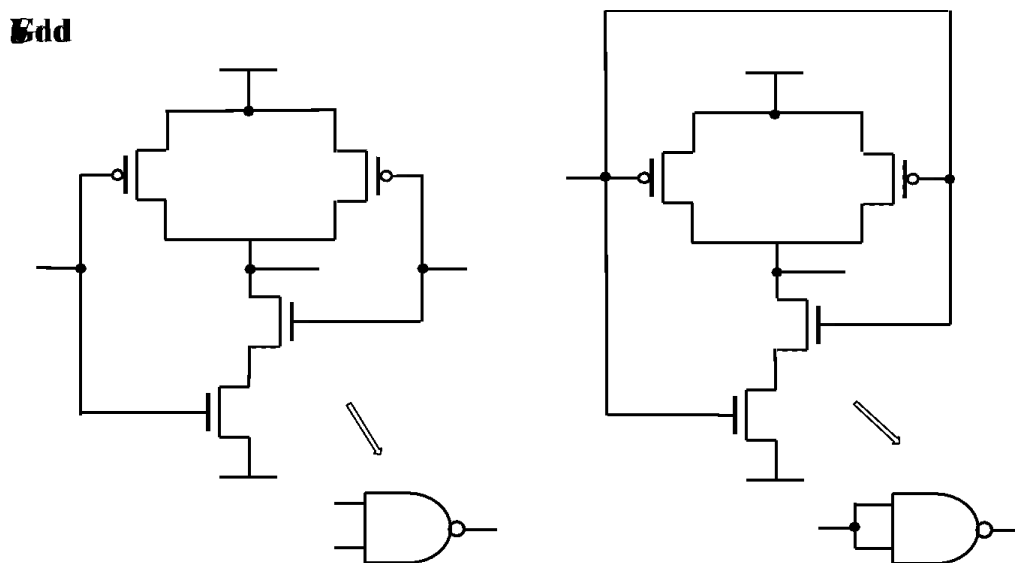


Рисунок 6. – Изоморфные КМОП реализации элемента И-НЕ: нормального вида и с закороченными входами

Помеченные графы  $G = (V, E, f)$  и  $H = (W, F, g)$ , заданные множествами  $V$  и  $U$  вершин,  $E$  и  $F$  ребер и пометками вершин  $f: V \rightarrow P$  и  $g: W \rightarrow P$ , являются изоморфными, если такая биекция  $\Pi: V \rightarrow W$ , что для любых вершин  $v_i, v_j \in V$  их образы  $\Pi(v_i)$  и  $\Pi(v_j)$  смежны в  $H$ , если и только если они смежны в  $G$ , и если  $w = \Pi(v)$ , то  $f(v) = g(w)$ . Необходимым (но не достаточным) условием изоморфизма двух графов является равенство числовых характеристик графов, называемых инвариантами. В качестве таких инвариантов вершин принимаются их степени (полустепени) и число вершин, отстоящих от данной на определенном расстоянии.

В каждом классе попарно изоморфных графов выбирается один граф, называемый каноническим видом любого графа из этого класса. Проверка изоморфизма пары графов сводится к построению и сличению их канонических видов, получаемых путем канонизации графов. Канонизация графа производится путем упорядочения вершин графа в соответствии с

инвариантами его вершин, не зависящими от исходной нумерации вершин. Канонизация графа подсхемы и построение по нему представления в формате SPICE, позволили свести проверку изоморфизма двух подсхем к проверке совпадения их текстовых описаний.

На этапе распознавания изоморфных графов благодаря отмеченным выше особенностям графов, представляющих транзисторные схемы, удалось предложить метод, позволяющий канонизировать их до конца. Вместо задания канонизированного графа в формате SPICE для него вычислялся хеш, представляющий каноническую маркировку вектора степеней вершин графа как последовательности чисел в битовую строку, имеющую длину слова. Графы изоморфны, если хэши их канонизированных форм совпадают.

*Испытания программы.* Выполнено две группы экспериментов. Для первого эксперимента использовались схемы, полученные автоматическим синтезом в САПР. В этом случае известна технологическая библиотека. В эксперименте использовалось около десятка комбинационных и последовательностных схем сложностью от сотни до 10 тыс. транзисторов. Наблюдалось стопроцентное покрытие схемы транзисторного уровня логическими вентилями.

Во втором эксперименте использовалось около десятка схем из практики проектирования, извлеченных из топологии. Устройства содержали от 5 до 50 тыс. транзисторов. В некоторых схемах кроме МОП-транзисторов имелись и другие примитивные элементы (биполярные транзисторы и RC элементы). Здесь наблюдалось покрытие схемы транзисторного уровня логическими вентилями на уровне 60–70 процентов.

Верификация правильности декомпиляции осуществлялась программой netgen [14]. Часть декомпилированных схем была верифицирована Mentor Graphics Calibre nmLVS. Во всех случаях декомпилированная схема успешно проходила проверку LVS.

На компьютере с четырёхядерным процессором Intel i5-4460 3.20GHz с оперативной памятью 16,0 ГБ декомпиляция схемы с 50 тыс. транзисторов выполнялась за время меньше минуты.

*Заключение.* В работе решены сложные в комбинаторном плане задачи, позволившие создать программные средства для решения задачи декомпиляции КМОП-схемы на уровне транзисторов в иерархическую схему из логических элементов. Практическая значимость работы состоит в том, что разработанные алгоритмы и программные средства могут быть использованы для решения важных практических задач из области проектирования СБИС: 1) функциональной верификации результатов физического проектирования; 2) ускорения процесса моделирования схем на транзисторном уровне; 3) перепроектирования (reengineering) СБИС на новый технологический базис.

### **Список литературы**

- [1]. Logic Gate Recognition in Guardian LVS – Silvaco. In [http://www.silvaco.com/content/appNotes/iccad/2-003\\_LogicGates.pdf](http://www.silvaco.com/content/appNotes/iccad/2-003_LogicGates.pdf) (access date: 4.11.2019).
- [2]. Seok, Moon Gi. Framework for simulation of the Verilog/SPICE mixed model: Interoperation of Verilog and SPICE simulators using HLA/RTI for model reusability / Moon Gi Seok, Daejin Park, Geun Rae Cho, Tag Gon Kim // 22nd International Conference of Very Large Scale Integration (VLSI-SoC), Mexico, Oct. 6–8, 2014, DOI:10.1109/VLSI-SoC.2014.7004185
- [3]. Nian, Zhang. The subcircuit extraction problem / Zhang Nian, D.C. Wunsch, F. Harary // IEEE Potentials. – 2003. – Vol. 22. – № 3. – P. 22. – 25.
- [4]. Hunt, V.D. Reengineering: Leveraging the Power of Integrated Product Development / V.D. Hunt. – Wiley. – 1993. – 283 p.
- [5]. Yang, L. FROSTY: A program for fast extraction of high-level structural representation from circuit description for industrial CMOS circuits / L. Yang, C-J.R. Shi // Integration, the VLSI Journal. – 2006. – V. 39. – № 4. – P. 311–339.
- [6]. Krasilnikova, L.V. Partition of a transistor circuit into library modules from a given library / L.V. Krasilnikova, Yu.V. Pottosin // Proceedings of the Second Intern. Conf. on Computer-Aided Design of Discrete Devices (CAD DD'97), Minsk, Republic of Belarus, November 12–14, 1997. – 1997. – Vol. 1. – P.94-97.

- [7]. Черемисинов, Д.И. Извлечение сети логических элементов из КМОП схемы транзисторного уровня / Д.И. Черемисинов, Л.Д. Черемисинова // Микроэлектроника. – 2019. – Том 48. – № 3. – С. 224–234. DOI: 10.1134/S0544126919030037.
- [8]. Рабаи, Ж.М. Цифровые интегральные схемы. Методология проектирования, 2-е издание: Пер с англ. / Ж.М. Рабаи, А. Чандраксан, Н. Борижович. – М.: ООО «И.Д. Вильямс». – 2007. – 912 с.
- [9]. Baker, R. J. CMOS Circuit Design, Layout, and Simulation / R. J. Baker. –Wiley-IEEE Press, Third Edition. – 2010. – 1214 p.
- [10]. Bushnell, M. Essentials of Electronic Testing for Digital, Memory and Mixed-Signal VLSI / M. Bushnell, V. Agrawal. – Springer Science & Business Media. – 2006. – 690 p.
- [11]. Черемисинова, Л.Д. Синтез и оптимизация комбинационных структур СБИС / Л.Д. Черемисинова. – Минск: ОИПИ НАН Беларуси. – 2005. – 235 с.
- [12]. Junttila, T. Engineering an Efficient Canonical Labeling Tool for Large and Sparse Graphs / T. Junttila T., P. Kaski. – Proceedings of the Meeting on Algorithm Engineering & Experiments, New Orleans, Louisiana, January 6, 2007. – P. 135–149.
- [13]. Закревский, А.Д. Логические основы проектирования дискретных устройств / Закревский А.Д., Поттосин Ю.В., Черемисинова Л.Д. – М.: Физматлит. – 2007. – 589 с.
- [14]. Netgen version 1.5 Tutorial. In URL: <http://opencircuitdesign.com/netgen/tutorial/tutorial.html> (access date: 16.02.2018).

## **DECOMPILATION OF CMOS CIRCUIT AT THE TRANSISTOR LEVEL INTO A HIERARCHICAL CIRCUIT OF LOGIC ELEMENTS**

***D.I. Cheremisinov***

*Leading researcher of uiip of nas of belarus, candidate of technical sciences, associate professor*

***L.D. Cheremisinova***

*Principal researcher of UIIP of NAS of Belarus, doctor of technical sciences, professor*

*United Institute of Informatics Problems of National Academy of Sciences of Belarus, Republic of Belarus {cher, cld}@newman.bas-net.by*

**Abstract.** The problem of decompilation of a flat CMOS transistor circuit in SPICE format into a hierarchical network of logic elements is considered. The problem arises in VLSI layout verification as well as in the circuit reengineering. The most general and theoretically complex case is considered when the library of logic elements is not known. This situation occurs when the transistor-level circuits are redesigned to a new technology basis. The result of the decompilation procedure is a hierarchical SPICE description consisting of logic element descriptions corresponding to recognized transistor subcircuits. The method and its program implementation for recognizing subcircuits that are CMOS gates are proposed.

**Keywords:** CMOS circuits, verification, circuit reengineering, SPICE format, subcircuit recognition, graph canonization.

УДК 004.932.2

## МОДИФИКАЦИЯ ПРОЦЕССОВ РАСПОЗНАВАНИЯ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ В ЗАДАЧАХ КОМПЬЮТЕРНОЙ ДИАГНОСТИКИ



**Л.Ю. Шилин**  
Декан ФИТиУ БГУИР,  
доктор технических  
наук, профессор



**А.А. Навроцкий**  
Заведующий кафедрой  
ИТАС БГУИР, кандидат  
физико-математических  
наук, доцент



**Р. В. Козарь**  
Аспирант кафедры ИТАС  
БГУИР

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь  
E-mail: dekfitu@bsuir.by, navrotsky@bsuir.by, pozitr0n.kozarroman@gmail.com

### **Л.Ю. Шилин**

Окончил Минский радиотехнический институт. Декан факультета информационных технологий и управления БГУИР, доктор технических наук, профессор. Является автором более 170 научных и научно-методических работ. Проводит научные исследования по анализу и синтезу дискретных систем фазовой синхронизации.

### **А.А. Навроцкий**

Окончил Белорусский государственный университет. Заведующий кафедрой информационных технологий автоматизированных систем БГУИР, кандидат физико-математических наук, доцент.

### **Р.В. Козарь**

Окончил Белорусский государственный университет информатики и радиоэлектроники. Проводит научные по распознаванию и анализу медицинских изображений.

**Аннотация.** Рассматриваются особенности распознавания медицинских изображений

**Ключевые слова:** Распознавание, оператор Собеля, медицинские изображения

**Введение.** В последнее время одним из актуальных направлений развития информационных технологий в медицине становится обработка и последующий анализ цифровых изображений. При разработке систем распознавания медицинских изображений, необходимо решить ряд вопросов: определить, какую цветность изображения лучше использовать для распознавания, какой классификатор будет более точным для обрабатываемых изображений, какие методы использовать для фильтрации и сегментации медицинских изображений и др.

Распознавание медицинских изображений имеет ряд особенностей [1], однако общими являются:

- необходимость улучшение качества изображения;
- определение подходов к распознаванию отдельных элементов;
- применение машинного обучения для последующего анализа и обработки.

В медицине, как правило, выделяются объекты, являющиеся отражением патологических процессов в организме, а классификация позволяет установить диагноз [2].

Для классификации объектов используются методы машинного обучения, такие как метод опорных векторов, дискриминантного анализа и др.

С развитием оснащённости медицинских центров задача автоматизированной диагностики патологических процессов становится все более актуальной. Для повышения качества классификации объектов на медицинских изображениях требуется выделить оптимальные методы параметрического описания этих объектов, а также определить оптимальные методы классификации и анализа с минимальным процентом ложноположительного и ложноотрицательного определения патологических процессов.

*Критерии сравнения.* В процессе анализа существующих методов были сформулированы критерии для их последующего сравнения.

Первый критерий – форма представления результата обработки изображения. Методы можно разделить на две группы: методы, результатом работы которых является обработанное изображение (алгоритмы Собеля и Канни) и методы, результатом работы которых выступают выявленные в результате обработки исходного изображения признаки и атрибуты [3].

Второй критерий сравнения – это точка приложения анализа, определяющая какой из параметров исходного изображения (например, яркостная, цветовая, градиентная или текстурная информация) будет взят для последующей обработки.

Третий критерий сравнения – применяемый метод или алгоритм для обработки.

Четвертый критерий сравнения – ограничения по цветности обрабатываемых изображений. Часть методов могут работать только с полутоновыми изображениями, что требует предварительной обработки изображения [4].

Результаты сравнения используемых для распознавания медицинских изображений методов приведены в таблице 1.

*Таблица 1. – Обзор методов распознавания медицинских изображений*

Метод	Результат	Точка приложения анализа	Используемый для обработки алгоритм	Ограничения входных параметров
Алгоритм Канни	Изображение	Разрывность яркости, интенсивность	Пороговая обработка	Нет ограничений
Алгоритм Собеля	Изображение	Разрывность яркости	Вычисление градиентов от функции яркости	Полутоновое изображение
Размерность Реньи	Число	Фрактальная размерность	Обученная нейросеть	Нет ограничений
Метод локальной бинарной текстуры	Число	Локальный бинарный шаблон	Геометрическая интерпретация, евклидово расстояние	Полутоновое изображение
Метод Харалика	Число	Текстурные признаки статистики разностей серого тона в двух соседних точках Харалика по квадратному окну		Полутоновое изображение

Отличительными особенностями медицинских изображений (например, полученных методом оптической эндоскопии) являются невысокое разрешение, наличие искажений и низкая контрастность. Основная трудность при распознавании таких изображений заключается в необходимости получении четкого контура исследуемой области. Для вычисляющий приближение градиента яркости изображения используется дискретный дифференциальный оператор Собеля, который является одним из лучших алгоритмов выделения границ.

Результатом применения оператора Собеля в каждой точке изображения является либо вектор градиента яркости в этой точке, либо его норма. Результат работы оператора показывает, насколько «резко» или «плавно» меняется яркость изображения в каждой точке, а значит, определяет вероятность нахождения точки на грани, а также ориентацию границы. Для определения нахождения границы в данной точке изображения устанавливается пороговое значение. Чем меньше порог, тем больше границ будет найдено, однако шумы будут сильно искажать результат. Высокий порог позволит избавиться от шумов, однако могут быть не определены слабые края.

Выделение границ использует два порога фильтрации: если значение пикселя выше верхней границы – он принимает максимальное значение (граница считается достоверной), если ниже – пиксель подавляется, точки со значением, попадающим в диапазон между порогов, принимают фиксированное среднее значение [5].

Типовой алгоритм Собеля использует область изображения 3x3. Для выделения контуров объектов на изображении может быть использован метод линейной фильтрации, основанный на вычислении аperiодической свертки фрагмента изображения со специальным ядром в пространственной области.

Соответственно, используемое для вычислений ядро будет определять тип линейного фильтра. Для выделения контуров методом линейной фильтрации может использоваться фильтр Собеля:

$$y_{ij} = \sqrt{\left[ (a_{i,j} - a_{i,j+2}) + 2 \times (a_{i+1,j} - a_{i+1,j+2}) + (a_{i+2,j} - a_{i+2,j+2}) \right]^2 + \left[ (a_{i,j} - a_{i+2,j}) + 2 \times (a_{i,j+1} - a_{i+2,j+1}) + (a_{i,j+2} - a_{i+2,j+2}) \right]^2}$$

Градиентные фильтры могут быть также записаны в матричной форме. Для окна размером 3x3 маска фильтра Собеля имеет следующий вид:

Для оси (ox):

$$[-1 \ -2 \ 1 \ 0 \ 0 \ 0 \ 1 \ 2 \ 1]$$

Для оси (oy):

$$[-1 \ 0 \ 1 \ -2 \ 0 \ 2 \ -1 \ 0 \ 1]$$

Фильтр Собеля основан на обработке изображения специальными масками для различных осей, действие которых фактически равносильно дискретному дифференцированию функции распределения интенсивностей пикселей, находящихся под ней [6].

Указанные выше матрицы взаимно транспонированные. Изображение представляет собой матрицу интенсивностей пикселей. Первая матрица собирает информацию вокруг пикселя  $a_{ij}$  на строке  $i$  и строке  $j$  изображения в виде суммы:

$$X = \sum_{k=1}^3 \sum_{m=1}^3 A_{1,km} b_{km}$$

Аналогичную сумму, имеющую смысл дискретной производной по направлению, ортогональному первому собирает вторая транспонированная матрица:

$$Y = \sum_{k=1}^3 \sum_{m=1}^3 A_{2,km} b_{km}$$

*Улучшенный альтернативный вариант.* Для более качественного выделения краев на медицинских изображениях предложен альтернативный диагональный вариант масок.

Для оси (оx):

$$[-1 \ -2 \ 0 \ -2 \ 0 \ 2 \ 0 \ 2 \ 1]$$

Для оси (оy):

$$[0 \ 2 \ 1 \ -2 \ 0 \ 2 \ -1 \ -2 \ 0]$$

Выделение дискретных производных здесь производится не вдоль осей (оx) и (оy), а под углом 45 градусов к ним. Матрицы при этом не будут являться взаимно транспонированными. Следовательно, интенсивность линий в модифицированном фильтре больше, а точность контура выше.

Для более точного сопоставления результатов вносится корректировочный коэффициент, равный  $\frac{4}{5}$  для всех пикселей изображения. За меру сравнения качества фильтра Собеля с различными видами масок можно использовать параметр концентрации, имеющий смысл среднего минимума интенсивности по  $h$  строкам. Число столбцов матрицы интенсивности изображения равно  $w$  (предполагается, что черный цвет имеет интенсивность равную нулю).

$$\gamma = \frac{1}{h} \sum_i^k \alpha_{ij}$$

*Заключение.* Предложенная модификация оригинального метода Собеля учитывает особенности медицинских изображений, поэтому позволяет получать лучшие результаты при распознавании.

#### **Список литературы**

- [1]. Бондаренко, А. Н. Нейросетевая классификация медицинских изображений на основе спектра размерностей Ренье [Текст] / А.Н. Бондаренко, А. В. Кацук // Сборник научных трудов НГТУ, 2005. № 1. С.1-4
- [2]. Методы распознавания медицинских изображений для задач компьютерной автоматизированной диагностики. [Электронный ресурс]. – Режим доступа: <https://www.science-education.ru/ru/article/view?id=14414>.
- [3]. Анализ медицинских изображений [Электронный ресурс] – Режим доступа: <https://postnauka.ru/faq/80995>
- [4]. Р. Каллан. Основные концепции нейронных сетей. – М.: Изд. дом «Вильямс», 2003.
- [5]. Форсайт Дэвид А., Понс Джин. Компьютерное зрение. Современный подход = Computer Vision: A Modern Approach. – М.: Вильямс, 2004.
- [6]. Диагностика медицинских изображений при помощи машинного обучения [Электронный ресурс] – Режим доступа: <http://www.vechnayamolodost.ru/articles/drugie-nauki-o-zhizni/analiz-meditsinskikh-izobrazheniy/>



## **RECOGNITION OF MEDICAL IMAGING IN THE COMPUTER DIAGNOSTIC PROBLEMS**

**L. SHYLIN**

*Dean of the Faculty of Information Technologies and Control in BSUIR, Doctor of Technical Sciences, Professor*

**A. NAVROTSKY**

*Head of the Department of Information Technologies in Automated Systems in BSUIR, Candidate of Physics and Mathematics, Associate Professor*

**R. KOZAR**

*Postgraduate student of Information Technologies in Automated Systems in BSUIR*

*Belarusian State University of informatics and radioelectronics, Republic of Belarus  
E-mail: dekfitu@bsuir.by, navrotsky@bsuir.by, pozitr0n.kozarroman@gmail.com*

**Abstract.** The features of medical image recognition are considered.

**Keywords:** recognition, Sobel operator, medical images.

УДК [005.95/.96:378]

## **ГЛУБОКИЙ АНАЛИЗ СИСТЕМЫ ПОДБОРА КАДРОВ В УЧРЕЖДЕНИЯХ ВЫСШЕГО ОБРАЗОВАНИЯ**



**А.Г. Зенкевич**  
Директор ИПК и ПК  
УО «БГУТ»



**Т.В. Казак**  
Профессор кафедры ИПиЭ БГУИР

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь  
Институт повышения квалификации и переподготовки руководителей и специалистов  
транспортного комплекса Республики Беларусь УО «Белорусский государственный университет  
транспорта», Республика Беларусь  
E-mail: sasha\_zenkevich@mail.ru*

### **А.Г. Зенкевич**

*Окончил Белорусский государственный институт инженеров железнодорожного транспорта.  
Соискатель БГУИР. Проводит научные исследования в области психологического обеспечения работы с  
кадровым резервом руководителей учреждения высшего образования.*

### **Т.В. Казак**

*Доктор психологических наук, профессор, член-корреспондент Международной академии  
психологических наук*

**Аннотация.** Рассматриваются технологии подбора и расстановки кадров в учреждении высшего образования. Предпринята попытка решения проблемы формирования качественного, высокопрофессионального состава педагогических работников, приводятся результаты анализа психологического исследования по подбору кадрового резерва руководителей в учреждении высшего образования. Даны рекомендации по совершенствованию действующей системы управления персоналом.

**Ключевые слова:** педагогические кадры, технологии подбора и расстановки кадров в учреждении высшего образования, формирование кадрового состава, кадровый резерв руководителей.

**Введение.** Осуществление коренных преобразований в экономике Республики Беларусь, расширение международных связей, необходимость неотложного решения социальных проблем требуют изменений в системе управления кадровыми ресурсами во всех сферах общественной деятельности, в том числе и педагогической.

Совершенствование подбора и расстановки кадров в учреждениях высшего образования выступает важным условием успешного решения стоящих перед обществом задач и укрепления системы высшего образования в нашей республике. Именно учреждения высшего образования и квалифицированные педагогические кадры могут воздействовать на развитие происходящих в стране изменений, обеспечивать реализацию решений, направленных на модернизацию жизни белорусского общества.

В то же время проблема формирования качественного, высокопрофессионального состава работников в любой организации является одной из сложнейших проблем любой государственной системы. Здесь поднимаются вопросы совершенствования отбора кадров, разработки научных критериев их оценки, научный подход к анализу потребностей в

педагогических кадрах, формирование эффективных технологий расстановки и продвижения кадров, повышение обоснованности кадровых решений и др.

*Материалы и методы.* Вопросами исследования человеческих ресурсов, их потенциальных возможностей успешно занимались как зарубежные, так и отечественные ученые: Г. Беккер [1], Э. Денисон [2], Дж. Кендрик [3], С. Фишер [4], С. А. Дятлов [5], Р. И. Капелюшников [6], М. М. Критский [7] и др.

По мнению Э. Денисона, «в современных условиях человеческий капитал формирует значительно больше прибавочной стоимости, чем обычный капитал» [2]. Г. Беккер рассматривал человеческий капитал как «совокупность врожденных способностей, приобретенных знаний, навыков и мотиваций, которые могут быть использованы в течение определенного времени для производства товаров и услуг и стать источниками дохода для человека, семьи, организации и общества» [1]. М. М. Критский человеческий капитал рассматривал как «основанное на экономии труда обогащение жизнедеятельности» [7, с. 23]. С. Фишер определял человеческий капитал как меру воплощенной в человеке способности приносить доход, отметив при этом, что «человеческий капитал создается тогда, когда человек (возможно с помощью своих родителей) инвестирует в самого себя, оплачивая образование и приобретение квалификации. Инвестиции в человеческий капитал со временем окупаются, давая отдачу в виде более высокой заработной платы или способности выполнять работу, приносящую большее удовлетворение» [4, с. 303]. Дж. Кендрик подчеркивал, что невещественные инвестиции в человеческий капитал, включающие затраты на образование и спецподготовку, умножая опыт и знания людей, способствуют росту производительности капитала, воплощенного в людях [3]. С. А. Дятлов выделял биологический (уровни здоровья населения и физических способностей к выполнению трудовых операций) и культурный человеческий капитал (используемая в трудовой деятельности совокупность интеллектуальных способностей, образованности, умений, навыков, квалификационной подготовки) как его составляющие [5].

Э. Денисон в качестве основы человеческого капитала рассматривает образовательный уровень совокупной рабочей силы, влияющий на рост ее производительности труда [2]. Именно «образование и подготовка на производстве увеличивают объем человеческого капитала» [6, с. 7], что должно составлять суть кадровой работы.

Среди недостаточно освещенных направлений работы с кадрами остаются новые подходы к формированию персонала, особенно методы и средства изучения личности при подборе и расстановке кадров, не систематизирован опыт оценки и продвижения персонала. Все это обуславливает потребность в проведении дополнительных теоретических исследований в области кадрового обеспечения, в поиске прогрессивных технологий кадровой работы.

Основные недостатки в подборе кадров обусловлены, нежеланием использовать и подбирать кадры, руководствуясь деловыми и нравственными критериями.

Многие организации для заполнения вакансий стремятся использовать собственных работников, прежде чем нанимать лиц со стороны. Как отмечается в [8, с. 9], внутренние источники полезны, так как дают работнику возможность продвинуться по службе или избежать увольнения, от чего он выигрывает. Нужны они и организации, поскольку она знает своих сотрудников лучше, чем претендентов извне, а кроме того, может извлекать выгоду, избегая затрат на обучение новичков, максимально используя при этом внутренние резервы.

Ранее отбор персонала для работы в государственных учреждениях, в том числе и в учреждениях высшего образования, считался простой процедурой [9, с. 93]. Руководитель беседовал с претендентами и сам их распределял, опираясь на свою интуицию. В подобных решениях содержалась большая доля субъективизма. Сегодня отбор кадров рассматривается как серьезная процедура, которая проводится компетентными специалистами. При этом чтобы достичь цели – осуществить выбор из числа претендентов такого работника, который был бы

в состоянии наиболее эффективным способом решать поставленные перед ним задачи, – должны быть проработаны методология и формы проведения этой процедуры.

Сегодня весьма важную роль при отборе персонала играют сведения о деловых качествах кандидатов на должность, запрашиваемых с прежнего места работы претендентов. Наряду с этим для оценки кандидатов на замещение вакантных должностей в государственных организациях широко используются анализ анкетных данных (конкурс документов) и собеседование (кадровое интервью). Анкетирование позволяет установить, насколько образование претендента, имеющийся у него опыт работы, наличие у него каких-либо ограничений на выполнение должностных обязанностей соответствуют квалификационным требованиям и характеру предполагаемой деятельности. Благодаря такой оценке проводится первоначальный отсев менее подходящих кандидатов.

Критерии отбора следует формировать так, чтобы они всесторонне характеризовали работника: образование, опыт, профессиональные и личные характеристики. «Эталонные» уровни требований по каждому критерию разрабатываются исходя из характеристик уже работающих в учреждении высшего образования работников, хорошо справляющихся со своими обязанностями. Отбор может быть затруднен, если список требований к работнику со стороны учреждения будет содержать слишком много необоснованных пунктов [10].

В учреждении образования «Белорусский государственный университет транспорта» было проведено исследование по подбору кадрового резерва руководителей. В эксперименте приняло участие 105 респондентов. При проведении психологического тестирования использовался пакет психодиагностического инструментария «Персонал», разработанный Екатеринбургским НИИ охраны труда и предназначенный для диагностики деловых и личностных качеств, учитывающий критерии, по которым можно оценить деловые, личностные и профессионально важные качества отбираемых лиц.

Все респонденты были разделены на 3 группы в зависимости от занимаемой в настоящий момент должности: высший уровень руководства, средний и низовой. Количество респондентов в этих группах составило 17, 24 и 64 человека соответственно.

Для каждой группы респондентов был выполнен поисковый факторный анализ с целью выявления структуры взаимосвязей между фиксируемыми в исследовании переменными. В качестве таких переменных выступили: *потребность в достижениях, способность к планированию деятельности, уровень активности, организаторские склонности, результаты опросника Томаса «Стиль поведения в конфликте», способность к эффективной управленческой деятельности.*

При проведении факторного анализа в качестве метода извлечения использовался метод главных компонент, а также применено вращение факторов по методу Varimax. На основании анализа собственных значений факторов, было установлено, что для описания факторной структуры респондентов высшего уровня руководства необходимо использовать шестифакторную модель, а для среднего и низового уровня достаточно использования пяти факторов. Ниже приведены построенные факторные модели.

В таблице 1 приведена факторная модель, построенная по результатам тестирования респондентов высшего уровня руководства. Мы построили шестифакторную модель, объясняющую 88,216% всей дисперсии.

Таблица 1. –Факторная матрица для высшего уровня руководства

Переменные	Факт. 1	Факт. 2	Факт. 3	Факт. 4	Факт. 5	Факт.6
Потребность в достижениях	-0,125	-0,521	-0,263	<b>-0,652</b>	-0,109	-0,031
Способность к планированию деятельности	0,114	-0,202	-0,053	0,085	-0,044	<b>0,925</b>
Уровень активности	0,259	<b>-0,594</b>	-0,317	0,382	-0,452	-0,274
Организаторские склонности	-0,096	-0,112	-0,182	<b>0,868</b>	-0,121	0,116
Стратегии поведения в конфликтной ситуации (опросник Томаса)						
Соперничество	0,005	0,060	<b>-0,914</b>	-0,090	0,286	0,003
Сотрудничество	-0,095	<b>-0,928</b>	-0,082	0,069	0,068	0,119
Компромисс	-0,028	-0,171	0,194	0,105	<b>-0,901</b>	0,113
Избегание	0,235	<b>0,776</b>	-0,132	0,150	0,299	-0,218
Приспособление	-0,080	0,153	<b>0,915</b>	-0,134	0,131	-0,017
Способность к эффективной управленческой деятельности						
Умение управлять собой	<b>0,724</b>	-0,016	0,159	0,513	0,259	-0,150
Личные ценности	<b>0,791</b>	0,146	-0,129	0,266	-0,013	-0,157
Личные цели	<b>0,925</b>	0,010	-0,071	-0,092	-0,115	-0,159
Саморазвитие	<b>0,953</b>	0,122	-0,010	0,170	-0,033	0,037
Навыки решать проблемы	<b>0,855</b>	0,088	0,128	-0,081	0,175	0,373
Творческий подход	<b>0,816</b>	0,165	-0,133	-0,246	0,253	-0,060
Влияние на людей	<b>0,825</b>	-0,008	-0,163	-0,303	0,059	0,047
Понимание особенностей управленческого труда	<b>0,801</b>	0,087	0,051	0,252	0,494	0,074
Навыки руководства	<b>0,782</b>	-0,129	-0,014	0,280	0,009	0,314
Умение обучать	0,530	-0,037	0,292	0,303	<b>0,548</b>	0,380
Способность формировать коллектив	<b>0,820</b>	0,085	0,166	-0,059	-0,216	0,384
Собственные значения	7,373	2,301	2,165	2,164	2,007	1,633
Доля объясненной дисперсии	36,865	11,505	10,825	10,818	10,036	8,166

В первый, наиболее мощный фактор (36,865%) вошли все составляющие способности к эффективной управленческой деятельности, за исключением умения обучать. Заметим, что для переменной, характеризующей умение обучать факторные нагрузки, соответствующие первому и пятому факторам почти совпадают (0,53 и 0,548), что позволяет отнести эту переменную как к первому фактору, так и к пятому. Итак, первый фактор определим как **способность к эффективной управленческой деятельности**.

Для четырех следующих факторов доля объясненной дисперсии почти совпадает, находясь в диапазоне от 11,5 до 10 процентов.

Второй фактор включает в себя с отрицательными знаками уровень активности (-0,594) и стремление к сотрудничеству (-0,928) и с положительным знаком стремление к избеганию (0,776). Это позволяет нам определить второй фактор как **избегание взаимодействия, уход от сотрудничества**.

Следующий фактор объединяет с противоположными знаками соперничество и (-0,914) и приспособление (0,915). На наш взгляд, его можно определить как **избегание конфронтации**.

В четвертый фактор, также с противоположными знаками вошли организаторские склонности (0,868) и потребность к достижениям (-0,652). Таким образом, для респондентов

высшего звена руководства характерны стремление к эффективной организации текущей работы при снижении стремления к достижениям. Наша интерпретация этого фактора – **организация текущей работы, решение текущих вопросов, сосредоточенность на текущей работе.**

В пятый фактор вошли стремление к компромиссу (-0,901) и умение обучать (0,548). Противоположные знаки факторных нагрузок этих переменных позволяют нам определить фактор как **авторитарность /непоколебимость.**

И шестой фактор (8,166% дисперсии) включает в себя только одну переменную – **способность к планированию деятельности.** Таким образом, для руководителей высшего звена эта переменная слабо коррелирует с остальными выделенными факторами и выступает в качестве самостоятельного фактора.

В таблице 2 приведена факторная модель, построенная для среднего уровня руководства, объясняющая 83,94% всей дисперсии.

Таблица 2. – Факторная матрица для среднего уровня руководства

Переменные	Фактор 1	Фактор 2	Фактор 3	Фактор 4	Фактор 5
Потребность в достижениях	-0,023	0,151	0,283	-0,065	<b>-0,862</b>
Способность к планированию деятельности	0,444	-0,212	0,352	-0,066	<b>0,613</b>
Уровень активности	0,022	<b>0,733</b>	0,367	-0,264	-0,137
Организаторские склонности	<b>0,630</b>	0,523	-0,008	0,221	-0,284
Стратегии поведения в конфликтной ситуации (опросник Томаса)					
Соперничество	0,014	-0,145	0,073	<b>0,926</b>	0,035
Сотрудничество	0,239	0,168	<b>0,835</b>	-0,051	-0,038
Компромисс	0,227	<b>0,725</b>	-0,215	-0,270	-0,081
Избегание	-0,094	0,196	<b>-0,834</b>	-0,110	0,063
Приспособление	-0,301	<b>-0,739</b>	0,141	-0,494	0,024
Способность к эффективной управленческой деятельности					
Умение управлять собой	<b>0,760</b>	0,250	0,278	-0,244	0,166
Личные ценности	<b>0,784</b>	0,110	0,178	-0,442	-0,043
Личные цели	<b>0,863</b>	0,098	0,222	-0,063	0,107
Саморазвитие	<b>0,849</b>	0,170	-0,124	-0,111	0,203
Навыки решать проблемы	<b>0,943</b>	0,159	-0,081	0,053	0,115
Творческий подход	<b>0,819</b>	0,238	0,281	-0,063	0,308
Влияние на людей	<b>0,939</b>	0,165	0,116	0,122	0,024
Понимание особенностей управленческого труда	<b>0,665</b>	0,227	0,272	0,301	0,424
Навыки руководства	<b>0,927</b>	0,004	0,141	0,210	-0,038
Умение обучать	<b>0,894</b>	0,022	0,181	-0,013	-0,013
Способность формировать коллектив	<b>0,949</b>	0,059	0,009	0,121	-0,001
Собственные значения	8,905	2,316	2,204	1,759	1,604
Доля объясненной дисперсии	44,524	11,581	11,019	8,795	8,019

Как видим, первый фактор этой модели, объясняющий наибольшую часть дисперсии (44,524%) включает в себя все переменные, характеризующие способность к эффективной управленческой деятельности. Так и определим этот фактор: **способность к эффективной управленческой деятельности.**

Следующий фактор, объясняющий 11,581% дисперсии включает с положительными знаками уровень активности (0,733) и стремление к компромиссу (0,835) и с отрицательным

знаком тенденцию к приспособлению (-0,739). Это позволяет нам определить фактор как **активную коммуникацию**.

Третий фактор, объясняющий 11,019% дисперсии включает в себя с противоположными знаками две стратегии поведения в конфликтной ситуации сотрудничество (0,835) и избегание (-0,834). Наша интерпретация этого фактора – **сотрудничество**.

Следующий фактор (8,795% дисперсии) содержит только одну переменную – **соперничество**. Так его и определим.

И последний, пятый фактор включает в себя с противоположными знаками потребность в достижениях (-0,862) и способность к планированию (0,613). На наш взгляд, это сочетание отражает **исполнительность, направленность на решение текущих вопросов**.

Факторная модель, для респондентов низового уровня руководства приведена в таблице 3. Построенная модель описывает 72,523% общей дисперсии.

Таблица 3. – Факторная матрица для низового уровня руководства

Переменные	Фактор 1	Фактор 2	Фактор 3	Фактор 4	Фактор 5
Потребность в достижениях	-0,050	0,127	0,084	-0,181	<b>0,817</b>
Способность к планированию деятельности	0,179	-0,199	0,092	0,402	<b>0,506</b>
Уровень активности	<b>0,542</b>	0,153	0,302	0,096	0,155
Организаторские склонности	<b>0,527</b>	0,177	0,351	0,091	0,357
Стратегии поведения в конфликтной ситуации (опросник Томаса)					
Соперничество	0,222	<b>0,820</b>	0,034	-0,300	-0,036
Сотрудничество	0,095	0,172	<b>0,743</b>	-0,179	0,350
Компромисс	-0,200	-0,161	0,129	<b>0,848</b>	-0,194
Избегание	-0,175	-0,053	<b>-0,786</b>	-0,173	0,036
Приспособление	-0,024	<b>-0,878</b>	-0,143	-0,120	-0,089
Способность к эффективной управленческой деятельности					
Умение управлять собой	<b>0,851</b>	0,125	0,092	-0,129	0,006
Личные ценности	<b>0,827</b>	-0,027	0,034	-0,250	0,069
Личные цели	<b>0,849</b>	0,066	0,145	-0,105	0,095
Саморазвитие	<b>0,721</b>	-0,212	0,033	-0,084	0,255
Навыки решать проблемы	<b>0,914</b>	0,011	0,018	0,013	0,081
Творческий подход	<b>0,851</b>	0,138	0,044	-0,039	-0,171
Влияние на людей	<b>0,865</b>	0,156	0,127	-0,085	-0,045
Понимание особенностей управленческого труда	<b>0,904</b>	0,078	0,049	-0,005	-0,147
Навыки руководства	<b>0,875</b>	0,051	0,083	0,082	0,050
Умение обучать	<b>-0,065</b>	0,311	-0,445	<b>0,534</b>	0,159
Способность формировать коллектив	<b>0,870</b>	-0,031	0,131	0,055	0,138
Собственные значения	8,036	1,828	1,713	1,497	1,430
Доля объясненной дисперсии	40,181	9,139	8,564	7,487	7,152

В первый, наиболее мощный фактор этой модели, объясняющий 40,181% дисперсии, вошли все переменные, описывающие способность к эффективной управленческой деятельности, за исключением умения обучать, а также уровень активности и организаторские склонности. Данный фактор мы можем определить как **способность к активной и эффективной управленческой деятельности**.

Следующий фактор объясняет 9,139% дисперсии и включает в себя соперничество (0,82) и приспособление (-0,878). Высокое стремление к соперничеству и избегание приспособления позволяет нам определить этот фактор как **доминирование (или авторитарность)**.

В третий фактор, объясняющий 8,564% дисперсии, вошли с противоположными знаками еще две стратегии поведения к конфликтной ситуации: сотрудничество (0,743) и избегание (-0,786). Мы определили этот фактор как **стремление к конструктивному взаимодействию**.

Четвертый фактор, объясняющий 7,487% дисперсии, составляют стремление к компромиссу (0,848) и умение обучать (0,534). На наш взгляд, эти переменные отражают **коммуникативные способности**.

Пятый фактор, объясняющий 7,152% дисперсии образуют потребность в достижениях (0,817) и способность к планированию деятельности (0,506). Мы его определили как **стремление и способность к достижению цели**.

Интересно отследить положение переменной «способность к планированию деятельности» в выделенных факторах для групп респондентов разных уровней руководства.

Для респондентов высшего уровня руководства, способность к планированию деятельности выступает в качестве отдельного фактора, практически не коррелируя ни с одной из остальных переменных.

Для респондентов среднего уровня руководства способность к планированию деятельности уже отрицательно коррелирует с потребностью в достижениях (эти переменные образуют пятый фактор). Это дает основания полагать, что респонденты среднего уровня руководства свои способности к планированию деятельности направляют, в основном, на решение текущих повседневных задач, не стремясь к более высоким достижениям.

Для респондентов низового уровня руководства эта переменная относится к пятому фактору и положительно коррелирует с потребностью в достижениях. На наш взгляд, сочетание этих качеств (способность к управлению своей деятельностью и потребность в достижениях) позволит работникам низового уровня руководства добиться значительных успехов в своей деятельности и, возможно, приведет к карьерному росту.

**Результаты.** Практическая значимость основных результатов проведенного психологического исследования заключается в том, что они могут быть использованы в процессе текущего и стратегического управления расстановкой кадров, а именно:

- улучшение коммуникаций в данном учреждении, четкое распределение полномочий и ответственности;
- широкое внедрение активных методов поиска и целенаправленной подготовки нужных для данного учреждения работников;
- совершенствование системы разработки критериев подбора: комплексный анализ требований и инструментов оценки кандидатов;
- оптимизация ресурсов при подборе персонала.

**Заключение.** В дальнейшем результаты проведенного психологического исследования позволят принять и развить новую стратегию управления персоналом в данном учреждении высшего образования.

#### **Список литературы**

- [1.] Беккер, Б. Человеческий капитал (главы из книги) / Г. Беккер // США: экономика, политика, идеология. - 1993. - № 11. - С. 109-119; № 12. - С. 86-104.
- [2.] Денисон, Э. Ф. Вклад знаний в экономический рост: межстрановой анализ / Э. Ф. Денисон // Советско-американский симпозиум экономистов : М., 8-21 июня 1976 г. ; предисл. и общ. ред. акад. Т. С. Хачатурова. - М. : Прогресс, 1978. - С. 39-46.
- [3.] Кендрик, Дж. Экономический рост и формирование капитала / Дж. Кендрик // Вопр. экономики // 1976. - № 11. - С. 141-154.
- [4.] Фишер, С. Экономика : пер. с англ. / С. Фишер, Р. Дорнбуш, Р. Шмалензи. - М. : Дело, 1998. - 864 с.



- [5.] Дятлов, С. А. Основы теории человеческого капитала / С. А. Дятлов. - СПб. : СПбУЭФ, 2004. - 469 с.
- [6.] Капелюшников, Р. И. Сколько стоит человеческий капитал России? / Р. И. Капелюшников. - М. : изд. дом Высш. школы экономики, 2012. - 76 с.
- [7.] Критский, М. М. Человеческий капитал / М. М. Критский. - Л. : Изд-во ЛГУ, 1991. - 120 с.
- [8.] Круден, Г. Дж. Зарубежный опыт управления персоналом. Ч. II : Подбор и расстановка кадров : пер. с англ. / Г. Дж. Круден, А. У. Шерман. - М. : Ин-т повышения квалификации гос. служащих Рос. акад. гос. службы при Президенте Рос. Федерации, 2001. - 111 с.
- [9.] Джавланов, О. Т. Номенклатура: эволюция отбора. Историко-политологический анализ / О. Т. Джавланов, В. А. Михеев. - М. : Луч РАУ, 1993. - 138 с.
- [10.] Колесникова, Е. Оценка эффективности социально-экономической политики предприятия / Е. Колесникова, Ю. Дуванова // Упр. персоналом. - 2007. - № 17. - С. 43-45.

## **CLASSIFICATION OF ERYTHROCYTES BY THE SPECTRAL ESTIMATES OF THEIR SURFACES' AFM-IMAGES**

**A.G. ZENKEVICH**  
*BSUIR applicant, director of  
IPK and PK educational  
institutions «Belarusian State  
University of Transport»*

**T.V. KAZAK,**  
*Doctor of Psychological Sciences,  
Professor, Corresponding Member of the  
International Academy of Psychological Sciences,  
Professor, Department of Engineering Psychology  
and Ergonomics, Belarusian State University of  
Informatics and Radioelectronics*

*Department of Engineering Psychology and Ergonomics of the Belarusian State University of Informatics and Radioelectronics, Republic of Belarus  
Institute for Advanced Studies and Retraining of Heads and Specialists of the Transport Complex of the Republic of Belarus «Belarusian State University of Transport», Republic of Belarus  
E-mail: sasha\_zenkevich@mail.ru*

**Abstract.** The technologies of selection and placement of personnel in a higher education institution are considered. An attempt is made to solve the problem of the formation of a high-quality, highly professional composition of teachers, the results of the analysis of psychological research on the selection of personnel reserve of leaders in higher education institutions are given. Recommendations are given on improving the existing personnel management system.

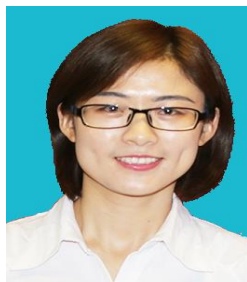
**Keywords:** teaching staff, technologies for the selection and placement of personnel in higher education institutions, the formation of personnel, the personnel reserve of managers.

УДК [004.932]

## МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И АЛГОРИТМ СКЕЛЕТИЗАЦИИ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ОРТА И ZHANG-SUEN



**Ц. Ма**  
Аспирант БГУИР  
Гражданин Китай



**С.Х. Жэнь**  
Аспирант БГУИР  
Гражданин Китай



**В.Ю. Цветков**  
заведующий кафедрой  
инфокоммуникационных  
технологий БГУИР  
доктор технических  
наук, доцент



**В.К. Конопелько**  
Профессор кафедрой  
инфокоммуникационных  
технологий БГУИР,  
доктор технических  
наук, профессор

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь  
E-mail: volos@bsuir.by

**Аннотация.** Рассматривается задача скелетизации бинарных изображений. Для построения предельно тонких связанных скелетов бинарных изображений с низкой вычислительной сложностью предложены математическая модель и алгоритм ОРСА (One-Pass Combination Algorithm) одноподитерационной скелетизации на основе комбинации и упрощения моделей одноподитерационной ОРТА и двухподитерационной Zhang-Suen (ZS) скелетизации. Эксперименте показано, что ОРСА алгоритм позволяет повышение скорости скелетизации, уменьшение избыточности связей между пикселями скелета в несколько раз.

**Ключевые слова:** скелетизации изображений, одноподитерационной скелетизации, двухподитерационной скелетизации, алгоритм ОРТА, алгоритм ZS, алгоритм ОРСА.

**Введение.** Одним из способов распознавания объектов на изображении является структурный подход, основанный на построении скелета объектов. Существует довольно много подходов, позволяющих скелетизировать изображение. В качестве наиболее известных можно взять параллельные алгоритмы ОРТА и ZS. Основная идея алгоритма ZS заключается в том, что на каждом шаге, пробегая по изображению рамкой  $3 \times 3$ , проверяется принадлежность каждого пикселя к границе заданной связной области. Если условия проверки выполняются, то пиксель удаляется из области. И вне зависимости от количества выполненных шагов область останется связной, в предельном случае она выродится в линию толщиной в один пиксель. Наиболее высокое качество скелетов обеспечивают итерационные параллельные алгоритмы. Они могут реализовываться с использованием одной или нескольких подитераций, на каждой из которых происходит удаление избыточных элементов, окрестности которых удовлетворяют определенным условиям. Для многих одноподитерационных алгоритмов характерно нарушение связности и формирование избыточных фрагментов скелета. Наиболее качественные скелеты формирует известный одноподитерационный алгоритм ОРТА, основанный на 18-ти бинарных масках, но он чувствителен к контурному шуму и имеет высокую вычислительную сложность. Благодаря относительной простоте широкую известность получил двухподитерационный алгоритм Zhang и Suen (ZS), основанный на 6-ти логических условиях, но он размывает диагональные линии толщиной 2 пикселя и удаляет области размером  $2 \times 2$  пикселя. Оба алгоритма не обеспечивают достижение минимальной толщины линий скелета [1-5].

Проблемами многих одноподитерационных алгоритмов являются нарушение связности и формирование избыточных фрагментов скелета. Для их решения в некоторых алгоритмах каждая итерация разбивается на две и более подитераций. Благодаря относительной простоте, согласно анализу [6], широкую известность получил двухподитерационный алгоритм ZS (Zhang-Suen) [7], основанный на 6-ти логических условиях. Как установлено в [8], основными недостатками алгоритма ZS являются размытие диагональных линий толщиной 2 пикселя и удаление областей размером 2x2 пикселя. Для устранения данных недостатков и развития алгоритма ZS предложены следующие его модификации: изменение порогового значения в первом логическом условии [6]; использование горизонтальных и вертикальных прямоугольных окон [9-10]. Использование расширенных окон из 11 пикселей [11-13]; расширение алгоритма на полутоновые изображения [14]; расширение условий удаления пикселей на второй подитерации для устранения чрезмерной эрозии [17]. Повышение качества скелетов в данных модификациях достигается за счет роста вычислительной сложности. Многие двухподитерационные алгоритмы, как и одноподитерационные, ориентированы на скелетизацию символов сканированных документов.

Целью работы является формирование связанных скелетов минимальной толщины для объектов произвольной формы на бинарных изображениях с низкой вычислительной сложностью.

Для бинарного изображения  $I = \|i(y, x)\|_{(y=0, \overline{Y-1}, x=0, \overline{X-1})}$  размером  $Y \times X$ , пиксели которого имеют значения 1 или 0 в зависимости от принадлежности площадному объекту или фону соответственно, алгоритмы скелетизации формируют матрицу скелетизации, значения элементов которой 1 или 0 указывают на фрагменты скелета и фона соответственно.

В алгоритме ОРТА на каждой итерации окрестность единичного элемента  $s_{ОРТА}(y, x)$  матрицы  $S_{ОРТА}$  скелетизации проходит две проверки.

В алгоритме ZS каждая итерация состоит из двух подитераций и приводит к обнулению части единичных элементов  $s_{ZS}(y, x)$  матрицы  $S_{ZS}$  скелетизации, смежные элементы в окрестности которых удовлетворяют определенным условиям.

На первой подитерации обнуляются элементы  $s_{ZS}(y, x)$ , смежные элементы которых удовлетворяют условиям

$$2 \leq \sum_{k=2}^9 p(k) \leq 6 \quad (1)$$

$$A(p(1)) = 1 \quad (2)$$

$$p(2) p(4) p(6) = 0 \quad (3)$$

$$p(4) p(6) p(8) = 0 \quad (4)$$

где  $A(p(1))$  – число комбинаций 01, встречающихся среди элементов  $p(2) - p(9)$ .

На второй подитерации обнуляются элементы  $s_{ZS}(y, x)$ , смежные элементы которых удовлетворяют условиям (1), (2) и условиям

$$p(2) p(4) p(8) = 0 \quad (5)$$

$$p(2) p(6) p(8) = 0 \quad (6)$$

Итерации продолжаются до тех пор, пока элементы  $s_{zs}(y, x)$  обнуляются.

Превосходя в качестве скелетов, алгоритм ОРТА имеет большую вычислительную сложность в сравнении с алгоритмом ZS. Это подтверждается теоретической оценкой сложности алгоритмов.

*Предлагаемая математическая модель и алгоритм скелетизации.* Для построения предельно тонких связанных скелетов одноподитерационной скелетизации на основе комбинации и упрощения моделей одноподитерационной ОРТА и двухподитерационной ZS скелетизации. Согласно предлагаемой модели на каждой итерации единичный элемент  $s_{OPCA}(y, x)$  матрицы  $S_{OPCA}$  скелетизации обнуляется (перед первой итерацией значения пикселей бинарного изображения  $I$  переносятся в матрицу  $S_{OPCA}$  скелетизации), если значения элементов выборки  $P = \left\| p(k) \right\|_{(k=2,11)}$ , формируемой из элементов окрестности  $s_{OPCA}(y, x)$ , как показано на рис. 1,  $(p(1) = s_{OPCA}(y, x))$ , удовлетворяют условию (1), а также условиям:

$$\sum_{k=2}^9 |p(k) - p(\text{mod}_8(k-2+1)+2)| = 2, \quad (7)$$

$$\neg((p(4)=1) \wedge (p(8)=0) \wedge (p(10)=0) \vee (p(6)=1) \wedge (p(2)=0) \wedge (p(11)=0)) \quad (8)$$

После выполнения всех итераций, когда не удаляется ни одного элемента скелета на основании условий (1), (7) и (8), полученная матрица  $S_{OPCA}$  скелетизации проходит еще одну проверку, в результате которой удаляются единичные элементы  $s_{OPCA}(y, x)$ , удовлетворяющие условию

$$(p(k)=0) \wedge (p(\text{mod}_8(k-2+3)+2)=1) \wedge (p(\text{mod}_8(k-2+5)+2)=1) \text{ при } k = \{3, 5, 7, 9\} \quad (9)$$

На основе рассматриваемой модели предлагается алгоритм ОРСА скелетизации на основе комбинации и упрощения алгоритмов ОРТА и ZS. Сущность алгоритма ОРСА состоит в формировании матрицы  $S_{OPCA}$  скелетизации путем переноса в нее значений пикселей обрабатываемого бинарного изображения; итеративной обработке матрицы  $S_{OPCA}$  скелетизации для удаления ее избыточных единичных элементов с использованием условий (1), (7), (8); прекращении итеративной обработки при отсутствии избыточных единичных элементов матрицы  $S_{OPCA}$  скелетизации, для которых выполняются условия (1), (7), (8); удалений избыточных единичных элементов матрицы  $S_{OPCA}$  скелетизации с использованием условия (9). В результате выполнения алгоритма ОРСА формируется матрица  $S_{OPCA}$  скелетизации, единичные элементы которой указывают на линии скелетов, соответствующих бинарным объектам на исходном изображении  $I$ .

Произведено сравнение предложенного алгоритма ОРСА с известными алгоритмами одноподитерационной ОРТА [13] и двухподитерационной ZS [15] скелетизации, скорость скелетизации  $TS$  (число удаляемых элементов в секунду; чем значение  $TS$  больше, тем скорость утоньшения выше).

Значения оценок  $ST$ ,  $TS$ ,  $MA$  для 3-ти тестовых изображений  $I(1) - I(3)$  (рис. 1) и алгоритмов ОРТА, ZS и ОРСА приведены в табл. 1. Результаты скелетизации  $S(1) - S(3)$  тестовых изображений с помощью алгоритмов ОРТА, ZS и ОРСА приведены на рисунке 1.

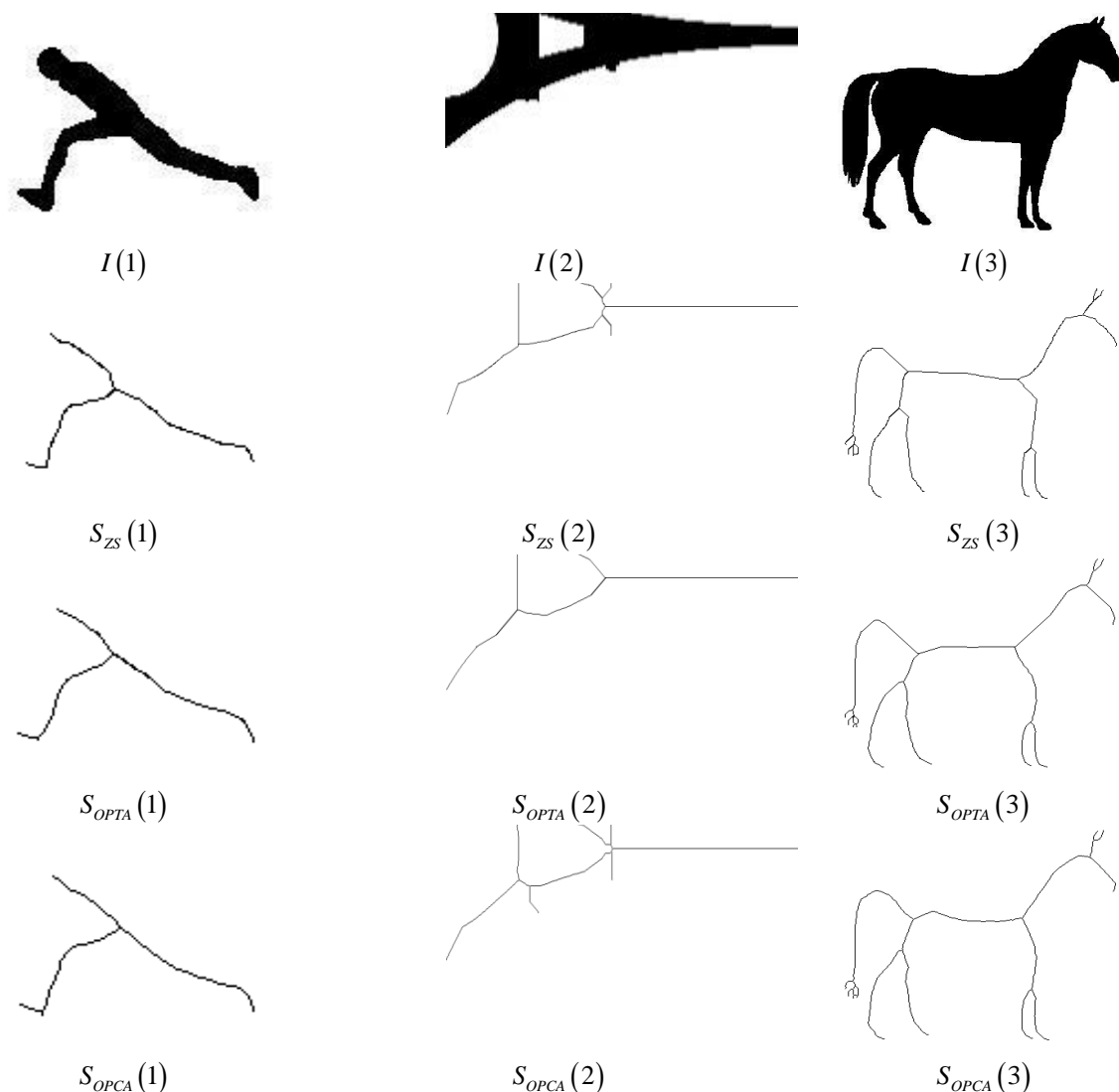


Рисунок 1. – Тестовые бинарные изображения и результаты скелетизации тестовых изображений с помощью алгоритма ОРТА, ZS и ОРСА.

**Таблица 1. – Оценки результатов скелетизации тестовых изображений  $I(1) – I(3)$**

Изображение	Алгоритм	ST	TS	MA	SC
I(1)	ОРТА	0,8631	33,83	0,8521	0,1237
	ZS	0,7833	75,67	0,9208	0,1849
	ОРСА	0,9877	89,64	0,9263	0,0183
I(2)	ОРТА	0,8965	10,75	0,9115	0,0952
	ZS	0,8784	23,597	0,9602	0,1233
	ОРСА	0,9993	36,251	0,9606	0,0032
I(3)	ОРТА	0,9796	10,846	0,8941	0,0303
	ZS	0,8350	23,023	0,9502	0,1500
	ОРСА	0,9991	34,571	0,9544	0,0058

Все оценки, приведенные в табл. 1, свидетельствуют о более высокой эффективности алгоритма ОРСА в сравнении с алгоритмами ОРТА и ZS. Предложенная оценка SC избыточности связей между пикселями скелета также как оценка ST характеризует толщину линий скелета, но обеспечивает большую точность.

*Заклучение.* Предложены математическая модель и алгоритм ОРСА одноподитерационной скелетизации на основе комбинации и упрощения моделей одноподитерационной ОРТА и двухподитерационной ZS скелетизации. ОРСА отличается от ОРТА исключением масок, предназначенных для удаления избыточных элементов на горизонтальных и вертикальных прямых линиях скелета, использованием упрощенного условия для удаления пикселей в точках изломов линий скелета, исключением масок, предназначенных для удаления избыточных концевых элементов скелета. ОРСА отличается от ZS исключением всех условий удаления пикселей, кроме двух основных. По сравнению с ОРТА и ZS алгоритм ОРСА обеспечивает уменьшение толщины скелета, повышение скорости скелетизации, и уменьшение избыточности связей между пикселями скелета.

#### **Список литературы**

- [1.] Lam, L. Thinning methodologies-a comprehensive survey / L. Lam AI // Pattern Analysis and Machine Intelligence, IEEE Transactions. – 1992. – Vol. 14. – P. 869-885.
- [2.] Jagna, A. An efficient image independent thinning algorithm. International Journal of Advanced Research in Computer and Communication Engineering. – 2014. – Vol. 3.
- [3.] S, S. I. A fast parallel thinning algorithm for the binary image skeletonization / I.S. S. Deng. Weian, N. E. Brener // International Journal of High Performance Computing Application. – 2000. – Vol. 14. – P. 65-81.
- [4.] Rosenfeld, A. Digital Picture Processing / A. Rosenfeld, A. Kak // Elsevier. – 1976
- [5.] Tarabek, P. A robust parallel thinning algorithm for pattern recognition / P. Tarabek // 7th IEEE International Symposium on Applied Computational Intelligence and Informatics. – 2012. – P. 75-79.
- [6.] Harous, S. Handwritten Character-Based Parallel Thinning Algorithms: A Comparative Study / S. Harous, Elnagar // A University of Sharjah Journal of Pure & Applied Sciences. – 2009. – Vol. 6, № 1. – P. 81-101.
- [7.] Zhang, T.Y. A fast parallel algorithm for thinning digital patterns / T.Y. Zhang, C.Y. Suen // Comm. ACM, – 1984. – Vol. 27, № 3. – P. 236-239.
- [8.] Lu, H.E. A Comment on “A Fast Parallel Algorithm for Thinning Digital Patterns”. Communications of the ACM / Lu H.E., P. Wang, S.P. – 1986. – Vol. 29, № 3. – P. 239-242.
- [9.] Abdulla, W.H. A preprocessing algorithm for handwritten character recognition. Pattern Recognition Letters / W.H. Abdulla, A.O.M. Saleh, A.H. Morad. – 1988. – Vol. 7, № 1. – P. 13-18.
- [10.] Sossa, J.H. An improved parallel algorithm for thinning digital patterns / J.H. Sossa // Pattern Recognition Letters. – 1989. – Vol. 10. – P. 77-80.
- [11.] Guo, Z. Parallel thinning with two-subiteration algorithms. Communications of the ACM / Z. Guo, R.W. Hall. – 1989. – Vol. 32, № 3. – P. 359-373.
- [12.] Guo, Z. Fast fully parallel thinning algorithms / Z. Guo, R.W. Hall // CVGIP: Image Understanding. – 1992. – Vol. 52, № 3. – P. 317-328.
- [13.] Zhang, Y.Y. A parallel thinning algorithm with two-subiteration that generates one-pixel-wide skeletons / Y.Y. Zhang, P.P. Wang // International Conference on Pattern Recognition, Vienna, Austria, – 1996. – Vol. 4. – P. 457-461.
- [14.] Kundu, M.K. A Parallel Graytone Thinning Algorithm (PGTA) / M.K. Kundu, B.B. Chaudhuri, D.D. Majumder // Pattern Recognition Letters. – 1991. – Vol. 12, № 8. – P. 491-494.

- [15.] Kwon J.S. An enhanced thinning algorithm using parallel processing / J.S. Kwon, J. W. Gi, E. K. Kang // Proceeding of the International Conf. on Image Processing. – 2001. – P. 752-755.
- [16.] Jang B.K. One-pass parallel: Analysis, properties, and quantitative evaluation / B.K. Jang, R.T. Chin // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1992. – Vol. 14. – P.1129 - 1140.
- [17.] Boudaoud, L.B. A new thinning algorithm for binary images / L.B. Boudaoud, A. Tari, A. Sider // 3rd International Based Conference on Control, Engineering & Information Technology. – 2015. – P. 1-6.

## **SINGLE-ITERATION IMAGE SKELETONIZATION BASED ON OPTA AND ZHANG-SUEN ALGORITHMS**

<b>MA JUN</b> <i>Postgraduate student of the BSUIR</i>	<b>REN XUN HUAN</b> <i>Postgraduate student of the BSUIR</i>	<b>Y.YU. TSVLATKOU</b> <i>Doctor of Engineering Sciences and Header of Infocommunication Technologies Department BSUIR, Associate professor</i>	<b>V.K.KONOPELKO</b> <i>Doctor of Engineering Sciences Professor of Infocommunication Technologies Departmen BSUIR, professor</i>
---	---	--	--

*Belarusian State University of Informatics and Radio Electronics, Republic of Belarus  
E-mail: volos@bsuir.by*

**Abstract.** This paper is focused on the field of the skeletonization of the binary image. Mathematical model and OPCA algorithm which based on a combination and simplification of single-iterative OPTA and two-iterative ZS are proposed for constructing extremely thin bound skeletons of binary images with low computational complexity. The experiment showed that the OPCA algorithm allows increasing the skeletonization speed, reducing the redundancy of the bonds between the skeleton pixels several times.

**Keywords:** skeletonization image, single-iteration skeletonization, double-iteration skeletonization, algorithm OPTA, algorithm ZS, algorithm OPCA.

УДК 004.93

## СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ



**А.С. Прокопеня**

*Аспирант кафедры ЭВС, БГУИР*



**И.С. Азаров**

*Доцент, доктор технических наук,  
заведующий кафедрой ЭВС*

*Белорусский государственный университет информатики и радиоэлектроники  
ул. П. Бровки, 6, БГУИР, каф. ЭВС, 220013, Минск, Беларусь, тел. +375 17 2938805,  
E-mail: azarov@bsuir.by*

### **Прокопеня А.С.**

*Окончил БНТУ в 2017 году, квалификация «Педагог-программист». Защитил магистерскую диссертацию по педагогике (БНТУ, 2018). Сфера научных интересов: цифровая обработка изображений, распознавание изображений.*

### **Азаров И.С.**

*В 2002 г. окончил БГУ. В 2009 г. защитил кандидатскую, а в 2015 диссертацию в БГУИР по специальности 05.13.17 «Теоретические основы информатики». Сфера научных интересов: цифровая обработка сигналов, изображений, анализ и синтез речи.*

**Аннотация.** Цель работы, результаты которой представлены в рамках статьи, заключалась в исследовании современных архитектур сверточных нейронных сетей для распознавания изображений. В статье рассмотрены такие архитектуры как AlexNet, ZFnet, VGGNet, GoogleNet, ResNet. Характеристикой о качестве распознавания изображения для нейронной сети является ошибка top-5. На основе полученных результатов было выявлено, что на данный момент сетью с наиболее точным результатом является свёрточная сеть ResNet с показателем точности в 3,57%. Преимуществом данного исследования является то, что приведенная статья дает краткую характеристику свёрточной нейронной сети, а также дает представление о современных архитектурах свёрточных сетей, их строением и качественными показателями.

**Ключевые слова:** свертка, фильтр, структура, подвыборка, функция активации.

**Введение.** Повсеместное развитие и распространение технологий компьютерного зрения (computer vision) влечет за собой изменение других профессиональных областей жизнедеятельности человека. Свёрточные нейронные сети применяются в системах распознавания объектов и лиц, специальном медицинском ПО для анализа снимков, навигации автомобилей, оснащенных автономными системами, в системах защиты, и других сферах. С ростом вычислительной мощности персональных компьютеров, а также появлением баз изображений стало возможным обучать глубокие нейронные сети (deep neural networks). В задаче распознавания изображений применяются свёрточные нейронные сети (Convolutional Neural Networks). Цель статьи – обзор современных архитектур свёрточных нейронных сетей для задачи классификации изображений.

Одна из задач машинного обучения – это задача классификации изображений. Классифицировать объект на изображении – значит указать номер, к которому относится распознаваемый объект. Для оценки алгоритмов машинного обучения обычно используются размеченные базы данных изображений, например, CIFAR-10, ImageNet, PASCAL VOC. Из-



за того, что на изображениях, например в базе изображений ImageNet может присутствовать несколько объектов, а размечен(аннотирован) только один, основным критерием ошибки является top-5 ошибка. Т.е. считается, что алгоритм не ошибся, если правильная категория объекта находится среди пяти категорий, выданных алгоритмом как наиболее вероятные. Следовательно многие нейронные сети для классификации изображений оцениваются с помощью ошибки top-5 [1].

Свёрточные нейронные сети (СНС) применяются для оптического распознавания образов, классификации изображений, детектирования предметов, семантической сегментации и других задач. Основы современной архитектуры СНС были заложены в одной из первых сетей – LeNet-5 Яна ЛеКуна.

### Структура сверточной нейронной сети

Сеть свертки представляет собой многослойный перцептрон (перцептрон, англ. perceptron от лат. perceptio – восприятие [2]) – математическая или компьютерная модель восприятия информации мозгом, созданный для распознавания 2D-поверхностей с высокой степенью устойчивости к масштабированию, преобразованиям и другим видам деформации. Обучение решению такой задачи осуществляется с подкреплением, при этом используются сети вида, архитектура которых соответствует следующим ограничениям [3].

Извлечение признаков. Каждый нейрон получает входной сигнал от локального рецептивного поля в предыдущем слое, извлекая его локальные признаки. Как только признак извлечен, его местоположение не имеет значения, т.к. приблизительно установлено его расположение относительно других признаков.

Отображение признаков. Каждый вычислительный слой сети состоит из множества карт признаков. Каждая карта признаков имеет форму плоскости, на которой все нейроны должны совместно использовать одно и то же множество синаптических весов. Эта форма структурных ограничений имеет преимущества.

Инвариантность к смещению. Инвариантность к смещению реализуется посредством карт признаков с использованием свертки с ядром небольшого размера, которая выполняет функцию "сплющивания".

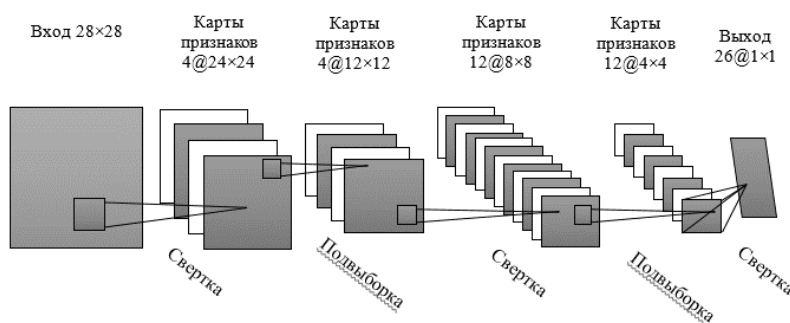


Рисунок 1. – Сеть свертки для обработки изображений

Подвыборка. За каждым слоем свертки следует вычислительный слой, осуществляющий локальное усреднение (local averaging) и подвыборку (subsampling). Посредством локального усреднения достигается уменьшение разрешения для карт признаков. Такая операция приводит к уменьшению чувствительности выходного сигнала оператора отображения признаков, а также к смещению и другим формам преобразований.

На рисунке 1 представлена схема свёрточной сети, состоящей из одного входного, четырех скрытых и одного выходного слоя нейронов. Эта сеть была создана для обработки изображений, в частности при распознавании рукописного текста. Входной слой, состоящий из матрицы 28x28 сенсорных узлов, получает изображения различных символов, которые



### Формулы для обновления весовых коэффициентов

Обновление весовых коэффициентов ( $w$ ), где  $i$  – это индекс итерации,  $v$  – переменная импульса, а  $\epsilon$  – скорость обучения показано на схеме. Скорость обучения подбиралась равной для всех слоев, а также корректировалась вручную в процессе всего обучения. Следующий шаг заключался в делении скорости обучения на 10, когда количество ошибок при валидации переставало уменьшаться. AlexNet показывает результат top-5 ошибок – 15,3% соответственно.

ZFNet – победитель ILSVRC 2013 с top-5 ошибкой 11,2%. Основную роль в этом играет настройка гиперпараметров, а именно размер и количество фильтров, размер пакетов, скорость обучения и т. д. М. Цилер и Р. Фергюс предложили систему визуализации ядер, весов и скрытого представления изображений. Система получила название DeconvNet.

Сетевая архитектура ZFNet, практически идентична архитектуре сети AlexNet. Существенные различия между ними в архитектуре заключаются в следующем:

- размер фильтра ZFNet и шаг в первом сверточном слое в AlexNet равен  $11 \times 11$ , шаг равен 4; в ZFNet –  $7 \times 7$ , шаг равен 2;
- количество фильтров в чистых сверточных слоях сети (3, 4, 5).

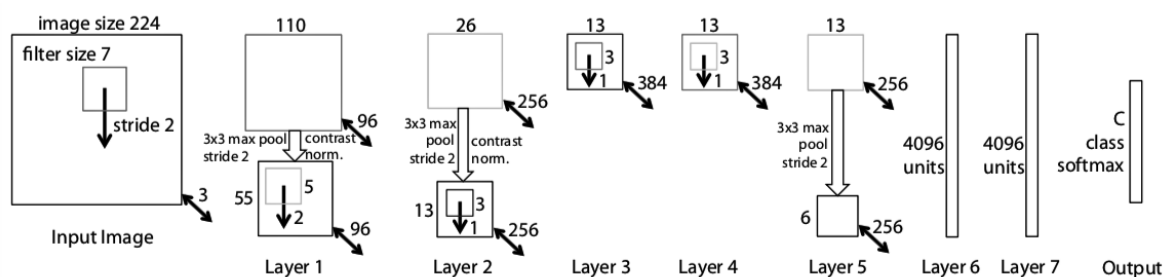


Рисунок 3 – Архитектура ZFNet

### VGGNet

В 2014 году К. Симонян и Э. Циссерман из Оксфордского университета предложили архитектуру нейронной сети, называемую VGG (Visual Geometry Group).

VGG16 является улучшенной версией AlexNet, в которой большие фильтры (размером 11 и 5 в 1-м и 2-м сверточном слое) заменены на несколько фильтров размером  $3 \times 3$ , следующих один за другим.

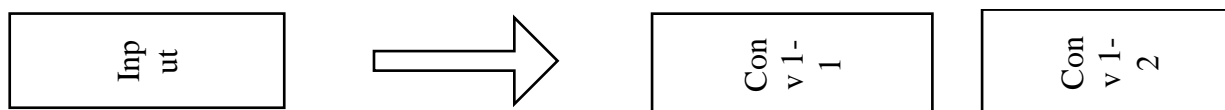


Рисунок 4. – Архитектура сети VGGNet

Во время обучения вход в ConvNets (Convolutional networks – сверточные нейронные сети) представляет собой изображение RGB фиксированного размера  $224 \times 224$  px. На следующем шаге изображение пропускается через стопку сверточных слоев размером  $3 \times 3$ . В одной из конфигураций сети VGGNet используются фильтры  $1 \times 1$ , которые можно рассматривать как линейное преобразование входных каналов.

Шаг свертки фиксируется на значении 1 пиксель. Пространственное дополнение (padding) входа сверточного слоя выбирается таким образом, чтобы пространственное разрешение сохранялось после свертки, то есть дополнение равно 1 для  $3 \times 3$  сверточных слоев.

Пространственный пулинг осуществляется при помощи пяти max-pooling слоев, которые следуют за одним из сверточных слоев (не все сверточные слои имеют последующие max-pooling). Операция max-pooling выполняется на окне размера 2x2 пикселей с шагом 2 [13].

После стека сверточных слоев идут 3 полносвязных слоя: первые два слоя имеют по 4096 каналов, третий слой – 1000 каналов (т.к. в соревновании ILSVRC необходимо распределить объекты по 1000 категориям). Последний слой – softmax. Все скрытые слои снабжены функцией активации ReLU.

Авторы продемонстрировали, что с помощью стандартных блоков можно добиться определенных результатов в конкурсе ImageNet. Число ошибок top-5 сократилось до 7,3 % [6].

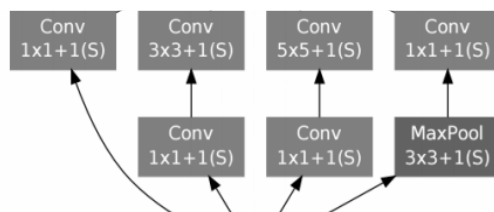


Рисунок 5. – Архитектура GoogleNet

Сверточная сеть от компании Google (GoogLeNet), известная как Inception-v1 – победитель ILSVRC 2014 с top-5 ошибкой 6,7 % [Szegedy et al., 2015].

Все свертки в сети, в том числе внутри модулей Inception, используют прямолинейную линейную активацию. Сеть насчитывает 22 слоя при подсчете только слоев с параметрами. Общее количество слоев, используемых для построения сети, составляет 100. Переход от полностью связанных слоев к среднему пулу улучшил точность топ-1 примерно на 0,6%, однако использование отсева оставалось необходимым даже после удаления полностью связанных слоев.

Учитывая глубину сети, способность распространять градиенты обратно по всем слоям была эффективной задачей. Высокая производительность более мелких сетей при выполнении этой задачи позволяет предположить, что функции, создаваемые слоями в середине сети, должны быть очень дискриминационными. При добавлении вспомогательных классификаторов, связанных с этими промежуточными уровнями, ожидалось различие на более низких ступенях в классификаторе. Это помогло побороть проблему исчезающего градиента, обеспечивая регуляризацию. Классификаторы в сети принимают форму небольших сверточных сетей, размещенных поверх выходных данных модулей Inception. Во время обучения их потеря добавляется к общей потере сети с весом. Во время вывода эти вспомогательные сети отбрасываются. Более поздние контрольные эксперименты показали, что влияние вспомогательных сетей относительно невелико (около 0,5%) и что для достижения такого же эффекта требуется только одна из них.

Точная структура дополнительной сети, включая вспомогательный классификатор, выглядит следующим образом:

- средний объединяющий слой с размером фильтра  $5 \times 5$  и шагом 3, что приводит к выходу  $4 \times 4 \times 512$  для (4a) и  $4 \times 4 \times 528$  для (4d) каскада;
- свертка  $1 \times 1$  со 128 фильтрами для уменьшения размеров и прямой линейной активации;
- полностью связанный слой с 1024 единицами и прямой линейной активацией;
- отбрасываемый слой с 70% -ным соотношением отброшенных выходов;
- линейный слой с потерями softmax в качестве классификатора (предсказывает те же 1000 классов, что и основной классификатор, но удаляется во время вывода).

В архитектуре GoogLeNet используется модуль Inception, построение сети осуществляется на основе модулей такого типа [1].

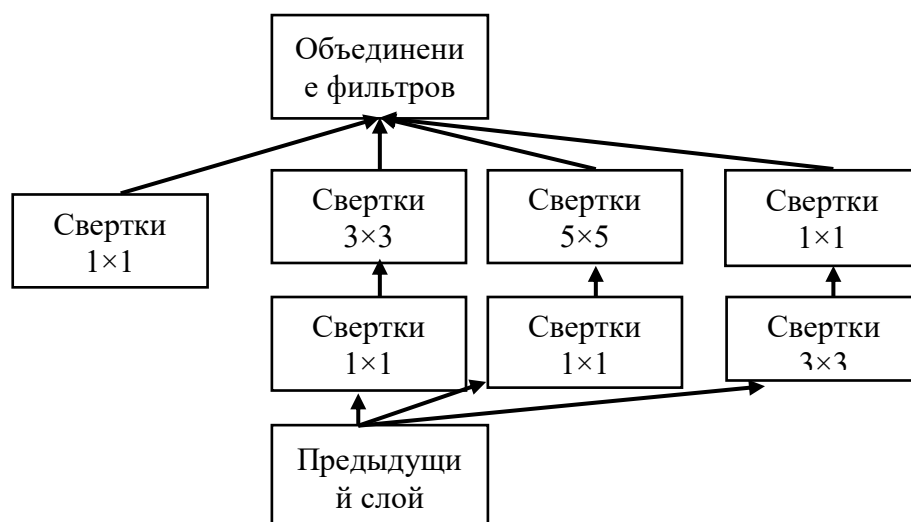


Рисунок 6 – Модуль Inception

### Модуль Inception.

Inception использует несколько ветвей (параллельных), вычисляющих различные свойства на основании одинаковых входных данных, а впоследствии объединяет полученные результаты.

Свертка размером  $1 \times 1$  является способом сокращения размерности карты свойств. Такой тип свёрточных слоев представлен в работе «Сеть» в сети М. Лина. В результате такая архитектура позволяет сократить число ошибок для top-5 категорий еще на 0.5% – до значения 6,7%.

### Модуль Inception-v2 и Inception-v3

В следующей итерации модуля Inception (Inception-v2 [7]), декомпозируется слой с фильтром  $5 \times 5$  на два слоя  $3 \times 3$ . Следующий этап – использование Batch Normalization [Ioffe, Szegedy, 2015], позволяющее увеличить скорость обучения посредством нормализации распределения выходов слоёв внутри сети. В той же статье авторы предложили концепцию модуля Inception-v3. В модуль Inception-v3, заложен принцип декомпозиции фильтров, а именно декомпозирование фильтра размером  $N \times N$  двумя последовательными фильтрами  $1 \times N$  и  $N \times 1$ . Также в Inception-v3 используется RMSProp (Метод адаптивного скользящего среднего градиентов) [Hinton, Srivasta, Swersky, 2012], вместо градиентного спуска используется усечение градиентов [Pascanu et al., 2013], которое используется для повышения стабильности обучения. Объединение из четырёх модулей Inception-v3 показал результат в категории top-5 ошибку 3,58 % на ILSVRC 2015, Inception-v2 – результат top-5 – 5.60%.

**ResNet.** ResNet – сокращенное название для Residual Network (дословно – «остаточная сеть»).

Простая Сеть. Базовые линии (рис. 7, в центре) основаны на философии сетей VGG [12] (рис. 7, слева). Сверточные слои имеют фильтры размером  $3 \times 3$ , и следуют правилам проектирования:

- при одинаковом размере карты выходных объектов, слои имеют одинаковое количество фильтров;
- если размер карты объектов уменьшится вдвое, число фильтров напротив, удваивается для того, чтобы сохранить сложность времени для слоя.

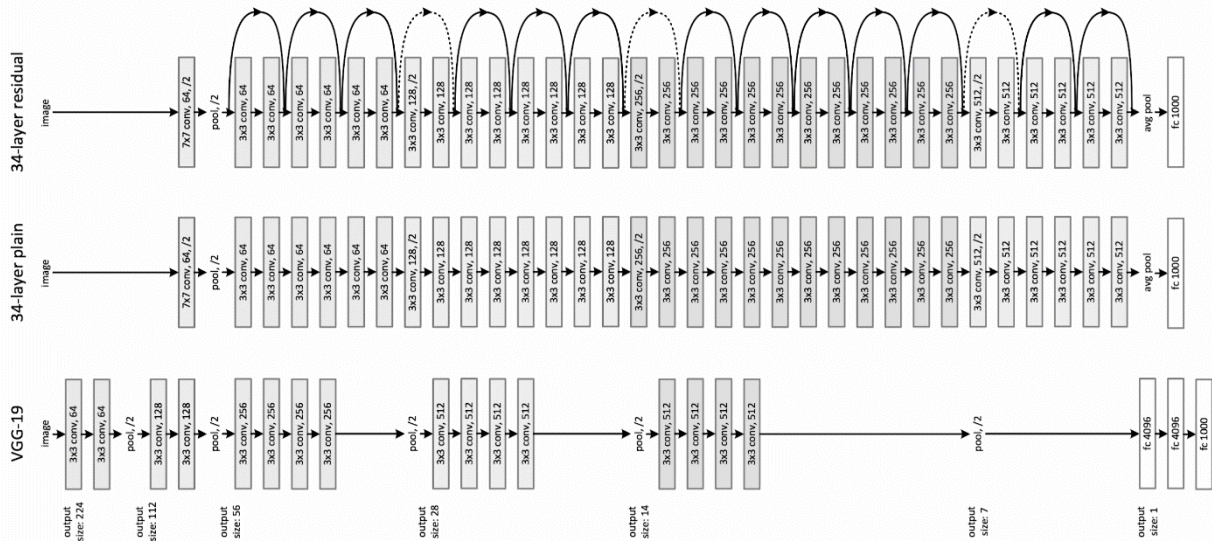


Рисунок 7. – Пример сетевой архитектуры для ImageNet. Слева: модель VGG-19. Посередине: простая сеть с 34 слоями. Справа: ResNet с 34 слоями.

**ResNet:** на основе описанной простой сети добавлено быстрое соединение (рис. 7, справа), которое превращает сеть в ее остаточную версию. Идентификационные быстрые соединения  $F(x \{W\} + x)$  могут использоваться непосредственно, когда вход и выход имеют одинаковые размерности (быстрые соединения сплошной линией на рис. 7).

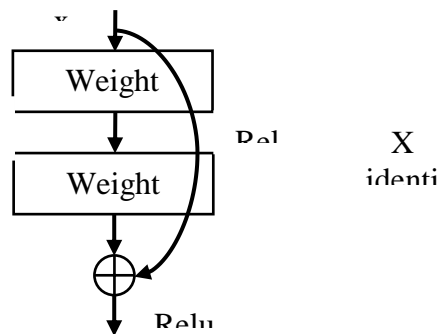


Рисунок 8. – Быстрое соединение

При увеличении размерности (пунктирные линии на рисунке 7), быстрое соединение рассматривает 2 варианта:

1. Быстрое соединение выполняет сопоставление идентификаторов с дополнительными нулями, которые добавлены для того, чтобы увеличить размерность. Такой подход не вводит дополнительных параметров.

2. Проекция быстрого соединения в  $F(x \{W\} + x)$  используется для сопоставления размерностей (выполнено с помощью  $1 \times 1$  сверток).

Для любой из опций, если быстрые соединения идут по картам объектов двух размерностей, они выполняются с шагом 2.

Таблица 1. – Характеристики CNN ResNet

Layer name	Output size	18-layer	34-layer	50-layer	101-layer	152-layer
Conv1	112×112	7×7, 64, Stride 2				
		3×3 max pool, stride 2				
Conv2_x	56×56	$\begin{bmatrix} 3 \times 3,64 \\ 3 \times 3,64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3,64 \\ 3 \times 3,64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,64 \\ 3 \times 3,64 \\ 1 \times 1,256 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 3,64 \\ 3 \times 3,64 \\ 1 \times 1,256 \end{bmatrix} \times 3$	$\begin{bmatrix} 3 \times 3,64 \\ 3 \times 3,64 \\ 1 \times 1,256 \end{bmatrix} \times 3$
Conv3_x	28×28	$\begin{bmatrix} 3 \times 3,128 \\ 3 \times 3,128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3,128 \\ 3 \times 3,128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 8$
Conv4_x	14×14	$\begin{bmatrix} 3 \times 3,256 \\ 3 \times 3,256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3,256 \\ 3 \times 3,256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,1024 \end{bmatrix} \times 36$
Conv5_x	7×7	$\begin{bmatrix} 3 \times 3,512 \\ 3 \times 3,512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3,512 \\ 3 \times 3,512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-dfc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

Каждый блок ResNet имеет два уровня глубины (используется в небольших сетях, таких как ResNet 18, 34) или 3 уровня (ResNet 50, 101, 152) (Таблица 1).

**50-слойная ResNet:** каждый 3-слойный блок заменяется в 34-слойной сети этим 3-слойным узким местом, в результате получается 50-слойная ResNet (см. Таблицу 1). Они используют вариант 2 для увеличения размерностей. Эта модель имеет 3,8 миллиарда FLOPs.

**ResNet с 101 и 152 слоями:** они создают ResNet с 101 и 152 слоями, используя больше 3-слойных блоков (см. Таблицу 1). После увеличения глубины 152-слойная сеть ResNet (11,3 миллиарда FLOP) имеет меньшую сложность, чем сети VGG-16 и VGG-19 (15,3 / 19,6 миллиарда FLOPs). ResNet – 152 достигает результата top-5 – 3.57%.

#### Сравнение моделей свёрточных нейронных сетей

Для оценки показателей моделей свёрточных нейронных сетей указывают вид ошибки (top-5). На изображениях в базе ImageNet может присутствовать множество объектов, однако аннотирован, только один из них. Основным критерием ошибки, является ошибка top-5. Результаты сравнения результатов различных свёрточных нейронных сетей представлены в таблице 2.

Таблица 2. – Сравнение показателей CNN в задачах распознавания изображений

Нейронная сеть	Топ-5
AlexNet	15,3 %
ZF Net	11,2 %
VGG Net	7,3 %
GoogleLeNet	6,7 %
Inception-v2	5,60 %
Inception-v3	3,58 %
ResNet-152	3,57 %

### Заключение

Распространение и развитие технологий компьютерного зрения влечет за собой изменение других профессиональных областей жизнедеятельности человека. Свёрточные нейронные сети (СНС) применяются в системах распознавания объектов и лиц, специальном медицинском ПО для анализа снимков, навигации автомобилей, оснащенных автономными системами, в системах защиты, и других сферах. С ростом вычислительной мощности компьютеров, появлением баз изображений стало возможным обучать глубокие нейронные сети. Одной из главных задач машинного обучения является задача классификации изображений. СНС применяются для оптического распознавания образов и объектов, детектирования предметов, семантической сегментации и т.д. В данной статье были рассмотрены наиболее распространенные архитектуры свёрточных нейронных сетей для задачи распознавания изображений, их строение и особенности.

В результате проведенного анализа архитектур выявлено, что свёрточная нейронная сеть ResNet-152 показала наилучший результат в задаче распознавания изображений, с показателем top-5 равным 3,57%, что говорит о достаточно точном определении объекта. Особенностью архитектуры ResNet является то, что свёрточные слои имеют фильтры 3×3, а также то, что в сеть добавлено быстрое соединение, которое превращает сеть в ее остаточную версию.

### Список литературы

- [1.] Сикорский, О.С. Обзор свёрточных нейронных сетей для задачи классификации изображений / О. С. Сикорский // Новые информационные технологии в автоматизированных системах – Москва, 2017. – № 20. – С. 37–42.
- [2.] Википедия [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki/Перцептрон>. – Дата доступа: 24.05.2019.
- [3.] LeCun Y. and Y. Bengio. "Convolutional networks for images, speech and time series", in M.A. Arbib, ed., The Hand book of Brain Theory and Neural Networks, Cambridge, MA: MIT Press, 1995.
- [4.] Нейронные сети: полный курс, 2-е издание.: Пер. с англ. – М. Издательский дом «Вильямс», 2006. – 1104 с.: ил. – Парал. тит. англ.
- [5.] Википедия [Электронный ресурс]. – Режим доступа: [http://ainews.ru/2018/11/alexnet\\_svertochnaya\\_nejroset\\_dlya\\_klassifikacii\\_izobrazhenij.html](http://ainews.ru/2018/11/alexnet_svertochnaya_nejroset_dlya_klassifikacii_izobrazhenij.html) – Дата доступа: 24.05.2019.
- [6.] Karen Simonyan, Andrew Zisserman, 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- [7.] Sergey Ioffe, Christian Szegedy, 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.
- [8.] Geoffrey Hinton, Nitish Srivasta, Kevin Swersky. 2012. «Lecture 6a Overview of Mini – Batch Gradient Descent» [www.cs.toronto.edu/~tijmen/csc321/slides/lecture\\_slides\\_lec6.pdf](http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf). Accessed 21 Mar. 2017.
- [9.] Dual Path Networks [Электронный ресурс]. – <https://paperswithcode.com/paper/dual-path-networks> – Дата доступа: 24.05.2019.



[10.] Deep Residual Learning for Image Recognition / Kaiming He [et al.]. – In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.

[11.] Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree / Chen-Yu Lee [et al.]. – In Artificial Intelligence and Statistics, pages 464–472, 2016.

[12.] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[13.] Neurohive [Электронный ресурс]. – <https://neurohive.io/ru/vidy-nejrosetej/vgg16-model/> – Дата доступа: 24.05.2019.

[14.] ImageNet Classification with Deep Convolutional Neural Networks / Alex Krizhevsky [et al.]. – November 2013, 9 S.

## **OVERVIEW OF CONVOLUTIONAL NEURAL NETWORKS FOR IMAGE RECOGNITION**

***A.S. Prokopenya***

*Postgraduate Student, Department of ECT,  
BSUIR*

***I.S. Azarov***

*assistant professor,  
Doctor of Technical Sciences, Head of the  
Department of ECT*

*Belarusian state University of Informatics and Radioelectronics*

*6, P. Brovki str., BGUIR, KAF. EMU, 220013, Minsk, Belarus, tel. +375 17 2938805,*

*E-mail: azarov@bsuir.by*

**Abstract.** The purpose of the work, the results of which are presented in the article, was to study modern architectures of convolutional neural networks for image recognition. This article discusses such architectures as AlexNet, ZF net, Get, Google Net, Reset. The characteristic about the image recognition quality for a neural network is the top-5 error. Based on the results obtained, it was found that at the moment the network with the most accurate result is the RESNET convolutional network with an accuracy rate of 3.57%. The advantage of this study is that this article provides a brief description of the convolutional neural network, as well as gives an idea of modern architectures of convolutional networks, their structure and quality indicators.

**Keywords:** convolution, filter, structure, subsample, activation function

УДК 004.021

## ВЫБОР АЛГОРИТМОВ КЛАСТЕРИЗАЦИИ



**А.С. Чочиева**  
Магистрант БГУИР



**И. И. Пилецкий**  
Доцент кафедры информатики БГУИР,  
кандидат физико-математических наук,  
старший научный сотрудник

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь  
E-mail: sonfanann@gmail.com, ianmenski@gmail.com

**А.С. Чочиева**  
Магистрант БГУИР.

**И. И. Пилецкий**  
Кандидат физико-математических наук, доцент БГУИР. В сфере ИТ более 47 лет. Участие в разработке нескольких десятков крупных проектов: главный конструктор проекта, главный архитектор ПО и информационного обеспечения, руководитель проекта, начальник отдела, заведующий лабораторией (НИИ ЭВМ, АН ИМ Беларуси, ИВА, БГУИР). Автор десятков научных исследований и публикаций.

**Аннотация.** Выбор алгоритма машинного обучения для решения некоторой задачи является проблемой. В данном докладе рассматриваются алгоритмы кластерного анализа и методика их выбора для эффективного решения прикладных задач.

**Ключевые слова:** кластеризация, алгоритмы машинного обучения, быстродействие и эффективность алгоритмов.

### 1. Введение

Алгоритмы кластерного анализа позволяют определить группы (кластеры) данных более схожих друг с другом, чем с остальными данными и выявить ранее незамеченные закономерности. Эти алгоритмы относятся к классу задач обучения без учителя. Метки классов в этих алгоритмах заранее неизвестны как в алгоритмах классификации, которым они иногда предшествуют.

Название «кластерный анализ» происходит от английского слова cluster – гроздь, скопление. Впервые в 1939 был определен предмет кластерного анализа и сделано его описание исследователем Трионом (Tryon) [1].

### 2. Анализ работы алгоритмов кластерного анализа

Краткая характеристика алгоритмов кластерного анализа.

Методы кластеризации можно поделить на:

- Методы разбиения: здесь сходство рассматривается по отношению к центру масс кластера, т.е. среднему значению координат объектов кластера в пространстве данных;
- Иерархические методы: постепенно строят кластеры, менее чувствительны к шуму;
- Плотностные методы: обнаруживают плотные объединённые компоненты данных, гибки по форме данных, менее чувствительны к аномалиям;

- Сетевые методы. Общая идея - пространство объектов разбивается на конечное число ячеек, образующих сетевую структуру, в рамках которой выполняются все операции кластеризации;

- Модельные методы. Методы этого семейства предполагают, что имеется некоторая математическая модель кластера в пространстве данных и стремятся максимизировать сходство этой модели и имеющихся данных. Часто при этом используется аппарат математической статистики [2];

- Концептуальная кластеризация - определяет кластеры как группы объектов, относящейся к одному классу или концепту – определённому набору пар атрибут-значение. [2].

### **Методы разбиения. K-means**

Методы разбиения также могут называться центроидными методами. Примером алгоритма, применяющего метод разбиения, является алгоритм K-means. Он разделяет точки данных на K не пересекающихся кластеров путём нахождения K центральных точек (центроидов) и назначения каждой точке кластера, соответствующего ближайшему центроиду. В алгоритме k-means сходство рассматривается по отношению к центру масс кластера – среднему значению координат объектов кластера в пространстве данных.

Алгоритм [3]:

Шаг 1. Алгоритм начинается с генерирования K случайных точек — центроидов. Часто берутся случайные точки из набора данных, на котором проводится обучение.

Шаг 2. Каждой точке из набора данных присваивается ближайший центроид (близость можно определять евклидовым расстоянием).

Шаг 3. Координаты центроидов пересчитываются — берётся среднее от координат назначенных соответствующему центроиду точек.

Шаг 4. Шаги 2-3 повторяются до выполнения одного из критериев остановки:

- Координаты центроидов не меняются;
- Точки остаются в тех же самых кластерах;
- Было достигнуто максимальное число итераций.

Этот метод вычислительно быстрее, масштабируем, быстрее для данных с меньшим количеством признаков. Но он даёт хороший результат только на данных с хорошей, близкой к сферической форме, которой и можно назначить центроид. Плохо работает с данными, в которых есть аномалии и шум [4].

Чтобы сократить влияние шума и обособленных точек пространства на результат кластеризации, алгоритм K-medoids, в отличие от K-means, использует для представления центра кластера не центр масс, а представительный объект – один из объектов кластера.

Алгоритм [2]:

Шаг 1. Сначала выбираются K случайных представительных объектов, как и в методе k-means.

Шаг 2. Каждая из оставшихся точек объединяется в кластер с ближайшим представительным объектом.

Шаг 3. Итеративно для каждого представительного объекта производится его замена произвольным непредставительным объектом пространства данных. Процесс замены продолжается до тех пор, пока улучшается качество результирующих кластеров.

Качество кластеризации определяется суммой отклонений между каждым объектом и представительным объектом соответствующего кластера, которую метод стремится минимизировать. То есть, итерации продолжаются до тех пор, пока в каждом кластере его представительный объект не станет медоидом – наиболее близким к центру кластера объектом [2].

### Иерархические методы

Иерархические методы группируют данные в дерево кластеров (дендрограмму), представляющее собой последовательность вложенных кластеров. Есть два подхода для построения этого дерева:

- агломерационный (снизу вверх);
- дивизивный (сверху вниз).

Алгоритм.

При агломерационном подходе:

Шаг 1. Анализ начинается с индивидуального кластера (одна точка, синглетон).

Шаг 2. К синглетону рекурсивно добавляются два или более подходящих кластера, поднимаясь по иерархии. Процесс можно остановить, когда будет достигнуто нужное количество кластеров равное  $K[5]$ .

При дивизивном подходе:

Шаг 1. Анализ начинается с одного кластера, содержащего все точки.

Шаг 2. Этот кластер рекурсивно разделяется вниз по иерархии, пока не будет достигнуто  $K$ -тое количество кластеров.

Слияние и деление кластеров основывается на их близости (схожести). При построении дерева кластеров используются меры расстояния между кластерами (англ. *sorting strategies, linkage criteria*; связи, правила объединения), при помощи которых определяется степень сходства между парами кластеров.

Пусть  $n$  - количество наблюдений,  $D_{KL}$  – расстояние между кластерами  $C_K$  и  $C_L$ ,  $d(x,y)$  – расстояние между двумя векторами  $x$  и  $y$  (вычисленное с помощью любой выбранной меры расстояния),  $N_K$  и  $N_L$  - количество наблюдений в кластерах  $C_K$  и  $C_L$  соответственно. Из наиболее часто встречающихся методов определения расстояния между кластерами можно перечислить следующие методы[6].

Метод одиночной связи (Single Linkage, метод ближайшего соседа)

Расстояние между двумя кластерами берётся равным минимальному расстоянию между двумя экземплярами из разных кластеров.

$$D_{KL} = \min_{i \in C_K} \min_{j \in C_L} d(x_i, x_j),$$

где  $x_i$  и  $x_j$  соответственно  $i$ -ый и  $j$ -ый экземпляры (векторы значений признаков) кластеров  $C_K$  и  $C_L$  соответственно.

Метод полной связи (Complete Linkage, метод дальнего соседа)

Расстояние между двумя кластерами берётся равным максимальному расстоянию между двумя экземплярами из разных кластеров.

$$D_{KL} = \max_{i \in C_K} \max_{j \in C_L} d(x_i, x_j),$$

где  $x_i$  и  $x_j$  соответственно  $i$ -ый и  $j$ -ый экземпляры (векторы значений признаков) кластеров  $C_K$  и  $C_L$  соответственно.

Метод средней связи (Average Linkage)

Расстояние между двумя кластерами берётся равным среднему расстоянию между экземплярами этих кластеров.

$$D_{KL} = \frac{1}{N_K N_L} \sum_{i \in C_K} \sum_{j \in C_L} d(x_i, x_j),$$

где  $x_i$  и  $x_j$  соответственно  $i$ -ый и  $j$ -ый экземпляры (векторы значений признаков) кластеров  $C_K$  и  $C_L$  соответственно.

Центроидный метод

Расстояние между двумя кластерами берётся равным расстоянию между центроидами этих кластеров.

$$D_{KL} = \|\bar{x}_K - \bar{x}_L\|^2, \text{ где } \bar{x}_K \text{ и } \bar{x}_L - \text{центроиды кластеров } C_K \text{ и } C_L \text{ соответственно.}$$

Метод минимальной дисперсии Уорда (Ward's Minimum-Variance Method)

$$D_{KL} = B_{KL} = \frac{\|\bar{x}_K - \bar{x}_L\|^2}{\frac{1}{N_K} + \frac{1}{N_L}},$$

где  $B_{KL} = W_M - W_K - W_L$ ,  $C_M = C_K \cup C_L$ ,

$$W_K = \sum_{i \in C_K} \|x_i - \bar{x}_K\|^2,$$

$\bar{x}_K$  и  $\bar{x}_L$  - центроиды кластеров  $C_K$  и  $C_L$  соответственно.

Для применения первых трёх методов необходимо выбрать меру расстояния, такие как (для векторов  $x$  и  $y$ ,  $N$  - количество признаков/размерность векторов):

Евклидово расстояние [7]:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{j=1}^N (x_j - y_j)^2}$$

- также взвешенное Евклидово расстояние, когда каждому признаку даётся некоторый вес, пропорциональный степени важности признака в задаче:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{j=1}^N w_j (x_j - y_j)^2}$$

Скалярное произведение [7]:

$$d(x, y) = 1 - x * y = 1 - \|x\| \|y\| \cos(x, y)$$

Мера относительно нормы  $L_1$  (Манхэттан):

$$d(x, y) = \sqrt{\sum_{j=1}^N |x_j - y_j|}$$

Мера относительно нормы  $L_\infty$  (расстояние Чебышева) [7]:

$$d(x, y) = \max_j (|x_j - y_j|)$$

Метрика Махаланобиса, обобщающая евклидову меру:

$$D^2 = (x - m)^T \cdot C^{-1} \cdot (x - m),$$

где  $D^2$  — квадрат расстояния Махаланобиса,  $x$  — вектор признаков (координат) наблюдения,  $m$  — средние значения независимых переменных (признаков, среднее по столбцам выборки) множества, до которого высчитывается расстояние;  $C^{-1}$  — обратная ковариационная матрица независимых переменных [8];

Степенная метрика:

$$d(x, y) = \left( \sum_{j=1}^N |x_j - y_j|^p \right)^{1/r},$$

где  $r$  и  $p$  — определяемые пользователем параметры. Также является обобщением Евклидовой меры. Поведение данной меры выглядит следующим образом: Параметр  $r$  контролирует вес разностей по отдельным компонентам, параметр  $p$  контролирует вес придаваемый расстоянию между объектами в целом. Если  $r$  и  $p$  равны 2, то это расстояние равно Евклидову расстоянию [9].

Иерархические методы изначально не нуждаются во входном  $K$  (количестве кластеров). Они рассчитывают полную иерархию кластеров, что приводит к встроенной гибкости по отношению к детализации, а дендрограмму можно использовать для визуального представления иерархии [10]. Из-за сложности, иерархические методы медленны на больших наборах данных, особенно при использовании дивизивного подхода [5]. Также, объединение/слияние кластеров необратимы и ошибку, которая накопится в будущем, невозможно рассчитать. Этим алгоритмам сложно справиться с кластерами разных размеров и с выпуклыми формами [4].

Примерами иерархических алгоритмов являются Complete-link (Com-link), BIRCH — Они и будут рассматриваться в данной работе. BIRCH — это агломерационный алгоритм, который работает динамически и инкрементативно и подходит для больших наборов данных, но может работать только с численными данными [11]. Com-link один из link алгоритмов, в которых слияние кластеров основывается на их расстоянии, в данном случае на наибольшем расстоянии. В простейшем случае у него машинное время  $O(n^3)$ , тогда как у BIRCH  $O(n)$  [11].

#### **Плотностные методы**

Примером плотностного алгоритма является DBSCAN [12]. Он группирует вместе близкие друг к другу точки, основываясь на расстоянии (обычно евклидовым) и на минимальном количестве точек в группе. Он также маркирует точки, лежащие в областях с низкой плотностью, как аномалии [12]. Для этого ему нужно два параметра:

-  $\varepsilon$  – максимальное расстояние между точками для того, чтобы они считались соседями и частью одного кластера;

- минимальное количество точек, которые формируют плотную область.

При рассмотрении этого алгоритма важно также рассмотреть следующие определения [2].

*Корневой точкой (корневым объектом)* называется точка,  $\varepsilon$ -окрестность которого содержит не менее некоторого минимального числа  $MinPts$  точек.

*Точка  $p$  непосредственно плотно-достижима* из точки  $q$  если  $p$  находится в  $\varepsilon$ -окрестности  $q$  и  $q$  является корневой.

*Точка  $p$  плотно-достижима* из точки  $q$  при заданных  $\varepsilon$  и  $MinPts$ , если существует последовательность точек  $p_1, \dots, p_n$ , где  $p_1 = q$  и  $p_n = p$ , такая что  $p_{i+1}$  непосредственно плотно достижима из  $p_i$ ,  $1 \leq i \leq n$ .

*Точка  $p$  плотно-соединена* с точкой  $q$  при заданных  $\varepsilon$  и  $MinPts$ , если существует точка  $o$  такая, что  $p$  и  $q$  плотно-достижимы из  $o$ .

Для поиска кластеров алгоритм DBSCAN проверяет  $\varepsilon$ -окрестность каждой точки. Если  $\varepsilon$ -окрестность точки  $p$  содержит больше точек чем  $MinPts$ , то создаётся новый кластер с корневой точкой  $p$ . Затем DBSCAN итеративно собирает точки непосредственно плотно-достижимые из корневых точек, которые могут привести к объединению нескольких плотно-достижимых кластеров. Процесс завершается, когда ни к одному кластеру не может быть добавлено ни одной новой точки.

Иерархическая кластеризация и особенно кластеризация разбиением подходят для нахождения сферических или выпуклых кластеров и часто чувствительны к аномалиям и шуму. Реальные же данные могут принимать различные формы.

На рисунке 1 изображён результат, выданный алгоритмом K-Means на наборе данных, имеющем 2 овальных, 2 линейных кластера и 1 плотный кластер [13]. Как можно заметить, результат неудовлетворительный.

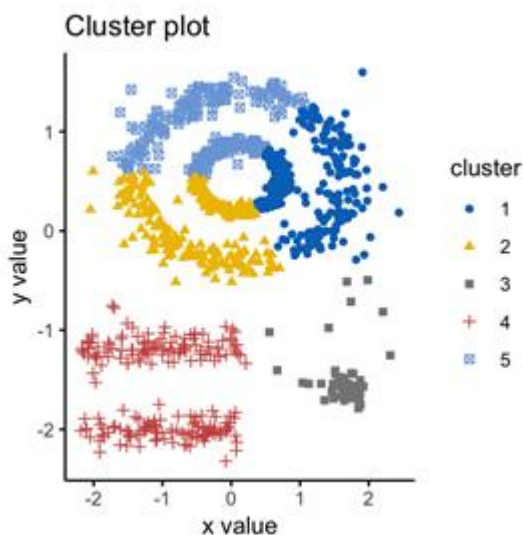


Рисунок 1. – Результат от алгоритма K-means

В то время как DBSCAN гораздо лучше справился с этим набором данных, правильно отделив все кластеры (рисунок 2)

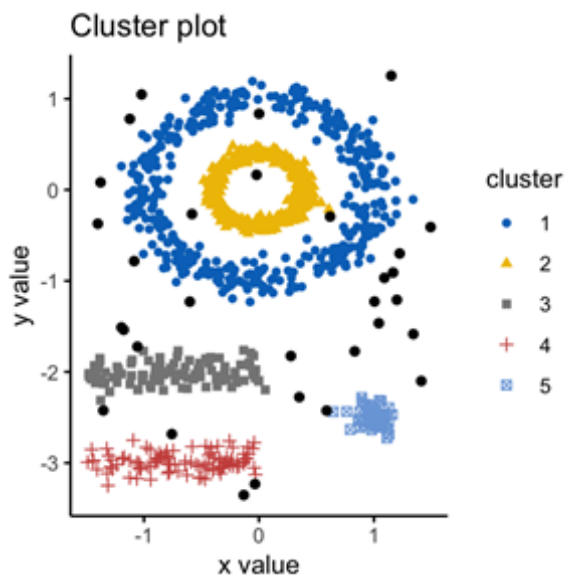


Рисунок 2. – Результат от алгоритма DBSCAN

Но результат сильно зависит от правильного подбора параметров алгоритма. И если плотность кластеров в наборе данных сильно меняется в разных областях или плотность в наборе данных в целом слишком маленькая, то это может привести к плохому результату. Также этот алгоритм не поддерживает мультипроцессорную работу [3].

Алгоритм OPTICS [14].

Из-за высокой зависимости результатов DBSCAN от параметров (в частности параметра  $\epsilon$ ), в качестве представителя плотностных методов брался алгоритм OPTICS, который является модификацией алгоритма DBSCAN. Он выбирает основной образец данных с высокой плотностью и, отталкиваясь и расширяясь от него, находит кластеры. Лучше подходит для больших наборов данных чем текущая реализация DBSCAN в Scikit-Learn.

#### Сетевые методы

Главное достоинство методов этой группы в малом времени выполнения, которое обычно не зависит от количества объектов данных, а зависит только от количества ячеек в каждом измерении пространства.

Примером реализации сетевого метода является алгоритм CLIQUE [15]. Он адаптирован под кластеризацию данных высокой размерности (применяет подпространственную кластеризацию - Subspace clustering - рассмотрена ниже). Метод основан на том предположении, что если в многомерном пространстве данных распределение объектов не равномерно – встречаются регионы плотности и разрежения, то проекция региона плотности в подпространство с меньшей размерностью будет частью региона плотности в этом подпространстве [2].

Подпространственные методы кластеризации (Subspace clustering) это расширение более традиционных методов кластеризации, стремящееся найти кластеры в различных подпространствах набора данных. Часто в данных с высокой размерностью (большим количеством признаков), многие признаки не являются важными и могут шумом скрыть существующие кластеры. Метод производит выбор признаков (Feature selection), анализируя весь набор данных, и убирает ненужные и излишние признаки. Подпространственные методы локализуют поиск важных признаков, что позволяет им найти кластеры, существующие в нескольких, возможно пересекающихся подпространствах [16].

Существуют два основных направления сетевой кластеризации по стратегии поиска:



- сверху вниз – метод сначала кластеризует по всем признакам, затем обрабатывает подпространства каждого кластера, итеративно улучшая результат;
- снизу вверх – метод сначала находит плотные области в низкоразмерных подпространствах и затем объединяет их для образования кластеров.

#### **Модельные методы [17].**

Метод K-means, как и многие другие методы, делает довольно сильные предположения о данных. Например, предполагается, что кластеры данных имеют сферическую форму. Модельные методы же более гибкие. Модель кластеризации может быть адаптирована под некоторые предположения о распределении данных.

EM-алгоритм (Expectation-Maximization algorithm) является одной из наиболее часто используемых реализаций модельного метода. Это итеративный алгоритм, максимизирующий  $L(D|\Theta)$  (целевая функция, оценивающая качество кластеризации;  $D$  – данные,  $\Theta$  — параметры модели, в случае K-means – это центроиды). Алгоритм K-means является частным случаем EM алгоритма.

EM-алгоритм может быть использован с множеством различных видов вероятностного моделирования. Этот алгоритм схож с K-means в том, что он переключается между шагом ожидания (expectation step, сравним с переназначением точек кластерам) и шагом максимизации (maximization step, сравним с пересчётом центроидов, т. е. это пересчёт самих параметров модели  $\Theta$ ).

#### **Концептуальная кластеризация.**

Алгоритм COBWEB (описан в 1987 году) – классический метод инкрементальной концептуальной кластеризации. Он создаёт иерархическую кластеризацию в виде дерева классификации: каждый узел этого дерева ссылается на концепт и содержит вероятностное описание этого концепта, которое включает в себя вероятность принадлежности концепта к данному узлу и условные вероятности вида:  $P(A_i = v_{ij}/C_k)$ , где  $A_i = v_{ij}$  – пара атрибут-значение,  $C_k$  – класс концепта [2].

### **3. Исследование работы алгоритмов кластерного анализа**

В данной работе в п. 2 был проведен теоретический анализ алгоритмов кластеризации. В этом разделе практически на наборах данных проанализированы и рассмотрены первые три метода. Результаты анализа приведены ниже.

**Постановка задачи:** Создание программного обеспечения, способного рекомендовать пользователям алгоритмы машинного обучения в соответствии с требуемыми параметрами (на пример: планируемый объём данных, требуемая точность, скорость выполнения, потребление ресурсов).

В данной работе основные усилия направлены на исследования алгоритмов — K-means, Birch, K-means.

#### **Методика решения.**

*Подход к решению.* Для решения проблемы подбора оптимального алгоритма нужно сделать следующее:

- Провести сбор или генерирование наборов данных для кластеризации;
- Провести тестирование интересующих алгоритмов;
- Создать набор данных из результатов тестирования;
- Провести моделирование (интерполяцию) результатов при других начальных условиях методом регрессии;
- Разработать рекомендательную систему, которая по заданным начальным условиям будет рекомендовать наиболее подходящий алгоритм.

Далее будет разобран каждый из пунктов подробнее.

*Сбор данных.* Часть данных генерировалась и часть бралась из интернет источника [18]. Данные брались с метками – для последующей оценки качества результата – и без меток (тогда

точность относительно меток не учитывалась – значение в результатах указывалось равным -1). При тестировании алгоритмов кластеризации важно разнообразие форм данных, так как многие алгоритмы лучше справляются с глобулярными данными.

*Тестирование.* Исследуемый алгоритм тестируется несколько раз. На вход алгоритма подаются различные наборы данных, с различным количеством признаков, наблюдений, различных кластеров и с различием форм этих кластеров. Для тестирования используется специально разработанное ПО. Один тест проводит тестирование сразу на множестве наборов данных и записывает результаты в csv файл.

Все тесты и сама программа-рекомендер разрабатывались на языке Python. При тестировании использовались библиотека Scikit-learn и PySpark. Для мониторинга потребляемых ресурсов использовался пакет psutil.

В ходе тестирования оцениваются следующие параметры:

- Скорость работы;
- Точность;
- Силуэтный коэффициент;
- Количество потребляемых ресурсов.

Для оценки качества результата использовались две метрики — точность (accuracy), считаемая через `homogeneity_score` из библиотеки `scikit-learn`, и силуэтный коэффициент.

Метрика `homogeneity_score` (метрика однородности), по сути, оценивает насколько метки, поставленные алгоритмом, соответствуют настоящим меткам (`true labels`), взятым из набора данных. Эта метрика не даёт достаточно достоверный результат для алгоритмов, в которых заранее не задаётся количество искомых кластеров, что в особенности относится к плотностным методам.

У кластеров есть два важных свойства: точки внутри кластера схожи друг с другом и расстояние/различие между кластерами должно быть как можно больше [3]. Эти свойства может адресовать силуэтный коэффициент. В отличие от предыдущей метрики, он не опирается на настоящие метки, а лишь на результат самого алгоритма. Согласно официальной документации `scikit-learn` [19], силуэтный коэффициент вычисляется, используя среднее внутри-кластерное расстояние ( $a$ ) и среднее расстояние до ближайшего кластера ( $b$ ), для каждого образца/наблюдения.

$$SI = \frac{b - a}{\max(a, b)}$$

*Сбор тестовых данных.* Все результаты тестов собираются из csv файлов воедино с помощью библиотеки `pandas`. Результаты тестов со значением параметра `accuracy` равным -1 (когда набор данных был без настоящих меток) не учитываются при построении модели этого параметра.

*Моделирование.* На основе результатов тестирования строились и обучались модели по каждому параметру, для каждого алгоритма. На основе этих моделей выполняется рекомендация пользователю алгоритма согласно его заданным параметрам: количество признаков и наблюдений и назначенный на каждый тестируемый параметр вес.

Не для всех параметров подходит одна и та же модель и метод регрессии. Опытным путём проводится подбор регрессоров для получения наиболее точного прогнозирования значения параметра. На данный момент было найдено, что лучше всего подходят деревья решений, но это может измениться в ходе дальнейшего сбора данных.

*Рекомендация.* После получения нормализованных предсказанных значений, они перемножаются на соответствующие каждому параметру заданные пользователем веса и суммируются для каждого алгоритма. Эти суммы сортируются, и алгоритм с наибольшей

суммой рекомендуется пользователю.

На данный момент больше всего данных собрано для алгоритма K-means (529 тестовых результатов), что несколько повысило точность предсказаний относительно двух остальных алгоритмов. Изначально для всех алгоритмов было собрано 76 тестовых результатов. Самая лучшая точность предсказаний была получена для accuracy:

```
KMeans:Python Sklearn (avg_accuracy)
test accuracy :0.7381377080017326
train accuracy: 0.9367138015666745
```

Для силуэтного коэффициента, а также для других параметров, точность на данный момент ниже:

```
KMeans:Python Sklearn (avg_silhouette_score)
test accuracy :0.38682883929587886
train accuracy: 0.43454342071548047
```

Для улучшения точности прогнозирования требуется большее количество тестовых данных, а также возможно продолжить подбор моделей регрессии. Возможно также нужно ввести новый признак — численную оценку качества тестируемых данных, формы кластеров и степени их близости/пересечения. Данный вывод был сделан из приведенных ниже визуализаций тестовых данных — данные весьма шумные, для наборов с одинаковым количеством признаков и наблюдений значения полученных тестовых параметров могут отличаться.

На этих графиках вертикальная ось - это значение параметра. Горизонтальные оси - это количество признаков (predictors) и количество наблюдений (observations). Каждая точка соответствует результату теста. Для большей части тестируемых наборов данных количество признаков варьируется от 2 до 512. Для K-means были сгенерированы наборы данных с шагом количества признаков равным 1000 (1000, 2000, 3000), что заметно на графиках. Особенно шумным выглядит график для силуэтного коэффициента (см. рисунки 3, 4, 5, 6, 7).

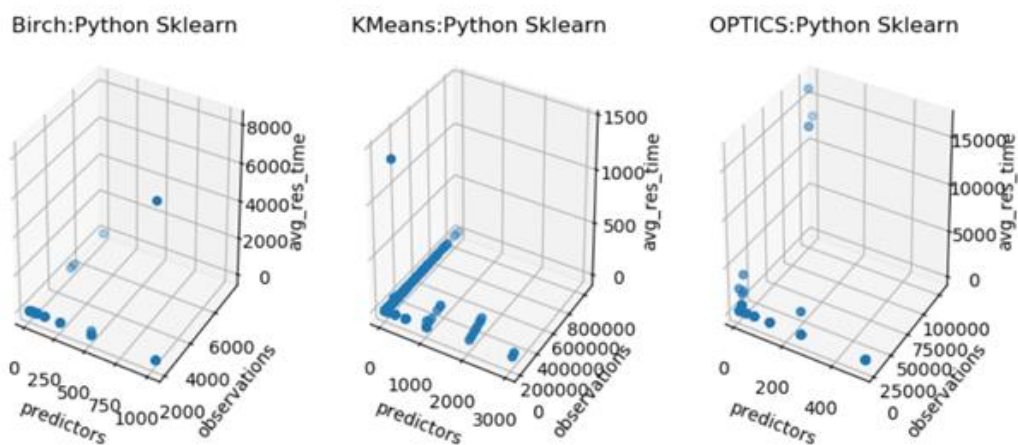


Рисунок 3. – Визуальное представление тестовых данных по количеству признаков, наблюдений и значениям параметра времени выполнения

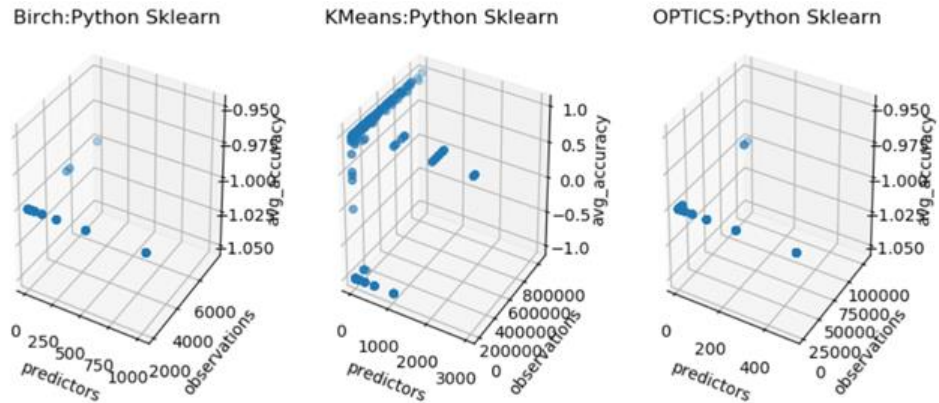


Рисунок 4. – Визуальное представление тестовых данных по количеству признаков, наблюдений и значениям параметра accuracy

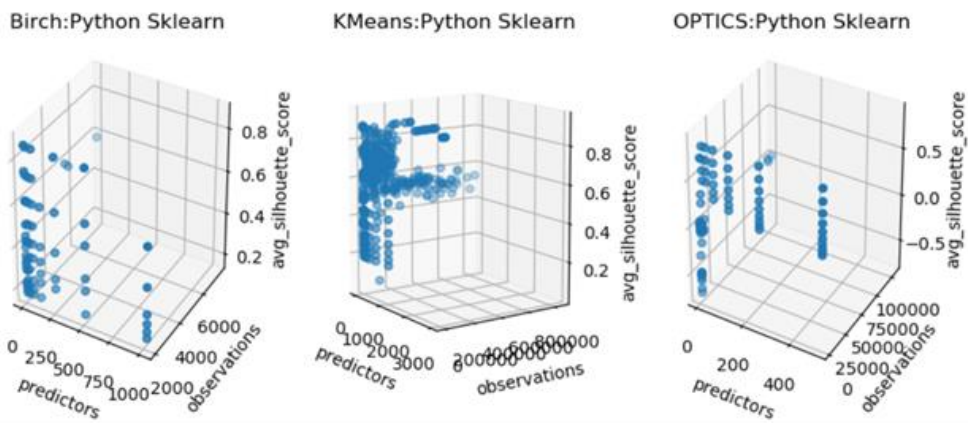


Рисунок 5. – Визуальное представление тестовых данных по количеству признаков, наблюдений и значениям силуэтного коэффициента.

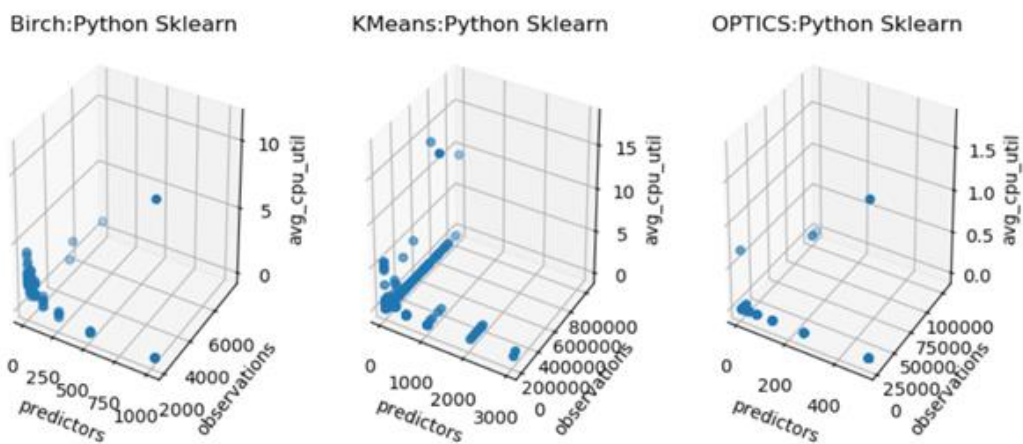


Рисунок 6. – Визуальное представление тестовых данных по количеству признаков, наблюдений и значениям процента использования ЦП

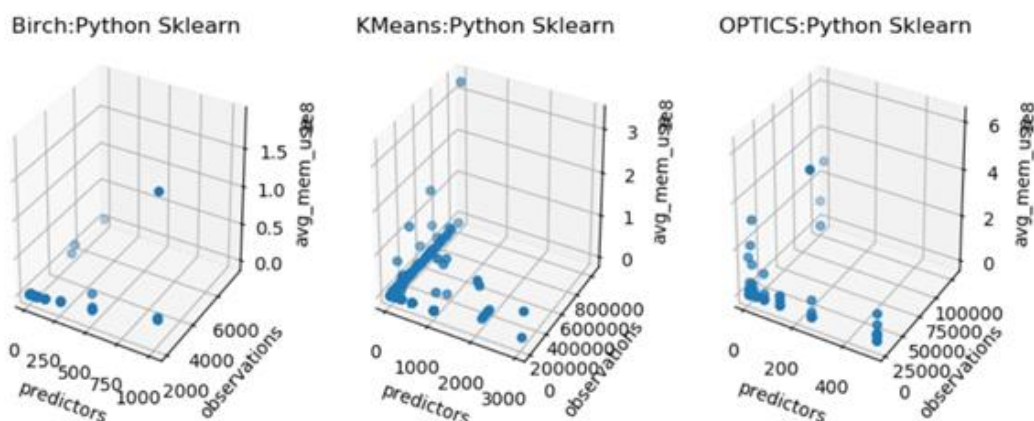


Рисунок 7. – Визуальное представление тестовых данных по количеству признаков, наблюдений и значениям параметра использования оперативной памяти

#### 4. Заключение

Для того чтобы создать программное обеспечение, способное рекомендовать алгоритмы машинного обучения, были собраны данные об их эффективности на наборах данных различного рода и размера. Были написаны для них тесты производительности и подобраны или сгенерированы различные наборы данных. Программа-рекомендер использует собранные от тестов данные для создания моделей, и с помощью них даётся рекомендация пользователю в соответствии с установленными им приоритетами-весами (скорость, точность и т.д.).

Для получения более точных моделей, и, соответственно, более качественных рекомендаций требуется получить больше тестовых данных от большего количества различных наборов данных.

#### Список литературы

- [1]. Кластерный анализ. [Электронный ресурс]: Национальная библиотека им. Н. Э. Баумана. Последнее изменение страницы: 19:52, 24 декабря 2016. - Режим доступа: [https://ru.bmstu.wiki/Кластерный\\_анализ](https://ru.bmstu.wiki/Кластерный_анализ). Дата доступа: 05.12.2019.
- [2]. “Обзор алгоритмов кластеризации числовых пространств данных”, 30 декабря 2012. [Электронный ресурс] - Режим доступа: <https://habr.com/ru/post/164417/>. Дата доступа: 20.02.20.
- [3]. “The Most Comprehensive Guide to K-Means Clustering You’ll Ever Need”, by Pulkit Sharma, . [Электронный ресурс]: Analytics Vidhya, August 19, 2019 — Режим доступа: <https://ijcset.net/docs/Volumes/volume6issue4/ijcset2016060404.pdf>. Дата доступа: 14.12.19.
- [4]. “Performance Evaluation of Clustering Algorithm Using Different Datasets”. [Электронный ресурс]: International Journal of Advance Research in Computer Science and Management Studies, Volume 3, Issue 1 (Январь 2015) —Режим доступа: <http://www.ijarcsms.com/docs/paper/volume3/issue1/V3I1-0058.pdf>. Дата доступа: 20.02.20.
- [5]. “Performance Evaluation of Clustering Algorithms”. [Электронный ресурс]: International Journal of Engineering Trends and Technology (IJETT)-Volume 4, Issue 7 (Июль 2013) —Режим доступа: <https://pdfs.semanticscholar.org/a5f8/02501295bd79290b6d7768050588b714bd6e.pdf> — Дата доступа: 05.12.2019.
- [6]. “Clustering Methods”. [Электронный ресурс]: SAS/STAT(R) 9.2 User's Guide, Second Edition — Режим доступа: [https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug\\_cluster\\_sect012.htm](https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_cluster_sect012.htm) — Дата доступа: 05.12.2019.
- [7]. Осовский С., “Нейронные сети для обработки информации”, 2002 г., стр. 228-229.
- [8]. «Mahalanobis Distance – Understanding the math with examples (python)». [Электронный ресурс]: MachineLearning+: Simple and straightforward tutorials on machine learning in R and Python. —Режим доступа: <https://www.machinelearningplus.com/statistics/mahalanobis-distance/> — Дата доступа: 05.12.2019.
- [9]. «Степенное расстояние». [Электронный ресурс]: Life-prog.ru —Режим доступа: [https://life-prog.ru/2\\_89638\\_stepennoe-rasstoianie.html](https://life-prog.ru/2_89638_stepennoe-rasstoianie.html) — Дата доступа: 05.12.2019.

- [10]. Amol Bhagat et al, “Penalty Parameter Selection for Hierarchical Data Stream Clustering”. [Электронный ресурс]: Procedia Computer Science 79, стр. 24–31 (2016) —Режим доступа: <https://www.sciencedirect.com/science/article/pii/S1877050916001368>. Дата доступа: 20.02.20.
- [11]. “A survey of hierarchical clustering algorithms”. [Электронный ресурс]: The Journal of Mathematics and Computer Science, Vol .5 No.3, стр. 229–240 (2012)—Режим доступа: <https://www.isr-publications.com/jmcs/articles-417-a-survey-of-hierarchical-clustering-algorithms>. Дата доступа: 20.02.20.
- [12]. Kelvin Salton do Prado, статья ”How DBSCAN works and why should we use it?”, 2 апреля 2017. [Электронный ресурс]: Сайт ”Towards Data Science”. —Режим доступа: <https://towardsdatascience.com/how-dbscan-works-and-why-should-i-use-it-443b4a191c80>. Дата доступа: 20.02.20.
- [13]. “DBSCAN: Density-Based Clustering Essentials”. [Электронный ресурс]: DataNovia: Online Data Science Courses (2018). — Режим доступа: <https://www.datanovia.com/en/lessons/dbscan-density-based-clustering-essentials/>. Дата доступа: 20.02.20.
- [14]. “sklearn.cluster.OPTICS”. [Электронный ресурс]: Официальная документация Scikit-Learn версии 0.22.1. (2020 год) —Режим доступа: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.OPTICS.html>. Дата доступа: 20.02.20.
- [15]. Agrawal, R. Automatic “Subspace Clustering of High Dimensional Data for Data Mining Applications” / R. Agrawal, J. Gehrke, D. Gunopulos, P. Raghavan // In Proc. ACM SIGMOD Int. Conf. on Management of Data, Seattle, Washington, 1998. -С.94-105.
- [16]. «Subspace Clustering for High Dimensional Data: A Review», Lance Parsons, Ehtesham Haque, Huan Liu. [Электронный ресурс]: Sigkdd Explorations, Volume 6, Issue 1 - Page 105, June 2004. —Режим доступа: [https://www.kdd.org/exploration\\_files/parsons.pdf](https://www.kdd.org/exploration_files/parsons.pdf). Дата доступа: 20.02.20.
- [17]. «Model-based clustering», Christopher D. Manning, Prabhakar Raghavan & Hinrich Schütze, Cambridge University Press, 2008. [Электронный ресурс] Сайт, сопутствующий книге “Introduction to Information Retrieval». —Режим доступа: <https://nlp.stanford.edu/IR-book/html/htmledition/model-based-clustering-1.html>. Дата доступа: 20.02.20.
- [18]. P. Fränti and S. Sieranoja, K-means properties on six clustering benchmark datasets Applied Intelligence, 48 (12), 4743-4759, December 2018 [Электронный ресурс] — Режим доступа: <http://cs.joensuu.fi/sipu/datasets/>. Дата доступа: 20.02.20.
- [19]. Документация на силуэтный коэффициент из Scikit-learn. [Электронный ресурс]: Документация Scikit-learn — Режим доступа: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html). Дата доступа: 20.02.20.

## **CHOICE OF CLUSTERING METHODS**

***A.S.Chochieva***  
*Master’s student of BSUIR*

***I.I. Piletski***  
*PhD*  
*Associate Professor of*  
*Informatics Department*  
*of the BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus*  
*E-mail: sonfanann@gmail.com, ianmenski@gmail.com*

**Abstract.** There is an existing problem of choosing a machine learning algorithm as a solution to a task. In this study, a review of numerous clustering algorithms is conducted and a method for choosing a clustering algorithm for efficient solving of practical problems is developed and attempted.

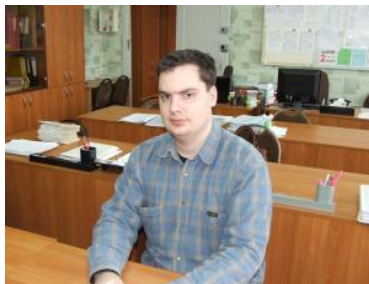
**Keywords:** Clustering, machine learning algorithms, algorithm performance and efficiency

УДК 004.032.26:61

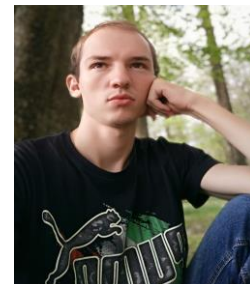
## ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ В ОБРАБОТКЕ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ



**А.А. Каракулько**  
магистрант кафедры инженерной психологии и эргономики БГУИР



**О.С. Медведев**  
магистр техники и технологии, инженер-программист, кафедра инженерной психологии и эргономики БГУИР



**П.А. Мороз**  
аспирант кафедры инженерной психологии и эргономики БГУИР



**О.Н. Дудич**  
кандидат медицинских наук, доцент, кафедра офтальмологии, БелМАПО



**В.Л. Красильникова**  
доктор медицинских наук, профессор, кафедра офтальмологии БелМАПО



**В.С. Осипович**  
кандидат технических наук, доцент, кафедра инженерной психологии и эргономики БГУИР



**К.Д. Яшин**  
Заведующий кафедрой инженерной психологии и эргономики БГУИР, кандидат технических наук, доцент

Белорусский государственный университет информатики и радиоэлектроники», Республика Беларусь, Минск

Белорусская медицинская академия последипломного образования, отделение офтальмологии, Республика Беларусь, Минск

E-mail: o.med@bsuir.by

### **А.А. Каракулько**

В 2019 году окончил БГУИР, специальность «Информационные системы и технологии (в обеспечении промышленной безопасности)». Квалификация инженер-системотехник.

### **О.С. Медведев**

В 2013 году окончил БГУИР, специальность «Инженерно-психологическое обеспечение информационных технологий». Квалификация инженер-системотехник. В 2019 году окончил магистратуру по специальности «Управление безопасностью производственных процессов» с присвоением академической степени магистра техники и технологии. Ведет курсы «CISCO».

### **П.А. Мороз**

В 2016 году окончил БНТУ, специальность «Программное обеспечение информационных технологий». Квалификация инженер-программист. В 2018 году окончил магистратуру по специальности «Инженерная психология, эргономика» с присвоением академической степени магистра технических наук.

**О.Н. Дудич**

*Окончила Саратовский государственный медицинский институт по специальности «Лечебное дело». Защитила диссертацию на соискание ученой степени кандидата медицинских наук на тему «Профилактика послеоперационных воспалительных осложнений в хирургии возрастной катаракты с использованием инъекционной формы кислоты ацетилсалициловой». Является автором 62 печатных работ, 1 патента, 10 учебно-методических пособий, 5 инструкций по применению. Основной исполнитель научного проекта «Разработка и внедрение изделия на основе нерезорбируемых материалов для устранения посттравматических дефектов и деформаций глазницы и глазничного орбитального комплекса с использованием технологий 3D моделирования и прототипирования».*

**В.Л. Красильникова**

*Окончила Гродненский государственный медицинский институт по специальности «Лечебное дело». Защитила диссертацию на соискание ученой степени кандидата медицинских наук. В 2007 г. защитила диссертацию на соискание ученой степени доктора медицинских наук на тему «Медико-социальная реабилитация пациентов с анофтальмом с помощью композиционного офтальмологического имплантата: клинкоэкспериментальное исследование». Является автором более 130 печатных работ, 5 патентов, 10 учебно-Пятая Международная научно-практическая конференция «BIG DATA and Advanced Analytics. BIG DATA и анализ высокого уровня», Минск, Республика Беларусь, 13-14 марта 2019 года 55 методических пособий, 3 инструкций по применению. Под ее руководством защищена 1 диссертационная работа на соискание ученой степени кандидата медицинских наук. Научный руководитель проекта «Разработка и внедрение изделия на основе нерезорбируемых материалов для устранения посттравматических дефектов и деформаций глазницы и глазничного орбитального комплекса с использованием технологий 3D моделирования и прототипирования».*

**К.Д. Яшин**

*Руководитель научно-исследовательской группы НИГ – 7.1 «Системы и приборы экологического мониторинга в управлении безопасностью жизнедеятельности».*

**В.С. Осипович**

*В 2004 году окончил Белорусский государственный университет информатики и радиоэлектроники по специальности «микроэлектроника», а в 2005 году получил степень магистра по той же специальности. В 2010 году защитил диссертацию на соискание ученой степени кандидата технических наук по приборам, системам и медицинской технике. Он является автором 115 публикаций, 2 патентов, 12 учебных пособий. Ежегодно он устно выступает на международных научных форумах. Основные направления научной деятельности: разработка алгоритмов и технологий обработки больших данных, исследования и разработки в области программной обработки медицинских изображений.*

**Аннотация.** Представлены результаты исследований по разработке и реализации в виде программного средства поиска глазных яблок и костных структур глазниц. Для достижения цели было использовано глубокое обучение нейронной сети. Показано, что погрешность расчёта объёмов глазниц на основе результатов распознавания орбит нейронной сетью составляет 4 – 8 %.

**Ключевые слова:** глазное яблоко, костные структуры глазницы, экстраокулярные мышцы, ретробульбарная клетчатка (ее смещение).

*Введение*

Методы, основанные на машинном обучении (ML) и особенно глубоком обучении (DL), способны выявлять, локализовать и количественно определять патологические особенности практически при любом заболевании, в частности, и при патологии глазницы.

В то же время, анализ изображений, полученных с применением микроспирального компьютерного томографа MSCT [5-7] в формате DICOM [8] и трехмерная реконструкция костей лицевого черепа, позволяет хирургу более надежно оценить анатомические особенности отдельного пациента, локализацию, границы и распространенность патологического процесса, и планировать объем операции [9, 10].

Для работы нейронной сети с полученными результатами работы компьютерного томографа за основной элемент программной обработки были взяты биомаркеры.

Определение биомаркеров – основные элементы глазницы и глазного яблока, которые должна распознать нейросеть после ее обучения.



*Цель настоящей работы* – получение программного средства, позволяющее однозначно идентифицировать глазные яблоки и глазницы

### **Методы и материалы**

*Исходные данные.* В качестве исходных данных для обучения нейронной сети были использованы результаты микроспиральной компьютерной томографии 70 пациентов с переломами костей глазницы различной тяжести. По каждому пациенту анализировался набор изображений в формате DICOM, полученных с использованием результате Микроспирального компьютерного томографа Siemens Emotion 6 (Германия).

Подготовка изображений для обучения нейронной сети осуществлялась путём послойной разметки этих изображений. До процесса разметки была осуществлена конвертация DICOM изображений в RGB изображения. В качестве инструмента для разметки данных использовали приложение VGG Image Annotato, это приложение для ручного аннотирования изображений с возможностью осуществлять множественную разметку. Результатом разметки являются файлы в формате csv и json, которых хранят информацию о координатах точек, ограничивающих полигон (результат разметки глазницы) в привязке к названию файла. Рисунок 1 отражает пример разметки исходных файлов.

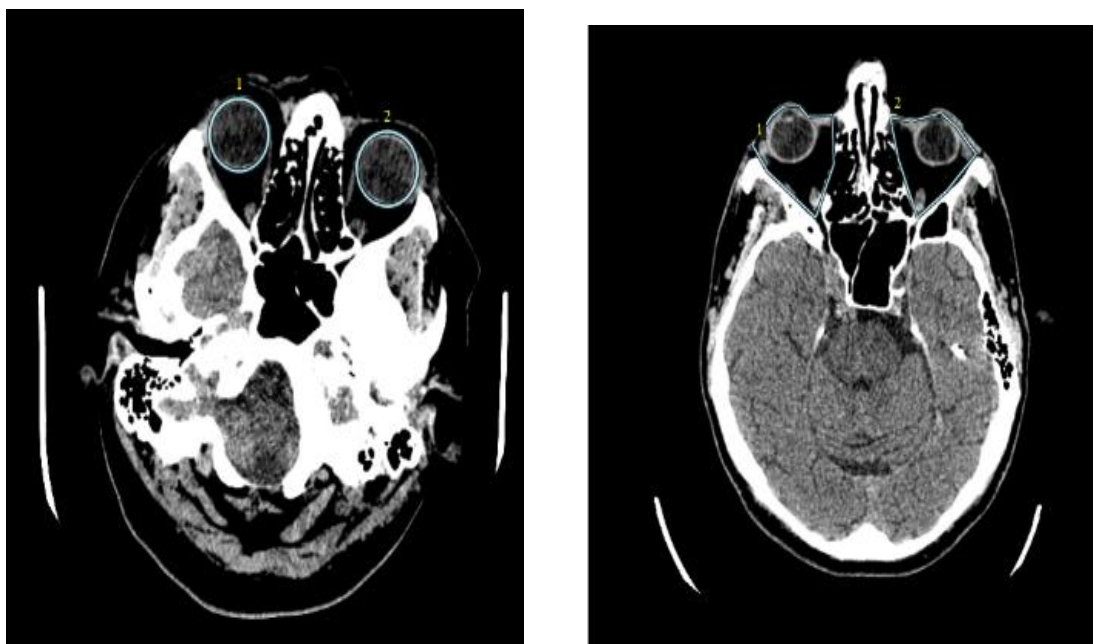


Рисунок 1. – Результат разметки исходных данных в одном слое

Все исходные данные были разделены на тренировочные и тестовые в процентном соотношении 80 и 20 % соответственно. Тестовые исходные данные использовались для проверки работы нейронной сети после обучения. Кроме того, тестовые исходные данные использовались в качестве контрольного эксперимента для сравнения результатов расчёта объёма глазниц размеченных нейронной сетью с объёмом глазниц, размеченных вручную.

*Обучение нейронной сети.* Решение поставленной задачи было осуществлено средствами языка программирования Python с использованием платформы Anaconda. В качестве основных фреймворков для работы с нейронными сетями мы использовали Tensorflow и Keras. На основе архитектуры U-net была разработана новая архитектура нейронной сети. Обучение нейронной сети происходило со следующими параметрами: итераций на эпоху – 100, learning rate – 0,001, регулязация – 0,0001, минимальная вероятность при детектировании – 0,95.

Количество изображений, использованных для обучения нейронной сети: снимков с биомаркерами 800, снимков глазниц 1633. Аугментация в нейронной сети осуществлялась при помощи алгоритма, заложенного в фреймворке Keras, что позволило в 20 раз увеличить наборы изображения для обучения нейронной сети.

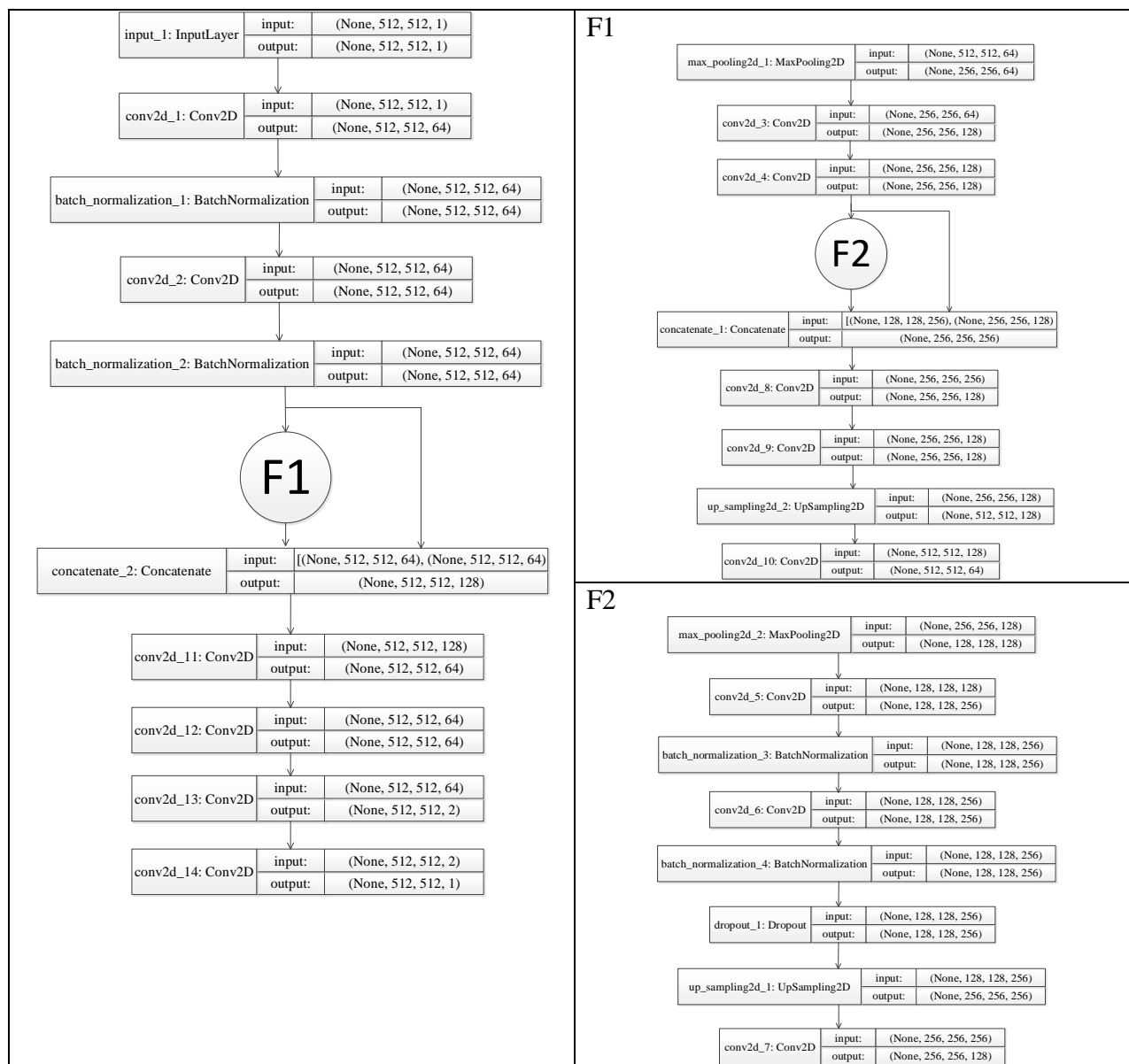


Рисунок 2. – Архитектура алгоритма работы нейронной сети

Процедурная составляющая нейронной сети представлена на рисунке 2 в виде алгоритма работы нейронной сети.

*Алгоритм работы программного средства.* Разработанное программное средство может работать в двух режимах: режиме обучения нейронной сети и режиме поиска контуров глазниц, глазных яблок и расчёта их объёмов.

глазницы, глазного яблока и расчёта их объёмов состоит из следующих трёх процедур:

- распознавание нейронной сетью по входным изображениям в формате DICOM контуров глазниц и их разметка;

- проверка верности разметки правой и левой глазниц, а также глазных яблок коррективная результатов разметки;

–расчёт объёмов глазниц, глазных яблок и вывод результатов через интерфейс пользователя.

Разработка дополнительной процедуры для однозначной идентификации правой и левой глазниц потребовалась в связи с тем, что в результате работы нейронной сети в некоторых слоях правая глазница и левая глазница были отмечены с ошибкой: левая как правая, а правая как левая. На рисунке 3 отражена блок-схема алгоритма корректировки результатов распознавания нейронной сетью контуров глазниц.

Для расчёта объёма глазниц использовали следующую формулу:

$$V = N \times a^2 \times h \quad (1)$$

где  $N$  – количество пикселей, размеченных, как часть глазницы;

$a$  – длина стороны пикселя, мм;

$h$  – расстояние между слоями, мм.

Значения длины стороны пикселя  $a$  и расстояния между слоями  $h$  зависят от технических характеристик и разрешающей способности аппарата микроспиральной компьютерной томографии. В нашем случае разрешение исходных DICOM файлов составляет  $512 \times 512$  пикселей и  $a = 0,455$  мм, а расстояние между слоями  $h = 0,625$  мм.

Реализованные средствами языка программирования Python процедуры были апробированы на 14 наборах DICOM файлов с целью оценки погрешности разметки нейронной сетью глазниц.

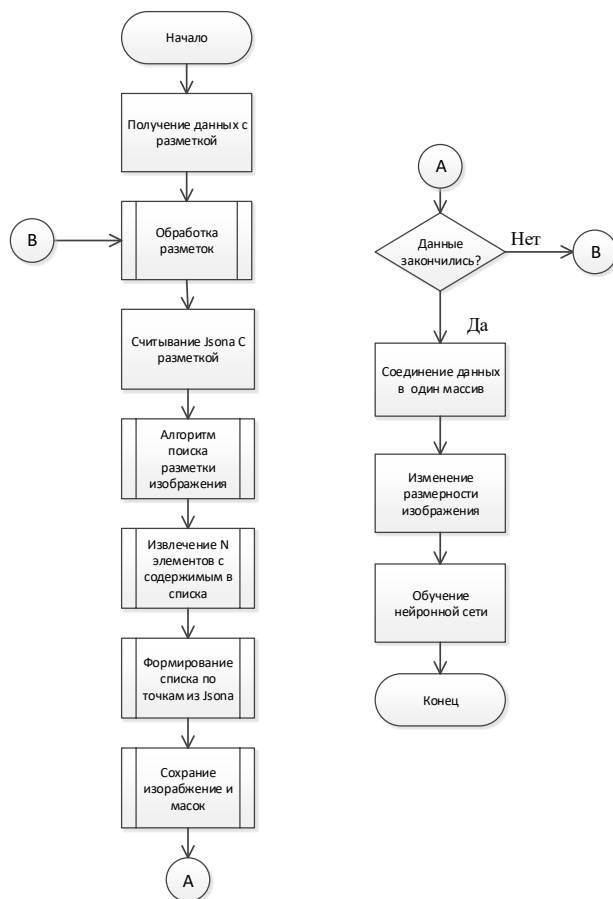


Рисунок 3. – Блок-схема алгоритма корректировки результатов поиска контуров глазниц

### Результаты и обсуждение

В результате обучения нейронной сети была получена матрица весовых коэффициентов для каждого слоя нейронной сети на основе которых происходила разметка глазниц в тестовых наборах данных.

Зависимости значения метрики достоверности и уровня потери данных от времени обучения нейронной сети отражены на рисунках 4 и 5, соответственно.

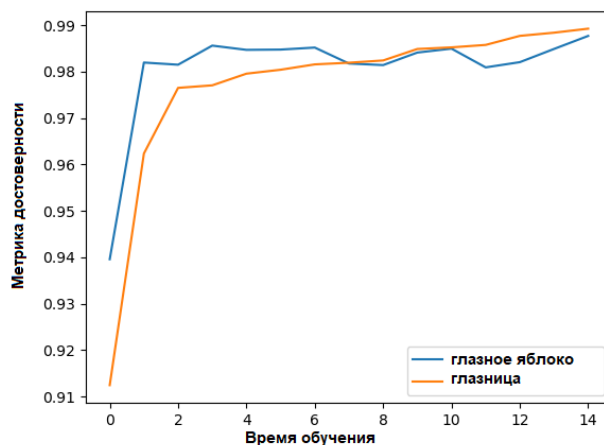


Рисунок 4. – Зависимость значения метрики достоверности от времени обучения (количества итераций)

$$Accuracy = \frac{tp + tn}{tp + fp + fn + tn} \quad (2)$$

где Accuracy – метрика достоверности;

*tp* (True Positive) – истинноположительный результат. Классификатор решил, что цвет совпадает, и он совпал;

*fp* (False Positive) – ложноположительный результат. Классификатор решил, что цвет совпадает, но он не совпал. Это так называемая ошибка первого рода;

*fn* (False Negative) – ложноотрицательный результат. Классификатор решил, что цвет не совпадает, а он совпал. Это так называемая ошибка второго рода. Обычно при создании модели желательно минимизировать ошибку второго, даже увеличив тем самым ошибку первого рода;

*tn* (True Negative) – истинноотрицательный результат. Классификатор решил, что цвет не совпадает, и он не совпал;

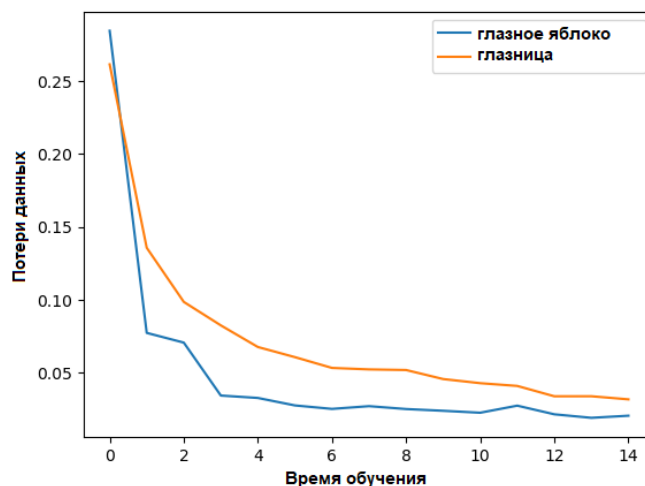


Рисунок 5. – Зависимость потерь данных от времени обучения нейронной сети (количества итераций)

На рисунке 5 четко прослеживается сокращение потерь биомаркеров за время обучения нейронной сети.

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^N (y * \log(\hat{y}_i) + (1 - y) * \log(1 - \hat{y}_i)) \quad (3)$$

где  $y$  – метка класса;

$\hat{y}$  – то что предсказывает модель;

$L$  – loss (функция нейронной сети) необходимая для идентификации положительных принципов обучения нейронной сети.

В результате обработки изображений тестовых наборов данных с использованием обученной нейронной сети мы получаем бинарную маску изображения (вероятности принадлежности к классу). На рисунке 7 отражён результат поиска контуров глазных яблок, построенных нейронной сетью используя исходное изображение слоя, поступавшего на вход программного средства.

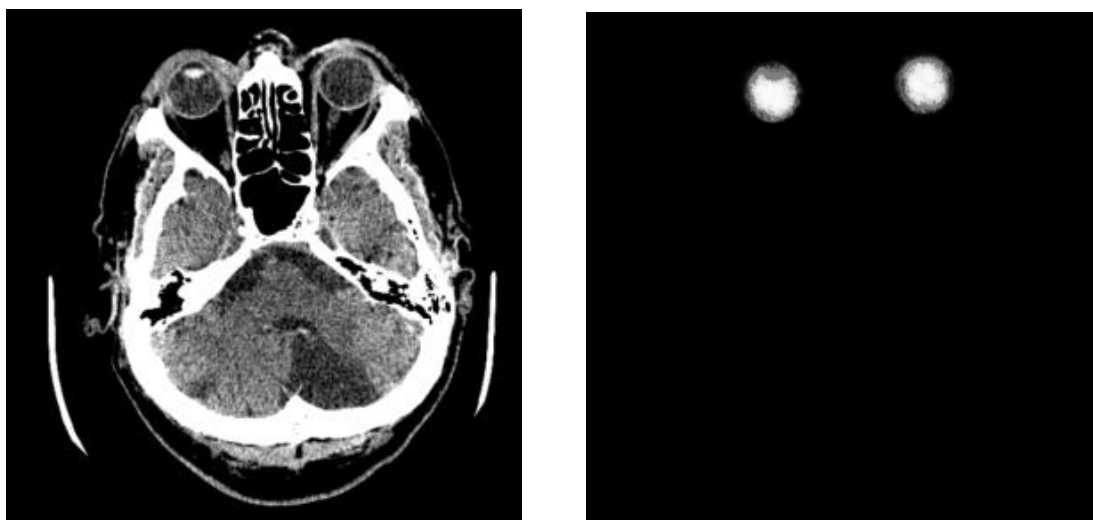


Рисунок 6. – Результат поиска глазного яблока нейронной сетью после обучения

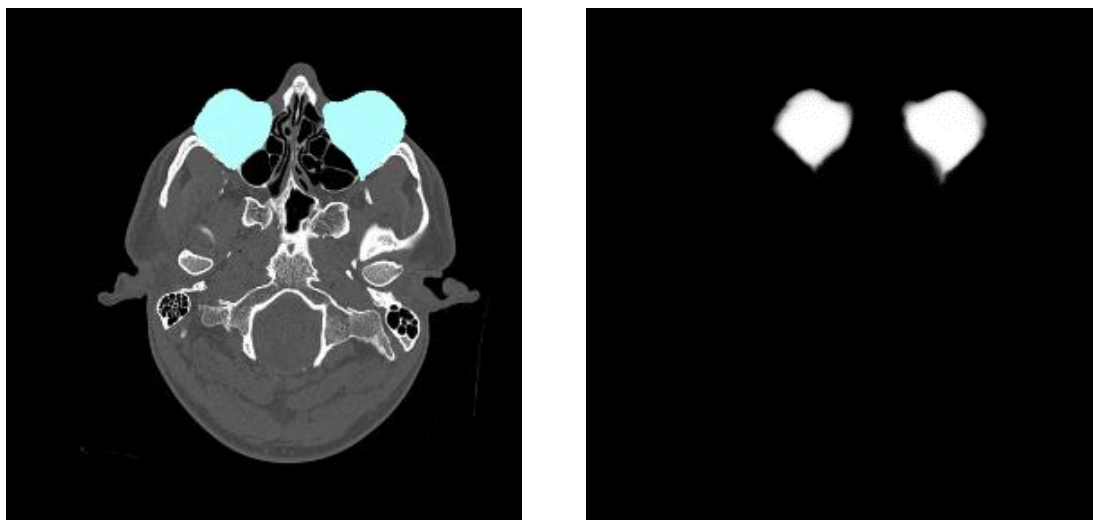


Рисунок 7. – Результат поиска глазниц нейронной сетью после обучения

Далее на основании наборов файлов с контурами глазниц осуществлялся расчёт объёма правой и левой глазниц по формуле (1).

Результат сравнения объёмов глазниц, рассчитанных по результатам разметки нейронной сетью, с объёмами глазниц, рассчитанных по результатам ручной разметки, показал, что разница составляет 4-8%. Это свидетельствует о высокой точности разметки глазниц нейронной сетью. Предложенное программное средство целесообразно использовать для автоматизации процесса расчёта объёма глазниц на этапе подготовки к операции и при оценке результатов операции по замещению тонких костей глазницы.

#### *Заключение*

Установлено, что погрешность расчётов объёма глазниц на основе биомедицинских изображений (результатов компьютерной томографии) с использованием нейронной сети составляет 4-8%.

Разработано и апробировано программное средство, позволяющее сократить затраты времени на подготовку к операции по замещению тонких костей глазницы на 30-40 минут.

#### *Список литературы*

- [1] Levchenko O.V., Krylov V.V., Davydov D.V., Lezhnev D.A., Mikhayliukov V.M., Sharifullin F.A., Kostenko D.I. The computer tomography for estimation of surgical reconstruction efficacy for treatment of posttraumatic defects and orbit deformations. Russian journal of neurosurgery. 2014, (1): Pp. 29-33.
- [2] Levchenko O.V., Mikhailiukov V.M., Davydov D.V. Frameless navigation system for surgical treatment of posttraumatic defects and cranioorbital deformations. Russian journal of neurosurgery. 2013, (3): Pp. 9-14.
- [3] Davydov D.V., Lezhnev D.A., Kostenko D.I. Diagnostic MSCT and Planning for Surgical Treatment for Patients with Orbital-Wall Injuries and Post-Traumatic Deformities. Doctor.ru. Gastroenterology 1(118), 2016, Pp. 116-120
- [4] IFMBE at <https://studylib.ru/doc/2399666/v.a.-stuchilov-%E2%80%A2-a.a.-nikitin---moskovskij-oblastnoj-nauchno>
- [5] Masalitina N.N., Kurochka K.S. (2017) The automated classification of computer tomography results for lumbar spine geometric modeling. Doklady BGUIR. (3), Pp. 12-19.
- [6] Patent 2000125015/14 RF A61B 17/80, A61B 17/56, A61F 2/28. Rybalchenko's plate for reconstructing damaged bone structures of cheekbone, eye-socket and upper jaw complex and the way of reconstructing damaged bone structures of cheekbone, eye-socket and upper jaw complex with this plate. / Rybalchenko G.N.; Closed Joint Stock Company "KONMET Incorporated", Rybalchenko G.N. – Application № 2202302, 04.10.2000; published 20.04.2003.
- [7] Patent 2000116451/14 RF A61 F2/28. Method manufacturing an individual precision implant for compensation of complex subtotal polyostotic eye-socket defect. / Shalunov A. – S.Z.; Shalunov A. – S.Z. – Application № 2164392, 27.06.2000; published 27.03.2001.

[8] Digital imaging and communications in medicine [Электронный ресурс]. – Режим доступа: <https://www.dicomstandard.org/current/> (дата обращения: 25.01.2020).

[9] Aggregation Network for Instance Segmentation [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1803.01534> (дата обращения: 25.01.2020).

[10] Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow [Электронный ресурс]. – Режим доступа: [https://github.com/matterport/Mask\\_RCNN/](https://github.com/matterport/Mask_RCNN/) (дата обращения: 25.01.2020).

## **USE OF THE NEURAL NETWORK IN PROCESSING OF MEDICAL IMAGES AND CALCULATION OF VOLUME OF THE EYE**

**A. Karakulko**

*Graduate student of Human Engineering and Ergonomics*

**O. Medvedev**

*Engineer of Human Engineering and Ergonomics, Master of Technical Science*

**P. Moroz**

*PG, student of Human Engineering and Ergonomics and Radioelectronics*

**A. Dudzich**

*M.D., associate professor, Department of Ophthalmology, Belarusian medical academy of postgraduate education.*

**V. Krasilnikova**

*M.D., Professor, Department of Ophthalmology, Belarusian medical academy of postgraduate education*

**V. Osipovich, PhD**

*Associate Professor, the Chair of Engineering Psychology and Ergonomics*

**K. Yashin, PhD**

*Head of the Department of Human Engineering and Ergonomics, BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus, Minsk  
Belarus Medical Academy of Postgraduate Education, Republic of Belarus, Minsk  
E-mail: o.med@bsuir.by*

**Abstract.** The results of studies on the development and implementation in the form of a software tool for searching eyeballs and bone structures of the orbits are presented. To achieve the goal, deep neural network training was used. It is shown that the error in calculating the orbit volumes based on the results of the recognition of orbits by the neural network is 4-8%.

**Keywords:** eyeball, bone structures of the orbit, extraocular muscles, retrobulbar fiber (its displacement).

УДК 339.138:004

## ЦИФРОВАЯ ТРАНСФОРМАЦИЯ В МАРКЕТИНГОВОЙ ДЕЯТЕЛЬНОСТИ: ОТ АВТОМАТИЗАЦИИ К АЛГОРИТМИЧЕСКОМУ МАРКЕТИНГУ



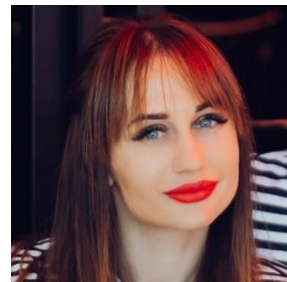
**Пархименко В.А.**  
к.э.н., доцент,  
заведующий  
кафедрой  
экономики БГУИР



**Савчик О.А.**  
м.э.н., Product  
Marketing Manager,  
Rvolt



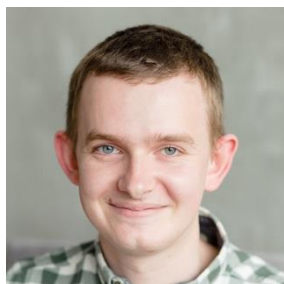
**Верняховская В.В.**  
м.э.н., заместитель  
декана ИЭФ БГУИР



**Ланко М.В.**  
м.э.н., Senior  
Software Testing  
Engineer, EPAM  
Systems



**Князева Л.П.**  
к.ф.-м.н., доцент,  
декан ИЭФ БГУИР



**Дершень А.Г.**  
м.э.н., Senior  
Developer,  
ITranstion



**Дершень В.В.**  
м.э.н.,  
преподаватель  
кафедры экономики  
БГУИР

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

ЗАО «Итранзишэн», Республика Беларусь

ИООО «ЭПАМ СИСТЭМЗ», Республика Беларусь

ООО Rvolt, Федеративная Республика Германия

E-mail: [parkhimenko@bsuir.by](mailto:parkhimenko@bsuir.by), [sosna.victoria@gmail.com](mailto:sosna.victoria@gmail.com), [dershen95@gmail.com](mailto:dershen95@gmail.com),  
[mary.putilina@gmail.com](mailto:mary.putilina@gmail.com), [volha.savchuk@gmail.com](mailto:volha.savchuk@gmail.com), [verniahovskaya@bsuir.by](mailto:verniahovskaya@bsuir.by), [knyazeva@bsuir.by](mailto:knyazeva@bsuir.by)

**Аннотация.** Данная статья имеет двойственный характер: с одной стороны, это обзор тех объективных процессов цифровой трансформации, которые протекают и наблюдаются в сфере маркетинговой деятельности, с другой, – обзор некоторых достижений группы исследователей, работающих и работавших на кафедре экономики Белорусского государственного университета информатики и радиоэлектроники и пытающихся эти объективные процессы зафиксировать и проанализировать.

**Ключевые слова.** Цифровая трансформация. Информатизация и автоматизация маркетинга. Высокотехнологичный маркетинг. Алгоритмический маркетинг. Интернет-маркетинг. Цифровой маркетинг.



## 1. Авторская классификация этапов цифровой трансформации маркетинга

Цифровая трансформация – не только модный термин в современной научной и публицистической литературе, но и объективный факт изменений, свершающихся на наших глазах в настоящее время.

Маркетинг как сфера практической деятельности не избежал ощутимого воздействия со стороны мощного прогресса в информационных технологиях и их повсеместного внедрения в бизнес-процессы коммерческих организаций.

Для фиксации эволюции маркетинга в данном контексте авторы предлагают концептуальную модель, классифицирующую 4 основных этапа цифровой трансформации маркетинга (рисунок 1).



Рисунок 1. – Авторская классификация этапов цифровой трансформации маркетинга

Подобная классификация, как и в случае всех других классификаций, во многом условна и отражает прежде всего предпочтения авторов, нежели однозначные и четко отграниченные во времени этапы объективного процесса. Все же представляется, что следующее выделение существенных, узловых точек цифровой трансформации маркетинга, является полезным для теории и практики.

Рассмотрим более подробно каждый из выделяемых этапов.

**Этап теоретического осмысления и разработки математического аппарата моделей**, по мнению авторов (см. например, исторический обзор в [1]), начался уже в конце 1950-х годов, когда стали появляться научные публикации об использовании математических моделей в маркетинге. Отправной точкой в этом контексте следует считать публикацию классической монографии Филипа Котлера «Принятие маркетинговых решений: модельный подход» [2], изданной в 1971 году и охватывающей основные существующие на тот момент математические модели в маркетинге. В монографии Ф. Котлер делает очевидное (однако на тот момент сугубо теоретического характера) утверждение, что, имея в распоряжении математическую модель и набор ограничений, можно поручить компьютеру поиск решения, оптимизирующего целевую функцию, определяемую владельцем бизнеса. Монография переиздавалась с существенными дополнениями и изменениями в 1983 [3] и в 1992 [4] годах, при этом Ф. Котлер, став к тому времени признанным "отцом-основателем" современного маркетинга, уступил место ведущего автора Гэри Лилиэну. Последний в дальнейшем (в конце 1990-х – начале 2000-х гг.) создал свое собственное новое направление в маркетинговой науке – маркетинговую инженерию (marketing engineering), т.е. методологию построения и использования математических моделей и компьютерных технологий для принятия научно-обоснованных решений в маркетинговой деятельности. Г. Лилиэн организовал консалтинговую компанию “DecisionPro”, разработал специализированное программное

обеспечение и издал ряд учебников по маркетинговой инженерии в 2004 г. [5] и в 2017 г. [6].

Начиная с 1989 года неоднократно переиздавалась еще одна классическая монография «Модели реакции рынка: эконометрический анализ и временные ряды» [7]. В издании 2006 года авторы впервые отметили, что разрабатываемая десятилетиями теория и математические модели стали наконец-то активно использоваться на практике. Фактором этого послужило бурное развитие информационных технологий, позволивших накапливать большие объемы данных о популяционном поведении с дальнейшим их анализом и выходом на оптимизационные задачи.

Серьезными обзорными работами в последние десятилетия были два издания (2008, 2017 гг.) коллективной монографии «Справочник по моделям принятия маркетинговых решений» под редакцией Б. Виеренга и Р. Ван дер Ланса [8, 9]. Отметим, что авторы монографии также указывают на ведающую роль информационных технологий в прогрессе своей области научных исследований, что, помимо прочего, привело к появлению новых 13 глав (из 17 глав всего, т.е. 76% было обновлено!) во втором издании монографии, т.е. всего за 9 лет произошли существенные изменения.

В Беларуси математическими моделями в маркетинге серьезно занимались и опубликовали обзорные книги Н.Н. Анохина [10], А.В. Сак и В.А. Журавлев [11]. Стоит отметить, что отечественные исследования лишь косвенно были направлены на проблему автоматизации принятия маркетинговых решений посредством компьютерных программ.

Следующий этап – этап **информатизация и автоматизация отдельных маркетинговых операций** – по мнению авторов, характеризуется внедрением информационных технологий (ИТ) в маркетинговую деятельность для автоматизации отдельных наиболее простых процессов.

М.В. Лапко (Путилина) в рамках своего диссертационного исследования изучила многочисленные источники, как отечественные, так и зарубежные. Тем не менее, наибольшее влияние на разработанную ей классификацию оказал подход А. Вороного [12], а также классификации, предложенные исследовательскими компаниями Gartner и Forrester. Также были использованы данные практического опыта повседневной работы отделов маркетинга различных компаний, выявленные в ходе масштабного анкетирования. Предлагаемый автором перечень включает 20 категорий ИКТ, используемых в маркетинге, в частности: CRM-системы; MRM-системы; HelpDesk/ServiceDesk-системы; BI-системы (в области анализа продаж и продуктов); Системы интернет-статистики/аналитики; Системы аналитики мобильных приложений и мобильных сайтов; Системы поддержки креатива и рекламной деятельности; Системы связи оффлайна с онлайн; Геоинформационные системы (ГИС); Маркетинговые пакеты «полного цикла»; Сервисы автоматизации e-mail рассылок; «Офисные» системы; Системы управления контентом сайта (Web Content Management Systems); Системы мониторинга бренда и социальных медиа (онлайн-репутации, sentimentanalysis); Пакеты создания визуального контента (графика, видео, gif-файлы); Автоматизированный онлайн-помощник (Automated online assistant); Рекомендательные системы (Recommender system); Сервисы для размещения контекстной рекламы; Дополненная реальность (Augmented reality); Инструменты для разработки новых продуктов (см. например, [13, 14]).

Указанные процессы нашли свое отражение и в научных публикациях других авторов. Так, в последние десятилетия был создан большой пласт литературы по автоматизации маркетинга (Marketing automation), которая, правда, до недавнего времени

большой акцент делает на внедрении система класса CRM, автоматизации внутрифирменных маркетинговых процессов (в том числе документооборота), а не использовании математико-статистических моделей.

**Этап возникновения высокотехнологичного маркетинга** – это, по мнению

авторов, новое качество маркетинга, который на данном этапе характеризуется добавлением более сложных («высоких») технологий, а также аппаратных средств, и более тотальным (всеохватывающим) внедрением информационных технологий на всех уровнях маркетинговой деятельности [15, 16].

Авторы полагают, что на данном этапе маркетинг из сугубо управленческой (по сути, гуманитарной) науки, базирующейся во многом на интуиции и здравом смысле лица, принимающего решения, а также его компетенции в сфере психологии потребителей, постепенно получает существенный технический уклон и становится высокотехнологичной прикладной дисциплиной [16].

Согласно определению экономического словаря, высокие технологии – это технологии, появляющиеся в ходе научно-технического развития. Часто используют заимствованное из английского языка выражение хай-тек (high-tech), обозначающее процессы с использованием передовых технологий [17]. Высокие технологии требуют масштабного задействования научных и материально-технических ресурсов, представляют собой передовой рубеж развития науки и техники, воплощают в жизнь самые свежие открытия и изобретения [18].

К высоким технологиям в настоящее время можно отнести: беспроводные технологии; нанотехнологии; робототехнику; электронику; программное обеспечение, в частности исследования в области искусственного интеллекта; системы безопасности; навигационные технологии; экологически чистые технологии (альтернативные источники энергии и переработка отходов); социальные технологии (системы распространения новостей, когнитивистика); биотехнологии [19].

Далеко не все перечисленные выше высокие технологии используются в маркетинге сегодня (скажем, нанотехнологии), но программное обеспечение и беспроводные технологии, а также технологии когнитивистики и инструменты искусственного интеллекта нашли активное применение во множестве конкретных ситуаций.

Отметим, что интернет-маркетинг и вообще феномен электронной коммерции – это отдельная область (относительно новая в истории развития маркетинга), которую, по всей видимости, все-таки можно рассматривать и как феномен высокотехнологического маркетинга.

Следует отметить также, что границы высокотехнологичного маркетинга постоянно смещаются. Это заложено в определении высоких технологий: то, что было ново и актуально несколько лет назад, сегодня может перейти в разряд обычных технологий.

С точки зрения авторов данной статьи, в настоящий момент новыми в маркетинге в первую очередь выступают технологии, связанные с большими данными (big data) и интеллектуальным анализом данных (data mining, knowledge discovery, predictive analytics). Так, некоторые инструменты высокотехнологичного маркетинга прямо базируются на сборе и анализе больших данных, тогда как другие не имеют непосредственной связи с этими процессами, но могут быть связаны с Big Data. Например, QR-коды и NFC-метки могут приводить пользователей в места, где данные будут собраны.

Визуализация связи высокотехнологичного маркетинга с большими данными, предложенная В.В. Дершень в 2017 г., изображена на рисунке 2.

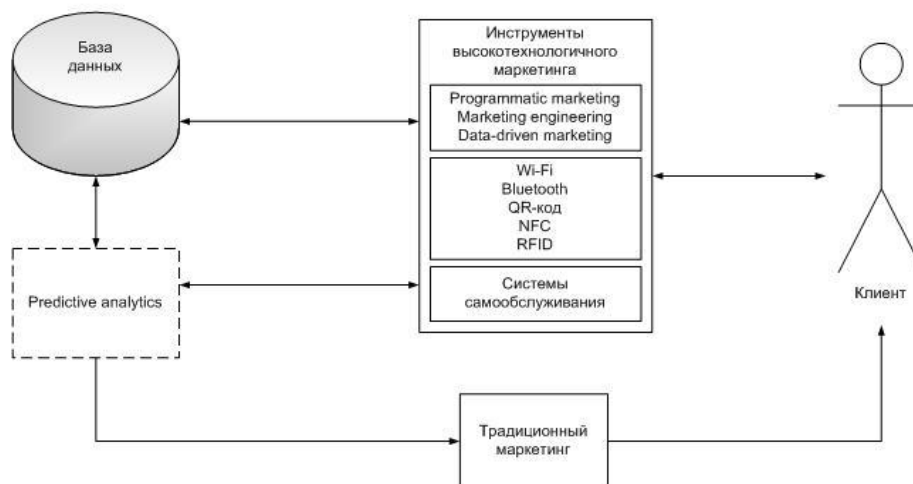


Рисунок 2. – Система высокотехнологичного маркетинга на основе больших данных [20]

Инструменты высокотехнологичного маркетинга могут использоваться как в сборе информации о потребителях для дальнейшей работы с ней, так и для обратного воздействия по результатам анализа данных. Кроме того, собранные с помощью высокотехнологичных решений данные могут использоваться для воздействия на потребителей инструментами традиционного маркетинга.

**Этап алгоритмического маркетинга** – естественное продолжение эволюции маркетинга, этот этап подразумевает интеллектуализацию и автономизацию процессов принятия решений, т.е. их делегирование алгоритмам.

В самом широком смысле алгоритмический маркетинг можно определить, как маркетинговую деятельность, осуществляемую автоматически, без участия (или при самом минимальном участии) человека [21, 22, 23]. Информационные системы (компьютерное оборудование и программное обеспечение) используют данные о поведении покупателей (в первую очередь об онлайн-поведении), самостоятельно принимают маркетинговые решения, реализуют их и оценивают последствия для достижения целей, сформулированных владельцами бизнеса, топ-менеджерам и/или специалистами по маркетингу.

С точки зрения авторов, выход в 2017 году книги Ильи Кацова «Введение в алгоритмический маркетинг: искусственный интеллект для маркетинговых операций» [24], если не по содержанию, то по духу, можно считать особой вехой в рассмотренной истории научных исследований. В книге постулируется необходимость синтеза достижений различных наук для создания информационных систем, способных принимать миллионы «микро-решений» (micro-decisions) в маркетинговой деятельности.

Анализ научных публикаций позволяет выделить следующие основные области алгоритмического маркетинга, в которых в настоящее время ведутся интенсивные научные исследования и прикладные разработки [23]: Прогнозирование выбора потребителя (Customer Choice Prediction); Динамическое ценообразование (Dynamic Pricing); Прогнозирование оттока клиентов (Churn Prediction); Рекомендательные системы (Recommendation Systems); Автоматическая закупка рекламы (Programmatic Advertising); Анализ потребительской корзины (Market Basket Analysis); прогнозирование популярности контента (Content Popularity Prediction) и др.

Очевидно, что аппарат алгоритмического маркетинга еще будет развиваться, а ИТ-

компании будут предлагать готовые технические решения (информационные системы и модули), позволяющие реализовать автоматическое принятие маркетинговых решений в практике существующих компаний (в первую очередь, онлайн-магазинов). Важной задаче выступает включение изучения основ алгоритмического маркетинга в учебный процесс подготовки менеджеров, маркетологов, экономистов [22].

Можно ли осуществить полную алгоритмизацию (т.е. делегирование всех решений компьютерным алгоритмам и системам) всех областей маркетинга – вопрос, открытый для будущих исследований и технологического развития. На текущий момент ответ скорее отрицательный, чем положительный, так как поведение покупателей (зачастую иррациональное, импульсивное и нелогичное) остается слишком сложным объектом для моделирования и предсказания.

## **2. Авторская классификация используемых технологий в маркетинге**

Дать системный и комплексный обзор в области используемых и перспективных технологий в маркетинге – задача непростая и на сегодняшний день, вероятно, не имеющая однозначного решения.

Тем не менее авторы хотят высказать некоторые свои мысли по этому поводу, предложив следующую укрупненную классификацию (без учета описанных выше этапов цифровой трансформации маркетинга) уже широко используемых подходов, технологий и инструментов.

С нашей точки зрения, можно выделить такие основные группы маркетинга, базирующиеся на специфических группах технологий:

- маркетинг, использующий беспроводные технологии (Bluetooth, Wi-Fi, RFID, NFC);
- маркетинг, использующий технологии Big Data, Data Mining, AI (Data-driven marketing, programmatic marketing, marketing engineering, dynamic pricing);
- интернет-маркетинг с его набором специфических digital-инструментов;
- маркетинг, в котором отдельные операции автоматизированы;
- маркетинг, использующий технологии виртуальной и дополненной реальности;
- маркетинг, использующий нейрофизиологические технологии (нейромаркетинг).

Рассмотрим каждый из упомянутых выше видов более подробно, предложив краткое описание наиболее типичных технологий. Подчеркнем при этом, однако, то, что четких границ между выделяемыми направлениями не существует и многие технологические решения являются взаимодополняющими.

**Маркетинг, использующий беспроводные технологии.** К этой группе относятся инструменты, позволяющие передавать информацию на расстоянии, чаще небольшим, например, в личной беспроводной сети (Bluetooth), в локальной беспроводной сети (Wi-Fi), идентифицировать объекты посредством радиосигналов на расстоянии до 300 м (RFID) и до 10 см (NFC), а также QR-коды, которые не являются сами по себе беспроводной технологией, однако дают возможность производить различные операции без проводного взаимодействия, используя лишь сканер (например оплачивать услуги, получать доступ к информации, регистрировать заказы и проч.). Данные инструменты и технологии имеют большое значение в системе современного маркетинга, так как помогают взаимодействовать с пользователями и, таким образом, связаны или могут быть связаны со сбором и анализом больших данных (см. рисунок 2).

Таблица 1 – Некоторые беспроводные технологии, используемые в маркетинге

Наименование	Краткое описание	Пример использования
Bluetooth	Bluetooth-маркетинг – способ реализации маркетинговых коммуникаций с использованием технологии Bluetooth в непосредственной близости от целевой аудитории [8].	iBeacon-маячки, способные организовать навигацию по помещению или выступать в качестве маркетингового канала
Wi-Fi	В маркетинге Wi-Fi можно использовать для рекламной коммуникации, сбора информации о клиентах, информировании о новинках, скидках, актуальных предложениях, программах лояльности, а также, благодаря собранным файлам cookie, поощрять дальнейшие визиты и активность.	Страницы загрузки при подключении к сети Wi-Fi в ресторане
RFID, NFC	Радиочастотная идентификация (RFID) – это современная технология, используя которую информация, необходимая для уникальной идентификации конкретного объекта, дистанционно записывается или считывается с наклеенной или встроенной в объект метки, с помощью радиоволн. NFC (Near Field Communication) – это технология беспроводной высокочастотной связи малого радиуса действия (до 10 см), позволяющая осуществлять бесконтактный обмен данными между устройствами, расположенными на небольших расстояниях: например, между считывающим терминалом и сотовым телефоном или пластиковой смарт-картой. Технология NFC базируется на RFID-технологии [9]. В маркетинге технологии RFID и NFC применяются на рекламных мероприятиях для распространения промо-контента участниками мероприятия в социальных сетях [10]. Кроме того, возможно считывание меток телефоном для получения информации, участия в акциях, получения доступа к Wi-Fi.	Использование RFID-меток и считывающих устройств для записи заказа в сети ресторанов Marketplace; Использование NFC-меток для загрузки игры в ресторане McDonalds
QR-код	QR-код (quick response) – матричный код (двумерный штрих-код), разработанный и представленный японской компанией Denso-Wave в 1994 году [11]. Основное достоинство QR-кода – это лёгкое распознавание сканирующим оборудованием, что дает возможность использования в торговле, производстве, логистике. Возможностей применения QR-кодов в маркетинге очень много: получение дополнительной информации при сканировании кода; вовлечение клиентов для участия в розыгрышах и акциях; оформление заказа или помещение товара в корзину при сканировании QR-кода; распространение контактной информации; получение обратной связи от потребителя; привлечение внимания клиентов с помощью вирусных кампаний, основанных на любопытстве.	Подарок при сканировании QR-кода в ресторане The Melt, США ( <a href="http://lpgenerator.ru/blog/2014/03/17/10-primerov-pravilnogo-i-nepravilnogo-ispolzovaniya-qr-kodov/">http://lpgenerator.ru/blog/2014/03/17/10-primerov-pravilnogo-i-nepravilnogo-ispolzovaniya-qr-kodov/</a> )

**Маркетинг, использующий технологии Big Data, Data Mining, AI.** Это направление подразумевает в первую очередь обработку и анализ тех данных, которые

получены по итогам взаимодействия с потребителем. Технологий работы с большими массивами данных позволяют оптимизировать и автоматизировать внутренние и внешние процессы компании, отслеживать различные виды данных, и тем самым прогнозировать дальнейшее развитие ситуации, контролировать и управлять результатами компании.

Таблица 2 – Некоторые подходы, технологии и инструменты маркетинга, использующего технологии Big Data, Data Mining, AI

Наименование	Краткое описание	Пример использования
Маркетинг, основанный на данных (Data-driven marketing, Marketing engineering)	Data-driven marketing основывается на анализе массивов потребительских данных. В данном случае работа заключается в максимальной автоматизации и оптимизации внутренних и внешних процессов компании, и отслеживании различных видов данных, таких как коэффициент оттока клиентов, уровень удовлетворенности клиентов, доля привлеченных потребителей, пожизненная ценность клиента, конверсия, прибыль, внутренняя норма доходности, окупаемость и т.д. Собранные данные позволяют прогнозировать, контролировать и управлять результатами компании [1]. Marketing engineering – системный подход к сбору данных и знаний для принятия эффективных маркетинговых решений в различных сферах [2]. В маркетинговой инженерии решения принимаются на основе данных и знаний, что позволяет избежать субъективности и эмоциональных факторов при выборе стратегии [3].	Google Music использует данные для создания персонализированных плейлистов. В качестве источников данных берутся время дня, года, предпочтения пользователя, история прослушиваний и количество прослушиваний ( <a href="https://maddata.agency/whats-new/data-driven-marketing-cto-v-trende">https://maddata.agency/whats-new/data-driven-marketing-cto-v-trende</a> )
Программатик-реклама	Программатик-реклама – это автоматическое предложение цены на показы рекламы в режиме реального времени; совокупность методов закупки рекламы в интернете с использованием автоматизированных систем и алгоритмов для принятия решений о сделке без участия человека на основе социально-демографических и поведенческих данных о пользователях, имеющихся в распоряжении как площадки, так и рекламодателя [4]. Также в литературе можно встретить такие термины, как programmatic advertising, programmatic buying. Основным преимуществом данной технологии является значительное улучшение таргетинга, т.к. можно подобрать формат и содержимое в соответствии с ситуацией, в которой пользователь находится прямо сейчас [5].	Рекламные кампании Яндекс.Директ и Google Ads
Рекомендательные системы	Рекомендательные системы – программы, которые на основе собранных данных пытаются предсказать, какие объекты или товары будут интересны клиенту. Можно выделить два основных типа рекомендательных систем. В первом случае пользователю рекомендуются объекты, похожие на те, что этот пользователь уже употребил. Во втором случае для рекомендации также используются оценки и других пользователей, что часто дает лучший результат [6].	Разделы рекомендаций в интернет-магазинах («Возможно, Вам понравится», «Если Вам понравилось, не пропустите», «С этим товаром также покупают» и т.д.)

Продолжение таблицы 2

Dynamic pricing (динамическое ценообразование)	Динамическое ценообразование – динамическая корректировка цен для потребителей в зависимости от ценности, которую эти клиенты приписывают продукту или услуге [7]. Динамическое ценообразование позволяет менять цену из-за колебания спроса или ситуации в конкурентной среде.	Belavia
--	---	---------

**Интернет-маркетинг с его набором специфических digital-инструментов.** Перенос аспектов традиционного маркетинга в интернете невозможен без учета специфики сети и работы в ней. Все инструменты преследуют одну цель – получить максимальный желаемый эффект от пользователей. При этом digital-инструменты используются в комбинации друг с другом для наибольшего эффекта.

Таблица 3 – Некоторые инструменты интернет-маркетинга

Наименование	Краткое описание	Пример использования
Системы управления контентом (Content Management System, CMS)	Основная задача CMS — предоставление возможности самостоятельного наполнения веб-портала информацией без привлечения технических специалистов. От пользователей движков, как правило, не требуется владения специальными знаниями веб-технологий и программирования (однако для эксплуатации систем управления сайтом необходимо освоить принципы работы и детально ознакомиться с руководством по использованию CMS).	WordPress, Joomla, Drupal
Поисковая оптимизация (Search Engine Optimization)	Поисковая оптимизация (SEO) – комплекс мер по внутренней и внешней оптимизации, для поднятия позиций сайта в результатах выдачи поисковых систем по определённым запросам пользователей, с целью увеличения сетевого трафика и потенциальных клиентов (для коммерческих ресурсов) и последующей монетизации этого трафика.	
Контекстная реклама (Contextual Advertising)	Контекстная реклама – тип интернет-рекламы, при котором рекламное объявление показывается в соответствии с содержанием, контекстом интернет-страницы (лат. contextus – соединение, связь).	Яндекс.Директ, Google Ads
Таргетированная реклама	Таргетированная реклама – это текстовые, медийные или мультимедийные объявления, которые демонстрируются только тем пользователям.	
Веб-аналитика	Веб-аналитика (англ. Web analytics) — система измерения, сбора, анализа, представления и интерпретации информации о посетителях веб-сайтов с целью их улучшения и оптимизации. Основной задачей веб-аналитики является мониторинг посещаемости веб-сайтов, на основании данных которого определяется аудитория сайта и изучается поведение посетителей для принятия решений по развитию и расширению функциональных возможностей веб-ресурса [18]. Системы веб-аналитики в зависимости от сложности дают суммарную информацию по посещениям, выбранным по некоторому срезу (измерениям), заданному пользователем, информацию по просмотрам страниц внутри каждого посещения, детализацию с	Google Analytics, Яндекс.Метрика



Продолжение таблицы 3

	возможностью просмотра всех действий посетителей: движений, мыши, кликов, нажатий клавиш и т. д. По собранной поведенческой информации строятся отчеты в виде карт активности посетителей на странице	
Программы для автоматизации e-mail рассылок	Правильно оптимизированная, автоматическая рассылка способствует развитию бизнеса, помогает укрепить лидерство на рынке и сохранить устойчивый поток дохода. Программы для автоматизации рассылки позволяют создавать рассылки по готовым шаблонам, добавлять в рассылку интерактивные материалы, отправлять письма по локальному времени получателя, сегментируют посетителей сайта и т.д.	UniSender, AMS Enterprise

**Маркетинг, в котором отдельные операции автоматизированы.** В данном разделе описаны инструменты, которые частично автоматизируют работу маркетологов, уменьшая количество трудоемких процессов анализа информации (BI-системы, контент-анализ, геоинформационные системы), упрощают планирование (программы медиа-планирования, MRM-системы, маркетинговые пакеты «полного цикла») и облегчают работу с клиентами, при этом обычно повышая их удовлетворенность (CRM-системы, автоматизированные системы самообслуживания).

Таблица 4 – Некоторые программы автоматизации маркетинговых операций

Наименование	Краткое описание	Пример использования
BI-системы	Business intelligence (сокращённо BI) — обозначение компьютерных методов и инструментов для организаций, обеспечивающих перевод транзакционной деловой информации в человекочитаемую форму, пригодную для бизнес-анализа, а также средства для массовой работы с такой обработанной информацией [17]. Системы бизнес-аналитики позволяют объединить накопленные компанией в различных системах фактические данные и превратить их в полезные знания для принятия эффективных решений для бизнеса.	Power BI, Klipfolio
Программы для работы с текстовыми массивами (контент-анализ)	Программы для работы с текстовыми массивами позволяют анализировать большое количество текстовой информации для выделения основных фактов и тенденций. Маркетологу они помогают исследовать рынок, контролировать деятельность конкурентов, выявлять источники угроз со стороны внешнего окружения, проводить мониторинг СМИ и т.д.	ЛЕКТА
Геоинформационные системы	Геоинформационные системы позволяют визуализировать большие объемы статистической информации и анализировать продажи, потребителей, конкурентов – в привязке к их территориальному расположению.	
Программы медиа-планирования	Программы по медиа-планированию предоставляют широкие возможности по созданию и редактированию расписаний выхода рекламы, используют уникальную технологию прогноза аудитории для созданных расписаний [20].	Galileo, SuperNova

Продолжение таблицы 4

CRM-системы (Customer Relationship Management)	CRM-система — это прикладное ПО для предприятий, которое предназначено для упрощения и автоматизации взаимодействия с клиентами. Основной целью использования таких программ является увеличение уровня продаж за счет улучшения качества обслуживания заказчиков и оптимизации схемы маркетинга. CRM-система позволяет сохранять актуальную информацию о клиентах, устанавливать бизнес-процессы и анализировать результаты своей работы [15].	amoCRM, Bitrix24, Salesforce
MRM-системы (Marketing Resource Management)	MRM (Marketing Resource Management) — специализированные программные продукты, предназначенные для управления маркетинговыми ресурсами и являющиеся платформой для автоматизации маркетинга. Основные задачи MRM-систем — это повышение общей эффективности маркетинга за счёт оптимизации, автоматизации и ускорения типовых рабочих процессов. MRM, как правило, охватывают область планирования и реализации рекламных кампаний, работу с рекламными материалами и корпоративным стилем. При этом автоматизации подвергаются процессы и данные, обычно не охватываемые другими системами автоматизации — такими как, например, CRM или ERP [16].	BrandMaker
Маркетинговые пакеты «полного цикла»	Обеспечивают автоматизацию всех маркетинговых функций компании, включая учет маркетинговой деятельности, маркетинговый анализ, обработку результатов маркетинговых исследований (внешних и внутренних факторов), процесс разработки стратегического и оперативного планов маркетинга, построение и контроль маркетинговых бюджетов (бюджет продаж и коммерческих расходов) [19].	Marketing Analytic
Автоматизированные системы самообслуживания	Под системами самообслуживания понимаются как специальные сервисы на веб-сайтах [42], позволяющие клиентам самостоятельно получить необходимую информацию и выполнить некоторые операции, так и аппаратное обеспечение в офисах, магазинах, ресторанах, позволяющее клиентам делать заказы, совершать покупки и выполнять операции без помощи работников компании.	Интернет-служба сервиса абонента (ИССА) компании velcom; офлайн-система самообслуживания применяется в филиалах Беларусбанк для обмена валют

**Маркетинг, использующий технологии виртуальной и дополненной реальности.** Технологии виртуальной и дополненной реальности помогают привлечь и удержать клиентов, когда все другие методы уже кажутся обыденностью для потенциальных потребителей. Вовлеченность в таком случае гораздо выше, так как клиенты оказываются в центре событий и полностью погружаются в новый, виртуальный мир, где они могут взаимодействовать с окружающей средой.

Таблица 5 – Некоторые технологии виртуальной и дополненной реальности

Наименование	Краткое описание	Пример использования
Виртуальная реальность	Виртуальная реальность (virtual reality, VR) – созданный техническими средствами мир (объекты и субъекты), передаваемый человеку через его ощущения: зрение, слух, обоняние, осязание и другие [12].	Рекламная кампания Marriott для молодых семей, планирующих медовый месяц, позволяющая в условиях виртуальной реальности оказаться на Гавайях или в Лондоне; Рекламная кампания IKEA «Исследуй будущую кухню от лица своего ребенка»
Дополненная реальность	Дополненная реальность (augmented reality, AR) – результат введения в поле восприятия любых сенсорных данных с целью дополнения сведений об окружении и улучшения восприятия информации [13]. Дополненная реальность, в отличие от виртуальной, совмещает реальный мир и виртуальную накладку, которая в прямом смысле накладывается на реальность. Устройства дополненной реальности способны проектировать в режиме реального времени программы и системы, используя пространство реального мира.	Мобильное приложение MSQRD, позволяющее накладывать маски и анимированные эффекты к человеческой внешности
Смешанная реальность	Является следствием объединения реального и виртуальных миров для созданий новых окружений и визуализаций, где физический и цифровой объекты сосуществуют и взаимодействуют в реальном времени.	
Фиджитал (Phygital)	Фиджитал – вид цифрового маркетинга, который объединяет виртуальные и реальные коммуникации, причем не последовательно, а одновременно. Фиджитал использует сильные стороны цифрового опыта человека и офлайна, переносит наши удобные онлайн-привычки в реальный мир.	

### **Маркетинг, использующий нейрофизиологические технологии (нейромаркетинг).**

*Нейромаркетинг.* Нейромаркетинг в первую очередь означает объединение двух отраслей науки: маркетинга и нейробиологии. Задача нейромаркетинга – лучше понять потребителя и его реакцию на маркетинговые раздражители путем прямого измерения процессов в мозге и повысить эффективность методов маркетинга, изучая реакцию мозга [45].

Инструменты нейромаркетинга являются инструментами нейронауки. Их можно разделить на 3 группы. Психофизические инструменты регистрируют физиологические реакции различных частей тела, за исключением мозга и показывают увеличение или снижение неврологических функций. К ним относят измерение гальванической реакции кожи, электромиографию, окулографию, анализ частоты сердечных сокращений и давления.

Инструменты, фиксирующие электрическую активность мозга, отмечают и контролируют изменение неврологической функции, происходящее в течение миллисекунд (электроэнцефалография, магнитная энцефалография, транскраниальная магнитная стимуляция). Инструменты, устанавливающие метаболическую активность мозга, обеспечивают высокую степень пространственных измерений, позволяют определять активность в определенных структурах мозга, как правило, с точностью до миллиметров (функциональная магнитно-резонансная томография, позитронно-эмиссионная томография). Кроме того, возможно использование инструментов разных групп одновременно для получения более точных данных.

*Таблица 6 – Инструменты нейробиологии, применяемые в нейромаркетинге*

Инструмент	Описание
Гальваническая реакция кожи	Измерение электрической чувствительности кожи. Регистрируются изменения в сопротивлении кожи при пропускании слабого электрического тока. Показатели увеличиваются с увеличением потоотделения, что четко коррелирует с состоянием возбужденности симпатической нервной системы.
Электромиография (ЭМГ)	Метод исследования биоэлектрических потенциалов, возникающих в скелетных мышцах животных и человека при возбуждении мышечных волокон с помощью поверхностных или игольчатых электродов.
Окулография (eye-tracking)	Отслеживание движения глаз относительно определённой плоской поверхности (например, экрана). Прежде всего, регистрируются моменты задержки взгляда в определённом месте, а также высокая или низкая скорость движения взгляда.
Анализ частоты сердечных сокращений и давления	Измерение показателей ЧСС и кровяного давления. Повышение показателей говорит о возбуждении.
Электроэнцефалография (ЭЭГ)	Неинвазивный метод исследования функционального состояния головного мозга путём регистрации его биоэлектрической активности. В процессе формируется электроэнцефалограмма – график электрической активности головного мозга.
Магнитная энцефалография (МЭГ)	Бесконтактная регистрация магнитных полей на поверхности скальпа, отражающих протекание токов в нервных клетках коры головного мозга человека, позволяющая рассчитывать интенсивность и локализацию источников тока, формируя в динамике карты активности нейронов.
Транскраниальная магнитная стимуляция (ТМС)	Метод, позволяющий стимулировать кору головного мозга при помощи коротких магнитных импульсов. Магнитные стимуляторы, используемые сегодня, способны генерировать магнитное поле интенсивностью до 2 Тесла, что позволяет стимулировать элементы коры головного мозга на глубине до 2 см. В зависимости от конфигурации ТМС может активировать различные по площади участки коры.
Функциональная магнитно-резонансная томография (фМРТ)	Функциональная магнитно-резонансная томография, позволяющая в сочетании со структурной томографией мозга получать карты деоксигинации гемоглобина, отражающие локальную активацию нервных клеток во время психической деятельности (последовательность BOLD).

Продолжение таблицы 6

Позитронно-эмиссионная томография	В основе этого метода лежит возможность при помощи ПЭТ-сканера отслеживать распределение в организме биологически активных соединений, меченных позитрон-излучающими радиоизотопами – радиофармпрепаратов (РФП). Выбор подходящего РФП позволяет изучать метаболизм, транспорт веществ, экспрессию генов и т.д.
-----------------------------------	---

**Список литературы**

- [1]. Пархименко, В. А. О некоторых вопросах алгоритмического маркетинга / В. А. Пархименко // Информационные технологии и системы 2018 (ИТС 2018) : материалы междунар. науч. конф. (Республика Беларусь, Минск, 25 октября 2018 года) / редкол. : Л. Ю. Шилин [и др.]. - Минск : БГУИР, 2018. - С. 17-20.
- [2]. Philip Kotler. Marketing decision making: a model building approach. – Holt, Rinehart and Winston, 1971 – Business & Economics – 720 p.
- [3]. Gary L. Lilien, Philip Kotler. Marketing decision making: a model-building approach. – Harper & Row, 1983 – Business & Economics – 875 p.
- [4]. Gary L. Lilien, Philip Kotler, K. Sridhar Moorthy. Marketing Models. Prentice-Hall, 1992 - Marketing - 803 p.
- [5]. Gary L. Lilien, Arvind Rangaswamy. Marketing Engineering: Computer-assisted Marketing Analysis and Planning. – DecisionPro, 2004 – Business & Economics – 518 p.
- [6]. Gary L. Lilien, Arvind Rangaswamy, Arnaud De Bruyn. Principles of Marketing Engineering and Analytics. – DecisionPro, Apr 17, 2017 – Business & Economics – 328 p.
- [7]. Dominique M. Hanssens, Leonard Parsons, and Randall L. Schultz. Market Response Models: Econometric and Time Series Analysis. – Springer Science & Business Media, Apr 11, 2006 - Business & Economics - 502 p.
- [8]. Berend Wierenga. Handbook of Marketing Decision Models. – Springer Science & Business Media, Sep 5, 2008 - Business & Economics - 630 p.
- [9]. Berend Wierenga, Ralf van der Lans. Handbook of Marketing Decision Models. – Springer, Jul 12, 2017 – Business & Economics – 598 p.
- [10]. Анохина Н.Н. Математические модели маркетинга : Учеб. пособие для вузов / Н.Н.Анохина. – Мн. : БГЭУ, 2002. – 63 с.
- [11]. Сак А.В., Журавлев В.А. Оптимизация маркетинговых решений. – Мн.: Издательство Гревцова, 2010. – 304 с.
- [12]. Вороной А. Инструментарий маркетолога и аналитика: информационно-аналитические системы и программные продукты. [Электронный ресурс]. Режим доступа [http://www.infowave.ru/publications/2marketolog/20-21\\_09\\_07\\_konf/](http://www.infowave.ru/publications/2marketolog/20-21_09_07_konf/). – Дата доступа: 16.02.2020.
- [13]. Путилина, М. В. Классификация средств автоматизации маркетинговой деятельности на предприятии. // Международная научно-техническая конференция, приуроченная к 50-летию МРТИ-БГУИР (Минск, 18-19 марта 2014 года) : материалы конф. В 2 ч. Ч. 2. - Минск, 2014. - С. 292-293.
- [14]. Путилина, М. В. Использование современных информационно-коммуникативных технологий в маркетинге // Наука и инновации. – 2017. – № 3. - С. 47-49.
- [15]. Пархименко, В. А. Вызовы и пределы высокотехнологичного маркетинга / В. А. Пархименко // Веб-программирование и интернет-технологии. WebConf 2015 : материалы 3-й Междунар. науч.-практ. конф., Респ. Беларусь, Минск, 12–14 мая 2015 г. – Минск : БГУ, 2015. – С. 3–6.
- [16]. Пархименко В.А., Салахов Р.А. Вызовы и пределы высокотехнологичного маркетинга // Маркетинг в России и за рубежом. - 2015. - №6. - С. 3-8.
- [17]. Блэк, Дж. Экономика. Толковый словарь / Дж. Блэк. – М. : Инфра-М, 2000. – 848 с.
- [18]. Высокие технологии – понятие и классификация [Электронный ресурс]. – Режим доступа : <http://xbb.uz/Hi-Tech/Vysokie-tehnologii-ponjatie-i-klassifikacija/>.
- [19]. Высокие технологии – Википедия [Электронный ресурс]. – Режим доступа : [https://ru.wikipedia.org/wiki/Высокие\\_технологии/](https://ru.wikipedia.org/wiki/Высокие_технологии/).
- [20]. Дершень В.В., Пархименко В.А. Система высокотехнологичного маркетинга на основе больших данных // BIG DATA and Advanced Analytics: collection of materials of the third international scientific and practical conference. (Minsk, Belarus, May 3 - 4, 2017) / editorial board: M. Batura [etc.]. - Minsk, BSUIR, 2017. - С. 282-287.
- [21]. Хацкевич Г.А., Пархименко В.А. Преподавание основ алгоритмического маркетинга - необходимость современного бизнес-образования // Актуальные проблемы бизнес-образования: матер. XVII Межд. науч.-практ. конф., 19-20.04.2018, Минск / БГУ, ИБМТ, Ассоциация бизнес-образования; [редкол.: В.В.Апанасович (гл. ред и др. ]. - Минск: Нац. библ. Беларуси, 2018.- С.236-240.

- [22]. Пархименко В.А., Хацкевич Г.А. Задачи и перспективы алгоритмического маркетинга // Веб-программирование и интернет-технологии WebConf2018 : тез. докл. 4-й Междунар. науч.-практ. конф., Минск, 14-18 мая 2018 г. / Белорус. гос. ун-т. - Минск : БГУ, 2018. - С. 5-6.
- [23]. Пархименко В.А. Основные области алгоритмического маркетинга // Экономический рост Республики Беларусь: глобализация, инновационность, устойчивость : материалы XI Междунар. науч.-практ. конф. (Минск, 17 мая 2018 г.). - Минск : БГЭУ, 2018. - С. 356-357.
- [24]. Пуя Katsov. Introduction to Algorithmic Marketing: Artificial Intelligence for Marketing Operations. – Dec 2, 2017 - 506 p.
- [25]. Пролетарский, А. В. Беспроводные сети Wi-Fi / А. В. Пролетарский, И. В. Баскаков, Д. Н. Чирков. – М. : БИНОМ, 2007. – 178 с.
- [26]. Росс, Дж. Wi-Fi. Беспроводная сеть / Дж. Росс. – М. : НТ Пресс, 2007. – 178 с.
- [27]. Bluetooth – Википедия [Электронный ресурс]. – Режим доступа : <https://ru.wikipedia.org/wiki/Bluetooth/>.
- [28]. Bluetooth-маркетинг [Электронный ресурс]. – Режим доступа : <http://www.e-executive.ru/wiki/index.php/Bluetooth-маркетинг/>.
- [29]. Власов, М. RFID. 1 технология - 1000 решений. Практические примеры использования RFID в различных областях / М. Власов. – М. : Альпина Паблишер, 2014. – 218 с.
- [30]. Что такое NFC [Электронный ресурс]. – Режим доступа : <https://faqhard.ru/base/17/01.php/>.
- [31]. Путилина, М. RFID – новое слово в маркетинге?.. / М. Путилина, В. Пархименко // Маркетинг: идеи и технологии. – №1 (51), 2013. – С. 9–14.
- [32]. QR-код – Википедия [Электронный ресурс]. – Режим доступа : <https://ru.wikipedia.org/wiki/QR-код/>.
- [33]. Джеффри, М. Маркетинг, основанный на данных. 15 ключевых показателей, которые должен знать каждый / М. Джеффри. – М. : Манн, Иванов и Фербер, 2013. – 505 с.
- [34]. Marketing engineering – Wikipedia [Электронный ресурс]. – Режим доступа : [https://en.wikipedia.org/wiki/Marketing\\_engineering/](https://en.wikipedia.org/wiki/Marketing_engineering/).
- [35]. Principles of Marketing Engineering [Электронный ресурс]. – Режим доступа : <http://smallbusiness.chron.com/principles-marketing-engineering-77416.html/>.
- [36]. What is Programmatic Marketing? [Электронный ресурс]. – Режим доступа : <http://www.smartinsights.com/internet-advertising/internet-advertising-targeting/what-is-programmatic-marketing/>.
- [37]. Что такое программатик: модный термин или работающая технология? [Электронный ресурс]. – Режим доступа : <http://digitalbee.com/blog/digital-marketing/programmatic-dlya-chaynikov-cto-takoe-programmatik-i-kak-on-rabotaet/>.
- [38]. Как работают рекомендательные системы. Лекция в Яндексе [Электронный ресурс]. – Режим доступа : <https://habrahabr.ru/company/yandex/blog/241455/>.
- [39]. Дершень, А. Г. Проблемы и перспективы внедрения методов динамического ценообразования в Республике Беларусь : автореф. дисс. ... магистра экономических наук : 1-25 80 04/ А. Г. Дершень; науч. рук. В. А. Пархименко. - Минск : БГУИР, 2019. – 8 с.
- [40]. Дершень, А. Г. Динамическое ценообразование: определение, сущность / А. Г. Дершень // Экономика, Бизнес, Инновации. Сборник статей VIII Международной научно-практической конференции. – Пенза: Наука и просвещение, 2019. – С. 192–194
- [41]. Дершень, А. Г. Система понятий динамического ценообразования / А. Г. Дершень // Проблемы экономики и информационных технологий: материалы 55-й научной конференции аспирантов, магистрантов и студентов. (Минск, 23-27 апреля 2019 г.). – Минск: БГУИР, 2019.
- [42]. Котлер, Ф. Маркетинг менеджмент. Экспресс-курс. 3-е изд. / Ф. Котлер, К. Л. Келлер. – СПб. : Питер, 2014. – 480 с.
- [43]. ИССА. velcom [Электронный ресурс]. – Режим доступа : <https://my.velcom.by/>.
- [44]. Терминал валютно-обменный Automated Currency Exchange Machine – ИВА Беларусь [Электронный ресурс]. – Режим доступа : <http://iba.by/products/banksystems/ATM/>.
- [45]. Дершень, В. В. Инструменты нейромаркетинга: проблемы и перспективы / В. В. Дершень // Наука и инновации. – 2018. – №6 . – С. 18–22.

## **DIGITAL TRANSFORMATION OF MARKETING: FROM "PATCHWORK" AUTOMATION TO ALGORITHMIC MARKETING**

**V. A. Parkhimenko**

*Ph. D., Associate Professor,  
Head of the Department of  
Economics of BSUIR*

**V. V. Dershen**

*Master of economics, lecturer  
at the Department of  
Economics of BSUIR*

**A. G. Dershen**

*Master of economics, Senior  
Developer, ITransition*

**M. V. Lapko**

*Master of economics, Senior  
Software Testing Engineer,  
EPAM Systems*

**O. A. Savchyk**

*Master of economics, Product  
Marketing Manager, Rvolt*

**V. V. Vernyakhovskaya**

*Master of economics, Deputy  
Dean of the faculty of  
engineering and Economics*

**L. P. Knyazeva**

*Ph. D., Associate Professor,  
Dean of the faculty of  
engineering and Economics  
of BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus*

*ITransition, Republic of Belarus*

*EPAM Systems, Republic of Belarus*

*Rvolt, Federal Republic of Germany*

*E-mail: parkhimenko@bsuir.by, sosna.victoria@gmail.com, dershen95@gmail.com,  
mary.putilina@gmail.com, volha.savchyk@gmail.com, verniahovskaya@bsuir.by, knyazeva@bsuir.by*

**Abstract.** This article has a dual character: on the one hand, this is a review of those real processes of digital transformation that occur and are observed in the sphere of marketing, on the other hand, it is a review of some achievements of a research group who work (worked) at the Department of Economics of Belarusian State University of Informatics and Radioelectronics and try these real processes to describe and analyze.

**Keywords.** Digital transformation. Informatization and marketing automation. High-tech marketing. Algorithmic marketing. Internet marketing. Digital marketing.

УДК [304.2]

## АНАЛИЗ И ОСОБЕННОСТИ ТЕХНОЛОГИИ BIGDATA В УПРАВЛЕНИИ ЧЕЛОВЕЧЕСКИМИ РЕСУРСАМИ



**В.В. Шаталова**  
заместитель декана  
ФКП,  
канд. техн. наук, доцент



**Т.В. Казак**  
член-корреспондент  
Международной академии  
психологических наук, профессор  
кафедры ИПиЭ, доктор  
психологических наук, профессор



**Лихачевский Д.В.**  
декан ФКП БГУИР  
кандидат технических  
наук, доцент

Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь  
E-mail: shatalova@bsuir.by, kazak@bsuir.by

### **В.В. Шаталова**

Окончила Белорусский государственный университет информатики и радиоэлектроники, факультет компьютерного проектирования. Работает в должности заместителя декана факультета компьютерного проектирования.

### **Т.В. Казак**

Доктор психологических наук, профессор, член-корреспондент Международной академии психологических наук, профессор кафедры инженерной психологии и эргономики Белорусского государственного университета информатики и радиоэлектроники

**Аннотация.** В данной статье рассмотрена значимость применения технологии Big Data в инновационных стратегиях управления, технологий и методов управления людьми в профессиональной деятельности в условиях цифрового общества, проанализированы перспективы применения данной технологии для менеджмента, в том числе и для системы образования. Показаны возможности применения технологий Big Data для современных технологий и методов управления в условиях современного рынка.

**Ключевые слова:** Управление человеческими ресурсами, HR, BigData, информационное обеспечение, интеллектуальные и человеческие ресурсы.

Для современного менеджмента характерно возрастание интереса к роли человеческих и интеллектуальных ресурсов и полной реализации их профессионального потенциала. Это вызвано высоким уровнем разделения труда, существенным ростом удельной доли высококвалифицированного труда как в производственной, так и в непроизводственной сферах и повышением роли каждого сотрудника в успешной работе организации. Поэтому задача разработки современных технологий и качественного инструментария для обеспечения оперативного и эффективного подбора персонала становится одной из наиболее значимых, ключевых в работе служб персонала. Современные условия деятельности большинства организаций требуют, чтобы подбор персонала осуществлялся целеориентированно, оперативно, необходимо учитывать не только конкретные должностные обязанности и требования рабочего места, но и тактические и стратегические цели организации, а также возможность профессионального развития сотрудника, коммуникативные и иные качества



сотрудника, необходимые для профессиональной деятельности и успешной адаптации к новым реалиям ИТ-общества. Такой многофакторный подход к оценке и подбору персонала осуществим лишь при условии интенсивного использования информационных и компьютерных технологий. Именно использование современных информационных технологий создает в организациях, предприятиях, компаниях предпосылки для многократного повышения эффективности производства, деятельности отделов кадров и служб персонала и обеспечит технологически грамотное осуществление управленческой деятельности и функционирование организации в целом [1, с. 95].

Управление человеческими ресурсами (Human resource management – HRM, или просто HR) является политикой организации или внутренней комплексной функцией организации, которая сосредоточена на наборе, управлении и обеспечении деятельности людей, которые работают в организации. Одна из проблем низкой конкурентоспособности современной образовательной организации заключается в недостаточном использовании интеллектуальных ресурсов и недостаточном управлении человеческими ресурсами. Это мотивирует к совершенствованию методов и технологий управления человеческими ресурсами. Управление человеческими ресурсами особенно важно для сферы образования, в которой интеллектуальный капитал и человеческий капитал играют большую роль. Особенностью современной стадии управления человеческими ресурсами в образовании является широкое применение информационных технологий для управления персоналом. Для управления человеческими ресурсами применяют специальный математический и организационный аппарат [2].

Основные функции процесса управления человеческими ресурсами, реализуемые на разных уровнях системы управления организацией, - выработка решений и контроль за их исполнением. Необходимость обеспечения выполнения этих функций дает возможность рассматривать управление человеческими ресурсами как информационный процесс, т. е. функционально включающий получение, передачу, обработку, хранение и использование информации, а саму иерархическую систему управления – как информационную систему. Информационное обеспечение системы управления человеческими ресурсами представляет собой совокупность реализованных решений по объему, размещению и формам организации информации, циркулирующей в системе управления при ее функционировании. Информационное обеспечение включает оперативную, нормативно-справочную информацию, классификаторы технико-экономической информации и системы документации (унифицированные и специальные). К разработке информационного обеспечения службы управления человеческими ресурсами предъявляется ряд организационно-методических требований: рациональная интеграция обработки информации при минимальном дублировании информации в информационной базе, сокращение числа форм документов; возможность машинной обработки информации, содержащейся в документах и во внутримашинной сфере; необходимая избыточность информационного обеспечения, позволяющая пользователям различного уровня получать информацию с различной степенью детализации [1, с. 95].

Рост информации в геометрической прогрессии, из-за активного развития ИТ, оцифровке данных и автоматизации различных бизнес-процессов, приводит к необходимости обработки большого количества документов на ежедневной основе. Фактически они вплотную сталкиваются с проблемой больших данных. Инструменты и методы обработки большого массива информации называются Big Data. Специальные алгоритмы способны не только упростить производственные процессы, но и решить вопрос с подбором подходящих сотрудников. Они позволяют анализировать информацию вне зависимости от ее объема. Долгое время технология больших данных не использовалась активно в управление персоналом, так как специалисты данной области изначально сосредоточены на локальных

процессах, развитии персонала, построении заработных плат, подборе или документообороте, нежели аналитических инструментах [3].

Применение Big Data для работы HR-специалистами в настоящее время представляет огромный интерес. При надлежащей обработке они создают все большее количество инструментов, значительно облегчающих деятельность работников HR-отдела. Big Data – это комплекс инструментов и методов обработки большого массива данных для предоставления их в виде наглядных и воспринимаемых человеком результатов работы отдельных подразделений компании, на основе полученной информации принимаются управленческие решения. Именно поэтому Big Data приобретает в последние годы огромную популярность в управлении человеческими ресурсами [4].

Главные выгоды применения Big Data в управлении персоналом: автоматизация и обобщения функционирующих подходов к поиску и отбору сотрудников, оценка эффективности работы сотрудников; улучшение качества работы персонала и повышение производительности труда, решение задач в области командообразования путем соотнесения качеств людей. Сегодня Big Data в управлении персоналом, в основном, связана с совершенствованием технологий подбора персонала и управления талантами.

В последнее время широкое распространение получила технология «управления талантами, Talent Management», которая входит в сферу HRM [5]. Talent Management используется для привлечения, развития и удержания сотрудников. Термин используется для обозначения особо талантливых сотрудников или сотрудников с высоким потенциалом. Управление талантами часто используется как синоним HRM, хотя поле управления талантами значительно меньше, чем поле HRM. Управление человеческими ресурсами становится также сферой научного исследования, которая пользуется большим успехом в области управления и в области организационной психологии [2].

Анализируемые данные включают в себя опыт работы, пол, социальное положение и связи, личные качества и т.д. На основе этих массивов данных строится модель, которая в будущем служит основой для отбора подходящих сотрудников и подбора для них рабочих мест, а также с помощью Big Data можно получить информацию о том, как будет развиваться карьера сотрудника еще до того, как он выйдет на работу. Big Data позволяет измерить управленческий опыт, предсказать и предотвратить профессиональное выгорание, быстро найти подходящего кандидата и сформировать высокоэффективную команду [6]. При этом HR-данные можно разделить на 2 группы:

- *внешние* находятся за пределами информационного пространства компании – профили кандидатов в социальных сетях, резюме соискателей и описания вакансий на специализированных сайтах по поиску работы;

- *внутренние* сведения о сотрудниках, их обязанностях и показателях деятельности из корпоративных систем, Excel-таблиц и регламентирующих документов, а также неформальные отзывы коллег и руководства [7].

Технологии Big Data позволяют осуществить комплексный сбор и всесторонний анализ человеческого капитала с учетом общей эффективности бизнеса и отдельных направлений деятельности предприятия, а именно:

- ключевые навыки сотрудников;

- привлечение и управление развитием талантов, формирование корпоративной модели компетенций и профессиограмм;

- загруженность работников: соотношение времени, затрачиваемое на создание дополнительной ценности и времени, уходящего на решение административных задач; анализ выполняемых задач и времени, отводящегося для их исполнения; общая загруженность персонала;

–текучесть кадров, прогнозирование потребности в персонале, расчет индексов удовлетворенности и вовлеченности сотрудников;

–аналитика корпоративной культуры, правил поведения и взаимодействия в компании и своевременное выявление нежелательных изменений;

–аналитика каналов привлечения персонала: корпоративного сайта, рекрутинговых площадок, устных рекомендаций – анализ их результативности, сравнительной скорости развития и потенциала;

–оценка вклада руководителей с помощью анализа результативности всей организации и обратной связи от сотрудников, полученной методом анонимных опросов, фокус-групп и личных интервью;

–индивидуальная результативность сотрудников, для поощрения каждого работника в зависимости от его вклада, а также определить зависимость между показателями деятельности специалиста и результатами оценки его навыков и компетенций;

–аналитика группового взаимодействия и взаимовлияния для формирования эффективных команд с учетом зависимости общей результативности группы от входящих в ее состав индивидов [8].

На сегодняшний день технологии больших данных используют новые направления в практике управления человеческими ресурсами: рекрутинг (recruiting), переподготовка кадров, анализ трудовых отношений и прибыль от работы с персоналом.

Рекрутинг и e-recruiting – анализ и подбор наилучших и талантливых кандидатов в большом объеме неструктурированных данных: профили в социальных сетях, фотографии, комментарии, отзывы с предыдущих мест работы и т.д. Использование сетевых технологий (e-recruiting) позволяет HR с помощью электронных рекрутинговых инструментов размещать вакансии на рабочие места и отслеживать претендентов на тысячи рабочих мест в различных местах. Целый ряд характеристик претендентов мгновенно можно увидеть в Интернете: обратная связь, профессиональные качества, научная активность, интервью, судимость, наркозависимость и многое другое. Это помогает специалистам по персоналу отслеживать вакансии быстрее и проще, позволяет устранить ограничения географического положения. Предложения о вакансиях стали доступны всем, у кого есть выход в Интернет. Появились специальные рекрутинговые порталы. [9]. В социальных сетях специалисты HRM могут построить бренд компании путем размещения новостей о компании и фотографий веселых корпоративных мероприятий. Все это имеет прямое отношение и к образованию.

Обучение и развитие профессиональной подготовки создает условия для непрерывного образования и профессионального развития. Это достигается с помощью обучающих программ, оценки эффективности деятельности и программ вознаграждения персонала [2].

### **Список литературы**

[1.] Дейнека А.В. Управление человеческими ресурсами: Учебник для бакалавров / А. В. Дейнека, В. А. Беспалько. — М.: Издательско-торговая корпорация «Дашков и К<sup>о</sup>», 2013. - 392 с.

[2.] Бутко Е.Я. Управление человеческими ресурсами // Образовательные ресурсы и технологии. – 2016. - №5(17), - С.3-9 [Электронный ресурс] <https://www.muiiv.ru/vestnik/pp/>

[3.] Большие данные BigData для HR. Как увидеть личность за цифрой? [Электронный ресурс] <https://hr-media.ru/bolshie-dannye-bigdata-dlya-hr-kak-uvidet-lichnost-za-tsifroj/>.

[4.] Как Big Data помогает в подборе персонала: технологии для хедхантеров [Электронный ресурс] <https://aiconference.ru/ru/article/kak-big-data-pomogaet-v-podbore-personala-tehnologii-dlya-hedhanterov-93666>.

[5.] Армстронг М. Практика управления человеческими ресурсами / Пер. с англ. 8-е изд. – СПб.: Питер, 2004.

[6.] Big Data в кадровом менеджменте хедхантеров [Электронный ресурс] <https://www.hr-director.ru/article/67163-big-data-v-menedjmente-18-m6>

[7.] HR аналитика и «BIG DATA»: что происходит сейчас и чего ожидать завтра? [Электронный ресурс] <https://www.ipm.by/publications/publectors/hr-analytics/>

[8.] Как и зачем HR использует Big Data: технологии больших данных в управлении человеческими ресурсами. [Электронный ресурс] <https://www.bigdataschool.ru/bigdata/>

[9.] Lepak D. P., Scott A. S. Virtual HR: Strategic Human Resource Management in the 21st Century // Human Resources Management Review. 1998. Vol. 8. Iss. 3. P. 215–234.

## **FEATURES OF BIG DATA TECHNOLOGY IN HUMAN RESOURCE MANAGEMENT**

***V. V. Shatalova***

*Deputy Dean of the Faculty of  
Computer-Aided Design,  
PhD, Associate Professor*

***T. V. Kazak***

*Doctor of Psychological Sciences,  
Professor, Corresponding Member of the  
International Academy of Psychological Sciences,  
Professor, Department of Engineering Psychology  
and Ergonomics, Belarusian State University of  
Informatics and Radioelectronics*

*Belarusian State University Informatics and Radioelectronics, Republic of Belarus*

*E-mail: shatalova@bsuir.by*

**Abstract.** This article discusses the importance of using Big Data technology in innovative management strategies, technologies and methods of managing people in professional activities in a digital society, and analyzes the prospects for using this technology for management, including for the education system. The possibilities of using Big Data technologies for modern technologies and management methods in the modern market are shown.

**Keywords:** Human resource management, HR, BigData, information support, intellectual and human resources.

УДК 004.023:629.3.072.1-022.233

## УСКОРЕНИЕ ВСТРЕЧНОГО ПОИСКА КРАТЧАЙШИХ ПУТЕЙ НА БОЛЬШИХ ДИНАМИЧЕСКИХ ГРАФАХ МЕТОДОМ БУТСТРЭПИНГА



**М.П. Ревотюк**

Доцент кафедры ИТАС БГУИР,  
кандидат технических наук, доцент



**Н.В. Хаджинова**

Старший преподаватель кафедры  
ИТАС БГУИР

Белорусский государственный университет информатики и радиоэлектроники,  
Республика Беларусь  
E-mail: rmp@bsuir.by

### **М.П. Ревотюк**

Окончил Минский радиотехнический институт. Доцент кафедры информационных технологий автоматизированных систем БГУИР. Направления научных исследований: моделирование и оптимизация управления дискретными процессами в реальном времени, защита информации, системное и объектно-ориентированное программирование и проектирование.

### **Н.В. Хаджинова**

Окончила Белорусский государственный университет информатики и радиоэлектроники. Заместитель декана по учебно-методической работе факультета инфокоммуникаций, старший преподаватель кафедры информационных технологий автоматизированных систем БГУИР. Направления научных исследований: моделирование и оптимизация управления взаимодействующими процессами, объектно-ориентированное программирование и проектирование.

**Аннотация.** Предлагаются приемы ускорения встречного поиска кратчайших путей на больших динамических графах, когда порядок порождаемых деревьев путей существенно меньше порядка графа. Однократная инициализация области переменных состояния и расширение деревьев кратчайших путей методом бутстрэппинга снижает сложность поиска путей до линейной зависимости от объема сканируемого пространства.

**Ключевые слова:** транспортные сети, кратчайшие пути, вычислительная сложность.

**Постановка задачи.** Известно, что при поиске кратчайших путей на нагруженном ориентированном графе  $G(N, A)$ , где  $N$  – множество вершин,  $A$  – множество дуг с весовой функцией  $W : A \rightarrow R^+$ , время построения дерева путей одним из лучших для подобной задачи алгоритмом Дейкстры растет в первом приближении по закону  $x \cdot \log_2 x$  с увеличением расстояния  $x$  от корня дерева [1]. Реализация процедур поиска с отображением очередей анализируемых вершин на кучи Фибоначчи характеризуется сложностью  $O(m + n \cdot \log_2 n)$ , где  $m = |A|$ ,  $n = |N|$ . Отображение очереди вершин на вектор размером  $L_{\max}$  позволяет снизить сложность до величины  $O(m + n \cdot L_{\max})$ , где  $L_{\max}$  – максимальная длина дуги графа [1,2].

В случае поиска пути между двумя заданными вершинами графа целесообразно организовать процесс поиска путем построения двух встречно растущих деревьев. В результате объем анализируемых данных сокращается в два раза [3]. Дерево из конечной вершины должно строиться на графе с обратным направлением дуг, поэтому модель сети задается расширенным графом – объединением исходного графа и его инверсии.

Реализация алгоритма Дейкстры на графах с изменяемой структурой порождает проблему выбора способа представления графа и пространства состояний процесса поиска решения. Нагруженный граф  $G(N, A)$ , представляющий транспортную сеть, обычно имеет незначительную степень связности вершин, поэтому практически задается списком дуг и полностью определяется тернарным отношением  $G(x, y, w)$ , где  $x, y$  – концевые вершины дуги с весом  $w$ . В таких случаях описание графа задано легко модифицируемым списком дуг с расширяемыми множествами вершин и дуг.

Информация о возмущениях транспортной сети может быть задана списком дуг с изменившимися весами, представленным отношением  $R(x, y, w)$ , где  $x, y$  – концевые вершины дуги с весом  $w$  (значение  $w = \infty$  означает запрос удаления дуги  $x \rightarrow y$ ). При этом не обязательно выполнение условия  $x, y \in N$ , что означает отсутствие запрета на коррекцию любой части графа транспортной сети.

Очевидно, что учет возмущений транспортной сети в терминах операций над отношениями  $G(x, y, w)$  и  $R(x, y, w)$  – рутинная операция в системах баз данных.

Реляционная модель графа и пространства поиска решения – основа эффективной по памяти реализации схемы Дейкстры для поиска дерева кратчайших путей методом бутстрэппинга [4]. Идея такого метода основана на аналогии складывания ветвей дерева от корня до последнего постоянно включаемого в дерево узла, что ограничивает потребность в памяти параметрами результата поиска, а не модели исходного графа.

Предмет обсуждения – реализация метода бутстрэппинга для встречного поиска дерева кратчайших путей на графе транспортной сети. В отличие от известных постановок подобных задач, здесь будет учитываться технологичность реализации такого метода на реляционных базах данных, что обеспечит открытость модели транспортных операций для расширения с целью учета различных ограничений.

*Модель графа и пространства поиска решения.* Рассмотрим вначале форму задания графа, а затем модель переменных состояния поиска кратчайших путей. Здесь будут использованы понятия отношения и представлений (view) таблиц в терминах реляционных баз данных.

Пространство состояний поиска решения в традиционных реализациях алгоритма Дейкстры соответствует графу всей транспортной сети и включает  $D$  – массив расстояний от корня дерева,  $P$  – массив номеров предшествующих вершин,  $Q$  – очередь вершин, элементы которой упорядочены по текущему значению расстояния от корня дерева [1,2]. При этом  $|D|=|N|$ ,  $|P|=|N|$ ,  $|Q|\leq|N|$ . Реально возникающие транспортные задачи обычно не нуждаются в наличии описания всего графа сети.

Метод бутстрэппинга всегда работает с подграфом исходного графа. В случае решения задач поиска кратчайших путей на множестве вершин некоторой компактной части графа возможна несложная модификация алгоритма Дейкстры, когда поиск проводится на расширяемом подграфе с последовательно включаемыми вершинами создаваемого дерева кратчайших путей от заданной корневой вершины.

Для реализации алгоритма Дейкстры для каждой вершины  $x$  требуется представление множества непосредственно достижимых смежных вершин  $x', x' = \{k \mid w(x, k) > 0\}$ , где  $w(x, k)$  – вес дуги  $x \rightarrow k$ ,  $x, k \in N$ . На основе таблицы отношения  $G(x, y, w)$  легко построить его виртуальное представление  $G_x(x, y, w)$  в форме FSF (Forward Star Form). Индексации кортежей такой таблицы по ключу  $x$  формально определяет множество связей  $\{(x, y, w), y \in x'\}$ ,  $x \in N$ . После этого проверка условия  $x \in N$  состоит в оценке успеха поиска кортежа с ключом  $x$  в таблице  $G_x$ .

Обозначим для каждой вершины  $y$  множество входных смежных вершин  $'y$ ,  $'y = \{k \mid w(k, y) > 0\}$ , где  $k, y \in N$ . Виртуальное представление  $G_y(x, y, w)$  в форме BSF (Backward Star Form) инверсии графа образуется после индексации таблицы отношения *Graph* по ключу  $y$ . Такое представление определяет множество кортежей  $\{(x, y, w), x \in 'y\}$ ,  $y \in N$ . Успех поиска кортежа с ключом  $y$  в таблице  $G_y$  означает истинность условия  $y \in N$ .

Таким образом, имеем  $N = N_x \cup N_y$ , что означает отсутствие необходимости явного отдельного определения множества  $N$ . Операция расширения графа сети  $G(N, A)$  реализуется на логическом уровне модели данных без использования понятия адресной функции. Это влечет отсутствие проблем комплексирования существующих моделей транспортной сети из разных источников и разной степени детализации. Кроме этого, доступно на высоком описании ограничений на условия перемещения на сети в терминах функций как отношений.

С целью определения расширенного графа, пригодного для организации процесса ветвления на общей очереди, обозначим исходный граф через  $G^+ = G(N^+, A^+)$ , а граф с инвертированием направления дуг –  $G^* = G(N^*, A^*)$ . Между вершинами таких графов должно быть взаимно однозначное соответствие. Для его задания необходимо использовать симметричную функцию отображения номеров вершин  $x^+ \leftrightarrow x^*$ ,  $x^+ \in N^+$  и  $x^* \in N^*$ . Учитывая, что динамический граф транспортной сети задан списком дуг  $G(x, y, w)$ , а отображение номеров вершин  $x^+ \leftrightarrow x^*$  потребуется лишь для реализации схемы Дейкстры, предлагается в операциях с узлами дерева кратчайших путей номер вершины исходного графа дополнить признаком принадлежности к графам  $G^+$  или  $G^*$ .

Пусть  $s$  и  $t$  – начальная и конечная вершины искомого кратчайшего пути на исходном графе  $G^+$ . Так  $N^+ \cap N^* = \emptyset$ , то встречный поиск можно проводить синхронным движением волны от корней деревьев на несвязном графе  $G^+ \cup G^*$ . Для этого достаточно начать процесс ветвления из вершин  $s^+ \in N^+$  и  $t^* \in N^*$ . Последнее соответствует формальному объединению графов фиктивной дугой  $s^+ \rightarrow t^*$ , для которой  $w(s^+, t^*) = \infty$ . Синхронное движение оказывается оптимальным. Таким образом, расширенный граф  $G^+ \cup G^*$  оказывается проекцией двух виртуальных графов на список дуг  $G(x, y, w)$ .

Результат поиска кратчайшего пути от вершины  $s$  к вершине  $t$  при встречном поиске будет представлен деревьями, где узлы идентифицируются конкатенацией номеров вершин исходного графа и признаками принадлежности к виртуальным графам  $G^+$  или  $G^*$ :  $x^s = x^+s^+$ ,  $x \in N^s$ ;  $x^t = x^+t^+$ ,  $x \in N^t$ . Здесь множества вершин  $N^s \subseteq N^+$  и  $N^t \subseteq N^*$  встречных деревьев формируются рекурсивным переносом на очередном шаге ветвления буквенными символами признаков принадлежности определенному дереву (направления поиска) от корней из начального состояния:  $s^s = s^+s^+$ ;  $t^t = t^+t^+$ .

Независимо от направления поиска, дерево кратчайших путей или любое его поддереву – связный граф по определению. Если  $s$  – исходная вершина, а  $x$  – произвольный узел или лист дерева путей,  $s, x \in N$ , то текущий или кратчайший путь  $s \rightarrow x$  можно восстановить обратным движением из листа  $x$ :

$$p(x, s) = \{x, P(x), P(P(x)), \dots, s\}. \quad (1)$$

Элементы (1) упорядочены по невозрастанию значений расстояния от корня дерева путей. Альтернативы формирования дерева кратчайших путей отражаются листьями, путь от

корня до которых не обязательно кратчайший, но восстанавливается по правилу построения  $p(x, s)$ . Далее они будут ключами доступа к переменным состояниям поиска.

Пусть  $L_k$  – множество листьев текущего дерева на этапе  $k$ ,  $L_k^*$  – подмножество листьев без постоянной пометки. Очевидно, что в любой момент построения дерева кратчайших путей его узлы можно отобразить на элементы множества

$$L_k = \bigcup \{p(s, x), x \in L_k\}. \quad (2)$$

Расширение дерева кратчайших путей происходит только из некоторого листа  $x_k$  без пометки,  $x_k = \arg \min_x \{d(x), x \in L_k^*\}$ . Лист  $x_k$  превращается в узел ветвления после операции  $L_{k+1}^* = L_k^* \setminus \{x_k\}$  и до окончания поиска становятся пассивным (получает постоянную метку). Ветвление из узла  $x_k$  может привести к включению новых или коррекции в сторону уменьшения расстояния от корня  $s$  дерева существующих в  $L_{k+1}^*$  листьев из множества  $\{x_k'\}$ . Так как каждому элементу  $x, x \in L_k$ , соответствует  $d(x)$  – длина пути до корня  $s$ , то элементы множества  $L_k^* = \{x \mid d(x) \geq d(x_k), x \in L_k\}$  представляют приоритетную очередь [1-3].

Очевидно, элементы  $x \in L_k \setminus L_k^*$  остаются упорядоченными по невозрастанию значений  $d(x)$ . Таким образом, необходимо отображение ассоциаций  $x \rightarrow p(x)$ ,  $x \rightarrow d(x)$  и  $d(x) \rightarrow x$ . Для этого представим переменные состояния процесса построения дерева кортежами отношения  $T(p, q, d, o, z)$ , где  $p$  – номер предшествующего узла дерева не обязательно кратчайшего пути в узел  $q$ ;  $d$  – расстояние от корня до узла  $q$ ,  $o$  – признак направления поиска ( $o \in \{ 's', 't' \}$ );  $z$  – признак постоянной пометки узла дерева.

Состояние процесса построения деревьев встречного поиска на любом его этапе  $k$  должно быть представлено четверками  $(x, p(x), d(x), o(x)), x \in L_k^*$ . Начальное состояние процесса построения деревьев соответствует четверкам  $(s, s, 0, 's')$  и  $(t, t, 0, 't')$ . Условие завершения процесса –  $(L_k^* = \emptyset)$  или фиксация постоянной пометки вершины некоторого дерева, когда сопряженная вершина уже является постоянно помеченной [3].

После остановки в вершине  $x$  остается достроить путь до конечной вершины в исходном графе. Так как остановка может быть обнаружена в любом из встречно растущих деревьев, результат поиска необходимо получить лишь для дерева из исходной вершины.

*Представления модели поиска решения.* Список дуг графа транспортной сети может быть задан вариантами представлений отношения *Graph*, отличающимися законами упорядочения кортежей:

$G_0$  – размещение строк таблицы  $G(x, y, w)$  в памяти на физическом уровне;

$G_s$  – справочник выходных дуг вершин графа (ключ  $x$ );

$G_t$  – справочник входных дуг вершин графа (ключ  $y$ ).

Дерево кратчайших путей, согласно определению (2), в табличной форме задается вариантами представлений отношения *Trace*:

$T_0$  – размещение строк таблицы  $T(p, q, d, o, z)$  в памяти на физическом уровне;

$T_v$  – справочник узлов деревьев кратчайших путей (ключ  $q + o$ );

$T_d$  – элементы приоритетной очереди (ключ  $d$ ).

Виртуальное представление любого отношения может определяться разными способами, конкретизируемыми в среде программирования баз данных. Например, поддержка понятия структурных индексов в системах семейства FoxPro [5] обеспечивает возможность отображения нескольких законов упорядочения одной таблицы отношения:



```

proc test && Пример процедуры встречного поиска на графах
para filein && Тестовый файл в формате DIMACS
close data
set safe off
create table Graph (a C(1), x C(8), y C(8), w N(8)) && Список дуг графа сети
append from (m->filein) delimited with blank for a='a'
index on x tag s && Представление исходного графа
index on y tag t && Представление инвертированного графа
create table Trace (p C(8), q C(8), d N(12), o C(1), z C(1)) && Трасса поиска
index on q+o tag v && Представление узлов дерева кратчайших путей
index on d tag d && Представление приоритетной очереди
priv infinity
infinity=val(replicate('9',fsize("d","Trace")))
* . . .
? sptp(vertex,verty) && Поиск кратчайшего пути от vertex до verty
*
close data
retu

```

Здесь приведен фрагмент программы на языке программирования dBase для тестирования рассматриваемых далее процедур встречного поиска на примерах моделей транспортной сети, заданных в формате DIMACS[6].

*Реализация встречного поиска.* Формирование на прямом и инвертированном графах транспортной сети деревьев кратчайших путей с корнями в вершинах  $s$  и  $t$  проводится по схеме Дейкстры в представлении  $T_d$ :

```

proc sptp && Реализация встречного поиска
para s,t
insert into Trace (p,q,d,o) values (s,s,0,'s') && Исходная вершина в очереди
insert into Trace (p,q,d,o) values (t,t,0,'t') && Конечная вершина в очереди
priv f && Длина искомого кратчайшего пути
m.f=infinity
do prct && Построение деревьев кратчайших путей
retu m.f

```

Сканирование приоритетной очереди и пометка выбираемых вершин преследует цель построения дерева кратчайших путей, определяемого представлением  $T_v$ :

```

proc prct && Построение деревьев кратчайших путей
priv q,d,o,r
select Trace
set order to d && Контекст приоритетной очереди
scan all && Сканирование очереди вершин
repl z with '=' && Вершина постоянно фиксирована в дереве |
scatter memvar fields q,d,o
m.r=recno()
set order to v t && Контекст списка узлов дерева
if !iif(m.o='s',expr(),expr())&& Контроль первой встречи деревьев
do tail && Завершение поиска
exit && Кратчайший путь найден
endi
select Trace
set order to d && Контекст приоритетной очереди
go record m.r
ends
retu

```

Контроль первой встречи деревьев выполняется функциями, различаемыми принадлежностью выбираемой вершины прямому или инвертированному графу.

Ветвление в узле исходного графа обеспечивает перебор выходных дуг узла:

```
proc exps && Ветвление в узле исходного графа
  if seek(m.q+'t').and.(z='') && Сопряженная вершина постоянно помечена
  | retu .f. && Кратчайший путь найден
  endi
  select Graph
  set order to s && Контекст исходного графа
  seek m.q && Позиция узла ветвления
  scan rest while x=m.q && Сканирование выходных дуг
  do plan with Graph.y && Попытка расширения пути
  ends
  retu
```

Ветвление в узле инвертированного графа обеспечивает перебор входных дуг узла:

```
proc expt && Ветвление в узле инвертированного графа
  if seek(m.q+'s').and.(z='') && Сопряженная вершина постоянно помечена
  retu .f. && Кратчайший путь найден
  endi
  select Graph
  set order to t && Контекст инвертированного графа
  seek m.q && Позиция узла ветвления
  scan rest while y=m.q && Сканирование входных дуг
  do plan with Graph.x && Попытка расширения пути
  ends
  retu
```

Использование отдельных реализаций процедур ветвления позволяет сократить количество переключений представлений графа до одного. Для каждой дуги исходного или инвертированного графа выполняются попытки расширения пути:

```
proc plan && Попытка расширения пути
  para v && Вершина продолжения пути
  priv e
  m.e=m.d+w && Оценка расстояния до корня графа
  select Trace
  if seek(m.v+m.o) && Вершина уже в очереди
  if d>m.e
  repl p with m.q, d with m.e && Обновление пути в вершину
  endi
  else && Включение вершины в очередь
  insert into Trace (p,q,d,o) values (m.q,m.v,m.e,m.o)
  endi
  select Graph
  retu
```

Отказ от ветвления в некотором узле исходного или инвертированного графа приводит к переходу к завершению поиска. Необходимость перехода выявляется приведенными выше функциями ветвления проверкой факта постоянной пометки сопряженной вершины. Однако сумма расстояний от корней встречных деревьев в этот момент гарантированно является лишь верхней границей длины кратчайшего пути. Отклонение от оптимального значения может достигать значения  $L_{\max}/2$ .

Уточнение длины кратчайшего пути проводится посредством сканирования остатка приоритетной очереди, где остаются узлы деревьев кратчайших путей без постоянной пометки. В результате определяется узел встречи деревьев поиска, от которого сумма расстояний от корней встречных деревьев минимальна:

```

proc tail && Завершение поиска
  priv u,o
  select Trace
  m.f=m.d+d && Верхняя граница длины кратчайшего пути
  m.u=q && Узел сопряжения деревьев поиска
  set order to d
  goto record m.r
  scan rest && Сканирование остатка приоритетной очереди
  m.r=recno()
  scatter memvar fields d,q,o
  set order to v
  if seek(m.q+iif(m.o='s','t','s')).and.(m.f>m.d+d)
    m.f=m.d+d && Уточнение длины кратчайшего пути
    m.u=q && Узел сопряжения деревьев поиска
  endi
  set order to d
  goto record m.r
ends
do back && Инвертирование вершин пути к конечной вершине
retu

```

Можно заметить, что до перехода к завершению поиска бесполезная оценка длины кратчайшего пути не проводится.

Использование процедуры инвертирования вершин пути от вершины встречи деревьев к конечной вершине позволяет представить результат поиска в стандартной форме представления пути алгоритмом Дейкстры:

```

proc back && Инвертирование вершин пути к конечной вершине
  priv p,q,d,e,w
  set order to v
  seek m.u+'s' && Узел сопряжения дерева от исходной вершины
  m.e=d
  seek m.u+'t' && Узел сопряжения дерева от конечной вершины
  scatter memvar fields p,q,d
  do while (m.q!=t).and.seek(m.p+'t') && Сканирование пути к конечной вершине
    m.w=m.d-d
    scatter memvar fields p,d
    repl d with m.e+m.w, p with m.q
    m.e=d
    m.q=q
  endd
retu

```

Рассмотренный вариант ветвлений в узлах исходного и инвертированного графа допускает получение списка и параметров смежных дуг узла по мере потребности. Таким образом, для решения локальных транспортных задач нет необходимости наличия громоздкого описания полной модели динамически меняющейся транспортной сети.

*Реоптимизация результатов встречного поиска.* Изменения дуг графа транспортной сети, определяемые отношением  $R(x, y, w)$ , должны быть учтены в новой версии графа  $G(x, y, w)$ . Очевидно, что кратчайшие пути от корней  $s$  и  $t$  вне изменений сохраняются.

Отсюда следует целесообразность определения на основе отношения *Trace* значений

$$\begin{aligned}
 x_{\min} &= \arg \min_x \{d(x) \mid (x \in N_s) \wedge (x, ?, ?) \in R(x, ?, ?)\}, d_{\min} = d(x_{\min}); \\
 x_{\max} &= \arg \max_x \{d(x) \mid (x \in N_s) \wedge (x, ?, ?) \in R(x, ?, ?)\}, d_{\max} = d(x_{\max}).
 \end{aligned}$$

Далее необходимо уточнить деревья встречного поиска для узлов, которые удалены от корня  $s$  на расстояниях из интервала  $(d_{\min}, d_{\max})$ . Не составляет труда восстановить состояние

поиска по известным значениям  $x_{\min}$  и  $x_{\max}$ , а затем продолжить сканирование приоритетной очереди до остановки встречного поиска.

Такая процедура характеризуется надежностью учета любых условий ветвления и ограничений на структуру кратчайших путей, а ее вычислительная сложность в среднем в два раза лучше повторного выполнения встречного поиска. Однако она может быть заменена одной из известных процедур реоптимизации текущего дерева кратчайших путей (в случае выполнения необходимых условий их корректного применения) [7].

*Заключение.* Таким образом, процесс построения кратчайших путей на графах давно изученным алгоритмом Дейкстры представлен в форме, согласованной с возможностями современных технологий быстрого проектирования приложений. Предложенная структуризация процедуры встречного поиска непосредственно проецируется на понятия объектно-ориентированного моделирования и программирования, идиомы систем управления баз данных для создания открытых для расширения и детализации систем. Динамическое выделение области поиска решения ограничивает потребность в памяти и вычислительную сложность поиска деревьев кратчайших путей на больших графах размерностью локальных транспортных задач.

#### **Список литературы**

- [1]. Deo, N. Shortest-Path Algorithm: Taxonomy and Annotation/Deo N., Chi-yin Pang//Networks, 1984. – Vol. 14. – P. 275–323.
- [2]. Fredman, M.L. Fibonacci heaps and their uses in improved network optimization algorithms/ M.L. Fredman, R.E. Tarjan// Journal of the Association for Computing Machinery, vol. 34, 1987. – P. 596–615.
- [3]. Nicholson, T.A. J. Finding the Shortest Route between Two Points in a Network / T.A.J. Nicholson // The Computer Journal, 9(3), 1966. – P. 275–280.
- [4]. Ревотюк, М. П. Многократный поиск кратчайших путей на больших графах методом бутстрэппинга / М. П. Ревотюк, Н. В. Хаджинова // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сборник материалов V Международной научно-практической конференции, Минск, 13–14 марта 2019 г. В 2 ч. Ч. 1 / Белорусский государственный университет информатики и радиоэлектроники; редкол. : В. А. Богущ [и др.]. – Минск, 2019. – С. 301 – 310.
- [5]. Клепинин, В. Visual FoxPro в подлиннике/ В. Клепинин, Т. Агафонова //СПб.: БХВ-Петербург, 2007. – 1216 с.
- [6]. 9th DIMACS Implementation Challenge – Shortest Paths [Electronic resource]. – Access mode: <http://users.diag.uniroma1.it/challenge9/download.shtml> – Date of access: 11.01.2020.
- [7]. Shortest paths on dynamic graphs: a survey/ Ferone D. [et al]//Pesquisa Operacional. – 2017. Vol. 37, iss. 3. – P. 487-508.

## **POINT-TO-POINT SEARCHING OF THE SHORTEST PATHS ON THE BIG DYNAMIC GRAPHS BY THE BOOTSTRAPPING METHOD**

***M.P. REVOTJUK, PhD***

*Department of Information Technologies  
of Automated Systems of BSUIR,  
Associate Professor*

***N.V. KHAJNOVA***

*Department of Information Technologies  
of Automated Systems of BSUIR,  
Senior Lecturer*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus  
E-mail: rmp@bsuir.by*

**Abstract.** On the classical problem of the point-to-point searching off the shortest paths in massive dynamic graphs considered the possibility of accelerating the search procedure by incorporating a priori information about the search space. Global initialization of state variables predefined search and selection solutions can improve performance of multiple procedures to find paths to a linear dependence on the volume of the scanned area.

**Keywords:** transport networks, shortest paths in graphs, computational complexity.

УДК 612.15, 616.5, 57.087

## КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ПОВЕРХНОСТНЫХ БИОТКАНЕЙ ЧЕЛОВЕКА С МИКРОЦИРКУЛЯТОРНЫМ РУСЛОМ ПОД ВОЗДЕЙСТВИЕМ ЛАЗЕРНОГО ИЗЛУЧЕНИЯ



**Р.А. Лащётко**  
Магистрант кафедры ЭТТ  
БГУИР



**М.М. Меженная**  
Доцент кафедры ИПиЭ  
БГУИР, кандидат  
технических наук, доцент



**Т.В. Калилец**  
Аспирантка кафедры  
ИПиЭ БГУИР



**С.К. Дик**  
Доцент кафедры ЭТТ  
БГУИР, кандидат физико-  
математических наук,  
доцент



**И.В. Кишкевич**  
Магистрантка кафедры  
ИПиЭ БГУИР



**Е.Н. Рункевич**  
Магистрантка кафедры  
ИПиЭ БГУИР

Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Республика Беларусь

E-mail: [ruslanlaschetko@gmail.com](mailto:ruslanlaschetko@gmail.com), [mezhennaya@bsuir.by](mailto:mezhennaya@bsuir.by), [t.kalilets@bsuir.by](mailto:t.kalilets@bsuir.by), [sdick@bsuir.by](mailto:sdick@bsuir.by), [kishkevich.inna@gmail.com](mailto:kishkevich.inna@gmail.com), [runkevich8@gmail.com](mailto:runkevich8@gmail.com)

### **Р.А. Лащётко**

Окончил БГУИР по специальности Медицинская электроника. Обучается в магистратуре, занимается научными исследованиями в области моделирования поверхностных биотканей человека с микроциркуляторным руслом под воздействием оптического излучения.

### **М.М. Меженная**

Окончила БГУИР по специальности Медицинская электроника. Доцент кафедры инженерной психологии и эргономики БГУИР, руководит научными исследованиями в области цифровой обработки и анализа биомедицинских сигналов и изображений.

### **Калилец Т.В.**

Окончила БГУИР по специальности Электронно-оптические системы и технологии. Обучается в аспирантуре, занимается научными исследованиями в области разработки оптических методов и технических средств диагностики и лечения заболеваний кожи, является научным руководителем гранта БРФФИ Министерства образования РБ «Исследование микроциркуляции поверхностных биотканей человека методом спекл-визуализации и разработка соответствующего методического и программного обеспечения».

**Дик С.К.**

*Окончил Минский радиотехнический институт по специальности Радиотехника, руководит научными исследованиями в области лазерной медицины и биомедицинской оптики.*

**Кишкевич И.В.**

*Окончила БГУИР по специальности Медицинская электроника, является исполнителем гранта БРФФИ Министерства образования РБ «Исследование микроциркуляции поверхностных биотканей человека методом спекл-визуализации и разработка соответствующего методического и программного обеспечения».*

**Рункевич Е.Н.**

*Окончила БГУИР по специальности Медицинская электроника, является исполнителем гранта БРФФИ Министерства образования РБ «Исследование микроциркуляции поверхностных биотканей человека методом спекл-визуализации и разработка соответствующего методического и программного обеспечения».*

**Аннотация.** В статье описаны результаты моделирования поверхностных биотканей человека с микроциркуляторным руслом под воздействием лазерного излучения средствами программного пакета COMSOL Multiphysics. Результатом моделирования являются две экспериментальные модели (статическая и динамическая), применимые в исследованиях спекл-изображений при различной скорости движения эритроцитов, что позволит установить зависимость между скоростью и контрастом спеклов и использовать данную информацию после апробации модели на практике для обратной задачи – оценки скорости на основе реальных данных о контрасте.

**Ключевые слова:** микроциркуляция крови, кровотока, спекл-визуализация, спекл-изображение, моделирование поверхностного кровотока.

**Введение.** Одним из перспективных направлений в изучении системы микроциркуляции является лазерная спекл-визуализация, основанная на использовании лазерного излучения для исследования биоспеклов кожи. В рамках данного подхода анализируются параметры динамического спекл-поля, которое образуется в результате интерференции отраженного или рассеянного биообъектом когерентного излучения. Спекл-поле в плоскости наблюдения формирует картину, состоящую из множества спеклов (пятен), интенсивность света и форма которых меняются при наличии в объекте движущихся рассеивателей. При этом быстрота смены новых реализаций спекл-картины (динамика спеклов) зависит от скорости движения рассеивателей (клетки покровной ткани – дермиса, эпидермиса – и клетки крови) [1].

Целью данной работы является компьютерное моделирование поверхностных биотканей человека с микроциркуляторным руслом под воздействием лазерного излучения. Планируемое моделирование позволит решить две основные задачи: наглядное отображение процесса формирования спекл-изображения; создание модели, применимой впоследствии в исследовании влияния скорости кровотока на формирование спекл-изображения.

**Материалы и методы.** Для реализации поставленных задач в качестве основного рабочего инструмента выбран пакет для мультифизического моделирования COMSOL Multiphysics. Средствами данного пакета создана экспериментальная модель для изучения процесса формирования спекл-картины. Рабочее окно пакета COMSOL Multiphysics представлено на рисунке 1. В процессе изучения результирующие спекл-изображения обрабатывались в специально разработанном авторами программном обеспечении [2].

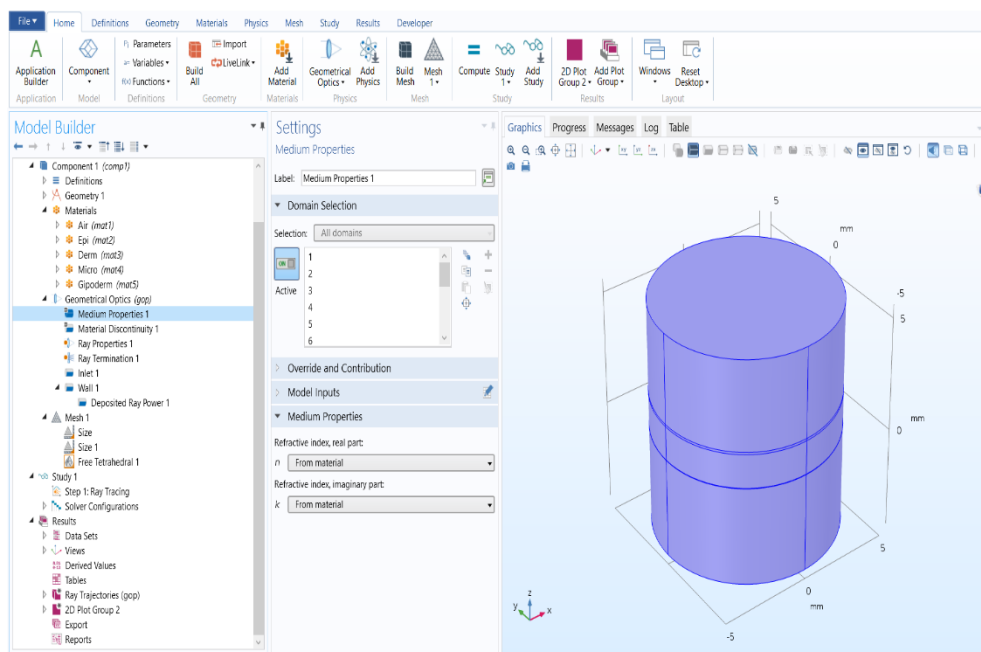


Рисунок 1. – Рабочее окно программы COMSOL Multiphysics с отображением многослойной модели воздух-эпидермис-дерма-гиподерма

В качестве теоретической основы компьютерного моделирования использована разработанная авторами ранее биофизическая модель поверхностных биотканей человека [3].

Процесс создания моделей в пакете COMSOL Multiphysics в общем случае состоит из следующих этапов:

- 1) Создать геометрию исследуемой области;
- 2) Назначить слоям геометрии материалы, которые будут отражать соответствующие физические свойства;
- 3) Подключить и настроить требуемые модули физики для эксперимента, проводимого над данной моделью;
- 4) Построить вычислительную сетку исходя из требуемой точности расчётов;
- 5) Задать параметры проведения эксперимента, такие как длительность, шаг вычисления, возможные модификации подключенных модулей физики.

Несмотря на общую схожесть модели обладают рядом значимых отличий, т.к. призваны решить разные задачи.

Для рассматриваемой задачи геометрия создавалась из примитивов цилиндрической формы. Моделируемая область имеет диаметр 10 мм, соответствующий реальному размеру освещаемой части биоткани в процессе регистрации спеклов. Размеры примитивов соответствуют размерам слоёв биотканей.

Для моделирования распространения лучей внутри модели для каждого из слоёв определены материалы со своими физическими параметрами. Для исследования распространения лазерного излучения внутри модели используется модуль физики «Геометрическая оптика», его настройка заключается в определении оптических условий среды и свойств объектов.

В настройках модуля «Геометрическая оптика» следует определить несколько ключевых параметров: показатель преломления внешней области; максимальное количество моделируемых лучей; используемую длину волны; источник излучения и его параметры; приёмник лучей, на поверхности которого будет рассчитываться накопление яркости/интерференционная картина.

В процессе построения вычислительной сетки можно воспользоваться предлагаемым шагом для используемого модуля физики либо определить его самостоятельно. Для получения более детальной картины отдельно была определена точность локального участка сетки в области приёмника отражённого излучения.

Настройка параметров проведения эксперимента заключалась в определении шагов исследования. В данном случае используется единственный шаг «Трассировка лучей», для которого был определен временной шаг для проводимых вычислений, через которые будет производиться просчёт значений ячеек вычислительной сетки.

Для модели было создано несколько примеров геометрии, с целью изучения влияния топологии на результирующую спекл-картину. Внешний вид простейшей из них представлен на рисунке 2.

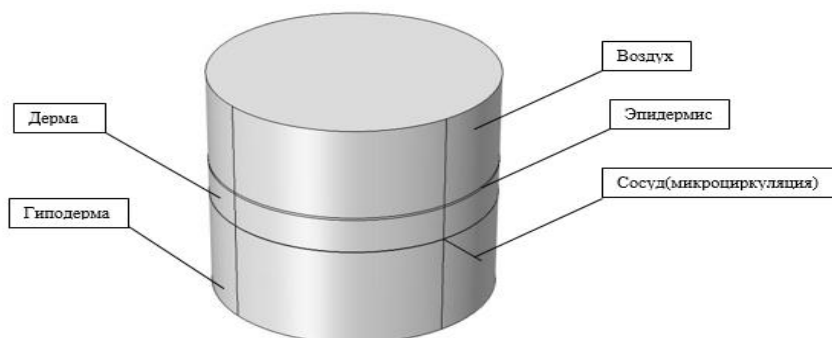


Рисунок 2. – Внешний вид геометрической модели, созданный средствами COMSOL Multiphysics

Процесс распространения лазерного излучения внутри исследуемой области изображён на рисунке 3. Красным цветом обозначены первичные лучи, а синим – вторичные, т.е. отражённые на границе слоёв с разными показателями преломления.

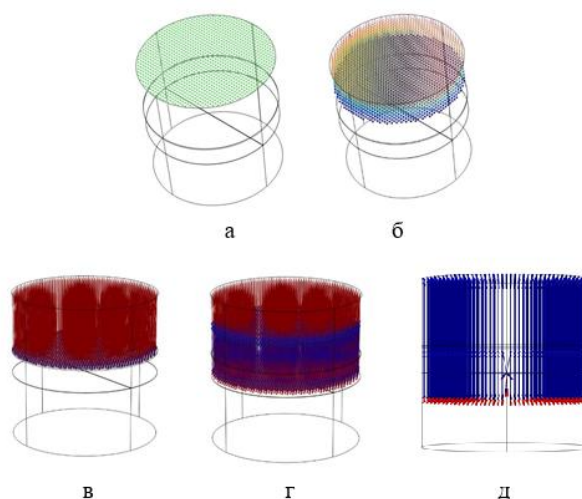


Рисунок 3. – Процесс распространения лучей плоского волнового фронта внутри исследуемого объекта (красным показаны первичные лучи, синим – вторичные/отраженные): начальный момент времени (а), процесс прохождения через воздушный слой (б), достижение биоткани (в), прохождение через биоткани (г), заключительный момент времени (д)



**Результаты.** С целью изучения поведения рассеиваемого света и зависимости образуемых спекл картин от объектов, попадающих в исследуемую область биоткани, было проведено моделирование ряда ситуаций:

1) Моделирование одиночного сосуда, полностью заполненного светотражающими частицами – эритроцитами, расположенного внутри исследуемой области (рисунок 4);

2) Моделирование идеализированной (равномерной) картины распределения светотражающих частиц на уровне микроциркуляции, расположенной внутри исследуемой области (рисунок 5);

3) Моделирование случайной схемы распределения светотражающих частиц на уровне микроциркуляции, расположенной внутри исследуемой области (рисунок 6);

4) Моделирование топологии сосудов произвольной формы (рисунок 7).

Для большей наглядности моделирование проводилось с увеличенными в 10 раз размерами геометрии.

На рисунке 4 видно, что наибольшая интенсивность отражённого излучения приходится на область с высокой концентрацией светотражающих элементов (или содержащих их сосудов).

При моделировании участка микроциркуляции результирующая спекл-картина способна отразить топологию сосудистой сети. Пример спекл-изображения для случайного участка микроциркуляции представлен на рисунке 7. Используя подобное моделирование можно проследить отдельные факторы, влияющие на образование спекл-картины.

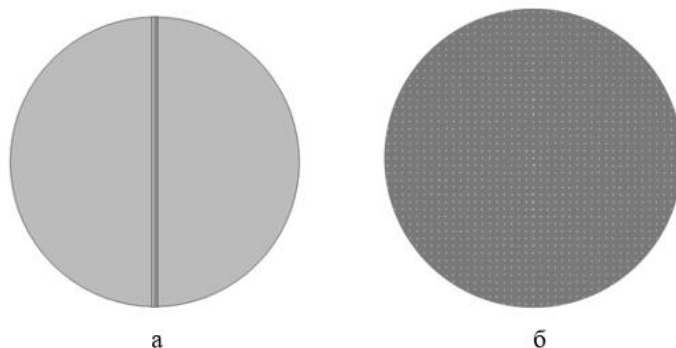


Рисунок 4. – Исследуемый участок модели, имитирующий одиночный сосуд (а), и результирующая спекл-картина (б)

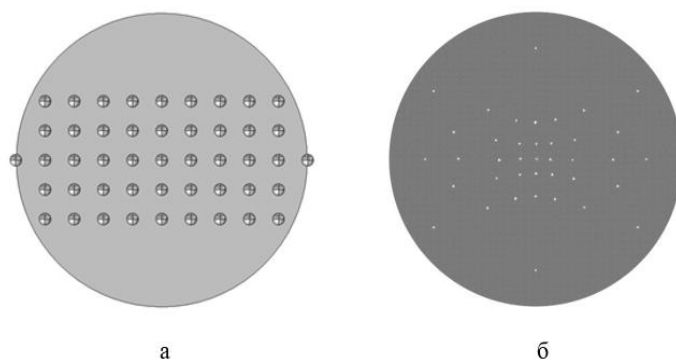


Рисунок 5. – Исследуемый участок модели, имитирующий идеализированную картину распределения эритроцитов в анализируемой области (а), и результирующая спекл-картина (б)

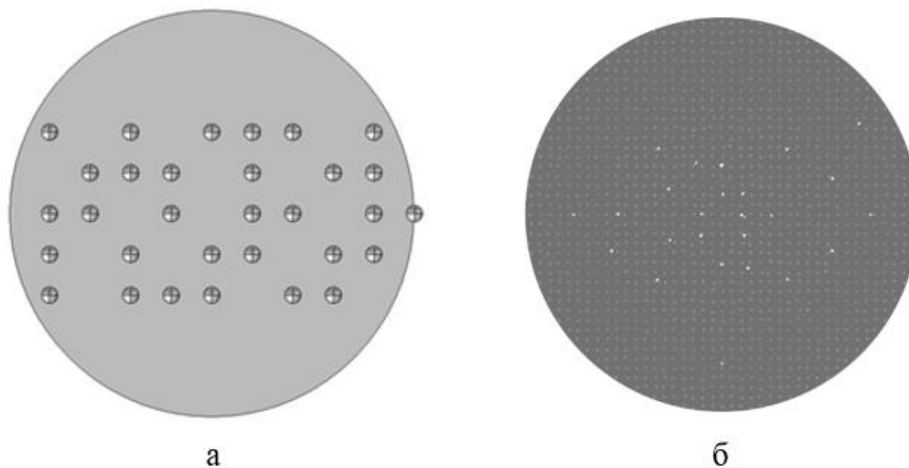


Рисунок 6. – Исследуемый участок модели, имитирующий случайную картину распределения эритроцитов в анализируемой области (а), и результирующая спекл-картина (б)

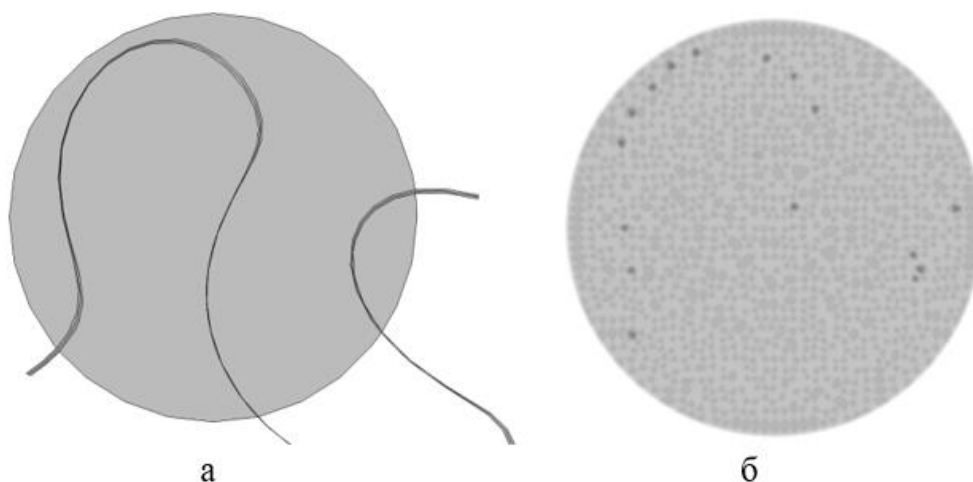


Рисунок 7. – Исследуемый участок модели, имитирующий случайный участок микроциркуляции в анализируемой области (а), и результирующая спекл-картина (б)

Для адаптации описанной модели к исследованию динамических процессов исходная (статическая) модель была модернизирована. Ключевым отличием новой модели является наличие дополнительного модуля физики, позволяющего задать скорость движения для части геометрии.

Геометрия подверглась переработке, чтобы соответствовать динамической структуре. Слой микроциркуляции, представлявший собой набор светоотражающих частиц, был отделён от основной части модели, также ему была присвоено свойство «линейной скорости». Теперь в процессе выполнения шагов исследования проводится имитация процесса протекания крови по заданному направлению.

Свойства подключённого модуля «Геометрическая оптика» подверглись изменениям. Основное отличие от предыдущего исследования заключается в моделировании непрерывного лазерного излучения, в то время как для получения одиночной спекл-картины достаточно было одной итерации испускания лазерного пучка. Свойства самого лазера остались прежними.

Настройка шагов исследования была преобразована к виду, отражающему процесс работы высокочувствительной камеры: получаемые изображения рассчитывались с частотой 120 кадров в секунду, при этом моделирование движения не прекращалось.

Так же, как и при исследовании процесса образования спекл-изображений, для большей наглядности получаемых данных было сохранено увеличение геометрических размеров в 10 раз.

Для испытания данного подхода было проведено моделирование процесса прохождения группы светотражающих частиц через центр исследуемого участка. Результирующие видеофайлы были обработаны в программном обеспечении авторов для расчёта спекл-контраста [2]. Начальная и конечная позиции группы светотражающих частиц, а также результат обработки полученных данных, представлены на рисунках 8 и 9 соответственно.

Как видно из представленного на рисунке 9 изображения на границах областей прохождения светотражающих частиц образуется неоднородность, выражающиеся в высоких значениях спекл-контраста. Данный результат подтверждает перспективность применения компьютерного моделирования в исследовании метода спекл-визуализации.

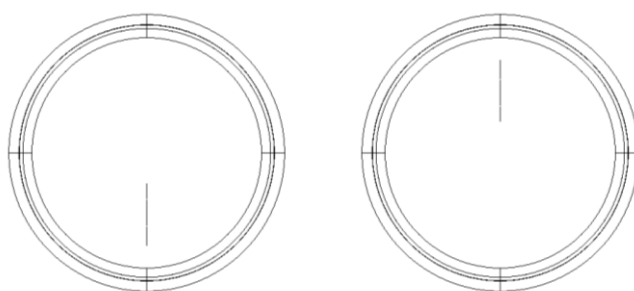


Рисунок 8. – Начальная и конечная позиции моделирования процесса прохождения группы светотражающих частиц – эритроцитов – через центр исследуемого участка

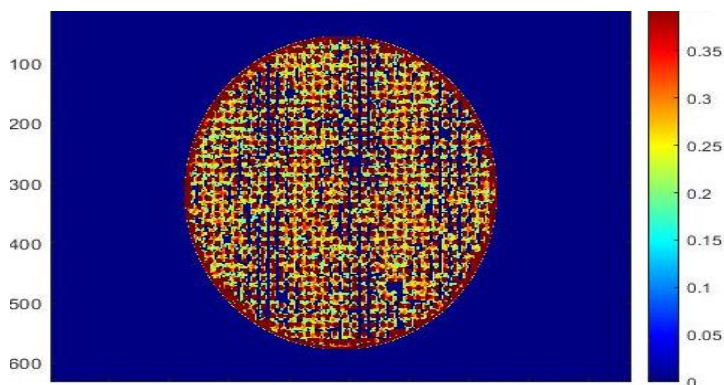


Рисунок 9. – Результаты обработки полученных в примере из рисунка 8 данных

**Заключение.** Полученная динамическая модель позволяет получать видеофайлы, содержащие в себе процесс накопления отражённого излучения при условии изменяющейся геометрии (движения светотражающих частиц внутри исследуемого участка). К недостаткам данной модели следует отнести высокие требования к качеству модели, а также необходимость её преобразования для каждого нового образца сосудистой топологии. Результаты данного моделирования могут быть обработаны специальным программным обеспечением для расчёта спекл-контраста.

На основе полученных результатов планируется провести моделирование образования спекл-картины при различной скорости движения эритроцитов, что позволит установить зависимость между скоростью и контрастом спеклов и использовать данную информацию после апробации модели на практике для обратной задачи – оценки скорости на основе реальных данных о контрасте.

#### **Список литературы**

[1].Виленский М.А., Агафонов Д.Н., Зимняков Д.А., Тучин В.В., Задражевский Р.А. Спекл-корреляционный анализ микрокапиллярного кровотока ногтевого ложа / Квантовая электроника, Т.41, №4 (2011) – С.324-328.

[2].Меженная, М.М. Расчет контраста спекл-изображений: методическое обеспечение и программная реализация / М.М. Меженная, Е.Н. Рункевич, И.В. Кишкевич, Т.В. Калилец, С.К.Дик Д.В. Лихачевский, Р.А. Лащётко // Научный журнал «Доклады БГУИР» / редкол.: В.А. Богущ [и др.]. Мн.: БГУИР, №7 (117), 2018. Стр. 139-143.

[3].Лащётко, Р. А. Биофизическая модель поверхностных тканей человека с микроциркуляторным руслом / Р. А. Лащётко // Электронные системы и технологии: сборник тезисов 55 научной конференции аспирантов, магистрантов и студентов / Минск : БГУИР, 2019. – С. 334.

### **MODELING OF HUMAN SURFACE TISSUES WITH MICROVASCULATURE UNDER THE INFLUENCE OF OPTICAL RADIATION**

**R.A. Lashchetko**  
*Master student BSUIR*

**M.M. Mezhennaya**  
*Ph.D., Associate Professor of BSUIR*

**T.V. Kalilec**  
*Postgraduate Student of BSUIR*

**S.K. Dzik**  
*Ph.D., Associate Professor of BSUIR*

**I.V. Kishkevich**  
*Researcher of BSUIR*

**K.N. Runkevich**  
*Researcher of BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus  
E-mail: ruslanlaschetko@gmail.com*

**Abstract.** The article describes the results of modeling human surface tissues with a microvasculature under the influence of optical radiation using the COMSOL Multiphysics software package. The simulation result is two experimental models applicable in the study of speckle images.

**Keywords:** blood microcirculation, blood flow, speckle visualization, speckle image, contrast, modeling of human surface tissues with microvasculature.

УДК [621.39+ 004.77]

## ГИГАБИТНЫЕ ПАССИВНЫЕ ОПТИЧЕСКИЕ СЕТИ С РАСШИРЕННОЙ ЗОНОЙ ДОСЯГАЕМОСТИ



**А.Д. Тусупов**  
PhD докторант  
специальности  
Информационные  
системы, Евразийского  
национального  
университета им.  
Л.Н.Гумилева



**Н.И. Листонад**  
Доктор технических наук,  
профессор, заведующий кафедрой  
Информационных  
радиотехнологий Белорусского  
Государственного Университета  
Информатики и  
Радиоэлектроники



**А.Т. Тохметов,**  
кандидат физико-  
математических наук,  
доцент кафедры  
Информационных систем  
Евразийского  
национального  
университета имени  
Л.Н.Гумилева

Кафедра Информационных систем Евразийского национального университета им. Л.Н.Гумилева  
Евразийский национальный университет им. Л.Н.Гумилева, г.Нур-Султан, Республика Казахстан  
E-mail: tussupov@gmail.com

### **А.Д. Тусупов**

Выпускник Технологического института Карлсруэ (г.Карлсруэ, Германия) по специальности информационная и коммуникационная инженерия. Имеет академическую степень магистра наук, PhD докторант специальности Информационные системы, Евразийского национального университета им. Л.Н.Гумилева.

### **Н.И. Листонад**

Доктор технических наук, профессор, заведующий кафедрой информационных радиотехнологий Белорусского государственного университета Информатики и Радиоэлектроники (с 2016 года по настоящее время), руководитель Офиса программы Европейской Комиссии Темпус (в настоящее время - программы - Erasmus+). Защитил диссертацию на соискание ученой степени доктора технических наук по специальности «Телекоммуникационные системы и компьютерные сети» (2001 г.). Обладатель почетных грамот и почетного знака «Отличник образования Республики Беларусь».

### **А.Т. Тохметов**

Кандидат физико-математических наук, доцент кафедры информационных систем Евразийского национального университета им. Л.Н.Гумилева (с 2016 года по настоящее время). Окончил аспирантуру в Ленинградском физико-техническом институте имени А.Ф.Иоффе АН СССР – диплом кандидата физико-математических наук по специальности «01.04.07–Физика твердого тела» (1989 г.), доцент по специальности «Информатика и информационные системы» (2002 г.). Обладатель почетных грамот Министерства образования и науки Республики Казахстан.

**Аннотация.** Гигабитные пассивные оптические сети (GPON) являются наиболее передовой технологией. Скорость передачи данных составляет 2,5 Гбит/с для прямого и обратного потоков. Но данная архитектура сети имеет ограниченную физическую длину сети в 20 км. Это связано с высокими бюджетными потерями сети. Такое ограничение доступа делает сеть труднодоступной для абонентов, расположенных далеко от объектов телеком-оператора и покрытие удаленных населенных пунктов является достаточно инвестиционно-ёмким, тем самым усложняет устранение цифрового неравенства между городом, пригородом и селом. Для

решения данной проблемы предлагается использование оптических усилителей, которые расширят зону досягаемости GPON до 60 км, что является пределом для логической длины по текущим протоколам.

**Ключевые слова:** GPON, последняя миля, оптический усилитель, гибридная WDM/TDM GPON.

*Введение.* В наши дни в целях обеспечения качественной цифровой связью, а также предоставления доступа к инновационным цифровым услугам, таких как электронное правительство, цифровое телевидение, облачные решения, и развитие умных городов, образовательных, медицинских и государственных учреждений должна быть возможность к подключению высокоскоростного интернета. К сожалению, телекоммуникационные инфраструктуры и широкополосной доступ к сети Интернет труднодоступны для абонентов, расположенных далеко от объектов телеком-оператора. Покрытие удаленных населенных пунктов является достаточно инвестиционно-ёмким, тем самым усложняя устранение цифрового неравенства между городом, пригородом и селом. «Волокно до дома» FTTH (Fiber-to-the-Home) – это понятие описывает общий подход к организации кабельной инфраструктуры сети доступа, при которой от узла связи до абонента используется оптическое волокно. Это позволяет заменить существующую медно-кабельную инфраструктуру сетей абонентского доступа на оптическое волокно, с высокой пропускной способностью, многократно превышающую пропускную способность всех других систем связи и может измеряться терабитами в секунду. Одной из перспективных технологий FTTH для систем оптоволоконного доступа являются пассивные оптические сети PON (Passive Optical Network). Пассивные оптические сети (PON) могут решить проблему последней мили, в результате чего информация дойдет до пользователя со скоростью передачи данных в гигабит в секунду.

*1 Архитектура и стандарты пассивной оптической сети.* Архитектура PON, изображенная на рис. 1, состоит из одного приемопередающего модуля в OLT (Optical Line Terminal), расположенного в центральном узле связи, пассивного оптического разветвителя и множества абонентских устройств ONU (Optical Network Unit) или ONT (Optical Network Terminal). ONU используется в терминологии IEEE (института инженеров по электротехнике и радиоэлектронике), а ONT в терминологиях ITU-T (международного союза электросвязи), Технология PON достигает физической длины до 20 километров от узла связи до абонента и к оптическому сплиттеру (разветвителю) можно подключить от 32-128 абонентских узлов (теоретическое число абонентских узлов может быть настолько большим, насколько позволяет бюджет мощности и максимальная скорость приемопередающей аппаратуры). Передача потока информации от центрального узла OLT к абонентскому узлу ONU называется прямым (нисходящим) потоком (ПП), а потоки данных от разных абонентских узлов ONU в центральный узел OLT, называется обратным (восходящим) потоком (ОП).

PON состоит из трех основных стандартов: Broadband PON (BPON), Ethernet PON (EPON) и Gigabit PON (GPON). Эти стандарты имеют различные характеристики и их свойства перечислены в Табл. 1. Технология EPON является стандартом IEEE 802ah [1] и использует протокол Ethernet. Скорость передачи данных по прямым и обратным потокам составляют 1000 Мбит/сек. Физическая длина сети 10 км с соотношением разветвления на 1:16 или 1:32. Технология BPON является стандартом ITU-T G.983 [2] и использует протокол асинхронной передачи данных ATM (Asynchronous Transfer Mode). Скорость передачи данных по прямым потокам составляет 622 Мбит/сек и по обратным потокам 155 Мбит/сек. Физическая длина сети 20 км с соотношением разветвления 1:32. Технология GPON является стандартом ITU-T G.984 [3] и использует протоколы Ethernet или мультиплексирование с разделением времени TDM (Time Division Multiplexing). Скорость передачи данных по прямым потокам составляет 2488 Мбит/сек и по обратным потокам 1244 Мбит/сек. Физическая длина сети 20 км, с соотношением разветвления на 1:32 или 1:64.

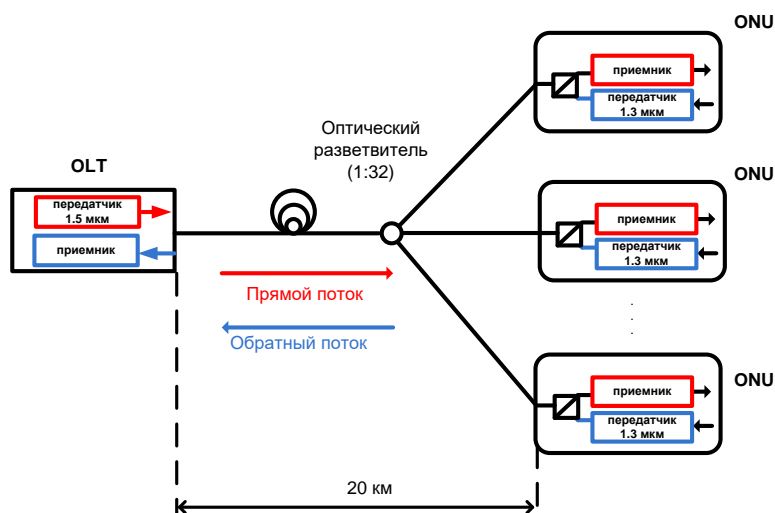


Рисунок 1. – Архитектура пассивной оптической сети PON.

Таблица 1. – Основные характеристики и свойства PON [4]

Характеристика	EPON	BPON	GPON
Стандарт	IEEE 802.3ah	ITU-T G.983	ITU-T G.984
Протокол	Ethernet	ATM	Ethernet, TDM
Скорость передачи, (Мбит/сек)	1000, ПП и ОП	622 ПП, 155 ОП	2488 ПП, 1244 ОП
Длина сети (км)	10	20	20
Соотношение разветвления	16 или 32	32	32 или 64

Сравнивая вышеперечисленные стандарты пассивно оптических сетей, мы определяем, что GPON является самой передовой технологией, поскольку данная технология предоставляет высокую скорость передачи данных 2,5 Гбит/сек и 1,2 Гбит/сек для прямого и обратного потоков соответственно. GPON поддерживает различные протоколы, возможности увеличения оптического соотношения разветвления и является последним стандартом для пассивных оптических сетей.

По рекомендациям ITU-T G.984, GPON поддерживает оптический бюджет мощности в 28 дБ, что практически ограничивается фактором соотношения разветвления 1:32 и длины сети не более 20 км. Если ограничения на физическом уровне могут быть преодолены, тогда можно поддерживать логическую длину сети до 60 км и логическое разветвление до 128 абонентских узлов.

*2 Гибридная WDM/TDM GPON с расширенной зоной досягаемости.* Гибридная пассивная оптическая сеть TDM/WDM является эффективным решением для оптических сетей доступа. Этот метод доступа сочетает преимущества технологий временного мультиплексирования (TDM) и спектрального уплотнения (WDM), что позволяет повышения спектральной эффективности сети. Это же сочетание позволяет передавать данные, уплотненные во времени, по каждому спектрально уплотненным каналом, что приводит к

увеличению количество абонентов.

Как уже отмечалось выше, архитектура технологий GPON из-за бюджетных потерь в сети в 28 дБ имеет ограниченную физическую длину 20 км. Данное ограничение длины сети обуславливается трудностью обслуживания абонентов, расположенных далеко от центрального узла связи. Но бюджет мощность сети может быть увеличен путем добавления в сеть оптических усилителей. В таком случае, логическая длина сети GPON может достичь длины более 60 км. Рис. 2 иллюстрирует расширенную зону досягаемости гибридной сети WDM/TDM GPON с использованием встроенного оптического усилителя до логической длины, превышающей 60 км.

Расширения зоны досягаемости гибридной WDM/TDM GPON сети имеет свои преимущества:

- 1) Увеличенная длина сети:
  - а) объединяет тысячи центральных узлов связи в несколько городских узлов связи;
  - б) достигает абонентов, которые расположены далеко от центрального узла связи и неравномерно распределены по географическим зонам;
  - в) инфраструктура, сокращенных местных и региональных центральных узлов связи, может быть предназначена для установки оптических усилителей;
  - г) значительная экономия капитальных и эксплуатационных расходов [5];
- 2) В гибридной WDM/TDM GPON архитектуре количество пользователей может превышать более 1000 пользователей на OLT порт [6];
- 3) Высокая скорость прямого и обратного потоков трафика и имеется возможность легкого усовершенствования до 40 Гбит/сек;
- 4) Высокая пропускная способность для каждого пользователя;
- 5) Единый интерфейс для оптоволоконна;
- 6) Простая масштабируемость и возможность модернизации;
- 7) Мульти-эксплуатация (оптоволоконная инфраструктура, эксплуатируемая несколькими операторами);
- 8) Централизованное управление и мониторинг;
- 9) Отказоустойчивость и балансировки трафика.

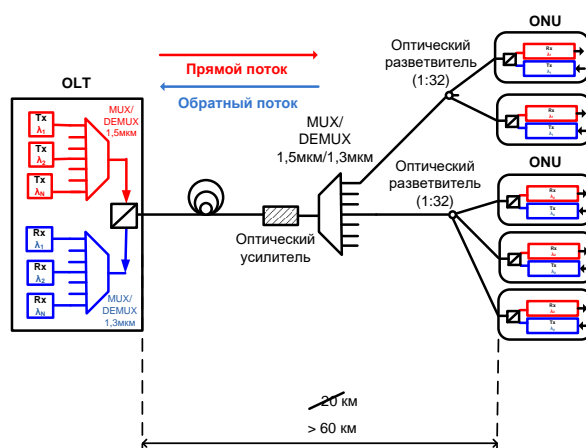


Рисунок 2. – Гибридная WDM/TDM GPON с использованием встроенного оптического усилителя с расширенной зоной досягаемости, с логической длиной выше чем 60 км

3 Требования к оптическим усилителям для GPON с расширенной зоной досягаемости. Для расширения зоны досягаемости сети WDM/TDM GPON требуются использование оптических усилителей. Существуют различные типы оптических усилителей, которые



показывают высокий потенциал в качестве линейных усилителей:

- Оптоволоконные усилители легированные редкоземельными элементами такие, как: эрбий (Er); неодим (Nd); иттербий (Yb); тулий (Tm); празеодим (Pr);
- Полупроводниковые оптические усилители (ППОУ): объемные ППОУ и ППОУ с наноразмерной гетероструктурой такие, как ППОУ на квантовых ямах (Quantum Well - QW), ППОУ на квантовых точках (Quantum Dots - QD);
- Рамановский усилитель.

Чтобы выбрать лучшие оптические усилители для расширения зоны досягаемости сети eGPON, в первую очередь, усилители должны выполнять требования PON и соответствовать стандартам GPON ITU-T. Усилители должны быть дешевые по цене, поскольку стоимость оптического усилителя делиться между всеми абонентов сети PON. А также оптические усилители должны быть приспособлены к приложениям используемые в будущем.

Требования к оптическим усилителям для прямого потока (Таблица 2):

- Функционирование в диапазоне длин волн 1,5 мкм;
- Передача данных со скоростью выше 2,5 Гбит / с;
- Усиление всех спектрально уплотненных каналов на одинаковом уровне и с низкими межканальными перекрестными помехами.

Для передачи сигнала по обратному потоку требования гораздо выше. Потери перед усилителем намного выше, чем по прямому потоку в основном из-за оптического разветвления. Также абоненты сети географически распределены так, что расстояние до усилителя сильно варьируется.

Требования к оптическим усилителям для обратного потока (Таблица 2):

- Функционирование в диапазоне длин волн 1,3 мкм;
- Высокий коэффициент усиления для компенсации больших потерь;
- Высокий динамический диапазон для входных мощностей (ДДВМ) необходимые для покрытия вариаций уровней мощностей от расстояния абонентов;
- Высокая устойчивость для режима пульсирующего трафика, которая является необходимым условием для связи с разным уровнем мощности сигнала.

Основные требования к усилителям для гибридной WDM/TDM GPON с расширенной зоной досягаемости приведены в Таблице 2 в сравнительной форме.

Таблица 2. – Требования к расширителям досягаемости (усилителям) GPON, разделенные для прямого и обратного потока

Прямой поток	Обратный поток
Скорость передачи данных: сегодня - 2.5 Гбит/сек; в будущем – выше 40 Гбит/сек	Высокий динамический диапазон для входных мощностей (ДДВМ)
Технологии CWDM/DWDM	Высокая устойчивость для режима пульсирующего трафика
Плоский коэффициент усиления спектрально уплотненных каналов	Высокий коэффициент усиления
	Низкие межканальные перекрестные помехи

4 Проекты по GPON с расширенной зоной досягаемости. В настоящее время ведутся активные разработки для расширение зоны досягаемости пассивных оптических сетей. Рассмотрим основные исследования в этой области, исследующие эффективность различных технологий оптических усилителей.

#### 4.1 Оптические усилители EDFA (Проект PIEMAN)

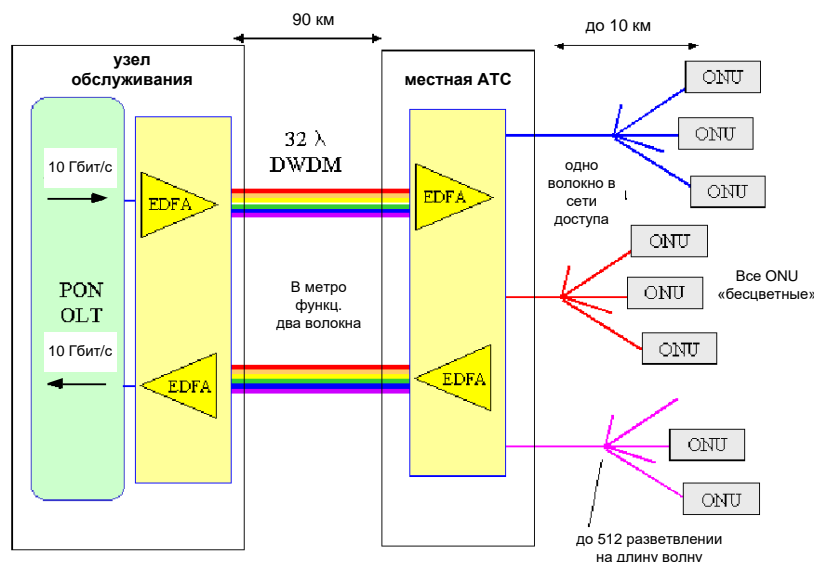


Рисунок 3. – Архитектура сети проекта PIEMAN [7]

Проект будущего широкополосного оптического доступа PIEMAN (Photonic Integrated Extended Metro and Access Network) является проектом Европейского Союза (ЕС) с общей стоимостью в € 3,9 млн. [7].

Проект PIEMAN выполняет исследования на физическом уровне для нового поколения PON с использованием спектрального уплотнения DWDM на 32 канала, где каждый канал имеет пропускную способность в 10 Гбит/с прямого и обратного потоков до каждого абонента. Каждый канал на определенной длине волны разветвляется на 512 абонентов. Проект имеет гибридную WDM/TDM структуру длиной в 100 км и использует усилители EDFA для прямого и обратного потоков. Архитектура сети проекта PIEMAN показана на рисунке 3.

Обратный поток в проекте PIEMAN не соответствует требованиям PON ITU-T. Волоконные оптические усилители, особенно для диапазона 1,3 мкм, являются очень дорогими и неразвитыми технологиями. Обратный поток находится в режиме пульсирующего трафика и требует принятия дополнительных мер к использованию волоконных усилителей. Из-за неполного подавления шумов в диапазоне возникают перекрестные помехи и, поэтому могут накладываться существенные ограничения на фактор разветвления [8].

4.2 Рамановский усилитель. Iannone и др. [9] экспериментально использовали четыре общепринятых TDM PON приема-передатчика с симметричными трафиками в 2,5 Гбит/с с использованием гибридного ППОУ-рамановского усилителя, для достижения расширенной зоной досягаемости в 60 км с общей инфраструктурой для обслуживания 128 абонентов.

Экспериментальная установка показана на рисунке 4.

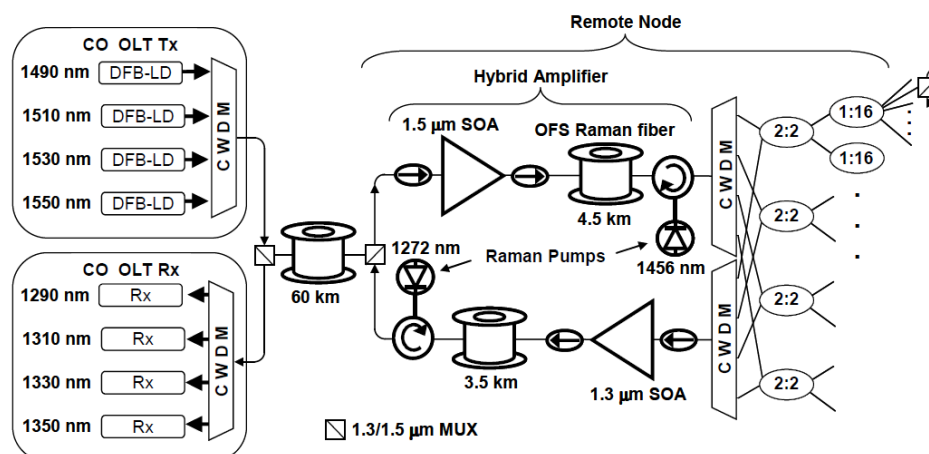


Рисунок 4. – Экспериментальная установка расширенной PON с гибридным ППОУ-рамановским усилителем

Гибридный ППОУ-рамановский усилитель обеспечивает широкую ширину спектра усиления в 150 нм, но эта технология очень дорога по цене. Рамановский усилитель неэффективен (требует мощный лазер накачки) и показывает низкий ДДВМ для обратного потока.

4.3 Полупроводниковый оптический усилитель. Nasset и др. [10] разработали ППОУ на длине волны 1,3 мкм для достижения расширение GPON. ППОУ показывает высокий коэффициент усиления (30 дБ) и умеренный коэффициент шума (7 дБ). Они экспериментально продемонстрировали GPON с расширенной зоной досягаемости с одним каналом для прямого и обратного потоков с оптическим разветвлением 1:32 и длиной сети в 60 км. Разработанный ППОУ имеет медленный эффект усиления, которые ограничивают ДДВМ.

4.4 Полупроводниковые оптические усилители на квантовых точках. Мы [11] экспериментально продемонстрировали расширение зоны досягаемости сети WDM/TDM GPON до 60 км с использованием полупроводникового оптического усилителя на квантовых точках. Все полученные сигналы не превышали ошибочный порог.

ППОУ с активной областью на квантовых точках (ППОУ-КТ) имеют локализацию носителей в нульмерном пространстве. КТ реализуются как когерентно напряженные, бездефектные островки при самоорганизованном эпитаксиальном выращивании гетероструктур несогласованных полупроводников [12]. В принципе, в ППОУ-КТ временная характеристика может быть 1000 раз быстрее, чем у ППОУ-КЯ. Это делает усилитель перспективным для использования в технологиях со скоростью 1 Тбит/сек [12]. Среда усиления в КТ показывает уникальные свойства, как сверхбыстрое восстановление усиления в порядке ( $\sim 1$  пс) [13], значительно широкую спектральную ширину усиления ( $\sim 120$  нм) [14], высокий коэффициент усиления ( $> 25$  дБ) [15], низкий коэффициент шума ( $> 5$  дБ) [16], возможность для неохлажденной работы [17], высокий динамический диапазон для входных мощностей (ДДВМ), высокая устойчивость для пульсирующего трафика [18], и поляризационную нечувствительность [19].

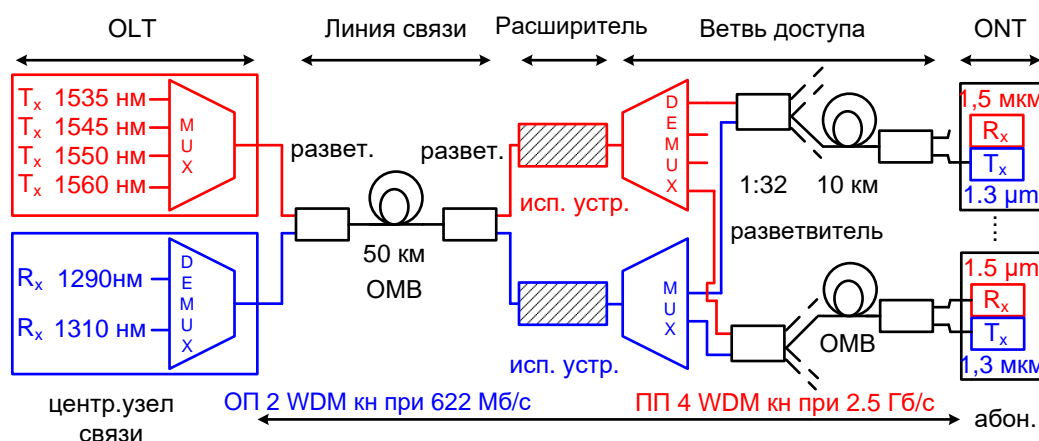


Рисунок 5. – Испытательный стенд расширенной WDM/TDM GPON с четырьмя каналами прямого потока и с двумя каналами обратного потока, где каждый канал обслуживает 32 абонента и с длиной сети в 60 км используя оптические усилители для двунаправленного усиления

Установка (рисунок 5) состоит из центрального узла с приемопередающим модулем OLT, 50 км оптического волокна (SMF-28), расширительной коробки для усилителей с испытываемым устройством, устройства мультиплектора/демультиплектора, пассивного оптического разветвителя 1:32 и абонентского узла ONT. OLT в центральном узле оборудован с четырьмя WDM каналами передатчиками (1535 нм, 1545 нм, 1550 нм, 1560 нм) для прямого потока со скоростью данных в 2,5 Гбит/с, и двумя WDM приемниками (1290 нм и 1310 нм). Четыре абонентских узлов ONT расположены на дистанции 10 км от расширительной коробки. Абонентские узлы ONT содержат четыре приемника, но из-за недостатка оборудования только два из четырех ONT оборудованы передатчиками, где генерируют поток для обратной связи со скоростью в 622 Мбит/с. Выпускаемая мощность передатчиков OLT и ONU составляет +5 дБм. Приемники идентичны с теми, которые использовались для измерения удлинение бюджета. Трафик в настоящем эксперименте является непрерывный. Бюджет потерь для 1,5 мкм прямого потока составляет 45 дБ и 48 дБ для 1,3 мкм обратного потока. Проблема поляризации ППОУ-КТ были учтены с помощью поляризационных контроллеров, настроенные на максимальную выходную мощность.

Отсутствие ошибочных порогов и перекрестных помех между каналами прямых и обратных потоков наблюдаются даже тогда, когда система работает со всеми каналами одновременно.

### **Список литературы**

- [1] Стандарт Института инженеров по электротехнике и радиоэлектронике IEEE Ethernet First Mile Standards 802.3ah-2004, included in 802.3-2005
- [2] Рекомендации международного союза электросвязи ITU-T Recommendations G.983
- [3] Рекомендации международного союза электросвязи ITU-T Recommendations G.984
- [4] Белая книга: Full service broadband with GPON, 2008
- [5] Davey R., Kani J., Bourgart F. and Mccammon K. Options for future optical access networks // IEEE Commun. Mag. 44 – 2006 — Vol. 6 – P. 50–56
- [6] Lázaro J.A., Prat J., Chanclou P., Tosi Beleffi G.M., Teixeira A., Tomkos I., Soila R. and Koratzinos V Scalable Extended Reach PON // OFC/NFOEC – Feb. 2008 – paper OThL2
- [7] Photonic Integrated Extended Metro and Access Network URL: <http://www.ist-pieman.org/>
- [8] Baekelandt B., Mélange C., Ossieur P., Bauwelinck J., De Ridder T., Qiu X., Vandewege J., Smith D., Davey R. Impact of crosstalk on multi-wavelength high split PON networks // BBEurope '07 – 2007 – Antwerp, Belgium.
- [9] Iannone P.P., Lee H.H., Reichmann K.C., Zhou X., Du M., Pálsdóttir B., Feder K., Westbrook P, Brar K., Mann J. and Spiekman L. Hybrid CWDM Amplifier Shared by Multiple TDM PONs // National Fiber Optic Engineers Conference (NFOEC) – 2007 – paper PDP13
- [10] Nasset D., Kelly T., Appathurai S. and Davey R. Extended Reach GPON Using High Gain Semiconductor Optical Amplifiers // 33rd European Conference and Exhibition of Optical Communication OFC/NFOEC – 2008
- [11] Bonk R., Brenot R., Meuer C., Vallaitis T., Tussupov A., Rode J. C., Sygletos S., Vorreau P., Lelarge F., Duan G. H., Krimmel H.-G., Pfeiffer Th., Bimberg D., Freude W., Leuthold J. 1.3/ 1.5  $\mu\text{m}$  QD-SOAs for WDM/TDM GPON with Extended Reach and Large Upstream/Downstream Dynamic Range // Proc. Opt. Fiber Communication Conf. (OFC'09), San Diego, USA; OWQ1, accepted for publication
- [12] Bhattacharya P., Bimberg D., Arakawa Y. Special Issue on Optoelectronic Devices Based on Quantum Dots // IEEE – Sep. 2007 – vol. 95 – No. 9, P.1718-1722.
- [13] Vallaitis T., Koos C., Bonk R., Freude W., Laemmlin M., Meuer C., Bimberg D., Leuthold J. Slow and fast dynamics of gain and phase in a quantum dot semiconductor optical amplifier // Opt. Express – Jan. 2008 – Vol. 16 – No. 1 – P. 170-178.
- [14] Brenot R., Lelarge F., Legouezigou O., Pommereau F., Poingt F., Legouezigou L., Derouin E., Drisse O., Rousseau B., Martin F., Duan G. H. Quantum Dots Semiconductor Optical Amplifier with a -3dB Bandwidth of up to 120 nm in Semi-Cooled Operation // The Proceedings of the Optical Fiber Communication Conference – 24-28 February 2008 – San Diego, CA, USA –paper OTuC1.
- [15] Akiyama T., Sugawara M., Arakawa Y. Quantum-Dot Semiconductor Optical Amplifiers // IEEE – Sep. 2007 – vol. 95 – No. 9 – P.1757-1766.
- [16] Bimberg, D. Quantum dot based nanophotonics and nanoelectronics // Electr. Letters – 2008 – vol.44 – Issue 3 – P.168.
- [17] Wang H., Aw E.T., Xia M., Thompson M.G., Penty R.V., White I.H. Temperature Independent Optical Amplification in Uncooled Quantum Dot Optical Amplifiers // OFC, OSA Technical Digest (CD) –24-28 February 2008 – San Diego, CA, USA – paper OTuC2
- [18] Bonk R., Meuer C., Vallaitis T., Sygletos S., Vorreau P., Ben-Ezra S., Tsadka S., Kovsh A. R., Krestnikov I.L., Laemmlin M., Bimberg D., Freude W., Leuthold J. Single and Multiple Channel Operation Dynamics of Linear Quantum-Dot Semiconductor Optical Amplifier // ECOC'08 – Sept. 2008 – Brüssel – paper Th1.C2,
- [19] Yasuoka N., Kawaguchi K., Ebe H., Akiyama T., Ekawa M., Morito K., Sugawara M., Arakawa Y. Quantum-Dot Semiconductor Optical Amplifiers With Polarization-Independent Gains in 1.5 $\mu\text{m}$  Wavelength Bands // ECOC'08 – Sept. 2008 – Brüssel, Th1.C1

## **OPTICAL AMPLIFIERS FOR REACH EXTENSION IN GIGABIT PASSIVE OPTICAL NETWORKS**

**A. TUSSUPOV**  
*PhD student of the L.N.  
Gumilyov Eurasian  
National University*

**N.I Listopad.,**  
*Doctor of Technical Sciences  
Professor, Head of the Department  
of “Information Radio  
Technologies”, Belarusian State  
University of Informatics and Radio  
Electronics*

**A.T Tokhmetov,**  
*Ph.D., associate professor  
of the L.N. Gumilyov  
Eurasian National  
University*

*L.N.Gumilyov Eurasian National University  
Nur-Sultan, Kazakhstan  
E-mail: tussupov@gmail.com*

**Abstract.** Gigabit Passive Optical Networks (GPON) are the most advanced technology. The bit rates is 2.5 Gbps for upstream and downstream. But this network architecture has a limited physical network length of 20 km. This is due to high budget losses of the network. This restriction of access makes the network inaccessible for subscribers located far from the telecom operator’s facilities and coverage of remote settlements is quite investment-intensive, thereby complicating the elimination of the digital divide between city, suburbs and villages. To solve this problem, the use of optical amplifiers is proposed, which will expand the reach of the GPON to 60 km, which is the limit for the logical length according to current protocols.

**Keywords:** GPON, last mile, optical amplifier, hybrid WDM / TDM GPON

УДК 004.94:504.054

## **ДИНАМИЧЕСКИЙ ВЫЧИСЛИТЕЛЬНЫЙ ГРАФ ДЛЯ ОБОБЩЕНИЯ РАЗРОЗНЕННЫХ НАУЧНЫХ ДАННЫХ В УНИВЕРСАЛЬНУЮ МОДЕЛЬ НА ПРИМЕРЕ ПОВЕДЕНИЯ ЦЕЗИЯ В СИСТЕМЕ «ПОЧВА-РАСТЕНИЕ»**



***А.Н. Никитин***

*Заведующий лабораторией радиоэкологии ГНУ  
«Институт радиобиологии Национальной академии  
наук Беларуси», кандидат сельскохозяйственных наук*

*Государственно научное учреждение «Институт радиобиологии Национальной академии наук  
Беларуси», Республика Беларусь  
E-mail: nikitinale@gmail.com*

### ***А.Н. Никитин***

*Заведующий лабораторией радиоэкологии Институт радиобиологии НАН Беларуси. Кандидат сельскохозяйственных наук. Проводит научные исследования в области моделирования поведения техногенных радионуклидов в естественных экосистемах.*

**Аннотация.** Разработка эффективных методов синтеза новых научных знаний посредством создания моделей, обобщающих накопленные разнородные данные является актуальной задачей для областей с высокими затратами на получение каждого обучающего примера. В статье обосновывается использование для этих целей динамического вычислительного графа с известными уравнениями в узлах и измеряемыми или виртуальными переменными на гранях. Обучение графа осуществляется методом стохастического градиентного спуска и обратного распространения ошибки. Для вычисления ошибки используются как целевая, так и промежуточные переменные. Показано, что данный подход позволяет получить легко интерпретируемую и валидируемую модель для обучения которой используется весь набор имеющихся разнородных данных, описывающих исследуемый объект или явление.

**Ключевые слова:** динамический вычислительный граф, обобщающая модель, агрегация данных, радиоэкология, цезий.

Данная работа посвящена проблеме синтеза нового научного знания на основе обработки накопленных разрозненных данных. Продуктом научной деятельности являются не только новые знания, но и набор исходных и прошедших математическую и статистическую обработку данных. Сегодня многие издательства научных журналов и финансирующие исследования организации стимулируют размещение в общем доступе исходных данных, собранных при проведении исследований. С одной стороны, это увеличивает воспроизводимость исследований и позволяет верифицировать результаты независимыми учеными. С другой стороны, за счет обобщения и обработки объединенного массива данных могут быть синтезированы новые научные знания. С учетом того, что получение исходных данных нередко сопряжено с серьезными финансовыми затратами, такой подход позволяет снизить издержки на научную деятельность. А увеличение объема обрабатываемых данных повышает надежность результата и уровень обобщения. Синтезу нового научного знания с помощью обработки накопленных ранее данных препятствует отсутствие единства и унифицированности в доступных наборах. Даже при решении идентичных научных задач

набор данных (измеряемых параметров, условий эксперимента и т.п.) может различаться. А если стоит задача собрать максимально возможный объем данных, связанных с заданным объектом или явлением, то разнообразие наборов возрастает многократно, и степень пересечения между ними может различаться от полной до пустого множества.

Одним из наиболее распространенных подходов для синтеза нового научного знания из набора опубликованных результатов исследований является мета-анализ. Данный подход можно отнести, скорее, к полуколичественным. Используются в нем, как правило, конечные результаты опубликованных исследований, а не промежуточные показатели или исходные данные. Построить сколько-нибудь сложную модель исследуемого объекта с помощью методов метаанализа практически невозможно, а область ее применения лишь в незначительной мере расширяется относительно исходных публикаций. В области радиозкологии примерами подобных подходов могут быть обобщающие публикации по коэффициентам перехода радионуклидов из почвы в растения [1;2] или метаанализ воздействия ионизирующих излучений на живые организмы [3-5].

Несколько в стороне стоит другой подход, заключающийся в объединении простых моделей в качестве подмоделей в единую, более сложную супермодель. Супермодель имеет более широкую область применения и может переноситься на условия, не исследовавшиеся в некоторых из подмоделей. В области радиозкологии примером такого подхода может быть модель Absolom [6; 7] и ее оптимизированная версия [8]. Для использования данного подхода необходимо наличие уже готовых субмоделей или наборов данных, позволяющих их построение. Основное ограничение заключается в том, что часто субмодели строятся по результатам разрозненных экспериментов, различающихся по условиям, контролируемым параметрам, объектам, и адекватность их объединения в супермодель нередко вызывает вопросы.

Таким образом, возникает потребность в разработке методического подхода, позволяющего создавать модели объектов или явлений на основе объединения разрозненных наборов данных, обладающих различной степенью полноты, различающихся по условиям и методом сбора, с максимально полным учетом каждого из фрагментов. Использование подобного подхода позволит создавать новые научные знания на основе собранных различными исследовательскими группами данных. Степень обобщения, надежность и область применения таких моделей может быть существенно шире по сравнению с агрегацией, осуществленной с использованием иных подходов.

Наша гипотеза состоит в том, что поставленная задача может быть решена с использованием динамического вычислительного графа. В узлах графа могут использоваться известные или вновь предлагаемые уравнения, описывающие связь между отдельными параметрами системы, а на ребрах – измеряемые или виртуальные переменные. Параметры уравнений в узлах вычислительного графа могут оптимизироваться с использованием стандартных методов машинного обучения с обратным распространением ошибки [9]. Но при обучении используются все возможные для данного примера ошибки: как финального в вычислительном графе значения, так и промежуточных, представленных в примерах из набора данных.

Одной из областей научных знаний, нуждающейся в эффективном подходе для обобщения накопленных научных данных в единой модели, является радиозкология. Данные в этой области собраны либо в регионах, подвергшихся радиоактивному загрязнению в результате аварий, либо в лабораторных условиях. Выявленные количественные закономерности лишь в ограниченной мере применимы для отличающихся природно-климатических условий. Это вызывает серьезные затруднения при организации мероприятий по ликвидации последствий инцидентов с выбросом радиоактивных веществ в окружающую среду. Универсальная модель поведения основных дозообразующих веществ в окружающей среде, формирования доз облучения человека и биоты, а также оценки эффективности



различных мер ограничения облучения позволит повысить степень готовности к реагированию на аварийные ситуации.

Обобщение разрозненных данных о поведении радионуклидов в окружающей среде в универсальной модели должно основываться на использовании известных физических, химических и биологических закономерностей, определяющих интенсивность перехода и перераспределения изотопов между компонентами экосистем. Для моделирования поведения радиоактивных веществ в системе "почва-растение", среди прочего, необходимо учитывать физико-химические формы радиоактивных выпадений, процессы изменения этих форм в почве, взаимодействие радиоактивных изотопов с ионами, органическими молекулами и кристаллическими структурами в почве, влияние на эти процессы микроорганизмов и корневых выделений растений, механизмы корневого поглощения элементов, физиологически и молекулярно-биохимические процессы в растениях. Учет и анализ всех этих факторов в рамках одного эксперимента практически невозможен. Но объединение результатов разнообразных экспериментов, посвященных данным проблемам, имеет существенное фундаментальное и прикладное значение.

Для проверки выдвинутой гипотезы нами выбрана очень упрощенная (минимально жизнеспособная) модель поведения  $^{137}\text{Cs}$  в системе "почва-растение" (рис. 1). Анализ доступных наборов данных по данному предмету указывает на то, что в них могут присутствовать как все входные и выходные переменные (очень редко), так и лишь отдельные элементы в различных сочетаниях.

Многие классические работы, посвященные проблеме накопления радиоактивных изотопов цезия растениями содержат данные о валовом содержании в почве  $^{137}\text{Cs}$  или  $^{134}\text{Cs}$  и обменного калия. Ряд авторов не уделял достаточное внимание роли  $\text{K}$  в корневом поглощении цезия, поэтому они обнародовали данные только по содержанию радиоактивного изотопа в почве и растениях. Значительное количество работ посвящено химическим процессам, влияющим на переход цезия в растворимую форму, в них мы можем найти данные по соотношению различных форм нахождения  $^{137}\text{Cs}$  в почве и факторам, влияющим на него. Среди рассматриваемых переменных наиболее трудоемкой в определении является содержание  $^{137}\text{Cs}$  в почвенном растворе, поэтому можно найти публикации, включающий практически полный набор данных, но без этой переменной. Кроме того, проведено достаточно большое количество лабораторных экспериментов с гидропонной культурой, посвященных исследованию интенсивности поступления в растения радиоактивных изотопов цезия из раствора и влиянию на нее содержания в растворе  $\text{K}^+$  и других ионов.

Можно использовать несколько различных способов включения этих данных в единую модель. Мы сразу отбрасываем вариант с фильтрацией только тех примеров, которые содержат полный набор переменных. Это сильно ограничивает набор данных для обучения модели и, по сути, не привносит ничего нового.

Второй способ заключается в фильтрации данных, содержащих полный набор входных и выходных переменных, значимых с практической точки зрения (валовое содержание  $^{137}\text{Cs}$  в почве, концентрация обменного калия и удельная активность  $^{137}\text{Cs}$  в растении, в нашем случае) и обучении модели только на них. Поскольку связь между входными и выходными переменными может быть довольно сложной, для построения модели целесообразно использовать искусственную нейронную сеть, метод опорных векторов или модели, основанные на композициях решающих деревьев. Недостатком такого способа является отказ от определенной доли данных, довольно существенной в большинстве случаев, при одновременной потребности в большом их объеме, поскольку подобные модели имеют значительное количество обучаемых параметров.

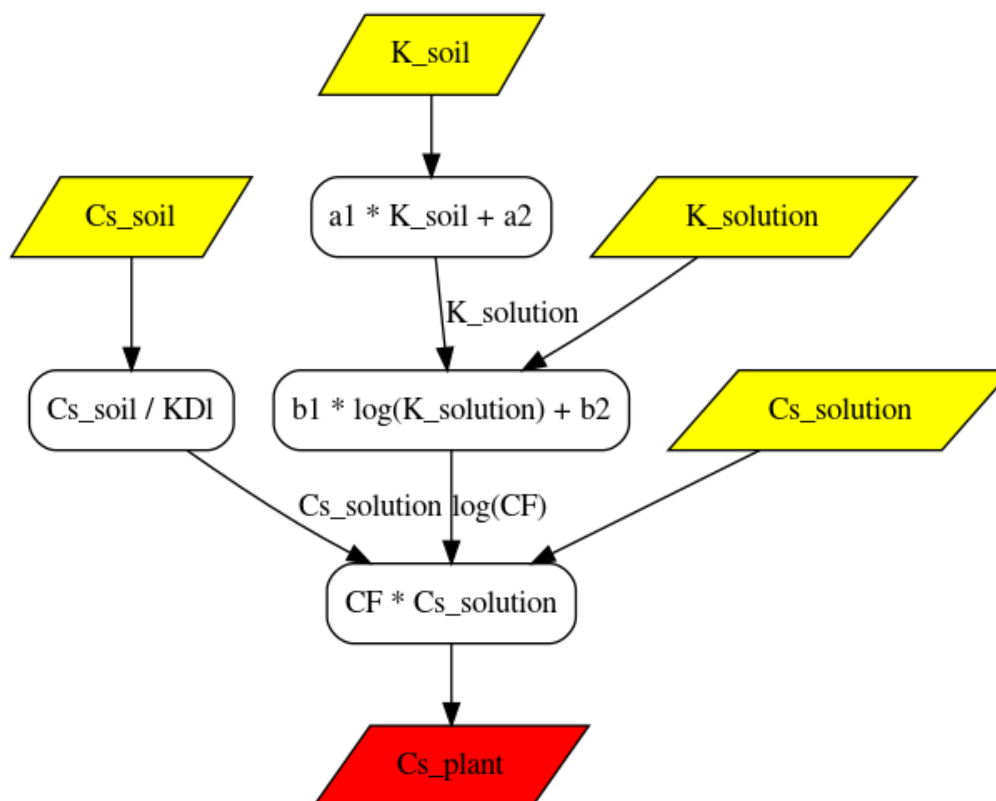


Рисунок 1. – Упрощенная модель поведения цезия в системе "почва-растение". Условные обозначения:  $K_{soil}$  – содержание обменного калия в почве,  $K_{solution}$  – концентрация калия в почвенном растворе,  $Cs_{soil}$  – удельная активность  $^{137}Cs$  в почве,  $Cs_{solution}$  – удельная активность  $^{137}Cs$  в почвенном растворе,  $Cs_{plant}$  – удельная активность  $^{137}Cs$  в растении, CF (коэффициент концентрирования  $^{137}Cs$ ), KD1 (коэффициент распределения  $^{137}Cs$  между твердой фазой и почвенным раствором),  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$  – обучаемые параметры модели. Желтые параллелограммы – входные переменные, Красный параллелограмм – выходная переменная.

В настоящее время наиболее распространенным подходом является обучение субмоделей для отдельных звеньев модели на доступных для этого поднаборах данных. Можно также обучать объединения из нескольких субмоделей в случае пропусков в промежуточных переменных. При объединении в модель необходимо провести усреднение обученных параметров с весами, пропорциональными надежности используемых данных. Недостатком этого способа является потенциальная несовместимость субмоделей, обученных на наборах данных, собранных в сильно отличающихся условиях. Кроме того, могут возникнуть трудности в обучении и последующем объединении композиций субмоделей, обученных на различных наборах данных, с различными вариантами пропуска промежуточных переменных.

Обучение динамического вычислительного графа на полном наборе собранных данных является способом, практически лишенным перечисленных выше недостатков. При работе с хорошо изученной проблемой в узлах вычислительного графа могут использоваться известные функции, связывающие отдельные переменные, а для инициализации – параметры этих функций, оцененные с той или иной степенью достоверности. Эти знания являются надежной отправной точкой для обучения модели.

Поскольку реальные численные значения переменных, используемых в модели, могут отличаться на порядки и такой же диапазон варьирования у параметров уравнений, обучение модели может быть крайне неэффективным. Для преодоления этой проблемы следует

использовать стандартный подход машинного обучения – стандартизация значений. Но прибегать к нормализации ( $\mu = 0$ ,  $\sigma = 1$ ) в данном случае неприемлемо, поскольку многие переменные, описывающие реальные состояния и качества, никогда не принимают отрицательные значения и в используемых уравнениях могут использоваться такие функции, как логарифм или квадратный корень. Оптимальным решением в данном случае может быть деление значения на максимальное или среднее в ряду данных. При выборе такого решения легко переводить параметры уравнений в стандартизированный вид и обратно.

Библиотека PyTorch является удобным средством для решения поставленной задачи. Динамический вычислительный граф в ней может быть сформирован в несколько строчек кода. Он легко подстраивается под имеющийся набор данных и имеет встроенные функции вычисления градиентов и обучения методом обратного распространения ошибки. Имеется возможность сложения нескольких ошибок, что немаловажно в данном случае, поскольку перед нами стоит задача использовать для обучения как выходные, так и промежуточные переменные.

Для апробации данного подхода в PyTorch был построен динамический вычислительный граф с архитектурой, представленной на рисунке 1, с возможностью подстраиваться под каждый обучающий пример в зависимости от входящих в его состав переменных и сложением ошибок как от выходной переменной (удельная активность  $^{137}\text{Cs}$  в растении), так и от промежуточных (содержание  $\text{K}$  и  $\text{Cs}$  в почвенном растворе). При этом ошибкам от промежуточных значений придавались в пять раз более низкие веса по сравнению с финальными, в соответствии с практической важностью данных переменных. В качестве функции ошибки использовано среднее квадратическое отклонение. Метод обучения – стохастический градиентный спуск; методы основанные на одновременном обучении на всем наборе данных или подвыборках здесь неприменимы, поскольку граф должен подстраиваться под каждый пример.

В испытаниях метода использованы искусственно сгенерированные реалистичные данные, полученные с помощью модели [8] и добавлением шума различной интенсивности к каждой переменной и параметру. Данные были стандартизированы делением на максимальное значение. После этого случайным образом удалено 5, 10 и 20% значений каждой переменной. Набор данных включал 1000 примеров, которые перед обучением разделялись на обучающую и проверочную выборки в соотношении 4:1.

Значения параметров модели были инициализированы единицами. Следует отметить, что предлагаемая модель практически не нуждается в регуляризации, поскольку учет ошибок от промежуточных переменных позволяет удерживать значения параметров в приемлемом диапазоне.

Предварительный анализ показал, что для обучения достаточно 500–1000 эпох. Среднеквадратичная ошибка модели на валидационной выборке при обучении на полном наборе данных, при удалении 5, 10 и 20% случайных значений каждой переменной оказалась практически идентичной – 0,012–0,013. Можно лишь отметить, что увеличение степени разрозненности данных в выборке несколько снижает скорость обучения модели (рисунки 2 и 3). Полученный результат свидетельствует о высокой эффективности предложенного подхода для обобщения разнородных данных в единую вычислительную модель.

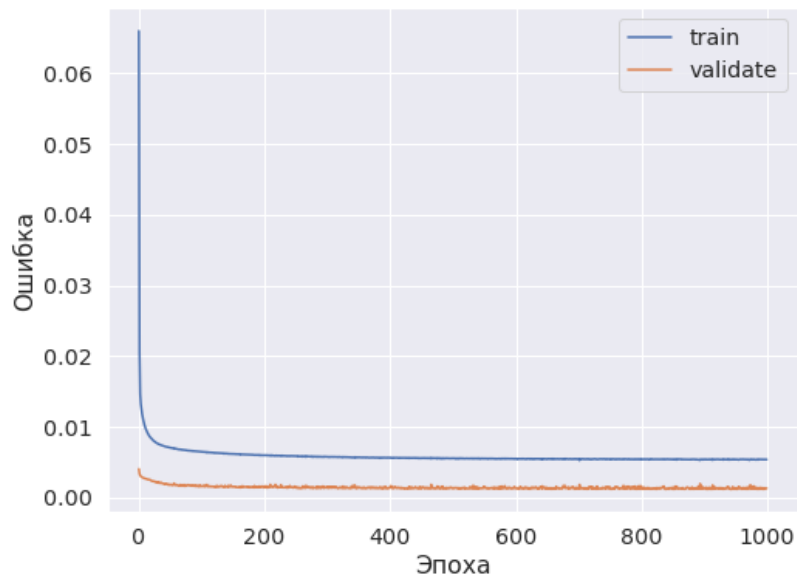


Рисунок 2. – Динамика обучения модели на данных без пропусков. Примечание: ошибка на обучающей выборке является взвешенной суммой среднеквадратичных ошибок по выходной и промежуточным переменным, ошибка на валидационной выборке представляет среднеквадратичную ошибку только на выходной переменной

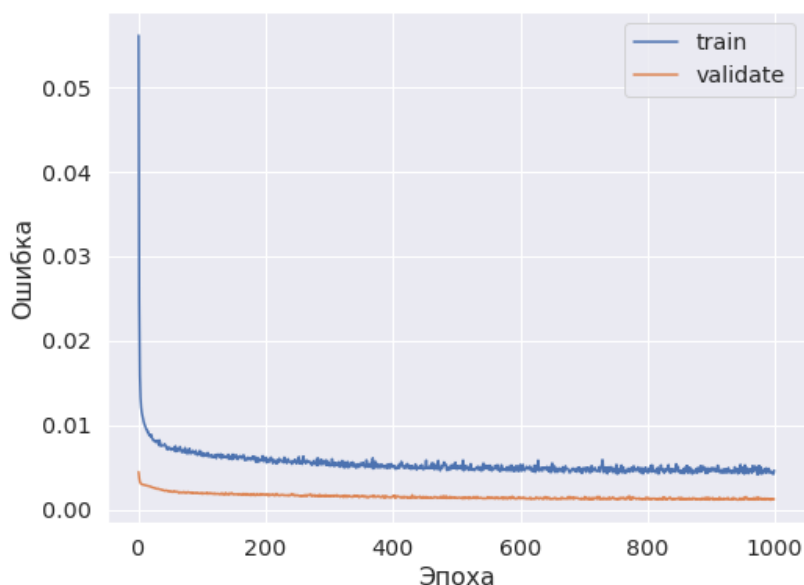


Рисунок 3. – Динамика обучения модели на данных при удалении 20% случайных значений в каждой из переменных. Примечание: ошибка на обучающей выборке является взвешенной суммой среднеквадратичных ошибок по выходной и промежуточным переменным, ошибка на валидационной выборке представляет среднеквадратичную ошибку только на выходной переменной

Таким образом, предложен подход, позволяющий создавать количественную модель, обучаемую на наборе разрозненных данных. Данный подход актуален для решения научных и прикладных задач, где получение обучающих примеров сопряжено с существенными финансовыми или иными издержками, что заставляет в максимальной степени использовать

накопленные ранее данные. Причем получение этих данных могло быть сопряжено с решением отличающихся друг от друга задач. Следует также сказать, что подобные модели используют ограниченное количество обучаемых параметров, что накладывает менее жесткие нижние ограничения на объем данных.

Использование динамического вычислительного графа позволяет обучать модель на всех данных, практически без исключения. При этом для обучения используются не только все доступные примеры, но и промежуточные значения в примерах. Ошибкам по наиболее важным с практической или иной точки зрения переменным придаются более высокие веса при обучении. Обученная модель обладает высокой степенью согласованности составных частей, а область ее определения и степень обобщения шире, чем у каждого из наборов данных в отдельности. В области радиоэкологии подобные модели могут использоваться при загрязнении радионуклидами территорий с разнообразными природно-климатическими условиями, а также применяться для прогноза поведения радионуклидов на фоне изменяющихся погодно-климатических условий.

Облегчает обучение описанного динамического вычислительного графа возможность использования известных параметров уравнений в качестве отправной точки при обучении. Состояние близкое к минимуму ошибки повышает вероятность нахождения глобального минимума и снижает риск переобучения. После обучения полученные параметры можно сравнить с опубликованными значениями и сделать дополнительное заключение об адекватности модели.

В отличие от многих других методов машинного обучения, предлагаемая модель полностью интерпретируема. Она может использоваться как целиком, так и отдельными фрагментами, давая на выходе значения, соответствующие переменным, характеризующим реальные явления или предметы.

Созданная модель может включаться в качестве составной части в более крупную, приобретая возможность обучаться на еще более широком наборе данных. В частности, модель поведения радионуклидов в системе "почва-растение" может включаться в модель миграции радиоактивных изотопов по пищевым цепям, формирования доз облучения населения и влияния инкорпорированных радионуклидов на здоровье. Подобные инструменты востребованы при проектировании аварийных ситуаций и защитных мер.

Мы полагаем, что данная модель содержит в себе дополнительную информацию, которая может быть использована для ее дальнейшего совершенствования. Ошибки по промежуточным переменным на валидационной выборке указывают на слабые узлы вычислительного графа, в которых можно попытаться изменить уравнения для получения более надежных результатов. Сумма квадратов градиентов на обучаемых параметрах указывает на надежность их оценки на использованном наборе валидационных данных. Внедрение этих приемов является ближайшей задачей развития предложенного подхода.

### **Список литературы**

[1.] IAEA. Quantification of radionuclide transfer in terrestrial and freshwater environments for radiological assessments / IAEA. – Vienna: International atomic energy agency, 2009. – 616 p.

[2.] IAEA. Handbook of parameter values for the prediction of radionuclide transfer in terrestrial and freshwater environments. Technical report No. 472. / IAEA. – Vienna: International atomic energy agency, 2010. – 197 p.

[3.] Sazykina T.G. Non-parametric estimation of thresholds for radiation effects in vertebrate species under chronic low-dose exposures / T.G. Sazykina, A.I. Kryshev, K.D. Sanina // Radiation and Environmental Biophysics. – 2009. – Vol. 48. – № 4. – P. 391-404.

[4.] Fesenko S. Comparative radiation impact on biota and man in the area affected by the accident at the chernobyl nuclear power plant / S. Fesenko [et al.] // Journal of Environmental Radioactivity. – 2005. – Vol. 80. – № 1. – P. 1-25.

[5.] Real A. Effects of ionising radiation exposure on plants, fish and mammals: Relevant data for environmental radiation protection / A. Real [et al.] // Journal of Radiological Protection. – 2004. – Vol. 24. – № 4A. – P. A123-A137.

- [6.] Absalom J.P. Predicting soil to plant transfer of radiocesium using soil characteristics / J.P. Absalom [et al.] // Environmental Science & Technology. – 1999. – Vol. 33. – № 8. – P. 1218-1223.
- [7.] Absalom J. Predicting the transfer of radiocaesium from organic soils to plants using soil characteristics / J. Absalom [et al.] // Journal of Environmental Radioactivity. – 2001. – Vol. 52. – № 1. – P. 31-43.
- [8.] Tarsitano D. Evaluating and reducing a model of radiocaesium soil-plant uptake / D. Tarsitano, S. Young, N. Crout // Journal of Environmental Radioactivity. – 2011. – Vol. 102. – № 3. – P. 262-269.
- [9.] Rojas R. The backpropagation algorithm / R. Rojas // Neural networks: A systematic introduction : vols. – Berlin, Heidelberg: Springer Berlin Heidelberg, 1996. – P. 149-182.

## **DYNAMIC COMPUTATIONAL GRAPH FOR GENERALIZATION BITTY SCIENTIFIC DATA IN UNIVERSAL MODEL USING BEHAVIOR OF CESIUM IN «SOIL-PLANT» SYSTEM AS EXAMPLE**

*A. N. Nikitin*

*Head of the Laboratory of Radioecology in the Institute of Radiobiology of the National Academy of Sciences of Belarus*

*State Scientific Institution «Institute of Radiobiology of the National Academy of Sciences of Belarus», Republic of Belarus  
E-mail: nikitinale@gmail.com*

**Abstract.** The development of effective methods for synthesizing new scientific knowledge by creating models that summarize the accumulated heterogeneous data is an urgent task for areas with high costs for obtaining each training example. The article suggest dynamic computational graph with known equations in nodes and measured or virtual variables on the edges for this purpose. The computational graph is trained by the method of stochastic gradient descent and back-propagation of the error. The error is calculated both on the target and intermediate variables. The article prove that proposed approach allows to obtain an easily interpreted and validated model for training which uses the entire set of available heterogeneous data describing the object or phenomenon being studied.

**Keywords:** dynamic computation graph, generalized model, data aggregation, radioecology, cesium.

УДК 004.032.26:581.522-047.44

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ БАНКОВСКИХ РЕШЕНИЙ ПРИ ВЫДАЧЕ КРЕДИТОВ



**Д.С. Сенькович**  
Магистрант кафедры  
информатики БГУИР, инженер-  
программист



**А.В. Жвакина**  
Доцент кафедры информатики  
БГУИР, кандидат технических  
наук, доцент

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

E-mail: dmitrysenkovich@gmail.com, zhvakina@bsuir.by

### **А.В. Жвакина**

*Образование:* Белорусский государственный университет информатики и радиоэлектроники; кандидат технических наук, доцент. *Научные интересы:* нейронные сети, системы поддержки принятия решений, языки программирования, веб-программирование, бизнес-анализ, разработка требований к программному обеспечению, включая UML-моделирование, разработка приложений с графическим интерфейсом для решения прикладных задач. Более 80 публикаций, в том числе 14 учебных пособий.

### **Д.С. Сенькович**

*Образование:* Белорусский государственный университет, Факультет Прикладной Математики и Информатики; *Место работы:* программист, “NES FinTech”; *Научные интересы:* нейронные сети, веб-программирование, облачные вычисления, распределенные системы, микросервисы.

**Аннотация.** Обсуждаются различные подходы к анализу информации о клиентах банка с целью принятия решения о выдаче кредита. Исследованы различные способы моделирования данной задачи, оценена их точность и время получения результатов.

**Ключевые слова:** анализ данных, нейронные сети, кредитование.

Залогом выгодного кредитования является оценка потенциального заемщика с точки зрения его платежеспособности и социальной благонадежности. Необходимо проанализировать возможность возврата кредита, опираясь на данные о предыдущей кредитной истории клиента, его платежеспособности и социальных особенностях.

Для решения данной задачи используются различные подходы. Один из них базируется на применении кредитного скоринга [1]. При этом рассматривается информация об имевшихся кредитах, и на основании статистической или математической обработки данных из базы кредитного регистра определяются:

- класс рейтинга;
- скорбалл;
- вероятность просрочки более 90 дней в год на определенную сумму.

Класс рейтинга присваивается на основании оценки в баллах (скорбаллов), характеризующей вероятность того, что кредит не будет возвращен, и вероятности просрочки выплат по кредиту. Чем ниже скорбалл, тем выше вероятность нарушения своевременного погашения кредита. Важными факторами являются время, прошедшее после заключения

первого договора о кредитной сделке, количество таких сделок, а также овердрафтовых и потребительских договоров, суммах и продолжительности просрочек.

Недостатком данного подхода является то, что его нельзя использовать по отношению к лицам, которые вообще не брали кредиты или брали их более 5 лет назад, заключали сделки о кредитах с лизинговыми или микрофинансовыми организациями.

Таким образом, чтобы принять решение о выдаче кредита, недостаточно опираться лишь на сведения о кредитном рейтинге, необходимо оценить также множество факторов, влияющих на возврат денег заемщиком: платежеспособность, которую характеризует уровень дохода и надежность компании-работодателя, демографические и социальные данные [2, 3].

В качестве параметров, используемых для принятия решения о выдаче кредита, выбраны следующие:

- размер кредита (amount);
- срок кредитования (term);
- сумма ежемесячного дохода (income);
- род занятий (occupation\_type);
- наличие отчислений в пенсионный фонд(pension\_contributions);
- судимость – наличие, отсутствие (criminal\_record)
- семейный статус – женат, холост (marital\_status)
- сумма пенсионного дохода(add\_income\_pension)
- звонок перед визитом(add\_info\_call\_before\_visit)
- мошенничество (add\_info\_fraud)
- недостаточность информации (add\_info\_inadequate)
- плохой внешний вид (add\_info\_poor\_appearance)
- наличие задолженности (had\_arrears)
- раньше были кредиты (had\_credits\_before)
- имеет активный кредит (has\_active\_credit)
- ежемесячный платеж (monthly\_payment)
- официальный доход (pt\_income)

Использовалось 1000 обезличенных наборов данных, описывающих реальных потенциальных заемщиков и принятое банком решение.

Так как исходные данные представлены разными типами, то предварительно необходимо было привести их в числовой вид и обработать: булевы значения (например, наличие пенсионных отчислений – да или нет) заменить на 1 и 0, данные, принимающие возможное значение из списка – на номер в данном списке; признаки, которые не содержат полезной информации, например, с одинаковыми значениями для всех клиентов отбросить.

Обрабатываемые данные случайным образом разделены на две группы: использованные для тренировки, и те, на которых проводилась независимая оценка качества, в соотношении 4:1, т.е. 800 и 200 записей. В свою очередь тренировочный набор данных разделен на выборки для обучения и валидации в соотношении 3:2.

Для прогнозирования вероятности погашения кредита использовались логистическая регрессия, метод опорных векторов (SVM), алгоритм случайного леса (*random forest*), нейронные сети.

Задача усложняется тем, что является несбалансированной – в наборе данных из 1000 клиентов только 90 получили отказ в выдаче кредита, поэтому при обучении подсчитывается максимальная оценка F1, выбирается модель с наилучшим ее значением, параметры, соответствующие данной модели, используются для оценки на тестовой выборке (*F1 score*).

В случае логистической регрессии исследовались алгоритмы оптимизации, представленные в таблице 1, использовано 100 различных случайных значений коэффициента регуляризации в интервале от 0 до 1.



Таблица 1. – Логическая регрессия

Алгоритм оптимизации	Вид регуляризации
liblinear	l1
sag	l1
newton-cg	l2
lbfgs	l2
sag	l2
saga	l2

SVM обучена с различными видами ядер и 100 различными случайными значениями коэффициента регуляризации в интервале от 0 до 1. И использованные ядра: *rbf*, *poly*, *sigmoid*. Линейное ядро не было использовано, потому что этот способ использования SVM сводится к логистической регрессии.

Случайный лес обучался с использованием 100 случайных значений количества простейших алгоритмов (деревьев) в интервале от 50 до 500.

Недостатком вышерассмотренных подходов является то, что при аппроксимации нелинейных зависимостей количество параметров в таких моделях растет слишком быстро, что не характерно для нейронных сетей.

В экспериментах использовались полносвязная нейронная сеть (рисунок 1) и сеть для распознавания образов (*patternnet*), представленная на рисунке 3.

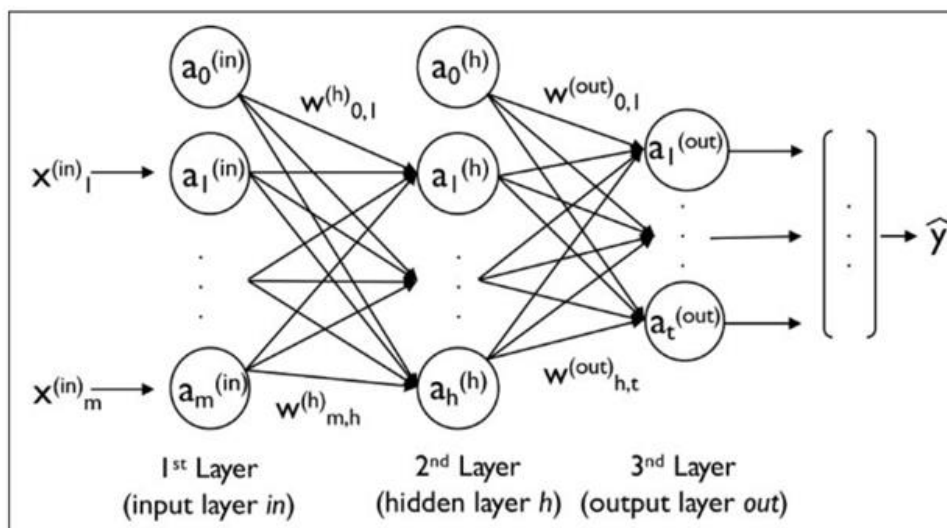


Рисунок 1. – Полносвязная нейронная сеть

В экспериментах использованы модели с различным количеством внутренних слоев: от 1 до 7.

В качестве функции активации выбрана *ReLU*, которая имеет формулу  $f(x) = \max(0, x)$  и представлена на рисунке 2.

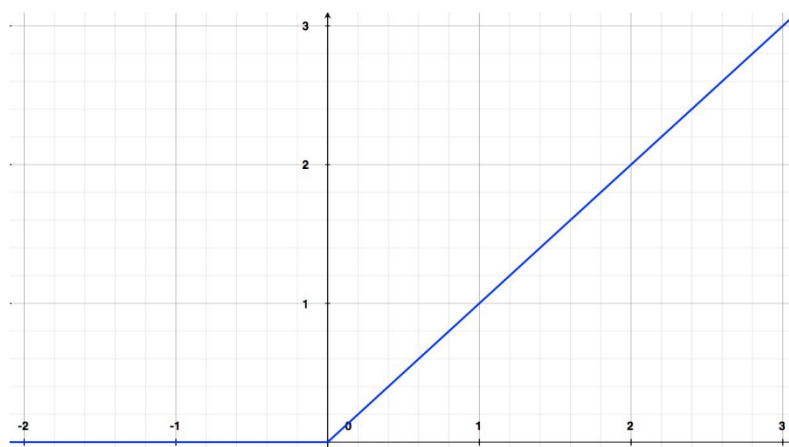


Рисунок 2. – Функции активации

Достоинством данной функции по сравнению, например, с классической сигмоидой, является ее простота, так как это ускоряет вычисления [8]. Функция активации *ReLU* используется во всех экспериментах в полносвязных слоях.

Выходной слой модели имеет функцию активации *softmax*.

Инициализация весов нейронной сети может быть реализована простой выборкой небольших чисел, с небольшим разбросом около 0.01. Однако существуют исследования [5], в которых обосновывается выбор начальных весов из того или иного распределения в зависимости от количества нейронов на слое, функции активации и тому подобных параметров. Такие способы позволяют достичь лучших результатов на более ранних этапах обучения.

Примером может служить *glorot uniform* инициализатор. Это один из самых распространенных инициализаторов, использующихся для глубоких нейронных сетей [9, 10]. Особенно хорошие результаты в сходимости функция дает при использовании функции активации гиперболический тангенс. Вместо инициализации весов случайными числами из некоторого интервала данная схема инициализации выбирает значения из следующего интервала:

$$[-\sqrt{6 / (fan\_in + fan\_out)}; \sqrt{6 / (fan\_in + fan\_out)}],$$

где *fan\_in* и *fan\_out* длина векторов входных и выходных данных соответственно.

Математически доказано Хавьером Глоротом, что такая схема инициализации позволяет избежать возможной ситуации, при которой градиенты будут либо чересчур маленькими или вовсе нулевыми, что может привести к медленному обучению, а в худшем случае и отсутствию прогресса [7].

Другой довольно популярный выбор инициализатора - *He normal*. Этот способ выбирает значения из нормального распределения с математическим ожиданием 0 и разбросом  $\sqrt{2 / fan\_in}$ , где *fan\_in* длина векторов входных данных. Этот способ лучше зарекомендовал себя с функцией активации ReLU, которая и используется в экспериментах.

Для каждой нейронной сети количество нейронов выбрано от 16 до 64 случайным образом. Каждый слой сети содержит именно это количество нейронов.

Каждая из нейронных сетей обучена в течение 50 эпох. Эпоха – изменение весов сети после прямого и обратного распространения на всех векторах выборки. Слишком маленькое значение ведет к более высокой вероятности получения более простой модели, неспособной выявлять сложные нелинейные зависимости. Слишком большие значения количества эпох ведет к возможному переобучению модели, проблемам *vanishing* и *exploding gradients*, когда

производные к ходе обратного распространения настолько малы, что обучение впадает в стагнацию и т.д. [4].

Коэффициент обучения задает скорость изменения параметров, ту часть градиента, которая используется для обновления весов. Любой алгоритм машинного обучения никогда не достигает абсолютного оптимума, а изменяется в некоторой его окрестности. При меньших значениях коэффициента алгоритм будет сходиться медленнее, но в какой-то момент позволит приблизиться ближе к оптимуму функции потерь, в то время как большие значения коэффициента обучения позволяют быстрее обучать модель на первых порах, ухудшая результаты в будущем. Для каждой из нейронных сетей генерируется случайное значение коэффициента обучения из интервала от 0.1 до 0.001 по 10 на каждый из видов нейронных сетей.

Классическим алгоритмом обучения является *SGD* – стохастический градиентный спуск, оперирующий на подвыборках целой выборки для более быстрого прогресса в обучении [5, 6]. Возможным улучшением является использование метода *Momentum* – учет предыдущих значений параметров при очередном изменении. В этом помогает экспоненциально взвешенное скользящее среднее:

Экспериментально подтверждено, что стохастический градиентный спуск с *Momentum* в большинстве случаев лучше обычного [7].

*RMSprop* – еще один алгоритм обучения сети. Со временем параметры могут стать очень большими или очень маленькими, что приведет к очень большим и очень маленьким значениям градиентов соответственно (проблемы *gradient exploding* и *gradient vanishing* соответственно). *RMSprop* «выравнивает» значения параметров:

Приведенные выше алгоритмы обучения достаточно хорошо себя зарекомендовали на практике. Однако сейчас чаще используется еще один алгоритм – *Adam* [7]. Он сочетает в себе оба подхода, описанных выше, и в экспериментах показал наилучший результат.

Важным моментом для таких сетей является борьба с переобучением. Использовались два подхода:

- *L2*-регуляризация
- *dropout* регуляризация.

Смысл *L2*-регуляризации заключается в «штрафовании» модели за слишком большие веса, ограничивая таким образом значения каждого отдельного веса в сети:

$$D_{i,j}^{(l)} := \frac{1}{m} (\Delta_{i,j}^{(l)} + \lambda \Theta_{i,j}^{(l)}) \text{ , if } j \neq 0$$
$$D_{i,j}^{(l)} := \frac{1}{m} \Delta_{i,j}^{(l)} \text{ if } j = 0$$

*Dropout* регуляризация заключается в «выключении» некоторой части нейронов в слое. Как следствие, нейронная сеть не может «полагаться» на какие-то определенные нейроны, что результирует в регуляризационном эффекте.

В экспериментах использована *dropout* регуляризация. Для каждой нейронной сети генерируется *dropout* слой со случайным коэффициентом *dropout rate* – части «выключенных» нейронов – от 0.1 до 0.5. Этот коэффициент, по сути, вероятность «выключения» нейрона.

Исследовалась также нейронная сеть *patternnet*, которая при проверке на тестовых данных показала точность 96,7% (рисунок 4), общая точность – 92,4.

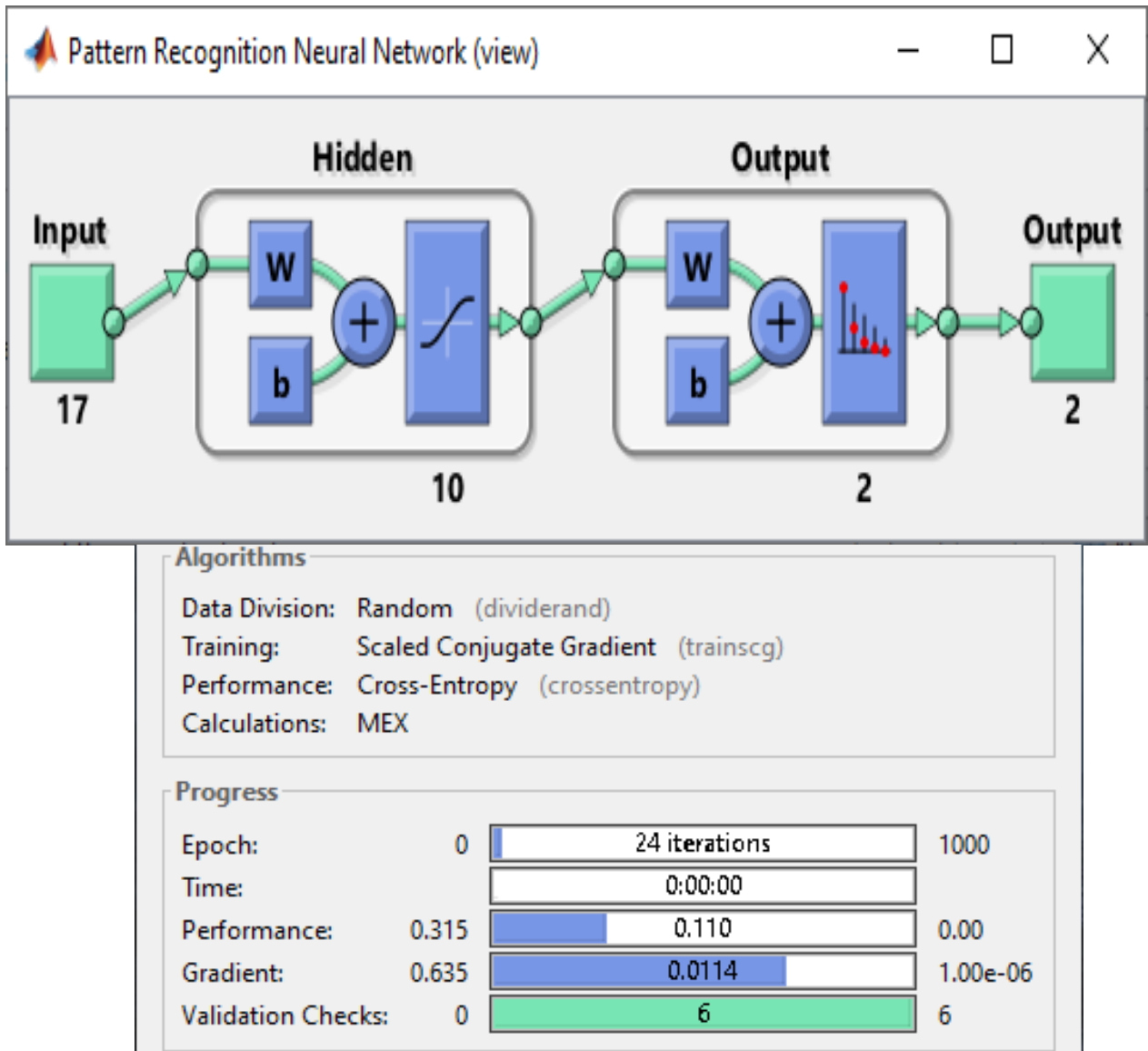


Рисунок 3. – Сеть для распознавания образов

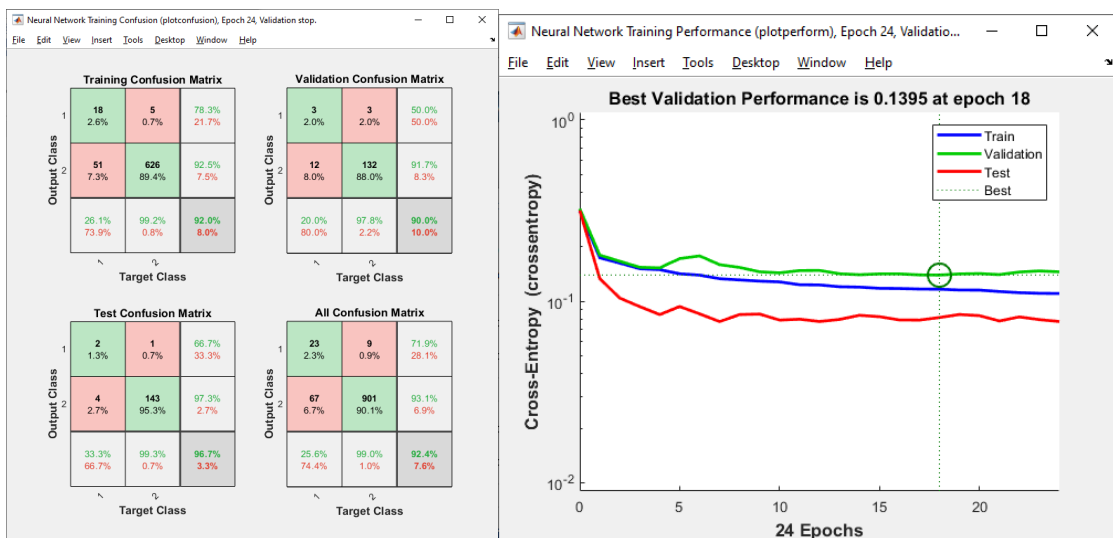


Рисунок 4. – Нейронная сеть *patternnet* (1)

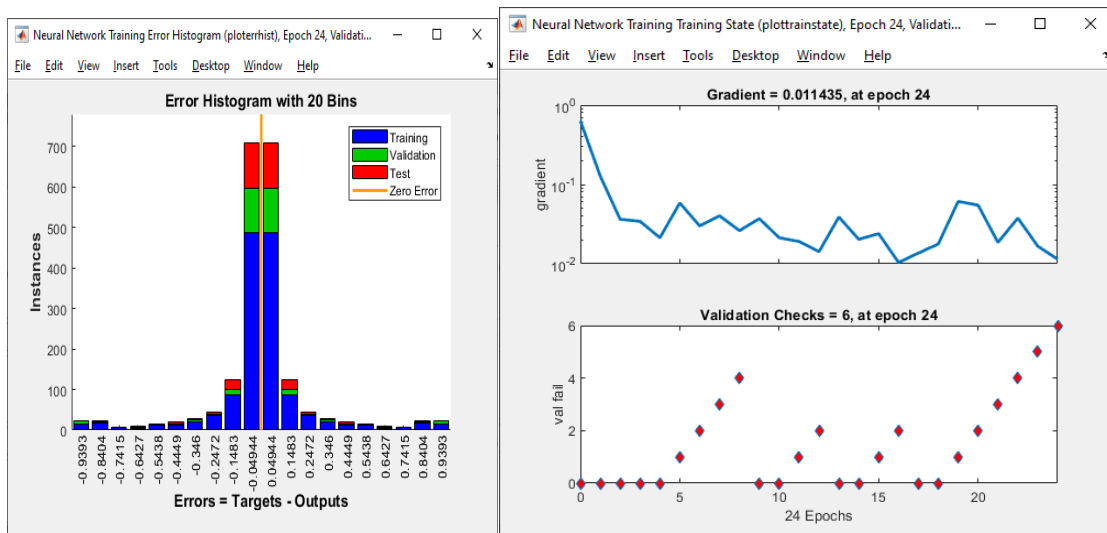


Рисунок 4. – Нейронная сеть *patternnet* (2)

Во втором слое использовалась функция softmax (рисунок 5)

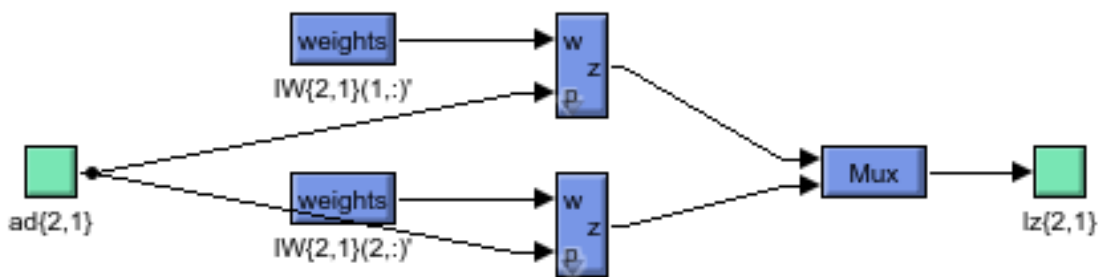


Рисунок 5. – Функция softmax

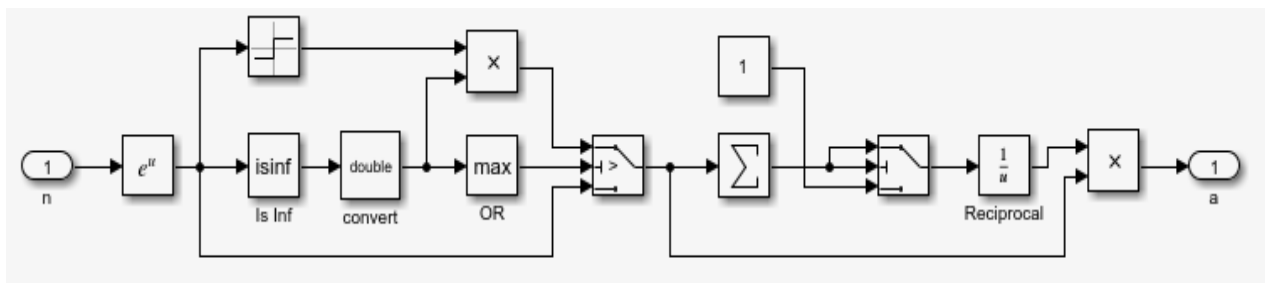


Рисунок 6. – Сеть

При экспериментах с сетью *lvqnet* (learning vector quantization neural network), представленной на рисунок 6, получено более длительное время обучения (8 сек.) и значения точности хуже (91%), что демонстрирует нецелесообразность использования данной сети для решения поставленной задачи.

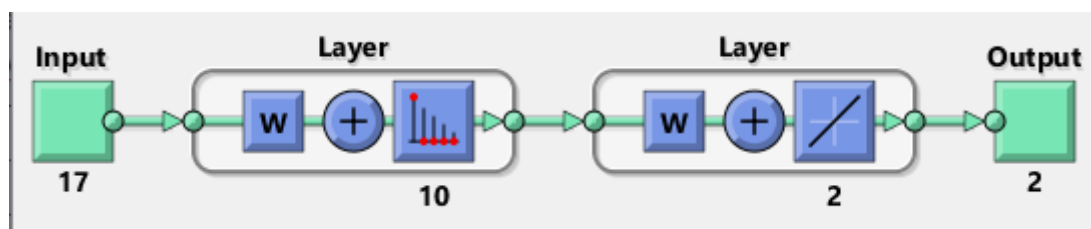


Рисунок 7. – Длительное время обучения

В таблице 2 представлены лучшие результаты обучения с лучшими параметрами у каждой из моделей, включая время обучения на все эксперименты для всех параметров.

Таблица 2. – Лучшие результаты обучения

Модель	Параметры	Время обучения	Время обучения лучшей модели	F1 оценка	Точность
Логистическая регрессия	Алгоритм оптимизации lbfgs, L2 регуляризация, C 0.23	40.94 секунд	0.0149 секунды	0.9595015576	90.5%
SVM	Ядро sigmoid, C 0.17	3.02 секунды	0.005 секунды	0.9565217391	91.1%
Случайный лес	265 базовых алгоритмов, построение деревьев до конца	32.95 секунды	0.34 секунды	0.9620253165	91.5%
Нейронная сеть	Оптимизатор adam, кол-во слоев 1, количество нейронов 61, dropout rate 0.4625, коэффициент обучения 0.0741, обучение в течение 50 эпох	23611.46 секунд ≈ 6.5 часов	4.925 секунды	0.9565217391	95.5%

Таким образом, в результате исследования различных подходов к поддержке принятия решения о выдаче кредита банку: логистической регрессии, метода опорных векторов (SVM), алгоритма случайного леса (random forest), нейронных сетей, наибольшую точность показали нейронные сети. Использование нейронной сети позволит прогнозировать кредитоспособность потенциальных заемщиков и снизить кредитные риски банковских организаций.

#### Список литературы

- [1]. Что такое кредитный скоринг? [Электронный ресурс]. – Режим доступа: <https://creditregister.by/Help/WhatIsCreditScoring> - Дата доступа: 16.02.2020
- [2]. Как банки принимают решение о выдаче кредита [Электронный ресурс]. – Режим доступа: [https://mycreditinfo.ru/kak\\_banki\\_prinimaut\\_reshenie\\_o\\_vydache\\_kredita](https://mycreditinfo.ru/kak_banki_prinimaut_reshenie_o_vydache_kredita) - Дата доступа: 16.02.2020
- [3]. Как банки проверяют заемщиков [Электронный ресурс]. – Режим доступа: <https://www.sravni.ru/enciklopediya/info/kak-banki-proverjajut-zajomshhikov/> - Дата доступа: 16.02.2020
- [4]. Simon Haykin. Neural Networks: A Comprehensive Foundation. (2nd Edition) / Simon Haykin. –Pearson Education, (South Asia :). 2006. – 823 p.
- [5]. Laurene V. Fausett. Fundamentals of Neural Networks : Architectures, Algorithms and Applications. /Laurene V. Fausett. – Delhi, India: Pearson Education India, 2018. – 480 p.

[6]. Каллан Р. Нейронные сети : краткий справочник / Р. Каллан ; пер. с англ. и ред. А. Г. Сивака. – М.: Вильямс, 2018. - 279 с.

[7] Optimizers Explained - Adam, Momentum and Stochastic Gradient Descent / mlfromscratch.com [Электронный ресурс]. – Режим доступа: <https://mlfromscratch.com/optimizers-explained/#/> – Дата доступа: 23.02.2020.

[8] What are the benefits of using ReLU over softplus as activation functions? / Quora [Электронный ресурс]. – Режим доступа: <https://www.quora.com/What-are-the-benefits-of-using-ReLU-over-softplus-as-activation-functions> – Дата доступа: 23.02.2020.

[9] Neural Networks and Deep Learning / Coursera [Электронный ресурс]. – Режим доступа: <https://www.coursera.org/learn/neural-networks-deep-learning> – Дата доступа: 23.02.2020.

[10] Deep Learning Specialization / Coursera [Электронный ресурс]. – Режим доступа: <https://www.coursera.org/specializations/deep-learning> – Дата доступа: 23.02.2020.

## **NEURAL NETWORK DECISION SUPPORT SYSTEM ABOUT LENDING**

***D.S. Senkovich***

*Master's Department  
Informatics BSUIR, Software Engineer*

***A.V.Zhvakina***

*Associate Professor, Department of Informatics,  
BSUIR, Candidate of Technical  
Sciences, Associate Professor*

*Belarusian State University of Informatics and Radioelectronics*

*E-mail: [dmitrysenkovich@gmail.com](mailto:dmitrysenkovich@gmail.com), [zhvakina@bsuir.by](mailto:zhvakina@bsuir.by)*

**Abstract.** The possibilities of various approaches to analyzing information about bank customers with the aim of making a decision on granting a loan are discussed. Various methods of modeling this problem are investigated; their accuracy and time of obtaining results are estimated.

**Keywords:** data analysis, neural networks, lending.

УДК 519.862:336.7

## МОДЕЛИРОВАНИЕ ДИНАМИКИ МАКРОЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ДЕЙСТВИЙ ЗАЕМЩИКОВ КОММЕРЧЕСКОГО БАНКА



**А.И. Слаута,**  
Магистрант ГрГУ,  
специалист ОАО «БНБ-Банк»



**О.Б. Цехан,**  
Заведующий кафедрой математического и  
информационного обеспечения  
экономических систем УО "ГрГУ им.Янки  
Купалы", кандидат физико-  
математических наук, доцент.

Учреждение образования "Гродненский государственный университет имени Янки Купалы",  
Республика Беларусь  
ОАО «БНБ-Банк», Республика Беларусь  
E-mail: anya.slauta@mail.ru, tsekhan@grsu.by

### **А.И. Слаута**

Окончила Учреждение образования "Гродненский государственный университет имени Янки Купалы" в 2018 г. Студент магистратуры по специальности «Прикладной компьютерный анализ данных». Работает в ОАО «Белорусский народный банк» в должности специалиста отдела анализа и контроля кредитных рисков.

### **О.Б. Цехан**

Область научных интересов: математическая теория управления, моделирование динамических систем, теория принятия решений.

**Аннотация.** Построены адекватные эконометрические модели для прогнозирования макроэкономических показателей ВВП, ИПЦ, курс белорусского рубля к доллару с целью применения их в дальнейшем для прогнозирования вероятности дефолта экономических субъектов коммерческого банка Республики Беларусь. В качестве инструмента анализа и прогнозирования использовалась программная среда для выполнения статистических вычислений R.

**Ключевые слова:** макроэкономические показатели, вероятность дефолта, заемщики, ARIMA, прогнозирование, R

**Введение.** Кредитная деятельность является основным источником дохода банков. Поэтому важной для успешной деятельности банков является задача мониторинга рисков, рисков невозврата долга, а также прогнозирование вероятности дефолта на будущие периоды. Прогнозирование развития возможных отрицательных процессов в деятельности банка позволит обеспечить высокую эффективность кредитной деятельности банка, избежать банкротства. Отметим, что наличие оценки вероятности дефолта относится к основным требованиям Базельского комитета к использованию внутренних рейтингов [1].

Анализ исследований отечественных и зарубежных авторов позволяет говорить об эффективности применения процесса моделирования для формирования достоверных прогнозов на основе анализа исторических данных о развитии процессов и влияющих на них



факторов. В работе [2] представлен обзор основных моделей для оценки вероятности дефолта заемщика.

С целью построения моделей для прогнозирования вероятности дефолта заемщиков коммерческого банка в качестве признаков выбраны вероятности дефолта физических (Total RB) и юридических лиц (Total LE), рассчитанные по кредитным портфелям.

Известно [3-4], что учет макроэкономических переменных существенно улучшает качество модели, позволяет улучшить прогнозное качество вероятности дефолта. Макроэкономические показатели имеют огромное значение для принятия тех или иных решений в политической, социальной, экономической сферах, являются ориентиром для предпринимателей, отдельных граждан, зарубежных партнеров в их хозяйственной деятельности. Они отражают реальное положение экономики в стране, формируют представление о финансовом положении экономических субъектов. При моделировании дефолта заемщиков макроэкономические факторы характеризуют эффекты внешней среды [5], которые одинаково влияют на все группы заемщиков в определенное календарное время.

*Материалы и методы.* Одними из главных макроэкономических показателей являются валовой внутренний продукт (ВВП) и индекс потребительских цен (ИПЦ). В качестве макропеременных часто используют темп роста ВВП и уровень инфляции как опережающие индикаторы банковского кризиса. Также значимым показателем остаётся обменный курс белорусского рубля к доллару [6]. Информация о будущих значениях этих показателей может использоваться банками при анализе кредитного риска и оценке финансовой стабильности.

Целью настоящего исследования является построение моделей, способных давать качественный прогноз макроэкономических показателей (ВВП, ИПЦ, курс белорусского рубля к доллару), с целью применения их в дальнейшем для прогнозирования вероятности дефолта экономических субъектов. В качестве инструмента анализа и прогнозирования использовалась программная среда для выполнения статистических вычислений R.

С целью анализа и прогнозирования динамики выбранных показателей наряду с рядами исходных данных рассматривались также ряды, характеризующие их динамические характеристики – приросты. На рисунке 1 отражены изменения динамических характеристик исследуемых временных рядов макропоказателей.

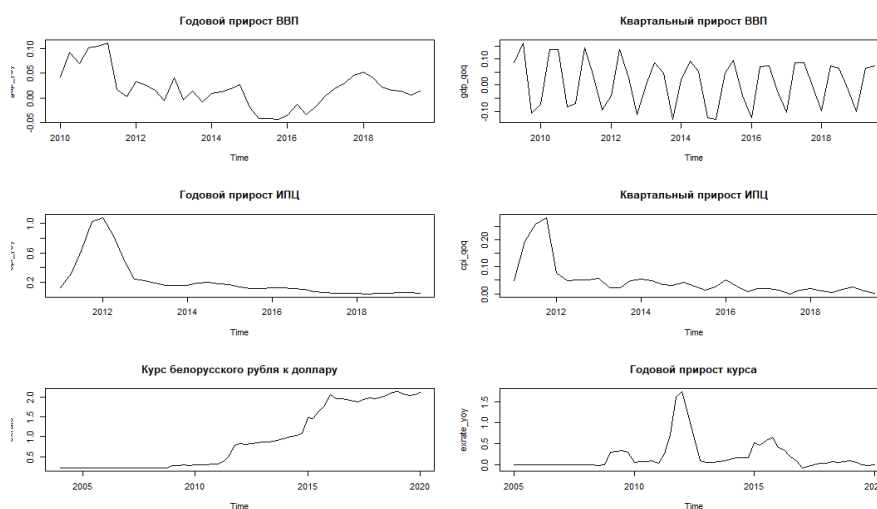


Рисунок 1. – Графическое представление динамики временных рядов макропоказателей ВВП, ИПЦ и курса белорусского рубля к доллару

Как видно из рисунка 1, квартальный прирост ВВП имеет выраженную сезонность. Динамика показателей годового и квартального приростов ИПЦ имеет убывающую тенденцию после 2012 года. Период до 2012 года исключим из рассмотрения, т.к. виден

сильный скачек в приросте показателя, что не характерно для общей динамики показателя и требует отдельного анализа. Годовой прирост курса белорусского рубля к доллару также имеет несколько выбросов, что может вызвать затруднения в построении модели.

Отличительной особенностью проводимого исследования является использование множественного корреляционно-регрессионного анализа. Поскольку показатели, отобранные в этой работе для построения моделей, в дальнейшем будут использоваться для построения модели анализа и прогнозирования вероятности дефолта экономических субъектов основе прогноза макропоказателей, определенный приоритет имеют ряды, которые в меньшей степени коррелируют друг с другом и в большей – с вероятностями дефолта. С целью отбора показателей рассчитаны коэффициенты корреляции между макропоказателями и вероятностями дефолта физических (Total RB) и юридических лиц (Total LE) (таблица 1). Расчет выполнен на основе квартальных данных о показателях ВВП (2011-2019гг.), индекса потребительских цен (2012-2019гг.), курса белорусского рубля к доллару (2004-2019 гг.), а также вероятностей дефолта портфеля юридических и физических лиц коммерческого банка.

В таблице 1 приняты следующие обозначения:

Total LE – вероятность дефолта портфеля юридических лиц;

Total RB – вероятность дефолта портфеля физических лиц;

GDP уоу – годовой прирост ВВП;

GDP қоқ – квартальный прирост ВВП;

CPI уоу – годовой прирост ИПЦ;

CPI қоқ – квартальный прирост ИПЦ;

ExRate – среднеквартальный курс белорусского рубля к доллару;

ExRate уоу – годовой прирост курса белорусского рубля к доллару;

ExRate қоқ – квартальный прирост курса белорусского рубля к доллару.

Таблица 1. – Коэффициенты корреляции между макроэкономическими показателями и показателями уровня дефолта портфеля юр. и физ. лиц

	Total LE	Total RB	GDP уоу	GDP қоқ	CPI уоу	CPI қоқ	ExRate	ExRate уоу	ExRate қоқ
Total LE	1								
Total RB	0.67	1							
GDP уоу	-0.42	-0.68	1						
GDP қоқ	-0.22	0.16	0.04	1					
CPI уоу	0.62	0.73	-0.87	-0.21	1				
CPI қоқ	0.29	0.26	-0.49	-0.72	0.67	1			
ExRate	-0.72	-0.83	0.53	0.02	-0.72	-0.33	1		
ExRate уоу	0.49	0.65	-0.81	-0.14	0.83	0.59	-0.66	1	
ExRate қоқ	0.34	0.23	-0.34	-0.48	0.56	0.62	-0.51	0.59	1

Из таблицы 1 можно видеть, что показатели Total LE и Total RB имеют наиболее тесную связь с GDP уоу, CPI уоу и ExRate. Однако факторы GDP уоу и CPI уоу сильно связаны (коэффициент корреляции равен -0.87), поэтому в дальнейшем при построении моделей регрессии для Total LE и Total RB для исключения мультиколлинеарности эти факторы не могут быть учтены в моделях одновременно. В связи с этим для прогнозирования выбраны годовой прирост ВВП, квартальный прирост ИПЦ и фактическое среднее значение курса белорусского рубля к доллару за квартал.

В качестве метода прогнозирования выбран класс моделей ARIMA [7]. Класс моделей авторегрессионного и скользящего среднего хорошо работает со стационарными временными рядами. Модели ARIMA(p,d,q)(P,D,Q)<sub>n</sub> можно передавать на вход нестационарной

ряд указывая параметры  $d$  (порядок интегрированности ряда) и  $D$  (порядок сезонных разностей). Рекомендуется не использовать модели с порядком  $d > 2$ . Если при  $d > 2$  ряд является нестационарным, есть необходимость взятия сезонных разностей. Иногда взятия сезонных разностей достаточно, чтобы ряд стал стационарным.

*Результаты.* Перед построением моделей проведем предварительный эконометрический анализ временных рядов на основе визуализации динамики показателей и поведения их автокорреляционных (ACF) и частных автокорреляционных (PACF) функций.

Годовой прирост ВВП показывает изменение показателя в процентах за год относительно выбранного периода в процентах. Динамика временного ряда ВВП, а также автокорреляционная и частичная автокорреляционная функции представлены на рисунке 2.

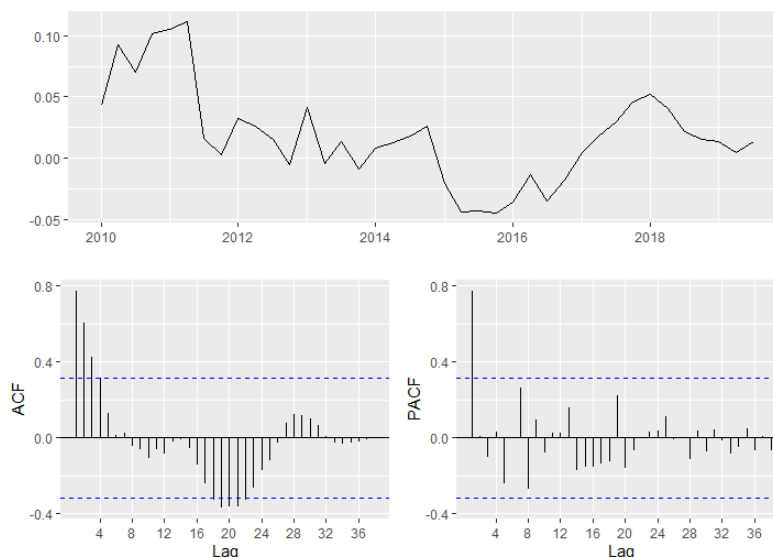


Рисунок 2. – Динамика временного ряда годового прироста ВВП и его автокорреляционные функции

На рисунке 2 можно заметить, что ряд приростов ВВП не является стационарным. Это также подтверждается статистическими тестами Дики-Фуллера ( $p$ -value = 0.58) и Квятковского – Филлипса – Шмидта – Шина ( $p$ -value = 0.084). Взятие первых разностей ряда приводит его к стационарному виду.

Стационарный временной ряд ВВП может подаваться на вход функции `arima()` в пакете R для построения модели. Лучшая модель для годового прироста ВВП – ARIMA(1,0,0)(0,1,1)<sub>4</sub>.

Коэффициенты модели являются значимыми. Остатки модели распределены нормально и не автокоррелированы, о чем говорят тесты Шапиро-Уилка (0.44) и Бройша-Годфри (0.33). Таким образом, модель считается адекватной и может быть использована для прогнозирования. Уравнение модели принимает вид:

$$\text{GDP}_{\text{р}oy_t} = \frac{0.7916}{8.294 \cdot 10^{-12}} \cdot \text{GDP}_{\text{р}oy_{t-1}} + \text{GDP}_{\text{р}oy_{t-4}} - \frac{0.7916}{8.294 \cdot 10^{-12}} \cdot \text{GDP}_{\text{р}oy_{t-5}} + \varepsilon_t + \frac{0.8382}{0.007453} \varepsilon_{t-4}. \quad (1)$$

Под коэффициентами модели (1) указаны  $p$ -значения. Полученная модель говорит о зависимости текущего значения от предыдущих с лагом 1, 4 и 5.

Динамика квартального прироста индекса потребительских цен и функции ACF, PACF отражены на рисунке 3.

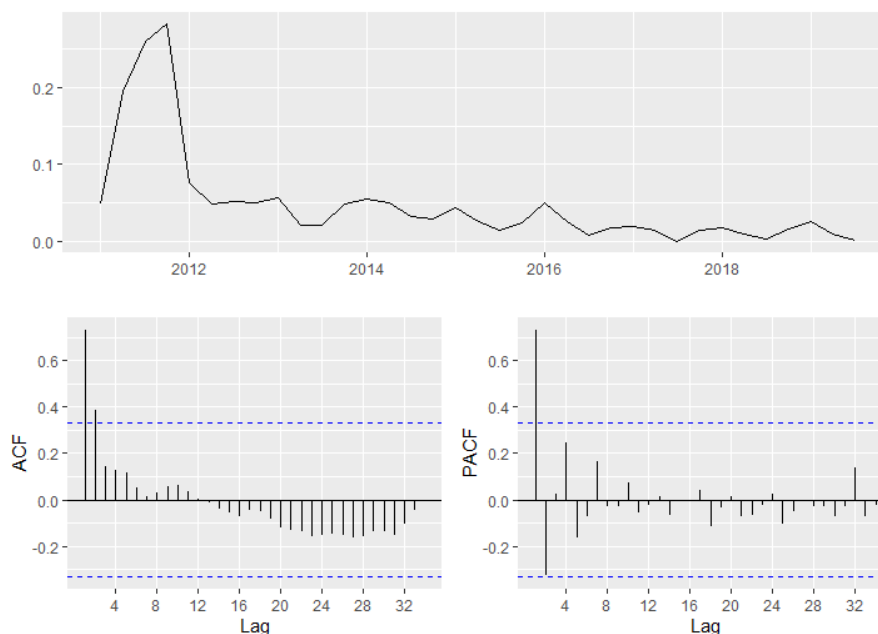


Рисунок 3. – Динамика временного ряда квартального прироста ИПЦ и его автокорреляционные функции

Временной ряд квартального прироста ИПЦ также не является стационарным. Для преобразования ряда к стационарному виду достаточно взятия первых разностей.

Лучшая модель для временного ряда ИПЦ –  $ARIMA(1,1,1)(1,0,0)_4$ .

Коэффициенты модели являются значимыми. Остатки модели распределены нормально и не автокоррелированы, о чем говорят тесты Шапиро-Уилка (0.52) и Бриша-Годфри (0.81). Таким образом, модель считается адекватной и может быть использована для прогнозирования. Запись модели в явном виде:

$$\begin{aligned}
 CPIqoq_t = & 1.5056CPIqoq_{t-1} - 0.5056CPIqoq_{t-2} - 0.8719CPIqoq_{t-4} + \\
 & + 0.8719CPIqoq_{t-5} - 0.4408CPIqoq_{t-6} + \varepsilon_t - 0.6257\varepsilon_{t-1}.
 \end{aligned} \quad (2)$$

В формуле (2) под коэффициентами указаны их р-значения. Текущее значение прироста ИПЦ зависит от прошлых значений с лагами 1,2,4 и 5.

Динамика курса белорусского рубля к доллару представлена на рисунке 4.

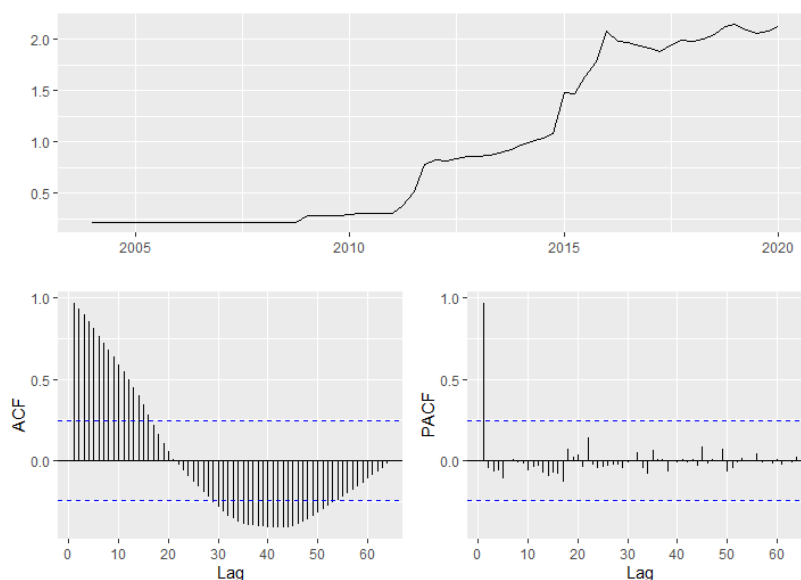


Рисунок 4. – Динамика временного ряда курса белорусского рубля к доллару и его автокорреляционные функции

Временной ряд имеет ярко выраженный тренд, а значит не может быть стационарным. Для преобразования его к стационарному виду необходимо взять первые разности. Функция пакета R Auto.arima() построила лучшую модель ARIMA(2,1,0)(0,1,1)<sub>4</sub>.

Остатки модели стационарны и не имеют автокорреляции, однако тест показал, что нормальное распределение отсутствует. Уравнение модели принимает вид:

$$ExRate_t = 1.1577ExRate_{t-1} + 0.0804ExRate_{t-2} - \frac{0.2381}{0.06334}ExRate_{t-3} + ExRate_{t-4} - 1.1577ExRate_{t-5} - 0.0804ExRate_{t-6} + \frac{0.2381}{0.06334}ExRate_{t-7} + \varepsilon_t + \frac{0.9012}{8.663 \cdot 10^{-8}}\varepsilon_{t-4}. \quad (3)$$

Прогнозные значения макроэкономических показателей представлены в таблице 2.

Таблица 2. – Прогноз макропоказателей на четыре периода.

Показатель	Период	Точечный прогноз	Нижняя граница прогнозного интервала	Верхняя граница прогнозного интервала
ВВП	4Q2019	1.50%	-0.50%	3.49%
	1Q2020	2.26%	-0.28%	4.80%
	2Q2020	2.10%	-0.72%	4.93%
	3Q2020	1.25%	-1.74%	4.24%
ИПЦ	4Q2019	1.28%	0.42%	2.15%
	1Q2020	1.94%	0.91%	2.96%
	2Q2020	0.92%	-0.18%	2.02%
	3Q2020	0.52%	-0.62%	1.66%
Курс бел.руб. к доллару	4Q2019	2.101724	2.04405	2.15939
	1Q2020	2.126378	2.03816	2.21460
	2Q2020	2.165767	2.04539	2.28615
	3Q2020	2.224655	2.07645	2.37286

Фактические и смоделированные временные ряды макропоказателей представлены на рисунке 5.

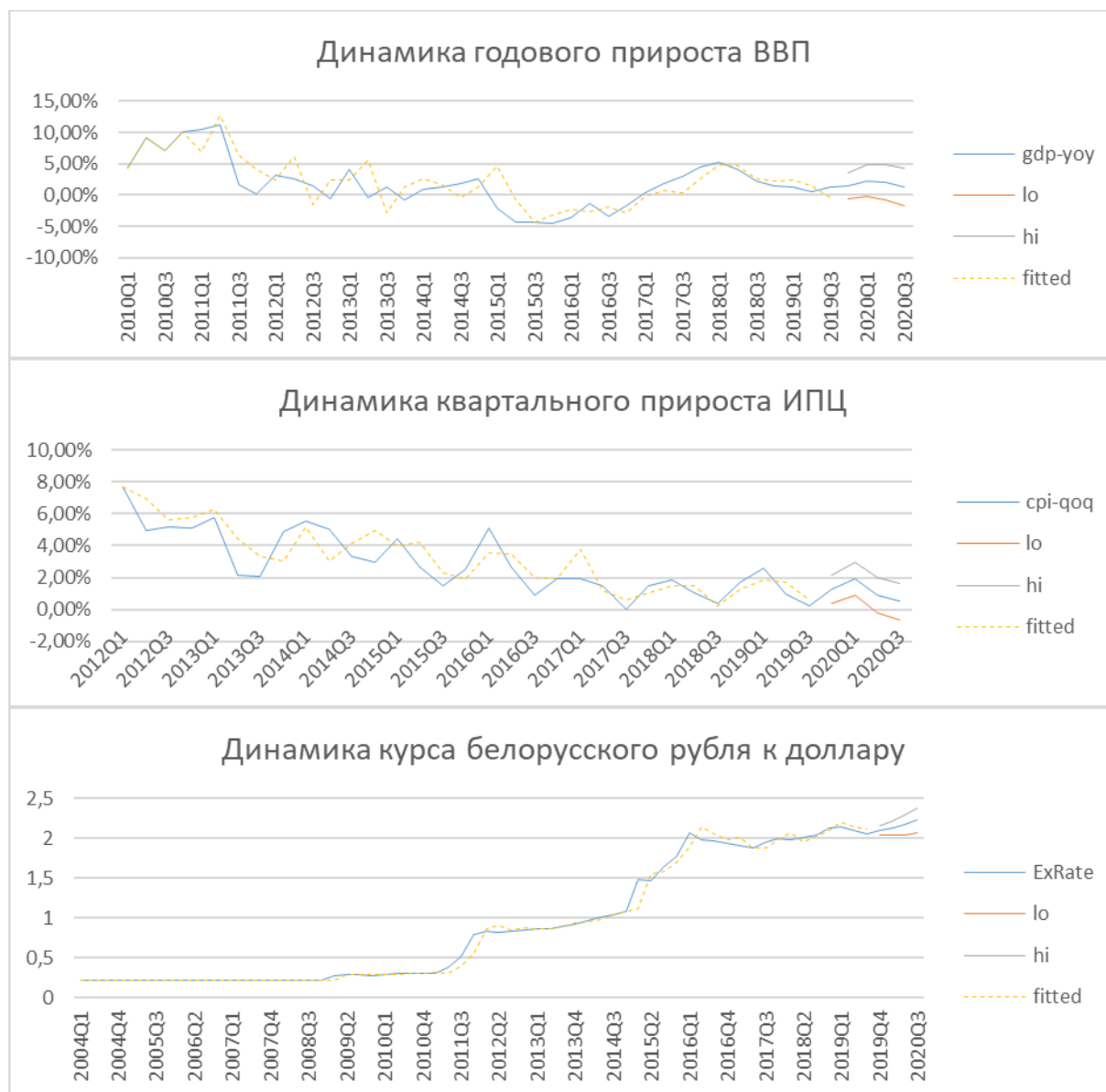


Рисунок 5. – Динамика фактического и смоделированного ряда

Точечный прогноз показывает, что динамика показателей сохранится и на будущие четыре периода. Прирост индекса потребительских цен имеет убывающую тенденцию, а курс белорусского рубля к доллару – возрастающую. Прогнозный интервал отражает степень уверенности в прогнозе.

Отметим, однако, что выводы остаются верными при сохранении условий развития анализируемых процессов. Прогнозные значения макропоказателей, рассчитанные на основе построенных в данной статье моделей, в целом согласуются с прогнозными значениями, рассчитанными Исследовательским центром ИПМ [8].

**Заключение.** Построенные модели способны давать качественный краткосрочный прогноз макроэкономических показателей ВВП, ИПЦ, курса белорусского рубля к доллару. Они могут использоваться банками при анализе кредитного риска, оценке финансовой

стабильности, а также с целью применения их в дальнейшем для прогнозирования вероятности дефолта заемщиков коммерческого банка.

#### **Список литературы**

- [1]. International Convergence of Capital Measurement and Capital Standards. – Basel Committee on Banking Supervision, 2004.
- [2]. Тотьмянина, К.М. Обзор моделей вероятности дефолта / К.М. Тотьмянина // Управление финансовыми рисками. – 2011. – №1. – С.12–24
- [3]. Карминский А.М., Пересецкий А.А., Петров А.Е. Рейтинги в экономике: методология и практика / под ред. А.М. Карминского. М.: Финансы и статистика, 2005.
- [4]. Карминский, А.М. Моделирование вероятности дефолта российских банков: расширенные возможности / А.В. Костров, А.М. Карминский // Журнал новой экономической ассоциации. – 2013. – №1(17). – С.64-86
- [5]. Яблонский, О.Л. Применение математических моделей для стресс-тестирования кредитных рисков / А.Н. Бузулуцкая, О.Л. Яблонский // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня: сб. материалов V Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 13–14 марта 2019 года). В 2 ч. Ч. 1/ редкол. : В.А. Богуш[и др.]. – Минск:БГУИР, 2019. – С. 331-336.
- [6]. Национальный статистический комитет Республики Беларусь [Электронный ресурс]. – Режим доступа: [https://www.belstat.gov.by/ofitsialnaya-statistika/ssrd-mvf\\_2/natsionalnaya-stranitsa-svodnyh-dannyh/](https://www.belstat.gov.by/ofitsialnaya-statistika/ssrd-mvf_2/natsionalnaya-stranitsa-svodnyh-dannyh/). – Дата доступа: 05.01.2020.
- [7]. Вох. G.E.P. Time Series Analysis: Forecasting and Control / G.M. Jenkins, G.C. Reinsel, G.E.P. Voh. – Prentice Hall, 1994. – 3rd edition. – 614p.
- [8]. Макроэкономический прогноз для Беларуси [Электронный ресурс] // <http://www.research.by/webroot/delivery/files/bro2020r1.pdf>. – Дата доступа: 15.02.2020

## **MODELING DYNAMICS OF MACROECONOMIC INDICATORS FOR FORECASTING THE PROBABILITY OF DEFAULT OF A COMMERCIAL BANK BORROWERS**

***H.I. Slauta,***

*Master student of GrSU,  
specialist PJSC “BNB-Bank”*

***V.B. Tsekhan,***

*PhD in Physico-mathematical  
Sciences, Head of department of  
Mathematical and information department.*

*Educational Institution “Yanka Kupala State University of Grodno”, Republic of Belarus  
PJSC “BNB-Bank”, Republic of Belarus  
E-mail: [anya.slauta@mail.ru](mailto:anya.slauta@mail.ru), [tsekhan@grsu.by](mailto:tsekhan@grsu.by)*

**Abstract.** Adequate econometric models have been built to predict macroeconomic indicators of GDP, the CPI, the Belarusian ruble against the dollar in order to use them in the future to predict the likelihood of default of economic entities. As a tool for analysis and forecasting, we used a software environment for performing statistical calculations of R.

**Keywords:** macroeconomic indicators, probability of default, borrowers, ARIMA, forecasting, R

УДК 004.421.6

## ФРЕЙМВОРКИ ЯЗЫКА ПРОГРАММИРОВАНИЯ PYTHON ДЛЯ РАСПОЗНАВАНИЯ ОБЪЕКТОВ В РЕЖИМЕ ФОТОГРАФИИ



**В.С. Маркевич**

Студент ГГУ имени Франциска Скорины



**Л.Н. Марченко**

Заведующий кафедрой фундаментальной и прикладной математики факультета математики и технологий программирования

Учреждение образования «Гомельский государственный университет имени Франциска Скорины», Республика Беларусь  
E-mail: [vladismark@gmail.com](mailto:vladismark@gmail.com), [lmarchenko@gsu.by](mailto:lmarchenko@gsu.by)

### **В.С. Маркевич**

Студент 3-го курса специальности «Экономическая кибернетика» факультета математики и технологий программирования

### **Л.Н. Марченко**

Заведующий кафедрой фундаментальной и прикладной математики факультета математики и технологий программирования, кандидат технических наук, доцент,

**Аннотация.** Нейронные сети открыли новые возможности в области распознавания образов. В связи с этим появляется возможность разработки новых методов и алгоритмов для анализа объектов, в частности, фотографий. В работе разработан алгоритм и приложение для распознавания объектов, в котором пользователи могут загружать фотографии и получать заданные готовые изображения с выделенными на нем объектами. Здесь использовался ряд фреймворков языка программирования Python для распознавания образов на фотографии, что в данный момент актуально.

**Ключевые слова:** нейронные сети, фреймворки, распознавание образов, режим фотографии, библиотеки языка программирования Python, искусственный интеллект.

**Введение.** Нейронные сети широко используются для распознавания образов. Существует множество алгоритмов, решающих конкретные задачи в этой сфере. Для распознавания объектов на фотографии обычно используются библиотеки NumPy, SciPy, H5Py, ImageAI, которые используют компьютерное зрение и облегчают создание алгоритмов.

Работа посвящена разработке приложения для распознавания объектов в режиме фотографии. Достижение данной цели предопределило создание алгоритма и кода на языке программирования Python для распознавания объектов, в котором пользователи смогут загружать фотографии и получать заданные готовые изображения с выделенными на нем объектами с использованием библиотек Numpy, SciPy, Pillow, Matplotlib, H5Py, Keras, ImageAI, Tensorflow, OpenCV, которые обладают более мощными возможностями по сравнению с их предшественниками. Приложение должно содержать: подключенные библиотеки и главную функцию их реализации.



*Материалы и методы.* Рассмотрим инструменты, технологии и другие решения, которые позволяют оптимальным образом удовлетворить описанные ранее требования обычно используются при решении таких задач [2].

Приложение для распознавания объектов в режиме фотографии реализовано на языке программирования Python, который широко применяется для разработки алгоритмов нейронных сетей. Для управления программными пакетами использовался менеджер пакетов `pip` [1]. Система управления пакетами – это набор программного обеспечения, который позволяет управлять процессом удаления, установки, настройки и обновления различных компонентов программного обеспечения. Программное обеспечение представлено в виде специальных пакетов, содержащих дистрибутив программного обеспечения, набор определённых метаданных, которые могут включать в себя имя, номер версии, описание пакета, имя разработчика, контрольную сумму, отношения с другими пакетами. Метаданные сохраняются в системной базе данных пакетов. Большинство дистрибутивов Python уже содержат `pip` [3]. Одно из главных преимуществ `pip` – это простота интерфейса командной строки, которая позволяет установить пакеты Python простой командой: `pip install some-package-name`, также легко удалять пакеты: `pip uninstall some-package-name`. `Pip` предоставляет возможность управлять всеми пакетами и их версиями с помощью файла `requirements.txt`.

`NumPy` – библиотека языка Python, предоставляющая поддержку больших многомерных массивов и матриц, и дающая большой спектр высокоуровневых математических функций. В результате любой алгоритм, который может быть выражен в виде последовательности операций над массивами и реализованный с использованием `NumPy`, работает достаточно эффективно. `SciPy` – это библиотека для языка программирования Python, созданная для выполнения научных и инженерных расчетов, которая решает задачи поиска экстремумов функций, вычисления интегралов функций, поддержку специальных функций, обработку сигналов, обработка изображений, работу с генетическими алгоритмами, решение обыкновенных дифференциальных уравнений и др [3]. Библиотека `SciPy` может взаимодействовать с `PyTables` – иерархической базой данных, разработанной для управления большими объемами данных.

Для визуализации результатов расчётов применялась библиотека `Python Imaging Library`, `Matplotlib`, `HippoDraw`, `Chaco`, `Biggles`, `MayaVi` [4].

Библиотека `PIL` (`Python Imaging Library`) предназначена для работы с растровой графикой. В разработанном проекте использовался форк `Pillow`. Возможности библиотеки – это поддержка бинарных, индексированных изображений; разных форматов, таких как: `BMP`, `ESP`, `GIF`, `JPEG`, `PDF`, `PNG`, `TIFF`; преобразование изображений одного формата в другой [5].

Библиотека `MatPlotLib` представляет собой пакет, предназначенный для визуализации данных, который вместе с `NumPy` и `SciPy` представляет возможности для решения задач технических вычислений. Данный пакет поддерживает многие виды графиков и диаграмм. Также с помощью `MatPlotLib` можно создавать анимированные изображения. Набор форматов изображений такой же, как и у библиотеки `PIL`.

`Keras` – это нейросетевая библиотека, предоставляющая настройку над фреймворком `TensorFlow` и направленная на работу с сетями глубокого обучения. `Keras` предоставляет высокоуровневый набор абстракций, благодаря которому проще формировать нейронные сети. Эта библиотека содержит многочисленные реализации блоков нейронных сетей.

`TensorFlow` – это программная библиотека для машинного обучения. Она создана для решения задач и построения тренировки нейронной сети с целью автоматического нахождения образов. `TensorFlow` является системой машинного обучения второго поколения. Данная библиотека работает на параллельных процессорах, как `GPU`, так и `CPU`. Вычисления происходят в виде потоков данных через граф состояний. `TensorFlow` хорошо подходит для автоматизированной аннотации изображений [7].

TensorFlow – это продолжение закрытой библиотеки DistBelief. Закрытая система машинного обучения DistBelief разрабатывалась Google Brain для внутренних проектов для работы с нейронными сетями и использовалась во многих проектах группы фирм холдинга Alphabet. После упрощения и оптимизации коды библиотеки увеличили надёжность и удобство и представляют собой TensorFlow [8].

Библиотека ImageAI – это библиотека Python, созданная для создания приложений и систем с автономными возможностями глубокого обучения и компьютерного зрения [9-11].

Библиотека OpenCV – это библиотека алгоритмов компьютерного зрения общего назначения, то есть – это набор данных, классов, функций для обработки изображения. Основные модули библиотеки: ядро, которое содержит основные данные и алгоритмы, базовые операции, матричную алгебру и многое другое; Highgui – модуль для ввода-вывода изображений; Svaux – экспериментальные функции для пространственного зрения; CV – модуль обработки изображения и компьютерного зрения, содержащий множество алгоритмов обработки изображений [12].

#### *Проектирование и реализация приложения «Распознавание объектов».*

Для работы нейронных сетей были подключены рассмотренные библиотеки для реализации алгоритмов распознавания объектов. Для правильной работы приложения алгоритм загружен в корень программы с помощью библиотеки ImageAI достать данный алгоритм и поместить его внутрь (рисунок 1).

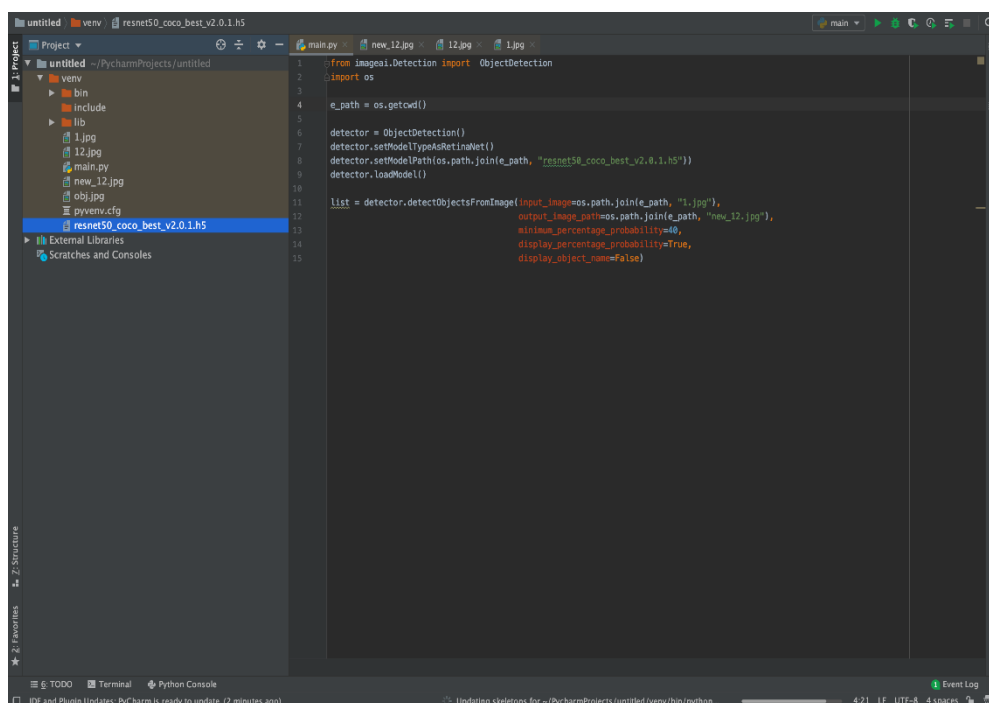


Рисунок 1. – Размещение алгоритма с помощью библиотеки ImageAI

Следует отметить, что алгоритм, который называется resnet50 не способен читаться в текстовых редакторах, он читается на бинарном уровне. Данный алгоритм несет в себе:

- разрезание фотографии на слои через матрицы;
- разбивание точек матрицы на кластеры;
- алгоритм проводит множество прямых через кластеры, тем самым придавая очертание образам, которые желает видеть программист;
- добавление весовых коэффициентов для более точного анализа данных;
- создание функциональных пирамид.

Разрезание фотографий состоит из 106-ти свёрточных слоев и лучше детектирует небольшие объекты. Основная особенность состоит в том, что на выходе есть три слоя, каждый из которых рассчитан на обнаружение объектов разного размера. На рисунке 2 приведен пример такого схематического устройства:

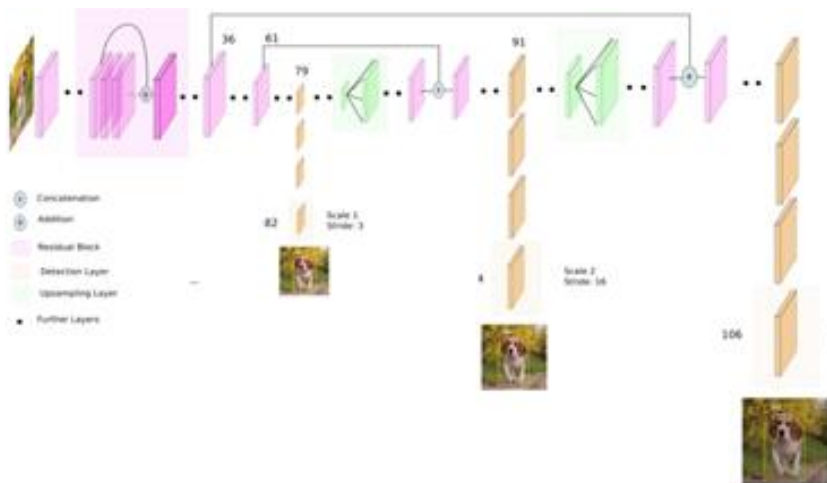


Рисунок 2. – Разрезание фотографий

Рассмотрим этапы работы программы. Загружаем фотографии в корень разработанной программы, называем их «12», «obj» и «1» (рисунок 3).

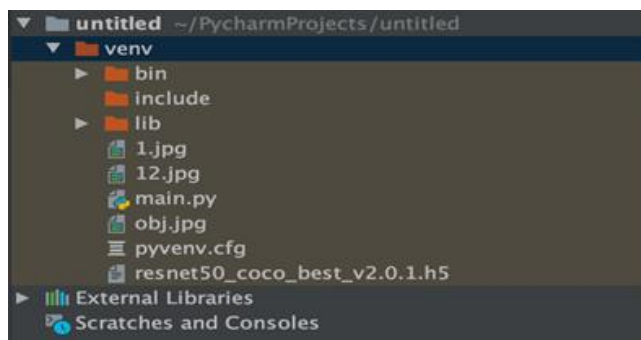


Рисунок 3. – Корень разработанной программы

Все подключенные библиотеки находятся в пакете lib. На рисунке 3 представлены описанные выше библиотеки, которые были задействованы при создании программы.

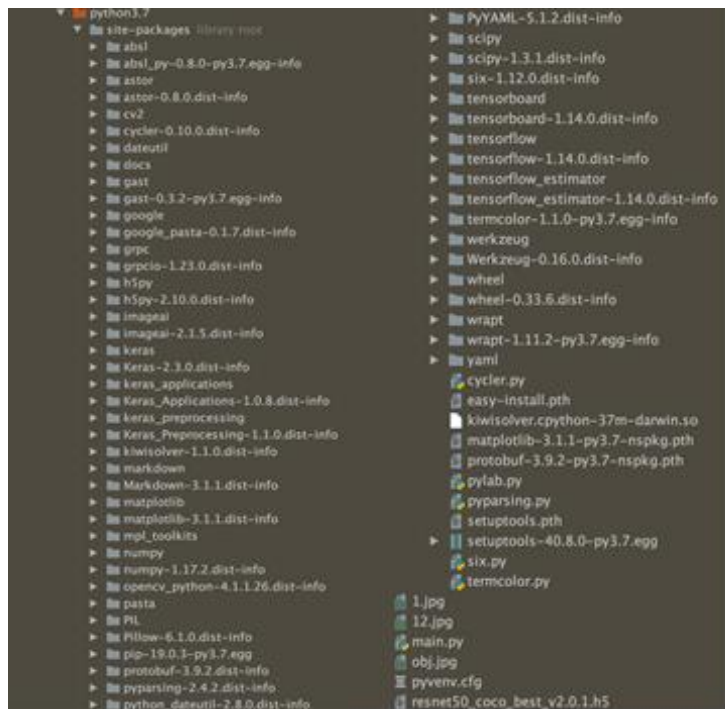


Рисунок 4. – Библиотеки, задействованные при создании программы

Главный класс называется main, который содержит главные методы, загружающие и обрабатывающие фотографии (рисунок 4. )

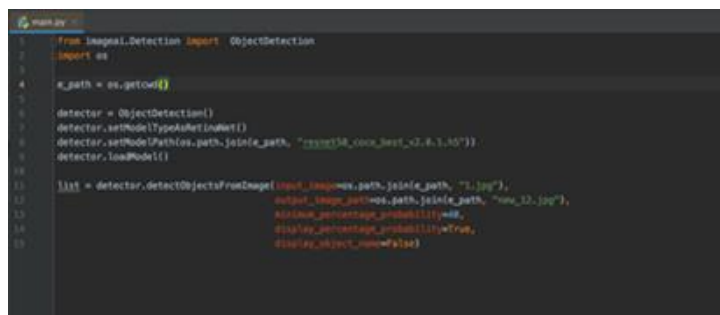


Рисунок 5. – Главный класс main

Когда приложение начинает обработку фотографии, то вначале оно включает все библиотеки, после загружает фотографию и алгоритм начинает «пробегать» по всем слоям (рисунок 5). В результате создается фотография, которая называется «new\_12».

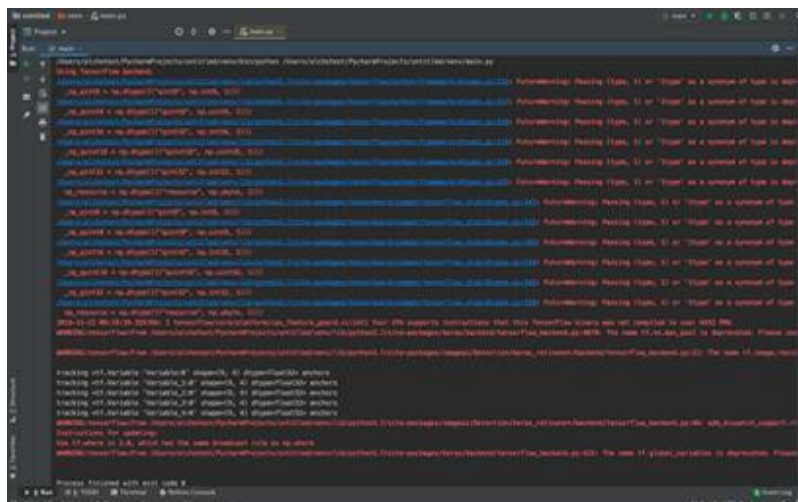


Рисунок 6. – Алгоритм работы программы

Для верификации приложения использовались личные фотографии. На рисунке 6 представлены фотографии до обработки и после обработки: Вероятность распознавания после обучения нейронной сети составила 60 %.

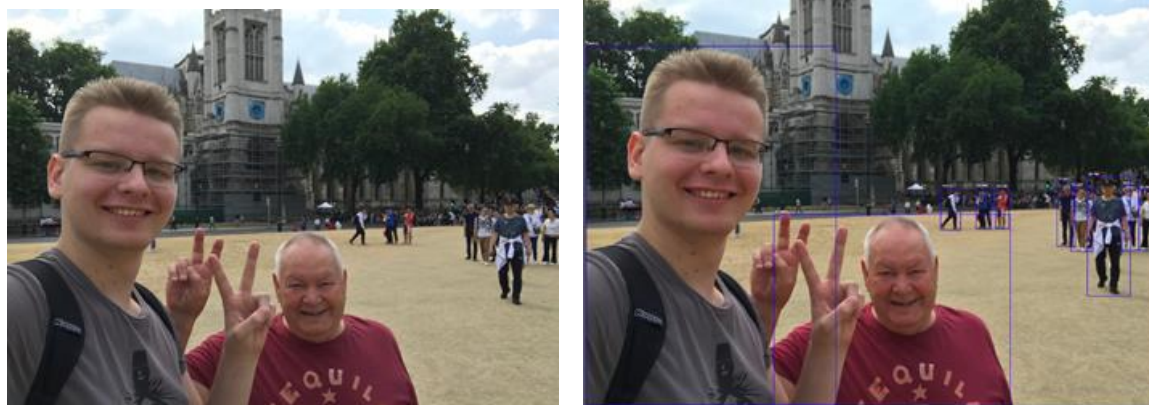


Рисунок 7. – Фотографии до обработки и после

**Заключение.** Распознавание объектов на основе нейронных сетей используется в различных сферах деятельности. Приложение распознавания объектов в режиме фотографии реализовано с помощью библиотек языка программирования Python: Numpy, SciPy, Pillow, Matplotlib, N5Py, Keras, ImageAI, Tensorflow, OpenCV. Приложение удобно и просто в использовании.

#### Список литературы

- [1.] Мюллер, А. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. Пер с англ. / А. Мюллер, С. Гвидо. – СПб.: ООО «Альфа-книга», 2017. – 480 с.
- [2.] Вандер Плас, Дж. Питон для сложных задач. Наука о данных и машинное обучение / Дж. Вандер Плас. – СПб.: Питер, 2019. – 576 с.
- [3.] Нуньес-Иглесиас, Х. Элегантный SciPy. Научное программирование на Python / Х. Нуньес-Иглесиас, Ш. Уолт, Х. Дэшноу. – М.: ДМК Пресс, 2018. – 266 с.
- [4.] Blanco-Silva, F.J. Learning SciPy for Numerical and Scientific Computing. – Packt Publishing, 2013. – 150 p.
- [5.] Ninad Sathaye. Python Multimedia. – Packt Publishing, 2010. – 292 p.
- [6.] Blanco-Silva, F.J. Mastering SciPy. – Packt Publishing, 2015. – 404 p.
- [7.] Джулли, А. Библиотека Keras / А. Джулли, С. Пал. – М.: ДМК Пресс, 2017. – 294 с.

[8.] Жерон, О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow. Концепции, инструменты и техники для создания интеллектуальных систем / О. Жерон. – М.: Вильямс, 2018. – 688 с.

[9.] Гудфеллоу, Я. Глубокое обучение. Пер с англ. А.А. Слинкина / Я. Гудфеллоу, И. Бенджио, А. Курвилль // 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.: цв. ил.

[10.] Шапиро, Л. Компьютерное зрение / Л. Шапиро, Дж. Стокман. – М., 2013. – 752 с.

[11.] Форсайт, Д. Компьютерное зрение. Современный подход / Д. Форсайт, Ж. Понс. – М.: Вильямс, 2004. – 928 с.

[12.] Bradski, G., Kaehler, A. Learning OpenCV 3: Computer Vision with the OpenCV library. - Packt Publishing: OReilly Media, 2008. – 580 p.

[13.] A - Markevich, V. Application development «Recognition of objects in the photo» on programming language Python. / V. Markevich // 8th Podlasie Conference on Mathematics 8PCM (5-8 December 2019, Bialystok). – Poland, 2019. – P. 100.

## **DEVELOPING THE OBJECT RECOGNITION APPLICATION IN PHOTO MODE**

*Francisk Skorina Gomel State University, Republic of Belarus*

**Abstract.** Neural networks have opened up new possibilities in the field of image recognition. In this regard, it is possible to develop new methods and algorithms for analyzing objects, in particular, photos. The work has developed an algorithm and an application for object recognition, in which users can upload photos and get specified ready-made images with objects highlighted on it. Here we used a number of Python programming language frameworks for image recognition in photos, which is currently relevant.

**Keywords:** neural networks, frameworks, pattern recognition, photography mode, Python programming language libraries, artificial intelligence.

УДК 004.4' 242 : 004.6-027.45

## АНАЛИЗ И ОЦЕНКА НАДЁЖНОСТИ ПРИКЛАДНЫХ КОМПЬЮТЕРНЫХ ПРОГРАММ



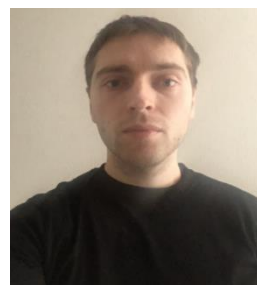
**С.М. Боровиков**  
Доцент кафедры ПИКС БГУИР,  
кандидат технических наук



**С.С. Дик**  
Аспирант кафедры ПИКС БГУИР



**В.Т. Лэ**  
Магистрант кафедры ПИКС БГУИР



**К.И. Клинов**  
Студент ФКП БГУИР

Белорусский государственный университет информатики и радиоэлектроники,  
Республика Беларусь  
E-mail: bsm@bsuir.by

### **С.М. Боровиков**

Доцент кафедры проектирования информационно-компьютерных систем БГУИР. Основная область научных интересов: прикладные математические методы в проектировании изделий радиоэлектроники, включая алгоритмы статистического прогнозирования надёжности изделий электронной техники и оценку надёжности прикладного программного обеспечения на ранних этапах его разработки. Руководитель разработки программных комплексов по автоматизированному расчёту и обеспечению надёжности электронных устройств: система АРИОН (2008-2009 гг.), система АРИОН-плюс (2011-2015 гг.).

### **С.С. Дик**

Окончил Белорусский государственный университет информатики и радиоэлектроники (2016 г.), в настоящее время является аспирантом этого университета, магистр технических наук. Работает в ООО «Itransition» в должности Project Manager. Проводит научные исследования по созданию и внедрению программного обеспечения в процессы подготовки спортсменов.

### **Ван Там Лэ**

Гражданин Вьетнама. Окончил Белорусский государственный университет информатики и радиоэлектроники (2018 г.), в настоящее время является магистрантом кафедры проектирования информационно-компьютерных систем этого университета. Выполняет магистерскую диссертацию по разработке методики, предназначенной для оценки надёжности прикладных программных средств на ранних этапах их проектирования.

### **К.И. Клинов**

Окончил филиал Белорусского национального технического университета «Борисовский государственный политехнический колледж» (факультет «Приборостроение»). В настоящее время является

*студентом заочной формы обучения Белорусского государственного университета информатики и радиоэлектроники. Работает разработчиком программного обеспечения в ЧП «Гузикс» (г. Минск).*

**Аннотация.** Развитие и совершенствование технологий Big Data обуславливают необходимость разработки новых прикладных компьютерных программ по обработке больших объемов данных. Заинтересованные специалисты хотели бы знать ожидаемый уровень надёжности компьютерных программ до начала написания их кода на языке программирования. В работе предложен возможный подход к оценке надёжности прикладных компьютерных программ, основываясь на прогнозном объеме будущей программы (числе строк кода) и области её применения. На основе рассмотрения экспериментальных данных об эксплуатационной надёжности используемых компьютерных программ предложена модель оценки ожидаемой надёжности планируемых к разработке компьютерных программ. Модель учитывает основные характеристики будущей компьютерной программы, квалификацию программистов, выполнение процедуры тестирования, а также основные факторы, определяющие эксплуатационную надёжность компьютерных программ: быстродействие используемых компьютеров, изменчивость входных данных и рабочую нагрузку, которую эксплуатационная среда будет оказывать на компьютерную программу.

**Ключевые слова:** прикладные компьютерные программы, модель надёжности, оценка надёжности до написания кода.

**Введение.** При создании аналитических систем для анализа больших данных (big data) возникает необходимость выполнения ряда этапов, среди которых важнейшими являются следующие:

- выбор целевой переменной (англоязычный вариант – target), описывающей эффективность деятельности компаний по организации и оказанию управленческих, образовательных, коммерческих, медицинских и других услуг, производству продукции;
- определение набора признаков (англоязычный вариант – features), позволяющих предсказывать (прогнозировать) значения целевой переменной с высокой достоверностью и минимальными последующими рисками для компании или организации;
- получение данных в виде значений признаков и целевой переменной; эти данные будут обрабатываться с помощью аналитической системы, а по результатам обработки будет прогнозироваться целевая переменная, которую можно будет использовать для оперативного принятия решений в соответствующей сфере деятельности людей;
- выбор алгоритма анализа данных для принятия решений;
- выбор имеющихся приложений или разработка специальных прикладных программ для обработки данных;
- оценка эффективности разработанной аналитической системы.

Отметим, что все указанные этапы создания аналитических систем важны и от успешного выполнения этапа зависит в конечном итоге эффективность создаваемой системы обработки больших данных. Например, подбор данных для обработки и алгоритмов анализа может стать не меньшей проблемой, чем другие этапы, так как отсутствует понимание, какие данные следует собирать и хранить, а какие можно игнорировать.

Некоторые возможные подходы к созданию аналитических систем обработки больших данных и прогнозирования целевой переменной были рассмотрены в [1, 2]. Со специальными методами и алгоритмами анализа больших данных можно ознакомиться в [3]. Используя рекомендации специалистов в области обработки больших данных, в конечном итоге можно определиться с данными и алгоритмами их обработки. Далее возникает следующая непростая задача: выбор приложений для обработки данных и/или разработка прикладного программного средства. Причём, используемые компьютерные программы должны отвечать высоким требованиям надёжности, поскольку надо быстро обрабатывать большие данные (big data), получать достоверные результаты и оперативно их использовать для управленческих и других действий.

В данной работе рассмотрены подходы к оценке надёжности прикладных программ на ранних этапах их разработки, до написания кода на языке программирования.



*Актуальность.* Аналитические системы, используемые для обработки больших данных относятся к классу сложных информационно-компьютерных систем, в которых вклад программного обеспечения в ненадёжность систем может составлять 40 и более процентов [4]. Для принятия решения о целесообразности создания и эффективности функционирования аналитических систем необходимо на ранних этапах их проектирования оценить ожидаемую надёжность разрабатываемого программного обеспечения. Методы оценки надёжности программного обеспечения, описанные в технической литературе, исходят из того, что написана компьютерная программа, устранены ошибки, обусловленные нарушениями правил языка программирования, и имеются определённые данные о тестировании программного средства (компьютерной программы и документации к ней). Но разработчиков аналитических систем и программного обеспечения к ним во многих случаях интересует прогнозный уровень надёжности компьютерных программ ещё до написания кода на языке программирования. Метод оценки надёжности компьютерных программ до написания их кода был предложен в [5, 6]. Метод использует статистические модели оценки ожидаемого числа возможных ошибок в компьютерной программе (некоторые специалисты используют термин «дефекты проектирования программы»). Однако метод нуждается в уточнении и адаптации к быстродействию используемых компьютеров, учёту изменчивости входных данных, а также рабочей нагрузки, которую будет воспринимать компьютерная программа со стороны эксплуатационной среды (ввод-вывод данных и нахождение этих операций в очереди и наличие состояний ожиданий, загрузка-выгрузка программы и/или её модулей из памяти и т.д.).

*Исходные положения и предпосылки.* В соответствии с ГОСТ 27.205-1997 [7] о прогнозной надёжности разрабатываемой компьютерной программы будем судить по интенсивности проявления ошибок ( $\lambda$ ), обусловленных дефектами её проектирования. Поскольку проявление ошибки при использовании программы приводит к невыполнению задачи пользователя (то есть равносильно отказу программы), то далее будем употреблять термин «интенсивность отказов программы». Использование  $\lambda$  в качестве характеристики надёжности программных средств удобно тем, что её можно оценивать по прогнозируемой плотности ошибок  $F$  в компьютерной программе [8]. Плотность ошибок  $F$  это отношение вида

$$F = \frac{\text{Количество возможных ошибок в компьютерной программе}}{\text{Число строк кода компьютерной программы}}.$$

Прогнозирование плотности отказов не требует знания среды исполнения компьютерной программы и, следовательно, подходит для ранних этапов разработки программного обеспечения. Когда информация о предполагаемой среде исполнения становится доступной, прогнозируемая плотность отказов может быть преобразована в прогнозируемую интенсивность отказов.

Вероятность безошибочного выполнения компьютерной программы ( $P$ ) в течение компьютерного (процессорного) времени  $t_{\text{комп}}$  обработки данных может быть определена в предположении экспоненциального распределения времени проявления ошибок:

$$P(t_{\text{комп}}) = \exp(-\lambda t_{\text{комп}}),$$

где  $\lambda$  – ожидаемая интенсивность отказов компьютерной программы; представляет собой среднее число появлений ошибок в единицу времени; в качестве единицы времени в данном случае следует принимать один час компьютерного (процессорного) времени.

Согласно техническому отчёту [8], для оценки ожидаемой интенсивности отказов компьютерной программы может быть использована модель

$$\lambda = K_{\text{тр}} \frac{N_{\text{ош}}}{L} = K_{\text{тр}} F, \quad (1)$$

где  $\lambda$  – ожидаемая интенсивность отказов компьютерной программы (размерность: 1/ч);  $K_{\text{тр}}$  – коэффициент, показывающий, как при использовании компьютерной программы по назначению скрытые дефекты проектирования программы количественно трансформируются (преобразуются) в интенсивность отказов (размерность: строка кода/ошибок в час);  $N_{\text{ош}}$  – число оставшихся ошибок в компьютерной программе;  $L$  – объём текста кода компьютерной программы в строках кода;  $F$  – ожидаемая плотность оставшихся ошибок в компьютерной программе (размерность: ошибок/строка кода).

В [8] приводятся экспериментальные данные, на основе которых получены значения коэффициента преобразования  $K_{\text{тр}}$  в зависимости от назначения, области применения компьютерной программы (таблица 1).

Таблица 1. – Экспериментальные данные о надёжности компьютерных программ

Область применения компьютерной программы	Число программ	Суммарное число строк кода	Средняя плотность ошибок $F$ на одну программу, ошибок/строка кода	Средняя эксплуатационная интенсивность отказов $\lambda$ , 1/ч	Значение $K_{\text{тр}}$ , строка кода/ошибок в час
1. Авиация	7	540617	0,013	0,08	6,28
2. Мониторинг и обеспечение безопасности	21	1793831	0,009	0,0109	1,2
3. Телекоммуникации, мобильные электронные устройства	5	88252	0,008	0,108	13,8
4. Управление производственными процессами	2	140090	0,002	0,0076	3,8
5. Автоматизированные системы управления (в различных областях и сферах деятельности людей)	12	2575427	0,009	0,198	23
6. Инструменты разработки программ, моделирование, испытательные стенды, обучение	6	193435	0,014	0,236*	16,9
Среднее	53	5331652	0,0094	0,1	10,6

Примечание. \* – получено пересчётом средней интенсивности отказов во время тестирования, составляющей 11,8 ч<sup>-1</sup>.

Значения коэффициентов преобразования  $K_{\text{тр}}$ , полученные по данным таблицы 1, приведены в последнем столбце таблицы. Эти значения рассчитаны на основе выражения (1) как  $K_{\text{тр}} = \lambda / F$ . Например, для строки «Среднее» коэффициент преобразования  $K_{\text{тр}}$  равен  $0,1 / 0,0094 \approx 10,6$ .

Приведённые в таблице 1 коэффициенты  $K_{\text{тр}}$  соответствуют случаю выполнения программ на компьютере с 32-разрядным процессором, быстродействие: 10 миллионов

операций в секунду (англоязычный вариант: 10 million instructions per second, кратко – 10 MIPS).

Механизм преобразования между плотностью ошибок и интенсивностью отказов программы основан на следующем. Оператор с ошибками не приведёт к отказу ни при каких обстоятельствах, пока он не будет выполнен, то есть до тех пор, пока он не повлияет на содержимое памяти или на управление компьютером. Будем предполагать, что в больших компьютерных программах (объём: тысячи–десятки тысяч строк кода) ошибки всегда существуют. Тогда на вероятность возникновения отказа влияют три характеристики среды выполнения программы:

- производительность компьютера (быстродействие, пропускная способность);
- изменчивость данных и состояний управления;
- рабочая нагрузка, которую эксплуатационная среда будет оказывать на компьютерную программу.

При прочих равных условиях программа, постоянно выполняющаяся на компьютере с большим быстродействием, будет показывать более высокую интенсивность отказов, чем та же программа, выполняемая на более медленном компьютере. Показатели отказов программ, выраженные с учётом компьютерных или процессорных часов, являются наиболее полезными показателями надёжности в данной эксплуатационной среде. Но следует не забывать, что подверженность программ проявлениям ошибок (отказам) зависит от числа прогонов, а не от течения времени. Таким образом, если один прогон программы с заданным набором данных занимает 1 секунду на компьютере *A* и 0,1 секунды на компьютере *B*, то проявление ошибки в единицу времени (интенсивность отказов) для программы, выполняемой на быстром компьютере (*B*), будет в десять раз больше, чем в случае её выполнения на медленном компьютере (*A*). При прочих равных условиях можно ожидать, что интенсивность отказов (выраженная в единицах общего времени) для компьютера *B* будет в десять раз больше, чем при запуске программы на компьютере *A*.

Неисправная программа, выполняемая даже на очень быстром компьютере, не будет иметь отказов программного обеспечения, если она постоянно работает с набором данных, который не приводил к появлению ошибок. С другой стороны, введение преднамеренной изменчивости во входные данные, как это делают при тестировании, ускорит возникновение отказов программы.

Рабочая нагрузка компьютерной системы влияет на частоту отказов программного обеспечения. Было обнаружено, что при очень высокой рабочей нагрузке интенсивность отказов может увеличиться более чем на порядок по сравнению с базовой, низкой рабочей нагрузкой программы. Экспериментально установлено, что между нагрузкой компьютерной программы и интенсивностью отказов имеет место очень тесная положительная корреляция [8]. Базовая концепция, лежащая в основе этого явления, заключается в том, что в условиях большой рабочей нагрузки встречаются более необычные ситуации (загрузка и выгрузка программы из памяти, нахождение операций ввода-вывода в очереди, наличие состояний ожидания и т.д.). И программист компьютерной программы, возможно, не предвидел все эти ситуации. Кроме того, системное программное обеспечение будет чаще выходить из строя при более частом его использовании.

*Результаты.* Интенсивность отказов компьютерной программы до начала этапа тестирования может быть определена на основе модели Муса по формуле [9]

$$\lambda_0 = CVN_{\text{нач}} = C \frac{R}{LE_R} N_{\text{нач}}, \quad (2)$$

где  $C$  – коэффициент пропорциональности;  $V$  – возможное непрерывное число прогонов компьютерной программы (команд программы) в течение одного часа;  $N_{\text{нач}}$  – начальное число ошибок компьютерной программы (до начала тестирования);  $R$  – производительность компьютера (быстродействие);  $L$  – объём компьютерной программы в строках кода;  $E_R$  – коэффициент расширения кода (увеличения числа команд программы), зависит от используемого языка программирования (для Assembler  $E_R = 1,0$ ; для Macro Assembler,  $E_R = 1,5$ ; для Си  $E_R = 2,5$ ; для Fortran, Cobal  $E_R = 3,0$ ; для Ada  $E_R = 4,5$ ; для Си++  $E_R = 6,0$ ).

Авторы работ в области надёжности программного обеспечения называют коэффициент пропорциональности  $C$  по-разному. В работах [9, 10] коэффициент  $C$  назван коэффициентом проявления ошибок программы и приводятся его примерные численные значения: от  $1,5 \cdot 10^{-7}$  до  $4 \cdot 10^{-7}$ . В работе [11] этот коэффициент называют постоянной, связанной с динамическим строением программы и её выполнением на различных платформах вычислительных систем. Указывается диапазон возможных значений от  $1,4 \cdot 10^{-7}$  до  $10,6 \cdot 10^{-7}$  и отмечается, что по умолчанию принимают значение  $C = 4,2 \cdot 10^{-7}$ . Размерность коэффициента  $C$ : 1/ошибка.

С учётом того, что для числа оставшихся в компьютерной программе ошибок  $N$  справедливо равенство

$$N = LF, \quad (2)$$

выражение (2) для  $\lambda_0$  можно представить в виде

$$\lambda_0 = C \frac{R}{E_R} F, \quad (3)$$

где  $F$  – ожидаемая плотность оставшихся ошибок в компьютерной программе до выполнения тестирования программного средства.

При выполнении тестирования компьютерной программы и устранения части скрытых ошибок уменьшается ожидаемая плотность оставшихся ошибок  $F$  выражения (3). С учётом этого при окончании тестирования значение  $\lambda_0$  трансформируется в эксплуатационную интенсивность отказов  $\lambda_{\text{эсп}}$ .

Коэффициент  $C$  обычно считают постоянным для конкретной компьютерной программы на всех этапах её разработки.

Из сравнения выражений (1) и (3) видно, что

$$K_{\text{тр}} = C \frac{R}{E_R}.$$

Используя данные таблицы 1, определим средние значения коэффициента пропорциональности  $C$  для компьютерных программ различных отраслей (областей применения). Будем считать, что программы были написаны на языке программирования Си, поэтому примем  $E_R = 2,5$ . Согласно [8], при выполнении (прогоне) программ использовались компьютеры с быстродействием процессора  $R = 10$  MIPS (million instructions per second – миллионов операций в секунду). Подсчитанные значения коэффициента  $C$  для разных областей применения компьютерных программ приведены в таблице 2.

Таблица 2. – Значения коэффициента пропорциональности  $C$

Область применения компьютерной программы	Авиация	Мониторинг и обеспечение безопасности	Телекоммуникации, мобильные устройства	Управление производственными процессами	Автоматизированные системы управления	Разработка программ, моделирование, обучение	Среднее
Коэффициент $C$ , 1/ошибка	4,36E-10	8,33E-11	9,58E-10	2,64E-10	1,60E-09	1,17E-09	7,36E-10
Коэффициент $K_{\Sigma}$	5,2	1,0	11,5	3,2	19,2	14,1	8,84

Значения коэффициента  $C$  оказались разными в зависимости от назначения компьютерной программы. Это является закономерным поскольку коэффициент  $C$  учитывает влияние на интенсивность отказов программ степени изменчивости входных данных и рабочую нагрузку на программу.

Представим коэффициент  $C$  в виде произведения

$$C = C_B \cdot K_{вх} \cdot K_n = C_B \cdot K_{\Sigma}, \quad (4)$$

а величины, входящие в выражение (4), будем называть как:  $C_B$  – базовое значение коэффициента проявления ошибок программы;  $K_{вх}$  – коэффициент увеличения интенсивности отказов, обусловленный изменчивостью входных данных;  $K_n$  – коэффициент увеличения интенсивности отказов, обусловленный рабочей нагрузкой на программу;  $K_{\Sigma}$  – суммарный коэффициент увеличения интенсивности отказов, обусловленный совместным действием изменчивости входных данных и рабочей нагрузки на компьютерную программу.

В качестве базового значения коэффициента проявления ошибок  $C_B$  примем значение коэффициента  $C$  для компьютерных программ, используемых для мониторинга и обеспечения безопасности. Пользуясь таблицей 2 (3-й столбец), можно записать:  $C_B = 8,33 \cdot 10^{-11}$  1/ошибка. Для программ этого назначения суммарный коэффициент увеличения интенсивности отказов принят  $K_{\Sigma} = 1,0$ .

Для определения ожидаемой надёжности компьютерной программы после выполнения её тестирования будем предполагать, что процедура тестирования выполняется общепринятыми методами, и при этом её суммарная продолжительность составляет не менее 40 процентов общего времени, отводимого на разработку компьютерной программы. Согласно работе [8], экспериментальные данные показали, что в таких случаях интенсивность отказов компьютерных программ за счёт выполнения тестирования уменьшается примерно в семь раз относительно начальной интенсивности отказов  $\lambda_0$ . Тогда модель для определения ожидаемой интенсивности отказов компьютерной программы, прошедшей тестирование, может быть представлена в виде

$$\lambda_{\text{экс}}^{(i)} = 0,14 C_B V N_{\text{нач}} K_{\Sigma}^{(i)}, \quad (5)$$

где  $\lambda_{\text{экс}}^{(i)}$  – ожидаемая (для  $i$ -й области применения, см. таблицу 2) эксплуатационная интенсивность отказов компьютерной программы, прошедшей тестирование;  $K_{\Sigma}^{(i)}$  – суммарный коэффициент увеличения интенсивности отказов для компьютерных программ  $i$ -й области применения (см. таблицу 2); множитель 0,14 – коэффициент, учитывающий

уменьшение интенсивности отказов в семь раз за счёт выполнения процедуры тестирования компьютерной программы ( $1/7 \approx 0,14$ ).

Прогнозное число возможных прогонов программы (команд программы) в течение одного часа определяется в соответствии с выражением (2):

$$V = \frac{R}{LE_R}.$$

Начальное число ошибок компьютерной программы (до начала тестирования)  $N_{\text{нач}}$  можно определить в соответствии с работой [2] как

$$N_{\text{нач}} = F \cdot L = A \cdot D \cdot S \cdot L, \quad (6)$$

где  $A$  – базовая плотность дефектов для компьютерных программ, используемых в данной отрасли;  $D$  – метрика, учитывающая специфику организации, разрабатывающей программное обеспечение, а также опыт и квалификацию программистов;  $S$  – метрика, учитывающая характеристики программного средства (категорию сложности, степень новизны, язык программирования, используемую операционную систему, компьютерную сеть, степень использования стандартных модулей).

Выбор и/или определение величин  $A$ ,  $D$ , и  $S$  формулы (6) рассмотрены в работе [2].

**Заключение.** Предложенный подход определения ожидаемой надёжности компьютерных программ основан на использовании статистических (экспериментальных) данных о свойствах используемых программ и даёт возможность приближённо оценить ожидаемую надёжность планируемого к разработке программного средства, в том числе используемого для анализа больших данных. Такой расчёт полезен, так как позволяет ещё до разработки программного средства получить представление о его ожидаемой надёжности и при необходимости разработать мероприятия по обеспечению требуемого уровня надёжности.

#### Список литературы

[1]. Batura M., Dzik S., Tsyrelchuk I., Borovikov S. Big Data Volumes and Some Approaches to the Creation of Corporate Analytical Systems / BIG DATA and Advanced Analytics. Использование BIG DATA для оптимизации бизнеса и информационных технологий: сб. материалов II Междунар. науч.-практ. конф. (Минск, Республика Беларусь, 15–17 июня 2016 года) / редкол. : М.П. Батура [и др.]. – Минск : БГУИР, 2016. – С. 74-80.

[2]. Боровиков, С. М. Большие данные и принципы разработки аналитических систем / С. М. Боровиков, С. К. Дик, С. С. Дик // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сборник материалов V Международной научно-практической конференции, Минск, 13–14 марта 2019 г. В 2 ч. Ч. 2 / Белорусский государственный университет информатики и радиоэлектроники; редкол. : В. А. Богуш [и др.]. – Минск, 2019. – С. 167-171.

[3]. Что такое Big Data (большие данные) в маркетинге: проблемы, алгоритмы, методы анализа [Электронный ресурс]. – Режим доступа : <http://lpgenerator.ru/blog/2015/11/17/chto-takoe-big-data-bolshie-dannye-v-marketinge-problemy-algoritmy-metody-analiza/> (дата обращения: 27.02.2020).

[4.] Боровиков, С. М. Метод оценки ожидаемой надёжности прикладных компьютерных программ систем медицинской электроники / С. М. Боровиков, С. С. Дик // Доклады БГУИР. – 2018. – № 7 (117). – С. 112-117.

[5]. Боровиков, С. М. Прогнозирование ожидаемой надёжности прикладных программных средств с использованием статистических моделей их безотказности / С. М. Боровиков, С. С. Дик // BIG DATA Advanced Analytics: collection of materials of the fourth international scientific and practical conference, Minsk, Belarus, May 3 – 4, 2018 / editorial board: M. Batura [etc.]. – Minsk, BSUIR, 2018. – P. 348 - 354.

[6]. Боровиков, С. М. Возможный подход к оценке надёжности прикладных программных средств для технологий Big Data / С. М. Боровиков, Лэ Ван Там, С. С. Дик // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сборник материалов V Международной научно-практической конференции, Минск, 13–14 марта 2019 г. В 2 ч. Ч. 2 / Белорусский государственный университет информатики и радиоэлектроники; редкол. : В. А. Богуш [и др.]. – Минск, 2019. – С. 77-83.

[7]. ГОСТ 27.205-97. Надёжность в технике. Проектная оценка надёжности сложных систем с учётом технического и программного обеспечения и оперативного персонала. Основные положения. – Минск : Госстандарт Республики Беларусь, 2005. – 22 с.

[8]. Software reliability, measurement and testing guidebook for software reliability measurement and testing: RL-TR-92-52, Vol II (of two) Final technical report April 1992/ Science Applications International Corp. (SAIC), Research Triangle Institute (RTI). Rome Laboratory Air Force Systems Command Griffiss Air Force Base NY 13441-5700.

[9]. Программирование и основы алгоритмизации. Исследование методов оценки и повышения надёжности программного обеспечения [Электронный ресурс]. – Режим доступа : <http://www.azovikdip.ru/index18.php> (дата обращения: 27.02.2020).

[10]. Модели надёжности программного и информационного обеспечения [Электронный ресурс]. – Режим доступа : <https://refdb.ru/look/2279745-pall.html> (дата обращения: 27.02.2020).

[11]. Модели оценки надёжности программных средств [Электронный ресурс]. – Режим доступа : <https://helpiks.org/4-73503.html> (дата обращения: 27.02.2020).

## RELIABILITY OF APPLIED SOFTWARE

**S.M. Borovikov,**

*PhD, associate professor of the department of Information and Computer Systems Design of the Belarusian State University of Informatics and Radioelectronics*

**S.S. Dick,**

*master of engineering, PG student of the Belarusian state university of informatics and Radioelectronics*

**Van Tam LE**

*master student of the department of Information and Computer Systems Design of the Belarusian State University of Informatics and Radioelectronics*

**K.I. Klinov**

*student of the Belarusian State University of Informatics and Radioelectronics*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus  
E-mail: bsm@bsuir.by*

**Abstract.** The development and improvement of Big Data technologies necessitates the development of new applied computer programs for processing large amounts of data. Interested specialists would like to know the expected level of reliability of computer programs before writing their code in a programming language. A possible approach to assessing the reliability of applied computer programs based on the predicted volume of the future program (number of lines of code) and the area of its application is proposed. Based on the consideration of experimental data on the operational reliability of the computer programs used, a model for evaluating the expected reliability of computer programs planned for development is proposed. The model takes into account the main characteristics of the future computer program, the qualifications of programmers, and the implementation of the testing procedure..

**Keywords:** applied computer programs, reliability model, reliability assessment before writing code.

УДК 004.93'1-021.131

## СИСТЕМА РАСПОЗНАВАНИЯ ОБЪЕКТОВ В ВИДЕОПОТОКЕ НА ОСНОВЕ ТЕХНОЛОГИЙ ВИРТУАЛИЗАЦИИ И ОБЛАЧНЫХ ВЫЧИСЛЕНИЙ



**Е.А. Пац**

*Магистрант БГУИР, Cloud Operations Engineer Netcracker Technology*



**Е.В. Насуро**

*Доцент кафедры электронных вычислительных машин БГУИР, кандидат технических наук*

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь  
E-mail: egor.pats@gmail.com*

### **Пац Е.А.**

*Окончил Белорусский государственный университет информатики и радиоэлектроники. Магистрант БГУИР. Работает в Netcracker Technology в должности Cloud Operations Engineer. Проводит научные исследования в области облачных технологий и DevOps методологий.*

### **Е.В. Насуро**

*Доцент кафедры электронных вычислительных машин БГУИР, основные направления деятельности: проектирование взаимодействия пользователя и сложных программно-аппаратных продуктов, научное руководство магистрантами.*

**Аннотация.** Облачные технологии – это перспективный и широко применяемый способ обработки и хранения данных. Использование технологий виртуализации позволяет повысить отказоустойчивость предоставляемых сервисов, уменьшить время доставки кода и улучшить утилизацию вычислительных ресурсов. Таким образом возникает потребность использования облачных технологий для решения различных задач. Одной из таких задач является обработка видеопотока. В работе предложен способ построения архитектуры облачной платформы для обработки большого количества потоков в режиме реального времени. Так же рассмотрены инструменты для конфигурации и развертывания системы.

**Ключевые слова:** Облачные технологии, виртуализация, IaaS, Terraform, IaC, configuration management systems, Chef, Puppet, Ansible, Salt, PaaS, Kubernetes.

**Введение.** В современном мире наблюдается очень быстрый рост технологий. Это связано с увеличением потребностей человека и желанием автоматизировать рутинные процессы. Вследствие чего, увеличивается количество сфер человеческой деятельности, где есть возможность применения информационных технологий.

На сегодняшний день одним из перспективных направлений информационных технологий являются облачные вычисления и предоставляемые ими современные сервисы создания, хранения, обработки и поиска данных.

Современные облачные технологии способны существенно сократить расходы, и сегодня многие компании все чаще переносят свои корпоративные системы и бизнес-приложения в облако. Данные системы позволяют в значительной мере снизить нагрузку на инфраструктуру, оптимизировать утилизацию вычислительных ресурсов, повысить доступность критически важных данных и сервисов, улучшить информационную



безопасность компонентов и систем. Перечисленные преимущества делают очевидным выбор облачных технологий большинством компаний.

Одним из наиболее удачных примеров использования облачных технологий является использование микросервисной архитектуры [1] программного обеспечения. Данная архитектура подразумевает использование отдельных сервисов для решения определенных задач. Таким образом убирается единая точка отказа приложения, что увеличивает доступность сервиса даже при отсутствии части функционала.

Зачастую при употреблении термина «видеопоток» подразумевается наличие видеокamеры, которая используется для его записи. Таким образом возникает ряд проблем, которые необходимо решить во время проектирования архитектуры системы. Проблема масштабируемости системы, необходимо предусмотреть возможность многократного увеличения обрабатываемых видеопотоков. Вторая, немаловажная проблема – это проблема изолированности данных. Под данным термином понимается, недоступность данных видеопотока сторонним процессам системы, в том числе и процессам, обрабатывающим соседние потоки данных.

Задача обнаружения и мониторинга объектов в видеопотоке на данный момент, не имеет общепринятых решений. Для анализа видеопоследовательности зачастую используются высококвалифицированные сотрудники, но с развитием технологий появляется возможность оптимизировать данный процесс. Этап обнаружения и распознавания объектов в видеопотоке решается с помощью автономных систем искусственного интеллекта.

При расширении количества обрабатываемых потоков данных необходимо позаботиться о распределении нагрузки, масштабируемости и отказоустойчивости всех ключевых компонентов сервиса, данные задачи позволяет решить разработанная облачная система [2-3]. Зачастую, при развертывании программного обеспечения, отсутствует возможность запуска каждого экземпляра на отдельном, изолированном сервере. Одной из особенностей сложных систем выступает утилизация ресурсов. На данный момент, разработка подобных систем очень актуальна. Большинство компаний, которые предоставляют программное обеспечение как сервис, хотят получить максимальную доступность и минимальное время отклика предоставляемого сервиса. Таким образом, система, разрабатываемая в рамках диссертации, находится на пике популярности, так как отвечает всем основным принципам высоконагруженных, отказоустойчивых, масштабируемых систем.

Стандартные способы предоставления приложений являются дорогостоящими, так же для предоставления доступа к ресурсу в формате 24/7 необходимо экспертное сопровождение серверов, на которых размещено приложение. Данный подход постепенно перестает использоваться большинством компаний из-за его сложности и дороговизны. В качестве решения данной проблемы были предложены технологии виртуализации, которые позволяют собрать необходимую виртуальную инфраструктуру, которая будет соответствовать бизнес требованиям компаний. Такой подход к созданию и использованию инфраструктуры получил название – облачные технологии.

Так как востребованность облачных технологий значительно выросла за последние годы, большинство крупных компаний разработали методы предоставления виртуальной инфраструктуры как сервиса (IaaS) [4]. Таким образом появилась возможность переноса критически важных инфраструктурных решений в облачную платформу с гарантированной доступностью более 99,9% [5].

**Материалы и методы.** Большинство современных облачных провайдеров имеют более одного центра обработки данных. Самыми крупными компаниями предоставляющими виртуальную инфраструктуру являются Google (GCP), Amazon (AWS), Microsoft (Azure). Данные компании имеют центры обработки данных, расположенные на разных континентах и в разных странах. Каждый провайдер облачных технологий стремится увеличить время доступности инфраструктуры для конечного потребителя. Таким образом для сравнения лидеров, предоставляющих возможность создания и поддержания виртуальной инфраструктуры, был создан рейтинг, для которого используются две линейные прогрессивные экспертные шкалы: полнота видения (completeness of vision) и способность реализации (ability to execute). На сегодняшний день компания Amazon с облачным решением под названием Amazon Web Services занимает лидирующую позицию по доступности инфраструктурных решений (рисунок 1).



Рисунок 1. – Оценка качества предоставляемых облачных услуг на основе доступности инфраструктурных решений

Большинство современных приложений и бизнес ресурсов имеют особые требования к компьютерной безопасности. У многих компаний есть желание перенести свои инфраструктурные решения в облачную платформу, но по причинам безопасности, либо иным бизнес требованиям, нет желания использовать публичные облака и переводить данные и сервисы на сервера облачных провайдеров. Для таких случаев были разработаны облачные платформы для частного использования. Наиболее популярной частной облачной платформой является OpenStack.

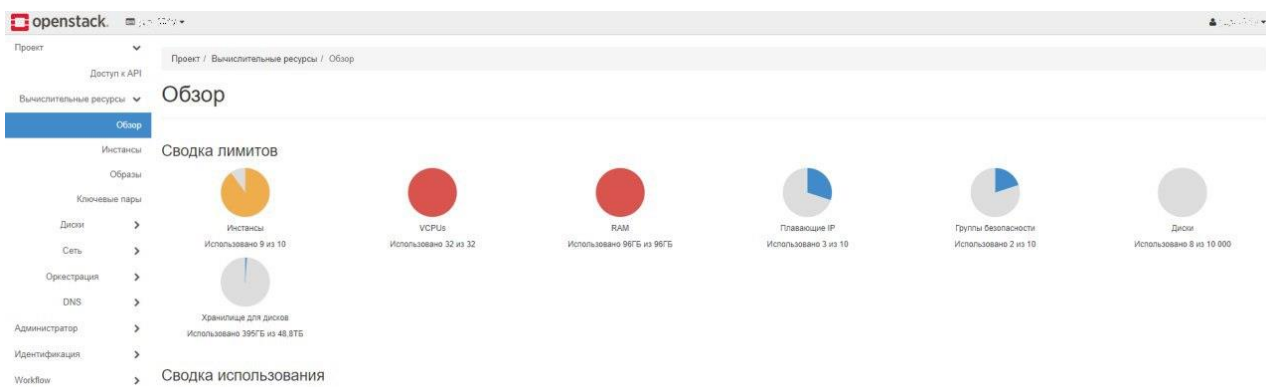


Рисунок 2. – Веб-интерфейс облачной платформы OpenStack

Облачные платформы позволяют объединять сервера с большими вычислительными мощностями в кластер, что означает, предоставление мощностей нескольких серверов, как единого пула ресурсов. Таким образом, предоставляется возможность, спроектировать архитектуру оптимальным образом, что значительно снижает расходы на аренду или обслуживание железных серверов. Использование данной технологии позволяет использовать один сервер с множеством ролей. Количество ролей зависит от количества виртуальных машин, запущенных на данном сервере.

Так как все вычислительные ресурсы в облаке составляют единый пул, это позволяет гибко масштабировать инфраструктуру. Таким образом, если планируется временная нагрузка на систему, существует возможность добавить к кластеру еще несколько вычислительных узлов для распределения нагрузки. В то же время при выходе одного из серверов из строя, виртуальные машины будут автоматически перенесены на свободные вычислительные узлы. Что позволяет быть уверенными в доступности инфраструктуры.

Большое количество облачных провайдеров создает конкуренцию в предоставлении данной услуги. Каждая из облачных платформ имеет отдельный интерфейс пользователя. Используя данный интерфейс пользователю предоставляется возможность в создании и контроле инфраструктуры решений в режиме реального времени. Но, зачастую, архитектура инфраструктуры разрабатывается заранее, просчитывается количество необходимых узлов, процессорного времени, оперативной памяти, дискового пространства и так далее. Для конфигурации определенной инфраструктуры, использование пользовательского интерфейса, значительно замедляет время развертывания и настройки основных компонентов. Таким образом большинство провайдеров виртуальной инфраструктуры используют API (Application Program Interface).

Использование API облачных платформ значительно ускоряют создание необходимых компонентов инфраструктуры. Для оптимизации данного процесса был разработан инструмент конфигурации под названием Terraform (рисунок 2). Terraform позиционируется как IaS (Infrastructure as code), позволяя описывать виртуальную инфраструктуру кодом. Так же, используя данный инструмент, появляется возможность поддержания и конфигурации инфраструктуры. Terraform является одним из ключевых инструментов для эффективного развертывания и обновления инфраструктурных решений. Таким образом, Terraform отлично подходит для реализации DevOps методологий.

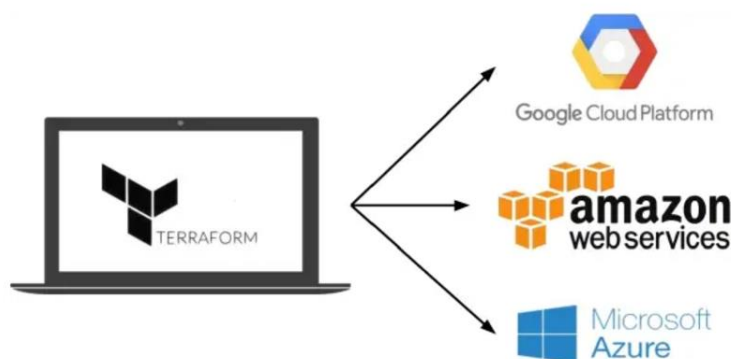


Рисунок 3. – Возможность использования Terraform, для разворачивания виртуальной инфраструктуры, используя различные облачные провайдеры

При развертывании инфраструктуры, используя Terraform (рисунок 3), появляется возможность сохранения состояния ее компонентов. Таким образом, при изменении одного или нескольких компонентов, инструмент конфигурации применит необходимые настройки только для определенных структурных частей инфраструктуры.

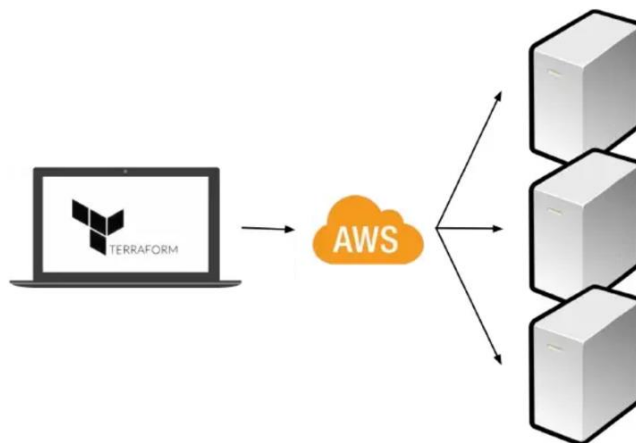


Рисунок 4. – Получение виртуальной инфраструктуры используя Terraform с API облачного провайдера AWS

В рамках научной работы был разработан Terraform модуль, позволяющий развернуть инфраструктуру для дальнейшей установки системы оркестрации контейнеров под названием Kubernetes.

После разворачивания виртуальной инфраструктуры, появляется необходимость в ее настройке. Одной из основных концепций DevOps является разделение инфраструктуры на «домашних животных» и «крупного рогатого скота». Разница этих подходов в моделях обслуживания. В случае с «домашними животными» каждый сервер критически важен, и отказ любого из серверов может вывести из строя всю систему. Обычно размер такой инфраструктуры достаточно мал, что позволяет не прибегать к инструментам автоматизации и конфигурации. При использовании данной модели обслуживания все действия выполняются вручную. Совершенно другой подход к обслуживанию инфраструктуры применяется при модели «крупного рогатого скота». В такой модели размер кластера может достигать до сотен вычислительных узлов, и отказ одного или нескольких серверов никак не отразится на состоянии предоставляемого сервиса. Но настройка таких систем вручную невозможна, таким образом возникает потребность применения систем управления конфигурациями.

Системы управления конфигурациями представляют из себя программное обеспечение, позволяющее получить доступ одновременно к большому количеству вычислительных узлов, с целью применения определенных настроек. Так как с развитием облачных технологий применение данных систем становится все более актуальным, количество данных систем так же увеличивается. Каждая из систем управления конфигурациями разработана для определенных задач и имеет различный порог вхождения и поддержания разработанных решений. К наиболее актуальным можно отнести четыре системы, такие как Chef, Puppet, Ansible, Salt.

Chef – система управления конфигурациями, клиентская часть написана на языке программирования Ruby, серверная часть разработана на Erlang. Система используется для упрощения и автоматизации процессов настройки и конфигурации большого количества серверов (рисунок 4), имеет возможность интеграции с различными облачными решениями. Для использования Chef на вычислительном узле должен быть установлен клиентская часть данной системы с необходимыми правами, что означает, что система не имеет возможности работать в Push-режиме, то есть управлять конфигурациями неподготовленных серверов. Код, написанный для данной системы, называется «рецепт». Рецепты представляют из себя описание состояния вычислительного узла, включая установленные пакеты, запущенные службы, созданные файлы. Chef проверяет каждый из ресурсов инфраструктуры и поддерживает их в эталонном состоянии.

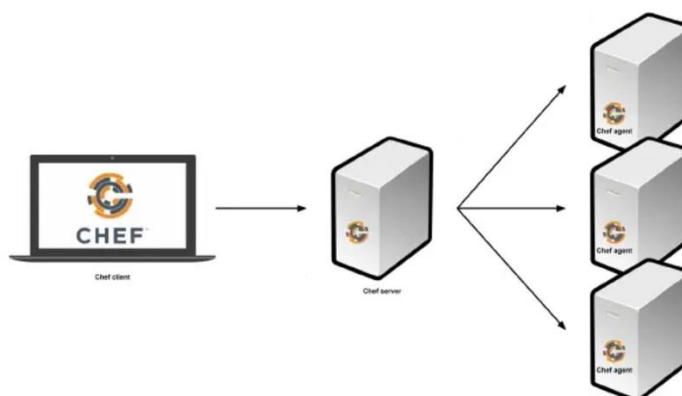


Рисунок 5. – Схема доставки конфигураций системой Chef

Puppet – кроссплатформенная система управления конфигурациями, работающая в режиме клиент-сервер. Система позволяет централизованно управлять настройками операционных систем большого количества вычислительных узлов, включая запущенное на них программное обеспечение. Puppet постоянно опрашивает узлы кластера и поддерживает их в актуальном состоянии. Наравне с Chef, Puppet получил большое распространение и используется большинством компаний на поддержания инфраструктурных решений. Puppet так же не поддерживает Push-режим, таким образом, вычислительные узлы нуждаются в первоначальной подготовке для использования данной системы управления конфигурациями.

Ansible – система управления конфигурациями написанная на языке Python. Данная система имеет достаточно низкий порог вхождения для профессионального использования. Код конфигурации, пишется используя формат `yaml`, и описывает задачи, которые должны быть выполнены на удаленных серверах. Ansible единственный на данный момент инструмент конфигурации, который не требует клиентского приложения на управляемых вычислительных узлах. Это означает, что данная система работает исключительно в Push-режиме. Подключение к серверам, осуществляется используя SSH. Таким образом требуется чтобы пользователь, от имени которого идет подключения обладал необходимыми правами и состоял в нужных группах.

Для получения списка узлов Ansible должен получить на вход инвентарный файл. В данном файле описывается инфраструктура и способы подключения к каждому из серверов. Так же присутствует возможность выделять группы серверов и определять переменные для определенных групп. Ansible при подключении к серверу, первоначально получает полную информацию о его состоянии, после чего переходит к конфигурации. Если определенное действие уже было выполнено, инструмент конфигурации не будет выполнять его повторно. Все эти достоинства выделяют Ansible из списка самых популярных систем управления конфигурациями.

SaltStack – система управления конфигурациями и удаленного выполнения команд. Данное программное обеспечение, так же, как и Ansible, написано на языке программирования Python. Данная система, наравне с Chef и Puppet, требует предварительной установки агентов на вычислительные узлы. Salt Master является центральной службой к которой подключаются клиенты – Salt Minion. Модули состояния систем описываются в формате `yaml`.

Таблица 1. – Сравнение основных параметров систем управления конфигурациями

	Source	Cloud	Type	Infrastructure	Language	Architecture	Community	Maturity
<b>Chef</b>	Open	All	Config Mgmt	Mutable	Procedural	Client/Server	Large	High
<b>Puppet</b>	Open	All	Config Mgmt	Mutable	Declarative	Client/Server	Large	High
<b>Ansible</b>	Open	All	Config Mgmt	Mutable	Procedural	Client-only	Large	Medium
<b>Salt</b>	Open	All	Config Mgmt	Mutable	Declarative	Client/Server	Medium	Medium

*Результаты.* Для выбора инструмента управления конфигурациями был проведен анализ самых популярных систем. Исходя из таблицы 1 был можно сделать вывод, что оптимальным инструментом, для решения поставленной задачи, является Ansible. Данная система управления конфигурациями имеет открытый исходный код, что означает возможность ее бесплатного использования. Ansible имеет большое сообщество, это означает, что большое количество тривиальных задач уже имеют оттестированные решения. Одним из ключевых преимуществ использования Ansible является отсутствие программ-агентов. Для конфигурации узлов достаточно иметь доступ к ним. Написание кода для конфигурации системы не вызывает затруднений, так как формат yaml интуитивно понятен.

В рамках работы над магистерской диссертацией был разработан Ansible playbook для подготовки всех узлов кластера, и последующего разворачивания системы оркестрации docker контейнеров – Kubernetes.

В работе были проанализированы основные провайдеры облачной инфраструктуры, был проведен их сравнительный анализ и принято архитектурное решение. Особое внимание было уделено типам и принципам построения инфраструктурных решений, были рассмотрены основные методы и способы поддержки кластеров. Разработаны модули автоматического развертывания, согласно концепции DevOps. Был проведен разбор инструментов автоматического управления конфигурациями.

#### Список литературы

- [1]. Ричардсон К. Микросервисы. Паттерны разработки и рефакторинга – Санкт-Петербург: Издательский дом «Питер», 2020. – 544 с.
- [2]. Yevgeniy Brikman Terraform: Up & Running: Writing Infrastructure as Code – O’Reilly Media, 2019 – 319 с.
- [3]. Lorin Hochstein Ansible: Up and Running – O’Reilly Media, 2015 – 299 с.
- [4]. Сайфан Джиджи Осваиваем Kubernetes. Оркестрация контейнерных архитектур. – СПб.: Питер, 2019. – 400 с.
- [5]. Gartner Report: Magic quadrant for Cloud Infrastructure as a Service [Электронный ресурс] – 2020. – Режим доступа: <https://pages.awscloud.com/Gartner-Magic-Quadrant-for-Infrastructure-as-a-Service-Worldwide.html>. – Дата доступа: 25.02.2020

## **SYSTEM OF RECOGNITION OF OBJECTS IN VIDEO FLOW ON THE BASIS OF VIRTUALIZATION TECHNOLOGIES AND CLOUD COMPUTING**

***E.A. Pats***

*undergraduate student of the BSUIR,  
Cloud Operations Engineer Netcracker  
Technology*

***K.V. Nasuro***

*assistant professor at BSUIR,  
Candidate of Technical Sciences*

*Belarusian state university of informatics and radio electronics, Republic of Belarus  
E-mail: egor.pats@gmail.com*

**Abstract.** Cloud technologies is a promising and widely used way to process and store data. Using virtualization technologies can increase the fault tolerance of the services provided, reduce the time of code delivery and improve the utilization of computing resources. Thus, there is need to use cloud technology to solve various problems. Video stream processing is one of such tasks. A method for building a cloud platform architecture for processing a large number of streams in real time is proposed. The tools for configuring and deploying the system were discussed.

**Keywords.** Cloud technologies, virtualization technologies, IaaS, Terraform, IaC, configuration management systems, Chef, Puppet, Ansible, Salt, PaaS, Kubernetes.

УДК 608.2

## ТЕХНОЛОГИИ ГЛУБОКОЙ НЕЙРОННОЙ СЕТИ МНОГОМАСШТАБНОГО ДЕТЕКТИРОВАНИЯ ЛИЦ



**А.В. Суша**  
Аспирант кафедры ЭВС, БГУИР



**М.И. Вашкевич**  
Доцент кафедры ЭВС БГУИР,  
к.т.н., доцент

Белорусский государственный университет информатики и радиоэлектроники  
Минск, Республика Беларусь  
E-mail: isushik94@bsuir.by, vashkevich@bsuir.by

### **А.В. Суша**

Окончил Белорусский государственный университет информатики и радиоэлектроники. Аспирант кафедры ЭВС, БГУИР, магистр технических наук.

### **М.И. Вашкевич**

Окончил Белорусский государственный университет информатики и радиоэлектроники (БГУИР) по специальности «Электронные вычислительные средства» в 2008 г. В 2013 г. защитил кандидатскую диссертацию по специальности 05.13.05 «Элементы и устройства вычислительной техники и систем управления». Работает доцентом кафедры ЭВС в БГУИР. Научные интересы: обработка сигналов, методы машинного обучения.

**Аннотация.** Целью настоящей работы являлось проектирование глубокой искусственной нейронной сети для детектирования лиц. Основное внимание при проектировании было уделено обеспечению высокой производительности и уменьшению требуемых вычислительных затрат за счет: 1) факторизации операции свертки; 2) применения точечных сверток; 3) комбинирования поканальных и точечных сверток. Разработанный детектор сравнивался со схожими детекторами лиц, полученными на основе широко распространенных архитектур нейронных сетей MobileNet и NasNet. Предложенная архитектура детектора лиц имеет вычислительную сложность 5.1 MFLOPs, что в два раза меньше, чем у MobileNet (11.7 MFLOPs) и в четыре раза меньше, чем у NasNet (22 MFLOPs). Соответственно время детектирования на изображении 416×416 составило 5.12 мс (или 195 FPS) с видеокарты GeForce 1080 Ti, а также 65.4 мс (или 15 FPS) на одном ядре процессора Intel Core i7-8700K. При этом точность нашей архитектуры равна 85% и уступает MobileNet лишь на 4%, а NasNet – на 9.5%.

**Ключевые слова:** детектирование лиц, глубокие нейронные сети, сверточные нейронные сети

**Введение.** В настоящее время разработано множество способов детектирования лиц, большинство из которых основано на парадигме скользящего окна и применении классических методов компьютерного зрения и машинного обучения (например, метод Виолы-Джонса на базе каскадов Хаара и HOG+SVM). Преимуществом этих подходов является высокая скорость детектирования, что позволяет их использовать в режиме реального времени. Существенным недостатком данных методов является большое число ложных срабатываний и слабая устойчивость к множеству факторов, затрудняющих детектирование: разный масштаб лиц, частичное перекрытие лиц другими объектами, ракурс, выражения лиц, засветка и прочее [1]. Применение глубоких искусственных нейронных сетей (далее – ИНС) для детектирования лиц на изображении позволяет повысить точность и достичь большую



инвариантность к указанным факторам. Однако, подход, использующий скользящее окно в связке с ИНС, имеет большую вычислительную сложность в сравнении с классическими методами. Для уменьшения сложности нейросетевых детекторов лиц и, в общем случае, объектов, были предложены новые архитектуры ИНС.

На смену парадигме скользящего окна пришла *парадигма двухэтапного детектирования*. Известным примером реализации этой парадигмы является детектор объектов R-CNN [2], который впоследствии был несколько раз модифицирован для повышения точности и скорости детектирования. Суть этого метода заключается в выделении на первом этапе т. н. *регионов интереса* (сегментация изображения на регионы объекта или фона), которые сокращают число кандидатов для следующего этапа классификации регионов. Такой подход позволил понизить вычислительную сложность детектора, так как сократилось число запусков «тяжелого» классификатора.

Следующей парадигмой является *парадигма одноэтапного детектирования*, которая возникла в стремлении ускорить детектирование объектов посредством ИНС. В некотором смысле она является переосмыслением парадигмы скользящего окна со стороны ИНС. Детектирование осуществляется за один вычислительный проход ИНС, в котором осуществляется одновременная классификация всех ячеек (окон) равного размера, на которые разбивается изображение. Реализацией этой парадигмы явился метод YOLO (You Only Look Once) [3, 4], преимущество которого заключается в том, что изображение анализируется сверточной ИНС только один раз. Однако у такого подхода имеется один существенный недостаток: он проводит детектирование объектов одного масштаба. Алгоритм многомасштабного детектирования предложен в работе SSD (Single Shot multibox Detector) [5]. Он был усовершенствован в методе FPN (Feature Pyramid Network) [6] путем передачи признаков одного масштаба на другой для осуществления классификации.

Целью настоящей работы является проектирование архитектуры ИНС для многомасштабного детектирования лиц методами, позволяющими снизить требования к памяти и вычислительным ресурсам. При этом уменьшение точности детектирования должно быть пропорционально меньшим, чем уменьшение параметров и операций в ИНС.

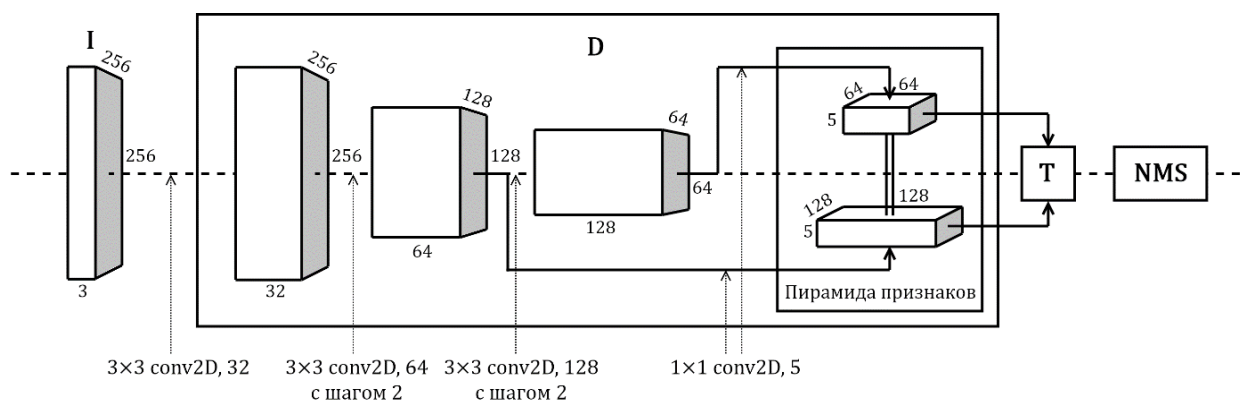


Рисунок 1. – Структура детектора лиц

*Описание архитектуры детектора лиц.* На рисунке 1 показана обобщенная структура детектора лиц, реализующего парадигму одноэтапного детектирования. Изображение  $I$  подается на сверточную ИНС  $D$ , последний слой которой генерирует карту детектируемых лиц. Сама ИНС построена по принципу сети с пирамидой признаков [6]. Пирамида признаков позволяет проводить детектирование лиц разного размера за один проход ИНС путем генерирования карт разного размера. После вычисления всех карт детектируемых лиц осуществляется формирование списка рамок, найденных на изображении лиц (операция  $T$  на

рисунке 1). Затем для этого списка применяется алгоритм подавления немаксимумов (NMS – non maximum suppression) [7], который устраняет дубликаты и схожие рамки из списка.

В общем виде процесс детектирования можно записать следующими выражениями:

$$\begin{aligned} \{[B_{i_r j_r}]_r\} &= D(I), \quad i_r = 1, \dots, N_r, \quad j_r = 1, \dots, M_r, \quad r = \{1, 2, 3\}, \\ \{B_{r, v_r, q_r}\} &= T(\{[B_{i_r j_r}]_r\}), \quad v_r \in 1, \dots, N_r, \quad q_r \in 1, \dots, M_r, \\ \{B_k\} &= NMS(\{B_{r, v_r, q_r}\}), \quad k = 1, \dots, |\{B_{r, v_r, q_r}\}|, \end{aligned}$$

где  $I$  – входное изображение;  $D(\cdot)$  – оператор, выполняющий детектирование лиц;  $r$  – номер карты детектируемых лиц;  $N_r$  и  $M_r$  – число строк и столбцов карты детектируемых лиц под номером  $r$ ;  $B_{i_r j_r}$  – значение карты детектируемых лиц под номером  $r$  в  $i_r$ -й строке и  $j_r$ -м столбце;  $T(\cdot)$  – операция преобразования карты детектируемых лиц в список рамок лиц, при этом осуществляется фильтрация рамок на основе вероятности обнаружения лица;  $B_{r, v_r, q_r}$  – элемент списка найденных лиц, соответствующий в  $v_r$ -й строке и  $q_r$ -м столбце карты под номером  $r$ ;  $B_k$  – элемент списка оставшихся лиц после применения  $NMS$ , соответствующий в  $k$ -му элементу из списка найденных лиц.

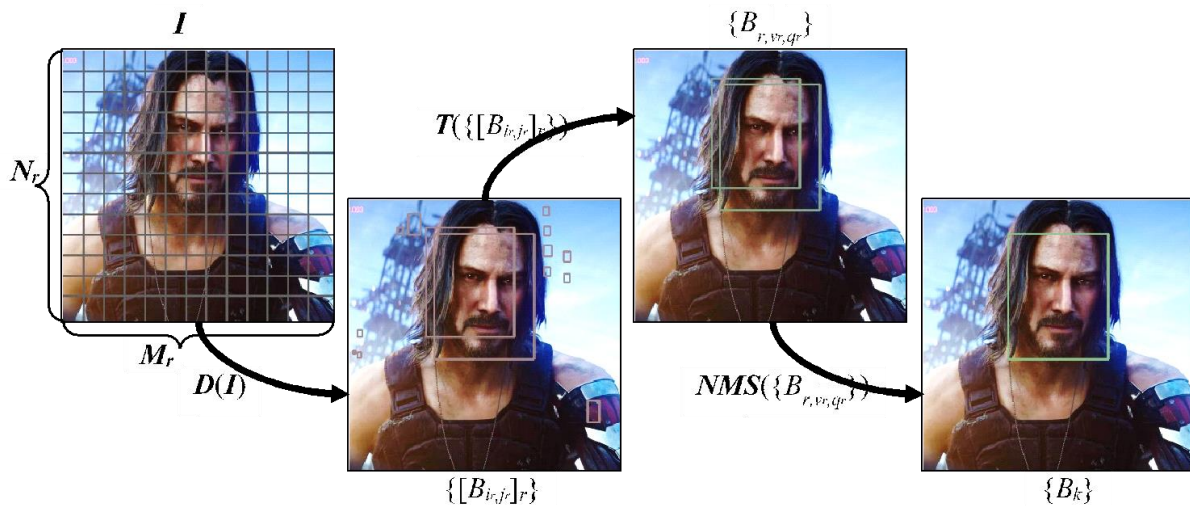


Рисунок 2. – Пример процесса детектирования лица

Процесс детектирования лица на примере поэтапной обработки одного изображения представлен на рисунке 2. На первом этапе ИНС генерирует карту размером  $N_r \times M_r$ , где  $r \in \{1, 2, 3\}$ , элементы которой отвечают за некоторый небольшой участок на входном изображении. Если разбить входное изображение на ячейки, формирующие сетку  $N_r \times M_r$ , то каждая ячейка будет соответствовать одному элементу карты. Элементом  $B_{i_r j_r}$  в  $i_r$ -й строке и  $j_r$ -м столбце карты под номером  $r$  является 5-мерный вектор, который имеет следующий формат:

$$B_{i_r j_r} = [p, x, y, w, h]^T, \quad (1)$$

где  $p$  – вероятность обнаружения лица,  $p = \sigma(l_p)$ , где  $l_p$  – 1-й линейный выход сети в ячейке,  $\sigma(\cdot)$  – функция логистического сигмоида,  $x, y$  – относительное смещение центра рамки лица в границах ячейки, при этом:  $x = \sigma(l_x)$  и  $y = \sigma(l_y)$ , где  $l_x$  и  $l_y$  – 2-й и 3-й линейные выходы сети в ячейке;  $w, h$  – нормированная ширина и высота рамки относительно размера

ячейки, при этом:  $w = \exp(l_w)$  и  $h = \exp(l_h)$ , где  $l_w$  и  $l_h$  – 4-й и 5-й линейные выходы сети в ячейке.

*Оптимизация архитектуры ИНС.* Анализ научных статей позволил выделить следующие методы повышения производительности ИНС, суть которых заключается в уменьшении числа выполняемых машинных операций в различных типовых блоках:

– первую операцию свертки необходимо производить с некоторым шагом [8], например шаг окна равный 2, что в 4 раза уменьшает число операций, проводимых в этом слое;

– регулирование числа каналов признаков операцией точечной свертки [9];

– операции свертки можно факторизовать на несколько последовательных операций с меньшим числом параметров [10]. К примеру, свертка с окном  $3 \times 3$  выполняется двумя свертками  $3 \times 1$  и  $1 \times 3$ , что уменьшает число параметров сети, а также число операций умножений с накоплением;

– традиционная операция свертки может быть заменены на последовательные операции поканальной и точечной свертки [11], что уменьшает число параметров с  $n = c \times h \times w \times k$  на  $m = c \times h \times w + c \times k$ , где  $c$  – число входных каналов,  $k$  – число выходных каналов,  $h \times w$  – размер ядра свертки.

Пусть  $W$  – тензор параметров операции свертки;  $x$  – некоторый трехмерный тензор;  $i$  и  $j$  – индексы строки и столбца тензора соответственно;  $m$  и  $n$  – индексы канала входного тензора и канала выходного тензора соответственно;  $s_r$  и  $s_c$  – шаги выборки элемента тензора по строкам и по столбцам;  $K$  и  $L$  – высота и ширина окна свертки;  $M$  – число окон (каналов). Тогда математически операции свертки можно записать следующим образом:

$$\begin{aligned} conv2d(x, W, s_r, s_c) &= [y_{i,j,n}] = \left[ \sum_{k,l,m}^{K,L,M} x_{(s_r \cdot i + k - \frac{K}{2}, s_c \cdot j + l - \frac{L}{2}, m)} \cdot W_{k,l,m,n} \right], \\ dwConv2d(x, W, s_r, s_c) &= [y_{i,j,m}] = \left[ \sum_{k,l}^{K,L} x_{(s_r \cdot i + k - \frac{K}{2}, s_c \cdot j + l - \frac{L}{2}, m)} \cdot W_{k,l,m} \right], \\ pwConv2d(x, W, s_r, s_c) &= [y_{i,j,n}] = \left[ \sum_m^M x_{(s_r \cdot i, s_c \cdot j, m)} \cdot W_{m,n} \right], \end{aligned}$$

где  $conv2d$  – операция свертки,  $dwConv2d$  – поканальная операция свертки и  $pwConv2d$  – точечная операция свертки.

Для симулирования поведения обычной свертки с помощью поканальных сверток применяется замыкающая точечная свертка, что математически записывается следующим образом:

$$sepConv2d(x, W_{dw}, W_{pw}, s_r, s_c) = pwConv2d(dwConv2d(x, W_{dw}, s_r, s_c), W_{pw}, 1, 1),$$

где  $W_{dw}$  – тензор параметров поканальной свертки и  $W_{pw}$  – тензор параметров точечной свертки.

В работе, описывающей VGG-16 [12] показано, что использование нескольких последовательных сверток  $3 \times 3$  (например, две или три) имеют то же рецептивное поле как и одна свертка  $5 \times 5$  или  $7 \times 7$ . При этом такая последовательность имеет меньшее число параметров, что уменьшает требования к памяти и число вычислительных операций. Более того, применение нелинейной функции активации после каждой свертки в последовательности позволяет улучшить аппроксимационные свойства ИНС, так как в работе

[13] было показано, что многослойная ИНС с нелинейными функциями активации является универсальным аппроксиматором.

При проектировании архитектуры ИНС для детектирования лиц было решено провести факторизацию свертки  $3 \times 3$  на две свертки  $3 \times 1$  и  $1 \times 3$ . Таким образом уменьшилось затраты памяти на для хранения параметров сети на треть. Дополнительно после каждой новой свертки применяется функция активации  $ReLU(x) = \max(x, 0)$  [14]. Выбор этой функции активации обусловлен тем, что она требует выполнения только одной операции: сравнения с нулем – которая сочетает в себе простоту линейной функции, а также необходимое свойство нелинейности. Также, помимо высокой скорости вычислений при прямом проходе, эта функция активации имеет простую функцию производной первого порядка, что позволяет обучать ИНС быстрее.

*Описание архитектуры многомасштабного детектора лиц.* Для простоты описания предлагаемой архитектуры введем понятие блока свертки. Блок свертки – последовательные операции свертки, которые являются факторизацией некоторой одной операции свертки с тем же рецептивным полем. На рисунке 3 показаны два варианта блока свертки (*block* и *sepBlock*), применяемые в предлагаемой архитектуре сверточной ИНС. Здесь и далее операции свертки записываются в следующем формате: размер окна и/или число выходных каналов, название операции и в скобках аргументы операции, входной тензор и тензор параметров опускаются.

Следующей структурной единицей является операция *bottleneck* (дословно – «бутылочное горлышко») [15]. Суть операции заключается в преднамеренном понижении или повышении числа каналов карт активаций внутри. Эта операция включает в себя блок свертки, определенный ранее. Соответственно применяется два варианта этой операции: *block bottleneck* и *sepBlock bottleneck*.

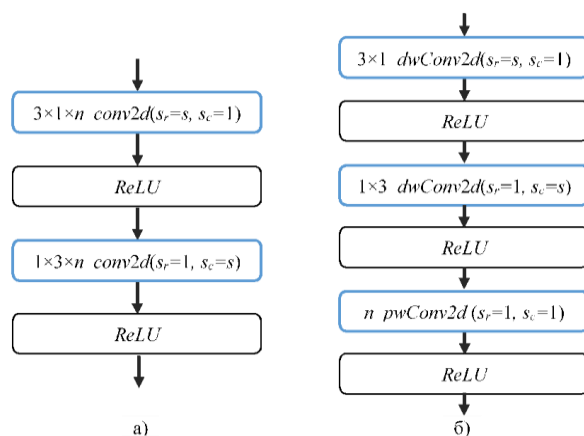


Рисунок 3. – Схема применяемых блоков: а) – *block*, б) – *sepBlock*; где  $n$  – число выходных каналов;  $s$  – шаг выборки окна свертки

Рисунок 4 содержит схему операции *bottleneck* в общем виде, где *batch normalization* – операция нормализации по обучающему мини-пакету [16]. Также применяется остаточная связь (англ. – residual connection) [15] для случая, когда ширина и высота тензора на входе блока и на выходе одинаковы. При разном числе каналов применяется операция точечной свертки для выравнивания числа каналов входного тензора.

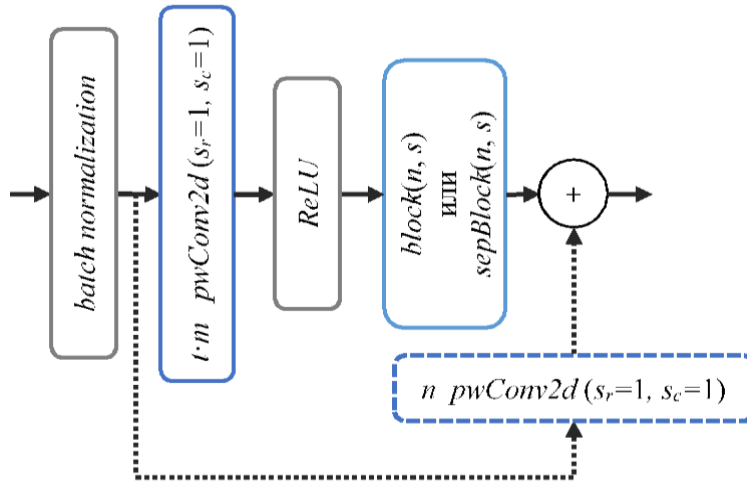


Рисунок 4. – Схема операции *bottleneck*; где  $m$  – число входных каналов;  $n$  – число выходных каналов;  $t$  – коэффициент расширения/сжатия числа каналов

На рисунке 5 приведена архитектура спроектированной сверточной ИНС. Выражением вида « $\times N$ » обозначается число повторений операции в блоке. Также на схеме имеется дополнительная операция *upsample  $m \times n$* , которая выполняет пространственное увеличение размера карты признаков методом ближайшего соседа, где  $m$  – число повторов элемента карты по вертикали, а  $n$  – число повторов элемента карты по горизонтали.

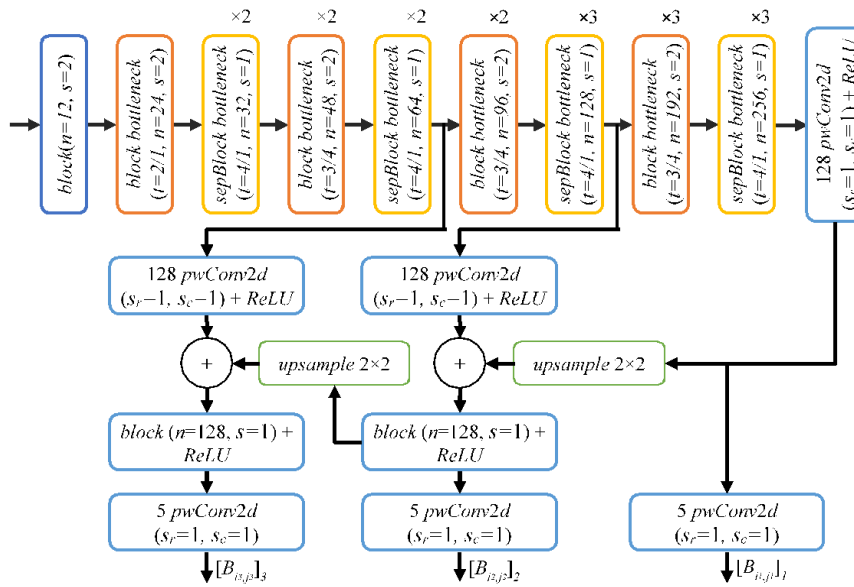


Рисунок 5. – Архитектура ИНС

*Процесс обучения.* Для обучения использовалось несколько функций ошибки для каждого из элементов вектора  $V_{i,r,j_r}$  в выражении (1). Для обучения выхода, предсказывающего вероятность обнаружения лица использовалась средняя взвешенная бинарная перекрестная энтропия:

$$L_p = -\frac{1}{n} \sum_{i=1}^n k_p \cdot p_i \cdot \log \log (\hat{p}_i) + k_n \cdot (1 - p_i) \cdot \log \log (1 - \hat{p}_i),$$

где  $n$  – число обучающих примеров,  $p_i$  – целевое значение вероятности обнаружения объекта  $i$ -го примера,  $\hat{p}_i$  – предсказанное значение вероятности обнаружения объекта  $i$ -го примера,  $k_p$  и  $k_n$  – весовые коэффициенты для положительных и отрицательных примеров наличия объекта, которые определяются по формулам:  $k_p = \frac{n_{max}}{n_p}$  и  $k_n = \frac{n_{max}}{n_n}$ , где  $n_p$  – число положительных примеров,  $n_n$  – число отрицательных примеров и  $n_{max} = \max(n_p, n_n)$ .

Функцией ошибки для выходов, предсказывающих параметры рамки лица, является сумма средних квадратических ошибок координат центра рамки и сумма средних квадратических ошибок логарифмов ширины и высоты рамки:

$$L_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2,$$

$$L_{wh} = \frac{1}{n} \sum_{i=1}^n (\log \log (w_i) - \log \log (\hat{w}_i))^2 + (\log \log (h_i) - \log \log (\hat{h}_i))^2,$$

где  $x_i, y_i, w_i, h_i$  – целевые значения смещения центра и размера рамки объекта  $i$ -го примера;  $\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i$  – соответствующие предсказанные параметры для объекта  $i$ -го примера.

Для вычисления функций ошибки параметров рамки лица используются только те примеры, для которых целевое значение вероятности обнаружения объекта равно 1. В обучении применяется линейная комбинация приведенных функций ошибок, которая записывается следующим образом:

$$L = a_p \cdot L_p + a_{xy} \cdot L_{xy} + a_{wh} \cdot L_{wh}, \quad (2)$$

где  $a_p, a_{xy}, a_{wh}$  – весовые коэффициенты.

Функция ошибки (2) минимизировалась методом стохастической оптимизации Adam [17]. При обучении применялись следующие значения коэффициентов:  $a_p = 1, a_{xy} = 10, a_{wh} = 5$ . Скорость обучения равнялась 0.001, и уменьшался линейно в 1,3 раз каждые 50 000 итераций. Обучение длилось 200 000 итераций.

*Результаты экспериментов.* Обучение детектора проводилось на базе данных WIDER [18]. Эта база содержит 393 703 размеченных лиц на 32 203 изображениях. Изображения в базе разбиты на 61 группу в зависимости от масштаба, позы, перекрытия и освещения и проч. На рисунке 6 показаны образцы изображений лиц из этой базы.



Рисунок 6. – Пример групп изображений из базы WIDER

Помимо ИНС изображенной на рисунке 5 были обучены широко известные архитектуры: NasNet [19] и MobileNet V2 [20]. Эти архитектуры на сегодняшний день

являются одними из самых экономных в вычислительном плане, что позволяет использовать их для сравнения.



Рисунок 7. – Пример изображений из базы Fddb с нанесенной разметкой

Тестирование детекторов лиц проводилось на базе Fddb [21], содержащей 2845 изображений с 5171 вручную размеченными лицами. Пример изображений из базы приведен на рисунке 7. Для проведения тестирования изображения из этой базы масштабировались таким же образом, как и для обучения. В качестве метрик использовались точность, полнота и F1-мера [22]. Пороговое значение вероятности детектирования – 0.5, пороговое значение  $IoU$  (англ. – intersection over union) для алгоритма подавления не максимумов – 0.4 и для фиксирования правильно сдетектированного лица – 0.4. Для сравнения вычислительной сложности оценивалось количество выполняемых операций с плавающей точкой каждой из тестируемых ИНС – FLOPs (float point operations). Результат тестирования приведен в таблице 1.

Таблица 1. – Результаты тестирования детекторов лиц на базе Fddb

Детектор	Размер изображения	FLOPs	Точность	Полнота	F1-мера	GPU, мс
NasNet	416×416	22 044 117	94.58	94.88	94.73	10.04±0.14
MobileNet V2	416×416	11 706 405	89.16	96.09	92.49	6.73±1.03
Предлагаемый детектор	416×416	5 155 823	84.93	96.70	90.44	5.12±1.02

Измерение времени прямого прохода (колонка GPU в таблице 1) проводилось путем сбора статистики о времени работы детектора на одиночных изображениях базы из Fddb. Изображения приводились к размеру 416×416. За один проход сети осуществлялось детектирование всех лиц на изображении. По собранным данным о времени работы детектора производился расчет математического ожидания и среднеквадратического отклонения, которые и представлены в таблице. Для замера времени использовалась видеокарта Nvidia GeForce 1080 Ti.

**Заключение.** Полученные результаты тестирования детектора лиц показывают, что при существенном уменьшении числа выполняемых операций сети, точность детектирования уменьшается в меньшей степени, в то время как полнота детектирования остается на том же уровне. Можно сделать вывод, что предложенная архитектура ИНС для детектирования лиц является оптимальным вариантом для приложений, работающих в условиях ограниченных ресурсов.

#### Список литературы

[1.] A survey on face detection in the wild: Past, present and future / S. Zafeiriou, C. Zhang, Z. Zhang // Computer Vision and Image Understanding, 2015. – Vol. 138. – PP. 1-24.

- [2.] Rich feature hierarchies for accurate object detection and semantic segmentation / R. Girshick, J. Donahue, T. Darrell, J. Malik // *Computer Vision and Pattern Recognition*, 2014. – PP. 580-587.
- [3.] You Only Look Once: Unified, Real-Time Object Detection / J. Redmon, S. Divvala, R. Girshick, A. Farhadi // *Computer Vision and Pattern Recognition*, 2016. – PP. 779-788.
- [4.] YOLO9000: Better, Faster, Stronger / J. Redmon, A. Farhadi // *Computer Vision and Pattern Recognition*, 2017. – PP. 6517-6525.
- [5.] SSD: Single Shot MultiBox Detector / W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, A.C. Berg // *European Conference on Computer Vision*, 2016. – PP. 21-37.
- [6.] Feature Pyramid Networks for Object Detection / Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie // *Computer Vision and Pattern Recognition*, 2017. – PP. 936-944.
- [7.] Soft-NMS – Improving Object Detection with One Line of Code / N. Bodla, B. Singh, R. Chellappa, L. Davis // *IEEE International Conference on Computer Vision*, 2017. – PP. 5562-5570.
- [8.] ImageNet Classification with Deep Convolutional Neural Networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // *Neural Information Processing Systems*, 2012. – Vol. 25. – PP. 1-9.
- [9.] Going Deeper with Convolutions / C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich // *Computer Vision and Pattern Recognition*, 2015. – PP. 1-9.
- [10.] Rethinking the Inception Architecture for Computer Vision / C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna // *Computer Vision and Pattern Recognition*, 2016. – PP. 2818-2826.
- [11.] Rigid-Motion Scattering for Texture Classification / L. Sifre, S. Mallat // *arXiv.org. Preprint: 1403.1687*.
- [12.] Very Deep Convolutional Networks for Large-Scale Image Recognition / K. Simonyan, A. Zisserman // *arXiv.org. Preprint: 1409.1556*.
- [13.] Approximation by Superpositions of a Sigmoidal function / G.V. Cybenko // *Mathematics of Control, Signals and Systems*, 1989. – Vol. 2. – PP. 303-314.
- [14.] Rectified Linear Units Improve Restricted Boltzmann Machines / V. Nair, G. Hinton // *International Conference on Machine Learning*, 2010. – PP. 807-814.
- [15.] Deep Residual Learning for Image Recognition / K. He, B. Singh, R. Chellappa, L. Davis // *Computer Vision and Pattern Recognition*, 2015. – PP. 770-778.
- [16.] Batch normalization: accelerating deep network training by reducing internal covariate shift / S. Ioffe, C. Szegedy // *International Conference on Machine Learning*, 2015. – Vol. 37. – PP. 448-456.
- [17.] Adam: A Method for Stochastic Optimization / D.P. Kingma, L.J. Ba // *International Conference on Learning Representations*, 2015. – PP. 1-13.
- [18.] WIDER FACE: A Face Detection Benchmark / S. Yang, P. Luo, C. Loy, X. Tang // *Computer Vision and Pattern Recognition*, 2016. – PP. 5525-5533.
- [19.] Learning Transferable Architectures for Scalable Image Recognition / B. Zoph, V. Vasudevan, J. Shlens, Q. Le // *Computer Vision and Pattern Recognition*, 2017. – PP. 8697-8710.
- [20.] MobileNetV2: Inverted Residuals and Linear Bottlenecks / M. Sandler, A. Howard, V. Zhu, A. Zhmoginov, L. Chen // *Computer Vision and Pattern Recognition*, 2018. – PP. 4510-4520.
- [21.] FDDB: A Benchmark for Face Detection in Unconstrained Settings // V. Jain, E. Learned-Miller // *Technical Report UM-CS-2010-009*, Dept. of CS, University of Massachusetts, 2010.
- [22.] Precision and recall [Электронный ресурс] / Wikipedia – Режим доступа : [https://en.wikipedia.org/wiki/Precision\\_and\\_recall](https://en.wikipedia.org/wiki/Precision_and_recall) – Дата доступа : 01.02.2020.



## **DEEP MULTI-SCALE FACE DETECTOR BASED ON DEEP NEURAL NETWORK**

**A.V. Susha**

*Postgraduate student of the  
BSUIR*

**M.I. Vashkevich,**

*Assistant Professor of the BSUIR*

*Belarusian State University of Informatics and Radioelectronics*

*Minsk, Republic of Belarus*

*E-mail: isushik94@bsuir.by, vashkevich@bsuir.by*

**Abstract.** The main objective of this work was a development of a deep artificial neural network for face detection purposes. The focus of its design was made on providing of the high performance of the detector and lowering of its computational power requirements by using: 1) factorization of convolution; 2) pointwise convolution; 3) combination of depthwise and pointwise convolution. The detector was compared with similar face detectors based on other well-known neural network architectures MobileNet and NasNet. The proposed face detector has a computational complexity equalling 5.1 MFLOPs, which is two times less than MobileNet's one (11,7 MFLOPs) and four times less than NasNet's one (22 MFLOPs). The detection time for  $416 \times 416$  image was 5.12 ms (or 195 FPS) using GPU GeForce 1080 Ti, and 65.4 ms (or 15 FPS) using one processor core of Intel Core i7-8700K. The precision of our design is 85% and less on 4% than MobileNet has, and less on 9.5% than NasNet has.

**Keywords:** face detection, deep neural networks, convolutional neural networks.

УДК 004.93

## **ПРОБЛЕМЫ СТАТИСТИЧЕСКИХ ОЦЕНОК ПРИ АНАЛИЗЕ НЕСБАЛАНСИРОВАННЫХ КЛАССОВ БОЛЬШИХ ДАННЫХ**



***В.В. Старовойтов***

*Главный научный сотрудник ОИПИ  
НАН Беларуси,  
доктор технических наук, профессор*

*Объединенный институт проблем информатики национальной академии наук Беларуси,  
Республика Беларусь  
E-mail:valerystar @ mail.ru*

### ***В.В. Старовойтов***

*Главный научный сотрудник ОИПИ НАН Беларуси, лауреат Государственной Премии Республики Беларусь (2002г.).*

**Аннотация.** В статье утверждается что прикладная статистика в настоящее время не готова к анализу и обработке больших данных. Вычислять средние значения, дисперсию и прочие статистические характеристики для многочисленных и разнообразных классов объектов, относящихся к категории больших данных, бессмысленно и бесполезно. Одной из актуальных задач является классификация множества объектов на существенно различные по объему классы. К ним относятся реальные задачи разделения людей на заболевших некоторой болезнью и здоровых, сортировка электронной почты на спам и обычные сообщения и т.п. Разработано множество методов классификации данных. Результаты их работы описываются матрицами ошибок. По этим матрицам можно оценить качество классификации и выбрать лучший метод классификации определенных данных. До настоящего времени для оценки качества результатов классификации данных чаще всего используются функции Accuracy, Sensitivity, Specificity и F1. В результате экспериментальных исследований установлено, что указанные функции искажают истинные результаты классификации в случае существенного дисбаланса классов. Показано, что для оценки бинарной классификации из известных функций наиболее инвариантной к дисбалансу классов является функция AUC, которая вычисляет площадь под ROC-кривой. В случае бинарной классификации она равна среднему арифметическому значению функций Sensitivity и Specificity.

**Ключевые слова:** прикладная статистика, классификация несбалансированных данных, Accuracy, Sensitivity, Specificity, F1, AUC.

**Введение.** Шестидесять лет назад цифровые компьютеры сделали информацию читаемой. Двадцать пять лет назад Интернет сделал ее доступной. С 2000-х поисковики создали единую (всемирную) базу данных. В 21-м веке самым ценным ресурсом на Земле становится информация, содержащаяся в цифровых данных различного вида. Таких данных накопилось столь много, что 10 лет назад стали широко применяться понятие «большие данные» или Big Data. Формального определения понятия «большие данные» нет. Неформально этим термином обозначают плохо структурированные данные огромных объемов, прирастающих с огромной скоростью.

Профессор Де Мауро проанализировал различные определения понятия “Big Data” и предложил следующее: «Большие данные – это информационный багаж, характеризующийся таким большим объемом, скоростью изменений и разнообразием, что для его преобразования

в нечто ценное требуются специальные технологии и аналитические методы» [1]. Понятие «большие данные» включает не только наборы данных гигантского объема до  $10^{15}$ - $10^{18}$  байт, но и методы сбора таких данных, их хранения и обработки. Для сравнения скажем, что масса Земли равна примерно  $6 \cdot 10^{23}$  граммов.

Как описывать, как использовать такие данные, анализировать их и извлекать содержащуюся в них информацию?

**Прикладная статистика и большие данные.** Математическая статистика – это область математики, в которой исследуются задачи анализа множеств количественных и качественных данных. В статистике хорошо проработаны задачи анализа небольших групп данных: несколько десятков или сотен единиц. В еще докомпьютерные времена был разработан ряд математически обоснованных критериев. Имеется закон больших чисел. Он гласит, что среднее значение конечной выборки из фиксированного распределения стремится к математическому ожиданию этого распределения. В статистике предложены методы описания и анализа множеств данных различной природы, проверки гипотез равенства средних, дисперсий, наличия линейной зависимости между данными двух множеств и т.п. Статистика продолжает развиваться в основном, используя вероятностный подход к анализу данных небольших объемов.

В начале 1980-х годов в СССР заговорили о появлении прикладной статистики [2-3]. Она нацелена на решение реальных задач. В неё входят ориентированные на прикладную деятельность статистические методы анализа данных, а также методологию организации статистического исследования и организацию компьютерной обработки данных, в том числе разработку и использование баз данных и статистического ПО.

Возьмем частный пример из области прикладной статистики: математические методы диагностики [4]. Они делятся на параметрические и непараметрические. Первые основаны на предположении, что классы описываются распределениями из некоторых параметрических семейств. Обычно рассматривают многомерные нормальные распределения, при этом зачастую принимают гипотезу о том, что ковариационные матрицы для различных классов совпадают. На таких предположениях сформулирован классический дискриминантный анализ Фишера. Однако обычно нет оснований считать, что наблюдения извлечены из нормального распределения. Поэтому более корректными, чем параметрические, считаются непараметрические методы диагностики. Идея таких методов основана на лемме Неймана-Пирсона. Согласно этой лемме решение об отнесении вновь поступающего объекта к одному из двух классов принимается на основе отношения плотностей распределения двух классов. Если плотности распределения неизвестны, то применяют их непараметрические оценки плотностей, построенные по обучающим выборкам, а диагностическое решение принимают по их отношению. Таким образом, для решения задачи диагностики достаточно построить непараметрические оценки плотности для выборок объектов произвольной природы. Как это сделать в условиях больших данных никто знает.

В 2019 году в Беларуси провели республиканскую акцию, в которой участвовали стоматологи, онкологи и оториноларингологи [5]. Врачи оценивали статистику рака кожи в области лица, полости рта, на небных дужках, языке. Первичная выявляемость онкозаболеваний составила 6%, зато третью–четвертую стадии онкозаболеваний они диагностировали в 80% случаев. Эти результаты свидетельствуют о трудностях различия здорового и больного, т.е. о сложности задачи выделения информативных признаков и разбиения образцов исследуемых данных даже на два класса: здоровый и больной.

**Современная статистика не готова анализировать большие данные.** Методы классической статистики практически не применимы к анализу больших данных. Причинами этого (кроме их объема) являются децентрализованные способы хранения, плохая структуризация и взаимосвязь данных. Это привело к тому, что к настоящему моменту большие данные не могут быть статистически описаны.

В базе данных ImageNet (<http://www.image-net.org>) собрано более 14 миллионов аннотированных изображений, разбитых на 21 841 категорию. Например, породы собак представлены 120 классами и это лишь одна из категорий объектов. Как статистически обрабатывать такие множества данных пока не ясно. Вычислять средние значения, дисперсию и прочие статистические характеристики для подобных разнообразных данных бессмысленно и бесполезно. Аналогом бесполезности является величина средней зарплаты по стране.

Майский номер за 2018 год журнала *Statistics & Probability Letters* был целиком посвящен обсуждению проблемы «Роль статистики в эпоху больших данных» [6]. Было опубликовано 36 статей, написанных ведущими учеными-статистиками мира, однако кроме вывода о том, что что-то нужно делать никаких конкретных предложений они не дали. Таким образом статистика как наука существенно отстала от бурно увеличивающегося объема данных, которые требуется анализировать.

На сегодняшний день основным инструментом работы с большими данными являются глубокие нейронные сети. Их особенностью является принцип: чем больше данных использовано при обучении, тем лучше результат. Разработала, реализовала и положила начало практическому применению таких сетей компания Google. Она сначала обучила свои нейронные сети отличать кошек от собак, а в настоящее время уверенно распознавать их породы. С каждым годом расширяются области применения сетей этого типа. Глубокие нейронные сети уверенно распознают людей по их фотографиям, а также многие другие категории объектов, представленных большими объемами данных разных типов. При этом объемы обучающих выборок составляют миллионы объектов.

В интернете можно найти множество библиотек с открытыми кодами для применения глубоких нейронных сетей к решению самых разных задач. Очень интересным и полезным для проверки разных идей анализа и распознавания данных является платформа [kaggle.com](https://www.kaggle.com), принадлежащая компании Google LLC. На нее выкладываются различные задачи анализа данных, чаще всего не имеющие удовлетворительных решений, дается срок 2-3-4 месяца и объявляется конкурс на лучшее решение. Участники открыто обмениваются идеями и опытом, а компании, объявившие конкурс, получают быстрые прототипы решений научно-исследовательских задач.

Одним из важных инструментов статистического анализа данных является определение наличия корреляции между группами объектов. В книге Шонбергера и Кукера «Большие данные: революция, которая изменит то, как мы живем, работаем и думаем», 2013 года [7] сделано важное заявление: *«корреляция не равна причинно-следственной связи, она может быть ошибочной»*. Например, маленьким детям говорят, что аисты приносят детей. Этому нашлось реальное подтверждение. Изучалась корреляция между числом аистов, свивших гнезда в южных районах Швеции, и рождаемостью в соседних поселениях. Корреляция между этими явлениями составила около 0.8. Оказалось, что синхронные изменения числа аистов и детей объяснялись изменением среднего уровня жизни жителей. При исключении этой искажающей переменной прежней корреляции уже не наблюдалось. Вывод: причинная зависимость не может быть выведена ни из какого наблюдаемого совместного изменения явлений.

Реальные связи и закономерности могут быть гораздо более глубокими и не очевидными. Например, Google предложила предсказывать возникновение эпидемий гриппа посредством анализа данных поисковых запросов об этой болезни [8]. Компания показала, что относительная частота определенных запросов сильно коррелирует с процентом визитов к врачу пациентов с симптомами, подобными гриппу. Поисковая система Google обработала порядка 450 миллионов математических моделей с целью уточнения условий поиска, сравнивая результаты с фактическими данными о случаях гриппа за 2007-2208 годы. Были найдены 45 условий поиска, использование которых давало коэффициент корреляции между

их прогнозом и официальными данными 97%. Это позволяет оценить уровень еженедельной активности гриппа в каждом регионе США с точностью до одного дня.

Поэтому в противовес классической статистике в марте 2008 года Питер Норвиг, директор исследований в Google, заявил на одной из конференций: **«Все модели ошибочны и все чаще вы можете добиться успеха без них».**

*Классические оценки классификации несбалансированных данных работают некорректно.* Актуальными направлениями анализа больших данных являются предсказание чего-либо в результате анализа данных и классификация данных. Очень важно корректно оценить результаты классификации, особенно в случае несбалансированных классов данных, и выбрать точнее работающий классификатор.

Анализ публикаций, посвященных классификации данных различной природы показывает, что в подавляющем большинстве из них используются стандартные статистические оценки, вычисляемые по матрицам ошибок, такие как Accuracy (правильность), Sensitivity (она же Recall или чувствительность), Specificity (специфичность), F1 (гармоническое среднее), area under ROC curve (AUC или площадь под ROC кривой). Причем первая оценка используется в более чем в 80% публикаций.

Приведем пример. [9] Эта статья посвящена анализу модели прогнозирования для диагностики рака молочной железы, как доброкачественной, так и злокачественной, на ранней стадии, поскольку это увеличивает шансы на успешное лечение. Сравниваются алгоритм классификации данных на базе SVM, Naive Bayes, k-NN, Decision Tree, а результаты оцениваются с помощью таких статистических мер, как точность, чувствительность, специфичность, значение положительного прогноза, значение отрицательного предсказания и площадь под кривой ROC. Исходные данные собраны из цифровых изображений и представлены в базе данных <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>. Данные представлены 32 признаками и размечены на два класса: злокачественная или доброкачественная опухоль. Всего дано 569 образцов: 357 доброкачественных, 212 злокачественных. Визуальное представление 10 первых признаков показано на Рис.1. Справа и слева от вертикальных отрезков показаны распределения признаков доброкачественных и злокачественных опухолей. Из рисунка видно существенное пересечение признаков обоих классов что означает невозможность 100% разделения двух классов. Основной оценочной функцией результатов классификации была функция Accuracy. Лидеры соревнования добились значений порядка 95-98%, а лучшими были признаны классификаторы, использовавшие PCA и SVM.

Приведем еще один пример. На сайте kaggle.com в конкурсе по определению мошенничества с банковскими транзакциями данные представляли собой 284807 корректных транзакций и 492 ложные (0,172% от всех операций). Дисбаланс классов составил 578:1. Применим этим данным простейший (необученный) классификатор, который относит все операции к классу корректных. Он будет имеет очень высокое значение функции Accuracy (99,827 %) и низкое значение Precision (практически 0,0 %), однако такой классификатор не выявит ни одной мошеннической транзакции. Какова польза от такого «правильного», т.е. почти безошибочного классификатора? Поэтому очень актуальным является вопрос: с помощью какой функции можно корректнее оценить результаты классификации несбалансированных данных? На данный момент прикладная статистика предложений не имеет.

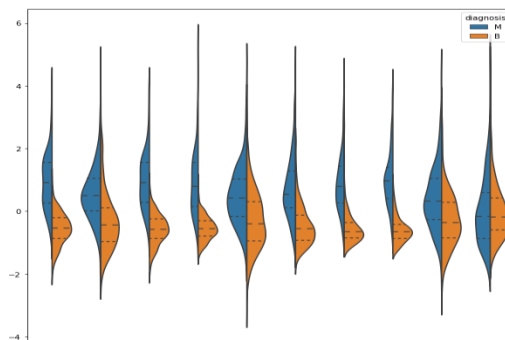


Рисунок 1. – Первые 10 признаков из описания данных Breast Cancer Wisconsin Data Set. Все признаки двух классов существенно пересекаются

Приведем еще один пример. На сайте [kaggle.com](https://www.kaggle.com) в конкурсе по определению мошенничества с банковскими транзакциями данные представляли собой 284807 корректных транзакций и 492 ложные (0,172% от всех операций). Дисбаланс классов составил 578:1. Применим этим данным простейший (необученный) классификатор, который относит все операции к классу корректных. Он будет иметь очень высокое значение функции Accuracy (99,827 %) и низкое значение Precision (практически 0,0 %), однако такой классификатор не выявит ни одной мошеннической транзакции. Какова польза от такого «правильного», т.е. почти безошибочного классификатора? Поэтому очень актуальным является вопрос: с помощью какой функции можно корректнее оценить результаты классификации несбалансированных данных? На данный момент прикладная статистика предложений не имеет.

В результате приведенного анализа и приведенных примеров можно сделать вывод о том, что современная статистика оказалась не готова обрабатывать реально большие данные. В настоящее время подобными задачами занимается только *data science* или наука о данных, относящаяся к искусственному интеллекту.

**Оценки результатов классификации на базе набора KDDCUP99.** Рассмотрим еще один пример множества несбалансированных данных. Это широко используемая и доступная база данных KDDCUP99 [10]. Она использовалась на 3-м международном конкурсе инструментариев для обнаружения знаний и интеллектуального анализа данных, который проводился Пятой международной конференцией по обнаружению знаний и интеллектуальному анализу данных в 1999 году. Задача конкурса заключалась в создании детектора несанкционированных вторжений в сеть на базе прогнозирующей модели, способной различать четыре типа вторжений и «нормальные» соединения. Все атаки разделены на четыре категории:

- DOS- сетевые атаки (Denial of Service attacks);
- R2L ((Remote-to-Local): несанкционированный доступ с удаленного компьютера;
- U2R (User-to-Root): несанкционированный доступ к данным сетевого администратора;
- Probing атаки – сканировании сетевых портов для получения конфиденциальной информации.

В таблице 1 приведены данные дисбаланса между классами этой базы. Самый большой дисбаланс между нормальными соединениями и атаками типа U2R. Данных первого класса в 1295,1 раз больше чем второго.

Таблица 1. – Дисбаланс между классами в базе KDD Train

класс	Normal	Dos	Probe	R2L	U2R	всего
<b>размер</b>	<b>67 343</b>	<b>45 927</b>	<b>11 656</b>	<b>995</b>	<b>52</b>	<b>125 973</b>
Normal		1,47	5,78	67,68	1 295,1	1,87
Dos			3,94	46,16	883,21	2,74
Probe				11,71	224,15	10,81
R2L					19,13	126,6
U2R						2 422,6

На рисунке 2 из статьи [11] показаны проекции образов исследуемых классов на две главные оси. Видно, что классы практически не разделимы без ошибок.

Оценим пять вариантов бинарной классификации на два класса U2R и Normal, по матрицам ошибок приведенным в Таблице 2. Первый вариант соответствует наилучшей классификации (всего 2 ошибки). Затем идут второй и третий. Четвертый и пятые самые плохие. В меньшем классе правильно определен лишь один объект. Два последних варианта соответствуют одному и тому же результату классификации, отличаются лишь перестановкой классов.

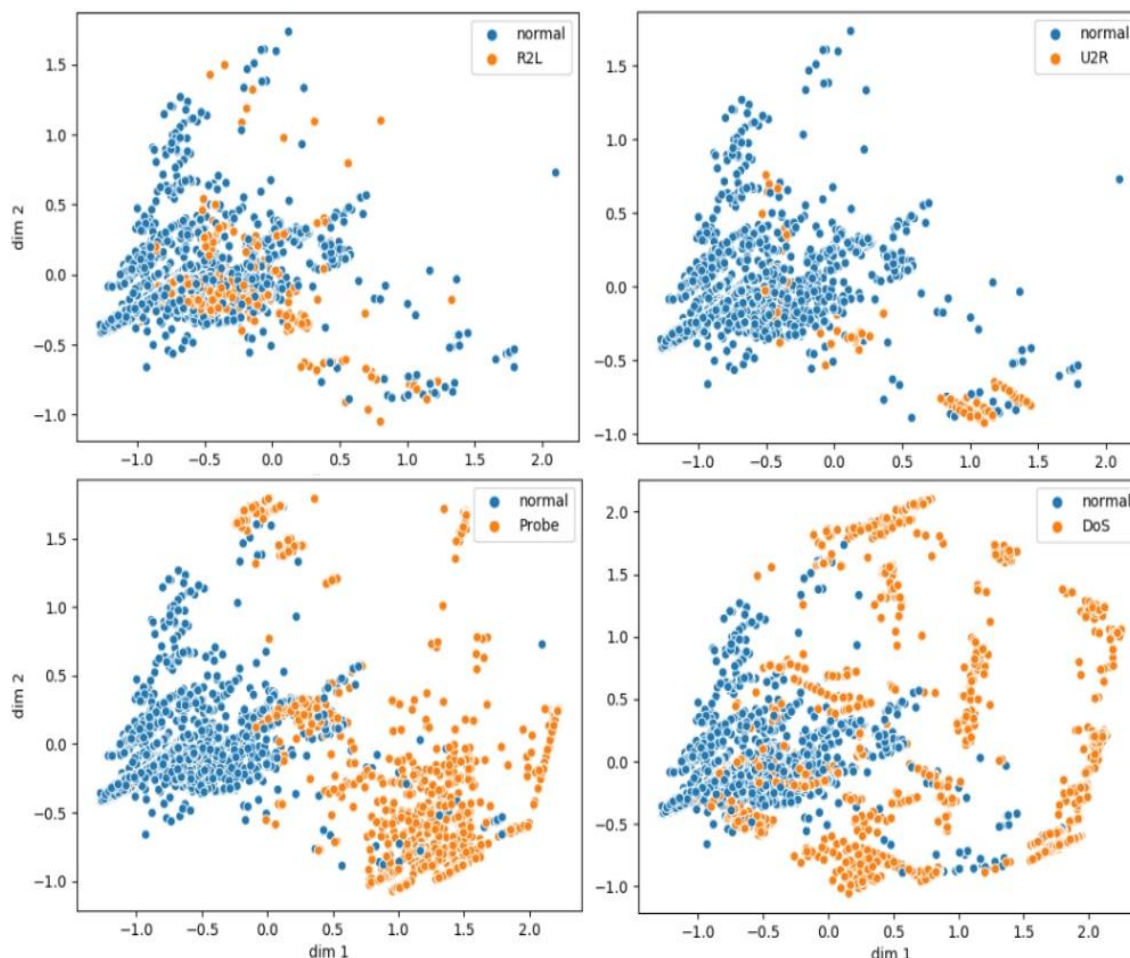


Рисунок 2. – Проекция пар классов на 2 главные оси, выбранные методом главных компонент (PCA). Классы не разделимы

Таблица 2. – Матрицы пяти вариантов ошибок классификации объектов классов U2R и Normal

U2R (52 объекта)		предсказанные классы	
		Yes	No
реальные классы, вариант 1	Yes	51	1
	No	1	67342
вариант 2	Yes	41	10
	No	10	67332
вариант 3	Yes	32	1
	No	20	67342
вариант 4	Yes	1	1
	No	51	67342
вариант 5	Yes	67342	51
	No	1	1

Нами было исследовано около 20 известных функций оценки качества бинарной классификации со сбалансированными и несбалансированными классами данных [12]. Часть из них представлена в Таблице 3.

Вычислим функции из Таблицы 3 к матрицам ошибок и запишем в Таблицу 4.

Таблица 3. – Основные функции оценки качества бинарной классификации

№	Обозначение функции	Формула
1	Sensitivity и Recall, чувствительность	$\frac{tp}{tp + fn}$
2	Specificity, специфичность	$\frac{tn}{tn + fp}$
3	Precision, точность	$\frac{tp}{tp + fp}$
4	Аccuracy, правильность	$\frac{tp + tn}{n}$
5	Jaccard index, индекс Жаккара	$\frac{tp}{tp + fn + fp}$
6	F1, гармоническое среднее	$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ , при $\beta = 1$
7	F2	$\frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}$ , при $\beta = 2$
8	геометрическое среднее, Geometric mean (GM)	$\sqrt{\text{Precision} \times \text{Recall}}$
9	area under ROC curve (AUC), площадь под ROC кривой	$\frac{\text{Sensitivity} + \text{Specificity}}{2}$
10	Cohen's kappa, каппа Коэна	$\frac{(tp + fp)(tp + fn) + (fn + tn)(fp + tn)}{n^2}$
11	MCC, коэффициент корреляции Мэтьюса	$\frac{(tp \times tn - fp \times fn)}{\sqrt{(tp + fp) \times (tp + fn) \times (tn + fp) \times (tn + fn)}}$

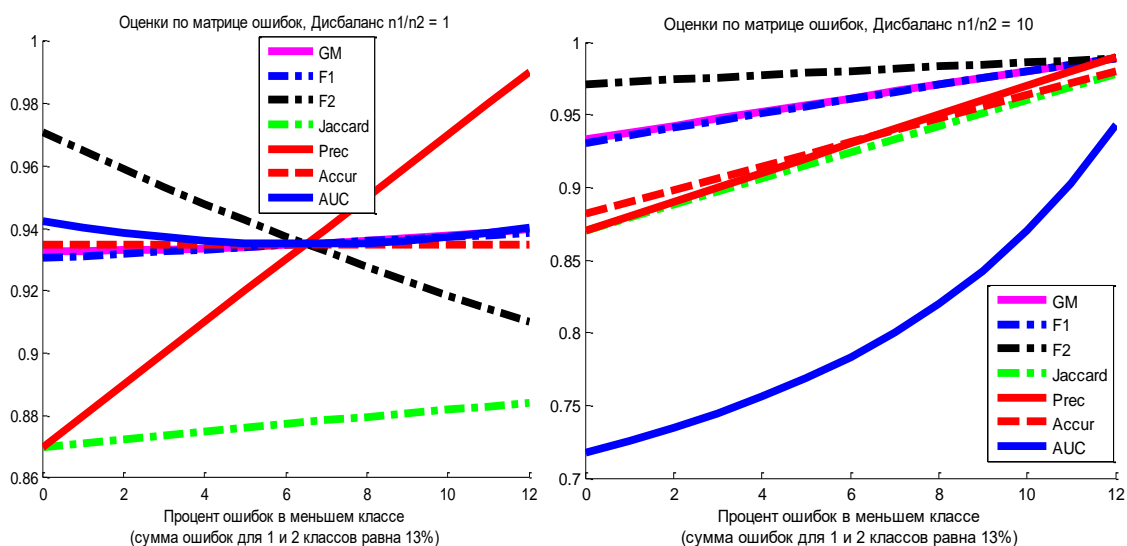
где  $n = tp + fp + fn + tn$ .



Таблица 4. – Сравнительные результаты оценок классификации

номер	функция	вар. 1	вар. 2	вар. 3	вар. 4	вар. 5
1	Sensitivity	0.9808	0.8039	0.9697	0.5000	0.9992
2	Specificity	1.0000	0.9999	0.9997	0.9992	0.5000
3	Precision	0.9808	0.8039	0.6154	0.0192	1.0000
4	Accuracy	1.0000	0.9997	0.9997	0.9992	0.9992
5	Jaccard	0.9623	0.6721	0.6038	0.0189	0.9992
6	F1	0.9808	0.8039	0.7529	0.0370	0.9996
7	F2	0.9808	0.8039	0.8696	0.0833	0.9994
8	GM	0.9808	0.8039	0.7725	0.0981	0.9996
9	AUC	0.9904	0.9019	0.9847	0.7496	0.7496
10	Cohen's kappa	0.9904	0.9019	0.8764	0.5185	0.5185
11	MCC	0.9904	0.9019	0.8862	0.5490	0.5490

Проанализируем результаты пяти вариантов классификации, собранные в Таблице 4. Первые четыре функции – это популярные меры качества, вычисляемые по матрице ошибок классификации. Корректные оценки должны быть инвариантны к перестановке классов в матрице ошибок, однако функции Sensitivity, Specificity, Precision, индекс Jaccard, F1 и F2 не инвариантны к такой перестановке, т.е. это неудовлетворительные меры. Accuracy инвариантна, но она не учитывает дисбаланс классов (см. варианты 4 и 5 в Таблице 4). При 100% необнаружении объектов меньшего класса (атак типа U2R) и одной ошибке классификации нормальных объектов Accuracy равна 0.9992, что означает почти 100% правильность классификации. Это называется парадоксом Accuracy. Данный пример показывает, что функция Accuracy не годится для оценки качества классификации несбалансированных данных.



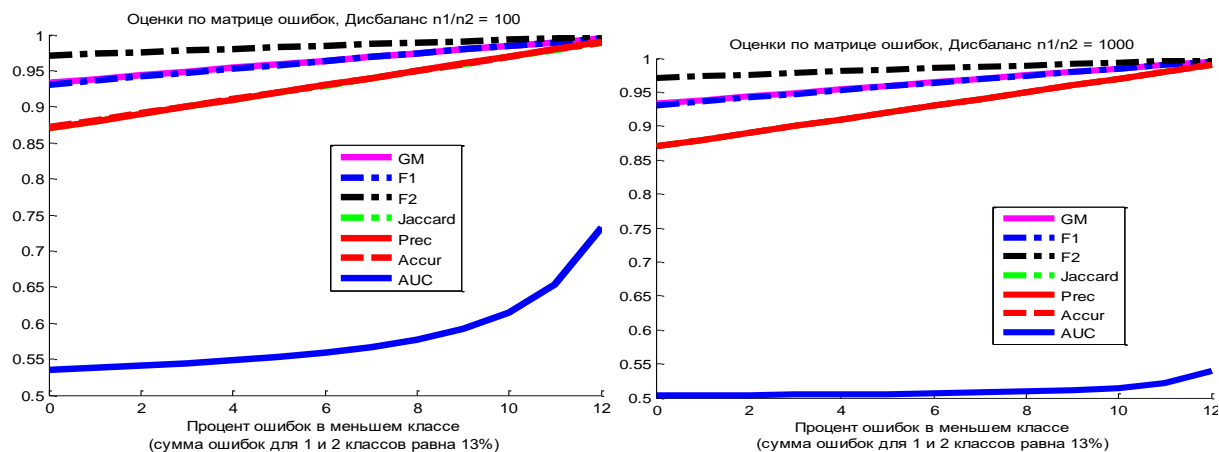


Рисунок 3. – Графики семи оценочных функций при суммарных ошибках в двух классах равным 13%. Слева вверху без дисбаланса, справа вверху с дисбалансом 1:10, слева внизу с дисбалансом 1:100, справа внизу с дисбалансом 1:1000

На Рисунке 3 показано как изменяются значения семи функций из Таблицы 3 без дисбаланса и при дисбалансе классов 1:1000, процент ошибок в меньшем классе изменяется от 0 до 13%. При отсутствии дисбаланса функция Accuracy имеет постоянное значение, а GM, F1 и AUC имеют примерно такие же значения. При увеличении дисбаланса функция GM практически совпадает с F1. Значения функций Precision, Accuracy и индекса Jaccard также практически совпадают. Значения всех функций кроме AUC выше 0.87 и стремятся к 1 с возрастанием числа ошибок в меньшем классе. Значения 1 всех функций соответствуют абсолютно точной классификации данных. График функции AUC демонстрирует наиболее адекватную оценку качества классификации. Значения уменьшаются при росте дисбаланса.

Функции MCC, Cohen's kappa, Youden\_index, AUC инвариантны к перестановке классов в матрице ошибок. Отметим что две последних не инвариантны к транспонированию матрицы ошибок. Инвариантность к транспонированию позволяет произвольно записывать значения ошибок I и II рода.

Бинарная классификация часто применяется при разделении на классы больных и здоровых, нормальных и имеющих отклонения, и т.п. Выполним разделение данных базы KDD Train на два класса normal и anomaly. Матрица ошибок приведена в Таблице 5. Дисбаланс классов невелик и равен 1,33.

Таблица 5. –Матрица ошибок разбиения базы KDD Train на два класса: нормальные соединения и угрозы

	anomaly	normal
anomaly	9362	3471
normal	298	9413
число объектов в классе	9660	12884
1: 1.33 дисбаланс		

Вычислим 12 оценочных функций, приведенных выше. Результаты приведены в Таблице 6. Поскольку дисбаланс невелик функция Accuracy, показавшая 83,28%, достаточно близка к реальной оценке результата классификации. Функции F1, GM, AUC, kappa, MCC имеют примерно такие же значения. Однако с учетом результатов предыдущего эксперимента рекомендуется применять AUC, либо kappa и MCC.

Таблица 6. – Оценки классификации, вычисленные по матрице ошибок из Таблицы 5

функция	значение	функция	значение
Sensitivity (recall)	0.9692	F2	0.9094
Specificity	0.7306	GM	0.8408
Precision	0.7295	AUC	0.8499
Accuracy	0.8328	Карра	0.8361
Jaccard	0.7130	MCC	0.8496
F1	0.8324		

Функция AUC не идеальна, но она самая простая с точки зрения вычислений и согласно исследованиям, описанным в работе [12] наиболее устойчива к дисбалансу классов анализируемых данных. Авторы работы [13] пришли к аналогичному заключению. Они исследовали четыре варианта представления данных для классификации и пришли к выводу что, если в обучающей выборке присутствует шум или другие искажения, то меры, основанные на вероятности, не очень хороши для анализа результатов классификации, при этом функция AUC является лучшим вариантом оценки качества классификации.

**Заключение.** В работе затронута важная проблема: прикладная статистика в настоящее время не имеет методов обработки больших данных. Одной из актуальных прикладных задач анализа больших данных является разделение множеств объектов на существенно различные по объему классы, т.е. имеющие дисбаланс. Продемонстрировано, что при оценке результатов классификации данных с дисбалансом не стоит использовать классические функции, такие как Accuracy, Sensitivity, Specificity и F1. Для оценки качества классификации из известных наиболее инвариантной к дисбалансу классов является функция AUC, вычисляющая площадь под ROC кривой. В случае классификации на два класса она равна среднему арифметическому значению функций Sensitivity и Specificity.

#### Список литературы

- [1.]De Mauro A, Greco M, Grimaldi M. A formal definition of Big Data based on its essential features // Library Review.- 2016.- Vol. 65.- No. 3.- С. 122-135.
- [2.]Орлов А. И. О развитии прикладной статистики. — В сб.: Современные проблемы кибернетики (прикладная статистика). — М.: Знание, 1981, с.3-14.
- [3.]Математическая и прикладная статистика: учеб. пособие / Ю.С. Харин, Е.Е. Жук.-Мн.: БГУ, 2005.- 279с.
- [4.]Орлов А.И. Прикладная статистика, М.: Экзамен, 2004.-656с.
- [5.]Рак полости рта у белорусов чаще всего находят на последних стадиях. Почему?
- [6.]The role of Statistics in the era of big data // Statistics and Probability Letters.- Vol.136.- 2018.- 170p.
- [7.]Mayer-Schönberger V., Cukier K. Big data: A revolution that will transform how we live, work, and think. – Houghton Mifflin Harcourt, 2014.
- [8.]Ginsberg, J., et.al. Detecting influenza epidemics using search engine query data // Nature.- 2009.- Vol.457 (7232) .- С.1012-1014.
- [9.]Sinha A, Sahoo B, Rautaray SS, Pandey M. Analysis of Breast Cancer Dataset Using Big Data Algorithms for Accuracy of Diseases Prediction // Int. Conf. on Computer Networks and Inventive Communication Technologies, 2019.- С. 271-277.
- [10.] KDD Cup 1999 Data. <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>.
- [11.] Odiathevar M., Seah W. K. G., Fream M. A Hybrid Online Offline System for Network Anomaly Detection // 28th Int. Conf. on Computer Communication and Networks, IEEE, 2019. – С.1-9.
- [12.] Старовойтов В.В., Голуб Ю.И. Сравнительный анализ оценок качества бинарной классификации // Информатика.- 2020.- Т.17.- №1.
- [13.] Ferri C., Hernández-Orallo J., Modroiu R. An experimental comparison of performance measures for classification // Pattern Recognition Letters.- 2009.- Vol.30.- №1.- С.27-38.

## **PROBLEMS OF STATISTICAL EVALUATIONS IN ANALYSIS OF IMBALANCED CLASSES OF BIG DATA**

**V. V. Starovoitov,**  
*Doctor of Engineering Sciences  
Chief Researcher, Professor*

*United Institute of Informatics Problems of the National Academy of Sciences of Belarus,  
Republic of Belarus  
E-mail: valerystar @ mail.ru*

**Abstract.** The article stated that applied statistics is currently not ready for analysis and processing of big data. It is senseless and useless to calculate the average values, variance, and other statistical characteristics for numerous and diverse classes of objects that belong to the category of big data. One of the actual tasks is classification of object sets into classes that are significantly different in volume. There are real problems to divide people into those who are sick and healthy, sorting emails into spam and regular messages, etc. Many data classification methods have been developed. Their results are described by confusion matrices. Using these matrices, one can evaluate quality of classification and choose the best method. To date, the basic functions of classification quality assessment are Accuracy, Sensitivity, Specificity, and F1. As a result of experimental studies, it was found that these functions distort the true assessment of the classification quality in the case of a significant class imbalance. We have shown that to evaluate the binary classification the AUC function is the best among well-known functions. It calculates the area under the ROC curve. In the case of binary classification, it is equal to the mean value of the Sensitivity and Specificity.

**Keywords:** applied statistics, classification of imbalanced data, Accuracy, Sensitivity, Specificity, F1, AUC.

## AUTOR INDEX

<b>A</b>		<b>L</b>	
Abbood N. A.	183	Liauchuk V. A.	71
Abdul-Azalova M. Y.	96	Lupersolsky A.	11
Adamov A.	33	<b>M</b>	
Alhamd M. W.	137,183	Mamasoatov D. R.	213
Alqaisi Z. A.	183	Mamatova N. M.	53,96
Altinoglu F. F.	162	Melqonyan M.	137
Asimov R. M.	78	Mezianaya K. N.	128
Asipovich V. S.	78	<b>N</b>	
Aydin A.	162	Narziyev N. B.	147
<b>B</b>		Nishanov A. K.	147
Bahromov A. A.	195	<b>P</b>	
Beknazarova S. S.	118,156	Paulenka D. A.	71
Buranbaeva A. I.	33	Pechkovsky E. I.	71
<b>D</b>		Pilecki I. I.	90, 170
Davydovsky A. G.	176	Pisetsky Yu. V.	165
Dzik C. S.	170	Podelko A.	19
<b>G</b>		Polat G.	162
Gindin S. I.	33	Prihozhy A. A.	202
Glesner M.	33	<b>R</b>	
Gorovoy V. G.	123	Radyna Ya. M	112
Guletskii N. V.	112	<b>S</b>	
<b>H</b>		Saidov F. F.	213
Horton I.	33	Salman T. M.	137
Hutkouski U. N.	90	Sidorovich A. S.	123
<b>I</b>		Skliarova I.	33
Ibodullayev S. N.	195	Sklyarov V.	33
<b>J</b>		Snezhko E. V.	71
Jaumitbayeva M. K.	156	Soo P. I.	165
Jiyanbekov Kh. R.	165	<b>T</b>	
Jurayev G. P.	147	Trubin I. A.	53
<b>K</b>		Tsyrelchuk I. N.	96,165
Karaneuski K. M.	128	<b>V</b>	
Karasik O. N.	202	Valevich S. V.	78
Keevallik A.	33	Varlamova L. P.	176
Khomonenko A. D.	33	Vaskouski M.	105
Klyuyeva Y. G.	33	<b>Y</b>	
Konovalova M. A.	128	Yakhshibaev D. S.	165
Kovalev V. A.	71	Yashin K. D.	128,176
Kruse I.	78	Yavorskiy V. V.	33
Kuchynski A. V.	90	<b>Z</b>	
Kurbanov S. K.	118	Zibitsker B.	11,19

<b>А</b>		<b>М</b>	
Азаров И. С.	271	Маркевич В. С.	375
Ахмедиярова А. Т.	59	Марченко Л. Н.	375
<b>Б</b>		Медведев О. С.	294
Байдикова Н. В.	59	Меженная М. М.	332
Батура М. П.	220	Мороз П. А.	294
Боровиков С. М.	384	<b>Н</b>	
<b>В</b>		Навроцкий А. А.	252
Вашкевич М. И.	400	Насуро Е. В.	391
Верняховская В. В.	303	Никитин А. Н.	350
Войчик В.	59	<b>О</b>	
Волорова Н. А.	220	Осипович В. С.	294
<b>Д</b>		<b>П</b>	
Дершень А. Г.	303	Пархименко В. А.	303
Дершень В. В.	303	Пац Е. А.	391
Дик С. С.	384	Пилецкий И. И.	220, 281
Дик С. К.	334	Прокопеня А. С.	271
Дудич О. Н.	294	Прытков В. А.	220
<b>Е</b>		<b>Р</b>	
Есмагамбетова М. М.	59	Ревотюк М. П.	324
<b>Ж</b>		Рункевич Е. Н.	332
Жвакина А. В.	358	<b>С</b>	
Жэнь С. Х.	265	Савчик О. А.	303
<b>З</b>		Сенькович Д. С.	358
Зенкевич А. Г.	257	Слаута А. И.	367
<b>К</b>		Сонькин М. А.	59
Казак Т. В.	257,319	Старовойтов В. В.	409
Калилец Т. В.	332	Суша А. В.	399
Калимолдаев М. Н.	59	<b>Т</b>	
Каракулько А. А.	294	Тохметов А. Т.	340
Касымова Д. Т.	59	Тусупов А. Д.	340
Кишкевич И. В.	332	<b>У</b>	
Клинов К. И.	382	Утепбергенов И. Т.	59
Клюева Е. Г.	59	<b>Х</b>	
Князева Л. П.	303	Хаджинова Н. В.	324
Козарь Р. В.	252	<b>Ц</b>	
Конопелько В. К.	265	Цветков В. Ю.	265
Красильникова В. Л.	294	Цехан О. Б.	367
<b>Л</b>		<b>Ч</b>	
Лапко М. В.	304	Черемисинов Д. И.	242
Лащётко Р. А.	332	Черемисинова Л. Д.	242
Листопад Н. И.	341	Чочиева А. С.	281
Лихачевский Д. В.	319	<b>Ш</b>	
Лэ.В.Т.	382	Шаталова В. В.	319
<b>М</b>		Шилин Л. Ю.	252
Ма Ц.	265	<b>Я</b>	
		Яворский В. В.	59
		Яшин К. Д.	294

---

**СПОНСОР КОНФЕРЕНЦИИ**



**ИОТ инжиниринг, машинное обучение, математическая  
обработка данных,  
разработка программного обеспечения**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Syberry – резидент Парка Высоких Технологий. Наша  
специализация – профессиональная разработка ПО для  
заказчиков из США.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Внедрение ИТ-решений, Корпоративные системы,  
ИТ-консалтинг, Разработка и внедрение Веб-приложений,  
Разработка Интернет-вещей**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Компания по разработке заказного ПО в Восточной Европе,  
предоставляющая ИТ-решения и консалтинговые услуги.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Разработка программного обеспечения любой сложности**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Разработка, тестирование и поддержка программного  
обеспечения. Анализ, модернизация и интеграция бизнес-  
решений**



---

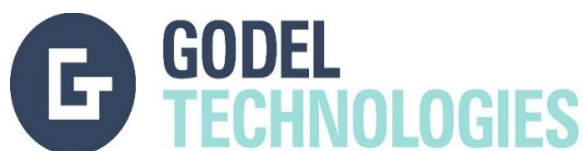
**СПОНСОР КОНФЕРЕНЦИИ**



**Поставка, монтаж/демонтаж и обслуживание лифтов, эскалаторов и траволаторов, запасных частей к ним, наладка систем диспетчеризации, пусконаладочные работы.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**«Godel Technologies» - спектр услуг, предоставляемых компанией, включает предпроектные исследования (в сфере наукоемких технологий), бизнес-анализ, разработку информационной стратегии, разработку заказного программного обеспечения, тестирование и внедрение информационных систем.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Успешная ИТ-компания, осуществляющая деятельность в сфере разработки и продвижения мобильных игр, резидент белорусского Парка высоких технологий**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Компания ICS предоставляет полный цикл услуг по внедрению и поддержке Dynamics AX/B365FO на предприятии**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**XB Software – один из лидеров веб-разработки как собственных продуктов – Webix, GanntPRO, KUKU.io и др., – так и индивидуальных решений для компаний США, Великобритании, Канады и стран Западной Европы.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Компания по разработке программного обеспечения.  
Уникальная интеллектуальная платформа с передовыми  
вычислительными возможностями.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Компания по разработке разработке веб и мобильных приложений, а также различных веб-сервисов.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Компания-поставщик услуг по разработке программного обеспечения, тестированию ПО и консалтингу в сфере ИТ.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**СИСТЕМНЫЕ  
ТЕХНОЛОГИИ**

**Компания - крупнейший поставщик комплексных ИТ-решений для банковской, финансовой и промышленной сфер.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Один из крупнейших банков Беларуси. В отличие от других крупных белорусских банков, специализируется на кредитовании предприятий реального сектора экономики.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Унитарное предприятие по оказанию услуг – провайдер телекоммуникационных, ИКТ- и контент-услуг в Беларуси.**

---

**СПОНСОР КОНФЕРЕНЦИИ**



**Разработчик IT-решений для Западной Европы и США.**

*Научное издание*

## **BIG DATA AND ADVANCED ANALYTICS**

## **BIG DATA И АНАЛИЗ ВЫСОКОГО УРОВНЯ**

СБОРНИК МАТЕРИАЛОВ  
VI МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ

(Республика Беларусь, Минск, 20-21 мая 2020 года)

В трех частях

Часть 1

В авторской редакции  
Ответственный за выпуск *С.К. Дик*  
Компьютерная верстка *И.В. Андриалович*

Подписано в печать 14.05.2020. Формат 60×84 1/8.  
Гарнитура «Таймс». Усл. печ. л. 49,76. Уч.-изд. л. 25,22.  
Тираж 50 экз. Заказ 99.

Издатель и полиграфическое исполнение  
УП «Бестпринт». Свидетельство о государственной регистрации издателя,  
изготовителя, распространителя печатных изданий №1/160 от 27.01.2014.  
Ул. Филатова д. 9, к. 1. 220026, г. Минск.

ISBN 978-985-90533-7-5



9 789859 053375 >

