III International Conference TACSIT-2019 Proceedings

Theoretical and Applied Computer Science and Information Technologies

TACSIT'19
May 7-8, 2019
Severodonetsk, Ukraine


1920

Ministry of Education and Science of Ukraine
Volodymyr Dahl East Ukrainian National University

Theoretical and Applied Computer Science and Information Technologies

III International Conference TACSIT-2019
Proceedings

May 7-8, 2019
Severodonetsk, Ukraine

Severodonetsk, 2019

UDK 004

Theoretical and Applied Computer Science and Information Technology: Proceedings of the III International Conference TACSIT-2019, May 7-8 2019. / I. Skarga-Bandurova (Ed.). – Severodonetsk: Volodymyr Dahl East Ukrainian National University, 2019. - 47 p.

ISSN 2522-4387 ISSN 2522-4395 (electronic)

INTERNATIONAL PROGRAM COMMITTEE

ORGANIZING COMMITTEE

Abstracts are published on the author's originals

CONTENTS

An Investigation on Security Framework of UIDAI 7
Arpana Chaturvedi

Analytical and Simulation Approaches in Computer Network Design 9
Lina Barbaruk, Olha Lavrinenko, Ivan Kvasov

Color Image Control Using Direct Memory Access 11
Maksym Hoha

Network Communication for Remote Parkinson's Disease Monitoring System 13
Oleksandr Berezhnyi, Tetiana Biloborodova, Inna Skarga-Bandurova, Viktoriia Derevianchenko

Design of Algorithms for Technological Processes Control 15
Volodymyr Kardashuk

Efficient Streaming on Multi-Core of Shared-Memory Computing 17
Dmytro Nedzelskyi

Methodology of Fake News Detection 19
Mykyta Davidenko, Tetiana Biloborodova

Hyperparameters Optimization in CNN for Image Recognition 21
Rostyslav Siriak, Inna Skarga-Bandurova, Tetiana Biloborodova

Justification and Practical Application of Routing Protocols 23
Lina Barbaruk, Alice Mikhaylova, Denis Bakitko

Mathematical Model for Assigning Routes of Public Transport 26
Maryna Derkach, Inna Skarga-Bandurova

Modeling of the Water Diffusion Mechanism 29
Ksenia Hulevska, Larisa Shumova, Viktoria Mokhonko

Methods Overview of Open-Source Speech Recognition Software 31
Viktoriia Derevianchenko, Inna Skarga-Bandurova, Polina Fursa, Mark Koverha

Processing Technique for Biomedical Image Analysis 33
Vyacheslav Lyashenko, Oleg Kobylin, Oleksandr Ryazantsev, Ivan Ryazantsev

Query Optimization in Database Systems 35
Maksym Nesterov, Denis Bakitko, Alice Mikhaylova

Real-Time Data Analytics for the Internet of Things 37
Yana Krytska, Tetiana Biloborodova, Inna Skarga-Bandurova

Real-Time Wearable System for Monitoring Cardiovascular Disease 40
Tetiana Biloborodova, Inna Skarga-Bandurova, Viktoriia Derevianchenko

Statistical Analysis of Optical Image for Information System 42
Andrii Riazantsev, Ganna Khoroshun, Oleksandr Ryazantsev

Wireless Sensor Network Conditions Monitoring and Diagnosis 44
Gennady Krivoulya, Vladislav Sergienko

Author Index 46

An Investigation on Security Framework of UIDAI

Arpana Chaturvedi
Department of Information Technology,
Jagannath International Management School,
New Delhi, India
ac240871@gmail.com

***Abstract*—In this paper AESXTS encryption mechanism and digital signature technology is used with AODV (Ad hoc On-Demand Distance Vector) routing protocol to get rid of various issues like DoS (Denial of Service), eavesdropping, imitation, coaxing etc. The theoretical analysis is shown using NS2 simulator and implementation assures to provide better data security, reliability, transmission and energy efficiency. This implementation shows that SAODV (Secure-AODV) routing protocol when used in routing layer in these applications and system, it increases its own self defensive ability to fight against various challenging hidden security issues.**

Keywords— AODV, SAODV, WSN, AES-XTS, UIDAI, DoS

I. INTRODUCTION

The UIDAI (Unique Identification Authority of India) data, both at rest or on move, stored at CIDR (Central Identities Data Repository) is at high risk. Due to the advent of new Hacking technologies and advanced tools, the risk factor has increased. The government is enforcing citizen of India to link Aadhaar number with different citizen centric services so that right citizen can avail the benefits. There is need to safeguard the data stored in different data centers which might be secured through a strict encryption technique applied on application process in parallel mode. The proposed approach is AES in XTS Mode in Map Reduce paradigm which supports parallel programming in the distributed environment. The findings after reviews of results show that it provides better security against external attacks and overcomes the shortcomings of Kerberos. Encryption followed by compression on various datasets provides better result and protection from vulnerabilities and threats. Here we integrate the AES encryption standard in XTS mode and digital signature technology to improve classic AODV (Ad hoc On-Demand Distance Vector) routing protocol. The resultant SAODV (Secure Ad hoc On-Demand Distance Vector) routing protocol provides better information security and achieves energy efficiency as well.

II. PROPOSED ENCRYPTION ALGORITHM- AES (XTS)-MR

There are so many technologies and algorithms available which might be appropriate to handle security issues. The study concludes that the implementation of AES XTX with map reduce parallel programming will be a cost-effective solution to process such a large user-generated vital and sensitive data.

The combination of Advanced Encryption Standards (AES) and Map Reduce (MR) is termed as AES-MR. It is the proposed algorithm for this research work which is related to the case study done for UIDAI to provide better data security. This encryption algorithm is proposed to provide Data level security. Map Reduce is a parallel programming language and AES is the encryption algorithm suitable for longer messages. It is suggested that the best features of both techniques should be used together to provide much stricter security features in the introduced security layer.

Map Reduce with AES-XTS has the capability to compose applications that generates endless data during runtime. It has the ability to adapt non-critical failures and can perform better planning, testing of information. The failed jobs if encountered in the clusters of machines, it re-executes them.

In this proposed technique AES is used with XTS mode which is supported by IEEE 1619-2007 standards [5]. The XTS modes contain XEX-TCB-CTS (XTS) mode where XTS stands the XEX Tweakable Block Cipher with Cipher Text Stealing. The XTS mode performs parallel executions and allows pipelining in respective executions. Data Encryption Standards which were used earlier is vulnerable to Brute Force attacks. It was due to the small size of the key (53 to 2054 bits) DES uses for encryption. US Government Agency NIST (National Institute of Standards and Technology) selected Rijndael's Algorithm as Advanced Encryption Standard. It is being a better security standard now becoming an Industry Standard.

AES is designed to accept 128, 192, 256 bits' size of keys. These various sizes of keys are capable of encrypting different types and variety of information in bulk. The performance of AES algorithms varies on different 32 bit and 64-bit CPU's based on key sizes. This technique will provide better security (Fig.1) in case of the UIDAI system where the large data sets are generated for storing sensitive information of residents and generating UID numbers of a citizen of India.

The block contents can perform parallel processing in various modes of operations. These operations can handle fixed block as well as variable block encryption with the help of single key or different keys.

The Electronic Cook Book (ECB) and XTX are used with AES to increases the efficiency of an algorithm and can be improved by the use of it. [1] The XTX supports parallel encryption mode with Symmetric Block Cipher encryption mode. It was designed to protect data lying at rest on storage devices. It uses a fixed size of data units to perform cryptographic protection of data at rest. The operation of AES -XTX Mode with two different keys is shown in Fig. 2 [1,

2]. The XTS-AES mode is an enhanced concept of Rogaway's XEX (XOR Encrypt XOR) Tweakable Block Cipher, improved with a method called "Cipher Text Stealing". It expands the range of possible different types of data inputs. XEX can only encrypt sequences of complete blocks of any data type. This input data should be an integer and necessarily be a multiple of 128 bits. In XTS-AES, the data string consists of one or more complete blocks which are followed by a single, non-empty partial block ().

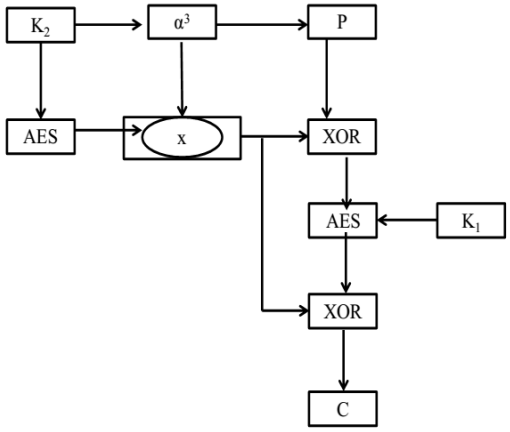

Fig. 1. Operation of AES-XTX Mode with two different keys

The working of AES-XTS is shown in Fig. 2.

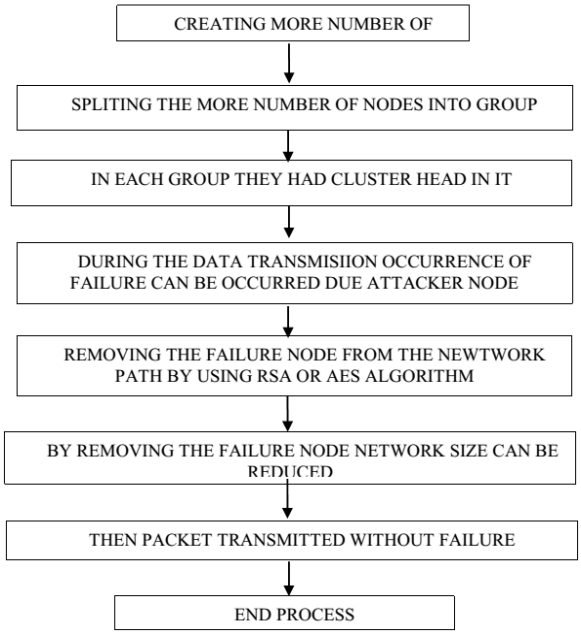

Fig. 2. Flow Chart to show steps used in Implementation Work

The XTS-AES is composed of two keys, first one is an encryption key and second one is tweak key. It incorporates the logical position of the data block into the encryption [3]. The output produced by XTS is independent which leads to parallelization. XTS, an instantiation of the tweakable Block Cipher class. It is capable to implement ciphers in parallel and pipeline modes. It enables the encryption of the last incomplete block also.

CONCLUSION

The AES-MR encryption technique to be implemented to solve the issue of data security at the storage level i.e. on HDFS (Hadoop Distributed File System is suggested. It is done by encrypting the data using AES-MR(XTX) along with compression at Mapper and Reducer phase. AES-MR(XTX) will not only enhance the security of important data at HDFS level but with the help of parallel processing, we can do it with a faster manner as well.

With the help of simulation using NS2 simulator, it is concluded that as the security is key anxiety of the wireless network which transfers data from nodes towards others. In the existing work, the performance of MANET is considered by using different parameters such as Network Overhead, Delay and Throughput. There are a few nodes which increase the system transparency, Delay and reduce Network throughput they are famous as malicious nodes. In proposed work, we used the AES algorithm in WSN and then encryption to expand the demand or of the network. In the future this approach can also be used to get better security in other networks like VANET, SPANs, etc. it has been concluded that AES encryption standard, when used in XTS mode, will give better results in terms of security, Speed of transmission, better filtration of packets and Nodes. This technology can be used in various services provided by the government nowadays where it became mandatory to use Aadhaar card to link with the services to avail benefits. The block size used in AES is comparatively larger and varied, depending upon the requirement. The encryption process of XTS mode in AES is also complex that hackers cannot easily hack the transmitted crucial information. We can conclude that as AES-MR (XTS) helps to attain all the levels of security that too at a faster speed, it is a good approach. There is a problem of mixed traffic and transit routes, typical of medium and small cities. Moreover, public transport vehicles often do not have a permanently fixed route. Therefore, to automate the procedure of fixing the route, the mathematical model has been developed that allows assigning the route as a result of the arrival of vehicles in certain segments.

[1] Public Comments-Modes Development - Block Cipher Techniques, https://csrc.nist.gov/Projects/Block-Cipher-Techniques/BCM/Public-Comments-ModesDevelopment,Comments submitted to Encryption Modes @nist.gov.

[2] Liskov M., Mine Matsu K. Comments on XTS-AES. September 2, 2008 This is a comment in response to the request for comment on XTS-AES, as specified in IEEE Std. 1619-2007 September 2, 2008, https://csrc.nist.gov/csrc/media/ projects /block-ciphertechniques/ documents/ bcm/ comm. ents/xts/xts_co mments-liskov_minematsu .pdf.

[3] Philip Derbeko, Shlomi Dolev, Ehud Gudes, Shantanu Sharma Security and Privacy Aspects in map reduce on Clouds: A Survey, www.https://arxiv.org/abs/1605.00677, www.iiste.org, ISSN 2224-610X (Paper) ISSN 2225-0603 (Online) Vol 2, No.2, 2012.

Analytical and Simulation Approaches in Computer Network Design

Lina Barbaruk
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
barbaruk.angelina@gmail.com

Olha Lavrinenko
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
ya.olia13@gmail.com

Ivan Kvasov
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
ivankvasov65@ukr.net

***Abstract*—This paper proposed a methodological approach to modeling computer network for business center. It describes the main problems that can be encountered when designing a network. Developed by meth-od will help to minimize costs, quickly implement the designed net-work to the enterprise and save time. When selecting a program, all criteria were taken into account, and only after that was NetCracker selected.**

Keywords—simulation modeling, NetCracker, design, statistical data

I. Introduction

From year to year information technologies are developing more and more in all fields of human's activity. Such field as data transmission networks, without which in-formation exchange of data is practically impossible, isn't left without attention. In such a situation, analysis of technical characteristics remains the main most urgent task in the field of information technologies.

Defining and solving a problem with simulation modeling of networks is actual too, it allows to plan prematurely, analyze the possibility of problems' existence in the process of development and exploitation, conduct validation of correctness and ability of network to work on various functional positions, conduct the research of new technologies and mechanisms, but the main thing – is to save money.

The goal is to define all possible problems and factors through namely modeling computer network with help of program NetCracker [1].

The objectives are:

- size of network (number of users, users' location/equipment's location);
- number of data, which was transferred or received;
- data's safety;
- another important factor as requirements to a connection (wired, wireless or combined), cable type, type of used computers/devices and their location;
- network assignments – exchange of information/data between computers, programs/database hosting.

In order to automate the work of business center, the design of an information system consists of:

- distance-vector;
- cable network and active network equipment;
- computer peripheral equipment;
- storage equipment (server);
- system (operation systems, database management systems), special (monitoring and network management systems) and application software.

In order of reducing the time and financial costs of developing computer networks on a local scale, it is necessary to develop a methodological approach to designing a computer network for a business center through modeling.

The network design is a hard process that involves topological design. In the network design process, it is necessary to determine where to place the components and how connect them. The designed network must necessarily meet the needs of a subscriber and operator. Therefore, to achieve this goal, it was necessary to follow clear rules for network design. However, there are two important elements that affect network design – bandwidth and data transfer delay.

Bandwidth in computer network refers to the speed at which data can be transmitted through a network connection or interface. When processing various types of network data, an important element of speed is data transfer delay. Utilities such as ping and traceroute typically estimate delay by determining the time it takes a network packet to move from source to destination and back.

II. Approach to Design Computer Network of Business Center

Using approaches and principles of organization and functioning of networks, a methodological approach to the design of a computer network of business center has been proposed, which involves the implementation of the main stages:

1) *analysis and description* of the design area;

2) determination of the overall structure of the business center;

3) implementation of the topological synthesis of the optimal network structure;

4) formation of the description of the technical equipment of the business center;

5) research of the local computer network using method of simulation modeling.

According to the first design stage, the analysis and description of the design area, the business center, was carried out.

At the second stage, the general structure of the local network is determined.

The third is the choice of topology. The optimal topology for this network is a tree (Fig.1).

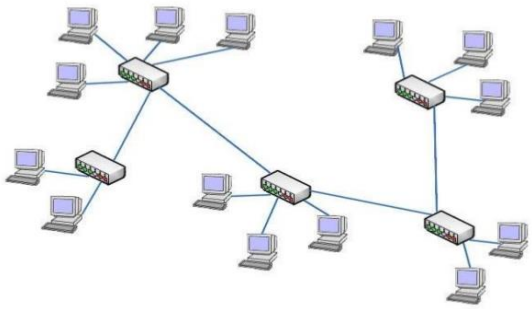

Fig. 1. Tree topology

At the fourth stage, following actions were performed to build the network. Following elements are selected: Digital PC 5510 short-tower-300MHz, ProFast Switch/DES-810, StepServer VL PCI и Cisco 803-BT1. A connection has been created between clients and switches, between switches and router, as well as between a server and a switch. The characteristics of the channel indicate the type of cable (twisted pair), its length and the maximum transmission rate (100 Mbps).

III. Results of Experiment

At the fifth stage experiment in developed simulation model was conducted.

To obtain statistical data on the distribution of the percentage of packets' traffic used by various devices, a traffic distribution law was established with the following data:

1) The exponential law of traffic distribution.

2) The time between transactions (batches) is 0,05 seconds.

3) The packet size transferred between clients and the server was set to 50 Mbps.

4) Simulation time is 10 seconds.

Traffic type has been set: LAN peer-to-peer traffic; FTP; E-Mail (SMTP); HTTP.

After the completion of the simulation experiment, statistical data were obtained. The results of simulation modeling of the department's local network indicate that the average load on communication channels, depending on the operating modes, is 10-50 Mbit/s, and the channels are used by no more than 50%, which meets the requirements.

The calculation was made according to the formula:

$$U = \frac{\max(Q_{max}, Q_{gen})}{Q_{max}} * 100\% , \quad (1)$$

where U is the utilization rate of the mono channel;

Q_{max} - maximum throughput without taking into account the cost of waiting between packet transmission [bit/s];

Q_{gen}- bandwidth, demanded from the mono channel by the station [bit/s];

Q_{gen} = k*Qi, where Qi - loads from each class of the application, [bit/c].

Conclusion

The methodological approach to the design of a computer network of a business center has been proposed. On its basis, a local network was designed, which would allow individual users to easily and quickly interact with each other. The main tasks performed by this network:

- joint work with documents;
- archiving and saving work performed on the server, to save space on the hard disk;
- simplified workflow.

Also, simulation network modeling was done in various conditions. The results of the study indicate that using the proposed approach allows designing in cases where analytical calculations are cumbersome, and sometimes impossible, to identify problem areas of networks and take appropriate measures at the design stage, to select active network and cable equipment.

[1] Yakubova M., Serikov T., "Development and imitating modeling in the developed network consisting of several knots removed among themselves on Netcracker 4.1," 17th International Conference on Micro/Nanotechnologies and Electron Devices EDM, Novosibirsk, Russia. 2016. pp. 210-213.

Color Image Control Using Direct Memory Access

Maksym Hoha
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
maksym.hoha@nure.ua

***Abstract*—The approach of the control the color informational display using modern technologies to simplify the implementation of the data transfer protocol and minimize the use of hardware costs in the conditions of instantaneous and extra-large data transfer in continuous mode was considered in the paper. The software and hardware implementation of the control based on STM32F103C8T6 microcontroller using ARM Cortex-M3 technology, WS2812B color LED ribbon and direct memory access technology was proposed.**

Keywords—data transfer protocol, non-return-to-zero line code, direct memory access, hardware abstraction layer WS2812B, Cortex-M technology, STM32 microcontroller's unit

I. Introduction

Currently, information technologies are widely used in various solutions related to software and hardware implementations: control devices, hardware platforms for applications, and software and hardware implementation of data transfer.

Color display controls the information and, in particular, implements data transfer. This implementation is one of the best, as it contains a small but comprehensive stack of modern technologies that allows you to significantly speed up the data transfer and, thereby, release the main device (microcontroller) from the temporary memory buffer monitoring.

The reason of the inefficient use of hardware resources is follow. The LED strip on the WS2812B has only one digital input – DIN, connected to the first LED on the strip. It is supplied with a special pulse sequence encoding the bits. Each LED needs to transmit 24 bits (8 bits for each color: red, green and blue). Thus, in order to light all LEDs, 24*N bits must be transmitted, where N is the number of LEDs on the strip.

When LEDs have accepted the bits, they are light up and statically glow until receiving a new bit sequence.

The bits are encoded by rather short pulses with tight tolerances. To generate them using microcontroller resources and software delays principle it's necessary to disable all interrupts to avoid the reset or the failed bit forming. The processor time resources are also wasted irrationally.

II. Direct Memory Access

Direct memory access (DMA) is a feature that allows certain hardware subsystems to access main system memory (random-access memory – RAM), independent of the microcontroller unit (MCU). This feature is useful at any time that the MCU cannot keep up with the rate of data transfer, or when the MCU needs to perform work while waiting for a relatively slow input/output (I/O) data transfer [1].

DMA need additional logic to work with the same data, and, in this case, using of DMA is impractical.

The advantage of DMA is to minimize the use of microcontroller resources or completely eliminate its participation in data transfer to interact with other microcontroller's systems (timers, interrupts, input-output ports, analog-to-digital converter (ADC), etc.).

This feature does not affect the MCU at all, because it shares with DMA shared access to RAM and system bus. But measurements have shown that in this case this slowdown does not exceed 0.2%.

III. Software Implementation

A. Protocol Implementation

For the software implementation of the data transfer protocol performed follow settings:

1. Enable the microcontroller timer in the PWM measurement mode.

2. Depending on the clock frequency of the microcontroller bus on which the connection with the timer is located, set the overflow counter with a certain coefficient that depends on the timer prescaler (a special parameter that divides the timer frequency by the specified integer coefficient) and satisfies the condition:

timer frequency(MHz) / 0.8 MHz = overflow factor

The 0.8 MHz is a constant condition for the correct signal transfer intervals implementation according to the protocol. The signal transfer period lasts 1.25μs, which corresponds to:

$$1 / 1.25*10^{-6}\ s = 0.8*10^{5}\ MHz$$

It should be noted that the coefficient of the timer overflow counter is a direct indication of the pulses number, which are needed to transmit a signal. This will help to set the duty cycle and specify the number of pulses to transmit a logical one and a logical zero (need to specify the number of pulses + 1 to bring the signal transfer time closer to the time parameters in the documentation) with the necessary intervals of time respectively.

Also, need to specify the periods number of data transfer reset, which is 50μs, and to create a buffer of a certain size to transfer the generated data, which contains information about the number of LEDs, the reset time before each packet transfer and the packet length.

It is necessary to given the fact that this protocol need to implement the data transfer via the G-R-B sequence and not R-G-B. This fact is described in the WS2812B specification.

B. Data Transfer Implementation Using the DMA

It is necessary to implement a mechanism for transferring data from the MCU to the LED strip to image data in the form of color. Need to use DMA technology for the best decision. The use of technology is based on a pair of interrupts that control the data transfer process. The code description is made using the HAL (Hardware Abstraction Layer) library.

At the time of the data transfer start, it is necessary to notify the microcontroller about the event, which is a prerequisite for calling the event/interrupt handler.

The event (the role of the event is provided to the data transfer start function using DMA) should receive a reference to the used timer, the timer channel (the channels are designed to perform the most frequently used tasks in generating pulse signals and to measure the parameters of external pulse signals and can operate in the output or input mode accordingly.

Then an event / interrupt handler is automatically called which the data transfer must be completed in. Otherwise, an endless transfer of data from the buffer will follow. Continuous data flow will not allow to change the color on the LED strip further.

The completion of the transfer is also a function that plays the role of an event. It notifies the microcontroller when the transfer is over without losing and duplicating information. The parameters of this function are the reference to the timer and its used channel.

CONCLUSION

The use of the modern STM32 family microcontroller with built-in DMA technology allows significantly to increase the speed and volume of information processing, in particular, data transfer, which makes this device the leader in the implementation of software and hardware solutions, both in terms of software and hardware and in material costs comparing with such families as AVR or PIC.

[1] Duyck A. Pushing the Limits of Kernel Networking. URL: https://events.static.linuxfound.org/sites/events/files/slides/pushing-kernel-networking.pdf

Network Communication for Remote Parkinson's Disease Monitoring System

Oleksandr Berezhnyi
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
berejnoi201@gmail.com

Tetiana Biloborodova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
beloborodova.t@gmail.com

Inna Skarga-Bandurova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
skarga_bandurova@ukr.net

Viktoriia Derevianchenko
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
derevianchenkovs@ukr.net

***Abstract* — The earlier treatment of Parkinson's disease (PD)can prevent the disease from developing and to prolongate the diseases prodromal phase. In this context, home monitoring services are potentially powerful tools for remote diagnosis and can improve healthcare services. Tremor is the most common symptom of a PD disorder and it has several advantages for continuous PD symptoms monitoring. The connection between the smartphone application and cloud platform for smartphone sensors data transmission for early tremor symptoms detection is developed. It includes developing of a configuration of a smartphone application for sensor data transmission and developing of a configuration of a cloud platform for tremor symptoms monitoring. The connection settings developed for the system proved to be efficient when sensor data transmitted from the smartphone to cloud storage.**

Keywords — health monitoring system, smartphone application, Parkinson's disease, data tramsmission

I. INTRODUCTION

In recent years, it has become clear that some symptoms of Parkinson's disease (PD) occur decades before the development of motor symptoms and clinical diagnosis, and that monitoring these symptoms may provide earlier PD detection. This may enable earlier treatment to prevent the disease from developing and to prolongate the diseases prodromal phase [1].

Using data from wearable devices, such as smartphone sensors, for movement measurements, the earlier detection of motor symptoms will become possible [2].

Tremor is the most common symptom of a movement disorder [3] and it has followed advantages for continuous PD monitoring. Tremor is the most common symptom of movement disorder, appears at the early stage of diseases, the most obvious symptom, and easy to detect.

Outside the clinic, patient monitoring is used to monitor tremors. Home monitoring services are potentially powerful tools for remote diagnosis and can improve healthcare services.

Smartphone sensors and web-based cloud platform enable remote monitoring, evaluation and daily monitoring of person with PD and earlier detection of PD symptoms.

The proposed solution consists of a smartphone application, a cloud platform that receives and processes daily motion information of a monitored user, gathered from a data of smartphone sensors.

The system aims to provide a tool for the objective and efficient monitor the status of a monitored user in Parkinson's disease context.

The goal of the research is developing a connection between mobile application and cloud platform for smartphone sensors data transmission for early tremor symptoms detection. The objectives are developing of a configuration of a smartphone application for sensor data transmission and developing of a configuration of a cloud platform for tremor symptoms monitoring.

II. SYSTEM ARCHITECTURE

The proposed system allows the monitoring of PD symptoms remotely by using an application installed on the users' own smartphone. The most significant benefit using of smartphone-based approach is high accessibility. Smartphone-based approach helps to enable automated acquisition, transmission, processing and analyzing of the monitored user data. The core of the smartphone application is the testes to monitor and assess PD symptoms, which consists of active tests, that are initiated and self-administered by the participants at various times during the day. The accelerometer and touchscreen used in these tests. The tests are designed to measure symptom of motor function, such as tremor. The system can detect and quantify the motor symptom of PD such as tremor. The system is used Back4App platform to data collection, the processing will be done in the cloud while allowed users can view and manage

all the information related to the monitored user using a web browser.

The system architecture include smartphone, communication channel, cloud storage.

The Back4App platform is offered as cloud storage [4]. The platform provides the follow function: Live Query, JSON Import / Export, Manage Parse Server Versions, Parse Command Line Tool. Back4App platform allows monitoring of several users.

The mobile application of the personal health monitoring system is used to obtain data on the presence or absence of tremor symptoms. Data acquisition occurs when conducting two tests using the built-in mobile phone sensors described above. The tremor test is performed using an accelerometer and a capacitive touchscreen.

The mobile application transmitted data to the cloud using a wireless Wi-Fi or a mobile 2G network. The minimum system requirements of the mobile application have been determined.

III. IMPLEMETATION

In the current release, the personal mobile sensing system includes two tests for evaluating motor symptoms of Parkinson's disease. They are tremor tests using an accelerometer and tremor test using a touchscreen.

Consider the developing of settings of a configuration for a smartphone application for data transmission. At this stage, Android Studio [5] is used. In order to demonstrate the developing of settings of a configuration for data transmission, the accelerometer tremor test is used.

The next one the free cloud storage platform [4] is used. The configuration settings [6] are used. For correct connecting of the smartphone application to cloud storage and data visualization, the libraries [7] must be used.

IV. TESTING AND EVALUATION

In order to demonstrate the connection reliability of the monitoring system, the active tests provided on 2 healthy individuals with no previously diagnosed motor conditions.

As a comparison, the tests were also run an additional 2 times with simulated Parkinson's disorder, such as tremor. The data was transmitted from smartphone application using Wi-Fi network to cloud storage. The time delay of data transmission is up to 1 seconds. The dashboard on Back4Upp platform used for real-time data visualization.

CONCLUSION

The settings of configuration connection for the continuous home assessment of PD symptoms, such as tremor, using smartphone embedded sensors are developed.

The connection settings developed for the system proved to be efficient when sensor data transmitted from the smartphone to cloud storage. The period of time required to transfer data to the cloud equal to the period of time less than one second.

[1] Brundin, P. and Bloem, B.R., 2018. The Times They Are a-Changin': Parkinson's Disease 20 Years from Now. Journal of Parkinson's disease, 8(Suppl 1), p.S1.

[2] Son, H., Park, W.S. and Kim, H., 2018. Mobility monitoring using smart technologies for Parkinson's disease in free-living environment. Collegian, 25(5), pp.549-560.

[3] Jankovic, J., 2008. Parkinson's disease: clinical features and diagnosis. Journal of neurology, neurosurgery & psychiatry, 79(4), pp.368-376. DOI:10.1136/jnnp.2007.131045.

[4] Back4App [Electronic resource]. – Avaible at : https://www.back4app.com.

[5] Android.com [Electronic resource]. – Avaible at : https://developer.android.com/studio/.

[6] Github.com [Electronic resource]. – Avaible at : https://github.com/beloborodova-t/ALIOT-47/tree/master/pack.

[7] Github.com [Electronic resource]. – Avaible at : https://github.com/beloborodova-t/ALIOT-47/tree/master/libs

Design of Algorithms for Technological Processes Control

Volodymyr Kardashuk
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
kardashuk1@gmail.com

***Abstract*—The results of designing the program library of algorithmic elements for the creation of algorithms and construction of control schemes for technological processes are presented in order to accelerate the formation of the final structure of the system.**

Keywords—software library, algorithmic element, technological process, control system

I. INTRODUCTION

Algorithmic design is to develop algorithms for the functioning and creation of software for computer systems.

Acceleration of the creation of control algorithms, the construction of schemes of interconnections of the system, the identification of priority control channels - is a topical issue in the design of control systems of technological processes.

Circuits of control and executive mechanisms, which constantly interact with the software system, receive commands from the controller, which provides centralized control of the technological process.

Modern design technologies for the creation of control systems for technological processes involve the use of pre-drawn templates that are made out in software libraries of algorithmic elements.

The purpose of the project is to create a program library of algorithmic elements for accelerating the design of control systems, and, if necessary, rapid modernization of algorithms.

II. PROPOSED APPROACH

The process of creating the finite control algorithms at the design stages consists in solving complex problems - from the creation of adaptive parametric models to the construction of a work process using computer technology.

When implementing control algorithms, the software library needs constant updating and support.

The proposed software library of algorithmic elements used in the design systems of process control circuits contains:

- combinational logic elements (AND, NOT, OR, MOD2, encoders, decoders, multiplexers, etc.);
- elements of a serial type (RS-trigger, D-trigger, registers, counters, etc.);
- dynamic elements (elements of integration, differentiation and filtering);
- static elements (adder, elements of subtraction, division, square root product, etc.)
- elements of nonlinearity (elements of restriction zones, insensitivity, elements of the relay with hysteresis, etc.);
- elements of regulation (regulators of the laws of regulation).

In more complex technological schemes, such as the digital proportional–integral–derivative controller (PID controller), which is more versatile than other regulators, it is possible to use one of the well-known regulatory laws.

In the XZ, the digital PID controller analyzes the XOS feedback and issues the control signal Y in case of deviation (Fig. 1).

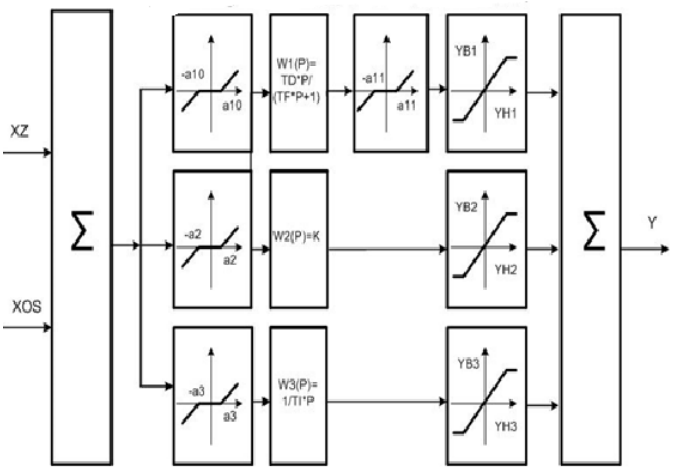

Fig. 3. Digital PID controller

The output signal of the digital controller is determined by three components: the proportional, differential and integral parts of the equation.

The presented regulator is a program element that implements a discrete part of the equation:

$$W_{PID}\ (p) = K_p(1 + \frac{1}{T_i \cdot p} + T_d \cdot p) \qquad (1)$$

Graphic editor of functional schemes is used to solve the following tasks of the technological chain:

- constructing / editing a graphic image of a single element of a functional schema;
- build / edit a functional schema.

The development of graphic images for constructing algorithms is carried out with the help of Altium Designer software company Altium, which, according to the developers of the company, reduces the time for designing the final solution in 2 times [1].

III. Implementation

For a program description of the functional purpose of algorithmic elements, the programming language C ++ [2] is used, which allows the use of software in various operating systems.

At the final stage of design, a set of automata (algorithms) of control on the basis of algorithmic elements is combined into a single program, loaded into the controller's memory and launched for execution in sequence in the order of the set queue or when receiving control in accordance with the conditions.

In case of necessity the introduction of changes in the technological scheme is carried out by replacing one element with another or by adding the necessary element / elements, which reduces the time for the program implementation of the control system.

Conclusion

The advantages of using such software libraries are the study of many options for using the model at any stage of the design without the restructuring of the control system, rapid upgrading of algorithms, thus creating precise models of objects consisting of dozens of elements.

As a result of implementing the goal of designing a database of existing graphic images (Schematic Library Document), stored in * .PsbSch format, is supplemented with new elements that allow the implementation of elements of integration, differentiation, adder, etc.

[1] Simple, modern, powerful. Available at https://www.altium.com/altium-designer/ru.

[2] Schemes creating. Available at https://www.altium.com/altium-designer/ru/playlists/schematic-design.

Efficient Streaming on Multi-Core of Shared-Memory Computing

Dmytro Nedzelskyi
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
nedzelsky946@gmail.com

***Abstract*—The article investigates the dependence of the number of effectively executed programs in multi-core computers with shared memory on the parameters of programs and computers. All computer cores execute parallel streams of a single program developed in accordance with the OpenMP API. There are no interactions between streams program. Conflicts can occur only when cores are accessed into shared memory. Suggested the necessary models. Analytical expressions are obtained for the dependence of the number of effectively executed programs on the properties of programs, core and shared memory parameters. The main reason for limiting the number of effectively executed programs is the overload of the shared memory. The threshold value of the number of effectively executed programs has been determined.**

Keywords—multi-core computer, core, shared memory, acceleration factor, efficiency, stream

I. INTRODUCTION

Efficient utilization of the cores in the computers with a high number of cores is one of the most important problems for designers of software for these systems [1]. Multi-core computers performance is defined using acceleration rate when executing an application with parallel streams or a set of applications without parallel streams when utilizing n cores.

Experimental approaches do not generalize well and discovered accelerators are not always applicable or even partially applicable to the next application. Also, the features which cause acceleration and the reasons for the limit of acceleration where an increase in a number of cores does not increase the performance any longer did not analyze earlier.

Existing works in multi-core computer systems do not have an analytical approach estimating of the number of effectively used cores.

The purpose of the article is to study the efficiency of using cores (threads) in multi-core computers with shared memory depending on the number of cores, properties program performed by the cores, core and shared memory parameters when executing parallel programs developed in accordance with the OpenMP API.

II. EFFICIENT UTILIZATION OF A MULTI-CORE COMPUTER WITH N CORES EXECUTING SINGLE PARALLEL APPLICATION

Efficient utilization of multi-core computer is calculated using the acceleration rate. The acceleration rate of a multi-core computer is defined as:

$$S = \frac{T_1}{T_n} = \frac{\mathrm{Pr}_n}{\mathrm{Pr}_1}, \quad (1)$$

where T_1 – application execution time of a single core, T_n – application execution time using *n* cores where the application consists of exactly *n* streams, Pr_1 - performance of a single core when there is no influence from other cores, Pr_n - performance of an *n* core computer.

Since all *n* cores perform the same stream of a parallel application the core utilization rate is the same for all cores and utilization rate of the multi-core computer in this case equals the single core utilization rate $H_k = H_i$, where H_i - single core utilization rate, when all *n* cores are executing a parallel program; H_k - multi-core computer utilization rate when all *n* cores are executing a parallel program.

Multi-core computer performance where all n cores are utilised is defined as

$$\mathrm{Pr}_n = \sum_{i=1}^{n} H_i * \mathrm{Pr}_1 = \mathrm{Pr}_1 * \sum_{i=1}^{n} H_i = \mathrm{Pr}_1 * n * H_k \ .$$

Multi-core computer acceleration rate in this case equals

$$S = \frac{\mathrm{Pr}_n}{\mathrm{Pr}_1} = \frac{\mathrm{Pr}_1 * n * H_k}{\mathrm{Pr}_1} = n * H_k .$$

III. PROPOSED MODEL OF A MULTI-CORE COMPUTER

Multi-core computers can be described as a 2-phase system serving a high volume of requests. Phase 1 comprises of n cores executing applications and generating requests to the shared memory. Phase 2 occurs in shared memory itself. Buffers with capacity for k requests are placed between phases.

Phase 1 of the equivalent model consists of an equivalent core. The equivalent core performance (frequency of generated requests) is assumed to be a sum of performances of all original model cores

$$\lambda_{eqv} = \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} \frac{\omega_{MEM}^i P_{MEM}^i}{t_{INS}^i}.$$

The equivalent model has a single buffer and its size is the sum of the original model buffer sizes.

Phase 2 of the equivalent model is memory. Its performance (execution frequency) is equal to that of the original model. Buffer selection and core blocking logic are identical in both models.

Hence, from the performance point of view, the original and simplified models of a multi-core computer are equivalent.

The overall load rate of the shared memory for the n-core computer is calculated according to the formula

$$\rho_{MEM}^{\Sigma} = \frac{\sum_{i=1}^{n} \lambda_i}{\mu_{MEM}} = \frac{\sum_{i=1}^{n} \frac{\omega_{MEM}^{i} P_{MEM}^{i}}{t_{INS}^{i}}}{\mu_{MEM}} = \sum_{i=1}^{n} \rho_{MEM}^{i},$$

where ρ_{MEM}^{Σ} - memory load rate from all cores; ρ_{MEM}^{i} - memory load rate from i-th core.

Solving simultaneous equations for selected overall load rate ranges we obtain the expressions for the model utilization rates: equivalent core utilization rate and shared memory utilization rate.

To simplify calculations, we assume an infinite buffer size between a core and the shared memory. The limit of core utilization rate with an infinite buffer size for selected ranges of shared memory load rates is defined below:

$$\lim_{n\to\infty} H_k = \frac{1}{\rho_{MEM}^{\Sigma}} \quad \text{- if} \quad \rho_{MEM}^{\Sigma} \succ 1$$

$$\lim_{n\to\infty} H_k = 1 \quad \text{- if} \quad \rho_{MEM}^{\Sigma} \leq 1.$$

There is an upper limit (threshold) on the number of cores after which the performance of a multi-core computer does not improve.

This threshold (practical number of cores (*n*)) for parallel applications designed using OpenMP API defines the overall shared memory load rate generated by all cores of the multi-core computer.

$$\rho_{MEM}^{\Sigma} = \sum_{i}^{n} \rho_{MEM}^{i} = 1.$$

The multi-core computer acceleration rate (S) while executing parallel applications can be calculated using defined above overall shared memory load rate:

$$S = n \quad \text{- when} \quad \rho_{MEM}^{\Sigma} \leq 1;$$

$$S = \frac{n}{\rho_{MEM}^{\Sigma}} \quad \text{- when} \quad \rho_{MEM}^{\Sigma} \succ 1.$$

CONCLUSION

The proposed methodology can be used to estimate from above the number of threads and, accordingly, the number of cores when developing parallel applications in accordance with the OpenMP API.

Shared memory overload is the main cause of a multi-core computer sublinear acceleration rate when executing the parallel application using OpenMP API. The overload will cause the cores to be idle for a significant percentage of the time.

[1] Pakhomov S. Investigation of the effectiveness of multi-core systems using the applications of molecular dynamics LAMMPS and NAMD. August 30, 2016 http://designinfun.blogspot.com/2016/08/lammps-namd.html.

Methodology of Fake News Detection

Mykyta Davidenko
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
nickita.davidenko@gmail.com

Tetiana Biloborodova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
beloborodova.t@gmail.com

***Abstract*— Most existing approaches require labeled fake news (FN) train dataset to train a model. It is important to consider scenarios where limited or no labeled FN items are available in which semi-supervised or unsupervised models can be applied. The unsupervised technique to identify FN without the training data is designed and developed. It includes data preprocessing, text feature extraction, converting words to vectors and clustering using the k-means algorithm. The clustering is evaluated through several parameters. The results of the Silhouette clustering coefficient as the main evaluation method, we can conclude that the clustering model is accurate. Follow clustering evaluation, the word2vec model showed quality results of vectorizing and searching similar words to build distances between objects for the follow clustering analysis..**

Keywords*— *fake news detection, text mining, clustering, word2vec

I. INTRODUCTION

Nowadays, fake news (FN) detection is an actual research problem. Big social networking services companies are developed solutions for FN recognizing. For example, Facebook allows users scoring the news that is possibly suspicious [1].

The research approaches for FN detection can be defined in the following categories [2]: data-oriented, feature-oriented, model-oriented, application-oriented.

We detailed investigation and analysis of the model-oriented FN research. Most approaches include extracting various features and follow using these features into supervised classification models [3, 4]. However, the accuracy of FN detection, is still challenging, due to the dynamic nature of social media, and the complexity and diversity of online text data. Also, in the absence of high-quality training data is a problem for the creation of detection models.

One of the major challenges for FN detection is the fact that each feature, such as source trustworthiness, style of news text, or social response, has some restriction to directly predict FN singly. The process of obtaining a reliable FN dataset is difficult for the following reasons [5]: the real-world online dataset is usually big, incomplete, unstructured, unlabeled, and noisy; everyday a large amount of false information with diverse intentions and different linguistic characteristics is created via social media.

Also, most existing approaches require a labeled FN train dataset to train a model. It is necessary to design an approach which is able to identify FN even without the training samples. It is important to design effective, automatic and applicable approaches for high-quality online FN detection.

The goal is effectiveness model-oriented approaches for FN detection. The objectives can be defined as follow: study and analyzed related work for FN detection and feature extraction; designed and development of unsupervised technique for FN detection consisting of data preprocessing, text feature extraction, converting words to vectors and clustering using k-means algorithm; clustering evaluation.

II. CLUSTERING FOR FAKE NEWS DETECTION

According to the model-oriented approach for FN detection can be used unsupervised and supervised machine learning technique. We study classification as supervised learning, and perform clustering as unsupervised learning.

The clustering in the context of FN detection can be defined as follow [6]. Given a corpus of FN $J = \{j_1, j_2, j_3, ..., j_n\}$ with size of n where each document $\overrightarrow{n_i}$ is a vector of terms in a dictionary, $\sum = \{t_1, t_2, t_3, \dots, t_T\}$ with size of $T = |\sum|$. The problem is clustering of documents based on their terms into homogeneous classes with respect to FN categories. To this end, we first cluster documents based on appearance positions of each term in an article and its correlations with other terms (Spatial relation extraction) following by designing an automatic ensemble co-clustering to cluster documents according to their positions in different factors among various low-rank decompositions.

III. FEATURE EXTRACTION

The wor2vec model is a popular technique for feature extraction in text mining. It can be describing as follow. The word2vec model takes a large text corpus as input and maps each word to a vector, giving the coordinates of the words in the output. At the first stage, it creates a dictionary through learning on the input text data, and then calculates the vector of words. The vector is based on contextual proximity: words found in the text follow to identical words (therefore having similar meanings) in the vector have close coordinates of word vectors. The parameters and corpus sizes are affected to the model accuracy. Accuracy increases overall if the number of

words used increases, and if the number of dimensions increases [7].

IV. Implementation Result

The dataset by George McIntire [8] was used. The dataset was prepared in 2017. It consists of data from 5279 articles. The articles came from media organizations such as the New York Times, WSJ, Bloomberg, NPR, and the Guardian and were published in 2015 or 2016. The dataset consists of the headline and text of a news article as input variables and output variable with the two classes: FAKE or REAL.

The FN detection was conducted as follow.

The raw data was preprocessed. The bad characters, tokenize and stop words are removed.

After training the model with the data generated from the example sentence above, we can see that the model can output most of the similar words for each word as an input word.

The follow steps after word2vec are converting words to vectors and using K-Means algorithm for clustering.

There are many parameters are used for clustering evaluation. The follow parameters are used in our experiment: homogeneity, completeness, V-measure, silhouette, Adjusted Rand index and Silhouette coefficient.

The results of evaluations parameters are:

Homogeneity: 1.000
Completeness: 1.000
V-measure: 1.000
Adjusted Rand Index: 1.000
Silhouette Coefficient: 0.326

Conclusion

The most of approaches require labeled data for accurate FN detection. The models created by supervised classification methods may be more accurate given a high-quality dataset for training. However, unsupervised models don't require labeled data and can be more practical because unlabeled datasets are more available to obtain. The unsupervised technique to identify FN without the training data is designed and development. It consists of follow steps: data preprocessing, text feature extraction, converting words to vectors and clustering using the k-means algorithm.

We provide clustering evaluation through several parameters: homogeneity, completeness, V-measure, silhouette, Adjusted Rand index and Silhouette coefficient. Based on the results of the Silhouette clustering coefficient as the main evaluation method, we can conclude that the clustering model is accurate.

Follow clustering evaluation, the word2vec model showed quality results of vectorizing and searching similar words to build distances between objects for the follow clustering analysis.

[1] News Feed fyi. Addressing hoaxes and fake news [Electronic resource]. – Avaible at : https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/. Accessed: 2019-04-15.

[2] Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H., 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), pp.22-36.

[3] Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J. and Stein, B., 2017. A stylometric inquiry into hyperpartisan and fake news. arXiv preprint arXiv:1702.05638.

[4] Tacchini, E., Ballarin, G., Della Vedova, M.L., Moret, S. and de Alfaro, L., 2017. Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506

[5] Zhang, X. and Ghorbani, A.A., 2019. An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management

[6] Hosseinimotlagh, S. and Papalexakis, E.E., 2018. Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. MIS2, Marina Del Rey, CA, USA

[7] Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

[8] Github.com [Electronic resource]. – Avaible at : https://github.com/ricknta/fake-news. Accessed: 2019-04-15

Hyperparameters Optimization in CNN for Image Recognition

Rostyslav Siriak
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
hashem.r@gmail.com

Inna Skarga-Bandurova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
skarga_bandurova@ukr.net

Tetiana Biloborodova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
beloborodova.t@gmail.com

***Abstract*—The necessity of creating a model of recognition of gestures based on convolutional neural network that effective not only in pattern recognition, but also in terms of learning speed and resource intensity, is substantiated. In this regard, the work solved the problem of optimization of hyperparameters and the selection of the best optimizer backpropagation errors. To implement the tasks, a model was created that can recognize hand gestures, both from a single image and from streaming video. When choosing an optimizer, two adaptive methods were tested - Adadelta and Adam. The experiments confirmed the high efficiency of Adadelta, however, when compared with Adam, it showed more than twice as long network training.**

Keywords—hyperparameter, convolutional neural network (CNN) , adaptive methods

I. INTRODUCTION

Model optimization is one of the challenging tasks in the development and implementation of machine learning solutions. Hyperparameters are settings that can be configured to control the behavior of a machine learning algorithm. A feature of the hyperparameters is that they are specific to the type of machine learning model that needs to be optimized. In some cases, the parameter is modeled as a hyperparameter, because it cannot be studied from the training set. Examples of hyperparameters are the following: speed of model learning; the number of hidden units; convolutional kernel width, etc.

The criteria for defining the hyperparameters are very abstract and flexible. Data science specialists usually spend a lot of time setting up hyperparameters to achieve the best performance for a particular model. In many cases, the hyperparameter optimization is considered as a global optimization of a black-box error function f whose evaluation is expensive [1]. Solving this problem is very challenging due to high complexity of the function f and depends on the current task.

In this context, the goal is to create a gesture recognition model based on a convolutional neural network (CNN), effective not only in pattern recognition but also in terms of learning speed and resource intensity. In this connection, the optimization problems of hyper-parameters and the selection of the best back-propagation optimizer were solved.

II. EXPERIMENTAL SETUP

CNN solves two problems: (1) determining the attributes of objects (2) determining the probability of the object belonging to a particular class.

In the convolutional core layer, a linear transformation is applied to each pixel of the image, revealing the characteristic features of this class of images. The use of three convolutional layers seems to be optimal in terms of the quality of training to a small number of parameters. The number of parameters for a convolutional layer is defined as:

$$p = d * k * w + b \ , \tag{1}$$

where d is the depth of the input data (the number of channels), k is the core, w is the weights, b is bias.

The result of the convolution must be passed through the nonlinear activation function.

The feature maps formed on each layer are fed to the max-pooling sub-sampling layer.

The max-pooling layer with a 2x2 filter follows each convolution layer, enhances the features identified on the previous layer and reduces the dimension of the input data and, as a result, the number of parameters.

The size of the output from the max-pooling layer is defined as:

$$n_{out} = \left\lfloor \frac{n_{in} - k}{s} \right\rfloor + 1 \ , \tag{7}$$

where n_{in} is the size of the input data from the previous layer, k is the size of the core, s is the step size with which the core is shifted in the image.

Max-pooling has no options for training. In order to prevent retraining, in which the trained model too closely matches a given data set, losing the ability to generalize, a dropout is used in the developed neural network [2].

The second part of the created network is a multi-layered perceptron and is intended for training the hand gesture classifier on the identified features. Data from the last down

sampling layer is fed to the Flatten layer, thereby transforming into a one-dimensional vector.

The last layer of the network with the number of outputs equal to the number of recognized categories, implements the SoftMax activation function.

SoftMax assigns a value, represented by a non-negative real number, to each class, expressing the probability of belonging.

The created CNN consists of two parts: a) three blocks of alternating convolutional and subsampling layers that form the input feature vector for learning; b) three fully connected layers.

Convolutional layers apply a 3×3 convolution kernel and form 16, 32, and 64 feature maps, respectively. Feature maps pass through subsampling layers with a max-pooling of 2×2, each time halving the data dimension.

The classifying part of the network consists of a flatten layer, each node of which corresponds to one value of the feature vector, and two fully connected layers dense. The last layer is the output and implements the SoftMax function.

CNN training. Network training is based on the backpropagation algorithm. Backpropagation can be viewed as a differentiation of a complex function with the search for derived loss with respect to variables.

Depending on the results of calculating the loss function, the parameters change at each iteration. Changing the values of weights will be carried out through the gradient descent. Thus, the task of training the classifier S is to minimize the loss function in the space of weights.

The next is choosing an optimizer. When choosing an optimizer, two adaptive methods were tested - Adadelta [3] and Adam [4].

Adadelta is essentially an extension of another AdaGrad optimizer, which had a problem with reducing learning speed. The problem arose due to the accumulation of the sum of squares of gradients.

In Adadelta, instead of the total sum of updates, the averaged square of the gradient is used, for which the exponentially decaying running average is used.

The second tested optimizer of gradient descent was the Adam adaptive inertia method. It calculates adaptive learning rates based on the first and second gradient moment scores. The estimate of the first moment is calculated from the previously obtained partial derivative values, like a moving average of gradients.

A significant increase in speed help to reduce the effectiveness of network training and increased the number of incorrect answers. Decreasing speed made the network long-learning.

The experiments performed show a sufficiently high Adadelta performance, however, when compared with Adam, it showed more than twice the network's long learning ability.

Having error values, we calculate the partial derivative of the objective function E with respect to each output of the neuron. To calculate the loss derivatives for the variables in the embedded equation, the chain rule is applied.

For a convolutional layer, the backpropagation procedure is used.

In the max-pooling layer, an error from the previous layer passes through a single maximum value. Since this layer does not train the network, the error passes through it unchanged.

III. Result

During the experiments, we used our own database containing six gestures: "fist", "one", "palm", "letterSH", "two", "zero." The total number of video files is 7674. The model was built using the Keras library and the TensorFlow framework. In the process of learning, preprocessing was carried out: the image of the hand was segmented and transferred to the grayscale format, after which the Canny edge detector was used to highlight the hand contours. The processed images were normalized and fed to the input of CNN with three coagulation layers and three fully connected. The last layer was the Softmax classifier whose number of neurons is equal to the number of object classes. When choosing an optimizer, two adaptive methods were tested - Adadelta and Adam. The experiments confirmed the high efficiency of Adadelta, however, when compared with Adam, it showed more than twice as long network training. As a result of experiments on a test subset with a size of 1,535 images, recognition accuracy of 94% was achieved.

Conclusion

In this work, we performed optimization hyperparameter manually that is exhausting and unexpanded. To address these challenges, the algorithms such as Grid Search and Random Search that automatically infer a potential set of hyperparameters and attempt to optimize them could be used. In the course of testing with gesture recognition, a high degree of correct network responses was obtained in real-time on a webcam, including people who did not participate in the creation of an image database. Future work will focus on the development of the effective method for automatic optimizing hyperparameters of deep learning algorithms.

[1] Ilievski I., Ahkar T., Feng J., Shoemaker Ch.A. Efficient Hyperparameter Optimization of Deep Learning Algorithms Using Deterministic RBF Surrogates. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17). P. 822-829. 2017.

[2] Hinton, Geoffrey E., et al. "Improving neural networks by preventing co-adaptation of feature detectors." arXiv preprint arXiv:1207.0580 (2012).

[3] Zeiler, Matthew D. "ADADELTA: an adaptive learning rate method." arXiv preprint arXiv:1212.5701 (2012).

[4] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

Justification and Practical Application of Routing Protocols

Lina Barbaruk
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
barbaruk.angelina@gmail.com

Alice Mikhaylova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
alicedemoran@gmail.com

Denis Bakitko
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
bakitko_denis@ukr.net

***Abstract*—This article analyzes the work of the OSPF routing protocol. Considered and justified the need for its use, identified the advantages and disadvantages. Analyzed the main features. A comparative analysis of the protocols OSPF and RIP. Based on the work done, conclusions formulated.**

Keywords—OSPF, RIP, routing protocol, convergence, link redundancy, metrics, interface cost, routing table, packet hops, Cisco Packet Tracer

I. INTRODUCTION

Network configuration is a complex and often multi-level process, the consequences of which have a huge impact on the future operation of the network. One of the main tasks that need to addressed when setting up a network is the choice of a routing protocol.

Dynamic routing protocols are usually divided by the type of algorithms used in them into:

- distance-vector;
- link-state;
- mixed type.

In fact, the protocols, taking into account the state of the channels, have replaced the distance vector protocols. The following dynamic routing protocols exist RIP, OSPF, EIGRP, BGP, IS-IS.

In this article, the main characteristics and operation principles of the protocol will be examined taking into account the state of the OSPF channel based on its comparison with the distance vector protocol RIP. A comparison will be made using the Cisco Packet Tracer simulation program.

II. COMPARATIVE ANALYSIS OF OSPF AND RIP PROTOCOLS BY MAIN CRITERIA

Comparison of protocols carried out on the basis of key indicators, including security, load balancing, type of algorithm, and some others. The protocol metrics and the routing table configuration were considered in most detail. In addition, the issue was considered the issue of convergence. Routing protocols are often evaluated by the convergence rate, which means that the faster the convergence performed, the more efficient the protocol is. OSPF, as a relatively new protocol, provides the fastest convergence, which makes it one of the most preferred.

A. Convergence

Network convergence is a necessary characteristic, without which the network is not fully operational. Convergence refers to the process of agreement between all routers on the best routes. Routing algorithms with poor convergence lead to creating routing loops or network failure.

One of the main principles of OSPF operation is that each router inside a zone stores the full topology of its zone. The time of bringing the network into this state is called convergence [4].

B. Protocol metrics

The metrics used by the routing protocols have a direct impact on creating the best route with the least number of hops. OSPF and RIP use different metrics, which described in more detail below.

In OSPF, when choosing the best route, a metric called cost used. It said that each link has a cost, respectively, if the route passes through several links, then their cost summed up. But the best route is the one whose cost is the lowest. The cost of an interface is inversely proportional to its bandwidth [2].

Cisco provides the following costing options:

- Cost calculated as the inverse of the link speed value.
- The cost for each link set by the user manually, based on personal ideas about the quality of this link.

The cost calculated by the formula (1):

$$\rho = \alpha / \beta \tag{1}$$

where ρ is the cost, α is the specified bandwidth, and β is the bandwidth of interface.

RIP as a metric used packet jump - this is the number of routers through which a packet can pass along a given route. Each hop in the path from the source to the destination assigned a value of the number of hops, which is usually 1. When the router receives an update of the routing information that has a new or modified destination record, the router adds 1 to the metric value specified in the update and writes the network to

routing table the sender's IP address is used as the next hop. A directly connected network to the router has a metric of zero; the inaccessible network has a metric 16 [4]. A more detailed analysis conducted in Table I.

TABLE I. TABLE OF COMPARISON

Indicator	OSPF	RIP
Security	Open password or key authentication MD5	Open password or key authentication MD5
Load balancing	Same metrics	-
Algorithm Type	Link-state	Distance vector
Combining routes	+	-
Variable-length subnet masks	+	+
Maximum number of routers in the network	65534	15
Accounting in the metric of various characteristics of the path	One main and three additional	One main
Update routing information	Only changes	Whole table
Availability of implementation	Open	Open
IPv6 support	+	-

C. *Configuring Routing*

To configure dynamic routing, the following commands used: router and network. The router command that starts the routing process has the form:

Router(config)# *router protocol [keyword]*

where protocol is one of the routing protocols and keyword is an optional parameter.

The network command has the form:

Router (config-router)# *network network-number [keyword]*

where network-number is the identifier of the connected network and keyword is an optional parameter [5].

When using the OSPF protocol, when the router configured to work with the protocol, it begins to explore the environment through the following initialization stages [6]:

- Use hello to find
- The phase of first exchange between route bases.
- Sending route information with later confirmation.
- Compiling the routing table.
- Transition to full status.

When using the RIP protocol, the routing process connects the interfaces with the addresses that correspond to them and begins processing packets in specified networks.

III. SIMULATION AND COMPARISON

A. *Cisco Packet Tracer Features*

This simulator was developed by Cisco and recommended for use in the process of studying telecommunication networks and network equipment. But the Cisco Packet Tracer features are suitable not only for training but also for setting up any network at the planning stage.

It includes the following features:

- Workspace, which used to create a network.
- Simulation, both in simulation mode and real-time.
- A graphical interface that used to interact with the user during the configuration process.

With the help of commands that entered into the command line of devices, you can display a variety of information about the network, including the routing table [7].

B. *Network modeling*

For a more detailed analysis of operating the OSPF protocol, a network modeled. Simulations performed using the Cisco Packet Tracer simulator. The network was configured in two ways: using OSPF and RIP. An example of packet transmission shown in Fig. 1

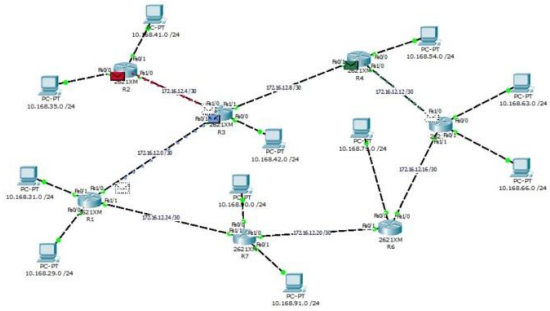

Fig. 1. Packet transmission.

C. *Routing Configuration Comparison*

Configuring network routing using OSPF (2) and RIP (3) presented below. For example, the router R1 taken. Setup made in the command line.

```
R1>ena
R1#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
R1(config)#router ospf 1
R1(config-router)#network 10.168.29.0 0.0.0.255 area 0
R1(config-router)#network 10.168.31.0 0.0.0.255 area 0
R1(config-router)#network 172.16.12.1 0.0.0.3 area 0
R1(config-router)#network 172.16.12.26 0.0.0.3 area 0
```

Fig. 2. Network configuration setup (OSPF).

```
R1>ena
R1#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
R1(config)#router rip
R1(config-router)#network 10.168.29.0
R1(config-router)#network 10.168.31.0
R1(config-router)#network 172.16.21.1
R1(config-router)#network 172.16.12.25
```

Fig. 3. Network configuration setup (RIP)

The difference in configuration setting is only in the advanced parameters — the inverse subnet mask and indication of the zone used in the OSPF protocol. Further on Fig. 4-5 shows the routing tables of two networks:

```
     10.0.0.0/24 is subnetted, 11 subnets
C       10.168.29.0 is directly connected, FastEthernet0/0
C       10.168.31.0 is directly connected, FastEthernet0/1
O       10.168.35.0 [110/3] via 172.16.12.2, 00:57:06, FastEthernet1/0
O       10.168.41.0 [110/3] via 172.16.12.2, 00:57:06, FastEthernet1/0
O       10.168.42.0 [110/2] via 172.16.12.2, 00:57:06, FastEthernet1/0
O       10.168.54.0 [110/3] via 172.16.12.2, 00:57:06, FastEthernet1/0
O       10.168.63.0 [110/4] via 172.16.12.2, 00:57:06, FastEthernet1/0
                    [110/4] via 172.16.12.25, 00:57:06, FastEthernet1/1
O       10.168.66.0 [110/4] via 172.16.12.2, 00:57:06, FastEthernet1/0
                    [110/4] via 172.16.12.25, 00:57:06, FastEthernet1/1
O       10.168.79.0 [110/3] via 172.16.12.25, 00:57:06, FastEthernet1/1
O       10.168.90.0 [110/2] via 172.16.12.25, 00:57:06, FastEthernet1/1
O       10.168.91.0 [110/2] via 172.16.12.25, 00:57:06, FastEthernet1/1
    172.16.0.0/30 is subnetted, 7 subnets
C       172.16.12.0 is directly connected, FastEthernet1/0
O       172.16.12.4 [110/2] via 172.16.12.2, 00:57:06, FastEthernet1/0
O       172.16.12.8 [110/2] via 172.16.12.2, 00:57:06, FastEthernet1/0
O       172.16.12.12 [110/3] via 172.16.12.2, 00:57:06, FastEthernet1/0
O       172.16.12.16 [110/3] via 172.16.12.25, 00:57:06, FastEthernet1/1
O       172.16.12.20 [110/2] via 172.16.12.25, 00:57:06, FastEthernet1/1
C       172.16.12.24 is directly connected, FastEthernet1/1
```

Fig. 4. Routing Table (OSPF)

```
    172.16.0.0/30 is subnetted, 7 subnets
C       172.16.12.0 is directly connected, FastEthernet1/0
R       172.16.12.4 [120/1] via 172.16.12.1, 00:00:06, FastEthernet1/0
R       172.16.12.8 [120/1] via 172.16.12.1, 00:00:06, FastEthernet1/0
R       172.16.12.12 [120/2] via 172.16.12.1, 00:00:06, FastEthernet1/0
R       172.16.12.16 [120/2] via 172.16.12.25, 00:00:00, FastEthernet1/1
R       172.16.12.20 [120/1] via 172.16.12.25, 00:00:00, FastEthernet1/1
C       172.16.12.24 is directly connected, FastEthernet1/1
C     10.168.29.0/24 is directly connected, FastEthernet0/0
C     10.168.31.0/24 is directly connected, FastEthernet0/1
R     10.168.35.0/24 [120/2] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.41.0/24 [120/2] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.42.0/24 [120/1] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.54.0/24 [120/2] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.63.0/24 [120/3] via 172.16.12.25, 00:00:00, FastEthernet1/1
                     [120/3] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.66.0/24 [120/3] via 172.16.12.25, 00:00:00, FastEthernet1/1
                     [120/3] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.79.0/24 [120/2] via 172.16.12.25, 00:00:00, FastEthernet1/1
R     10.168.90.0/24 [120/1] via 172.16.12.25, 00:00:00, FastEthernet1/1
R     10.168.91.0/24 [120/1] via 172.16.12.25, 00:00:00, FastEthernet1/1
```

Fig. 5. Routing Table (RIP)

Both routing tables show directly connected devices and the paths along which devices that at distance from the router are connected, as well as the ports used for this.In addition, we can see the difference in the administrative distance, which is 110 for OSPF and 120 for RIP.The designation O is the address formation by the OSPF protocol, and R is the address formation by the RIP protocol, respectively.

D. Comparing Metrics

An important difference between OSPF and RIP could be observed in the metrics that they use. These data could be viewed, and in some cases even changed, when you configure the network in Cisco Packet Tracer. In Fig. 6 marked packet hop for RIP:

```
    172.16.0.0/30 is subnetted, 7 subnets
C       172.16.12.0 is directly connected, FastEthernet1/0
R       172.16.12.4 [120/1] via 172.16.12.1, 00:00:06, FastEthernet1/0
R       172.16.12.8 [120/1] via 172.16.12.1, 00:00:06, FastEthernet1/0
R       172.16.12.12 [120/2] via 172.16.12.1, 00:00:06, FastEthernet1/0
R       172.16.12.16 [120/2] via 172.16.12.25, 00:00:00, FastEthernet1/1
R       172.16.12.20 [120/1] via 172.16.12.25, 00:00:00, FastEthernet1/1
C       172.16.12.24 is directly connected, FastEthernet1/1
C     10.168.29.0/24 is directly connected, FastEthernet0/0
C     10.168.31.0/24 is directly connected, FastEthernet0/1
R     10.168.35.0/24 [120/2] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.41.0/24 [120/2] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.42.0/24 [120/1] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.54.0/24 [120/2] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.63.0/24 [120/3] via 172.16.12.25, 00:00:00, FastEthernet1/1
                     [120/3] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.66.0/24 [120/3] via 172.16.12.25, 00:00:00, FastEthernet1/1
                     [120/3] via 172.16.12.1, 00:00:06, FastEthernet1/0
R     10.168.79.0/24 [120/2] via 172.16.12.25, 00:00:00, FastEthernet1/1
R     10.168.90.0/24 [120/1] via 172.16.12.25, 00:00:00, FastEthernet1/1
R     10.168.91.0/24 [120/1] via 172.16.12.25, 00:00:00, FastEthernet1/1
```

Fig. 6. Hop records (RIP)

In addition, on the network using OSPF it is possible to find out and change the cost of interface, as one of the main protocol metrics. To do this, use the command:

R1#*show ip ospf interface*

The results shown in Fig. 7:

```
R1#show ip ospf interface
FastEthernet0/0 is up, line protocol is up
  Internet address is 10.168.29.1/24, Area 0
  Process ID 1, Router ID 172.16.12.26, Network Type BROADCAST, Cost: 1
  Transmit Delay is 1 sec, State DR, Priority 1
  Designated Router (ID) 172.16.12.26, Interface address 10.168.29.1
  No backup designated router on this network
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    No Hellos (Passive interface)
  Index 1/1, flood queue length 0
  Next 0x0(0)/0x0(0)
  Last flood scan length is 1, maximum is 1
  Last flood scan time is 0 msec, maximum is 0 msec
  Neighbor Count is 0, Adjacent neighbor count is 0
  Suppress hello for 0 neighbor(s)
```

Fig. 7. Interface Cost (OSPF)

Changes in cost will make adjustments to the formation of routes.

The Cisco Packet Tracer simulator does not provide for such commands for the RIP protocol since the protocol has a completely different principle of route generation. Accordingly, introduction of changes by the user is also not provided. At the same time, the software product allows you to consider most of the features of the protocols, which reflected in the work.

CONCLUSION

According to the results of the study, we can conclude that the OSPF protocol is one of the most popular in the modern world, but this, for the most part, concerns networks with a large number of nodes. Many factors contribute to this, in particular, the high time of network convergence It can be concluded that the results of the two protocols may be about the same, provided that the network being studied is small. In addition, the OSPF protocol has more functionality than the older RIP protocol. The analysis and comparison carried out in this paper allow us to draw conclusions about feasibility of using the OSPF protocol in certain conditions.

[1] Yury Semenov, “Algorithms of telecommunication networks. Protocols and Routing Algorithms ”, ISBN: 978-5-94774-707-2, No. 2, 2014.

[2] Lan Wang, Cheng Yi, Adam Alyyan, "OSPFN: An OSPF Based Routing Protocol for Named Data Networking", NDN-0003, July 25, 2012.

[3] J. Krouz, Ross K., “Computer networks. A downward approach,” ISBN: 9785699780907, 2016.

[4] Olifer V., Olifer N., “Computer Networks: Principles, Technologies, Protocols. Tutorial ”, ISBN: 9785496019675, 2016

[5] Routing Basics: RIP and OSPF. https://setevik.com.ua/2017/11/07/basics-rip-and-ospf/

[6] Configure the OSPF protocol on the Cisco router. https://easy-network.ru/51-urok-31.html

[7] Cisco Packet Tracer Tutorial. https://www.scribd.com/document/239146133/ Tutorial-PoCisco-Packet-Tracer

Mathematical Model for Assigning Routes of Public Transport

Maryna Derkach
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
derkach@snu.edu.ua

Inna Skarga-Bandurova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
skarga-bandurova@snu.edu.ua

***Abstract*— In this paper, the task of assigning a route is solved based on the results of the arrival of vehicles in certain segments. A mathematical model of the problem is presented, which allows you to identify unique sections of the road, i.e. segments of the city roads, which belong only to a single route. The graph of the transport system is based on the example of the city of Severodonetsk.**

Keywords— procedure, assigning route, certain segment, graph, transport system

I. INTRODUCTION

The increase in vehicles affects the overall dynamics of the transport system; cities of medium and small size usually have mixed traffic [1], or several transit routes. For various reasons, vehicles of urban transport are not assigned to a specific route. Every day at the beginning of the working day each vehicle is given a new route. To solve this problem [2], there are two approaches: daily manual update, or automatic route assignment. In the latter case, there is a need for a unique procedure for assigning routes.

II. MATHEMATICAL MODEL

The mathematical model of the city's transport network is given by a graph:

$$G=(U,E), \tag{1}$$

with set of nodes:

$$U = U_1 \cup U_2 \cup U_3 \cup U_4, u = |U|, \tag{2}$$

set of arcs:

$$e = |E|, \tag{3}$$

where U_1, U_2, U_3, U_4 - set of nodes first, second, third and fourth type, u - number of nodes, e - number of arcs, |.| - power sign set.

Arcs are sections of roads between nodes, which in turn form a route [3]. Routes are given by the matrix:

$$\begin{aligned} &R = \left\| r_{i,j} \right\|_{u \times u} \\ &i, j = \overline{1,u} \\ &i \neq j, \end{aligned} \tag{4}$$

where r – route from the node i in the node j.

Consider the transport network of the city Severodonetsk for public transport routes.

The first type set includes nodes:

$$U_1=\{u_1, u_3, u_4, u_6, u_7, u_9, u_{10}, u_{18}, u_{20}, u_{21}, u_{23}, u_{26}, u_{27}, u_{28}, u_{29}, u_{43}\}.$$

The second type set includes nodes:

$$U_2=\{u_{12}, u_{13}, u_{15}, u_{16}, u_{17}, u_{19}, u_{22}, u_{24}, u_{25}, u_{30}, u_{31}, u_{33}, u_{34}, u_{35}, u_{37}, u_{38}, u_{39}, u_{40}, u_{41}, u_{42}, u_{44}\}.$$

The third type set includes nodes:

$$U_3=\{u_2, u_5, u_8, u_{11}, u_{14}, u_{32}\}.$$

Appropriately, the fourth type set includes nodes:

$$U_4=\{u_{36}\}.$$

Received a matrix of routes of the city:

$$R = \begin{Vmatrix} u_1 & u_4 & u_7 & u_{10} & u_{18} & u_{23} & u_{26} & u_{27} & u_{28} & u_{29} & u_{30} & u_{31} & u_{32} & u_{33} & u_{34} & u_{35} & u_{36} & u_{37} & u_{25} & u_{22} \\ u_2 & u_5 & u_8 & u_{11} & u_{20} & u_{30} & u_{31} & u_{32} & u_{33} & u_{34} & u_{35} & u_{36} & u_{37} & u_{25} & u_{22} & & & & & \\ u_2 & u_5 & u_8 & u_{11} & u_{12} & u_{13} & u_{14} & u_{15} & u_{16} & u_{17} & u_{19} & u_{24} & u_{36} & u_{38} & u_{44} & u_{43} & u_{42} & u_{41} & u_{40} & u_{39} \\ u_2 & u_5 & u_8 & u_{11} & u_{12} & u_{13} & u_{14} & u_{15} & u_{16} & u_{17} & u_{19} & u_{24} & u_{36} & u_{38} & u_{44} & & & & & \\ u_3 & u_6 & u_9 & u_{14} & u_{21} & u_{32} & u_{42} & u_{41} & u_{40} & u_{39} & & & & & & & & & & \end{Vmatrix}.$$

Since the routes of the city turned out to be transit [4], there are segments of the road belonging to several routes at once:

$$r_{1,22} \cap r_{2,22} = \{ u_{30}, u_{31}, u_{32}, u_{33}, u_{34}, u_{35}, u_{36}, u_{37}, u_{25}, u_{22} \}$$
$$r_{2,39} \cap r_{2,44} = \{ u_2, u_5, u_8, u_{11}, u_{12}, u_{13}, u_{14}, u_{15}, u_{16}, u_{17}, u_{19}, u_{24}, u_{36}, u_{38}, u_{44} \}$$
$$r_{2,39} \cap r_{3,39} = \{ u_{42}, u_{41}, u_{40}, u_{39} \}$$

$$r_{2,22} \cap r_{2,39} \cap r_{2,44} = \{ u_2, u_5, u_8, u_{11} \};$$
$$r_{1,22} \cap r_{2,22} \cap r_{3,39} = \{ u_{32} \}$$

the procedure for setting the route is reduced to finding unique segments, that is, a segment of the road that belongs to only one route:

$$r_{2,22} \setminus r_{1,22} \setminus (r_{2,22} \cap r_{2,39} \cap r_{2,44}) = \{ u_{20} \}$$
$$r_{2,39} \setminus r_{2,44} \setminus r_{3,39} = \{ u_{43} \}.$$

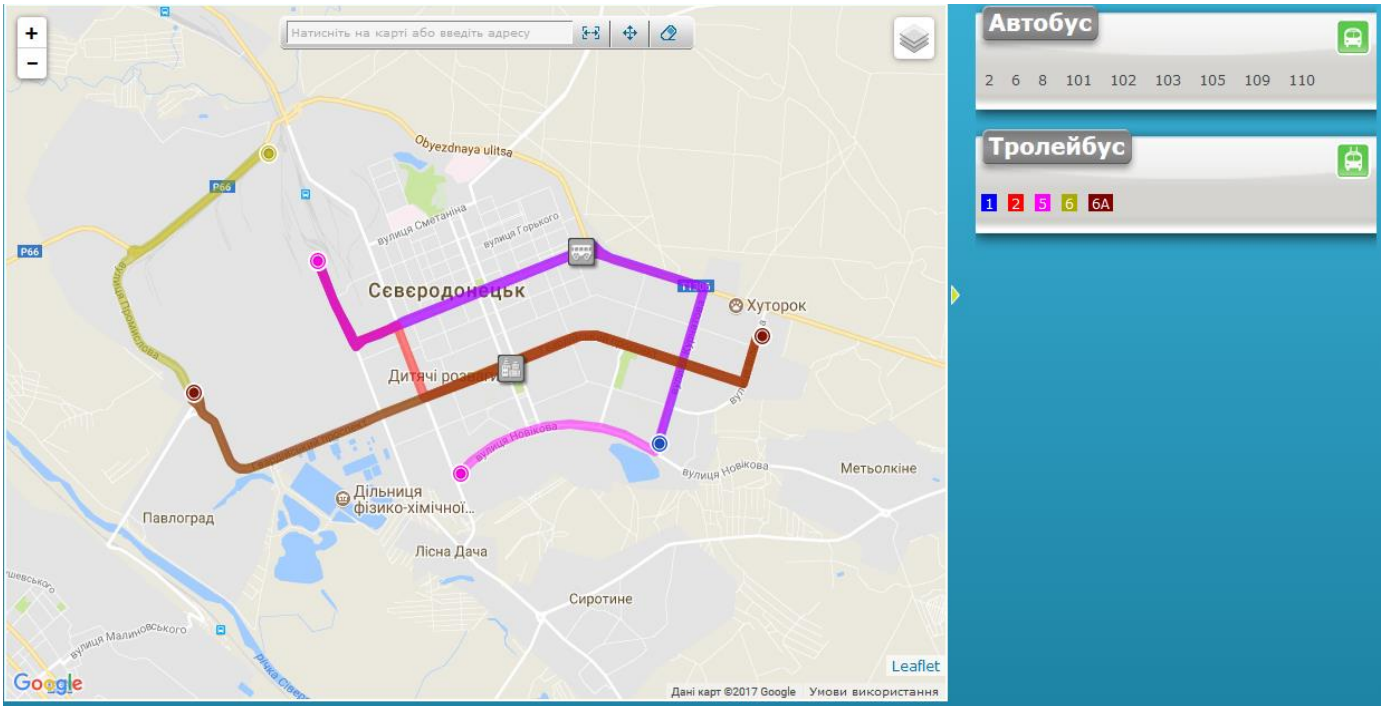

Fig. 1. Public transport routes of the city of Severodonetsk

The Fig. 1 clearly shows that the public transport routes of the city of Severodonetsk are really transit routes. Using the proposed mathematical model allowed us to obtain unique control points, which makes it possible to unambiguously determine the route.

CONCLUSION

There is a problem of mixed traffic and transit routes, typical of medium and small cities. Moreover, public transport vehicles often do not have a permanently fixed route. Therefore, to automate the procedure of fixing the route, the mathematical

model has been developed that allows assigning the route as a result of the arrival of vehicles in certain segments.

[1] Hua X. Bus arrival time prediction using mixed multi-route arrival time data at previous stop / X. Hua, W. Wang, Y. Wang, M. Ren // Transport. – 2018. – Vol. 33(2). – pp. 543-554.

[2] Bai C. Dynamic Bus Travel Time Prediction Models on Road with Multiple Bus Routes / C. Bai, Z.-R. Peng, Q.-C. Lu, J. Sun // Computational Intelligence and Neuroscience. – 2015. – Vol. 2015, doi:10.1155/2015/432389.

[3] Yu B. Bus arrival time prediction at bus stop with multiple routes / B. Yu, W. H. K. Lam, M. L. Tam // Transportation Research Part C: Emerging Technologies. – 2011. – Vol. 19(6). – pp. 1157–1170.

[4] Zheng C.-J. Improved iterative prediction for multiple stop arrival time using a support vector machine / C.-J. Zheng, Y.-H. Zhang, X.-J.Feng // Transport. – 2012. – Vol. 27(2). – pp. 158–164.

Modeling of the Water Diffusion Mechanism

Ksenia Hulevska
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
makotoyorigaka@gmail.com

Larisa Shumova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
shumova@ukr.net

Viktoria Mokhonko
Department of Chemical Engineering and Ecology
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
maxtory@gmail.com

***Abstract*—This paper consider the development of a tool for modeling a diffusion process in a porous medium with absorption of a diffusing substance on the surface of pores and its introduction into an automated control system for sewage treatment plants to ensure the rational use of water.**

Keywords—simulation modeling, automated systems, wastewater, water purification, adsorption, software

I. INTRODUCTION

The rational use of water have of paramount importance and is one of the most important tasks of ensuring environmental safety and sustainable development of the country's economy.

Optimal water supply can be achieved as a result of the introduction of water-saving technologies and technologies with low water consumption or as a result of improved water supply systems [1].

The most current way to achieve the required quality of water purification is adaptive methods of managing the purification process. The implementation of adaptive control requires the creation and use of a mathematical modeling tool.

The results of solving problems of modeling water preparation processes are present. The adsorption purification method is considered. This method is used for deep treatment of wastewater from dissolved organic substances after biochemical treatment, as well as at local sewage treatment plants of industrial effluents, which are multicomponent mixtures [2]. In this method, the molecules of the dissolved pollutant are transferred from the solution to the surface of a highly absorbing substance (adsorbent) under the action of the force field of the surface. Based on the studies presented in [3], adsorption can be considered as a process of stationary diffusion with the absorption on the surface of particles in regions with a fine-grained random boundary.

The purpose of the study is to improve the local wastewater treatment system based on adaptive control methods using simulation modeling of the diffusion process in a porous medium with absorption of a diffusing substance on the surface of pores.

II. MATHEMATICAL FOUNDATION

On the basis of the Dirichlet problem, a limit model of diffusion in a porous medium was obtained [4] with absorption on the surface of pores:

$$\begin{cases} -D\Delta u(x)+C(x,u)=f(x), x\in\Omega, \\ u(x)=0, \quad x\in\partial\Omega \end{cases}, \qquad (1)$$

where D is the diffusion coefficient, Δ is the Laplace operator, C(x, u) describes the limiting absorption of the system, f(x) is the given function of external sources.

The required function u(x) sets the concentration of a substance that diffuses at every point of the Ω area.

The finite difference method was applied to the model (1).

Jacobi's iterative method is used to solve a system of difference equations. The computational algorithm of the method is that the new value at each grid point is determined as the average of the previous values of the four neighboring points (left, right, top and bottom). This process is repeated to the last point.

For the resulting system of difference equations a computational algorithm was compiled. This algorithm consists of performing the following formal procedures:

1) Set $u^0[i,j,k]=0$; $i,j,k=\overline{0,\ldots,n}$;
2) Calculate $u^1[i,j,k]$, $u^2[i,j,k],\ldots,u^n[i,j,k]$ bye $|u^{n-1}[i,j,k]-u^n[i,j,k]|<\varepsilon$, where ε - accuracy is 0.001.

The final stage of mathematical modeling is the development of a modeling tool - a set of programs that implement the presented methods and computational algorithms. A visual interface has been created to present the results of computational experiments using the Delphi development environment. The results illustrate the effectiveness of the developed algorithms for determining the optimal parameters of the wastewater treatment process.

CONCLUSION

The implementation of the developed simulator into the control system of local water treatment plants will improve the accuracy of calculations of the required volume of the adsorbent, ensure the implementation of control system adaptation algorithms to the changing characteristics of wastewater, as well as obtain high quality water treatment.

[1] Ковшун Н. Е. Проблемы обеспечения устойчивого водопользования в Украине. / Н. Е. Ковшун // Стратегія і тактика державного управління. Збірник наукових праць. – Рівне, 2014. – Вип. 1. – С. 88-92.

[2] Зайнуллин Р.Р. Проблема очистки городских сточных вод. / Р.Р. Зайнулин, А.А. Галяутдинов // Инновационная наука. – 2016. - № 6. – С. 68-69.

[3] Хруслов Е. Я. Нелинейная задача Робена в областях с мелкозернистой случайной границей / Е. Я. Хруслов, Л. А. Хилькова // Доповіді НАНУ. – 2017. – № 9. – С. 3-8.

[4] Гулевська К.В. Використання імітаційного моделювання при відображенні фізичних процесів. / К.В. Гулевська, Л.О. Хількова // Актуальні проблеми науково-промислового комплексу регіонів. Матеріали IV Всеукраїнської науково-практичної конференції. - Рубіжне: видавець О. Зень, - 2018. - С. 267-269..

Methods Overview of Open-Source Speech Recognition Software

Viktoriia Derevianchenko
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
derevianchenkovs@ukr.net

Inna Skarga-Bandurova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
skarga_bandurova@ukr.net

Polina Fursa
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
polinaphursa@gmail.com

Mark Koverha
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
healthunder@gmail.com

***Abstract*— Experiments of speech recognition software for controlling a computer using voice, dictating texts and human-computer interaction or communicating with a computer were conducted. In this paper a series of experiments for study the open-source software for voice actuated control is presented..**

Keywords— voice, speech recognition, software, human-computer interaction

I. INTRODUCTION

At present, it is planned to introduce chatbots capable of recognizing human speech and searching for information by voice commands. Telephony will be robotic based on chatbots: companies will be able to automatically receive and make calls when interacting with customers or partners, while the robot will be able to conduct complex dialogs, moving from topic to topic. In general, speech recognition comes down to three types: recognition of separately pronounced words, so-called commands, recognition of continuous speech in a large dictionary, speech-based identification.

The task of automatic speech recognition in real conditions is relevant, given the variability of the source of the speech signal and acoustic noise, which hides the original sequence of audio segments. In recent years, significant progress has been made in this area, and there are commercial voice-independent applications that quite successfully recognize speech when processing voice commands, in interactive systems, in shorthand systems. The recognition accuracy of speech units in these systems has reached the required threshold.

In this work, we provide an analysis of open source speech recognition software for continuous speech recognition.

II. EXPERIMENT DESIGN

There are 2 types of licenses in most common speech recognition systems:

• BSD (Berkeley Software Distribution) license [1] includes audio speech recognition products: CMU Sphinx, PocketSphinx, Julius, etc.

• GPL (General Public License) [2] includes: Simon software, iATROS, RWTH ASR, etc.

Simon [3] is a software for Internet surfing, mailing, managing multimedia applications that can be adapted to the needs of older people. For testing, the acoustic model from [4] was used.

As alternative speech recognition software, Deepspeech [5] and PocketSphinx [6] were reviewed and tested. DeepSpeech is an open source software product for converting speech to text. For training, a model trained in machine learning methods is used, based on Baidu's in-depth speech research. PocketSphinx also needs language model files, acoustic model files, and a pronunciation dictionary (phonetic dictionary). These applications were tested with data from Kaggle [7] and with their own recorded files.

Recognition of continuous speech on new sets. Source data were recorded and tested using customized voice data [8].

The speech recognition was evaluated using two criteria, they are WER and Latency.

1) Word Error Rate (WER) - A word error rate measures word level mismatch: it compares the words issued by the recognizer with those that the user actually uttered. Each error (replacement, insertion or deletion) is counted in the recognizer. WER can be calculated as follows:

$$WER = \frac{Number\ of\ Substitution + Insertions + Deletions}{Total\ number\ of\ words} \cdot 100\%$$

2) Delay (en. Latency)

Delay is defined as the total time (in seconds) required to perform speech recognition. More precisely, we set the delay

as the time from the moment when the recording ends, until the recognition results appear on the screen.

III. Result

Using Simon software to control mouse and Internet surfing using standard libraries and new phrases. To do this we worked with two scenarios:

(1) controlling mouse to click on the youtube icon via voice;

(2) controlling browser and testing the following commands: "page up", "page down", and searching keywords on the page by the command "48".

In a result, almost all voice commands were performed successfully except "48". Adaptation to the user's voice for key phrases/words of this script is successful - it does not return any errors Sometimes when recording words, it reacts to the voice loudness and the intelligibility of pronunciation. We had to repeat some words for several times. In case the word was said louder than usual, the program did not work and asked to overwrite. But this happened rarely, and mostly it successfully fixes the phrases/words.

Conclusion

The analysis of the software for recognition of continuous speech showed that at present there is no universal system for recognizing continuous speech that would be capable of self-learning, would be speaker-independent, resistant to noise, and have a low error rate. The considered software solutions at the moment are not universal and accurate, the speech recognition error greatly depends on the presence of extraneous medium and high-frequency noise, as well as on the microphone quality.

According this, it should be noted that the tasks of the development of the PCP, the development and implementation of software and information solutions in this area are relevant..

[1] "Original BSD license". Various Licenses and Comments about Them. Free Software Foundation. Available at: https://www.gnu.org/licenses/license-list.html#OriginalBSD.

[2] Montague B. Comparing the BSD and GPL Licenses on Technology Innovation Management Review by Available at: https://timreview.ca/article/67 [Accessed 2 May. 2019].

[3] Simon https://simon.kde.org/ [Accessed 2 May. 2019].

[4] Available at: http://www.speech.cs.cmu.edu/sphinx/models. [Accessed 2 May. 2019].

[5] Deepspeech https://github.com/mozilla/DeepSpeech [Accessed 2 May. 2019].

[6] PocketSphinx https://github.com/cmusphinx/pocketsphinx/blob/master/include/pocketsphinx.h [Accessed 2 May. 2019].

[7] Common-voice Data. Kaggle Available at: https://www.kaggle.com/mozillaorg/common-voice [Accessed 2 May. 2019].

[8] Customized data, Google drive. Available at: https://drive.google.com/open?id=11pNpJOK-Pr1AAb22jOW0FIhYJ9p1ekAE [Accessed 2 May. 2019].

Processing Technique for Biomedical Image Analysis

Vyacheslav Lyashenko
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
lyashenko.vyacheslav@gmail.com

Oleg Kobylin
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
oleg.kobylin@nure.ua

Oleksandr Ryazantsev
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
a_ryazantsev@ukr.net

Ivan Ryazantsev
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
ivan.riazantsev@nure.ua

***Abstract*—Image processing methods are used in all areas of research. These methods provide additional information, a better understanding of the object that is being studied. Among the areas of using image processing methods, medicine occupies a special place. Biomedical data allow us to assess human health, to identify diseases in the early stages. Images of cellular structures of cytological preparations are one of the examples of biomedical data. Based on image analysis methods, we can isolate various components of cellular structures of cytological preparations. To do this, we apply the methods of wavelet analysis for different color components of the input image. Applying morphological analysis, we can identify individual cellular structures. The results are shown on the example of images of cellular structures of cytological preparations.**

Keywords—image processing, biomedical data, cell structures, wavelet analysis, color space, cytological preparations

I. INTRODUCTION

The analysis of biomedical data is of practical importance. Such an analysis helps to make an initial preliminary assessment of the state of human health. At the same time, we can identify possible diseases in the early stages of their development. This allows for timely treatment and save the patient from a possible disease.

Biomedical data can be presented as a time series of data, descriptive statistics, or biomedical image. One of the most common methods for recording biomedical data is images obtained under a microscope.

However, we must take into account the specifics of biomedical images. Therefore, it is important not to lose information as a result of the implementation of individual procedures for analyzing the original image. Also, one of the tasks is to obtain additional information as a result of applying various image processing procedures.

The main objective of this study is to consider such a sequence of procedures for analyzing the original image, as a result of which we can obtain additional information. This is an important task for the study of cellular structures whose images were obtained under a microscope.

II. RESEARCH RESULTS

One of the formats for representing color images is RGB. The image in RGB color system consists of three color channels - R (red), G (green) and B (blue) [8, 9]. Each color channel allows you to consider a specific frequency region of the image. This is very important for image analysis using wavelets. This allows you to take into account all the features of the image.

Thus, we propose to consider the original image $B(i, j) = k$ as a combination of three images:

$BR(i, j) = rk$ – original image in color channel R, where rk is the brightness value of the image at the point $BR(i, j)$;

$BG(i, j) = gk$ – original image in color channel G, where gk is the brightness value of the image at the point $BG(i, j)$;

$BB(i, j) = bk$ – original image in color channel B, where bk is the brightness value of the image at the point $BB(i, j)$.

Then we apply the wavelet analysis procedure for each color channel of the original image. We get three images that define the set of points of difference of brightness values in each color space. This provides additional information about the image we are analyzing. Based on this information, we can concretize various areas of interest.

III. RESULT OF EXPERIMENTAL STUDIES

The image of cellular structure consist of cells megaloblastic anemia, segments in the nucleus of neutrophils and erythrocytes. We considered the original image without dividing it into separate color channels. To do this, we converted the original image to a black-white image. Then we applied wavelet processing.

Thus, we have additional information. This allows for more accurate analysis. For this you can use morphological image analysis [1] (see Fig. 1).

In Fig. 1 presents the results of morphological analysis. As a result of this analysis, we identified erythrocytes.

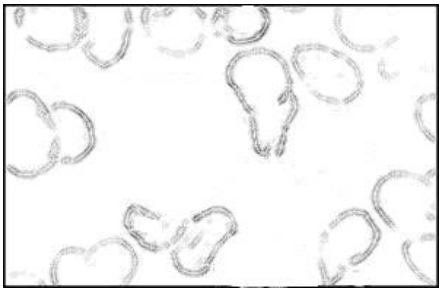

Fig. 1. Results of morphological analysis

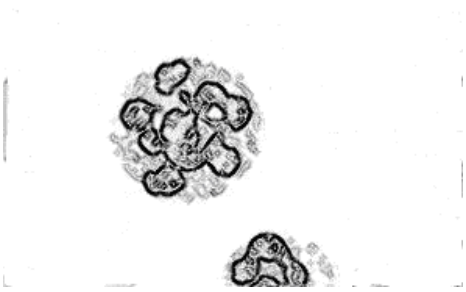

Fig. 2. Results of morphological analysis

In Fig. 2 presents the results of morphological analysis, which allow to isolate the cells of megaloblastic anemia.

We can specify a different morphology to identify the desired area of interest.

CONCLUSION

For more information about the objects of interest in the image, we used the wavelet methodology and the technique of decomposing the image into colored components. With the help of morphological analysis, we have identified various components of images of cellular structures of cytological preparations. For further research, it is necessary to automate the process of identifying components of images cell structures the cytological preparations.

[1] K. Lee, A. Shamsoddini, X. Li, J. C. Trinder, and Z. Li, “Extracting hurricane eye morphology from spaceborne SAR images using morphological analysis,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 117, pp. 115-125, 2016.

Query Optimization in Database Systems

Maksym Nesterov
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
maksym.nesterov@gmail.com

Denis Bakitko
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
bakitko_denis@ukr.net

Alice Mikhaylova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
alicedemoran@gmail.com

***Abstract*—This article reviewed the methods directed at optimizing the database. The goal was to find the most suitable method for the quick execution of queries.**

Keywords—database, indexing, clusters, optimization, partitioning, statistics, SQL

I. INTRODUCTION

The processing time of the request also determines the speed of the database, which characterized by amount of time per request processing. The execution time of a request may also depend on complexity of the request.

By optimizing database queries, you can increase the speed of the application. The request processing time is a very important criterion for evaluating database performance. When designing a system, it is important to predict an increase in database queries, as well as an increase in data volumes. The increase in performance consists of the execution time of queries, the speed of information retrieval in non-indexed fields, the greatest number of parallel access to data.

II. METHODS OF SETTING PRODUCTIVITY

Methods such as indexing, query code optimization, clustering, and partitioning will be considered.

Methods such as indexing, query code optimization, clustering, and partitioning will be considered.

Indexes are used to speed up data retrieval, which retrieved faster by reducing the number of disk I/O operations for which pointers are used.

Indexing is very effective if the number of records sought does not exceed 10-15%. If the proportion of the files you are looking for is too large, the index file will be used too often, which will lead to a losing strategy for using it. In these cases, it is necessary to execute a direct search, without using indexing [1].

Each index must be unique. Unique indexes exclude the same data. Creating uniquely makes sense only if the data itself could unique. In the case of duplication of data, there will be an error and it will not be possible to save the entered information.

Indexing allows you to not view the entire table for data retrieval, thereby increasing database performance.

Partitioning meant breaking a large table into less to promote the execution of necessary queries. Partitioning capabilities include partition independence, which makes it possible to carry out backup, restore, and index creation operations specifically on a partition, rather than on the entire large table, which will significantly reduce database idle time [2].

Partitioning improves the performance of information processing in large tables, but it does not protect against poor-quality queries.

There are the following types of building cluster systems [3]:

- With shared disks.
- With shared memory.

Each node of such systems specifically serves its own database fragment. Such systems lack shared memory and storage devices.

When a database uses shared memory and I /O devices, it called Shared everything.

Such systems characterized by high-speed between nodes and have shared access to storage devices. The strength of the system is fault tolerance, parallel processing, and ability of the network to expand.

A major disadvantage of the system is competition of nodes for input devices and memory. These shortcomings manifest themselves when the network is busy and when executing INSERT, UPDATE, DELETE commands, which need many processor resources to execute.

The Shared Nothing model fixed all the flaws of Shared Everything.

To optimize execution of data change statements, use the EXPLAIN PLAN statement, which allows you to view the execution plan of an SQL statement.

EXPLAIN PLAN allows you to see an execution plan that the analyzer can use to execute an expression [4]. The constructed expression plan is written to the table without saving the SQL expression.

The optimizer's query execution plan runs faster with checking the column index stores than using index row stores [5]. The selection of columns is based on a lower cost value than on rows.

During the execution of each SQL query, optimizer looks for the best solution for its execution. Optimizer relies on statistical data, which includes information about distributed data, characteristics of tables and indexes. [6]

III. CONDUCTING AN EXPERIMENT

A table was created in the database, which consists of 101158 rows. The table consists of columns:

- ID - identifier.

- Users - workers.
- Age - their age.
- City - city of residence.
- Work_with - the date from which the employee got a job.

A query for sampling dates taken not from the main table, but from its sections, which means that execution of the query is faster than comparing all the records of the main table with the condition in the query.

TABLE I. QUERY TIME

	Index	**No Index**	**Partitioning**
Query time (sec)	0.0007	0.0055	0.0007

The query indexed and partitioned tables executed in 0.0007 seconds, which is 7.85 times faster than to the usual similar table.

As a result of the analysis of productivity tuning methods, an index method chosen. This method allows you to quickly process a query to the table due to less load on system resources and the ability not to compare each row of the table with the condition in the query, but to directly access the index files. Index files allow you to process a query faster and their implementation in practice is easier than partitioning a table. Indexing should be applied to those rows that are not confirmed by data changes. In the example, this table does not change; it serves to collect company statistics on employees. And also indexing reduces resource consumption when executing a query.

CONCLUSION

According to the results of the analysis of database productivity methods, it could be concluded that each method has its own advantages and may affect the speed of the query to the database. However, each method implies that to successfully carry out the method, it is necessary to carefully work out the structure of the database in the required subject area. Great emphasis should be placed on the ability to update the columns and operate on their data.

In the analysis performed, the method that most suitable for the database optimization problem was chosen.

[1] Sadhana J. Kamatkar, Ajit Kamble, Amelec Viloria, Lissette Hernández-Fernandez and Ernesto García Cali: Database Performance Tuningand Query Optimization, 2018, 4.

[2] Sanjay Agrawal, Vivek Narasayya, Beverly Yang: Integrating Vertical and Horizontal Partitioning into Automated Physical Database Design, 2004, 1-9.

[3] Database cluster. http://www.r-it.su/solutions/san/nas-db-cluster/.

[4] Khaled Yagoub, Pete Belknap, Benoit Dageville, Karl Dias, Shantanu Joshi, and Hailing Yu: Oracle's SQL Performance Analyzer, 2008, 3.

[5] Per-Åke Larson, Cipri Clinciu, Eric N. Hanson, Artem Oks, Susan L. Price, Srikumar Rangarajan, Aleksandras Surna, Qingqing Zhou: SQL Server Column Store Indexes, 2011, 2-8.

[6] Ziauddin, M., Das, D., Su, H., Zhu, Y., & Yagoub, K.: Optimizer plan change management, 2008, 2-5.

Real-Time Data Analytics for the Internet of Things

Yana Krytska
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
kritskayana@gmail.com

Tetiana Biloborodova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
beloborodova.t@gmail.com

Inna Skarga-Bandurova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
skarga_bandurova@ukr.net

***Abstract*—Data mining (DM) is one of the most valuable technologies enable to identify unknown patterns and make Internet of Things (IoT) smarter. The current survey focuses on IoT data and knowledge discovery processes for IoT. In this paper, we present a systematic review of various DM models and discuss the DM techniques applicable to different IoT data. Some data specific features were analyzed, and algorithms for knowledge discovery in IoT data were considered. Challenges and opportunities for mining multimodal, heterogeneous, noisy, incomplete, unbalanced and biased data as well as massive datasets in IoT are also discussed.**

Keywords— Data Mining, Internet of Things, IoT, Knowledge Discovery in Database, KDD, massive data set

I. INTRODUCTION

IoT applications generate more than 2.5 quintillion data bytes daily. To convert this data into knowledge, data mining systems are increasingly in demand. Data mining (DM) enables to find and discover novel, interesting, and useful patterns from large data sets and generate new knowledge from information obtained from IoT devices. However, basic data mining algorithms and technologies are not quite sufficient for IoT framework. So, it becomes a great challenge to collect, analyze and manage IoT data as well as to generate and update data mining algorithms for IoT purposes. The DM approaches applicable to IoT data are discussed. An important aspect of DM of the IoT-based system is the effective structure of the system, which should take into account security, data privacy, data sharing mechanisms, scalability, etc. Such a DM system for IoT includes data acquisition devices, raw data properties, extraction levels, processing, data analysis, it is necessary to take into account the properties of the IoT devices when planning DM for IoT [1].

This survey focuses on IoT data and knowledge discovery processes for IoT. Our main contribution in this paper is that we targeted on data specific features and selected some well-known algorithms best suited for knowledge discovery in different IoT applications.

II. DATA MINING FOR IOT

A. IoT Data Characters

IoT data can be: multimodal and heterogeneous, noisy and incomplete, unbalanced and biased, dependent on time and location, dynamic, different data quality, almost always require real-time analysis.

Given that IoT data is the basis for extracting knowledge, it is important to have high quality information. This condition can directly affect the accuracy of knowledge extraction.

Figure 1 shows an overall level for transformation of data and depicts a level of services where big DM for IoT is applicable.

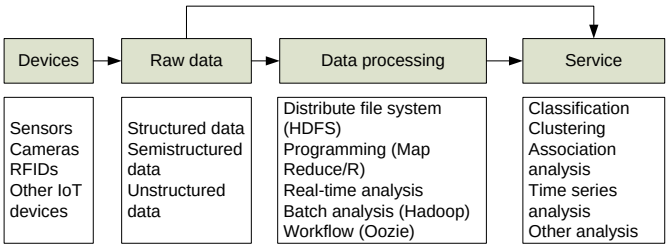

Fig. 1. Big data mining based on IoT (Adapted from [2])

B. Basic idea of using data mining for IoT

The main characteristics of the source data of IoT-based system are the following [2]:

1. They are really big data.

2. Heterogeneity of the sources being combined and the types of data.

3. The complexity of recoverable knowledge: due to heterogeneity and a large amount of data when extracting knowledge, it is necessary to analyze their properties and the interrelation of various data sources.

In the process of extracting useful knowledge, there are the following issues.

1. Data extraction: data can be combined from various sources, they are diverse and heterogeneous, and noisy.

2. Uncertainty and incompleteness of data: compliance with data security and confidentiality causes uncertainty and incompleteness of data in the extraction of useful knowledge.

To solve these problems, approaches and methodologies are being developed that try to minimize their consequences. Tracking and detection of data errors, preprocessing filtering, and data reduction mechanisms are used. To combine data from several sources, parallel programming models are used, for which classical approaches to DM are adapted.

The DM process for IoT is similar to the base one, but there are some big differences. The process of extracting useful patterns from raw data is known as Knowledge discovery in databases (KDD).

Based on the DM and IoT overview, the data mining in IoT process is as follows (see Fig. 2): DM for IoT begins with the first step of capturing data generated from IoT devices which includes: Sensor networks, Actuators, Wireless Sensor Network (WSN), Wireless Sensor and Actuator Network (WSAN), Radio Frequency Identification (RFID) Tags, Cameras, GPS etc.

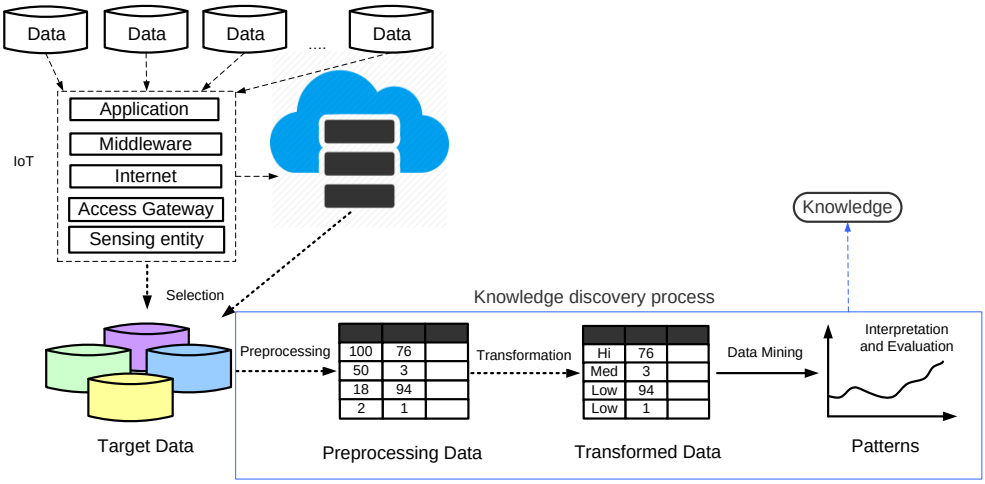

Fig. 2. KDD process for IoT data

C. Applying data mining algorithms for IoT Data

To determine which algorithm to use for a particular task, we need to first define the task and aim of analysis. Some of tasks include finding unusual data points, predicting values or categories, structures discovery, feature extraction and more.

Table 1 shows some examples of using data mining algorithms for IoT data.

TABLE I. APPLICATION OF DM ALGORITHMS FOR IOT DATA

DM algorithm	Goal	Data Source
Classification	Device recognition; Traffic event detection; Parking lot management; Inhabitant action prediction; Inhabitant action prediction; Inhabitant action prediction; Physiology signal analysis	RFID; GPS, smart phone, and vehicle sensor; Passive infrared sensor; RFID, sensor, video camera, microphone, wearable kinematic sensor, and so on; Video camera; Microphone; Wireless ECG sensor
Clustering	Network performance enhancement; Inhabitant action prediction; Provisioning of the needed services; Housekeeping; Managing the plant zones; Relationship in a social network	Wireless sensor; X10 lamp and home application; Raw location tracking data; Vacuum sensor; GPS and sensor for agriculture; RFID, smart phone, PDA, and so on
Frequent Pattern	RFID tag management; Spatial colocation pattern analysis; Purchase behavior analysis; Inhabitant action prediction	RFID; GPS and sensor; RFID and sensor; RFID and sensor
Anomaly Detection	Smart Traffic; Smart Environment; Traffic Prediction Finding Anomalies in Power Dataset	GPS, smart phone, and vehicle sensor; Wireless sensor, smart phone; GPS, smart phone, and vehicle sensor; RFID, wireless sensor
Hybrid	Inhabitant action prediction	RFID and sensor

D. Mining of Massive Datasets

IoT systems include multiple heterogeneous networked embedded devices that generate massive amounts of data. Massive Data IoT leads to different issues in processing and DM. The large amount of data, the high transfer rate and the variety of properties of large IoT data necessitate a new requirement for intelligent analysis of such data and the diversity in data sources is also a problem.

The massive data are generally collected from different heterogeneous sources (e.g., video cameras, sensors, RFID, other IoT devices, people, etc.) providing heterogeneous sensing data (e.g., text, video, sound). In this context, heterogeneous data processing (e.g., fusion, classification) brings new challenges and open new possibilities for systems. Obviously, these random variables from heterogeneous sensors have different probability distributions.

Define z_n as the data from the n-th sensor and $\mathrm{Z} := \{z_n\}_{n=1}^{N}$ as the heterogeneous data set, the margins $\{z_n\}_{n=1}^{N}$ are generally differently or heterogeneously distributed.

In many IoT applications, datasets are often modeled as multi-sensor data fusion, distribution estimation or distributed detection. For detection, this tasks joint probability density function $f(\mathrm{Z})$ of the heterogeneous data set Z is needed to get from the marginal probability density function $\{f(z_\partial)\}_{n=1}^{N}$.

In these cases, one often uses simple models such as the product model or multivariate Gaussian model, which lead to suboptimal solutions [3]. Other approaches are based on copula theory, to tackle heterogeneous data processing in IoT. In copula theory, it is the copulas function that couples' multivariate joint distributions to their marginal distribution functions, mainly thanks to the Sklar theorem.

Sklar' theorem can be present as follow. Let F be an N-dimensional cumulative distribution function with continuous marginal probability density function $F_1, F_2, ..., F_N$. Then there is a unique copulas function C such that for all $z_1, z_2, ..., z_N$ in $[-\infty, +\infty]$

$$F(z_1, z_2, ..., z_N) = C(F_1(z_1), F_2(z_2), ..., F_N(z_N)) \quad (1)$$

Next, the probability density function can be obtained by the N-order derivative of (1)

$$\begin{aligned} f(z_1, z_2, ..., z_N) &= \frac{\partial^N}{\partial_{z_1}, \partial_{z_2}, ..., \partial_{z_N}} C(F_1(z_1), F_2(z_2), ..., F_N(z_N)) \\ &= f_p(z_1, z_2, ..., z_N) c(F_1(z_1), F_2(z_2), ..., F_N(z_N)) \end{aligned} \quad (2)$$

where $f(z_1, z_2, ..., z_N)$ is the product of the marginal probability density function $\{ f(z_\partial) \}_{n=1}^{N}$ and c(·) is the copula density weights the product distribution appropriately to incorporate dependence between the random variables.

E. Map Reduce

In IoT applications, mass data processing such as MapReduce is constructed for parallel and distributed data processing [4]. Querying and reasoning for data can be adapted to large data is a more flexible approach.

One of the most popular parallel processing methods in cloud platform is MapReduce and its open source implementation Hadoop for cloud-based parallel or distributed data processing. For the parallelization, scalability, load balancing, and fault-tolerance is MapReduce is widely used in cloud platforms for query processing for data analysis.

For problems decision of optimizing parallel data mining, a heuristic cloud bursting algorithm, Maximally Overlapped Bin packing driven Bursting (MOBB), is developed. It considers the time overlap to improve data mining parallelization.

In large IoT data environment, data can be defined by types, state and analysis tasks. Parallel and particle data processing framework is needed to enable the execution MapReduce pattern in dynamic cloud infrastructures, in contrast with centralized master server implementations. These re-build and execution data mining algorithm are not applicable for big data analysis system. Despite its evident merits such as scalability, fault-tolerance, ease programming, and flexibility, MapReduce has limitation in interactive or real-time processing on handling IoT data processing and is not a uniform decision for every large-scale analytical task. Its high communication cost and redundant processing is an IoT application problems.

CONCLUSION

The DM technique is top-of-the-agenda in the IoT concept that arises from the need to manage and analyze big sensors data. With that, DM algorithm selection for IoT is not a huge challenge itself, it mainly depends on the task and also the type of data that we are dealing with. Instead, many other issues should be resolved, they are cleaning the data; transforming all data in a unified format; struggling with missing values and/or reducing massive data sets; understanding the informational content of the data or data interestingness rate; establishing whether the data is sufficient to the purpose of DM or not.

The using of DM techniques for IoT are directs to map reduce, finding similar items. It helps to develop, control and monitor the IoT-based application in different areas.

[1] Lee, D. and Lee, H., 2018. IoT service classification and clustering for integration of IoT service platforms. The Journal of Supercomputing, pp.1-17.

[2] Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A.V. and Rong, X., 2015. Data mining for the internet of things: literature review and challenges. International Journal of Distributed Sensor Networks, 11(8), p.431047.

[3] D. Mari and S. Kotz, Correlation and Dependence. London, U.K.: Imperial College Press, 2001.

[4] Cai, H., Xu, B., Jiang, L., & Vasilakos, A. V. (2016). IoT-based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges. IEEE Internet of Things Journal, 1–1. doi:10.1109/jiot.2016.2619369.

Real-Time Wearable System for Monitoring Cardiovascular Disease

Tetiana Biloborodova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
beloborodova.t@gmail.com

Inna Skarga-Bandurova
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
skarga_bandurova@ukr.net

Viktoriia Derevianchenko
Computer Science and Engineering Department
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
derevianchenkovs@ukr.net

***Abstract*—In this paper, a wearable ECG monitoring system based on IoT is proposed. Wearable devices design employs few optimal components for the acquisition of acceptable ECG signal. The R peaks corresponding to each heartbeat, and T waves, a morphological feature of the ECG are detected. It enables to perform heart rate and heart rate variability analyses, as well as extract, store and analyze the long term ECG measurements.**

Keywords— health monitoring system, wearable device, ECG signal acquisition, QRS-complex detection

I. INTRODUCTION

Cardiovascular diseases (CVD) are the leading cause of fatalities representing 30% of all global deaths. Due to inadequate preventive measures, CVD related fatalities continue to rise. Electrocardiogram (ECG) is widely used to monitor heart function.

The recent technologies as a Big Data, Internet of Things (IoT), mobile technologies, and wearable devices step-by-step take up the art of cardiovascular diseases diagnostic to a new level. In this context, the goal is development of wearable ECG monitoring systems.

In spite of the wide variety of academic works, the ECG monitoring systems still require improvements to deliver the best quality of monitoring and data processing.

In this paper, the IoT wearable system for ECG monitoring with signal acquisition, processing for QRS-complex detection are proposed.

II. HEALTH MONITORING SYSTEM DESIGN

The architecture of a wearable biomedical information monitoring system is represented by three main components.

The first necessary component is various types of wearable devices and sensors that perceive the physical signals of the human body, which are directly wearable devices.

Biomedical sensors can also be directly integrated into garments and accessories, thus forming a wearable body area network (WBAN) [1].

WBAN is a subset of wearable biomedical sensors and systems that can be used to control, stimulate, treat, and replace the biological and physical functions of the human body [2].

The second component of the system — external storage, processing, and data transfer devices. Data can be transmitted directly from sensors or using external devices with more computational power (e.g. a smartphone).

The third component is represented by cloud storages, servers. Since wearable biomedical information monitoring devices and base stations have limited re-sources, the resulting data is usually sent to servers, cloud storage for processing and long-term storage.

A. Wearable device configuration

A wearable ECG device is responsible for collecting ECG data from human skin, and then transmitting this data to an access point through a wireless channel. The ECG monitoring node in our system mainly includes: (1) a sensor module; (2) controller module with wireless adapter; and (3) power module.

(1) Sensor module is the core component of the monitoring device, which is responsible for obtaining human ECG data.

Recorded physiological signals usually consist of a source signal and noise. Noise occurs at each stage of data collection, before they are digitized. Noise of the power supply module, muscular noise, noise of the analog-digital converter suppresses ECG signals.

(2) A controller module with a wireless adapter is used to process and transmit the received ECG signal. In the controller module, it is possible to set the necessary parameters for additional signal processing, its buffering, packaging for further transmission via a wireless channel.

(3) Power module provides reliable power supply of each module in the ECG monitoring device.

B. Detection of a QRS complex and measurement of heart rate

Detail waves of ECG signals are important information for health/medical application. Usually, the following elements can be distinguished on an ECG, among them are the P, Q, R, S, T wave. The QRS complex can be identified using a common method for determining ECG parameters. R-peak is easier to distinguish from noisy components, as it has large amplitude (Fig. 1).

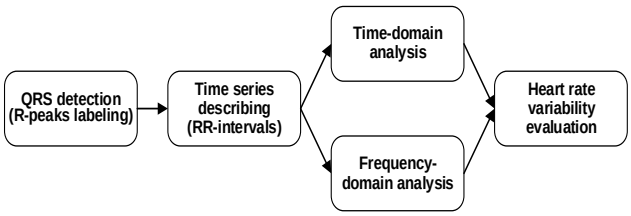

Fig. 1. The general stages of heart rate variability evaluation process

After the pre-processing, variable threshold method can be used to further detect the R-peak. The R-peaks allow heart rate and heart rate variability analyses to be performed, extracting the wanted long term physiological measurements. The formula for variable threshold value is defined as follows.

$$V_{TN} = [x(n) - x(n-1)] \cdot 70\%$$

The threshold makes it possible to differentiate R peak from the baseline, which is corresponding to 70% of ECG peak data detection. We were able to find QRS complex based on the detected R-peak. Detection of QRS complex is particularly important in ECG signal processing. In our system, we used a robust real-time QRS detection algorithm [3]. This algorithm reliably detects QRS complexes using slope, amplitude, and other information. The information obtained from QRS detection, temporal information of each beat and QRS morphology information can be further used for the other ECG parameter detection. In order to detect QRS complex, the signal is initially passed through a band-pass filter. It is composed of cascaded high-pass and low-pass filters. Subsequent processes are five-point derivative (Eq. 1), square (Eq. 2), moving window integrator (Eq. 3), and detection.

$$y(nT) = \frac{2x(nT) + x(nT-T) - x(nT-3T) - 2x(nT-4}{8} \quad (1)$$

$$y(nT) = [x(nT)]^2 \quad (2)$$

$$y(nT) = \frac{1}{N}[x(nT-(N-1)T) + x(nT-(N-2)T) + \ldots + x(n \quad (3)$$

The instantaneous heart rate computed directly from R-R interval. In clinical settings, heart rate is measured in beats per minute (bpm). So, the formula for determining heart rate from RR interval is given below (Eq. 4).

$$Heart\_rate = \frac{60,000}{RRInterval(ms)} \quad (4)$$

III. Implementation Result

Using proposed system design the EEG wearable device was developed. The ECG signal obtained from the wearable device. The recording includes signals from several electrodes on the skin.

For the time-series analysis of the heart rate signal, are using intervals between the heartbeats and its variation over time. The most important of ECG signals analysis is to determine the position of the R-peaks since they correspond to the heartbeat and the position of other peaks and waves can be found relative to the defined R-peak. The R-peaks are detected using moving window integrator.

From a signal processing the position of all R-peaks, RR-intervals between R-peaks and differences between adjacent RR-intervals are found. From these parameters, the analysis of the heart health state can be implemented and the potential heart state disorders will be predicted.

Thus, relatively stable real-time ECG signal is successfully monitored and diagnostic ECG elements, such as R-peaks, are detected.

Conclusion

In summary, we developed the wearable system for outpatient ECG monitoring. The architecture of ECG monitoring system is based on the three-level scheme. The configuration of ECG wearable device to provide healthcare service was designed. The acquisition and processing of ECG signal are presented. The diagnostic ECG elements, such as R-peaks, are detected. Analysis of continuous heart rate signal and procedures for extracting of heart rate parameters are described.

[1] Slideshare.net. (2019). Dynamic Data Analytics for the Internet of Things: Challenges and Opp…. [online] Available at: https://www.slideshare.net/PayamBarnaghi/dynamic-data-analytics-for-the-internet-of-things-challenges-and-opportunities/8 [Accessed 17 Jan. 2019].

[2] Kim, Y. K., Wang, H., & Mahmud, M. S. (2016). Wearable body sensor network for health care applications. Smart Textiles and Their Applications, 161–184. doi:10.1016/b978-0-08-100574-3.00009-6

[3] J. Pan, & W. J. Tompkins, A real-time QRS detection algorithm, IEEE Transactions on Biomedical Engineering, Vol. 32, No. 3, pp.230-236, March 1985.

Statistical Analysis of Optical Image for Information System

Andrii Riazantsev
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
drew.ryazancev@gmail.com

Ganna Khoroshun
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
an_khor@i.ua

Oleksandr Ryazantsev
V. Dahl East Ukrainian National University
Severodonetsk, Ukraine
a_ryazantsev@ukr.net

***Abstract*—The surface Kirchhoff integral is numerically calculated by the fourth-order Newton – Cotes method with high accuracy. The intensity is normalized. Implemented quantization of the signal with a unit step for different distances of observation is carried out. The cross section of the central topological object intensity and its characteristics are shown. Variation curves and cumulates are constructed for intensity distributions at different observation distances. The basic statistical parameters of distributions are determined, such as, the mode, the mean, the median and the standard deviation.**

Keywords—optical image, statistical analysis, information optical system

I. INTRODUCTION

Information technology [1-3] is the broad subject concerned with all aspects of managing and processing information. It includes different levels — from the physical hardware to the operating systems, applications, databases, storage, servers etc. Database is set of independent materials, presented in an objective form and systematized in such a way that these materials can be found and processed using a computer. In the research laboratory usually a big volume of data is analyzed. The functions of automated measuring information optical system (AMIOS) include: data collection and analysis, storage and retrieval of information, protection and transmission of information on laser radiation characteristics. An optical image contains topological objects such as: maxima, minima and zeroes of intensity. Based on them we can define the systematization method for database creating. Computer science is focused entirely on efficiently programming computers using mathematical algorithms

II. CALCULATION OF DATA AND ITS STATISTICAL ANALYSIS

The problem of passing laser beam through a set of optical elements can be solved by the Kirchhoff integral:

$$A(x_1, y_1, z) \sim \iint A_0(x, y) \exp(i\Phi_0(x, y) + if(x - x_1, y - y_1))dxdy \quad (1)$$

where $A_0(x, y)$ - initial amplitude in the plane of the diffraction element XY, $\Phi_0(x, y)$ - the initial phase in the plane of the diffraction element, $f(x - x_1, y - y_1)$ - the function of describing the diffraction phenomenon, which is observed at a distance *z* from the screen on the plane X_1Y_1 .

For the scalar light field an optical image has information parameters: beam intensity and phase; identifying parameters: wavelength of laser radiation, beam waste radius, initial distribution of phase and beam intensity; as well as the structural parameter - the image observation distance z from the exit window of the laser.

We normalize the received data for intensity, so that the maximum value is 255, which corresponds to the maximum value for experimental images obtained with a CCD camera. The next step is quantization of the data with a unit interval. The systematization method for database creating is based on coordinates of topological objects and their parameters, such as core ellipticity and angle of inclination.

After this image is ready for implementation of statistical analysis methods. First of all, carry out a data grouping using the variation series. Variation curves obtained at different distances are analogous to each other. The central part of variation curves is similar to the normal distribution with a positive excess. It has a high narrow peak, indicating the accumulation of frequencies in the middle, the typicality and reliability of the average value. The mean value $\bar{I}$ is calculated by the formula:

$$\bar{I} = \frac{\sum_1^N I_i}{N}, \quad (2)$$

where I_i – the observed intensity in the *i*-th group N – the number of quantized intensities. The standard deviation σ represents the degree of scatter of data around the mean value and is defined by the

$$\sigma = \sqrt{\frac{\sum_1^n (I_i - \bar{I})}{N}}. \quad (3)$$

In result, we get the following statistical values for the image at z=40cm: the arithmetic mean is 139, the mode is 141 and the standard deviation is about 48 are changed slowly with a distance.

Another method of data representation is a cumulative curve. The curve for smaller distance has a more expressive

form, due to the fact that a light field decreases with the growing of observation distance.

CONCLUSION

The data obtained by the theoretical investigation of the diffracted field is performed by the surface Kirchhoff integral, which is numerically calculated, using the fourth-order Newton – Cotes method. The intensity is normalized and quantized with a unit step for different distances of observation. The cross-section of the central topological object intensity and its characteristics are shown. Variation curves and cumulates are constructed for intensity distributions. The basic statistical parameters of distributions are determined, such as the mode and the median from the graph, the mean and the standard deviation by calculations.

[1] M. Z. Zgurovsky, I. I. Kovalenko, V. M. Mikhailenko, Introduction to Computer Information Technologies, Teach. manual .- K .: View of the European Unitary Enterprise, 2002.- 265 p. ISBN 966-7508-74-9.

[2] D. S. Yadav Foundation of Information Technology, New Age International Publisher; 3rd edition, 2006

[3] V. Rajaraman, Introduction to information technology, Prentice hall India PVT., Limited, 2003, 384 p.

Wireless Sensor Network Conditions Monitoring and Diagnosis

Gennady Krivoulya
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
krivoulya@i.ua.

Vladislav Sergienko
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
vladislav.serhienko@nure.ua

***Abstract*—The paper discusses the adaptive neuro-fuzzy inference system ANFIS for intellectual diagnostics of large-scale wireless sensor networks. The solution for functional diagnostics of wireless sensor network is realized by the expert system designed on the knowledge base in the form of a neuron-fuzzy network.**

Keywords—ANFIS, wireless sensor network, diagnostic parameters, neural network, fault detection, fault tolerance

I. INTRODUCTION

Last advances in low-power wireless technologies have enabled us to expand wireless sensor networks (WSN) on different new networked systems. Besides many algorithmic studies that focus on designing efficient schemes or protocols to coordinate large-scale sensor networks, there are also systematic studies that make efforts in optimizing sensor networks in practice, which are usually tested on lab-scale testbeds or small scale deployments.

WSN node faults are usually due to the following causes [1]: the failure of modules (such as communication and sensing module) due to fabrication process problems, environmental factors, enemy attacks and so on; battery power depletion; being out of the communication range of the entire network.

Existing approaches to diagnosing sensor networks are generally sink-based and grounded on actively pulling state information from all sensor nodes to the central point what is referred to as “centralized analysis”. However, the sink-based diagnosis tools incur huge communication overhead to the traffic sensitive sensor networks. Also, due to the unreliable wireless communications, sink often obtains incomplete and sometimes suspicious information, leading to highly inaccurate judgments.

The main purpose of this work is the development of automated methods for intelligent functional diagnosis for WSN.

In the process of achieving the main goal, the following tasks are formulated and solved: conducting a continuous analysis of the technical state of the WSN in the process of functioning without disturbing functional links; operational receipt of information about the technical state of the WSN at an arbitrary time; elimination of the need for additional stimulus signals for WSN in the diagnostic process; possibility of predicting deviations of the technical state of the WSN from normal in the process of obtaining current data from sensors.

II. INTELLIGENCE DIAGNOSTIC OF WSN

The information part of the intelligent diagnostic system (IDS) [2, 3] provides accumulation, storage and transfer of information to other parts of it, and also implements the interface of the end-user. Data from sensors is unstructured and requires further processing. The need for real-time decision-making results in the fact that the number of decision trees constructed according to incoming data should be equal to the number of counts (analog of conveyor data processing). Tree decision trees for each time interval require significant memory costs for the IDS, so averaging for input data is usually used to reduce such costs. However, information on current changes in data from sensors over a period of time may be lost, which is a significant disadvantage of the methods for calculating averages. The problem of a significant amount of complex object data can be solved by using these data as a training sample for a neuro-fuzzy knowledge base.

There are mainly two sorts of node faults in WSN. The primary type is function fault, in which the sensor node cannot convey the data packet suitably. The second type is data fault, in which the node can convey the information bundle effectively yet the information gathered by sensor node is off base.

To evaluate the health of a sensor node, a binary logic function X is often used with a set of its values {0,1}. At the same time, if x is 1, then the node is operational and if x is 0, then the node is inoperable. However, fuzzy X values are required to use in the IDS.

For the fuzzification of input the crisp values convert of the input variables X and output variables Y into fuzzy, select the appropriate distribution functions and the number and values terms.

We consider a crisp variable between “0” and “1” and it has five terms with the following limits. – "0-0.1" - "Very Close to 0" ; – "0.1-0.2" - "Close to 0"; – "0.2-0.8" - "Average Value"; – “0.8-0.9” - “Close to1”; – "0.9-1" - "Very Close to 1".

We apply this fuzzification for all input variables and for output variables. Figure 1 is shown five membership functions of the input variable X.

To implement the diagnostic algorithms, we use the MatLab and ANFIS system.

The adaptive network-based fuzzy inference systems (ANFIS) is used to solve problems related to parameter identification. ANFIS is basically a graphical network

representation of Sugeno-type fuzzy systems endowed with the neural learning capabilities. The network is comprised of nodes with specific functions collected in layers. ANFIS is able to construct a network realization of IF / THEN rules. All computations can be presented in a diagram form. ANFIS normally has 5 layers of neurons of which neurons in the same layer are of the same function family.

The experiment of IDS functioning was provided. The possible measures of numerical values for 24 diagnostic parameters (DP1, ..., DP24) are estimated. The values of the sensor readings are obtained at discrete moments of time t_0, t_1, t_2, ..., t_i. The time interval (t_i+1-t_i) between two adjacent dimensions is selected taking into account the speed of the change of diagnostic parameters. All 24 characteristics will play the role of diagnostic parameters in the process of intellectual diagnosis.

CONCLUSION

The IDS considered in this paper, along with the use of traditional knowledge base, allows us to use the neural networks and to formalize the above practical problems that arise during the operation of various technical objects to achieve the main goal of work.

The ANFIS algorithm is used in order to improve the efficiency and accuracy of diagnostic sensor node. Because of using different ways to train and simulate ANFIS data within a single wireless sensor node, we generate a kind of intelligent system.

[1] Harshitha, Dr. K N Shreenath. Fault Detection Mechanisms In Wireless Sensor Networks-A Review. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 7, ISSUE 3, MARCH 2018. pp 78-83.

[2] Krivoulya G. Expert diagnosis of computer systems using neuro-fuzzy knowledge base/ Krivoulya G., A. Lipchansky, Ye. Sheremet EWDTS'2014: Proceding. of international conference., 15-17 September, 2016. – Erevan, – P. 619-622

[3] Krivoulya, G. Expert evaluation model of the computer system diagnostic features / G. Krivoulya, A. Shkil, D. Kucherenko, A. Lipchansky, Ye. Sheremet // EWDTS'2014: Proceeding of international conf.,26-29 September,2014.– Kiev,Ukraine,2014 –P. 286-289.

Author Index

Bakitko, Denis 23, 35
Barbaruk, Lina 9, 23
Berezhnyi, Oleksandr 13
Biloborodova, Tetiana 13, 19, 21, 37, 40
Chaturvedi, Arpana 7
Davidenko, Mykyta 19
Derevianchenko, Viktoriia 13, 31, 40
Derkach, Maryna 26
Fursa, Polina 31
Hoha, Maksym 11
Hulevska, Ksenia 29
Kardashuk, Volodymyr 15
Khoroshun, Ganna 42
Kobylin, Oleg 33
Koverha, Mark 31
Krivoulya, Gennady 44
Krytska, Yana 37
Kvasov, Ivan 9
Lavrinenko, Olha 9
Lyashenko, Vyacheslav 33
Mikhaylova, Alice 23, 35
Mokhonko, Viktoria 29
Nedzelskyi, Dmytro 17
Nesterov, Maksym 35
Riazantsev, Andrii 42
Ryazantsev, Ivan 33
Ryazantsev, Oleksandr 33, 42
Sergienko, Vladislav 44
Shumova, Larisa 29
Siriak, Rostyslav 21
Skarga-Bandurova, Inna 13, 26, 31, 37, 40

Наукове видання

THEORETICAL AND APPLIED COMPUTER SCIENCE AND INFORMATION TECHNOLOGIES

III International Conference TACSIT-2019 Proceedings

Редактор Скарга-Бандурова І.С.
Комп'ютерна верстка Білобородова Т.О.

Підп. до друку 25.05.2019. Формат 60×84/16. Папір офсет. Гарнітура "Times New Roman". Обл. друк. арк. 2,73.
Наклад 60 прим.

V. Dahl East Ukrainian National University Publishing

pr. Central 59-a, 93400, Severodonetsk, Ukraine
e-mail: uni@snu.edu.ua,
uni.snu.edu@gmail.com