ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

PROGRESSIVE INFORMATION TECHNOLOGIES

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

UDC 621.396.93:004.032.26

THE CLOUD GNSS DATA FUSION APPROACH BASED ON THE MULTI-AGENT AUTHENTICATION PROTOCOLS' ANALYSIS IN THE CORPORATE LOGISTICS MANAGEMENT SYSTEMS

Burlachenko I. S. – Senior Lecturer of the Department of Computer Engineering, Petro Mohyla Black Sea National University, Mykolaiv, Ukraine.

Savinov V. Yu. – PhD, Associate Professor of the Department of Computer Engineering, Petro Mohyla Black Sea National University, Mykolaiv, Ukraine.

Tohoiev O. R. – Postgraduate Student of the Department of Computer Engineering, Petro Mohyla Black Sea National University, Mykolaiv, Ukraine.

Zhuravska I. M. – Dr. Sc., Professor, Professor of the Department of Computer Engineering, Petro Mohyla Black Sea National University, Mykolaiv, Ukraine.

ABSTRACT

Context. Modern corporate logistics management systems or tracking systems consist of wireless positioning systems. Typically, mobile technologies use signal receivers of satellites of GNSS. However, there is the general problem of data transmitting issues to analytic centers for further in the corporate logistics management systems. The object of the study was to develop the solutions alternative to GNSS.

Objective. The goal of the work is the increasing vehicle location access control accuracy, based on the multi-agent authentication protocols' analysis.

Method. The study proposes the cloud data fusion platform offers to collect and archive data about all movable objects that can be on the road between departments inside the same warehouse area. Since there are operations with different wireless base stations' participation, so, the movable object can don't enough time for a movement trajectory analysis in real-time. Thus, data processing must fulfill in the cloud dispatching center and data flow fusion is needed. An equations system that identifies the vehicles' location based on the method of multi-agent authentication process analysis after the GPS signal loss, was proposed. The multi-sectional configuration of the recursive neural network and the usage of cloud data fusion made it possible to increase the accuracy of vehicle location determination.

Results. The developed method has been implemented in software and investigated for solving the problems of vehicle location control accuracy in the corporate logistics management systems.

Conclusions. The conducted experiments have confirmed the proposed approach and allow recommending it for use in practice for solving the problems of improving the efficiency of vehicle location determination via the role-based access control integrated with multi-agent authentication service.

KEYWORDS: cloud technologies, data fusion, multi-agent analysis, authentication protocol, recursive neural network, management system, corporate logistics.

ABBREVIATIONS

AP is an access point;
API is an application programming interface;
AT is an access token;
BS is a base station;
CDN is a content delivery network;
CP is a control point;
DF is a data fusion;
GNSS is a global navigation satellite system;
GPU is a graphics processing unit;
GRU is a gated recurrent unit;
LSTM is a long short-term memory;
MAAS is a multi-agent authentication service;
MS is a mobile station;

DOI 10.15588/1607-3274-2021-4-9

NS is a navigation system;
NSA is a navigation system antenna;
RBAC is a role-based access control;
RNN is a recursive neural network;
RO is a resource owner;
RTLS is a real-time locating system;
SSO is a single sign-on;
URI is a uniform resource identifier;
UA is a user-agent.

NOMENCLATURE

A_{AT}^{RO} is a set of MAAS agents for serving a protected resource to which the RO has access using ATs;

A_{AT}^{AS} is a set of MAAS agents that verify the validity of the authorization grant, and then sends the AT to other agents;

A_{AT}^{RS} is a set of MAAS agents that provide access to the protected resource to other agents if AT is validated;

A_{RT}^{AS} is a set of MAAS agents that verify the authorization grant is correct and returns the requested refresh token to the agent;

A_{UA}^{RS} is a set of MAAS agents that redirect the RO using the UA to the authentication interface of the authorization server;

C_A is an authorization code;

CH_{cmpr}^{TL} is a tasks' log compressed at the vehicle host's storage;

$DF_{p(\ell n,\ell t)}^{GPS}$ is a qualitative characteristic of the DF at the vehicle's position;

dt is a time offset between the MS and CP;

G_{AU} is an authorization grant;

IT_E is an access error;

$k_i \mid i = 1\ldots N$ are priority coefficients that are configured according to the vehicles' tasks;

KP_{cur}^{A} are key points that planned on the tasks' route;

$p(ln, lt)$ is a vehicle's position during the task completion;

PC_{RO} are credentials;

P_R is a problem range;

R_{AU} is a request to access the required resource;

R_{pld} is a number of a successful request for data;

Rx is a distance from the satellite to NSA;

rx is a distance from BS to NSA;

T_A is an access token;

T_{AC} is an access token with a long time to live;

T_R is a refresh token;

T_r is a task route;

Tx is a time of signal arrival on the MS;

tx is a time of signal arrival on the CP;

v is a value of the speed of radio waves.

INTRODUCTION

Modern corporate logistics management systems or tracking monitoring systems consist of wireless positioning systems. Typically, mobile technologies include signal receivers of global satellites of GNSS. Technical complexity aims to improve GNSS technology were investigated. However, it is the general source of data transmitting issues. The developments with the main goal of GNSS alternative solutions have been reviewed. Thus, the cloud data fusion platform offers to collect and archive data potential techniques from all movable objects that can be on the road and ongoing processing in the cloud dispatching center. There are operations with several times performance and different base stations' participation. The movable object does not have time for a considerable distance movement between departments inside the same warehouse area, so data flow fusion real-time performance is needed.

To automate the determination of mobile object location in the corporate logistics management systems, it is necessary to have an analytic approach of a decision this task after the GPS signal loss.

The object of study is the process of determination of mobile object location building by the solutions alternative to GNSS.

The subject of study is the methods based on the multi-agent authentication protocols' analysis.

The purpose of the work is to increase the vehicle location access control accuracy.

1 PROBLEM STATEMENT

The problem of each company, which manages a large number of vehicles, is to ensure the rhythm of transportation. Besides, goods must be delivered to the destination at the specified time. To make the dispatching management effective, it is necessary to obtain promptly the coordinates (x, y) on the whereabouts of a particular vehicle at any time and, if necessary, to correct T_r. Modern traffic control systems that use satellite-based coordinates can allow quick track of vehicles' movement information, such as forklifts, even without the intervention of drivers. While, the Rx, rx, dt, Tx, tx, v are known.

The problem is the low quality of GPS signal during the management of the barge's construction in the shipyard shown in Figure 1.

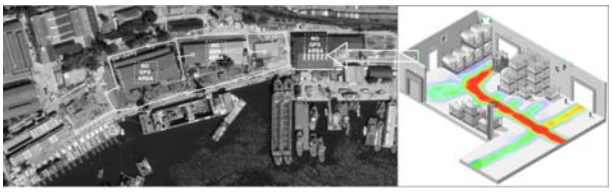


Figure 1 – Warehouses where forklifts cannot be identified by GPS tracked ID

The purpose of the research is to demonstrate the possibilities of vehicle identification, various methods and

DOI 10.15588/1607-3274-2021-4-9

techniques of database analysis, the geo-distributed data, and the selection of parameters visualization in particular.

They can be used to reduce IT_E and improve the operations of corporate logistics systems via applying of the MAAS agents A_{AT}^{RO}, A_{AT}^{AS}, A_{AT}^{RS}, A_{RT}^{AS}, A_{UA}^{RS} and others with analysis of the multi-agent authentication process, determine their tokens T_A, T_R, T_{AC}, etc., and the redefinition T_r as a result.

2 REVIEW OF THE LITERATURE

In recent years, the world has been actively conducting research to improve GPS technology. One of them related to counteraction to generating noise interference technologies and anti-satellite weapons (ASAT, etc.) [1]. Research is also underway to improve satellite protection against space radiation. The development of so-called "radiation tolerance technologies" contributes not only to military communication reliability. However, it can also help to provide missions for commercial use and environmental monitoring [2, 3]. There are a lot of military technologies that lie in civil areas.

In addition to the research aim of GPS improving technology, there is the main goal to replace it with alternative solutions. We should note that such solutions will differ significantly for indoor or outdoor tasks. In the future, external objects can be used for such navigation Earth's magnetic field space orientation systems as MAGNAV, QuASAR, etc. [4, 5]. Nowadays we have devices, which network user coordinates can have 10 meters accuracy determination [6].

Noise-immune solutions [7, 8] will allow the corporate networks' objects interaction. They belong to remote segments connected by heterogeneous solutions [9, 10]. In such a situation, the main issue is the network authentication of remote corporate clients' geography. The devices should have a focus on a web interface that does not have a built-in authorizing. In this case, it is necessary to develop the OAuth 2.0 [11] authorization standard basis for a corporate network authentication service.

At the same time, it is important to improve the accuracy of determining the location of a network object. Precise gyroscopes with a nearly symmetrical mechanical resonator [12] navigation system are the most promising technology for vehicles. However, this method is necessary for the moving hosts of corporate network special equipment.

Most smartphones indeed contain gyroscopes, but their accuracy is low. Therefore, the more relevant and budgetary solution today is indoor Wi-Fi tracking usage. In this case, the minor error occurs – from 2.5 m to 1.3 m [13, 14] – improved with a multi-agent approach [15].

We should take into account the spatial structure of the internal environment. It has more complexity than the external one. It occurs due to the potential conception of people in the premises. In this case, the corporate network hosts' positioning algorithms should have the ability to process large amounts of data [16].

Scientists have systematized user identification, the same as other ways of protection against unauthorized access. Such scientists as V. Makhonin, V. Chudnikov, I. Rudakov have proposed RTLS mobile subscribers coordinates determining method.

To solve the problem of warehouses where forklifts cannot be identified by GPS the large data sets with the help of software tools analysis need to be processed. The developed approach will process the large amounts of data associated with barge construction scenarios.

The software tools should also be able to work with fast-flowing large volume data, the same as with structured and poorly structured data as well. There are a lot of methods and techniques for big data analysis which we consider effective, but machine learning and artificial neural networks will be applied.

3 MATERIALS AND METHODS

The vehicle (car, vessel, drone, etc.) is equipped with an NSA that is used to receive signals from satellites.

This antenna is covered by an optical casing, sealed, protected from the weather. The receiver of the NS installed in the car determines its own coordinates and periodically transmits the data to the CP by the means of the special radio channel. In addition, a display can be installed in the driver's cab, which displays a map of the area of the vehicle's location. In addition, the car position coordinates and speed vector can be indicated on the map.

The proposed system scheme of the cloud GNSS data fusion approach is illustrated in Figure 2.

The process of calculating the coordinates and route displaying is constant, the driver's intervention, as a rule, is not required. By the way, other information can be transmitted to the CP along with the navigation data, for example, information about the fuel level in the tank or tank, a driver's workplace presence status or engine condition, whether the engine switched off or started, etc. Error notification messages can also be displayed automatically, e.g., signals about the unexpected opening of the vehicle door (while driving), messages about the unauthorized transported containers.

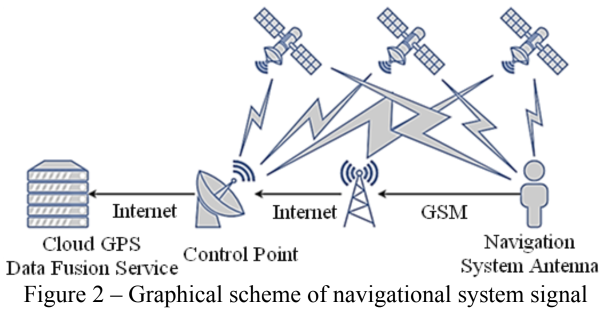


Figure 2 – Graphical scheme of navigational system signal searching process

Thus, the collecting and archiving data from all vehicles on the road is always processing in the dispatching center. The dispatcher can determine the location of each

DOI 10.15588/1607-3274-2021-4-9

vehicle at any time, and make the traffic overall picture or view the schedule of a particular car for any period.

The system components: MS, three satellites, and CP. The coordinates of CP and BS are already known. The clocks of the MS and the measuring station are not synchronized. The first BS sends a pulse to the NS and the CP. On NS and CP the time of arrival of a signal is noted; the time value specified by the NS is sent to the CP (a possible option when the calculations are performed on the NS itself). This operation is performed at least three times with the participation of different BS. Of course, everything needs to be done quickly so that the NS does not have time to move a considerable distance. Once the signal time data has been collected, calculations can begin. To calculate the coordinates of the subscriber it is necessary to solve a system of formula (1):

$$\begin{cases} R1 - r1 = v(T1 - t1 + dt), \\ R2 - r2 = v(T2 - t2 + dt), \\ R3 - r3 = v(T3 - t3 + dt). \end{cases} \quad (1)$$

As we see, in the system there are three unknown quantities (*Rx* can be expressed in terms of *x, y*) – the coordinates of MS (*x, y*) and *dt*, so that the calculation of coordinates becomes a solvable task.

In Figure 3 the diagram demonstrates a general data flow in the corporate cloud platform with data fusion integration. Navigation receiver (so-called "AP") consists of three main nodes: radio frequency receiver unit; a computing unit in which information is processed and, in fact, finding the coordinates of the receiving point; information display unit.

Figure 4 shows cases when P_R of GPS occurred. The technical characteristics of the navigation receiver have been influenced by the vehicle's workflow. Therefore, currently produced various AP, designed to work in navigation systems and differ in functionality, characteristics, cost, and appearance. For example, the receivers are placed in moisture-proof enclosures, the receivers are additionally combined with a sounder and receivers for vehicles with space communication devices so that you can send and receive messages.

To analyze the data fusion process, the surface formula (2) was proposed:

$$DF^{GPS}_{p(\ell n,\ell t)} = \sum_{i=1}^{N-1} k_i \cdot f(CH^{TL}_{cmpr}, KP^{A}_{cur} \in T_R) + k_N \cdot \sum_{0}^{t_f} VT_{t_h}. \quad (2)$$

Every vehicle performs current activities as KP^{A}_{cur} that are planned in the T_R.

Figure 5 shows the analytical surface with GPS tracker signals' losses.

If navigation receivers are designed to work as part of more complex systems, they can be made in the form of boards that are installed in the racks of the main equipment. The specific implementation of such AP depends on where and by whom they will be operated, and each customer can choose a receiver based on their own requirements.

To solve the peaks problem the approach based on analysis of multi-agent authentication protocols was proposed.

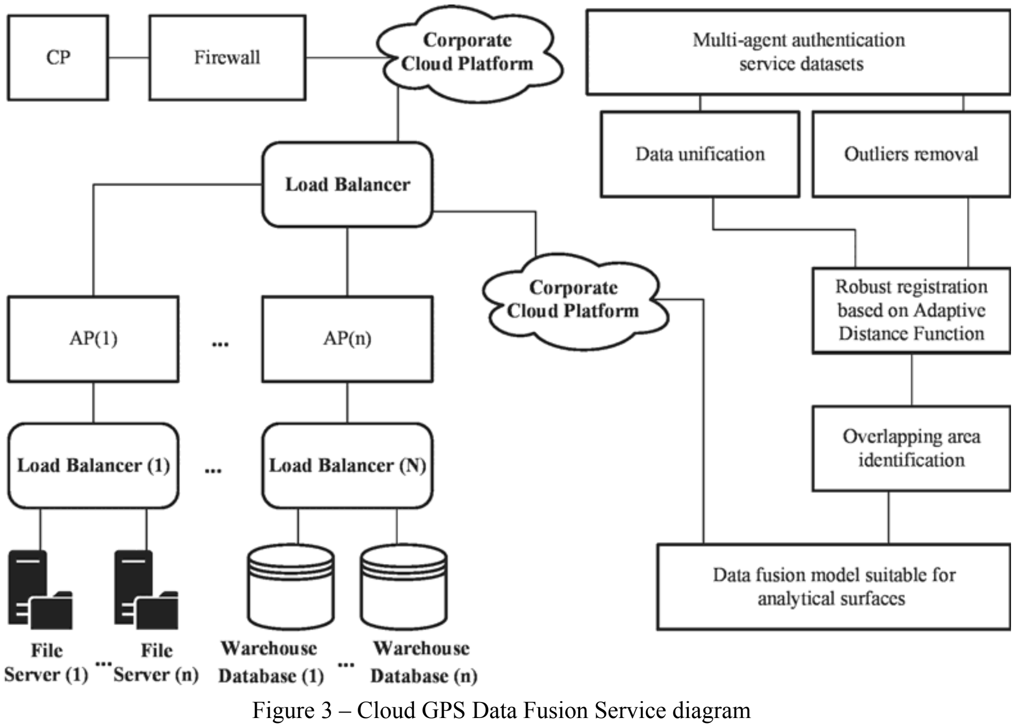


Figure 3 – Cloud GPS Data Fusion Service diagram

DOI 10.15588/1607-3274-2021-4-9

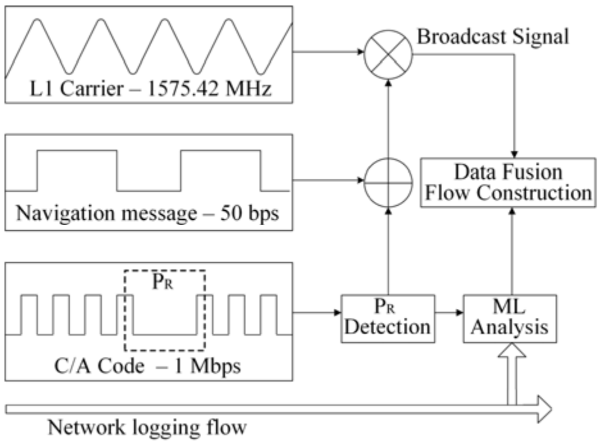


Figure 4 – Detailed GPS signal analysis by Cloud GPS Data Fusion Service

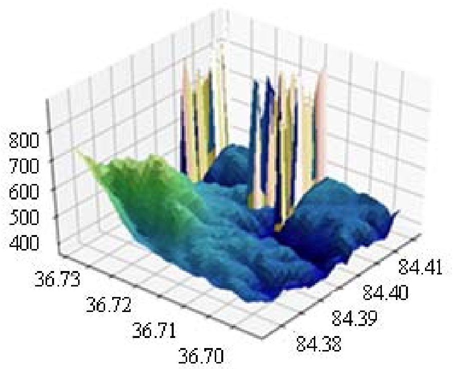


Figure 5 – Data fusion surface with peaks when GPS signal lost

4 EXPERIMENTS

The data analysis can be fulfilled via a multi-agent authentication protocols service.

A MAAS is quite complex, but if software architecture to manage the barge's construction in the shipyard with many services is complex, then OAuth 2.0 will be a base when developing the authentication services need to identify the departments that were involved in the shipbuilding process. The problem of authorization in dozens of services was encountered. This problem was solved with a MAAS. It helped implement seamless authentication across various services and migrate observable agents' data to separate databases to improve the efficiency of the vehicles monitoring during the barge's construction.

The MAAS service has three main tasks:

1) Single point authentication (so-called "Single Sign-On" or SSO) for all system services. Services do not store credentials but trust this to one dedicated service.
2) Safe and granular access to resources. Safe because passwords are stored in one place and are as secure as possible. Granular, since service agents can configure access to resources, based on the data that came from the authentication service.
3) Centralized agent and access management because all information about the agents is stored in the authentication service.

The MAAS defines several roles. A resource owner is an agent that has access rights to a protected resource. An agent can be a user or some kind of system. A protected resource is an agent that handles routes of HTTP endpoints, which can be: API endpoint, file on the CDN, and web service. A resource server is an agent that stores a protected resource to which the resource owner has access. Observer agent is an application that requests access to a protected resource on behalf of the resource owner and with his permission – with authorization. An authorization server is an agent that issues a token to a client to access a protected resource after the successful authorization of the resource owner.

Each agent in the interaction can combine several roles. For example, an agent can be a resource owner at the same time and request access to their own resources. In Figure 6 agents' interaction scheme was considered.

An abstract sequence diagram of MAAS RBAC of interaction between agents includes several objects. The Observer agent sends a request R_{AU} to access the required resource owner agent A_{AT}^{RO}. The A_{AT}^{RO} gives back to the observer agent an authorization grant G_{AU}, which confirms the identity of the A_{AT}^{RO} and rights to the resource that the observer agent is requesting access to. Depending on the flow, this can be a token or credentials. The Observer agent sends the G_{AU} obtained in the previous step to the authorization agent A_{AT}^{AS}, expecting an access token T_{AC} from it to access the protected resource. The A_{AT}^{AS} verifies the validity of the G_{AU}, and then sends the T_{AC} back to the observer agent. After receiving the T_{AC}, the observer agent requests the protected resource from the resource agent A_{AT}^{RS}. The A_{AT}^{RS} verifies the correctness of the T_{AC} and then provides access to the protected resource. Observer agent as part of MAAS can interact with another authentication service of another department inside the same warehouse during the shipbuilding stages.

GPS tracker signal with vehicle identifier is not available, so in Figure 6 agent interaction includes also the observer agent coming with a G_{AU}, to the authorization agent A_{RT}^{AS} and asks to provide the access token T_A and refresh token T_R. The A_{RT}^{AS} makes sure that everything is fine with the G_{AU} and returns the requested T_A and T_R to the observer agent. The observer agent with a T_A requests a protected resource until it receives the first invalid token error IT_E from A_{RT}^{RS}.

After receiving an access error IT_E, the observer agent goes to the A_{RT}^{AS} with a T_R and asks to replace the expired T_A with a new one. In response, the observer agent receives a new T_A, as well as a new T_R, or the lifetime of the old T_R is extended.

The observer agent implemented in the vehicle receives some sort of successful authentication identifier, such as a string, which is associated with the data in the

DOI 10.15588/1607-3274-2021-4-9

database for further analysis. Grant refers to the data that represents the successful authorization of the client by the owner of the resource, used by the client to obtain an access token.

There are several ways to get a grant. Authorization code is used for confidential clients or web services. Clients' credentials are used for confidential clients that request access to their resources or resources previously agreed with the authorization server. Implicit is used by public clients that know how to work with redirection the URIs.

Resource owner password credentials and device authorization are used to authorize devices that can work over the Internet without browsers.

Proposed MAAS can handle described grant types shown in Figure 7.

According to the current MAAS RBAC implementation the first part with resource owner password credentials flow is recommended to be used at agent's interactions that are purely informational but used as an identification criterion of the vehicle inside the warehouse.

Interaction between agents is performed as follows. Resource owner agent A_{CGF}^{RO} passes credentials PC_{RO} to the observer agent. The observer agent uses PC_{RO} to obtain a T_A and T_R. The observer agent redirects the resource owner agent A_{ACGF}^{RO} using the agent A_{UA}^{RS} to the authorization agent A_{AC}^{AS} and specifies the client ID I_{CL} and the redirection URI U_R. Interacting with the A_{AC}^{AS} through the A_{UA}^{RS}, the A_{CGF}^{RO} is authenticated on the A_{AC}^{AS}. A_{CGF}^{RO} checks the rights requested by the observer agent and allows issuance. The A_{CGF}^{RO} is returned to the observer agent using the A_{UA}^{RS} back to the U_R that was specified. As a query parameter, code C_A will be added. With C_A, the observer agent is sent to the A_{AC}^{AS} instruction to receive an access token T_A in response (and a refresh token, if required). The A_{AC}^{AS} validates the C_A, making sure that the T_A is correct, and issues a T_A and optionally a T_R to the observer agent. The observer agent will be able to access the resource that will be used for vehicle location monitoring.

The A_{UA}^{RS} has one requirement: it must be able to work with HTTP redirects. Without this, the A_{CGF}^{RO} will not be able to get to the A_{AC}^{AS} and return with a grant.

Also, MAAS data analysis can use LSTM. According to the proposed MAAS RBAC diagrams, the CH_{cmpr}^{TL} is described by formula (3):

$$CH_{cmpr}^{TL}=\begin{bmatrix}\left\{\frac{R_{AU}}{G_{AU}},\frac{G_{AU}}{T_{AC}},\frac{R_{pld}}{T_{AC}},R_{pld}\right\}\\ \left\{\frac{T_{(A)U(R)}}{G_{AU}},\frac{R_{pld}}{T_A},\frac{IT_E}{T_A},R_{pld}\right\}\\ \left\{\frac{PC_{RO}}{R_{pld}},\frac{T_{(A)U(R)}}{PC_{RO}},R_{pld}\right\}\\ \left\{\frac{C_A}{I_{CL}},\frac{C_A}{U_R+U_A},\frac{T_A}{C_A+U_R},\frac{R_{pld}}{I_{CL}},R_{pld}\right\}\end{bmatrix}, \quad (3)$$

where R_{pld} needs to be used in LSTM to identify the vehicle's location.

The block diagram of the algorithm that implements the proposed approach is shown in Figure 8.

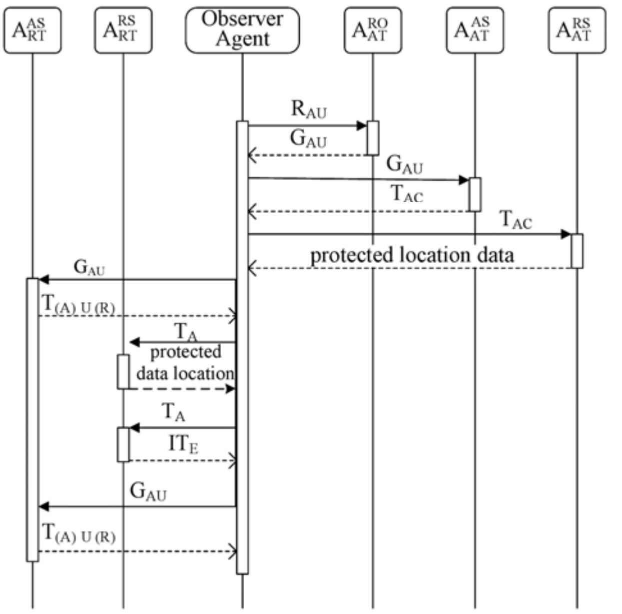


Figure 6 – RBAC scheme in multi-agent authentication service

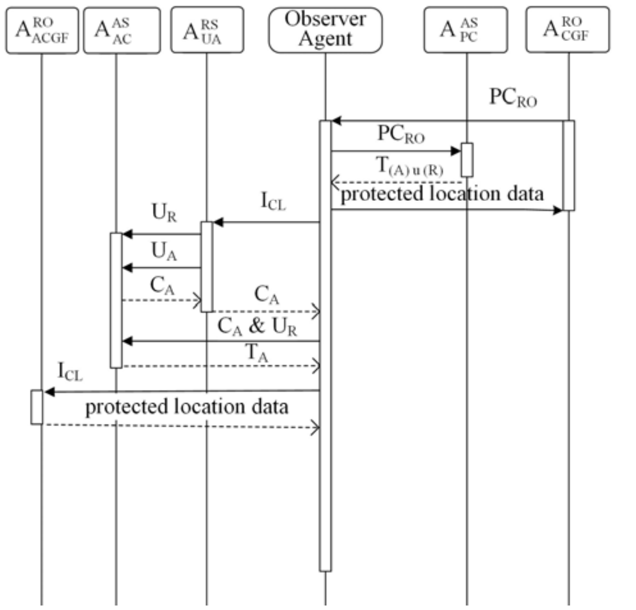


Figure 7 – Communication process between services to analyze protected location's data using MAAS RBAC

DOI 10.15588/1607-3274-2021-4-9

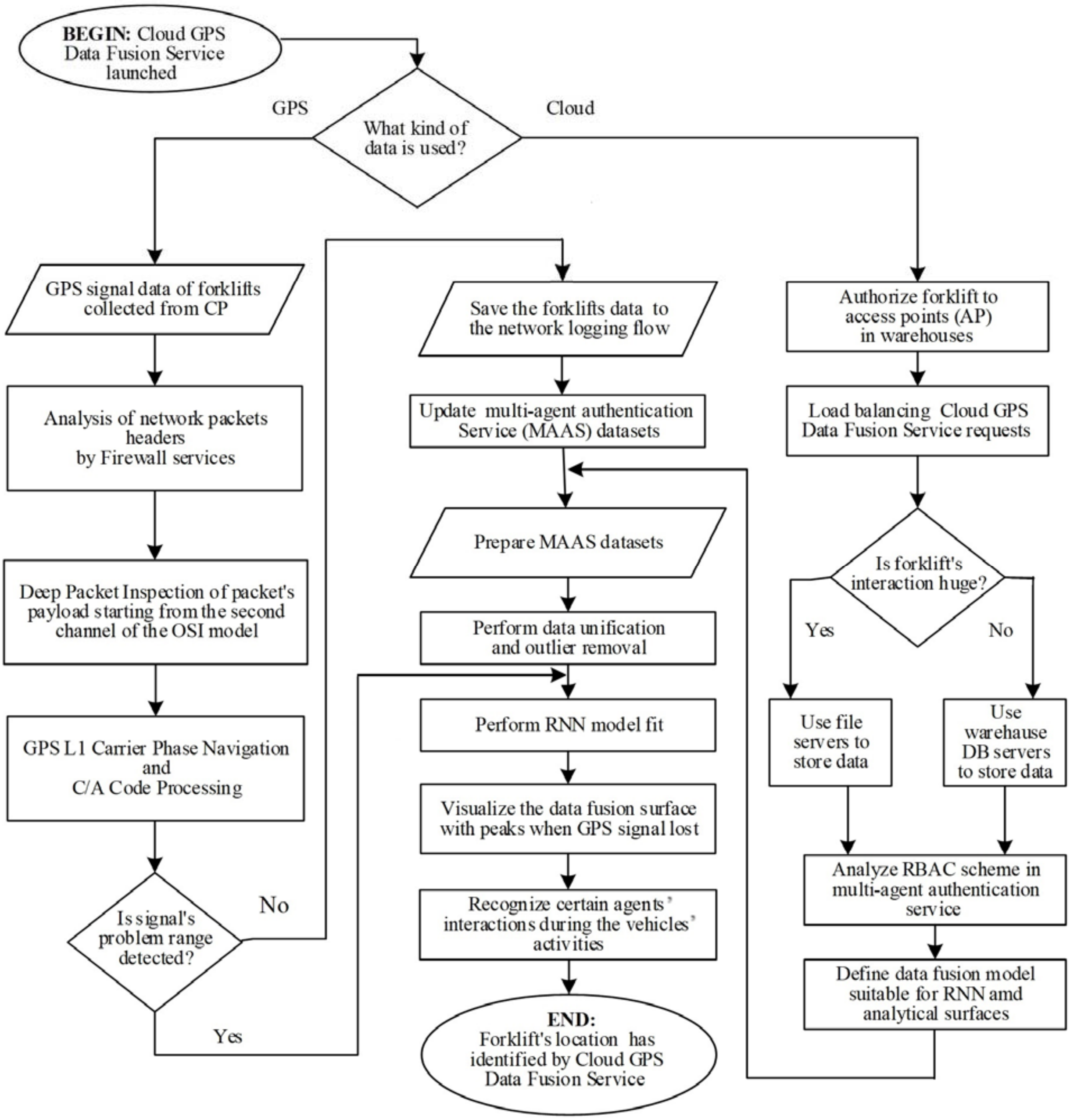


Figure 8 – Block diagram of the algorithm that implements the proposed approach

5 RESULTS

As a result of an experiment, a modification was described that solved the problem of long-term memory of the proposed RNN: when neurons remember recently received information well, but they can’t store in memory for a long time something that they processed many cycles ago, no matter how important that information may be.

This turns out to be convenient when several vehicles are simultaneously located within the same warehouse. In LSTM networks, internal neurons are equipped with a complex system of gates, as well as the concept of a cell state, which is a kind of long-term memory. The gate determines what information will enter the cellular state, what will be erased from it, and what will affect the result that the RNN will give at an iteration.

In Figure 9 the inputs of the *raw_LocCruteria* section described the raw quantitative values refers to agents’ interactions during the vehicles’ activities on the planned in the tasks’ route T_R. The section *relationCruteria* includes samples with CH_{cmpr}^{TL}.

DOI 10.15588/1607-3274-2021-4-9

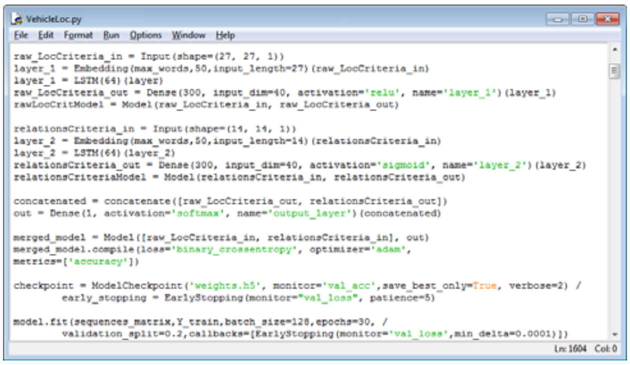
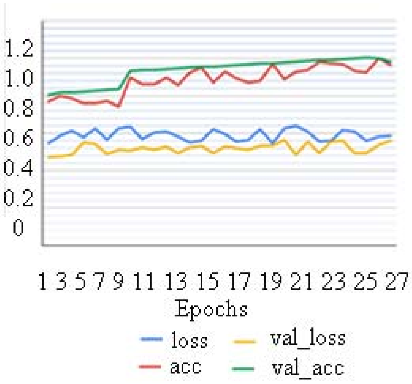
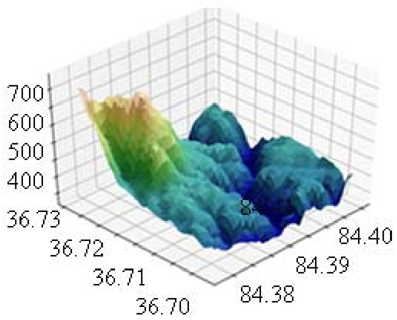


Figure 9 – Recursive neural network layers for searching vehicle's location based on MAAS protocols logging

LSTM networks are trained using the backpropagation through time algorithm, the idea of which is to expand the computation graph in time. From LSTM networks, we can build multilayer neural networks, passing the output sequence of the next layer to the next one.



a



b

Figure 10 – The accuracy plot of the agent's classification models (a) and smoothed activities surface (b) of forklift vehicle

It should be noted that such active attention to the considered family of models is currently due, in particular, to their high performance in many tasks. Like other recurrent neural networks, LSTM and GRU, especially two- and multilayer ones, are characterized by a rather complex learning procedure. The GPUs can significantly speed up the learning processes of deep neural networks, which is clearly demonstrated by the active implementation (and optimization) of the described recurrent models for GPU computing.

The location identifying uses the proposed method based on analysis of the multi-agent authentication process. As a result, shown in Figure 10b, the analytical surface has the following smoothed form without peaks.

6 DISCUSSION

As evident from Figure 10, the multi-sectional configuration of the recursive neural network and the usage of data made vehicle location possible to achieve an accuracy of 97.8% (Figure 10a) of the classification of forklift vehicles in the warehouse. Current MAAS uses the RBAC scheme, which is based on roles. Usually, the agent's role is given full access to a specific service, instead of being tied to specific functionality. A multi-agent authentication service describes how communication between services should be implemented to ensure secure authorization. Many nuances are described in sufficient detail, for example, the flow of interaction of nodes with each other, but some are left to the mercy of a specific implementation.

We should also note that a significant number of mobile technical systems include various positioning systems. Typically, mobile technologies include signal receivers of GLONASS/GPS navigation systems global satellites. There are several Earth orbit satellites of the satellite navigation system. They include ground system management and subscriber devices (so-called "APs") as special navigation receivers. Their main task is to receive signals from the satellite and determine the distance to the one. Subscriber receivers measure the distance to several currently visible satellites. Their coordinates are extremely accurate and can be reached at any time. The coordinates have equal distance surfaces intersection point calculation.

Three satellites' distance makes possible a single point of Earth's surface coordinates' determination. In this case, the receiving device's three equation data is very important. The subscriber receiver performs instantly automatic measurements and calculations. It increases the calculating of the coordinates' receiving point accuracy.

The proposed system uses several additional measures. This study investigated an equations system, which is used to calculate the current position of a moving object in real-time. As the input data for the calculations, the coordinates of the three satellites relative to the Earth and their distance to the desired object are used. The accurate determination of the moving object coordinates requires the real-time data constantly coming from NS, which is not always feasible due to possible signal interruptions.

Separating the entity of the agent and the application requesting access allowed to analyze data fusion ability. According to asynchronous interactions, developed software can manage vehicle navigation system rights separately from agents' rights. Instead of the credential, which has a certain set of rights and lifetime, we get access to metadata resources generated during the vehicle's movements between departments inside the same warehouse during one technological process, e.g., vessel building stages.

DOI 10.15588/1607-3274-2021-4-9

In this context, the research should also take into account, signals from satellites are emitted at two frequencies. If we compare the time of receiving signals, we can determine and take into account the ionosphere radio signals' delay. They increase the accuracy of measurements. The potential accuracy of determining the coordinates can be increased to several tens of centimeters. Modern navigation receivers can have 12 or more operating channels. They do not only increase the accuracy of coordinate determination but also speeds it up. As a rule, such receivers calculate the speed and direction of movement of moving objects.

However, the large amounts of traffic, receiving signals' difficulties, or data absence can cause functionality losses. Therefore, such conditions (long-distance industrial facilities, quarries, mines, areas with complex terrain, etc.), alternative positioning methods are needed to determine the coordinates and identify mobile objects. One of them is the reception of the signals from the fourth satellite. It can help to eliminate the systematic receiving device and satellite time difference error.

CONCLUSIONS

The actual problem of mathematical support development has been solved to automate defining the trajectory in real-time of the movable object during movement between departments inside the same warehouse area.

The scientific novelty of obtained results is that the method of data flow fusion real-time performance is firstly proposed. An equations system that identifies the vehicles' location based on the method of multi-agent authentication process analysis after the GPS signal loss, was proposed. The multi-sectional configuration of the recursive neural network and the usage of cloud data fusion allow increasing the accuracy of vehicle location determination.

The practical significance of obtained results is that the efficiency of the role-based access control integrated into multi-agent authentication service was improved. as well as experiments to study their properties are conducted. The experimental results allow us to recommend the proposed approach for use in practice, as well as to determine vehicles' location in the warehouse with an accuracy of 97.8%. The location of the forklift is determined by a cloud-based GPS fusion service. Time estimates of the costs of sending data to data centers in comparison with the costs of processing, archiving, and merging these data will be approximately equal. Time costs for data transfer will not take up most of the time, since the cloud infrastructure will be deployed for the local enterprise territory. Local computing and storage resources will be enough to locate the forklift in a time range of up to 15–20 minutes after the GPS signal has been lost.

Prospects for further research are to study the proposed cloud data fusion about the location of vehicles, being at various territorially remote warehouses but providing material support for a common logistical chain.

ACKNOWLEDGEMENTS

The work is supported by NIBULON Company (Mykolaiv, Ukraine) having expanded logistical capabilities. Without the provided datasets of its Integrated Shipyard Management System, without superior knowledge and experience, it would not be possible to obtain qualitative results in the research and therefore their support was very essential.

REFERENCES

1. Koplow D. A. An inference about interference: A surprising application of existing international law to prohibit anti-satellite weapons, *Georgetown Law Faculty Publications And Other Works,* 2014, Vol. 35, pp. 737–827.
2. Navigation and guidance systems, *BAE Systems,* 2019, URL: https://www.baesystems.com/en/product/navigation-and-guidance-systems.
3. Blyukher B., Ivanova L., Malakhov E., Sitnikov V. Probabilistic risk assessment of environmental hazards, *American Society of Mechanical Engineers, Pressure Vessels and Piping Division (PVP),* 2000, Vol. 400, pp. 69–71.
4. Canciani A. J. Absolute positioning using the Earth's magnetic anomaly field : Theses and Dissertations of PhD, 2016. URL: https://scholar.afit.edu/etd/251.
5. Kramer D. DARPA looks beyond GPS for positioning, navigating, and timing, *Physics Today*, 2014, Vol. 10, Issue 23, P. 67. DOI: 10.1063/PT.3.2543.
6. Retianza D. V. Duivenbode J. V., Huisman H., Lomonova E. A. Fault-tolerant controller and failure analysis of automotive electromagnetic suspension systems, *Ecological Vehicles and Renewable Energies (EVER 2019), Proceedings of the IEEE 14th International Conference*. Monte-Carlo, Monaco,·08–10 May 2019, 2019, No. 8813514. DOI: 10.1109/EVER.2019.8813514.
7. Rumiankov D., Zhuravska I., Solobuto L. , Musiyenko M. Reduction of noise similar to solar interference in computer networks based on power line communication, *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2017) : Proceedings of the IEEE 9th International Conference*. Bucharest, Romania, 21–23 September 2017, 2017, Vol. 1, pp. 215–221. DOI: 10.1109/IDAACS.2017.8095079.
8. Ukhina H. Sytnikov V., Streltsov O., Stupen P., Yakovlev D. Stability evaluation based on the sustainability triangle application for transfer functions above 2nd order, *Computer Science and Application Engineering (CSAE 2019), Proceedings of the 3rd International Conference,* October 2019, Article-No. 6, pp. 1–5. DOI: 10.1145/3331453.3361305.
9. Ovchar K., Borodin A., Burlachenko I., Krainyk Y. Automated recognition and sorting of agricultural objects using multi-agent approach, *CEUR Workshop Proceedings*, 2019, Vol. 2546, pp. 76–86.
10. Burlachenko I., Zhuravska I., Tohoiev O., Ukhan Ye., Tiutiunyk Ye. Multi-agent monitoring system for heat loss mapping of multi-story buildings, *CEUR Workshop Proceedings*, 2019, Vol. 2516, pp. 218–225.
11. RFC 8628: OAuth 2.0 Device Authorization Grant, 2019, URL: https://oauth.net/2/device-flow/.
12. McAlpine K. Small, precise and affordable gyroscope for navigating without GPS, Michigan Engineering, 23 March 2020.

DOI 10.15588/1607-3274-2021-4-9

13. Sivers M., Fokin G., Dmitriev P., et al. Wi-Fi based indoor positioning system using inertial measurements, *Next Generation Wired/Wireless Networking : Proceedings of the International Conference, 28–30 August 2017*, 2017, DOI: 10.1007/978-3-319-67380-6_69.
14. Asher V., Thakkar H., Tambe S., Bhavathankar P. Indoor object tracking using Wi-Fi access points, *Inventive Communic. and Computational Technol., LNNS,* 2020, Vol. 89, pp. 687–698. DOI: 10.1007/978-981-15-0146-3_64.
15. Tohoiev O., Burlachenko I., Zhuravska I., Savinov V. The monitoring system based on a multi-agent approach for moving objects positioning in wireless networks, *CEUR Workshop Proceedings*, 2020, Vol. 2608, pp. 79–90.
16. Berg K., Spil T. A. M., Effing R. The privacy paradox of utilizing the Internet of Things and Wi-Fi tracking in smart cities, *IFIP Advances in Information and Communication Technology*, 2019, Vol. 558, pp. 364–381. DOI: 10.1007/978-3-030-20671-0_25.

Received 04.08.2021.
Accepted 01.10.2021.

УДК 621.396.93:004.032.26

ХМАРНИЙ ПІДХІД ЗЛИТТЯ ДАНИХ GNSS НА ОСНОВІ МУЛЬТИАГЕНТНОГО АНАЛІЗУ ПРОТОКОЛІВ АУТЕНТИФІКАЦІЇ В КОРПОРАТИВНИХ ЛОГІСТИЧНИХ СИСТЕМАХ УПРАВЛІННЯ

Бурлаченко І. С. – старший викладач кафедри комп’ютерної інженерії Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

Савінов В. Ю. – канд. техн. наук, доцент кафедри комп’ютерної інженерії Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

Тогоєв О. Р. – аспірант кафедри комп’ютерної інженерії Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

Журавська І. М. – д-р техн. наук, професор, професор кафедри комп’ютерної інженерії Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

АНОТАЦІЯ

Актуальність. Сучасні корпоративні логістичні системи управління або системи стеження за положенням складаються з бездротових систем позиціонування. Як правило, мобільні технології використовують приймачі сигналів супутників GNSS. Однак, існує загальна проблема передачі даних до аналітичних центрів для подальшого використання у системах корпоративної логістики. Об’єктом дослідження є процес розробки рішень, альтернативних для GNSS.

Мета. Підвищення точності контролю доступу до місцезнаходження транспортного засобу на основі мультиагентного аналізу протоколів автентифікації.

Метод. У дослідженні запропоновано хмарну платформу злиття даних, яка передбачає збирати та архівувати дані про всі рухомі об’єкти, що можуть знаходитися на дорозі між відділами всередині однієї складської території. Оскільки існують операції з участю різних бездротових базових станцій, тож, рухомому об’єкту може не вистачити часу для аналізу траєкторії руху в режимі реального часу. Таким чином, обробка даних повинна виконуватися в хмарному диспетчерському центрі, і необхідне злиття потоків даних. Запропонована система рівнянь, яка визначає місцезнаходження транспортних засобів на основі методу мультиагентного аналізу процесу автентифікації після втрати сигналу GPS. Багатосекційна конфігурація рекурсивної нейронної мережі та використання злиття хмарних даних дозволили підвищити точність визначення місцезнаходження транспортного засобу.

Результати. Розроблений метод реалізований програмно та проведені дослідження для вирішення проблем точності контролю розташування транспортного засобу в корпоративних логістичних системах управління.

Висновки. Під час дослідження ефективність визначення місцезнаходження транспортного засобу була покращена за допомогою керування доступом на основі ролей, інтегрованого з мультиагентною службою автентифікації.

КЛЮЧОВІ СЛОВА: хмарні технології, злиття даних, мультиагентний аналіз, протокол автентифікації, рекурсивна нейронна мережа, система управління, корпоративна логістика.

УДК 621.396.93:004.032.26

ОБЛАЧНЫЙ ПОДХОД СЛИЯНИЯ ДАННЫХ GNSS НА ОСНОВЕ МУЛЬТИАГЕНТНОГО АНАЛИЗА ПРОТОКОЛОВ АУТЕНТИФИКАЦИИ В КОРПОРАТИВНЫХ ЛОГИСТИЧЕСКИХ СИСТЕМАХ УПРАВЛЕНИЯ

Бурлаченко И. С. – старший преподаватель кафедры компьютерной инженерии Черноморского национального университета имени Петра Могилы, Николаев, Украина.

Савинов В. Ю. – канд. техн. наук, доцент кафедры компьютерной инженерии Черноморского национального университета имени Петра Могилы, Николаев, Украина.

Тогоев А. Р. – аспирант кафедры компьютерной инженерии Черноморского национального университета имени Петра Могилы, Николаев, Украина.

Журавская И. Н. – д-р техн. наук, профессор, профессор кафедры компьютерной инженерии Черноморского национального университета имени Петра Могилы, Николаев, Украина.

DOI 10.15588/1607-3274-2021-4-9

АННОТАЦИЯ

Актуальность. Современные корпоративные логистические системы управления или системы слежения за положением состоят из беспроводных систем позиционирования. Как правило, мобильные технологии используют приемники сигналов спутников GNSS. Однако, существует общая проблема передачи данных в аналитические центры для дальнейшего использования в системах корпоративной логистики. Объектом исследования является процесс разработки решений, альтернативных для GNSS.

Цель. Повышение точности контроля доступа к местонахождению транспортного средства на основе мультиагентного анализа протоколов аутентификации.

Метод. В исследовании предложено облачную платформу слияния данных, предложено собирать и архивировать данные обо всех движущихся объектах, которые могут находиться на дороге между отделами внутри одной складской территории. Поскольку существуют операции с участием различных беспроводных базовых станций, поэтому, подвижному объекту может не хватить времени для анализа траектории движения в режиме реального времени. Поэтому обработка данных должна выполняться в облачном диспетчерском центре, и необходимо слияние потоков данных. Предложена система уравнений, которая определяет местонахождение транспортных средств на основе метода мультиагентного анализа процесса аутентификации после потери сигнала GPS. Многосекционная конфигурация рекурсивной нейронной сети и использования слияния облачных данных позволили повысить точность определения местоположения транспортного средства.

Результаты. Разработанный метод реализован программно, проведены исследования для решения проблем точности контроля местоположения транспортного средства в корпоративных логистических системах управления.

Выводы. В ходе исследования эффективность определения местоположения транспортного средства была улучшена с помощью управления доступом на основе ролей, интегрированного с мультиагентной службой аутентификации.

КЛЮЧЕВЫЕ СЛОВА: облачные технологии, слияние данных, мультиагентный анализ, протокол аутентификации, рекурсивная нейронная сеть, система управления, корпоративная логистика.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Koplow D. A. An inference about interference: A surprising application of existing international law to prohibit anti-satellite weapons / D. A. Koplow // Georgetown Law Faculty Publications And Other Works. – 2014. – Vol. 35. – P. 737–827.
2. Navigation and guidance systems // BAE Systems. – 2019. – URL: https://www.baesystems.com/en/product/navigation-and-guidance-systems.
3. Probabilistic risk assessment of environmental hazards / [B. Blyukher, L. Ivanova, E. Malakhov, V. Sitnikov] // American Society of Mechanical Engineers, Pressure Vessels and Piping Division (PVP). – 2000. – Vol. 400. – P. 69–71.
4. Canciani A. J. Absolute positioning using the Earth's magnetic anomaly field : Theses and Dissertations of PhD / A. J. Canciani. – 2016. – URL: https://scholar.afit.edu/etd/251.
5. Kramer D. DARPA looks beyond GPS for positioning, navigating, and timing / D. Kramer // Physics Today. – 2014. – Vol. 10, Issue 23. – P. 67. DOI: 10.1063/PT.3.2543.
6. Fault-tolerant controller and failure analysis of automotive electromagnetic suspension systems / [D. V. Retianza, J. V. Duivenbode, H. Huisman, E. A. Lomonova] // Ecological Vehicles and Renewable Energies (EVER 2019) : Proceedings of the IEEE 14th International Conference, Monte-Carlo, Monaco,·08–10 May 2019. – 2019. – No. 8813514. DOI: 10.1109/EVER.2019.8813514.
7. Reduction of noise similar to solar interference in computer networks based on power line communication / [D. Rumiankov, I. Zhuravska, L. Solobuto, M. Musiyenko] // Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2017) : Proceedings of the IEEE 9th International Conference, Bucharest, Romania, 21–23 September 2017. – 2017. – Vol. 1. – P. 215–221. DOI: 10.1109/IDAACS.2017.8095079.
8. Stability evaluation based on the sustainability triangle application for transfer functions above 2nd order / [H. Ukhina, V. Sytnikov, O. Streltsov et al.] // Computer Science and Application Engineering (CSAE 2019) : Proceedings of the 3rd International Conference. – October 2019. – Article No. 6. – P. 1–5. DOI: 10.1145/3331453.3361305.
9. Automated recognition and sorting of agricultural objects using multi-agent approach / [K. Ovchar, A. Borodin, I. Burlachenko, Y. Krainyk] // CEUR Workshop Proceedings. – 2019. – Vol. 2546. –P. 76–86.
10. Multi-agent monitoring system for heat loss mapping of multi-story buildings / [I. Burlachenko, I. Zhuravska, O. Tohoiev et al.] // CEUR Workshop Proceedings. – 2019. – Vol. 2516. – P. 218–225.
11. RFC 8628: OAuth 2.0 Device Authorization Grant. – 2019. – URL: https://oauth.net/2/device-flow/.
12. McAlpine K. Small, precise and affordable gyroscope for navigating without GPS / K. McAlpine // Michigan Engineering. – 23 March 2020.
13. Sivers M. Wi-Fi based indoor positioning system using inertial measurements / [M. Sivers, G. Fokin, P. Dmitriev, et al.] // Next Generation Wired/Wireless Networking : Proceedings of the International Conference, 28–30 August 2017. – 2017. – DOI: 10.1007/978-3-319-67380-6_69.
14. Indoor object tracking using Wi-Fi access points / [V. Asher, H. Thakkar, S. Tambe, P. Bhavathankar] // Inventive Communic. and Computational Technol., LNNS. – 2020. – Vol. 89. – P. 687–698. DOI: 10.1007/978-981-15-0146-3_64.
15. The monitoring system based on a multi-agent approach for moving objects positioning in wireless networks / [O. Tohoiev, I. Burlachenko, I. Zhuravska, V. Savinov] // CEUR Workshop Proceedings. – 2020. – Vol. 2608. – P. 79–90.
16. Berg K. The privacy paradox of utilizing the Internet of Things and Wi-Fi tracking in smart cities / K. Berg, T. A. M. Spil, R. Effing // IFIP Advances in Information and Communication Technology. – 2019. – Vol.558. – P. 364–381. DOI: 10.1007/978-3-030-20671-0_25.

DOI 10.15588/1607-3274-2021-4-9

УДК 004.9

ТЕХНОЛОГІЯ ВІЗУАЛЬНОЇ СИМУЛЯЦІЇ ПАСАЖИРОПОТОКІВ У СФЕРІ ГРОМАДСЬКОГО ТРАНСПОРТУ SMART CITY

Литвин В. В. – д-р техн. наук, професор, завідувач кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

Бублик М. І. – д-р екон. наук, професор, професор кафедри менеджменту і міжнародного підприємництва, Національний університет «Львівська політехніка», Львів, Україна.

Висоцька В. А. – канд. техн. наук, доцент, доцент кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

Мацелюх Ю. Р. – магістр кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

АНОТАЦІЯ

Актуальність. На сьогоднішній день проблема візуальної симуляції пасажиропотоків у сфері громадського транспорту є важливою при створенні інформаційних систем для розвитку сучасних Smart City. В умовах Industry 4.0 важливим є створення технологій, засобів та інструментів для впровадження єдиної саморегульованої інтелектуальної системи обміну даними при наданні відповідних послуг перевезення пасажирів у сфері громадського транспорту. Наприклад, наочно відображати проблемні ділянки на маршрутах в Smart City, формувати множити та індентифікувати основні зупинки в часових зрізах з найбільшим пасажирообміном, формувати пропозиції щодо потреби модернізації маршрутів з врахуванням збільшенням завантаженності громадського транспорту на певних ділянках Smart City та отримувати результати прогнозування пасажиропотоків при внесенні відповідних змін на основі методів машинннного навчання.

Метою дослідження є розроблення технології візуальної симуляції пасажиропотоків у сфері громадського транспорту для підвищення якості надання послуг пасажирських перевезень в Smart City.

Метод. Удосконалено імітаційну модель для обчислення пасажиропотоку при зміні кількості рухомого складу на маршруті, де на відміну від відомих, додано прогнозування на основі розробленої нейронної мережі. Удосконалено механізм візуальної симуляції пасажиропотоків з використанням карт GoogleMaps та динамічне переміщення по них з контролем швидкості відображення симуляції. Запропоновано нейронну мережу з повнозв’язними шарами з використанням оптимізаційного алгоритму з адаптивним рівнем навчання Adam для прогнозування пасажиропотоку між зупинками за визначений період доби. Визначені критерії деталізації даних про пасажиропотоки на міських маршрутах, включаючи загальні показники відношення пасажирообміну на певній зупинці до поточного періоду доби. При проектуванні інтелектуальної системи зазнав подальшого розвитку підхід зміни місткості рухомого складу громадського транспорту у Smart City, де на відміну від відомих, зміна місткості обмежена наявними транспортними засобами. Зазнав подальшого розвитку метод розрахунку комплексу показників пасажиропотоків на зупинках та перегонах з врахованням додаткових локальних графіків та специфіки роботи транспорту на окремих індивідуальних маршрутах.

Результати. Розроблено інтелектуальну систему візуального моделювання пасажирських перевезень на основі нейронної мережі та машинного навчання, яка дозволяє оптимізувати роботу пасажирських перевезень громадським транспортом в Smart City. Це подання даних надає можливість оцінити рентабельність додавання нового транспортного засобу на маршрут чи необхідно скоригувати розклад інших транспортних засобів для кращого покриття завантажених ділянок у час пік. Для роботи програмного засобу використовується загальновідомий стандарт подання даних про громадський транспорт – GTFS. Це дозволяє адаптувати розроблений програмний продукт універсальним, а не специфічним для конкретного міста чи країни. Проведено порівняння отриманих результатів на множині даних тролейбусних маршрутів (біля 2000 записів, з зібраних на основі експериметнальних маркетингових досліджень) в місті Львів (Україна) для формування прогнозу зміни пасажиропотоків на певних ділянках в різні проміжки часу.

Висновки. Виявлено, що спрогнозовані нейронною мережею пасажиропотоки у порівнянні з фактичними, призводять до їх зростання у середньому на 28% на критичних перегонах в годину пік. Ці отриманні результати дозволяють обґрунтувати доцільність додавання розкладу нового транспортного засобу для кращого покриття завантажених ділянок у час пік. Порівняння зміни пасажиропотоків, розподілених за перегонами у часі доби з 19:00 до 20:00, за фактичними даними та після роботи нейронної мережі вказує на їх збільшення у середньому в 70% перегонів, які були спрогнозовані, що дасть змогу прийняти обґрунтоване рішення про запуск додаткового транспорту на відповідні маршрути.

КЛЮЧОВІ СЛОВА: візуальна симуляція пасажиропотоків, візуальне моделювання, інтелектуальна система, Smart City, GoogleMaps, нейронна мережа, прогнозування пасажиропотоку, машинне навчання, інформаційні технології, опрацювання даних.

АБРЕВІАТУРА

ІС – інформаційна система;

ІТ – інформаційна технологія;

ПЗ – програмне забезпечення.

НОМЕНКЛАТУРА

S – система візуального стимулювання пасажиропотоків;

I – множина вхідних даних;

O – множина вихідних даних;

R – основні правила опрацювання потоку вхідних даних;

U – додаткові параметри опрацювання вхідних даних;

N – нейронна мережа;

α – оператор скачування вхідних даних;

β – оператор опрацювання вхідних даних;

DOI 10.15588/1607-3274-2021-4-10

γ – оператор візуалізації пасажиропотоків;

μ – оператор опрацювання GTFS даних;

χ – оператор формування даних про маршрут;

ω – оператор формування даних про пасажиропотоки;

λ – оператор прогнозування пасажиропотоків нейронною мережею;

i_1 – множина посилань на GTFS дані;

i_2 – множина даних загальних налаштувань;

i_3 – дані про кількість пасажирів;

i_4 – множина розкладів додаткового транспортних засобів;

o_1 – схематично візуально симульовані пасажиропотоки;

o_2 – візуально симульовані пасажиропотоки на карті;

o_3 – схематично візуально симульовані прогнозовані пасажиропотоки;

r_1 – правила GTFS формату;

r_2 – правила форми Excel формату;

u_1 – номери маршруту;

u_2 – типи маршруту;

u_3 – види візуальної симуляції;

DB_{GTFS} – база GTFS даних;

DB_{SDMM} – множина вихідних даних з процесу формування даних про маршрут;

d_1^s – інформація про розклад транспорту;

d_2^s – Google карти;

d_3^s – дані про маршрут;

DB_{OPDS} – множина вихідних даних з процесу формування даних про пасажиропотоки;

d_1^o – розмір датасету для навчання;

d_2^o – сформовані пасажиропотоки;

DW_{GTFS} – результат скачування GTFS даних;
PL_{GTFS} – заповнення GTFS даними бази даних;
CG_{DWCR} – перевірка правильності скачування;
UD_{DWPB} – оновлення прогресбару скачування;
SV_{FL} – результат збереження файлів;
CG_{FLIG} – перевірка цілісності файлів;
CG_{ETDB} – перевірка існування бази GTFS даних;
CR_{NWDB} – нова хешована база GTFS даних;
CL_{ETDB} – очищення наявної бази GTFS даних;
CR_{TBD} – множина таблиць у базі даних;
FG_{TBD} – заповнення таблиць GTFS даними;
DW_{GGMP} – множина завантажених Google карт;
AR_{DTDW} – пришвидшення завантаження даних;
CR_{ELFL} – створений Excel файл для заповнення»;
CR_{ADFL} – створений файл додаткового транспортного засобу;
CG_{CDAB} – результат перевірки наявності карт;
CG_{CDIG} – результат перевірки цілісності карт;
CR_{HRSR} – створена ієрархія для зберігання;
UD_{PBDW} – оновлення прогресбару скачування карт;
SV_{MP} – результат збереження карт;
CG_{CDIG} – результат перевірки цілісності карт;
PR_{GTFS} – результат підготовки GTFS даних;
GR_{STRT} – множина даних про вибраний маршрут;
DG_{RTSD} – визначення розкладу для маршруту;
CR_{ELTM} – створений шаблон Excel файлу;
FL_{ELFL} – заповнення даними Excel файлу;
CR_{FL} – результат створення файлу;
FL_{TMFL} – заповнений шаблон файлу згідно правил;
FM_{PFSD} – розклад пасажиропотоків;
CR_{HRPT} – множина погодинних пасажиропотоків;
CR_{TSSD} – розклад зупинок транспорту;
CR_{SPSD} – створення розкладу зупинок;
FG_{SPSD} – заповнення розкладу зупинок;
CG_{CCSD} – перевірка коректності розкладу;
GR_{PTSD} – дані з розкладу для пасажиропотоків;
DG_{DSSZ} – розмір датасету;
CT_{PRFL} – обчислення пасажиропотоків;
DG_{VSTP} – визначення виду візуальної симуляції;
RG_{VSTF} – запуск візуальної симуляції;
GR_{VSDT} – множина даних для візуальної симуляції;
RG_{VSTP} – зчитування виду візуальної симуляції;
OV_{MP} – результат накладання карти;
RG_{VSMP} – запуск візуальної симуляції на карті;
TR_{NRNT} – результат навчання нейронної мережі;
SV_{NRNT} – результат збереження нейронної мережі;
CL_{NWPF} – обчислення нових пасажиропотоків;
DM_{NNSZ} – визначення розміру нейронної мережі;
CR_{NNML} – створення моделі нейронної мережі;
LG_{NRNT} – процес навчання нейронної мережі;
CG_{STNN} – результат перевірки наявності збереженої нейронної мережі;
DT_{LSNN} – результат видалення файлу минулої нейронної мережі;
WT_{NNFL} – результат запис нейронної мережі у файл;
VD_{DT} – результат перевірки правильності даних;
RG_{NRNT} – результат зчитування нейронної мережі;
RG_{ADSD} – зчитування додаткового розкладу;
FG_{NWPF} – прогнозування нових пасажиропотоків;
RN_{VSN} – результат схематичного запуску візуальної симуляції нових пасажиропотоків.

ВСТУП

Пасажиропотоки залежать не тільки від особливостей прокладення маршрутів, а й певних основних точок найбільшого пасажирообміну у Smart City [1–2]. Пасажиропотоки є найбільш важливим аспектом, від якого необхідно відштовхуватися при створенні нових транспортних маршрутів і сполучень, оновленні чи зміні існуючих у Smart City. На даний момент ця проблема з дослідження та візуалізації пасажиропотоків не є вирішеною та досить актуальною для великих Smart City [3–4]. Проблема візуальної симуляції набуває гострої актуальності в умовах глобалізації світу [5–6]. Зростає потреба в побудові симуляційних моделей реального світу [7–

DOI 10.15588/1607-3274-2021-4-10

8], зокрема для пасажиропотоків у сфері громадського транспорту [9–12]. У [1–2] запропоновано динамічну імітаційну модель розподілу пасажиропотоків на транспортних мережах із врахуванням розкладу та затримки поїздів, громадського транспорту великих та середніх міст, яку покладено в основу алгоритму візуальної симуляції пасажиропотоків. У результаті моделювання в [1, 12–15] формують статистичні показники, включаючи обсяг пасажиропотоку кожного транспортного засобу та зупинки, які анімовано симулюються в ІС. В [1] модель надає кількісний приклад для ілюстрації розробленого ПЗ. Запропонований підхід в [2] моделювання фактичної продуктивності рішення фокусується на оптимізації теоретичної цільової функції та здійснює вдосконалення цільової функції. В [16–18] побудовано зв'язок між оптимізацією математичних моделей і отриманим фактичним позитивним ефектом удосконалення цільової функції. В [2] описано ПЗ, що використовує власний механізм візуальної симуляції і дозволяє застосовувати реальну карту, а також динамічне переміщенні по ній з контролем швидкості здійснення симуляції [19–23]. З метою удосконалення організації обслуговування пасажирів на маршруті в [3] запропоновано використовувати раціональний розподіл транспортних засобів із врахування їх пасажиромісткості протягом доби перевезення. У [4] сформовано основи простої теорії руху, яка пояснює походження та призначення різних типів поїздок у міських районах, де автор вперше використав поняття частоти поїздки, місця поїздки та способу подорожі. Отже, проблема підвищення якості пасажирських перевезень безпосередньо пов'язана з пасажиропотоками у сфері громадського транспорту, а саме з їх оцінюванням та аналізуванням і, як наслідок, візуальною симуляцією [19–23]. На даний момент проблема є малодослідженою серед відомих інформаційних систем.

Метою дослідження є розроблення інформаційної технології візуальної симуляції пасажиропотоків у сфері громадського транспорту для підвищення якості надання послуг пасажирських перевезень в Smart City. Для досягнення мети були поставлені такі завдання:

– удосконалити імітаційну модель для обчислення пасажиропотоку при зміні кількості рухомого складу на маршруті;

– розробити нейронну мережу з повнозв'язними шарами з використанням оптимізаційного алгоритму з адаптивним рівнем навчання Adam для прогнозування пасажиропотоку між зупинками за певний період доби;

– розробити та описати програмне забезпечення візуальної симуляції пасажиропотоків у сфері громадського транспорту Smart City;

– здійснити аналіз результатів експериментальної апробації запропонованої технології візуальної симуляції пасажиропотоків у сфері громадського транспорту Smart City.

1 ПОСТАНОВКА ПРОБЛЕМИ

Систему візуального стимулювання пасажиропотоків S подано кортежем:

$$S = < I, O, R, U, N, \alpha, \beta, \gamma > ,$$

де $I = \{i_1, i_2, i_3, i_4\}$, $O = \{o_1, o_2, o_3\}$, $R = \{r_1, r_2\}$, $U = \{u_1, u_2, u_3\}$.

Основними процесами візуального стимулювання пасажиропотоків є «Опрацювання GTFS даних», «Формування даних про маршрут», «Формування даних про пасажиропотоки» та «Прогнозування пасажиропотоків нейронною мережею». Процес опрацювання GTFS даних опишемо суперпозицією: $DB_{GTFS} = \beta \circ \alpha$, тобто $DB_{GTFS} = \beta(\alpha(i_1, r_1))$.

Процес формування даних про маршрут опишемо суперпозицією: $DB_{SDMM} = \beta \circ \alpha$, тобто

$$DB_{SDMM} = \beta(DB_{GTFS}, r_1, \alpha(i_2, r_2, u_1, r_2)) ,$$

де $DB_{SDMM} = \{d_1^s, d_2^s, d_3^s\}$.

Процес формування даних про пасажиропотоки опишемо суперпозицією: $DB_{OPDS} = \gamma \circ \beta$, тобто

$$DB_{OPDS} = \gamma(d_3^s, u_3, \beta(i_3, d_1^s, d_2^s)) ,$$

де $DB_{OPDS} = \{o_1, o_1, d_1^o, d_2^o\}$.

Процес прогнозування пасажиропотоків нейронною мережею опишемо суперпозицією:

$$o_3 = \gamma \circ \beta \text{, тобто } o_3 = \gamma(N, \beta(i_4, d_1^o, d_2^o)) .$$

2 АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

Можливості візуалізації пасажиропотоків у сфері громадського транспорту надає компанія «A+C Україна» (рис. 1), а також наявна три місячна візуалізація даних з продажу електронного квитка у місті Житомирі (Україна) на сайті http://texty.org.ua/. Компанія «A+C Україна» не розкриває своїх методів і концентрується на окремих містах чи маршрутах, не підтримує міжнародні стандарти подання маршрутів, і надає перевагу розробленню рішень у кожній окремій ситуації. Також зосереджена на організації збору інформації, ніж на візуальну симуляцію пасажиропотоків.

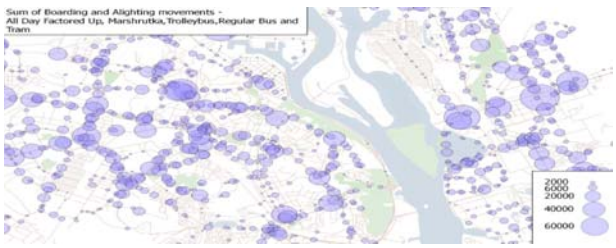


Рисунок 1 – Програма компанії «A+C Україна»

DOI 10.15588/1607-3274-2021-4-10

e-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2021. № 4
p-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2021. № 4

Візуалізації даних з електронного квитка на сайті http://texty.org.ua/ не оновлювалася з січня 2019 року (рис. 2). Зараз на сайті наведено візуалізацію пасажиропотоків тільки за період у три місяці. Варто зазначити, що дані, зібрані з електронного квитка, є вже частково спотворені, оскільки не всі пасажири купують квиток. Дана система не використовує дані про справжні пасажиропотоки, а надає тільки узагальнену схематичну картину оплачених квитків, не враховуючи поточної завантаженості транспорту на перегонах.

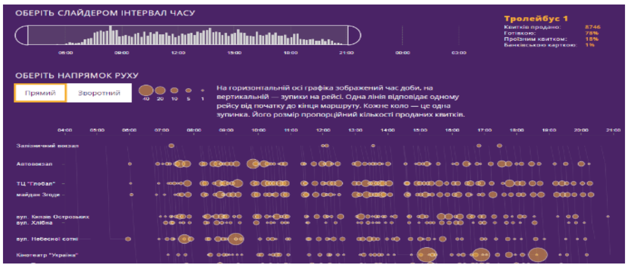


Рисунок 2 – Сайт texty.org.ua

Подібними ПЗ є PTV Visum та PTV Vissim, представлені компанією PTV Group. Один із цих продуктів – PTV Visum (рис. 3), створений як ПЗ для планування руху, при проектуванні та плануванні транспортних маршрутів з метою розширення можливостей транспортної мережі міст. PTV Vissim розроблений за сучасними вимогами і є гнучким ПЗ для моделювання руху транспорту у Smart City, але має ряд недоліків (табл. 1).

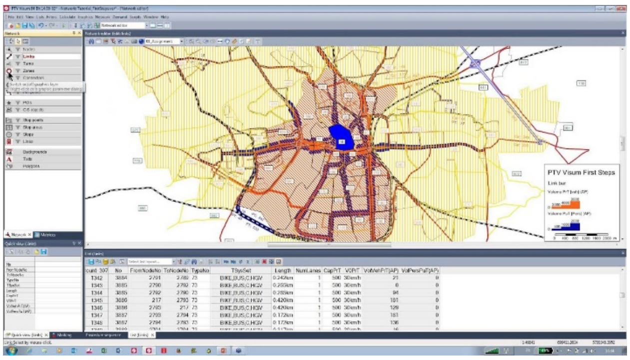


Рисунок 3 – Програмний продукт PTV Visum

Таблиця 1 – Порівняльний аналіз аналогів

Характеристики	А+С Україна	texty.org.ua	PTV Visum
Ціна	Окрема під проблему	Безкоштовноа	~2 млн. грн за 100 км
Мова	Українська	Українська	Англійська
Демо-версія	Немає	Немає	30 днів
Тип продукту	Невідомо	Сайт	Настільна програма
Наявність документації	Немає	Немає	Платні книжки
Кількість міст	11	1	2500
Актуальність інформації	Проводиться за запитом	Останні дані січень 2019р.	Постійна
Тип візуалізації	3D	2D	2D та 3D
Використання нейронних мереж	Невідомо	Немає	Немає
Виконавець роботи	Компанія	Невідомо	Користувач

DOI 10.15588/1607-3274-2021-4-10

Відповідно до аналізу одні ПЗ дуже дорогими та не адаптованими для пересічного громадянина окремої держави, але надають великі можливості для опрацювання даних. Паралельно є незрозумілими користувачу і вимагають великих зусиль на розуміння документації. Інші продукти є тільки компаніями, які надають послуги з дослідження пасажиропотоків, тобто користувачу немає необхідності щось робити, але заплатити він зобов'язаний. Або в третьому випадку, абсолютно безкоштовні, але які не можливо використовувати, бо вони розроблені суто для одного міста і для одного типу даних. І лише надають інформацію для загального розуміння в себе на сайті, причому ця інформація має актуальність на січень 2019 року. Крім того. всі розглянуті аналоги є суто комерційними проектами, тому невідома ні їх архітектура, ні які методи були використанні при їх розробленні. Враховуючи поточні потреби в дослідженні та візуальній симуляції пасажиропотоків у Smart City з численним ростом населення та обсягів транспорту випливає висновок, що цей напрям потребує нових підходів до розв'язання, що і обумовлює актуальність постановки задачі роботи.

3 МАТЕРІАЛИ ТА МЕТОДИ

Удосконалимо імітаційну модель системи візуального стимулювання пасажиропотоків V для обчислення пасажиропотоку при зміні кількості рухомого складу на маршруті:

$$V = < S, DB_{GTFS}, DB_{SDMM}, DB_{OPDS}, DB_{FPTN}, \mu, \chi, \omega, \lambda >$$

або $O = \gamma \circ \lambda \circ \omega \circ \beta \circ \chi \circ \mu \circ \alpha$.

Модель модуля опрацювання GTFS даних DB_{GTFS}:

$$DB_{GTFS} = < DW_{GTFS}, PL_{GTFS}, i_1, r_1, DW_{GTFS}, \alpha, \beta, \mu >,$$
$$DB_{GTFS} = \beta \circ \mu \circ \alpha,\ DB_{GTFS} = \beta(\mu(DW_{GTFS}, r_1, \alpha(i_1, r_1))).$$

Для опрацювання GTFS-даних використовується унікальний алгоритм авторської розробки (рис. 4) для перетворення наданих даних про зупинки та основні точки на маршруті у GTFS-форматі у коректне зображення на реальній карті з поділом маршруту на дві частини: прямий та зворотній напрями.

Модель модуля скачування GTFS даних DW_{GTFS}:

$$DW_{GTFS} = < CG_{DWCR}, UD_{DWPB}, SV_{FL}, CG_{FLIG}, \alpha, \mu_1 >,$$
$$DW_{GTFS} = \mu_1 \circ \alpha,\ DW_{GTFS} = \mu_1(\alpha(i_1, r_1)).$$

Модель заповнення GTFS даними БД PL_{GTFS}:

$$PL_{GTFS} = < DW_{GTFS}, CG_{ETDB}, CR_{NWDB}, CL_{ETDB}, CR_{TBD}, FG_{TBD}, \beta, \mu_2 >,$$
$$PL_{GTFS} = \beta \circ \mu_2,$$
$$PL_{GTFS} = \beta(CG_{ETDB}, CR_{NWDB}, CL_{ETDB}, CR_{TBD}, FG_{TBD}, \mu_2(DW_{GTFS}, r_1,)).$$

Модель формування даних про маршрут DB_{SDMM}:

$$DB_{SDMM} = < DB_{GTFS}, DW_{GGMP}, AR_{DTDW}, CR_{ELFL}, CR_{ADFL}, r_1, \alpha, \beta, \chi >,\ DB_{SDMM} = \beta \circ \chi \circ \alpha,$$

$$DB_{SDMM} = \beta(DW_{GGMP}, AR_{DTDW}, CR_{ELFL}, CR_{ADFL}, r_1, \chi(DW_{GGMP}, i_2, r_1, r_2, u_1, u_2, \alpha(i_2,r_2,u_1,u_2))).$$

Модель модуля скачування Google карт DW_{GGM}:

$$DW_{GGM} = < DB_{GTFS}, CG_{CDAB}, CG_{CDIG}, CR_{HRSR}, UD_{PBDW}, SV_{MP}, CG_{CDIG}, i_2, u_1, u_2, d_2^s, \alpha, \chi_1 >,$$

$$DW_{GGM} = \chi_1\circ\alpha,\ DW_{GGM} \subseteq d_2^s,$$

$$DW_{GGM} = \chi_1(DB_{GTFS}, CG_{CDAB}, CG_{CDIG}, CR_{HRSR}, UD_{PBDW}, SV_{MP}, CG_{CDIG}, \alpha(i_2, u_1,u_2)).$$

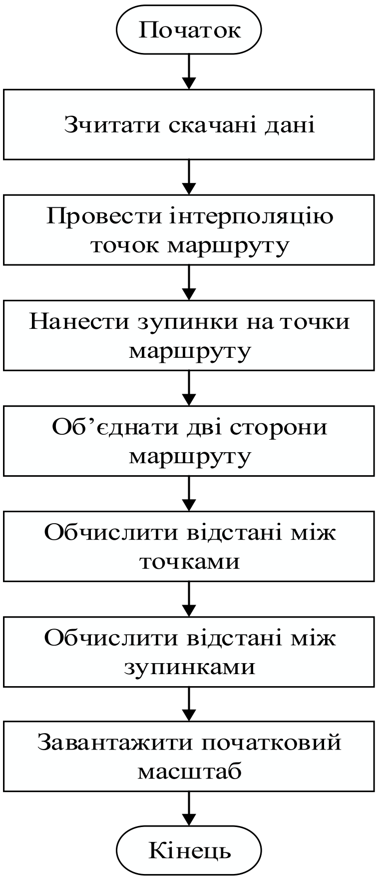


Рисунок 4 – Алгоритм опрацювання GTFS-даних

Модель пришвидшення завантаження карт AR_{DTDW}:

$$AR_{DTDW} = < DB_{GTFS}, PR_{GTFS}, GR_{STRT}, DG_{RTSD}, i_2, u_1, u_2, d_1^s, d_3^s, \alpha, \chi_2 >,$$

$$AR_{DTDW} = \chi_2\circ\alpha,\ AR_{DTDW} \subseteq \{ d_1^s, d_3^s \},$$

$$AR_{DTDW} = \chi_2(DB_{GTFS}, PR_{GTFS}, GR_{STRT}, DG_{RTSD}, \alpha(i_2,u_1,u_2)).$$

Модель модуля створення Excel файлу для заповнення CR_{ELFL} подамо кортежем:

$$CR_{ELFL} = < CR_{ELTM}, FL_{ELFL}, AR_{DTDW}, r_2, \chi_3, \alpha>,$$
$$CR_{ELFL} = \chi_3\circ\alpha,$$
$$CR_{ELFL} = \chi_3(CR_{ELTM}, FL_{ELFL}, \alpha(r_2, AR_{DTDW})).$$

Модель модуля створення файлу додаткового транспортного засобу CR_{ADFL} подамо кортежем:

$$CR_{ADFL} = < AR_{DTDW}, CR_{FL}, FL_{TMFL}, r_1, \chi_4, \alpha, \beta >,$$

$$CR_{ADFL} = \beta\circ\chi_4\circ\alpha,$$
$$CR_{ADFL} = \beta(FL_{TMFL}, \chi_4(CR_{FL}, \alpha(r_1, AR_{DTDW}))).$$

До загально відомих використаних алгоритмів в розробленій системі належать алгоритм для зчитування та парсингу CSV-файлів, а також алгоритм рисування графічних примітивів.

Модель модуля формування даних про пасажиропотоки DB_{OPDS} подамо кортежем:

$$DB_{OPDS} = < DB_{SDMM}, CR_{HRPT}, CR_{TSSD}, i_3, d_1^s, d_2^s, d_3^s, i_3, u_3, o_1, o_2, d_1^o, d_2^o, \beta, \gamma, \omega >,$$

$$DB_{OPDS} = \gamma\circ \omega\circ\beta,\ DB_{OPDS} \subseteq \{ o_1, o_2, d_1^o, d_2^o \},$$

$$DB_{OPDS} = \gamma (CR_{HRPT}, CR_{TSSD}, d_3^s, u_3, \omega(DB_{SDMM}, \beta (i_3, d_1^s, d_2^s))).$$

Модель модуля формування розкладу з пасажиропотоками FM_{PFSD} подамо кортежем:

$$FM_{PFSD} = < DB_{SDMM}, CR_{SPSD}, FG_{SPSD}, CG_{CCSD}, i_3, d_1^s, d_2^s, \beta, \omega_1 >,\ FM_{PFSD} = \omega_1\circ\beta,$$

$$FM_{PFSD}=\omega_1(DB_{SDMM},CR_{SPSD},FG_{SPSD},CG_{CCSD},\beta(i_3, d_1^s, d_2^s))$$

Модель модуля створення погодинних пасажиропотоків CR_{HRPT} подамо кортежем

$$CR_{HRPT} = < FM_{PFSD}, GR_{PTSD}, DG_{DSSZ}, CT_{PRFL}, DG_{VSTP}, RG_{VSTF}, u_3, o_1, d_1^o, d_2^o, \omega_2 >,$$

$$CR_{HRPT} = \omega_2\circ\beta,\ CR_{HRPT} = \subseteq \{ o_1, d_1^o, d_2^o \},$$

$$CR_{HRPT} = \omega_2(GR_{PTSD}, DG_{DSSZ}, CT_{PRFL}, DG_{VSTP}, RG_{VSTF}, u_3, \beta(FM_{PFSD},)).$$

Для вирішення поставлених завдань спеціально розроблено унікальний алгоритм візуальної симуляції, який надає можливість наочно відобразити рух транспорту та існуючі пасажиропотоки як на реальній карті, так і схематично (рис. 5).

Модель модуля створення розкладу зупинок транспорту CR_{TSSD} подамо кортежем:

$$CR_{TSSD} = < GR_{VSDT}, RG_{VSTP}, OV_{MP}, RG_{VSMP}, u_3, d_2^s, o_2, \beta, \gamma, \omega_3 >$$

$$CR_{TSSD} = \gamma\circ\omega_3\circ\beta,\ DB_{OPDS} \subseteq o_2,$$

$$CR_{TSSD} = \gamma(GR_{VSDT}, RG_{VSTP}, OV_{MP}, \omega_3(u_3, \beta(d_2^s, FM_{PFSD}))).$$

У розроблюваній ІС використані різноманітні методи, які виконують функції, необхідні для забезпечення його коректної роботи (табл. 2).

Модель модуля прогнозування пасажиропотоків нейронною мережею DB_{FPTN} подамо кортежем:

$$DB_{FPTN} = < TR_{NRNT}, SV_{NRNT}, CL_{NWPF}, N, i_4,\gamma, d_1^o, d_2^o, \beta, \lambda >,\ DB_{FPTN} = \gamma\circ\lambda\circ\beta,\ DB_{FPTN} \subseteq o_3,$$

$$DB_{FPTN} = \lambda (SV_{NRNT}, CL_{NWPF}, \gamma (N, TR_{NRNT}, \beta(N, i_4, d_1^o, d_2^o))).$$

DOI 10.15588/1607-3274-2021-4-10

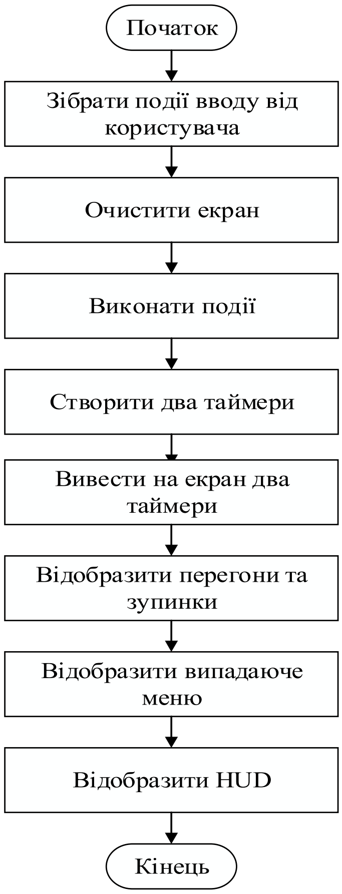


Рисунок 5 – Алгоритм створення схематичного режиму

Таблиця 2 – Використані методи та інструменти

Назва	Призначення
Інтерполяція	Створення додаткових точок для відображення маршруту
Найменша відстань від точки до прямої	Проекція зупинки на маршрут
Найменша відстань між двома точками	Загальна довжина маршруту, та відстань між зупинками
Світова геодезична система WGS 84	Передача GPS координат
Проекція Меркатора	Перетворення координат GPS у світові координати
Координати пікселя на картах Google	Перетворення світових координат на координати пікселя
Нейронна мережа	Прогнозування зміни пасажиропотоків на маршруті
Метод Karney (2013)	Відстань між двома точками заданих GPS координатами

Модель навчання нейронної мережі TR_{NRNT}:

$$TR_{NRNT} = < DM_{NNSZ}, CR_{NNML}, LG_{NRNT}, N, d_1^o, d_2^o, \beta, \lambda_1 >,$$

$$TR_{NRNT} = \lambda_1 \circ \beta,\ TR_{NRNT} = \lambda_1 (N, \beta(d_1^o, d_2^o))).$$

Модель збереження нейронної мережі SV_{NRN}:

$$SV_{NRNT} = < TR_{NRNT}, CG_{STN}, DT_{LSNN}, WT_{NNF}, N, d_1^o, \beta, \lambda_2 >,\ SV_{NRNT} = \lambda_2 \circ \beta,$$

$$SV_{NRNT} = \lambda_2 (N, CG_{STN}, DT_{LSNN}, WT_{NNF}, \beta(TR_{NRNT}, d_1^o))).$$

Модель обчислення нових пасажиропотоків CL_{NWPF} подамо кортежем:

$$CL_{NWPF} = < VD_{DT}, RG_{NRNT}, RG_{ADSD}, FG_{NWPF}, RN_{VSN}, SV_{NRN}, N, i_4, \gamma, \beta, \lambda_3 >$$

$$CL_{NWPF} = \gamma \circ \lambda_3 \circ \beta,\ CL_{NWPF} \subseteq o_3,$$

$$CL_{NWPF} = \gamma(SV_{NRNT}, \lambda_3(N, \beta(i_4))).$$

Для виконання функції прогнозування зміни пасажиропотоків використано можливості нейронної мережі з повнозв'язними шарами, яка побудована на основі оптимізаційного алгоритму з адаптивним рівнем навчання Adam (рис. 6).

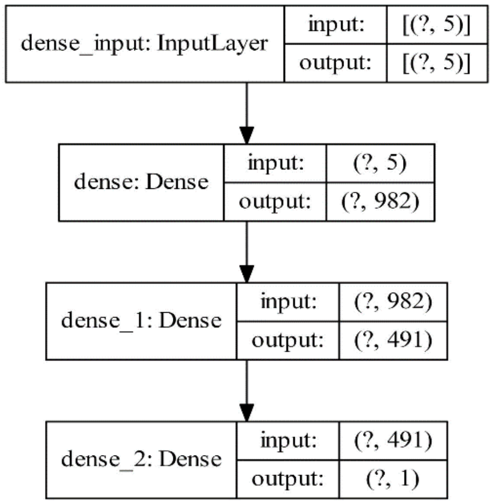


Рисунок 6 – Приклад структури нейронної мережі

Розмір нейронної мережі визначається динамічно і залежить від розміру датасету, на якому відбуватиметься навчання.

4 ЕКСПЕРИМЕНТИ

Мінімальні вимоги для запуску ІС – персональний комп'ютер або ноутбук на базі процесора Intel Core i5-6500 або еквівалентного йому, інших виробників. При менших технічних характеристиках ПЗ не тестувався (таблиця 3).

Таблиця 3 – Продуктивність ПЗ

Час від запуску	Максимальна кількість кадрів
0 хв	318
5 хв	306
10 хв	312
15 хв	301
20 хв	305
25 хв	304
30 хв	309

Однак, з програмної сторони можливе зменшення частоти кадрів у режимі візуальної симуляції на слабших процесорах, а також можливе збільшення часу на проведення навчання нейронної мережі. Всі інші функцій будуть працювати коректно. Продуктивність ПЗ тримається майже на однаковому рівні протягом довгої сесії візуальної симуляції у режимі «на карті», що обґрунтовує твердження, що дана ІС оптимізована і немає проблем з втратою продуктивності чи засмічуванням пам'яті.

DOI 10.15588/1607-3274-2021-4-10

Вхідні дані для ІС подаються кількома способами. Перш за все, дані про транспортні маршрути скачуються з відповідних серверів міста та подаються у міжнародному стандарті GTFS. Дані ж про пасажиропотоки подаються у спеціальному шаблоні Excel формату (рис. 7). Дані про новий транспортний засіб подають теж у стандарті GTFS, але не вимагають повного набору файлів, тобто використовується власна полегшена версія цього стандарту (рис. 8). Дані про маршрути передаються у форматі архіву zip, у якому містяться текстові документи txt.

Дані про пасажиропотоки подаються у файлі з розширенням xls. Дані про нову транспортну одиницю, хоч і підтримують стандарт GTFS, але не потребують розміщення у архіві і використовують тільки текстовий файл txt.У ІС база даних призначена для зберігання даних у форматі GTFS (рис. 9).

	A	B	C	D	E	F	G	H
1	Зупинка	Час прибуття	Час відправлення	Увійшло	Вийшло		Конструктивна місткість транспортного засобу	Максимальна місткість транспортного засобу
2	Університет (78)	07:17:50	07:18:10	39	0		105	116
3	Собор Святого Юра (771)	07:20:50	07:21:10	3	0			
4	Кропивницького (46)	07:23:50	07:24:10	11	4			
5	Привокзальний ринок (110)	07:25:50	07:26:10	10	8			
6	Скриня (113)	07:27:50	07:28:10	15	11			
7	Кульпарківська (95)	07:29:50	07:30:10	9	9			
8	Народна (96)	07:30:50	07:31:10	8	14			

Рисунок 7 – Excel файл для заповнення пасажирообміном

stop_times2: Блокнот

Файл Редагування Формат Вигляд Довідка

```
trip_id,arrival_time,departure_time,stop_id,stop_sequence,shape_dist_traveled,timepoint
12420_0_1,16:25:00,16:25:00,4795,2,,1
12420_0_1,16:26:00,16:26:00,4796,3,,1
12420_0_1,16:27:00,16:27:00,4797,4,,1
```

Рисунок 8 – Файл з додатковим транспортом

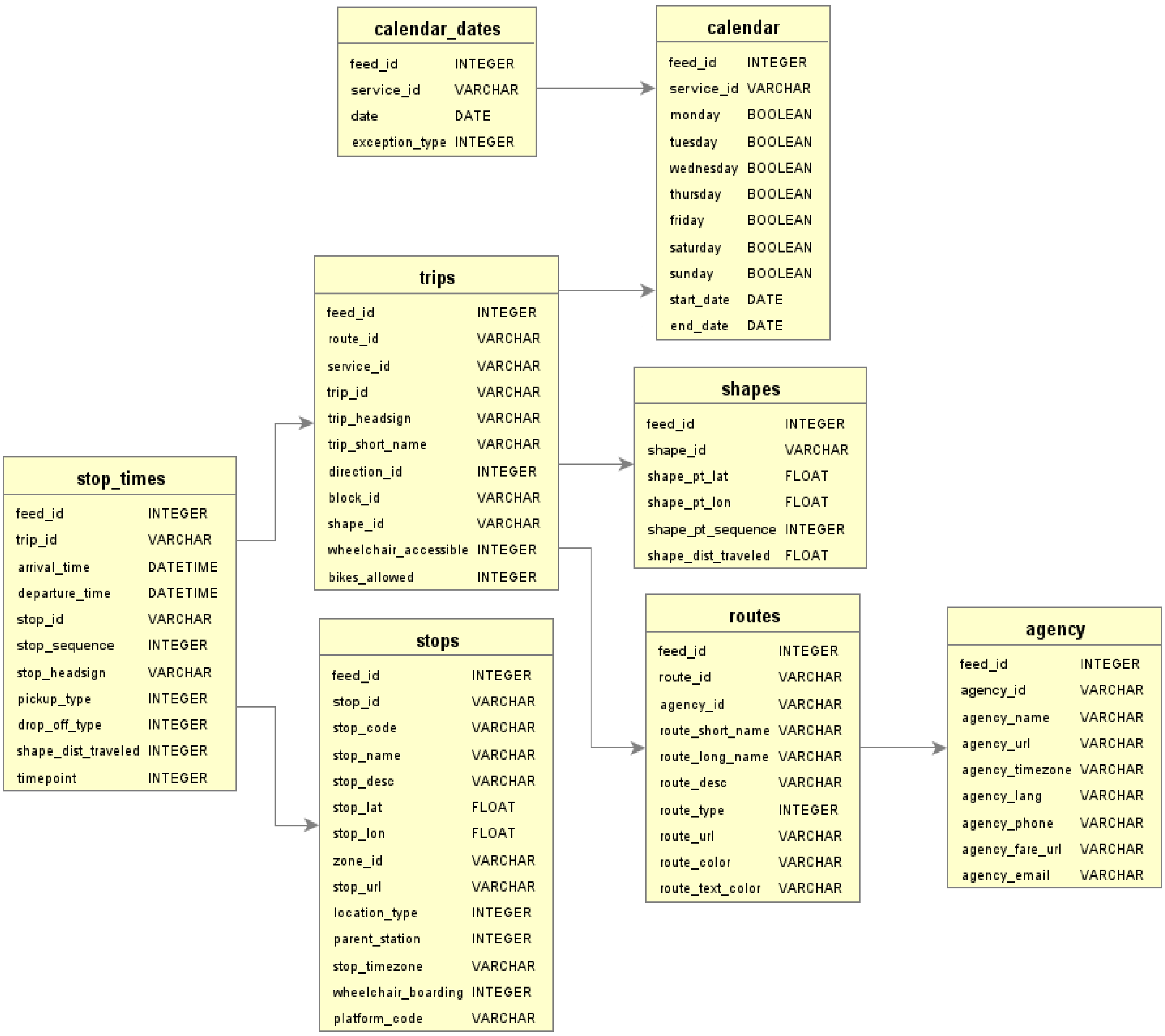


Рисунок 9 – Схема бази даних для GTFS даних

DOI 10.15588/1607-3274-2021-4-10

5 РЕЗУЛЬТАТИ

Для належного функціонування ПЗ необхідно мати встановлене наступне програмне забезпечення, а саме: операційну система Windows, інтерпретатор Python та необхідний набір модулів до нього. Розроблений програмний засіб написаний на Python. Для взаємозв’язку з базою даних використовуються SQL запити. Програмний продукт формує вихідні дані у двох форматах. Першим з яких є Excel файл, у якому створений шаблон для заповнення пасажирообміну на зупинках (рис. 7). Іншим – є наочне візуальне представлення, яке подається у вигляді візуальної симуляції у режимі «на карті» (рис. 10) або схематично (рис. 11). ІС надає можливість прогнозувати зміни пасажиропотоків. Це відбувається за допомогою спеціально розробленої нейронної мережі. Для отримання прогнозу за допомогою нейронної мережі потрібно надати розклад нового транспортного засобу, який планується для додавання на поточний маршрут. Зміна пасажиропотоків відбудеться на перегонах і може залежати від часу доби та завантаженості поточного маршруту. Спрогнозовані зміни візуалізуються у схематичному режимі роботи візуальної симуляції (рис. 11). До функціональних обмежень можна віднести можливість зміни тільки одного маршруту за один раз. Після успішного заповнення файлу Excel даними про кількість пасажирів на зупинках, можна обирати будь який режим роботи, чи схематичний чи «на карті». У схематичному режимі зупинки позначені одна за одною відповідно до їхнього розташування на маршруті у прямому та зворотному напрямках. Тут підтримується прокручування вверх-вниз мишкою для перегляду всіх зупинок, навігація стрілками, щоб вибирати наступні чи попередні проміжки часу інтервалом в одну годину. Інтерфейс даного режиму використовує різні кольори для позначення різних показників, станів та явищ. Це зумовлено потребою зробити інтерфейс програми максимально інтуїтивно зрозумілим (рис. 12). При використанні клавіші D можна потрапити у перегляд загальної інформації за весь день, а не погодинно. Кожна зупинка підписана назвою та номером, який використовується на справжніх зупинках у місті на відповідних дорожніх знаках. Під цими даними зображено кількість пасажирів, що зайшли і вийшли на зупинці за конкретний період, а ще нижче – суму всього пасажирообміну на зупинці. Розмір і сегменти кола визначаються динамічно. Чим більший радіус кола – тим більше людей на цій зупинці зайшло і вийшло (відносно максимальної кількості в кожному годинному інтервалі протягом дня), сектори позначені різними кольорами для відображення співвідношення тих хто зайшов і тих хто вийшов на зупинці. Між зупинками пасажиропотік позначається широкою лінією, розмір якої змінюється динамічно. Цифра на ній позначає скільки людей було перевезено протягом цієї години на даному перегоні між зупинками.

Колір верхнього підкреслення залежить від коефіцієнту наповненості транспортного засобу відносно його конструктивної місткості і може відображатися трьома кольорами (рис. 13–14):

– зелений (норма): менше 0.9;

– оранжевий (допустиме перевищення норми): 0.95 – 0.99;

– червоний (критичне перевищення) 1.00 і більше.

Колір нижнього підкреслення у свою чергу залежить від коефіцієнту наповненості транспортного засобу відносно його реальної місткості і теж може відображатися трьома кольорами (рис. 13):

– зелений (норма): коефіцієнт конструктивної місткості менший 1.00;

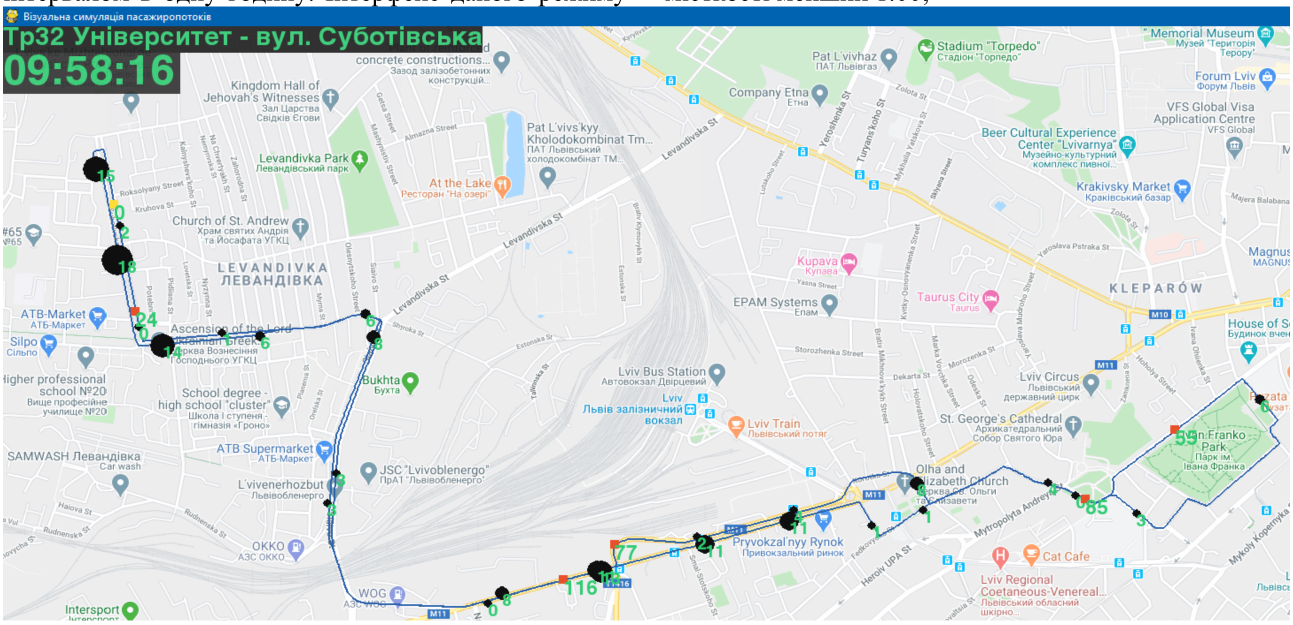


Рисунок 10 – Візуальна симуляції у режимі «на карті»

DOI 10.15588/1607-3274-2021-4-10

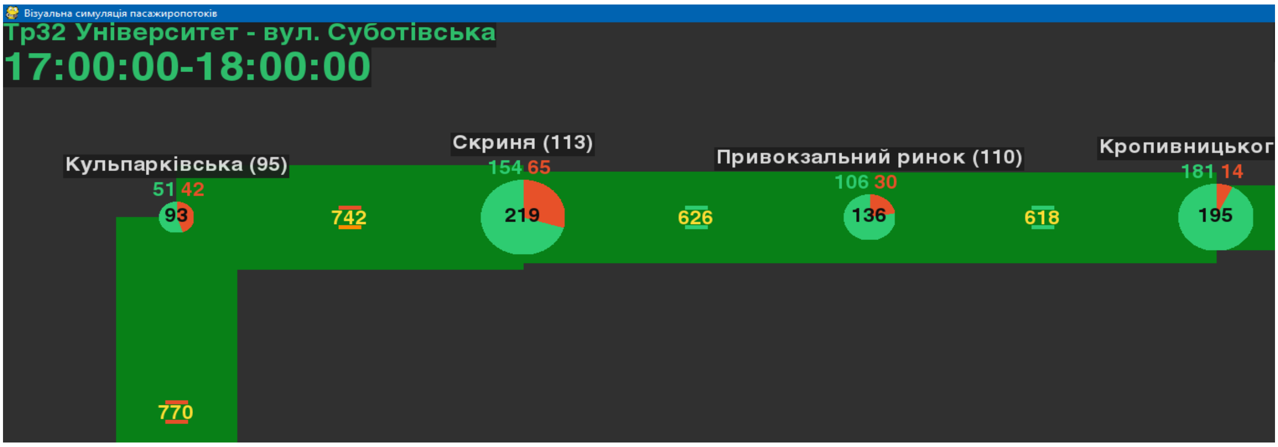


Рисунок 11 – Візуальна симуляції після прогнозування

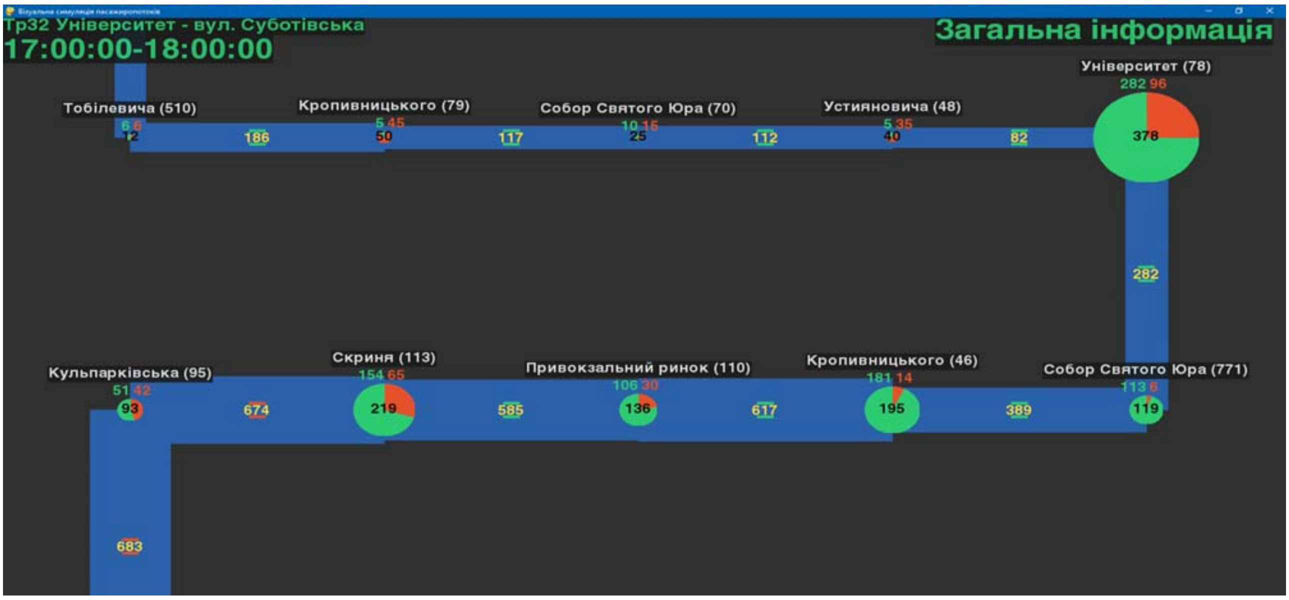


Рисунок 12 – Схематичний режим

– оранжевий (допустиме перевищення норми): коефіцієнт конструктивної місткості більший 1.00, але коефіцієнт реальної місткості менший 0.95;

– червоний (критичне перевищення): інші.

У цьому режимі також можна натиснути на зупинку і отримати детальнішу інформацію (рис. 14).

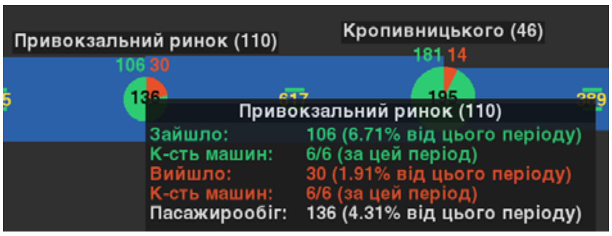


Рисунок 13 – Вікно даних про зупинку

Перегони між зупинками теж підтримують функцію додаткового вікна з даними, в якому наведено максимальну пропускну здатність перегону за кількома параметрами, де всі перевантаження зображуються наочно зміною кольорів (рис. 14).

Загальна інформація про маршрут виводиться при натисканні на кнопку «Загальна інформація», яка розміщена у правому верхньому куті (рис. 15).

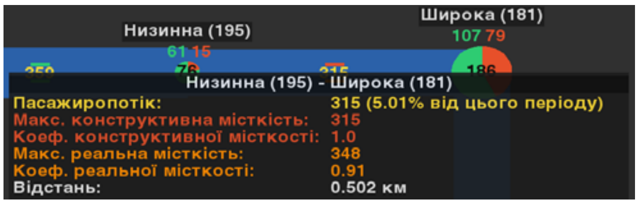


Рисунок 14 – Вікно даних про перегін між зупинками

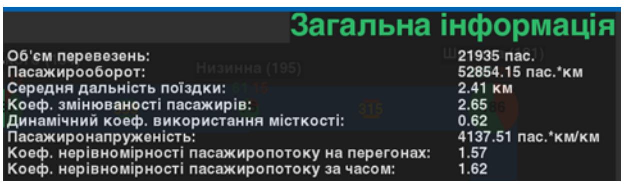


Рисунок 15 – Вікно загальної інформації

Наступним доступним режимом є режим візуальної симуляції «на карті» (рис. 16–17), де

DOI 10.15588/1607-3274-2021-4-10

відповідно до отриманих даних про час руху транспорту та пасажирів відбувається точна симуляція руху протягом дня від зупинки до зупинки разом із підбиранням і висаджуванням пасажирів на них. Даний режим найбільше підходить для наочного відображення місця де реально на карті знаходять ймовірно перевантажені пересадкові вузли чи певні значущі об’єкти міста, які притягують до себе людей.

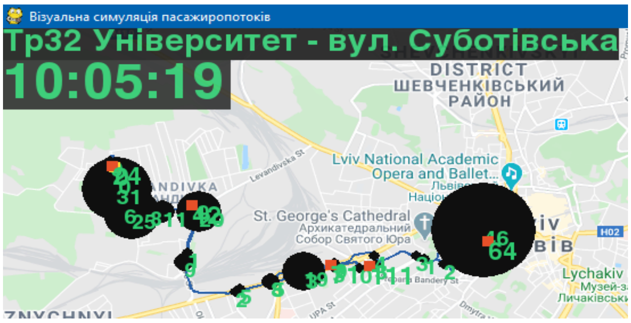


Рисунок 16 – Найменший масштаб

Оскільки на екрані зображується точний час доби, то можна побачити скільки транспорту знаходиться у даний час на маршруті, відслідкувати, що з ними відбувається, встановити перебування транспортного засобу на обідній перерві, а також визначити чи транспорт засіб рухався занадто повільно, ніж було заплановано розкладом руху. У цьому режимі також підтримується збільшення та зменшення масштабу (рис. 17–18), уповільнення швидкості симуляції у 2 та 3 рази від звичайної. За допомогою пробілу можна призупини симуляцію і дослідити даний момент часу.

Червоним кольором позначається активний транспортний засіб, що знаходиться на маршруті, жовтим – якщо він на обідній перерві. На кожній зупинці позначається кількість пасажирів, що чекають на транспорт, і від цього динамічно залежить розмір кола, котре позначає зупинку (рис. 19–20). Чим більша кількість пасажирів, тим більший розмір кола. Розмір кола змінюється динамічно між приїздами транспорту, також при посадці і висадці. Переміщення по карті у всі сторони відбувається стрілками на клавіатурі, а зміна масштабу – коліщатком мишки. Стандартна швидкість анімації розрахована так, що за 1 секунду реального часу симулюється 1 хвилина віртуального. Дану швидкість можна зменшити в 2 і в 3 рази клавішею PageDown, а збільшити – PageUp.

Для запуску можливості прогнозуванням зміни пасажиропотоків за допомогою додаткового транспортного засобу використовується окремий блок у правій частині головного меню. ІС створює файл за шаблоном подання інформації у GTFS форматі, який повинен бути заповнений користувачем інформацією про новий розклад руху додаткового транспортного засобу. Далі відбувається перерахунок даних про пасажиропотоки, який виводиться у схематичному режимі (рис. 21).

У даному режимі зміною кольорів перегонів між зупинками позначається результат прогнозування зміни пасажиропотоків. Синім кольором показується відсутність змін, або ділянка, яка не була охоплена роботою додаткового транспортного засобу. Зеленим – позначається те, що пасажиропотік виріс, а червоним – його зменшення. Це подання даних надає можливість оцінити чи буде рентабельним додавання даного транспортного засобу, чи можливо потрібно скоригувати його розклад для кращого покриття завантажених ділянок у час пік.

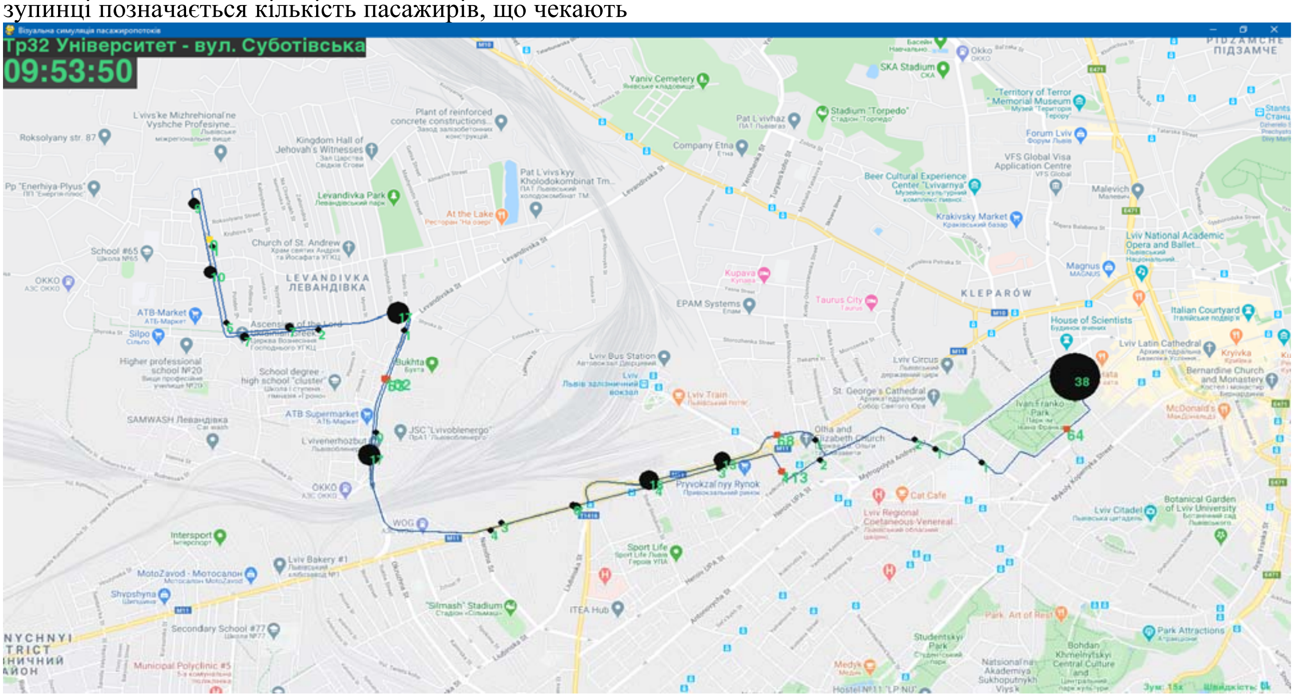


Рисунок 19 – Режим «на карті»

DOI 10.15588/1607-3274-2021-4-10

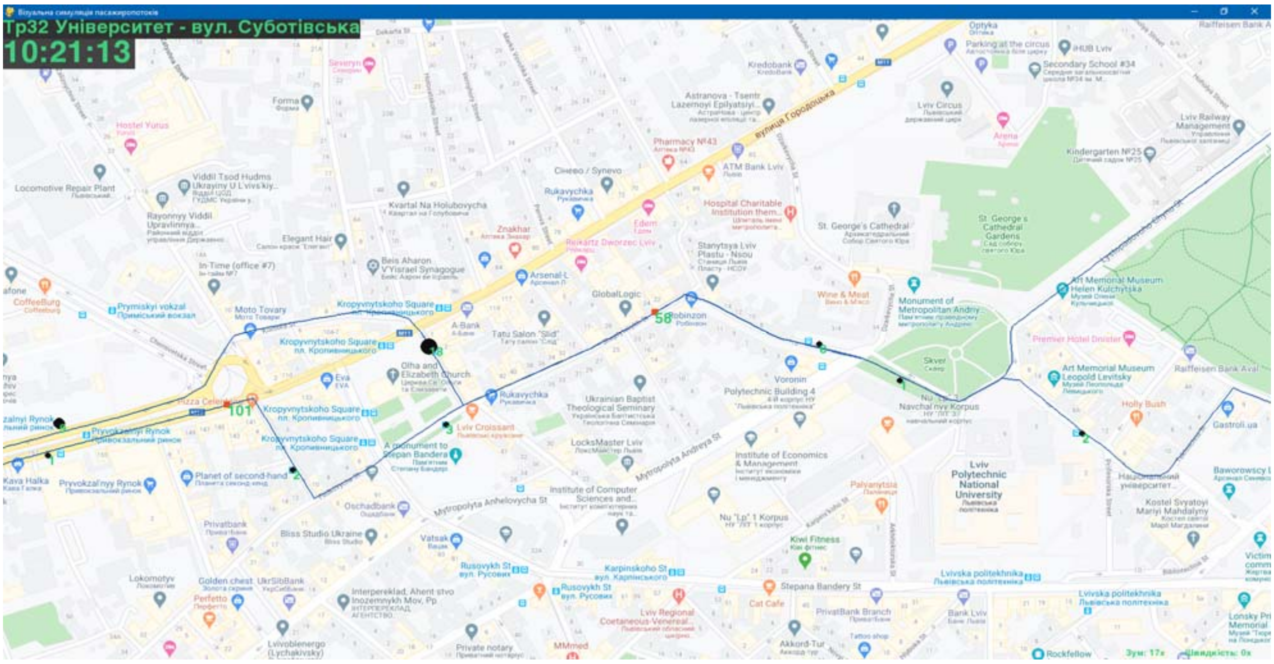


Рисунок 20 – Найбільший масштаб

Рисунок 21 – Змінені пасажиропотоки

6 ОБГОВОРЕННЯ

Зростання пасажиропотоків наведено на рис. 22, де верхньою оранжевою лінією позначено спрогнозовані нейронною мережею. З рис. 22 видно, що з 6-го до 13-ого перегону відбувається зростання пасажиропотоків в середньому на 28%, а на перегонах з 1 по 5 змін майже не відбувається.

У табл. 4 проведено порівняння зміни пасажиропотоків, розподілених за перегонами у часовому проміжку з 19:00 до 20:00, за фактичними даними та після прогнозування нейронною мережею.

З 13 перегонів, значні зміни є тільки на 9 перегонах, інші 4 залишилися без змін (або з мінімальними змінами). Це обгрунтовує прийняти рішення про запуск додаткового транспорту.

DOI 10.15588/1607-3274-2021-4-10

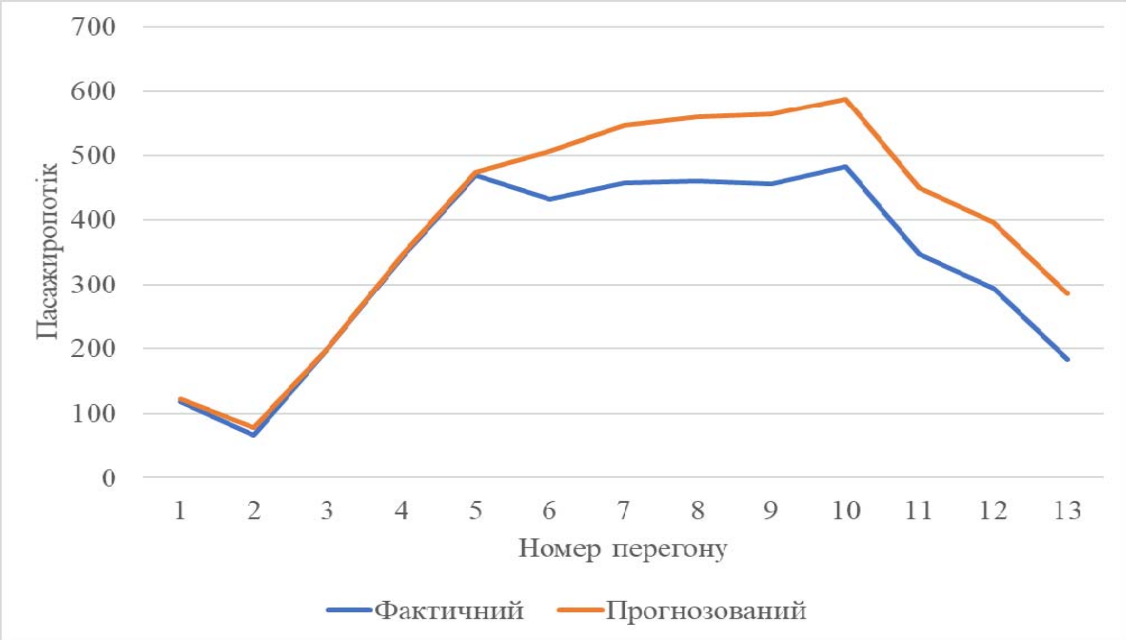


Рисунок 22 – Графік навантаження перегонів пасажиропотоками

Таблиця 4 – Аналіз зміни пасажиропотоків

Номер перегону	Фактичний	Прогнозований	Абсолютна зміна	Відносна зміна
1	118	123	5	4.24%
2	67	79	12	17.91%
3	201	201	0	0.00%
4	344	346	2	0.58%
5	469	473	4	0.85%
6	432	506	74	17.13%
7	457	546	89	19.47%
8	460	559	99	21.52%
9	456	564	108	23.68%
10	483	587	104	21.53%
11	348	450	102	29.31%
12	294	397	103	35.03%
13	183	286	103	56.28%

ВИСНОВКИ

Проблема візуальної симуляції пасажиропотоків у сфері громадського транспорту, що досліджується у роботі, є актуальною для розвитку сучасних Smart City. З метою підвищення якості надання послуг пасажирських перевезень у межах Smart City створено ІС, яка дає змогу візуально симулювати пасажиропотоки за реальних умов та спрогнозувати їх зміни за допомогою роботи нейронної мережі при додаванні розкладу додаткового транспортного засобу. Для роботи програмного засобу використовується загальновідомий стандарт подання даних про громадський транспорт – GTFS. Це дозволяє зробити даний ПЗ універсальним, а не специфічним для конкретного міста чи країни.

Запропоновано та розроблено власні алгоритми для поділу маршрути на пряму і зворотну сторону, розміщення зупинок на маршруті, а також унікальні алгоритми візуальної симуляції у режимах «на карті» та схематичному. Для прогнозування зміни пасажиропотоків після додавання розкладу додаткового транспортного засобу використано можливості повнозв'язної нейронної мережі. Вона надає можливість спрогнозувати зміну пасажиропотоку у відповідний час на даному сегменті маршруту. Дана нейронна мережа дає змогу оптимізувати роботу пасажирських перевезень громадським транспортом. Спрогнозовані нейронною мережею пасажиропотоки, у порівнянні з фактичними, призводять до їх зростання у середньому на 28% на критичних перегонах в годину пік. Ці отриманні результати дозволяють обґрунтувати доцільність додавання розкладу нового транспортного засобу для кращого покриття завантажених ділянок у час пік. Порівняння зміни пасажиропотоків, розподілених за перегонами у часі доби з 19:00 до 20:00, за фактичними даними та після роботи нейронної мережі вказує на їх збільшення у середньому в 70% перегонів, які були спрогнозовані.

ІС візуальної симуляція пасажиропотоків у сфері громадського транспорту дозволить:

DOI 10.15588/1607-3274-2021-4-10

1. Оцінити та наочно відобразити проблемні ділянки (перегони) на маршрутах Smart City.
2. Встановити основні зупинки з найбільшим пасажирообміном та вузли з найбільшим числом пересадок.
3. Приймати ефективне рішення щодо потреби модернізації маршрутів Smart City.
4. Спрогнозувати зміни пасажиропотоків при внесенні коректив до маршрутів Smart City.

ПОДЯКИ

Роботу виконано в рамках держбюджетної теми «Методи та засоби функціонування систем підтримки прийняття рішень на основі онтологій» (ID:839 2017-05-15 09:20:01 (2459-315)). Дослідження провадилось в межах спільних наукових досліджень кафедри інформаційних систем та мереж НУ «Львівська політехніка» на тему «Дослідження, розроблення і впровадження інтелектуальних розподілених інформаційних технологій та систем на основі ресурсів баз даних, сховищ даних, просторів даних та знань з метою прискорення процесів формування сучасного інформаційного суспільства». Наукові дослідження провадилися також в рамках ініціативної тематики досліджень кафедри ІСМ НУ «Львівська політехніка» на тему «Розроблення інтелектуальних розподілених систем на основі онтологічного підходу з метою інтеграції інформаційних ресурсів».

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Li W. A dynamic simulation model of passenger flow distribution on schedule-based rail transit networks with train delays / W. Li, W. Zhu // Journal of Traffic and Transportation Engineering. – 2016. – Vol. 3(4). – P. 364–73. DOI: 10.1016/j.jtte.2015.09.009
2. Bachelet B. Enhancing theoretical optimization solutions by coupling with simulation / B. Bachelet, L. Yon // Modeling & Simulation (OICMS) : 1st Open International Conference, 13–15 June 2005: proceedings. – Clermont-Ferrand: Blaise Pascal University, 2005. – P. 331–342.
3. Рябов И. М. Совершенствование организации облуживания пассажиров на маршруте путем использования автобусов разной вместимости / И. М. Рябов, Р. Я. Кашманов // Вестник Сибирской государственной автомобильно-дорожной академии. – 2019. –№ 16(3). – С. 264–275.
4. Voorhees A. M. A General Theory of Traffic Movement / A. M. Voorhees // Transportation. – 2013. – Vol. 40(6). – P. 1105–1116. DOI: 0.1007/s11116-013-9487-0
5. Short-term prediction of bus passenger flow based on a hybrid optimized LSTM network / [Y. Han, C. Wang, Y. Ren et al.] // ISPRS International Journal of Geo-Information. – 2019. – Vol. 8(9). – P. 366. DOI: 10.3390/ijgi8090366
6. Cronin J. J. Measuring Service Quality: a Reexamination and Extension / J. J. Cronin, S. A. Taylor // Journal of Marketing. – 1992. – Vol. 56(3). – P. 55–68.
7. Parasuraman A. SERVQUAL: a Multiple Item Scale for Measuring Consumer Perceptions of Service Quality / A. Parasuraman, V. A. Zeithaml, L. L. Berry // Journal of Retailing. – 1988. – Vol. 64 (1). – P. 12–37.
8. Beirão G. Enhancing Service Quality in Public Transport Systems / G. Beirão, J. S. Cabral // WIT Transactions on The Built Environment. – 2006. – Vol. 89. – P. 837–845. doi:10.2495/UT060811
9. Friman M. Service Supply and Customer Satisfaction in Public Transportation: the Quality Paradox / M. Friman, M. Fellesson // Journal of Public Transportation. – 2009. – Vol. 12(4). – P. 57–69. DOI: 10.5038/2375-0901.12.4.4
10. Local public transport, service quality and tendering contracts / [E. Marcucci, E. Valeri, A. Stathopoulos, V. Gatta] // Urban Sustainable Mobilità. – Milan: FrancoAngeli. – 2012. – Vol. 1. – P. 161–172.
11. Service Quality and Customer Satisfaction of PublicTransport on Cape Coast-Accra Route, Ghana / [T. K. Ojo, D. O. Mireku, S. Dauda, R. Y. Nutsogbodo] // Developing Country Studies. – 2014. – Vol. 4(18). – P. 142–149.
12. Optimization of multiple vehicle routing problems using approximation algorithms / [R. Nallusamy, K. Duaiswamy, R. Dhanalaksmi, P. Parthiban] // International Journal of Engineering Science and Technology. – 2009. – Vol. 1(3). – P. 129–135.
13. Optimizing bus transit network with parallel ant colony algorithm / [B. Yu, Z. Yang, C. Cheng, C. Liu] // Proceedings of the Eastern Asia Society for Transportation Studies. – 2005. – Vol. 5. – P. 374–389.
14. Afandizadeh Sh. Bus fleet optimization using genetic algorithm a case study of Mashhad / Sh. Afandizadeh, H. Khaksar, N. Kalantari // International Journal of Civil Engineering. – 2013. – Vol. 11(1). – P. 43–52.
15. Zhongzhen Ya. A parallel ant colony algorithm for bus network optimization / Ya. Zhongzhen, Yu. Bin, Ch. Chuntian // Computer-Aided Civil and Infrastructure Engineering. – 2007. – Vol. 22(1). – P. 44–55. DOI: 10.1111/j.1467-8667.2006.00469.x
16. Matseliukh Y., Intelligent System of Visual Simulation of Passenger Flows / Y. Matseliukh, V. Vysotska, M. Bublyk // Computational Linguistics and Intelligent Systems (COLINS 2020) : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 906–920.
17. Intelligent System of Passenger Transportation by Autopiloted Electric Buses in Smart City / [M. Bublyk, Y. Matseliukh, U. Motorniuk, M. Terebukh] // Computational Linguistics and Intelligent Systems (COLINS 2020) : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 1280–1294.
18. Analysis of the Demand for Bicycle Use in a Smart City Based on Machine Learning / [D. Koshtura, M. Bublyk, Y. Matseliukh et al.] // Modern Machine Learning Technologies and Data Science (MoMLeT+DS 2020) : 2nd International Workshop, Lviv-Shatsk, 2–3 June 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2631. – P. 172–183.
19. Optimization Model of the Buses Number on the Route Based on Queueing Theory in a Smart City / [L. Podlesna, M. Bublyk, I. Grybyk et al.] // Modern Machine Learning Technologies and Data Science (MoMLeT+DS 2020) : 2nd International Workshop, Lviv-Shatsk, 2–3 June 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2631. – P. 502–515.
20. Bayesian Networks' Development Based on Noisy-MAX Nodes for Modeling Investment Processes in Transport / [V. Lytvynenko, N. Savina, J. Krejci et al.] // Modern

DOI 10.15588/1607-3274-2021-4-10

Machine Learning Technologies and Data Science (MoMLeT&DS-2019) : 8th International Conference on "Mathematics. Information Technologies. Education", Shatsk, 2–4 June 2019 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2019. – Vol. 2386. – P. 1–10.
21. Fomenko A. Specialized Automated System for Control and Support of Rail Cargo Transportation / A. Fomenko, V. Vyshnia // International Journal of Computing. – 2019. – Vol. 18(2). – P. 191–200. DOI: 10.47839/ijc.18.2.1417
22. Mochurad L. Parallel Approach of the Algorithm of Finding the Optimal Solution of the Transport Problem by the Method of Potentials / L. Mochurad, N. Boyko, M. Bortnikova // Computational Linguistics and Intelligent Systems (COLINS 2020) : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 952–963.
23. Development of models and means of the server part of the system for passenger traffic registration of public transport in the "smart" city / [O. Y. Boreiko, V. M. Teslyuk, A. Zelinskyy, O. Berezsky] // Eastern-European Journal of Enterprise Technologies. – 2017. – Vol. 1 (2–85). – P. 40–47. DOI: 10.15587/1729-4061.2017.92831

Стаття надійшла до редакції 25.04.2021
Після доробки 01.10.2021.

УДК 004.9

ТЕХНОЛОГИЯ ВИЗУАЛЬНОЙ СИМУЛЯЦИИ ПАССАЖИРОПОТОКА В СФЕРЕ ОБЩЕСТВЕННОГО ТРАНСПОРТА SMART CITY

Лытвын В. В. – д-р техн. наук, профессор, заведующий кафедры «Информационные системы и сети», Национальный университет «Львовская политехника», Украина.

Бублык М. И. – д-р эконом. наук, профессор, профессор кафедры менеджмента и международного предпринимательства, Национальный университет «Львовская политехника», Львов, Украина.

Высоцкая В. А. – канд. техн. наук, доцент, доцент кафедры «Информационные системы и сети», Национальный университет «Львовская политехника», Украина.

Мацелюх Ю.Р. – магистр кафедры «Информационные системы и сети», Национальный университет «Львовская политехника», Украина.

АННОТАЦИЯ

Актуальность. На сегодняшний день проблема визуальной симуляции пассажиропотоков в сфере общественного транспорта является важной при создании информационных систем для развития современных Smart City. В условиях Industry 4.0 важным является создание технологий, средств и инструментов для внедрения единой саморегулируемой интеллектуальной системы обмена данными при предоставлении соответствующих услуг перевозки пассажиров в сфере общественного транспорта. Например, наглядно отображать проблемные участки на маршрутах в Smart City, форматы умножать и идентифицировать основные остановки во временных срезах с самым пассажирообменом, формировать предложения о необходимости модернизации маршрутов с учетов увеличением загруженности общественного транспорта на определенных участках Smart City и получать результаты прогнозирования пассажиропотоков при внесении соответствующих изменений на основе методов машинного обучения.

Целью исследования разработка технологии визуальной симуляции пассажиропотоков в сфере общественного транспорта для повышения качества предоставления услуг пассажирских перевозок в Smart City.

Метод. Усовершенствована имитационную модель для вычисления пассажиропотока при изменении количества подвижного состава на маршруте, где в отличие от известных, добавлено прогнозирования на основе разработанной нейронной сети. Усовершенствован механизм визуальной симуляции пассажиропотоков с использованием карт GoogleMaps и динамическое перемещение по ним с контролем скорости отображения симуляции. Предложено нейронную сеть с полносвязную слоями с использованием оптимизационного алгоритма с адаптивным уровнем обучения Adam для прогнозирования пассажиропотока между остановками за определенный период времени. Определены критерии детализации данных о пассажиропотоки на городских маршрутах, включая общие показатели отношения пассажирообмена на определенной остановке к текущему времени суток. При проектировании интеллектуальной системы потерпел дальнейшего развития подход изменения вместимости подвижного состава общественного транспорта в Smart City, где в отличие от известных, изменение емкости ограничена имеющимися транспортными средствами. Потерпел дальнейшее развитие метод расчета комплекса показателей пассажиропотоков на остановках и гонке с учетом дополнительных локальных графиков и специфики работы транспорта на отдельных индивидуальных маршрутах.

Результаты. Разработано интеллектуальную систему визуального моделирования пассажирских перевозок на основе нейронной сети и машинного обучения, которая позволяет оптимизировать работу пассажирских перевозок общественным транспортом в Smart City. Это представление данных позволяет оценить рентабельность добавления нового транспортного средства на маршрут необходимо скорректировать расписание других транспортных средств для лучшего покрытия загруженных участков в час пик. Для работы программного средства используется общеизвестный стандарт представления данных об общественном транспорте – GTFS. Это позволяет адаптировать разработан программный продукт универсальным, а не специфическим для конкретного города или страны. Проведено сравнение полученных результатов на множестве данных троллейбусных маршрутов (около 2000 записей, с собранных на основе экспериментальных маркетинговых исследований) в городе Львов (Украина) для формирования прогноза изменения пассажиропотоков на определенных участках в разные промежутки времени.

Выводы. Выявлено, что спрогнозированы нейронной сетью пассажиропотоки по сравнению с фактическими, приводят к их росту в среднем на 28% в критических гонке в час пик. Эти полученные результаты позволяют обосновать целесообразность добавления расписания нового транспортного средства для лучшего покрытия загруженных участков в час пик. Сравнение изменения пассажиропотоков, распределенных по гонкой во времени суток с 19:00 до 20:00, по

DOI 10.15588/1607-3274-2021-4-10

фактическим данным и после работы нейронной сети указывает на их увеличение в среднем в 70% гонок, которые были спрогнозированы, что позволит принять обоснованное решение о запуске дополнительного транспорта на соответствующие маршруты.

КЛЮЧЕВЫЕ СЛОВА: визуальная симуляция пассажиропотоков, визуальное моделирование, интеллектуальная система, Smart City, GoogleMaps, нейронная сеть, прогнозирование пассажиропотока, машинное обучение, информационные технологии, обработка данных.

UDC 004.9

VISUAL SIMULATION TECHNOLOGY FOR PASSENGER FLOWS IN THE PUBLIC TRANSPORT FIELD AT SMART CITY

Lytvyn V. – PhD, Professor, Head of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

Bublyk M. – Dr. Sc., Professor, Professor of Management and International Business Department, Lviv Polytechnic National University, Lviv, Ukraine.

Vysotska V. – PhD, Associate Professor of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

Matseliukh Y. – Postgraduate student student of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

ABSTRACT

Context. Today, the problem of visual simulation of passenger flow in public transport is essential in creating information systems for the development of modern Smart City. In Industry 4.0, it is crucial to develop technologies, means, and tools for implementing a single self-regulatory intelligent data exchange system in the provision of appropriate passenger transportation services in public transport. Today the following is essential: to visually display problem areas on routes in Smart City; to form and identify the main stops in time sections with the largest passenger exchange; to create proposals on the need to modernise routes taking into account the increase in public transport congestion in certain areas of Smart City, and to obtain results of passenger flow forecasting when making appropriate changes based on machine learning methods.

Objective of the study is to develop a technology for visual simulation of passenger traffic in the field of public transport to improve the quality of passenger services in Smart City.

Method. They have improved the simulation model for calculating **passenger flow** when changing the number of rolling stock on the route, in contrast to the known, added forecasting based on the developed neural network. The mechanism of visual simulation of passenger flows using GoogleMaps maps and dynamic movement on them with control of simulation display speed has been improved. A neural network with fully connected layers utilising an optimisation algorithm with an adaptive level of learning Adam to predict the flow of passengers between stops for a certain period of the day is proposed. Criteria for detailing passenger flows on urban routes are defined, including general indicators of the ratio of passenger traffic at a specific stop to the current period of the day. When designing the intelligent system, changing the capacity of public transport rolling stock in Smart City was further developed. Unlike the known ones, the available vehicles limit the change of power. The method of calculating a set of indicators of passenger traffic at stops and races, taking into account different local schedules and the specifics of transport on individual routes, has undergone further development.

Results. An intelligent system of visual modelling of passenger traffic based on a neural network and machine learning has been developed, allowing optimising passenger traffic by public transport in Smart City. This data presentation makes it possible to assess the profitability of adding a new vehicle to the route or adjusting the schedule of other cars to cover the loaded areas during peak hours better. The well-known standard of public transport data presentation – GTFS is used for the operation of the software. It allows you to adapt the developed software product to the universal, rather than specific to a particular city or country. It was provided with a comparison of the obtained results on a data set of trolleybus routes (about 2000 records, collected based on experimental marketing research) in Lviv (Ukraine) to form a forecast of changes in passenger flow on certain sections at different times.

Conclusions. It was found that the passenger flows predicted by the neural network in comparison with the actual ones lead to their growth by an average of 28% in critical races at rush hour. These results allow us to justify adding a schedule of a new vehicle for better coverage of loaded areas during peak hours. A comparison of changes in passenger traffic distributed by races during the day from 19:00 to 20:00, according to actual data and after the operation of the neural network indicates an increase in their average 70% of races that were predicted, which will allow a reasonable decision to launch additional transport on appropriate routes.

KEYWORDS: visual simulation of passenger flows, visual modelling, intelligent system, Smart City, GoogleMaps, neural network, passenger flow forecasting, machine learning, information technology, data processing.

REFERENCES

1. Li W., Zhu W. A dynamic simulation model of passenger flow distribution on schedule-based rail transit networks with train delays, *Journal of Traffic and Transportation Engineering*, 2016, Vol. 3(4), pp. 364–73. DOI: 10.1016/j.jtte.2015.09.009
2. Bachelet B., Yon L. Enhancing theoretical optimization solutions by coupling with simulation, *Modeling & Simulation (OICMS), 1st Open International Conference, 13–15 June 2005, proceedings*. Clermont-Ferrand, Blaise Pascal University, 2005, pp. 331–342.
3. Ryabov I. M., Kashmanov R. YA. Improving the organization of servicing passengers on the route by using buses of different capacities, *Bulletin of the Siberian State Automobile and Road Academy*, 2019, Vol. 16(3), pp. 264–275.

DOI 10.15588/1607-3274-2021-4-10

4. Voorhees A. M. A General Theory of Traffic Movement, *Transportation*, 2013, Vol. 40(6), pp. 1105–1116. DOI: 0.1007/s11116-013-9487-0
5. Han Y., Wang C., Ren Y., Wang S., Zheng H., Chen G. Short-term prediction of bus passenger flow based on a hybrid optimized LSTM network, *ISPRS International Journal of Geo-Information*, 2019, Vol. 8(9), P. 366. DOI: 10.3390/ijgi8090366
6. Cronin J. J., Taylor S. A. Measuring Service Quality: a Reexamination and Extension, *Journal of Marketing*, 1992, Vol. 56(3), pp. 55–68.
7. Parasuraman A., Zeithaml V. A., Berry L. L. SERVQUAL: a Multiple Item Scale for Measuring Consumer Perceptions of Service Quality, *Journal of Retailing*, 1988, Vol. 64 (1), pp. 12–37.
8. Beirão G., Cabral J. S. Enhancing Service Quality in Public Transport Systems, *WIT Transactions on The Built Environment*, 2006, Vol. 89, pp. 837–845. DOI: 10.2495/UT060811
9. Friman M., Fellesson M. Service Supply and Customer Satisfaction in Public Transportation: the Quality Paradox, *Journal of Public Transportation*, 2009, Vol. 12(4), pp. 57–69. DOI: 10.5038/2375-0901.12.4.4
10. [Marcucci E. , Valeri E., Stathopoulos A., Gatta V. Local public transport, service quality and tendering contracts, *Urban Sustainable Mobilità*. Milan, FrancoAngeli, 2012, Vol. 1, pp. 161–172.
11. Ojo T. K., Mireku D. O., Dauda S., Nutsogbodo R. Y. Service Quality and Customer Satisfaction of PublicTransport on Cape Coast-Accra Route, Ghana, *Developing Country Studies*, 2014, Vol. 4(18), pp. 142–149.
12. Nallusamy R., Duaiswamy K., Dhanalaksmi R., Parthiban P. Optimization of multiple vehicle routing problems using approximation algorithms, *International Journal of Engineering Science and Technology*, 2009, Vol. 1(3), pp. 129–135.
13. Yu B., Yang Z., Cheng C., Liu C. Optimizing bus transit network with parallel ant colony algorithm, *Proceedings of the Eastern Asia Society for Transportation Studies*, 2005, Vol. 5, pp. 374–389.
14. Afandizadeh Sh., Khaksar H., Kalantari N. Bus fleet optimization using genetic algorithm a case study of Mashhad, *International Journal of Civil Engineering*, 2013, Vol. 11(1), pp. 43–52.
15. Zhongzhen Ya., Bin Yu, Chuntian Ch. A parallel ant colony algorithm for bus network optimization, *Computer-Aided Civil and Infrastructure Engineering,* 2007, Vol. 22(1), pp. 44–55. DOI: 10.1111/j.1467-8667.2006.00469.x
16. Matseliukh Y., Vysotska V., Bublyk M. Intelligent System of Visual Simulation of Passenger Flows, *Computational Linguistics and Intelligent Systems (COLINS 2020), 4th International Conference, Lviv, 23–24 April 2020, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 906–920.
17. Bublyk M., Matseliukh Y., Motorniuk U., M. Terebukh Intelligent System of Passenger Transportation by Autopiloted Electric Buses in Smart City, *Computational Linguistics and Intelligent Systems (COLINS 2020), 4th International Conference, Lviv, 23–24 April 2020, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 1280–1294.
18. Koshtura D., Bublyk M., Matseliukh Y., Dosyn D., Chyrun L., Lozynska O., Karpov I., Peleshchak I., Maslak M., Sachenko O. Analysis of the Demand for Bicycle Use in a Smart City Based on Machine Learning, *Modern Machine Learning Technologies and Data Science (MoMLeT+DS 2020), 2nd International Workshop, Lviv-Shatsk, 2–3 June 2020, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020,Vol. 2631, pp. 172–183.
19. Podlesna L., Bublyk M., Grybyk I., Matseliukh Y., Burov Y., Kravets P., Lozynska O., Karpov I., Peleshchak I., R. Peleshchak] Optimization Model of the Buses Number on the Route Based on Queueing Theory in a Smart City, Modern Machine Learning Technologies and Data Science (MoMLeT+DS 2020), 2nd International Workshop, Lviv-Shatsk, 2–3 June 2020, CEUR workshop proceedings. Aachen, CEUR-WS.org, 2020, Vol. 2631, pp. 502–515.
20. Lytvynenko V., Savina N., Krejci J., Voronenko M., M. Yakobchuk, O. Kryvoruchko Bayesian Networks' Development Based on Noisy-MAX Nodes for Modeling Investment Processes in Transport, *Modern Machine Learning Technologies and Data Science (MoMLeT&DS-2019) : 8th International Conference on "Mathematics. Information Technologies. Education", Shatsk, 2–4 June 2019 : CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2019, Vol. 2386, pp. 1–10.
21. Fomenko A., Vyshnia V. Specialized Automated System for Control and Support of Rail Cargo Transportation, *International Journal of Computing*, 2019, Vol. 18(2), pp. 191–200. DOI: 10.47839/ijc.18.2.1417
22. Mochurad L., Boyko N., Bortnikova M. Parallel Approach of the Algorithm of Finding the Optimal Solution of the Transport Problem by the Method of Potentials, *Computational Linguistics and Intelligent Systems (COLINS 2020), 4th International Conference, Lviv, 23–24 April 2020, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 952–963.
23. Boreiko O. Y., Teslyuk V. M., Zelinskyy A., Berezsky O. Development of models and means of the server part of the system for passenger traffic registration of public transport in the "smart" city, *Eastern-European Journal of Enterprise Technologies,* 2017, Vol. 1(2–85), pp. 40–47. DOI: 10.15587/1729-4061.2017.92831

DOI 10.15588/1607-3274-2021-4-10

UDC 519.688: 004.75:004.421

ANALYSIS OF THE USE OF MULTITHREADED COMPUTING TECHNOLOGIES TO FACTORIZE OF NUMBERS BY A BINARY ALGORITHM

Prots'ko I. – Dr. Sc., Associate Professor, Department of Automated Control Systems, Lviv National Polytechnic University, Lviv, Ukraine.
Rykmas R. – Software Developer, LtdC "Uniservice", Lviv, Ukraine.

ABSTRACT

Context. Providing high-speed computation by computer systems of factorization of number into prime factors requires the development of effective algorithmic methods using computational technologies. Fast computation of factorization of numbers is used in such applications as, protection of information data, in algorithms of discrete transforms for transition from one to multidimensional computations and others.

Objective. The purpose of the work is to analyze the implementation of technologies of multithreaded computation of factorization of integer value by the binary algorithm of the method of trial divisions using computer systems with multi-core processors and graphics accelerators.

Method. A binary algorithm of trial divisions that uses the remainders of each digit of the binary representation of a number to perform a divisibility check on prime factors of the canonical factorization of number in parallel.

Results. The analysis and comparison of multithreaded computations of software implementations of factorization of number by binary algorithm using hyper-threading, AMP C++, CUDA technologies in computer systems with multi-core processors and graphics accelerators. The results of the process of number factorization for multithreaded computing technologies using the same parallel core function are analyzed.

Conclusions. In the study of realizations of number factorization by the binary algorithm in the multithreaded mode, the technology of hyper-threading calculations using multicore processors is most effectively performed. Heterogeneous computing using AMP C++ or CUDA technologies on computer systems and graphics accelerators requires consideration of GPU microarchitecture features for parallel computing core functions.

KEYWORDS: factorization, prime factors, multithreading, heterogeneous computation, parallel computation, remainder.

ABBREVIATIONS

AMP is an accelerated massive parallelism;
CPU is a central processing unit;
CMT is a chip multi-threading;
CS is a computer systems;
CUDA is a compute unified device architecture;
CTPL is a C++ thread pool library;
FN is a factorization of number;
HT is a hyper-threading;
GPU is a graphics processing unit;
GPGPU is a general-purpose computing on GPU;
IDE is an integrated development environment;
TD is a trial divisions.

NOMECLATURE

i is the number of binary digit (i = 0,1,… k–1);
k is the bit size of a number of factorization;
M_t is a remainder;
N is an integer number of factorization;
n_i is a binary digit of the number of factorization;
p is a prime number;
p_i is a prime factor;
T is the periodicity of remainders;
s_k is a degree of a prime factor;
t is the number of the vertical ruler of M_t.

INTRODUCTION

The fast execution of factorization, as a process of decomposition of an integer number into prime factors, has demand in many applications. FN is used to move from one-dimensional to multidimensional representation of sequences of the discrete transforms, which is widely used in multidimensional data processing tools [1]. In the Good-Thomas algorithm of calculating the discrete Fourier transform of a composite size N, which uses the Chinese remainder theorem for integers, it is not necessary to perform products on return factors if the size N the sequence of the transform is decomposed into relatively prime factors. The Agarwal-Cooley method converts the calculation of a one-dimensional N-point convolution into a multidimensional one, provided that the size is decomposed into relatively prime factors [2]. FN is used to find one of the parameters of cryptosystems in asymmetric algorithms for encrypting information data [3].

Since the emergence of new approaches in encryption algorithms in the late 70's, a number of FN algorithms have been developed. These include ρ – method, (p–1) – Pollard's method, elliptic curve method, square sieve method, numerical field sieve method and others [4]. Based on these methods, software modules and modifications of classical algorithms for solving various applied problems are developed [5].

The object of study is the process of developing algorithmic and software for FN into prime factors using the technologies of parallel computations.

The subject of the study is multi-threading computations the binary algorithm of the method of TD, which

DOI 10.15588/1607-3274-2021-4-11

performs a check for divisibility by a prime number and its degree using the remainders of each bit of the number.

The purpose of the work is to parallelize the computations of FN according to the binary algorithm of the method of TD using CS with multi-core processors and graphics accelerators for efficient determination the values of prime factors.

1 PROBLEM STATEMENT

Suppose an integer N is given, which we factorize into prime factors in the form

$$N = p_1^{s1}\, p_2^{s2} \ldots p_n^{sn}, \quad (1)$$

where $p_i < p_{i+1}$, i=1, 2,…n.

The simplest solution for factorization is the method of TD. The method is well parallelized using division tests without the remainder of the number N on a set of divisors from the set of prime numbers p_i = {2, 3, 5, 7, 9, 11,…}, i = 1, 2,…. However, the sequential execution of division operations to check for multiplicity in modern universal CS is performed using a division firmware that has the highest weight among arithmetic operations [6]. The problem of increasing the speed of calculation FN is solved in the direction of replacing the test division operation with a set of parallel computations with operations of adding numbers less than p_i and comparing the obtained sum with the corresponding prime number.

To quickly determine the values of simple decomposition factors, the possibility of parallelization of FN by the binary algorithm of the TD method using HT technologies in CS with multi-core processors and graphics accelerators is used.

2 REVIEW OF THE LITERATURE

Algorithmic means of IF, which characterized by regularity, modularity, simplicity and information independence of their components, are effectively implemented by computing means in combination with computing information technologies. The software implementation of algorithms of FN in modern CS is analyzed from the standpoint of the possibility of parallel execution using multithreaded organization [7, 8, 9].

Factorization algorithms based on ρ and (p − 1) methods of Pollard [10], Sherman-Lehman method [11], Shanks method [12], Lenstra elliptic curves method [13] are investigated in the direction of the possibility of parallelization. The main idea of these methods is to choose a random value of a number for a single parallel process, which is close to the element of the factorization. However, the execution of the algorithm by several simultaneous significant flows to find the factors of the number of the factorization does not always give the expected result.

In the known Fermat method, the main idea of the number $N = p_1 p_2$ is to find such pairs of natural numbers that satisfy a certain condition. However, in the case when the factors p_1 and p_2 are close in value to 1 and N, the algorithm will perform worse than the method of TD [14].

In the work [15] the improvement of computations of FN according to the binary algorithm of the TD method due to parallel computations and efficient use of CS computing resources is considered.

Comparative analysis of FN using CMT and CS with graphics accelerators is relevant to improve the software implementation of integer factorization algorithms.

3 MATERIALS AND METHODS

The well-known factorization algorithm is TD, which consists in checking the divisibility of the number N for a sequence of prime numbers, to a value less than or equal to the square root of the number N.

The work [16] describes a binary algorithm of the method of TD, in which the representation of the integer value of the number of the factorization in the binary number system is used to determine the divisibility of a decimal number on a prime number.

$$N = (n_{k-1}\, 2^{k-1} + n_{k-2}\, 2^{k-1} + \ldots + n_1\, 2^1 + n_0\, 2^0). \quad (2)$$

The values of the digits n_i (i = 0,1,… k–1) are equal to 0 or 1.

As a result, the check for the divisibility of N by a prime number p is reduced to the determination modulo p of each weighting factor 2^i (i = 0,1,…, k–1) of the number N, which has the form

$$\begin{aligned} N \bmod p &= (n_{k-1}\, 2^{k-1} + n_{k-1}\, 2^{k-2} + \ldots + n_0\, 2^0) \bmod p = \\ &= n_{k-1}\, (2^{k-1} \bmod p) + n_{k-1}\, (2^{k-2} \bmod p) + \ldots \\ &\quad + n_1\, (2^1 \bmod p) + n_0\, (2^0 \bmod p)\,. \end{aligned} \quad (3)$$

After accumulating the values of the remainders on the selected weights (2^i mod p) (i = 0,1,… k–1) for n_i (i = 0,1,…, k–1), equal to 1, compare the accumulated amount with a prime number p. In the case when the accumulated value of remainders is greater than p – again from the previously obtained accumulated value is carried out according to formula (3) the accumulation of remainders. In the case when the accumulated value of the residuals is equal to the number p – the decomposition element p is output and the transition to rechecking the divisibility of the power of this prime number is performed. In the case of comparison, when the accumulated value of the residuals is less than p, we move on to the next value from the sequence of prime numbers. Therefore, we obtain a set of prime factors of the factorization (1) of the number N.

To analyze the parallel organization, consider the work of the binary algorithm of FN using a block diagram (Fig. 1). The structure chart of FN into prime factors Pi consists the following functional blocks: 1 – the multiplexer, 2 – the block of lines of memory of the remainders, 3 – the multi-input adder, 4 – the first comparison unit, 5 – the control unit, 6 – the memory block of prime factors, 7 – the shift register, 8 – the second block of comparison.

The memory block 2 of the periods of the remainders of prime numbers and their powers contains T values of the remainders M_t (t = 0,1,…, T–1) for each of $p_i^{j,}$ which are partially shown in Table 1.

DOI 10.15588/1607-3274-2021-4-11

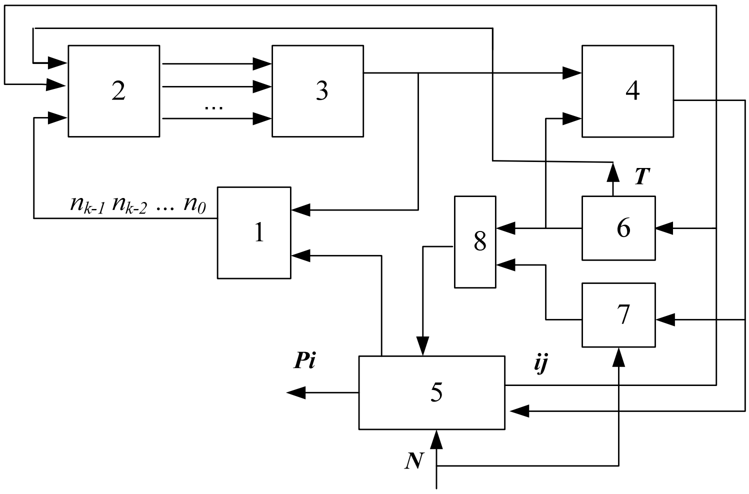


Figure 1 – The structure chart of the FN into factors p_i

Table 1 – Table of the remainders

i j	…M_t	M_9	M_8	M_7	M_6	M_5	M_4	M_3	M_2	M_1	M_0	mod $p_i^{\,j}$	*T*
11										2	1	mod 3^1	2
12						5	7	8	4	2	1	mod 3^2	6
13	25	26	13	20	10	5	16	8	4	2	1	mod 3^3	18
…	…	…	…	…	…	…	…	…	…	…	…	…	
21								3	4	2	1	mod 5^1	4
22	24	12	6	3	14	7	16	8	4	2	1	mod 5^2	20
23	24	12	6	3	64	32	16	8	4	2	1	mod 5^3	100
…	…	…	…	…	…	…	…	…	…	…	…	…	
31									4	2	1	mod 7^1	3
32	44	22	11	30	15	32	16	8	4	2	1	mod 7^2	21
33	338	169	256	128	64	32	16	8	4	2	1	mod 7^3	147
…	…	…	…	…	…	…	…	…	…	…	…	…	
41		6	3	7	9	10	5	8	4	2	1	mod 11^1	10
42	56	28	14	7	64	32	16	8	4	2	1	mod 11^2	110
…	…	…	…	…	…	…	…	…	…	…	…		
51	10	5	9	11	12	6	3	8	4	2	1	mod 13^1	12
…	…	…	…	…	…	…	…	…	…	…	…	…	

If there is a value for the n_i binary digit (2) of the number N equal 1 from the memory block 2 of periods of remainders of prime numbers and their powers, the values of the remainder M_t by ij – sequence number of prime numbers and their powers and number k are read in parallel for each bit. For example, determining the value of the remainders for the number of binary digit i = 28 (i = 0,1,… k–1) of the number of the FN, when checking for redivisibility by p = 5, that is mod $p_i^{\,j}$ = mod 5^2. According to Table 1, this corresponds to the horizontal ruler ij = 22 (with the periodicity of remainders T = 20) and the vertical ruler M_8, where t = (28 mod 20) = 8, at the intersection of ruler is the value of the remainder M_8 = 6.

The maximum value of the repetition period T of the remainders is equal $T_{\max} = (p_i^{\,j} - 1)$. The horizontal rulers for $p_i^{\,j} = (2^j - 1)$, which are prime Mersenne numbers, have a minimum value of $T_{\min} = j$. Thus, due to the periodic repetition of the remainders, with increasing bit size of the decomposition number N, the number t of the vertical ruler M_t, t =0,1,2,…T–1, is determined by the formula

$$t = i \bmod T,\ (i = 0,1,\ldots\ k-1). \qquad (4)$$

The organization of the memory block 2 of the periods of the remainders of prime numbers and their powers allows you to read in parallel the remainders for each of the bits ($n_{k-1}\ n_{k-2}\ldots\ n_0$) of the number.

If there are two values of bits equal to 1 in the number N, which are control inputs for the initial multi-input adders 3 of remainders, the value of the sum of remainders ($M_k + M_{k+1}$) will be set at the output of the adders of remainders. If at the control inputs for the initial multi-input adders 3 one of the two values of the bits of the number N is 0 another 1, then at output of initial multi-input adders the value of the remainders will be set, which will correspond to the remainders Mk for the 1 bit of the number N. For the case when at the control inputs for the initial multi-input adders two values of the bits of the number N are equal to 0, then the output of the initial adder will be set the value of remainder equal to zero. Similarly, with the help of logical operations OR, in the second stage of accumulation of remainders, depending on the values of the result of the operation OR, the corresponding value of the remainder will be set. The largest delay in the parallel accumulation of remainders will be in the case when all binary bits (2) will be equal to 1.

DOI 10.15588/1607-3274-2021-4-11

The multi-input adder 3, in parallel by adding the value of the remainders, from the outputs pass result in first comparison unit 4 for comparison with the value p_i. According to the specifier at the output of first comparison unit 4, if the value is "less" – the prime number p_i or its power is not an element of the canonical factorization and there is a transition to the next ruler *ij*. According to the specifier at the output of unit 4, if the value is "greater" and "equal", one of the two directions of calculations is selected. In the case of "greater", the multiplexer 1 is switched to the output value of the multi-input adder 3, in the case of "equal" – the p_i element of the factorization is memorized and the transition to determining the its degree *j* takes place. For the specifier "equal" at the output of first comparison unit 4 in the shift register 7 provides a shift by one digit to the left of the binary value of the initial number N, thereby reducing the allowable value of the prime number or its power in the analysis of the factorization. The allowable value is compared with a prime number or its power in the second comparison block 8 and in the case of comparison greater than or equal to its output is a sign of completion of the canonical factorization.

The binary algorithm of TD can perform a divisibility check in parallel for each of the prime numbers. In this case, the memory unit 2 contains the values of the remainders only for each of the prime numbers and their powers. Otherwise, it is possible to divide the memory unit 2 into parts for groups of prime numbers and their powers, which corresponds to coarse-grained parallelism.

4 EXPERIMENTS

The computation of FN on prime factors p_i based on binary algorithm of TD is implemented in the IDE Visual C ++ 2019 using technologies HT (Intel) and AMP C ++ (MS), CUDA (NVIDIA) of GPGPU.

Software implementation of parallel computation of FN using multi-threading technology CMT is achieved by creating threads and proportional distribution of their performance between the cores of the microprocessor in the CS. For efficient use of computing resources, a pool of threads is used and the organization of interaction between them is entrusted to the function of the CTPL library [17], which is an add-on to the standard STL library and has the ability to work with system threads.

Software implementations of parallel computation of FN using AMP C ++, CUDA technologies perform a similar algorithm in the kernel function as in CMT technology.

The developed programs are implemented to perform the following actions:

1. Enter the number *N* of factorization.
2. Construction of a sequence of remainders on the prime numbers p_i by the number *T*.
3. Analysis for the divisibility of the number *N* prime number p_i.
4. Output to the file of the factors p_i for number *N*.
5. Go to the next prime p_{i+1} with the constraint of the search set of prime numbers.

DOI 10.15588/1607-3274-2021-4-11

The configuration of the test system is as follows:
CPU: Intel Core i5-9300H (4 cores/8 threads) @2,4 GHz;
GPU: NVIDIA GeForce GTX 1650, 4.0 GB, 896 32-bit core CUDA @1,45GHz;
Memory: 8.0 GB, DDR4-2667 MHz.

The chosen evaluation method involves determining the time spent on computing of FN. The whole process of parallel FN includes both the main part of multithreaded computations, and other actions performed by the program to allocate and free memory, output.

5 RESULTS

Testing the effectiveness of technologies to FN is determined by the computation time. To compare and analyze the efficiency of the use of technologies for computing the FN by the binary algorithm of the method of TD, a number of tests were performed. The time of execution of factoring computations for the developed programs threadsSplitNumber and two programs ampSplitNumber and cudaSplitNumber was determined using GPGPU.

The test results are presented in Table 2, which contains data on the average time (microseconds) of FN in 32-bit mode, which is computed 5000 times.

Table 2– The average time of FN

N	Average time (μs)		
	Threads SplitNumber	Cuda SplitNumber	Amp SplitNumber
15361*14537 = 223302857	126	132	212
15107*18959 = 286413613	83	125	226
30829*22433 = 691586957	75	122	200
29717*52433 = 1558151461	138	129	219

The values of number *N* to FN are chosen to be equal to the product of two prime numbers. Created threads for analysis for divisibility are implemented by computing cores in 32-bit mode. Table 1 shows the time relationship (μsec) for the number of threads 10 in the program threadsSplitNumber [15].

6 DISCUSSION

Almost all modern operating systems support of threads control. The application program implements threads control with the help of special libraries that allow to achieve hardware acceleration of the canonical FN.

Comparison of the results of Table 2 shows the increase in execution time in the binary algorithm of TD in the order of threadsSplitNumber, cudaSplitNumber, ampSplitNumber.

In the program threadsSplitNumber with the optimal number of threads 10 we have the shortest execution time. The threads will be divided equally to process the Table of the remainders (Table 1). Each thread processes the table of the remainders and the binary form of the number *N* to decide on the divisibility of the corresponding prime number and its power. To organize the use of the pool of threads and the interaction between them, the function of the CTPL library is used [17], which is an add-on to the standard STL library, which has the ability to work with system threads.

CudaSplitNumber uses NVIDIA's CUDA technology [18]. This technology ushered a new era of improved performance for many applications as programming GPUs became simpler: archaic terms such as texels, fragments, and pixels were superseded with threads, vector processing, data caches and shared memory. Accordance GCGPU, every problem that can be parallelized on GPU always passes the following phases:

1. Transferring data from CPU to GPU.
2. Calling kernel for each thread.
3. Transferring data from GPU to CPU.

Time execution of parallel FN (Table 2) includes 2 and 3 phases. Therefore, the main part of multithreaded computing is the actions performed by the program to select, free of memory and transferring data.

The ampSplitNumber program uses technology for GPGPU called C++ AMP. It accelerates the execution of C ++ code by taking advantage of the GPUs present on video cards with DirectX11 support. But unlike CUDA, that are more oriented in C code, C++ AMP looks like STL library. Both languages extend C ++ with special keywords. For CUDA the syntax is "<<< >>>" and C++ AMP is using the keyword *restrict* (*amp*) to run kernels. It is obvious that code written on C++ AMP is less cluttered and hence easier to read than on CUDA. However, the results of running cudaSplitNumber are better than ampSplitNumber (Table 2). This is because the time to phase call a C++ AMP kernel is much longer than CUDA and a C++ AMP technology has some performance lags compare to 5-years old CUDA technology.

Time execution of parallel FN (Table 2) using GPGPU includes 2 and 3 phases. Therefore, the main part of multithreaded computing is the actions performed by the program to select, free of memory and transferring data.

CONCLUSIONS

Virtually any modern software code execution platform, whether a full-fledged operating system or a virtual machine, contains a set of APIs designed to manage threads and create parallel programs. The FN of binary algorithm of TD with use of parallel technologies of multithreaded computations is considered in the work. The software solution is implemented in the IDE Visual C ++ 2019 with the help of special libraries that allow you to perform parallelization using CS with multi-core processors and graphics accelerators.

The scientific novelty lies in the study of the use of parallel technologies for the binary algorithm of TD, which allows to take into account the peculiarities of the organization of multithreaded computations in problems with an average level of parallelism.

The practical significance of the application of the developed software code of FN in CS with multicore processors and graphics accelerators has shown that the efficiency of computations depends not only on hardware resources, but also on the selected technologies of parallelisation. The obtained results are important for improving the organization of efficient computation of FN for information data protection systems, discrete transforms and other applications.

The direction of further research will be the development of software with the expansion of the bits number of binary representation *N*, which will use the microarchitectural features of specific types of GPU memory, the use of atomic operations and other innovations.

ACKNOWLEDGEMENTS

The work is supported by the state budget scientific research project of Lviv National Polytechnic University "Experimental system of neural network cryptographic protection in real-time data transmission using codes imitating barker codes" (state registration number 0121U109503).

The authors are grateful to Myroslav Mishchuk of the student of project of Lviv National Polytechnic University for participation in testing and discussing the results.

REFERENCES

1. Knuth D. E. The art of computer programming. 3-ed. Vol. 1: Seminumerical Algorithms, Menlo Park. California, Addison-Wesley, 1998, 762 p.
2. McClellan J. H., Rader C. M.. Number Theory in Digital Signal Processing. Englewood Clis, NJ, Prentice-Hall, 1979, 276 p.
3. Stalling W. Cryptography and Network Security: Principles and Practice. 6-ed. Upper Saddle River, NJ, Prentice Hall, 2013, 672 p.
4. Samuel S., Wagstaff Jr. The Joy of Factoring, Providence, RI: American Mathematical Society, 2013, pp. 138–141.
5. Brent R. P. Some parallel algorithms for integer factorisation, *European Conference on Parallel Processing, Toulouse, 1999, Part of the Lecture Notes in Computer Science book series*. Berlin, Springing-Verlag, 1999, Vol. 1685, pp. 1–22.
6. Hindriksen V. How expensive is an operation on a CPU [Electronic resource], Access mode: https://streamcomputing.eu/blog/2012-07-16/how-expensive-is-an-operation-on-a-cpu
7. Mishra M., Chaturvedi U., Saibal K., Pal S. K. A multithreaded bound varying chaotic firefly algorithm for prime factorization, *Advance Computing Conference: 4th IEEE International conference, Gurgaon, India. February 2014: proceedings*. Published by IEEE Computer Society, 2014, pp. 1322–1325. DOI: 10.1109/IAdCC. 2014.6779518
8. Makarenko A. V., Pykhteyev A. V., Yefimov S. S. Parallel'naya realizatsiya i sravnitel'nyy analiz algoritmov faktorizatsii s raspredelennoy pamyat'yu [Electronic resource]. Access mode: http://cyberleninka.ru/article/n/parallelnaya-realizatsiya-i-sravnitelnyy-analiz-algoritmov-faktorizatsii-v-sistemah-s-raspredelyonnoy-pamyatyu
9. Huang L. New prime factorization algorithm and its parallel computing strategy, *Information Technologies in Education and Learning: International Conference (ICITEL 2015), Atlantis, 2015, proceedings*. Published by Atlantis Press, 2015, pp. 1–4.

DOI 10.15588/1607-3274-2021-4-11

10. Koundinya A. K., Harish G., Srinath N. K. et al. Performance Analysis of parallel Pollard's Rho algorithm, *International Journal of Computer Science & Information Technology (IJCSIT),* 2011, Vol. 5, No. 2, pp. 157–163. DOI: 10.5121/ijcsit.2013.5214
11. Kim D. K., Choi P., Lee M.-K. et al. Design and analysis of efficient parallel hardware prime generators, *Journal of semiconductor technology and science,* 2016, Vol. 16, No. 5, pp. 564–581. DOI: 10.5573/ JSTS. 2016.16.5.564
12. Gower J. E., Wagstaff S. S. Square form factorization, Mathematics of Computation, 2008, Vol. 77, No. 261, pp. 551–588.
13. Meulenaer de G. Gosset F., Dormale de G. M., Quisquater J. J. Integer factorization based on elliptic curve method: towards better exploitation of reconfigurable hard-ware, *Field-Programmable Custom Computing Machines: IEEE Symposium FCCM, Napa, California, 23–25 April 2007, proceedings.* Published by IEEE Computer Society, 2007, pp. 197–206.
14. Ishmukhametov S. T. Metody faktorizatsii natural'nykh chisel: uchebnoye posobiye. Kazan', Kazan'skiy universitet, 2011, 190 p.
15. Prots'ko I. O., Gryschuk O. V. Computation factorization of number at chip multithreading mode, *Radio, Electronics, Computer Science, Control,* 2019, No. 3, pp. 117–122. DOI 10.15588/1607-3274-2019-3-13
16. Prots'ko I.O. Binarno-bitovi alhorytmy: prohramuvannya i zastosuvannya. L'viv, Triada plyus, 2020, 120 p.
17. Library CTPL [Electronic resource]. Access mode: https://github.com/vit-vit/CTPL.
18. Kirk D. B., Hwu Wen-mei W. Programming Massively Parallel Processors: A Hands-on Approach, Second Edition, Waltham. Morgan Kaufmann, 2012, 518 p.
19. C++ AMP: Language and Programming Model, Version 1.0, August 2012 [Electronic resource]. Access mode: https://download.microsoft.com

Received 21.06.2021.
Accepted 01.10.2021.

УДК 519.688: 004.75:004.421

АНАЛІЗ ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ БАГАТОПОТОКОВИХ ОБЧИСЛЕНЬ ДЛЯ ФАКТОРИЗАЦІЇ ЧИСЕЛ ЗА БІНАРНИМ АЛГОРИТМОМ

Процько І. О. – д-р техн. наук, доцент кафедри автоматизованих систем управління Національного університету «Львівська політехніка», Львів, Україна.
Рикмас Р. В. – розробник, ТзОВ «Uniservice», Львів, Україна.

АНОТАЦІЯ

Актуальність. Забезпечення високої швидкодії обчислення комп'ютерними системами розкладу цілочисельного значення на прості множники вимагає розробки ефективних алгоритмічних методів з використанням обчислювальних технологій. Швидке обчислення факторизації чисел використовується в таких застосуваннях, як захист інформаційних даних, в алгоритмах дискретних перетворень для переходу від одного до багатовимірних обчислень та інших.

Метою роботи є аналіз впровадження технологій багатопотокового обчислення факторизації цілочисельного значення за бінарним алгоритмом методу пробних ділень з використанням комп'ютерних систем з багатоядерними процесорами та графічними прискорювачами.

Метод. Бінарний алгоритм пробних ділень, що використовує залишки кожного розряду двійкового подання числа, для здійснення паралельної перевірки подільності на прості множники для канонічного розкладання числа.

Результати. Проведено аналіз та порівняння програмних реалізацій багатопотокових обчислень факторизації числа за двійковим алгоритмом із використанням технологій гіперпоточності, AMP C++, CUDA в комп'ютерних системах з багатоядерними процесорами та графічними прискорювачами. Проаналізовано результати процесу факторизації чисел для багатопоточних обчислювальних технологій, що використовують однотиповий алгоритм для функції паралельного ядра.

Висновки. При дослідженні реалізації розкладання чисел за бінарним алгоритмом у багатопотоковому режимі найбільш ефективно виконується технологія гіперпоточних обчислень із використанням багатоядерних процесорів. Гетерогенні обчислення за допомогою технологій AMP C++ або CUDA на комп'ютерних системах та графічних прискорювачах вимагають врахування особливостей мікроархітектури графічного процесора для паралельного виконання функцій ядра.

КЛЮЧОВІ СЛОВА: розкладання на множники, прості множники, багатопотоковість, гетерогенні обчислення, паралельні обчислення, залишок.

УДК 519.688: 004.75:004.421

АНАЛИЗ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ МНОГОПОТОЧНЫХ ВЫЧИСЛЕНИЙ ДЛЯ ФАКТОРИЗАЦИИ ЧИСЕЛ ЗА БИНАРНЫМ АЛГОРИТМОМ

Процько И. Е. – д-р техн. наук, доцент кафедры автоматизированных систем управления Национального университета «Львовская политехника», Украина.
Рыкмас Р. В. – разработчик, ТзОВ «Uniservice», Львов, Украина.

АННОТАЦИЯ

Актуальность. Обеспечение высокого быстродействия вычисления компьютерными системами разложения целочисленного значения на простые множители требует разработки эффективных алгоритмических методов с использованием вычислительных технологий. Быстрое вычисление факторизации чисел используется в таких приложениях, как защита ин-

DOI 10.15588/1607-3274-2021-4-11

формационных данных, в алгоритмах дискретных преобразований для перехода от одного к многомерным вычислений и других.

Целью работы является анализ внедрения технологий многопоточного вычисления факторизации целочисленного значения за бинарным алгоритмом метода пробных делений с использованием компьютерных систем с многоядерными процессорами и графическими ускорителями.

Метод. Бинарный алгоритм пробных делений, что использует остатки каждого разряда двоичного представления числа, для осуществления параллельной проверки делимости на простые множители для канонического разложения числа.

Результаты. Проведен анализ и сравнение программных реализаций многопоточных вычислений факторизации числа за двоичным алгоритмом с использованием технологий гиперпоточности, AMP C++, CUDA в компьютерных системах с многоядерными процессорами и графическими ускорителями. Проанализированы результаты процесса факторизации чисел для многопоточных вычислительных технологий, использующих однотипные алгоритм для функции параллельного ядра.

Выводы. При исследовании реализации разложения чисел за бинарным алгоритмом в многопоточном режиме наиболее эффективно выполняется технология гиперпоточных вычислений с использованием многоядерных процессоров. Гетерогенные вычисления с помощью технологий AMP C++ или CUDA на компьютерных системах и графических ускорителях требуют учета особенностей микроархитектуры графического процессора для выполнения параллельных вычислений функций ядра.

КЛЮЧЕВЫЕ СЛОВА: разложение на множители, простые множители, многопоточность, гетерогенные вычисления, параллельные вычисления, остаток.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Knuth D. E. The art of computer programming. 3-ed. / D. E. Knuth. – Vol. 1: Seminumerical Algorithms. – Menlo Park, California : Addison-Wesley, 1998. – 762 p.
2. McClellan J. H. Number Theory in Digital Signal Processing /J. H. McClellan, C. M. Rader. – Englewood Clis, NJ: Prentice-Hall, 1979. – 276 p.
3. Stalling W. Cryptography and Network Security: Principles and Practice. 6-ed./ W. Stalling. – Upper Saddle River, NJ : Prentice Hall, 2013. – 672 p.
4. Samuel S. The Joy of Factoring / S. Samuel, Jr. Wagstaff. – Providence, RI: American Mathematical Society, 2013. – P. 138–141.
5. Brent R. P. Some parallel algorithms for integer factorisation / R. P. Brent // European Conference on Parallel Processing, Toulouse, 1999: Part of the Lecture Notes in Computer Science book series, Berlin : Springing-Verlag, 1999. – Vol. 1685. – P. 1–22.
6. Hindriksen V. How expensive is an operation on a CPU [Electronic resource] / V. Hindriksen. – Access mode: https://streamcomputing.eu/blog/2012-07-16/how-expensive-is-an-operation-on-a-cpu
7. A multithreaded bound varying chaotic firefly algorithm for prime factorization / [M. Mishra, U. Chaturvedi, K. Saibal, S. K. Pal] // Advance Computing Conference: 4th IEEE International conference, Gurgaon, India. February 2014: proceedings. – Published by IEEE Computer Society, 2014. – P. 1322–1325. DOI: 10.1109/IAdCC. 2014.6779518
8. Макаренко А. В. Параллельная реализация и сравнительный анализ алгоритмов факторизации с распределенной памятью [Електронний ресурс] / А. В. Макаренко, А. В. Пыхтеев, С. С. Ефимов. – Режим доступу: http://cyberleninka.ru/article/n/parallelnaya-realizatsiya-i-sravnitelnyy-analiz-algoritmov-faktori-zatsii-v-sistemah-s-raspredelyonnoy-pamyatyu
9. Huang L. New prime factorization algorithm and its parallel computing strategy / L. Huang // Information Technologies in Education and Learning: International Conference (ICITEL 2015), Atlantis, 2015: proceedings. – Published by Atlantis Press, 2015. – P. 1–4.
10. Koundinya A. K. Performance Analysis of parallel Pollard's Rho algorithm / A. K. Koundinya, G. Harish, N. K. Srinath et al. // International Journal of Computer Science & Information Technology (IJCSIT). – 2011. – Vol. 5, No. 2. – P. 157–163. DOI: 10.5121/ijcsit.2013.5214
11. Kim D. K. Design and analysis of efficient parallel hardware prime generators / [D. K. Kim, P. Choi, M.-K. Lee et al.] // Journal of semiconductor technology and science. – 2016. – Vol. 16, No. 5. – P. 564–581. DOI: 10.5573/ JSTS. 2016.16.5.564
12. Gower J. E. Square form factorization / J. E. Gower, S. S. Wagstaff // Mathematics of Computation. – 2008. – Vol. 77, No. 261. – P. 551–588.
13. Meulenaer G. Integer factorization based on elliptic curve method: towards better exploitation of reconfigurable hardware / [G. de Meulenaer, F. Gosset, G. M. de Dormale, J. J. Quisquater] // Field-Programmable Custom Computing Machines: IEEE Symposium FCCM, Napa, California, 23–25 April 2007: proceedings. – Published by IEEE Computer Society, 2007. – P. 197–206.
14. Ишмухаметов Ш. Т. Методы факторизации натуральных чисел: учебное пособие / Ш. Т. Ишмухаметов. – Казань : Казанький университет, 2011. – 190 с.
15. Процько І. О. Обчислення факторизації числа в мультипотоковому режимі на кристалі / І. О. Процько, О. В. Грищук // Радіоелектроніка, інформатика, управління. – 2019. – № 3. – С. 117–122. DOI 10.15588/1607-3274-2019-3-13
16. Процько І. О. Бінарно-бітові алгоритми: програмування і застосування : навчальний посібник / І. О. Процько. – Львів : Тріада плюс, 2020. – 120 с.
17. Бібліотека CTPL [Електронний ресурс]. – Режим доступу: https://github.com/vit-vit/CTPL
18. Kirk D. B. Programming Massively Parallel Processors: A Hands-on Approach, Second Edition / David B. Kirk, Wenmei W. Hwu. – Waltham : Morgan Kaufmann, 2012. – 518 p.
19. C++ AMP : Language and Programming Model — Version 1.0 : August 2012 [Електронний ресурс]. – Режим доступу: https://download.microsoft.com

DOI 10.15588/1607-3274-2021-4-11

UDC 004.412:519.237.5

A NONLINEAR REGRESSION MODEL TO ESTIMATE THE SIZE OF WEB APPS CREATED USING THE CAKEPHP FRAMEWORK

Prykhodko S. B. – Dr. Sc., Professor, Head of the Department of Software of Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

Shutko I. S. – Post-graduate student of the Department of Software of Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

Prykhodko A. S. – Student of the Department of Software of Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

ABSTRACT

Context. The problem of estimating the software size in the early stage of a software project is important because a software size estimate is used for predicting the software development effort, including Web apps created using the CakePHP framework. The object of the study is the process of estimating the size of Web apps created using the CakePHP framework. The subject of the study is the nonlinear regression models to estimate the size of Web apps created using the CakePHP framework.

Objective. The goal of the work is the building the nonlinear regression model with three predictors for estimating the size of Web apps created using the CakePHP framework on the basis of the Box-Cox four-variate normalizing transformation to increase the confidence in early size estimation of these apps.

Method. The model, confidence and prediction intervals of multiply nonlinear regression to estimate the size of Web apps created using the CakePHP framework are constructed based on the Box-Cox multivariate normalizing transformation for non-Gaussian data with the help of appropriate techniques. The techniques to build the models, confidence, and prediction intervals of nonlinear regressions are based on the multiple nonlinear regression analysis using the multivariate normalizing transformations. The techniques allow taking into account the correlation between dependent and independent variables in the case of normalization of multivariate non-Gaussian data. In general, this leads to a reduction of the mean magnitude of relative error, the widths of the confidence, and prediction intervals in comparison with nonlinear models constructed using univariate normalizing transformations.

Results. Comparison of the constructed model with the nonlinear regression models based on the decimal logarithm and the Box-Cox univariate transformation has been performed.

Conclusions. The nonlinear regression model with three predictors to estimate the size of Web apps created using the CakePHP framework is constructed on the basis of the Box-Cox four-variate transformation. This model, in comparison with other nonlinear regression models, has a larger multiple coefficient of determination, a smaller value of the mean magnitude of relative error and smaller widths of the confidence and prediction intervals. The prospects for further research may include the application of other multivariate normalizing transformations and data sets to construct the nonlinear regression model to estimate the size of Web apps created using the other frameworks.

KEYWORDS: software size estimation, Web app, nonlinear regression model, normalizing transformation, non-Gaussian data.

ABBREVIATIONS

DIT is a depth of inheritance tree;
KLOC is a thousand lines of code;
LB is a lower bound;
MMRE is a mean magnitude of relative error;
MRE is a magnitude of relative error;
PHP is a hypertext preprocessor;
PRED is a percentage of prediction;
SMD is a squared Mahalanobis distance;
UB is an upper bound.

NOMENCLATURE

$\hat{\mathbf{b}}$ is a estimator for vector of linear regression equation parameters, $\mathbf{b} = \{b_1, b_2, \ldots, b_k\}^T$;

$\hat{b}_i$ is a estimator for the *i*-th parameter of linear regression equation;

k is a number of predictors (independent variables);

N is a number of data points;

$N(0,1)$ is a Gaussian distribution with zero mathematical expectation and unit variance;

P is a non-Gaussian random vector, $\mathbf{P} = \{Y, X_1, X_2, \ldots, X_k\}^T$;

R^2 is a multiple coefficient of determination;

$\mathbf{S}_N$ is a sample covariance matrix, $\mathbf{S}_N = [S_{ij}]$;

T is a Gaussian random vector, $\mathbf{T} = \{Z_Y, Z_1, Z_2, \ldots, Z_k\}^T$;

$t_{\alpha/2,\nu}$ is a quantile of student's *t*-distribution with ν degrees of freedom and $\alpha/2$ significance level;

X_1 is a number of classes;

X_2 is a average number of methods per class;

X_3 is a DIT mean value per class;

Y is an actual software size in KLOC;

Z_j is a *j*-th standard Gaussian variable that is obtained by transforming variable X_j, $Z_j \sim N(0,1)$, $j = 1,2,\ldots,k$;

Z_Y is a standard Gaussian variable that is obtained by transforming variable Y, $Z_Y \sim N(0,1)$;

$\bar{Z}_Y$ is a sample mean of the Z_Y values;

DOI 10.15588/1607-3274-2021-4-12

$\hat{Z}_Y$ is a prediction result by linear regression equation for normalized data;

α is a significance level;

β_1 is a multivariate skewness;

β_2 is a multivariate kurtosis;

ε is a Gaussian random variable which defines residuals, $\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$;

ν is a number of degrees of freedom;

σ_ε is a standard deviation of ε;

$\boldsymbol{\psi}$ is a vector of multivariate normalizing transformation, $\boldsymbol{\psi} = \{\psi_Y, \psi_1, \psi_2, \ldots, \psi_k\}^T$;

$\mathbf{z}_X^+$ is a vector with components $Z_{1_i} - \overline{Z}_1$, $Z_{2_i} - \overline{Z}_2$, $\ldots$, $Z_{k_i} - \overline{Z}_k$ for *i*-row.

INTRODUCTION

Early software size estimation is one of the project managers' significant problems in evaluating software development efforts using mathematical models like COCOMO II [1]. Now many Web apps are created using the CakePHP framework making app development faster. However, today some software size estimation models that base metrics that can be measured from the class diagram are known [2–10]. There are only some regression equations and models, both linear [3, 4] and nonlinear [7, 8] ones, for estimating the software size of information open-source PHP-based systems. This demands the construction of the models for early size estimation of Web apps created using the CakePHP framework.

The object of study is the process of estimating the size of Web apps created using the CakePHP framework.

The subject of study is the regression models to estimate the size of Web apps created using the CakePHP framework.

The purpose of the work is to increase the confidence in early size estimation of Web apps created using the CakePHP framework.

1 PROBLEM STATEMENT

Suppose given the original sample as the four-dimensional non-Gaussian data set: actual software size in the thousand lines of code (KLOC) *Y*, the total number of classes X_1, the average number of methods per class X_2, the average of Depth of Inheritance Tree (DIT) per class X_3 in class diagram from *N* Web apps. Suppose that there are bijective five-variate normalizing transformation of non-Gaussian random vector $\mathbf{P} = \{Y, X_1, X_2, X_3\}^T$ to Gaussian random vector $\mathbf{T} = \{Z_Y, Z_1, Z_2, Z_3\}^T$ is given by

$$\mathbf{T} = \boldsymbol{\psi}(\mathbf{P}) \qquad (1)$$

and the inverse transformation for (1)

$$\mathbf{P} = \boldsymbol{\psi}^{-1}(\mathbf{T}). \qquad (2)$$

It is required to build the nonlinear regression model in the form $Y = Y(X_1, X_2, X_3, \varepsilon)$ based on the transformations (1) and (2).

2 REVIEW OF THE LITERATURE

In paper [3] the linear regression equations were proposed for estimating the software size of open-source PHP- and Java-based information systems. These equations are constructed on the basis of three metrics that can be measured from conceptual data model based a class diagram: a total number of classes, a total number of relationships, and an average number of attributes per class. However, there are four basic assumptions that justify the use of linear regression models, one of which is normality of the error distribution [11–13]. But this assumption is valid only in particular cases. Therefore, in papers [7] and [8], the nonlinear regression models were constructed using the same above metrics for estimating the software size of PHP- and Java-based information systems, respectively. But the size of Web apps may depend on other metrics. That is why in [9] the nonlinear regression model was constructed for estimating the size of Web apps created using the Laravel framework. This model depends on three factors (predictors), namely the total number of classes, the average number of methods per class, and the sum of average afferent coupling and average efferent coupling per class. However, the size of Web apps created using the CakePHP framework may depend on other metrics, and the model might have other parameters. This leads to the need of building the nonlinear regression model to estimate the size of Web apps created using the CakePHP framework.

A normalizing transformation is often a good way to construct nonlinear regression models [14–19]. According to [16], transformations are made for essentially four purposes, two of which are: firstly, to obtain approximate normality for the distribution of the error term (residuals), secondly, to transform the response and/or the predictor in such a way that the strength of the linear relationship between new variables (normalized variables) is better than the linear relationship between initial dependent and independent variables.

Well-known techniques to construct the nonlinear regression models are based on the univariate normalizing transformations (such as, the decimal logarithm, Box-Cox transformation), and do not take into account the correlation between dependent and independent variables. The use of such univariate normalizing transformations for constructing the nonlinear regression models does not always lead to good normality and linear relationship between normalized variables. This leads to the need to apply multivariate normalizing transformations.

In [7] the techniques to build the models, confidence, and prediction intervals of nonlinear regressions based on the bijective multivariate normalizing transformations were proposed. However, according to [20], there may be

DOI 10.15588/1607-3274-2021-4-12

data sets for which the results of creating nonlinear regression models depend on, firstly, which normalizing transformation is used, univariate, or multivariate, and, secondly, are there any outliers in the data set. That is why, in [20] the technique to build nonlinear regression models based on the multivariate normalizing transformations and prediction intervals was considered. In this technique the prediction intervals of nonlinear regressions are used to detect the outliers in the process of constructing the nonlinear regression models. We apply the above technique for building the nonlinear regression model with three predictors to estimate the size of Web apps created using the CakePHP framework.

3 MATERIALS AND METHODS

According to [20], the technique to build nonlinear regression models based on the multivariate normalizing transformations and prediction intervals consist of four steps. In the first step, multivariate non-Gaussian data are normalized using a multivariate normalizing transformation (1). To do this, we use the Box-Cox multivariate transformation.

In the second step, the nonlinear regression model is constructed based on the multivariate normalizing transformation (1) as in [7]. Before that, we first determine whether one data point of a multivariate non-Gaussian data set is a multidimensional outlier. To do this, we apply the statistical technique based on the normalizing transformations and the Mahalanobis squared distance (MSD) as in [7, 8, 20]. If there is a multidimensional outlier in a multivariate non-Gaussian data set, then we discard the one, and return to step 1, else build the linear regression model for normalized data based on the transformation (1) in the form

$$Z_Y = \hat{Z}_Y + \varepsilon = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \ldots + \hat{b}_k Z_k + \varepsilon , \qquad (3)$$

ε is a Gaussian random variable which defines residuals, $\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$.

After that the nonlinear regression model is built on the basis of the linear regression model (3) and the transformations (1) and (2) as

$$Y = \psi_Y^{-1}\left(\hat{Z}_Y + \varepsilon\right) . \qquad (4)$$

In the third step, the prediction interval of nonlinear regression is defined [20]

$$\psi_Y^{-1}\left(\hat{Z}_Y \pm t_{\alpha/2,\nu} S_{Z_Y} \left\{1 + \frac{1}{N} + \left(\mathbf{z}_X^+\right)^T \mathbf{S}_Z^{-1} \left(\mathbf{z}_X^+\right)\right\}^{1/2}\right) , \qquad (5)$$

where $t_{\alpha/2,\nu}$ is a student's *t*-distribution quantile with $\alpha/2$ significance level and ν degrees of freedom; $\nu = N - k - 1$; k is a number of independent variables (in our case, k is 3); $\mathbf{z}_X^+$ is a vector with components $Z_{1_i} - \bar{Z}_1$, $Z_{2_i} - \bar{Z}_2$, …, $Z_{k_i} - \bar{Z}_k$ for *i*-row; $\bar{Z}_j = \frac{1}{N}\sum_{i=1}^{N} Z_{j_i}$, $j = 1,2,\ldots,k$; $S_{Z_Y}^2 = \frac{1}{\nu}\sum_{i=1}^{N}\left(Z_{Y_i} - \hat{Z}_{Y_i}\right)^2$, $\nu = N - k - 1$; $\mathbf{S}_Z$ is a $k \times k$ matrix

$$\mathbf{S}_Z = \begin{pmatrix} S_{Z_1Z_1} & S_{Z_1Z_2} & \cdots & S_{Z_1Z_k} \\ S_{Z_1Z_2} & S_{Z_2Z_2} & \cdots & S_{Z_2Z_k} \\ \cdots & \cdots & \cdots & \cdots \\ S_{Z_1Z_k} & S_{Z_2Z_k} & \cdots & S_{Z_kZ_k} \end{pmatrix} . \qquad (6)$$

In (6) $S_{Z_qZ_r} = \sum_{i=1}^{N}\left[Z_{q_i} - \bar{Z}_q\right]\left[Z_{r_i} - \bar{Z}_r\right]$, $q, r = 1,2,\ldots,k$.

In the fourth step, we check if there are data that exit the bounds of the prediction interval. And if we detect the outliers, we discard them and repeat all the steps starting with the first for new data without outliers, else nonlinear regression model construction is completed.

We constructed a nonlinear regression model to estimate the size of Web apps created using the CakePHP framework by the above technique from 38 apps hosted on GitHub (https://github.com). The data set was obtained using the PhpMetrics tool (https://phpmetrics.org/) around following variables: actual software size in the thousand lines of code (KLOC) *Y*, the total number of classes X_1, the average number of methods per class X_2, and the average of Depth of Inheritance Tree (DIT) per class X_3. Table 1 contains that data set. We chose the above predictors X_1, X_2, and X_3 for two reasons. Firstly, these predictors can be obtained from the class diagram, and, secondly, there is no multicollinearity between these predictors according to [21, 22] since variance inflation factors for predictors X_1, X_2, and X_3 are equal to 1.08, 1.03, and 1.11, respectively.

We checked the four-dimensional data from Table 1 for multivariate outliers. But before that, we tested the normality of multivariate data from Table I because well-known statistical methods (for example, multivariate outlier detection based on the squared Mahalanobis distance (SMD)) are used to detect outliers in multivariate data under the assumption that the data is described by a Gaussian distribution [18, 23]. We applied a multivariate normality test proposed by Mardia and based on measures of the multivariate skewness β_1 and kurtosis β_2 [24, 25]. According to this test, the distribution of four-dimensional data from Table I is not Gaussian since the test statistic for multivariate skewness $N\beta_1/6$ of this data, which equals to 161.59, is greater than the quantile of the Chi-Square distribution, which is 40.00 for 20 degrees of freedom and 0.005 significance level.

DOI 10.15588/1607-3274-2021-4-12

e-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2021. № 4
p-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2021. № 4

Similarly, the test statistic for multivariate kurtosis β_2, which equals to 44.38, is greater than the value of the Gaussian distribution quantile, which is 29.79 for 24 mean, 5.05 variance, and 0.005 significance level. Because we used the statistical technique [26] to detect multivariate outliers in the four-dimensional non-Gaussian data from Table I based on the multivariate normalizing transformations and the SMD for normalized data. To normalize the data from Table 1, we applied the four-variate Box-Cox transformation with components [18]

$$Z_j = x(\lambda_j) = \begin{cases} \left(X_j^{\lambda_j} - 1\right) / \lambda_j, & \text{if } \lambda_j \neq 0; \\ \ln(X_j), & \text{if } \lambda_j = 0. \end{cases} \quad (7)$$

Here Z_j is a Gaussian variable; λ_j is a parameter of the Box-Cox transformation, $j = 1,2,3$. The variable Z_Y is defined analogously (7) with the only difference that instead of Z_j, X_j, and λ_j should be put respectively Z_Y, Y, and λ_Y.

Table 1 – The data set and SMD values

No	Y	X_1	X_2	X_3	SMD	SMD_Z
1	0.448	4	4.75	1.40	2.07	4.02
2	7.846	90	3.86	1.71	0.08	1.01
3	4.345	42	4.19	1.81	0.11	0.63
4	2.717	60	2.20	2.11	3.84	6.24
5	2.954	45	3.13	2.13	2.34	3.42
6	1.717	26	2.54	1.44	2.09	5.11
7	0.212	1	7.00	2.00	4.31	6.07
8	1.149	14	3.86	1.73	0.30	0.56
9	0.477	8	2.63	2.20	3.84	3.55
10	61.269	487	4.93	1.79	19.41	5.77
11	0.124	2	2.50	2.00	2.60	4.23
12	0.358	2	3.50	2.00	1.22	**16.19**
13	0.349	8	2.25	1.38	3.15	5.86
14	3.486	29	3.62	2.00	1.05	4.67
15	1.538	11	5.82	1.80	1.51	1.71
16	0.365	4	4.00	1.50	1.08	2.53
17	2.332	23	4.00	1.79	0.21	0.26
18	24.347	328	3.59	1.24	7.78	4.59
19	12.433	66	6.52	1.79	3.08	3.82
20	0.948	10	4.30	1.78	0.28	0.84
21	1.826	20	4.15	1.60	0.47	0.28
22	0.882	9	3.56	1.60	0.72	1.52
23	3.567	23	6.83	1.24	7.30	6.55
24	3.735	49	3.57	1.70	0.22	0.53
25	47.61	289	5.33	1.33	18.86	7.55
26	4.029	58	3.38	1.54	0.61	0.72
27	3.5	40	4.18	1.63	0.23	0.55
28	0.131	3	2.00	1.67	2.73	6.61
29	2.227	23	4.22	1.46	1.11	0.83
30	0.521	2	8.50	3.00	**20.51**	12.06
31	1.494	22	3.41	1.35	1.94	2.30
32	0.906	5	7.80	1.60	7.59	5.31
33	0.471	5	3.80	2.00	0.97	1.26
34	24.04	303	3.67	1.75	5.34	3.74
35	2.681	23	4.78	1.71	0.39	0.39
36	20.327	308	3.94	1.20	9.35	6.58
37	24.66	367	3.36	1.64	11.55	4.01
38	0.142	3	3.00	2.00	1.76	10.13

The parameter estimates of the four-variate Box-Cox transformation for the data from Table 1 are calculated by the maximum likelihood method according to [18] and are $\hat{\lambda}_Y = -0.004265$, $\hat{\lambda}_1 = 0.012715$, $\hat{\lambda}_2 = -0.267106$, $\hat{\lambda}_3 = -0.610377$.

Table 1 contains the SMD for normalized data (SMD_Z), which is transformed using the four-variate BCT. The SMD_Z values from Table 1 indicate there is one multivariate outlier in four-dimensional non-Gaussian data since the SMD_Z value for row 12 is greater than the quantile of the Chi-Square distribution, which equals to 14.86 for the 0.005 significance level. In Table 1, the row numbers with the outliers are highlighted in bold. Also, Table 1 contains the SMD values for data without normalization. Note, for data without normalization, row 30 is the multivariate outlier since the SMD value for row 30 is greater than the quantile of the Chi-Square distribution for the 0.005 significance level.

Then, we built the linear regression model for data from Table 1 in the form

$$Y = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \varepsilon, \quad (8)$$

where $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$, and $\hat{b}_3$ are parameter estimates, $\hat{b}_0 = -4.4155$, $\hat{b}_1 = 0.09974$, $\hat{b}_2 = 1.02683$, $\hat{b}_3 = -0.05258$, ε is a error term, which must be a Gaussian random variable to describe residuals, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, σ_ε is a standard deviation with the estimate $\hat{\sigma}_\varepsilon$ of 4.764.

According to [14], for error term in a linear regression model "the assumption of normality may be checked by examining the residuals." The null hypothesis H_0 that the observed frequency distribution of the ε values in (8) is the same as the normal distribution (there is no difference between the distributions) was checked by the Pearson Chi-Squared test. We rejected the null hypothesis H_0 with the 0.05 significance level since the χ^2 test statistic, which equals to 30.06, surpasses the critical value from the Chi-Squared distribution that is 7.81 for 0.05 significance level and 3 degrees of freedom. That is why there is no justification for the use of a linear regression model for mathematical modeling of the size of Web apps created using the CakePHP framework because the error distribution is not the Gaussian one. This creates a need to apply the nonlinear regression models in our case.

The nonlinear regression model with three predictors for estimating the size of Web apps created using the CakePHP framework is constructed based on the four-variate Box-Cox transformation for 37 data rows from Table 1 (without row 12) according to [20] and has the form [27]

$$Y = \left[\hat{\lambda}_Y \left(\hat{Z}_Y + \varepsilon\right) + 1\right]^{1/\hat{\lambda}_Y}, \quad (9)$$

DOI 10.15588/1607-3274-2021-4-12

where ε is a Gaussian random variable, $\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$, with the estimate $\hat{\sigma}_\varepsilon$ of 0.1286; $\hat{Z}_Y$ is a prediction result by the linear regression equation $\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3$ for normalized data, which are transformed by the four-variate Box-Cox transformation with components (7); $\hat{b}_0 = -4.27326$, $\hat{b}_1 = 1.06624$, $\hat{b}_2 = 1.17959$, $\hat{b}_3 = 0.37559$, $\hat{\lambda}_Y = -0.02762$, $\hat{\lambda}_1 = -0.02795$, $\hat{\lambda}_2 = 0.020293$, $\hat{\lambda}_3 = -0.67526$.

According to [20], after constructing a model (9), we have to find the nonlinear regression prediction interval by (5).

In the second iteration, for the data normalized by the four-variate Box-Cox transformation from 37 Web apps (without row 12), the matrix (6) is following

$$\mathbf{S}_Z = \begin{pmatrix} 79.91 & -1.12 & -2.60 \\ -1.12 & 4.50 & 0.049 \\ -2.60 & 0.049 & 0.594 \end{pmatrix}.$$

As we observe, there are two values of *Y* for Web applications 14 and 38 that are out of the prediction intervals computed by (5) for a significance level of 0.05. In Table 2, we marked the prediction interval lower (LB) and upper (UB) bounds calculated in the second iteration as LB_2, and UB_2.

Next, we erased data in Web applications 14 and 38. After that, we used data from the remaining 35 apps to construct the model. In nonlinear regression model (9) with parameters' estimates acquired from 35 rows of data, it appeared that two values of *Y* for rows 6 and 22 exceed the prediction interval. After five iterations, we saved 32 Web applications from Table 1 (excluding rows 6, 12, 14, 22, 25, and 38). There were no outliers in the fifth iteration. We completed the stages' iterations, and constructed nonlinear regression model (9) with 32 Web applications data. In Table 2, we marked the prediction interval LB and UB calculated in the third and fifth iterations as LB_3 and UB_3, and as LB_5 and UB_5, respectively. We highlighted the row numbers with the data outliers in bold, and a dash (–) shows the exception of the relevant numbers of data at the corresponding iteration.

In the fifth iteration, the parameter estimates for the model (9) constructed by the four-variate Box-Cox transformation from 32 Web apps are $\hat{\lambda}_Y = -0.05722$, $\hat{\lambda}_1 = -0.03731$, $\hat{\lambda}_2 = -0.10586$, $\hat{\lambda}_3 = -0.74223$, $\hat{b}_0 = -4.43109$, $\hat{b}_1 = 1.05416$, $\hat{b}_2 = 1.39398$, $\hat{b}_3 = 0.51635$, the estimate $\hat{\sigma}_\varepsilon$ is 0.05174.

In this case, the matrix (6) is following

$$\mathbf{S}_Z = \begin{pmatrix} 66.35 & -1.72 & -1.88 \\ -1.72 & 2.86 & 0.0605 \\ -1.88 & 0.0605 & 0.485 \end{pmatrix}.$$

We checked the multivariate normality of 32 rows of normalized data from Table I in the fifth iteration with a test proposed by Mardia [24].

Table 2 – LB and UB of nonlinear regression prediction intervals in various iterations

No	LB_2	UB_2	LB_3	UB_3	LB_5	UB_5
1	0.327	0.578	0.370	0.542	0.389	0.488
2	5.943	10.764	6.270	9.713	6.711	8.686
3	3.153	5.635	3.344	5.083	3.593	4.602
4	2.054	3.783	2.127	3.287	2.267	2.936
5	2.421	4.386	2.550	3.902	2.768	3.560
6	0.993	1.766	1.062	1.582	–	–
7	0.140	0.247	0.161	0.234	0.177	0.221
8	0.951	1.661	1.030	1.516	1.096	1.379
9	0.363	0.640	0.406	0.596	0.435	0.547
10	43.538	84.434	49.177	83.162	53.748	73.141
11	0.086	0.150	0.106	0.151	0.112	0.139
13	0.268	0.476	0.305	0.447	0.309	0.390
14	1.864	3.320	–	–	–	–
15	1.240	2.199	1.336	1.992	1.435	1.819
16	0.274	0.478	0.313	0.453	0.330	0.411
17	1.637	2.886	1.750	2.610	1.873	2.373
18	17.263	32.698	18.271	29.783	18.985	25.330
19	8.599	15.940	9.276	14.712	9.962	13.073
20	0.781	1.363	0.852	1.250	0.912	1.145
21	1.448	2.546	1.553	2.308	1.644	2.079
22	0.545	0.948	0.602	0.878	–	–
23	2.797	5.212	2.988	4.675	3.085	4.022
24	2.958	5.274	3.120	4.729	3.320	4.245
25	26.354	50.163	29.090	48.006	–	–
26	3.170	5.667	3.332	5.063	3.504	4.488
27	2.917	5.189	3.094	4.681	3.286	4.198
28	0.095	0.167	0.115	0.166	0.118	0.148
29	1.649	2.921	1.766	2.641	1.852	2.352
30	0.377	0.705	0.412	0.627	0.465	0.598
31	1.182	2.103	1.271	1.898	1.316	1.671
32	0.768	1.387	0.830	1.245	0.882	1.122
33	0.347	0.603	0.391	0.566	0.422	0.526
34	18.389	34.579	19.637	31.872	21.267	28.262
35	2.014	3.567	2.152	3.229	2.299	2.923
36	18.081	34.434	19.288	31.626	19.989	26.765
37	19.411	36.531	20.576	33.428	22.101	29.389
38	0.159	0.275	–	–	–	–

According to Mardia's test, the distribution of 32 rows of normalized data from Table I (excluding rows 6, 12, 14, 22, 25, and 38) is Gaussian since the test statistic for multivariate skewness $N\beta_1/6$ of this data, which equals to 14.05, is less than the quantile of the Chi-Square distribution, which is 40.00 for 20 degrees of freedom and 0.005 significance level. Similarly, the test statistic for multivariate kurtosis β_2, which equals to 23.80, is less than the quantile of the Gaussian distribution, which is 30.31 for 24 mean, 6.0 variance, and 0.005 significance level.

DOI 10.15588/1607-3274-2021-4-12

4 EXPERIMENTS

For comparison of the model (9) with other nonlinear regression models with three predictors, two nonlinear regression models are built based on 32 data rows from Table 1 (without rows 6, 12, 14, 22, 25, and 38) using the Box-Cox univariate transformation and the decimal logarithm univariate one.

The nonlinear regression model based on the linear regression model (3) for the normalized data and the decimal logarithm univariate transformation has the form

$$Y = 10^{\varepsilon+\hat{b}_0} X_1^{\hat{b}_1} X_2^{\hat{b}_2} X_3^{\hat{b}_3}, \qquad (10)$$

where the estimators for parameters are: $\hat{b}_0 = -1.80484$, $\hat{b}_1 = 0.981865$, $\hat{b}_2 = 1.186738$, $\hat{b}_3 = 0.319789$. The estimate $\hat{\sigma}_\varepsilon$ is 0.029761.

The nonlinear regression model based on the Box-Cox univariate transformation is analogously (9) with the only difference that the data for variables are normalized by the Box-Cox univariate transformation using the maximum likelihood method [18]. The estimators for parameters of the Box-Cox univariate transformation for each from variables Y, X_1, X_2, and X_3 are $\hat{\lambda}_Y = -0.054396$, $\hat{\lambda}_1 = -0.04861$, $\hat{\lambda}_2 = -0.15995$, $\hat{\lambda}_3 = -0.70493$. The parameter estimators of the linear regression model for normalized data by the Box-Cox univariate transformation are $\hat{b}_0 = -4.55626$, $\hat{b}_1 = 1.09337$, $\hat{b}_2 = 1.50882$, $\hat{b}_3 = 0.529486$. The estimate $\hat{\sigma}_\varepsilon$ is 0.055974.

The computer program implementing the constructed models (9) and (10) was developed to conduct experiments. The program was written in the sci-language for the Scilab system. Scilab (http://www.scilab.org) is the free and open source software, the alternative to commercial packages for system modeling and simulation packages such as MATLAB and MATRIXx [28].

5 RESULTS

The prediction results $\hat{Y}$ of nonlinear regression models (9) and (10) for values of predictors from Table 1 (without rows 6, 12, 14, 22, 25, and 38) and values of MRE are shown in the Table 3. The prediction results by model (9) and values of MRE are shown in the Table 3 for two cases: Box-Cox univariate and four-variate normalizing transformations.

Table 3 – The prediction results and confidence intervals of multiple regressions

No	the four-variate Box-Cox transformation				Univariate transformations							
					the decimal logarithm transformation				the Box-Cox transformation			
	$\hat{Y}$	MRE	LB	UB	$\hat{Y}$	MRE	LB	UB	$\hat{Y}$	MRE	LB	UB
1	0.436	0.0274	0.419	0.454	0.433	0.0344	0.409	0.457	0.435	0.0285	0.417	0.455
2	7.631	0.0274	7.401	7.869	7.667	0.0229	7.399	7.943	7.701	0.0185	7.451	7.960
3	4.065	0.0645	3.958	4.174	4.072	0.0629	3.947	4.201	4.149	0.0450	4.032	4.270
4	2.578	0.0510	2.445	2.720	2.826	0.0400	2.643	3.021	2.592	0.0461	2.446	2.746
5	3.138	0.0622	3.012	3.269	3.247	0.0992	3.085	3.418	3.201	0.0836	3.063	3.346
7	0.197	0.0686	0.188	0.207	0.197	0.0710	0.184	0.211	0.189	0.1077	0.180	0.199
8	1.229	0.0699	1.204	1.255	1.238	0.0775	1.204	1.273	1.254	0.0918	1.226	1.283
9	0.488	0.0230	0.468	0.508	0.490	0.0262	0.462	0.519	0.493	0.0328	0.471	0.515
10	62.657	0.0227	58.937	66.626	54.581	0.1092	51.153	58.238	59.540	0.0282	55.807	63.537
11	0.125	0.0041	0.119	0.130	0.115	0.0757	0.108	0.122	0.121	0.0213	0.116	0.127
13	0.347	0.0053	0.331	0.364	0.350	0.0040	0.328	0.374	0.346	0.0097	0.328	0.364
15	1.615	0.0500	1.566	1.665	1.611	0.0474	1.549	1.675	1.642	0.0676	1.589	1.697
16	0.368	0.0083	0.356	0.381	0.361	0.0119	0.344	0.378	0.368	0.0078	0.355	0.382
17	2.107	0.0964	2.061	2.154	2.126	0.0884	2.069	2.185	2.156	0.0755	2.105	2.208
18	21.916	0.0998	20.734	23.170	22.596	0.0719	21.294	23.978	21.199	0.1293	19.984	22.493
19	11.406	0.0826	10.903	11.934	10.687	0.1404	10.152	11.250	11.475	0.0771	10.933	12.045
20	1.021	0.0774	0.999	1.045	1.021	0.0766	0.991	1.051	1.040	0.0972	1.015	1.066
21	1.848	0.0119	1.809	1.888	1.868	0.0229	1.816	1.921	1.887	0.0333	1.844	1.931
23	3.521	0.0130	3.326	3.727	3.567	0.0000	3.332	3.819	3.553	0.0040	3.343	3.777
24	3.753	0.0047	3.660	3.848	3.839	0.0280	3.725	3.958	3.819	0.0225	3.717	3.924
26	3.964	0.0161	3.858	4.073	4.114	0.0211	3.979	4.253	4.018	0.0028	3.901	4.137
27	3.712	0.0607	3.628	3.799	3.743	0.0694	3.639	3.849	3.785	0.0814	3.692	3.880
28	0.132	0.0098	0.126	0.139	0.124	0.0563	0.115	0.133	0.129	0.0174	0.122	0.136
29	2.086	0.0631	2.031	2.144	2.122	0.0470	2.050	2.197	2.128	0.0445	2.067	2.191
30	0.527	0.0115	0.494	0.563	0.558	0.0702	0.505	0.615	0.517	0.0078	0.481	0.555
31	1.482	0.0080	1.431	1.536	1.539	0.0299	1.472	1.609	1.507	0.0087	1.451	1.566
32	0.994	0.0971	0.950	1.040	1.013	0.1176	0.952	1.077	0.990	0.0926	0.943	1.040
33	0.471	0.0002	0.458	0.485	0.463	0.0166	0.445	0.482	0.475	0.0086	0.460	0.490
34	24.502	0.0192	23.361	25.702	23.957	0.0034	22.720	25.261	23.885	0.0065	22.702	25.132
35	2.591	0.0336	2.531	2.653	2.588	0.0346	2.514	2.665	2.648	0.0124	2.581	2.716
36	23.116	0.1372	21.792	24.526	23.475	0.1549	22.067	24.973	22.397	0.1018	21.034	23.854
37	25.471	0.0329	24.281	26.723	25.506	0.0343	24.174	26.912	24.616	0.0018	23.395	25.904

DOI 10.15588/1607-3274-2021-4-12

The MRE values for the model (9) based on the Box-Cox four-variate transformation are smaller than for the model (9) based on the Box-Cox univariate transformation for 17 from 32 rows of data (rows 1, 5, 7–11, 13, 15, 18, 20, 21, 24, 27, 28, 31, 33). Also, the MRE values for the model (9) based on the Box-Cox four-variate transformation are less than for the model (10) based on the decimal logarithm univariate transformation for 21 from 32 rows of data (rows 1, 5, 7–11, 16, 19, 21, 24, 26–28, 30–33, 35–37).

To evaluate the prediction accuracy of the nonlinear regression models we applied the standard metrics R^2, MMRE, and PRED(0.25). MMRE and PRED(0.25) are accepted as standard evaluations of prediction results by regression models. These metrics are applied in software engineering too [29, 30]. The acceptable values of MMRE and PRED(0.25) are not more than 0.25 and not less than 0.75 respectively. The values of R^2, MMRE and PRED(0.25) equal respectively 0.9963, 0.0425, and 1.0 for model (9) based on the Box-Cox four-variate transformation, and equal respectively 0.9961, 0.0442 and 1.0 for the model (9) based on the Box-Cox univariate transformation, and equal respectively 0.9871, 0.0552 and 1.0 for the model (10) for the decimal logarithm univariate transformation. The MMRE and R^2 values are better for the model (9) based on the Box-Cox four-variate transformation.

The confidence and prediction intervals of nonlinear regression are defined for the data from Table 1 (without rows 6, 12, 14, 22, 25, and 38). Table 3 contains the lower (LB) and upper (UB) bounds of the confidence intervals of nonlinear regressions based on the univariate and four-variate transformations respectively for 0.05 significance level. We defined the confidence intervals of the sample mean of size using (5) with the only difference that in the sum in curly brackets, there is not 1. The widths of the confidence interval of nonlinear regression based on the Box-Cox four-variate transformation are less than for nonlinear regression based on the Box-Cox univariate transformation for all 32 rows of data. Also, the widths of the confidence interval of nonlinear regression based on the Box-Cox four-variate transformation are less than for nonlinear regression based on the decimal logarithm univariate transformation for 31 from 32 rows of data (except row 10). Approximately the same results are obtained for the prediction intervals of nonlinear regressions.

Table 4 contains the lower (LB) and upper (UB) bounds of the prediction intervals of non-linear regressions based on the univariate and multivariate transformations respectively for 0.05 significance level. The widths of the prediction interval of nonlinear regression based on the Box-Cox four-variate transformation are less than for nonlinear regression based on the Box-Cox univariate transformation for all 32 rows of data. Also, the widths of the prediction interval of nonlinear regression based on the Box-Cox four-variate transformation are less for nonlinear regression based on the decimal logarithm univariate transformation for 31 from 32 rows of data (except row 10).

Note, a more significant advantage of the model (9) constructed by the four-variate Box-Cox transformation compared with the two above models based on the univariate transformations is the smaller widths of the confidence and prediction intervals. Such, the width of the nonlinear regression prediction interval for the four-variate Box-Cox transformation is less than after the univariate Box-Cox transformation for all 32 data rows (with the difference up to 11%) and less than after decimal logarithm univariate transformation for 31 (with the difference up to 49%) from 32 data rows (except row 10 with the difference of 9%).

Table 4 – The bounds of prediction interval

No	Y	univariate				four-variate	
		decimal logarithm		Box-Cox		Box-Cox	
		LB	UB	LB	UB	LB	UB
1	0.448	0.369	0.507	0.385	0.493	0.389	0.488
2	7.846	6.586	8.924	6.707	8.853	6.711	8.686
3	4.345	3.501	4.735	3.633	4.744	3.593	4.602
4	2.717	2.403	3.323	2.255	2.982	2.267	2.936
5	2.954	2.777	3.797	2.796	3.669	2.768	3.560
7	0.212	0.167	0.232	0.168	0.214	0.177	0.221
8	1.149	1.065	1.439	1.108	1.421	1.096	1.379
9	0.477	0.418	0.574	0.435	0.558	0.435	0.547
10	61.269	46.450	64.135	50.568	70.205	53.748	73.141
11	0.124	0.098	0.135	0.108	0.137	0.112	0.139
13	0.349	0.298	0.412	0.305	0.392	0.309	0.390
15	1.538	1.383	1.877	1.445	1.867	1.435	1.819
16	0.365	0.309	0.421	0.326	0.415	0.330	0.411
17	2.332	1.829	2.470	1.898	2.451	1.873	2.373
18	24.347	19.271	26.495	18.181	24.751	18.985	25.330
19	12.433	9.140	12.496	9.922	13.287	9.962	13.073
20	0.948	0.878	1.187	0.920	1.177	0.912	1.145
21	1.826	1.607	2.171	1.662	2.143	1.644	2.079
23	3.567	3.031	4.197	3.081	4.101	3.085	4.022
24	3.735	3.302	4.464	3.346	4.362	3.320	4.245
26	4.029	3.536	4.787	3.518	4.593	3.504	4.488
27	3.500	3.220	4.350	3.318	4.321	3.286	4.198
28	0.131	0.105	0.146	0.114	0.145	0.118	0.148
29	2.227	1.824	2.470	1.871	2.422	1.852	2.352
30	0.521	0.467	0.666	0.451	0.593	0.465	0.598
31	1.494	1.319	1.795	1.325	1.716	1.316	1.671
32	0.906	0.863	1.188	0.869	1.128	0.882	1.122
33	0.471	0.397	0.540	0.422	0.536	0.422	0.526
34	24.04	20.478	28.028	20.527	27.827	21.267	28.262
35	2.681	2.226	3.009	2.327	3.016	2.299	2.923
36	20.327	20.001	27.551	19.173	26.199	19.989	26.765
37	24.66	21.797	29.846	21.150	28.685	22.101	29.389

Also, the width of the nonlinear regression prediction interval for the four-variate Box-Cox transformation is less than after the univariate Box-Cox transformation for all 32 data rows (with the difference up to 11%) and less than after decimal logarithm univariate transformation for 31 (with the difference up to 59%) from 32 data rows (except row 10 with the difference of 8%).

6 DISCUSSION

We apply four-variate normalizing transformations to build the nonlinear regression model for estimating the size of Web apps created using the CakePHP framework by appropriate techniques [20] since the error distribution of the linear regression model is not Gaussian what the chi-squared test result indicates. Also, there are outliers in the data from Table 1. Moreover, the four-variate distribution of the data from Table 1 is not Gaussian what the Mardia multivariate normality test based on measures of the multivariate skewness and kurtosis indicates. Because we use the statistical technique [26] to detect multivariate outliers in the four-dimensional non-Gaussian data from Table I based on the multivariate normalizing transformations and the SMD for normalized data. Note, we have other four-variate outliers for the data from Table 1 without applying normalization compared to outlier detection results using the above technique [26].

Also note that in our case for the data from Table 1, the poor normalization of multivariate non-Gaussian data using the Johnson univariate transformation leads to an increase in the widths of the confidence and prediction intervals of multiple nonlinear regression for a larger number of data rows compared to the Box-Cox four-variate transformation.

The widths of the confidence and prediction intervals of multiple nonlinear regression based on the Box-Cox four-variate transformation are smaller for more data rows than for multiple nonlinear regressions following the univariate transformations, both the decimal logarithm and the Box-Cox ones. Also the MMRE value is smaller for the model (9) for the Box-Cox four-variate transformation in comparison with all other nonlinear models based on univariate transformations. This may be explained best four-variate normalization of non-Gaussian data from Table 1 using the Box-Cox four-variate transformation.

The obtained results and results from [9] indicate that constructing a multiple nonlinear regression model to estimate the size (in KLOC) of Web apps using the specific framework (CakePHP in our case and Laravel in [9]) leads to an increase of confidence in estimating.

CONCLUSIONS

The important problem of increase of confidence in estimating the size of Web apps created using the CakePHP framework is solved.

The scientific novelty of obtained results is that three-factors nonlinear regression model to estimate the size of Web apps created using the CakePHP framework is firstly constructed on the basis of the Box-Cox four-variate transformation. This model, in comparison with other nonlinear regression models, has a smaller value of the mean magnitude of relative error, smaller widths of the confidence and prediction intervals of three-factors nonlinear regression.

The practical significance of obtained results is that the software realizing the constructed model is developed in the sci-language for Scilab. The experimental results allow to recommend the constructed model for use in practice.

Prospects for further research may include the application of other multivariate normalizing transformations and data sets to construct multiple nonlinear regression models for estimating the size of Web apps created using the specific frameworks.

REFERENCES

1. Boehm B. W., Abts C., Brown A. W. et al. Software cost estimation with COCOMO II. Upper Saddle River, NJ, Prentice Hall PTR, 2000, 506 p.
2. Kaczmarek J., Kucharski M. Size and effort estimation for applications written in Java, *Information and Software Technology*, 2004, Vol. 46, Issue 9, pp. 589–601. DOI: 10.1016/j.infsof.2003.11.001
3. Tan H. B. K., Zhao Y., Zhang H. Estimating LOC for information systems from their conceptual data models, *Software Engineering : the 28th International Conference (ICSE ‘06)*. Shanghai, China, May 20–28, 2006, proceedings, pp. 321–330. DOI: 10.1145/1134285.1134331
4. Tan H. B. K., Zhao Y., Zhang H. Conceptual data model-based software size estimation for information systems, *Transactions on Software Engineering and Methodology*, 2009, Vol. 19, Issue 2, October 2009, Article No. 4. DOI: 10.1145/1571629.1571630
5. Zifen Y. An improved software size estimation method based on object-oriented approach, *Electrical & Electronics Engineering : IEEE Symposium EEESYM*’12, Kuala Lumpur, Malaysia, 24–27 June, 2012, proceedings, *IEEE*, 2012, pp. 615–617. DOI: 10.1109/EEESym.2012.6258733.
6. Kiewkanya M., Surak S. Constructing C++ software size estimation model from class diagram, *Computer Science and Software Engineering, 13th International Joint Conference, Khon Kaen*. Thailand, July 13–15, 2016, proceedings, IEEE, 2016, pp. 1–6. DOI: 10.1109/JCSSE.2016.7748880
7. Prykhodko N. V., Prykhodko S. B. Constructing the non-linear regression models on the basis of multivariate normalizing transformations, *Electronic modeling*, 2018, Vol. 40, No. 6, pp. 101–110. DOI: 10.15407/emodel.40.06.101
8. Prykhodko N. V., Prykhodko S. B. The non-linear regression model to estimate the software size of open source Java-based systems, *Radio Electronics, Computer Science, Control,* 2018, No. 3 (46), pp. 158–166. DOI: 10.15588/1607-3274-2018-3-17
9. Prykhodko S. B., Prykhodko N. V., Vorona M. V. et al. Nonlinear regression model for estimating the size of web applications created using the Laravel framework, *Information technology and computer engineering,* 2021, Vol. 50, No. 1, pp. 115–121. DOI: 10.31649/1999-9941-2021-50-1-115-121 [Published in Ukrainian]
10. Nassif A. B., AbuTalib M., Capretz L. F. Software effort estimation from use case diagrams using nonlinear regression analysis, *Electrical and Computer Engineering : IEEE Canadian Conference CCECE’20.,* London, ON, Canada, 30 Aug.–2 Sept., 2020, proceedings, IEEE, 2020, pp. 1–4. DOI: 10.1109/CCECE47787.2020.9255712.
11. Montgomery D. C., Peck D. C., Introduction to linear regression analysis / D. C. Montgomery, // 2nd edn. – New York: John Wiley & Sons, 1992. – 544 p.
12. Seber G.A.F., Lee A. J. Linear regression analysis, 2nd edn. New York, John Wiley & Sons, 2003, 582 p.
13. Weisberg, S. Applied linear regression, 4th edn. New York, John Wiley & Sons, 2013, 368 p.
14. Bates D. M., Watts D. G. Nonlinear regression analysis and its applications. New York, John Wiley & Sons, 1988, 384 p. DOI:10.1002/9780470316757
15. Seber G.A.F., Wild C. J. Nonlinear regression. New York: John Wiley & Sons, 1989, 768 p. DOI: 10.1002/0471725315

DOI 10.15588/1607-3274-2021-4-12

16. Ryan T.P. Modern regression methods. New York, John Wiley & Sons, 1997, 529 p. DOI: 10.1002/9780470382806
17. Drapper N. R., Smith H. Applied regression analysis. New York, John Wiley & Sons, 1998, 736 p.
18. Johnson R. A., Wichern D. W. Applied multivariate statistical analysis. Pearson Prentice Hall, 2007, 800 p.
19. Chatterjee S., Simonoff J. S. Handbook of regression analysis. New York, John Wiley & Sons, 2013, 236 p. DOI: 10.1002/9781118532843
20. Prykhodko S., Prykhodko N. Mathematical modeling of non-Gaussian dependent random variables by nonlinear regression models based on the multivariate normalizing transformations, *Mathematical Modeling and Simulation of Systems : 15th International Scientific-practical Conference MODS'2020, Chernihiv, Ukraine, June 29 – July 01, 2020 : selected papers. Springer*, Cham., 2021, pp. 166–174. (Advances in Intelligent Systems and Computing, Vol. 1265). DOI: 10.1007/978-3-030-58124-4_16
21. Belsley D. A., Kuh E., Welsch R. E. Regression diagnostics: Identifying influential data and sources of collinearity / D. A. Belsley,. – New York: John Wiley, 1980. – 300 p. DOI:10.1002/0471725153
22. Chatterjee S., Price B. Regression analysis by example. New York, John Wiley & Son, 2012, 424 p.
23. Olkin I., Sampson A. R., Smelser N. J., Baltes P. B. Eds. Multivariate analysis: Overview, *International encyclopedia of social & behavioral sciences. 1st edn.* Elsevier, Pergamon, 2001, pp. 10240–10247.
24. Mardia K. V. Measures of multivariate skewness and kurtosis with applications, *Biometrika*, 1970, Vol. 57, pp. 519–530. DOI: 10.1093/biomet/57.3.519
25. Mardia K. V. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, *Sankhya: The Indian Journal of Statistics, Series B (1960–2002)*, 1974, Vol. 36, Issue 2, pp. 115–128.
26. Prykhodko S., Prykhodko N., Makarova L. , Pugachenko K. Detecting outliers in multivariate non-Gaussian data on the basis of normalizing transformations, *Electrical and Computer Engineering : the 2017 IEEE First Ukraine Conference (UKRCON) «Celebrating 25 Years of IEEE Ukraine Section»*, Kyiv, Ukraine, May 29 – June 2, 2017, proceedings, pp. 846–849. DOI: 10.1109/UKRCON.2017.8100366
27. Prykhodko S., Prykhodko N., Knyrik K. Estimating the efforts of mobile application development in the planning phase using nonlinear regression analysis, *Applied Computer Systems*, 2020, Vol. 25, No. 2, pp. 172–179. DOI: 10.2478/acss-2020-0019
28. Campbell S. L., Chancelier J.-P., Nikoukhah R. Modeling and simulation in Scilab/Scicos. Springer, 2005, 313 p.
29. Foss T., Stensrud E., Kitchenham B., Myrtveit I. A simulation study of the model evaluation criterion MMRE, *IEEE Transactions on software engineering*, 2003, Vol. 29, Issue 11, pp. 985–995. DOI: 10.1109/TSE.2003.1245300
30. Port D., Korte M. Comparative studies of the model evaluation criterions MMRE and PRED in software cost estimation research, *Empirical Software Engineering and Measurement, the 2nd ACM-IEEE International Symposium ESEM, Kaiserslautern*. Germany, October, 2008, proceedings. New York, ACM, 2008, pp. 51–60.

Received 04.09.2021.
Accepted 25.10.2021.

УДК 004.412:519.237.5

НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ВЕБ-ЗАСТОСУНКІВ, ЩО СТВОРЮЮТЬСЯ З ВИКОРИСТАННЯМ ФРЕЙМВОРКУ CAKEPHP

Приходько С. Б. – д-р техн. наук, професор, завідувач кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова, Миколаїв, Україна.

Шутко І. С. – аспірант кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова, Миколаїв, Україна.

Приходько А. С. – студент кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова, Миколаїв, Україна.

АНОТАЦІЯ

Актуальність. Проблема оцінювання розміру програмного забезпечення на ранній стадії програмного проекту є важливою, оскільки оцінювання розміру програмного забезпечення використовується для прогнозування трудомісткості розробки програмного забезпечення, включаючи веб-застосунки з відкритим кодом на PHP, що створені із використанням фреймворку CakePHP. Об'єктом дослідження є процес оцінювання розміру веб-застосунків з відкритим кодом на PHP, що створені із використанням фреймворку CakePHP. Предметом дослідження є нелінійні регресійні моделі для оцінювання розміру веб-застосунків з відкритим кодом на PHP, що створені із використанням фреймворку CakePHP.

Мета. Метою роботи є побудова нелінійної регресійної моделі з трьома предикторами для оцінювання розміру веб-застосунків, шо створюються із використанням фреймворку CakePHP на основі чотиривимірного нормалізуючого перетворення Бокса-Кокса, щоб підвищити достовірність раннього оцінювання розміру цих застосунків.

Метод. Модель, довірчі інтервали та інтервали передбачення багатовимірної нелінійної регресії для оцінювання розміру веб-застосунків з відкритим кодом на PHP, створених із використанням фреймворку CakePHP, побудовані на основі багатовимірного нормалізуючого перетворення Бокса-Кокса для негаусівських даних за допомогою відповідних методів. Методи побудови моделей, рівнянь, довірчих інтервалів і інтервалів передбачення нелінійних регресій засновані на множинному нелінійному регресійному аналізі з використанням багатовимірних нормалізуючих перетворень. Ці методи дозволяють враховувати кореляцію між залежними та незалежними змінними у разі нормалізації багатовимірних негаусівських даних. Загалом, це призводить до зменшення середньої величини відносної похибки, ширини довірчих інтервалів і інтервалів передбачення в порівнянні нелінійними моделями, побудованими з використанням одновимірних нормалізуючих перетворень.

Результати. Проведено порівняння побудованої моделі з нелінійними регресійними моделями на основі десяткового логарифму та одновимірного перетворення Бокса-Кокса.

DOI 10.15588/1607-3274-2021-4-12

Висновки. Модель нелінійної регресії з трьома предикторами для оцінювання розміру веб-застосунків, створених за допомогою фреймворку CakePHP, побудована на основі чотиривимірного перетворення Бокса-Кокса. Ця модель, у порівнянні з іншими нелінійними регресійними моделями, має більший множинний коефіцієнт детермінації, менше значення середньої величини відносної похибки та менші ширини довірчих інтервалів та інтервалів передбачення. Перспективи подальших досліджень можуть включати застосування інших багатовимірних нормалізуючих перетворень та наборів даних для побудови нелінійних регресійних моделей для оцінювання розміру веб-додатків, створених за допомогою інших фреймворків.

КЛЮЧОВІ СЛОВА: оцінка розміру програмного забезпечення, веб-додаток, нелінійна регресійна модель, нормалізуюче перетворення, негаусівські дані.

УДК 004.412:519.237.5

НЕЛИНЕЙНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ ДЛЯ ОЦЕНКИ РАЗМЕРА ВЕБ-ПРИЛОЖЕНИЙ, СОЗДАВАЕМЫХ С ИСПОЛЬЗОВАНИЕМ ФРЕЙМВОРКА CAKEPHP

Приходько С. Б. – д-р техн. наук, профессор, заведующий кафедрой программного обеспечения автоматизированных систем Национального университету кораблестроения имени адмирала Макарова, Николаев, Украина.

Шутко И. С. – аспирант кафедры программного обеспечения автоматизированных систем Национального университету кораблестроения имени адмирала Макарова, Николаев, Украина.

Приходько А. С. – студент кафедры программного обеспечения автоматизированных систем Национального университету кораблестроения имени адмирала Макарова, Николаев, Украина.

АННОТАЦИЯ

Актуальность. Проблема оценки размера программного обеспечения на ранней стадии программного проекта является важной, поскольку оценки размера программного обеспечения используется для прогнозирования трудоемкости разработки программного обеспечения, включая веб-приложения с открытым кодом на PHP, созданных с использованием фреймворка CakePHP. Объектом исследования является процесс оценки размера веб-приложений с открытым кодом на PHP, созданных с использованием фреймворка CakePHP. Предметом исследования является нелинейные регрессионные модели для оценки размера веб-приложений с открытым кодом на PHP, созданных с использованием фреймворка CakePHP.

Цель. Целью работы является построение нелинейной регрессионной модели с тремя предикторами для оценки размера веб-приложений, создаваемых с использованием фреймворка CakePHP на основе четырехмерного нормализующего преобразования Бокса-Кокса, чтобы повысить достоверность раннего оценивания размера этих приложений.

Метод. Модель, доверительные интервалы и интервалы предсказания многомерной нелинейной регрессии для оценки размера веб-приложений с открытым кодом на PHP, созданных с использованием фреймворка CakePHP, построены на основе многомерного нормализующего преобразования Бокса-Кокса для негауссовских данных с помощью соответствующих методов. Методы построения моделей, уравнений, доверительных интервалов и интервалов предсказания нелинейных регрессий основаны на множественном нелинейном регрессионном анализе с использованием многомерных нормализующих преобразований. Эти методы позволяют учитывать корреляцию между зависимыми и независимыми переменными в случае нормализации многомерных негауссовских данных. В общем, это приводит к уменьшению средней величины относительной погрешности, ширины доверительных интервалов и интервалов предсказания по сравнению нелинейными моделями, построенными с использованием одномерных нормализующих преобразований.

Результаты. Проведено сравнение построенной модели с нелинейными регрессионными моделями на основе десятичного логарифма и одномерного преобразования Бокса-Кокса.

Выводы. Модель нелинейной регрессии с тремя предикторами для оценки размера веб-приложений, созданных с помощью фреймворка CakePHP, построена на основе четырехмерного преобразования Бокса-Кокса. Эта модель, по сравнению с другими нелинейными регрессионными моделями, имеет больший множественный коэффициент детерминации, меньшее значение средней величины относительной погрешности и меньшие ширины доверительных интервалов и интервалов предсказания. Перспективы дальнейших исследований могут включать применение других многомерных нормализующих преобразований и наборов данных для построения нелинейных регрессионных моделей для оценки размера веб-приложений, созданных с помощью других фреймворков.

КЛЮЧЕВЫЕ СЛОВА: оценка размера программного обеспечения, веб-приложение, нелинейная регрессионная модель, нормализующее преобразование, негауссовские данные.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Boehm B.W. Software cost estimation with COCOMO II / [B. W. Boehm, C. Abts, A. W. Brown et al.]. – Upper Saddle River, NJ : Prentice Hall PTR, 2000. – 506 p.
2. Kaczmarek J. Size and effort estimation for applications written in Java / J. Kaczmarek, M. Kucharski // Information and Software Technology. – 2004. – Vol. 46, Issue 9. – P. 589–601. DOI: 10.1016/j.infsof.2003.11.001
3. Tan H. B. K. Estimating LOC for information systems from their conceptual data models / H. B. K. Tan, Y. Zhao, H. Zhang // Software Engineering : the 28th International Conference (ICSE ‘06), Shanghai, China, May 20–28, 2006 : proceedings. – P. 321–330. DOI: 10.1145/1134285.1134331
4. Tan H. B. K. Conceptual data model-based software size estimation for information systems / H. B. K. Tan, Y. Zhao, H. Zhang // Transactions on Software Engineering and Methodology. – 2009. – Vol. 19. – Issue 2. – October 2009. – Article No. 4. DOI: 10.1145/1571629.1571630
5. Zifen Y. An improved software size estimation method based on object-oriented approach / Y. Zifen // Electrical & Electronics Engineering : IEEE Symposium EEESYM’12, Kuala Lumpur, Malaysia, 24–27 June, 2012 : proceedings. – IEEE, 2012. – P. 615–617. DOI: 10.1109/EEESym.2012.6258733.

DOI 10.15588/1607-3274-2021-4-12

6. Kiewkanya M. Constructing C++ software size estimation model from class diagram / M. Kiewkanya, S. Surak // Computer Science and Software Engineering : 13th International Joint Conference, Khon Kaen, Thailand, July 13–15, 2016 : proceedings. – IEEE, 2016. – P. 1–6. DOI: 10.1109/JCSSE.2016.7748880
7. Prykhodko N.V. Constructing the non-linear regression models on the basis of multivariate normalizing transformations / N. V. Prykhodko, S. B. Prykhodko // Electronic modeling. – 2018. – Vol. 40, No. 6. – P. 101–110. DOI: 10.15407/emodel.40.06.101
8. Prykhodko N.V. The non-linear regression model to estimate the software size of open source Java-based systems / N. V. Prykhodko, S. B. Prykhodko // Radio Electronics, Computer Science, Control. – 2018. – No. 3 (46). – P. 158–166. DOI: 10.15588/1607-3274-2018-3-17
9. Приходько С.Б. Нелінійна регресійна модель для оцінювання розміру Web-застосунків, що створюються з використанням фреймворку Laravel / [С. Б. Приходько, Н. В. Приходько, М. В. Ворона та ін.] // Інформаційні технології та комп'ютерна інженерія. – 2021. – Том 50, № 1. – С. 115–121. DOI: 10.31649/1999-9941-2021-50-1-115-121
10. Nassif A. B. Software effort estimation from use case diagrams using nonlinear regression analysis / A. B. Nassif, M. AbuTalib, L. F. Capretz // Electrical and Computer Engineering : IEEE Canadian Conference CCECE'20, London, ON, Canada, 30 Aug.–2 Sept., 2020 : proceedings. – IEEE, 2020. – P. 1–4. DOI: 10.1109/CCECE47787.2020.9255712.
11. Montgomery D.C. Introduction to linear regression analysis / D. C. Montgomery, D. C. Peck, // 2nd edn. – New York: John Wiley & Sons, 1992. – 544 p.
12. Seber G.A.F. Linear regression analysis / G. A. F. Seber, A. J. Lee // 2nd edn. – New York: John Wiley & Sons, 2003. – 582 p.
13. Weisberg, S. Applied linear regression / S. Weisberg // 4th edn. – New York: John Wiley & Sons, 2013. – 368 p.
14. Bates D.M. Nonlinear regression analysis and its applications / D. M. Bates, D. G. Watts. – New York: John Wiley & Sons, 1988. – 384 p. DOI:10.1002/9780470316757
15. Seber G.A.F. Nonlinear regression / G.A.F. Seber, C. J. Wild. – New York : John Wiley & Sons, 1989. – 768 p. DOI: 10.1002/0471725315
16. Ryan T.P. Modern regression methods / T. P. Ryan. – New York: John Wiley & Sons, 1997. – 529 p. DOI: 10.1002/9780470382806
17. Drapper N.R. Applied regression analysis / N. R. Drapper, H. Smith. – New York: John Wiley & Sons, 1998. – 736 p.
18. Johnson R.A. Applied multivariate statistical analysis / R. A. Johnson, D. W. Wichern. – Pearson Prentice Hall, 2007. – 800 p.
19. Chatterjee S. Handbook of regression analysis / S. Chatterjee, J. S. Simonoff. – New York : John Wiley & Sons, 2013. – 236 p. DOI: 10.1002/9781118532843
20. Prykhodko S. Mathematical modeling of non-Gaussian dependent random variables by nonlinear regression models based on the multivariate normalizing transformations / S. Prykhodko, N. Prykhodko // Mathematical Modeling and Simulation of Systems : 15th International Scientific-practical Conference MODS'2020, Chernihiv, Ukraine, June 29 – July 01, 2020 : selected papers. – Springer, Cham., 2021. – P. 166–174. – (Advances in Intelligent Systems and Computing, Vol. 1265). DOI: 10.1007/978-3-030-58124-4_16
21. Belsley D. A. Regression diagnostics: Identifying influential data and sources of collinearity / D. A. Belsley, E. Kuh, R. E. Welsch. – New York : John Wiley, 1980. – 300 p. DOI:10.1002/0471725153
22. Chatterjee S. Regression analysis by example / S. Chatterjee, B. Price. – New York : John Wiley & Son, 2012. – 424 p.
23. Olkin I. Multivariate analysis: Overview / I. Olkin, A. R. Sampson // N. J. Smelser, P. B. Baltes, Eds. International encyclopedia of social & behavioral sciences. 1st edn. – Elsevier, Pergamon, 2001. – P. 10240–10247.
24. Mardia K. V. Measures of multivariate skewness and kurtosis with applications / K. V. Mardia // Biometrika. – 1970. – Vol. 57. – P. 519–530. DOI: 10.1093/biomet/57.3.519
25. Mardia K.V. Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies / K. V. Mardia // Sankhya: The Indian Journal of Statistics, Series B (1960–2002). – 1974. – Vol. 36, Issue 2. – P. 115–128.
26. Detecting outliers in multivariate non-Gaussian data on the basis of normalizing transformations / [S. Prykhodko, N. Prykhodko, L. Makarova, K. Pugachenko] // Electrical and Computer Engineering : the 2017 IEEE First Ukraine Conference (UKRCON) «Celebrating 25 Years of IEEE Ukraine Section», Kyiv, Ukraine, May 29 – June 2, 2017 : proceedings. – P. 846–849. DOI: 10.1109/UKRCON.2017.8100366
27. Prykhodko S. Estimating the efforts of mobile application development in the planning phase using nonlinear regression analysis / S. Prykhodko, N. Prykhodko, K. Knyrik // Applied Computer Systems. – 2020. – Vol. 25, No. 2. – P. 172–179. DOI: 10.2478/acss-2020-0019
28. Campbell S.L. Modeling and simulation in Scilab/Scicos / S. L. Campbell, J.-P. Chancelier, R. Nikoukhah. – Springer, 2005. – 313 p.
29. A simulation study of the model evaluation criterion MMRE / [T. Foss, E. Stensrud, B. Kitchenham, I. Myrtveit] // IEEE Transactions on software engineering. – 2003. – Vol. 29, Issue 11. – P. 985–995. DOI: 10.1109/TSE.2003.1245300
30. Port D. Comparative studies of the model evaluation criterions MMRE and PRED in software cost estimation research / D. Port, M. Korte // Empirical Software Engineering and Measurement : the 2nd ACM-IEEE International Symposium ESEM, Kaiserslautern, Germany, October, 2008 : proceedings. – New York : ACM, 2008. – P. 51–60.

DOI 10.15588/1607-3274-2021-4-12

UDC 004.93

ANDROID SOFTWARE AGING AND REJUVENATION MODEL CONSIDERING THE BATTERY CHARGE

Yakovyna V. S. – Dr. Sc., Full Professor of Artificial Intelligence Department, Lviv Polytechnic National University, Lviv, Ukraine; Faculty of Mathematics and Computer Science, University of Warmia and Mazury in Olsztyn, Poland.

Uhrynovskyi B. V. – Assistant of Software Department, Lviv Polytechnic National University, Lviv, Ukraine.

ABSTRACT

Context. A feature of mobile systems is their dependence on battery charge, which is an important factor when planning various processes, in particular when planning time of performing software rejuvenation procedure.

Objective. The goal of this article is to develop a model of software aging process with performing rejuvenation procedure for the Android operating system considering the factor of battery charge.

Method. A complex model based on Continuous-Time Markov Chains is proposed, which combines the software aging and rejuvenation model, the user behavior model and consider battery charge factor. A graph of states and transitions describing a complex model is constructed. Based on the formed graph the system of differential equations is written. The system was calculated using the 4th order Runge-Kutta method. The optimal time for the rejuvenation procedure can be determined when rejuvenation will not interfere with the user and will be performed before the battery is fully discharged, ie when the probability of the system being in these states is the lowest.

Results. The simulation of the developed model for test values of transition rates is performed. Considering the battery charge model allows to avoid planning the rejuvenation procedure at a time when the mobile device is likely to have a low charge or be completely discharged.

Conclusions. The proposed model based on the Markov chain allows to predict the start time of software rejuvenation procedure, considering both user behavior and battery level, which can have a significant impact on the predicted time. Also, the early implementation of the rejuvenation procedure may have the effect of reducing the system workload and delaying the discharge of the device, which should be checked in further studies. The expediency and importance of the consideration of battery charge factor and the need for further study of the proposed software aging and rejuvenation model are substantiated.

KEYWORDS: software aging, software rejuvenation, Markov Chains, Android.

ABBREVIATIONS

OS is an operating system;

UI is a user interface;

AY is an “Active Young” state;

RA is a “Rebirth Active” state;

ARe is an “Active Recovering” state;

AYS(L)p is an “Active Young Stable (Low) Power” state;

AOS(L)p is an “Active Old Stable (Low) Power” state;

AReS(L)p is an “Active Recovering Stable (Low) Power” state;

ARS(L)p is an “Active Rebirth Stable (Low) Power” state;

SYS(L)p is a “Sleep Young Stable (Low) Power” state;

SOS(L)p is a “Sleep Old Stable (Low) Power” state;

SReS(L)p is a “Sleep Recovering Stable (Low) Power” state;

SRS(L)p is a “Sleep Rebirth Stable (Low) Power” state;

Op is an “Off Power” state.

NOMENCLATURE

λ is a transition rate between states;

λ_{AS} is a transition rate from “Active” state to “Sleep” state;

λ_{SA} is a transition rate from “Sleep” state to “Active” state;

λ_{YO} is a transition rate from “Young” state to “Old” state;

λ_{ORe} is a transition rate from “Old” state to “Recovering” state;

λ_{ReY} is a transition rate from “Recovering” state to “Young” state;

λ_{YR} is a transition rate from “Young” state to “Rebirth” state;

λ_{RY} is a transition rate from “Rebirth” state to “Young” state;

λ_{HpLp} is a transition rate from “High Power” state to “Low Power” state;

λ_{LpCh} is a transition rate from “Low Power” state to “Charging” state;

λ_{ChHp} is a transition rate from “Charging” state to “High Power” state;

λ_{LpOp} is a transition rate from “Low Power” state to “Off Power” state;

λ_{SpLp} is a transition rate from “Stable Power” state to “Low Power” state;

λ_{LpSp} is a transition rate from “Low Power” state to “Stable Power” state;

t is a time;

P(t) is the probabilities vector of the system in different states at time *t*;

$P_{AYS(L)p}(t)$ is the probability of a system being at time t in the “Active Young Stable (Low) Power” state;

$P_{AOS(L)p}(t)$ is the probability of a system being at time t in the “Active Old Stable (Low) Power” state;

DOI 10.15588/1607-3274-2021-4-13

$P_{AReS(L)p}(t)$ is the probability of a system being at time t in the “Active Recovering Stable (Low) Power” state;

$P_{ARS(L)p}(t)$ is the probability of a system being at time t in the “Active Rebirth Stable (Low) Power” state;

$P_{SYS(L)p}(t)$ is the probability of a system being at time t in the “Sleep Young Stable (Low) Power” state;

$P_{SOS(L)p}(t)$ is the probability of a system being at time t in the “Sleep Old Stable (Low) Power” state;

$P_{SReS(L)p}(t)$ is the probability of a system being at time t in the “Sleep Recovering Stable (Low) Power” state;

$P_{SRS(L)p}(t)$ is the probability of a system being at time t in the “Sleep Rebirth Stable (Low) Power” state;

$P_{Op}(t)$ is the probability of a system being at time t in the “Off Power” state;

T_{avg} is an average transition time.

INTRODUCTION

Software aging process is a phenomenon of performance deterioration and failures rate increase caused by the accumulation of software and system errors that are unpredictable and their effects can occur with delays in systems that run for a long time without reboots [1–7]. Aging-related errors are memory leaks, rounding errors, etc. The phenomenon of software aging has been widely studied in various systems and applications, such as Linux, Apache server, middleware [3–6], Android OS [8–10].

Software rejuvenation procedure is a proactive technique to prevent and delay software aging [2]. This approach involves the planned regular or irregular cleaning of the system internal state in such a way as to reduce the number of accumulated errors. Software rejuvenation can be applied to different levels of the system, such as system processes and user applications, data objects within processes, or the entire operating system.

Mobile systems have specific characteristics of operation and use by the user, compared to server or personal computer systems. The previous article [10] describes some features of mobile platforms, as well as the relevance of studying the phenomenon of software aging in mobile systems. Software aging in Android OS manifests as slow response of the user interface after prolonged use of a mobile phone, which is common issue for most users. Also, Android is one of the most popular systems for mobile phones and occupies 84.1% of the market as of 2020 [11]. Due to the urgency of the software aging problem in mobile systems and the popularity of Android OS, an important task is to develop mathematical and algorithmic tools to combat the phenomenon of software aging, in particular modeling of the phenomenon taking into account different factors allows to plan the rejuvenation procedure.

The object of the study is the process of software aging.

The subject of the study is a model of the software aging process for mobile OS with performing rejuvenation procedure.

The purpose of the work is to develop a model of software aging process for Android OS with performing rejuvenation procedure, which takes into account the model of mobile device usage by the users and battery charge level.

DOI 10.15588/1607-3274-2021-4-13

1 PROBLEM STATEMENT

The rejuvenation procedure is used to counteract the phenomenon of software aging. Performing this procedure requires predicting the occurrence of software aging and planning the time of rejuvenation based on the predicted data. The model of the rejuvenation process should take into account the specific factors and features of the Android OS, as well as data that will objectively reflect the status and dynamics of changes in system performance to be able to make accurate predictions.

Analytical description of software aging process with rejuvenation procedure will be presented in the form of Continuous-Time Markov Chains. Steps needed to build a model and perform calculations to determine the optimal rejuvenation time:

– build and describe the model of software aging process and rejuvenation procedure, the model of the battery charge states and the model of mobile phone user activity;

– identify the states in which the system may be according to considered models and describe the transition rates between these states;

– build a graph of state transitions and describe a system of differential equations based on this graph;

– calculate a system of differential equations using a test set of transition rates between states and analyze the results of calculations.

Calculating the system of differential equations will determine the probabilities of the mobile OS in different states at a particular time. To determine the optimal time to perform the rejuvenation procedure, proposed model allows to find the time when the probability of the system being in a state of active usage and low battery is minimal and aging-related effects are observed.

2 RELATED WORKS

To counteract the phenomenon of software aging in the Android mobile system, various approaches are proposed, both at the level of optimizing the architecture of the developed software [12] and by performing the rejuvenation procedure [13–14].

The article [15] proposes a model, which is a combination of the user behavior model and the software aging model with performing rejuvenation procedure. These models and their combinations are constructed in the form of stochastic Petri nets [16], based on which Continuous-Time Markov Chains were generated [17].

The model of user behavior is shown in Figure 1 and is described by two possible states: “Active” (the user is actively using the phone) and “Sleep” (the phone is in standby mode). Determining whether the device is in a certain state, as well as determining the transition rate between states, is proposed by accumulating statistics on user activity over time.

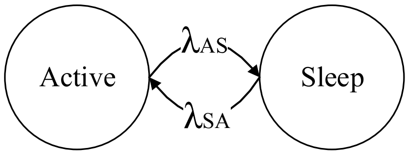


Figure 1 – Mobile device usage model

Software aging model with performing rejuvenation procedure consists of four states: "Young", "Old", "Recovering" and "Rebirth". It is assumed that when the device is turned on, the system is productive and the probability of aging-related failure is very low, i.e., the system is at "Young" state. After prolonged use of the phone, the system goes into "Old" state. The probability of such a transition can be determined by monitoring certain system metrics (such as UI response time and memory usage) and certain thresholds that will be indicators of the transition from "Young" to "Old" state. Switching from "Old" state to "Recovering" state means restarting the mobile device by the user due to critically poor performance or device rebooting due to system aging-related failure. Restart of the device returns the system to "Young" state. The rejuvenation procedure takes place in "Rebirth" state, which can be accessed from "Young" state.

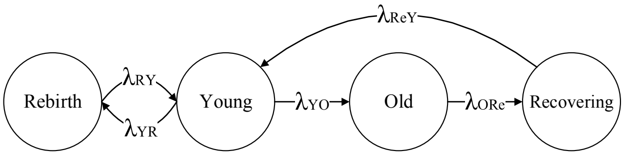


Figure 2 – Software aging model with rejuvenation procedure

The main goal of the model when planning the rejuvenation procedure is to determine the optimal time of transition to "Rebirth" state, which will precede the transition to "Old" state. An important factor in choosing the optimal time is to restrict the performance of the rejuvenation procedure if the user is currently using the phone, i.e., the phone is in "Active" state. This is important limitation because the rejuvenation procedure may have negative impact on user experience or cause a failure when user uses a mobile device, for example, if the rejuvenation procedure involves rebooting system or critical system components.

The model proposed by the authors [15] allowed them to determine the optimal time for the rejuvenation procedure (approximately 30 hours after turning on the phone), although the input data for the model were selected with many simplifications and assumptions. Proposed model can be expanded and improved. This paper proposes new model which considers the battery charge factor, as mobile devices depend on the battery capacity and are vulnerable to increase of intensity of the battery discharge.

3 MATERIALS AND METHODS

The model discussed in the previous section is proposed to be extended by an additional model that describes the battery charge. Two variants of battery charge models can be considered: general and simplified, which will be used in this work. The general model of a battery charge level can be described by means of four states (Fig. 3):

1) High Power – high battery level (for example, battery charge interval 100–25%);
2) Low Power – critically low battery charge at which it is impractical to perform rejuvenation procedure (for example, less than 25%);
3) Charging – the device is charging;
4) Off Power – the device discharged and turned off.

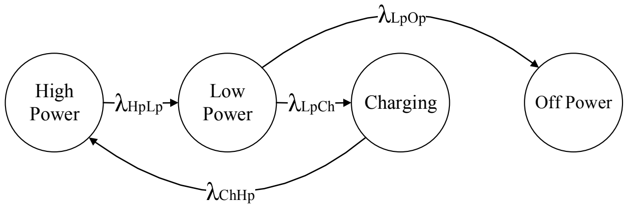


Figure 3 – General model of battery charge

State "Charging" does not actually affect the choice of rejuvenation, as it is impossible to predict the transition to this state or the duration of stay in this state, so the choice to rejuvenate or not still depends on the current battery status. Therefore, it is proposed to simplify this model and replace "High Power" and "Charging" states with the "Stable Power" state, which will mean a sufficient charge of the phone battery for the rejuvenation procedure. It is worth noting that "Charging" state can still be taken into account to adjust and delay the rejuvenation procedure after the initial planning based on the model. The final version of the simplified model is shown in Figure 4.

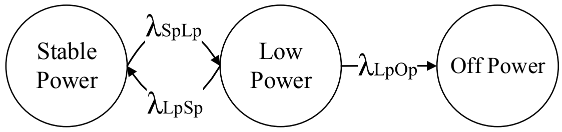


Figure 4 – Simplified model of battery charge

The transitions between states in this model can be calculated based on information about the current state of battery charge and the estimated time to complete discharge. The transition threshold between "Stable" and "Low", "Low" and "Off" states can be expressed as predictions of the discharge or charging time to the appropriate level, for example, 12 hours before the battery is fully discharged, or 3 hours before the battery is fully charged.

The software aging model with rejuvenation, the user behavior model and the model of battery charge were combined into one model, which is shown in Figure 5 and is represented by a Continuous-Time Markov Chains. An important feature of the new model is the impossibility of transition to "Rebirth" state from "Active" and "Low Power" states.

DOI 10.15588/1607-3274-2021-4-13

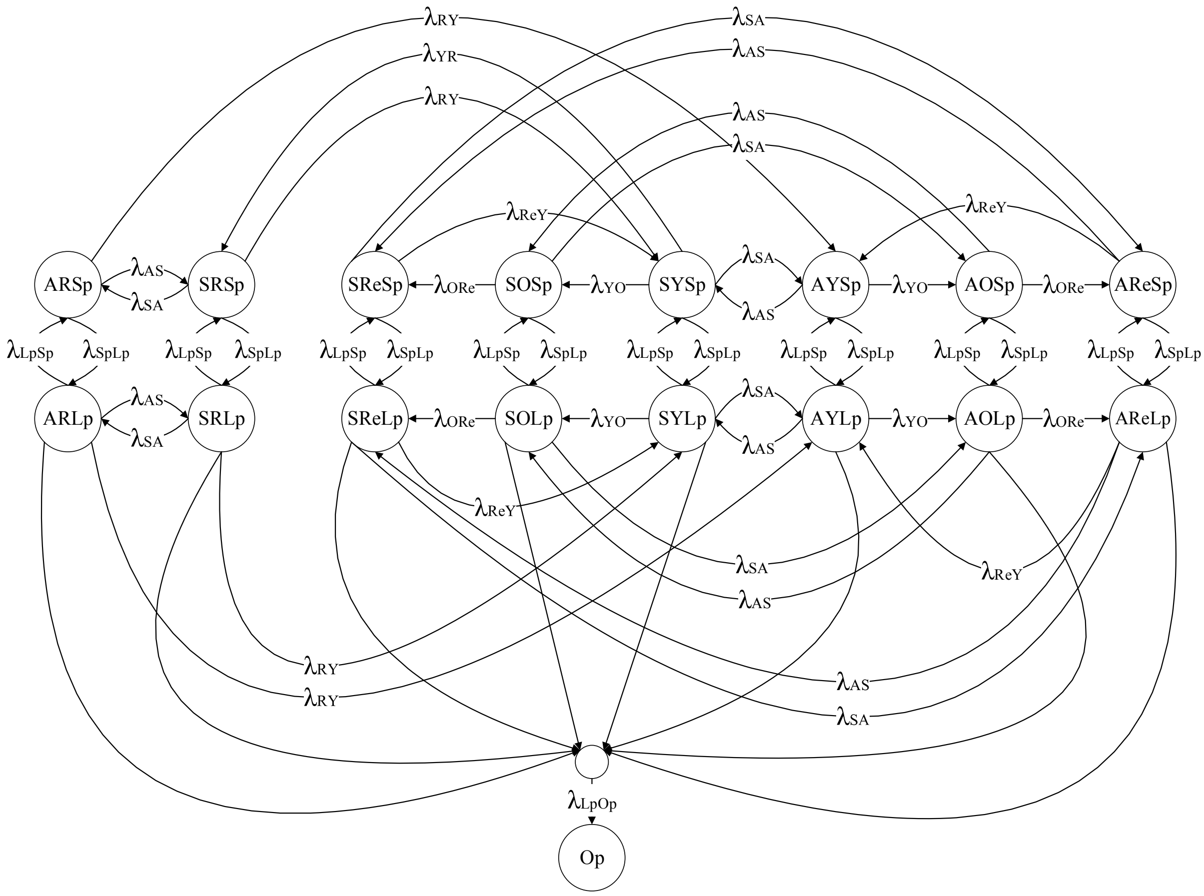


Figure 5 – Continuous-Time Markov Chain for the model of software aging and rejuvenation which takes into account the battery charge

Given the matrix of transition rates between states and the probabilities vector of the system in different states at the initial time, we can calculate the probabilities vector at any subsequent time (1):

$$P(t)=(P_{S\,\mathrm{Re}\,Sp}(t),P_{SOSp}(t),P_{SYSp}(t),P_{SRSp}(t),\\ P_{ARSp}(t),P_{AYSp}(t),P_{AOSp}(t),P_{A\,\mathrm{Re}\,Sp}(t),P_{S\,\mathrm{Re}\,Lp}(t),\\ P_{SOLp}(t),P_{SYLp}(t),P_{SRLp}(t),P_{ARLp}(t),P_{AYLp}(t),\\ P_{AOLp}(t),P_{A\,\mathrm{Re}\,Lp}(t),P_{Op}(t)). \qquad (1)$$

A system of differential equations (2) is written based on the Markov Chain, which can be calculated by numerical methods, in particular by the Runge-Kutta method.

$$\frac{dP_{S\,\mathrm{Re}\,Sp}(t)}{dt}=-(\lambda_{SA}+\lambda_{\mathrm{Re}\,Y}+\lambda_{SpLp})P_{S\,\mathrm{Re}\,Sp}(t)+\lambda_{SA}P_{A\,\mathrm{Re}\,Sp}(t)+\lambda_{O\,\mathrm{Re}}P_{SOSp}(t)+\lambda_{LpSp}P_{S\,\mathrm{Re}\,Lp}(t);$$

$$\frac{dP_{SOSp}(t)}{dt}=-(\lambda_{SA}+\lambda_{O\,\mathrm{Re}}+\lambda_{SpLp})P_{SOSp}(t)+\lambda_{AS}P_{AOSp}(t)+\lambda_{YO}P_{SYSp}(t)+\lambda_{LpSp}P_{SOLp}(t);$$

$$\frac{dP_{SYSp}(t)}{dt}=-(\lambda_{SA}+\lambda_{YO}+\lambda_{YR}+\lambda_{SpLp})P_{SYSp}(t)+\lambda_{AS}P_{AYSp}(t)+\lambda_{S\,\mathrm{Re}\,Sp}P_{\mathrm{Re}\,Y}(t)+\lambda_{SRSp}P_{RY}(t)+\lambda_{LpSp}P_{SYLp}(t); \qquad (2)$$

$$\frac{dP_{SRSp}(t)}{dt}=-(\lambda_{SA}+\lambda_{RY}+\lambda_{SpLp})P_{SRSp}(t)+\lambda_{AS}P_{ARSp}(t)+\lambda_{YR}P_{SYSp}(t)+\lambda_{LpSp}P_{SRLp}(t);$$

$$\frac{dP_{ARSp}(t)}{dt}=-(\lambda_{AS}+\lambda_{RY}+\lambda_{SpLp})P_{ARSp}(t)+\lambda_{SA}P_{SRSp}(t)+\lambda_{LpSp}P_{ARLp}(t);$$

DOI 10.15588/1607-3274-2021-4-13

$$\frac{dP_{AYSp}(t)}{dt} = -(\lambda_{AS} + \lambda_{YO} + \lambda_{SpLp})P_{AYSp}(t) + \lambda_{SA}P_{SYSp}(t) + \lambda_{RY}P_{ARSp}(t) + \lambda_{\mathrm{Re}Y}P_{A\mathrm{Re}Sp}(t) + \lambda_{LpSp}P_{AYLp}(t);$$

$$\frac{dP_{AOSp}(t)}{dt} = -(\lambda_{AS} + \lambda_{O\mathrm{Re}} + \lambda_{SpLp})P_{AOSp}(t) + \lambda_{SA}P_{SOSp}(t) + \lambda_{YO}P_{AYSp}(t) + \lambda_{LpSp}P_{AOLp}(t);$$

$$\frac{dP_{A\mathrm{Re}Sp}(t)}{dt} = -(\lambda_{AS} + \lambda_{\mathrm{Re}Y} + \lambda_{SpLp})P_{A\mathrm{Re}Sp}(t) + \lambda_{SA}P_{S\mathrm{Re}Sp}(t) + \lambda_{O\mathrm{Re}}P_{AOSp}(t) + \lambda_{LpSp}P_{A\mathrm{Re}Lp}(t);$$

$$\frac{dP_{S\mathrm{Re}Lp}(t)}{dt} = -(\lambda_{SA} + \lambda_{\mathrm{Re}Y} + \lambda_{LpSp} + \lambda_{LpOp})P_{S\mathrm{Re}Lp}(t) + \lambda_{AS}P_{A\mathrm{Re}Lp}(t) + \lambda_{O\mathrm{Re}}P_{SOLp}(t) + \lambda_{SpLp}P_{S\mathrm{Re}Sp}(t);$$

$$\frac{dP_{SOLp}(t)}{dt} = -(\lambda_{SA} + \lambda_{O\mathrm{Re}} + \lambda_{LpSp} + \lambda_{LpOp})P_{SOLp}(t) + \lambda_{AS}P_{AOLp}(t) + \lambda_{YO}P_{SYLp}(t) + \lambda_{SpLp}P_{SOSp}(t);$$

$$\frac{dP_{SYLp}(t)}{dt} = -(\lambda_{SA} + \lambda_{YO} + \lambda_{LpSp} + \lambda_{LpOp})P_{SYLp}(t) + \lambda_{AS}P_{AYLp}(t) + \lambda_{S\mathrm{Re}Lp}P_{\mathrm{Re}Y}(t) + \lambda_{SRLp}P_{RY}(t) + \lambda_{SpLp}P_{SYSp}(t);$$

$$\frac{dP_{SRLp}(t)}{dt} = -(\lambda_{SA} + \lambda_{RY} + \lambda_{LpSp} + \lambda_{LpOp})P_{SRLp}(t) + \lambda_{AS}P_{ARLp}(t) + \lambda_{SpLp}P_{SRSp}(t);$$

$$\frac{dP_{ARLp}(t)}{dt} = -(\lambda_{AS} + \lambda_{RY} + \lambda_{LpSp} + \lambda_{LpOp})P_{ARLp}(t) + \lambda_{SA}P_{SRLp}(t) + \lambda_{SpLp}P_{ARSp}(t);$$

$$\frac{dP_{AYLp}(t)}{dt} = -(\lambda_{AS} + \lambda_{YO} + \lambda_{LpSp} + \lambda_{LpOp})P_{AYLp}(t) + \lambda_{SA}P_{SYLp}(t) + \lambda_{RY}P_{ARLp}(t) + \lambda_{\mathrm{Re}Y}P_{A\mathrm{Re}Lp}(t) + \lambda_{SpLp}P_{AYSp}(t)$$

$$\frac{dP_{AOLp}(t)}{dt} = -(\lambda_{AS} + \lambda_{O\mathrm{Re}} + \lambda_{LpSp} + \lambda_{LpOp})P_{AOLp}(t) + \lambda_{SA}P_{SOLp}(t) + \lambda_{YO}P_{AYLp}(t) + \lambda_{SpLp}P_{AOSp}(t);$$

$$\frac{dP_{A\mathrm{Re}Lp}(t)}{dt} = -(\lambda_{AS} + \lambda_{\mathrm{Re}Y} + \lambda_{LpSp} + \lambda_{LpOp})P_{A\mathrm{Re}Lp}(t) + \lambda_{SA}P_{S\mathrm{Re}Lp}(t) + \lambda_{O\mathrm{Re}}P_{AOLp}(t) + \lambda_{SpLp}P_{A\mathrm{Re}Sp}(t);$$

$$\frac{dP_{Op}(t)}{dt} = \lambda_{LpOp}(P_{SYLp}(t) + P_{AYLp}(t) + P_{SOLp}(t) + P_{AOLp}(t) + P_{RALp}(t) + P_{RSLp}(t) + P_{A\mathrm{Re}Lp}(t) + P_{S\mathrm{Re}Lp}(t)).$$

The main goal of the model considered in this paper is to predict the optimal time to perform the rejuvenation procedure. Since the time distribution in models based on the Continuous-Time Markov Chain is exponential, the average transition time to a certain state is determined by the formula:

$$T_{avg} = \frac{1}{\lambda}. \qquad (3)$$

By calculating a system of differential equations (2) of model for different λ_{YR} values, the optimal rejuvenation time can be determined when the probability of being in the Active" and "Low Power" states is the lowest, i.e., when the rejuvenation procedure will not interfere with the user and will be performed until the battery is fully discharged.

4 EXPERIMENTS

To verify and analyze the proposed model of software aging with rejuvenation, the system of differential equations (2) was calculated by the 4th order Runge-Kutta method. The calculation of the system of equations for the original model [15] is also performed and it allows us to compare the values of 1/ λ_{YR} of both models and determine the influence of the factor on forecasting the rejuvenation procedure. Transition rates between all states of the system are approximate values which allow to simulate aging process and perform experimental calculations.

In work [15] two periods of user activity are considered, namely day (from 9 a.m. till 10 p.m.) and night (from 10 p.m. till 9 a.m.). Rates of transitions between "Active" and "Sleep" states for these intervals can be determined based on the collected user activity data. For example, you can divide the day into 15-minute time intervals, which will be denoted as "Active" or "Sleep" depending on the average number of events of the device sensors collected over many previous days [15]. In this study, 1 minute is used as a unit of time. The period of user activity in the day for analysis has no fundamental differences, so this paper assumes the verification of the model for the day period. Thus, we assume that the transition rates of λ_{AS} and λ_{SA} are 0.05 (average time of phone usage before entering sleep state is 20 minutes) and 0.02 (average time in sleep state is 50 minutes), respectively.

Transition rates from "Young" to "Old" state and from "Old" to "Recovering" state in [15] are considered equal and are based on increasing the average delay time of the user interface. These transition rates in a real system can be calculated based on the observed aging-related metrics, which show a performance degradation or an increase

DOI 10.15588/1607-3274-2021-4-13

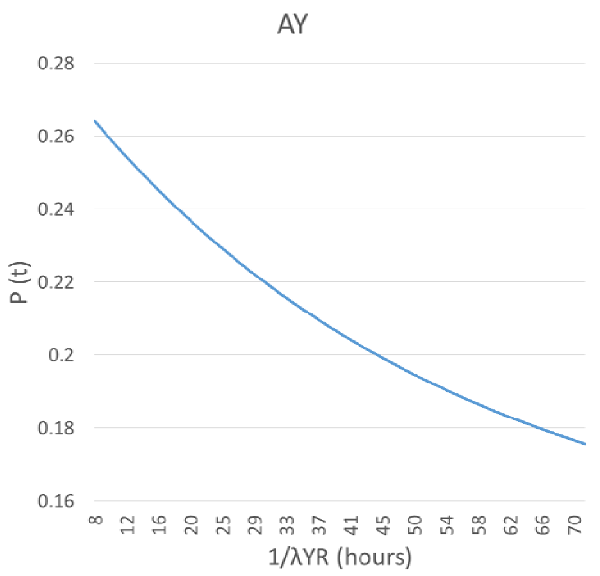
usage of system resources. However, in this paper, it is assumed that the transitions from "Young" to "Old" and from "Old" to "Recovering" last 100 hours, so $\lambda_{YO} = \lambda_{ORe} = 0.00017$. Rejuvenation procedure and recovery of the system after aging can be considered as a process of device rebooting, which lasts about one minute, so $\lambda_{ReY} = \lambda_{RY} = 1$.

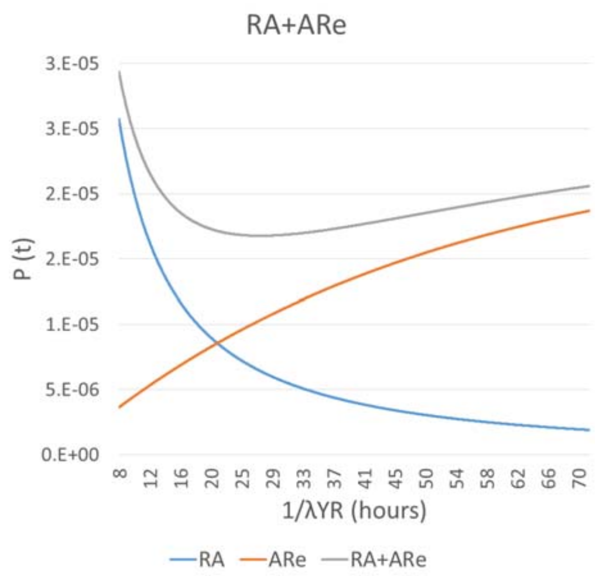
Transitions rates of battery charge model can be obtained based on remaining charge predictions at the time of calculations on the mobile device. In this work, it is assumed that the discharge time of the battery to a low level, and then to complete shutdown is 24 and 1 hour, respectively, i.e., $\lambda_{SpLp} = 0.0007$, $\lambda_{LpOp} = 0.017$. The time required to charge the battery to full level is assumed to be equal to 2 hours, i.e., $\lambda_{LpSp} = 0.0083$.

5 RESULTS

Calculations results of differential equations system of the original model without considering the battery charge are shown in Figure 6. Fig. 6 (a) shows the probability of being in the state AY, and Fig. 6 (b) shows the sum of the probabilities of being in the RA and ARe states.

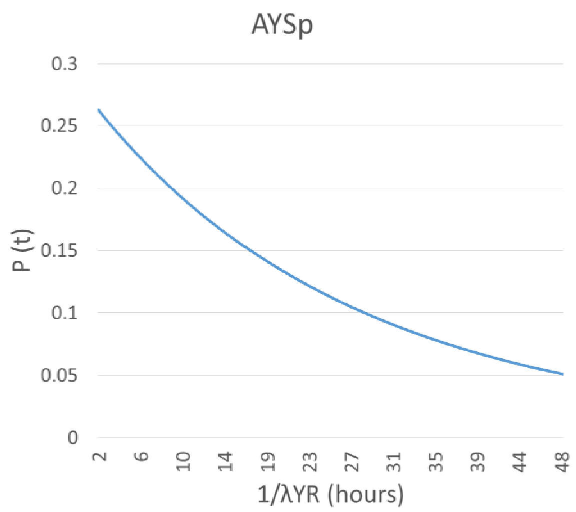
Figure 7 shows the results of calculations of model considering the battery charge. Fig. 7 a shows the probability of being in the state of AYSp, and Fig. 7 b shows the sum of the probabilities of being in the RALp and AReLp states, as well as the individual probabilities of being in these states.

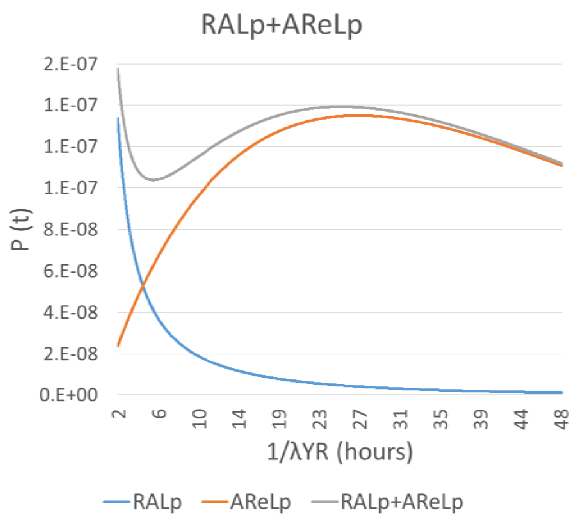




a b

Figure 6 – Calculations results of the model which doesn't take into account the battery charge:
a – the probability of state AY; b – the total probability of states RA and ARe





a b

Figure 7 – Calculations results of the model which considers the battery charge:
a – the probability of AYSp state; b – the total probability of RALp and AReLp states

DOI 10.15588/1607-3274-2021-4-13

Figure 8 shows the probability that the mobile device is in a discharged state.

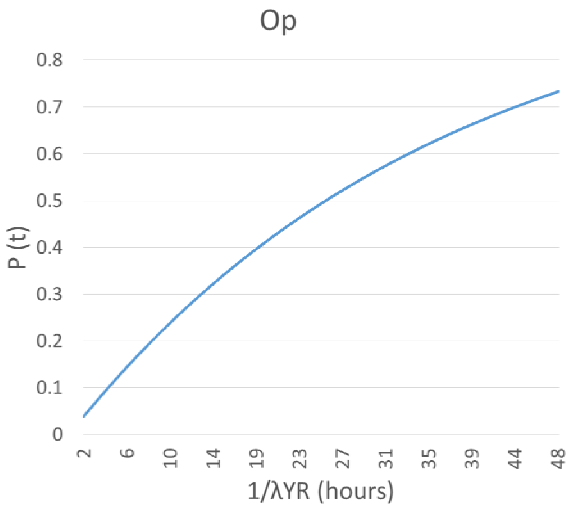


Figure 8 – The probability of state Op

6 DISCUSSION

Figures 6 (a) and 7 (a) show that as the time to the rejuvenation procedure increases, the probability of the system being in a young state decreases both considering the battery charge and without taking it into account.

The result of original aging model calculating (Fig. 6 (b)) shows that the optimal time to perform the rejuvenation procedure is 1 / λ_{yr} = 1620 minutes (27 hours), i.e., when the sum of the probabilities of being in the RA and ARe states is the lowest. In turn, the calculation of the model which considers the battery charge (Fig. 7 (b)) show that the rejuvenation procedure can be performed in 350 minutes (5 hours 50 minutes) after device start up. Comparing the results of calculations of both models, we can conclude that factor of battery charge has a significant impact on predicting the time of the rejuvenation procedure.

The idea of considering the battery charge factor is that performing a rejuvenation procedure is inefficient and impractical at a time when the battery level is critically low, or device is completely discharged. The formulated hypothesis is confirmed by the obtained calculations. If we do not consider the time to full discharge of the battery, the forecast of the original model is impractical, because after 27 hours (Fig. 8) the device is more likely to be discharged. The influence of the battery discharge can also be observed on Fig. 7 (b) where 1/ λ_{yr}>27.

It is worth noting that planning the rejuvenation procedure based on the battery level may not necessarily lead to performing this procedure at the selected time (5 hours 50 minutes), because at this time the system may still be in the state “Young”, and the battery may be recharged. Therefore, it is worth considering additional checks and new calculations of the model at scheduled time.

It is also necessary to note the reasons why the probabilities of being in the states “Recovering” and “Rebirth” are in the range of 10^{-7}. These two states are procedures that perform rebooting and restoring the system, so after performing them, the system immediately goes to the next state. At the same time, the probability of "Rebirth" is lower, because this state can be performed only if the mobile device is not used by the user and has a high battery charge. A high probability of being in these states would mean that the system is constantly or often in the process of rebooting, which is unacceptable for the user of the mobile device.

CONCLUSIONS

The paper describes software aging and rejuvenation model for mobile device which considers the factor of battery charge level.

The scientific novelty of the proposed mathematical model is to consider the new factor of battery charge level. The developed model is represented by a Continuous-Time Markov Chain, which combines the user behavior model, the model of software aging with the rejuvenation procedure, as well as the battery charge model.

The practical significance of using the battery level in the model to predict the rejuvenation time is to consider real mobile device usage scenarios, in particular, rejuvenation performing will not be effective or will not be performed at all in cases when the mobile device may have a very low charge or be completely discharged. In turn, early rejuvenation can delay the transition to a low battery level state, as rejuvenation will reduce the system workload and power consumption.

The future research is to investigate proposed model for different conditions and real data, because in this work a simulation was performed using test transition rates between states. Also, it is worth testing the hypothesis of the effect of the rejuvenation procedure performing before transition to the low battery state on the delay of the battery discharge by reducing the system workload.

ACKNOWLEDGEMENTS

The work is supported by the Ukrainian national state budget research project number 0121U109527.

REFERENCES

1. Parnas D. L. Software aging, *16th International Conference on Software Engineering, 1994, proceedings,* pp. 279–287. https://doi.org/10.1109/ICSE.1994.296790
2. Huang Y., Kintala C., Kolettis N., Fulton N. D. Software rejuvenation: Analysis, module and applications, *25th Symposium on Fault Tolerant Computing, 27–30 June 1995, proceedings*. Pasadena, California, 1995, pp. 381–390. https://doi.org/10.1109/FTCS.1995.466961
3. Grottke M., Jr R. M., Trivedi K. S. The fundamentals of software aging, *IEEE 19th International Symposium on Software Reliability Engineering : proceedings*, 2008, pp. 1–6.
4. Cotroneo D., Natella R., Pietrantuono R., Russo S. A Survey of Software Aging and Rejuvenation Studies, *ACM Journal on Emerging Technologies in Computing Systems,* 2014, Vol. 10, №1, Article 8. https://doi.org/10.1145/2539117
5. Valentim N. A., Macedo A., Matias R. A Systematic Mapping Review of the First 20 Years of Software Aging and Rejuvenation Research, *IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW),* 2016. https://doi.org/10.1109/ISSREW.2016.42

DOI 10.15588/1607-3274-2021-4-13

6. Yakovyna V. S., Uhrynovskyi B. V. Software aging in the context of its reliability: a systematic review, *Scientific Bulletin of UNFU*, 2019, №29 (5), pp. 123–128. https://doi.org/10.15421/40290525
7. Ahamad S. Study of software aging issues and prevention solutions, *The International Journal of Computer Science and Information Security*, 2016, Vol. 14, No. 8, pp. 307–313.
8. Maji A. K., Hao K., Sultana S., Bagchi S. Characterizing failures in mobile OSes: a case study with Android and Symbian, *International Symposium on Software Reliability Engineering. IEEE Computer Society,* 2010, pp. 249–258. https://doi.org/10.1109/ISSRE.2010.45
9. Cotroneo D., Fucci F., Iannillo A. K., Natella R., Pietrantuono R. Software aging analysis of the android mobile os, *IEEE 27th International Symposium on Software Reliability Engineering*, 2016, pp. 478–489. https://doi.org/10.1109/ISSRE.2016.25
10. Yakovyna V.S., Uhrynovskyi B. V. Software aging in the context of reliability: a review of the issue, *Scientific Bulletin of UNFU,* 2020, No. 30 (2), pp. 107–112. https://doi.org/10.36930/40300219
11. Smartphone OS market share [Electronic resource]. Access mode: https://www.idc.com/promo/smartphone-market-share/os
12. Wu H., Wolter K. Software aging in mobile devices: Partial computation offloading as a solution, *IEEE International Symposium on Software Reliability Engineering Workshops*, 2015, pp. 125–131. https://doi.org/10.1109/ISSREW.2015.7392057
13. Weng C. Zhao D., Lu L., Xiang J., Yang C., Li D. A Rejuvenation Strategy in Android, *IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW),* 2017, pp. 273–279. https://doi.org/10.1109/ISSREW.2017.50
14. Guo C., Wu H., Hua X., Lautner D., Ren S. Use Two-Level Rejuvenation to Combat Software Aging and Maximize Average Resource Performance, *IEEE 12th International Conference on Embedded Software and Systems*, 2015, pp. 1160–1165.
15. Xianga J., Wenga C., Zhaoa D., Tiana J., Xionga S., Lia L., A. Andrzejak A New Software Rejuvenation Model for Android, *IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), 2018*. https://doi.org/10.1109/ISSREW.2018.00021
16. Stochastic Petri net [Electronic resource]. Access mode: https://en.wikipedia.org/wiki/Stochastic_Petri_net
17. Norris J. R. Markov Chains, Cambridge University Press, 1997.

Received 30.08.2021.
Accepted 16.10.2021.

УДК 004.93

МОДЕЛЬ ПРОЦЕСУ СТАРІННЯ ТА ОМОЛОДЖЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ОПЕРАЦІЙНОЇ СИСТЕМИ ANDROID З УРАХУВАННЯМ ФАКТОРУ РІВНЯ ЗАРЯДУ БАТАРЕЇ

Яковина В. С. – д-р техн. наук, професор, професор кафедри Систем штучного інтелекту, Національний університет «Львівська політехніка», Львів, Україна; Факультет математики та комп'ютерних наук, Вармінсько-Мазурський Університет в Ольштині, Польща.

Угриновський Б. В. – асистент кафедри Програмного забезпечення, Національний університет «Львівська політехніка», Львів, Україна.

АНОТАЦІЯ

Актуальність. Особливість мобільних систем полягає в їх залежності від рівня заряду батареї, що є важливим чинником під час планування різного роду процесів, зокрема виконання процедури омолодження програмного забезпечення для зменшення впливу ефектів старіння цього програмного забезпечення.

Мета роботи. Розроблення моделі процесу старіння та омолодження програмного забезпечення для операційної системи Android з урахуванням чинника рівня заряду батареї.

Метод. Запропоновано комплексну модель на основі ланцюга Маркова з неперервним часом, яка об'єднює модель старіння із виконанням процедури омолодження програмного забезпечення, модель використання мобільного пристрою користувачем та фактор рівня заряду батареї. Побудовано граф станів та переходів, що описує об'єднані моделі. На основі діаграми написано систему диференційних рівнянь, яку обчислено з допомогою методу Рунге-Кутти 4-го порядку. Оптимальний час виконання процедури омолодження можна визначити за умов, коли її виконання не заважатиме користувачу та буде виконуватись завчасно до настання можливого повного розряду батареї, тобто тоді, коли ймовірність перебування системи в цих станах є найнижчою для певного значення часу виконання процедури омолодження.

Результати. Виконано симуляція розробленої моделі для тестових значень інтенсивностей переходів. Врахування моделі рівня заряду батареї дозволяє уникнути планування виконання процедури омолодження в час, коли мобільний пристрій з великою ймовірністю може мати низький заряд чи бути повністю розрядженим.

Висновки. Розроблена модель на основі ланцюга Маркова дозволить виконувати прогнозування часу початку процедури омолодження програмного забезпечення, враховуючи як поведінку користувача, так і рівень заряду батареї, який може мати значний вплив на прогнозований час. Також, раннє виконання процедури омолодження може мати вплив на зменшення навантаження на систему та відтермінування розряду пристрою, що варто перевірити в подальших дослідженнях. Обґрунтовано доцільність і важливість врахування чинника рівня заряду батареї і необхідність подальшого дослідження розробленої моделі старіння та омолодження із урахуванням нового чинника.

КЛЮЧОВІ СЛОВА: старіння програмного забезпечення, омолодження програмного забезпечення, ланцюг Маркова, операційна система Android.

УДК 004.93

МОДЕЛЬ ПРОЦЕССА СТАРЕНИЯ И ОМОЛОЖЕНИЯ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ОПЕРАЦИОННОЙ СИСТЕМЫ ANDROID С УЧЕТОМ ФАКТОРА УРОВНЯ ЗАРЯДА БАТАРЕИ

Яковина В. С. – д-р техн. наук, профессор, профессор кафедры Систем искусственного интеллекта, Национальный университет «Львовская политехника», Львов, Украина; Факультет математики и компьютерных наук, Варминско-Мазурский Университет в Ольштыне, Польша.

DOI 10.15588/1607-3274-2021-4-13

Угриновський Б. В. – ассистент кафедры Программного обеспечения, Национальный университет «Львовская политехника», Львов, Украина.

АННОТАЦИЯ

Актуальность. Особенность мобильных систем заключается в их зависимости от уровня заряда батареи, что является важным фактором при планировании различного рода процессов, в частности выполнения процедуры омоложения программного обеспечения для уменьшения влияния эффектов старения программного обеспечения.

Цель. Разработка модели процесса старения и омоложения программного обеспечения для операционной системы Android с учетом фактора уровня заряда батареи.

Метод. Предложено комплексную модель на основе цепи Маркова с непрерывным временем, которая объединяет модель старения с выполнением процедуры омоложения программного обеспечения, модель использования мобильного устройства пользователем и фактор уровня заряда батареи. Построен граф состояний и переходов, который описывает объединенную модель. На основе графа написано систему дифференциальных уравнений, которую вычислено с помощью метода Рунге-Кутты 4-го порядка. Оптимальное время выполнения процедуры омоложения можно определить в условиях, когда ее выполнение не будет мешать пользователю и будет выполняться заблаговременно до наступления возможного полного разряда батареи, то есть тогда, когда вероятность нахождения системы в этих состояниях является самой низкой для определенного значения времени выполнения процедуры омоложения.

Результаты. Выполнено симуляцию разработанной модели для тестовых значений интенсивностей переходов. Учет уровня заряда батареи позволяет избежать планирования выполнения процедуры омоложения в то время, когда мобильное устройство с большой вероятностью может иметь низкий заряд или быть полностью разряженным.

Выводы. Разработанная модель на основе цепи Маркова позволяет выполнять прогнозирования времени начала процедуры омоложения программного обеспечения, учитывая как поведение пользователя, так и уровень заряда батареи, который может оказать значительное влияние на прогнозируемое время. Также, раннее выполнение процедуры омоложения может влиять на уменьшение нагрузки на систему и отсрочку разряда устройства, что стоит проверить в дальнейших исследованиях. Обоснована целесообразность и важность учета фактора уровня заряда батареи и необходимость дальнейшего исследования разработанной модели старения и омоложения программного обеспечения с учетом нового фактора.

КЛЮЧЕВЫЕ СЛОВА: старение программного обеспечения, омоложение программного обеспечения, цепь Маркова, операционная система Android.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Parnas D. L. Software aging / D. L. Parnas // 16th International Conference on Software Engineering, 1994: proceedings. – P.279–287. https://doi.org/10.1109/ICSE.1994.296790
2. Software rejuvenation: Analysis, module and applications / [Y. Huang, C. Kintala, N. Kolettis, N. D. Fulton] // 25th Symposium on Fault Tolerant Computing, 27–30 June 1995: proceedings. – Pasadena, California, 1995. – P. 381–390. https://doi.org/10.1109/FTCS.1995.466961
3. Grottke M. The fundamentals of software aging / M. Grottke, R. M. Jr, K. S. Trivedi // IEEE 19th International Symposium on Software Reliability Engineering : proceedings. – 2008. – P.1–6.
4. Cotroneo D. A Survey of Software Aging and Rejuvenation Studies / D. Cotroneo, R. Natella, R. Pietrantuono, S. Russo // ACM Journal on Emerging Technologies in Computing Systems, 2014. – Vol. 10, № 1. – Article 8. https://doi.org/10.1145/2539117
5. Valentim N. A. A Systematic Mapping Review of the First 20 Years of Software Aging and Rejuvenation Research / N. A. Valentim, A. Macedo, R. Matias // IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), 2016. https://doi.org/10.1109/ISSREW.2016.42
6. Яковина В.С. Старіння програмного забезпечення в контексті його надійності: огляд проблематики / В. С. Яковина, Б. В. Угриновський // Науковий вісник НЛТУ України. – 2019. – №29 (5). – С. 123–128. https://doi.org/10.15421/40290525
7. Ahamad S. Study of software aging issues and prevention solutions / S. Ahamad // The International Journal of Computer Science and Information Security. – 2016. – Vol. 14, № 8. – P. 307–313.
8. Characterizing failures in mobile OSes: a case study with Android and Symbian / [A. K. Maji, K. Hao, S. Sultana, S. Bagchi] // International Symposium on Software Reliability Engineering. IEEE Computer Society. – 2010. – P. 249–258. https://doi.org/10.1109/ISSRE.2010.45
9. Software aging analysis of the android mobile os / [D. Cotroneo, F. Fucci, A. K. Iannillo et al.] // IEEE 27th International Symposium on Software Reliability Engineering. – 2016. – P. 478–489. https://doi.org/10.1109/ISSRE.2016.25
10. Яковина В. С. Старіння програмного забезпечення мобільних додатків: аналіз проблематики / В. С. Яковина, Б. В. Угриновський // Науковий вісник НЛТУ України. – 2020. – №30 (2). – С. 107–112. https://doi.org/10.36930/40300219
11. Smartphone OS market share [Electronic resource]. – Access mode: https://www.idc.com/promo/smartphone-market-share/os
12. Wu H. Software aging in mobile devices: Partial computation offloading as a solution / H. Wu, K. Wolter // IEEE International Symposium on Software Reliability Engineering Workshops. – 2015. – P. 125–131. https://doi.org/10.1109/ISSREW.2015.7392057
13. Weng C. A Rejuvenation Strategy in Android / [C. Weng, D. Zhao, L. Lu et al.] // IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW). – 2017. – P. 273–279. https://doi.org/10.1109/ISSREW.2017.50
14. Guo C. Use Two-Level Rejuvenation to Combat Software Aging and Maximize Average Resource Performance / [C. Guo, H. Wu, X. Hua et al.] // IEEE 12th International Conference on Embedded Software and Systems. – 2015. – P. 1160–1165.
15. Xianga J. A New Software Rejuvenation Model for Android / [J. Xianga, C. Wenga, D. Zhaoa et al.] // IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW). – 2018. https://doi.org/10.1109/ISSREW.2018.00021
16. Stochastic Petri net [Electronic resource]. – Access mode: https://en.wikipedia.org/wiki/Stochastic_Petri_net
17. Norris J. R. Markov Chains / J. R. Norris // Cambridge University Press. – 1997.

DOI 10.15588/1607-3274-2021-4-13

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

UDC 519.2

ENTROPY APPROACH IN SYSTEM RESEARCH OF DIFFERENT COMPLEXITY OBJECTS TO ASSESS THEIR CONDITION AND FUNCTIONALITY

Kozulia T. V. – Professor, Dr. Sc., Professor of Software Engineering and Information Technology Management department National Technical University "Kharkiv Polytechnic Institute".

Sviridova A. S. – Master of Software Engineering and Information Technology Management department National Technical University "Kharkiv Polytechnic Institute".

Kozulia M. M. – PhD, Associated Professor of Software Engineering and Information Technology Management department National Technical University "Kharkiv Polytechnic Institute".

ABSTRACT

Context. Consideration of comprehensively studied object in the form "system – environment" to obtain an approximate accurate real situation reflection.

Objective. Search for solutions to problematic research issues based on the entropy approach for systems "object – environment" of different nature and complexity, studying them and obtaining knowledge (stable information) and providing them as a set of complex system tasks modulated by different entropy functions.

Method. The following criteria are used to assess the sustainability of the development of a system object: integrity – the failure of the trajectory of development of the object at a certain forecast time interval from a set of safe states; monotony of growth of indicators of development of object on a certain time interval with the subsequent preservation of them in the set intervals of admissible values; compliance of the development trajectory with the target changes according to the requirements of safety and sustainable development, resistance to disturbance, including asymptotic stability of the program trajectory and structural stability of the system.

In the conditions of nonlinear development of events and spontaneity of processes "object – external systems" at stable structure of system object of research it is expedient to apply the entropic approach and knowledge from the field of the theory of stability developed for technical and cybernetic systems.

Results. The proposed entropy approach to analysis is determined by the fact that the object is characterized from the standpoint of compliance with acceptable regulatory constraints and processes regarding the acceptability of the object of the external environment or the possibility of resolving the situation of coexistence "object – environment".

Within the analysis of a system object, this means that for both stationary and dynamic conditions, their state is described by a certain function, the changes of which indicate the approach to a certain point of homeostatic relations with the environment.

The practical application of the provided methodological proposal for finding solutions in conditions of uncertainty of a certain kind is considered on the example of determining measures to influence the course of positive development of the child's body in the situation of diagnosis of cerebral palsy in the form of information and software application at realization of the appointments of medical character applied to them (factors of influence of emergency).

Conclusions. The proposed entropy approach to the choice of decision-making problems for determining the state and changes as a result of process transformations in system objects of the type "studied system – environment" in conditions of uncertainty does not require additional conditions characteristic of known estimates by criteria in common mathematical means of decision making.

KEYWORDS: entropic approach, entropy-information estimates, software application, complex objects, WindowsForms technology.

ABBREVIATIONS

ApEn is an entropy of approximation;
BESA is a Burg entropy spectral analysis;
BG is a blood glucose;
CESA is a configurational entropy spectral analysis;
CP is a cerebral palsy;

DOI 10.15588/1607-3274-2021-4-14

DM is a decision making;
EEG is an electroencephalogram;
EnDEMO is an ensemble double entropy and multi-purpose optimization;
EP is an episyndrome;
FFA is a frequency floods analysis;
HD is a Halphen distribution;
IGH is an internal gastric hemorrhage;
ME is a maximum entropy;
MSE is a multiscale entropy approach;
PV is a periventricular;
PVI is a periventricular ischemia;
PVL is a periventricular leukomalacia;
RMSE is a root mean square error;
SampEn is a sampling entropy algorithm;
SpecEn is a spectral entropy.

NOMENCLATURE

ξ is a variable of system state information;

$\mathrm{MS}(\xi \mid \eta)$ is a variable of mathematical expectation of conditional entropy;

$\mathrm{M\Delta S}(\xi \mid \eta)$ is a variable of change of conditional entropy at variable ξ;

$\mathrm{N}(\mathrm{Y}_a)$ is variable of the number of the set Y_x elements;

ΔN_i is a variable of changes in the number of elements in the object of the *i*-th type in the total number of changes in the system ΔN;

$S(\xi \mid \eta)$ is a variable of conditional entropy η at $\xi = x$;

$\Delta S(\xi \mid \eta)$ is a variable of changes in conditional entropy η at $\xi = x$;

k is a proportionality coefficient, Boltzmann constant, $k = 1.38\ 10^{-23}$ J / K;

$\Delta\Omega$ – changes in the number of possible microstates (methods) by which you can make a given macroscopic state of the system, identifying the number of system microstates, provided that all microstates are equally likely.

INTRODUCTION

Scientific activity of deep study and knowledge of the world in general is based on the use of achievements of domestic and world science and technology to meet social, economic, cultural and other needs, regulated in Ukraine by the Law of Ukraine “On scientific and scientific-technical activities” of 26.11.2015 № 848-VIII. The difficulty of matching the data obtained by calculations or modeling to the real states of complex systems is that most of the complexity is considered in terms of dimensionality or structural properties of the studied objects. As the experience of modern monitoring (information-analytical) systems related to the study of complex objects, such as different in the hierarchy of ecological systems [1–4], the complexity is generally appropriate to consider with the peculiarities of the interaction of such an object with the environment, which consists of systems set of different nature, in direct or indirect contact with the components and elements of such an object. The study of such system objects is based on the accuracy of their model representation and the clarity of the definition of the target position for structuring as a complex complex of cooperative nature, ie independent functioning systems structurally linked to achieve this goal.

For most of the designed systems in modelling are problematic issues of the cognitive process, which have two directions of scientific research evolution, namely obtaining a result from less to more meaningful, from less to more general knowledge. Establishing a unanimous answer, making a meaningful decision at the maximum breadth of reality is based on the unity of knowledge, due to the unity of the world, to obtain an adequate model of the surrounding reality. Complex system objects and attempts to solve problems require a special interdisciplinary unifying approach, taking into account the only function of the state of any system – the entropy function. The only approach in the systemological analysis of the initial monitoring data and uncertainty about the relationship of such systems with the environment in relation to the interaction between them and changes accordingly properties and state of systems is defined as an entropic approach.

In previous works, the authors of the article, for example [5], proposed a combination of methods of information-entropy type to resolve complex study issues of complex objects, taking into account their interaction with the environment. The research proposes to introduce in quality problems not so much a complex methodological entropy basis for studying complexly structured systems, but to use the so-called entropy measure of conformity on the basis of functions of different entropies and to connect the phenomenological component to the analytical system of state and complexes functionality in the form of knowledge-oriented databases, where the data is existing knowledge and obtained in the research course [6–8].

Thus, the **purpose** of finding solutions to problematic issues in any field of research based on the entropy approach calculations is determined by the need to consider a comprehensive “object – environment” for systems of different nature and complexity, studying them and obtaining knowledge (stable information) should be provided in as a set of problems of a complex system, modulated by different entropy functions.

The paper proposes to consider the application of the entropy approach calculation to modeling states and processes for the system object “system – environment”:

1) theoretical substantiation of the system of entropy functions as a methodological basis of knowledge-oriented analysis of monitoring data to find target solutions for the study of system objects taking into account the process phenomena of internal and external nature;

2) testing of the proposed entropy analytics of system analysis for stabilization of the human body state.

Models and methods of complex system research associated with the environment is an article **object**.

DOI 10.15588/1607-3274-2021-4-14

Subject is entropy approach in system research of different complexity objects to assess their condition and functionality

Methodological basis of a comprehensive study of any object as a systemic formation of the type "object – environment – (object – environment) – processes of internal and external nature – (object – environment)^2 – object^3" it is advisable to build on a universal approach that covers all stages of systematic analysis of monitoring data and allows you to state the acquisition of knowledge about the object after the cognitive process.

1 PROBLEM STATEMENT

Within the framework of the analysis of the system object "researched system – environment" the entropic estimation of a condition of the specified components, their mutual influence and processes is offered. This means that for both stationary and dynamic conditions the state is described by an entropy function, changes in which indicate the approach to a certain point of homeostatic relations with the environment ($S_\varepsilon(K)$).The input data are the results of monitoring information regarding the determination of the characteristic parameters of the situation description in the system "object – environment" and "object system – internal object environment" ($C_\varepsilon(K)$) that meet the objectives of environmental research. In any case, the interaction between systems ξ (set X) and η (set Y) regarding the requirements for the space of their existence $X \times Y$ implemented on many possible pairs U at $a \in X$ for those y from Y_a provided $(a, y) \in U$. Then the information entropy from the description of the study object is determined as follows

$$\Delta S(y \mid a) = \log_2 N(Y_a), \quad (1)$$

where $N(Y_a)$ – the number of elements of the set Y_x, that is, those elements that perceive the action and have a reaction to it within the objectives of the study: $a = \overline{1, n}$, $n \in N$

System state information ξ (in x) regarding available information about η (relatively y) is set by the formula

$$I(x : y) = \Delta S(y) - \Delta S(y \mid x). \quad (2)$$

If there is data for the system object from different monitoring media sources, which consists of unrelated or weakly related information (indicators, factors, measurement parameters, etc.), subject to certain probabilistic laws, in practice it is advisable to introduce a probabilistic approach. Within its limits mixing of probabilities and frequencies is allowed at considerable on time and volume of supervision; formation of mathematical expectation for entropy $MS_W(y \mid x)$ and information $MI_W(x : y)$, the value of which can take an excellent value (in the combinatorial approach, it is always a positive value, which should be when you imagine the amount of information). According to A. N. Kolmogorov [25], the true measure of the amount of information is the average value $I_W(x, y)$, characterizing the tightness of the connection between the systems ξ and η, state parameters x, y symmetrically: $S_W(x \mid x) = 0$, $I_W(x : x) = S_W(x)$, $I_W(x, y) = MI_W(x : y) = MI_W(y : x)$ despite the fact that $S_W(y \mid x)$ and $I_W(x : y)$ are functions from x. For system analysis, values are introduced in accordance with the following accepted definitions.

The amount of information to establish the exact value ξ in the presence of a known and sufficient amount of values $\eta = y_j$ equals

$$S(\xi \mid \eta = y_j) = -\sum_i p_{i|j} \log_2 p_{i|j},$$

(in the proposed version of the analysis according to fig. 2 $\Delta S(\xi \mid \eta = y_j) = -\sum_i p_{i|j} \log_2 p_{i|j}$, where $p_{i|j}$ incompatible probability distribution $P\{\xi = x_i, \eta = y_j\} = p_{ij}$, and the difference between the probability of obtaining data / information about the studied object and the actual one with clearly defined properties or target, ie

which is on average

$$MS(\xi \mid \eta) = -\sum_j P(\eta = y_j) \sum_i p_{i|j} \log_2 p_{i|j},$$

on offer

$$M\Delta S(\xi \mid \eta) = -\sum_j P(\eta = y_j) \sum_i p_{i|j} \log_2 p_{i|j}, \; p_{i|j} = \Delta p_{i|j} \quad (4)$$

[24]–[25].

The amount of information relatively ξ, contained in the results of the previous stage (task) or determined by the characteristics of the system η, is equal to the difference

$$I(\xi \mid \eta) = S(\xi) - MS(\xi \mid \eta);$$

and informational changes about the studied object will be $\Delta I(\xi \mid \eta) = \Delta S(\xi) - M\Delta S(\xi \mid \eta)$ according to the obtained observation data

$$I(\eta \mid \xi) = -\sum_i p_{ij} \log_2 \frac{p_{ij}}{P\{\xi = x_i\} P\{\eta = y_j\}}, \quad \Delta I(\eta \mid \xi) = f(\Delta p_{ij}) \quad (5)$$

According to equations (2)–(5) we have the final calculated definitions of the information background of the studied questions:

DOI 10.15588/1607-3274-2021-4-14

$$\Delta S_W(x) = -\sum_x \Delta p(x) \cdot \log_2 \Delta p(x);$$
$$\Delta S_W(y \mid x) = -\sum_i \Delta p(y \mid x) \cdot \log_2 \Delta p(y \mid x); \quad (6)$$
$$\Delta I_W(x : y) = \Delta S_W(y) - \Delta S_W(y \mid x).$$

Thus, when determining an object state according to statistical data, sets of microstates of the system are considered according to the statistical function of Boltzmann entropy in order to determine the probability of obtaining a macroscopic system realization. When studying the system "object – environment" it is advisable to have information on the change in the probability of supporting the implementation of the macro-situation, which is responsible for the target equilibrium in the space of states and processes in such a system (see Fig. 2):

$$\Delta S = k \cdot \ln \Delta \Omega . \quad (7)$$

Subject to changes in the system, when the object interacts with the environment, certain results are obtained N-probable consequences (outputs) p_n, of which interesting are the modified ΔN, determining the realization of equilibrium or purpose in accordance with changes in the Hartley entropy:

$$\Delta S = \log_2 \Delta N . \quad (8)$$

This sequence of obtaining a chain of entropy changes in the analysis of states and processes allows us to move in the direction of probable trajectories in space "object – environment", leading to the goal or desired development in such a system or object in its current conditions. Thus, the level of knowledge about the system state (final decision) is defined as Shannon's information entropy for independent changes in events X, which corresponds ΔN-altered probabilistic states described by probabilities $p_n \left(n = \overline{1, N}\right)$:

$$\Delta S_x = -\sum_{n=1}^{N} p_n \cdot \log_2 p_n = -M\left[\log_2 \Delta p_n\right]. \quad (9)$$

The final result of the search for a solution is evaluated as follows:

– the minimum probability of change, which corresponds to the implementation of the required macrostate with a single probability;

– the maximum probability of change, which leads to a uniform distribution of the probability of realization of states close to the solution $\Delta n = \overline{1, \Delta N}, \Delta p_n = 1/\Delta N$;

– zero for other cases of changes in statistical entropy according to Shannon:

$$\Delta S_x = -\sum_{s=1}^{l} \frac{\Delta N_i}{\Delta N} \cdot \ln \frac{\Delta N_i}{\Delta N} . \quad (10)$$

The final state is determined by the structural entropy, according to which the system's desire for equilibrium is formed ($\Delta S \to 0, S_1 \to \min \to \Delta S > 0, S_2 \to \max$)

For all solutions, the system or object must reach the *attractor* – a structure (function) that sets (determines) the stable system state (synergetics, nonlinear thinking). Thus, for any statement of the problem of ecological assessment of the situation for the studied object "system – environment" it is important to establish the current situation as a starting point for establishing the changes occurring, by probabilistic estimation through the entropy function taking into account the interaction between object components at the level of "impact – response" to increase the information field of the study and achieve real and accurate results.

2 REVIEW OF THE LITERATURE

Complex systems study, which are open and multidimensional formations, is based on the creation of models taking into account their stochastic nature of behavior and the relationship with the elements. System analysis of such objects is proposed to be carried out on the so-called entropy model, which is a promising direction for studying complex systems with ambiguous or probabilistic behavior and building methods for their study. The use of the entropy function as a universal function of the systems state and processes is often considered in the development of methods for identifying the properties of studied objects and in their modeling to study and create projects for the use of natural resources for economic purposes [9–12].

The authors of recent generalized publications on the modeling of complex systems argue that complexity is an integral feature of the system. In general, any object of study consists of elements set, and the more elements it contains, the more complex it is. However, the number of elements that make up the system is not the only factor of complexity. Not only the number of elements is a factor, but also the number of interactions between these elements, as well as how the intensity of these interactions should be taken into account when analyzing the complexity of any formation level. The number of interactions with the system environment and their intensity should be taken into account in the analytical consideration. An important factor is the functions delivered by the system and their status (ie stability in the performance of activities and security of the system) [2]. There are two main types of system complexity:

– internal complexity: refers to the system structural complexity – is a function of the elements number, the number of interactions and the intensity (or strength) of these interactions;

– external complexity: refers to the system / complexity of the interface environment – is a function of the interactions number between the system and its environment, the intensity of these interactions and their probabilistic expected performance functions.

DOI 10.15588/1607-3274-2021-4-14

In such studies, Shannon's information entropy is proposed as an indicator of complexity. Its value depends on the detail level of the object study, the number of elements in a particular system and the relationship between them, the hierarchy levels number, and so on. As shown in the articles, according to Shannon's information entropy, smaller sets mean less complexity. The choice of aggregation allows you to deal with subsets separately to determine the complexity. To do this, it is proposed to focus on the complex system properties and the needs and goals of the study of complex systems through the process of modeling at the stage of problem identification and focus on the diagnosis of the interests system.

Since the complexity requires an unambiguous structure, it is proposed to clearly define the properly measured basic indicators of system properties based on calculations of the model complexity of the form Higraph. This model allows you to display the complexity of the system in accordance with the relationship of the Hygraph components and simplify it by aggregating in a useful way according to the constraints without losing the content of the studied system.

Despite the developed appropriate proposals for working with complex systems, it remains a problem to organize the elements at each level of the hierarchy or study stage so that the results correspond not only to the holistic model of the analyzed complex object, but also to real monitoring or practice. This applies to the need to maintain the amount of information load on the system element in any research schedule, ie not to provide conditions for restrictions, but to take into account the internal connections and interactions with the surrounding environmental systems.

In the study of ecological systems and the use of natural resources, an entropy-based index is proposed to establish the optimal type of hydraulic system and its condition for water softening to avoid natural and man-made hazards [11–12]. Entropy as a measure of the system state of natural and man-made species is proposed by the authors as a modified indicator to assess the suitability of water distribution network, where there are hydraulic uncertainties, such as flow rate in the pipes, uncertainty due to mechanical parameters, the probability of failure that consider simultaneosly [11]. The entropy measure in the proposed method of determining the quality of network demand is calculated by establishing a coefficient based on the ratio of nodal requirements to the total costs of all parts of the network. The probability of failure of hydraulic systems is also based on the entropy function assessment of the hydraulic network state based on the probability of their failure in any specified hazard scenario using the penalty function, which is satisfactorily inserted into the existing entropy function. This comprehensive approach in solving problems of quality assessment of a complex system object based on one entropy function allowed to combine the assessment of mechanical behavior of units in a functioning hydraulic network, finding the optimal hydraulic scheme for designing a new system, deciding on the best water softening plan for existing network, exposed to various hazards of natural and man-made origin. For this purpose, the entropy-based index is proposed as an effective tool for comprehensive assessment of the entire hydro network state, which, however, does not provide information about the causes of hazards and their consequences. Therefore, it is still necessary to position the object of study in an appropriate manner to the environment and interaction with external systems, taking into account changes over time in the object.

Similarly, Canadian scientists [12] propose to overcome the uncertainties arising from data processing such as time series when modeling situations from the state of hydrometric networks based on the use of entropy as a methodology of so-called EnDEMO, which takes into account the uncertainties from the generation of the input data ensemble. The proposed offers for the Endemo approach provided a more reliable result by creating an ensemble to account for uncertainty in the simulated time series datasets, provided opportunities to identify more reliable locations for potential stations.

To solve the problem of environmental pollution and rational use of environmental resources in the socio-economic aspect, entropy is considered as a component of analysis methods, in accordance with the common thermodynamic nature of ecological, socio-economic systems and the uniqueness of entropy maximization in achieving such targets or equilibrium state. it is proposed to introduce an entropy function to determine compliance with the safety requirements of the analyzed natural and man-made objects both separately and in their interaction [13–15].

Thus, scientists at Wuhan University [13] proposed solutions to the rational use of water resources and environmental and man-made safety through the application of the principle of ME for frequency floods analysis (FFA), which is the basis for determining the size of hydraulic structures. In this study, entropy properties as a measure of chaos / equilibrium of the system are proposed to be considered for a comprehensive analysis of the studied object state in combination with Halphen distribution (HD) in Monte Carlo modeling was used to evaluate descriptive and predictive abilities of suggested procedure HD/ME which have good descriptive ability due to its flexible shape and excellent tail properties, least RMSE and higher efficiency compared to the procedure in several simulated cases. The proposed HP / ME procedure is well suited for estimating project floods in two selected areas and is offered as an alternative candidate for hydrological frequency analysis, but in addressing flood behavior, this approach is simplified and static for situation analysis.

Similar proposals for the introduction of entropy in the analytical system of research were provided by scientists from the Department of Biological and Agricultural Engineering at the University of Texas and the College of Water Resources and Architectural Design of Northwestern University in China [14]. Water flows are complex systems of stochasticity, seasonality, and periodicity that are proposed to be studied using configurational entropy spectral analysis (CESA), which combines spectral analy-

DOI 10.15588/1607-3274-2021-4-14

sis and time series analysis to characterize the periodically obtained picture of the situation and the corresponding manifestations of stochasticity.

In the field of water resources management, irrigation and estimation for forecasting monthly groundwater levels [15], it is proposed to introduce the domain of entropy determination into the methodological support of spectral analysis – this is the frequency domain in which three types of entropy are known. This entropies from the basis for three types of spectral analysis in research predicting monthly and annual groundwater levels, which is a stochastically nonequilibrium system, namely (1) BESA, (2) CESA and (3) relative spectral analysis. In general, the proposal concerns not so much the complexity of entropy research as the possibility of comparing different parameters of groundwater state based on their representation as an entropy function.

Special natural biological objects of ecological research include the human body, which is considered as a complex stochastic system formation, in which the systems are functionally and corporately combined to ensure human life in certain environmental conditions. To establish the state of such a systemic object as an organism, it is proposed to use various entropic functions and methods of entropy analysis in order to diagnose changes and predict the consequences of effects on its functionality at the violations risk [16–22].

Quantitative assessment of irregularity (lack of order) in the obtained measurement data or monitoring observations of the system state or system object is overcome by entropy spectral analysis, which in recent decades has been innovatively improved to remove some limitations for studying biological signals such as entropy approximation ApEn and SampEn. The latter are used to study endocrine-metabolic functionality and heart rate, EEG, vocals and hemodynamics of time series.

Stochastic and unpredictable components of the system order for diagnosing the state of body systems due to EEG are proposed to be considered when estimating changes in the amplitude component of the EEG power spectrum at each frequency as probabilities in entropy calculation and as information on temporal changes in EEG signal. Such propositions apply to so-called SpecEn, based on Shannon entropy in the form of Fourier transform peaks as a reflection of the regularity of the relative spectral power density of the EEG. In statistical studies to obtain short and noise data sets, such as biomedical signals, the ApEn is used to study time series for similar data, taking into account N samples (points). ApEn (m, r, N) measures the logarithmic probability of close (within r) sections of m models of continuous observations. Thus, in the entropy analysis of background EEG activity in patients with Alzheimer's disease, ApEn remain close within the same tolerance width r in the following comparisons. For a sequence with larger response values corresponding to greater complexity or irregularity in the data, a non-negative number is assigned to avoid the appearance of ln (0) in the calculations. To avoid the disadvantages of such data analysis, the calculation of SampEn is proposed. SampEn (m, r, N) for certain parameters of the path length m and the tolerance r as the negative logarithm of the conditional probability that two sequences similar to m points remain similar in the next paragraph, where the coincidence is not included in the probability calculations. Thus it is possible to avoid dependence on the length of the EEG recording and to obtain relative consistency under certain circumstances, to reduce the computation time.

Thus, Shannon's entropy formula is accordingly normalized and considered convenient as a way to quantify the distribution of spectral power in spectral analysis to identify objects of study, such as the state of the body in Alzheimer's disease [16]. Quantitative determinations of traditional BG values based on linear spectral analysis are performed using a nonlinear MSE method, which determines the signal complexity using a popular SampEn to measure its inequalities in various time scale [17].

In the system data processing of system state in spectral analysis uses the so-called optimal threshold to distinguish between AD and normal EEG signals, which is best obtained by the concept of fuzzy entropy defined in a complex wavelet with multiple resolution [18]. This methodology allows to obtain a high signal-noise ratio and a lower root mean square error than with the traditional approach.

Socio-environmental issues in the study of safety monitoring are proposed to be considered through the prism of entropy to overcome its increase in the infection control system to find internal and external, micro- and macromotivators that maintain a stable institutional focus and optimal allocation of resources [19]. Prioritization, new destabilization factors and lack of resources accompany the monitoring of objects as a control system of specific intervention in the presence of the inevitable tendency to increase entropy in an endless cycle of periodic activity, modernization of the system in a constant feedback cycle with the external environment.

Thus, overcoming the disorder and nonlinear system signals (responses) of the systems studied on the aspect of studying their state and functionality is considered in accordance with the properties of changes in entropy function of various kinds, which allows to obtain heterogeneous knowledge of the tasks. Due to the combinatorial nature of entropy, cognition in the study of objects concerns not only the statics of its existence, but also the processes that support it or lead to its reconstruction. In this case, we are dealing with the enthalpic part of the change in free energy, the modulation of thermal oscillations as the fraction of free energy in entropy. The process of studying stochastic interactions between more elements in the system and between their systems requires modeling a portrait of events within the object, a complete picture of interactions with entropy forces as a universal feature of self-organization of systems and the use of several areas of knowledge that often remain unconnected. obtaining the integrity of the phenomena perception. This applies to the study of both micro- and macrosystems, such as protein structures and human behavior, which is a conse-

DOI 10.15588/1607-3274-2021-4-14

quence of emotional states in areas of the brain [20–21]. The results of the review show that a combination of applied spectral, entropy and time functions has positive results in such works, which can provide and transmit reliable biomarkers for determining the profiles of SS, ES and TS for a complete description of a person and intervention in his brain.

Thus, the experience of research of complex systems to study their properties and some process phenomena has shown the benefits of using the entropy approach, which will be appropriate in the study of system objects of the type "system – environment" for learning the system and DM regarding the achievement of the target state by this system, ie regulation as a directed influence from environmental systems.

3 MATERIALS AND METHODS

Software applications have been developed for entropy analysis, for example, for statistical monitoring studies on entropy analysis of time series, EZ Entropy in MATLAB environment in object-oriented style with graphical user interface has been proposed [22]. This application offers various stages of analysis, including (1) data recording processing, (2) batch processing of several data files, (3) calculation windows, (4) revocation, (5) display of intermediate data and final results, (6) adjustment input parameters and (7) export the results of calculations after startup or in real time during the analysis.

In accordance with the proposed entropy approach to the study of the object "system – environment" introduced an experimental model of transition from baseline to target in accordance with the identified factors of significant influence on the development of events in the system with internal changes and external interaction with the environment, namely "system – (system – environment) – process – the situational state of the system", based on logic, knowledge and experience [23–24].

In the building process of a complex study object model, one of the important issues is to substantiate the probability of its implementation, compliance with functional sustainable development "system – environment", which is achieved if the system model is stable and itself in response to environment, ie maintaining equilibrium behavior. The latter is characterized by the reflection of the real situation through a logical sequence of causal relationships, ie a system of dependent events, processes, phenomena, states of use of conditional entropy. In this case, the interconnected systems and processes are determined by combining into a hypersystem or system formation, which is the "memory" of the system by estimating the rate of information loss, the degree of chaos of the system, namely Kolmogorov entropy or K-entropy. Gradual changes in the time of the system are not related and do not determine information about the current and future state, it is only a reflection of gradual movement due to transformations of a certain degree of determinism or stochasticity.

For static analysis, it is natural to evolve the Boltzmann entropy, defined as a measure of the information incompleteness about the system, into Shannon's information entropy, which is the basis of quantitative information theory [9]. The study of the functionality over time of complex system formations, processes in them and external interactions cannot be considered simply on the basis of their modeling by studying discrete random events in the form of probability distribution and information to determine the required event in such space – Hartley entropy.

The system movement in the space of time and states in accordance with the situation of its environment occurs to a certain position of immutability or to a certain target state, naturally conditioned or desired or planned. Achieving such a position will be determined by the Rainier entropy according to the definitions of the Kulbak-Leibler distance as information divergence, relative entropy, distance from the true distribution of events or states. If in ecology these are diversity indices, then in the case of studies of a damaged system or a system that moves away from natural development (for example, the human body or its individual systems), it is a measure of the complexity of the situation according to the introduced range indicator, divergence.

System approximation to the natural target occurs through certain microstates, a number of which are not additive for entropy generalization, as indicated in the Tsallis entropy. This allows the study of complex system objects to provide statistical observations as descriptions of states, which take into account the interaction of system elements with neighboring neighbors, the system as a whole and its components.

Thus, the representation of a complex system in the form of an entropy model allows you to fully take into account statistics about it and consistently resolve uncertainties about the presence of ignorance of internal processes and existing interactions with external systems. With this approach to obtaining information about the studied object, the entropy of the multidimensional stochastic system "object – environment (external systems)" is calculated, which consists of two components – the boundary entropy, which determines the independence of the object integrity in the system at available internal systems (additivity of the entropy state function) and the entropy of the relationships degree between the system elements, ie processes in the space of coexistence of the elements of the stochastic system (fig. 1).

The following criteria are used to assess the sustainability of a system object:

1) integrity – the failure of the trajectory of object development at a certain forecast time interval from the set of safe states;

2) monotony of growth of object development indicators on a certain time interval with the subsequent preservation of them in the set intervals of admissible values;

3) compliance of the development trajectory with targeted changes in terms of security and sustainable development;

DOI 10.15588/1607-3274-2021-4-14

4) resistance to perturbation, including asymptotic stability of the program trajectory and structural stability of the system.

In the conditions of nonlinear development of events and processes spontaneity "object – external systems" at stable structure of system object of research it is expedient to apply the entropic approach and knowledge from the field of the stability theory developed for technical and cybernetic systems. In this case, the stochasticity and uncertainty of the situation is overcome by consistent presentation and analysis of qualitative information to obtain results on the conditions for maintaining structural stability in the study system "object – environment" in accordance with changes in entropy function ΔS from analysis "state – processes". According to the results obtained with this approach, it is likely to establish spontaneous processes of equilibrium regulation in the system formation or transition to new states of equilibrium with changes that are associated with increasing entropy in the studied system (Fig. 2).

The changes that occur in the object under the influence of factors influencing its elements of the surrounding environmental systems, in accordance with the entropy approach have a similar sequence and effectiveness (Fig. 3).

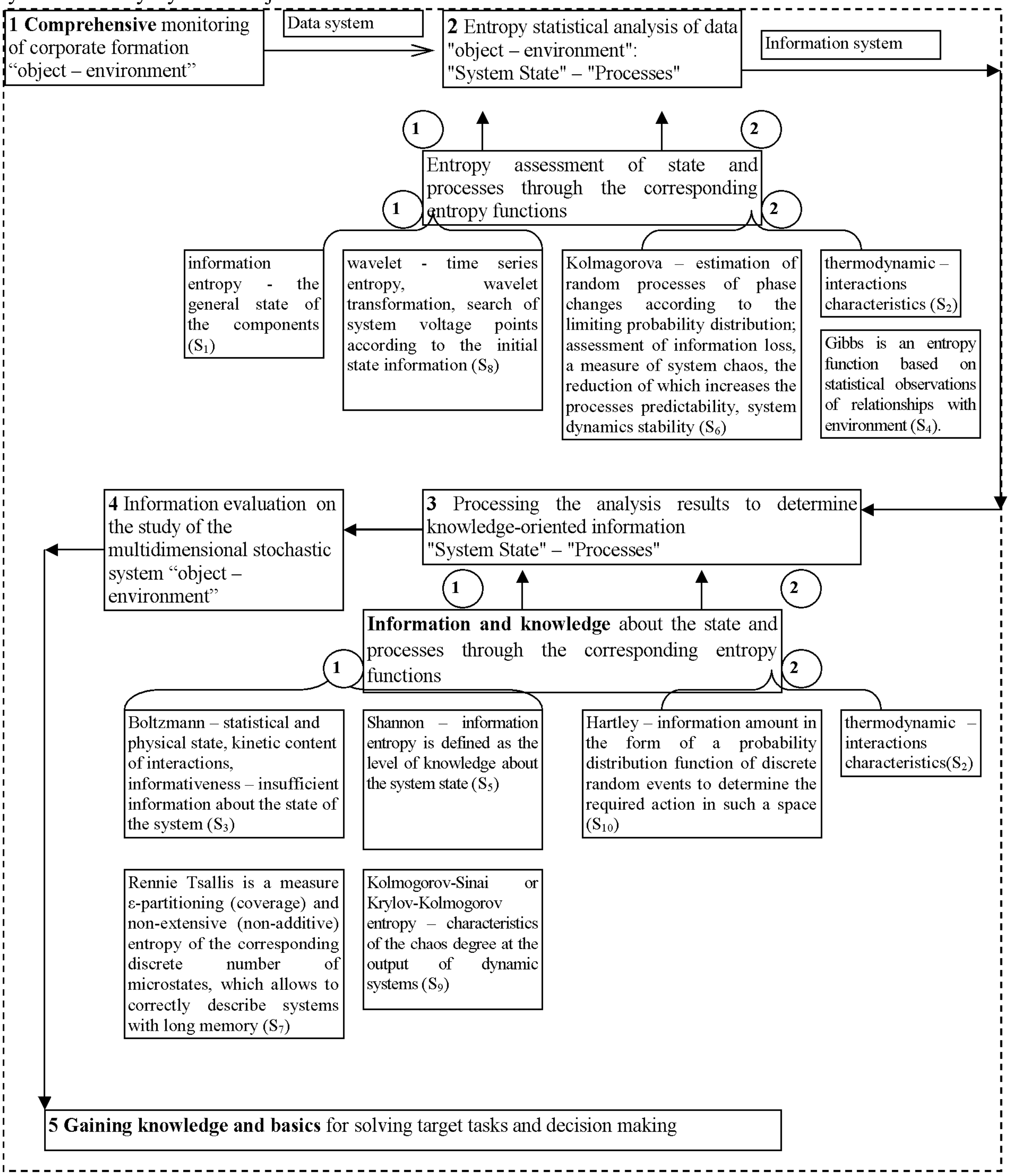
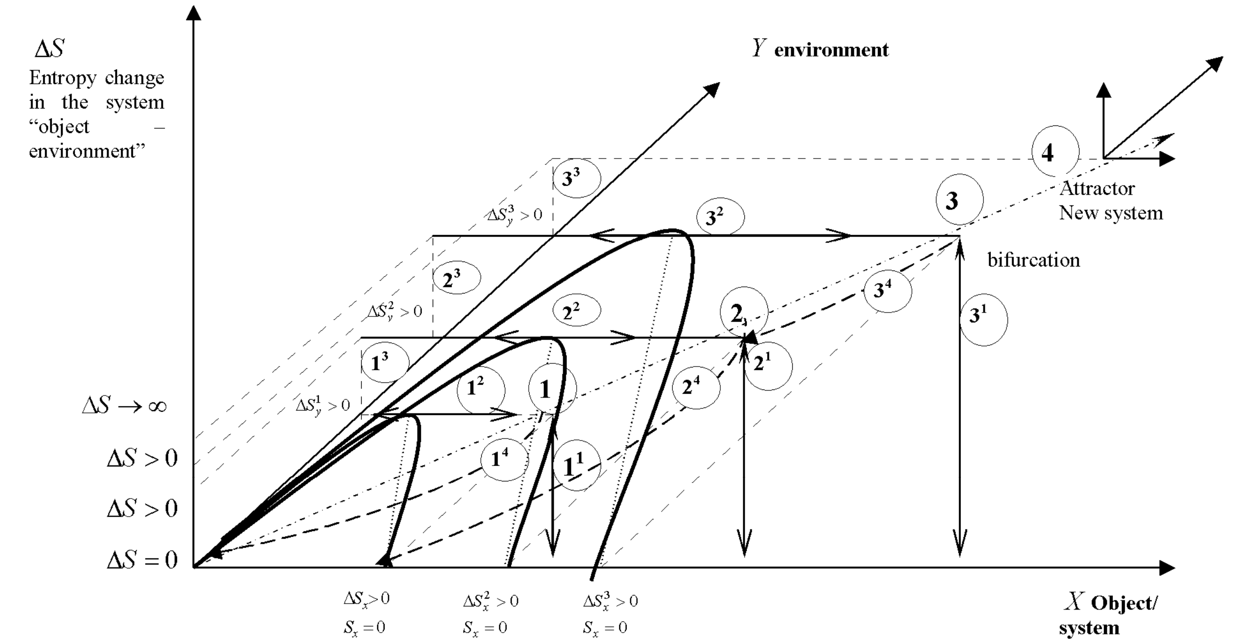


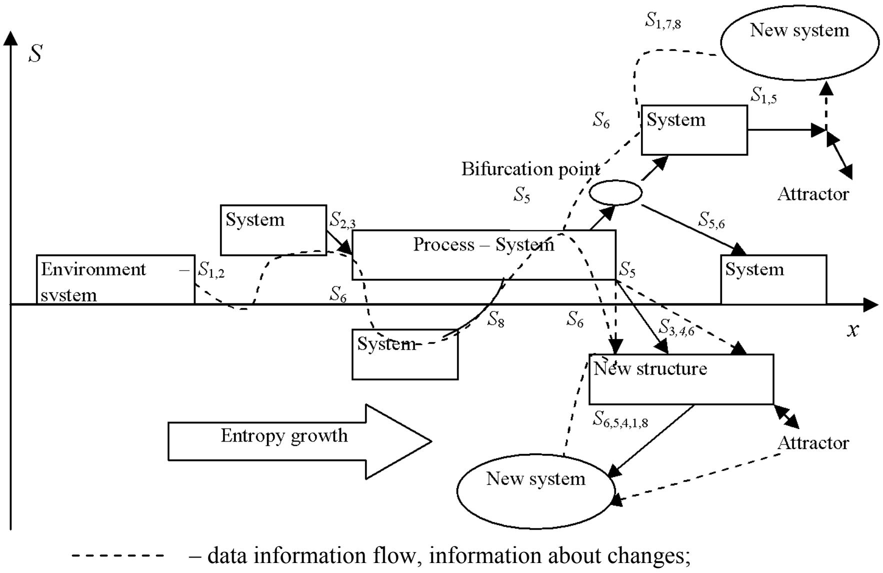
Figure 1 – Scheme of entropy approach implementation in the study of complex system formations "object – environment" (the author's proposal)

DOI 10.15588/1607-3274-2021-4-14



– Information about return to the previous (initial) state;

Figure 2 – Scheme for assessing the state of the studied system "object – environment" on the entropy approach to the analysis of "state – process" (the author's proposal): 1, 2, 3, 4 – states of the system in accordance with the conditions of interaction "object – environment"; 1^1, 2^1, 3^1 – arbitrary inverse processes of returning the system to its initial state, ie $S_x = 0$; 1^2, 2^2, 3^2 – arbitrary processes of changes from the interaction "object – environment", which leads to $\Delta S > 0$; 1^3, 2^3, 3^3 – changes in the system "object – environment" in relation to the state of external systems, leading to new states; 1^4, 2^4, 3^4 – processes of self-regulation between the object and the environment within the studied system, maintaining balance in the system with its allowable changes



– data information flow, information about changes;
– data physical flow, physical changes.

Figure 3 – Scheme of changes in the state of the system according to entropy-information estimates equilibrium search: $S_{1,2,3,4,5,6,7,8}$ – information entropy, thermodynamic, Boltzmann, Gibbs, Shannon, Kolmagorov, Rainy-Tsalles, wavelet entropy (author)

According to Figure 2, the systems state is assessed by qualitative and quantitative characteristics of entropy change, and Figure 3 shows the individual trajectory of the final structuring of the object in the space of interaction with the environment through the prism of entropy function. To obtain an unambiguous answer to the need to adjust the relationship in the studied system between the object and environmental systems, it is proposed to use not the value of the entropy function, but its changes as characteristics of entropy shift state or process from the established trajectory of their development (natural, reference, standard, boundary, etc).

DOI 10.15588/1607-3274-2021-4-14

4 EXPERIMENTS

The practical application of the provided methodological proposal for finding solutions in conditions of uncertainty of a certain kind is considered on the example of determining measures to influence the course of positive development of the child's body in the situation of cerebral palsy (CP) diagnosis in the form of information and software application at realization of the appointments of medical character applied to them (environment influence factors).

The condition of the child's body is assessed on the basis of qualitative observations of children groups of different gestational periods: healthy (19) (health groups), cerebral palsy (68), hydrocephalus (11) to study the effects of internal gastric hemorrhage (IGH of certain degree (dg), 2dg, 3dg, 4dg) on the result of "walking" (includes the category of help_walk) or not walking in combination with parallel disorders in the body of PV, PVL 1dg, 2dg, 3dg; PVI, EP [26].

According to the obtained monitoring data in the medical institution of health level of the above children groups, it is proposed to determine their condition according to the probability of walking opportunities for children of different ages and genders. The target point of the treatment process is the maximum walking of the child, which is fixed by the absence of changes in entropy function, ie during the implementation of decisions to return the natural movement of the child (see Fig. 2). In this case, the probability of deviations absence of these parameters of qualitative observations by children groups of different gestational periods from the health indicators of the walking group without restrictions is achieved, the following set of changes of possible microstates is considered:

$$\Delta\Omega = F\{\text{IGH1dg, IGH2dg, IGH3dg, IGH4dg, PVL1dg, PVL2dg, PVL3dg, PVI, EP}\}, \quad (11)$$

In order to determine the body moving trajectory affected by cerebral palsy to the most healthy state based on the results of monitoring the child's body during treatment, it is proposed to automatically establish the state according to this algorithm.

Step 1. The implementation of the macro-situation is determined by five groups of indicators of the state of the organism (IGH, PVL, PVI and episyndrome), which corresponds to certain changes in the entropy of the system (see equation (7)).

Step 2. Calculated the total number of changes in the system ΔN as the probability of changing the deviations number of healthy movement indicators in the observation groups walking, auxiliary walking – $\Delta n = \overline{1, \Delta N}, \Delta p_n = 1/\Delta N$.

Step 3. The system achievement of the desired result at this stage is determined – walking without restrictions (see Hartley entropy (8)).

Step 4. Establishment of the information characteristic concerning determinations of results of changes in in groups of supervision walking, auxiliary walking (see the equation (5)).

Step 5. The general characteristic of changes in the investigated object concerning positive development of child organism in a situation of the diagnosis of a cerebral palsy (see realization of consecutive calculations (6)) is given (Fig. 4).

The proposed entropy approach to analysis is determined by the fact that the object is characterized from the standpoint of compliance with acceptable regulatory constraints and processes regarding the acceptability of the object of the external environment or the possibility of resolving the situation of coexistence "object – environment".

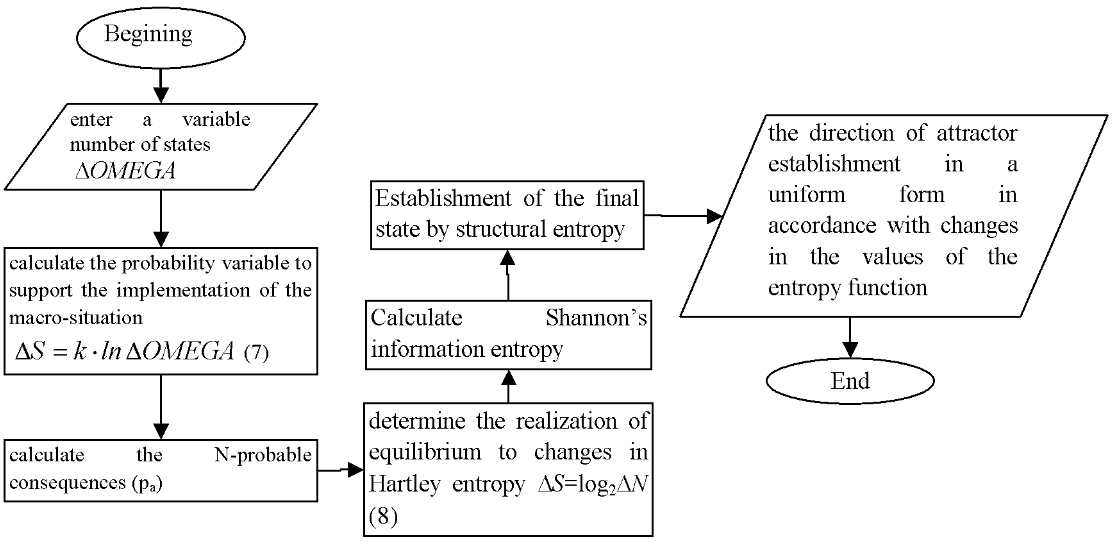


Figure 4 – Algorithm for finding a solution to achieve the system state "gait without restrictions"

DOI 10.15588/1607-3274-2021-4-14

5 RESULTS

According to the proposed entropy approach, the above dynamics of the system in the environment of other systems is proposed to identify the direction of the attractor in a uniform form in accordance with changes in entropy function in the sequence "system state – excitation – changes – process – state of stabilized system". This makes it possible to study biological systems on the basis of a complementary methodology with an idea of current instantaneous events.

During the implementation of experimental studies on the proposed algorithm for entropy assessment of states and processes C# software application using Windows-Forms technology is developed. As a result of working with the software application, a graph is obtained, which shows four groups of children: healthy, walking without restrictions, auxiliary walking and non-walking, who have the appropriate quantitative IGH, PVI, PVL and episyndrome (Fig. 5).

The Oy axis reflects the probability of one of the four parameters for a particular child state. The Ox axis reflects week on which the child was born. The left part of Figure 5 provides recommendations for increasing or decreasing the value of each of the indicators for children to achieve the state "walking without restrictions".

6 DISCUSSION

The proposed entropy approach to the choice of decision-making problems for determining the state and changes as a result of process transformations in system objects of the type "studied system – environment" in conditions of uncertainty does not require additional conditions characteristic of known estimates by criteria in common mathematical means of DM. The application of the entropy approach ensures full compliance with the methodology of the stability theory, according to which the result of data processing invariant with respect to the processing method corresponds to reality, while using known criteria the processing result depends on the processing method and reflects the subjectivity of the researcher but not objective ratio.

At the same time, it should be noted that when processing monitoring data for the transition to information results, there are difficulties in the transition to the probabilistic characteristics of the state and dynamic phenomena. Therefore, in this paper, the entropy estimate is closely related to statistical data processing (Fig. 6).

CONCLUSIONS

The following tasks have been solved:

1) Theoretically proved that for DM to achieve a certain target function that is responsible for the state and functionality of the studied object in certain surrounding conditions by environment systems, it is advisable to use not the function S itself as the implementation of the complex system macrostate through certain experiments of microstates, but its changes in the system "object – environment" in relation to the state of external systems ΔS^{123}_{y}(see Fig. 2) using for certain situations different entropy functions with their processing in relation to the

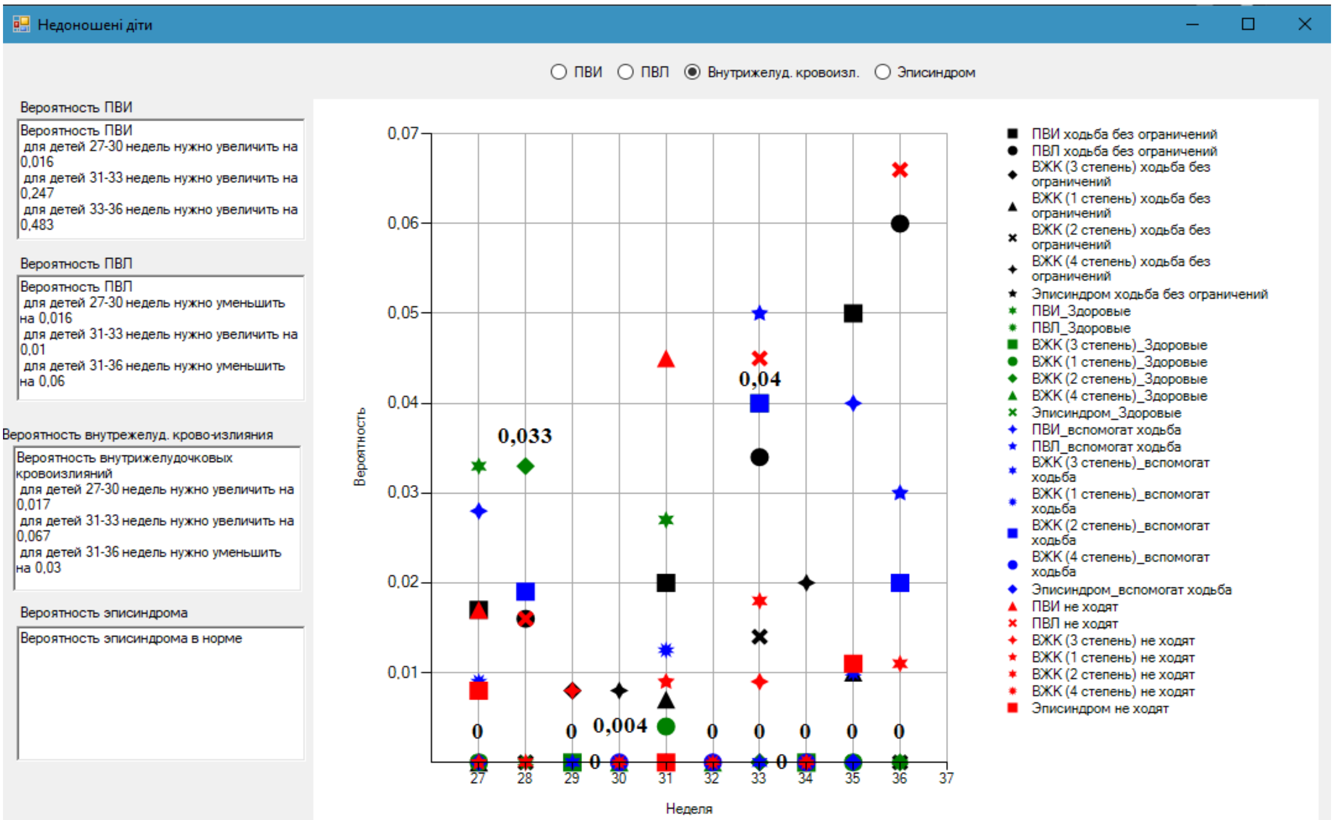


Figure 5 – Graphic interpretation of child's condition determination and decisions to improve it

DOI 10.15588/1607-3274-2021-4-14

Initial data for the analysis of the patients condition in the "walking" group

	1 NewVa	2 NewVa	3 IGH	4 IGH I	5 IGH I	6 IGH V	7 PVL	8 PVL I	9 PVL II	10 PVL III	11 EП	12 27_boy	13 31_boy	14 34_boy	15 27_girl	16 31_girl	17 34_girl
1	0,275	0,05	0	0	0,008	0,008	0,008	0,008	0	0,01		0,05	0,075	0,15	0,107	0,107	0,179
2		0,075	0,00	0,014	0	0	0	0,034	0	0,02		0	0,007	0,012	0	0	0
3		0,15	0,01	0	0	0,023	0,01	0,035	0,023	0,04		0	0,014	0	0	0,015	0
4	0,393	0,107	0	0	0,015	0	0,01	0,031	0	0,04		0,08	0	0	0,015	0	0
5		0,107	0	0,015	0	0	0	0,031	0,015	0,04		0,008	0	0,023	0	0	0
6		0,179	0	0	0	0	0,01	0,036	0,036	0,08		0,008	0	0,012	0,015	0	0,018
7												0,008	0,034	0,035	0,031	0,031	0,036
8												0	0	0,023	0	0,015	0,036
9												0,017	0,02	0,046	0,046	0,046	0,089

New Var 1 – probability of boys and girls getting into the "walking" group;

New Var 2 – probability distribution by gestational age

Factor analysis results of the "walking" state factors

Variable	Factor Loading Extraction: Pri (Marked loadin Factor 1
NewVar2	-0,991
IGH I	-0,194
IGH II	0,408
IGH III	0,372
IGH IV	-0,242
PVL I	-0,677
PVL II	-0,647
PVL III	-0,928
PVI	-0,9077
Expl.Var	3,9464
Prp.Totl	0,4385

Variable	Factor Loadings (Un Extraction: Principal (Marked loadings are Factor 1	Factor 2
NewVar2	-0,991	-0,048
IGH I	-0,194	-0,557
IGH II	0,408	-0,792
IGH III	0,372	0,827
IGH IV	-0,242	-0,086
PVL I	-0,677	0,706
PVL II	-0,647	-0,484
PVL III	-0,928	-0,150
PVI	-0,908	0,148
Expl.Var	3,946	2,409
Prp.Totl	0,438	0,268

Variable	Factor Loadings (Unrotated) (Extraction: Principal compone (Marked loadings are > .7000 Factor 1	Factor 2	Factor 3
NewVar	-0,991	-0,048	0,074
IGH I	-0,194	-0,557	-0,751
IGH II	0,408	-0,792	0,439
IGH III	0,372	0,827	0,003
IGH IV	-0,242	-0,086	-0,939
PVL I	-0,677	0,706	-0,128
PVL II	-0,647	-0,484	0,215
PVL III	-0,928	-0,150	0,083
PVI	-0,908	0,148	0,387
Expl.Var	3,946	2,409	1,863
Prp.Totl	0,438	0,268	0,207

Factor load in the group "walking"

Value	Eigenvalues (walking.sta) Extraction: Principal components Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	3,946	43,849	3,946	43,849
2	2,409	26,762	6,355	70,611
3	1,863	20,699	8,218	91,310

Figure 6 – Fragment of statistical analysis of experimental data

proposed approaches (equations (3)–(10)); this, in turn, allows for a consistent analysis of uncertainties and their resolution to establish the conditions for stabilizing the object or achieving the goal of regulating the situation with the development of certain solutions;

2) An example of solving problematic issues regarding the stabilization of the disturbed organism in certain conditions by probabilistic entropy, namely its changes for children of different ages and conditions with pathological deviations to the function of "walking", which allowed to offer a software application that allows you to automatically provide an answer to the child's condition (assignment to a certain group of health on the move) and a decision on the maximum approach of the system to the state "walking without restrictions" (see Fig. 5).

Scientific novelty – the entropy-information representation of the compound of complex systems of the form "object – environment" at the level of state and resources of its change when interacting with the environment.

Further development is the immitational modeling of real dynamic systems based on the entropy and informational presentation of their states and dynamics.

ACKNOWLEDGEMENTS

The work was performed at the Department of Software Engineering and Information Technology Management at National Technical University "Kharkiv Polytechnic Institute".

DOI 10.15588/1607-3274-2021-4-14

REFERENCES

1. Shuangshuang Li, Sokchhay Heng, Sokly Siev, Chihiro Yoshimura, Oliver Saavedra, Sarann Ly Multivariate interpolation and information entropy for optimizing raingauge network in the Mekong River Basin, *Hydrological Sciences Journal*, 2019, Vol. 64, Issue 12, pp. 1439–1452. DOI: https://doi.org/10.1080/02626667.2019.1646426
2. Hadi Mahmoudi-Meimand, Sara Nazif, Ali Abbaspour Rahim, Sabokbar Hasanali Faraji An algorithm for optimisation of a rain gauge network based on geostatistics and entropy concepts using GIS, *Journal of Spatial Science*, 2016, Vol. 61, Issue 1, pp. 233–252. DOI: https://doi.org/10.1080/14498596.2015.1030789
3. Azadeh Gholamia, Bonakdari Hossein, Mohammadianb Majid, Zajia Amir Hossein, Gharabaghi Bahram Assessment of geomorphological bank evolution of the alluvial threshold rivers based on entropy concept parameters, *Hydrological Sciences Journal/Journal des Sciences Hydrologiques*, 2019, Vol. 64, Issue 7, pp. 856–872. DOI: https://doi.org/10.1080/02626667.2019.1608995
4. Hao L., Su X., Singh V. P., Ayantobo O. O. Spatial Optimization of Agricultural Land Use Based on Cross-Entropy Method, *Entropy,* 2017, № 19, pp. 592–607. DOI: https://doi.org/10.3390/e19110592
5. Kozulia T., Kozulia M. Integrated information system assessment of complex objects environmental safety level, *Visnyk NTU «HPI». Serija Systemnyj analiz, upravlinnja ta informacijni tehnologii'*. Harkiv, NTU «HPI», 2017, № 55 (1276), pp. 39–44. DOI: https://doi.org/10.20998/2079-0023.2017.55.07
6. Kozulia T. V., Bilova M. A., Kozulia M. M., Fonta N. G. Informacionnoe obespechenie analiza bezopasnosti prirodno-tehnogennyh ob#ektov v kontekste ih vzaimodejstvija s okruzhajushhej sredoj, *International scientific journal*, Tbilisi, 2017, №3 (53), pp.72–80
7. Kozulia T. V., Kozulia M. M. Using graph-analytical methods modeling of system objects to determine integrated assessment of their state, *Problems of Atomic Science and Technology*, 2019, No. 3(121), pp. 116–123.
8. Kozulia T., Kozulia M., Didmanidze I. Comprehensive study of the systemic formation «object – environment» safety state, *Technogenic and Ecological Safety*, 2020, No. 7(1/2020), pp. 3–12.
9. Tyrsin A. N. Jentropijnoe modelirovanie mnogomernyh stohasticheskih sistem. Monografija. Voronezh, «Nauchnaja kniga», 2016, 156 p.
10. Aboutaleb Hycham, Monsuez Bruno Entropy in Design Phase: a Higraph-Based Model Approach, *2017 IEEE International Conference on Information Reuse and Integration (IRI), 4–6 Aug. 2017: proceedings. San Diego, IEEE,* 2017, pp. 526–534. DOI: https://doi.org/10.1109/iri.2017.51
11. Hosseini M., Emamjomeh H. Entropy-based Serviceability Assessment of Water Distribution Networks, Subjected to Natural and Man-made Hazards, *International Journal of Engineering*, 2014, Vol. 27, No. 5, pp. 675–688. DOI: https://doi.org/10.5829/idosi.ije.2014.27.05b.02
12. Keum Jongho, Frezer Seid Awol, Ursulak Jacob, Coulibaly Paulin Introducing the Ensemble-Based Dual Entropy and Multiobjective Optimization for Hydrometric Network Design Problems: EnDEMO, *Entropy,* 2019, No. 21, P. 947 DOI: https://doi.org/10.3390/e21100947
13. Xiong Feng, Guo Shenglian, Chen Lu, Jiabo Yin, Pan Liu Flood Frequency Analysis Using Halphen Distribution and Maximum Entropy, *Journal of Hydrologic Engineering*, 2018, 23(5), 04018012. DOI: https://doi.org/10.1061/(asce)he.1943-5584.0001637
14. Zhou Zhenghong, Juanli Ju, Xiaoling Su, Vijay P. Singh, Gengxi Zhang Comparison of Two Entropy Spectral Analysis Methods for Streamflow Forecasting in Northwest China, *Entropy,* 2017, №19 (11), P. 597. DOI: https://doi.org/10.3390/e19110597
15. Cui Huijaun, Vijay P. Singh Entropy Spectral Analyses for Groundwater Forecasting, *Journal of Hydrologic Engineering*, 2017, №22(7), 06017002. DOI: https://doi.org/10.1061/(asce)he.1943-5584.0001512
16. Abásolo D., Hornero R, Espino P, Álvarez D, Poza J. Entropy analysis of the EEG background activity in Alzheimer's disease patients, *Physiological Measurement* 2006, №27(3), pp. 241–53. DOI:10.1088/0967-3334/27/3/003
17. Fabris C., Sparacino G, Sejling AS, Goljahani A, Duun-Henriksen J, Remvig LS, Juhl CB, Cobelli C. Hypoglycemia-related electroencephalogram changes assessed by multiscale entropy, *Diabetes Technology & Therapeutics,* 2014, №16(10), P. 688–94. DOI: https://doi.org/10.1089/dia.2013.0331
18. Lazar P., Jayapathy R, Torrents-Barrena J, Mol B, Mohanalin, Puig D. Fuzzy-entropy threshold based on a complex wavelet denoising technique to diagnose Alzheimer disease, *Healthcare Technology Letters* 2016, №3(3), pp. 230–238. DOI: https://doi.org/10.1049/htl.2016.0022
19. Baracco G J. Battling Entropy in Infection Control Systems, *Infection Control & Hospital Epidemiology,* 2017, №38(12), pp. 1428–1429. DOI: https://doi.org/10.1017/ice.2017.225
20. Kornev A. P. Self-organization, entropy and allostery, *Biochemical Society Transactions*, 2018, №46(3), pp. 587–597. DOI: https://doi.org/10.1042/bst20160144
21. Al-Qazzaz NK. Sabir MK, Ali SHBM, Ahmad SA, Grammer K. Electroencephalogram Profiles for Emotion Identification over the Brain Regions Using Spectral, Entropy and Temporal Biomarkers, *Sensors (Basel).* 2019, №20(1), P. 59. DOI: https://doi.org/10.3390/s20010059
22. Li Peng Entropy: a software application for the entropy analysis of physiological time-series, *BioMedical Engineering OnLine,* 2019, №18(1), P. 30. DOI: https://doi.org/10.1186/s12938-019-0650-5
23. Lennikov R. V. Informacionno-jentropijnye modeli innovacionnyh processov v jekonomicheskih sistemah, *Izvestija Tul'skogo gosudarstvennogo universiteta Estestvennye nauki. Informatika,* 2011, Vyp. 1, pp. 167–176.
24. Kozulia T. V., Kozulia M. M. Informacijno-programne zabezpechennja obrobky ta analizu stanu skladnyh ekologichnyh ob'jektiv, *Informacijna bezpeka ta informacijni tehnologii': monografija*. Harkiv, TOV «DISA PLJuS», 2019, pp. 202–231.
25. Kolmogorov A. N. Teorija informacii i teorija algoritmov. Moscow, Nauka, 1987, 304 p.
26. Litovchenko T. A., Vareshnjuk E. V., Sharonova N. V., Kozulia T. V. Matematicheskij analiz i prognozirovanie razvitija bol'shih motornyh funkcij u detej, rozhdennyh nedonoshennymi, *Nauchnye vedomosti. Serija Medicina. Farmacija.* 2014, № 11 (182). Vyp. 2, pp. 109–112.

Received 30.03.2021.
Accepted 04.10.2021.

УДК 519.2

ЕНТРОПІЙНИЙ ПІДХІД У СИСТЕМНИХ ДОСЛІДЖЕННЯХ ОБ'ЄКТІВ РІЗНОЇ СКЛАДНОСТІ З ОЦІНКИ ЇХ СТАНУ ТА ФУНКЦІОНАЛЬНОСТІ

Козуля Т. В. – професор, доктор технічних наук, професор кафедри Програмної інженерії та інформаційних технологій управління, Національний Технічний Університет «Харківський Політехнічний Інститут».

DOI 10.15588/1607-3274-2021-4-14

Свірідова А. С. – магістр кафедри Програмної інженерії та інформаційних технологій управління, Національний Технічний Університет «Харківський Політехнічний Інститут».

Козуля М. М. – канд. техн. наук, доцент кафедри Програмної інженерії та інформаційних технологій управління, Національний Технічний Університет «Харківський Політехнічний Інститут».

АНОТАЦІЯ

Актуальність. Розгляд комплексно дослідженого об'єкта у вигляді «система – навколишнє середовище» для отримання наближено точного відображення реальної ситуації.

Мета. Пошук рішень проблемних питань наукових досліджень на базі ентропійного підходу для систем «об'єкт – навколишнє середовище» різної природи і складності, вивчення їх і отримання знань (стійкої інформації) та надання їх у вигляді комплексу завдань складної системи, модульованих завдяки різним ентропійним функціям.

Метод. Для оцінки стійкості розвитку системного об'єкта використовують такі критерії: цілісність – невихід траєкторії розвитку об'єкта на певному прогнозному інтервалі часу з множини безпечних станів; монотонність зростання індикаторів розвитку об'єкта на певному інтервалі часу з наступним збереженням їх у заданих інтервалах припустимих значень; відповідність траєкторії розвитку цільовим змінам за вимогами безпеки і сталого розвитку;стійкість до збурювання, у тому числі, асимптотична стійкість програмної траєкторії й структурна стійкість системи.

В умовах нелінійного розвитку подій і самодовільності процесів «об'єкт – зовнішні системи» при стійкій структурі системного об'єкта дослідження доцільним є застосування ентропійного підходу та знань із області теорії стійкості, розробленої для технічних і кібернетичних систем.

Результати. Запропонований ентропійний підхід щодо аналізу визначається тим, що об'єкт характеризується з позицій дотримання допустимих нормативних обмежень і процеси стосовно прийнятності об'єктом впливів зовнішнього оточення чи можливості урегулювання ситуації співіснування «об'єкт – навколишнім середовищем».

У межах аналізу системного об'єкта це означає, що і для стаціонарних, і для динамічних умов їх стан описується певною функцією, зміни якої вказують на наближення до певної точки гомеостатичних відносин з навколишнім середовищем.

Практичне застосування наданої методологічної пропозиції з пошуку рішень в умовах невизначеності певного роду розглянуто на прикладі визначення заходів впливу на хід позитивного розвитку організму дитини в ситуації діагнозу дитячий церебральний параліч (ДЦП) у вигляді інформаційно-програмного додатку щодо імовірності віднесення дитини до групи ходьба або не ходьба при реалізації застосованих до них призначень лікувального характеру (фактори впливу НС).

Висновки. Запропонований ентропійний підхід до вибору розв'язку задач прийняття рішень щодо визначень стану і змін в наслідок процесних перетворень в системних об'єктах виду «досліджена система – навколишнє середовище» в умовах певної невизначеності не вимагає виконання додаткових умов, характерних для відомих оцінок за критеріями в поширених математичних засобах ПР.

КЛЮЧОВІ СЛОВА: ентропійний підхід, ентропійно-інформаційні оцінки, програмне забезпечення, складні об'єкти, технологія WindowsForms.

УДК 519.2

ЭНТРОПИЙНЫЙ ПОДХОД В СИСТЕМНОМ ИССЛЕДОВАНИИ ОБЪЕКТОВ РАЗЛИЧНОЙ СЛОЖНОСТИ ПО ОЦЕНКЕ ИХ СОСТОЯНИЯ И ПРОИЗВОДИТЕЛЬНОСТИ

Козуля Т. В. – профессор, доктор технических наук, профессор кафедры Программной инженерии и информационных технологий управления, Национальный Технический Университет «Харьковский Политехнический Институт».

Свиридова А. С. – магистр кафедры Программной инженерии и информационных технологий управления, Национальный Технический Университет «Харьковский Политехнический Институт».

Козуля М. М. – канд. техн. наук, доцент кафедры Программной инженерии и информационных технологий управления, Национальный Технический Университет «Харьковский Политехнический Институт».

АННОТАЦИЯ

Актуальность. Рассмотрение комплексно исследованного объекта в виде «система – окружающая среда» для получения приближенно точного отражения реальной ситуации.

Цель. поиск решений проблемных вопросов научных исследований на базе энтропийного подхода для систем «объект – окружающая среда» различной природы и сложности, изучение их и получения знаний (устойчивой информации) и предоставления их в виде комплекса задач сложной системы, модулированных благодаря различным энтропийным функциям.

Метод. Для оценки устойчивости развития системного объекта используют следующие критерии: целостность – невыход траектории развития объекта на определенном прогнозном интервале времени из множества безопасных состояний; монотонность роста индикаторов развития объекта на определенном интервале времени с последующим сохранением их в заданных интервалах допустимых значений; соответствие траектории развития целевым изменениям по требованиям безопасности и устойчивого развития; устойчивость к возмущению, в том числе, асимптотическая устойчивость программной траектории и структурная устойчивость системы.

В условиях нелинейного развития событий и самодовильности процессов «объект – внешние системы» при устойчивой структуре системного объекта исследования целесообразно применение энтропийного подхода и знаний из области теории устойчивости, разработанной для технических и кибернетических систем.

Результаты. Предложенный энтропийный подход к анализу определяется тем, что объект характеризуется с позиций соблюдения допустимых нормативных ограничений и процессы относительно приемлемости объектом воздействий внешнего окружения или возможности урегулирования ситуации сосуществования «объект – окружающая среда».

В рамках анализа системного объекта это значит, что и для стационарных, и для динамических условий их состояние описывается определенной функцией, изменения которой указывают на приближение к определенной точке гомеостатических отношений с окружающей средой.

Практическое применение предоставленной методологической предложения по поиску решений в условиях неопределенности своего рода рассмотрен на примере определения мер воздействия на ход позитивного развития организма ребенка в ситуации диагноза детский церебральный паралич (ДЦП) в виде информационно-программного приложения по вероятности отнесения ребенка к группе ходьба либо не ходьба при реализации примененных к ним назначений лечебного характера (факторы влияния НС).

DOI 10.15588/1607-3274-2021-4-14

Выводы. Предложенный энтропийный подход к выбору решении задач принятия решений относительно определений состояния и изменений вследствие процессных преобразований в системных объектах вида «исследована система – окружающая среда» в условиях некоторой неопределенности не требует выполнения дополнительных условий, характерных для известных оценок по критериям в распространенных математических средствах ПР.

КЛЮЧЕВЫЕ СЛОВА: энтропийный подход, энтропийно-информационные оценки, программное приложение, сложные объекты, технология WindowsForms.

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. Shuangshuang Li Multivariate interpolation and information entropy for optimizing raingauge network in the Mekong River Basin / [Li Shuangshuang, Heng Sokchhay, Siev Sokly et al.] // Hydrological Sciences Journal . – 2019 – Vol. 64, Issue 12. – P. 1439–1452. DOI: https://doi.org/10.1080/02626667.2019.1646426
2. An algorithm for optimisation of a rain gauge network based on geostatistics and entropy concepts using GIS / [Mahmoudi-Meimand Hadi, Nazif Sara, Rahim Ali Abbaspour, Hasanali Faraji Sabokbar] // Journal of Spatial Science. – 2016. – Vol. 61, Issue 1. – P. 233–252. DOI: https://doi.org/10.1080/14498596.2015.1030789
3. Azadeh Gholamia Assessment of geomorphological bank evolution of the alluvial threshold rivers based on entropy concept parameters. / [Azadeh Gholamia, Hossein Bonakdari, Majid Mohammadianb et al.] // Hydrological Sciences Journal/Journal des Sciences Hydrologiques. – 2019. – Vol. 64, Issue 7. – P. 856–872. DOI: https://doi.org/10.1080/02626667.2019.1608995
4. Spatial Optimization of Agricultural Land Use Based on Cross-Entropy Method. / [L. Hao, X. Su, V. P. Singh, O. O. Ayantobo] // Entropy. – 2017. – № 19. – P. 592–607. DOI: https://doi.org/10.3390/e19110592
5. Kozulia T. Integrated information system assessment of complex objects environmental safety level / T. Kozulia, M. Kozulia // Visnyk NTU «HPI». Serija Systemnyj analiz, upravlinnja ta informacijni tehnologii’. Harkiv: NTU «HPI», 2017. – № 55 (1276). – P. 39–44. DOI: https://doi.org/10.20998/2079-0023.2017.55.07
6. Kozulia T. V. Informacionnoe obespechenie analiza bezopasnosti prirodno-tehnogennyh ob#ektov v kontekste ih vzaimodejstvija s okruzhajushhej sredoj / [T. V. Kozulia, M. A. Bilova, M. M. Kozulia, N. G. Fonta] // International scientific journal, Tbilisi, – 2017, – №3 (53). – P.72–80.
7. Kozulia T.V. Using graph-analytical methods modeling of system objects to determine integrated assessment of their state. / T. V. Kozulia, M. M. Kozulia // Problems of Atomic Science and Technology. – 2019. – № 3 (121). – P. 116–123.
8. Kozulia T. Comprehensive study of the systemic formation «object–environment» safety state / T. Kozulia, M. Kozulia, I. Didmanidze // Technogenic and Ecological Safety. – 2020. – №7(1/2020). – P. 3–12.
9. Tyrsin A. N. Jentropijnoe modelirovanie mnogomernyh stohasticheskih sistem. Monografija. / A. N. Tyrsin. – Voronezh : «Nauchnaja kniga», 2016. – 156 p.
10. Aboutaleb Hycham. Entropy in Design Phase: a Higraph-Based Model Approach. / Hycham Aboutaleb, Bruno Monsuez // 2017 IEEE International Conference on Information Reuse and Integration (IRI). – P. 526–534. DOI: https://doi.org/10.1109/iri.2017.51
11. Hosseini M. Entropy-based Serviceability Assessment of Water Distribution Networks, Subjected to Natural and Man-made Hazards. / M. Hosseini, H. Emamjomeh // International Journal o f Engineering, (May 2014) – Vol. 27, No. 5. – P. 675–688. DOI: https://doi.org/10.5829/idosi.ije.2014.27.05b.02
12. Keum Jongho Introducing the Ensemble-Based Dual Entropy and Multiobjective Optimization for Hydrometric Network Design Problems: EnDEMO / [Jongho Keum, Seid Awol Frezer, Jacob Ursulak, Paulin Coulibaly] // Entropy. – 2019. – № 21. – 947 p. DOI: https://doi.org/10.3390/e21100947
13. Xiong Feng. Flood Frequency Analysis Using Halphen Distribution and Maximum Entropy / [Feng Xiong, Shenglian Guo, Lu Chen et al.] // Journal of Hydrologic Engineering. – 2018. – 23(5). – 04018012. DOI: https://doi.org/10.1061/(asce)he.1943-5584.0001637
14. Zhou Zhenghong. Comparison of Two Entropy Spectral Analysis Methods for Streamflow Forecasting in Northwest China / [Zhou Zhenghong, Ju Juanli, Su Xiaoling et al.] // Entropy. – 2017. – №19 (11). – P. 597. DOI: https://doi.org/10.3390/e19110597
15. Cui Huijaun Entropy Spectral Analyses for Groundwater Forecasting / Cui Huijaun, P. Singh Vijay // Journal of Hydrologic Engineering. – 2017. – №22(7). – 06017002. DOI: https://doi.org/10.1061/(asce)he.1943-5584.0001512
16. Entropy analysis of the EEG background activity in Alzheimer’s disease patients / [D. Abásolo, R. Hornero, P. Espino et al.] // Physiological Measurement. – 2006 – № 27(3). – P. 241–53. DOI:10.1088/0967-3334/27/3/003
17. Hypoglycemia-related electroencephalogram changes assessed by multiscale entropy. / C Fabris, G Sparacino, AS Sejling et al.] // Diabetes Technology & Therapeutics. – 2014. – №16(10). – P. 688–94. DOI: https://doi.org/10.1089/dia.2013.0331
18. Lazar P. Fuzzy-entropy threshold based on a complex wavelet denoising technique to diagnose Alzheimer disease. / [P. Lazar, R. Jayapathy, J. Torrents-Barrena et al.] // Healthcare Technology Letters 2016, – №3(3). – P. 230–238. DOI: https://doi.org/10.1049/htl.2016.0022
19. Baracco GJ. Battling Entropy in Infection Control Systems / GJ. Baracco // Infection Control & Hospital Epidemiology. – 2017 Dec. – № 38(12). – P. 1428–1429. DOI: https://doi.org/10.1017/ice.2017.225
20. Kornev A. P. Self-organization, entropy and allostery. / A. P. Kornev // Biochemical Society Transactions. – 2018. – Jun 19. – № 46(3) – P. 587–597. DOI: https://doi.org/10.1042/bst20160144
21. Al-Qazzaz NK. Electroencephalogram Profiles for Emotion Identification over the Brain Regions Using Spectral, Entropy and Temporal Biomarkers / [NK Al-Qazzaz, MK Sabir, SHBM Ali et al.] // Sensors (Basel). – 2019. – Dec. 20. – № 20 (1). – P. 59. DOI: https://doi.org/10.3390/s20010059
22. Li Peng Entropy: a software application for the entropy analysis of physiological time-series. / Li Peng // BioMedical Engineering OnLine. – 2019. – Mar 20. – № 18 (1). – P. 30. DOI: https://doi.org/10.1186/s12938-019-0650-5
23. Lennikov R. V. Informacionno-jentropijnye modeli innovacionnyh processov v jekonomicheskih sistemah / R. V. Lennikov // Izvestija Tul’skogo gosudarstvennogo universiteta Estestvennye nauki. Informatika. – 2011. – Vyp. 1. – P. 167–176.
24. Kozulia T. V. Informacijno-programne zabezpechennja obrobky ta analizu stanu skladnyh ekologichnyh ob’jektiv / T. V. Kozulia, M. M. Kozulia // Informacijna bezpeka ta informacijni tehnologii’: monografija. – Harkiv : TOV «DISA PLJuS», 2019. – P. 202–231.
25. Kolmogorov A. N. Teorija informacii i teorija algoritmov / A. N. Kolmogorov. – M. : Nauka, 1987. – 304 p.
26. Matematicheskij analiz i prognozirovanie razvitija bol’shih motornyh funkcij u detej, rozhdennyh nedonoshennymi. / [T. A. Litovchenko, E. V. Vareshnjuk, N. V. Sharonova, T.V. Kozulia] // Nauchnye vedomosti. Serija Medicina. Farmacija. – 2014. – № 11 (182), Vyp. 2. – P. 109–112.

DOI 10.15588/1607-3274-2021-4-14

Наукове видання

Радіоелектроніка,

інформатика,

управління

№ 4/2021

Науковий журнал

Оригінал-макет підготовлено у редакційно-видавничому відділі НУ «Запорізька політехніка»

Свідоцтво про державну реєстрацію
КВ № 24220-14060 ПР від 19.11.2019.

Підписано до друку 07.12.2021. Формат 60×84/8.
Папір офс. Різогр. друк. Ум. друк. арк. 19,07.
Тираж 300 прим. Зам. № 962.

69063, м. Запоріжжя, НУ «Запорізька політехніка», друкарня, вул. Жуковського, 64

Свідоцтво суб'єкта видавничої справи
ДК № 6952 від 22.10.2019.