Information systems modeling

UDC 004.04:004.67:004.77 **doi: 10.20998/2522-9052.2021.1.02**

Mykhailo Mozhaiev[1], Pavlo Buslov[2]

Hon. Prof. M.S. Bokarius Kharkiv Research Institute of Forensic Examinations, Kharkiv, Ukraine
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

METHOD OF MODELING OF A SOCIAL PROFILE USING BIG DATA STRUCTURE TRANSFORMATION OPTIMIZATION

Abstract. The object of the research are methods and algorithms of optimizing of the Big Data transformation to build a social profile model, the subject of the research are methods of constructing of a social profile. For decision-making person, the problem of scientific methodological and instrumental re-equipment is relevant for the effective fulfillment of a set of managerial tasks and confronting of fundamentally new challenges and threats in society. This task is directly related to the problem of building of a model of the social profile of both the individual and the social group as a whole. Therefore, the problem of optimizing of methods of constructing of a mathematical model of a social profile is certainly relevant. During the research, methods of the mathematical apparatus of graph theory, database theory and the concept of non-relational data stores, Big Data technology, text analytics technologies, parallel data processing methods, methods of neural networks' using, methods of multimedia data analyzing were used. These methods were integrated into the general method, called the method of increasing of the efficiency of constructing of a mathematical model of a social profile. The proposed method improves the adequacy of the social profile model, which will significantly improve and simplify the functioning of information systems for decision-making based on knowledge of the social advantages of certain social groups, which will allow dynamic correction of their behavior. The obtained results of testing the method make it possible to consider it as an effective tool for obtaining of an objective information model of a social portrait of a social group. This is because the correctness of setting and solving of the problem ensured that adequate results were obtained. Unlike the existing ones, the proposed modeling method, which uses an oriented graph, allows to improve significantly the quality and adequacy of this process. Further research should be directed towards the implementation of proposed theoretical developments in real decision-making systems. This will increase the weight of automated decision-making systems for social climate analysis.

Keywords: social profile model; information decision-making systems; graph theory; Big Data technologies; social portrait of a social group.

Introduction

In the context of development of digital economy and informatization of most social processes, the tasks of managing of social and economic systems acquire a new specificity, both in terms of goals and in terms of methods and tools used. For the decision-making person, the problem of scientific, methodological and instrumental re-equipment is relevant for the effective fulfillment of a set of managerial tasks and confronting fundamentally new challenges and threats in society. Mathematical, informational and technical support is needed to enable collection, storage, monitoring, and analysis of large amounts of heterogeneous data generated by all available Internet resources and services (information and news portals, personal blogs, thematic forums, etc.) with the possibility of social profiling (individuals, various groups, social phenomena, events, processes), establishing explicit and hidden relationships between personal and group profiles, as well as identifying causal dependencies in the information field of the society.

Accordingly, new opportunities open for solving very complex and ambiguous problems related to assessment of social tension in the context of individual focus groups, forecasting the consequences of introduction of certain innovations in the life of society, identification and verification of the digital personality (to prevent compromising people by attackers using unauthorized access to personal accounts, counterfeiting accounts using technical means and artificial intelligence methods), developing service-oriented intelligent Internet technologies, and so on.

A special promise is integration of the latest information technologies related to collection and processing of large unstructured data (Big Data), intelligent data analysis (Data Mining) and distributed databases.

1. Research object and its technological audit

Big Data [1], a modern IT industry, includes a range of methods, tools, and technologies for collecting and analyzing of large amounts of different structured and unstructured data in order to achieve efficient results in the context of information distribution across multiple nodes of a computing network and its continuous updating. It is also permissible to apply the said period to the data itself, which are processed using this technology. Initially, the term Big Data was used in the academic environment to solve problems related to growth and diversity of data.

The first Big Data-based solutions appeared in the second half of the zero years and were considered as alternative ones to the classic relational DBMS in Business Intelligence [2].

The technology in question can be applied in cases where there is too much data to be processed by conventional means, in particular a server with a relational DBMS. However, the exact quantitative values of information at which data becomes "great" are not determined. The application of the technology depends on the computing hardware power and the number of records in the database, that is, in some cases, gigabytes of data can already be considered as Big Data, since the computer system does not cope with their timely processing, and in others, petabytes of information are acceptable processed using classical methods and, therefore, are not "large" ones. Big Data is often associated with the Hadoop framework [3, 4], although this technology is also based on the concept of NoSQL data stores [5, 6].

The main sources of big data are sensors, measuring devices, cellular network equipment, audio-video recording devices, message streams inside social networks, blogs and news feeds, satellites. According to Cisco [7], mobile traffic in 2014 amounted to 2.5 Exabyte's per month, and it increased by more than 9.5 times in 2019. The development and widespread dissemination of electronic sources of information has led to introduction of BD technologies in various areas of activity. This is due, among other things, to an increase in accumulation of information in the Internet, which has become especially noticeable with an increase in the popularity of social networks. In the early 2010s, many developers of the IT product market began to use the technology in question directly in the context of big data processing and to produce their own analytical solutions.

Despite the seeming advantage over classic tools, social media analytics of the Internet has a number of problems. The main disadvantages of existing methods of social analytics are the following [8-12]:

- Low prevalence of technologies to handle large amounts of data, resulting in inefficient use of computing resources and additional economic costs;
- Focus of existing solutions on a narrow range of tasks related largely to statistical social research and marketing;
- Difficulties in dealing with information structures that are highly unstructured and heterogeneous;
- Presence in social data of such properties as inconsistency, insufficiency, linguistic inaccuracy, emotional coloring, which interfere with their objective assessment and interpretation;
- Question of legality of use of obtained information by third persons;
- Lack of integration of various information technologies for analysis of text and multimedia data within a single system;
- Complexity of information storage and protection.

To eliminate the above drawbacks, it is necessary to create new and refine existing methods for processing personalized social data, compile algorithms for analyzing heterogeneous text and multimedia information, storage and presentation models of initial, intermediate analyzed and resulting data.

As a tool to increase the efficiency of methods for processing personalized social data, graph theory and an algorithm for enumerating of sections were chosen.

2. The purposes and tasks of the research

Within the framework of this research, the need to study and use a number of technological and theoretical tools is revealed: the mathematical apparatus of graph theory, database theory and the concept of non-relational data stores, Big Data technology, text analytics technologies, parallel data processing methods, and methods of neural networks' using, methods for analyzing multimedia data.

The purpose of this article is to improve the methods of automated transformation of unstructured social data from public sources into a structured picture of a person or group of persons by developing algorithms for enumerating of sections, which contain methods of formation of new sections, method of formation of new sections from already built ones, a procedure for eliminating excess elements and a search order.

To achieve this purpose, the research needs to solve the following private tasks:

1. To investigate methods and tools used in automated transformation of unstructured social data.
2. To develop algorithms of sections' enumeration that improve Big Data conversion efficiency.

3. Research of existing solutions of the problem

Social profiling results based on analysis of public sources of information should have a number of properties, such as explicit structuring, scalability, presentation both in the form of tables and images (graphs, diagrams) [13-18]. This is necessary to enable them to be used in social research applications. Currently, the most relevant areas of application of social profiles are marketing and law enforcement [9-12]. The first group includes collaborative filtering tasks, the solution of which allows improving the quality of advertising campaigns by taking into account the interests of the target audience and predicting its reaction. As a result, it allows us to optimize the costs of marketing research (surveys' conducting, target groups' forming, etc.) and to promote goods and services purposefully, as well as to increase customer loyalty. In the field of law enforcement, the task of building of social profiles intersects with open source intelligence (OSINT). Law enforcement agencies use personal and group profiles to assess the level of social tension and possible reactions of society to provocative actions, to create the "risk groups," and to collect the files for investigative and judicial measures.

4. Research methods

Within the framework of this research, the need to study and use of the following number of technological and theoretical tools is revealed: the mathematical apparatus of graph theory, database theory and the concept of non-relational data stores, Big Data technology, text analytics technologies, parallel data processing methods, methods of neural networks' using, methods of multimedia data analyzing. Modern developments in the developed countries of the world in the field of information, processing and transmission have led to study of oriented information networks, which have wide practical applications.

These networks can often be mathematically described as an oriented graph. We call (s, t) a network when it is necessary to distinguish it from a network with another source and drain. Let us denote it by G(X, G). In this case

$$X = s \cup t \cup \{x_i\}_{i=1}^{N} -$$

is a set of vertices of the network, G – the reflection of X in X, which generates the arcs of the network.

The vertex of the orgraph will be called as terminal in one of two cases:

– No arc enters it;

– Not a single arc comes out of it. In the sense of this definition, the network vertices s and t are terminal.

The arcs and vertices of network G are called network elements; we denote their set as E.

For network G, we introduce the ratio of the continuity of elements in this way, if i is the vertex of the network, than σ(i) is the set of arcs & from it, and if i is an arc, than σ(i) is the vertex into which this arc enters.

Let it be $i \in \sigma^{-1}(j)$, if $j \in \sigma(i)$. In other words, $\sigma^{-1}(j)$ it is a plurality of elements, for which element j is a successor. Let us note that the continuity ratio of σ elements can be entered on any oriented graph without loops. A network G with a plurality of elements E, for which the ratio σ is given, is denoted through G(E, σ).

Let us consider that the network elements are set by numbers.

We call the correct numbering of elements (by analogy with the correct numbering of vertices) such one that if $j \in \sigma(i)$, than $i < j$, where i, j are the numbers of the network elements.

Let's note that the elements of the finite orgraph G G(E, σ) without loops can always be correctly numbered. Indeed, it is possible to construct a one-to-one mapping (of graph G(E, σ) to graph F(X, G)), in which the set of vertices X corresponds to the set E of elements of graph G, and the relation of the contiguity of vertices G is the ratio of the continuity of elements σ on graph G. It is known that on graphs of the form (X, G), for vertices it is always possible to set the correct numeration [13-20]. Therefore, it can be set on the set E of elements of the graph G(E, σ), which follows from the mutual uniqueness of the mapping ψ. Let us call with a mixed (s, t) section or a mixed terminal section such a set of P-elements of the network that if they are removed from it, then there will be no terminal path.

The element l is called as redundant in section P if the set of elements {P\\l} is also a section [21].

A section that does not contain redundant features is minimal, and a section that contains redundant features is redundant. In an excess section view, there is always a feature that can be deleted without violating a specific section property. Let us note that from one excess section we can sometimes form several minimal sections, except various excess vertices.

As a source section, we take a set of elements

$$P_0 = \sigma(s).$$

Clearly, there can be no redundant elements in it. In the original section, the elements are arranged in ascending order.

From each section $P = \{l_1, l_2, \dots, l_n\}$, which was formed, we can create $l \le$ n new sections.

If $t \overline{\in} \sigma(l_i)$, than

$$P_i = \left(\left(P \setminus l_i\right) \cup \sigma\left(l_i\right)\right) \setminus I_i, \qquad (1)$$

where I_i – is the set of redundant elements in P_i.

In each newly constructed section, we will tidy up the elements by number growth, and replace the elements with the same numbers with others. All formed sections are entered in the general list of sections in lexicographic order by growth of numbers of composite elements. Let us note that when creating cuts with P, replacing various elements l_i with $\sigma(l_i)$ the same minimum section can be formed several times, which should be listed only for the first time.

Due to the correct numbering of the network elements, each newly constructed section is placed in a lexicographically ordered list after the section on which it is built.

Let us consider a way to exclude redundant elements from each P-built section

$$\tilde{P} = \left(P \setminus l_i\right) \cup \sigma\left(l_i\right). \qquad (2)$$

In the section, will $\tilde{P}$ some element r_j be redundant if all (s, r_j) paths or all (r_j, t) paths pass through some element $r_k \in P$. But for the accepted method of sections' creating on (s, r_j) paths of similar elements there can not be r_k in section $\tilde{P}$. Therefore, the search for redundant elements should be done, considering only segments of (r_j, t) paths.

We will put in order the elements by number growth in the generated section $\tilde{P}$. Then, to check for redundancy r_j concerning overshoot we can take into account the pieces of (r_j, t) paths, on which the number of the last element is no more than the number of the last element in the section $\tilde{P}$.

Let us denote as

$$\tilde{P} = \{r_1, r_2, \dots, r_m\}$$

a section, that is formed from a section

$$P = \{l_1, l_2, \dots, l_n\}$$

by replacing l_i with $\sigma(l_i)$. The procedure of excluding of redundant elements is the following.

The viewed sections are selected from the list of constructed sections in lexicographic order by their component numbers. When we view a section from it, new sections are generated by replacing those elements l_i for which $\sigma(l_i)$ do not have a t drain.

Let us consider how to convert a section list to a network. In a formalized form, this task can be formulated as follows.

There is a list μ, consisting of M rows $S_1, \ldots, S_M$, which has the following features.

1. The row $S_i = (a_{i1}, \ldots, a_{im})$ contains m_i non-repeating alphabet characters $A = \{a_i\}_{i=1}^{N}$.
2. Any character permutation in a row is allowed.
3. There are no S_i, $S_j \in \mu$, such, that $S_i \le S_j$.
4. There is at least one character pair a_i, a_j, for which no row $S_k \in \mu$ exists, such that $a_i \in S_k$, $a_j \in S_k$. In this case, we will say that the symbols a_i i a_j are incompatible.

We need to build the minimum orgraph *R* (*Y*, *F*) by the number of vertices, which has such properties.

In fact, this means that we need to define the order of characters in the list rows in order to combine the maximum number of matching characters in the rows and the incompatible characters would be in one row.

Let's select the three characters a_i, a_j i a_k such, that symbol a_i belongs to two rows at the same time S_p i S_q.

Let us introduce such concepts of form as direct form:

$$f = a_i(a_j,\, a_k)\,, \tag{3}$$

and inverse form

$$\overline{f} = (a_j,\, a_k)a_i\,. \tag{4}$$

The forms f_1 i f_2 we call as equivalent if $f_1 = a_i(a_j, a_k)$, a $f_2 = a_i(a_k, a_j)$. The forms f_1 i f_2 we call as alternative to each other if $f_1 = \overline{f_2}$ or $f_2 = \overline{f_1}$. Let's talk that f_i contradicts f_j, if the general system of inequalities formed by forms f_i i f_j, is non-essential. It is obviously, that the form $f = a_i(a_j, a_k)$ can only conflict with the following forms:

$$f_1 = a_j(a_i,\, a_l)\,;\;\; f_2 = a_k(a_i,\, a_l)\,;$$

$$f_3 = (a_j,\, a_l)a_i\,;\; f_4 = (a_k,\, a_1)\,, \tag{5}$$

where $\ell \in \left(\overline{1,N}\right)$.

Let's compose the initial system of forms $\Sigma = \Sigma_0$, that contains straight forms for all such triples of characters. Let us number the forms of this system (let them be Q pieces) and compose a matrix of contradictions of forms.

5. Research results

We will change the system of forms Σ, replacing, where it is necessary, the forms that contain it, with alternative ones to receive a system $\Sigma = \Sigma^*$ containing a minimum number of conflicting forms.

The value $\delta_i > 0$ is the indicator of the need to replace the form f_i with f_i^*.

We introduce auxiliary vectors D^+, D^- i Δ into consideration, in which, when replacing the form f_i with $\overline{f_i}$ i-row i and i-column of matrix V, is changed, as well as some components of the vectors D^+, D^- i Δ according to the following rule:

1. We replace v_{ij} with $-v_{ij}$ in the i-column of V matrix, $j = 1, \ldots, Q$, at the same time

$$\begin{gathered} d_i^+(new) = d_i^-(old), \quad d_i^-(new) = d_i^+(old), \\ \delta_i(new) = -\delta_i(old). \end{gathered} \tag{6}$$

2. We replace v_{ki} with $-v_{ki}$ in the i-column of V matrix, $k = 1, \ldots, Q$.

Doing the same for those f_i, in which $\delta_i > 0$, we gradually reduce the total number of contradictions in the system Σ, until it turns out that all $\delta_i \le 0$, $i = 1, \ldots, Q$, i and get a system of forms $\Sigma = \Sigma^*$ with a minimum number of contradictions. It will give a partial (or complete) order on the set of characters A, which will allow combining the maximum number of characters in rows to construct a graph of the necessary form.

6. SWOT analysis of research results

Strengths. The strengths of this research are that a comparative analysis of methods of personalized social data processing was carried out. Besides, the compilation of algorithms of analyzing of heterogeneous text and multimedia information, storage and presentation models of initial, intermediate analyzed and resulting data were made. It was established that the most promising is the use of the mathematical apparatus of graph theory, database theory and the concept of non-relational data stores, Big Data technology, text analytics technologies, parallel data processing methods, methods of neural networks' using, methods of multimedia data analyzing.

As a result, an oriented orthograph was built, which is a mathematical model of an oriented information system.

Weaknesses. The weaknesses of this research are related to the fact that the result is only a model of an oriented information system, which does not fully reflect the entire variety of existing and promising systems of processing of personal data and building of a social profile of both individuals and social groups of various differentiations.

Opportunities. The additional possibilities when using the above results in systems, as well as identification and verification of a digital image of a person are to optimize the required time and to improve the quality of such systems.

Threats. The options proposed in this paper are theoretical. Practical processing and justification prior to possible implementation solutions are required.

Conclusions

1. To improve the efficiency of automated conversion of unstructured social data, it is desirable to use OSINT technologies and supplement them using methods and algorithms to optimize Big Data transformation by building oriented graphs.

The rationale for using such an apparatus is the correct application of the theory of processing of large amounts of data and the theory of oriented graphs. The result of implementation of the algorithm is the ability to build an oriented graph, which is a mathematical model of oriented information networks that are quite often used in practice.

2. Thus, we have gained a new opportunity to further improve the functioning of computer information networks by optimizing of data structures by converting the information list into a corresponding oriented graph.

3. Further research should be directed towards the implementation of proposed theoretical developments in real decision-making systems. This will increase the weight of automated decision-making systems for analyzing of the social climate of society.

REFERENCES

1. Mayer-Schönberger, Viktor and Cukier, Kenneth (2013), *Big Data: A Revolution that Will Transform how We Live, Work, and Think*, Business & Economics, Houghton Mifflin Harcourt, 242 p.
2. White, T. (2015), *Hadoop: The Definitive Guide*, O'Reilly Media, Inc., 235 p.
3. Tiwari, G. (2011), *NoSQL: What It Is and Why You Need it > Definition and Introduction*, Professional NoSQL, Packt Publishing, 384 p.
4. Saravana, Balaji B., Karthikeyan, N.K. and Raj Kumar, R.S., (2018), “Fuzzy service conceptual ontology system for cloud service recommendation”, *Computers & Electrical Engineering*, Vol. 69, pp. 435–446, DOI: https://doi.org/10.1016/j.compeleceng.2016.09.013.
5. Merlac, V., Smatkov, S., Kuchuk, N. and Nechausov A. (2018), “Resourses Distribution Method of University e-learning on the Hypercovergent platform”, *Conf. Proc. of 2018 IEEE 9th International Conference on Dependable Systems, Service and Technologies. DESSERT’2018*, Ukraine, Kyiv, May 24-27, pp. 136-140, DOI: http://dx.doi.org/10.1109/DESSERT.2018.8409114
6. Mozhaiev, M., Kuchuk, N. and Usatenko, M. (2019), “The method of jitter determining in the telecommunication network of a computer system on a special software platform”, *Innovative technologies and scientific solutions for industries*, Vol. 4 (10), pp. 134-140. doi: https://doi.org/10.30837/2522-9818.2019.10.134.
7. Kuchuk, G., Kovalenko, A., Komari, I.E., Svyrydov, A. and Kharchenko, V. (2019), “Improving big data centers energy efficiency: Traffic based model and method”, *Studies in Systems, Decision and Control*, vol 171, Kharchenko, V., Kondratenko, Y. (Eds.), Springer Nature Switzerland AG, pp. 161-183, DOI: http://doi.org/10.1007/978-3-030-00253-4_8
8. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and A Byers, Hung (2011), “Big data techniques and technologies”, *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, pp. 27-31.
9. Donets, V., Kuchuk, N. and Shmatkov, S. (2018), “Development of software of e-learning information system synthesis modeling process”, *Advanced Information Systems*, Vol. 2, No 2, pp. 117–121, DOI: https://doi.org/10.20998/2522-9052.2018.2.20.
10. Dhivakar, B., Saravanan, S.V., Sivaram, M. and Krishnan R.A. (2012), “Statistical Score Calculation of Information Retrieval Systems using Data Fusion Technique”, *Computer Science and Engineering*, Vol. 2, Issue 5, pp.43-45, doi: http://doi.org/10.5923/j.computer.20120205.01
11. Sivaram, M., Yuvaraj, D., Amin Salih, Mohammed, Porkodi, V. and Manikandan V. (2018), “The Real Problem Through a Selection Making an Algorithm that Minimizes the Computational Complexity”, *International Journal of Engineering and Advanced Technology*, Vol. 8, iss. 2, 2018, pp. 95-100.
12. Buslov, P. (2017), “Analysis of using the features of communication models in social groups and virtual communities”, *Ukrainian Scientific Journal of Information Security*, vol. 23, issue 1, pp. 39-44.
13. Swami, M. and Thulasiraman, K. (1984), Graphs, networks and algorithms, Mir, Moscow, 455 p.
14. Kovalenko, A. and Kuchuk H. (2018), “Methods for synthesis of informational and technical structures of critical application object’s control system”, *Advanced Information Systems*, Vol. 2, No. 1, pp. 22–27, DOI: https://doi.org/10.20998/2522-9052.2018.1.04
15. Kuchuk G.A., Mohammad A.S. and Kovalenko, A.A. (2011), “The parametric data transmission control method for modifying the transport protocols of wireless networks”, *Information Processing Systems*, No. 8 (98), pp. 211-218.
16. Kuchuk G.A., Mohammad A.S. and Kovalenko, A.A. (2011), “Method for reducing data transmission time in a wireless network”, *Control, navigation and communication systems*, No. 3 (19), pp. 209-213.
17. Ruban, I., Kuchuk, H. and Kovalenko A. (2017), “Redistribution of base stations load in mobile communication networks”, *Innovative technologies and scientific solutions for industries*, No 1 (1), P. 75–81, doi : https://doi.org/10.30837/2522-9818.2017.1.075.
18. Nechausov, A., Mamuziĉ, I. and Kuchuk, N. (2017), “Synthesis of the air pollution level control system on the basis of hyperconvergent infrastructures”, *Advanced Information Systems*, Vol. 1, No. 2, pp. 21–26, DOI: https://doi.org/10.20998/2522-9052.2017.2.04 .
19. Zubov, V.S. (1999), *Programmer's handbook. Basic methods of solving graph problems and sorting*, IIID House "Filin". Moscow, 256 p.
20. Shamraev, A., Shamraeva, E., Dovbnya, A., Kovalenko, A. and Ilyunin, O. (2017), “Green microcontrollers in control systems for magnetic elements of linear electron accelerators”, *Studies in Systems, Decision and Control*, 74, pp. 283-305, DOI: https://doi.org/10.1007/978-3-319-44162-7_15
21. Ore, O. (1968), *Graph theory*, Nauka, Moscow, 336 p.

Received (Надійшла) 28.10.2020
Accepted for publication (Прийнята до друку) 27.01.2021

ABOUT THE AUTHORS / ВІДОМОСТІ ПРО АВТОРІВ

Можаєв Михайло Олександрович – кандидат технічних наук, завідувач сектором комп'ютерно-технічних, телекомунікаційних досліджень та досліджень відео-, звукозапису Харківського науково-дослідного інституту судових експертиз імені засл. проф. М. С. Бокаріуса, Харків, Україна;
Mykhailo Mozhaiev – Candidate of Technical Sciences, Head of Department of Computer Engineering, Telecommunications, Video- and Audio-recording Research, Hon. Prof. Bokarius Kharkiv Research Institute of Forensic Examinations, Kharkiv, Ukraine;
e-mail: mikhail.mozhayev@hniise.gov.ua; ORCID ID: http://orcid.org/0000-0003-1566-9260.

Буслов Павло Володимирович – аспірант, Харківський національний університет радіоелектроніки, Харків, Україна;
Pavlo Buslov – postgraduate student, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: p.buslov@ukr.net; ORCID ID: http://orcid.org/0000-0002-2558-4984.

Метод моделювання соціального профілю з використанням оптимізації перетворення структур Big Data

М. О. Можаєв, П. В. Буслов

Анотація. Об'єктом дослідження є методи та алгоритми оптимізації перетворення Big Data для побудови моделі соціального профілю, предметом дослідження – методи побудови соціального профілю. Для особи, яка приймає рішення, актуальна проблема науково методологічного і інструментального переоснащення для ефективного виконання комплексу управлінських завдань і протистояння принципово новим викликам і загрозам у суспільстві. Це завдання безпосередньо пов'язана з проблемою побудови моделі соціального профілю як індивідуума, так і соціальної групи в цілому. Тому задача оптимізації методів побудови математичної моделі соціального профілю є безумовно актуальною. В ході дослідження використовувалися методи математичного апарату теорії графів, теорії баз даних і концепції нереляційних сховищ даних, технології Big Data, технологій текстової аналітики, методів паралельної обробки даних, методів застосування нейронних мереж, методів аналізу мультимедіа даних. Дані методи були інтегровані в загальний метод, названий методом покращення ефективності побудови математичної моделі соціального профілю. Запропонований метод дозволяє покращати адекватність моделі соціального профілю, що дозволить суттєво покращати та спростити функціонування інформаційних систем прийняття рішень, які ґрунтуються на знаннях про соціальні переваги тих чи інших суспільних груп, що дозволить здійснювати динамічну корекцію їх поведінки. Отримані результати апробації методу дозволяють вважати його ефективним інструментом отримання об'єктивної інформаційної моделі соціального портрету соціальної групи. Це пов'язано з тим, що коректність постановки і рішення задачі забезпечила отримання адекватних результатів. На відміну від існуючих, запропонований метод моделювання, який використовує орієнтований граф дозволяє суттєво покращати якість та адекватність цього процесу. Подальші дослідження необхідно спрямувати на реалізацію пропонованих теоретичних розробок в реальні системи прийняття рішень. Це дозволить збільшити вагу автоматизованих систем прийняття рішень для аналізу соціального клімату суспільства.

Ключові слова: модель соціального профілю; інформаційні системи прийняття рішень; теорія графів; технології Big Data; соціальний портрет соціальної групи.

Метод моделирования социального профиля с использованием оптимизации преобразования структуры Big Data

М. А. Можаев, П. В. Буслов

Аннотация. Объектом исследования являются методы и алгоритмы оптимизации преобразования Big Data для построения модели социального профиля, предметом исследования - методы построения социального профиля. Для лица, принимающего решения, актуальна проблема научно методологического и инструментального переоснащение для эффективного выполнения комплекса управленческих задач и противостояния принципиально новым вызовам и угрозам в обществе. Эта задача непосредственно связана с проблемой построения модели социального профиля как индивидуума, так и социальной группы в целом. Поэтому задача оптимизации методов построения математической модели социального профиля является безусловно актуальной. В ходе исследования использовались методы математического аппарата теории графов, теории баз данных и концепции нереляционных хранилищ данных, технологии Big Data, технологий текстовой аналитики, методов параллельной обработки данных, методов применения нейронных сетей, методов анализа мультимедиа данных. Данные методы были интегрированы в общий метод, названный методом повышения эффективности построения математической модели социального профиля. Предложенный метод позволяет улучшить адекватность модели социального профиля, что позволит существенно улучшить и упростить функционирования информационных систем принятия решений, основанные на знаниях о социальных преимуществах тех или иных общественных групп, что позволит осуществлять динамическую коррекцию их поведения. Полученные результаты апробации метода позволяют считать его эффективным инструментом получения объективной информационной модели социального портрета социальной группы. Это связано с тем, что корректность постановки и решения задачи обеспечила получение адекватных результатов. В отличие от существующих, предложенный метод моделирования, который использует ориентированный граф позволяет существенно улучшить качество и адекватность этого процесса. Дальнейшие исследования необходимо направить на реализацию предлагаемых теоретических разработок в реальные системы принятия решений. Это позволит увеличить вес автоматизированных систем принятия решений для анализа социального климата общества.

Ключевые слова: модель социального профиля; информационные системы принятия решений; теория графов; технологии Big Data; социальный портрет социальной группы.