

The Intelligent Monitoring of Messages on Social Networks

Serhii Holub¹ [0000-0002-5523-6120] and Nataliia Khymytsia² [0000-0003-4076-3830]
and Maria Holub¹ [0000-0002-0553-8163], and Solomiia Fedushko² [0000-0001-7548-5856]

¹Cherkasy State Technological University, Shevchenko Boulevard 460, Cherkasy, Ukraine.

²Lviv Polytechnic National University, Lviv 79013, Ukraine

s.holub@chdtu.edu.ua, nhymytsa@gmail.com,
sashasokolovksa92@gmail.com, solomiia.s.fedushko@lpnu.ua

Abstract. The results of research on the processes of use of information technology of multilevel intellectual monitoring for classification of text internet-messages in social networks are presented. The selection of messages was carried out by experts. Two textbook classes were formed. The first class included texts of malefactors. The second class was formed on the basis of posts that, in the opinion of the authors, are not malefactors. The processes of decomposition of text messages of social networks, the adaptive formation of dictionaries of signs, use of the agent approach to the construction of the classifier, and its use for recognition of malefactors are investigated in the work. Experimentally confirmed hypotheses about the existence in Ukrainian-speaking social communities of participants who have a common style of presenting messages. This may be the result of their joint special training for this type of activity. The effectiveness of classifying agents was assessed by the number of correctly classified text messages. Conclusions about the belonging of messages to a certain class were made based on the results of the classification of observation points that describe individual texts. The number of correctly classified messages prepared by the agent was over 92%. The hypothesis about the possibility of using the method of adaptive formation of the dictionary of features in the technology of classification of text messages, with a volume of more than 100 characters, has been experimentally confirmed. This allows you to automate the processes of protection of the information space of Ukrainian-language Internet content of social networks.

Keywords: social networks, web communities, content, information, sources, monitoring, content analysis, intellectual agents, dictionary of features.

1 Introduction

Intelligent monitoring is an information technology that provides knowledge to decision-making processes by organizing continuous observations and processing their results. Information is obtained in the form of information about the properties

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). COAPSN-2020: International Workshop on Control, Optimisation and Analytical Processing of Social Networks

of the object of monitoring, knowledge is obtained in the form of identified patterns and trends. Monitoring intelligent systems (MIS) is a software implementation of information technology of multilevel intelligent monitoring. Features of application of this technology demand efficiency in the formation of new knowledge and their adaptability to change of properties of the environment. Therefore, MIS organizes continuous observations and implements their processing to build a model knowledge base (MKB) - as a means of obtaining information and storing knowledge after the transformation of the results of observations of the objects of monitoring. Building MKB using agent approaches based on cloud technologies allows quickly respond to changing monitoring tasks and adapt to changing properties of monitoring objects.

Information monitoring is one of the priorities in a hybrid war. Depending on the media, the task of converting the results of observations to a typical form of input data is performed. The most common media in social networks are text messages. The use of intelligent monitoring of text messages in social networks has certain features. These messages are short and presented in a special style. Therefore, conventional methods of analysis and previously constructed models for their development are not effective. In particular, the previously solved attribution problems, which allow classifying the authors of printed works, will not allow classifying the authors of messages on social networks.

Monitoring of internet-messages is ordered to provide information and knowledge of various decision-making processes. The most common are classification tasks. But the results of the tasks of identification and forecasting are useful. And the peculiarity is that to perform these tasks it is necessary to process text, video, and audio messages, which in their original state require the transformation into a typical form of an array of numerical characteristics. And often there is a need to combine information obtained from different messages of one author. Therefore, special importance is given to methods of converting messages from the primary form to a typical matrix of an array of numerical characteristics of informative features and the processes of consolidation of simulation results. The informativeness of this array must exceed the limit of informative sufficiency for the available means of model synthesis. The results of the consolidation of information from different messages should allow obtaining knowledge in the form of an emergent effect from the coordination of interactions of diverse models. Therefore, the construction of a dictionary of features for text, audio, or video messages should be considered as a solution to one of the local tasks in the process of fulfilling the global task, the consolidation of the interactions of models of monitoring objects in social networks.

This paper presents the results of research on the processes of intellectual analysis of text messages from social networks. The process of solving the problem of monitoring text messages is an element of information technology for intelligent monitoring of Internet content.

This technology is designed to identify people who have undergone the same type of training, who use social networks to influence the opinion of readers of the Ukrainian-language information environment; and can be classified using the Intelligent Monitoring System (IMS) [1].

2 Related Works

At the beginning of the XXI century, the issue of developing various methods of analyzing Internet content and developing new approaches to solving this problem is becoming the subject of interest of even more authoritative scientists.

In particular, practical approaches to the purification, indexing, and extraction of textual information are published in a collection of scientific papers edited by M. W. Berry [2]. Valuable for our study were primarily articles that address various aspects of algorithmic progress in discriminant analysis, investigate spectral clustering, describe trends in identifying trends and removing synonyms, describe case studies in web-mining, and analyze customer support logs to extract relevant topics and query characteristics. Another fundamental edition, prepared by Aggarwal, Charu C., Zhai, ChengXiang, focuses on text embedded in heterogeneous and multimedia data, making the extraction process much more complex; methods such as transfer training and multilingual mining are described [3].

It is important for a comprehensive study of the issues of intellectual monitoring is the analysis of different approaches to the practical application of content analysis techniques. In our opinion, it is necessary to note the work of O. Ivanov, which proposed a technical concept of the content of the analytical program - independent of the text being analyzed, able to identify significant links between words, determine the structure of analyzed texts [4]. No less relevant is the approach proposed by researchers I. Khomytska, V. Teslyuk, O. Morushko, A. Holovatyy. These scientists, in particular, present the results of the applied methods, models and software, which confirm that the author's attribution of the text at the phonological level is more effective [5].

Studies by A. Peleshchynshyn, V. Vus, O. Markovets, S. Albota analyze a special system of user activity indicators. The authors also explore the linguistic methods of influence that are purposefully used in online communication [6-8]. The work of O. Markovets, R. Pazderska, N. Dumanskyi, I. Dronyuk presents methods that will help determine the level of trust in the author, such as monitoring, attestation, organization and personification [9]. Y. Syerov, N. Shakhovska and S. Fedushko in the context of research on artificial intelligence, proposed a new method for determining the adequacy of data of personal medical profiles [10]. S. Fedushko and E. Benova's work on computer and language content analysis, authentication of personal data of user accounts, consolidation, and analysis of user information tracks and their behavioral models of communication in social areas of the Internet, proposed a comprehensive approach to identifying dangerous threats, which have a negative impact on Internet users and are the basis for creating countermeasures against information and communication threats [11]. Within the research work of N. Khymytsia, T. Ustyianovych, and I. Dronyuk several series of stages of search and processing of messages from web forums containing historiographical information by means of web scraping, data analysis, and big data analysis have been developed [12].

Among the numerous theoretical, methodological and practical studies that highlight the problems of building monitoring systems, in the context of our topic, it is important to use multilevel information-analytical monitoring, which is associated

with information analysis, model synthesis, expert evaluation, diagnosis and forecasting and organization of continuous observations of signs coming from different sources. The aim is to regularly provide information support for decision-making processes in various subject areas. Such approaches using intelligent monitoring are considered in the study of S. Kynytska, S. Holub where it is proposed to implement the methodology of creating information systems of multilevel intelligent monitoring by using an agent approach involving cloud technologies [1]. Also, in the works of S. Holyb, N. Khymytsya proposed new aspects of the practical application of monitoring information systems to identify similar historical periods in the economy [13, 14].

The results of information retrieval allow us to conclude that the information technology of intelligent monitoring can be used to solve the problem of classifying text messages on social networks according to the properties of their authors. The methods of text mining used by this technology were used to classify messages larger than 500 characters. The classification of much shorter messages requires the use of new approaches to the construction of classifiers, in particular the agent approach [1].

Thus, the purpose of this work is to study the processes of classification of texts of 100 characters on the example of messages from one of the social networks. The results of research should provide an opportunity to automate the processes of protection of the information space of Ukrainian-language content on the Internet and create conditions for the development of measures to combat the influence of attackers.

3 The research of the process of classification of text messages in social networks

To achieve this goal, several hypotheses were formulated.

1. Ukrainian-language social networks are influenced by persons who have undergone special training and can be classified according to a similar style of text messages;
2. The use of an adaptive formed dictionary of features based on the results of the deep decomposition of the text [15] will provide sufficient information on the array of numerical characteristics of text messages in social networks to perform the task of classifying their authors.

An experiment was performed to test these hypotheses. The problem of classifying the authors of text messages on the social network "Facebook" was solved.

46 messages were selected by experts, among which 34 messages formed a finite set T :

$$T = \{t_1, t_2, \dots, t_{34}\}, \quad (1)$$

representing a training sample. Expertly, two classes of texts of the set K were formed:

$$K = \{k_1, k_2\} \quad (2)$$

The class of texts, the authors of which were identified as generators of unfair influences, was called "Bots" (Class 1). Class 2 included the original messages of the authors. Class 2 is called "Original".

It was necessary to construct a classifier f to ensure the display of elements of the set

$$C = \{t_{35}, t_{36}, \dots, t_{46}\}, \quad (3)$$

that is, new messages in "Facebook", on the elements of the set K :

$$f: C \rightarrow K. \quad (4)$$

Figure 1 describes the process of machine learning of the model-classifier of text Internet-messages.

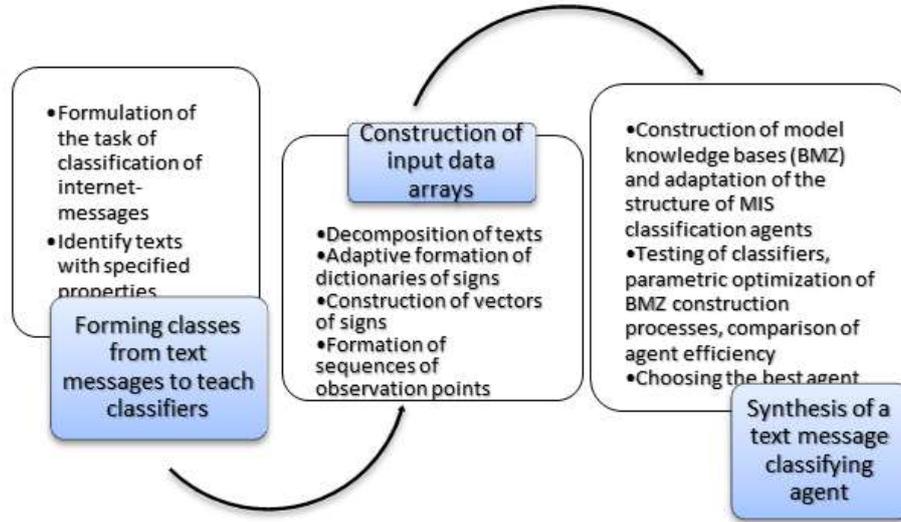


Fig. 1. Functional diagram of the construction of the classifier of Internet-messages

According to the method described in [15], text messages were divided into sections of 100 characters and converted into a vector form of sequences of observation points in the multidimensional feature space of the input data array (IDA). 159 vectors of numerical characteristics of text messages were formed.

Observation points (vectors of numerical characteristics), which are the result of transformations of texts from the class "Bots", were marked as "Our". Other vectors were designated as "Alien". 3 sequences of observation points were identified. Sequence "A" contained 73 observation points and was used to train models. Sequence "B" was used to implement the external criterion of model quality and contained 72 observation points.

The sequence "C" contained 14 points, was used to test the results of the classification of texts and did not participate in the synthesis of the model.

Table 1. shows a fragment of the array of input data.

Table 1. An array of input data of the Monitoring intelligent system.

Observation point	Class	Features (simulation variables)							
		а	б	в	г	д	е	є	...
	Y	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	...
Bot 1_ukr (0) [100]	100	6	3	3	0	2	1	2	...
Bot 2_ukr (0) [100]	100	3	2	3	3	2	5	0	...
Bot 2_ukr (1) [100]	100	3	0	3	3	1	5	0	...
Bot 2_ukr (2) [100]	100	6	3	3	1	1	3	1	...
Bot 3_ukr (0) [100]	100	9	1	7	0	3	1	1	...
Bot 3_ukr (1) [100]	100	2	1	4	0	3	8	1	...
Bot 3_ukr (2) [100]	100	4	2	2	1	1	1	2	...
...
Original 1_ukr (0) [100]	-100	1	3	6	0	2	4	0	...
Original 1_ukr (1) [100]	-100	5	5	0	1	3	4	0	...
Original 2_ukr (0) [100]	-100	5	1	3	2	2	6	1	...
Original 2_ukr (1) [100]	-100	9	1	6	0	2	4	1	...
Original 3_ukr (0) [100]	-100	8	1	5	3	3	4	0	...
Original 3_ukr (1) [100]	-100	3	5	4	0	3	5	0	...
Original 4_ukr (0) [100]	-100	9	0	4	0	2	3	1	
Original 4_ukr (1) [100]	-100	5	0	3	0	4	6	1	
Original 4_ukr (2) [100]	-100	6	3	3	0	2	4	2	
Original 4_ukr (3) [100]	-100	12	2	5	2	2	9	0	
...

Unlike typical content analysis, the adaptive formation of the dictionary of features involves the decomposition of texts of one class to the level of individual characters and their combinations, the choice of criteria for informative features, and setting the limit of informative sufficiency for the IDA. According to these parameters, the classification features of messages from social networks are selected. The dictionary of features is used to convert text elements (windows) into a vector form of an array of numerical characteristics.

Each feature of the dictionary forms an element of the vector of characteristics of the text. A separate vector depicts the observation point in a multidimensional feature space. The observation point is a row in the IDA matrix (Table 1). Symbols are column names. They are modeling variables in the synthesis of classifier models.

The IDA, which is formed by sequences "A" and "B" of observation points, was fed to the entrance of the MIS. Using the agent approach described in [1], MIS synthesized

an agent-classifier of text messages. In the process of classification of the classifier involved intelligent agents, the structure of which is based on the algorithms of MGCA [16], neural networks of several topologies, hybrid methods of model synthesis. The formation and adaptation of the structure of intellectual agents took place in conditions of competition - they were given: the same modeled indicator and quality criterion - the number of correctly classified texts. The highest quality was the agent, MKB which was based on the multi-row algorithm MGCA. After parametric optimization of the MKB synthesis process, it turned out that the quality criterion of the models used the condition of the minimum standard deviation of the simulation results from the actual values of the simulated indicator on the sequence of points "B":

$$S_y = \sqrt{\frac{\sum_{i=1}^{72} (y_i - y_i^*)^2}{72}}, \quad (5)$$

y_i – calculated value of the simulated class index, y_i^* – the actual value of the class indicator.

In Fig. 2 shows a fragment of the MKB agent-classifier of text Internet-messages.

$$\begin{aligned} Y = & +15,6622093047518 + 0,343662829679701 \cdot y_2 + 0,5 \\ & 75071591951118 \cdot x_8 - 3,07611074952025E- \\ & 5 \cdot y_2^2 + 4,38629067311242E- \\ & 5 \cdot x_8^2 + 0,0326096880710039 \cdot x_6^6 - \\ & 1,51997544750547E-5 \cdot x_6^2 + 9,71819962121924E- \\ & 5 \cdot y_2 \cdot x_8 - 1,98675405733841E- \\ & 8 \cdot y_2^3 + 8,76337274466889E-9 \cdot y_2 \cdot x_8^2 - \\ & 2,70919002926536E-8 \cdot x_8 \cdot y_2^2 - \\ & 1,04270819793911E-8 \cdot x_8^3 + 2,81175741291694E- \\ & 12 \cdot y_2^4 - 3,42114384513513E- \\ & 12 \cdot y_2^2 \cdot x_8^2 + 4,15458173320306E-12 \cdot x_8^4 - \\ & 6,10546690069651E- \\ & 5 \cdot x_6 \cdot x_8 + 1,14363884060127E- \\ & 9 \cdot x_6^3 + 8,19778350736378E- \end{aligned}$$

Fig. 2. A fragment of the model knowledge base of the agent-classifier of social community messages

MKB, a fragment of which is presented in fig. 2, is used as an algorithm to convert the results of observations of messages in social communities, presented in the form of a matrix of numerical characteristics in the conclusions about whether the author of the message is a wrecker. Y has a multilayer structure, which includes models of lower layers (y_1, y_2, \dots) and numerical characteristics of the features of the IDA, presented in fig. 2. The methodology of building models of model knowledge of intellectual agents [17] is a separate area of research and is not considered in this work.

The effectiveness of the classifier was assessed by the number of correctly classified observation points and the correctness of conclusions about the belonging

of new text messages in the social community to one of the classes. They are combined in the sequence "C". If the message was described by several observation points, the conclusion that the message belongs to one of the classes was made by most of the results of the classification of observation points of the sequence "C", which describe a single message.

4 Results

Table 2 presents the test results of the classifying agent.

Table 2. The classification of text messages from the test sequence "C"

Observation point	Class	The value of the classifier	Conclusion	Result
Bot 36_ukr (0) [100]	1	-50,82	Bot	1
Bot 37_ukr (0) [100]	1	-74,67	Bot	1
Bot 37_ukr (1) [100]	1	-34,48		
Bot 38_ukr (0) [100]	1	-52,53	Bot	1
Bot 39_ukr (0) [100]	1	-38,19	Bot	1
Bot 40_ukr (0) [100]	1	-67,75	Bot	1
Original 41_ukr (0) [100]	2	15,25	Not a bot	1
Original 42_ukr (0) [100]	2	37,86	Not a bot	1
Original 43_ukr (0) [100]	2	50,90	Not a bot	1
Original 44_ukr (0) [100]	2	22,02	Not a bot	1
Original 45_ukr (0) [100]	2	27,03	Not a bot	1
Original 46_ukr (0) [100]	2	-5,14	Bot	0
Original 47_ukr (0) [100]	2	53,30	Not a bot	1

Among the points of the test sequence "C", the agent singled out among the analyzed texts those whose authors were characterized by experts as malefactors. The agent error in test № 12 can be eliminated by adjusting the threshold value of the separating surface. An increase in the diversity of the classifier can be achieved by increasing the number of texts in the classroom and increasing the number of texts that are described by observation points in the examination sequence "C".

The test results of the classifier, shown in table 1, allow us to assert its adequacy. The number of classified text messages from the social community Facebook was over 92%.

As it was possible to teach the classifier of texts according to the formed text classes, it means that the authors of text messages are united by common properties. Therefore, Hypothesis 1 received its experimental confirmation.

Acceptable results of the classifier test prove the effectiveness of the used intelligent monitoring technology. Therefore, Hypothesis 2 should be considered experimentally confirmed.

5 Conclusions

For the first time, the results of the application of information technology of multilevel intellectual monitoring for classification of texts, the volume of 100 characters are received. This allows to expand the capabilities of monitoring intelligent systems and use them for intelligent analysis of Internet messages in social communities. It is possible to automate the processes of protection of the information space of Ukraine, to identify and analyze examples of information impact on readers of Ukrainian-language content.

6 References

1. Kynytska, S, Holub, S.: Multi-agent Monitoring Information Systems. In: Palagin A., Anisimov A., Morozov A., Shkarlet S. (eds) *Mathematical Modeling and Simulation of Systems. MODS 2019. Advances in Intelligent Systems and Computing*, vol 1019. pp 164-171. Springer, Cham (2019).
2. *Survey of Text Mining I: Clustering, Classification, and Retrieval*. Ed. by M. W. Berry. 2004. Springer, (2003).
3. Aggarwal, C. C., Zhai, C.: *Mining Text Data*. 527 p. Springer (2012).
4. Ivanov, O.,V.,: Computer content analysis: problems and perspectives of virishennya. *Methodology, theory and practice of sociological analysis of ordinary sou-spilstva*. pp. 335-340. (2009).
5. Khomytska, I., Teslyuk, V., Holovatyy, A., Morushko O.: Methods, models and means of the system for differentiation of phonostatistical structures of english functional styles. Development of methods, models and means of authorship attribution of a text. In: *Eastern-European Journal of Enterprise Technologies*. № 3/12 (93). pp. 41–46. (2018).
6. Peleshchyshyn, A., Vus V., Markovets O., Albota S.: Identifying Specific Roles of Users of Social Networks and Their Influence Methods. In: *Proceedings of the 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018*, pp. 39–42. Lviv. DOI: 10.1109/STC-CSIT.2018.8526635. (2018).
7. Peleshchyshyn, A., Vus, V., Albota, S., Markovets, O. A Formal Approach to Modeling the Characteristics of Users of Social Networks Regarding Information Security Issues. *Advances in Intelligent Systems and Computing*. Volume 902, 2020, Pages 485-494. 2nd International Conference of Artificial Intelligence, Medical Engineering, Education, AIMEE 2018; Moscow; Russian Federation; 6 October 2018 through 8 October 2018; Code 226259. (2018).
8. Trach, O., Peleshchyshyn, A.: Development of directions tasks indicators of virtual community life cycle organization, *International Scientific and Technical Conference "Computer Sciences and Information Technologies*, pp. 127-130 (2017).
9. Markovets, O., Pazderska, R., Dumanskyi, N., Dronyuk, I. Analysis of citizens' appeals in heterogeneous web services *CEUR Workshop Proceedings Volume 2392*, 2019, Pages 184-198 1st International Workshop on Control, Optimisation and Analytical Processing

of Social Networks, COAPSN 2019; Lviv; Ukraine; 16 May 2019 through 17 May 2019; Code 149063

10. Syerov, Y., Shakhovska, N., Fedushko, S.: Method of the Data Adequacy Determination of Personal Medical Profiles. Proceedings of the International Conference of Artificial Intelligence, Medical Engineering, Education (AIMEE2018). Advances in Artificial Systems for Medicine and Education II. Volume 902, 2019. pp. 333-343. https://doi.org/10.1007/978-3-030-12082-5_31. (2019).
11. Fedushko, S., Benova, E.: Semantic analysis for information and communication threats detection of online service users. The 10th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2019) November 4-7, 2019, Coimbra, Portugal. Procedia Computer Science, Volume 160, 2019, Pages 254-259. <https://doi.org/10.1016/j.procs.2019.09.465>. (2019).
12. Khymytsia, N., Ustyianovych, T., Dronyuk, I. Identification and Modeling of Historiographic Data in the Content of Web Forums. CEUR Workshop Proceedings Volume 2392, 2019, Pages 297-308. 1st International Workshop on Control, Optimisation and Analytical Processing of Social Networks, COAPSN 2019; Lviv; Ukraine; 16 May 2019 through 17 May 2019; Code 149063. (2019).
13. Holyb, S., Khymytsia N.: The use of multi-level modeling in the cliometric studies process. In: Proceedings of the 13th International Conference "Modern Problems of Radio Engineering, Telecommunications and Computer Science", TCSET 2016, pp. 733–735. Lviv-Slavske, Ukraine (2016).
14. Golub, S., Khymytsia, N.: The Method of Cliodinamik Monitoring. In: Proceedings of the 2nd International Conference on Data Stream Mining and Processing, TDSMP 2018, pp. 223–226. Lviv (2018).
15. Holub, M., S.: Form of mass input of recent tributes in the classification of texts in the technology of information monitoring. Mathematical machines and systems. 2018. No. 1. pp. 59-66. (2018).
16. Ivakhnenko, A.G.: Inductive method of self-organization of models of complex systems. Kiev, Naukova Dumka, 1981. 296 p. (1981).
17. Zhiryakova I.A., Holub S.V. New approach to conceptual knowledge. Technical science and technology. 2015, No. 2. pp. 78-82. (2015).