Prospects of information systems for detecting the propaganda texts

Tarasenko Yaroslav

Cherkasy State Technological University,
460 Shevchenko Blvd, Cherkasy, UA-18006, Ukraine,
yaroslav.tarasenko93@gmail.com

Abstract. *The work is devoted to searching the promising directions of developing the information systems for detecting the propaganda texts in order to increase the reliability of their functioning in the context of information confrontation and to ensure national security. As a result of researching the existing models and information systems for detecting the signs of psycholinguistic influence and propaganda, a promising direction of their development was found, which consists in applying the texts' automated classification. Based on the comparative analysis of information systems for morphological analysis in text categorization with the system for defining the morphological-syntactic and semantic category of the propagandist's psycholinguistic portrait, the perspective ways of developing the information systems for detecting the propaganda texts are determined.*

Keywords: psycholinguistic influence, propaganda detection, automated texts' classification, semantic category, quantum-semantic study.

I. INTRODUCTION AND PROBLEM STATEMENT

Today, more and more states are united in a single information space. In addition to the positive features, such as free experience exchange, quicker response in business, economics or politics, this phenomenon leads to introducing the concept of information warfare. Ukraine is also involved in information warfare, and based on [1], as no one else needs reliable tools to protect and counteract the usage of information weapons. The article also states that one of the means for conducting information war is purposeful informational influence on the population. It can be said for sure that this type of information weapon is one of the most dangerous. However, as noted in [2] there are subspecies of psychological weapons, such as information-psychological, linguistic, psychotronic, psychophysical, psychotropic and somatopsychological weapons. All types of informational-psychological weapons except the linguistic one are not difficult to detect due to the aggressiveness of the action and the inability to implement certain types in the information space. So, linguistic weapon is definitely the most dangerous.

There are modern information systems for detecting signs of information-psychological manipulation, based on models [3-4], which use different approaches to detect the influence on consciousness. However, even promotional text will have the following characteristics, but the real danger is the propaganda text. There are automated information systems for propaganda detection in textual data based on the transfer learning approach [5], or on the methods of propagandist's identification in the social network [6]. But a study of information systems for propaganda detection in textual data has shown that the mentioned above methods and information systems for propaganda detection are aimed at investigating the offender's behavior, identifying signs of misinformation, etc., and this is not sufficient for conducting a deep investigation of well-hidden propaganda that affects consciousness. Although all systems agree on a common approach which is in automated text analysis.

Therefore, there is a need to explore promising areas for the further development of existing information systems for detecting propaganda texts and, as a consequence, a need to develop approaches for their effectiveness and reliability improvement, in context of increasing danger from the information warfare and the national security providing as a whole.

II. PROBLEM SOLUTION AND RESULTS

Improving the reliability of propaganda detection information systems is possible by improving the efficiency of automated text analysis. Since only semantic analysis can determine the hidden meaning and its perception by the target audience, research should be directed to a more comprehensive concept of automated textual analysis, namely, automated text classification [7], which includes automated morphological or semantic analysis of textual information, as well as the syntactic and emotional analysis used in the mentioned models and systems of detecting the propaganda.

In addition, the article [7] presents the experimental analysis results of various automated information systems for classifying texts which are based on diverse approaches to the terms' detection and classification, including English terms. However, the task of identifying the propaganda has its own specificity, which requires special look on analyzing the text. Particular attention is paid to known morphological analysis algorithms and the results are presented.

In turn, in [8] it was carried out the work on the use of quantum-semantic research in the text's formation or modification according to propagandist's individual semantic function. At the same time, the morphological-syntactic and semantic category of the propagandist's psycholinguistic portrait is considered.

In the course of the comparative experiment, an automated information system for determining the coordinates of the semantic particle was implemented and the results of the morphological analyzer functioning when automatically classifying the texts were compared with the study results of the morphological-syntactic and semantic category of the propagandist's psycholinguistic portrait. The experiment was based on a set of 50 texts with signs of psycholinguistic influence written by different authors. Special elements of information propaganda were added in 30% of the texts manually using the most common methods of both explicit and

covert propaganda. The results of the comparison are shown in the table 1.

Table 1. The characteristics of morphological analysis in text categorization compared with the morphological-syntactic and semantic category definition

Characteristics	Morph. analysis (in categorization)				Morphological-syntactic category				Semantic category	
Transaction time (sec)	0,5-3				2-5				15-45	
Text size (thousands of characters)	1		5		1		5		1	5
Number of tests	15	50	15	50	15	50	15	50	50	50
Accuracy of propaganda definition	23 %	20 %	31 %	29 %	48 %	45 %	64 %	61 %	79 %	87 %

The analysis has shown that categorization by morphological units is not accurate. However, there was no detected efficiency increase caused by changing the specific morphological analysis algorithm or by the experiments' number. The main difference is in the fundamental approach to text analysis and in the direction of the further research. Improvement of accuracy is possible on the basis of analyzing the semantic category, especially in the context of considering the propagandist's psycholinguistic portrait.

Thus, based on the experiment results, it is possible to note the following perspective ways of developing the information systems for detecting the propaganda texts, which are considered to be able to improve the reliability and efficiency of the automated process of information propaganda detection.

1. Conducting semantic research to provide a process of deep analysis and to increase accuracy in the automated categorization of propaganda texts, taking into account the types of propaganda activities.

2. The automated morphological analysis algorithms adaptation for taking into account the features of the propagandist's psycholinguistic portrait, as well as the text's propaganda discourse and the offender's psycho-emotional characteristics.

3. The quantum-semantic analysis usage for taking into account the quantum nature of the perception the word's forms as well as semantic categories, which improves the accuracy of detecting the hidden propaganda.

4. Replacing popular approaches of text analysis by the neural networks with faster computer hermeneutic systems that do not require a long and painstaking process of training the neural network, the essence difficulty of which is in selection and classification a core sample test.

III. CONCLUSIONS

For developing the approaches to improve the efficiency of information systems for the detecting the propaganda texts, as well as to increase their reliability considering the increasing risk of hostile psycholinguistic influence, it is proposed to increase the efficiency of information systems for automated text classifying in such areas as focusing on semantic analysis, retraining the algorithms of morphological analysis, using the quantum-semantic approach and applying the computer hermeneutics in complex in order to improve the reliability of textual data categorization considering the informational propaganda.

REFERENCES

[1] Malyk Ya. Information war and Ukraine / Ya. Malyk // Academic papers collection «Democratic governance», 2015. – Issue 15. URL: http://www.lvivacademy.com/vidavnitstvo_1/visnyk15/fail/Malyk.pdf

[2] Makarenko S.I. Information confrontation and electronic warfare in network-centric wars at the beginning of the XXI century / S.I. Makarenko: monograph. – St. Petersburg: High technology, 2017. – 546 p.

[3] Goncharov I.V. Modeling the processes of information-psychological impact in social networks / I.V. Goncharov, P.A. Parinov, A.A. Sirota // Proceedings of Voronezh State University. Series: Systems analysis and information technologies, 2018. – № 2. – P. 93-104.

[4] Java S. Detection of Online Manipulation to Prevent Users Victimization / S. Java, F.L. Basheer, S. Riaz, M.J. Kaur, A. Mushtaq // Proceedings of Amity International Conference on Artificial Intelligence, Dubai, UAE, February 4-6, 2019. – P. 593-599.

[5] Aggarwal K. NSIT@NLP4IF-2019: Propaganda Detection from News Articles usingTransfer Learning / K. Aggarwal, A. Sadana // Proceedings of the 2nd Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, Hong Kong, China, November 4, 2019. – P. 143–147.

[6] Orlov M. Using Behavior and Text Analysisto Detect Propagandistsand Misinformers on Twitter / M. Orlov, M. Litvak // Proceedings of the 5th International Conference «Information Management and Big Data», Lima, Peru, September 3-5, 2018. – P. 67-74.

[7] Batura T.V. Automatic text classification methods / T.V. Batura // Software and Systems, 2017. – Vol. 30, № 1. – P. 85-99.

[8] Tarasenko Ya. Determining the coordinates of semantic particle in english text with a known psycholinguist portrait of propagandist / Ya. Tarasenko // Ukrainian Information Security Research Journal, 2019. – № 21 (3). – P. 168-174.