

NEURAL NETWORK ALGORITHMS AND METHODS FOR MONITORING THE PSYCHOLOGICAL STATE OF SOCIETY DURING EPIDEMICS

Aleksey Rogachev^{1,2*}, and *Tamara Mazaeva*^{1,2}

¹Volgograd State University, 26, Universitetskiy Ave., Volgograd, Russia

²Volgograd State Technical University, 28, Lenin Ave., Volgograd, Russia

Abstract. The article presents the issue of monitoring the socio-psychological state of society in the period of epidemics by means of neural network algorithms and methods. Publications of sociologists, psychologists, and philologists, who have created a number of methods for in-depth analysis of emotions and tonality of texts in the Internet media, including cognitive and interpretive decoding, are devoted to substantiating approaches and methods for studying the content of Internet content. The purpose of the study is to substantiate the methods and computer tools for studying the socio-psychological state of society in crisis situations, in particular epidemics, based on neural network technologies based on Internet resources. The article considers methodological approaches and particular methods of their computer implementation. It is shown that for the computer analysis of the psychological state of society in the context of the epidemic, it is necessary to adapt the methodology for designing neural network technologies, as well as systems for collecting and textual analysis of the content of electronic and Internet resources. An effective approach to creating such systems is embedding, which uses a dense vector representation of tokens in a multidimensional space, the dimension of which should be selected experimentally in the process of training and testing the developed artificial neural networks (ANN). For contextual neural network analysis, a multiclass-oriented ANN with regularization layers of the form "SpatialDropout1D" can be used. The neural network architecture can be built on fully connected layers with an activation function of the "ReLU" type. The scientific and applied significance of the results of neural network analysis based on Internet resources is the possibility of obtaining classified assessments and segmentation of target information about the psychological state of society during periods of epidemics. This information can be used to effectively counter information threats to society.

* Corresponding author: rafr@mail.ru

1 Introduction

Social tension in the Russian and global community, especially during periods of epidemics, requires in-depth socio-psychological research based on extensive relevant information. The new deadly disease COVID-19, caused by the SARS-CoV-2 virus, has spread to almost all countries and has been recognized as an epidemic. It has generated panic moods and information speculation, which require a comprehensive study for effective counteraction. The variety of publications about epidemics and pandemics, in particular, coronaviruses, necessitate the collection and analysis of huge amounts of information, which is constantly generated in exponentially increasing volumes and promptly posted mainly on the global Internet.

The article [1] states that "one of the main conditions for the innovative development of Russia is a radical change in the psychological state of our society". The methodology of modern data science increasingly uses social indicators based on aggregated quantitative assessments of the characteristics of society.

The term "social indicators" appeared in the United States in the early 1960s at the initiative of the American Academy of Arts and Science, which was commissioned by NASA. In the 1970s, the US Government began to regularly publish relevant data, and the journal Social Indicators Research was created. A similar approach has been adopted by international organizations such as the UN and the OECD. Then, in the 1980s, there was a slight decline in interest in social indicators, but in the 1990s it began to revive. Stepashin V. S. and other authors note that this happened as a result of the adoption of the sustainable development program by the international community. The various social indicators were replaced by composite indices, which include various components.

Various social indicators are actively used by international organizations such as the United Nations, the Statistical Office of the European Union (Eurostat), the OECD (Organization for Economic Cooperation and Development), the World Bank, and the European Commission. They are used by almost all European countries, as well as the United States, Canada, Japan, Australia, Latin America and South Africa. G. V. Osipov notes that "... the approach was supplemented by a subjective one that takes into account the psychological well-being of people, the concepts of quality of life and functional abilities (capabilities) appeared" [5]. The Institute of Psychology of the Russian Academy of Sciences has developed a Composite Index of the psychological State of society, and the dynamics of the psychological state of modern Russia identified on its basis, considered earlier and subjected to further monitoring.

Sociological research based on network methodology is conducted by such scientists as P. Ya. Aronson, S. Yu. Barsukova, E. R. Batkayeva, G. S. Baty-gin, A. A. Bashkarev, A. P. Vasilyeva, V. P. Vorobyov, L. Zhuzhuan, M. D. Kondratova, E. O. Leonova, O. V. Lylova, B. Wellman. The results of the study of virtual network structures are also widely presented (L. Adamik, G. V. Gradoselskaya, S. Dokuka, A. Semenov). In particular, Yu. V. Bondarenko and T. I. Barsukova identified quantitative and qualitative characteristics of social networks of low-income citizens, based both on the structural characteristics of support networks and on the analysis of the intensity of individual interactions [10]. In the dissertation research of M. A. Tronevskaya "Social identification of employees in social media" (Specialty 22.00.04 - Social structure, social institutions and processes, 2018), a network approach is presented analysis of the structure and features of communication of 29 virtual professional communities, including those operating on the «VK» and Facebook platforms. The collection, processing and analysis of the received information were carried out using the Igraph, Sna, and RSiena libraries of the R language for statistical calculations. An analysis of the functionality of software products is also presented: online panels on the service www.Anketa.ru, as well as the resources of MROC (Marketing Research Online

Communities), ESOMAR, RDS (Response Driven Sampling), ServyManky, which provide the collection of sociological information on the Internet. However, the known approaches are focused on statistical analysis of information without an in-depth semantic assessment of the studied texts.

Publications of Russian and foreign scientists-sociologists, psychologists, and philologists [7, 8, 9], who have created a number of methods for deep analysis of emotions and tonality of texts in the Internet media, including cognitive and interpretive decoding, are devoted to substantiating approaches and methods for studying the content of Internet content [8]. The significance of the development of algorithms and methods of monitoring and neural network analysis for the study of the socio-psychological state of society is determined by their ability to identify the deep internal characteristics of texts downloaded from numerous Internet resources using pre-trained neural networks.

The significance of the development of algorithms and methods of monitoring and neural network analysis for the study of the socio-psychological state of society is determined by their ability to identify the deep internal characteristics of texts extracted from numerous Internet resources using previously trained neural networks.

2 Materials and methods

The development of methods and computer tools for studying the psychological state of society during epidemics is based on data from Internet resources using neural network technologies implemented using computer systems. At the same time, the scientific tasks are to identify the specifics of the object of research, as well as the architecture of deep neural networks and integrate them with the means of automatic information search, focused on the socio-psychological state of society during crises and epidemics.

The basic methodology of the research is a system analysis and a set of specific methods for finding relevant information. Among them, the key ones are the formation of a system of indicators of the psychological state of society in the period of epidemics. Texts are pre-selected, and corpora of model and real published texts in natural language are formed using the methods of contextual analysis and synthesis. Using a probabilistic approach, "symbolic" models of natural language can be trained on a sufficiently large body of specialized texts. It is used to develop and configure means of automated computer downloading of information (parsing) from Internet resources that characterize the socio-psychological state of society during epidemics.

3 Results and discussion

Social tension in the Russian and international communities requires a comprehensive study based on extensive relevant information. A variety of publications about the Covid-Sars-19 coronavirus, primarily on the global Internet, concerning its sources and the degree of threat to humanity, ranging from long-standing knowledge to specialists, to modern conspiracy theories [2], necessitate the collection and analysis of huge amounts of information constantly generated in exponentially increasing volumes and operatively posted mainly.

3.1 The use of artificial neural networks in medical and sociological research

In the context of mass diseases, we note the observation of G. M. Zarakovsky [6] that from the late 1990s to the mid-2000s, the statistics of diseases in our country significantly deteriorated, in the etiology of which stress factors play a major role (diseases of the circulatory system and food organs), while the number of diseases with infectious and

parasitic diseases, on the contrary, decreased. The author explains this phenomenon in the light of two possibilities: 1) the divergence of adaptation to what is happening at the conscious and unconscious levels, 2) the psychophysiological costs of a more active lifestyle, in particular, multiple employment, etc., necessary for adaptation to new economic conditions.

The work [4] can be considered a systematic review of the well-known technologies for creating INS for processing text information, including the formation of case papers, pre-processing of source data, architecture and hyperparameters of artificial neural networks (ANN). It examines computer-based technologies for text information analysing, including language-adapting symbols and structures, new definitions, and contexts [7], using Python libraries such as Keras, ScikitLearn, NLTK, Gensim, spaCy, and NetworkX [6]. ANN researchers note the possibility of using neural network approaches for text processing in natural languages (NLP - Natural Language Processing) and artificial intelligence (AI) methods to identify the target content [9, 10].

The traditional approach to text processing is the analysis of the frequency of natural language words in the corpus of texts, called "frequency embedding", in which each word is associated with a certain number - the frequency of the word.

3.2 The construction of models of embedding terms

The traditional approach to text processing is the analysis of the frequency of natural language words in the text corpus, called "frequency embedding", in which each word is associated with a certain number - the frequency of the word.

More effective is the adjusted estimate of the frequency value - the inverse frequency of the words of the document or the inversion of the frequency with which a certain word occurs in the text body under study. This approach allows you to reduce the weight of the most frequently used words (prepositions, conjunctions, general concepts). The value of the inverse frequency indicator will be higher if a certain word is used with a high frequency in a particular text, but rarely in other documents.

Each word w_i in the training sample is discarded with a probability calculated by the formula (1). The value of the constant t in the dependence (1) is recommended to be equal to 10^{-5} .

$$P(w_i) = 1 - \sqrt{\frac{t}{f(w_i)}} \quad (1)$$

where $f(w_i)$ - is the frequency of the word w_i ; t - is an empirical constant.

Function (1) allows you to sample words whose frequency exceeds the value of t while maintaining the frequency ranking.

The use of adjusted word sets allows us to effectively automate semantic analysis, identifying the topics available in the text corpus, and classify texts by main topics.

To improve the efficiency of computer analysis, Tomas Mikolov proposed the locality hypothesis, according to which "words that occur in the same environments have similar meanings" [11]. To implement the locality hypothesis, word embeddings are constructed in a vector space, the dimension of which, regardless of the volume of the dictionary, can be on the order of $10^2 \dots 10^3$. In vector space, each word will correspond to a collection of several hundred numbers. Such embedding vectors can be added, multiplied by scalars, and angles and distances that have a certain meaning can be defined between them, as logical actions on certain words.

As a visual example, given in [8], the result (2) of the vector calculation

$$\text{vec}(\text{res}) = \text{vec}(\text{"Madrid"}) - \text{vec}(\text{"Spain"}) + \text{vec}(\text{"France"}) \quad (2)$$

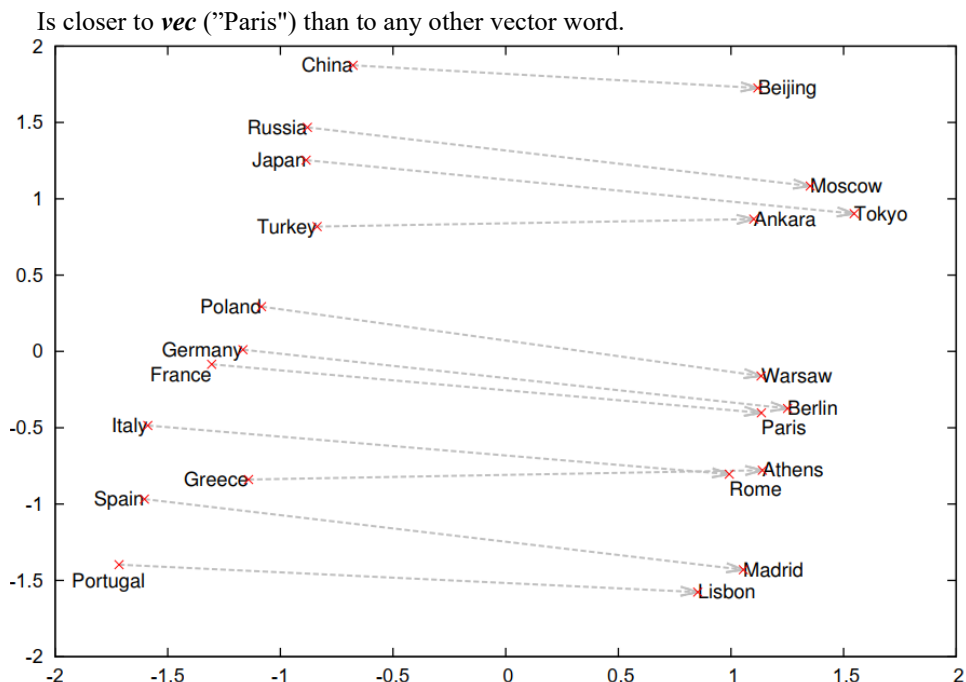


Fig. 1. Two-dimensional projection of multidimensional embeddings on the example of the capitals of certain countries [11].

The method of constructing embeddings, based on the probabilistic assessment of the joint use of a combination of words through artificial neural networks (ANN), trained on thematic text corpora, was called "word2vec".

The neural (associative) approach is based on the hypothesis that language units interacting with each other do not necessarily form a consistent context [2]. The neural network model is based on a structure of several components, including a vectorized representation of data, an input layer of neurons, hidden layers of various architectures, and an output layer with predicted values. The deep learning INS architecture is based on models such as Recurrent Neural Networks (RNN), Long ShortTerm Memory (LSTM), Recursive Neural Tensor Networks (RNTN), Convolutional Neural Networks (CNN or ConvNets), and generative-adversarial networks (Generative Adversarial Networks, GAN).

The architecture of the AN studied by the authors, focused on multi-class analysis on the example of 5 pre-formed categories, was based on a 16-dimensional model of the representation of the words "embedding". The subsequent regularization was implemented using a layer of the form "SpatialDropout1D". The neural network architecture is based on fully connected layers with the function of activating neurons of the "Relu" type. The ANN fragment in Python is presented below.

Specially prepared "text corpora" were used for ANN training. Constructing a corpus with source texts in a form suitable for creating an application, regardless of the method of data collection (by scrapping, extracting from RSS, or using some API), is a non-trivial task [4]. The Internet is not a medium for HTML files that are easy to process. It is a repository of information, where HTML files are often used as a means of visual representation.

Without being able to read various types of documents, including text, PDF, images, videos, emails, etc., researchers lose a significant part of the data [3].

```
modelEm = Sequential()  
modelEm.add(Embedding(maxWrdsCnt, 16, inp_length = xLn))  
modelEm.add(SpatialDropout1D(0.2))  
modelEm.add(Flatten())  
modelEm.add(BatchNormalization())  
modelEm.add(Dense(200, activation='relu'))  
modelEm.add(Dropout(0.2))  
modelEm.add(BatchNormalization())  
modelEm.add(Dense(5, activation='sigmoid'))
```

Fig. 2. Fragment of the source code of the program.

In addition, the language data coming from the source must be cleaned up and transformed into data structures suitable for analysis [2]. The method of web scraping is the collection of data by any means other than programs that use the API [3], most often carried out by a program that automatically requests the web server, receives data (HTML and other files that are placed on web pages), and then parses this data to extract the necessary information. To do this, you can use Web crawlers (web spiders), so called because they "crawl" on the Internet [3]. Their work is based on recursive traversal. They should extract the content of the page at the specified URL, examine that page for another URL, extract the page at the found URL, and so on.

The read data requires an in-depth semantic analysis based on symbols and their combinations, words (tokens) and their combinations (n-grams), sentences and whole paragraphs.

4 Conclusions

The analytical review made it possible to justify the following provisions.

1. For computer analysis of the psychological state of society in crisis conditions, including epidemics, it is necessary to adapt the methodology of designing and optimizing neural network technologies and systems for collecting and textual analysis in natural language of the content of electronic and Internet resources.

2. An effective approach to creating such systems is both frequency and vector embedding, which uses a vector representation of tokens in a multidimensional vector space, the dimension of which is several hundred or more and should be selected experimentally in the process of training and testing the developed ANNs.

3. For contextual neural network analysis, an ANN focused on multiclass analysis can be used, based on the "embedding" model with regularization layers of the "SpatialDropout1D" type. The neural network architecture can be built on fully connected layers with an activation function of the "ReLU" type.

4. The scientific significance and application of the results of neural network analysis based on Internet resources is the possibility of obtaining classified assessments and

segmentation of target information about the psychological state of society during epidemics.

References

1. Yurevich A.V., Yurevich M. A. Dynamics of the psychological state of Russian society: expert assessment. URL: <https://psyfactor.org/lib/social8.htm> (11.07.2020)
2. The coronavirus pandemic is a global hoax. URL: <https://fishki.net/anti/3261069-pandemija-koronavirusa-globalynyj-obman.html>
3. Ryan Mitchell. Scraping web sites with Python. Moscow: DMK Press, (2016)
4. Bengforth B., Bilbro R., Ojeda T. Applied analysis of text data in Python. Machine learning and building natural language processing applications. - St. Petersburg: Peter (2019)
5. Osipov G. V. Measurement of social reality. Moscow: ISPI RAS, (2011)
6. Zurakowski G. M. Quality of life of the population of Russia: psychological co-tavlaya. M: Meaning, (2009)
7. Developing Cognitive semiotic of data in computer-based communication (signs, concepts, discourse) / Olyanitch A. V., Khachmafova Z. R., S. R. Makerova, Akhidzhakova M. P. Ostrovskaya T. A. Communications in Computer and Information Science. (2019)
8. Surkova A.S., Chernobaev I.D. Comparison of neural network architectures in the task of automatic text classification. Modern informatization problems in the technological and telecommunication systems analysis and synthesis. MIP-2019'AS Proceedings of the XXIV-th International Open Science Conference. (2019)
9. Kim Y., Jernite Y., Sontag D., Rush A. Character-Aware Neural Language Models. arXiv Prepr. arXiv1508.06615. (2015)
10. Tarasov D. S. Deep Recurrent Neural Networks for Multiple Language Aspect based Sentiment Analysis of User Reviews Proceedings of the 21st International Conference on Computational Linguistics Dialog. (2015)
11. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. ICLR Workshop, (2013)
12. Rogachev, A. Fuzzy Set Modeling of Regional Food Security. Advances in Intelligent Systems and Computing 726 774-782 doi: 10.1007/978-3-319-90835-9_89 (2019).
13. Tokarev K.E.et al. The intelligent analysis system and remote sensing images segmentation engineering by using methods of advanced machine learning and neural network modeling. IOP Conference Series: Materials Science and Engineering. Krasnoyarsk Sci-ence and Technology City Hall of the Russian Union of Scientific and Engineering As-sociations. Krasnoyarsk, Russia. p. 12124 (2020).