# Neural network natural language processing tools for identifying personal priorities in the project performers selection in the field of smart agriculture



# Neural network natural language processing tools for identifying personal priorities in the project performers selection in the field of smart agriculture

**K A Polshchykov[1], A S Velikanova[1] and E V Igityan[1]**

[1]Belgorod State University, 85, Pobedy str., Belgorod, 308015, Russia

E-mail: polshchikov@bsu.edu.ru

**Abstract**. The project implementation effectiveness in creating digital smart agriculture systems depends on the correct selection of performers. Considering personal priorities makes it possible to increase the validity of decision-making regarding the employment of specific individuals for the implementation of IT projects in the agro-industrial sector. Personal priorities are internal, hidden characteristics that have an effect on the process of long-term joint work and interaction in various situations that arise in the team. The identification of the analyzed individual's personal priorities is proposed to be reduced to solving the classification problem based on the analysis of person's text Internet traces using neural network technologies of natural language processing. As a training sample, it is proposed to use a set of text document vectors and the corresponding marks of personal priority classes. In the process of identifying the personal priorities classes, it is required to create an appropriate text array based on parsing and processing of text messages published on the Internet by the analyzed person. Next, from the resulting text array, a text vector of the document must be formed, which then needs to be submitted to the input of the neural network. It is assumed that the mark of the analyzed individual's personal priority class will be displayed in the output layer of neurons.

## 1. Introduction

The smart agriculture concept is based on the development and application of innovative digital solutions to increase yields and reduce costs in food production [1, 2]. Successful project implementation to create effective digital management systems in the agro-industrial complex requires great efforts and hard work of each performer. When selecting the staff of project teams, it is necessary to take into account the personal priorities of a potential employee [3].

Personal priorities are the goals, values and meanings inherent in the inner world of individuals, determining their motivational orientation in life. These personal priorities are the determining factor of an individual's professional activity success and psychological well-being. Person's desire to adhere to moral standarts in behavior leads them to satisfaction with their work and life in general [4-6].

Taking into account personal priorities allows employers to increase the validity of decisions regarding the employment or appointment of specific persons to certain positions in companies and organizations [7]. Personal priorities are hidden from external observation, sometimes deliberately, but they have an effect on the process of long-term joint work and affects interaction in various

problematic situations that arise in the team. This makes it very difficult to get information about the personal priorities of strangers. To reduce the undesirable risk when hiring an employee, reference documents from reputable persons, in which the employer has high confidence, are traditionally used. However, many people do not have the opportunity to receive such recommendations, for example, due to little or no work experience. This makes it necessary to develop new methods for assessing an individual's personal priorities.

## 2. Materials and methods

In modern society, a characteristic feature of communication is the mass active use of Internet services. Users of IT gadgets generate a huge number of text and voice messages, tweets, microblogs, and comments on Internet resources every day. Digital traces and text content on public resources of Internet messengers, chats and social networks intellectual analysis is used to obtain the necessary information about the individual's interests, personal qualities and attitude to various phenomena in the social, cultural, political and economic public life areas.

The purpose of this study is to form a conceptual framework of application for the text data intellectual analysis technologies for Internet-users' personal priorities determination.

Based on the classification of life values, it is possible to distinguish the types of personal priorities that determine the professional and economic well-being of an individual. The first type is called moral and business priorities (MBP). It is characterized by the predominance of spiritual and moral, humanistic values, commitment to the observance of moral norms in behavior, collectivism, readiness to help and support other people, to develop themselves in order to ensure social well-being. The second type of personal priorities is selfish-prestigious (SPP). Individuals with this type of priorities prefer pragmatic values, individualism, their own prestige, personal achievements, high financial status, and egoistic isolation of their own person.

Identification of the analyzed individuals' personal priorities is reduced to solving the problem of classification, i.e., to classify their personal priorities as SPP or MBP. This problem is proposed to be solved on the basis of the person's text Internet traces analysis using neural network technologies of natural language processing (NLP) [8-13].

To implement this idea, first of all, it is necessary to select $K$ individuals whose classes of personal priorities are determined in advance on the basis of other people's expert assessments (for example, their managers, colleagues, partners, etc.). At the next stage, the Internet messages published by each selected person should be parsed and formed into $K$ composite text arrays.

Then, each received text array should be subjected to preparatory NLP processing, performing the procedures for segmenting the text into sentences, tokenizing them into individual words, excluding stop words and punctuation marks, and lemmatization. In the next step, for each document number $k$, using the statistical method to estimate the importance of words in the document (Term Frequency – Inverse Document Frequency , TF-IDF) [14], a vector $C_k$, the format of which is shown in Figure 1, should be created.

| $C_1$ | $C_2$ | $C_3$ | ... | $C_w$ | ... | $C_Z$ |
|---|---|---|---|---|---|---|

**Figure 1.** Format of text document vector.

The values of the text document vector elements $c_w$ are calculated based on the expression:

$$c_w = \frac{u_w}{\sum_{z=1}^{Z} u_z} \log \frac{K}{F_w}, \quad (1)$$

where $u_w$ is the amount of words number $w$ in the text document; $Z$ is the amount of words in the dictionary used; $u_z$ is the amount of words number $z$ in the text document; $F_w$ is the amount of text documents that contain the word number $w$.

To determine, based on the text document vector, which class (SPP or MBP) the personal priorities of the individual who published the corresponding set of Internet messages belong to, it is proposed to use the capabilities of a neural network. To solve NLP problems, Recurrent Neural Networks (RNN) [15], networks with long Short-Term Memory (LSTM) [16], and transformer networks (Generative Pre-trained Transformer, GPT-2, GPT-3) [17, 18] are most often used. This requires training of the neural network, as a result of which the neuron parameters are adjusted in such a way that classification errors are minimized.

As a training sample, it is proposed to use a set of $K$ text documents vectors and their corresponding marks of personal priority classes. The mark $l_k$ takes the value 0 if the individual's personal priorities belong to the SPP class. The mark $l_k$ equals 1 if the individual's personal priorities belong to the MBP class. An example of training sample fragments is grayed out in table 1.

**Table 1.** Format and example of training sample fragments.

| Number of the text document | Elements of the text document | | | | | | Marks of personal priorities classes |
|---|---|---|---|---|---|---|---|
| | $c_1$ | $c_2$ | … | $c_w$ | … | $c_Z$ | $l_k$ |
| 1 | 0 | 0.0024 | … | 0 | … | 0.0018 | 1 |
| 2 | 0 | 0 | … | 0.0004 | … | 0 | 0 |
| … | … | … | … | … | … | … | … |
| $k$ | 0 | 0 | … | 0,0002 | … | 0.0012 | 1 |
| … | … | … | … | … | … | … | … |
| $K$ | 0.0001 | 0 | … | 0 | … | 0 | 0 |

Once configured, the neural network can be used to determine the class of personal priorities of the analyzed individual. For this purpose, using parsing of open network resources, the Internet messages published by this person into a text array should be collected, then its preparatory NLP processing should be performed, and a vector of the corresponding text document with elements whose values are calculated using the formula (1) should be formed. Then, the generated vector will need to be submitted to the previously trained neural network input (Figure 2).

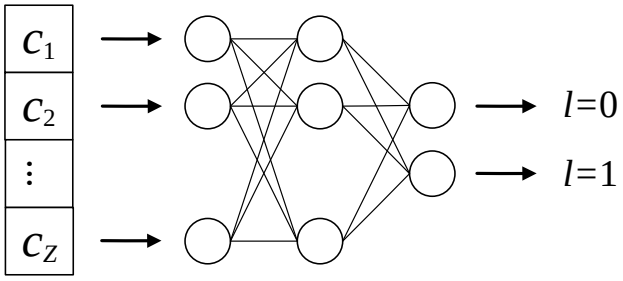


**Figure 2.** Input and output values of the neural network.

As a result, at the neural network output, a mark of the individual's personal priorities class will be obtained. The mark $l$ takes the value 0 if the individual's personal priorities belong to the SPP class. The mark $l$ equals 1 if the individual's personal priorities belong to the MBP class.

## 3. Conclusion

The results presented in the article are focused on the innovative projects performer selection for the smart agriculture digital systems development. Such projects are complex, they are very important for providing the population with food, therefore, it is necessary to include qualified specialists with moral and business personal priorities in their performers.

The paper proposes a conceptual framework for determining personal priorities using intelligent analysis of text messages published by Internet users. From the point of view of the individual's performance of tasks in the working collective effectiveness, the whole variety of personal priorities can be reduced to two main classes: SPP and MBP. The process of identifying the classes involves the following steps:

1) on the basis of parsing and NLP-processing of published on the Internet by the analyzed person text messages, the corresponding text array is compiled;

2) from the resulting text array, a text vector of the document, the elements of which are calculated using the TF-IDF method, is formed;

3) the text vector of the document is fed to the neural network input, due to the setting of which the mark of the analyzed individual's personal priorities class is displayed in the output layer of neurons.

The subject of further research is the justification of the neural network type and parameters choice for solving the problem of personal priorities choice classification. In addition, it will be necessary to develop and use software tools to implement the conceptual foundations proposed in the work and to obtain experimental results that allow to assess the possibility of identifying personal priorities based on the intellectual analysis of Internet users' text messages.

## 4. Acknowledgments

The reported study was funded by RFBR, project number 20-37-90083.

## References

[1] Sekaran K, Meqdad M N, Kumar P, Rajan S and Kadry S 2020 Smart agriculture management system using internet of things *Telkomnika* **18** 1275-1284

[2] Agrimonti C, Lauro M and Visioli G 2021 Smart agriculture for food quality: facing climate change in the 21st century *Critical Reviews in Food Science and Nutrition* **61** 971-981

[3] Velikanova A S, Polshchykov K A, Likhosherstov R V and Polshchykova A K 2021 The use of virtual reality and fuzzy neural network tools to identify the focus on achieving project results *J. Phys.: Conf. Ser.* **2060** 012017

[4] Gerasimova A and Oboznov A 2015 Normative orientation of the specialists of official activity and professionals of socionomic profile *Proceedings of 2nd Global Conference on Psychology Researches (GCPR-2014). Procedia Social and Behavioral Sciences* **190** 39-42

[5] Gerasimova A S 2013 Value-normative method of evaluation of educational motivation of students *Eksperimentalnaya psikhologiya* **6** 96-104

[6] Ovsyanikova E A, Mandibura N A, Gerasimova A S, Hudayeva M Y, Tkachenko N S and Godovnikova L V 2018 Modern status of research on the problem of psychological well-being of the person in the domestic and world psychological science *Revista Publicando* **5** 349-358

[7] Gerasimova A and Oboznov A 2014 Development of an altruistic orientation of the personality of students and expert of helping professions *International Multidisciplinary Scientific Conferences on Social Sciences and Arts. Psychology and Psychiatry, Sociology and Healthcare, Education* **1** 115-120

[8] Young T, Hazarika D, Poria S and Cambria E 2018 Recent Trends in Deep Learning Based Natural Language Processing *IEEE Computational Intelligence Magazine* **13** 55-75

[9] Stewart R and Velupillai S 2021 Applied natural language processing in mental health big data

*Neuropsychopharmacology* **46** 252-253
[10] Polshchykov K, Lazarev S, Polshchykova O and Igityan E 2019 The Algorithm for Decision-Making Supporting on the Selection of Processing Means for Big Arrays of Natural Language Data *Lobachevskii Journal of Mathematics* **40** 1831-1836
[11] Öztürk H, Özgür A, Schwaller P, Laino T and Ozkirimli E 2020 Exploring chemical space using natural language processing methodologies for drug discovery *Drug Disc Today* **25** 689-705
[12] Polshchykov K A, Lazarev S A, Konstantinov I S, Polshchykova O N, Svoikina L F, Igityan E V and Balakshin M S 2020 Assessing the Efficiency of Robot Communication *Russian Engineering Research* **40** 936-938
[13] Arts S, Hou J and Gomez J C 2021 Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures *Research Policy* **50** 104144
[14] Kim D, Seo D, Cho S and Kang P 2019 Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec *Inform Sciences* **477** 15-29
[15] Pitsilis G K, Ramampiaro H and Langseth H 2018 Effective hate-speech detection in Twitter data using recurrent neural networks *Applied Intelligence* **48** 4730–4742
[16] Qaisar S M 2020 Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory *2nd International Conference on Computer and Information Sciences (ICCIS)* pp 1-4
[17] Lee H H, Shu K, Achananuparp P, Prasetyo P K, Liu Y, Lim E-P and Varshney L R 2020 Generative Pre-training Based Cooking Recipe Generation and Evaluation System *WWW '20: Companion Proceedings of the Web Conference* pp 181-184
[18] Dehouche N 2021 Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3) *Ethics Sci Environ Polit* **21** 17-23