Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 190 (2021) 500–507

Procedia Computer Science

www.elsevier.com/locate/procedia

2020 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: Eleventh Annual Meeting of the BICA Society

A method for reducing the impact of information risks on a megaproject life cycle based on a semantic information field

Igor A. Kuznetsov[a,*], Matvey V. Koptelov[a], Dmitriy A. Kovtun[a], Anna I. Guseva[a]

[a]*National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 115409, Kashirskoe shosse, 31, Moscow, Russia*

Abstract

This paper consider a comprehensive method for devising loyalty programs based on the stages of a life cycle of an international megaproject. The method is based on the analysis of information risks and their management. The method is focused on aggregation and processing of data from various sources of textual information, which demonstrates the attitudes of key categories of individuals regarding the implementation of a megaproject at its numerous life cycle stages. The semantic information field is formed using hardware and software based on Neural Network Technologies. Authors examine the most popular neural network architectures that are used in the sentiment analysis. The paper describes a comparative analysis of classification accuracy of neural network architectures based on volume of texts and neural network profitability to sentiment analysis of large and small volumes of text. The method is aimed at managing and influencing information flow that accompanies the implementation of a megaproject stages. The application of semantic information field makes it possible to account for the informational risks of a megaproject and to prepare an effective set of measures to counteract these risks in a timely manner. This work was supported by RFBR grant № 20-010-00708\20.

Peer-review under responsibility of the scientific committee of the 2020 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: Eleventh Annual Meeting of the BICA Society

Keywords: decision support system; information risks; neural network; sentiment analysis; natural language processing.

* Corresponding author. Tel.: +7-495-788-56-99; fax: +7-495-788-56-99.
E-mail address: iakuznetsov@mephi.ru

1877-0509 © 2021 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 2020 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence: Eleventh Annual Meeting of the BICA Society
10.1016/j.procs.2021.06.108

1. Introduction

The decision to launch an international megaproject involves consideration of multiple risks, such as long implementation period, large financial investment and its significantly delayed return, international image and reputation of a Country and a Corporation involved, and the use of foreign territories. These risks accompany the project in course of its entire life cycle – from a planning stage to maintenance and decommissioning of a construction site. A special place in an implementation of megaprojects is given to a decision support system that has a wide range of applications and is capable of analyzing an entire array of risks. Generally, major types of risks associated with international projects are country-related, legal, technological, competitive, financial, market and information related and force majeure [1]. The impact of these risks on a project can be significantly reduced by developing and implementing a loyalty program for international megaprojects [2].

In a case when the main consumer of a megaproject's operations is one or more foreign countries or their public and private companies, the loyalty program should be follow the B2B structure [3]. Many existing loyalty programs are designed to operate for no more than 2-3 years. They do not correspond to scale and duration necessary for proper functioning of megaprojects; and, they do not possess the required complexity and systematization [4].

Within the framework of an NPP construction project, the risks can be divided into three groups: project risk, systematic risk and the risk of external influence [5]. The risk of external influence R_{ei} consists of two factors: force majeure R_{fm} and the information risk R_{inf}. Each of the risks is assigned a risk index – a measure of risk, which is a quantitative risk assessment, obtained using point estimates based on ordinal scales [6, 7]. The value of each risk index ranges from 0 to 1, depending on the developmental stages of a project. Further, using the same method of expert assessments, the relative weights of these indices W are determined. The sum of these indices should equal to one. The formula for calculating the risk of external influence index is as follows (see Formula 1).

$$R_{bi} = R_{fm} * W_{fm} + R_{\inf} * W_{\inf}, \quad W_{fm} + W_{\inf} = 1 \tag{1}$$

Today, information space that surrounds a project implementation has significant impact on economic indicators of that project, i.e. the importance of information risks increases [8]. The term "information risk" should be considered in the context of informational flow that is generated not only inside but also outside the system. Sometimes, the flow of information generated outside a project can have a significant impact on its implementation.

Information risks of an NPP construction project are presented in Table 1 (listed by construction stages).

Table 1. Information risks of an NPP construction.

Stages of a LC	Risk	Probability
1. Concept	Adverse socio-political situation	0,2
2. Design	Public impact on project implementation	0,4
3. Construction and commissioning	Decline in Corporation image as a vendor of nuclear technologies.	0,1
	Public impact on project implementation	0,4
4. Operation and maintenance	Decline in Corporation image as a vendor of nuclear technologies.	0,1
	Public impact on project implementation	0,2
5. Handling of SNF and RW	–	–
6. Decommission	Decline in Corporation image as a vendor of nuclear technologies.	0,1
	Public impact on project implementation	0,1
7. Severe accident management	Decline in Corporation image as a vendor of nuclear technologies.	0,9

In 2011, the construction of Indian Kudankulam NPP was delayed by 1.5 years and incurred serious economic losses as a result of mass protests that were actively supported by information technology [7]. Thus, it is important to take into account the impact of information technologies when making economic assessments of a project.

The proposed methods use the following key categories of individuals as the sources of information: statements and publications of senior officials and local authorities, reports and press releases of leading international organizations in the field of nuclear energy, electronic publications in media and on corporate information portals, posts and messages in social networks.

In order to form the semantic information field of a megaproject, we applied the principles of designing Digital Twin, segmentation of information groups and digital content providers, and intellectual processing of semi-structured text data [9].

2. Managing the Semantic Information Field of a Megaproject

Information risk management method is based on the concept of measuring risk index on a regular basis. As the risk assessments are performed and results are obtained, their analysis yields a set of measures that can be used to manage relevant events and activities in order to reduce negative impact of risks at each stage of an NPP construction.

In order to build a solution management model, the following steps must be implemented.

1 determine the input and output parameters of a system, establish links between them;
2 develop an approach to determine the level of influence that an information agenda exerts on the risks of a project lifecycle;
3 develop a risk management framework;
4 determine a risk identification algorithm based on the sentiment analysis of an information message;
5 prepare a set of events and activities for managing the risks.

2.1. Defining Input and Output Parameters

Input parameters can be described as a vector of initial information about an NPP construction stage (see Formula 2). These parameters are characterized by sets of different indicators such as terms of cooperation *S*, concluded agreements *D*, and international cooperation *M*.

$$f_0 = f(S_0, D_0, M_0) \tag{2}$$

As the sets of events and activities that influence key categories of individuals are performed, the changes can be described in a function of transition from primary (initial) information to a current state of informational flow (Formula 3).

$$\alpha = \alpha(S_1, D_1, M_1) \tag{3}$$

The result is the output parameter *R*, which represents a set of risk indicators for the completion of a particular stage of NPP construction.

2.2. Development of the Risk Index

Once the input parameters (the initial information about a certain stage of NPP construction) are received, the system generates risk indices based on processing and analysis of information about key categories of individuals that influence a successful completion of that stage of construction $I = \{i_1, i_2, \ldots, i_N\}$. Text data sentiment analysis is implemented using Deep Neural Network algorithms, which allow classifying the sentiment as positive, negative, or neutral. Within the framework of examined method, information messages can be classified according to a particular scale and assigned an appropriate value of a risk parameter in accordance with Table 2.

Table 2. Scale of correspondence between the message sentiment and the risk parameter.

Sentiment	Probability of positive words	Value of the risk parameter *Pi*
Positive	0.6 – 1	0
Neutral	0.4 – 0.6	0,3
Negative	0 – 0.4	0,9

As a result of sentiment analysis of N messages, a project information risk index is calculated using Formula 4.

$$R_{\text{inf}} = \sum_{i=1}^{N} \frac{R_i}{N},\ N > 1 \tag{4}$$

where R_{inf} – project information risk index, R_i – parameter value for the i-th information message, N – number of messages.

2.3. Developing a Risk Management Framework

The objects of management are the risks that arise at a particular stage of an NPP construction. The management mechanism is a system that develops a set of measures to combat those risks; and, the feedback is a set of results on how selected events and activities affect key categories of individuals. Before selected events and activities are carried out, there is an initial set of data on risk indices $R_{inf} = \{r_1, r_2, \ldots, r_j\}$, on the basis of which, a system generates necessary set of events and activities [10] according to Formula 5.

$$A = (a_1, a_2, \ldots, a_f) \tag{5}$$

Events and activities generation system forms a vector of management actions according to Formula 6.

$$L = (f_0, \alpha, A) \tag{6}$$

where f_0 – available information on a particular stage of construction; α – a corresponding transition function of specified parameters during the implementation of selected events and activities, calculated using the Formula 3; A – a set of events and activities.

Based on the results from the implementation of the set of events and activities used to influence key categories of individuals, the final vector is formed according to Formula 7.

$$E = (R, L, Z) \tag{7}$$

where Z – vector of unpleasant actions that are directed at the management object and characterize the influence of an uncertain information environment; R – output parameter that characterizes the set of generated risk indicators for the completion of a particular stage of an NPP construction; L – events and activities aimed at countering the risks.

2.4. Sentiment Analysis Based Algorithms for the Identification of Risks

Sentiment analysis of text messages of previously established key categories of individuals is one of the most important tasks in monitoring the information agenda. In order to successfully complete this task on a par with traditional machine learning methods such as the naive Bayesian classifier (NB) and the support vector machine (SVM), the following deep learning methods used:

- Convolutional Neural Networks (CNN);
- Recurrent Neural Networks (RNN):
  - Long short-term memory (LSTM);
  - Gated recurrent units (GRU);
- Deep Belief Networks (DBN);
- Combined neural networks.

One of the most common data sets used to sentiment analysis is the posts on social network Twitter, post comments, and movie reviews. To analyze such data sets in previously published works [11, 12], the authors use LSTM. The study in publication [12] also compares the accuracy of LSTM and SVM; and, as a result, deep learning method works more accurately. CNN is used on similar data sets in publications [13, 14]. Works [15, 16] are devoted to the sentiment analysis using DBN. In [16], DBN is compared to two machine learning methods for analyzing blogs and forums. The authors proved better accuracy of DBN in comparison to SVM and NB. In [17], two deep learning models and two machine learning methods are compared. As a result, it can be concluded that machine learning methods work better than neural network models. However, CNN doesn't give much way to LSTM when it comes to accuracy. The authors concluded that in order to improve the performance of neural networks, we need to change the parameters and collect more data for training. In [18], the authors conducted an independent analysis of several methods and proposed a combination of these methods to reduce the shortcomings of each method. Comparative results and methods are presented in Table 3.

Table 3. Quality of models used for sentiment analysis

Article #	Model	Data sets	Model quality, %
11	LSTM	Twitter	80,0
12	LSTM	Movie reviews	88,4
	SVM		82,9
13	CNN	Movie reviews	74,5
		Twitter	68,0
14	DBN	Movie reviews	71,3
		Book reviews	64,3
		Electronics reviews	73,6
		Kitchen appliances reviews	72,6
15	DBN	Internet site comments	80,0
	NB		75,2
	SVM		76,1
16	NB	Internet site comments	73,5
	SVM		72,4
	LSTM		61,7
	CNN		70,6
17	CNN	Internet site comments	79,1
	RNN		77,3
	SVM		72,6
	NB		72,1
18	LSTM	Twitter	85,7
	CNN		83,1
	NB		71,0
	SVM		79,0

2.5. Developing the Sets of Events and Activities

Based on already calculated risk indices, a set of events and activities is developed within a framework of risk management. The implementation of this set directly impacts the risks we aspire to manage. The set is based on expert data, and we can use the method of analyzing the hierarchy by T. Saaty to determine the priority of implementing various loyalty program activities and the degree of their impact on external influences [10].

Using regularly updated data of risk indices, we can assess the impact that sets of events and activities have on key categories of individuals in real time [19]. If necessary, sets can be quickly adjusted in case a constructed model does not have the expected effect and the level of risk remains extremely high.

3. Results

On the basis of management methods of a semantic information field of a megaproject, a general risk management framework is prepared (see Fig. 1).

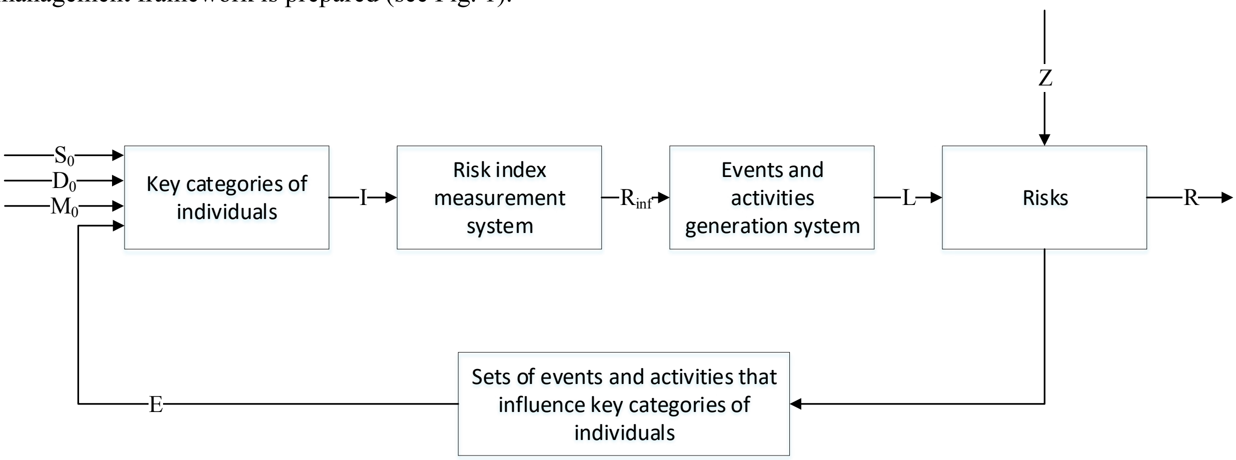

Fig. 1. Information risk management framework for an NPP construction.

Thus, the model of risk management at particular stages of an NPP life cycle can be presented in the form of Formula 8.

$$F = \{f_0, \alpha, A, R, L, Z\} \tag{8}$$

In order to identify risk based on the sentiment analysis, a Convolutional neural network architecture was used. For implementation CNN-based algorithm TensorFlow software library was used. The results of comparative analysis are presented in Table 3. To obtain these results, 8000 recordings of marked-up news feeds data sets were used. The neural network consisted of 4 hidden layers with Rectified Linear Unit (ReLU) as activation function, where the learning process included 10 epochs that classified the objects according to one of the three classes as output layer: positive, neutral, and negative. The results are shown as follows (see Fig. 2).

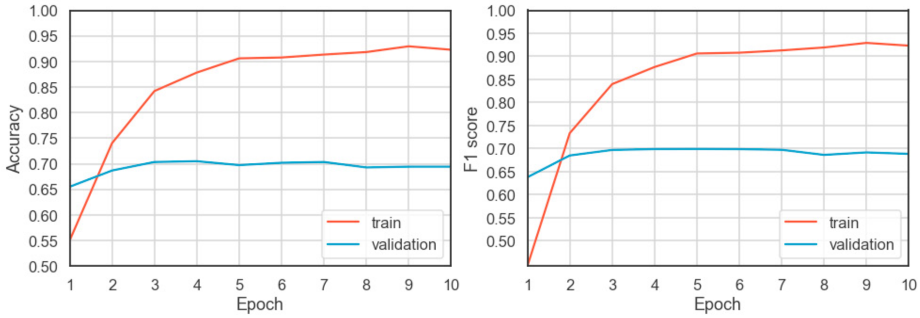

Fig. 2. Quality of neural network in terms of accuracy and F1 score.

According to obtained results, the quality of the algorithm is 70% accurate in classifying text data into three classes, which is comparable to other studies. The proposed model can be used as a part of general method to identify semantic information field management framework with uniform software and hardware solutions.

4. Discussion and Future Research

Method considered in this paper is focused on developing a set of events and activities and assessing their impact on semantic information field of a megaproject. The method is primarily intended to ease the use of informational flow in managing the loyalty of key categories of individuals.

Based on the sentiment analysis of the semantic field, we can construct a map that shows the dynamics of mentioning an analyzed subject. An example of such map is shown below (see Fig. 3), where blue color indicates an increase in level of loyalty, and the red color indicates its decline.

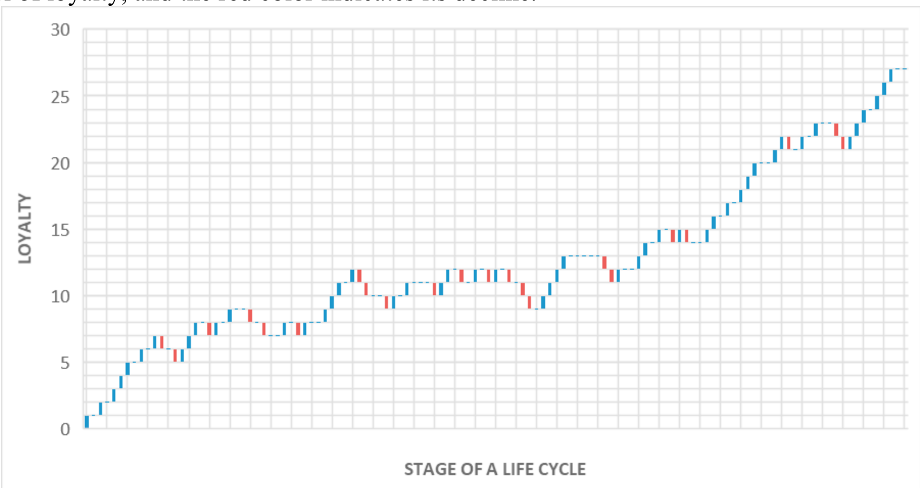

Fig. 3. Level of loyalty by the stages of a life cycle.

When there is a downward trend in the level of loyalty, a set of measures can be formed to smooth or compensate for the resulting effect. The effectiveness of implemented measures can be marked and tracked on the map.

The use of modern technologies in the field of data analysis, including machine learning methods and methods of intellectual processing of semi-structured text data, helps to improve the quality of decision-making in an implementation of international megaprojects. It may also allow for a timely impact on the sources of risks (key categories of individuals) at a particular stage of a project.

5. Summary

Preliminary studies have demonstrated the prospects of using the proposed method; and, using a deep learning model for the sentiment analysis shows high accuracy. Next steps in the development of the foregoing method are to improve the neural network model (to enable it work with Russian-language text data) and to test this model on data sets of other languages. Finally, in order to collect and store data from key categories of individuals in real or close to real time, specialized software and hardware solutions must be implemented.

6. Acknowledgments

This work was supported by RFBR grant № 20-010-00708\20.

References

[1] A. I. Guseva, M.V. Koptelov. (2018) "An Risk assessment of prospective investment projects for the construction of nuclear power plants abroad". International Journal of Engineering and Technology (UAE), 7 (2.23), 251-254.

[2] Nick J.F.Bombaij, Marnik G.Dekimpe. (2020) "When do loyalty programs work? The moderating role of design, retailer-strategy, and country characteristics". *International Journal of Research in Marketing*, 37 (1), 175-195.

[3] Vijay Viswanathan, F. JavierSese, Manfred Krafft. (2017) "Social Influence in the Adoption of a B2B Loyalty Program: The Role of Elite Status Members". *International Journal of Research in Marketing*, 34 (4) 901-918 (2017).

[4] A. N. Shtanko, S. D. Kulik. (2019) "Scientific personnel training in convolutional neural networks for the implementation of research projects of the MegaScience class". IOP Conference Series: Journal of Physics: Conference Series, 1406 (2019), 012014.

[5] Thijs Van de Graaf, Benjamin K. Sovacool. (2014) "Thinking big: Politics, progress, and security in the management of Asian and European energy megaprojects". *Energy Policy*, 74, 16-27.

[6] Mohamed Nabawy, Laila M. Khodeir. (2020) "A systematic review of quantitative risk analysis in construction of mega projects". *Ain Shams Engineering Journal*, 8.

[7] A. Guseva, M. Koptelov, D. Kovtun. (2018) "The Use of Automated Systems for Economic Analysis of NPP Construction Projects in Light of Risk Impact Assessment". *In: 2018 International Russian Automation Conference (RusAutoCon)*, pp. 1-5, Sochi).

[8] S. Kulik. (2016) "Factographic information retrieval for communication in multicultural society". Procedia - Social and Behavioral Sciences (International Conference on Communication in Multicultural Society, CMSC 2015, 6-8 December 2015, Moscow, Russian Federation), Vol. 236 (14 December 2016), 2016, pp. 29-33.

[9] A. I. Guseva, I.A. Kuznetsov. (2017) "The Use of Entropy Measure for Higher Quality Machine Learning Algorithms in Text Data Processing". *In: Proceedings 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pp. 47-52, Prague.

[10] Matrosova E.V., Tikhomirova A.N. (2016) "Peculiarities of expert estimation. Comparison methods". *In: Alexei V. Samsonovich, Valentin V. Klimov 7th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2016*, vol. 88, pp. 163-168. New York City, NY, USA.

[11] Merin Thomas, C.A. Latha. (2019) "Sentimental analysis using recurrent neural network". *International Journal of Engineering & Technology*, 7 (2.27), 88-92.

[12] Jyostna Devi Bodapati, N. Veeranjaneyulu, Shareef Shaik. (2019) "Sentiment Analysis from Movie Reviews Using LSTMs". *Ingenierie des Systemes d'Information*, 24 (1), 125-129.

[13] Shiyang Liaoa, Junbo Wangb, Ruiyun Yua, Koichi Satob, Zixue Cheng. (2017) "CNN for situations understanding based on sentiment analysis of twitter data". *In: Procedia Computer Science*, volume 111, pp. 376-381.

[14] Patrawut Ruangkanokmas, Tiranee Achalakul, and Khajonpong Akkarajitsakul. (2016) "Deep Belief Networks with Feature Selection for Sentiment Classification". *In: 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS).*

[15] Ahmed Al-Saffar, Suryanti Awang, Hai Tao, Nazlia Omar, Wafaa Al-Saiagh, Mohammed Al-bared. (2018) "Malay sentiment analysis based on combined classification approaches and Senti-lexicon algorithm". *In: PLoS ONE*, 13 (4).

[16] Jurgita Kapočiūtė-Dzikienė, Robertas Damaševičius, Marcin Woźniak. (2019) "Sentiment Analysis of Lithuanian Texts Using Traditional and Deep Learning Approaches". *In: the 24th International Conference on Information and Software Technologies (ICIST 2018)*.

[17] F. Yang, C. Du, L. Huang. (2019) "Ensemble Sentiment Analysis Method based on R-CNN and C-RNN with Fusion Gate". *In: International Journal of Computers Communications & Control*, vol. 14 (2), pp. 272-285.

[18] Shakeel Ahmad, Muhammad Zubair Asghar, Fahad M. Alotaibi, Irfanullah Awan. (2019) "Detection and classification of social media-based extremist affiliations using sentiment analysis techniques". In: Detection and classification of social media-based extremist affiliations using sentiment analysis techniques.

[19] Guseva A.I., Kuznetsov I.A., Bochkaryov P. V., Filippov S.A., Kireev V.S. (2020) "Development of Text Data Processing Pipeline for Scientific Systems". Advances in Intelligent Systems and Computing, 2020, Volume 948. pp. 124-136.