Intelligent tools for analyzing NON-factors of the project environment

O V Bulygina¹, A A Emelyanov¹ and O A Ivanova²

¹ Department of Information Technology in Economics and Management, Branch of the National Research University “MPEI” in Smolensk, 1 Energeticheskiy proyezd, Smolensk, 214013, Russia
² Department of Humanities, Pushchino State Institute of Natural Sciences, 3 Prospekt Nauki, Pushchino, 142290, Russia

E-mail: baguzova_ov@mail.ru

Abstract. The implementation of complex innovative projects to create high-tech products is usually associated with a high level of uncertainty, which is due to the lack of "quality" (accurate, complete, reliable, consistent, etc.) data on their internal and external environment. In this regard, special attention must be paid to the stage of project planning, which takes into account the results of identification and analysis of risk situations that may arise during the project implementation. In conditions of information uncertainty, it is advisable to use data mining methods to solve these problems. So, to identify the factors of the internal and external environment, which can negatively affect the project, it is proposed to use a neural fuzzy classifier. The analysis of the identified factors must be carried out taking into account the possibility of the simultaneous occurrence of several risks, i.e. the appearance of a systemic effect, which can be assessed using the Hartley emergence coefficient. In turn, the obtained values of this indicator determine the choice of analysis tools: in the case of a low value, it is proposed to use fuzzy logical inference according to the Mamdani algorithm, otherwise - fuzzy pyramidal networks.

1. Introduction

An important role in the process of managing complex projects to create high-tech products is played by information support for decision-making, which should be based on complete, reliable, accurate, consistent data. At the same time, the development of such projects is usually carried out in conditions of uncertainty, which is caused by a large number of participants, complex production and technological relations, a long implementation period [1].

In this regard, special importance in project planning should be given to identifying factors of the internal and external environment that may negatively affect its implementation [2]. However, for such projects, it is quite difficult to form a sufficient amount of "quality" information, which will be the basis for making timely and well-grounded management decisions.

To solve this problem, it was proposed to use an approach based on modeling NON-factors that do not possess one of the properties of classical knowledge models.

For the first time, the term "NON-factors" was introduced by A.S. Narinyani in the 1980s to describe "partial knowledge" as an integral element of the real system of knowledge. In this way, he denoted a set of factors determined in natural language and reflecting a negative assessment of the quality of knowledge about a process or object.
Further development of this concept is associated with the use of artificial intelligence methods for modeling real socio-economic and technical systems and supporting managerial decision-making.

To date, several dozen articles have been published on this topic. However, the concept of NON-factors has received the greatest development in the domestic literature. Table 1 shows the approaches to the definition and classification of NON-factors, which are proposed by Russian authors [3].

**Table 1.** The approaches to the definition and classification of NON-factors.

| Author          | Classes of NON-factors                                                                 | Comments                                                                 |
|-----------------|---------------------------------------------------------------------------------------|--------------------------------------------------------------------------|
| Narinyani A.S.  | basic (underdetermination, inaccuracy), systemic (ambiguity, fuzziness), classifiers (incorrectness), meta NON-factors (invisibility) | It is rather difficult to work with such a classification due to the fact that some NON-factors are a combination of others. |
| Borisov A.N.    | unknown, unreliability, ambiguity                                                     | Unknown is complete "ignorance" and does not fall into the Knowledge System.                  |
| Rybina G.V.     | uncertainty, underdetermination, inaccuracy, fuzziness                                 | In practice, it is difficult to separate uncertain and undetermined factors.                  |
| Tarasov V.B.    | informational (incompleteness, inaccuracy, uncertainty, fuzziness, inconsistency)    | Only informational NON-factors can be used for economic and mathematical modeling.             |
|                 | complex system development (irreversibility, instability, nonlinearity, nonequilibrium, openness) |                                                                                           |

The mentioned authors propose to carry out the NON-factor classification based on such characteristics as "defects" of both the knowledge itself and the Knowledge System.

There is another approach to the NON-factor definition that is based on the study of the subject area and the set scientific task. For example, Valkman Yu.R. proposes to carry out the classification in the space "NON-factor – object of research – modeling method". As a result, he distinguishes universal (common to all subject areas) and special (unique to a particular area) NON-factors.

In our opinion, the use of this approach makes it possible to more reasonably solve the problem of choosing tools for modeling NON-factors.

For example, Malcolm Beynon has explored the issues modeling inaccuracy and uncertainty of expert judgments in the analysis of project risks. So, he proposed a new way to assess the project viability, which consists of hybrid modeling of the project risk sources based on the combination of Saaty hierarchy analysis method and the Dempster-Shafer theory of belief and plausibility [4].

The works of Dubois D. and Prade A., dedicated to such NON-factors as inaccuracy and uncertainty, deserve special attention. So, they proposed to simulate these NON-factors based on quasi-measures of possibility and necessity using the theory of fuzzy sets.

The above confirms the relevance of the scientific task of developing the tools to identify and analyze NON-factors that must be taken into account in the implementation of complex projects to create high-tech products.

The purpose is to develop a new approach to the analysis of project risks based on modeling various factors of information uncertainty that arise in the internal and external environment, which can lead to a decrease in the effectiveness of project implementation.

The main contribution is the complex use of data mining for identification (neural fuzzy classifier) and analysis (fuzzy logic or fuzzy pyramidal networks, the choice of which is determined by the emergence coefficient) of NON-factors factors of different nature that pose a threat to the project implementation.
2. Systemic effect from a set of NON-factors
The task of identifying NON-factors is to build classifiers of factors of the internal and external environment of the innovative project to create high-tech products, which can lead to a significant decrease in the efficiency of its implementation.

As a mathematical apparatus for identifying NON-factors, it is proposed to use a neural fuzzy classifier, wherein fuzzy inference algorithms are implemented as a neural network with heterogeneous layers of neurons. This apparatus allows categorizing NON-factors based on the analysis of available information (its quantity and quality) and expert opinions.

In the process of analyzing NON-factors, it is necessary to take into account the fact that they can lead to negative project consequences, both individually and in combination. At the same time, it should be noted that situations of influence of one NON-factor are possible only for simple projects, usually aimed at the production of analog products. When implementing complex projects of an innovative nature, such situations occur extremely rarely.

The implementation of projects to create high-tech products is usually influenced by a large number of NON-factors of different nature, which, in turn, have complex interrelationships. With the simultaneous influence of several NON-factors to one degree or another (depending on their number), the emergence property may appear.

In system analysis, the emergence (or systemic effect) means the appearance of new properties in a complex system due to the interaction of its constituent elements (i.e., the properties of the system do not coincide with the properties of its subsystems).

As a result, an important role in the analysis of complex socio-economic system is played by the amount of new information about this system obtained as a result of combining its subsystems.

For the investigated tasks of project management, various sources of risk situations will be considered as NON-factors. Consequently, the risk of the innovative project will be the result of the appearance of a systemic effect from a certain set of NON-factors.

The maximum possible systemic effect from the influence of a certain set of NON-factors can be estimated using the principles of functioning of systems of the "boolean" type.

In general, the boolean is the set of all subsets of some set \( A \) (usually denoted as \( P(A) \)), for which \( \emptyset \in P(A) \) and \( A \in P(A) \). Thus, it represents the most powerful system with a hierarchical structure.

3. Application of the Hartley emergence coefficient to select tools for the NON-factor analysis
In the considered research task, classifiers of NON-factors will be used as complex systems. Combining such systems can lead to a situation where the power of the resulting complex system will exceed the sum of the power of the systems that make up this combination due to the occurrence of a systemic effect. This situation is possible only for a "boolean"-type system when "virtual" constituent elements appear "out of nowhere".

The possibility of the existence of such systems is shown in Hartley information theory, which states that the systemic effect consists of additional information arising in combining elements.

The Hartley emergence coefficient can be used to quantify the value of the systemic effect. So, for the case with one classifier, the emergence coefficient has the form:

\[
\phi = \frac{\log_2 \sum_{m=1}^{M} C_W^m}{\log_2 W},
\]

where \( \phi \) – the Hartley emergence coefficient; \( W \) – the number of basic indivisible elements in a complex system; \( m \) – the complexity of the constituent element of the system (i.e. the number of basic elements in the constituent element); \( M \) – maximum complexity of the subsystem (i.e. the maximum number of basic indivisible elements in the subsystem).

The Hartley emergence coefficient is defined in the range \([\phi_{\text{min}}, \phi_{\text{max}}]\), where \( \phi_{\text{min}} = 1 \) (minimum consistency);
\[ \varphi_{\text{max}} = \log_2 \left( \sum_{m=1}^{W} C_W^m \right) / \log_2 W \]  
(maximum consistency).

Formula (1) shows that the emergence coefficient characterizes the relative excess of the amount of information in the system, taking into account various systemic effects, over the amount of information determined without taking them into account. Thus, it is used to qualitatively reflect the level of consistency of the considered socio-economic system.

It should be noted that the maximum complexity of the subsystem \( M \) cannot exceed the number of basic indivisible elements \( W \), since the most complex element of the system is always an element that consists of all basic indivisible elements (i.e., the system itself).

For each number of basic indivisible elements of the system, there is a maximum level of consistency, which is not achievable in reality due to the rules associated with the peculiarities of forming the subsystems. So, there are almost always restrictions on the maximum and/or minimum number of basic indivisible elements as part of their some combination.

For example, if the system is formed from \( W \) basic indivisible elements, then there are the following restrictions on the complexity of its levels:

- the first level: the basic indivisible elements of the system are a generating set;
- the level: the constituent elements are formed by combinations of \( W \) by 2;
- the last level: the constituent elements are formed by combinations from \( W \) to \( M \).

In Hartley information theory, the following logarithmic measure is used to measure the amount of information contained in a message:

\[ I = L \cdot \log_2 W \text{ bit,} \]  
(2)

where \( I \) – the amount of information in the message (bits); \( L \) – the message length (the number of characters); \( W \) – the power of the alphabet used to encode the message (the number of characters).

When identifying an element of the set with power \( W \), performed for an equiprobable meeting, it is possible to exclude the multiplier \( L \), which characterizes the message length.

Based on the above and formula (2), it turns out that:

\[ I = \log_2 \left( \sum_{m=1}^{M} C_W^m \right) \text{ bit.} \]  
(3)

Let’s consider a practical example, when the work with two classifiers of NON-factors is carried out simultaneously. Suppose that there are system \( A \), which consists of basic elements \( K_\alpha \), and system \( B \), which consists of \( K_\beta \) basic elements.

In this case, the generalized emergence coefficient will be calculated by the formula:

\[ \mathcal{J}_{K_A \cup K_B} = \frac{\log_2 \left( \sum_{m=1}^{M} C_{K_A \cup K_B}^m \right)}{\log_2 \left[ \left( \sum_{m=1}^{M} C_{K_A}^m \cup \sum_{m=1}^{M} C_{K_B}^m \right) \right].} \]  
(4)

The generalized emergence coefficient (5) shows how much the amount of information obtained during the identification of one NON-factor increases as a result of their combination due to the occurrence of a systemic effect. In this case, additional virtual elements do not arise "out of nowhere".

The generalized emergence coefficient used to combine a set of systems \( \{K_\alpha\}_{\alpha \in A} \) has the form:

\[ \mathcal{J}_{\{K_\alpha\}} = \frac{\log_2 \left( \sum_{m=1}^{M} C_{\bigcup_{\alpha \in A} K_\alpha}^m \right)}{\log_2 \left( \bigcup_{\alpha \in A} \sum_{m=1}^{M} C_{K_\alpha}^m \right).} \]  
(5)
If the emergence coefficient (5) has a low value, then the project risk is motivated either by one NON-factor or by a simple combination of a small number of them. In this case, it is worth using the fuzzy logical inference according to the Mamdani algorithm, which allows obtaining the aggregate possibility of a risk event as a result of combining several NON-factors.

If the coefficient (5) has a high value, then it is necessary to combine all the NON-factors used for a specific risk into a single system. In this case, it is advisable to use a more complex approach, since the total risk may have properties that are absent in each of the combined NON-factors.

The border between low and high values of the emergence coefficient is determined by the middle of the definition area.

Consider an example of calculating the emergence coefficient. Suppose that 8 risk sources (i.e., NON-factors) were identified in the internal environment of a complex project.

Table 2 shows the results of calculating the emergence coefficient (5) when W=8 and the number of risk source combinations is varying from 1 to W.

Table 2. An example of calculating the emergence coefficient at W=8.

| m | m C<sub>W</sub> | M | \(\sum_{m=1}^{M} C^m_W\) | \(\log_2 \sum_{m=1}^{M} C^m_W\) | Emergence coefficient at \(\log_2 W = 3\) |
|---|---|---|---|---|---|
| 1 | 8 | 1 | 8 | 3.000 | 1.000 |
| 2 | 28 | 2 | 36 | 5.170 | 1.723 |
| 3 | 56 | 3 | 92 | 6.524 | 2.175 |
| 4 | 70 | 4 | 162 | 7.340 | 2.447 |
| 5 | 56 | 5 | 218 | 7.768 | 2.589 |
| 6 | 28 | 6 | 246 | 7.943 | 2.648 |
| 7 | 8 | 7 | 264 | 8.044 | 2.681 |
| 8 | 1 | 8 | 275 | 8.103 | 2.701 |

Columns 4, 5, 6 contain the maximum possible values. However, there are restrictions on the allowed combinations, i.e. their number is much less than the maximum. The example shows that all combinations of two risk sources are permissible, and the real values for \(M>2\) will be less.

For \(M=2\) the emergence coefficient is 1.723. This means that information about project risk due to the systemic effect of combination significantly exceeds the amount of information about permissible risk sources. The low value of the coefficient causes feasibility of using fuzzy logical inference according to the Mamdani algorithm [5].

4. Combining NON-factors using fuzzy pyramidal networks

To combine a set of NON-factors affecting the implementation of a complex innovative project in the aggregate, it is proposed to use the apparatus of fuzzy pyramidal networks. It is based on the algorithms of growing pyramidal networks developed at the Institute of Cybernetics of the Academy of Sciences of the USSR, which allow describing the structure of complex systems in natural language, and methods of fuzzy logic.

A distinctive advantage of this device is the automatic construction of the network, performed regardless of the initial data. Optimization of knowledge representation is carried out due to the adaptation of the network structure to the characteristics of the incoming information. Unlike most artificial intelligence methods, it is performed without introducing a priori redundancy.

The construction of a pyramidal network, which is an acyclic directed graph, consists of determining the structural relationships between the attribute description of the subject area. Its structural elements of the network are:
the receptors, i.e. the vertices located at the lower level of the network, which represent the values of indicators characterizing the sources of project risks (mainly expert assessments are used);

• the conceptors, i.e. the intermediate vertices, which are formed by combining the receptors (at least two);

• the outcomes, i.e. the vertices that represent complex assessments of some project risk.

The peculiarities of the considered research task lead to the need to modify the learning and recognition algorithms for the growing pyramidal network proposed by V.P. Gladun [6].

Thus, in the original learning algorithm, the control vertices were placed on the basis of the available statistics and the number of excited receptors. However, the unique nature of innovative projects does not allow the formation of a sufficient volume of statistical information for training the network. As a solution to this problem, it was proposed to use methods of fuzzy logic, which allows you to successfully process quasi-statistical data [7].

A distinctive feature of the proposed modification of the learning algorithm is taking into account the strength of the conceptor connectedness (i.e. the degree of influence of the vertex on the higher conceptors / outcomes), which is assessed using fuzzy inference according to the Larsen algorithm based on the analysis of the network structure, expert estimates, and available quasi-statistics. Another additionally introduced indicator is the significance of each vertex, which characterizes the degree of its consideration in the recognition process.

The recognition algorithm associated with the assessment of project risk is based on the comparison of three indicators calculated for each outcome: the cumulative possibility of occurrence; the general significance; the number of excited receptors located at the base of its pyramid.

5. Conclusion
The article proposes a new approach to assessing the risks of projects to create high-tech products. It is distinguished by the joint use of intelligent tools (neural fuzzy classifier, fuzzy logic, fuzzy pyramidal networks) to identify and analyze NON-factors of various nature. It is advisable to apply this approach when studying the prospects of complex innovative projects implemented in conditions of uncertainty and the absence of reliable mathematical methods for risk assessment.

Acknowledgments
The reported study was funded by the Grant of Russian Federation President to young Russian scientists state support MK-1534.2020.9 “Methods of modeling complex systems using fuzzy network algorithms and swarm intelligence”.

References
[1] Emelyanov A A, Bulygina O V, Emelyanova N Z and Yashin E S 2020 Simulation and fuzzy logic in import substitution risk management of high-tech equipment 2020 V Int. Conf. on Information Technologies in Engineering Education – Proc. (Moscow: IEEE Xplore)
[2] Dli M I, Bulygina O V, Emelyanov A A and Selyavskiy Yu V 2020 Intelligent analysis of complex innovative project prospects IOP Conf. Ser.: Materials Science and Engineering 919
[3] Bulygina O and Emelyanov A 2020 Analysis of NON-factors in innovative project management CEUR Workshop Proc. 2782 217-21
[4] Beynon M 2002 DS/AHP method: A mathematical analysis, including an understanding of uncertainty European Journal of Operational Research 140 148-64
[5] Mamdani E H 1974 Application of fuzzy algorithms for control of simple dynamic plants Proc. of the IEEE vol 121 12 pp 1585-1588
[6] Gladun V and Vaschenko N 2000 Analytical processes in pyramidal networks Int. J. “Information Theories and Applications” vol 7 3
[7] Bulygina O V and Chernovalova M V 2016 Application of fuzzy pyramidal network tools for
analyzing the IT-project feasibility *Proc. of the Int. Academic AMO-SPITSE-NESEFF* 89-90