Implementation of Cognitive Technologies in the Process of Joint Project Activities: Methodological Aspect

Aleksy Kwiliński
The London Academy of Science and Business, London, United Kingdom
ORCID: 0000-0001-6318-4001, e-mail: a.kwilinski@london-asb.co.uk

Jan Polcyn
Stanisław Staszic State University of Applied Sciences in Pila, Poland
ORCID: 0000-0001-7847-2743, e-mail: Jan.Polcyn@puss.pila.pl

Krzysztof Pająk
Stanisław Staszic State University of Applied Sciences in Pila, Poland
ORCID: 0000-0002-2011-1512, e-mail: krzysztofpajak@gmail.com

Sebastian Stępień
Poznan University of Economics and Business, Poland
ORCID: 0000-0001-9475-8418, e-mail: Sebastian.Stepien@ue.poznan.pl

DOI: 10.14595/CP/02/006

Abstract: The article defines the technical and logical-structural conditions for ensuring joint project activity on the basis of cognitive technologies. An algorithm was developed to support the approval of the decision on the outsourcing of project works on the terms of system outsourcing. A fuzzy logic algorithm for selecting a contractor for specific works within a specific project and providing support for the execution processes is proposed. The experimental part of the cognitive model is investigated on the basis of additive convolution of evaluation indicators. The pattern of calculation of relative degrees of advantage of indicators for estimation of potential outsourcers is created.

Key words: project management, project contractor, outsourcing, cognitive technologies, neural network modeling, fuzzy procedures, convolution of evaluations

JEL: O14

Introduction

Today, project management, as well as change management, is an area of rapidly evolving systems of management theory, and the results of its studies are widely used in the practice of project management. The high speed of change of external conditions, increasing the requirements in the area of quality of execution of works and tasks, complexity of achieving the results of implementation of projects lead to the fact that outsourcers are often involved in execution of the projects themselves, which leads to a search for new methods and practical algorithms of highly efficient and intelligently secured combination of the main project executor ("participant 1") and outsourcing firms ("participant 2").
The modern practice of joint project activity is increasingly inclined to using cognitive technologies based on the introduction of neural networks that can improve the organizational and technological efficiency and accuracy of interaction of "participant 1" and "participant 2".

The complexity of making a decision to outsource works is justified in the studies: [Atkinson, 1999; Anbari et al., 2008; Badewi, 2015; Ferreira & Otley, 2009; Turner, 1999; Bogachov et al., 2020; Dalevska et al., 2019; Dementyev & Kwilinski, 2020; Dzwigol 2019; 2020; Dzwigol & Dźwigoł-Barosz, 2018; 2020; Dzwigol et al., 2020; Kaźmierczyk & Chinalska, 2018; Kharazishvili et al., 2020; Kuzior et al., 2019; Kwilinski, 2018a; 2018b; 2019; Kwilinski et al., 2019; 2019a; Kwilinski & Kuzior, 2020; Lyulyov et al., 2020; 2021; Melnychenko, 2019; 2020; 2021; Miskiewicz, 2020a; 2020b; Saługa et al., 2020; Tkachenko et al., 2019a; 2019b; 2019c; 2019e]. At the same time, it should be taken into account that the alternative of outsourcing or in-house service to perform the project works does not directly influence the effectiveness of the project implementation, and to a greater extent, the change in efficiency is mainly due to the factors that determine this alternative [Hubbard 1990, Wit 1988]. The advantages of outsourcing in project management and support of joint participation are as follows: 1) opportunity for "participant 1" to focus on the main works within the project; 2) transferring the problem of searching for highly specialized personnel and human-related risks to the side of "participant 2"; 3) distribution of project risks between several parties; 4) speed and flexibility of obtaining the necessary competencies without the need for a “bloated” personnel of executors within the framework of the project; 5) reducing the cost of maintaining the infrastructure required to organize the process of project development and support; 6) possibility of applying technological innovations without time and financial costs. Taking into account the highlighted advantages, we conclude that, ultimately, the effective use of project outsourcing can improve the quality of products and services to fully meet the needs of the customer, while reducing costs and time to implement the project, as well as releasing resources for the highest priority areas of development.

At the same time, the use of outsourcing is associated with a number of negative aspects, which "participant 1" is obliged to control and level: 1) possibility of losing control over the development and support of separate project works; 2) possibility of reducing competitiveness due to the loss of technical knowledge and motivation of executors;
3) increased level of information risk associated with the presence of additional users of the project information system [Chih & Zwikael, 2015; Kerzner, 1987]. It is just this dialectical approach to the problems that forms the targeted and actual direction of studies on leveling negative consequences in project management on the basis of outsourcing with the use of intellectual cognitive models.

**Methodological Framework of the Study**

In terms of developing the methodological framework of the study, we proceed from the proposition that the technological aspect of management and the intellectual need in support of each project makes it impossible to iterate the phenomena that take place in the process of their implementation. Apart from the condition of project implementation, the determining factor of the external environment, which is influenced by a large number of accounted and unaccounted factors, is the economic dynamics, which forms the irreproducible conditions in which the project is being implemented. All of this implies the need for a polymethodic consideration of the scientific problem presented by the following methods.

**Interval-Probabilistic Method**

According to this approach, observations are presented not as numbers but as intervals reflecting the range of possible values of the observed values in the environment of the implemented project. Thus, in the works [De Marco & Narbaev, 2013; Liou, 2011] it is shown that, as a result of observation, not the elements of the sample \( x_j \) but the values \( y_j = x_j + \varepsilon_j \), where, \( \varepsilon_j \) – errors of measurements, studies and experiments, get known. Then the statistical distribution that the specified observations are subordinated is not \( f(x) \) but \( f(y) \), which differ as follows: 

\[
N_f(x) = \sup |f(y) - f(x)|
\]

(supremum is taken from the set of possible values of the error vector). The use of this technique through extending probabilistic values to interval-probabilistic ones allows to weaken the assumption of the classical probabilistic-statistical approach, which makes it possible to use it reasonably in the absence of homogeneous statistical observations in project management [Dvir et al., 2006].
Fuzzy-Logical Methods

The use of this approach is the most appropriate in cases of high complexity of the object under study, its nonlinearity, complexity of formalization, and in situations where the sources of information are interpreted qualitatively, inaccurately or indeterminately [Caron et al., 2013; Thamhain, 2014]. The basis of fuzzy-logical methods of calculating uncertainty is the concept of fuzzy sets. Under the fuzzy set A in the ground set \( X = \{x\} \), in the sum of pairs \( A = \{\mu_A(x), x\} \), \( \mu_A: X \rightarrow [0,1] \) is a membership function of the fuzzy set. Whereas, the value of this membership function \( \mu_A(x) \) for the element \( x \in X \) is called the measure of membership. Building membership functions is one of the main problems with the practical use of fuzzy sets. These methods can be either direct (assigning membership functions to a graph, table or formula) or indirect (statistical method, method of subtractive clustering, method based on expert estimates, parametric methods, etc.).

An important component of the fuzzy logic approach (method) is the notion of a linguistic variable that allows us to formalize fuzzy concepts of natural language [Afshari et al., 2014]. A linguistic variable is a set \( \beta = \{\beta, T, X, G, M\} \), where \( \beta \) is its name; \( T \) is a set of its values (terms), which are the names of fuzzy variables \( \{\alpha, X\alpha, A\} \) (where \( \alpha \) is the name of a fuzzy variable; \( A \) is a fuzzy set on the range of definition \( X\alpha \)); \( G \) is a syntactic procedure allowing to generate new terms; \( M \) is a semantic procedure allowing to convert terms that were generated by the syntactic procedure into fuzzy variables. At the same time, linguistic variables can be both numerical (then its term-set consists of fuzzy numbers) and non-numerical, which allows to reflect both physical (quantitative) and linguistic (qualitative) uncertainty.

Evaluation of Alternatives in Case of Additive Evaluation Indicators

Suppose we set \( n \) alternatives \( A_i, i=1,n \), which should be evaluated by \( m \) indicators \( x_j, j=1,m \), and the relative importance of each of them is set by a coefficient, \( w_j, j=1,m \). If these indicators are additive, then a weighted estimate of the \( i \)-th alternative is calculated by the formula:
where:

Rij is the estimate of the i-th alternative using the j-th indicator. If the estimates are normalized, the following formula is used:

(2) \[ R_i = \sum_{j=1}^{m} w_j R_{ij} \]

Since the estimates are fuzzy numbers, for the purpose of implementing the operations of multiplication and summation, it is required to use the above formula in accordance with one of the methods of fuzzy arithmetic and soft computing, particularly the interval method or based on the principle of Zadeh fuzzy generalization. If the estimates are expressed as triangular or trapezoidal numbers, then the algebraic summation will also result in triangular and trapezoidal numbers, respectively, and in the general case, a fuzzy arbitrary number becomes the result of multiplication and division. After calculating the weighted estimates Ri it is required to compare the alternatives Ai based on them. For this purpose, different methods of ordering of fuzzy numbers are used [Cicmil and Hodgson, 2006]. The alternative that ranked first in the ordered set is considered to be the best one.

**Fuzzy Logic of Processes**

The system of fuzzy inference is “the process of obtaining fuzzy conclusions about the required control over an object based on fuzzy conditions or preconditions, which are information about the current state of the object” [Carnall, 2007]. At their core, the systems of fuzzy inference have a knowledge base that is formed by experts in the subject area and is intended to formalize their empirical knowledge in the form of a set of fuzzy cognitive rules as follows: (i): Q,P,A→B,S,F,N, where (i) is the name of a fuzzy product; Q is the area of its application; P is the condition of applicability of its core; A→B is the core of a fuzzy product where: 1) A is the condition of the core (LHS), 2) B is embedding of the core (RHS), 3) → is the designation of a logical implication operation; S is method of determining the quantitative value of the degree of truth of the core creation; F is the confidence factor of a fuzzy product; N are post-conditions of project implementation [Coppin, 2004].
Core $A \rightarrow B$ is the central component of the system of fuzzy inference and is generally presented as a fuzzy predicative rule of the form: $P_i$. If $x$ is 1, then $y$ is $i$, where $x$ is an input variable, $y$ is an output variable, $A$ and $B$ are functions of a device defined on $x$ and $y$, respectively. The procedure of logical inference usually consists of the following main steps: introduction of fuzziness (fuzzification), aggregation of the degree of truth of preconditions of rules, activation of conclusions of rules, accumulation of the activated conclusions of rules and bringing to crispness (defuzzification). At the same time, the components of fuzzy models can have different implementation and the choice of specific implementation of one of the components often determines the choice of all other components.

**Method of Neural Network Modeling**

Artificial neural networks have their own limitations related to the inability to input the a priori information and the complexity of analysis of learning of such networks. These premises provide the basis for the framework of hybrid neural networks allowing to obtain the result based on a system of fuzzy inferences, and the parameters of these systems are tuned using the algorithms of neural network learning [Anbari, 2003; Russell and Norvig, 2016]. A hybrid neural network implies a neural network with clear signals, weights and activation function, but with the implementation of aggregation and activation operations using t-norm, t-conorm and other continuous operations. Hybrid networks are based on the concept of fuzzy neurons. The structure of fuzzy neurons "AND" and "OR" is presented in Fig. 1.

**Figure 1. Structure of Fuzzy Neurons "AND" and "OR"**

![Figure 1. Structure of Fuzzy Neurons "AND" and "OR"](source: own elaboration)

When using the neuron "AND", the following approaches can be used to implement triangular norm operation: min–conjunction, algebraic formation, boundary
formation, drastic multiplication of truth of preconditions of rules, etc. When using the neuron "OR", the following approaches are usually used to implement triangular norm operation: max–disjunction, algebraic sum, boundary sum, drastic sum of weights of truth of preconditions of rules.

**Main Findings**

**Algorithm for Support of Approval of a Decision on Outsourcing of Project Works (System Outsourcing)**

In case of technological and economic necessity, priority areas of “selective outsourcing” or “integrated outsourcing” should be selected, and at the same time, the task of evaluation of the expediency of outsourcing specific project works [Hatfield, 1995] becomes relevant. We reduce this problem to solving the task of multicriteria binary classification: depending on the results of the evaluation of the task using a set of indicators, it is required to refer it to one of two classes, that is, to perform a mapping such as \( f(R) : R \rightarrow Y \in \{C_1, C_2\} \), where \( C_1 \) is the class of works for which outsourcing is inappropriate, \( C_2 \) is the class of works for which it is advisable to involve third party executors [Ofer and Jack, 2019; Vakola et al., 2004]. The authors suggest the following set of indicators of such classification:

1. Impact of a work on information security of the project customer (\( z_1 \)):
   - the amount of losses as a result of the unavailability of work-related project services (\( z_{1,1} \)), USD;
   - the amount of losses as a result of the violation of data integrity through the use of work-related project services (\( z_{1,2} \)), USD;
   - the amount of losses as a result of the violation of data privacy through the use of work-related project services (\( z_{1,3} \)), USD;

2. Technical and economic characteristics of design works (\( z_2 \)):
   - term of performance of work by own efforts (\( z_{2,1} \)), days;
   - term of performance of work involving outsourcing (\( z_{2,2} \)), days;
   - cost of performance of work by own efforts (\( z_{2,3} \)), USD;
   - cost of performance of work involving outsourcing (\( z_{2,4} \)), USD;
   - number of actions (\( z_{2,5} \)) that are blocked by a task, units
3. Rating of a potential third party contractor of project works ($z_3$).

4. The readiness of the main contractor to perform project works independently ($z_4$):
   - number of similar works already performed by the contractor personnel ($z_{4,1}$), units;
   - proportion of employees with an academic degree on the staff ($z_{4,2}$), units;
   - proportion of R&D expenses in revenue ($z_{4,3}$), %;
   - average cost of previously completed projects ($z_{4,4}$), units;
   - number of similar projects completed ($z_{4,5}$), units;
   - compliance of processes with policy subdivisions and information security standards ($z_{4,6}$), units; linguistic estimate;
   - availability of necessary infrastructure and tools ($z_{4,7}$), linguistic estimate;
   - level of knowledge of necessary technologies ($z_{4,8}$), linguistic estimate;
   - skill level of the engineer team ($z_{4,9}$), linguistic estimate;
   - relevance of technological and instrument stack used ($z_{4,10}$), linguistic estimate;
   - prevalence of review practices and result audit ($z_{4,11}$), linguistic estimate;
   - stability of the use of automated tools for quality assurance of project works ($z_{4,12}$), linguistic estimate.

The use of the proposed set of indicators suggests that a potential third-party contractor (outsourcing firm) has already been selected [Eduardo et al., 2015]. The authors propose an algorithm for estimating the feasibility of outsourcing specific works within the project, a block diagram for which is presented in Fig. 2.

In the current practice of project management, a large number of algorithms for classification of compositions of project works is used, but most of them have a significant drawback – they work on the principle of "black box", that is, an attempt to explicitly interpret the patterns leading to the object attribution to one of the classes, and this results in some difficulties [Turner & Muller, 2003].
The classifiers based on cognitive models of knowledge representation, which underpin the rule base, are devoid of this drawback. Given that the works within a particular
project are usually unique and specific, the authors conclude that the “accurate” estimate of \( z_{1,1} - z_{1,3}, z_{2,1} - z_{2,4}, \) is difficult, and the use of probabilistic methods to remove uncertainty is complicated by a lack of relevant static information to confirm a specific distribution law. The indicators \( z_{4,6} - z_{4,12} \) have a qualitative nature and do not have a physical measurement scale [Einhorn et al., 2019].

Considering the above, the authors conclude that the rules of classification of the cognitive model should be fuzzy. The systems based on fuzzy inference work effectively even in situations where obtaining information is fraught with various difficulties and parameters and input data are not accurate and properly represented [Zwikael and Smyrk, 2011]. In this regard, it is suggested that an approach based on a fuzzy rule base can be used to evaluate the appropriateness of outsourcing the project [Freeman and Beale, 1992]. The structure of the cognitive model proposed by the authors to evaluate the feasibility of outsourcing a project is presented in Fig. 3. As we can see, a hierarchical structure is used. For integrated indicators, their own knowledge bases are created, which outputs are supplied at the input of a higher level knowledge base.

**Figure 3.** The Structure of a Cognitive Model for Evaluating the Feasibility of Outsourcing a Project

![Diagram](Source: own elaboration)

Such approach makes it possible to overcome the “curse of dimension” as the number of fuzzy rules contained in the base is significantly reduced. In the first stage of designing
a hierarchical system of fuzzy inference, it is required to specify the structure of a base of fuzzy lower-level cognitive rules. It is often difficult for a project manager to manually create rules as the range of definition of input variables is limited by the knowledge of the manager, and a large amount of input data leads to combinatorial explosion of the number of rules in the base. One such combination method is the method of subtractive clustering of experimental data, which is characterized by relative ease of use and the availability of software implementation [Fuchs et al., 2016]. In addition, as a result of using this method, rules that correspond to the areas of greatest concentration of data are generated, which eliminates the problem of combinatorial explosion and makes a fuzzy system more transparent for "participant 1" and "participant 2" within a separate project [White and Fortune, 2002]. Each center of the cluster \( V_1, V_2, \ldots, V_n \) found as a result of clustering where \( V_i = (z_1, z_2, \ldots, z_m, y); i = 1, n \) is matched with a fuzzy rule of the type:

\[
\text{Rule}(i): \text{If } \{z_i \text{ near } z_i^*\} \text{ then } \{y_i \text{ near } y_i^*\}.
\]

Membership functions are obtained through the process of designing the membership degrees of a respective cluster on the axis of variables, after which a set of membership degrees is approximated by the corresponding functions. As a result, we obtain the rules of a more conventional form:

\[
\text{(3)} \quad \text{Rule}(i): \text{If } z_i \text{ is } A_{1i} \text{ and } z_2 \text{ is } A_{2i} \text{ then } y \text{ is } B_i
\]

where:
- \( z_j, j = 1, n \) are input system parameters; \( y \) is an output variable;
- \( A_{ij}, B_i \) are membership functions of fuzzy sets defined for \( z_i \) and \( y \), respectively.

As a result of applying this transformation to the centers of all the clusters found, a knowledge base of a lower-level fuzzy system of inference is created within the project works. For the lower level of the hierarchy (which rule inputs are components of integral criteria), the rules will look as follows:

\[
\text{(4)} \quad \text{Rule}(j): \text{If } z_{i1} \text{ is } A_1 \text{ and } z_{i2} \text{ is } A_2 \text{ and } \cdots \text{ and } z_{ik} \text{ is } A_k \text{ then } z_j \text{ is } B_j
\]

where:
- \( z_{ik} \) – \( k \)-th criterion of \( i \)-th group of criteria of classification of project works;
- \( z_j \) – integral criterion of \( i \)-th group;
- \( A_k \) – membership function defined for \( z_{ik} \);
$B_j$ – membership function defined for $x_i$.

The method of subtractive clustering, despite its advantages, often builds suboptimal membership functions and does not have the ability to adjust the weights of rules. One of the most common optimization methods is to use hybrid neural-fuzzy networks [Schmidhuber, 2015]. The essence of this method is to build a neural network, which is isomorphic to the rules of the knowledge base, and its further learning. For the rules of Mamdani type, the M-ANFIS architecture of a hybrid neural-fuzzy network is generally used [Wauters and Vanhoucke, 2015]. The fragment of a network for the base of rules for the integral criterion "Impact of a work on the terms of performing a project as a whole, taking into account the activities of the outsourcer" is presented in Fig. 4. (the structure is similar for the other criteria).

**Figure 4.** Fragment of a Hybrid Neural-Fuzzy Network for the Criterion "Impact of a Work on the Terms of Performing a Project as a Whole, Taking into Account the Activities of the Outsourcer"

![Figure 4](image-url)

*Source: own elaboration*
The first layer of this network implements the fuzzification operation. The node outputs of this layer are the values of the membership functions at the given input values: 

\[ O_{i,j} = \mu_{A_{i,j}}(z_i) \]

It should be noted that in the case of using fuzzy input data, the membership degree of input data is a marker of the membership of one fuzzy set in another set and is calculated as the height of intersection of these fuzzy sets, which is illustrated in Fig. 5.

**Figure. 5.** The Calculation of the Membership Degree of Fuzzy Input Data When Evaluating the Feasibility of Outsourcing Project Works

The second layer implements the rules of a fuzzy system. Its outputs are the degrees of activation calculated using the T-norm: 

\[ O_{2,i} = w_i = \mu_{A_{i,1}}(z_1) \land \mu_{A_{i,2}}(z_2) \land \ldots \land \mu_{A_{i,n}}(z_n) \]

The third layer performs an implication operation \( O_{3,i} = w_i \circ C_i \). The fourth layer performs the aggregation operation \( O_{4,i} = \sum w_i \circ C_i \). The fifth layer implements the defuzzification of the output of the previous layer (usually based on the center of gravity method).

As a result of the learning of a hybrid neuro-fuzzy network, the parameters of membership functions are modified, which improves the quality of the system of inference when performing the project works. In addition, using the method of back propagation of error, you can adjust the appropriate weights of rules, which also has a positive effect on the quality of the model of cognitive project management [Ashurst et al., 2008].
Hybrid neuro-fuzzy networks for other integral criteria are built and trained in the same way. As a result, the outputs of these networks are supplied at the input of a higher-level rule base. The process of generating the rules of a higher-level knowledge base is similar to that process for the lower levels, but in this case the Sugeno algorithm of fuzzy inference is used [Deco and Schermann, 2000]. Each center of the cluster is matched with a rule of the type:

\[
\text{Rule}(j): \text{If } z_1 \text{ is } A_1 \text{ and } z_2 \text{ is } A_2 \text{ and } \ldots \text{ and } z_n \text{ is } A_n \text{ then } y \text{ is } C_m
\]

where:

- \( z_n \) – integral criterion of classification of project works;
- \( A_n \) – membership functions \( z_n \);
- \( y \) – membership of a project work in a class \( (C_1, C_2) \).

As a result of applying this transformation to the centers of all the clusters found, the knowledge base of a fuzzy system of inference is created. One of the most well-known methods for optimizing the rule base in classification tasks is a hybrid neuro-fuzzy classifier [Jin et al., 2017]. The basic node of the network for the task of classification of works within a project is presented in Fig. 6.

**Figure 6.** Hybrid Neuro-Fuzzy Classifier of Works within a Project

![Hybrid Neuro-Fuzzy Classifier of Works within a Project](source: own elaboration)
The first layer of the hybrid neuro-fuzzy classifier performs a defuzzification procedure, giving at the output the degrees of the membership of input variables in the respective fuzzy set $A_{ij}$. The neurons of the second layer, while implementing the operation of the T-norm, calculate the degree of activation of fuzzy rules. The neurons of the third layer perform a weighted summation of the degrees of activation of rules, and the fourth layer, based on the activation functions of the sigmoid type, calculates the degrees of membership of the object in one of the two classes [Jang, 1992]. The result of network learning is optimized membership and weight functions of cognitive rules.

The algorithm developed to support the approval of a decision on transferring separate project works on the principles of outsourcing and based on the analysis and classification of these works differs from the known ones by the application of such classification criteria as the impact of specific outsourced work on the system and reliability of the outsourcer, as well as the presentation of work characteristics in the form of antecedents of a hierarchical fuzzy-logical cognitive base of rules with further tuning of its parameters with the help of a neural network to take into account the tendencies of development of a project participant as an economic system [Globerson and Zwikael, 2002; Lin et al., 2019]. The application of the above algorithm as a tool of managerial activity makes it possible to increase the economic efficiency of project implementation through the involvement of responsible outsourcing structures.

**Fuzzy-Logical Algorithm for Selecting a Contractor for Specific Project Works**

The evaluation of the feasibility of outsourcing a project work involves the analysis of a potential outsourcer, and in the case of a decision on outsourcing, it is required to make a final choice. The authors suggest a cognitive algorithm of a fuzzy-logical procedure for selecting a third-party contractor for specific project works based on topic modeling technology to implement preliminary filtering of possible third-party contractors within a specific project (Fig. 7).

Probabilistic case models allow to work with documents described in the natural language of algorithms. This allows these models to be applied without the need for prior intelligent data processing. Thus, it was concluded that text descriptions of contractors and tasks ($d_1, ..., d_n$) are divided into semantically different components – 1) "title", 2) "basic
content", 3) "initial potency", respectively. To account for the specificity of each component (length, "noisy" background words), it is advisable to separate them at the input of the model, which was implemented by distinguishing three modalities in the topic model. In this case, the specificity of the components is taken into account by specifying a separate matrix of cognitive terms for each modality [Naeni et al., 2011; Shenhar & Wideman, 2000].

The disadvantage of probabilistic topical models is their fragmentary nature and algorithmic instability. One of the ways to solve this problem is through regularization [Pajares et al., 2011]. From this point of view, the approach of building topical models based on additive regularization is of the greatest interest.

**Figure 7.** The Procedure for Preliminary Filtration of the Set of Possible Third-Party Contractors of a Project
This approach is about superposition on the search for solution of additional constraints, each of which is formalized as a regulator of the optimization criterion $R_{\Phi, \Theta} = \max$, where $\Phi$ is a matrix of topic terms; $\Theta$ is a matrix of document topics. The weighted sum of such criteria $R(\Phi, \Theta)$ is maximized together with the basic plausibility criterion. The advantages of this approach include the simplicity of the mathematical methods, generalization of most known approaches, presence of software implementation. In view of the above, for preliminary filtering of the set of possible subcontractors of the project, the authors suggest to use additive regularization of topical models, and for selection of a regularization strategy, it is proposed to maximize the weighted normalized sum of internal criteria of quality of the topical model [Munns and Bjeirmi, 1996; Shipley and Johnson, 2009].

The result of building a topical model is a matrix $\Theta(\Theta(d_1)\ldots\Theta(d_n))$, which describes the distribution of topics by documents (textual descriptions of potential contractors), as well as a function $\varphi(d^*) \rightarrow \Theta(d^*)$, which allows to distribute topics in a new (not involved in the building of the model) document $d^*$. Given this, the task of the preliminary filtration of a set of outsourcers are reduced to finding a set of documents $(d_1, \ldots, d_n)$, which describes third party contractors, for which the distribution of topics is similar to the distribution of topics in a document describing the specific work of the project, i.e., documents, whose vectors $(\Theta(d_1)\ldots\Theta(d_n))$ are similar to the target vector $\Theta(d^*)$. After that, they should be ranked according to the degree of similarity, i.e., it is assumed that more similar documents are more relevant and should be displayed at the beginning of the list of recommendations within the performance of project works [Shenhar et al., 2001].

There are many known ways to determine the degree of similarity of vectors in a multidimensional space (Euclidean metric or distance, Manhattan distance, etc.). The general problem of estimating the proximity of vectors is the "curse of dimension" – in the case of a recommended system, for one target document, it is required to calculate the metric $n$ times ($n$ is the number of documents in the database excepting the target document). Particularly urgent problem arises in the conditions of intensive entry of requests for the development of project recommendations – in this case, the number of necessary calculations of the degree of similarity is determined by the formula:

$$
\Omega(\Omega(d_1)\ldots\Omega(d_n)) = \max_{i=1}^{n} \left( \varphi(d_i) \right) 
$$
where:

\( n \) is the number of documents in a collection.

One of the options to reduce the criticality of this problem is to use an inverted index, which is about storing a set of documents for each topic, where the probability distribution of the topic is positive, in the stage of creating a matrix \( \Theta \). When making recommendations for the \( d^* \) document, only those documents that are contained in the inverted index in those positions for which the probability distribution in the \( d^* \) document is positive are selected. Due to the fact that in the process of regularization the matrix \( \Theta \) is getting very thin, the number of calculations of the degree of similarity is significantly reduced (Wang, 2009). In spite of this, the amount of computing required can be large enough (for complex projects). In this regard, it is proposed to select in the inverted index for a topic only those documents, where the probability of this topic is higher than the average value.

\[
\Theta(d_i) > \sum_{j=1}^{n} \frac{\Theta(d_j)}{n}
\]

where:

\( i \) – \( i \)-th document of a collection;

\( \Theta(d_i) \) – probability value of a topic in the document \( i \);

\( n \) – number of documents in a collection.

It is advisable to calculate the average probability values for topics once in the learning process, in which case the use of this criterion will not affect the computational complexity of the algorithm for calculating similarities within the project.

To evaluate the similarity with the target document, the authors are asked to select documents only from the topics whose probability in the target document is higher than the average probability value in this document. Given that vectors \( \Theta(d_i) \) are normalized to one, the mean probability is a constant and inversely proportional to the number of topics of the model: \( \Theta(d_i) \propto 1/T \). Therefore, the application of this criterion does not affect the computational complexity of the algorithm for calculating similarity within a project [Zhang & Chen, 2018].

The application of the proposed selection methods allows to reduce the computational complexity of the algorithm for calculating the similarity of document vectors from the
database and the target document by reducing the number of required calculation operations. At the same time, excluding documents from a set of comparison does not have a significant negative effect on the quality of the recommendations, since only documents with a relatively low probability distribution of topics, which in the overwhelming majority of cases are not relevant to the target document, are excluded.

In spite of the availability of a great number of metrics for measuring the proximity of vectors in multidimensional space, the cosine degree of proximity is most commonly used. The cosine degree considers the attributes of the vector model to be independent and completely separate, although some topics in the model can often be similar to each other. In this regard, the authors suggest to use "soft" cosine degree for the calculation of similarity of the vectors of distribution of topics in documents, which takes into account the similarity between the topics (7):

\[
\text{soft\_cosine} = \sum_{i,j}^{N} s_{ij} \times d_{ti} \times d_{tj}
\]

where:
- \( s_{ij} \) – similarity between \( i \)-th and \( j \)-th topics;
- \( d_{ni} \) – occurrence probability of \( i \)-th topic in the document \( d_n \).

To calculate the similarity between the topics, it is suggested to use the cosine degree by distribution vectors of the terms in topics (8).

\[
\cosine(T_1, T_2) = \frac{\sum_{i}^{n} \Phi_i(T_1) \times \Phi_i(T_2)}{\sqrt{\sum_{i}^{n} [\Phi_i(T_1)]^2} \times \sqrt{\sum_{i}^{n} [\Phi_i(T_2)]^2}}
\]

where \( n \) is the number of tokens;
- \( \Phi_i(T) \) – probability of the topic \( T \) for \( i \)-th token.

It is suggested to calculate the degree of similarity between the topics only once in the stage of model learning. In this case, the use of this criterion will not adversely affect the computational complexity of the algorithm for calculating the proximity of documents within a project.
Experimental-Modular Part of the Cognitive Model

Over time, the preliminary filtration model database is supplemented with new documents. It seems impractical to retrain the model with each occurrence of a new document – it will put a great strain on the computation system, and the impact of a single document on the structure of the model is estimated as negligible (Cioffi, 2005). In this regard, the authors suggest to retrain the model (including building an inverted index and calculating the degree of similarity between the model topics) using one of the following patterns: 1) periodically (for example, once a week); when there is accumulation of critical mass of documents (for example, 3-4% of the documents involved) that were not involved in building the model; 2) when the external quality criteria of the model are reduced.

As an experiment, there was built a model with 20 topics within the project, three of which are background ones. The regularization strategy was selected using the method of maximizing the weighted normalized sum of internal quality criteria ($D$). The selected strategy allowed to improve the internal quality criteria of the model, which is shown in Table 1.

Table 1. Internal Quality Criteria of a Cognitive Model

<table>
<thead>
<tr>
<th>Modality</th>
<th>Criterion</th>
<th>No regularization</th>
<th>Optimal strategy</th>
<th>$D$, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Perplexity</td>
<td>222.88</td>
<td>280.15</td>
<td>26.13</td>
</tr>
<tr>
<td></td>
<td>Proportion of background words</td>
<td>0.06</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Core contrast</td>
<td>0.96</td>
<td>0.99</td>
<td>3.13</td>
</tr>
<tr>
<td></td>
<td>Core purity</td>
<td>0.46</td>
<td>0.57</td>
<td>23.91</td>
</tr>
<tr>
<td></td>
<td>Sparsity of $\Phi$</td>
<td>0.88</td>
<td>0.96</td>
<td>9.09</td>
</tr>
<tr>
<td>Title</td>
<td>Perplexity</td>
<td>1453.36</td>
<td>1748.12</td>
<td>20.28</td>
</tr>
<tr>
<td></td>
<td>Proportion of background words</td>
<td>0.31</td>
<td>0.31</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Core contrast</td>
<td>0.97</td>
<td>0.98</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>Core purity</td>
<td>0.18</td>
<td>0.22</td>
<td>22.2</td>
</tr>
<tr>
<td></td>
<td>Sparsity of $\Phi$</td>
<td>0.83</td>
<td>0.93</td>
<td>12.05</td>
</tr>
<tr>
<td>Basic content</td>
<td>Perplexity</td>
<td>123.37</td>
<td>150</td>
<td>21.59</td>
</tr>
<tr>
<td></td>
<td>Proportion of background words</td>
<td>0.04</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Core contrast</td>
<td>0.943</td>
<td>0.97</td>
<td>2.86</td>
</tr>
<tr>
<td></td>
<td>Core purity</td>
<td>0.4</td>
<td>0.49</td>
<td>22.5</td>
</tr>
<tr>
<td></td>
<td>Sparsity of $\Phi$</td>
<td>0.83</td>
<td>0.94</td>
<td>13.25</td>
</tr>
<tr>
<td>Initial potency</td>
<td>Perplexity</td>
<td>0.03</td>
<td>0.84</td>
<td>2700</td>
</tr>
</tbody>
</table>

Source: own elaboration
As a control sample, 10% of documents were randomly selected from a total of 5,200
documents. Table 2 presents a comparison of the number of required calculations of proximity
degree between document vectors calculated for each document of the control sample (using
the proposed previous selection and without it).

Table 2. The Number of Required Calculations of Proximity Degree of Documents within the
Cognitive Model of Selection

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Without previous selection</th>
<th>With previous selection</th>
<th>D, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>11,288</td>
<td>10,239</td>
<td>-9</td>
</tr>
<tr>
<td>Mode</td>
<td>4,495</td>
<td>3,632</td>
<td>-19</td>
</tr>
<tr>
<td>Median</td>
<td>9,653</td>
<td>7,288</td>
<td>-24.5</td>
</tr>
</tbody>
</table>

Source: own elaboration

Each document from the control sample was matched with a ranked list of
recommendations from 10 documents. Experts were asked to evaluate the relevance
of documents from the ranked list of the target documents using a binary scale (0 - not
relevant, 1 - relevant). As a result, the score of each expert is a set of tuples like {(10 ... 1),
(00...1)...(10...0)}. The potency of a large quantity is equal to the number of documents in the
control sample and the length of each tuple is equal to the length of the ranked list (in this
case - 10). There were provided mean scores with respect to each marker. “Mean average
precision at 10” appeared to be equal to 0.73 and “Mean reciprocal rank at 10” is equal
to 1.74. This suggests that the mean precision of the first 10 documents is 73%, and
at the same time, the project manager on average will likely find the first relevant document
at the position 1.74 (certainly, the positions themselves are indicated by natural numbers).

Preliminary filtration of a set of contractors must result in a ranked list of outsourcers.
After selecting the top-N outsourcers from the potential list (the value of N is determined
based on the volume of resources of a project customer available for detailed examination),
it is proposed to detail the information about them [Thomas and Mullaly, 2008]. In this stage,
it seems appropriate to clarify the ability and interest of each outsourcer in performing
the work, as well as to formalize the information needed for further evaluation of the
characteristics of a potential outsourcer. The authors propose the following criteria
for evaluating a potential outsourcer:

- number of joint projects (v1);
number of similar works completed (v2);
average cost of previously completed projects (v3);
time of activity in the project services market (v4);
proportion of R&D expenses in total revenue (v5);
proportion of employees with an academic degree on the staff (v6);
work completion time (v7);
cost of performing certain works within the project (v8);
contractor reputation (v9);
openness of the process for the customer (v10);
actuality of technologies and infrastructure used (v11);
quality of services provided within the project (v12);
transparency of contract conditions (v13);
process compliance with information security standards (v14);
financial stability (v15).

It is obvious that different stages of the life cycle of the project of development and support of works are dominated by certain characteristics that influence the process of making decisions on the use of project outsourcing, such as: the required quality of work delivery \((c_1)\); criticality of observance of terms of work delivery \((c_2)\); required level of information security \((c_3)\); flexibility degree of the financial plan \((c_4)\); frequency of progress control \((c_5)\); frequency of changes to the specification of requirements within the project \((c_6)\); amount of accumulated knowledge about the project system \((c_7)\); proportion of "non-standard" innovative works \((c_8)\), etc.

For example, at the stage of MVP (Minimum Viable Product) development, the quality of work is usually less important than at the stage of project development, and the amount of system knowledge accumulated increases as we move from stage to stage. In turn, the above characteristics influence the relative importance of evaluation criteria of potential third-party contractors [Sparrow, 2012]. So, the actuality of the technologies and infrastructure used is especially important in situations where high quality of work is required, and the presence of joint projects has a positive effect in situations where a large amount of system knowledge is available [Henderson, 2004]. The specified indirect influence
of a project stage on the relative importance of the evaluation criteria of third-party contractors creates a neural structure, the graphical model of which is presented in Fig. 8.

**Fig. 8.** Graphic Model of a Neural Network to Determine the Degree of Influence of a Project Stage on the Relative Importance of Criteria for Evaluation of Outsourcers

Based on the analysis of this model, the weighting factor of the $i$-th criterion at the $k$-th stage of the project is proposed to be calculated by the formula:

$$w_{k,i} = \sum_{j=1}^{n} W_{k,j}^C \times W_{i,j}^V$$

where:

- $W_{k,j}^C$ – importance of the $j$-th characteristic of the process of algorithm development in the $k$-th stage of the project;
- $W_{i,j}^V$ – relative degree of importance of the $i$-th criterion when evaluating the $j$-th characteristic of the process of development of a control project algorithm.

To calculate the values $W_{k,j}^C$ and $W_{i,j}^V$, it is proposed to use the Saati method. This method is based on paired comparisons of the degree of dominance of objects on a scale of intensity from 1 to 9. The diagram of this process (using the example of criteria for evaluation of potential outsourcers) is presented in Table 3.
It seems appropriate to determine the degree of violation of consistency of estimates. This usually involves the consistency ratio calculated by the formula:

\[
OS = \frac{(1-IS)}{CC}
\]

where:  
\( IS \) – consistency index calculated by the formula:

\[
IS = \frac{\lambda_{\text{max}} - n}{n-1}
\]

where:

\( \lambda_{\text{max}} \) – maximum proper number;

\( n \) - dimension of the matrix;

\( CC \) – random consistency (for a matrix with a dimension of 15CC equal to 1.59).

If the consistency ratio is greater than 0.1, the estimates should be revised.

The above process must be iterated for each stage of the project life cycle.

**Discussion**

In continuation of this research area, it is possible to determine the analogy of evaluating the degree of importance of the considered and substantiated criteria when evaluating the characteristics of different stages of the project. As a result, there will be obtained weighting factors of the criteria for evaluation of potential outsourcers, which take into account the variable of relative importance of the criteria at different stages of the project.
Some of the proposed criteria for evaluation of potential outsourcers (v9-v15) have a qualitative nature and do not have a physical scale of measurement. As noted, it is appropriate to use linguistic variables for evaluation of such criteria. For these criteria, it is proposed to introduce a linguistic variable like $\beta_i=\{B,T,X,G,M\}$, $i=9:15$ ($i$ is an ordinal number of the criterion) with triangular membership functions of terms and a range of definition $[0,1]$. Then, the result of the evaluation of each criterion ($v_9$-$v_{15}$) is a fuzzy triangular number corresponding to the term of the linguistic variable that it selected. Moreover, the "precision" evaluation of a number of proposed criteria ($v_7$-$v_8$) is complicated by the innovative nature of works, which makes it appropriate to use fuzzy numbers like $d=(d_1,d_2,d_3)$ for their evaluation.

Considering the above, it is suggested to use a fuzzy weighted additive convolution of evaluation criteria for evaluation of a potential third-party contractor:

\begin{equation}
    r = \Theta \sum_{i=1}^{15} W_{k,j} \times \overline{v}_i
\end{equation}

where:

$r$ – contractor rating;

$\overline{v}_i$ – normalized estimate of $i$-th criterion.

Qualitative criteria ($v_9$-$v_{15}$) are normalized since the range of membership functions of terms of linguistic variables is $[0,1]$. The criteria ($v_{1}$-$v_{6}$) implement the dependencies like "the larger the value of the criterion, the more preferred is the contractor" and their estimates are generally "crisp". These criteria are normalized by the formula:

\begin{equation}
    \overline{v}_i = \frac{v_i}{N}
\end{equation}

where:

$N$ is a number setting the upper limit of the criterion value (the maximum value of the criterion for the whole sample of outsourcers).

The criteria ($v_{7}$-$v_{8}$) implement the dependencies like "the smaller the value of the criterion, the more preferred is the contractor", and their estimates are expressed by "fuzzy" numbers. They are normalized by the formula:
where:

\[ M \text{ is the lower limit of the criterion (the minimum value of the criterion for the whole sample of outsourcers).} \]

For convenience of the convolution, it is suggested to represent the values of "crisp" estimates of criteria \( v_1 \text{-} v_6 \) by a fuzzy triangular number like \( (d,d,d,d_{i}) \), where «d» is a crisp value of the estimate.

The convolution result is also a triangular number expressing the integral estimate (rating) of a potential outsourcer. To select a potential outsourcer from the set of outsourcers, which were put in correspondence with the estimate at the previous stage, it is advisable to perform ranking using a triangular number. In this scientific work, we use the Jane method as a ranking method. According to this method, fuzzy estimates are normalized in accordance with the increase in the degree of their membership in a set of "big numbers" (15).

\[ Pos(r \in B) = \max_{\alpha} \min(r, \mu_{B}(v)) \]

where:

\[ B = (0,1,\infty,\infty) - "\text{big}" \text{ fuzzy number}; \]

\[ \mu_{B}(v) - \text{membership function.} \]

It is advisable to conclude a contract, within which all aspects of the interaction are clarified and the contract for the delivery of a project work is signed, with an outsourcer that get the highest-ranking position as a result of the ranking. If during the negotiation stage it is found that the interaction with this outsourcer is inappropriate, it is suggested to repeat the procedure with outsourcers that took the lower ranking positions until a positive result is achieved.

**Conclusion**

In the scientific work, a methodological guidance was developed to identify and technologically substantiate the priority areas for application of outsourcing in project development and support. It was suggested to introduce a fuzzy-logical model for attracting third-party contractors, which will allow to determine the conceptual possibility before
delivering project works based on the analysis of the level of information risk of project implementation and the technical and economic competence of outsourcing firms. To estimate the advisability of transferring specific project works to third-party contractors when using the direction of “selective outsourcing” or “integrated outsourcing”, an algorithm was developed to support an appropriate managerial decision based on the analysis and classification of project works.

A cognitive algorithm is proposed to select executors for specific works in project development and support, which is characterized by the availability of preliminary filtration of a set of possible project co-executors based on probabilistic topical modeling of unstructured text documents describing opportunities for participation of outsourcers and selection of the types of potential works. The application of this algorithm allows to take into account the variable importance of the criteria for evaluation of executors at different stages of project implementation under the conditions of deficiency and inconsistency of information caused by the of nature of its implementation.

References


