



Research Article

DeTreex Inductive Knowledge Acquisition System in Project Risk Assessment

Iwona Lapunka¹ and Iwona Pisz²

¹Opole University of Technology, Faculty of Production Engineering and Logistics, Department of Project Management, Opole, Poland

²Opole University, Faculty of Economics, Department of Logistics and Marketing, Opole, Poland

Correspondence should be addressed to: Iwona Lapunka; i.lapunka@po.opole.pl

Received date: 9 March 2015; Accepted date: 23 June 2015; Published date: 23 November 2016

Academic Editor: Ioana-Alexandra Bradea

Copyright © 2016. IwonaLapunka and IwonaPisz. Distributed under Creative Commons CC-BY 4.0

Abstract

The aim of this paper is to present the possibility of employing DeTreex inductive knowledge acquisition system in project risk assessment. DeTreex from AITECH Artificial Intelligence Laboratory is a system for knowledge acquisition implementing the inductive machine learning method. Induction in the knowledge acquisition system draws on the induction of decision trees proposed by Quinlan. DeTreex was used to carry out the analysis of data elaborated for the purposes of project risk assessment. The problem of acquisition of knowledge of potential project risk sources, the structuring and processing of such knowledge is, in the case of project management, a relatively new and not fully researched area of study. In the proposed approach, we aim at formulating certain rules in the relations between risk probability and its impact on the risk level of a project. The experience and knowledge of experts is required in order to elaborate effective decision rules. The representation of risk parameters in the form of IF-THEN rules constitutes the basis for a knowledge base in the architecture of the proposed expert system, dedicated to project risk assessment. Project risk assessment using an inductive knowledge acquisition system based on decision tree induction is an original approach to the problem.

Keywords: project risk management, inductive knowledge acquisition system, DeTreex, decision tree induction.

Introduction

Projects constitute an increasingly common type of business activity of individual companies, holdings, clusters and supply chains. The key goal of a project is to provide a particular product or service in compliance with adopted project constraints (Kerzner, 2009). Its aim is to fulfil the needs of stakeholders and thus it

should provide clients with a significant competitive edge. The ability to determine the tasks that must be accomplished, the estimation of time and costs of project execution, the choice of appropriate contractors, etc., constitute the basic problems faced in the project planning phase (Kerzner, 2013a). These quantities are defined in the presence of uncertainty (Knight, 1971). During project execution,

the degree of uncertainty in the planning of basic project parameters, i.e. time, costs, and scope, decreases with time. Uncertainty is inextricably linked with risk (Kallman, 2005; Knight, 1971). Practically speaking, the higher the uncertainty, the greater the risk. A substantial level of risk is one of the major features of projects – too many adverse events can result in project delays, excessive costs, unsatisfactory project results, or even lead to a complete failure of a project. Risks may have a negative or positive impact on a project (Kerzner, 2013b). The risk can be treated as a threat and a challenge (Hillson, 2002).

Risk is inherent in all projects undertakings. New and unique projects involve high risk. This is particularly true for projects characterized by a high degree of uncertainty, complexity and pace (Jaafari, 2001). The specificity of companies, along with their uniqueness leads to inherent risk. The degree of uncertainty varies, depending, among other factors, on the type, scope, size of a project, as well as on how it is carried out. Achieving success in project management requires the ability to control risk-bearing events and the operating conditions, which can influence the goals of a given project in a negative or positive manner. Project execution carries various opportunities and risks, for which a given company or supply chain should prepare and should be able to use them to its advantage. Risk is present in projects, mostly because they are innovative, unique, one-time actions and it is difficult to predict the direction of their execution under uncertain circumstances. In all elements of the project's environment whether closer or more distant, unexpected change can take place, which will directly or indirectly, and to a higher or lower degree, have bearing on the project. The level of risk grows exponentially with the extent of the planning horizon. The higher the uncertainties related to project outcomes, to not holding assumptions, expectations and formulated goals, the higher the project risk (Cooper et al., 2005). The probability, scope and impact of internal and external factors on the project also play a significant role (Elbrahimnejad et al.,

2010). Risk cannot be entirely eradicated. However, it can be kept to a minimum by taking appropriate actions, i.e. by introducing the guidelines, methods and mechanisms of project risk management into a company (Hillson, 2002; Kerzner, 2009). There are many different risk sources and some approaches have been suggested in the literature for classifying them. There are few sources of risks that face any projects. These risks include: schedule, scope, resources, quality, cost, technical, organizational, project management, customer, delivery (Elbrahimnejad et al., 2010).

The aim of this paper is to present the concept of project risk assessment using the DeTreex inductive knowledge acquisition system, based on decision tree induction introduced by Quinlan. The proposed approach can constitute the basis for building an expert system dedicated to the process of risk assessment of projects carried out in production and service enterprises, as well as in an entire supply chain. Hypothetically, it is assumed that gathered and structured expert knowledge of the relations between risk probability, risk impact on project activities and the risk level of each of these may facilitate the elaboration of effective rules for the knowledge base.

Project risk management

Project risk management involves decision-making and carrying out actions that lead to the achievement of an acceptable risk level by the project team. The knowledge of risk involved in the execution of a given project is one of the factors in project success (Lock, 2007). The ability to predict threats and accordingly preparing an effective contingency plan are key in project management. Risk management is focused on identifying and controlling events that can have a negative impact on the execution of a project; its main goal is to minimize the risk of project failure (WSDOT, 2014). Project risk management consists of several stages, as determined by the Project Management Institute (PMI, 2013). In practice, risk management is a process of identifying, analyzing and

assessing different types of risks, as well as of monitoring and controlling events that can affect a given project in a positive or negative manner. Risk management should be an integral part of project management (Jaafari, 2001). The main goal of risk management is to identify and assess the risk of a given project. The first step of risk management is risk identification. In this phase all potential risk sources are identified. Potential risk factors, which can have a particular impact on the project, are determined. A number of techniques of project risk identification are used in practice. Potential project risks can be determined through: brainstorming, checklists, questionnaires and interviews, the Delphi group method, cause-effect diagrams (Mojtahedi et al., 2010). Risk assessment constitutes the second stage of risk management. The main goal of project risk assessment is to measure the impact of risks identified for a particular project. To this effect, the following methods are employed: Event Tree Analysis (ETA), Faults Tree Analysis (FTA), the Monte Carlo method, Scenario Planning, Sensitivity Analysis, Expected Net Present Value (ENPV), Decision Tree, Program Evaluation and Review Technique (PERT), Estimations of System Reliability, Failure Mode and Effect Analysis (FMEA), and Fuzzy Set Theory (Elbrahimnejad et al., 2010; Nieto-Moroto, Ruz-Vila, 2011).

One of the terms that define risk and its impact on project management is the probability of the occurrence of an event, which would negatively impact a given

project goal. The quality and reliability of qualitative risk analysis requires defining different levels of risk probability and its impact. The general definitions of risk and impact levels are fine-tuned for an individual project in the course of risk management planning. Project risk assessment consists in the prioritization of identified risks in order to further analyze (usually through assessment and aggregation) the probability of their occurrence, as well as their impact. It is usually a complex and rather complicated process due to the associated uncertainty. Imprecise, incomplete, unobtainable and non-measurable information poses a fundamental difficulty. Under such circumstances, risk assessment can be at most approximate, and cannot be expected to yield precise estimates.

Proposed approach to project risk assessment

The proposed approach to project risk management involves several stages (Pisz, 2011). During the first stage, based on a work breakdown structure (WBS), risk factors are identified for each action within the project structure. WBS is a deliverable-oriented grouping of project elements, which defines and organizes the entirety of the project in order to achieve effective planning and comprehensive control (Fig. 1). The hierarchical structure defines tasks that can be carried out independently of other tasks, thus facilitating the allocation of resources and responsibilities, as well as project measurement and control.

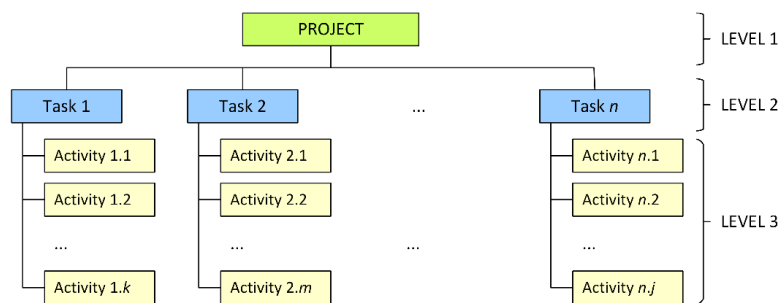


Figure 1: Work breakdown structure of a project

During the second stage, all factors that influence the execution of a project are defined by experts based on their knowledge and experience. Since each task/action can be affected by many different risks, all potential risks (R) and their sources (RS) along with identified risk factors (RF) should be accurately described and linked with a particular task or action in the WBS. Assessing project risk begins with the risk assessment of individual actions and tasks. The estimation of crucial data, i.e. the risk probability (RP) and risk impact (RI) of factors, usually begins with the use of linguistic terms, which are subsequently transformed into fuzzy numbers. Risk probability indicates the possibility of the occurrence of each

indicated type of risk. For each factor, risk probability and risk impact are defined based on the assumed five-level scale: very low $\in(0; 0.1)$, low $\in(0.1; 0.2)$, medium $\in(0.2; 0.4)$, high $\in(0.4; 0.8)$, and very high $\in(0.8; 1)$. The risk level of a project (RL of P) encompasses all potential risks that can affect project goals, i.e. costs, time, scope, and quality. Table 1 defines example negative effects, which can be used in the assessment of the impact of risk on the main project goals. Similar tables can be elaborated for positive effects. The table takes into account relative and numerical (here, nonlinear) approaches.

Table 1: Determining levels of impact of risk on the main project goals (Adapted from PMI, 2013)

Defined conditions for impact scales of a risk on major project objectives (Examples are shown for negative impacts only)					
Project objective	Relative or numerical scales are shown				
	Very low (0.05)	Low (0.10)	Moderate (0.20)	High (0.40)	Very high (0.80)
Cost	Insignificant cost increase	< 10% cost increase	10-20% cost increase	20-40% cost increase	> 40% cost increase
Time	Insignificant time increase	< 5% time increase	5-10% time increase	10-20% time increase	> 20% time increase
Scope	Scope decrease barely noticeable	Minor areas of scope affected	Major areas of scope affected	Scope reduction unacceptable to sponsor	Project end item is effectively useless
Quality	Quality degradation barely noticeable	Only very demanding applications are affected	Quality reduction requires sponsor approval	Quality reduction unacceptable to sponsor	Project end item is effectively useless

The third stage consists in transforming the linguistic risk parameters into fuzzy subsets and determining the appropriate membership functions for each input and output variable. Risk parameters are expressed as triangular and trapezoidal numbers mapped onto the set $(0, 1)$. The aim of this approach is to formulate certain rules governing the relations between risk probability (RP) and risk impact (RI) on the risk level of a project (RL of P). Expert experience and knowledge is required to elaborate effective decision rules. IF-THEN rules constitute the basis for a knowledge base in the architecture of the proposed expert system. Risk level (RL) is estimated for each project action/task, as well as for the entire project. RL is determined by the values of risk probability and the level of its impact. Risk level is defined on a five-level scale of risk: very low (VL), low (L), medium (M), high (H), and very high (VH), using a set of IF-THEN rules, for instance:

$$\begin{aligned}
 & \text{IF } (RI \text{ is } VH) \text{ AND } (RP \text{ is } L) \text{ THEN } RL \text{ is } H, \\
 & \text{IF } (RI \text{ is } M) \text{ AND } (RP \text{ is } L) \text{ THEN } RL \text{ is } M, \\
 & \text{IF } (RI \text{ is } VL) \text{ AND } (RP \text{ is } L) \text{ THEN } RL \text{ is } L.
 \end{aligned}
 \tag{1}$$

General definitions of probability and impact levels are fine-tuned for an individual project in the course risk management planning (Fig. 2). All factors that influence the execution of a project are

defined by experts, based on their knowledge and experience. Hence, among others, the need to acquire, structuralize and properly process knowledge during the entire course of project management.

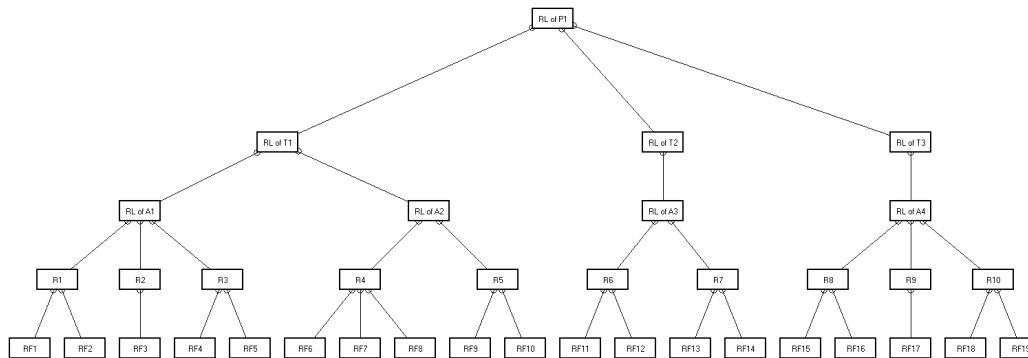


Figure 2: Hierarchical structure of project risk assessment

Project risk assessment using decision tree induction

The main problems that have distributed humanity at all times are the problems of origin and accuracy of knowledge. While the logic or science about how to think correctly and how to come to true conclusions was developed in the course of millenniums, the science of knowledge as the result of reasoning arose in our time within the framework of a scientific direction called 'artificial intelligence'. The use of computers in all spheres of our life requires studying and a fundamental understanding of how men extract knowledge from observations (Naidenova, 2010). The dynamic development of knowledge engineering creates great opportunities for constructing IT systems (termed expert systems) that could replace human experts in a given domain. Expert systems can be classified as: (1) expert advisory systems, which support decision-making, (2) expert control systems, which make decisions and (3) expert critiquing systems, whose main role is to analyze and interpret solutions. The increasing

complexity of decision-making and of problems that need solving in the contemporary world requires employing increasingly intelligent computer systems, which at the same time employ an increasing amount of computing power in order to process information contained in databases. Knowledge bases, which are capable of storing vast amounts of rules and facts, may thus be built. The aspects of their construction and the advantages of using intelligent systems for decision-making are particularly apparent to scientists in such fields as automatics, mathematics, industrial engineering, philosophy, pedagogy, and psychology. Nevertheless, the topic is steadily gaining popularity amongst scientists representing production engineering or management sciences.

Project risk assessment using an inductive knowledge acquisition system based on decision tree induction is an original approach to the problem. The process of knowledge acquisition has itself high significance due to the necessity of constructing a knowledge base, which

would then need to be verified in terms of its completeness, consistency and would undergo the elimination of redundant information. The proposed concept of an expert system employing knowledge acquisition (decision tree induction) is presented in Fig. 3. Expert system construction is a general concept of knowledge engineering. This pertains to both the methodology, and to the tools employed when building expert systems. Knowledge engineering encompasses the following processes: (1) acquiring knowledge from experts in a given field and structuring its representation, (2)

selecting inference methods appropriate for problems at hand and explaining their solution, (3) building suitable modules for the communication between the system and its user. The following basic capabilities are assumed for the proposed concept of expert system:

- problem analysis using rules,
- selection of facts contained in the knowledge base that are essential for problem analysis,
- explaining to the user how problem analysis has been carried out.

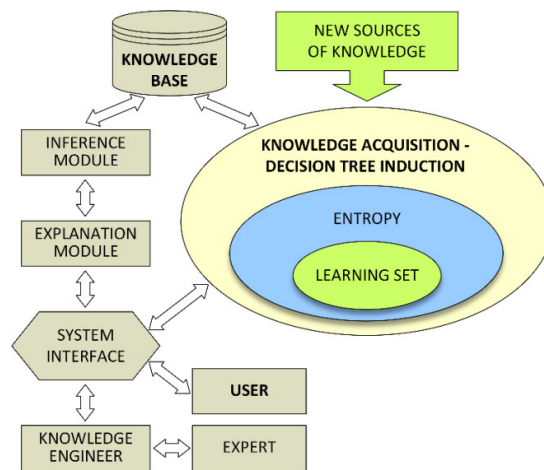


Figure3: Architecture of an expert system employing decision tree induction

Description of system components

A knowledge base in an expert system contains information and the entirety of knowledge required for solving a problem defined by the user, in the form of rules and facts stored in the base in a specific language for knowledge representation. Facts in the knowledge base determine the relations between certain objects and may be characterized by different features. Rules, on the other hand, are encountered in the majority of expert systems and assume the following form:

IF hypothesis THEN conclusion. (2)

For any expert system, implementing the widest possible spectrum of knowledge is of utmost significance, since the quality of the system depends more on the knowledge contained in the system

(knowledge base), than on the inference process the expert system employs. The acquired knowledge must be properly structured and, subsequently, processed in the system in order to achieve the assumed goal. The inference module constitutes another significant element of the entire system. Based on the knowledge collected in the knowledge base, it allows solving the problem by means of the inference procedures contained in the module. During the inference process, rule conditions are evaluated and actions corresponding to their conclusions are undertaken. In expert systems, inference is often carried out observing the rules of formal logic (*modus ponens*).

$$\frac{(P \Rightarrow Q), P}{Q} \quad (3)$$

A so-called explanation module should be implemented in the system as well. The role of this module is not only to present the inference results, but also to depict the entire inference process that leads to the conclusions drawn by the system. This allows to confirm if the system operates correctly, as well as to indicate whether or not the content of the knowledge base

should be corrected. Finally, the knowledge acquisition module is responsible for the process of acquiring and updating the knowledge recorded in the knowledge base. Procedures dedicated to the acquisition of new knowledge are implemented in this module (Fig. 4).

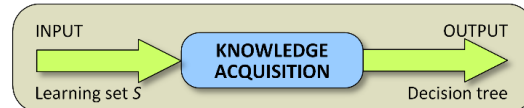


Figure4: Concept of operation of the knowledge acquisition module

Machine learning paradigm

Machine learning is the inference of unknown dependences between input data (independent variables, features, predictors) and output data (dependent data, response). The dependence between input (IN) and output (OUT) cannot, in general, be represented analytically or explicitly. Often it assumes the form of a so-called black box, operating according to a given algorithm acquired in the process of learning (Mitchell, 1997). In the case of data for which there exists no OUT, we talk about the inference of the probability of simultaneous occurrence of a certain

feature set. Decision trees are one of the most commonly employed methods of supervised learning. In supervised learning a certain data set, the so-called learning set, is specified, with input variables and the corresponding output variables known for members of the set. Based on this, implicit relations, linking IN and OUT are sought for the purpose of predicting OUT for data absent from the learning set. The proposed model is a method for finding such relations (using a suitably large learning set) – a machine learning method for the problem of project risk assessment (Fig. 5).



Figure 5: Inference model for the problem of project risk assessment

In order to carry out the analysis of data elaborated for the purposes of project risk assessment, we employed the machine learning induction. This approach allows representing decision trees in the form of rules, which can subsequently be used to build the knowledge base for the proposed expert system.

Induction in the knowledge acquisition system draws on the induction of decision trees proposed by Quinlan (1986). A decision tree is a tree structure, whose

each node represents a test result for one attribute value, whereas each leaf corresponds to a decision. A decision tree is grown by recursively dividing the dataset until each subset contains the elements of only one decision class (Liu, Lee, 2008). The top of the tree is termed the root, whereas each subsequent node is responsible for particular test of an attribute. Evaluating the test leads, through a tree edge, to the next node. At the bottom of a decision tree there are leaves corresponding to decisions. A decision tree

thus grown can be termed a knowledge model for explaining the structure of the knowledge implicit in a dataset, termed a learning set. The quality of the obtained decision tree is assessed by having it analyze cases contained in another set, the testing set. The decision-tree induction is employed mainly for object classification, while the generated knowledge, through the use of an expert system operation, aids decision-making. The major problem during the growing of a decision tree is determining a criterion for choosing the attribute for the root (Michalski, 1983). For this purpose, the so-called entropy is employed (Quinlan, 1986; Quinlan, 1996):

$$Entropy(S) = - \sum_{i=1}^{|S|} \frac{|S_i|}{|S|} \cdot \log_2 \left(\frac{|S_i|}{|S|} \right) \quad (4)$$

where:

$|S_i|$ – number of examples describing i -th object,

$|S|$ – number of examples in the learning set S .

The expected information value after the set of examples S is divided into subsets S_v , for which attribute A assumes the value v , is given by:

$$Gain(S, A) = Entropy(S) - \sum_{v \in A} \frac{|S_v|}{|S|} \cdot Entropy(S_v) \quad (5)$$

where:

v is the number of all possible values of attribute A ,

S_v – subset of set S , for which attribute A assumes the value v ,

$|S_v|$ – number of examples in set S_v ,

$|S|$ – number of examples in the learning set S .

Example implementation using DeTreex

DeTreex, an inductive knowledge acquisition system, was used to carry out the analysis of data elaborated for the purposes of project risk assessment. DeTreex is a system for knowledge acquisition implementing the inductive machine learning method. In addition, the system allows representing decision trees in the form of decision rules, which can be used in the knowledge base of the proposed expert system.

For the purposes of growing a decision tree, the set of learning examples must be saved in the learning file. Whereas, in order to test the accuracy of the classification of the tree a file containing the set of testing examples needs to be prepared. The file formats are identical, only the file extensions differ: learning file (*.lrn), testing file (*.tst). In the example under study the following areas of the problem were defined – attributes, such as critical and non-critical actions, risk probability, risk impact and their values, which determine risk level (Table 2). Due to the substantial size of the analyzed problem, the presented example of project risk assessment was considered only in terms of the time criterion.

Table 2: Determination of problem area – attributes and their values

B				
Critical activity	Non-critical activity	Risk probability	Risk impact	Risk level
Insignificant time increase	Time increase in limit of reserve	Very low $\in (0; 0.1)$	Very low $\in (0; 0.1)$	Very low risk
< 5% time increase	< 5% time increase over limit of reserve	Low $\in (0.1; 0.2)$	Low $\in (0.1; 0.2)$	Low risk
5-10% time increase	5-10% time increase over limit of reserve	Moderate $\in (0.2; 0.4)$	Moderate $\in (0.2; 0.4)$	Moderate risk
10-20% time increase	10-20% time increase over limit of reserve	High $\in (0.4; 0.8)$	High $\in (0.4; 0.8)$	High risk
> 20% time increase	> 20% time increase over limit of reserve	Very high $\in (0.8; 1)$	Very high $\in (0.8; 1)$	Very high risk

The following examples of events which may impact project execution in a given time horizon, were identified by experts as potential time-related risk factors:

- change of requirements in the course of project execution, significant elements not being taken into account, change of scope,
- incorrect assumptions or lack of understanding of project requirements,
- incorrect design assumptions, technological assumptions, construction brief, planning assumptions, and technical-organizational assumptions at each stage of project execution,
- delays in elaborating the construction, technologies, plans and schedules of project execution,
- insufficient amount of castings and forgings – long-term unavailability of main shapes,
- manufacturing errors – necessity of repairing or remaking certain shapes,
- delays in ordering products due to delays in elaborating construction, etc.,
- insufficient funds to cover immediate expenses,
- delays in component supplies, through a fault of suppliers or supply/logistics departments, at each stage of project execution, due to untimely payment by a contractor (e.g. lack of advance

payment at the commencement of contract), due to project changes or random incidents,

- discrepancies between material/component supplies and orders, damage to components,
- assembly errors,
- organizational problems,
- lack of suitable personnel (employees insufficiently trained or lacking experience), employee absenteeism, lack of commitment/motivation,
- unavailability of key employees during the execution of vital tasks,
- communication problems in the team,
- underestimation of labor intensity of the project.

For the purposes of generating a decision tree it is necessary to save a set of examples in a learning file (*.lrrn). Such a file can be prepared in a text editor or with a spreadsheet. Whereas, in order to verify the accuracy of classification of the tree, a test file (*.tst) containing a set of examples was prepared. Four basic attributes were defined, serving as the basis for project risk analysis in terms of time: (1) *critical_activity*, (2) *time_increase*, (3) *risk_probability*, and (4) *risk_impact*.

Table 3: Examples of risk level assessed by experts for critical activities

		Insignificant					5-10% time increase									
RI	VH	M	H	H	VH	VH	VH	H	H	VH	VH	VH	VH			
	H	M	M	H	H	VH	H	M	H	H	VH	VH	VH			
	M	L	M	M	H	H	M	L	M	M	H	VH	VH			
	L	VL	L	M	M	H	L	VL	L	M	H	H	H			
	VL	VL	VL	L	M	M	VL	VL	VL	L	M	H	H			
		VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH
		RP					RP									
		< 5% time increase					10-20% time increase									
RI	VH	H	H	H	VH	VH	VH	H	H	VH	VH	VH	VH			
	H	M	H	H	H	VH	H	M	H	H	VH	VH	VH			
	M	L	M	M	H	H	M	L	M	M	H	VH	VH			
	L	VL	L	M	H	H	L	VL	L	M	H	H	H			
	VL	VL	VL	L	M	H	VL	VL	VL	L	M	H	H			
		VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH
		RP					RP									

An excerpt from historical data, gathered based on the risk level assessed by experts

for individual project activities, is presented in Table 3 and Table 4.

Table 4: Fragment of historical data for risk level assessment in a learning file

No.	Critical activity	Time increase	Risk probability	Risk impact	Risk level
1	Yes	Insignificant	0.9	0.2	High risk
2	Yes	< 5%	0.05	0.6	Moderate risk
3	Yes	5-10%	0.9	0.5	Very high risk
4	Yes	10-20%	0.05	0.05	Very low risk
5	Yes	10-20%	0.3	0.05	Low risk
6	Yes	> 20%	0.3	0.5	High risk
7	Yes	> 20%	0.9	0.8	Very high risk
8	No	In limit of reserve	0.6	0.05	Moderate risk
9	No	< 5% over limit of reserve	0.05	0.05	Very low risk
10	No	5-10% over limit of reserve	0.1	0.7	High risk
...	No	10-20% over limit of reserve	0.2	0.5	High risk
110	No	> 20% over limit of reserve	0.05	0.1	Very low risk

The experience and knowledge of experts are necessary for elaborating the IF-THEN rules. In order to determine project risk level, a set of rules stored in a knowledge

base is used. Figure 6 presents the selected examples of rule representations in DeTreeX from AITECH Artificial Intelligence Laboratory.

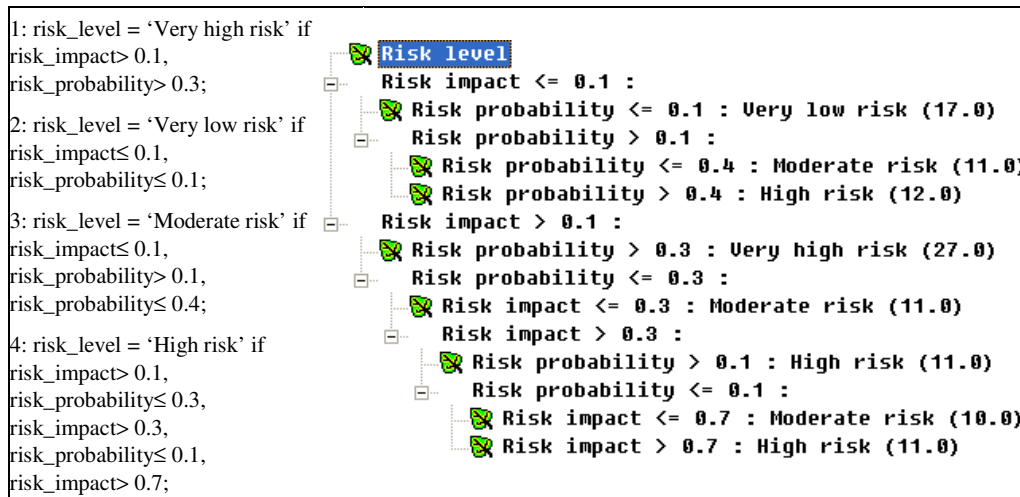
**Figure 6: Examples of rules for risk level assessment of project activities with time as a criterion**

Figure 7 presents a decision tree grown based on 110 historical risk assessments of project activities and outlines simplified relations between risk probability (RP),

risk impact (RI), and risk level (RL) $\in \{VL, L, M, H, VH\}$.

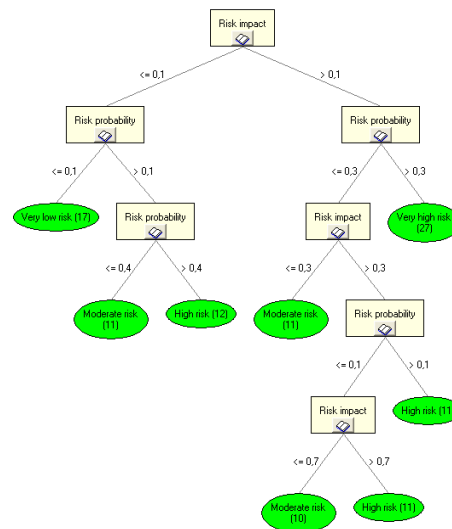


Figure 7: Decision tree for risk level assessment of project activities with time as a criterion

The value of entropy for the entire learning set, calculated in accordance with relations (4) and (5), indicated that the *RI* (*risk_impact*) attribute ensures maximum information increase in a node, leading to the choice of *RI* for the tree root. As a result, we obtained a tree with 8 leaves, represented as specific rules of knowledge representation.

$$Entropy(S) = 2,163 \quad (6)$$

$$\begin{aligned} Gain(S, critical_ & Gain(S, risk_pr \\ = 0,013 & = 0,067 \end{aligned} \quad (7)$$

$$\begin{aligned} Gain(S, time_inc & Gain(S, risk_im \\ = 0,135 & = 0,214 \end{aligned}$$

Decision tree induction is chiefly used for the classification of objects (problems), whereas the knowledge generated by inductive methods, and ultimately by means of an expert system, aids decision-making. A set of rules is significantly smaller than the set of data, from which the rules were created. The rules also aim to generalize the data, so as to enable the classification of new cases. The created sets of rules can be used to construct a knowledge base of an expert system dedicated to project risk assessment.

The machine learning paradigm for project risk assessment is not free of typical problems of machine learning. The issues that require particular attention and

warrant further research are, among others: (1) choice of an optimal method for determining the model, (2) suitable selection of significant features (attributes) of learning – elimination of redundant features, (3) suitable selection of an optimal and representative learning set, as well as (4) evaluation function for judging the acceptability of the result (expansion of tree-based induction by a testing set). The verification of knowledge obtained based on the test file, and pruning the decision tree branches according to a chosen parameter constitutes significant directions of further research. The role of the aforementioned parameter is to aid the determination of suitable error classification before and after the tree is pruned. If the error classification level after pruning the tree (or a node) is higher than the level prior to pruning the tree (or a node), pruning is not performed. The aim of subsequent studies will be to minimize error classification for the generated set of rules in a knowledge base.

Conclusion

Contemporary companies and supply chains operate in a turbulent and uncertain socio-economic environment. In the face of changes taking place in the surroundings, increasing numbers of practitioners and theoreticians are drawn to new concepts of

management, and particularly to integrated approaches. We believe that the consolidation of approaches to project management and knowledge management is the answer to the challenges faced by contemporary projects. At present, the project approach is indispensable in the execution of the majority of ventures. For management processes carried out in the spirit of this approach, one can observe, among other things, a team approach to project time management, as well as proactive and future-oriented reporting on the advancement of project. Since projects are developed progressively – at their commencement they are characterized by a high level of uncertainty. This stems from the fact that they deliver innovative value, and therefore are burdened with a significant level of uncertainty. Three types of uncertainty are differentiated in project planning and scheduling: the uncertainty in the time it takes to complete tasks, the uncertainty in temporal correlation and the uncertainty associated with resources. With so many unknowns at the beginning of execution, all estimates and assumptions may soon lose validity; for instance, it may take a week and not a day to complete a certain task, a product may cost twice as much, a novel technology may turn out to be more effective than it was assumed, or a new supplier may prove negligent.

The risk scale guidelines for achieving major project parameters, defined by PMI in the PMBOK Guide (2013), explicitly point to the significance of the time criterion for the success of every given project. Even the ability to determine risk factors which could result in the project deadline being exceeded by a small margin, is estimated at 0.05 on a $(0, 1)$ scale during project risk assessment. The risk of project failure, initially deemed to be very low, will gradually increase, depending on subsequent estimations. The probability of exceeding the deadline by as little as 5% necessitates assessing risk at the level of 0.1 (low), and, respectively, exceeding the deadline by 5-10% corresponds to medium risk (0.2); 10-20% – high risk (0.4), above 20% – very high risk (0.8). It must be observed that these data pertain only to risk factors influencing the

timeliness of project completion. Indicators for other project goals, such as cost, scope or quality, have not been taken into account here, whereas under real-life conditions the risk scales undergo multiplication with each possible factor taken into account.

The problem of knowledge representation is an important issue for expert systems dedicated to project risk assessment. Defining the relations between risk factors, the probability of their occurrence, their impact at the input, and the project risk level at the output of the inference model forms the basis for elaborating effective rules in the knowledge base of an expert system. The proposed approach can aid project managers in decision-making under the conditions of uncertainty, helping them comply with operating requirements and achieve project success. Due to the significance of project risk assessment to every stakeholder, the rule representation of expert knowledge is particularly important from the perspective of: (1) project team members, because it aids in identifying events that endanger project completion as planned and offers the means of effective monitoring; (2) end-users, since it contributes to fulfilling their needs and achieving an appropriate value to price ratio for employed means and resources; (3) suppliers and contractors, because a reasonable approach to risk in projects leads to better planning and improved results, both for sellers and buyers; (4) financial institutions, which offer loans for financing projects, as the conditions of the loan depend on risk; and (5) insurers, who require that the risk of failure is known, monitored and properly managed in the scope of a project, so as not to finance existing risks.

References

1. Cooper, D.F., Grey, S., Raymond, G. and Walker, P. (2005), Project risk management guidelines. Managing risk in large projects and complex procurements, John Wiley & Sons, Chichester.
2. Elbrahimnejad, S., Mosavi, S.M. and Seyrafiapour, H. (2010), 'Risk

- identification and assessment for build-operate-transfer projects: a fuzzy multi attribute decision making model', *Experts Systems with Applications*, 37, 575-586.
3. Hillson, D. (2002), 'Extending the risk process to manage opportunities', *International Journal of Project Management*, 20 (3), 235-240.
 4. Jaafari, A. (2001), 'Management of risk, uncertainties and opportunities on projects: time for a fundamental shift', *International Journal of Project Management*, 19, 89-101.
 5. Kallman, J. (2005), 'What is risk?', *Risk Management*, 52, 10.
 6. Kerzner, H. (2009), *Project management case studies*, John Wiley & Sons, New Jersey.
 7. Kerzner, H. (2013a), *Project management. A system approach to planning, scheduling, and controlling*, John Wiley & Sons, New Jersey.
 8. Kerzner, H. (2013b), *Project management. Metrics, KPIs and dashboards. A guide to measuring and monitoring project performance*, John Wiley & Sons, New Jersey.
 9. Knight, F.H. (1971), *Uncertainty and profit*, Cambridge.
 10. Liu, C.L. and Lee, C.H. (2008), 'Simplify multi-valued decision trees', *Proceedings of the 3rd International Symposium on Advances in Computation and Intelligence*, LNCS 5370, 581-590.
 11. Lock, D. (2007), *Project management*. 9th edition, Gower Publishing Ltd, Hampshire.
 12. Michalski, R.S. (1983), 'A theory and methodology of inductive learning', *Artificial Intelligence*, 20, 111-161.
 13. Mitchell, T. (1997), *Machine learning*, McGraw-Hill, New York.
 14. Mojtahedi, S.M.H., Mousavi, S.M. and Makui, A. (2010), 'Project risk identification and assessment simultaneously using multi-attribute group decision making technique', *Safety Science*, 48, 499-507.
 15. Naidenova, X. (2010), *Machine learning methods for commonsense reasoning processes: interactive models*, Information Science Reference, Hershey-New York.
 16. Nieto-Moroto, A., Ruz-Vila F., (2011), 'A fuzzy approach to construction project risk assessment', *International Journal of Project Management*, 29, 220-231.
 17. Pisz, I. (2011), 'Project risk assessment using Fuzzy Inference System', *Logistics and Transport*, 2 (13), 25-34.
 18. PMI Standards Committee (2013), *A guide to the project management body of knowledge*. 5th edition, Project Management Institute, Newtown Square.
 19. Quinlan, J.R. (1996), 'Improved use of continues attributes in C4.5', *Journal of Artificial Intelligence Research*, 4, 77-90.
 20. Quinlan, J.R. (1986), 'Induction of decision trees', *Machine Learning*, 1, 81-106.
 21. Washington State Department of Transportation, WSDOT (2014), *Project risk management guide*, Strategic Analysis and Estimating Office, Olympia.