Rough Sets Based Model as Project Success Support

Makitan Vesna*, Brtka Vladimir*, Brtka Eleonora*, Ivkovic Miodrag*
*University of Novi Sad/Technical Faculty "Mihajlo Pupin", Zrenjanin, Republic of Serbia
vesna@tfzr.uns.ac.rs, vbrtka@tfzr.uns.ac.rs, misa.ivkovic@gmail.com

Abstract—**This paper describes rough sets based model application during the entire project life cycle, starting from its initial phase, through realization and finally to closing. The main purpose of this model is data analysis support to successful project realization. Data analysis is based on the network diagram of particular project and implies calculation of a "path value" according to the selected weight, such as path's: duration, cost, scope, etc. For a chosen weight every project activity may has one of the three possible estimates. After selecting the set of estimates, model's heuristics finds a value of every path in the project network diagram and evaluates possible scenario of project realization. Model implementation may improve project realization, especially in projects with higher level of uncertainty, such as research and development ones.**

I. Introduction

The meaning of the project management concept is to help people run successful projects, those which are accomplished in a projected scope, time and cost and/or their sponsor is satisfied. That is why a lot of project management standards, methodologies, techniques, tools and other guidelines are available nowadays. Even more, the new ones are approaching offering guaranties of project success.

This paper represents one of the new approaches to successful project realization. It is the model that includes heuristics as a data analysis support to the project realization. In every phase of the project life cycle this model enables calculation of the "path value" according to the project duration, cost, scope or some other significant constraint. In that way, a project manager has an insight in the possible project scenarios caused by the most rigid constraint at specific moment.

Nowadays most of the project realizations are followed by high level of uncertainty caused by new technologies or the lack of resources, or raising risks, etc. In those cases activities that were not critical become critical and may cause project failure. Good project management should include tools and techniques that deal with various scenarios of project realization.

The model described in this paper enables prediction of "good" or "bad" paths in the project network diagram suggesting which one may become "critical" according to its duration, cost, scope, quality, etc.

As it will be seen in the next chapter, there are not so many papers related to this topic. Most of them are dealing with scheduling problems, heuristics based on experience or template usage. This model enables heuristics implementation in a wider range, including project time, cost, duration or any other significant constraint for which available values exist.

II. Related Work

There are a lot of examples of implemented heuristics in some areas of project management given by users, researches and software companies. The most used ones or so called "rules of thumb", project managers apply when planning and overseeing project tasks in order to respond to the complex environment [1]. Those heuristics are based on experience and should be applied and help in decision making when certain problem occurs [2, 3, 4].

The other heuristics implementation is mostly related to scheduling problems [5, 6, 7]. Useful categorization of a large number of heuristics for the resource-constrained project scheduling problem, their evaluation in a computational study and comparison with the proposed solution may be found in [8].

Other researches deal with application of heuristics in project duration assessment, such as [9] that showed several methods for coping with uncertain estimates of duration: PERT, fuzzy theory, and probabilistic computations.

Implementation of rough sets in project management may be found in [10]. This research describes so called critical success factors and the analysis of IT project according to these criteria. Rough sets are used to improve this analysis and increase IT projects success rate.

There is a textbook about some rough sets methods in IT project management [11]. It describes concepts of rough set theory that may be applied to the management of information technology implementation projects in order to assist project managers to place their emphasis on the "important" aspects of the project.

In [12] the principle and step of performance evaluation of project management based on rough set method are studied. Rough sets are used for project risk assessment, too [13, 14, 15].

The rough set theory based model in [16] is proven to be the most appropriate for data evaluation in many domains. The model is formed by numerous If-Then expressions which are readable and easy to understand. This model is frequently used in domains of: medicine, economy, management, education, etc.

Software solutions for project management support have many features and some of them enable evaluation of project duration, costs, resource availability, etc. then creation of "what if" scenarios, PERT analysis, template usage, etc. In that way, they provide heuristics implementation as well.

III. THE MODEL DESCRIPTION

In order to explain basic features of the model simple example of the project network diagram is chosen and presented in the Figure 1. It shows project activities (A, B, C,…, S), its events (1, 2, 3,…, 15) and relations. Possible value assigned to activities may be related to its: duration, cost, scope, quality, risk, etc. This value may be one of the three-element set contained of: optimistic, modal and pessimistic estimate. By choosing one of these values for every activity on the network diagram possible scenario of project realization is created. Existing paths of network diagram are then evaluated by the expert, which means that, every one of them is marked as a “good” or a “bad” one for a project realization. Bad means that path has activities, that could cause project failure and project managers should pay especial attention on those activities.

For model presentation duration is chosen to be constraint (weight) for which path value will be calculated. As it was said earlier every activity has PERT-like duration estimation, its: optimistic, modal and pessimistic time. Rough sets based part of the model calculates path value according to the chosen project scenario. It includes possible paths in the network diagram, its activities and for every one of them, one duration time assigned from the predefined set of values: optimistic, modal and pessimistic. In this case, when duration represents the most rigid constraint (selected weigh), critical/subcritical paths should be recognized as “bad” ones.

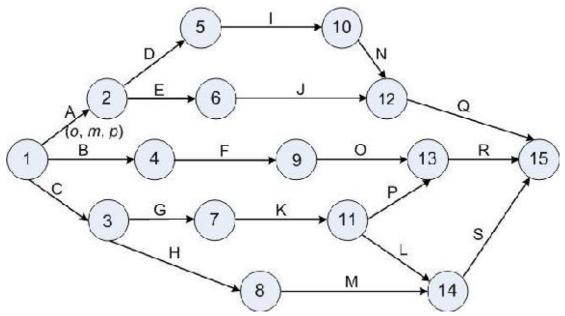

Figure 1. Network diagram of a sample project

Table 1 represents paths in the network diagram of a sample project with assigned value for every activity.

TABLE I.
POSSIBLE PATHS IN A SAMPLE PROJECT NETWORK DIAGRAM

Activities / Paths	A1	A2	A3	A4	A5	Decision
P1	1	2	3	1	3	1
P2	2	2	3	3	0	1
P3	2	2	3	3	0	2
P4	3	2	1	2	2	2
P5	1	1	2	3	1	1
P6	3	1	3	3	0	2

There are six possible paths in the network diagram. For example, path P1 contains activities: A, D, I, N and Q (denoted by A1, A2…) with durations: 1, 2, 3, 1 and 3, where 1 represents optimistic time, 2 modal and 3, pessimistic duration. Path P2 has only four activities: A, E, J and Q and that is why A5 has zero value. Decision about path value is given in the last column according to the expert judgment. Path may be a “good” one with value 1 or “bad” with value 2. Those values are assigned according to the path activities, selected durations for chosen scenario and other features known to the expert. Other features of the model are:

- Number of paths is arbitrary;
- Number of activities on the path is arbitrary (number of columns is equal to the number of activities on the longest path);
- Decision column may have more than two outputs. Any activity weight could be expressed by real number.

All of that enable project estimates in a wide range of possible situations and should improve decision making during the entire project life cycle.

IV. ROUGH SETS BASED PART OF THE MODEL

The Rough Set Theory (RST) is relatively new, it was introduced by Zdzislaw Pawlak in the early 1980`s [17, 18]. Basis of the RST is relation called indiscernibility relation. This relation is the mathematical basis of the RST, meaning that some information is associated with every object of the universe of discourse (data, knowledge) [18, 19]. The objects characterized by the same information are indiscernible in view of the available information about them. As mathematical basis for RST indiscernibility relation must be formally defined. First of all this relation is equivalence relation, it is reflexive, symmetric and transitive. As in [20], let U be a finite set of objects (universe), $Q=\{q_1,q_2,...,q_m\}$ is a finite set of attributes, V_q is the domain of attribute q and $V=\bigcup_{q\in Q}V_q$. The RST information system is the 4-tuple $S=\langle U,Q,V,f\rangle$ where $f=U\times Q\rightarrow V$ is a function such that $f(x,q)\in V_q$ for each $q\in Q, x\in U$, called information function. To every non–empty subset of attributes P is associated indiscernibility relation on U , denoted by I_P :

$$I_P=\{(x,y)\in U\times U : f(x,q)=f(y,q), \forall q\in P\} \qquad (1)$$

The family of all equivalence classes of the I_P is denoted by $U|I_P$, and class containing an element x by $I_P(x)$. The indiscernibility relation induces a partition of the universe into blocks of indiscernible objects called elementary sets. The information about the real world is given in the form of a decision system. The decision system is RST information system, but attributes which describe an object are divided so they form two sets: the set of condition attributes and the set of decision attributes, although there is usually one attribute called decision attribute while all other attributes are called condition attributes. So, the RST information system with defined set of condition attributes and decision attribute is called decision system. Table I represents the RST decision system for network diagram of a sample project. There are five condition attributes (A1, …, A5), and one binary decision attribute D. RST deals with the set of approximations: let X be a non–empty subset of U, and $\emptyset\neq P\subseteq Q$ (P is a non–empty subset of condition attributes). Set X is approximated by P–lower (2) and P–upper (3) approximations of X:

$$\underline{P}(X) = \{x \in U : I_P(x) \subseteq X\} \quad (2)$$

$$\overline{P}(X) = \bigcup_{x \in X} I_P(x) \quad (3)$$

The P–boundary of X is denoted by $Bn(X)$:

$$Bn(X) = \overline{P}(X) - \underline{P}(X) \quad (4)$$

If some object x belongs to lower approximation of *X*, it is certainly an element of *X*, but if *x* belongs to boundary region of *X*, then there is a probabilistic measure that *x* may belong to the set *X*. If this is a case, then nothing can be said with certainty about its belonging to the set *X*.

According to Table I, the approximation of the set *X*={P1, P2, P5}, i.e. the set *X* contains those objects (paths) for which decision is 1 (“good”), is given by: Lower approximation {P1, P5}, Boundary region {P2, P3}, Upper approximation {P1, P2, P3, P5}.

So, the set *X* is approximated, paths P1 and P5 are certainly “good”, but we are not sure about path P2. The exact reason why path P2 is the element of boundary region and, consequently upper approximation, is that there exists path P3 with same values of condition attributes while the value of decision attribute differs. Therefore, according to values of condition attributes of P2 and P3 we are not sure if this kind of path is “good” or “bad”. Set approximation enables the synthesis of the decision rules in the If-Then form: $\alpha \Rightarrow \beta$, where α is the antecedent of the rule while β is the consequent. As in [18] each object *x* that belongs to a decision system determines one decision rule:

$\bigwedge_{a \in C} a = a(x) \Rightarrow \bigwedge_{d \in D} d = d(x)$, here $a(x)$ stands for the value of attribute a of an object x. The expression $a = a(x)$ is called descriptor. If there is one decision attribute *d* we have: $\bigwedge_{a \in C} a = a(x) \Rightarrow d = d(x)$.

Based on lower approximation of set *X*, two rules are synthesized:

1. A1=1 ∧ A2=2 ∧ A3=3 ∧ A4=1 ∧ A5=3 ⇒ D=1, supported by {P1}
2. A1=1 ∧ A2=1 ∧ A3=2 ∧ A4=3 ∧ A5=1 ⇒ D=1, supported by {P5}

Based on boundary region of set *X*, one rule is synthesized:

3. A1=2 ∧ A2=2 ∧ A3=3 ∧ A4=3 ∧ A5=0 ⇒ D=1 ∨ D=2, supported by {P2, P3}

As it can be seen from this rule, logic of the RST accepts possibility of both values 1 and 2, for the same set of activities durations. This feature emulates real situations logic.

Rules for which the value of decision attribute is certainly 2 (“bad”) are synthesized according to “negative region” of set *X* that contains paths for which we are certain that the value of decision attribute is 2 (“bad”):

4. A1=3 ∧ A2=2 ∧ A3=1 ∧ A4=2 ∧ A5=2 ⇒ D=2, supported by {P4}
5. A1=3 ∧ A2=1 ∧ A3=3 ∧ A4=3 ∧ A5=0 ⇒ D=2, supported by {P6}

V. Model Evaluation

The data set presented in the Table I is used as a training set. After data analysis, previously mentioned five rules are synthesized. Rule synthesis is done by software system Rosetta – A Rough Set Toolkit for Analysis of Data. Rosetta system was developed as a cooperative effort involving the Knowledge Systems Group, Department of Computer and Information Science at NTNU, Norway, and the Logic Group, Institute of Mathematics at Warsaw University, Poland.

After the rule set creation, test set is prepared in order to evaluate the model. Test set is given in Table II.

TABLE II.
TEST SET FOR MODEL EVALUATION

Activities / Paths	A1	A2	A3	A4	A5	Decision
P1	1	2	3	1	3	1
P2	2	2	3	3	0	1
P3	2	2	3	3	0	2
P4	3	2	1	2	2	2
P5	1	1	2	3	1	1
P6	2	3	3	3	0	2
P7	2	2	1	1	3	1
P8	2	3	3	3	0	2
P9	2	2	3	3	3	2
P10	3	2	3	2	2	2
P11	2	1	2	3	1	1
P12	2	2	3	3	3	2
P13	1	2	3	1	1	1
P14	2	1	2	2	1	1
P15	2	2	3	3	1	2
P16	3	2	1	2	2	2
P17	1	2	1	2	1	1
P18	2	3	3	3	3	2
P19	1	2	3	1	1	1
P20	2	2	3	3	2	2
P21	2	2	3	3	0	1
P22	3	2	1	2	1	1
P23	1	1	1	3	1	1
P24	2	3	2	3	2	2

As test set contains 24 paths, while training set contains only six paths, the goal of this evaluation is to verify functionality of the model. Synthesized rules act as a classifier, five rule synthesized by training set classify test set paths. The performance of the classifier is measured by confusion matrix *C:* $|V_d| \times |V_d|$ matrix, where V_d is a set of possible values of decision attribute (“good” or “bad”). This matrix with integer entries summarizes the performance of rule set while classifying the set of paths. Entry: $C_{i,j} = \left|\{x \in U : d(x) = i, \overline{d}(x) = j\}\right|$, where $d(x)$ is the actual decision and $\overline{d}(x)$ is the predicted decision, counts the number of paths that really belong to class *i*, but were classified to class *j*. It is desirable for the diagonal entries to be as large as possible. The results are show in Figure 2. This means that four paths that are actually “good” are correctly classified as “good”, while two paths that are actually “bad” are correctly classified as “bad”. One path that is actually “bad” is classified as "good", while 17 paths were not classified at all. This clearly shows two things:

Manual

		Predicted			
		1	2	Undefined	
Actual	1	4	0	8	0.333333
	2	1	2	9	0.166667
	Undefined	0	0	0	Undefined
		0.8	1.0	0.0	0.25
ROC	Class	Undefined			
	Area	Undefined			
	Std. error	Undefined			
	Thr. (0, 1)	Undefined			
	Thr. acc.	Undefined			

Figure 2. The confusion matrix after test set classification by five rules without reduct sets computation

1. The training set is too small because there are 17 unclassified paths. This was expected because there are six paths in training set while there are 24 paths in test set.
2. The model evaluation considered as successful, there is only one misclassified path.

The evaluation showed that it is possible to use this kind of model to classify project network diagram paths, although the results have not been good enough so far.

In order to achieve better classification power it is possible to keep only those condition attributes of the paths that preserve indiscernibility relation and, consequently set approximation. This set of attributes is called reduct set. The rejected attributes are redundant since their removal cannot worsen the classification. Let P be a non-empty sub-set of condition attributes: $\emptyset \neq P \subseteq Q$, and a is some attribute from set P: $a \in P$. Attribute a is redundant (superfluous) in P if $I_P = I_{P-\{a\}}$. In other words, if attribute a is excluded from the set P while indiscernibility relation I, defined by (1) stays unchanged, then attribute a is redundant and can be omitted. Rosetta software system is capable of reduct set calculations. These calculations are often very complex, but in many practical applications it is not needed to calculate all the reducts, but only some of them. Since the training set is very small it is possible to employ exhaustive reduct sets calculation algorithm, which is implemented in Rosetta. This algorithm will calculate all possible reduct sets. Following reduct sets were calculated by exhaustive algorithm: R_1={A1}, R_2={A2, A3, A4}, R_3={A2, A5}.

Now, it is possible to synthesize rules according to reduct sets R_1, R_2 and R_3. Following 13 rules were synthesized:

1. A1=1 $\Rightarrow$ D=1
2. A1=2 $\Rightarrow$ D=1 OR D=2
3. A1=3 $\Rightarrow$ D=2
4. A2=2 $\wedge$ A3=3 $\wedge$ A4=3 $\Rightarrow$ D=1 $\vee$ D=2
5. A2=2 $\wedge$ A5=4 $\Rightarrow$ D=1 $\vee$ D=2
6. A2=2 $\wedge$ A3=3 $\wedge$ A4=1 $\Rightarrow$ D=1
7. A2=2 $\wedge$ A3=1 $\wedge$ A4=2 $\Rightarrow$ D=2
8. A2=1 $\wedge$ A3=2 $\wedge$ A4=3 $\Rightarrow$ D=1
9. A2=1 $\wedge$ A3=3 $\wedge$ A4=3 $\Rightarrow$ D=2
10. A2=2 $\wedge$ A5=3 $\Rightarrow$ D=1
11. A2=2 $\wedge$ A5=2 $\Rightarrow$ D=2
12. A2=1 $\wedge$ A5=1 $\Rightarrow$ D=1
13. A2=1 $\wedge$ A5=4 $\Rightarrow$ D=2

It is obvious that rules are synthesized according to three reduct sets, so some rules contain attribute A1 in the IF part, some rules contain attributes A2, A3 and A4, while some rules contain attributes A2 and A5. Figure 3 shows the classification power of 13 rules while classifying test set. The results are shown by confusion matrix.

Exhaustive

		Predicted		
		1	2	
Actual	1	11	1	0.916667
	2	8	4	0.333333
		0.578947	0.8	0.625
ROC	Class	Undefined		
	Area	3.402820e+038		
	Std. error	3.402820e+038		
	Thr. (0, 1)	3.402820e+038		
	Thr. acc.	3.402820e+038		

Figure 3. The confusion matrix after test set classification by 13 rules with exhaustive reduct sets computation

Now, there are 11 paths from test set that are correctly classified as "good" and four paths from test set, which are correctly classified as "bad". However, there are eight paths from test set that are actually "bad" but are incorrectly classified as "good" and one path from test set that is actually "good" but is incorrectly classified as "bad".

The most important fact is that there are no unclassified paths. It is possible to further increase classification power by usage of training set that contains more paths, which would result in a larger number of rules. It is also clear that reduct set computation contributes to the accuracy of classification.

VI. Conclusion and Further Work

The main purpose of the model showed in this paper is to improve decision making during the project life cycle, and in that way to provide project success. The model addresses PERT-like network diagram of a specific project in which every activity gets one of the three possible values. Those values are related to the project duration, cost, scope etc. By choosing activity values possible scenario of project realization is created. Then rough sets based part of the model evaluates every path of the network diagram and implies which ones are "bad". The "bad" paths are in some way critical and project manager should take that in account. Thanks to the fact that model enables different outputs for the same set of activity values, its usage is similar to the real-life decision making. More profound analysis may be provided by extending the set of Decision values of the model on three or five of them. In that way, paths would be evaluated on a larger scale. For a project manager who wants to be familiar with every possible scenario of his project realization and to reduce any raising risk, this could be very important knowledge. Everything abovementioned implies that the main purpose of described model is fulfilled. Further work will investigate the possibility of using actual activity weights and to discretize them via discretization process, which will result in finite number of classes. One more possible future research in this domain is to use real-life projects with some missing values and to apply and test algorithms for missing values completion.

ACKNOWLEDGMENT

This research is financially supported by Ministry of Education and Science of the Republic of Serbia under the project number TR32044 "The development of software tools for business process analysis and improvement", 2011-2014.

REFERENCES

[1] Purvis, R. L., McCray, G. E. and Roberts T. L, "The impact of project management heuristics to IS projects", IEEE Computer society, Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03), 0-7695-1874-5/03 , 2002.

[2] Agarwal, R., Tannru, M., & Dacruz, M. (1992). Knowledgebased support for combining qualitative and quantitative judgments in resource allocation decisions. Journal of Management Information Systems, 9 (1), 165–184.

[3] Bukszar, E., & Connolly, T. (1998). Hindsight bias and strategic choice: Some problems in learning from experience. Academy of Management Journal, 31 (3), 628–641

[4] Hogarth, R., & Einhorn, H. (July, 1990). Venture theory: A model of decision weights. Management Science, 36 (7), 780– 803

[5] Davis, E. W. and Patterson J. H., "A Comparison of Heuristic and Optimum Solutions in Resource-Constrained Project Scheduling", Management Science, vol. 21, no. 8, pp. 944-955, April 1975.

[6] Boctor, F. F., "Some efficient multi-heuristic procedures for resource-constrained project scheduling", Elsevier, European Journal of Operational Research, vol. 49, no. 1, pp 3-13, 1990.

[7] Tormos, P. & Lova, A., "A competitive heuristic solution technique for resource-constrained project scheduling", Springer, Annals of Operation Research no. 102, pp 65-81, 2001.

[8] Kolisch, R. and Hartmann S., "Experimental investigation of heuristics for resource-constrained project scheduling: An update", Elsevier, European Journal of Operational Research, vol. 174, no. 1, pp 23-37, October 2006.

[9] D. Pons, Does Reduced Uncertainty Mean Greater Certainty? – Project management with uncertain durations. Project Management Institute of New Zealand (PMINZ) 2006 Conference, Christchurch, New Zealand, 4-6 Oct. 2006.

[10] Peters, G. and Gordon Hunter M. "Disclosing Patterns in IT Project management – a rough set perspective", Springer-Verlag Berlin Heidelberg, PReMI2009, LNCS5909, pp 591-596, 2009.

[11] Peters, G., Lingras P., Slezak, D. and Yao, Y., "Rough sets: selected methods and applications in management and engineering", Springer-Verlag, London, 2012.

[12] Zhang, Q., "A rough set method for performance evaluation of project management", Second international conference on computer engineering and technology, pp 348-350, April, 2010.

[13] Zhengyuan, J. and Lihua, G. "The project risk assessment based on rough sets and neural network (RS-RBF)", 4th international conference on wireless communications, networking and mobile computing, pp 1-4 October, 2008.

[14] Ping, Z. and Shi-xiang, Y., "An approach to project risk analysis based on rough sets", Control and decision conference (CCDC), pp 1197-1202, May, 2010.

[15] Gang, X., Jinlong, Z., Lai, K.K. and Lean, Y., "Variable precision rough set for group decision-making: An application", Elsevier, International journal of approximate reasoning, vol. 49, no. 2, pp 331-343., October, 2008.

[16] Dobrilovic, D., Brtka, V., Berkovic, I., Odadzic, B., „Evaluation of the Virtual Network Laboratory Exercises Using a Method Based on the Rough Set Theory", Computer applications in engineering education, vol. 20, no. 1, pp. 29-37, 2012.

[17] Z. Pawlak, J. Grzymala-Busse, R. Slowinski and W. Ziarko, "Rough sets", Association for Computing Machinery. Communications of the ACM; Nov 1995; 38.

[18] Z. Pawlak, A. Skowron: Rudiments of rough sets, An International Journal of Information Sciences 177, pp. 3–27, 2007.

[19] Z. Pawlak, "Rough set approach to knowledge-based decision support", European Journal of OR, 1997; 99, pp. 48-57.

[20] S. Greco, M. Benedetto, R. Slowinski "New Dewelopments in the Rough Set Approach to Multi – Attribute Decision Analysis" Bulletin of International Rough Set Society, Volume 2, Number 2/3, September 1998; pp. 57–87.