

# Rough Rule-Based Systems for Sparse and Dense Data Analysis Used in Project Evaluation

Tadeusz A. GRZESZCZYK

Faculty of Management, Warsaw University of Technology  
 ul. Narbutta 85, 02-524 Warsaw, Poland  
 e-mail: tadeusz.grzeszczyk@pw.edu.pl

**Abstract**—Massive volumes of unstructured, multidimensional and heterogeneous data used in project management and evaluation processes generally rapidly increase. It causes and reinforces the need of searching for new, appropriate and efficient methods addressing the use of sparse and dense data analysis, and Big Data technology for these processes. Research on rule-based systems can lead to significant advances in project management and evaluation. The main objective of this paper is to discuss how rule-based systems can be used for sparse and dense data analysis applied in project evaluation. The obtained results indicated the great potential of such systems based on rough set theory. At the beginning of this paper, the core problems of rough rule-based systems are given. Then, sparse and dense data models build on decision tables are shortly shown. Subsequently, an exemplary of project classification using rough rule-based system based on sparse data is briefly characterized. Finally, the conclusions and recommendations which concern possible directions of project evaluation methods and systems development are presented.

**Keywords**—Project evaluation; Sparse and dense data analysis; Rough rule-based systems

## I. INTRODUCTION

The essence of project evaluation can be derived from classical, praxeological meta-theory which is characterized by the number of theories and project management schools [1]. The evaluation is also often referred to as a process of systematic testing of quality, benefits and value of some project [2]. The implementation of this process takes place on the basis of the search for answers to a number of research questions. Assessing quality, benefits and value of the projects is carried out as a result of orderly and systematic actions.

Volume, velocity and variety of data used in project evaluation processes generally increase, which causes the necessity of searching for new systems and technology addressing the use of sparse and dense data analysis. Research on rule-based systems, based on rough set theory, can lead to significant advances in project management and evaluation [3]. The main objective of this paper is to discuss how rule-based systems can be used for sparse and dense data analysis applied in project evaluation.

## II. ROUGH RULE-BASED SYSTEMS

Rule-based systems (a special type of expert systems) can be built by learning from empirical data and using expert knowledge. They can be divided among others

based on type of rule bases, type of structure, type of computing environment and type of logic [4]. Special expectations can be associated with logico-mathematically inspired tools. These systems can be based on soft computing and fuzzy logic [5,6,7] or rough set theory [8,9], which are significant mathematical tools useful in the processes of computational intelligence, knowledge representation development and have applications in many other areas.

Multi-criteria analysis of projects which are described by dense and sparse data models by using generalized rough set theory provides an opportunity to ensure adequate accuracy of defining vague concepts in the evaluation system based on knowledge. Such a concept may be described by two approximations: lower and upper, built from the collections of elementary sets (elementary concepts, knowledge granules). The accuracy of defining vague concepts rises with increasing the number of criteria that describe the individual projects.

In dense data models based on decision tables it can be relatively easy to save necessary data for operating information and knowledge acquired empirically. Understandable presentation of results in the form of decision rules simplifies the interpretation of the results obtained because the rules create a record of expert knowledge in the form of an adjustable base of useful knowledge, eg. in rules-based system for project evaluation.

Rough sets are also proper for processes of analysis of sparse data that are poorly structured, fuzzy, and are qualitative in nature. Rough set theory extends the capabilities of classical mathematical tools (based on the notion of conventional sets) for additional analysis of the imprecise concepts (that can be presented by using two sharp concepts).

It is possible to reduce redundant information and knowledge and build a smaller (then the original) set of criteria sufficient to adequately describe the analyzed projects. Minimizing also applies to the number of generated decision rules obtained by analyzing the criteria stored in sparse decision tables.

The most suitable methodical trend for rough rule-based systems for sparse data analysis used in project evaluation is based on domination principle. This trend is consistent with the most well-known method of multi-criteria optimization enabling searching solutions according to the Pareto principle. With the use of this principle, one can carry out relatively objective process of multiple-sorting, which makes the comparison of some

projects described by values of evaluation criteria. Approach based on domination relation, usually referred to as DRSA (Dominance-based Rough Sets Approach), is used to solve typical multi-criteria decision-making problems, ie. ranking, selection, sorting and more [10,11,12].

DRSA also ensures more readable form of knowledge represented by means of rules based on the dominance relation, because of their more complex syntax, as a result of replacing considerations about objects indiscernibility, testing greater number of relations: equality and outranking. Decision rules generated by using algorithms based on DRSA, represent a preference model discovered from sparse empirical data on specific project evaluation.

Newer DRSA, on the one hand, is characterized by all the benefits of older CRSA (Classical Rough Set Approach) ie. allows to analyze the data (in the conditions of risk and uncertainty) associated with problems solving: classification (sorting for the predetermined decision classes), ordering (ranking of alternatives) and objects selection. On the other hand, it constitutes a practical approach useful in the process of multi-criteria decision support, eg. in the field of comprehensive project evaluation [13]. This process is based, in its essence, on a number of preferentially ordered criteria. Compatible with DRSA analysis of sparse data (saved in decision tables) as a result leads to generating decision rules taking into account the preference dependence occurring in this data. Obtained decision rules may be useful in classification of new projects, which were not used in decision rules generation processes - that is, learning of the evaluation system.

Development of research on rough rule-based systems for sparse data analysis used in project evaluation can rely on trying to apply modified rough set

theory in this field. One of such modifications is the use of domination relation instead of indiscernibility relation. After such modification, in the effect of multicriteria classification and projects sorting, we can achieve linking projects to each individual decision classes in compliance with more useful domination principle.

**III. SPARSE AND DENSE DATA MODELS**

Application of rough rule-based systems gives the opportunity to build evaluation systems based on knowledge using dense data models and sparse representations, in which case there are doubts about the quality of reality described, accuracy of description, completeness and reliability of conducted analyzes. Data analysis based on sparse representations allow to carry out research under uncertainty reasoning, in cases of difficulties in the strict and precise describing complex phenomena and qualitative socio-economic and environmental impacts.

Dense data models relate to projects that are unique and described by many different criteria require computation with large matrices. For big data analytics, specially chosen approaches are needed [14]. It is desirable to build sparse data models with a small number of criteria and data, thereby simplifying the description and projects analysis. Thus, it is relatively easy to compare with each other and evaluate many projects. For practical reasons, it is not necessary to base in project evaluation processes on all possible criteria and there may be difficulties in collecting all the data necessary to build dense data models. After extracting a limited number of criteria, there is also the possibility of effective identification of parameters describing the projects and appropriate distinguishing one from another in evaluation processes, ie. to find similarities and differences between them, positive and negative features of both.

TABLE I. DECISION TABLE MODEL FOR KNOWLEDGE REPRESENTATION ABOUT EVALUATED PROJECTS

$U/C$	Criterion $c_1$	Criterion $c_2$	Criterion $c_3$	...	Criterion $c_m$	Decision Attribute
Project $x_1$	$f(x_1, c_1)$	$f(x_1, c_2)$	$f(x_1, c_3)$	...	$f(x_1, c_m)$	$d_1$
Project $x_2$	$f(x_2, c_1)$	$f(x_2, c_2)$	$f(x_2, c_3)$	...	$f(x_2, c_m)$	$d_2$
Project $x_3$	$f(x_3, c_1)$	$f(x_3, c_2)$	$f(x_3, c_3)$	...	$f(x_3, c_m)$	$d_3$
...	...	...	...	...	...	
Project $x_n$	$f(x_n, c_1)$	$f(x_n, c_2)$	$f(x_n, c_3)$	...	$f(x_n, c_m)$	$d_n$

In rough set theory, information/decision tables represent some gathered data from selected domain and expresses all the knowledge regarding observable universe. Rows of a table are called objects or examples [8]. Information about evaluated projects, described by the same specific set of evaluation criteria also can be saved in the form of a decision table. It may be unnecessarily large in the case of dense data.

Table 1 shows the use of the decision table model consisting of projects  $x_1, x_2, x_3, \dots, x_n$ , criteria ( $c_1, c_2, c_3, \dots, c_m$ ) and decision attribute. For individual projects occurring in table rows record the information function

reflecting relations between the different projects and criteria values used for their description and evaluation. It is possible to save in these rows the number of points given by experts for projects in evaluation processes.

**IV. PROJECT CLASSIFICATION USING ROUGH RULE-BASED SYSTEM**

Figure 1 shows the example of project classification using rough rule-based method based on empirical data from a set of sparse data input. Classification is carried out iteratively. Introduction of input data for a single

iteration ensures the functioning of the evaluation system for a single sequence of actions. In subsequent iterations, it is possible to modify the input parameters.

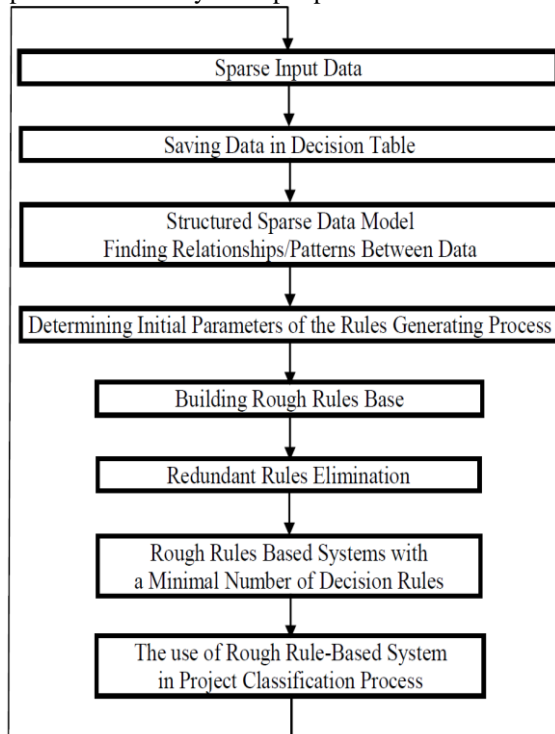


Figure 1. Project Classification Using Rough Set Theory (source: own research)

In the next stage, empirical data is collected in decision table. Then, the operations of structuring sparse data model and preliminary analysis of these data carried out for finding relationships / patterns between them is performed. The implementation of these operations allows finding the interpretation of data, finding relationships between data and determination of cumulative decision classes. For these classes, the following approximations are determined: lower and upper, which are necessary to determine decision rules.

Building rough rules base is made in the process of learning the system and decision rules generated by it. The process of building this knowledge base begins with determining initial parameters of the rules generating process: minimum value of rule support and maximum length of decision rules.

At the end of the process of preparing evaluation rules knowledge base, redundant rules elimination is carried out. The process of checking is performed in two steps: in conditional part of rules and in the whole set of rules. In the first step, rules containing redundant information in the conditional part are eliminated. In the second step, stronger rules are eliminated.

After the elimination of redundant rules, one can build rough rules based systems with a minimal number of decision rules possible to use in rough rule-based system for project classification processes realization. Thanks to the generalization of rough set theory, the execution of the process of knowledge discovery is also possible for empirical sparse data characterized by inconsistency.

Decision rules stored in the knowledge base are supplemented in subsequent iterations. With this approach, the dynamics of the system, continuous improvement of decision rules and the entire knowledge base are ensured.

## V. SUMMARY

Problems connected with project evaluation have a multi-faceted character and usually there are several various big data sources that can clarify the same phenomenon. Data collected within different scientific disciplines are interchangeable and useful. In the effect of analyzing existing solutions used for project evaluation, it was considered expedient to develop interdisciplinary research contributing to proposing new systems, methods and techniques which are an essential complement to the currently used ones. Still relatively little research is devoted to applications of sparse and dense data analysis in comprehensive project evaluation. Therefore, it is reasonable to continue research and search for new solutions in this range.

The conducted research conclusions presented in this paper show that building on the achievements in the field of knowledge engineering and artificial intelligence can significantly expand the range of possible solutions concerning the use of rule-based systems for sparse and dense data analysis applied in project evaluation.

Practical applications of evaluation system based on decision rules are limited so far. Evaluators with relevant competencies for conducting research, data gathering and analyzing, should familiarize themselves with new solutions resulting from knowledge based system development. These systems allow representing specific and important knowledge in qualitative and lexical forms which are typical for experts reasoning about project results. Despite these opportunities in comprehensive evaluation area, such systems are little known in the business. In public sector, these solutions are even less known.

On the one hand, studies on knowledge systems that can be used in the project evaluation process are characterized by significant application effects in the field of sparse and dense data analysis. On the other hand, the results are relatively difficult to wider use and demonstrate several disadvantages of the proposed solutions. For example, evaluation knowledge representation obtained in the form of patterns generated by learning intelligent systems does not always provide a clear and practically useful knowledge, the use of which can bring tangible business results. In some cases, it is worth to consider application of classical statistical modelling or adopted well known methods to large data analysis.

The considerations presented in this paper may inspire both project managers as well as evaluators and lead them to creation of new instruments of comprehensive project evaluation considering its difficult to measure, unstructured and qualitative elements. New evaluation methods based on rule-based systems can significantly enrich and improve sparse and dense data processes applied in project evaluation and also contribute to reducing subjectivity of evaluation

processes. These methods can provide significant support for the human in the processes of its non-standard, irregular and creative reasoning in evaluation context. Rule-based models have huge potential for quick development in the field of sparse and dense data analysis and can contribute to improving of computer processing and move project evaluation procedures towards replication of human thought and cognitive processes, which is not possible yet.

#### REFERENCES

- [1] L. Klein, C. Biesenthal, Towards a Praxeology of Resilient Project Management - A Conceptual Framework, Systemic Excellence Group, Berlin-Mitte 2013, p. 4.
- [2] The Joint Committee on Standards for Educational Evaluation. The program evaluation standards. How to assess evaluations of educational programs, Sage Publications, Thousand Oaks, CA 1994, p. 3.
- [3] T.A. Grzeszczyk, Application of the Rough Set Method for Evaluation of Structural Funds Projects, in: ICEIS 2006. Proceedings of the Eighth International Conference on Enterprise Information Systems. Artificial Intelligence and Decision Support Systems, Paphos, Cyprus 2006, pp. 202-207.
- [4] H. Liu, A. Gegov, M. Cocea, Rule Based Systems for Big Data: A Machine Learning Approach, Springer, 2015, pp. 1-5.
- [5] L.A. Zadeh, Fuzzy Sets, Information and Control, 8, 1965.
- [6] L.A. Zadeh, Fuzzy logic, neural networks and soft computing, Communications of the ACM, 37/1994.
- [7] M. Luo, F. Sun, H. Liu, Z. Li, A novel T-S fuzzy systems identification with block structured sparse representation, Journal of the Franklin Institute, Vol. 351, Issue 7, July 2014, pp. 3508-3523.
- [8] Z. Pawlak, Rough Sets. Theoretical Aspects of Reasoning About Data, Kluwer, Dordrecht 1991.
- [9] Z. Pawlak, A. Skowron, Rough sets: Some extensions, Information Sciences, 177, 2007.
- [10] Greco S., Matarazzo B., Słowiński R., The use of rough sets and fuzzy sets in MCDM, in: Gal T., Stewart T., Hanne T. (ed.), Advances in Multiple Criteria Decision Making, Chapter 14, Kluwer Academic Publishers, Boston 1999, 14.1–14.59.
- [11] Greco S., Matarazzo B., Słowiński R., Rough set theory for multicriteria decision analysis, European Journal of Operational Research, 2001, Vol. 129, No. 1, pp. 1–47.
- [12] Blaszczyński J., Greco S., Słowiński R., Multi-criteria classification – A new scheme for application of dominance-based decision rules, European Journal of Operational Research, 181(3), September 2007, pp. 1030–1044.
- [13] T.A. Grzeszczyk, Towards the model of comprehensive projects evaluation system, Faculty of Management Warsaw University of Technology, Warsaw 2013.
- [14] C. Ordóñez, Y. Zhang, W. Cabrera, The Gamma Matrix to Summarize Dense and Sparse Data Sets for Big Data Analytics, IEEE Transactions on Knowledge and Data Engineering, Vol. 28, Issue 7, 2016, pp. 1905-1918.