APPLICATION OF THE ROUGH SET METHOD FOR EVALUATION OF STRUCTURAL FUNDS PROJECTS

Tadeusz A. Grzeszczyk
Institute of Production Systems Organization, Warsaw University of Technology, ul. Narbutta 85, 02-524 Warsaw, Poland

Keywords: Structural funds, project proposals, evaluation, rough set theory.

Abstract: Main subject of the present paper is presentation of the concept for application of rough set theory in evaluation of structural funds projects. Author presents scheme of classification algorithms based on rough set approach. This algorithm can be used for the problem of project proposals classification.

1 INTRODUCTION

Structural funds of European Union are considered as a great chance for economical development of the entire Community. However, as results from practical experience, their absorption is still considered as a source of significant methodical difficulties for the beneficiaries as well as for institutions which hold management of these funds. One of the key aspects in management of these funds are the procedures of monitoring and evaluation of the project proposals. In this connection, the problem of effective evaluation of the projects to be co-financed by all the EU funds – including structural funds – is particularly important for solution of aforementioned methodical difficulties.

In order to support the process of EU funds absorption, special information systems are created. All the institutions dealing with management of EU funds as well as final beneficiaries make use of the above systems. The main tasks of a. m. systems are as follows (www.mf.gov.pl; www1.ukie.gov.pl/www/en.nsf):

– ensuring effective and transparent management of EU funds within the range of programs to be financially supported by EU,
– monitoring and management of projects beginning at the moment of preparing and sending application forms through all the stages of their realization, up to their final stage,
– monitoring and evaluation of financial indicators and effects of tasks carried out within the range of Community Support Framework and Operational Programmes,
– ensuring required reporting to European Commission, referred to implementation of EU structural funds and Cohesion Fund in Poland.

In the second point of the present paper, basic concepts useful for determining classification algorithm are presented. Rough set theory allows processing the experimentally obtained data. This theory and Case-Based Reasoning Technology are the fastest growing areas in the field of knowledge-based systems (Aamodt, 1994; Cao, Shiu, Wang, 2001; Kruse, Schwecke, Heinsohn, 1991). Rough sets are not yet as popular as fuzzy sets theory (Zadeh, 1965; Zadeh, 1994) which is in a sense complementary to the first one. Algorithm presented in point 3. can be applied at implementation of rules-based knowledge base for realization of evaluation system referred to EU financed project proposals.

Slowinski and Zopoundis were the first to apply rough set approach in the evaluation of corporate failure risk (Slowinski, Zopounidis, 1995). This method attempts to describe a set of enterprises by a set of multi-valued attributes. Grzeszczyk discusses the layout of conception referred to the use of artificial intelligence methods (neural networks) for prior appraisal of project proposals to be submitted by Polish enterprises to European Union in order to get financial assistance for investments from the EU structural funds and the state budget (Grzeszczyk, 2004). Main subject of the present paper is presentation of the concept for application of rough set theory in evaluation of structural funds projects.

A. Grzeszczyk T. (2006).
APPLICATION OF THE ROUGH SET METHOD FOR EVALUATION OF STRUCTURAL FUNDS PROJECTS.
In *Proceedings of the Eighth International Conference on Enterprise Information Systems - AIDSS*, pages 202-207
DOI: 10.5220/0002495702020207

The research connected with structural fund projects evaluation system for qualitative analysis was set in motion by the author in order to define the procedure of creating an information system (set of basic definitions is quoted in point 2.). At this stage of research it was defined what conditional attributes were as well as so-called universe. In this point of paper many definitions are presented (connected with rough set theory) being indispensable for further research works. Next step is announcing a way of creating decision table that includes a set of values of decision attribute, responding to given condition attributes. Preparation the classification algorithm finishes the works connected with creating the conception of qualitative analysis. In accordance with conception assumed by the author, determining of the values that decision attribute accepts, is possible in stage of tests and exploitation of evaluation system. Then an already created rules-based knowledge base (consisting of decision rules) is used. The following proceeding stages are assumed.

1) Qualification of basic concepts connected with analysis of information system (indispensable for further works on the use of rough sets theory at discovering decision rules):

- qualification of universe,
- information system and knowledge base (def. 1.-2.),
- indiscernibility relation and its properties (def. 3.),
- lower and upper approximations of set (def. 4.),
- concept of reduct and core set of attributes (def. 5.-6.).

2) Concepts used in analysis of decision system:

- decision table (def. 7.),
- dependence between attributes (def. 8.),
- dependence coefficient between sets of attributes (def. 9.).

3) Presentation of classification algorithm.

Mathematical tools (using the rough set theory) applied by author for analysis of decision table, on basis of which it is possible to define the decision rules, are presented in the next point of the present paper. The definitions are facilitating presentation of conception related to use of approximate reasoning in process of determining the value of decision attribute. Presenting of classification algorithm (see point 3. as well as fig. 1.) finishes this part of research connected with rough set method for evaluation in the process of structural funds projects preparation.

2 ANALYSIS OF ROUGH SET

A crucial role in understanding the rough set theory is played by a clear-cut defining of way for representation of knowledge. Information is kept in form of table, which unambiguously defines the studied information system. Lines of this table are making up the aforementioned objects, while columns make up next conditional attributes.

Now it is necessary to define the basic concepts to be useful in further considerations, definitions quoted after: (Damasio, Maluszynski, Vitoria, 2003; Pawlak, 1982; Pawlak, 1991).

Definition 1. Information system

Information system is a pair **K** = (*U, A*), where:

- U is a non-empty finite set of objects called a universe,
- A is a non-empty finite set of conditional attributes, where every attribute $a \in A$ is a function $a : U \to V_a$, where Va is a domain of attribute a.
- Every subset $X \subseteq U$ is called a concept in U.

Tools, which use approximate approach, are well useful to create a suitable knowledge representation, which is often called the ability of classification of a definite reality. Knowledge then, can be equivalent to the skill of conducting the process of classification (of divisions) within a given universe. Second definition describes knowledge base from the point of view of rough sets theory.

Definition 2. Knowledge base

If *U* is a universe and **R** means set (or a family) of equivalence relation then *W*=<*U*,**R**> is called knowledge base about *U.*

Another important concept is an indiscernibility relation. It is significant while solving the problem of reduction of knowledge.

Definition 3. Indiscernibility relation

If **K**=(*U, A*) is an information system and $B \subseteq A$, then in the universe set, the following binary indiscernibility relation can be defined, occurring between objects in system **K**: $IND(B) = \{(x, y) \in U \times U : \forall a \in B : a(x) = a(y)\}$.

Recapitulating, any two objects *x, y* (belonging to universe) fulfil the indiscernibility relation, if they have identical attribute „*a*" values, for the set of attributes under analysis.

Checking indiscernibility of objects is one of many basic actions that should be done while creating decision rules. The above mentioned definition is used, among others, at the stage of dividing universe into elementary concepts, that is sets of objects which are undistinguishable in respect of given attributes. A special case is division into decision classes (that is: sets of elements which are undistinguishable in respect of conditional attributes).

Approximation of sets is a basic activity in the rough sets theory. In this connection, one should qualify two approximations of set. The following definition serves to it.

Definition 4. Lower and upper approximation

Let **K**=(*U, A*), $X \subseteq U, B \subseteq A$.

B-lower approximation of *X*, then: $\underline{B}X = \{x \in U : I_B(x) \subseteq X\}$.

B-upper approximation of *X*, then: $\overline{B}X = \{x \in U : I_B(x) \cap X \neq \emptyset\}$.

B-borderline region of *X*, then: $BN_B(X) = \overline{B}X - \underline{B}X$.

There are additional remarks connected with the above presented definition 4:

- in relation to B-borderline region it is possible to apply the following, equivalent record $BN_B(X) = \overline{B}X \setminus \underline{BX}$,
- it is possible to qualify B-positive region of X (equal to lower approximation): $POS_B(X) = \underline{B}X$,
- in analogy to previous point, it is possible to record B-negative region) of X, as: $NEG_B(X) = U - \overline{B}X$, or $NEG_B(X) = U \setminus \overline{B}X$,
- concept X is B-definable for $\underline{B}X = \overline{B}X, BN_B(X) = \emptyset$,
- lower B-approximation of concept X is the maximum B-definable subset of universe U contained in X, and upper B-approximation is B-definable minimum subset of universe U containing X: $\underline{B}X \subseteq X \subseteq \overline{B}X$.

Next notion, of which applying is indispensable in case of conducting reduction of attributes, is reduct. It is a minimum subset of initial set of attributes, which guarantees the same set of elementary concepts as the initial set. It means that in case of reduct we have smaller number of arguments, and the same knowledge, characterizing the fragment of reality (universe) we are interested in.

Definition 5. Reduct of set of *B* attributes

Let **K**=(*U, A*) mean information system, set of attributes $B \subseteq A$. Another subset of attributes marked as $R \subseteq B$, is a reduct of a set of attributes *B*, if: set R is independent and IND(R)=IND(B).

Family of all reducts of set of attributes *B* is marked with symbol *RED(B)*.

The next defined concept is a core. It defines these attributes, which cannot be removed in any case. The core is a set of attributes contained in every reduct. Attributes of the core define the division of universe into elementary concepts, carrying in themselves desirable knowledge.

Definition 6. Core of set of *B* attributes

Let **K**=(*U, A*) mean information system, set of attributes $B \subseteq A$.

The core of set of attributes *B*, in information system **K**, marked with symbol *CORE(B)* is called the set of all indispensable attributes of this set.

It is obvious, that in every reduct there is a core. Sometimes, it can happen, that all conditional attributes are indispensable. In such case the reduct is equal to the core.

Information system added to one column, including record of three different values of decision attribute, can be called a decision table. Its formal definition is shown below.

Definition 7. Decision table (Kryszkiewicz, 1996)

Decision table is an information system K=$(U, A \cup \{d\})$, where:

- U is a universe,
- A is a set of conditional attributes,
- $d \notin A$ is a decision attribute.

The decision table presents dependence between conditional attributes and decision attribute. So it can be helpful at investigation of dependence in bases of knowledge. Interesting are e.g. dependencies occurring between sets of attributes. Knowledge represented by certain set of attributes can be mutually combined or result from knowledge characterising the other set. It seems to be indispensable to define in this place a dependence between sets of attributes: both total as well as partial ones.

Definition 8. Dependence between attributes

Let K=(U, A) mean information system, $B \subseteq A, C \subseteq A$.

$B \Rightarrow C$, which we read, set of attributes C depends on set of attributes B, when the following record is fulfilled: $IND(B) \subseteq IND(C)$.

Here it is necessary to define partial dependence, which can occur between sets of attributes.

Definition 9. Partial dependence between sets of attributes

Let K=(U, A) mean information system, $B \subseteq A, C \subseteq A$.

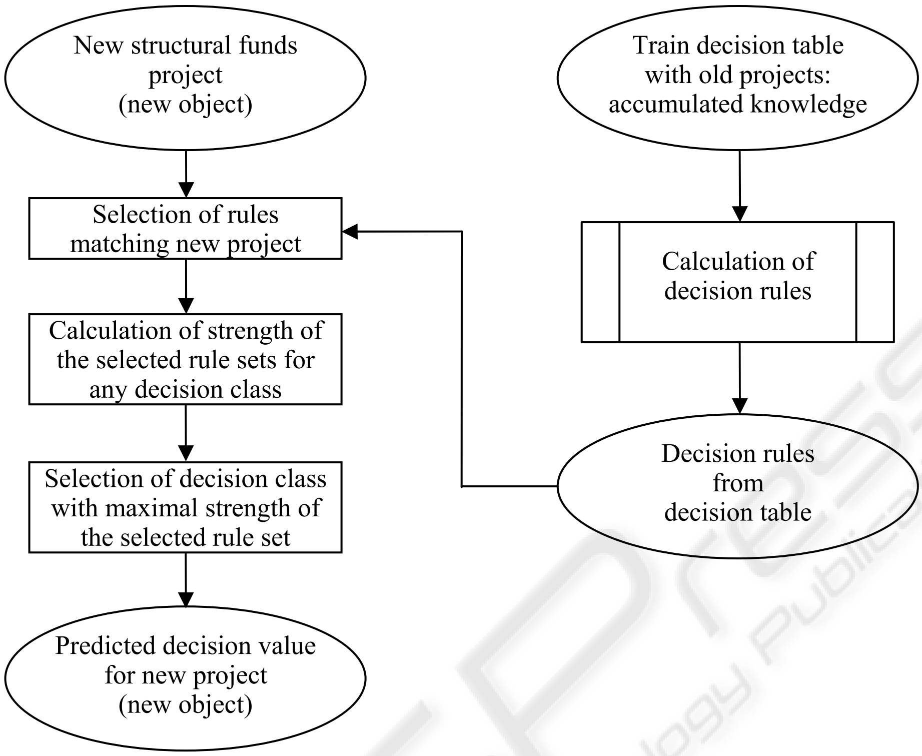

Figure 1: Algorithm for classification of structural funds project - source: author's own study on the basis of (Bazan, Hung Son Nguyen, Sinh Hoa Nguyen, Synak, Wroblewski 2000).

Set of attributes C depends on set of attributes B (in other words, knowledge contained in C depends on knowledge in B) in degree k, which can be written in symbols: $B \Rightarrow_k C$ when: $k = \gamma_B(C) = \frac{|POS_B(C)|}{|U|}$.

Coefficient "k" accepts values from interval $0 \le k \le 1$, it is defined as dependence coefficient between sets of attributes. k=1 means, that total dependence exists, marked as $B \Rightarrow C$. In this situation every object of universe belongs to positive region (i.e. lower approximation). There are no objects which would possess identical conditional attributes, and different decision attributes. In such case there is not contradictory information. On the other hand, for k=0 there is not any dependence between sets of attributes.

Above author gives only the basic concepts indispensable for further works on the usage of rough sets theory at discovering of decision rules. They are applicable for determining algorithm of creating decision rules. Detailed description of rough set theory can be found in: (Damasio, Maluszynski, Vitoria, 2003; Pawlak, 1982; Pawlak, 1991).

3 CLASSIFICATION ALGORITHM BASED ON ROUGH SET

In fig. 1. classification algorithm can be seen, which leads to formulating decision rules and to dealing with problem of resolving conflicts between decision rules classifying a new project to different classes and predicted decision value for new object (project). Its basic task is reduction of redundant objects and conditional attributes (total and local). The author of this work suggests, that the set of rules generated from decision tables is used to mark the value of decision attribute, influencing the projects proposals evaluation process.

The base of decision rules consisting of decision rules, determined with algorithm for example (Mrozek, 1992) – exemplifies basis for process of approximate reasoning. The consecutive values of decision attribute can be determined with its use, to decide which attribute is used to mark the value of decision attribute for the process of new project proposal evaluation.

4 CONCLUSION

Synthesising the presented results of research the author presents a hypothesis about possibilities and advisability of using the rough set theory in the process of structural funds projects evaluation. The main advantages of methods assisting evaluation, based on rough set theory– in relation to traditional statistical analyses – are first of all the features as below.

Rough set theory is an instrument serving for recording of experienced persons and experts experiences in the form of decision rules based on empirical materials as well as ensuring processing of information relatively easy.

There occurs a relatively high certainty, that no essential dependence between conditional attributes affecting decision attribute (so-called decision rule) will be omitted. However, at using traditional methods of statistical analysis even very essential dependencies occurring between attributes can be omitted. Since there is a lack of instruments enabling defining of such dependence. As an example, multifactor analysis of correlation makes possible qualification of numerical value of influence for individual attributes between themselves only. It does not create however, any possibility of defining connections between values of individual attributes.

Methods based on applying the rough set theory are using experts' experience and they make possible verification of their opinions as well as they are assuring relative easiness in interpretation of results. Thus the conclusions related to studied decision attribute are received. It is also easy to interpret their alternatively incorrect acting.

Dependences established thanks to use of rough set theory can be ranked in accordance with the degree of their importance. From additional description (based on empirical materials e.g. experts' opinions) an opinion of significance of decision rule as well as influence of definite attributes onto decision results can be made. There also exists large degree of possibility for verification of results, because every generated rule is accompanied with description including reference to empirical sources.

Redundancy of attribute is easy to prove with the use of division of decision table into elementary concepts. The results achieved can be authenticated by analysing a discernibility matrix.

Rough set theory makes possible carrying out analyses for different sets of conditional attributes. The discussed theory is well usable to investigation of low structured processes (especially socio-economic ones). It makes possible identification of decision rules, difficult to intuitive defining.

There exists relative easiness of modification in reference to decision table (by addition of new, not considered earlier conditional attributes). This makes possible creating different decision rules.

The SIMiK system (The Information System for Financial Monitoring and Controlling of the Structural Funds and the Cohesion Fund) - for details see (www.mf.gov.pl; www1.ukie.gov.pl/www/en.nsf) - recommended by Ministry of Finance in Poland, is intended to improve process of EU funds absorption. For potential project providers the application generator is the most important element of the system. It serves to put in all the necessary data. The evaluation system referred to project proposals should in future be operated not only with SIMiK but also with the other systems considered as the source of knowledge and information. The evaluation system provided with rules-base knowledge base, described in the present paper, can serve as one of the co-operating systems. Preliminary research works referred to application of classification algorithm also proved to be very promising.

REFERENCES

Aamodt, A., 1994. Explanation-Driven Case-Based Reasoning. In S. Wess, K. Althoff, M. Richter (eds.) Topics in case-based reasoning. Springer-Verlag.

Bazan, J. G., Hung Son Nguyen, Sinh Hoa Nguyen, Synak, P., Wroblewski, J., 2000. Rough Set Algorithms in Classification Problem. In L. Polkowski, S. Tsumoto, T. Y. Lin (eds.) Rough Set Methods and Applications. New Developments in Knowledge Discovery in Information Systems. Physica-Verlag, Heidelberg.

Cao, G., Shiu, S., Wang, X. 2001. A Fuzzy-Rough Approach for Case Base Maintenance. In D.W. Aha, I. Watson (eds.) Case-Based Reasoning Research and Development. 4th International Conference on Case-Based Reasoning, ICCBR 2001. Vancouver, BC, Canada. July 30 - August 2, Proceedings. Lecture Notes in Computer Science. Springer-Verlag, Heidelberg.

Damasio, C. V., Maluszynski, J., Vitoria, A., 2003. From Rough Sets to Rough Knowledge Bases. Fundamenta Informaticae, 57 (2-4), November 2003.

Grzeszczyk, T. A., 2004. Application of Neural Networks for Prior Appraisal of Structural Funds Project Proposals. In Proceedings of the Sixth International Conference on Enterprise Information Systems. Vol. 2. Portucalense University, Porto. Portugal.

Kruse, K., Schwecke, E., Heinsohn J., 1991. Uncertainty and Vagueness in Knowledge Based Systems. Springer-Verlag.

Kryszkiewicz, M., 1996. Boolean Reasoning and Aprori-like Technique for Rough Sets. ICS Research Report 26/96, Warsaw University of Technology, Warsaw.

Mrozek, A., 1992. A new method for discovering rules from examples in expert systems. Int. Man-Machine Studies, 36.

Pawlak, Z., 1982. Rough sets. International Journal of Computer and Information Sciences, 11/1982.

Pawlak, Z., 1991. Rough sets: Theoretical aspects of reasoning about data. Kluwer, Dordrecht.

Slowinski, R., Zopounidis, C., 1995. Application of the rough set approach to evaluation of bankruptcy risk, Intelligent Systems in Accounting, Finance and Management, No 4 /1995.

Zadeh, L. A., 1965. Fuzzy sets. Information and Control, 8/1965.

Zadeh, L. A., 1994. Fuzzy logic, neural networks and soft computing. Communications of the ACM, No 37.

www.mf.gov.pl

www1.ukie.gov.pl/www/en.nsf