Kazimierz Zaras

THE DOMINANCE-BASED ROUGH SET APPROACH (DRSA) APPLIED TO BANKRUPCY PREDICTION MODELING FOR SMALL AND MEDIUM BUSINESSES

Abstract

The preferential information given in the form of ranking or classification examples is more natural than those given in the form of functional parameters or the relational model of the preferences. Nevertheless, processing of these data cause certain difficulties related to a lack of coherence and contradictions in these examples. These contradictions often result from granularity of description language, inaccuracy or uncertainty of the information which makes the decision maker hesitate before the decision making. The model of the preferences will not correct or ignore these contradictions, but rather consider them to release a certain doubtful part of them. Then, exploitation of this model within the framework of decisional problems will lead to unquestionable and possible recommendations.

The Rough Set Theory takes into account this postulate making the contradiction analyze possible. This theory was introduced in the early 1980s by Polish researcher Z. Pawlak and developed by S. Greco, B. Matarazzo and R. Slowinski as the Data-based Rough Set Approach (DRSA).

In this proposal we will apply the DRSA to hybrid bankruptcy prediction modeling for small businesses. In this modeling the discrimination analysis results are used to explain the decision rules obtained from regional experts.

Keywords

Multi Criteria Decision Analysis (MCDA), Preference Modelling, Discriminate Analysis, Hybrid Model, Rough Set Theory, Dominance-Based Rough Set Approach (DRSA).

Introduction

There are many preference modelling methods where the model is adjusted to the decisional situation by determination of parameter values. In practice, the task of parameter values determination is not easy, because the DM does not understand the decisional situation in terms of parameters.

More realistic is the model construction from examples, called learning approach based on examples. Usually, referential activities are well known to the DM, and he is able to order them and express his preferences in this way. Simply, he shows us how he does his job.

However, processing of the information, coming from the DM creates certain difficulties because of the lack of consistency of examples and contradictions.

According, to Polish researcher Zdzislaw Pawlak the preference model should neither correct nor ignore these contradictions. It should rather consider them to induce certain and uncertain decision rules. The exploration of this model will allow us give to the DM two kinds of certain and possible recommendations.

The Rough Set Theory suggested by [Pawlak, 1982] respects the above principle. In 2001 Greco, Matarazzo and Slowinski have introduced Dominance-Based Rough Set Approach (DRSA) which is an extension of rough set theory for Multicriteria Decision Analysis (MCDA). The main change comparing to the classical rough sets is the substitution of the indiscernibility relation by a dominance relation, which permits to deal with inconsistencies typical for considerations of criteria and preferences. In DRSA, examples of decision making are presented in the form of a decision table.

1. Decision table

Formally, a decision table is the 4-tuplex T = (U,Q,V,f) where U is a finite set of objects, Q is a finite set of criteria, where Q is divided into non-empty condition criteria set C and the decision criterion d. Notice, that f(x,q) which belongs to V_q is an evaluation of the object x on criterion q which belongs to the set C, while f(x,d) is the class assignment (decision value) of the object.

Table 1

Example of decision table

***object* (*cand*)**	***q_1* *Piano***	***q_2* *Violin***	***q_3* *Trumpet***	***q_4* *Guitar***	***d* (*decision*)**
x_1	4	4	3	4	A
x_2	5	5	2	4	A
x_3	4	4	2	4	A
x_4	4	4	2	4	R
x_5	5	5	2	4	A
x_6	4	4	2	3	A
x_7	4	3	2	3	R

As an illustrative example, consider the problem selection of candidates to a high music school by the committee. The candidates are assigned to two disjunctive classes: accepted (A) or rejected (R) (see Table 1). The performance of each candidate is described by four criteria: level piano, violin, trumpet and guitar playing, each taking one of three possible values: 3;4;5 with respect to two first criteria and 2;3;4 with respect to the two second criteria. Criteria are ordered so that grater values are better.

The classical rough set approach allows us to obtain a partition of indiscernible classes of objects in the decision table. The objects are indiscernible if their performance is described by the same conjunction of the values with respect to the conditional criteria (C). The inconsistency of the first kind is identified by the classical rough set approach if two indiscernible objects correspond to two different disjunctive decision classes (d).

In our example, it is the case of $\{x_3,x_4\}$. The classical rough set approach doesn't allow us to identify the second kind of inconsistency where a principle of dominance is not respected. In our example, it is the case of the relation between candidates x_4 which dominates and x_6 which is dominated and the first one is rejected while the second one is accepted. This is why an extension of the classical rough set approach was suggested by [Greco, Matarazzo and Slowinski, 2001], called DRSA (Dominance-based Rough Set Approach).

2. Dominace-based Rough Set Approach (DRSA)

Let $\succsim_q$ be outranking relation such that:

$$x \succsim_q y \Leftrightarrow f(x,q) \geq f(y,q) \tag{1}$$

This relation is straightforward for gain-type (the more is better), for cost-type (the less, the better) can be satisfied by negating the values from V_q.

Dominance

We say that x **dominates** y with respect to $P \subseteq C$ denoted by xD_Py, if x is better than y on every criterion from P, $x \succsim_q y$, $\forall q \in P$.

Given $P \subseteq C$ and $x \in U$, let

$$\begin{aligned} D_P^+(x) &= \{y \in U : yD_Px\} \\ D_P^-(x) &= \{y \in U : xD_Py\} \end{aligned} \tag{2}$$

represent **P-dominating** and **P-dominated** sets for each $x \in U$, respectively.

Next, with respect to the decisional attribute we consider n disjoint classes CL=$\{CL_t, t \in T\}$, where $Cl_t = \{x \in U : f(x,d) = t\}$. Each object x$\in U$ is assigned to one and only one classCl_t. The classes are preference-ordered according to an increasing order of class indices. This is why for each class t cumulated decision classes are considered "at most" or "at least", defined respectively (3), as:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s \ et\ Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s \quad for\ each\ t \in T. \tag{3}$$

In DRSA, to identify inconsistency cases, we do **P-lower** and the **P-upper approximations** (4) of $Cl_t^{\geq}$ and$Cl_t^{\leq}$, t$\in T$ for each $P \subseteq C$ denoted as $\underline{P}(Cl_t^{\geq})$, $\overline{P}(Cl_t^{\geq})$ and $\underline{P}(Cl_t^{\leq})$, $\overline{P}(Cl_t^{\leq})$, respectively are defined:

$$\begin{aligned}
&\underline{P}(Cl_t^{\geq}) = \{x \in U : D_P^{+}(x) \subseteq Cl_t^{\geq}\},\\
&\overline{P}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_P^{+}(x),\\
&\underline{P}(Cl_t^{\leq}) = \{x \in U : D_P^{-}(x) \subseteq Cl_t^{\leq}\},\\
&\overline{P}(Cl_t^{\leq}) = \bigcup_{x \in Cl_t^{\leq}} D_P^{-}(x).
\end{aligned} \tag{4}$$

The **P-boundaries (**P-doubtful regions) (5) of $Cl_t^{\geq}$ and $Cl_t^{\geq}$ are defined as:

$$\begin{aligned}
&BN_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}),\\
&BN_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl)_t^{\leq}.
\end{aligned} \tag{5}$$

Coming back to our example we have only two classes, which will correspond to two cumulated classes: $Cl^{\geq} = Cl_A$ and $Cl^{\leq} = Cl_R$. We identify all **P-dominating** and **P-dominated** sets for each candidate and we are doing **P-lower** and **P-upper** approximations for each class.

$$\begin{aligned}
&\underline{P}(Cl_A) = \{x_1, x_2, x_5\},\\
&\overline{P}(Cl_A) = \{x_1, x_2, x_3, x_4, x_5, x_6\}\\
&\underline{P}(Cl_R) = \{x_7\},\\
&\overline{P}(Cl_R) = \{x_3, x_4, x_6, x_7\},\\
&BN_C(Cl_A) = BN_C(Cl_R) = \{x_3, x_4, x_6\}.
\end{aligned} \tag{6}$$

We can see that two P-doubtful regions allow us to identify three candidates where decision of committee was inconsistent. Two of them have been obtained the same evaluations, but x_3 was accepted and x_4 was rejected. The evaluation of sixth candidate is dominated by these of x_3 and x_4, but he was accepted. This example illustrates two kinds of inconsistencies which are identified by the DRSA.

3. Decision rules

On the basis of the approximations obtained by means of the dominance relations, it is possible to derive a generalized description of the preferential information contained in the decision table, in terms of decision rules.

In our example we have obtained the following rules:

Rule 1: If $q_1 \geq 4 \wedge q_3 = 3$ then $d = A$
Rule 2: If $q_1 = 5$ then $d = A$
Rule 3: If $q_2 = 3$ then $d = R$
Rule 4: If $q_1 = 4 \wedge q_2 = 4 \wedge q_3 = 2$ then $d = A \vee d = R$.

Forth rule covers three examples from the decision table where a decision of the committee was contradictory.

The reduced subset of the criteria, which give the same candidate classification as original set is composed of.

4. Application of DRSA to bankrupcy prediction modeling for small and medium businesses

Application of the DRSA in the prediction of the bankruptcy modelling of SMALL AND MEDIUM-SIZED BUSINESSES (SMB) in Abitibi--Témiscamingue was the object of the Master's essay of my student Mohamed Kaba in 2008. In this modelling we coupled results of the discriminate analysis [Altman, 1968] and of the DRSA analysis.

The discriminate analysis consists in calculation of the score Z according to the formula proposed by Altman:

$$Z = 1{,}2X_1+1{,}4X_2 + 3{,}3X_3 + 0{,}6X_4+0{,}999X_5 \qquad (7)$$

where:

X_1 – is working capital / total assets,
X_2 – is retained earnings / total assets,
X_3 – is earning before interest and taxes / total assets,
X_4 – is market value of equity / total liabilities,
X_5 – is sales / total assets.

The value of Z was calculated for seven manufacture companies chosen among thirty in Abitibi-Témiscamingue, the region of Quebec, for a period of five years. According to Altman the discriminate value of Z for the prosperous companies is equal to at least 2.67. The value of Z for any SMB in manufacture sector of Abitibi-Témiscamingue which went bankrupt was negative. The manufacture SMB whose Z was between 0 and 2.67, were classified as on the edge of bankruptcy.

Then, in the analysis the DRSA we considered three decisional classes of SMB: prosperous-(P), bankrupt-(B) and on the edge of bankruptcy-(M). These seven SMB were evaluated by four experts in charge of regional development with respect to five criteria on the scale of seven levels (Tables 2 and 3).

Table 2

Scale of evaluation

Criterion level	Signification
1	Very strong(Very developed)
2	Strong (Developed)
3	Strong enough (Developed enough)
4	Weak (Little developed)
5	Weak enough(Very little developed)
6	Very weak (Not developed)
7	I have no idea

The criteria were five capacities of the successful promoters (in according to Filion [1991]): leadership, vision, network of contacts, management abilities and differentiation. The experts had the perfect knowledge of the promoters, but they did not know results of the discriminate analysis.

Table 3

Decision Table

Prom.	Leader.	Vision	Network	Manage.	Dif	d
1	1	2	1	1	2	P
2	2	2	4	4	3	B
3	4	2	4	6	2	M
4	4	4	4	4	2	M
5	4	2	4	2	2	M
6	3	2	1	2	4	P
7	4	5	6	6	4	B

The calculations were done with software package 4eMka2 developed by Laboratory of Intelligent Decision Support System (IDSS) in the Institute of Computing Science, Poznan University of Technology. We identified three reduced subsets of criteria:

1. Management abilities, Network of contacts.
2. Management abilities, Leadership and Vision.
3. Management abilities, Leadership and Differentiation.

In fact, since Management abilities are in each reduced subset of criteria, this criterion cannot be ignored in the analysis without influencing the quality of approximation. In this case, the quality of approximation was not very high and equal to 0.57 what can be explain by difficulties of the experts to distinguish between the class of the bankruptcy (B) and the class on the edge of bankruptcy (M) (Rule 5).

We have obtained five decision rules:

Rule 1: **If** Network of contacts very developed **then** SMB are prosperous;

Rule 2: **If** Management abilities at least strong **then** SMB are at least on the edge of bankruptcy

Rule3: **If** Network of contacts at most little developed **then** SMB are at most on the edge of bankruptcy;

Rule 4: **If** Vision at most very little developed **then** bankruptcy.

Rule 5: **If** Management abilities at most weak and vision at least little developed **then** bankruptcy or on the edge of bankruptcy.

From these rules we can see that the criteria **Network of contacts** and **Vision** also are very important. If **Network of contacts** is very developed than SMB are prosperous, if this criterion is evaluated on the level at most little developed, SMB are on the edge of bankruptcy or on the level of the bankruptcy. **Vision** is discriminatory in the negative terms. If it is at most very little developed then SMB are bankrupt.

Conclusions

In this paper, the Dominance-Based Rough Set Approach is proposed as an operational tool for aid to the regional developing and assistance of Small and Medium Businesses. This problem was treated by [Dimitras et al., 1999; Slowinski and Zopounidas, 1995], but based on classical approach of rough sets and used for the prediction of corporate businesses failure in the particular bank of Greece. They considered only financial criteria. For regional development problems, it is important to consider quantitative and qualitative criteria. For this reason our proposal consists of hybrid model composed of quantitative and qualitative data. The quantitative part (Z-score) is used to identify decision classes of the rough set model which is based on qualitative criteria. The prediction model has the form of decision rules which are particularly useful for evaluation of new promoters.

Z-score method was adapted to evaluate SMB and it was validated for regional population of manufacture businesses. We observed that the discriminate values in the case of SMB are lower than these suggested by Altman. In particular, the lower value of Z, suggested by Altman to be close to 1.8, is rather close to zero. This explains the difficulties which experts had in distinguishing between two classes: bankruptcy and on the edge of bankruptcy.

References

Altman E. (1968): *Financial Ratios, Discriminate Analysis and the Prediction of Corporate Bankruptcy*. "Journal of Finance", 23, pp. 589-609.

Dimitras A.I., Slowinski R., Susmaga R., Zopounidas C. (1999): *Business Failure Prediction Using Rough Sets.* "European Journal of Operational Research", 114, (2), pp. 263-280.

Filion L.-J. (1991): *Visions et relations: clefs du succès de l'entrepreneur.* ADP/Éditions de l'entrepreneur, Montréal.

Greco S., Matarazzo B., Slowinski R. (2001): *Rough Set Theory for Multicriteria Decision Analysis.* "European Journal of Operational Research", 129, (1), pp. 1-47.

Kaba M. (2008): *Les causes de la faillite des PME de l'Abitibi-Témiscamingue: Cas des PME manufacturières de MRC Rouyn-Noranda.* Master's essay. Université du Québec en Abitibi-Témiscamingue, Québec.

Pawlak Z. (1982): *Rough Sets.* "International Journal of Information and Computer Science", Vol. 11, pp. 341-356.

Pawlak Z. (1991): *Rough Sets, Theoretical Reasoning about Data*. Kluwer Academic Publishers, Dordrecht-Boston-London.

Slowinski R. (2005): *Rough Set Based Decision Support. Chapter 16*. In: *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*. Edited by E.K. Burke and G. Kendall. Springer-Verlag, New York, pp. 475-527.

Slowinski R., Zoponidas C. (1995): *Application of the Rough Set Approach to Evaluation of Bankruptcy Risk.* "International Journal of Intelligent Systems in Accounting, Finance and Management", 4, pp. 27-41.

Zaras K. (2004): *Rough Approximation of a Preference Relation by a Multi-Attribute Dominance for Deterministic, Stochastic and Fuzzy Decision Problems*. "European Journal of Operational Research", 159, (1), pp. 196-206.