A New Rough Set Approach to Multicriteria and Multiattribute Classification

Salvatore Greco[1], Benedetto Matarazzo[1], and Roman Slowinski[2]

[1] Faculty of Economics, University of Catania,
55, Corso Italia, I-95129 Catania, Italy
[2] Institute of Computing Science, Poznan University of Technology,
3a, Piotrowo 60-965 Poznan, Poland

1 Introduction

As pointed out by Greco, Matarazzo and Slowinski [1] the original rough set approach does not consider *criteria*, i.e. attributes with ordered domains. However, in many real problems the *ordering properties* of the considered attributes may play an important role. E.g. in a bankruptcy evaluation problem, if firm A has a low value of the debt ratio (Total debt/Total assets) and firm B has a large value of the same ratio, within the original rough set approach the two firms are just discernible, but no preference is established between them two with respect to the attribute "debt ratio". Instead, from a decisional point of view, it would be better to consider firm A as preferred to firm B, and not simply "discernible", with respect to the attribute in question.

Motivated by the previous considerations, Greco, Matarazzo and Slowinski [2] proposed a new rough set approach to take into account the ordering properties of criteria. Similarly to the original rough set analysis, the proposed approach is based on approximations of a partition of objects in some pre-defined categories. However, differently from the original approach, the categories are ordered from the best to the worst and the approximations are built using dominance relations, being specific order binary relations, instead of indiscernibility relations, being equivalence relation. The considered dominance relations are built on the basis of the information supplied by condition attributes which are all criteria. In this paper we generalize this approach considering a set of condition attributes which are not all criteria.

The paper is organized in the following way. In the second section, the main concepts of the rough approximation based on criteria and attributes are introduced. In section 3 we apply the proposed approach to a didactic example to compare the results with the original rough set approach. Final section groups conclusions.

2 Multicriteria and multiattribute rough approximation

As usual, by an *information table* we understand the 4-tuple $S = \langle U, Q, V, f \rangle$, where U is a finite set of objects, Q is a finite set of *attributes*, $V = \bigcup_{q \in Q} V_q$

L. Polkowski and A. Skowron (Eds.): RSCTC'98, LNAI 1424, pp. 60–67, 1998.

and V_q is a domain of the attribute q, and $f : U \times Q \rightarrow V$ is a total function such that $f(x, q) \in V_q$ for every $q \in Q$, $x \in U$, called an *information function* (cf. Pawlak [4]).

Moreover, an information table can be seen as *decision table* assuming the set of attributes $Q = C \cup D$ and $C \cap D = \emptyset$, where set C contains so called *condition attributes*, and D, *decision attributes.*

In general, the notion of attribute differs from that of criterion, because the domain (scale) of a criterion has to be ordered according to a decreasing or increasing preference, while the domain of the attribute does not have to be ordered. We will use the notion of criterion only when the preferential ordering of the attribute domain is important in a given context. Formally, for each $q \in C$ which is a criterion there exists an *outranking* relation (Roy [6]) S_q on U such that xS_qy means "x is at least as good as y with respect to attribute q". We suppose that S_q is a total preorder, i.e. a strongly complete and transitive binary relation on U. Instead, for each attribute $q \in C$ which is not a criterion, there exists an *indiscernibility* relation I_q on U which, as usual in rough sets theory, is an equivalence binary relation, i.e. reflexive, symmetric and transitive. We denote by $C^>$ the subset of attributes being criteria in C and by $C^=$ the subset of attributes which are not criteria, such that $C^> \cup C^= = C$ and $C^> \cap C^= = \emptyset$. Moreover, for each $P \subseteq C$ we denote by $P^>$ the set of criteria contained in C, i.e. $P^> = P \cap C^>$, and by $P^=$ the set of attributes which are not criteria contained in C, i.e. $P^= = P \cap C^=$.

Let R_P be a reflexive and transitive binary relation on U, i.e. R_P is a a partial preorder on U, defined on the basis of the information given by the attributes in $P \subseteq C$. More precisely, for each $P \subseteq C$ we can define R_P as follows: $\forall x, y \in U$, xR_Py if xS_qy for each $q \in P^>$ (i.e. x outranks y with respect to all the criteria in P) and xI_qy for each $q \in P^=$ (i.e. x is indiscernible with y with respect to all the attributes which are not criteria in P). If $P \subseteq C^>$ (i.e. if all the attributes in P are criteria) and xR_Py, then x outranks y with respect to each $q \in P$ and therefore we can say that x dominates y with respect to P. Let us observe that in general $\forall x, y \in U$ and $\forall P \subseteq C$, xR_Py if and only if x dominates y with respect to $P^>$ and x is indiscernible with y with respect to $P^=$.

Furthermore let $\mathbf{Cl} = \{\mathrm{Cl}_t, t \in T\}$, $T = \{1, \cdots, n\}$, be a set of classes of U, such that each $x \in U$ belongs to one and only one $\mathrm{Cl}_t \in \mathbf{Cl}$. We suppose that $\forall r, s \in T$, such that $r > s$, the elements of Cl_r are preferred (strictly or weakly (Roy [6])) to the elements of Cl_s. More formally, if S is a comprehensive outranking relation on U, i.e. if $\forall x, y \in U$ xSy means "x is at least as good as y", we suppose

$$[x \in \mathrm{Cl}_r, y \in \mathrm{Cl}_s, r > s] \Rightarrow [xSy \text{ and not } ySx].$$

In simple words the classes $\mathbf{Cl}$ represent a comprehensive evaluation of the objects in U: the worst objects are in Cl_1, the best objects are in Cl_n, the other objects belong to the remaining classes Cl_r, according to an evaluation improving with the index $r \in T$. E.g. considering a credit evaluation problem we can have $T = \{1, 2, 3\}$, $\mathbf{Cl} = \{\mathrm{Cl}_1, \mathrm{Cl}_2, \mathrm{Cl}_3\}$ and Cl_1 represents the class of the "unacceptable” firms, Cl_2 represents the class of “uncertain” firms, Cl_3 represents the class of “acceptable” firms.

Starting from the classes in **Cl**, we can define the following sets:

$$\mathrm{Cl}_t^{\geq} = \bigcup_{s \geq t} \mathrm{Cl}_s,$$

$$\mathrm{Cl}_t^{\leq} = \bigcup_{s \leq t} \mathrm{Cl}_s.$$

Let us remark that $\mathrm{Cl}_1^{\geq} = \mathrm{Cl}_n^{\leq} = U$, $\mathrm{Cl}_n^{\geq} = \mathrm{Cl}_n$ and $\mathrm{Cl}_1^{\leq} = \mathrm{Cl}_1$. Furthermore $\forall t = 2, \cdots, n$ we have:

$$\mathrm{Cl}_{t-1}^{\leq} = U - \mathrm{Cl}_t^{\geq} \tag{1}$$

and

$$\mathrm{Cl}_t^{\geq} = U - \mathrm{Cl}_{t-1}^{\leq}. \tag{2}$$

For each $P \subseteq C$, let be

$$R_P^{+}(x) = \{y \in U : yR_Px\},$$

$$R_P^{-}(x) = \{y \in U : xR_Py\}.$$

Let us observe that, given $x \in U$, $R_P^{+}(x)$ represents the set of all the objects $y \in U$ which dominates x with respect to $P^{>}$ (i.e. the criteria of P) and are indiscernible with x with respect to $P^{=}$ (i.e. the attributes of P). Analogously $R_P^{-}(x)$ represents the set of all the objects $y \in U$ which are dominated by x with respect to $P^{>}$ and are indiscernible with x with respect to $P^{=}$.

We say that, with respect to $P \subseteq C$ and $t \in T$, $x \in U$ belongs to $\mathrm{Cl}_t^{\geq}$ *without any ambiguity* if $x \in \mathrm{Cl}_t^{\geq}$ and $y \in \mathrm{Cl}_t^{\geq}$ for all the objects $y \in U$ dominating x with respect to $P^{>}$ and indiscernible with x with respect to $P^{=}$.

Formally, remembering the reflexivity of R_P, we can say that $x \in U$ belongs to $\mathrm{Cl}_t^{\geq}$ without any ambiguity if $R_P^{+}(x) \subseteq \mathrm{Cl}_t^{\geq}$. Furthermore we say that, with respect to $P \subseteq C$ and $t \in T$, $y \in U$ *could belong* to $\mathrm{Cl}_t^{\geq}$ if there exists at least one object $x \in \mathrm{Cl}_t^{\geq}$ such that y dominates x with respect to $P^{>}$ and y is indiscernible with x with respect to $P^{=}$, i.e. $y \in R_P^{+}(x)$. Our definitions of lower and upper approximation are based on the previous ideas. Thus, with respect to $P \subseteq C$, the set of all the objects belonging to $\mathrm{Cl}_t^{\geq}$ without any ambiguity constitutes the lower approximation of $\mathrm{Cl}_t^{\geq}$, while the set of all the objects which could belong to $\mathrm{Cl}_t^{\geq}$ constitutes the upper approximation of $\mathrm{Cl}_t^{\geq}$.

Formally, $\forall t \in T$ and $\forall P \subseteq C$ we define the lower approximation of $\mathrm{Cl}_t^{\geq}$ with respect to P, denoted by $\underline{P}\mathrm{Cl}_t^{\geq}$, and the upper approximation of $\mathrm{Cl}_t^{\geq}$ with respect to P, denoted by $\overline{P}\mathrm{Cl}_t^{\geq}$, as:

$$\underline{P}\mathrm{Cl}_t^{\geq} = \{x \in U : R_P^{+}(x) \subseteq \mathrm{Cl}_t^{\geq}\},$$

$$\overline{P}\mathrm{Cl}_t^{\geq} = \bigcup_{x \in \mathrm{Cl}_t^{\geq}} R_P^{+}(x).$$

We say that, with respect to $P \subseteq C$ and $t \in T$, $x \in U$ belongs to $\mathrm{Cl}_t^{\leq}$ without any ambiguity if $x \in \mathrm{Cl}_t^{\leq}$ and $y \in \mathrm{Cl}_t^{\leq}$ for all the objects $y \in U$ dominated by x with respect to $P^{>}$ and indiscernible with x with respect to $P^{=}$.

Formally, remembering the reflexivity of R_P, we can say that $x \in U$ belongs to $\mathrm{Cl}_t^{\leq}$ without any ambiguity if $R_P^{-}(x) \subseteq \mathrm{Cl}_t^{\geq}$. Furthermore we say that with respect to $P \subseteq C$, $y \in U$ could belong to $\mathrm{Cl}_t^{\leq}$ if there exists at least one object $x \in \mathrm{Cl}_t^{\leq}$ such that x dominates y with respect to $P^{>}$ and y is indiscernible with x with respect to $P^{=}$, i.e. $y \in R_P^{-}(x)$. Thus, with respect to $P \subseteq C$, the set of all the objects belonging to $\mathrm{Cl}_t^{\leq}$ without any ambiguity constitutes the lower approximation of $\mathrm{Cl}_t^{\leq}$, while the set of all the objects which could belong to $\mathrm{Cl}_t^{\leq}$ constitutes the upper approximation of $\mathrm{Cl}_t^{\leq}$.

Formally, $\forall t \in T$ and $\forall P \subseteq C$, we define the lower approximation of $\mathrm{Cl}_t^{\leq}$ with respect to P, denoted by $\underline{P}\mathrm{Cl}_t^{\leq}$, and the upper approximation of $\mathrm{Cl}_t^{\leq}$ with respect to P, denoted by $\overline{P}\mathrm{Cl}_t^{\leq}$, as:

$$\underline{P}\mathrm{Cl}_t^{\leq} = \{x \in U : R_P^{-}(x) \subseteq \mathrm{Cl}_t^{\leq}\},$$

$$\overline{P}\mathrm{Cl}_t^{\leq} = \bigcup_{x \in \mathrm{Cl}_t^{\leq}} R_P^{-}(x).$$

The P-*boundary* (doubtful region) of $\mathrm{Cl}_t^{\geq}$ and $\mathrm{Cl}_t^{\leq}$ are respectively defined as

$$\mathrm{Bn}_P(\mathrm{Cl}_t^{\geq}) = \overline{P}\mathrm{Cl}_t^{\geq} - \underline{P}\mathrm{Cl}_t^{\geq},$$

$$\mathrm{Bn}_P(\mathrm{Cl}_t^{\leq}) = \overline{P}\mathrm{Cl}_t^{\leq} - \underline{P}\mathrm{Cl}_t^{\leq}.$$

$\forall t \in T$ and $\forall P \subseteq C$ we define the *accuracy* of the approximation of $\mathrm{Cl}_t^{\geq}$ and $\mathrm{Cl}_t^{\leq}$ as the ratios:

$$\alpha_P(\mathrm{Cl}_t^{\geq}) = \frac{\mathrm{card}(\underline{P}\mathrm{Cl}_t^{\geq})}{\mathrm{card}(\overline{P}\mathrm{Cl}_t^{\geq})},$$

$$\alpha_P(\mathrm{Cl}_t^{\leq}) = \frac{\mathrm{card}(\underline{P}\mathrm{Cl}_t^{\leq})}{\mathrm{card}(\overline{P}\mathrm{Cl}_t^{\leq})},$$

respectively. The coefficient

$$\gamma_P(\mathrm{Cl}) = \frac{\mathrm{card}(U - ((\bigcup_{t \in T} \mathrm{Bn}_P(\mathrm{Cl}_t^{\leq})) \cup (\bigcup_{t \in T} \mathrm{Bn}_P(\mathrm{Cl}_t^{\geq}))))}{\mathrm{card}(U)}$$

is called the *quality of approximation of partition* **Cl** by set of attributes P, or in short, *quality of classification*. It expresses the ratio of all P-correctly classified objects to all objects in the table. Each minimal subset $P \subseteq C$ such that $\gamma_P(\mathbf{Cl}) = \gamma_C(\mathbf{Cl})$ is called a *reduct* of **Cl** and denoted by $\mathrm{RED}_{\mathrm{Cl}}$. Let us remark that an information table can have more than one reduct. The intersection of all reducts is called the *core* and denoted by $\mathrm{CORE}_{\mathrm{Cl}}$.

3 An example

The following example (based on a previous example proposed by Pawlak [5]) illustrates the concepts introduced above. In Table 3, twelve warehouses are described by means of five attributes:

- A_1, capacity of the sales staff,
- A_2, perceived quality of goods,
- A_3, high traffic location,
- A_4, geographical region,
- A_5, warehouse profit or loss.

In fact, A_1, A_2 and A_3 are criteria, because their domains are ordered, A_4 is an attribute, whose domain is not ordered, and A_5 is a decision attribute, defining two ordered decision classes. More in detail we have that

- with respect to A_1 "high" is better than "medium" and "medium" is better than "low",
- with respect to A_2 "good" is better than "medium",
- with respect to A_3 "yes" is better than "no",
- with respect to A_5 "profit" is better than "loss".

Table 1. Example of an information table.

Warehouse	A_1	A_2	A_3	A_4	A_5
1	High	Good	no	A	Profit
2	Medium	Good	no	A	Loss
3	Medium	Good	no	A	Profit
4	Low	Medium	no	A	Loss
5	Medium	Medium	yes	A	Loss
6	High	Medium	yes	A	Profit
7	Medium	Medium	no	A	Profit
8	High	Good	no	B	Profit
9	Medium	Good	no	B	Profit
10	Low	Medium	no	B	Loss
11	Medium	Medium	yes	B	Profit
12	High	Medium	yes	B	Profit

3.1 The results from classical rough set approach

By means of the classical rough set approach we approximate the class Cl_1 of the warehouses making loss and the class Cl_2 of the warehouses making profit. It is clear that $C = \{A_1, A_2, A_3, A_4\}$ and $D = \{A_5\}$. The C-lower approximations, the C-upper approximations and the C-boundaries of sets Cl_1 and Cl_2 are respectively:

$\underline{C}\mathrm{Cl}_1 = \{4, 5, 10\}$, $\overline{C}\mathrm{Cl}_1 = \{2, 3, 4, 5, 10\}$, $\mathrm{Bn}_C(\mathrm{Cl}_1) = \{2, 3\}$, $\underline{C}\mathrm{Cl}_2 = \{1, 6, 7, 8, 9, 11, 12\}$, $\overline{C}\mathrm{Cl}_2 = \{1, 2, 3, 6, 7, 8, 9, 11, 12\}$, $\mathrm{Bn}_C(\mathrm{Cl}_2) = \{2, 3\}$. Therefore the accuracy of the approximation is 0.6 for the class of warehouses making loss and 0.78 for the class of warehouses making profit and the quality of classification is equal to 0.83. There is only one reduct which is also the core, i.e. $\mathrm{Red}(C) = \mathrm{Core}(C) = \{A_1, A_2, A_3, A_4\}$.

Using the algorithm LERS (Grzymala-Busse [3]) the following set of decision rules is obtained from the considered decision table 3 (Table 1) (within brackets there are the objects supporting the corresponding rules):

1. if $f(x, A_1)$ = high, then $x \in \mathrm{Cl}_2$ (1, 6, 8, 12)
2. if $f(x, A_1)$ = medium and $f(x, A_4) = B$, then $x \in \mathrm{Cl}_2$ (9, 11)
3. if $f(x, A_1)$ = medium and $f(x, A_2)$ = medium and $f(x, A_3)$ =no, then $x \in \mathrm{Cl}_2$ (7)
4. if $f(x, A_1)$ = medium and $f(x, A_2)$ = good and $f(x, A_4) = A$, then $x \in \mathrm{Cl}_1$ or $x \in \mathrm{Cl}_2$ (2, 3)
5. if $f(x, A_1)$ = low, then $x \in \mathrm{Cl}_1$ (4, 10)
6. if $f(x, A_1)$ = medium and $f(x, A_3)$ = yes and $f(x, A_4) = A$, then $x \in \mathrm{Cl}_1$ (5)

3.2 The results from approximations by dominance and indiscernibility relations

With this approach we approximate the class $\mathrm{Cl}_1^{\leq}$ of the warehouses at most making loss and the class $\mathrm{Cl}_2^{\geq}$ of the warehouses at least making profit. Since only two classes are considered, we have $\mathrm{Cl}_1^{\leq} = \mathrm{Cl}_1$ and $\mathrm{Cl}_2^{\geq} = \mathrm{Cl}_2$. When a larger number of classes is considered this equalities are not satisfied.

The C-lower approximations, the C-upper approximations and the C–boundaries of sets $\mathrm{Cl}_1^{\leq}$ and $\mathrm{Cl}_2^{\geq}$ are respectively: $\underline{C}\mathrm{Cl}_1^{\leq} = \{4, 10\}$, $\overline{C}\mathrm{Cl}_1^{\leq} = \{2, 3, 4, 5, 7, 10\}$, $\mathrm{Bn}_C(\mathrm{Cl}_1^{\leq}) = \{2, 3, 5, 7\}$, $\underline{C}\mathrm{Cl}_2^{\geq} = \{1, 6, 8, 9, 11, 12\}$, $\overline{C}\mathrm{Cl}_2^{\geq} = \{1, 2, 3, 5, 6, 7, 8, 9, 11, 12\}$, $\mathrm{Bn}_C(\mathrm{Cl}_2^{\geq}) = \{2, 3, 5, 7\}$. Therefore, the accuracy of the approximation is 0.33 for $\mathrm{Cl}_1^{\leq}$ and 0.6 for $\mathrm{Cl}_2^{\geq}$ while the quality of classification is equal to 0.67. There is only one reduct, which is also the core, i.e. $\mathrm{RED}_{\mathrm{Cl}}(C) = \mathrm{CORE}_{\mathrm{Cl}}(C) = \{A_1, A_4\}$.

The following minimal set of decision rules can be obtained from the considered decision table (within parentheses there are the objects supporting the corresponding rules):

1. if $f(x, A_1)$ is high, then $x \in \mathrm{Cl}_2^{\geq}$ (1, 6, 8, 12)
2. if $f(x, A_1)$ is at least medium and $f(x, A_4)$ is B, then $x \in \mathrm{Cl}_2^{\geq}$ (8, 9, 11, 12)
3. if $f(x, A_1)$ is low, then $x \in \mathrm{Cl}_1^{\leq}$ (4, 10)
4. if $f(x, A_1)$ is medium and $f(x, A_4)$ is A, then $x \in \mathrm{Cl}_1^{\leq}$ or $x \in \mathrm{Cl}_2^{\geq}$ (2, 3, 5, 7).

3.3 Comparison of the results

The advantages of the rough set approach based on dominance and indiscernibility relations over the original rough set analysis, based on the indiscernibility relation, can be summarized in the following points.

The results of the approximation are more satisfactory. This improvement is represented by a smaller reduct ($\{A_1, A_4\}$ against $\{A_1, A_2, A_3, A_4\}$). Let us observe that even if the quality of the approximation is deteriorated (0.67 vs. 0.83), this is another point in favour of the proposed approach. In fact, this difference is due to the warehouses 5 and 7. Let us notice that with respect to the attributes A_1, A_2, A_3, which are criteria, warehouse 5 dominates warehouse 7 and with respect to the attribute A_4, which is not a criterion, the warehouse 5 and 7 are indiscernible. However warehouse 5 has a comprehensive evaluation worse than warehouse 7. Therefore, this can be interpreted as an inconsistency revealed by the approximation by dominance and indiscernibility that cannot be pointed out when we consider the approximation by indiscernibility only.

From the viewpoint of the quality of the set of decision rules extracted from the information table by the two approaches, let us remark that the decision rules obtained from the approximation by dominance and indiscernibility relations give a more synthetic representation of knowledge contained in the information table. The minimal set of decision rules obtained from the new approach has a smaller number of rules (4 against 6), uses a smaller number of attributes and descriptors than the set of the decision rules obtained from the classical rough set approach, obtains rules supported by a larger number of objects. Furthermore, let us observe that the rules obtained from the original rough sets approach present some problems with respect to their interpretation. E.g. rule 3 obtained by the original rough set approach says that if the capacity of the sale staff is medium, the perceived quality of goods is medium and if the warehouse is not in a high traffic location then the warehouse makes profit. One can expect that improving the quality of the warehouse, e.g. considering a warehouse with the same capacity of the sales staff and the same quality of goods but located in a high traffic location the warehouse should also make profit. Surprisingly, the warehouse 5 of the considered decision table has these characteristics but it makes loss. Finally, let us remark that rule 4 from the new approach is an approximate rule, as well as rule 4 from the classical approach. However, rule 4 from the new approach is based on a small number of descriptors and supports a greater number of actions.

4 Conclusion

We presented a new rough set approach whose purpose is to approximate sets of objects divided in ordered predefined categories considering criteria, i.e. attributes with ordered domains, jointly with attributes which are not criteria. We showed that the basic concepts of the rough sets theory can be restored in the new context. We also applied the proposed methodology to an exemplary problem approached also with the classical rough set analysis. The comparison of the results proved the usefulness of the new approach.

Acknowledgments

The research of the first two authors has been supported by grant no. 96.01658. $ct10$ from Italian National Research Council (CNR). The third author wishes to acknowledge financial support from State Committee for Scientific Research, KBN research grant no. 8 T11C 013 13, and from CRIT 2 - Esprit Project no. 20288. For the task of typing this paper, we are indebted to the high qualification of Ms Silvia Angilella.

References

1. Greco S., Matarazzo, B., Slowinski, R.: Rough approximation of a preference relation by dominance relations, ICS Research Report **16**, Warsaw University of Technology, Warsaw, (1996) and to be published on European Journal of Operational Research.
2. Greco, S., Matarazzo, B. Slowinski, R.: A new rough set approach to evaluation of bankruptcy risk, in C. Zopounidis (ed.), Operational Tools in the Management of Financial Risks, Kluwer, Dordrecht, (1998), 121–136.
3. Grzymala–Busse, J.W: LERS - a system for learning from examples based on rough sets, in R. Slowinski, (ed.), Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers, Dordrecht, (1992), 3–18.
4. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, (1991).
5. Pawlak, Z.: Rough set approach to knowledge-based decision support, European Journal of Operational Research, **99**, (1997), 48–57.
6. Roy, B.: Méthodologie Multicritère d'Aide à la Décision, Economica, Paris, (1985).