Multicriteria Classification by Dominance-Based Rough Set Approach

♦ ♦ ♦

Methodological Basis of the 4eMka System

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

[1]Faculty of Economics, University of Catania, Corso Italia, 55,
95129 Catania, Italy

[2]Institute of Computing Science, Poznan University of Technology,
60-965 Poznan, Poland

Abstract. We are considering multicriteria classification that differs from usual classification problems since it takes into account preference orders in the description of objects by condition and decision attributes. The well-known methods of knowledge discovery do not use the information about preference orders in multicriteria classification. It is worthwhile, however, to take this information into account as many practical problems are involving evaluation of objects on preference ordered domains. To deal with multicriteria classification we propose to use a Dominance-based Rough Set Approach (DRSA). This approach is different from the Classical Rough Set Approach (CRSA) because it takes into account preference orders in the domains of attributes and in the set of decision classes. Given a set of objects partitioned into pre-defined and preference-ordered classes, the new rough set approach is able to approximate this partition by means of dominance relations (instead of indiscernibility relations used in the CRSA). The rough approximation of this partition is a starting point for induction of "*if..., then...*" decision rules. The syntax of these rules is adapted to represent preference orders. The DRSA keeps the best properties of the CRSA: it analyses only facts present in data and possible inconsistencies are not corrected. Moreover, the new approach does not need any prior discretization of continuous-valued attributes. The usefulness of the DRSA and its advantages over the CRSA are presented on a real study of evaluation of the risk of business failure.

This paper presents a conceptual and methodological basis of the software system called ***4eMka***, developed at the Laboratory of Intelligent Decision Support Systems of the Institute of Computing Science, Poznan University of Technology, by a group of students of Computer Science in the framework of their bachelor's diploma.

1 Multicriteria classification

In traditional meaning, classification concerns an assignment of a set of objects to a set of pre-defined classes. The objects are characterized by a set of attributes and the classes are not necessarily ordered according to a preference. In practice, however, very often the attribute domains and classes are preference-ordered. The attributes with preference-ordered domains are called *criteria*. For example, classification of bank clients from the viewpoint of bankruptcy risk may involve such characteristics as "return on equity (ROE)", "return on investment (ROI)" and "return on sales (ROS)". The domains of these attributes are not simply ordered but involve a preference order since, from the viewpoint of bank managers, greater values of ROE, ROI or ROS are better for clients being analysed for bankruptcy risk. Neglecting this information in knowledge discovery may lead to wrong conclusions. Consider, for example, two firms, *A* and *B*, evaluated by a set of attributes including ROE. If firm *A* has a high value of ROE while firm *B* has a low value of ROE, and evaluations of these firms on other attributes are equal, then, from a bankruptcy risk point of view, firm *A* is better than (dominates) firm *B*. If, however, in the data sample set firm *A* has been assigned to a class of higher risk than firm *B*, then this is obviously inconsistent. This inconsistency will not be detected by usual knowledge discovery methods and possible conclusions derived by them from these data could be: "*if* ROE of a firm is low, *then* the firm is safe" and "*if* ROE of a firm is high, *then* the firm is risky", which is paradoxical. In order to discover this inconsistency one should analyse the data sample set from the viewpoint of the *dominance principle* that requires that an object having a better (in general, not worse) evaluation on considered attributes cannot be assigned to a worse class.

The above deficiency of knowledge discovery methods in the context of multicriteria classification can be repaired by proposing concepts and algorithms respecting the dominance principle.

A knowledge discovery method that deals with multicriteria classification is the Dominance-based Rough Set Approach (DRSA) proposed in (Greco, Matarazzo, Slowinski, 1998, 1999). It generalizes the Classical Rough Set Approach (CRSA) (Pawlak, 1982, 1991) by substituting the indiscernibility relation, used in CRSA, by a dominance relation, enabling discovery of inconsistencies with respect to the dominance principle. DRSA prepares, moreover, a conceptual ground for discovering rules having syntax concordant with the dominance principle.

2 Dominance-based Rough Set Approach (DRSA)

As it is usual in knowledge discovery methods, in DRSA, information about objects is represented in a *data matrix*, in which rows are labelled by *objects* and represent the values of attributes for each corresponding object, whereas columns are labelled by *attributes* and represent the values of each corresponding attribute for the objects.

Let U denote a finite set of objects (universe), Q a finite set of attributes, V_q a domain of the attribute q, and $f(x,q)$ a function assigning to each pair object-attribute (x,q) a value from V_q. The set Q is, in general, divided into set C of *condition attributes* and a *decision attribute* d.

In multicriteria classification, condition attributes are *criteria*. The notion of criterion involves a preference order in its domain while the domains of attributes, usually considered in machine discovery, are not preference-ordered.

Furthermore, decision attribute d makes a partition of U into a finite number of classes $\boldsymbol{Cl}=\{Cl_t,\ t\in T\}$, $T=\{1,...,n\}$. Each $x\in U$ belongs to one and only one class $Cl_t\in\boldsymbol{Cl}$. The classes from $\boldsymbol{Cl}$ are preference-ordered according to increasing order of class indices, i.e. for all $r,s\in T$, such that $r>s$, the objects from Cl_r are preferred to the objects from Cl_s.

In multicriteria classification, due to the preference order in the set of classes ***Cl***, the sets to be approximated are not the particular classes but *upward unions* and *downward unions* of classes, respectively:

$$Cl_t^{\geq} = \bigcup_{s\geq t} Cl_s, \quad Cl_t^{\leq} = \bigcup_{s\leq t} Cl_s, \quad t=1,...,n.$$

Union $Cl_t^{\geq}$ is the set of objects belonging to class Cl_t or to a more preferred class, while $Cl_t^{\leq}$ is the set of objects belonging to class Cl_t or to a less preferred class.

Notice that for t=2,...,n we have $Cl_t^{\geq}=U\text{-}Cl_{t-1}^{\leq}$, i.e. all the objects not belonging to class Cl_t or better, belong to class Cl_{t-1} or worse.

Let us remark that in usual classification problems, knowledge discovery methods extract knowledge with respect to a given class Cl_t dividing the universe U into class Cl_t (set of positive examples) and its complement $U\text{-}Cl_t$ (set of negative examples), t=1,…,n. However, such bipartitions do not take into account the preference order among classes. Thus, in multicriteria classification we need another type of bipartitions that divide the universe into upward and downward unions of classes $Cl_t^{\geq}$ and $Cl_{t-1}^{\leq}$, t=1,…,n. In result of this division, each object from the upward union $Cl_t^{\geq}$ is preferred to each object from the downward union $Cl_{t-1}^{\leq}$. When extracting knowledge with respect to an upward union $Cl_t^{\geq}$, we consider as positive all objects belonging to $Cl_t^{\geq}$ and as negative all objects belonging to $Cl_{t-1}^{\leq}$. Analogously, when extracting knowledge with respect to a downward union $Cl_{t-1}^{\leq}$, we consider as positive all objects belonging to $Cl_{t-1}^{\leq}$ and as negative all objects belonging to $Cl_t^{\geq}$. In this approach to knowledge discovery the dominance principle is applied as follows.

Let $\succeq_q$ be a *weak preference relation* on U (often called outranking (see Roy, 1985)) representing a preference on the set of objects with respect to criterion q; $x\succeq_q y$ means "x is at least as good as y with respect to criterion q". We say that x *dominates* y with respect to $P\subseteq C$ (or, shortly, x *P-dominates* y), denoted by xD_Py, if $x\succeq_q y$ for all $q\in P$. Assuming, without loss of generality, that domains of all criteria are ordered such that preference increases with the value, xD_Py is equivalent to: $f(x,q) \geq f(y,q)$ for all $q\in P$. Observe that for each $x\in U$, xD_Px, i.e. P-dominance is reflexive.

Given $P\subseteq C$ and $x\in U$, the "granules of knowledge" used in DRSA for approximation of the unions $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are:

- a set of objects dominating x, called *P-dominating set*, $D_P^+(x)=\{y\in U: yD_Px\}$,

- a set of objects dominated by x, called *P-dominated set*, $D_P^-(x)=\{y\in U: xD_Py\}$.

Given a set of criteria $P\subseteq C$, the inclusion of an object $x\in U$ to the upward union of classes $Cl_t^{\geq}$, t=2,…,n, creates an *inconsistency in the sense of dominance principle* if one of the following conditions holds:

- x belongs to class Cl_t or better but it is P-dominated by an object y belonging to a class worse than Cl_t, i.e. $x\in Cl_t^{\geq}$ but $D_P^+(x)\cap Cl_{t-1}^{\leq}\neq\emptyset$,
- x belongs to a worse class than Cl_t but it P-dominates an object y belonging to class Cl_t or better, i.e. $x\notin Cl_t^{\geq}$ but $D_P^-(x)\cap Cl_t^{\geq}\neq\emptyset$.

If, given a set of criteria $P\subseteq C$, the inclusion of $x\in U$ to $Cl_t^{\geq}$, t=2,…,n, creates an inconsistency in the sense of dominance principle, we say that x belongs to $Cl_t^{\geq}$ *with some ambiguity*. Thus, x belongs to $Cl_t^{\geq}$ *without any ambiguity* with respect to $P\subseteq C$, if $x\in Cl_t^{\geq}$ and there is no inconsistency in the sense of dominance principle. This means that all objects P-dominating x belong to $Cl_t^{\geq}$, i.e. $D_P^+(x)\subseteq Cl_t^{\geq}$.

Furthermore, *x could belong to* $Cl_t^{\geq}$ with respect to $P\subseteq C$ if one of the following conditions holds:

1) according to decision attribute d, x belongs to $Cl_t^{\geq}$,
2) according to decision attribute d, x does not belong to $Cl_t^{\geq}$ but it is inconsistent in the sense of dominance principle with an object y belonging to $Cl_t^{\geq}$.

In terms of ambiguity, x could belong to $Cl_t^{\geq}$ with respect to $P\subseteq C$, if x belongs to $Cl_t^{\geq}$ with or without any ambiguity. Due to reflexivity of the dominance relation D_P, conditions 1) and 2) can be summarized as follows: *x could belong* to class Cl_t or better, with respect to $P\subseteq C$, if among the objects P-dominated by x there is an object y belonging to class Cl_t or better, i.e. $D_P^-(x)\cap Cl_t^{\geq}\neq\emptyset$.

For $P\subseteq C$, the set of all objects belonging to $Cl_t^{\geq}$ without any ambiguity constitutes the *P-lower approximation* of $Cl_t^{\geq}$, denoted by $\underline{P}(Cl_t^{\geq})$, and the set of all objects that could belong to $Cl_t^{\geq}$ constitutes the *P-upper approximation* of $Cl_t^{\geq}$, denoted by $\overline{P}(Cl_t^{\geq})$:

$$\underline{P}(Cl_t^{\geq})=\{x\in U: D_P^+(x)\subseteq Cl_t^{\geq}\},\qquad \overline{P}(Cl_t^{\geq})=\{x\in U: D_P^-(x)\cap Cl_t^{\geq}\neq\emptyset\},\qquad \text{for } t=1,...,n.$$

Analogously, one can define *P-lower approximation* and *P-upper approximation* of $Cl_t^{\leq}$ as follows:

$$\underline{P}(Cl_t^{\leq})=\{x\in U: D_P^-(x)\subseteq Cl_t^{\leq}\},\qquad \overline{P}(Cl_t^{\leq})=\{x\in U: D_P^+(x)\cap Cl_t^{\leq}\neq\emptyset\},\qquad \text{for } t=1,...,n.$$

All the objects belonging to $Cl_t^{\geq}$ and $Cl_t^{\leq}$ with some ambiguity constitute the *P-boundary* of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, denoted by $Bn_P(Cl_t^{\geq})$ and $Bn_P(Cl_t^{\leq})$, respectively. They can be represented in terms of upper and lower approximations as follows:

$$Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}), \qquad Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}), \qquad \text{for } t=1,...,n.$$

P-lower and *P*-upper approximations of unions of classes $Cl_t^{\geq}$ and $Cl_t^{\leq}$ have an important property of *complementarity*. It says that if object x belongs without any ambiguity to class Cl_t or better, it is impossible that it could belong to class Cl_{t-1} or worse, i.e. $\underline{P}(Cl_t^{\geq}) = U - \overline{P}(Cl_{t-1}^{\leq})$. Due to complementarity property, $Bn_P(Cl_t^{\geq}) = Bn_P(Cl_{t-1}^{\leq})$, for t=2,...,n, which means that if x belongs with ambiguity to class Cl_t or better, it also belongs with ambiguity to class Cl_{t-1} or worse.

From the knowledge discovery point of view, *P*-lower approximations of unions of classes represent *certain knowledge* provided by criteria from $P \subseteq C$, while *P*-upper approximations represent *possible knowledge* and the *P*-boundaries contain *doubtful knowledge*.

The above definition of rough approximations are based on a strict application of the dominance principle. However, when defining non-ambiguous objects, it is reasonable to accept a limited proportion of negative examples, particularly for large data matrices. Such extended version of DRSA is called Variable-Consistency DRSA model (VC-DRSA) (Greco *et al.* 2000a).

For every $P \subseteq C$, the objects being consistent in the sense of dominance principle with all upward and downward unions of classes are *P-correctly classified*. For every $P \subseteq C$, the *quality of approximation of multicriteria classification* ***Cl*** by set of criteria *P* is defined as the ratio between the number of *P*-correctly classified objects and the number of all the objects in the data sample set. Since the objects *P*-correctly classified are those ones that do not belong to any *P*-boundary of unions $Cl_t^{\geq}$ and $Cl_t^{\leq}$, t=1,...,n, the quality of approximation of multicriteria classification ***Cl*** by set of attributes *P*, can be written as

$$\gamma_P(\boldsymbol{Cl}) = \frac{\text{card}\left(U - \left(\bigcup_{t\in T} Bn_P\left(Cl_t^{\leq}\right)\right)\right)}{card\,(U)} = \frac{\text{card}\left(U - \left(\bigcup_{t\in T} Bn_P\left(Cl_t^{\geq}\right)\right)\right)}{card\,(U)}.$$

$\gamma_P(\boldsymbol{Cl})$ can be seen as a measure of the quality of knowledge that can be extracted from the data matrix, where *P* is the set of criteria and ***Cl*** is the considered classification.

Each minimal subset $P \subseteq C$ such that $\gamma_P(\boldsymbol{Cl}) = \gamma_C(\boldsymbol{Cl})$ is called a *reduct* of ***Cl*** and is denoted by $RED_{\boldsymbol{Cl}}$. Let us remark that a data sample set can have more than one reduct. The intersection of all reducts is called the *core* and is denoted by $CORE_{\boldsymbol{Cl}}$. Criteria from $CORE_{\boldsymbol{Cl}}$ cannot be removed from the data sample set without deteriorating the knowledge to be discovered. This means that in set *C* there are three categories of criteria:

1) *indispensable* criteria included in the core,

2) *exchangeable* criteria included in some reducts but not in the core,

3) *redundant* criteria being neither indispensable nor exchangeable, thus not included in any reduct.

3 Extraction of decision rules

The dominance-based rough approximations of upward and downward unions of classes can serve to induce a generalized description of objects contained in the data matrix in terms of "*if..., then...*" decision rules. For a given upward or downward union of classes, $Cl_t^{\geq}$ or $Cl_s^{\leq}$, the decision rules induced under a hypothesis that objects belonging to $\underline{P}(Cl_t^{\geq})$ or $\underline{P}(Cl_s^{\leq})$ are *positive* and all the others *negative*, suggest an assignment to "class Cl_t or better" or to "class Cl_s or worse", respectively. On the other hand, the decision rules induced under a hypothesis that objects belonging to the intersection $\overline{P}(Cl_s^{\leq}) \cap \overline{P}(Cl_t^{\geq})$ are *positive* and all the others *negative*, are suggesting an assignment to some classes between Cl_s and Cl_t ($s<t$).

In multicriteria classification it is meaningful to consider the following five types of decision rules:

1) *certain* $D_{\geq}$-*decision rules*, providing lower profile descriptions for objects belonging to $Cl_t^{\geq}$ without ambiguity: *if* $f(x,q_1)\geq r_{q1}$ *and* $f(x,q_2)\geq r_{q2}$ *and* $\ldots f(x,q_p)\geq r_{qp}$, *then* $x\in Cl_t^{\geq}$,

2) *possible* $D_{\geq}$-*decision rules*, providing lower profile descriptions for objects belonging to $Cl_t^{\geq}$ with or without any ambiguity: *if* $f(x,q_1)\geq r_{q1}$ *and* $f(x,q_2)\geq r_{q2}$ *and* $\ldots f(x,q_p)\geq r_{qp}$, *then* x could belong to $Cl_t^{\geq}$,

3) *certain* $D_{\leq}$-*decision rules*, providing upper profile descriptions for objects belonging to $Cl_t^{\leq}$ without ambiguity: *if* $f(x,q_1)\leq r_{q1}$ *and* $f(x,q_2)\leq r_{q2}$ *and* ... $f(x,q_p)\leq r_{qp}$, *then* $x\in Cl_t^{\leq}$,

4) *possible* $D_{\leq}$-*decision rules*, providing upper profile descriptions for objects belonging to $Cl_t^{\leq}$ with or without any ambiguity: *if* $f(x,q_1)\leq r_{q1}$ *and* $f(x,q_2)\leq r_{q2}$ *and* ... $f(x,q_p)\leq r_{qp}$, *then* x could belong to $Cl_t^{\leq}$,

5) *approximate* $D_{\geq\leq}$-*decision rules*, providing simultaneously lower and upper profile descriptions for objects belonging to $Cl_s\cup Cl_{s+1}\cup\ldots\cup Cl_t$ without possibility of discerning to which class: *if* $f(x,q_1)\geq r_{q1}$ *and* $f(x,q_2)\geq r_{q2}$ *and* ... $f(x,q_k)\geq r_{qk}$ *and* $f(x,q_{k+1})\leq r_{qk+1}$ *and* ... $f(x,q_p)\leq r_{qp}$, *then* $x\in Cl_s\cup Cl_{s+1}\cup\ldots\cup Cl_t$,

In the left hand side of a $D_{\geq\leq}$-decision rule we can have "$f(x,q)\geq r_q$" and "$f(x,q)\leq r'_q$", where $r_q\leq r'_q$, for the same $q\in C$. Moreover, if $r_q=r'_q$, the two conditions boil down to "$f(x,q)=r_q$".

Since a decision rule is an implication, by a *minimal* rule we understand such an implication that there is no other implication with the left hand side (LHS) of at least the same weakness (in other words, rule using a subset of elementary conditions or/and weaker elementary conditions) and the right hand side (RHS) of at least the same strength (in other words, rule assigning objects to the same union or sub-union of classes).

The rules of type 1) and 3) represent certain knowledge extracted from the data matrix, while the rules of type 2), 4) represent possible knowledge, and rules of type 5) represent doubtful knowledge.

The rules of type 1) and 3) are *exact*, if they do not cover negative examples, and they are *probabilistic*, otherwise. In the latter case, each rule is characterized by a confidence ratio, representing the probability that an object matching LHS of the rule matches also its RHS. Probabilistic rules are concordant with the VC-DRSA model mentioned above.

Let us comment application of decision rules to the objects described by criteria from C. When applying $D_{\geq}$-decision rules to object x, it is possible that x either matches LHS of at least one decision rule or does not match LHS of any decision rule. In the case of at least one matching, it is reasonable to conclude that x belongs to class Cl_t, being the lowest class of the upward union $Cl_t^{\geq}$ resulting from intersection of all RHS of rules covering x. Precisely, if x matches LHS of rules ρ_1, $\rho_2,\ldots,\rho_m$, having RHS $x\in Cl_{t1}^{\geq}$, $x\in Cl_{t2}^{\geq},\ldots$, $x\in Cl_{tm}^{\geq}$, then x is assigned to class Cl_t, where t=max$\{t1,t2,\ldots,tm\}$. In the case of no matching, it is concluded that x belongs to Cl_1, i.e. to the worst class, since no rule with RHS suggesting a better classification of x is covering this object.

Analogously, when applying $D_{\leq}$-decision rules to object x, it is concluded that x belongs either to class Cl_z, being the highest class of the downward union $Cl_t^{\leq}$ resulting from intersection of all RHS of rules covering x, or to class Cl_n, i.e. to the best class, when x is not covered by any rule. Precisely, if x matches the LHS of rules ρ_1, $\rho_2,\ldots,\rho_m$, having RHS $x\in Cl_{t1}^{\leq}$, $x\in Cl_{t2}^{\leq},\ldots$, $x\in Cl_{tm}^{\leq}$, then x is assigned to class Cl_t, where t=min$\{t1,t2,\ldots,tm\}$.

Finally, when applying $D_{\geq\leq}$-decision rules to object x, it is concluded that x belongs to the union of all classes suggested in RHS of rules covering x.

A set of decision rules is *complete* if it is able to cover all objects from the data matrix in such a way that consistent objects are re-classified to their original classes and inconsistent objects are classified to clusters of classes referring to this inconsistency. We call *minimal* each set of decision rules that is complete and non-redundant, i.e. exclusion of any rule from this set makes it non-complete.

One of three induction strategies can be adopted to obtain a set of decision rules (Stefanowski and Vanderpooten, 1994; Stefanowski, 1998):

- generation of a *minimal* description, i.e. a minimal set of rules,
- generation of an *exhaustive* description, i.e. all rules for a given data matrix,
- generation of a *characteristic* description, i.e. a set of rules covering relatively many objects each, however, all together not necessarily all objects from U.

In the following we present a rule induction algorithm, called DOMLEM (Greco, Matarazzo, Slowinski and Stefanowski, 2000b), built on the idea of LEM2 (Grzymala-Busse, 1992) and generating a *minimal* description.

In the algorithm, E denotes a complex (conjunction of elementary conditions e) being a candidate for LHS of the rule. Moreover, $[E]$ denotes a set of objects matching the complex E. Complex E is accepted as LHS of the rule iff $\emptyset \neq [E] = \bigcap_{e \in E} [e] \subseteq B$, where B is the considered approximation corresponding to RHS of the rule. For the sake of simplicity, in the following we present the general scheme of the DOMLEM algorithm for type 1) decision rules.

Procedure DOMLEM
(**input:** L– a family of P-lower approximations of upward unions of classes: { $\underline{P}(Cl_t^{\geq})$, $\underline{P}(Cl_{t-1}^{\geq})$, …, $\underline{P}(Cl_2^{\geq})$ }, where $P \subseteq C$;
output: $R_{\geq}$ set of $D_{\geq}$-decision rules);
begin
    $R_{\geq} := \emptyset$;
    for each $B \in L$ **do**
    begin
        E:=find_rules(B);
        for each rule $E \in$ **E do**
        if "if E, then x$\in Cl_s^{\geq}$" is a minimal rule **then** $R_{\geq} := R_{\geq} \cup E$;
    end
end.
Function find_rules
(**input:** a set B;
output: a set of rules **E** covering set B);
begin
    $G := B$; {a set of objects from the given approximation}
    E := $\emptyset$;
    while $G \neq \emptyset$ **do**
    begin
        $E := \emptyset$; {starting complex}
        $S := G$; {set of objects currently covered by E}
        while ($E = \emptyset$) **or not** ($[E] \subseteq B$) **do**
        begin
            best := $\emptyset$; {best candidate for elementary condition}
            for each criterion $q_i \in P$ **do**
            begin
                $Cond := \{(f(x,q_i) \geq r_{qi}) : \exists x \in S \ \ (f(x,q_i) = r_{qi})\}$;
                {for each positive object from S create an elementary condition}
                for each $elem \in Cond$ **do**
                if *evaluate({elem}∪E)* is_better_than *evaluate({best}∪E)* **then** *best*:=*elem*;
                    {evaluate if new condition is better than previous best};

end;

$E := E \cup \{best\}$;

S := S $\cap$ [*best*];

end; {while not ($[E] \subseteq$ B)}

for each elementary condition $e \in E$ **do**

if $[E - \{e\}] \subseteq B$ **then** $E := E - \{e\}$;

T := **T** $\cup$ $\{E\}$;

$G := B - \bigcup_{E \in \mathbf{E}} [E]$;

end; {while G $\neq \emptyset$}

create rules on the basis of **E**

end {function}

Let us comment the choice of the best condition using the function *evaluate(E)*. Complex *E*, being a candidate LHS for a rule can be evaluated by various measures. In the current version of DOMLEM the complex *E* with the highest ratio $|[E] \cap G| / |[E]|$ is considered the best. In case of tie, the complex *E* with the highest value of $|[E] \cap G|$ is chosen.

The procedure of rule extraction makes evidence of the utility of the concept of inconsistency in the sense of the dominance principle in knowledge discovery process. Decision rules are created by appending descriptors to a complex until a consistency is reached. For instance, in the case of type 1) decision rules, the descriptors are appended until there is no object dominating the complex while not belonging to the upward union of classes indicated in RHS of the rule being created. The concept of inconsistency is similarly applied in calculation of reducts. These remarks justify the use of DRSA in the discovery of rules and reducts even if there is no inconsistency in sample data set for the complete set of criteria *C*.

4 Example

To illustrate application of DRSA to multicriteria classification we will use a part of data provided by a Greek industrial bank ETEVA which finances industrial and commercial firms in Greece (Slowinski and Zopounidis, 1995). A sample composed of 39 firms has been chosen for the study in co-operation with the ETEVA's financial manager. The manager has classified the selected firms into three classes of the bankruptcy risk. The sorting decision is represented by decision attribute *d* making a trichotomic partition of the 39 firms:

- *d*=A means "acceptable",
- *d*=U means "uncertain",
- *d*=NA means "non-acceptable".

The partition is denoted by $\mathbf{Cl}=\{Cl_A, Cl_U, Cl_{NA}\}$ and, obviously, class Cl_A is better than Cl_U which is better than Cl_{NA}.

The firms were evaluated using the following twelve criteria ($\uparrow$ means preference increasing with value and $\downarrow$ means preference decreasing with value):

- A_1 = earnings before interests and taxes/total assets, ↑
- A_2 = net income/net worth, ↑
- A_3 = total liabilities/total assets, ↓
- A_4 = total liabilities/cash flow, ↓
- A_5 = interest expenses/sales, ↓
- A_6 = general and administrative expense/sales, ↓
- A_7 = managers' work experience, ↑ (very low=1, low=2, medium=3, high=4, very high=5),
- A_8 = firm's market niche/position, ↑ (bad=1, rather bad=2, medium=3, good=4, very good=5),
- A_9 = technical structure-facilities, ↑ (bad=1, rather bad=2, medium=3, good=4, very good=5),
- A_{10} = organization-personnel, ↑ (bad=1, rather bad=2, medium=3, good=4, very good=5),
- A_{11} = special competitive advantage of firms, ↑ (low=1, medium=2, high=3, very high=4),
- A_{12} = market flexibility, ↑ (very low=1, low=2, medium=3, high=4, very high=5).

The first six criteria are continuous (financial ratios) and the last six are ordinal. The data matrix is presented in table 1.

Table 1. Financial data matrix

Firm	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	d
F1	16.4	14.5	59.82	2.5	7.5	5.2	5	3	5	4	2	4	A
F2	35.8	67.0	64.92	1.7	2.1	4.5	5	4	5	5	4	5	A
F3	20.6	61.75	75.71	3.6	3.6	8.0	5	3	5	5	3	5	A
F4	11.5	17.1	57.1	3.8	4.2	3.7	5	2	5	4	3	4	A
F5	22.4	25.1	49.8	2.1	5.0	7.9	5	3	5	5	3	5	A
F6	23.9	34.5	48.9	1.7	2.5	8.0	5	3	4	4	3	4	A
F7	29.9	44.0	57.8	1.8	1.7	2.5	5	4	4	5	3	5	A
F8	8.7	5.4	27.4	3.3	4.5	4.5	5	2	4	4	1	4	A
F9	25.7	29.7	46.8	1.7	4.6	3.7	4	2	4	3	1	3	A
F10	21.2	24.6	64.8	3.7	3.6	8.0	4	2	4	4	1	4	A
F11	18.32	31.6	69.3	4.4	2.8	3.0	4	3	4	4	3	4	A
F12	20.7	19.3	19.7	0.7	2.2	4.0	4	2	4	4	1	3	A
F13	9.9	3.5	53.1	4.5	8.5	5.3	4	2	4	4	1	4	A
F14	10.4	9.3	80.9	9.4	1.4	4.1	4	2	4	4	3	3	A
F15	17.7	19.8	52.8	3.2	7.9	6.1	4	4	4	4	2	5	A
F16	14.8	15.9	27.94	1.3	5.4	1.8	4	2	4	3	2	3	A
F17	16.0	14.7	53.5	3.9	6.8	3.8	4	4	4	4	2	4	A
F18	11.7	10.01	42.1	3.9	12.2	4.3	5	2	4	2	1	3	A
F19	11.0	4.2	60.8	5.8	6.2	4.8	4	2	4	4	2	4	A
F20	15.5	8.5	56.2	6.5	5.5	1.8	4	2	4	4	2	4	A
F21	13.2	9.1	74.1	11.21	6.4	5.0	2	2	4	4	2	3	U
F22	9.1	4.1	44.8	4.2	3.3	10.4	3	4	4	4	3	4	U
F23	12.9	1.9	65.02	6.9	14.01	7.5	4	3	3	2	1	2	U

F24	5.9	-27.7	77.4	-32.2	16.6	12.7	3	2	4	4	2	3	U
F25	16.9	12.4	60.1	5.2	5.6	5.6	3	2	4	4	2	3	U
F26	16.7	13.1	73.5	7.1	11.9	4.1	2	2	4	4	2	3	U
F27	14.6	9.7	59.5	5.8	6.7	5.6	2	2	4	4	2	4	U
F28	5.1	4.9	28.9	4.3	2.5	46.0	2	2	3	3	1	2	U
F29	24.4	22.3	32.8	1.4	3.3	5.0	2	3	4	4	2	3	U
F30	29.7	8.6	41.8	1.6	5.2	6.4	2	3	4	4	2	3	U
F31	7.3	-64.5	67.5	-2.2	30.1	8.7	3	3	4	4	2	3	NA
F32	23.7	31.9	63.6	3.5	12.1	10.2	3	2	3	4	1	3	NA
F33	18.9	13.5	74.5	10.0	12.0	8.4	3	3	3	4	3	4	NA
F34	13.9	3.3	78.7	25.5	14.7	10.1	2	2	3	4	3	4	NA
F35	-13.3	-31.1	63.0	-10.0	21.2	23.1	2	1	4	3	1	2	NA
F36	6.2	-3.2	46.1	5.1	4.8	10.5	2	1	3	3	2	3	NA
F37	4.8	-3.3	71.9	34.6	8.6	11.6	2	2	4	4	2	3	NA
F38	0.1	-9.6	42.5	-20.0	12.9	12.4	1	1	4	3	1	3	NA
F39	13.6	9.1	76.0	11.4	17.1	10.3	1	1	2	1	1	2	NA

The main questions to be answered by the knowledge discovery process were the following:

- Is the information contained in table 1 consistent ?
- What are the reducts of criteria ensuring the same quality of approximation of the multicriteria classification as the whole set of criteria ?
- What decision rules can be extracted from table 1 ?
- What are the minimal sets of decision rules?

We have answered these questions using the **D**ominance-based **R**ough **S**et **A**pproach.

The **first result** of the DRSA is a discovery that the financial data matrix **is *consistent*** for the complete set of criteria *C*. Therefore, the *C*-lower approximation and *C*-upper approximation of $Cl_{NA}^{\leq}$, $Cl_{U}^{\leq}$ and $Cl_{U}^{\geq}$, $Cl_{A}^{\geq}$ are the same. In other words, the quality of approximation of all upward and downward unions of classes is equal to 1.

The **second discovery** is a set of 18 ***reducts*** of criteria ensuring the same quality of classification as the whole set of 12 criteria:

$RED_{Cl}^{1} = \{A_1, A_4, A_5, A_7\}$, $RED_{Cl}^{2} = \{A_2, A_4, A_5, A_7\}$, $RED_{Cl}^{3} = \{A_3, A_4, A_6, A_7\}$,

$RED_{Cl}^{4} = \{A_4, A_5, A_6, A_7\}$, $RED_{Cl}^{5} = \{A_4, A_5, A_7, A_8\}$, $RED_{Cl}^{6} = \{A_2, A_3, A_7, A_9\}$,

$RED_{Cl}^{7} = \{A_1, A_3, A_4, A_7, A_9\}$, $RED_{Cl}^{8} = \{A_1, A_5, A_7, A_9\}$, $RED_{Cl}^{9} = \{A_2, A_5, A_7, A_9\}$,

$RED_{Cl}^{10} = \{A_4, A_5, A_7, A_9\}$, $RED_{Cl}^{11} = \{A_5, A_6, A_7, A_9\}$, $RED_{Cl}^{12} = \{A_4, A_5, A_7, A_{10}\}$,

$RED_{Cl}^{13} = \{A_1, A_3, A_4, A_7, A_{11}\}$, $RED_{Cl}^{14} = \{A_2, A_3, A_4, A_7, A_{11}\}$, $RED_{Cl}^{15} = \{A_4, A_5, A_6, A_{12}\}$,

$RED_{Cl}^{16} = \{A_1, A_3, A_5, A_6, A_9, A_{12}\}$, $RED_{Cl}^{17} = \{A_3, A_4, A_6, A_{11}, A_{12}\}$, $RED_{Cl}^{18} = \{A_1, A_2, A_3, A_6, A_9, A_{11}, A_{12}\}$.

All above subsets of criteria are equally good and sufficient for perfect approximation of the classification performed by ETEVA's financial manager on the 39 firms. The core of ***Cl*** is empty ($CORE_{Cl}=\emptyset$) which means that no criterion is indispensable for the approximation. Moreover, all the criteria are exchangeable and no criterion is redundant.

The **third discovery** is the set of ***all*** decision rules. We obtained 74 rules describing $Cl_{NA}^{\leq}$, 51 rules describing $Cl_{U}^{\leq}$, 75 rules describing $Cl_{U}^{\geq}$ and 79 rules describing $Cl_{A}^{\geq}$.

The **fourth discovery** is the finding of ***minimal sets*** of decision rules. Several minimal sets were found; one of them is shown below (in parenthesis there is the number of objects supporting the rule):

1) ***if*** $f(x,A_3)\geq 67.5$ ***and*** $f(x,A_4)\geq -2.2$ ***and*** $f(x,A_6)\geq 8.7$, ***then*** $x\in Cl_{NA}^{\leq}$, (4),

2) ***if*** $f(x,A_2)\leq 3.3$ ***and*** $f(x,A_7)\leq 2$, ***then*** $x\in Cl_{NA}^{\leq}$, (5),

3) ***if*** $f(x,A_3)\geq 63.6$ ***and*** $f(x,A_7)\leq 3$ ***and*** $f(x,A_9)\leq 3$, ***then*** $x\in Cl_{NA}^{\leq}$, (4),

4) ***if*** $f(x,A_2)\leq 12.4$ ***and*** $f(x,A_6)\geq 5.6$, ***then*** $x\in Cl_{U}^{\leq}$, (14),

5) ***if*** $f(x,A_7)\leq 3$, ***then*** $x\in Cl_{U}^{\leq}$, (18),

6) ***if*** $f(x,A_2)\geq 3.5$ ***and*** $f(x,A_5)\leq 8.5$, ***then*** $x\in Cl_{U}^{\geq}$, (26),

7) ***if*** $f(x,A_7)\geq 4$, ***then*** $x\in Cl_{U}^{\geq}$, (21),

8) ***if*** $f(x,A_1)\geq 8.7$ ***and*** $f(x,A_9)\geq 4$, ***then*** $x\in Cl_{U}^{\geq}$, (27),

9) ***if*** $f(x,A_2)\geq 3.5$ ***and*** $f(x,A_7)\geq 4$, ***then*** $x\in Cl_{A}^{\geq}$, (20).

As the minimal set of rules is complete and composed of $D_{\geq}$-decision rules and $D_{\leq}$-decision rules only, application of these rules to the 39 firms will result in their exact re-classification to classes of risk.

Minimal sets of decision rules represent the most concise and non-redundant knowledge representations. The above minimal set of 9 decision rules uses 8 attributes and 18 elementary conditions, i.e. 3.85% of descriptors from the data matrix.

5 Comparison with other classification methods

None of machine discovery methods can deal with multicriteria classification because they do not consider preference orders in the domains of attributes and among the classes. Within multicriteria decision analysis there exist methods for multicriteria classification, however, they are not discovering classification patterns from data; they simply apply a preference model, like utility function in scoring methods (see e.g. Thomas *et al*., 1992), to a set of objects to be classified. In this sense, they are not knowledge discovery methods.

Comparing DRSA to CRSA, one can notice the following differences between the two approaches. CRSA extracts knowledge about a partition of *U* into classes which are not preference-ordered; the granules used for knowledge representation are sets of objects indiscernible by a set of condition attributes.

In case of DRSA and multicriteria classification, the condition attributes are criteria and classes are preference-ordered. The extracted knowledge concerns a collection of upward and downward unions of classes and the granules used for knowledge representation are sets of objects defined using dominance relation. This is the main difference between CRSA and DRSA.

There are three most remarkable advantages of DRSA over CRSA. The first one is the ability of handling criteria, preference-ordered classes and inconsistencies in the set of decision examples that CRSA is not able to discover – inconsistencies in the sense of violation of the dominance principle. In consequence, the rough approximations separate the certain part of information from the doubtful one, which is taken into account in rule induction. The second advantage is the analysis of a data matrix without any preprocessing of data, in particular, any discretization of continuous attributes. The third advantage of DRSA lies in a richer syntax of decision rules induced from rough approximations. The elementary conditions (criterion *rel.* value) of decision rules resulting from DRSA use *rel.*$\in\{\leq,=,\geq\}$, while those resulting from CRSA use *rel.*$\in\{=\}$. The DRSA syntax is more understandable to practitioners and makes the representation of knowledge more synthetic, since minimal sets of decision rules are smaller than minimal sets of decision rules resulting from CRSA.

6 Conclusion

Multicriteria classification differs from usual classification problems since it involves preference orders in domains of attributes and in the set of classes. This requires that a knowledge discovery method applied to multicriteria classification respects the dominance principle. As this is not the case of nowadays methods of Data Mining and Knowledge Discovery, they are not able to discover all relevant knowledge contained in the analysed data sample set and, even worse, they may yield unreasonable discoveries, because inconsistent with the dominance principle. These deficiencies are repaired in DRSA based on the concept of rough approximations consistent with the dominance principle.

Acknowledgement

The research of the first two authors has been supported by the Italian Ministry of University and Scientific Research (MURST). The third author wishes to acknowledge financial support from the State Committee for Scientific Research (KBN).

References and Further Reading

Greco, S., B. Matarazzo and R. Slowinski. 1996. "Rough Approximation of Preference Relation by Dominance Relations". *ICS Research Report* 16/96, Warsaw University of Technology and *European Journal of Operational Research*, **117**: 63-83, 1999.

Greco, S., B. Matarazzo and R. Slowinski. 1998. "A new rough set approach to evaluation of bankruptcy risk". In *Operational Tools in the Management of Financial Risk*, edited by C. Zopounidis, pp. 121-136. Dordrecht, Boston: Kluwer Academic Publishers.

Greco, S., B. Matarazzo and R. Slowinski. 1999. "The use of rough sets and fuzzy sets in MCDM". In *Advances in Multiple Criteria Decision Making*, edited by T. Gal, T. Hanne and T. Stewart, chapter 14, pp. 14.1-14.59. Dordrecht, Boston: Kluwer Academic Publishers. Detailed description of the Dominance-based Rough Set Approach to multicriteria classification and other multicriteria decision problems, i.e. choice and ranking problems. Heterogeneous information (qualitative and quantitative, ordered and non-ordered, crisp and fuzzy evaluations, as well as ordinal, quantitative and numerical non-quantitative scales of preference) can be processed within the new rough set approach.

Greco, S., B. Matarazzo, R. Slowinski and J. Stefanowski. 2000a. "Variable consistency model of dominance-based rough set approach". *Proc. RSCTC*'2000 *Conference*, Banff (to appear).

Greco, S., B. Matarazzo, R. Slowinski and J. Stefanowski. 2000b. "An algorithm for induction of decision rules consistent with the dominance principle". *Proc. RSCTC*'2000 *Conference*, Banff (to appear).

Grzymala-Busse, J.W. 1992. LERS - a system for learning from examples based on rough sets, In Slowinski, R., (ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, Dordrecht, 3-18.

Pawlak, Z. 1982. "Rough sets". *International Journal of Information & Computer Sciences* **11**: 341-356.

Pawlak, Z. 1991. "*Rough Sets. Theoretical Aspects of Reasoning about Data*". Dordrecht: Kluwer Academic Publishers.

Roy, B. 1985. "*Méthodologie Multicritère d'Aide à la Décision*". Paris: Economica.

Slowinski, R. (ed.) 1992. "*Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*". Dordrecht, Boston: Kluwer Academic Publishers.

Slowinski, R. and C. Zopounidis. 1995. "Application of the rough set approach to evaluation of bankruptcy risk". *Intelligent Systems in Accounting, Finance and Management* **4**: 27-41.

Stefanowski J. 1998. On rough set based approaches to induction of decision rules. In Polkowski L., Skowron A. (eds.), *Rough Sets in Data Mining and Knowledge Discovery*, Physica-Verlag, vol.1, 500-529.

Stefanowski, J. and Vanderpooten D. 1994. "A general two-stage approach to inducing rules from examples". In *Rough Sets, Fuzzy Sets and Knowledge Discovery (Banff, Canada)*, edited by W.P. Ziarko, pp. 317-325. London, Berlin: Springer-Verlag.

Thomas, L.C., J.N. Crook, and D. B. Edelman (eds.). 1992. *Credit Scoring and Credit Control*. Clarendon Press, Oxford.