

## ROUGH SET AND RULE-BASED MULTICRITERIA DECISION AIDING\*

Roman Slowinski<sup>1\*\*</sup>, Salvatore Greco<sup>2</sup> and Benedetto Matarazzo<sup>2</sup>

Received June 30, 2012 / Accepted July 28, 2012

**ABSTRACT.** The aim of multicriteria decision aiding is to give the decision maker a recommendation concerning a set of objects evaluated from multiple points of view called criteria. Since a rational decision maker acts with respect to his/her value system, in order to recommend the most-preferred decision, one must identify decision maker's preferences. In this paper, we focus on preference discovery from data concerning some past decisions of the decision maker. We consider the preference model in the form of a set of "if..., then..." decision rules discovered from the data by inductive learning. To structure the data prior to induction of rules, we use the Dominance-based Rough Set Approach (DRSA). DRSA is a methodology for reasoning about data, which handles ordinal evaluations of objects on considered criteria and monotonic relationships between these evaluations and the decision. We review applications of DRSA to a large variety of multicriteria decision problems.

**Keywords:** multicriteria decision aiding, ordinal classification, choice, ranking, Dominance-based Rough Set Approach, preference modeling, decision rules.

### 1 INTRODUCTION

In this paper, we review a multicriteria decision aiding methodology which employs decision maker's (DM's) preference model in form of a set of decision rules discovered from some preference data. Multicriteria decision problems concern a finite set of *objects* (also called alternatives, actions, acts, solutions, etc.) evaluated by a finite set of *criteria* (also called attributes, features, variables, etc.), and raise one of the following questions: (i) how to assign the objects to some ordered classes (*ordinal classification*), (ii) how to choose the best subset of objects (*choice* or *optimization*), or (iii) how to rank the objects from the best to the worst (*ranking*). The answer to everyone of these questions involves an aggregation of the multicriteria evaluations of objects, which takes into account preferences of the DM. In consequence, the aggregation formula is at

---

\*Invited paper

\*\*Corresponding author

<sup>1</sup>Institute of Computing Science, Poznan University of Technology, 60-965 Poznan, and Systems Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland. E-mail: Roman.Slowinski@cs.put.poznan.pl

<sup>2</sup>Department of Economics and Business, University of Catania, Corso Italia, 55, 95129 Catania, Italy. E-mails: {salgreco, matarazz}@unict.it

the same time the DM's *preference model*. Thus, any recommendation referring to one of the above questions must be based on the DM's preference model. The preference data used for building this model, as well as the way of building and using it in the decision process, are main factors distinguishing various multicriteria decision aiding methodologies.

In our case, we assume that the *preference data* includes either observations of DM's past decisions in the same decision problem, or examples of decisions consciously elicited by the DM on demand of an analyst. This way of preference data elicitation is called *indirect*, by opposition to *direct* elicitation when the DM is supposed to provide information leading directly to definition of all preference model parameters, like weights and discrimination thresholds of criteria, trade-off rates, etc. (see, e.g., Roy, 1996).

Past decisions or decision examples may be, however, *inconsistent* with the dominance principle commonly accepted for multicriteria decision problems. Decisions are inconsistent with the dominance principle if:

- in case of ordinal classification: object  $a$  has been assigned to a worse decision class than object  $b$ , although  $a$  is at least as good as  $b$  on all the considered criteria, i.e.  $a$  dominates  $b$ ;
- in case of choice and ranking: pair of objects  $(a, b)$  has been assigned a degree of preference worse than pair  $(c, d)$ , although differences of evaluations between  $a$  and  $b$  on all the considered criteria is at least as good as respective differences of evaluations between  $c$  and  $d$ , i.e. pair  $(a, b)$  dominates pair  $(c, d)$ .

Thus, in order to build a preference model from partly inconsistent preference data, we had an idea to structure this data using the concept of a *rough set* introduced by Pawlak (1982, 1991). Since its conception, rough set theory has often proved to be an excellent mathematical tool for the analysis of inconsistent description of objects. Originally, its understanding of inconsistency was different, however, than the above inconsistency with the dominance principle. The original rough set philosophy is based on the assumption that with every object of the universe  $U$  there is associated a certain amount of information (data, knowledge). This information can be expressed by means of a number of attributes. The attributes describe the objects. Objects which have the same description are said to be indiscernible (or similar) with respect to the available information. The *indiscernibility relation* thus generated constitutes the mathematical basis of rough set theory. It induces a partition of the universe into blocks of indiscernible objects, called elementary sets, which can be used to build knowledge about a real or abstract world. The use of the indiscernibility relation results in information *granulation*.

Any subset  $X$  of the universe may be expressed in terms of these blocks either precisely (as a union of elementary sets) or approximately. In the latter case, the subset  $X$  may be characterized by two ordinary sets, called the *lower* and *upper approximations*. A rough set is defined by means of these two approximations, which coincide in the case of an ordinary set. The lower approximation of  $X$  is composed of all the elementary sets included in  $X$  (whose elements, therefore, certainly belong to  $X$ ), while the upper approximation of  $X$  consists of all the elementary sets which have a non-empty intersection with  $X$  (whose elements, therefore, may belong to  $X$ ).

The difference between the upper and lower approximation constitutes the boundary region of the rough set, whose elements cannot be characterized with certainty as belonging or not to  $X$  (by using the available information). The information about objects from the boundary region is, therefore, inconsistent or ambiguous. The cardinality of the boundary region states, moreover, the extent to which it is possible to express  $X$  in exact terms, on the basis of the available information. For this reason, this cardinality may be used as a measure of vagueness of the information about  $X$ .

Some important characteristics of the rough set approach makes it a particularly interesting tool in a variety of problems and concrete applications. For example, it is possible to deal with both quantitative and qualitative input data and inconsistencies need not to be removed prior to the analysis. In terms of the output information, it is possible to acquire *a posteriori* information regarding the relevance of particular attributes and their subsets to the quality of approximation considered within the problem at hand. Moreover, the lower and upper approximations of a partition of  $U$  into decision classes, prepare the ground for inducing *certain* and *possible* knowledge patterns in the form of “*if... then...*” decision rules.

Several attempts have been made to employ rough set theory for decision aiding (Slowinski, 1993; Pawlak & Slowinski, 1994). The Indiscernibility-based Rough Set Approach (IRSA) is not able, however, to deal with preference ordered attribute scales and preference ordered decision classes. In multicriteria decision analysis, an attribute with a preference ordered scale (value set) is called a *criterion*.

An extension of the IRSA which deals with inconsistencies with respect to dominance principle, typical for preference data, was proposed by Greco, Matarazzo & Slowinski (1998a, 1999a,b). This extension, called the Dominance-based Rough Set Approach (DRSA) is mainly based on the substitution of the indiscernibility relation by a dominance relation in the rough approximation of decision classes. An important consequence of this fact is the possibility of inferring (from observations of past decisions or from exemplary decisions) the DM’s preference model in terms of decision rules which are logical statements of the type “*if..., then...*”. The separation of *certain* and *uncertain* knowledge about the DM’s preferences is carried out by the distinction of different kinds of decision rules, depending upon whether they are induced from lower approximations of decision classes or from the difference between upper and lower approximations (composed of inconsistent examples). Such a preference model is more general than the classical functional models considered within multi-attribute utility theory or the relational models considered, for example, in outranking methods (Greco *et al.*, 2002c, 2004; Slowinski *et al.*, 2002b).

This paper is a review based on previous publications. In the next section, we present some basics on the Indiscernibility-based Rough Set Approach (IRSA) as well as on its extension to similarity relation. In Section 3, we explain the need of replacing indiscernibility or similarity relation by dominance relation in the definition of rough sets, when considering preference data. This leads us to Section 4, where Dominance-based Rough Set Approach (DRSA) is presented with respect to multicriteria ordinal classification. This section also includes two special versions of DRSA: Variable Consistency DRSA (VC-DRSA) and Stochastic DRSA. Section 5 presents

DRSA with respect to multicriteria choice and ranking. In Sections 4 and 5, application of DRSA to all three categories of multicriteria decision problems is explained by the way of examples. Section 6 groups conclusions and characterizes other relevant extensions and applications of DRSA to decision problems. Finally, Section 7 provides information about additional sources of information about rough set theory and applications.

## 2 SOME BASICS ON INDISCERNIBILITY-BASED ROUGH SET APPROACH (IRSA)

### 2.1 Definition of rough approximations by IRSA

For algorithmic reasons, we supply the information regarding the objects in the form of a data table, whose separate rows refer to distinct objects and whose columns refer to the different attributes considered. Each cell of this table indicates an evaluation (quantitative or qualitative) of the object placed in that row by means of the attribute in the corresponding column.

Formally, a *data table* is the 4-tuple  $\mathcal{S} = \langle U, Q, V, f \rangle$ , where  $U$  is a finite set of *objects* (universe),  $Q = \{q_1, q_2, \dots, q_m\}$  is a finite set of *attributes*,  $V_q$  is the value set of the attribute  $q$ ,  $V = \cup_{q \in Q} V_q$  and  $f : U \times Q \rightarrow V$  is a total function such that  $f(x, q) \in V_q$  for each  $q \in Q$ ,  $x \in U$ , called the *information function*.

Each object  $x$  of  $U$  is described by a vector (string)

$$Des_Q(x) = [f(x, q_1), f(x, q_2), \dots, f(x, q_m)],$$

called the *description* of  $x$  in terms of the evaluations of the attributes from  $Q$ . It represents the available information about  $x$ .

To every (non-empty) subset of attributes  $P$  we associate an *indiscernibility relation* on  $U$ , denoted by  $I_P$  and defined as follows:

$$I_P = \{(x, y) \in U \times U : f(x, q) = f(y, q), \text{ for each } q \in P\}.$$

If  $(x, y) \in I_P$ , we say that the objects  $x$  and  $y$  are  $P$ -indiscernible. Clearly, the indiscernibility relation thus defined is an equivalence relation (reflexive, symmetric and transitive). The family of all the equivalence classes of the relation  $I_P$  is denoted by  $U|I_P$  and the equivalence class containing an object  $x \in U$  is denoted by  $I_P(x)$ . The equivalence classes of the relation  $I_P$  are called the *P-elementary sets* or *granules of knowledge* encoded by  $P$ .

Let  $\mathcal{S}$  be a data table,  $X$  be a non-empty subset of  $U$  and  $\emptyset \neq P \subseteq Q$ . The set  $X$  may be characterized by two ordinary sets, called the *P-lower approximation* of  $X$  (denoted by  $\underline{P}(X)$ ) and the *P-upper approximation* of  $X$  (denoted by  $\overline{P}(X)$ ) in  $\mathcal{S}$ . They can be defined, respectively, as:

$$\underline{P}(X) = \{x \in U : I_P(x) \subseteq X\}, \quad \overline{P}(X) = \{x \in U : I_P(x) \cap X \neq \emptyset\}.$$

The family of all the sets  $X \subseteq U$  having the same  $P$ -lower and  $P$ -upper approximations is called a *P-rough set*. The elements of  $\underline{P}(X)$  are all and only those objects  $x \in U$  which belong to the equivalence classes generated by the indiscernibility relation  $I_P$  contained in  $X$ . The elements of

$\overline{P}(X)$  are all and only those objects  $x \in U$  which belong to the equivalence classes generated by the indiscernibility relation  $I_P$  containing at least one object  $x$  belonging to  $X$ . In other words,  $\underline{P}(X)$  is the largest union of the  $P$ -elementary sets included in  $X$ , while  $\overline{P}(X)$  is the smallest union of the  $P$ -elementary sets containing  $X$ .

The lower and upper approximations can be written in an equivalent form, in terms of unions of elementary sets as follows:

$$\underline{P}(X) = \bigcup_{x \in U, I_P(x) \subseteq X} I_P(x), \quad \overline{P}(X) = \bigcup_{x \in X} I_P(x).$$

The  $P$ -boundary of  $X$  in  $\mathcal{S}$ , denoted by  $Bn_P(X)$ , is defined as:

$$Bn_P(X) = \overline{P}(X) - \underline{P}(X).$$

The term *rough approximation* is a general term used to express the operation of the  $P$ -lower and  $P$ -upper approximation of a set or of a union of sets. The rough approximations obey the following basic laws (cf. Pawlak, 1991):

- the inclusion property:  $\overline{P}(X) \subseteq X \subseteq \overline{P}(X)$ ,
- the complementarity property:  $\overline{P}(X) = U - \overline{P}(U - X)$ .

Directly from the definitions, we can also get the following properties of the  $P$ -lower and  $P$ -upper approximations (Pawlak, 1982, 1991):

- 1)  $\overline{P}(\emptyset) = \overline{P}(\emptyset)$ ,  $\underline{P}(U) = \overline{P}(U) = U$ ,
- 2)  $\overline{P}(X \cup Y) = \overline{P}(X) \cup \overline{P}(Y)$ ,
- 3)  $\underline{P}(X \cap Y) = \underline{P}(X) \cap \underline{P}(Y)$ ,
- 4)  $X \subseteq Y \Rightarrow \overline{P}(X) \subseteq \overline{P}(Y)$ ,
- 5)  $X \subseteq Y \Rightarrow \underline{P}(X) \subseteq \underline{P}(Y)$ ,
- 6)  $\underline{P}(X \cup Y) \supseteq \underline{P}(X) \cup \underline{P}(Y)$ ,
- 7)  $\overline{P}(X \cap Y) \subseteq \overline{P}(X) \cap \overline{P}(Y)$ ,
- 8)  $\underline{P}(\underline{P}(X)) = \overline{P}(\underline{P}(X)) = \underline{P}(X)$ ,
- 9)  $\overline{P}(\overline{P}(X)) = \underline{P}(\overline{P}(X)) = \overline{P}(X)$ .

Therefore, if an object  $x$  belongs to  $\underline{P}(X)$ , it is also *certainly* contained in  $X$ , while if  $x$  belongs to  $\overline{P}(X)$ , it is only *possibly* contained in  $X$ .  $Bn_P(X)$  constitutes the *doubtful region* of  $X$ : using the knowledge encoded by  $P$  nothing can be said with certainty about the inclusion of its elements in set  $X$ .

If the  $P$ -boundary of  $X$  is empty (i.e.  $Bn_P(X) = \emptyset$ ) then the set  $X$  is an ordinary set, called the  $P$ -exact set. By this, we mean that it may be expressed as the union of some  $P$ -elementary sets. Otherwise, if  $Bn_P(X) \neq \emptyset$ , then the set  $X$  is a  $P$ -rough set and may be characterized by means of  $\underline{P}(X)$  and  $\overline{P}(X)$ .

The following ratio defines an *accuracy* measure of the approximation of  $X(X \neq \emptyset)$  by means of the attributes from  $P$ :  $\alpha_P(X) = \frac{|P(X)|}{|\overline{P}(X)|}$ , where  $|Y|$  denotes the cardinality of a (finite) set  $Y$ . Obviously,  $0 \leq \alpha_P(X) \leq 1$ . If  $\alpha_P(X) = 1$ , then  $X$  is a  $P$ -exact set. If  $\alpha_P(X) < 1$ , then  $X$  is a  $P$ -rough set.

Another ratio defines a *quality* measure of the approximation of  $X$  by means of the attributes from  $P$ :  $\gamma_P(X) = \frac{|P(X)|}{|X|}$ . The quality  $\gamma_P(X)$  represents the relative frequency of the objects correctly assigned by means of the attributes from  $P$ . Moreover,  $0 \leq \alpha_P(X) \leq \gamma_P(X) \leq 1$ , and  $\gamma_P(X) = 0$  iff  $\alpha_P(X) = 0$ , while  $\gamma_P(X) = 1$  iff  $\alpha_P(X) = 1$ .

The definition of approximations of a subset  $X \subseteq U$  can be extended to a classification, i.e. a partition  $Y = \{Y_1, \dots, Y_n\}$  of  $U$ . The subsets  $Y_i, i = 1, \dots, n$ , are disjunctive classes of  $Y$ . By the  $P$ -lower and  $P$ -upper approximations of  $Y$  in  $S$  we mean the sets  $\underline{P}(Y) = \{\underline{P}(Y_1), \dots, \underline{P}(Y_n)\}$  and  $\overline{P}(Y) = \{\overline{P}(Y_1), \dots, \overline{P}(Y_n)\}$ , respectively. The coefficient  $\gamma_P(Y) = \frac{\sum_{i=1}^n |P(Y_i)|}{|U|}$  is called the *quality of approximation of classification*  $Y$  by the set of attributes  $P$ , or in short, the *quality of classification*. It expresses the ratio of all  $P$ -correctly classified objects to all objects in the data table.

The main issue in rough set theory is the approximation of subsets or partitions of  $U$ , representing *knowledge* about  $U$ , with other sets or partitions that have been built up using available information about  $U$ . From the perspective of a particular object  $x \in U$ , it may be interesting, however, to use the available information to assess the degree of its membership to a subset  $X$  of  $U$ . The subset  $X$  can be identified with the knowledge to be approximated. Using the rough set approach one can calculate the membership function  $\mu_X^P(x)$  (*rough membership function*) as

$$\mu_X^P(x) = \frac{|X \cap I_P(x)|}{|I_P(x)|}$$

The value of  $\mu_X^P(x)$  may be interpreted analogously as conditional probability and may be understood as the *degree of certainty* (credibility) to which  $x$  belongs to  $X$ . Observe that the value of the membership function is calculated from the available data, and not subjectively assumed, as it is in the case of membership functions of fuzzy sets.

Between the rough membership function and the rough approximations of  $X$  the following relationships hold:

$$\begin{aligned} \underline{P}(X) &= \{x \in U : \mu_X^P(x) = 1\}, & \overline{P}(X) &= \{x \in U : \mu_X^P(x) > 0\}, \\ Bn_P(X) &= \{x \in U : 0 < \mu_X^P(x) < 1\}, & \underline{P}(U - X) &= \{x \in U : \mu_X^P(x) = 0\}. \end{aligned}$$

In rough set theory there is, therefore, a close link between the granularity connected with the rough approximation of sets and the uncertainty connected with the rough membership of objects to sets.

A very important concept for concrete applications is that of the dependence of attributes. Intuitively, a set of attributes  $T \subseteq Q$  *totally depends* upon a set of attributes  $P \subseteq Q$  if all the values of the attributes from  $T$  are uniquely determined by the values of the attributes from  $P$ . In other words, this is the case if a functional dependence exists between evaluations by the attributes from  $P$  and by the attributes from  $T$ . This means that the partition (granularity) generated by the attributes from  $P$  is at least as “fine” as that generated by the attributes from  $T$ , so that it is sufficient to use the attributes from  $P$  to build the partition  $U|I_T$ . Formally,  $T$  totally depends on  $P$  iff  $I_P \subseteq I_T$ .

Therefore,  $T$  is totally (partially) dependent on  $P$  if all (some) objects of the universe  $U$  may be univocally assigned to granules of the partition  $U|I_T$ , using only the attributes from  $P$ .

Another issue of great practical importance is that of *knowledge reduction*. This concerns the elimination of superfluous data from the data table, without deteriorating the information contained in the original table.

Let  $P \subseteq Q$  and  $p \in P$ . It is said that attribute  $p$  is *superfluous* in  $P$  if  $I_P = I_{P-\{p\}}$ ; otherwise,  $p$  is *indispensable* in  $P$ .

The set  $P$  is *independent* if all its attributes are indispensable. The subset  $P'$  of  $P$  is a *reduct* of  $P$  (denoted by  $RED(P)$ ) if  $P'$  is independent and  $I_{P'} = I_P$ .

A reduct of  $P$  may also be defined with respect to an approximation of the classification  $Y$  of objects from  $U$ . It is then called a *Y-reduct* of  $P$  (denoted by  $RED_Y(P)$ ) and it specifies a minimal (with respect to inclusion) subset  $P'$  of  $P$  which keeps the quality of the classification unchanged, i.e.  $\gamma_{P'}(Y) = \gamma_P(Y)$ . In other words, the attributes that do not belong to a  $Y$ -reduct of  $P$  are superfluous with respect to the classification  $Y$  of objects from  $U$ .

More than one  $Y$ -reduct (or reduct) of  $P$  may exist in a data table. The set containing all the indispensable attributes of  $P$  is known as the *Y-core* (denoted by  $CORE_Y(P)$ ). In formal terms,  $CORE_Y(P) = \cap RED_Y(P)$ . Obviously, since the  $Y$ -core is the intersection of all the  $Y$ -reducts of  $P$ , it is included in every  $Y$ -reduct of  $P$ . It is the most important subset of attributes of  $Q$ , because none of its elements can be removed without deteriorating the quality of the classification.

## 2.2 Decision rules induced from rough approximations

In a data table the attributes of the set  $Q$  are often divided into *condition* attributes (set  $C \neq \emptyset$ ) and *decision* attributes (set  $D \neq \emptyset$ ). Note that  $C \cup D = Q$  and  $C \cap D = \emptyset$ . Such a table is called a *decision table*. The decision attributes induce a partition of  $U$  deduced from the indiscernibility relation  $I_D$  in a way that is independent of the condition attributes.  $D$ -elementary sets are called *decision classes*. There is a tendency to reduce the set  $C$  while keeping all important relationships between  $C$  and  $D$ , in order to make decisions on the basis of a smaller amount of information. When the set of condition attributes is replaced by one of its reducts, the quality of approximation of the classification induced by the decision attributes does not deteriorate.

Since the tendency is to underline the functional dependencies between condition and decision attributes, a decision table may also be seen as a set of *decision rules*. These are logical statements of the type “*if... then...*”, where the antecedent (condition part) specifies values assumed by one or more condition attributes (describing  $C$ -elementary sets) and the consequence (decision part) specifies an assignment to one or more decision classes (describing  $D$ -elementary sets). Therefore, the syntax of a rule can be outlined as follows:

if  $f(x, q_1)$  is equal to  $r_{q_1}$  and  $f(x, q_2)$  is equal to  $r_{q_2}$  and...  $f(x, q_p)$  is equal to  $r_{q_p}$ ,  
then  $x$  belongs to  $Y_{j_1}$  or  $Y_{j_2}$  or ...  $Y_{j_k}$ ,

where  $\{q_1, q_2, \dots, q_p\} \subseteq C$ ,  $(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$  and  $Y_{j_1}, Y_{j_2}, \dots, Y_{j_k}$  are some decision classes of the considered classification ( $D$ -elementary sets). If there is only one possible consequence, *i.e.*  $k = 1$ , then the rule is said to be *certain*, otherwise it is said to be *approximate* or *ambiguous*.

An object  $x \in U$  *supports* decision rule  $r$  if its description is matching both the condition part and the decision part of the rule. We also say that decision rule  $r$  *covers* object  $x$  if it matches at least the condition part of the rule. Each decision rule is characterized by its *strength* defined as the number of objects supporting the rule. In the case of approximate rules, the strength is calculated for each possible decision class separately.

Let us observe that certain rules are supported only by objects from the lower approximation of the corresponding decision class. Approximate rules are supported, in turn, only by objects from the boundaries of the corresponding decision classes.

Procedures for the generation of decision rules from a decision table use an *inductive learning* principle. The objects are considered as examples of decisions. In order to induce decision rules with a unique consequent assignment to a  $D$ -elementary set, the examples belonging to the  $D$ -elementary set are called *positive* and all the others *negative*. A decision rule is *discriminant* if it is consistent (*i.e.* if it distinguishes positive examples from negative ones) and *minimal* (*i.e.* if removing any attribute from a condition part gives a rule covering negative objects). It may be also interesting to look for *partly discriminant* rules. These are rules that, besides positive examples, could cover a limited number of negative ones. They are characterized by a coefficient, called the *level of confidence*, which is the ratio of the number of positive examples (supporting the rule) to the number of all examples covered by the rule.

The generation of decision rules from decision tables is a complex task and a number of procedures have been proposed to solve it (see, for example, Grzymala-Busse, 1992, 1997; Skowron, 1993; Ziarko & Shan, 1994; Skowron & Polkowski, 1997; Stefanowski, 1998; Slowinski, Stefanowski, Greco & Matarazzo, 2000). The existing induction algorithms use one of the following strategies:

- (a) The generation of a minimal set of rules covering all objects from a decision table.
- (b) The generation of an exhaustive set of rules consisting of all possible rules for a decision table.



- (c) The generation of a set of ‘strong’ decision rules, even partly discriminant, covering relatively many objects from the decision table (but not necessarily all of them).

To summarize the above description of IRSA, let us list particular benefits one can get when applying the rough set approach to analysis of data presented in decision tables:

- a characterization of decision classes in terms of chosen attributes through lower and upper approximation,
- a measure of the quality of approximation which indicates how good the chosen set of attributes is for approximation of the classification,
- a reduction of the knowledge contained in the table to a description by relevant attributes *i.e.* those belonging to reducts; at the same time, exchangeable and superfluous attributes are also identified,
- a core of attributes, being an intersection of all reducts, indicates indispensable attributes,
- a set of decision rules which is induced from the lower and upper approximations of the decision classes; this shows classification patterns which exist in the data set.

A tutorial example illustrating all these benefits has been given in (Slowinski *et al.*, 2005). For more details about IRSA and its extensions, the reader is referred to Pawlak (1991), Polkowski (2002), Slowinski (1992b) and many others (see Section 7). Internet addresses to freely available software implementations of these algorithms can also be found in the last section of this paper.

### 2.3 From indiscernibility to similarity

As mentioned above, the classical definitions of lower and upper approximations are based on the use of the binary indiscernibility relation which is an equivalence relation. The indiscernibility implies the impossibility of distinguishing between two objects of  $U$  having the *same* description in terms of the attributes from  $Q$ . This relation induces equivalence classes on  $U$ , which constitute the basic granules of knowledge. In reality, due to the imprecision of data describing the objects, small differences are often not considered significant for the purpose of discrimination. This situation may be formally modeled by considering similarity or tolerance relations (see *e.g.* Nieminen, 1988; Marcus, 1994; Slowinski, 1992a; Polkowski, Skowron & Zytkow, 1995; Skowron & Stepaniuk, 1995; Slowinski & Vanderpooten, 1995, 2000; Stepaniuk, 2000; Yao & Wong, 1995).

Replacing the indiscernibility relation by a weaker binary *similarity* relation has considerably extended the capacity of the rough set approach. This is because, in the least demanding case, the similarity relation requires reflexivity only, relaxing the assumptions of symmetry and transitivity of the indiscernibility relation.

In general, a similarity relation  $R$  does not generate a partition but a cover of  $U$ . The information regarding similarity may be represented using *similarity classes* for each object  $x \in U$ . More

precisely, the similarity class of  $x$ , denoted by  $R(x)$ , consists of the set of objects which are similar to  $x$ :

$$R(x) = \{y \in U : yRx\}.$$

It is obvious that an object  $y$  may be similar to both  $x$  and  $z$ , while  $z$  is not similar to  $x$ , *i.e.*  $y \in R(x)$  and  $y \in R(z)$ , but  $z \notin R(x)$ ,  $x, y, z \in U$ . The similarity relation is of course reflexive (each object is similar to itself). Slowinski & Vanderpooten (1995, 2000) have proposed a *similarity* relation which is only *reflexive*. The abandonment of the transitivity requirement is easily justifiable. For example, see Luce's paradox of the cups of tea (Luce, 1956). As for the symmetry, one should notice that  $yRx$ , which means "y is similar to x", is directional. There is a subject  $y$  and a referent  $x$ , and in general this is not equivalent to the proposition "x is similar to y", as maintained by Tversky (1977). This is quite immediate when the similarity relation is defined in terms of a percentage difference between evaluations of the objects compared on a numerical attribute in hand, calculated with respect to evaluation of the referent object. Therefore, the symmetry of the similarity relation should not be imposed. It then makes sense to consider the inverse relation of  $R$ , denoted by  $R^{-1}$ , where  $xR^{-1}y$  means again "y is similar to x".  $R^{-1}(x)$ ,  $x \in U$ , is the class of referent objects to which  $x$  is similar:

$$R^{-1}(x) = \{y \in U : xRy\}.$$

Given a subset  $X \subseteq U$  and a similarity relation  $R$  on  $U$ , an object  $x \in U$  is said to be *non-ambiguous* in each of the two following cases:

- $x$  belongs to  $X$  without ambiguity, that is  $x \in X$  and  $R^{-1}(x) \subseteq X$ ; such objects are also called *positive*;
- $x$  does not belong to  $X$  without ambiguity ( $x$  clearly does not belong to  $X$ ), that is  $x \in U - X$  and  $R^{-1}(x) \subseteq U - X$  (or  $R^{-1}(x) \cap X \neq \emptyset$ ); such objects are also called *negative*.

The objects which are neither positive nor negative are said to be *ambiguous*. A more general definition of lower and upper approximation may thus be offered (see Slowinski & Vanderpooten, 2000). Let  $X \subseteq U$  and let  $R$  be a reflexive binary relation defined on  $U$ . The lower approximation of  $X$ , denoted by  $\underline{R}(X)$ , and the upper approximation of  $X$ , denoted by  $\overline{R}(X)$ , are defined, respectively, as:

$$\underline{R}(X) = \{x \in U : R^{-1}(x) \subseteq X\},$$

$$\overline{R}(X) = \bigcup_{x \in X} R(x).$$

It may be demonstrated that the key properties – inclusion and complementarity – still hold and that

$$\overline{R}(X) = \{x \in U : R^{-1}(x) \cap X \neq \emptyset\}.$$

Moreover, the above definition of rough approximation is the *only one* that correctly characterizes the set of positive objects (lower approximation) and the set of positive or ambiguous objects

(upper approximation) when a similarity relation is reflexive, but not necessarily symmetric nor transitive.

Using a similarity relation, we are able to induce decision rules from a decision table. The syntax of a rule is represented as follows:

If  $f(x, q_1)$  is similar to  $r_{q_1}$  and  $f(x, q_2)$  is similar to  $r_{q_2}$  and...  $f(x, q_p)$  is similar to  $r_{q_p}$ , then  $x$  belongs to  $Y_{j_1}$  or  $Y_{j_2}$  or...  $Y_{j_k}$ ,

where  $\{q_1, q_2, \dots, q_p\} \subseteq C$ ,  $(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$  and  $Y_{j_1}, Y_{j_2}, \dots, Y_{j_k}$  are some classes of the considered classification ( $D$ -elementary sets). As mentioned above, if  $k = 1$  then the rule is *certain*, otherwise it is *approximate* or *ambiguous*. Procedures for generation of decision rules follow the induction principle described in point 2.2. One such procedure has been proposed by Krawiec, Slowinski & Vanderpooten (1998) – it involves a similarity relation that is learned from data. We would also like to point out that Greco, Matarazzo & Slowinski (1998b, 2000b) proposed a fuzzy extension of the similarity, that is, rough approximation of fuzzy sets (decision classes) by means of fuzzy similarity relations (reflexive only).

### 3 THE NEED OF REPLACING INDISCERNIBILITY RELATION BY DOMINANCE RELATION WHEN REASONING ABOUT PREFERENCE DATA

When trying to apply the rough set concept based on indiscernibility or similarity to reasoning about preference ordered data, it has been noted that IRSA ignores not only the preference order in the value sets of attributes but also the monotonic relationship between evaluations of objects on such attributes (called criteria) and the preference ordered value of decision (classification decision or degree of preference) (see Greco, Matarazzo & Slowinski, 1998a, 1999b, 2001a; Slowinski, Greco & Matarazzo, 2000a).

In order to explain how important is the above monotonic relationship for data describing multi-criteria decision problems, let us consider an example of a data set concerning pupils' achievements in a high school. Suppose that among the attributes describing the pupils there are results in *Mathematics (Math)* and *Physics (Ph)*. There is also a *General Achievement (GA)* result, which is considered as a classification decision. The domains of all three attributes are composed of three values: *bad*, *medium* and *good*. The preference order of the attribute values is obvious: *good* is better than *medium* and *bad*, and *medium* is better than *bad*. Such attributes are called *criteria* because they involve an evaluation. One can also notice a *semantic correlation* between the two criteria and the classification decision, which means that an improvement on one criterion should not worsen the classification decision, while the other criterion is unchanged. Precisely, an improvement of a pupil's score in *Math* or *Ph*, with other criterion value unchanged, should not worsen the pupil's general achievement (*GA*), but rather improve it. In general terms, this requirement is concordant with the dominance principle defined in the Introduction.

This semantic correlation is also called monotonicity constraint, and thus, an alternative name of the classification problem with semantic correlation between evaluation criteria and classification decision is *ordinal classification with monotonicity constraints*.

Two questions naturally follow consideration of this example:

- What classification rules can be drawn from the pupils' data set?
- How does the semantic correlation influences the classification rules?

The answer to the first question is: monotonic "if..., then..." decision rules. Each decision rule is characterized by a *condition profile* and a *decision profile*, corresponding to vectors of threshold values on evaluation criteria and on classification decision, respectively. The answer to the second question is that condition and decision profiles of a decision rule should observe the dominance principle (monotonicity constraint) if the rule has at least one pair of semantically correlated criteria spanned over the condition and decision part. We say that one profile *dominates* another if the values of criteria of the first profile are not worse than the values of criteria of the second profile.

Let us explain the dominance principle with respect to decision rules on the pupils' example. Suppose that two rules induced from the pupils' data set relate *Math* and *Ph* on the condition side, with *GA* on the decision side:

rule#1: if *Math=medium* and *Ph=medium*, then *GA=good*,

rule#2: if *Math=good* and *Ph=medium*, then *GA=medium*.

The two rules do not observe the dominance principle because the condition profile of rule #2 dominates the condition profile of rule #1, while the decision profile of rule #2 is dominated by the decision profile of rule #1. Thus, in the sense of the dominance principle, the two rules are inconsistent, *i.e.* they are wrong.

One could say that the above rules are true because they are supported by examples of pupils from the analyzed data set, but this would mean that the examples are also inconsistent. The *inconsistency* may come from many sources. Examples include:

- Missing attributes (regular ones or criteria) in the description of objects. Maybe the data set does not include such attributes as the *opinion of the pupil's tutor* expressed only verbally during an assessment of the pupil's *GA* by a school assessment committee.
- Unstable preferences of decision makers. Maybe the members of the school assessment committee changed their view on the influence of *Math* on *GA* during the assessment.

Handling these inconsistencies is of crucial importance for knowledge discovery about preferences. They cannot be simply considered as noise or error to be eliminated from data, or amalgamated with consistent data by some averaging operators. They should be identified and presented as uncertain rules.

If the semantic correlation was ignored in prior knowledge, then the handling of the above mentioned inconsistencies would be impossible. Indeed, there would be nothing wrong with rules #1 and #2. They would be supported by different examples discerned by considered attributes.

It has been acknowledged by many authors that *rough set theory* provides an excellent framework for dealing with inconsistencies in knowledge discovery (Grzymala-Busse, 1992; Pawlak, 1991; Pawlak, Grzymala-Busse, Slowinski & Ziarko, 1995; Polkowski, 2002; Polkowski & Skowron, 1999; Slowinski, 1992b; Slowinski & Zopounidis, 1995; Ziarko, 1998). As we have shown in Section 2, the paradigm of rough set theory is that of *granular computing*, because the main concept of the theory (rough approximation of a set) is built up of blocks of objects which are indiscernible by a given set of attributes, called *granules of knowledge*. In the space of regular attributes, the indiscernibility granules are bounded sets. Decision rules induced from rough approximation of a classification are also built up of such granules.

The authors have proposed an extension of the granular computing paradigm that enables us to take into account prior knowledge, either about evaluation of objects on multiple criteria only (Greco, Matarazzo, Slowinski & Stefanowski, 2002), or about multicriteria evaluation with monotonicity constraints (Greco, Matarazzo & Slowinski, 1998a, 1999b, 2000d, 2001a, 2002a, 2002b; Slowinski, Greco & Matarazzo, 2002a, 2009). The combination of the new granules with the idea of rough approximation is called the *Dominance-based Rough Set Approach (DRSA)*.

In the following, we present the concept of granules which permit us to handle prior knowledge about multicriteria evaluation with monotonicity constraints when inducing decision rules.

Let  $U$  be a finite set of objects (universe) and let  $Q$  be a finite set of attributes divided into a set  $C$  of *condition attributes* and a set  $D$  of *decision attributes*, where  $C \cap D = \emptyset$ . Also, let

$$X_C = \prod_{q=1}^{|C|} X_q \quad \text{and} \quad X_D = \prod_{q=1}^{|D|} X_q$$

be attribute spaces corresponding to sets of condition and decision attributes, respectively. The elements of  $X_C$  and  $X_D$  can be interpreted as possible evaluations of objects on attributes from set  $C = \{1, \dots, |C|\}$  and from set  $D = \{1, \dots, |D|\}$ , respectively. Therefore,  $X_q$  is the set of possible evaluations of considered objects with respect to attribute  $q$ . The value of object  $x$  on attribute  $q \in Q$  is denoted by  $x_q$ . Objects  $x$  and  $y$  are *indiscernible* by  $P \subseteq C$  if  $x_q = y_q$  for all  $q \in P$  and, analogously, objects  $x$  and  $y$  are *indiscernible* by  $R \subseteq D$  if  $x_q = y_q$  for all  $q \in R$ . The sets of indiscernible objects are equivalence classes of the corresponding *indiscernibility relation*  $I_P$  or  $I_R$ . Moreover,  $I_P(x)$  and  $I_R(x)$  denote equivalence classes including object  $x$ .  $I_D$  generates a partition of  $U$  into a finite number of decision classes  $Cl = \{Cl_t, t = 1, \dots, n\}$ . Each  $x \in U$  belongs to one and only one class  $Cl_t \in Cl$ .

The above definitions are valid for regular attributes, not involving monotonicity relationships between values of condition and decision attributes. In this case, the granules of knowledge are bounded sets in  $X_P$  and  $X_R$  ( $P \subseteq C$  and  $R \subseteq D$ ), defined by partitions of  $U$  induced by the indiscernibility relations  $I_P$  and  $I_R$ , respectively. Then, classification rules to be discovered are

functions representing granules  $I_R(x)$  by granules  $I_P(x)$  in the condition attribute space  $X_P$ , for any  $P \subseteq C$  and for any  $x \in U$ .

If value sets of some condition and decision attributes are preference ordered (*i.e.* they are evaluation criteria), and there are known monotonic relationships between value sets of these condition and decision attributes, then the indiscernibility relation is unable to produce granules in  $X_C$  and  $X_D$  that would take into account the preference order. To do so, the indiscernibility relation has to be substituted by a dominance relation in  $X_P$  and  $X_R$  ( $P \subseteq C$  and  $R \subseteq D$ ). Suppose, for simplicity, that all condition attributes in  $C$  and all decision attributes in  $D$  are criteria, and that  $C$  and  $D$  are semantically correlated.

Let  $\succeq_q$  be a weak preference relation on  $U$  (often called *outranking*) representing a preference on the set of objects with respect to criterion  $q \in \{C \cup D\}$ . Now,  $x_q \succeq y_q$  means “ $x_q$  is at least as good as  $y_q$  with respect to criterion  $q$ ”. On the one hand, we say that  $x$  *dominates*  $y$  with respect to  $P \subseteq C$  (shortly,  $xP$ -*dominates*  $y$ ) in the condition attribute space  $X_P$  (denoted by  $xDPy$ ) if  $x_q \succeq y_q$  for all  $q \in P$ . Assuming, without loss of generality, that the domains of the criteria are numerical (*i.e.*  $X_q \subseteq R$  for any  $q \in C$ ) and that they are ordered so that the preference increases with the value, we can say that  $xDPy$  is equivalent to  $x_q \geq y_q$  for all  $q \in P$ ,  $P \subseteq C$ . Observe that for each  $x \in X_P$ ,  $xDPx$ , *i.e.*  $P$ -dominance is reflexive. On the other hand, the analogous definition holds in the decision attribute space  $X_R$  (denoted by  $xDRy$ ), where  $R \subseteq D$ .

The dominance relations  $xDPy$  and  $xDRy$  ( $P \subseteq C$  and  $R \subseteq D$ ) are directional statements where  $x$  is a subject and  $y$  is a referent.

If  $x \in X_P$  is the referent, then one can define a set of objects  $y \in X_P$  dominating  $x$ , called the  $P$ -dominating set (denoted by  $D_P^+(x)$ ) and defined as  $D_P^+(x) = \{y \in U : yDPx\}$ .

If  $x \in X_P$  is the subject, then one can define a set of objects  $y \in X_P$  dominated by  $x$ , called the  $P$ -dominated set (denoted by  $D_P^-(x)$ ) and defined as  $D_P^-(x) = \{y \in U : xDPy\}$ .

$P$ -dominating sets  $D^+P(x)$  and  $P$ -dominated sets  $D_P^-(x)$  correspond to *positive* and *negative dominance cones* in  $X_P$ , with the origin  $x$ .

With respect to the decision attribute space  $X_R$  (where  $R \subseteq D$ ), the  $R$ -dominance relation enables us to define the following sets:

$$Cl_R^{\geq x} = \{y \in U : yDRx\}, \quad Cl_R^{\leq x} = \{y \in U : xDRy\}.$$

$Cl_{t_q} = \{x \in X_D : x_q = t_q\}$  is a decision class with respect to  $q \in D$ .  $Cl_R^{\geq x}$  is called the *upward union* of classes, and  $Cl_R^{\leq x}$  is the *downward union* of classes. If  $x \in Cl_R^{\geq x}$ , then  $x$  belongs to class  $Cl_{t_q}$ ,  $x_q = t_q$ , or better, on each decision attribute  $q \in R$ . On the other hand, if  $x \in Cl_R^{\leq x}$ , then  $x$  belongs to class  $Cl_{t_q}$ ,  $x_q = t_q$ , or worse, on each decision attribute  $q \in R$ . The downward and upward unions of classes correspond to the *positive* and *negative dominance cones* in  $X_R$ , respectively.

In this case, the granules of knowledge are open sets in  $X_P$  and  $X_R$  defined by dominance cones  $D_P^+(x)$ ,  $D_P^-(x)$  ( $P \subseteq C$ ) and  $Cl_R^{\geq x}$ ,  $Cl_R^{\leq x}$  ( $R \subseteq D$ ), respectively. Then, classification rules

to be discovered are functions representing granules  $Cl_R^{\geq x}$ ,  $Cl_R^{\leq x}$  by granules  $D_P^+(x)$ ,  $D_P^-(x)$ , respectively, in the condition attribute space  $X_P$ , for any  $P \subseteq C$  and  $R \subseteq D$  and for any  $x \in X_P$ .

#### 4 THE DOMINANCE-BASED ROUGH SET APPROACH (DRSA) TO MULTICRITERIA ORDINAL CLASSIFICATION

##### 4.1 Granular computing with dominance cones

When discovering classification rules, a set  $D$  of decision attributes is, usually, a singleton,  $D = \{d\}$ . Let us take this assumption for further presentation, although it is not necessary for the Dominance-based Rough Set Approach. The decision attribute  $d$  makes a partition of  $U$  into a finite number of classes,  $Cl = \{Cl_t, t = 1, \dots, n\}$ . Each object  $x \in U$  belongs to one and only one class,  $Cl_t \in Cl$ . The upward and downward unions of classes boil down, respectively, to:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$$

$$Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$$

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

Let us explain how the rough set concept has been generalized to the Dominance-based Rough Set Approach in order to enable granular computing with dominance cones (for more details, see Greco, Matarazzo & Slowinski (1998a, 1999b, 2000d, 2001a, 2002a), Slowinski, Greco & Matarazzo (2009), Slowinski, Stefanowski, Greco & Matarazzo (2000)).

Given a set of criteria,  $P \subseteq C$ , the inclusion of an object  $x \in U$  to the upward union of classes  $Cl_t^{\geq}$ ,  $t = 2, \dots, n$ , is *inconsistent with the 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  $D_P^-(x) \cap Cl_{t-1}^{\geq} \neq \emptyset$ .

If, given a set of criteria  $P \subseteq C$ , the inclusion of  $x \in U$  to  $Cl_t^{\geq}$ , where  $t = 2, \dots, n$ , is inconsistent with the 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 with the dominance principle. This means that all objects  $P$ -dominating  $x$  belong to  $Cl_t^{\geq}$ , i.e.  $D_P^+(x) \subseteq Cl_t^{\geq}$ . Geometrically, this corresponds to the inclusion of the complete set of objects contained in the positive dominance cone originating in  $x$ , in the positive dominance cone  $Cl_t^{\geq}$  originating in  $Cl_t$ .

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

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

In terms of ambiguity,  $x$  possibly belongs to  $Cl_t^{\geq}$  with respect to  $P \subseteq C$ , if  $x$  belongs to  $Cl_t^{\geq}$  with or without any ambiguity. Due to the reflexivity of the  $P$ -dominance relation  $D_P$ , the above conditions can be summarized as follows:  $x$  possibly belongs 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.*

$$D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset.$$

Geometrically, this corresponds to the non-empty intersection of the set of objects contained in the negative dominance cone originating in  $x$ , with the positive dominance cone  $Cl_t^{\geq}$  originating in  $Cl_t$ .

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 possibly belong to  $Cl_t^{\geq}$  constitutes the  $P$ -upper approximation of  $Cl_t^{\geq}$ , denoted by  $\overline{P}(Cl_t^{\geq})$ . More formally:

$$\underline{P}(Cl_t^{\geq}) = \{x \in U : D_P^+(x) \subseteq Cl_t^{\geq}\}$$

$$\overline{P}(Cl_t^{\geq}) = \{x \in U : D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\}$$

where  $t = 1, \dots, n$ . Analogously, one can define the  $P$ -lower approximation and the  $P$ -upper approximation of  $Cl_t^{\leq}$ :

$$\underline{P}(Cl_t^{\leq}) = \{x \in U : D_P^-(x) \subseteq Cl_t^{\leq}\}$$

$$\overline{P}(Cl_t^{\leq}) = \{x \in U : D_P^+(x) \cap Cl_t^{\leq} \neq \emptyset\}$$

where  $t = 1, \dots, n$ .

The  $P$ -lower and  $P$ -upper approximations of  $Cl_t^{\geq}$ ,  $t = 1, \dots, n$ , can also be expressed in terms of unions of positive dominance cones as follows:

$$\underline{P}(Cl_t^{\geq}) = \bigcup_{D_P^+(x) \subseteq Cl_t^{\geq}} D_P^+(x)$$

$$\overline{P}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_P^+(x).$$

Analogously, the  $P$ -lower and  $P$ -upper approximations of  $Cl_t^{\leq}$ ,  $t = 1, \dots, n$ , can be expressed in terms of unions of negative dominance cones as follows:

$$\underline{P}(Cl_t^{\leq}) = \bigcup_{D_P^-(x) \subseteq Cl_t^{\leq}} D_P^-(x)$$



$$\overline{P}(Cl_t^{\leq}) = \bigcup_{x \in Cl_t^{\leq}} D_P^-(x).$$

The  $P$ -lower and  $P$ -upper approximations so defined satisfy the following *inclusion properties* for each  $t \in \{1, \dots, n\}$  and for all  $P \subseteq C$ :

$$\begin{aligned} \underline{P}(Cl_t^{\geq}) &\subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq}), \\ \underline{P}(Cl_t^{\leq}) &\subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq}). \end{aligned}$$

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:

$$\begin{aligned} Bn_P(Cl_t^{\geq}) &= \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}), \\ Bn_P(Cl_t^{\leq}) &= \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}), \end{aligned}$$

where  $t = 1, \dots, n$ . The  $P$ -lower and  $P$ -upper approximations of the unions of classes  $Cl_t^{\geq}$  and  $Cl_t^{\leq}$  have an important *complementarity property*. It says that if object  $x$  belongs without any ambiguity to class  $Cl_t$  or better, then it is impossible that it could belong to class  $Cl_{t-1}$  or worse, *i.e.*

$$\underline{P}(Cl_t^{\geq}) = U - \overline{P}(Cl_{t-1}^{\leq}), \quad t = 2, \dots, n.$$

Due to the complementarity property,  $Bn_P(Cl_t^{\geq}) = Bn_P(Cl_{t-1}^{\leq})$ , for  $t = 2, \dots, n$ , which means that if  $x$  belongs with ambiguity to class  $Cl_t$  or better, then it also belongs with ambiguity to class  $Cl_{t-1}$  or worse.

Considering application of the lower and the upper approximations based on dominance  $D_P$ ,  $P \subseteq C$ , to any set  $X \subseteq U$ , instead of the unions of classes  $Cl_t^{\geq}$  and  $Cl_t^{\leq}$ , one gets upward lower and upper approximations  $\underline{P}^{\geq}(X)$  and  $\overline{P}^{\geq}(X)$ , as well as downward lower and upper approximations  $\underline{P}^{\leq}(X)$  and  $\overline{P}^{\leq}(X)$ , as follows:

$$\begin{aligned} \underline{P}^{\geq}(X) &= \{x \in U : D_P^+(x) \subseteq X\}, \\ \overline{P}^{\geq}(X) &= \{x \in U : D_P^-(x) \cap X \neq \emptyset\}, \\ \underline{P}^{\leq}(X) &= \{x \in U : D_P^-(x) \subseteq X\}, \\ \overline{P}^{\leq}(X) &= \{x \in U : D_P^+(x) \cap X \neq \emptyset\}. \end{aligned}$$

From the definition of rough approximations  $\underline{P}^{\geq}(X)$ ,  $\overline{P}^{\geq}(X)$ ,  $\underline{P}^{\leq}(X)$  and  $\overline{P}^{\leq}(X)$ , we can get also the following properties of the  $P$ -lower and  $P$ -upper approximations (see Greco, Matarazzo & Slowinski, 2007, 2012):

$$\begin{aligned} 1) \quad \underline{P}^{\geq}(\emptyset) &= \overline{P}^{\geq}(\emptyset) = \underline{P}^{\leq}(\emptyset) = \overline{P}^{\leq}(\emptyset) = \emptyset, \\ \underline{P}^{\geq}(U) &= \overline{P}^{\geq}(U) = \underline{P}^{\leq}(U) = \overline{P}^{\leq}(U) = U, \end{aligned}$$

- 2)  $\overline{P}^{\geq}(X \cup Y) = \overline{P}^{\geq}(X) \cup \overline{P}^{\geq}(Y),$   
 $\overline{P}^{\leq}(X \cup Y) = \overline{P}^{\leq}(X) \cup \overline{P}^{\leq}(Y),$
- 3)  $\underline{P}^{\geq}(X \cap Y) = \underline{P}^{\geq}(X) \cap \underline{P}^{\geq}(Y),$   
 $\underline{P}^{\leq}(X \cap Y) = \underline{P}^{\leq}(X) \cap \underline{P}^{\leq}(Y),$
- 4)  $X \subseteq Y \Rightarrow \overline{P}^{\geq}(X) \subseteq \overline{P}^{\geq}(Y),$   
 $X \subseteq Y \Rightarrow \overline{P}^{\leq}(X) \subseteq \overline{P}^{\leq}(Y),$
- 5)  $X \subseteq Y \Rightarrow \underline{P}^{\geq}(X) \subseteq \underline{P}^{\geq}(Y),$   
 $X \subseteq Y \Rightarrow \underline{P}^{\leq}(X) \subseteq \underline{P}^{\leq}(Y),$
- 6)  $\underline{P}^{\geq}(X \cup Y) \supseteq \underline{P}^{\geq}(X) \cup \underline{P}^{\geq}(Y)$   
 $\underline{P}^{\leq}(X \cup Y) \supseteq \underline{P}^{\leq}(X) \cup \underline{P}^{\leq}(Y),$
- 7)  $\overline{P}^{\geq}(X \cap Y) \subseteq \overline{P}^{\geq}(X) \cap \overline{P}^{\geq}(Y)$   
 $\overline{P}^{\leq}(X \cap Y) \subseteq \overline{P}^{\leq}(X) \cap \overline{P}^{\leq}(Y),$
- 8)  $\underline{P}^{\geq}(\underline{P}^{\geq}(X)) = \overline{P}^{\geq}(\underline{P}^{\geq}(X)) = \underline{P}^{\geq}(X)$   
 $\underline{P}^{\leq}(\underline{P}^{\leq}(X)) = \overline{P}^{\leq}(\underline{P}^{\leq}(X)) = \underline{P}^{\leq}(X),$
- 9)  $\overline{P}^{\geq}(\overline{P}^{\geq}(X)) = \underline{P}^{\geq}(\overline{P}^{\geq}(X)) = \overline{P}^{\geq}(X)$   
 $\overline{P}^{\leq}(\overline{P}^{\leq}(X)) = \underline{P}^{\leq}(\overline{P}^{\leq}(X)) = \overline{P}^{\leq}(X),$

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* provided by the criteria from  $P \subseteq C$ .

**4.2 Variable Consistency Dominance-based Rough set Approach**

The above definitions 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 tables. This extended version of the Dominance-based Rough Set Approach is called the Variable Consistency Dominance-based Rough Set Approach (VC-DRSA) model (Greco, Matarazzo, Slowinski & Stefanowski, 2001a).

For any  $P \subseteq C$ , we say that  $x \in U$  belongs to  $Cl_t^{\geq}$  with no ambiguity at consistency level  $l \in (0, 1]$ , if  $x \in Cl_t^{\geq}$  and at least  $l * 100\%$  of all objects  $y \in U$  dominating  $x$  with respect to  $P$  also belong to  $Cl_t^{\geq}$ , i.e.

$$\frac{|D_P^+(x) \cap Cl_t^{\geq}|}{|D_P^+(x)|} \geq l.$$

The term  $|D_P^+(x) \cap Cl_t^{\geq}|/|D_P^+(x)|$  is called *rough membership* and can be interpreted as conditional probability  $Pr(y \in Cl_t^{\geq}|y \in D_P^+(x))$ . The level  $l$  is called the *consistency level* because it controls the degree of consistency between objects qualified as belonging to  $Cl_t^{\geq}$  without any ambiguity. In other words, if  $l < 1$ , then at most  $(1 - l) * 100\%$  of all objects  $y \in U$  dominating  $x$  with respect to  $P$  do not belong to  $Cl_t^{\geq}$  and thus contradict the inclusion of  $x$  in  $Cl_t^{\geq}$ .

Analogously, for any  $P \subseteq C$  we say that  $x \in U$  belongs to  $Cl_t^{\leq}$  with no ambiguity at consistency level  $l \in (0, 1]$ , if  $x \in Cl_t^{\leq}$  and at least  $l * 100\%$  of all the objects  $y \in U$  dominated by  $x$  with respect to  $P$  also belong to  $Cl_t^{\leq}$ , i.e.

$$\frac{|D_P^-(x) \cap Cl_t^{\leq}|}{|D_P^-(x)|} \geq l.$$

The rough membership  $|D_P^-(x) \cap Cl_t^{\leq}|/|D_P^-(x)|$  can be interpreted as conditional probability  $Pr(y \in Cl_t^{\leq}|y \in D_P^-(x))$ . Thus, for any  $P \subseteq C$ , each object  $x \in U$  is either ambiguous or non-ambiguous at consistency level  $l$  with respect to the upward union  $Cl_t^{\geq} (t = 2, \dots, n)$  or with respect to the downward union  $Cl_t^{\leq} (t = 1, \dots, n - 1)$ .

The concept of non-ambiguous objects at some consistency level  $l$  leads naturally to the definition of  $P$ -lower approximations of the unions of classes  $Cl_t^{\geq}$  and  $Cl_t^{\leq}$  which can be formally presented as follows:

$$\underline{P}^l(Cl_t^{\geq}) = \left\{ x \in Cl_t^{\geq} : \frac{|D_P^+(x) \cap Cl_t^{\geq}|}{|D_P^+(x)|} \geq l \right\},$$

$$\underline{P}^l(Cl_t^{\leq}) = \left\{ x \in Cl_t^{\leq} : \frac{|D_P^-(x) \cap Cl_t^{\leq}|}{|D_P^-(x)|} \geq l \right\}.$$

Given  $P \subseteq C$  and consistency level  $l$ , we can define the  $P$ -upper approximations of  $Cl_t^{\geq}$  and  $Cl_t^{\leq}$ , denoted by  $\overline{P}^l(Cl_t^{\geq})$  and  $\overline{P}^l(Cl_t^{\leq})$ , respectively, by complementation of  $\underline{P}^l(Cl_{t-1}^{\leq})$  and  $\underline{P}^l(Cl_{t+1}^{\geq})$  with respect to  $U$  as follows:

$$\overline{P}^l(Cl_t^{\geq}) = U - \underline{P}^l(Cl_{t-1}^{\leq}),$$

$$\overline{P}^l(Cl_t^{\leq}) = U - \underline{P}^l(Cl_{t+1}^{\geq}).$$

$\overline{P}^l(Cl_t^{\geq})$  can be interpreted as the set of all the objects belonging to  $Cl_t^{\geq}$ , which are *possibly ambiguous* at consistency level  $l$ . Analogously,  $\overline{P}^l(Cl_t^{\leq})$  can be interpreted as the set of all the objects belonging to  $Cl_t^{\leq}$ , which are *possibly ambiguous* at consistency level  $l$ . The  $P$ -boundaries ( $P$ -doubtful regions) of  $Cl_t^{\geq}$  and  $Cl_t^{\leq}$  are defined as:

$$Bn_P(Cl_t^{\geq}) = \overline{P}^l(Cl_t^{\geq}) - \underline{P}^l(Cl_t^{\geq})$$

$$Bn_P(Cl_t^{\leq}) = \overline{P}^l(Cl_t^{\leq}) - \underline{P}^l(Cl_t^{\leq})$$

where  $t = 1, \dots, n$ . The VC-DRSA model provides some degree of flexibility in assigning objects to lower and upper approximations of the unions of decision classes. It can easily be demonstrated that for  $0 < l' < l \leq 1$  and  $t = 2, \dots, n$ ,

$$\underline{P}^l(CI_t^{\geq}) \subseteq \underline{P}^{l'}(CI_t^{\geq}) \quad \text{and} \quad \overline{P}^{l'}(CI_t^{\geq}) \subseteq \overline{P}^l(CI_t^{\geq}).$$

The VC-DRSA model is inspired by Ziarko’s model of the *variable precision* rough set approach (Ziarko, 1993, 1998). However, there is a significant difference in the definition of rough approximations because  $\underline{P}^l(CI_t^{\geq})$  and  $\overline{P}^l(CI_t^{\geq})$  are composed of non-ambiguous and ambiguous objects at the consistency level  $l$ , respectively, while Ziarko’s  $\underline{P}^l(CI_t)$  and  $\overline{P}^l(CI_t)$  are composed of  $P$ -indiscernibility sets such that at least  $l * 100\%$  of these sets are included in  $CI_t$  or have an non-empty intersection with  $CI_t$ , respectively. If one would like to use Ziarko’s definition of variable precision rough approximations in the context of multiple-criteria classification, then the  $P$ -indiscernibility sets should be substituted by  $P$ -dominating sets  $D_P^+(x)$ . However, then the notion of ambiguity that naturally leads to the general definition of rough approximations (see Slowinski & Vanderpooten (2000)) loses its meaning. Moreover, a bad side effect of the direct use of Ziarko’s definition is that a lower approximation  $\underline{P}^l(CI_t^{\geq})$  may include objects  $y$  assigned to  $CI_h$ , where  $h$  is much less than  $t$ , if  $y$  belongs to  $D_P^+(x)$ , which was included in  $\underline{P}^l(CI_t^{\geq})$ . When the decision classes are preference ordered, it is reasonable to expect that objects assigned to far worse classes than the considered union are not counted to the lower approximation of this union.

The VC-DRSA model presented above has been generalized in (Greco, Matarazzo & Slowinski, 2008b; Blaszczynski, Greco, Slowinski & Szelag, 2009). The generalized model applies two types of consistency measures in the definition of lower approximations:

- gain-type consistency measures  $f_{\geq t}^P(x), f_{\leq t}^P(x)$ :

$$\underline{P}^{\alpha_{\geq t}}(CI_t^{\geq}) = \{x \in CI_t^{\geq} > f_{\geq t}^P(x) \geq \alpha_{\geq t}\}$$

$$\underline{P}^{\alpha_{\leq t}}(CI_t^{\leq}) = \{x \in CI_t^{\leq} > f_{\leq t}^P(x) \geq \alpha_{\leq t}\}$$

- cost-type consistency measures  $g_{\geq t}^P(x), g_{\leq t}^P(x)$ :

$$\underline{P}^{\beta_{\geq t}}(CI_t^{\geq}) = \{x \in CI_t^{\geq} > g_{\geq t}^P(x) \geq \beta_{\geq t}\}$$

$$\underline{P}^{\beta_{\leq t}}(CI_t^{\leq}) = \{x \in CI_t^{\leq} > g_{\leq t}^P(x) \geq \beta_{\leq t}\}$$

where  $\alpha_{\geq t}, \alpha_{\leq t}, \beta_{\geq t}, \beta_{\leq t}$ , are threshold values on the consistency measures which are conditioning the inclusion of object  $x$  in the  $P$ -lower approximation of  $CI_t^{\geq}$ , or  $CI_t^{\leq}$ . Here are the consistency measures considered in (Blaszczynski, Greco, Slowinski & Szelag, 2009): for all  $x \in U$  and  $P \subseteq C$

$$\mu_{\geq t}^P(x) = \frac{|D_P^+(x) \cap CI_t^{\geq}|}{|D_P^+(x)|}, \quad \mu_{\leq t}^P(x) = \frac{|D_P^-(x) \cap CI_t^{\leq}|}{|D_P^-(x)|},$$

$$\begin{aligned} \bar{\mu}_{\geq t}^P(x) &= \max_{R \subseteq P} \frac{|D_R^+(z) \cap Cl_t^{\geq}|}{|D_R^+(z)|}, & \bar{\mu}_{\leq t}^P(x) &= \max_{R \subseteq P} \frac{|D_R^-(z) \cap Cl_t^{\leq}|}{|D_R^-(z)|}, \\ B_{\geq t}^P(x) &= \frac{|D_P^+(x) \cap Cl_t^{\geq}| |Cl_{t-1}^{\leq}|}{|D_P^+(x) \cap Cl_{t-1}^{\leq}| |Cl_t^{\geq}|}, & t &= 2, \dots, m, \\ B_{\leq t}^P(x) &= \frac{|D_P^-(x) \cap Cl_t^{\leq}| |Cl_{t+1}^{\geq}|}{|D_P^-(x) \cap Cl_{t+1}^{\geq}| |Cl_t^{\leq}|}, & t &= 1, \dots, m-1, \\ \varepsilon_{\geq t}^P(x) &= \frac{|D_P^+(x) \cap Cl_{t-1}^{\leq}|}{|Cl_{t-1}^{\leq}|}, & t &= 2, \dots, m, & \varepsilon_{\leq t}^P(x) &= \frac{|D_P^-(x) \cap Cl_{t+1}^{\geq}|}{|Cl_{t+1}^{\geq}|}, & t &= 1, \dots, m-1, \\ \varepsilon'_{\geq t}{}^P(x) &= \frac{|D_P^+(x) \cap Cl_{t-1}^{\leq}|}{|Cl_t^{\geq}|}, & t &= 2, \dots, m, & \varepsilon'_{\leq t}{}^P(x) &= \frac{|D_P^-(x) \cap Cl_{t+1}^{\geq}|}{|Cl_t^{\leq}|}, & t &= 1, \dots, m-1, \\ \varepsilon_{\geq t}^{*P}(x) &= \max_{r \leq t} \varepsilon_{\geq r}^P(x), & \varepsilon_{\leq t}^{*P}(x) &= \max_{r \geq t} \varepsilon_{\leq r}^P(x), \end{aligned}$$

with

$$\mu_{\geq t}^P(x), \mu_{\leq t}^P(x), \bar{\mu}_{\geq t}^P(x), \bar{\mu}_{\leq t}^P(x), B_{\geq t}^P(x), B_{\leq t}^P(x)$$

being gain-type consistency measures and

$$\varepsilon_{\geq t}^P(x), \varepsilon_{\leq t}^P(x), \varepsilon'_{\geq t}{}^P(x), \varepsilon'_{\leq t}{}^P(x), \varepsilon_{\geq t}^{*P}(x), \varepsilon_{\leq t}^{*P}(x)$$

being cost-type consistency measures.

To be concordant with the rough set philosophy, consistency measures should enjoy some monotonicity properties (see Table 1). A consistency measure is monotonic if it does not decrease (or does not increase) when:

- (m1) the set of attributes is growing,
- (m2) the set of objects is growing,
- (m3) the union of ordered classes is growing,
- (m4)  $x$  improves its evaluation, so that it dominates more objects.

As to the consistency measures  $\varepsilon_{\geq t}^P(x)$  and  $\varepsilon'_{\leq t}{}^P(x)$  which enjoy all four monotonicity properties, they can be interpreted as estimates of conditional probability, respectively:

$$Pr(y \in D_P^+(x) | y \in -Cl_t^{\geq}), \quad Pr(y \in D_P^-(x) | y \in -Cl_t^{\leq}).$$

They say how far the implications

$$y \in D_P^+(x) \Rightarrow y \in Cl_t^{\geq}, \quad y \in D_P^-(x) \Rightarrow y \in Cl_t^{\leq}$$

are **not** supported by the data.

**Table 1** – Monotonicity properties of consistency measures (Błaszczynski, Greco, Slowinski & Szlag, 2009).

Consistency measure	(m1)	(m2)	(m3)	(m4)
$\mu_{\geq t}^P(x), \mu_{\leq t}^P(x)$ (rough membership)	no	yes	yes	no
$\bar{\mu}_{\geq t}^P(x), \bar{\mu}_{\leq t}^P(x)$	yes	yes	yes	yes
$B_{\geq t}^P(x), B_{\leq t}^P(x)$ (Bayesian)	no	no	no	no
$\varepsilon_{\geq t}^P(x), \varepsilon_{\leq t}^P(x)$	yes	yes	no	yes
$\varepsilon_{\geq t}^{*P}(x), \varepsilon_{\leq t}^{*P}(x)$	yes	yes	yes	yes
$\varepsilon_{\geq t}^{\prime P}(x), \varepsilon_{\leq t}^{\prime P}(x)$	yes	yes	yes	yes

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

$$\begin{aligned} \gamma_P(Cl) &= \frac{\left| \left( U - \left( \bigcup_{t \in \{1, \dots, n\}} Bn_P(Cl_t^{\geq}) \right) \cup \left( \bigcup_{t \in \{1, \dots, n\}} Bn_P(Cl_t^{\leq}) \right) \right) \right|}{|U|} \\ &= \frac{\left| \left( U - \left( \bigcup_{t \in \{1, \dots, n\}} Bn_P(Cl_t^{\geq}) \right) \right) \right|}{|U|} \end{aligned}$$

$\gamma_P(Cl)$  can be seen as a measure of the quality of knowledge that can be extracted from the decision table, where  $P$  is the set of criteria and  $Cl$  is the considered classification.

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

- *indispensable* criteria included in the core,
- *exchangeable* criteria included in some reducts but not in the core,
- *redundant* criteria being neither indispensable nor exchangeable, thus not included in any reduct.

Note that reducts are minimal subsets of criteria conveying the relevant knowledge contained in the decision table. This knowledge is relevant for the explanation of patterns in a given decision table but not necessarily for prediction.

It has been shown in (Greco, Matarazzo & Slowinski, 2001d) that the quality of classification satisfies properties of set functions which are called *fuzzy measures*. For this reason, we can use the quality of classification for the calculation of indices which measure the relevance of particular attributes and/or criteria, in addition to the strength of interactions between them. The useful indices are: the value index and interaction indices of Shapley and Banzhaf; the interaction indices of Murofushi-Soneda and Roubens; and the Möbius representation. All these indices can help to assess the interaction between the considered criteria, and can help to choose the best reduct.

### 4.3 Stochastic dominance-based rough set approach

From a probabilistic point of view, the assignment of object  $x_i$  to “at least” class  $t$  can be made with probability  $Pr(y_i \geq t|x_i)$ , where  $y_i$  is classification decision for  $x_i$ ,  $t = 1, \dots, n$ . This probability is supposed to satisfy the usual axioms of probability:

$$Pr(y_i \geq 1|x_i) = 1, \quad Pr(y_i \leq t|x_i) = 1 - Pr(y_i \geq t + 1|x_i),$$

$$\text{and } Pr(y_i \geq t|x_i) \leq Pr(y_i \geq t'|x_i) \quad \text{for } t \geq t'.$$

These probabilities are unknown but can be estimated from data.

For each class  $t = 2, \dots, n$ , we have a binary problem of estimating the conditional probabilities  $Pr(y_i \geq t|x_i) = 1, Pr(y_i < t|x_i)$ . It can be solved by *isotonic regression* (Kotlowski, Dembczynski, Greco & Slowinski, 2008). Let  $y_{it} = 1$  if  $y_i \geq t$ , otherwise  $y_{it} = 0$ . Let also  $p_{it}$  be the estimate of the probability  $Pr(y_i \geq t|x_i)$ . Then, choose estimates  $p_{it}^*$  which minimize the squared distance to the class assignment  $y_{it}$ , subject to the monotonicity constraints:

$$\text{Minimize } \sum_{i=1}^{|U|} (y_{it} - p_{it})^2$$

$$\text{subject to } x_i \succeq x_j \rightarrow p_{it} \geq p_{jt} \quad \text{for all } x_i, x_j \in U$$

where  $x_i \succeq x_j$  means that  $x_i$  dominates  $x_j$ .

Then, stochastic  $\alpha$ -lower approximations for classes “at least  $t$ ” and “at most  $t - 1$ ” can be defined as:

$$\underline{P}^\alpha(CI_t^{\geq}) = \{x_i \in U : Pr(y_i \geq t|x_i) \geq \alpha\},$$

$$\underline{P}^\alpha(CI_{t-1}^{\leq}) = \{x_i \in U : Pr(y_i < t|x_i) \geq \alpha\}.$$

Replacing the unknown probabilities  $Pr(y_i \geq t|x_i), Pr(y_i < t|x_i)$ , by their estimates  $p_{it}^*$  obtained from isotonic regression, we get:

$$\underline{P}^\alpha(CI_t^{\geq}) = \{x_i \in U : p_{it}^* \geq \alpha\},$$

$$\underline{P}^\alpha(CI_{t-1}^{\leq}) = \{x_i \in U : p_{it}^* \leq 1 - \alpha\},$$

where parameter  $\alpha \in [0.5, 1]$  controls the allowed amount of inconsistency.

Solving isotonic regression requires  $O(|U|^4)$  time, but a good heuristic needs only  $O(|U|^2)$ .

In fact, as shown in (Kotlowski, Dembczynski, Greco & Slowinski, 2008), we don't really need to know the probability estimates to obtain stochastic lower approximations. We only need to know for which object  $x_i$ ,  $p_{it}^* \geq \alpha$  and for which  $x_i$ ,  $p_{it}^* \leq 1 - \alpha$ . This can be found by solving a linear programming (reassignment) problem.

As before,  $y_{it} = 1$  if  $y_i \geq t$ , otherwise  $y_{it} = 0$ . Let  $d_{it}$  be the decision variable which determines a new class assignment for object  $x_i$ . Then, reassign objects from union of classes indicated by  $y_{it}$  to union of classes indicated by  $d_{it}^*$ , such that the new class assignments are consistent with the dominance principle, where  $d_{it}^*$  results from solving the following linear programming problem:

$$\begin{aligned} &\text{Minimize} && \sum_{i=1}^{|U|} w_{y_{it}} |y_{it} - d_{it}| \\ &\text{subject to} && x_i \succeq x_j \rightarrow d_{it} \geq d_{jt} \quad \text{for all } x_i, x_j \in U \end{aligned}$$

where  $w_{it}$  are arbitrary positive weights and  $x_i \succeq x_j$  means that  $x_i$  dominates  $x_j$ .

Due to unimodularity of the constraint matrix, the optimal solution of this linear programming problem is always integer, i.e.  $d_{it}^* \in \{0, 1\}$ . For all objects consistent with the dominance principle,  $d_{it}^* = y_{it}$ . If we set  $w_0 = \alpha$  and  $w_1 = \alpha - 1$ , then the optimal solution  $d_{it}^*$  satisfies:  $d_{it}^* = 1 \Leftrightarrow p_{it}^* \geq \alpha$ . If we set  $w_0 = 1 - \alpha$  and  $w_1 = \alpha$ , then the optimal solution  $d_{it}^*$  satisfies:  $d_{it}^* = 0 \Leftrightarrow p_{it}^* \leq 1 - \alpha$ .

For each  $t = 2, \dots, n$ , solving the reassignment problem twice, we can obtain the lower approximations  $\underline{P}^\alpha(Cl_t^\geq)$ ,  $\underline{P}^\alpha(Cl_{t-1}^\leq)$ , without knowing the probability estimates!

**4.4 Induction of decision rules**

Using the terms of knowledge discovery, the dominance-based rough approximations of upward and downward unions of classes are applied on the data set in the pre-processing stage. In result of this stage, the data are structured in a way facilitating induction of "if..., then..." decision rules with a guaranteed consistency level. 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 are negative, suggests 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  $\overline{P}(Cl_s^\geq) \cap \overline{P}(Cl_t^\leq)$  are positive and all the others are negative, are suggesting an assignment to some classes between  $Cl_s$  and  $Cl_t (s < t)$ .

In the case of preference ordered data it is meaningful to consider the following five types of decision rules:

- 1) *Certain  $D_{\geq}$ -decision rules.* These provide lower profile descriptions for objects belonging to  $Cl_t^\geq$  without ambiguity: *if  $x_{q1} \succeq_{q1} r_{q1}$  and  $x_{q2} \succeq_{q2} r_{q2}$  and...  $x_{qp} \succeq_{qp} r_{qp}$ , then  $x \in Cl_t^\geq$ , where for each  $w_q, z_q \in X_q$ , " $w_q \succeq_q z_q$ " means " $w_q$  is at least as good as  $z_q$ ";*



- 2) *Possible  $D_{\geq}$ -decision rules.* Such rules provide lower profile descriptions for objects belonging to  $Cl_t^{\geq}$  with or without any ambiguity: if  $x_{q1} \succeq_{q1} r_{q1}$  and  $x_{q2} \succeq_{q2} r_{q2}$  and...  $x_{qp} \succeq_{qp} r_{qp}$ , then  $x$  possibly belongs to  $Cl_t^{\geq}$ ;
- 3) *Certain  $D_{\geq}$ -decision rules.* These give upper profile descriptions for objects belonging to  $Cl_t^{\leq}$  without ambiguity: if  $x_{q1} \preceq_{q1} r_{q1}$  and  $x_{q2} \preceq_{q2} r_{q2}$  and...  $x_{qp} \preceq_{qp} r_{qp}$ , then  $x \in Cl_t^{\leq}$ , where for each  $w_q, z_q \in X_q$ , " $w_q \preceq z_q$ ", means " $w_q$  is at most as good as  $z_q$ ";
- 4) *Possible  $D_{\leq}$ -decision rules.* These provide upper profile descriptions for objects belonging to  $Cl_t^{\leq}$  with or without any ambiguity: if  $x_{q1} \preceq_{q1} r_{q1}$  and  $x_{q2} \preceq_{q2} r_{q2}$  and...  $x_{qp} \preceq_{qp} r_{qp}$ , then  $x$  possibly belongs to  $Cl_t^{\leq}$ ;
- 5) *Approximate  $D_{\geq\leq}$ -decision rules.* These represent simultaneously lower and upper profile descriptions for objects belonging to  $Cl_s \cup Cl_{s+1} \cup \dots \cup Cl_t$  without the possibility of discerning the actual class: if  $x_{q1} \succeq_{q1} r_{q1}$  and...  $x_{qk} \succeq_{qk} r_{qk}$  and  $x_{qk+1} \preceq_{qk+1} r_{qk+1}$  and...  $x_{qp} \preceq_{qp} r_{qp}$ , then  $x \in Cl_s \cup Cl_{s+1} \cup \dots \cup Cl_t$ .

In the left hand side of a  $D_{\geq\leq}$ -decision rule we can have " $x_q \succeq_q r_q$ " and " $x_q \preceq_q r'_q$ ", where  $r_q \leq r'_q$ , for the same  $q \in C$ . Moreover, if  $r_q = r'_q$ , the two conditions boil down to " $x_q \sim_q r_q$ ", where for each  $w_q, z_q \in X_q$ , " $w_q \sim_q z_q$ " means " $w_q$  is indifferent to  $z_q$ ".

A *minimal* rule is an implication where we understand that there is no other implication with a left hand side which has at least the same weakness (which means that it uses a subset of elementary conditions and/or weaker elementary conditions) and which has a right hand side that has at least the same strength (which means, a  $D_{\geq}$ - or a  $D_{\leq}$ -decision rule assigning objects to the same union or sub-union of classes, or a  $D_{\geq\leq}$ -decision rule assigning objects to the same or larger set of classes).

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

The rules of type 1) and 3) are *exact* if they do not cover negative examples; they are *probabilistic*, otherwise. In the latter case, each rule is characterized by a confidence ratio, representing the probability that an object matching left hand side of the rule matches also its right hand side. Probabilistic rules concord with the Variable-Consistency Dominance-based Rough Set Approach model mentioned above.

#### 4.5 Rule-based classification algorithms

We will now comment upon the application of decision rules to some objects described by criteria from  $C$ . When applying  $D_{\geq}$ -decision rules to an object  $x$ , it is possible that  $x$  either matches the left hand side of at least one decision rule or it does not. In the case of at least one such match, it is reasonable to conclude that  $x$  belongs to class  $Cl_t$ , because it is the lowest class of the upward union  $Cl_t^{\geq}$  which results from intersection of all the right hand sides of the rules covering  $x$ . More

precisely, if  $x$  matches the left hand side of rules  $\rho_1, \rho_2, \dots, \rho_m$ , having right hand sides  $x \in C_{t1}^{\geq}, x \in Cl_{t2}^{\geq}, \dots, x \in Cl_{tm}^{\geq}$ , then  $x$  is assigned to class  $Cl_t$ , where  $t = \max\{t1, t2, \dots, tm\}$ . In the case of no matching, we can conclude that  $x$  belongs to class  $Cl_1$ , i.e. to the worst class, since no rule with a right hand side suggesting a better classification of  $x$  is covering this object.

Analogously, when applying  $D_{\leq}$ -decision rules to the object  $x$ , we can conclude that  $x$  belongs either to class  $Cl_z$  (because it is the highest class of the downward union  $Cl_t^{\leq}$  resulting from the intersection of all the right hand sides of the rules covering  $x$ ), or to class  $Cl_n$ , i.e. to the best class, when  $x$  is not covered by any rule. More precisely, if  $x$  matches the left hand side of rules  $\rho_1, \rho_2, \dots, \rho_m$ , having right hand sides  $x \in C_{t1}^{\leq}, x \in Cl_{t2}^{\leq}, \dots, x \in Cl_{tm}^{\leq}$ , then  $x$  is assigned to class  $Cl_t$ , where  $t = \min\{t1, t2, \dots, tm\}$ . In the case of no matching, it is concluded that  $x$  belongs to the best class  $Cl_n$  because no rule with a right hand side suggesting a worse classification of  $x$  is covering this object. Finally, when applying  $D_{\geq\leq}$ -decision rules to  $x$ , it is possible to conclude that  $x$  belongs to the union of all the classes suggested in the right hand side of the rules covering  $x$ .

A new classification algorithm has been proposed in (Blaszczynski, Greco & Slowinski, 2007). Let  $\varphi_1 \rightarrow \psi_1, \dots, \varphi_k \rightarrow \psi_k$ , be the rules matching object  $x$ . Then,  $R_t(x) = \{j : Cl_t \in \psi_j, j = 1, \dots, k\}$  denotes the set of rules matching  $x$ , which are recommending assignment of object  $x$  to a union including class  $Cl_t$ , and  $R_{\neg t}(x) = \{j : Cl_t \notin \psi_j, j = 1, \dots, k\}$  denotes the set of rules matching  $x$ , which are not recommending assignment of object  $x$  to a union including class  $Cl_t$ .  $||\varphi_j||, ||\psi_j||$  are sets of objects with property  $\varphi_j$  and  $\psi_j$ , respectively,  $j = 1, \dots, k$ . For a classified object  $x$ , one has to calculate the score for each candidate class:

$$score(Cl_t, x) = score^+(Cl_t, x) - score^-(Cl_t, x),$$

where

$$score^+(Cl_t, x) = \frac{\left| \bigcup_{j \in R_t(x)} (||\varphi_j|| \cap Cl_t) \right|^2}{\left| \bigcup_{j \in R_t(x)} ||\varphi_j|| \right| \times |Cl_t|}$$

and

$$score^-(Cl_t, x) = \frac{\left| \bigcup_{j \in R_{\neg t}(x)} (||\varphi_j|| \cap ||\psi_j||) \right|^2}{\left| \bigcup_{j \in R_{\neg t}(x)} ||\varphi_j|| \right| \times \left| \bigcup_{j \in R_{\neg t}(x)} ||\psi_j|| \right|}$$

$score^+(Cl_t, x)$  and  $score^-(Cl_t, x)$  can be interpreted in terms of conditional probability as a product of confidence and coverage of the matching rules:

$$score^+(Cl_t, x) = Pr(\{\varphi_j : j \in R_t(x)\} | Cl_t) \times Pr(Cl_t \{ \varphi_j : j \in R_t(x) \}),$$

$$score^-(Cl_t, x) = Pr(\{\varphi_j : j \in R_{\neg t}(x)\} | \neg Cl_t) \times Pr(\neg Cl_t \{ \varphi_j : j \in R_{\neg t}(x) \}).$$

The recommendation of the univocal classification  $x \rightarrow Cl_t$  is such that:

$$Cl_t = \arg \max_{t \in \{1, \dots, n\}} [score(Cl_t, x)].$$

A set of decision rules is *complete* if it is able to cover all objects from the decision table in such a way that consistent objects are re-classified to their original classes and inconsistent objects are classified to clusters of classes which refer to this inconsistency. Each set of decision rules that is complete and non-redundant is called *minimal*. Note that an exclusion of any rule from this set makes it non-complete.

In the case of the Variable-Consistency Dominance-based Rough Set Approach, the decision rules are induced from the  $P$ -lower approximations whose composition is controlled by the user-specified consistency level  $l$ . Consequently, the value of confidence  $\alpha$  for the rule should be constrained from the bottom. It is reasonable to require that the smallest accepted confidence level of the rule should not be lower than the currently used consistency level  $l$ . Indeed, in the worst case, some objects from the  $P$ -lower approximation may create a rule using all the criteria from  $P$  thus giving a confidence  $\alpha \geq l$ .

Observe that the syntax of decision rules induced from dominance-based rough approximations uses the concept of dominance cones: each condition profile is a dominance cone in  $X_C$ , and each decision profile is a dominance cone in  $X_D$ . In both cases the cone is positive for  $D_{\geq}$ -rules and negative for  $D_{\leq}$ -rules.

Also note that dominance cones which correspond to condition profiles can originate in any point of  $X_C$ , without the risk of being too specific. Thus, in contrast to granular computing based on indiscernibility (or similarity) relation, in case of granular computing based on dominance, the condition attribute space  $X_C$  need not be discretized (Greco, Matarazzo & Slowinski, 2007, 2008a, 2009).

Procedures for induction of rules from dominance-based rough approximations have been proposed in (Greco, Matarazzo, Slowinski & Stefanowski, 2001b; Blaszczyński, Slowinski & Szeląg, 2011).

The utility of decision rules is threefold: they *explain* (summarize) decisions made on objects from the dataset, they can be used to *make decisions* with respect to new (unseen) objects which are matching conditions of some rules, and they permit to *build up a strategy of intervention* (Greco, Matarazzo, Pappalardo & Slowinski, 2005). Attractiveness of particular decision rules can be measured in many different ways, however, the most convincing measures are Bayesian confirmation measures enjoying a special monotonicity property, as reported in (Greco, Pawlak & Slowinski, 2004).

In Giove, Greco, Matarazzo & Slowinski (2002), a new methodology for the induction of monotonic decision trees from dominance-based rough approximations of preference-ordered decision classes has been proposed.

It is finally worth noting that several algebraic models have been proposed for Dominance-based Rough Set Approach (Greco, Matarazzo & Slowinski, 2010a, 2010g, 2012) – the algebraic structures are based on bipolar disjoint representation (positive and negative) of interior and exterior of

a concept. These algebra models give elegant representations of basic properties of Dominance-based Rough Sets. Moreover, a topology for Dominance-based Rough Set Approach in a bitopological space has been proposed in (Greco, Matarazzo & Slowinski, 2010b).

#### 4.6 An illustrative example

To illustrate the application of the Dominanced Based Rough Set Approach to multicriteria classification, we will use a part of some data provided by a Greek industrial bank ETEVA which finances industrial and commercial firms in Greece (Slowinski & Zopounidis, 1995; Slowinski *et al.*, 2005). 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 bankruptcy risk. The classification 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  $Cl = \{Cl_A, Cl_U, Cl_{NA}\}$  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,  $\uparrow$
- $A_2$  = net income/net worth,  $\uparrow$
- $A_3$  = total liabilities/total assets,  $\downarrow$
- $A_4$  = total liabilities/cash flow,  $\downarrow$
- $A_5$  = interest expenses/sales,  $\downarrow$
- $A_6$  = general and administrative expense/sales,  $\downarrow$
- $A_7$  = managers' work experience,  $\uparrow$  (very low = 1, low = 2, medium = 3, high = 4, very high = 5)
- $A_8$  = firm's market niche/position,  $\uparrow$  (bad = 1, rather bad = 2, medium = 3, good = 4, very good = 5)
- $A_9$  = technical structure-facilities,  $\uparrow$  (bad = 1, rather bad = 2, medium = 3, good = 4, very good = 5)
- $A_{10}$  = organization-personnel,  $\uparrow$  (bad=1, rather bad = 2, medium = 3, good = 4, very good = 5)
- $A_{11}$  = special competitive advantage of firms,  $\uparrow$  (low =1, medium =2, high = 3, very high = 4)
- $A_{12}$  = market flexibility,  $\uparrow$  (very low = 1, low = 2, medium = 3, high = 4, very high = 5)

The first six criteria are cardinal (financial ratios) and the last six are ordinal. The data table is presented in Table 2.

Table 2 – Financial data table.

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 2 consistent?
- What are the reducts of criteria ensuring the same quality of approximation of the multi-criteria classification as the whole set of criteria?
- What decision rules can be extracted from Table 2?
- What are the minimal sets of decision rules?

We will answer these questions using the Dominance-based Rough Set Approach. The *first result* from this approach is a discovery that the financial data table is *consistent* for the complete set of criteria  $C$ . Therefore, the  $C$ -lower and  $C$ -upper approximations of  $C_{NA}^{\leq}$ ,  $C_U^{\leq}$  and  $C_U^{\geq}$ ,  $C_A^{\geq}$  are the same. In other words, the quality of approximation of all upward and downward unions of classes, as well as the quality of classification, 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:

$$\begin{array}{ll}
 RED_{CI}^1 = \{A_1, A_4, A_5, A_7\}, & RED_{CI}^2 = \{A_2, A_4, A_5, A_7\}, \\
 RED_{CI}^3 = \{A_3, A_4, A_6, A_7\}, & RED_{CI}^4 = \{A_4, A_5, A_6, A_7\}, \\
 RED_{CI}^5 = \{A_4, A_5, A_7, A_8\}, & RED_{CI}^6 = \{A_2, A_3, A_7, A_9\}, \\
 RED_{CI}^7 = \{A_1, A_3, A_4, A_7, A_9\}, & RED_{CI}^8 = \{A_1, A_5, A_7, A_9\}, \\
 RED_{CI}^9 = \{A_2, A_5, A_7, A_9\}, & RED_{CI}^{10} = \{A_4, A_5, A_7, A_9\}, \\
 RED_{CI}^{11} = \{A_5, A_6, A_7, A_9\}, & RED_{CI}^{12} = \{A_4, A_5, A_7, A_{10}\}, \\
 RED_{CI}^{13} = \{A_1, A_3, A_4, A_7, A_{11}\}, & RED_{CI}^{14} = \{A_2, A_3, A_4, A_7, A_{11}\}, \\
 RED_{CI}^{15} = \{A_4, A_5, A_6, A_{12}\}, & RED_{CI}^{16} = \{A_1, A_3, A_5, A_6, A_9, A_{12}\}, \\
 RED_{CI}^{17} = \{A_3, A_4, A_6, A_{11}, A_{12}\}, & RED_{CI}^{18} = \{A_1, A_2, A_3, A_6, A_9, A_{11}, A_{12}\}.
 \end{array}$$

All the eighteen subsets of criteria are equally good and sufficient for the perfect approximation of the classification performed by ETEVA’s financial manager on the 39 firms. The core of  $CI$  is empty ( $CORE_{CI} = \emptyset$ ) 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  $CI_{NA}^{\leq}$ , 51 rules describing  $CI_U^{\leq}$ , 75 rules describing  $CI_U^{\geq}$  and 79 rules describing  $CI_A^{\geq}$ .

The *fourth discovery* is the finding of *minimal sets* of decision rules. Several minimal sets were found. One of them is shown below. The number in parenthesis indicates the number of objects which support the corresponding rule, *i.e.* the rule strength:

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  $x \in CI_{NA}^{\leq}$ , (4),
2. if  $f(x, A_2) \leq 3.3$  and  $f(x, A_7) \leq 2$ , then  $x \in CI_{NA}^{\leq}$ , (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  $x \in CI_{NA}^{\leq}$ , (4),

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

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

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

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

$$8. \text{ if } f(x, A_1) \geq 8.7 \text{ and } f(x, A_9) \geq 4, \text{ then } x \in Cl_A^{\geq}, \quad (27),$$

$$9. \text{ if } f(x, A_2) \geq 3.5 \text{ and } f(x, A_7) \geq 4, \text{ then } x \in Cl_A^{\geq}, \quad (20).$$

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 criteria and 18 elementary conditions, *i.e.* 3.85% of descriptors from the data table.

The well-known machine discovery methods cannot deal with multicriteria classification because they do not consider preference orders in the domains of attributes and among the classes. There are multicriteria decision analysis methods for such classification. However, they are not discovering classification rules from data. They simply apply a preference model, like the utility function in scoring methods (see, *e.g.*, (Thomas, Crook & Edelman, 1992)), to a set of objects to be classified. In this sense, they are not knowledge discovery methods at all.

Comparing the Dominance-based Rough Set Approach to the Indiscernibility-based Rough Set Approach, we can notice the following differences between the two approaches. The Indiscernibility-based Rough Set Approach extracts knowledge about a partition of  $U$  into classes which are not preference-ordered. The granules used for knowledge representation are sets of objects which are indiscernible by a set of condition attributes.

In the case of the Dominance-based Rough Set Approach and multicriteria classification, the condition attributes are criteria and the 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 the dominance relation. This is the main difference between the Indiscernibility-based Rough Set Approach and the Dominance-based Rough Set Approach.

There are three notable advantages of the Dominance-based Rough Set Approach over the Indiscernibility-based Rough Set Approach. The *first* one is the ability to handle criteria, preference-ordered classes and inconsistencies in the set of decision examples that the Indiscernibility-based Rough Set Approach is simply not able to discover. Consequently, the rough approximations separate the certain information from the doubtful, which is taken into account in rule induction. The *second* advantage is the ability to analyze a data table without any pre-processing of data. The *third* advantage lies in the richer syntax of decision rules that are induced from rough approximations. The elementary conditions of decision rules resulting from

Dominance-based Rough Set Approach use relations from  $\{\leq, =, \geq\}$ , while those resulting from the Indiscernibility-based Rough Set Approach only use  $=$ . The Dominance-based Rough Set Approach syntax is more understandable to practitioners. The minimal sets of decision rules are smaller than the minimal sets which result from the Indiscernibility-based Rough Set Approach.

## 5 THE DOMINANCE-BASED ROUGH SET APPROACH TO MULTICRITERIA CHOICE AND RANKING

One of the very first extensions of the Dominance-based Rough Set Approach concerned preference-ordered data representing pairwise comparisons (*i.e.* binary relations) between objects on both, condition and decision attributes (Greco, Matarazzo & Slowinski, 1999a, 1999b, 2000d, 2001c). Note that while classification is based on the absolute evaluation of objects, choice and ranking refer to pairwise comparisons of objects. In this case, the decision rules to be discovered from the data characterize a comprehensive binary relation on the set of objects. If this relation is a preference relation and if, among the condition attributes, there are some criteria which are semantically correlated with the comprehensive preference relation, then the data set (serving as the learning sample) can be considered to be preference information for a decision maker in a multicriteria choice or ranking problem. In consequence, the comprehensive preference relation characterized by the decision rules discovered from this data set can be considered as a *preference model* for the decision maker. It may be used to explain the decision policy of the decision maker and to recommend a good choice or preference ranking with respect to new objects.

Let us consider a finite set  $A$  of objects evaluated by a finite set  $C$  of criteria. The best choice (or the preference ranking) in set  $A$  is semantically correlated with the criteria from set  $C$ . The preference information concerning the multicriteria choice or ranking problem is a data set in the form of a pairwise comparison table which includes pairs of some *reference objects* from a subset  $B \subseteq A \times A$ . This is described by preference relations on particular criteria and a comprehensive preference relation. One such example is a weak preference relation called the outranking relation. By using the Dominance-based Rough Set Approach for the analysis of the pairwise comparison table, we can obtain a rough approximation of the *outranking relation* by a dominance relation. The decision rules induced from the rough approximation are then applied to the complete set  $A$  of the objects associated with the choice or ranking. As a result, one obtains a four-valued outranking relation on this set. In order to obtain a recommendation, it is advisable to use an exploitation procedure based on the net flow score of the objects. We present this methodology in more detail below.

### 5.1 The pairwise comparison table as input preference information

Given a multicriteria choice or ranking problem, a decision maker can express the preferences by pairwise comparisons of the reference objects. In the following,  $xSy$  will denote the presence, while  $xS^c y$  denotes the absence of the outranking relation for a pair of objects  $(x, y) \in A \times A$ .

For each pair of reference objects  $(x, y) \in B \subseteq A \times A$ , the decision maker can select one of the three following possibilities:



- 1) object  $x$  is as good as  $y$ , *i.e.*  $xSy$ ,
- 2) object  $x$  is worse than  $y$ , *i.e.*  $xS^c y$ ,
- 3) the two objects are incomparable at the present stage.

A pairwise comparison table, denoted by  $S_{PCT}$ , is then created on the basis of this information. The first  $m$  columns correspond to the criteria from set  $C$ . The last, *i.e.* the  $(m + 1)$ -th column, represents the comprehensive binary preference relation  $S$  or  $S^c$ . The rows correspond to the pairs from  $B$ . For each pair in  $S_{PCT}$ , a difference between criterion values is put in the corresponding column. If the decision maker judges that two objects are incomparable, then the corresponding pair does not appear in  $S_{PCT}$ .

We will define  $S_{PCT}$  more formally. For any criterion  $g_i \in C$ , let  $T_i$  be a finite set of binary relations defined on  $A$  on the basis of the evaluations of objects from  $A$  with respect to the considered criterion  $g_i$ , such that for every  $(x, y) \in A \times A$  exactly one binary relation  $t \in T_i$  is verified. More precisely, given the domain  $V_i$  of  $g_i \in C$ , if  $v'_i, v''_i \in V_i$  are the respective evaluations of  $x, y \in A$  by means of  $g_i$  and  $(x, y) \in t$ , with  $t \in T_i$ , then for each  $w, z \in A$  having the same evaluations  $v'_i, v''_i$  by means of  $g_i$ ,  $(w, z) \in t$ . Furthermore, let  $T_d$  be a set of binary relations defined on set  $A$  (comprehensive pairwise comparisons) such that at most one binary relation  $t \in T_d$  is verified for every  $(x, y) \in A \times A$ .

The *pairwise comparison table* is defined as data table  $S_{PCT} = \langle B, C \cup \{d\}, T_G \cup T_d, f \rangle$ , where  $B \subseteq A \times A$  is a non-empty set of exemplary pairwise comparisons of reference objects,  $T_G = \bigcup_{g_i \in C} T_i$ ,  $d$  is a decision corresponding to the comprehensive pairwise comparison (comprehensive preference relation), and  $f : B \times (C \cup \{d\}) \rightarrow T_G \cup T_d$  is a total function such that  $f[(x, y), g_i] \in T_i$  for every  $(x, y) \in A \times A$  and for each  $g_i \in C$ , and  $f[(x, y), d] \in T_d$  for every  $(x, y) \in B$ . It follows that for any pair of reference objects  $(x, y) \in B$  there is verified one and only one binary relation  $t \in T_d$ . Thus,  $T_d$  induces a partition of  $B$ . In fact, the data table  $S_{PCT}$  can be seen as decision table, since the set of considered criteria  $C$  and the decision  $d$  are distinguished.

We are considering a pairwise comparison table where the set  $T_d$  is composed of two binary relations defined on  $A$ :

- $x$  outranks  $y$  (denoted by  $xSy$  or  $(x, y) \in S$ ), where  $(x, y) \in B$ ,
- $x$  does not outrank  $y$  (denoted by  $xS^c y$  or  $(x, y) \in S^c$ ), where  $(x, y) \in B$ , and  $S \cup S^c = B$ .

Observe that the binary relation  $S$  is reflexive, but not necessarily transitive or complete.

### 5.2 Rough approximation of preference relations specified in the pairwise comparison table

In the following we will distinguish between two types of evaluation scales of criteria: *cardinal* and *ordinal*. Let  $C^N$  be the set of criteria expressing preferences on a cardinal scale, and let  $C^O$ , be the set of criteria expressing preferences on an ordinal scale, such that  $C^N \cup C^O = C$  and

$C^N \cap C^O = \emptyset$ . Moreover, for each  $P \subseteq C$ , we denote by  $P^O$  the subset of  $P$  composed of criteria expressing preferences on an ordinal scale, i.e.  $P^O = P \cap C^O$ , and by  $P^N$  we denote the subset of  $P$  composed of criteria expressing preferences on a cardinal scale, i.e.  $P^N = P \cap C^N$ . Of course, for each  $P \subseteq C$ , we have  $P = P^N \cup P^O$  and  $P^N \cap P^O = \emptyset$ .

The meaning of the two scales is such that in the case of the cardinal scale we can specify the intensity of preference for a given difference of evaluations, while in the case of the ordinal scale, this is not possible and we can only establish an order of evaluations.

**5.2.1 Multigraded dominance**

We assume that the pairwise comparisons of reference objects on cardinal criteria from set  $C^N$  can be represented in terms of *graded preference relations* (for example “very weak preference”, “weak preference”, “strict preference”, “strong preference” and “very strong preference”), denoted by  $P_q^h$ : for each  $q \in C^N$  and for every  $(x, y) \in A \times A$ ,  $T_i = \{P_i^h, h \in H_i\}$ , where  $H_i$  is a particular subset of the relative integers and

- $xP_i^h y, h > 0$ , means that object  $x$  is preferred to object  $y$  by degree  $h$  with respect to criterion  $g_i$ ,
- $xP_i^h y, h < 0$ , means that object  $x$  is not preferred to object  $y$  by degree  $h$  with respect to criterion  $g_i$ ,
- $xP_i^0 y$  means that object  $x$  is similar (asymmetrically indifferent) to object  $y$  with respect to criterion  $g_i$ .

Within the preference context, the similarity relation  $P_i^0$ , even if not symmetric, resembles the indifference relation. Thus, in this case, we call this similarity relation “asymmetric indifference”. Of course, for each  $g_i \in C$  and for every  $(x, y) \in A \times A$ ,

$$[xP_i^h y, h > 0] \Rightarrow [yP_i^k x, k \leq 0], [xP_i^h y, h < 0] \Rightarrow [yP_i^k x, k \geq 0].$$

Let  $P = P^N$  and  $P^O = \emptyset$ . Given  $P \subseteq C$  ( $P \neq \emptyset$ ),  $(x, y), (w, z) \in A \times A$ , the pair of objects  $(x, y)$  is said to dominate  $(w, z)$  with respect to criteria from  $P$  (denoted by  $(x, y)D_P(w, z)$ ), if  $x$  is preferred to  $y$  at least as strongly as  $w$  is preferred to  $z$  with respect to each  $g_i \in P$ . More precisely, “at least as strongly as” means “by at least the same degree”, i.e.  $h \geq k$ , where  $h, k \in H_i, xP_i^h y$  and  $wP_i^k z$ , for each  $g_i \in P$ .

Let  $D_{\{i\}}$  be the dominance relation confined to the single criterion  $g_i \in P$ . The binary relation  $D_{\{i\}}$  is reflexive ( $(x, y)D_{\{i\}}(x, y)$ , for every  $(x, y) \in A \times A$ ), transitive ( $(x, y)D_{\{i\}}(w, z)$  and  $(w, z)D_{\{i\}}(u, v)$  imply  $(x, y)D_{\{i\}}(u, v)$ , for every  $(x, y), (w, z), (u, v) \in A \times A$ ), and complete ( $(x, y)D_{\{i\}}(w, z)$  and/or  $(w, z)D_{\{i\}}(x, y)$ , for all  $(x, y), (w, z) \in A \times A$ ). Therefore,  $D_{\{i\}}$  is a complete preorder on  $A \times A$ . Since the intersection of complete preorders is a partial preorder and  $D_P = \bigcap_{g_i \in P} D_{\{i\}}$ ,  $P \subseteq C$ , then the dominance relation  $D_P$  is a partial preorder on  $A \times A$ .

Let  $R \subseteq P \subseteq C$  and  $(x, y), (u, v) \in A \times A$ ; then the following implication holds:

$$(x, y)D_P(u, v) \Rightarrow (x, y)D_R(u, v).$$

Given  $P \subseteq C$  and  $(x, y) \in A \times A$ , we define the following:

- A set of pairs of objects dominating  $(x, y)$ , called the *P-dominating set*, denoted by  $D_P^+(x, y)$  and defined to be  $\{(w, z) \in A \times A : (w, z)D_P(x, y)\}$ ;
- A set of pairs of objects dominated by  $(x, y)$ , called the *P-dominated set*, denoted by  $D_P^-(x, y)$  and defined as  $\{(w, z) \in A \times A : (x, y)D_P(w, z)\}$ .

The *P-dominating sets* and the *P-dominated sets* defined on  $B$  for all pairs of reference objects from  $B$  are “granules of knowledge” that can be used to express *P-lower* and *P-upper* approximations of the comprehensive outranking relations  $S$  and  $S^c$ , respectively:

$$\underline{P}(S) = \{(x, y) \in B : D_P^+(x, y) \subseteq S\},$$

$$\overline{P}(S) = \bigcup_{(x,y) \in S} D_P^+(x, y).$$

$$\underline{P}(S^c) = \{(x, y) \in B : D_P^-(x, y) \subseteq S^c\},$$

$$\overline{P}(S^c) = \bigcup_{(x,y) \in S^c} D_P^-(x, y).$$

It has been proved in (Greco, Matarazzo & Slowinski, 1999a) that

$$\underline{P}(S) \subseteq S \subseteq \overline{P}(S), \underline{P}(S^c) \subseteq S^c \subseteq \overline{P}(S^c).$$

Furthermore, the following complementarity properties hold:

$$\underline{P}(S) = B - \overline{P}(S^c), \overline{P}(S) = B - \underline{P}(S^c),$$

$$\underline{P}(S^c) = B - \overline{P}(S), \overline{P}(S^c) = B - \underline{P}(S).$$

The *P-boundaries* (*P-doubtful regions*) of  $S$  and  $S^c$  are defined as

$$Bn_P(S) = \overline{P}(S) - \underline{P}(S), Bn_P(S^c) = \overline{P}(S^c) - \underline{P}(S^c).$$

From the above it follows that  $Bn_P(S) = Bn_P(S^c)$ .

The concepts of the quality of approximation, reducts and core can be extended also to the approximation of the outranking relation by multigraded dominance relations.

In particular, the coefficient

$$\gamma_P = \frac{|\underline{P}(S) \cup \underline{P}(S^c)|}{|B|}$$

defines the *quality of approximation* of  $S$  and  $S^c$  by  $P \subseteq C$ . It expresses the ratio of all pairs of reference objects  $(x, y) \in B$  correctly assigned to  $S$  and  $S^c$  by the set  $P$  of criteria to all the pairs of objects contained in  $B$ . Each minimal subset  $P \subseteq C$ , such that  $\gamma_P = \gamma_C$ , is called a *reduct* of  $C$  (denoted by  $RED_{S_{PCT}}$ ). Note that  $S_{PCT}$  can have more than one reduct. The intersection of all  $B$ -reducts is called the *core* (denoted by  $CORE_{S_{PCT}}$ ).

It is also possible to use the Variable Consistency Model on  $S_{PCT}$  (Slowinski, Greco & Matarazzo, 2002b), being aware that some of the pairs in the positive or negative dominance sets belong to the opposite relation, while at least  $l * 100\%$  of pairs belong to the correct one. Then the definition of the lower approximations of  $S$  and  $S^c$  boils down to:

$$\underline{P}(S) = \left\{ (x, y) \in B : \frac{|D_P^+(x, y) \cap S|}{|D_P^+(x, y)|} \geq l \right\}$$

$$\underline{P}(S^c) = \left\{ (x, y) \in B : \frac{|D_P^-(x, y) \cap S^c|}{|D_P^-(x, y)|} \geq l \right\}$$

**5.2.2 Dominance without degrees of preference**

The degree of graded preference considered above is defined on a cardinal scale of the strength of preference. However, in many real world problems, the existence of such a quantitative scale is rather questionable. This is the case with ordinal scales of criteria. In this case, the dominance relation is defined directly on evaluations  $g_i(x)$  for all objects  $x \in A$ . Let us explain this latter case in more detail.

Let  $P = P^O$  and  $P^N = \emptyset$ , then, given  $(x, y), (w, z) \in A \times A$ , the pair  $(x, y)$  is said to dominate the pair  $(w, z)$  with respect to criteria from  $P$  (denoted by  $(x, y)D_P(w, z)$ ), if for each  $g_i \in P$ ,  $g_i(x) \geq g_i(w)$  and  $g_i(z) \geq g_i(y)$ .

Let  $D_{\{i\}}$  be the dominance relation confined to the single criterion  $g_i \in P^O$ . The binary relation  $D_{\{i\}}$  is reflexive, transitive, but non-complete (it is possible that *not*  $(x, y)D_{\{i\}}(w, z)$  and *not*  $(w, z)D_{\{i\}}(x, y)$  for some  $(x, y), (w, z) \in A \times A$ ). Therefore,  $D_{\{i\}}$  is a partial preorder. Since the intersection of partial preorders is also a partial preorder and  $D_P = \bigcap_{g_i \in P} D_{\{i\}}$ ,  $P = P^O$ , then the dominance relation  $D_P$  is a partial preorder.

If some criteria from  $P \subseteq C$  express preferences on a quantitative or a numerical non-quantitative scale and others on an ordinal scale, *i.e.* if  $P^N \neq \emptyset$  and  $P^O \neq \emptyset$ , then, given  $(x, y), (w, z) \in A \times A$ , the pair  $(x, y)$  is said to dominate the pair  $(w, z)$  with respect to criteria from  $P$ , if  $(x, y)$  dominates  $(w, z)$  with respect to both  $P^N$  and  $P^O$ . Since the dominance relation with respect to  $P^N$  is a partial preorder on  $A \times A$  (because it is a multigraded dominance) and the dominance with respect to  $P^O$  is also a partial preorder on  $A \times A$  (as explained above), then the dominance  $D_P$ , being the intersection of these two dominance relations, is a partial preorder. In consequence, all the concepts introduced in the previous section can be restored using this specific definition of dominance.

### 5.3 Induction of decision rules from rough approximations of preference relations

Using the rough approximations of preference relations  $S$  and  $S^c$  defined in 5.2.1 and 5.2.2, it is possible to induce a generalized description of the preference information contained in a given  $S_{PCT}$  in terms of suitable decision rules. The syntax of these rules involves the concept of *upward cumulated preferences* (denoted by  $P_i^{\geq h}$ ) and *downward cumulated preferences* (denoted by  $P_i^{\leq h}$ ), having the following interpretation:

- $x P_i^{\geq h} y$  means “ $x$  is preferred to  $y$  with respect to  $g_i$  by at least degree  $h$ ”;
- $x P_i^{\leq h} y$  means “ $x$  is preferred to  $y$  with respect to  $g_i$  by at most degree  $h$ ”.

Exact definition of the cumulated preferences, for each  $(x, y) \in A \times A$ ,  $g_i \in C^N$  and  $h \in H_i$ , can be represented as follows:

- $x P_i^{\geq h} y$  if  $x P_i^k y$ , where  $k \in H_i$  and  $k \geq h$ ;
- $x P_i^{\leq h} y$  if  $x P_i^k y$ , where  $k \in H_i$  and  $k \leq h$ .

Let also  $G_i = \{g_i(x), x \in A\}$ ,  $g_i \in C^O$ . The decision rules have then the following syntax:

1)  $D_{\geq}$ -decision rules:

*If  $x P_{i1}^{\geq h(i1)} y$  and...  $x P_{ie}^{\geq h(ie)} y$  and  $g_{ie+1}(x) \geq r_{ie+1}$  and  $g_{ie+1}(y) \leq s_{ie+1}$  and...  $g_{ip}(x) \geq r_{ip}$  and  $g_{ip}(y) \leq s_{ip}$ , then  $x S y$ ,*

where  $P = \{g_{i1}, \dots, g_{ip}\} \subseteq C$ ,  $P^N = \{g_{i1}, \dots, g_{ie}\}$ ,  $P^O = \{g_{ie+1}, \dots, g_{ip}\}$ ,  $(h(i1), \dots, h(ie)) \in H_{i1} \times \dots \times H_{ie}$  and  $(r_{ie+1}, \dots, r_{ip}), (s_{ie+1}, \dots, s_{ip}) \in G_{ie+1} \times \dots \times G_{ip}$ . These rules are supported by pairs of objects from the  $P$ -lower approximation of  $S$  only;

2)  $D_{\leq}$ -decision rules:

*If  $x P_{i1}^{\leq h(i1)} y$  and...  $x P_{ie}^{\leq h(ie)} y$  and  $g_{ie+1}(x) \leq r_{ie+1}$  and  $g_{ie+1}(y) \geq s_{ie+1}$  and...  $g_{ip}(x) \leq r_{ip}$  and  $g_{ip}(y) \geq s_{ip}$ , then  $x S^c y$ ,*

where  $P = \{g_{i1}, \dots, g_{ip}\} \subseteq C$ ,  $P^N = \{g_{i1}, \dots, g_{ie}\}$ ,  $P^O = \{g_{ie+1}, \dots, g_{ip}\}$ ,  $(h(i1), \dots, h(ie)) \in H_{i1} \times \dots \times H_{ie}$  and  $(r_{ie+1}, \dots, r_{ip}), (s_{ie+1}, \dots, s_{ip}) \in G_{ie+1} \times \dots \times G_{ip}$ . These rules are supported by pairs of objects from the  $P$ -lower approximation of  $S^c$  only;

3)  $D_{\geq \leq}$ -decision rules:

*If  $x P_{i1}^{\geq h(i1)} y$  and...  $x P_{ie}^{\geq h(ie)} y$  and  $x P_{ie+1}^{\leq h(ie+1)} y \dots x P_{if}^{\leq h(if)} y$  and  $g_{if+1}(x) \geq r_{if+1}$  and  $g_{if+1}(y) \leq s_{if+1}$  and...  $g_{ig}(x) \geq r_{ig}$  and  $g_{ig}(y) \leq s_{ig}$  and  $g_{ig+1}(x) \leq r_{ig+1}$  and  $g_{ig+1}(y) \geq s_{ig+1}$  and...  $g_{ip}(x) \leq r_{ip}$  and  $g_{ip}(y) \geq s_{ip}$ , then  $x S y$  or  $x S^c y$ ,*

where  $O' = \{g_{i1}, \dots, g_{ie}\} \subseteq C$ ,  $O'' = \{g_{ie+1}, \dots, g_{if}\} \subseteq C$ ,  $P^N = O' \cup O''$ ,  $O'$  and  $O''$  are not necessarily disjoint,  $P^O = \{g_{if+1}, \dots, g_{ip}\}$ ,  $(h(i1), \dots, h(if)) \in H_{i1} \times \dots \times H_{if}$ ,  $(r_{if+1}, \dots, r_{ip}), (s_{if+1}, \dots, s_{ip}) \in G_{if+1} \times \dots \times G_{ip}$ . These rules are supported by pairs of objects from the  $P$ -boundary of  $S$  and  $S^c$  only.

**5.4 Use of decision rules for multicriteria choice and ranking**

The decision rules induced from a given  $S_{PCT}$  describe the comprehensive preference relations  $S$  and  $S^c$  either exactly ( $D_{\geq}$ - and  $D_{\leq}$ -decision rules) or approximately ( $D_{\geq\leq}$ -decision rules). A set of these rules covering all pairs of  $S_{PCT}$  represents a preference model of the decision maker who gave the pairwise comparison of reference objects. The application of these decision rules on a new subset  $M \subseteq A$  of objects induces a specific preference structure on  $M$ .

In fact, any pair of objects  $(u, v) \in M \times M$  can match the decision rules in one of four ways:

- At least one  $D_{\geq}$ -decision rule and neither  $D_{\leq}$ - nor  $D_{\geq\leq}$ -decision rules;
- At least one  $D_{\leq}$ -decision rule and neither  $D_{\geq}$ - nor  $D_{\geq\leq}$ -decision rules;
- At least one  $D_{\geq}$ -decision rule and at least one  $D_{\leq}$ -decision rule, or at least one  $D_{\geq\leq}$ -decision rule, or at least one  $D_{\geq\leq}$ -decision rule and at least one  $D_{\geq}$ - and/or at least one  $D_{\leq}$ -decision rule;
- No decision rule.

These four ways correspond to the following four situations of outranking, respectively:

- $uSv$  and *not*  $uS^c v$ , *i.e.* *true* outranking (denoted by  $uS^T v$ );
- $uS^c v$  and *not*  $uSv$ , *i.e.* *false* outranking (denoted by  $uS^F v$ );
- $uSv$  and  $uS^c v$ , *i.e.* *contradictory* outranking (denoted by  $uS^K v$ );
- *not*  $uSv$  and *not*  $uS^c v$ , *i.e.* *unknown* outranking (denoted by  $uS^U v$ ).

The four above situations, which together constitute the so-called *four-valued outranking* (Greco, Matarazzo, Slowinski & Tsoukias, 1998), have been introduced to underline the presence and absence of *positive* and *negative* reasons for the outranking. Moreover, they make it possible to distinguish contradictory situations from unknown ones.

A final *recommendation* (choice or ranking) can be obtained upon a suitable exploitation of this structure, *i.e.* of the presence and the absence of outranking  $S$  and  $S^c$  on  $M$ . A possible exploitation procedure consists of calculating a specific score, called the Net Flow Score, for each object  $x \in M$ :

$$S_{nf}(x) = S^{++}(x) - S^{+-}(x) + S^{-+}(x) - S^{-}(x),$$

where

$$S^{++}(x) = \text{card}(\{y \in M: \text{there is at least one decision rule which affirms } xSy\});$$

$$S^{+-}(x) = \text{card}(\{y \in M: \text{there is at least one decision rule which affirms } ySx\});$$

$$S^{-+}(x) = \text{card}(\{y \in M: \text{there is at least one decision rule which affirms } yS^c x\});$$

$$S^{-}(x) = \text{card}(\{y \in M: \text{there is at least one decision rule which affirms } xS^c y\}).$$

The recommendation in ranking problems consists of the total preorder determined by  $S_{nf}(x)$  on  $M$ . In choice problems, it consists of the object(s)  $x^* \in M$  such that  $S_{nf}(x^*) = \max_{x \in M} \{S_{nf}(x)\}$ .

The above procedure has been characterized with reference to a number of desirable properties in (Greco, Matarazzo, Slowinski & Tsoukias, 1998; Szelag, Greco & Slowinski, 2012).

Recently, Fortemps, Greco & Slowinski (2008) extended the Dominance-based Rough Set Approach to multicriteria choice and ranking on multi-graded preference relations, instead of uni-graded relations  $S$  and  $S^c$ .

It is also worth mentioning a machine learning approach to multicriteria choice and ranking using ensembles of decision rules. The approach presented by Dembczynski, Kotlowski, Slowinski & Szelag (2010) makes a bridge between stochastic methods of preference learning and DRSA for choice and ranking.

### 5.5 An illustrative example

Let us suppose that a company managing a chain of warehouses wants to buy some new warehouses. To choose the best proposals or to rank them all, the managers of the company decide to analyze first the characteristics of eight warehouses already owned by the company (reference objects). This analysis should give some indications for the choice and ranking of the new proposals. Eight warehouses belonging to the company have been evaluated by the following three criteria: capacity of the sales staff ( $A_1$ ), perceived quality of goods ( $A_2$ ) and high traffic location ( $A_3$ ). The domains (scales) of these attributes are presently composed of three preference-ordered echelons:  $V_1 = V_2 = V_3 = \{\text{sufficient, medium, good}\}$ . The decision attribute ( $d$ ) indicates the profitability of warehouses, expressed by the *Return On Equity (ROE)* ratio (in %). Table 3 presents a decision table which represents this situation.

**Table 3** – Decision table with reference objects.

Warehouse	$A_1$	$A_2$	$A_3$	$d$ (ROE%)
1	good	medium	good	10.35
2	good	sufficient	good	4.58
3	medium	medium	good	5.15
4	sufficient	medium	medium	-5
5	sufficient	medium	medium	2.42
6	sufficient	sufficient	good	2.98
7	good	medium	good	15
8	good	sufficient	good	-1.55

With respect to the set of criteria  $C = C^N = \{A_1, A_2, A_3\}$ , the following multigraded preference relations  $P_i^h, i = 1, 2, 3$ , are defined:

- $xP_i^0y$  (and  $yP_i^0x$ ), meaning that  $x$  is *indifferent* to  $y$  with respect to  $A_i$ , if  $f(x, A_i) = f(y, A_i)$ ;

- $xP_i^1y$  (and  $yP_i^{-1}x$ ), meaning that  $x$  is preferred to  $y$  with respect to  $A_i$ , if  $f(x, A_i) = \text{good}$  and  $f(y, A_i) = \text{medium}$ , or if  $f(x, A_i) = \text{medium}$  and  $f(y, A_i) = \text{sufficient}$ ;
- $xP_i^2y$  (and  $yP_i^{-2}x$ ), meaning that  $x$  is strongly preferred to  $y$  with respect to  $A_i$ , if  $f(x, A_i) = \text{good}$  and  $f(y, A_i) = \text{sufficient}$ .

Using the decision attribute, the comprehensive outranking relation was built as follows: warehouse  $x$  is at least as good as warehouse  $y$  with respect to profitability ( $xSy$ ) if

$$ROE(x) \geq ROE(y) - 2\%.$$

Otherwise, i.e. if  $ROE(x) < ROE(y) - 2\%$ , warehouse  $x$  is not at least as good as warehouse  $y$  with respect to profitability ( $xS^c y$ ).

The pairwise comparisons of the reference objects result in  $S_{PCT}$ . The rough set analysis of the  $S_{PCT}$  leads to the conclusion that the set of decision examples on the reference objects is inconsistent. The quality of approximation of  $S$  and  $S^c$  by all criteria from set  $C$  is equal to 0.44. Moreover,  $RED_{S_{PCT}} = CORE_{S_{PCT}} = \{A_1, A_2, A_3\}$ . This means that no criterion is superfluous.

The  $C$ -lower approximations and the  $C$ -upper approximations of  $S$  and  $S^c$ , obtained by means of multigraded dominance relations, are:

$$\begin{aligned} \underline{C}(S) &= \{(1,2), (1,4), (1,5), (1,6), (1,8), (3,2), (3,4), (3,5), (3,6), (3,8), (7,2), (7,4), (7,5), (7,6), (7,8)\} \\ \underline{C}(S^c) &= \{(2,1), (2,7), (4,1), (4,3), (4,7), (5,1), (5,3), (5,7), (6,1), (6,3), (6,7), (8,1), (8,7)\} \end{aligned}$$

All the remaining 36 pairs of reference objects belong to the  $C$ -boundaries of  $S$  and  $S^c$ , i.e.  $Bn_C(S) = Bn_C(S^c)$ .

The following minimal  $D_{\geq}$ -decision rules and  $D_{\leq}$ -decision rules can be induced from lower approximations of  $S$  and  $S^c$ , respectively (the figures within parentheses represent the pairs of objects supporting the corresponding rules):

- If  $xP_1^{\geq 1}y$  and  $xP_2^{\geq 1}y$ , then  $xSy$ ; ((1,6),(3,6),(7,6))
- If  $xP_2^{\geq 1}y$  and  $xP_3^{\geq 0}y$ , then  $xSy$ ; ((1,2),(1,6),(1,8),(3,2),(3,6),(3,8),(7,2),(7,6),(7,8))
- If  $xP_2^{\geq 0}y$  and  $xP_3^{\geq 1}y$ , then  $xSy$ ; ((1,4),(1,5),(3,4),(3,5),(7,4),(7,5))
- If  $xP_1^{\leq -1}y$  and  $xP_2^{\leq -1}y$ , then  $xS^c y$ ; ((6,1),(6,3),(6,7))
- If  $xP_2^{\leq 0}y$  and  $xP_3^{\leq -1}y$ , then  $xS^c y$ ; ((4,1),(4,3),(4,7),(5,1),(5,3),(5,7))
- If  $xP_1^{\leq 0}y$  and  $xP_2^{\leq -1}y$  and  $xP_3^{\leq 0}y$ , then  $xS^c y$ ; ((2,1),(2,7),(6,1),(6,3),(6,7),(8,1),(8,7))

Moreover, it is possible to induce five minimal  $D_{\geq \leq}$ -decision rules from the boundary of approximation of  $S$  and  $S^c$ :



- If  $xP_2^{\leq 0}y$  and  $xP_2^{\geq 0}y$  and  $xP_3^{\leq 0}y$  and  $xP_3^{\geq 0}y$ , then  $xSy$  or  $xS^c y$ :  $((1,1),(1,3),(1,7),(2,2),(2,6), (2,8),(3,1),(3,3),(3,7),(4,4),(4,5),(5,4),(5,5),(6,2),(6,6),(6,8),(7,1),(7,3),(7,7),(8,2),(8,6),(8,8))$
- If  $xP_2^{\leq -1}y$  and  $xP_3^{\geq 1}y$ , then  $xSy$  or  $xS^c y$ ;  $((2,4),(2,5),(6,4),(6,5),(8,4),(8,5))$
- If  $xP_2^{\geq 1}y$  and  $xP_3^{\leq -1}y$ , then  $xSy$  or  $xS^c y$ ;  $((4,2),(4,6),(4,8),(5,2),(5,6),(5,8))$
- If  $xP_1^{\geq 1}y$  and  $xP_2^{\leq 0}y$  and  $xP_3^{\leq 0}y$ , then  $xSy$  or  $xS^c y$ ;  $((1,3),(2,3),(2,6),(7,3),(8,3),(8,6))$
- If  $xP_1^{\geq 1}y$  and  $xP_2^{\leq -1}y$ , then  $xSy$  or  $xS^c y$ ;  $((2,3),(2,4),(2,5),(8,3),(8,4),(8,5))$

Using all the above decision rules and the Net Flow Score exploitation procedure on ten other warehouses proposed for purchase, the managers can obtain the result presented in Table 4. The dominance-based rough set approach gives a clear recommendation:

- For the *choice problem* it suggests the *selection of warehouse 2' and 6'*, having maximum score (11).
- For the *ranking problem* it suggests the *ranking* presented in the last column of Table 4, as follows:

$$(2', 6') \rightarrow (8') \rightarrow (9') \rightarrow (1') \rightarrow (4') \rightarrow (5') \rightarrow (3') \rightarrow (7', 10')$$

**Table 4** – Ranking of warehouses for sale by decision rules and the Net Flow Score procedure.

Warehouse for sale	$A_1$	$A_2$	$A_3$	Net Flow score	Ranking
1'	good	sufficient	medium	1	5
2'	sufficient	good	good	11	1
3'	sufficient	medium	sufficient	-8	8
4'	sufficient	good	sufficient	0	6
5'	sufficient	sufficient	medium	-4	7
6'	sufficient	good	good	11	1
7'	medium	sufficient	sufficient	-11	9
8'	medium	medium	medium	7	3
9'	medium	good	sufficient	4	4
10'	medium	sufficient	sufficient	-11	9

### 5.6 Summary

We briefly presented the contribution of the Dominance-based Rough Set Approach to multi-criteria choice and ranking problems. Let us point out the main features of the described methodology:

- The decision maker is asked for the preference information necessary to deal with a multicriteria decision problem in terms of exemplary decisions.
- The rough set analysis of preference information supplies some useful elements of knowledge about the decision situation. These are: the relevance of particular attributes and/or criteria, information about their interaction, minimal subsets of attributes or criteria (reducts) conveying important knowledge contained in the exemplary decisions and the set of the non-reducible attributes or criteria (core).
- The preference model induced from the preference information is expressed in a natural and comprehensible language of “if..., then...” decision rules. The decision rules concern pairs of objects and from them we can determine either the presence or the absence of a comprehensive preference relation. The conditions for the presence are expressed in “at least” terms, and for the absence in “at most” terms, on particular criteria.
- The decision rules do not convert ordinal information into numeric but keep the ordinal character of input data due to the syntax proposed.
- Heterogeneous information (qualitative and quantitative, ordered and non-ordered) and scales of preference (ordinal, cardinal) can be processed within the Dominance-based Rough Set Approach, while classical methods consider only quantitative ordered evaluations (with rare exceptions).
- No prior discretization of the quantitative domains of criteria is necessary.

## 6 RELEVANT EXTENSIONS OF DRSA

We introduced a knowledge discovery paradigm for multiattribute and multicriteria decision support, based on the concept of rough sets. Rough set theory provides mathematical tools for dealing with granularity of information and possible inconsistencies in the description of objects. Considering this description as an input data about a decision problem, the knowledge discovery paradigm consists of searching for rules in the data that facilitate an understanding of the decision maker’s preferences and that enable us to recommend a decision which is in line with these preferences. An original component of this paradigm is that it takes into account prior knowledge about preference semantics in the rules to be discovered.

Knowledge discovery from preference ordered data differs from usual knowledge discovery since the former involves preference orders in domains of attributes and in the set of decision classes. This requires that a knowledge discovery method applied to preference ordered data respects the dominance principle. As this is not the case for the well-known methods of data mining and knowledge discovery, they are not able to discover all relevant knowledge contained in the analyzed data sample and, even worse, they may yield unreasonable discoveries, because of inconsistency with the dominance principle. These deficiencies are addressed in the Dominance-based Rough Set Approach (DRSA). Moreover, this approach enables us to apply a rough set approach to multicriteria decision making. We showed how the approach could be used for multicriteria classification, choice and ranking. In more advanced papers, we have presented

many extensions of the approach that make it a useful tool for other practical applications. These extensions are:

- DRSA to decision under risk and uncertainty (Greco, Matarazzo & Slowinski, 2001e);
- DRSA to decision under uncertainty and time preference (Greco, Matarazzo & Slowinski, 2010c);
- DRSA handling missing data (Greco, Matarazzo & Slowinski, 1999c, 2000a);
- DRSA for imprecise object evaluations and assignments (Dembczynski, Greco & Slowinski, 2009);
- Dominance-based approach to induction of association rules (Greco, Matarazzo, Slowinski & Stefanowski, 2002);
- Fuzzy-rough hybridization of DRSA (Greco, Matarazzo & Slowinski, 1999b, 2000b, 2000c; Greco, Inuiguchi & Slowinski, 2002, 2003);
- DRSA as a way of operator-free fuzzy-rough hybridization (Greco, Inuiguchi & Slowinski, 2003, 2005, Greco, Matarazzo & Slowinski, 2007);
- DRSA to granular computing (Greco, Matarazzo & Slowinski, 2008a, 2009);
- DRSA to case-based reasoning (Greco, Matarazzo & Slowinski, 2008d);
- DRSA for hierarchical structure of evaluation criteria (Dembczynski, Greco, Slowinski, 2002);
- DRSA to decision involving multiple decision makers (Greco, Matarazzo & Slowinski, 2006, 2011);
- DRSA to interactive multiobjective optimization (Greco, Matarazzo & Slowinski, 2008c);
- DRSA to interactive evolutionary multiobjective optimization under risk and uncertainty (Greco, Matarazzo & Slowinski, 2010d).

The Dominance-based Rough Set Approach leads to a preference model of a decision maker in terms of decision rules. The decision rules have a special syntax which involves partial evaluation profiles and dominance relations on these profiles. The clarity of the rule representation of preferences enables us to see the limits of other traditional aggregation functions: the utility function and the outranking relation. In several studies (Greco, Matarazzo & Slowinski, 2001b, 2002c, 2003; Slowinski, Greco & Matarazzo, 2002b), we proposed an axiomatic characterization of these aggregation functions in terms of conjoint measurement theory and in terms of a set of decision rules. In comparison to other studies on the characterization of aggregation functions, our axioms do not require any preliminary assumptions about the scales of criteria. A side-result of these investigations is that the decision rule aggregation (preference model) is the most general among the known aggregation functions. The decision rule preference model fulfils, moreover, the postulate of transparency and interpretability of preference models in decision support.

Dealing with ordered data and monotonicity constraints makes also sense in general classification problems, where the notion of preference has no meaning. Even when the ordering seems irrelevant, the presence or the absence of a property have an ordinal interpretation. If two properties are related, one of the two: the presence or the absence of one property should make more (or less) probable the presence of the other property. A formal proof showing that the Indiscernibility-based Rough Set Approach is a particular case of the Dominance-based Rough Set Approach has been given in (Greco, Matarazzo & Slowinski, 2007). Having this in mind, DRSA can be seen as a general framework for analysis of classification data. Although it has been designed for ordinal classification problems with monotonicity constraints, DRSA can be used to solve a general classification problem where no additional information about ordering is taken into account.

The idea which stands behind this claim is the following (Błaszczynski, Greco & Slowinski, 2012). We assume, without loss of generality, that the value sets of all regular attributes are number-coded. While this is natural for numerical attributes, categorical attributes must get numerical codes for categories. In this way, the value sets of all regular attributes get ordered (as all sets of numbers are ordered). Now, to analyze a non-ordinal classification problem using DRSA, we transform decision table such that each regular attribute is cloned (doubled). It is assumed that the value set of each original attribute is ordered with respect to increasing preference (gain type), and the value set of its clone is ordered with respect to decreasing preference (cost type). Using DRSA, for each  $t \in \{1, \dots, n\}$ , we approximate two sets of objects from the decision table: class  $Cl_t$  and its complement  $\neg Cl_t$ . Obviously, we can calculate dominance-based rough approximations of the two sets. Moreover, they can serve to induce “if..., then...” decision rules recommending assignment to class  $Cl_t$  or to its complement  $\neg Cl_t$ . In this way, we reformulated the original non-ordinal classification problem to an ordinal classification problem with monotonicity constraints. Due to cloning of attributes with opposite preference orders, we can have rules that cover a subspace in the condition space, which is bounded from the top and from the bottom – this leads (without discretization) to more synthetic rules than those resulting from the Indiscernibility-based Rough Set Approach.

## 7 DRSA AND OPERATIONS RESEARCH PROBLEMS

DRSA is also a useful instrument in the toolbox of Operations Research (OR). DRSA has been applied to the following OR problems:

- 1) interactive multiobjective optimization (IMO-DRSA) (Greco, Matarazzo & Slowinski, 2008c);
- 2) interactive evolutionary multiobjective optimization under risk and uncertainty (Greco, Matarazzo & Slowinski, 2010d);
- 3) decision under uncertainty and time preference (Greco, Matarazzo & Slowinski, 2010c).

### 7.1 DRSA to interactive multiobjective optimization (IMO-DRSA)

DRSA to interactive multiobjective optimization (IMO-DRSA) (Greco, Matarazzo & Slowinski, 2008c) permits to deal with many optimization problems considered within OR (ranging from inventory management to scheduling, passing through portfolio management) in a way which is very much oriented towards interaction with the users. In fact, in IMO-DRSA a sample of representative solutions to a multiobjective optimization problem is presented to the DM who is asked to indicate a subset of relatively “good” solutions in the sample. Applying DRSA to the sample of representative solutions classified into “good” and “others” by the DM, a set of decision rules is induced in the form: “if objective  $f_{i1}(x) \geq \alpha_{i1}$  and...  $f_{ip}(x) \geq \alpha_{ip}$ , then  $x$  is a good solution”. The DM selects the rule that in his opinion is the most representative of his preference and the constraints coming from that rule are adjoined to the set of constraints imposed on the Pareto optimal set, in order to focus on a part interesting from the point of view of DM’s preferences in the next iteration. For example, if the DM selects the rule “if  $f_{i1}(x) \geq \alpha_{i1}$  and...  $f_{ip}(x) \geq \alpha_{ip}$ , then  $x$  is a good solution”, then the constraints  $f_{i1}(x) \geq \alpha_{i1}$  and...  $f_{ip}(x) \geq \alpha_{ip}$  are adjoined to the set of constraints of the multiobjective optimization problem, such that the new set of constraints implies a Pareto optimal set being the subset of the original Pareto optimal set. This subset satisfies the requirements of the selected rule, so that it is composed of solutions that are considered relatively good by the DM. The procedure continues iteratively until the DM is satisfied with one solution from the current sample – this is the most preferred solution.

The IMO-DRSA procedure can be analyzed from the point of view of input and output information. As to the input, the DM gives preference information by answering easy questions related to ordinal classification of some representative solutions into two classes (“good” and “others”). Very often, in multiple criteria decision analysis in general, and in interactive multiobjective optimization in particular, the preference information has to be given in terms of preference model parameters, such as importance weights, substitution rates and various thresholds. This information is specified in (Fishburn, 1967) with respect to Multiple Attribute Utility Theory, and in (Roy & Bouyssou, 1993; Figueira, Mousseau & Roy, 2005; Brans & Mareschal, 2005; Martel & Matarazzo, 2005) with respect to outranking methods. In case of multiobjective optimization, the preference information depends on the method; *e.g.*, the Geoffrion-Dyer-Feinberg method (Geoffrion-Dyer-Feinberg, 1972), the method of Zionts & Wallenius (Zionts & Wallenius, 1976, 1983) and the Interactive Surrogate Worth Tradeoff method (Chankong & Haimes, 1978, 1983) require information in terms of marginal rates of substitution; the reference point method (Wierzbicki, 1980) requires a reference point and weights to formulate an achievement scalarizing function; the Light Beam Search method (Jaszkiewicz & Slowinski, 1999) requires information in terms of weights and indifference, preference and veto thresholds, being typical parameters of ELECTRE methods. Eliciting such information requires a significant cognitive effort on the part of the DM. It is generally acknowledged that people rather prefer to make exemplary decisions than to explain them in terms of specific parameters. For this reason, the idea of inferring preference models from exemplary decisions provided by the DM is very attractive.

The output result of the analysis is the model of preferences in terms of “*if... then...*” decision rules which is used to reduce the Pareto optimal set iteratively, until the DM selects a satisfactory solution. The decision rule preference model is very convenient for decision support, because it gives argumentation for preferences in a logical form, which is intelligible for the DM, and identifies the Pareto optimal solutions supporting each particular decision rule. This is very useful for a critical revision of the original ordinal classification of representative solutions into the two classes of “good” and “others”. Indeed, decision rule preference model speaks the same language of the DM without any recourse to technical terms, like utility, tradeoffs, scalarizing functions and so on.

All this implies that IMO-DRSA has a transparent feedback organized in a learning oriented perspective, which permits to consider this procedure as a “glass box”, contrary to the “black box” characteristic of many procedures giving final result without any clear explanation. The information given by the decision rules is particularly intelligible for the DM, since they speak the language of the DM and permit him/her to identify the Pareto optimal solutions supporting each decision rule. Thus, decision rules give an explanation and a justification of the final decision, that does not result from a mechanical application of a certain technical method, but rather from a mature conclusion of a decision process based on active intervention of the DM.

Observe, finally, that the decision rules representing preferences are based on ordinal properties of objective functions only. Differently from methods involving some scalarization (almost all existing interactive methods), in any step the proposed procedure does not aggregate the objectives into a single value, avoiding operations (such as averaging, weighted sum, different types of distance, achievement scalarization) which are always arbitrary to some extent. Observe that one could use a method based on scalarization to generate the representative set of Pareto optimal solutions, nevertheless, the decision rule approach would continue to be based on ordinal properties of objective functions only, because the dialogue stage of the method operates on ordinal comparisons only. In the proposed method, the DM gets clear arguments for his/her decision in terms of “*if... then...*” decision rules and the verification if a proposed solution satisfies these decision rules is particularly easy. This is not the case of interactive multiobjective optimization methods based on scalarization. For example, in the methods using an achievement scalarizing function, it is not evident what does it mean for a solution to be “close” to the reference point. How to justify the choice of the weights used in the achievement function? What is their interpretation? Observe, instead, that IMO-DRSA operates on data using ordinal comparisons which would not be affected by any increasing monotonic transformation of scales, and this ensures the meaningfulness of results from the point of view of measurement theory (see, *e.g.*, Roberts, 1979).

With respect to computational aspects of the method, notice that the decision rules can be calculated efficiently in few seconds only using the algorithms presented in (Greco, Matarazzo, Slowinski & Stefanowski, 2001, 2002; Blaszczyński, Slowinski & Szlag, 2011). When the number of objective functions is not too large to be effectively controlled by the DM (say, seven plus or minus two, as suggested by Miller (1956)), then the decision rules can be calculated in

a fraction of one second. In any case, the computational effort grows exponentially with the number of objective functions, but not with respect to the number of considered Pareto optimal solutions, which can increase with no particularly negative consequence on calculation time.

## 7.2 DRSA to interactive evolutionary multiobjective optimization

Very often real life optimization problems are so complex that exact methods fail to find an optimal solution. In these cases some heuristics are to be applied. Within multiobjective optimization, Evolutionary Multiobjective Optimization (EMO) appeared to be particularly efficient; see, *e.g.*, (Coello Coello, Van Veldhuizen & Lamont, 2002; Deb, 2001).

The underlying reasoning behind the EMO search of an approximation of the Pareto-optimal frontier is that, in the absence of any preference information, all Pareto-optimal solutions have to be considered equivalent. On the other hand, if the DM (alternatively called user) is involved in the multiobjective optimization process, then the preference information provided by the DM can be used to focus the search on the most preferred part of the Pareto-optimal frontier. This idea stands behind Interactive Multiobjective Optimization (IMO) methods proposed long time before EMO has emerged.

Recently, it became clear that merging the IMO and EMO methodologies should be beneficial for the multiobjective optimization process (Branke, Deb, Miettinen & Slowinski, 2008). Several approaches have been presented in this context; see, *e.g.*, (Fonseca & Fleming, 1993; Deb, Sundar, Rao & Chaudhuri, 2006; Deb & Chaudhuri, 2010; Coello, Van Veldhuizen & Lamont, 2002; Branke, Kaußler & Schmeck, 2001; Greenwood, Hu & D'Ambrosio, 1997; Jaskiewicz, 2007; Phelps & Koksalan, 2003; Branke, Greco, Slowinski & Zielniewicz, 2009).

The methodology of interactive EMO based on DRSA (Greco, Matarazzo & Slowinski, 2010d, 2010e) involves application of decision rules in EMO, which are induced from easily elicited preference information by DRSA, proposing two general schemes, called DRSA-EMO and DRSA-EMO-PCT. This results in focusing the search of the Pareto-optimal frontier on the most preferred region. More specifically, DRSA is used for structuring preference information obtained through interaction with the user, and then a set of decision rules representing user's preferences is induced from this information. These rules are used to rank solutions in the current population of EMO, which has an impact on the selection and crossover.

Within interactive EMO, one can also apply DRSA for decision under uncertainty. This permits to take into account robustness concerns in the multiobjective optimization. In fact, two methods of robust optimization methods combining DRSA and interactive EMO have been proposed: DARWIN (Dominance-based rough set Approach to handling Robust Winning solutions in IN teractive multiobjective optimization) and DARWIN-PCT (DARWIN using Pairwise Comparison Tables). DARWIN and DARWIN-PCT can be considered as two specific instances of DRSA-EMO and DRSA-EMO-PCT, respectively.

The integration of DRSA and EMO is particularly promising for two reasons:

1. The preference information required by DRSA is very basic and easy to be elicited by the DM. All that the DM is asked for is to assign solutions to preference ordered classes, such as “good”, “medium” and “bad”, or compare pairs of non-dominated solutions from a current population in order to reveal whether one is preferred over the other. The preference information is provided every  $k$  iterations ( $k$  depends on the problem and the willingness of the user to interact with the system. In our studies,  $k$  ranges from 10 to 30).
2. The decision rules are transparent and easy to interpret for the DM. As explained in the previous subsection, the preference model supplied by decision rules is a “glass box”, while many other competitive multiple criteria decision methodologies involve preference models that are “black boxes” for the user. The “glass box” model improves the quality of the interaction and makes that the DM accepts well the resulting recommendation.

The integration of DRSA and EMO provides very general interactive EMO schemes which can be customized to a large variety of OR problems, from location and routing to scheduling and supply chain management.

### 7.3 DRSA to decision under uncertainty and time preference

DRSA can also be applied to preference modeling for decision under uncertainty with consequences distributed over time, using the idea of time-stochastic dominance, *i.e.* putting together the concept of time dominance and stochastic dominance (Greco, Matarazzo & Slowinski, 2010c). Preference information provided by the DM is a set of decision examples specifying the quality of some chosen acts, *i.e.* assigning these acts to preference-ordered classes. The resulting preference model expressed in terms of “*if... then...*” decision rules is much more intelligible than any utility function. Moreover, it permits to handle inconsistent preference information. Let us observe that the approach handles an additive probability distribution as well as a non-additive probability, and even a qualitative ordinal probability. Furthermore, in case the elements of sets of possible probability values and of time epochs were very numerous (like in real life applications in which very often they are infinite), it would be enough to consider a subset of the most significant probability values (*e.g.*, 0, 0.1, 0.2, ..., 0.9, 1) and a subset of the most significant epochs (*e.g.*, each month).

Applying DRSA to decision under uncertainty and time preference we get decision rules of the type:

*“if the cumulated outcome at  $t_1$  is at least  $-50$ , and the cumulated outcome at  $t_2$  is at least  $300$ , then act  $a_i$  is (at least) good”*

or

*“if the cumulated outcome at  $t_1$  is at most  $-100$  and the cumulated outcome at  $t_2$  is at most  $150$ , then act  $a_i$  is (at least) good”.*

This method can be extended on the case of pairwise comparisons (Greco, Matarazzo & Slowinski, 2010e) obtaining rules whose syntax is:



*“if the difference between the cumulated outcome of act a and act b is not smaller than 50 at  $t_1$ , and not smaller than 300 at  $t_2$ , then act a is at least weakly preferred to act b”.*

The above methodology can be very useful for dealing with many OR problems where uncertainty of outcomes and their distribution over the time play a fundamental role, such as portfolio selection, scheduling with time-resource interactions and inventory management. Indeed, putting together the decision rules produced by this methodology with IMO-DRSA and DRSA applied to EMO, provides an important tool for dealing with even more OR problems. An example of a recent application of this methodology to a typical OR problem, which is inventory control, can be found in (Greco, Matarazzo, Slowinski & Vaccarella, 2012).

## **8 SOURCES OF ADDITIONAL INFORMATION ABOUT ROUGH SET THEORY AND APPLICATIONS**

The community of researchers and practitioners interested in rough set theory and applications is organized in the International Rough Set Society (<http://roughsets.home.pl/www/>). The society's web page includes information about rough set conferences, about Transactions on Rough Sets published in a journal series of LNCS by Springer, and about International Journal of Granular Computing, Rough Sets and Intelligent Systems. This page also includes slides of tutorial presentations on rough sets. A database of rough set references can be found at <http://rsds.univ.rzeszow.pl>.

The following software is available free in the Internet:

RSES – Rough Set Exploration System <http://logic.mimuw.edu.pl/~rses>,

ROSE – ROugh Set data Explorer <http://idss.cs.put.poznan.pl/site/rose.html>,

jMAF – java Multi-criteria and Multi-attribute Analysis Framework  
<http://www.cs.put.poznan.pl/jblaszczyński/Site/jRS.html>, and

jRank – ranking generator using Dominance-based Rough Set Approach  
<http://www.cs.put.poznan.pl/mszelag/Software/jRank/jRank.html>.

## **ACKNOWLEDGEMENTS**

The first author wishes to acknowledge financial support from the Polish National Science Centre, grant no. NN519 441939.

## **REFERENCES**

- [1] BLASZCZYŃSKI J, GRECO S & SLOWINSKI R. 2007. Multi-criteria classification – a new scheme for application of dominance-based decision rules. *European J. Operational Research*, **181**(3): 1030–1044.

- [2] BLASZCZYNSKI J, GRECO S & SLOWINSKI R. 2012. Inductive discovery of laws using monotonic rules. *Engineering Applications of Artificial Intelligence*, **25**(2): 284–294.
- [3] BLASZCZYNSKI J, GRECO S, SLOWINSKI R & SZELAG M. 2009. Monotonic variable consistency rough set approaches. *International Journal of Approximate Reasoning*, **50**(7): 979–999.
- [4] BLASZCZYNSKI J, SLOWINSKI R & SZELAG M. 2011. Sequential covering rule induction algorithm for variable consistency rough set approaches. *Information Sciences*, **181**: 987–1002.
- [5] BRANKE J, DEB K, MIETTINEN K & SLOWINSKI R. (Eds.). 2008. *Multiobjective Optimization: Interactive and Evolutionary Approaches*, LNCS 5252, Springer, Berlin.
- [6] BRANKE J, KAUSSLER T & SCHMECK H. 2001. Guidance in evolutionary multi-objective optimization. *Advances in Engineering Software*, **32**: 499–507.
- [7] BRANKE J, GRECO S, SLOWINSKI R & ZIELNIEWICZ P. 2009. Interactive evolutionary multi-objective optimization using robust ordinal regression. In: EHRGOTT M, FONSECA CM, GANDBLEUX X, HAO J-K & SEVAUX M. (Eds.), *Evolutionary Multi-Criterion Optimization (EMO'09)*, LNCS 5467, Springer, Berlin, 554–568.
- [8] BRANS JP & MARESCHAL B. 2005. PROMETHEE Methods. Chapter 5. In: FIGUEIRA J, GRECO S & EHRGOTT M. (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, Berlin, 163–195.
- [9] CHANKONG V & HAIMES YY. 1978. The interactive surrogate worth trade-off (ISWT) method for multiobjective decision-making. In: ZIONTS S. (Ed.), *Multiple Criteria Problem Solving*, Springer-Verlag, Berlin, New York, pp. 42–67.
- [10] CHANKONG V & HAIMES YY. 1983. *Multiobjective Decision Making Theory and Methodology*, Elsevier Science Publishing Co., New York.
- [11] COELLO COELLO CA, VAN VELDTHUIZEN DA & LAMONT GB. 2002. *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer, Dordrecht.
- [12] DEB K. 2001. *Multi-Objective Optimization using Evolutionary Algorithms*, Wiley, Chichester, UK.
- [13] DEB K & CHAUDHURI S. 2010. I-MODE: An interactive multi-objective optimization and decision-making using evolutionary methods. *Applied Soft Computing*, **10**: 496–511.
- [14] DEB K, SUNDAR J, RAO NUB & CHAUDHURI S. 2006. Reference point based multi-objective optimization using evolutionary algorithms. *International Journal of Computational Intelligence Research*, **2**(3): 273–286.
- [15] DEMBCZYNSKI K, GRECO S & SLOWINSKI R. 2002. Methodology of rough-set-based classification and sorting with hierarchical structure of attributes and criteria. *Control & Cybernetics*, **31**: 891–920.
- [16] DEMBCZYNSKI K, GRECO S & SLOWINSKI R. 2009. Rough set approach to multiple criteria classification with imprecise evaluations and assignments. *European J. Operational Research*, **198**(2): 626–636.
- [17] DEMBCZYNSKI K, KOTLOWSKI W, SLOWINSKI R & M SZELAG. 2010. Learning of rule Ensembles for multiple attribute ranking problems. In: FÜRNRANZ J & HÜLLERMEIER E. (Eds.), *Preference Learning*, Springer, Berlin, pp. 217–247.

- [18] FIGUEIRA J, MOUSSEAU V & ROY B. 2005. ELECTRE Methods. Chapter 4. In: FIGUEIRA J, GRECO S & EHRGOTT M. (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, Berlin, pp. 133–162.
- [19] FIGUEIRA J, GRECO S & EHRGOTT M. (Eds.). 2005. *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer, Berlin.
- [20] FISHBURN PC. 1967. Methods of estimating additive utilities. *Management Science*, **13**(7): 435–453.
- [21] FONSECA CM & FLEMING PJ. 1993. Genetic algorithms for multiobjective optimization: Formulation, discussion, and generalization. In: *Proceedings of the Fifth International Conference on Genetic Algorithms*, pp. 416–423.
- [22] FORTEMPS PH, GRECO S & SLOWINSKI R. 2008. Multicriteria decision support using rules that represent rough-graded preference relations. *European J. Operational Research*, **188**(1): 206–223.
- [23] GEOFFRION A, DYER J & FEINBERG A. 1972. An interactive approach for multi-criterion optimization, with an application to the operation of an academic department. *Management Science*, **19**(4): 357–368.
- [24] GIOVE S, GRECO S, MATARAZZO B & SLOWINSKI R. 2002. Variable consistency monotonic decision trees. In: ALPIGINI JJ, PETERS JF, SKOWRON A & ZHONG N. (Eds.), *Rough Sets and Current Trends in Computing*. Springer LNAI 2475, Berlin, pp. 247–254.
- [25] GRECO S, INUIGUCHI M & SLOWINSKI R. 2002. Dominance-based rough set approach using possibility and necessity measures. In: ALPIGINI JJ, PETERS JF, SKOWRON A & ZHONG N. (Eds.), *Rough Sets and Current Trends in Computing*. Springer LNAI 2475, Berlin, pp. 85–92.
- [26] GRECO S, INUIGUCHI M & SLOWINSKI R. 2003. A new proposal for fuzzy rough approximations and gradual decision rule representation. *Transactions on Rough Sets II*. Springer LNCS 3135, Berlin, pp. 319–342.
- [27] GRECO S, INUIGUCHI M & SLOWINSKI R. 2005. Fuzzy rough sets and multiple-premise gradual decision rules. *International Journal of Approximate Reasoning*, **41**: 179–211.
- [28] GRECO S, MATARAZZO B, PAPPALARDO N & SLOWINSKI R. 2005. Measuring expected effects of interventions based on decision rules. *Journal of Experimental and Theoretical Artificial Intelligence*, **17**(1-2): 103–118.
- [29] GRECO S, MATARAZZO B & SLOWINSKI R. 1998a. A new rough set approach to evaluation of bankruptcy risk. In: ZOPOUNIDIS C. (Ed.), *Operational Tools in the Management of Financial Risk*. Kluwer, Dordrecht, pp. 121–136.
- [30] GRECO S, MATARAZZO B & SLOWINSKI R. 1998b. Fuzzy similarity relation as a basis for rough approximation. In: POLKOWSKI L & SKOWRON A. (Eds.), *Rough sets and Current Trends in Computing*. Springer LNAI 1424, Berlin, pp. 283–289.
- [31] GRECO S, MATARAZZO B & SLOWINSKI R. 1999a. Rough approximation of a preference relation by dominance relations. *European J. Operational Research*, **117**: 63–83.
- [32] GRECO S, MATARAZZO B & SLOWINSKI R. 1999b. The use of rough sets and fuzzy sets in MCDM. Chapter 14. In: GAL T, STEWART T & HANNE T. (Eds.), *Advances in Multiple Criteria Decision Making*. Kluwer, Dordrecht, pp. 14.1–14.59.

- [33] GRECO S, MATARAZZO B & SLOWINSKI R. 1999c. Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems. In: ZHONG N, SKOWRON A & OHSUGA S. (Eds.), *New Directions in Rough Sets, Data Mining and Granular-Soft Computing*. Springer LNAI 1711, 146–157.
- [34] GRECO S, MATARAZZO B & SLOWINSKI R. 2000a. Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems. In: ZANAKIS SH, DOUKIDIS G & ZOPOUNIDIS C. (Eds.), *Decision Making: Recent Developments and Worldwide Applications*. Kluwer, Dordrecht, pp. 295–316.
- [35] GRECO S, MATARAZZO B & SLOWINSKI R. 2000b. Rough set processing of vague information using fuzzy similarity relations. In: CALUDE CS & PAUN G. (Eds.), *Finite Versus Infinite – Contributions to an Eternal Dilemma*. Springer, Berlin, pp. 149–173.
- [36] GRECO S, MATARAZZO B & SLOWINSKI R. 2000c. Fuzzy extension of the rough set approach to multicriteria and multiattribute sorting. In: FODOR J, DE BAETS B & PERNY P. (Eds.), *Preferences and Decisions under Incomplete Knowledge*. Physica-Verlag, Heidelberg, pp. 131–151.
- [37] GRECO S, MATARAZZO B & SLOWINSKI R. 2000d. Extension of the rough set approach to multicriteria decision support. *INFOR*, **38**: 161–196.
- [38] GRECO S, MATARAZZO B & SLOWINSKI R. 2001a. Rough sets theory for multicriteria decision analysis. *European J. of Operational Research*, **129**: 1–47.
- [39] GRECO S, MATARAZZO B & SLOWINSKI R. 2001b. Conjoint measurement and rough set approach for multicriteria sorting problems in presence of ordinal criteria. In: COLORNI A, PARUCCINI M & ROY B. (Eds.), *A-MCD-A: Aide Multi-Critère à la Décision – Multiple Criteria Decision Aiding*. European Commission Report, EUR 19808 EN, Ispra, 2001b, pp. 117–144.
- [40] GRECO S, MATARAZZO B & SLOWINSKI R. 2001c. Rule-based decision support in multicriteria choice & ranking. In: BENFERHAT S & BESNARD P. (Eds.), *Symbolic & Quantitative Approaches to Reasoning with Uncertainty*. Springer LNAI 2143, Berlin, pp. 29–47.
- [41] GRECO S, MATARAZZO B & SLOWINSKI R. 2001d. Assessment of a value of information using rough sets and fuzzy measures. In: CHOCJAN J & LESKI J. (Eds.), *Fuzzy Sets and their Applications*. Silesian University of Technology Press, pp. 185–193.
- [42] GRECO S, MATARAZZO B & SLOWINSKI R. 2001e. Rough set approach to decisions under risk. In: ZIARKO W & YAO Y. (Eds.), *Rough Sets and Current Trends in Computing*. Springer LNAI 2005, Berlin, pp. 160–169.
- [43] GRECO S, MATARAZZO B & SLOWINSKI R. 2002a. Rough sets methodology for sorting problems in presence of multiple attributes and criteria. *European J. of Operational Research*, **138**: 247–259.
- [44] GRECO S, MATARAZZO B & SLOWINSKI R. 2002b. Multicriteria classification. In: KLOESGEN W & ZYTKOW J. (Eds.), *Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, chapter 16.1.9, pp. 318–328.
- [45] GRECO S, MATARAZZO B & SLOWINSKI R. 2002c. Preference representation by means of conjoint measurement & decision rule model. In: BOUYSSOU D, JACQUET-LAGRÈZE E, PERNY P, SLOWINSKI R, VANDERPOOTEN D & VINCKE P. (Eds.), *Aiding Decisions with Multiple Criteria-Essays in Honor of Bernard Roy*. Kluwer, Dordrecht, pp. 263–313.

- [46] GRECO S, MATARAZZO B & SLOWINSKI R. 2004. Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *European J. Operational Research*, **158**: 271–292.
- [47] GRECO S, MATARAZZO B & SLOWINSKI R. 2006. Dominance-based rough set approach to decision involving multiple decision makers. In: GRECO S, HATA Y, HIRANO S, INUIGUCHI M, MIYAMOTO S, NGUYEN HS & SLOWINSKI R. (Eds.), *Rough Sets and Current Trends in Computing (RSCTC 2006)*. Springer LNCS 4259, Berlin, pp. 306–317.
- [48] GRECO S, MATARAZZO B & SLOWINSKI R. 2007. Dominance-based Rough Set Approach as a proper way of handling graduality in rough set theory. *Transactions on Rough Sets VII*, Springer LNCS 4400, Berlin, pp. 36–52.
- [49] GRECO S, MATARAZZO B & SLOWINSKI R. 2008a. Granular computing for reasoning about ordered data: the dominance-based rough set approach. Chapter 15. In: PEDRYCZ W, SKOWRON A & KREINOVICH V. (Eds.), *Handbook of Granular Computing*. Wiley, Chichester, pp. 347–373.
- [50] GRECO S, MATARAZZO B & SLOWINSKI R. 2008b. Parameterized rough set model using rough membership and Bayesian confirmation measures. *International Journal of Approximate Reasoning*, **49**: 285–300.
- [51] GRECO S, MATARAZZO B & SLOWINSKI R. 2008c. Dominance-based Rough Set Approach to Interactive Multiobjective Optimization. Chapter 5. In: BRANKE J, DEB K, MIETTINEN K & SLOWINSKI R. (Eds.), *Multiobjective Optimization: Interactive and Evolutionary Approaches*. Springer LNCS 5252, Berlin, pp. 121–156.
- [52] GRECO S, MATARAZZO B & SLOWINSKI R. 2008d. Case-based reasoning using gradual rules induced from dominance-based rough approximations. In: WANG G, LI T, GRZYMALA-BUSSE JW, MIAO D, SKOWRON A & YAO Y. (Eds.), *Rough Sets and Knowledge Technology (RSKT 2008)*. Springer LNAI 5009, Berlin, pp. 268–275.
- [53] GRECO S, MATARAZZO B & SLOWINSKI R. 2009. Granular Computing and Data Mining for Ordered Data – the Dominance-based RoughSet Approach. In: MEYERS RA. (Ed.), *Encyclopedia of Complexity and Systems Science*, Springer, New York, pp. 4283–4305.
- [54] GRECO S, MATARAZZO B & SLOWINSKI R. 2010a. Algebra and Topology for Dominance-based Rough Set Approach. In: RAS ZW & TSAY L-S. (Eds.), *Advances in Intelligent Information Systems, Studies in Computational Intelligence*, vol. 265, Springer, Berlin, pp. 43–78.
- [55] GRECO S, MATARAZZO B & SLOWINSKI R. 2010b. On Topological Dominance-based Rough Set Approach. *Transactions on Rough Sets XII*, Springer LNCS 6190, Berlin, pp. 21–45.
- [56] GRECO S, MATARAZZO B & SLOWINSKI R. 2010c. Dominance-based rough set approach to decision under uncertainty and time preference. *Annals of Operations Research*, **176**: 41–75.
- [57] GRECO S, MATARAZZO B & SLOWINSKI R. 2010d. Dominance-based Rough Set Approach to Interactive Evolutionary Multiobjective Optimization. In: GRECO S, MARQUES PEREIRA RA, SQUILLANTE M, YAGER RR & KACPRZYK J. (Eds.), *Preferences and Decisions: Models and Applications*, Springer, Studies in Fuzziness 257, Berlin, pp. 225–260.
- [58] GRECO S, MATARAZZO B & SLOWINSKI R. 2010e. Interactive Evolutionary Multiobjective Optimization using Dominance-based Rough Set Approach. In: *PROC. OF WCCI 2010, IEEE WORLD CONGRESS ON COMPUTATIONAL INTELLIGENCE*, July 18–23, 2010, Barcelona, Spain, pp. 3026–3033.

- [59] GRECO S, MATARAZZO B & SLOWINSKI R. 2010g. Dominance-based Rough Set Approach to Granular Computing. In: YAO J. (Ed.), *Novel Developments in Granular Computing*, Hershey, New York, pp. 439–527.
- [60] GRECO S, MATARAZZO B & SLOWINSKI R. 2012. The Bipolar Complemented de Morgan Brouwer-Zadeh Distributive Lattice as an Algebraic Structure for the Dominance-based Rough Set Approach. *Fundamenta Informaticae*, **115**: 25–56.
- [61] GRECO S, MATARAZZO B, SLOWINSKI R & STEFANOWSKI J. 2001a. Variable consistency model of dominance-based rough set approach. In: ZIARKO W & YAO Y. (Eds.), *Rough Sets and Current Trends in Computing*. Springer LNAI 2005, Berlin, pp. 170–181.
- [62] GRECO S, MATARAZZO B, SLOWINSKI R & STEFANOWSKI J. 2001b. An algorithm for induction of decision rules consistent with dominance principle. In: ZIARKO W & YAO Y. (Eds.), *Rough Sets & Current Trends in Computing*. Springer LNAI 2005, Berlin, pp. 304–313.
- [63] GRECO S, MATARAZZO B, SLOWINSKI R & STEFANOWSKI J. 2002. Mining association rules in preference-ordered data. In: HACID M-S, RAS ZW, ZIGHED DA & KODRATOFF Y. (Eds.), *Foundations of Intelligent Systems*. Springer LNAI 2366, Berlin, pp. 442–450.
- [64] GRECO S, MATARAZZO B, SLOWINSKI R & TSOUKIAS A. 1998. Exploitation of a rough approximation of the outranking relation in multicriteria choice and ranking. In: STEWART TJ & VAN DEN HONERT RC. (Eds.), *Trends in Multicriteria Decision Making*. Springer LNEMS 465, Berlin, pp. 45–60.
- [65] GRECO S, MATARAZZO B, SLOWINSKI R & VACCARELLA G. 2012. Inventory Control using Interactive Multiobjective Optimization guided by Dominance-based Rough Set Approach, submitted for publication.
- [66] GRECO S, PAWLAK Z & SLOWINSKI R. 2004. Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence*, **17**(4): 345–361.
- [67] GREENWOOD GW, HU XS & D’AMBROSIO JG. 1997. Fitness functions for multiple objective optimization problems: combining preferences with Pareto rankings. In: BELEW RK & VOSE MD. (Eds.), *Foundations of Genetic Algorithms*, Morgan Kaufmann, pp. 437–455.
- [68] GOLDSTEIN WM. 1991. Decomposable threshold models. *Journal of Mathematical Psychology*, **35**: 64–79.
- [69] GRZYMALA-BUSSE JW. 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, Dordrecht, pp. 3–18.
- [70] GRZYMALA-BUSSE JW. 1997. A new version of the rule induction system LERS. *Fundamenta Informaticae*, **31**: 27–39.
- [71] JASZKIEWICZ A & SLOWINSKI R. 1999. The “Light Beam Search” approach – an overview of methodology and applications. *European Journal of Operational Research*, **113**: 300–314.
- [72] JASZKIEWICZ A. 2007. Interactive multiobjective optimization with the Pareto memetic algorithm. *Foundations of Computing and Decision Sciences*, **32**(1): 15–32.
- [73] KOTLOWSKI W, DEMBCZYNSKI K, GRECO S & SLOWINSKI R. 2008. Stochastic dominance-based rough set model for ordinal classification. *Information Sciences*, **178**(21): 4019–4037.

- [74] KRAWIEC K, SLOWINSKI R & VANDERPOOTEN D. 1998. Learning of decision rules from similarity based rough approximations. In: POLKOWSKI L & SKOWRON A. (Eds.), *Rough Sets in Knowledge Discovery*. Vol.2, Physica-Verlag, pp. 37–54.
- [75] LUCE RD. 1956. Semi-orders and a theory of utility discrimination. *Econometrica*, **24**: 178–191.
- [76] MARCUS S. 1994. Tolerance rough sets, Cech topologies, learning processes. *Bull. of the Polish Academy of Sciences, ser. Technical Sciences*, **42**(3): 471–487.
- [77] MARTEL JM & MATARAZZO B. 2005. Other outranking approaches. Chapter 6. In: FIGUEIRA J, GRECO S & EHRGOTT M. (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, Berlin, pp. 197–262.
- [78] MILLER GA. 1956. The magical number seven, plus or minus two: some limits in our capacity for processing information. *The Psychological Review*, **63**: 81–97.
- [79] NIEMINEN J. 1988. Rough tolerance equality. *Fundamenta Informaticae*, **11**: 289–296.
- [80] PAWLAK Z. 1982. Rough sets. *Int. J. Information & Computer Sciences*, **11**: 341–356.
- [81] PAWLAK Z. 1991. *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer.
- [82] PAWLAK Z, GRZYMALA-BUSSE JW, SLOWINSKI R & ZIARKO W. 1995. Rough sets. *Communications of the ACM*, **38**: 89–95.
- [83] PAWLAK Z & SLOWINSKI R. 1994. Rough set approach to multi-attribute decision analysis. *European J. of Operational Research*, **72**: 443–459.
- [84] POLKOWSKI L. 2002. *Rough Sets: Mathematical Foundations*. Physica-Verlag.
- [85] POLKOWSKI L & SKOWRON A. 1999. Calculi of granules based on rough set theory: approximate distributed synthesis and granular semantics for computing with words. In: ZHONG N, SKOWRON A & OHSUGA S. (Eds.), *New Directions in Rough sets, Data Mining and Soft-Granular Computing*. Springer LNAI 1711, pp. 20–28.
- [86] POLKOWSKI L, SKOWRON A & ZYTKOW J. 1995. Rough foundations for rough sets. In: LIN TY & WILDBERGER A. (Eds.), *Soft Computing. Simulation Councils, Inc., San Diego, CA*, pp. 142–149.
- [87] PHELPS S & KOKSALAN M. 2003. An interactive evolutionary metaheuristic for multiobjective combinatorial optimization. *Management Science*, **49**(12): 1726–1738.
- [88] ROBERTS F. 1979. *Measurement Theory, with Applications to Decision Making, Utility and the Social Sciences*, Addison-Wesley, Boston.
- [89] ROY B. 1996. *Multicriteria Methodology for Decision Aiding*. Kluwer Academic Publishers, Dordrecht.
- [90] ROY B & BOUYSSOU D. 1993. *Aide Multicritère à la Décision: Méthodes et Cas*, Economica, Paris.
- [91] SKOWRON A. 1993. Boolean reasoning for decision rules generation. In: KOMOROWSKI J & RAS ZW. (Eds.), *Methodologies for Intelligent Systems*. Springer LNAI 689, Berlin, pp. 295–305.
- [92] SKOWRON A & POLKOWSKI L. 1997. Decision algorithms: a survey of rough set-theoretic methods. *Fundamenta Informaticae*, **27**: 345–358.

- [93] SKOWRON A & STEPANIUK J. 1995. Generalized approximation spaces. In: LIN TY & WILDBERGER A. (Eds.), *Soft Computing*. Simulation Councils, Inc., San Diego, CA, pp. 18–21.
- [94] SLOWINSKI R. 1992a. A generalization of the indiscernibility relation for rough set analysis of quantitative information. *Rivista di matematica per le scienze economiche e sociali*, **15**: 65–78.
- [95] SLOWINSKI R. (ED.) 1992b. *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*. Kluwer, Dordrecht.
- [96] SLOWINSKI R. 1993. Rough set learning of preferential attitude in multi-criteria decision making. In: KOMOROWSKI J & RAS ZW. (Eds.), *Methodologies for Intelligent Systems*. Springer LNAI 689, Berlin, pp. 642–651.
- [97] SLOWINSKI R, GRECO S & MATARAZZO B. 2002a. Rough set analysis of preference-ordered data. In: ALPIGINI JJ, PETERS JF, SKOWRON A & ZHONG N. (Eds.), *Rough Sets and Current Trends in Computing*. Springer LNAI 2475, Berlin, pp. 44–59.
- [98] SLOWINSKI R, GRECO S & MATARAZZO B. 2002b. Mining decision-rule preference model from rough approximation of preference relation. In: PROC. 26<sup>th</sup> IEEE ANNUAL INT. CONF. ON COMPUTER SOFTWARE & APPLICATIONS, Oxford, UK, pp. 1129–1134.
- [99] SLOWINSKI R, GRECO S & MATARAZZO B. 2002b. Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle. *Control and Cybernetics*, **31**: 1005–1035.
- [100] SLOWINSKI R, GRECO S & MATARAZZO B. 2005. Rough set based decision support. Chapter 16. In: BURKE EK & KENDALL G. (Eds.), *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, Springer, New York, pp. 475–527.
- [101] SLOWINSKI R, GRECO S & MATARAZZO B. 2009. Rough Sets in Decision Making. In: MEYERS RA. (Ed.), *Encyclopedia of Complexity and Systems Science*, Springer, New York, pp. 7753–7786.
- [102] SLOWINSKI R, STEFANOWSKI J, GRECO S & MATARAZZO B. 2000. Rough sets based processing of inconsistent information in decision analysis. *Control and Cybernetics*, **29**: 379–404.
- [103] SLOWINSKI R & VANDERPOOTEN D. 1997. Similarity relation as a basis for rough approximations. In: WANG PP. (Ed.), *Advances in Machine Intelligence & Soft-Computing*, vol. IV, Duke University Press, Durham, NC, pp. 17–33.
- [104] SLOWINSKI R & VANDERPOOTEN D. 2000. A generalised definition of rough approximations. *IEEE Transactions on Data and Knowledge Engineering*, **12**: 331–336.
- [105] SLOWINSKI R & ZOPOUNIDIS C. 1995. Application of the rough set approach to evaluation of bankruptcy risk. *Intelligent Systems in Accounting, Finance & Management*, **4**: 27–41.
- [106] 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*. Vol. 1, Physica-Verlag, pp. 500–529.
- [107] STEPANIUK J. 2000. Knowledge Discovery by Application of Rough Set Models. In: POLKOWSKI L, TSUMOTO S & LIN TY. (Eds.), *Rough Set Methods and Application*, Physica-Verlag, Heidelberg, pp. 137–231.
- [108] SZELAG M, GRECO S & SLOWINSKI R. 2012. Variable Consistency Dominance-based Rough Set Approach to Preference Learning in Multicriteria Ranking. Submitted for publication.



- [109] THOMAS LC, CROOK JN & EDELMAN DB. (Eds.). 1992. *Credit Scoring and Credit Control*. Clarendon Press, Oxford.
- [110] TVERSKY A. 1977. Features of similarity. *Psychological Review*, **84**: 327–352.
- [111] YAO Y & WONG S. 1995. Generalization of rough sets using relationships between attribute values. In: PROC. 2<sup>nd</sup> ANNUAL JOINT CONFERENCE ON INFORMATION SCIENCES, Wrightsville Beach, NC, pp. 30–33.
- [112] WIERZBICKI AP. 1980. The use of reference objectives in multiobjective optimization. In: FANDEL G & GAL T. (Eds.), *Multiple Criteria Decision Making, Theory and Applications*, Springer, Berlin, pp. 468–486.
- [113] ZIARKO W. 1993. Variable precision rough sets model. *Journal of Computer and Systems Sciences*, **46**: 39–59.
- [114] ZIARKO W. 1998. Rough sets as a methodology for data mining. In: POLKOWSKI L & SKOWRON A. (Eds.), *Rough Sets in Knowledge Discovery*. Vol. 1, Physica-Verlag, pp. 554–576.
- [115] ZIARKO W & SHAN N. 1994. An incremental learning algorithm for constructing decision rules. In: ZIARKO WP. (Ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery*. Springer, Berlin, pp. 326–334.
- [116] ZIONTS S & WALLENUS J. 1976. An interactive programming method for solving the multiple criteria problem. *Management Science*, **22**: 652–663.
- [117] ZIONTS S & WALLENUS J. 1983. An interactive multiple objective linear programming method for a class of underlying nonlinear utility functions. *Management Science*, **29**: 519–523.