Roman Słowiński

Salvatore Greco

Benedetto Matarazzo

DOMINANCE-BASED ROUGH SET APPROACH TO MULTIPLE CRITERIA DECISION SUPPORT

Abstract

The utility of the rough set approach to multiple criteria decision support is related to the nature of both, the input preferential information available in decision analysis, and the output of the analysis. As to the input, the rough set approach requires a set of decision examples. This is convenient for the acquisition of preferential information from decision makers. Very often in multiple criteria decision support, this information has to be given in terms of preference model parameters, such as importance weights, substitution ratios and various thresholds. Producing such information requires a significant cognitive effort on the part of the decision maker. It is generally acknowledged that people often prefer to make exemplary decisions and cannot always explain them in terms of specific parameters. For this reason, the idea of inferring preference models from exemplary decisions provided by the decision maker is very attractive. Furthermore, the exemplary decisions may be inconsistent because of limited clear discrimination between values of particular criteria and because of hesitation on the part of the decision maker. These inconsistencies can convey important information that should be taken into account in the construction of the decision maker's preference model. The rough set approach is intended to deal with inconsistency and this is a major argument to support its application to multiple criteria decision analysis. The output of the analysis, i.e. the model of preferences in terms of "*if..., then...*" decision rules, is very convenient for decision support because it is intelligible and speaks the same language as the decision maker. The rough set approach adapted to multiple criteria decision support is called Dominance-based Rough Set Approach (DRSA). DRSA is concordant with the concept of granular computing, however, the granules are dominance cones in evaluation space and not bounded sets as it is the case in the basic rough set approach. It is also concordant with the paradigm of computing with words, as it exploits ordinal, and not necessarily cardinal, character of data. We present DRSA for multiple criteria classification, choice and ranking, as well as DRSA for decisions under risk. Finally, we compare DRSA with other decision support paradigms at an axiomatic level.

Keywords

Rough sets, multiple criteria decision support, decision under risk, knowledge discovery, preference model, decision rules.

INTRODUCTION

We present a knowledge discovery methodology for multiple attribute and multiple criteria decision support, which is based upon the concept of rough set proposed by Z. Pawlak [29, 30, 32]. Taking part in the development of rough set theory from the beginning, we adapted and extended its basic paradigm in many ways [31, 45, 46, 47]. For a long time, we also made attempts to employ rough set theory for decision support [33, 38, 39]. The standard rough set approach was not able, however, to deal with preference-ordered domains of attribute (then, called criteria) and preference-ordered decision classes, which are characteristic features of decision problems.

In the late 90's, adapting the standard rough set approach to knowledge discovery from preference-ordered data became a particularly challenging problem within the field of multiple criteria decision support. Why might it be so important? The answer is related to the nature of the input preferential information available in multiple criteria decision analysis and of the output of that analysis. As to the input, the rough set approach requires a set of decision examples. Such representation is convenient for the acquisition of preferential information from decision makers. Very often in multiple criteria decision analysis, this information has to be given in terms of preference model parameters, such as importance weights, substitution ratios and various thresholds. Producing such information requires a significant cognitive effort on the part of the decision maker. It is generally acknowledged that people often prefer to make exemplary decisions and cannot always explain them in terms of specific parameters.

For this reason, the idea of inferring preference models from exemplary decisions provided by the decision maker is very attractive. Furthermore, the exemplary decisions may be inconsistent because of limited clear discrimination between values of particular criteria and because of hesitation on the part of the decision maker. These inconsistencies cannot be considered as a simple error or as noise. They can convey important information that should be taken into account in the construction of the decision maker's preference model. The rough set approach is intended to deal with inconsistency and this is a major argument to support its application to multiple criteria decision analysis. The output of the analysis, i.e. the model of preferences in terms of decision rules, is very convenient for decision support because it is intelligible and speaks the same language as the decision maker.

An extension of the standard rough set approach which enables the analysis of preference-ordered data was proposed in [6, 7, 8, 11, 15]. 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 exemplary decisions) a 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 decision maker's preferences results from the distinction of different kinds of decision rules, 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 traditional functional models considered within multiattribute utility theory, or the relational models considered, for example, in outranking methods. This conclusion has been acknowledged by a thorough study of axiomatic foundations [16, 17, 42]. DRSA has also been used as a tool for inducing parameters of other preference models than the decision rules, like the relational outranking model used in multiple criteria choice problems [37].

As to the application side of the rough set approach, it has been used for discovering regularities in complex phenomena, like stormwater pollution [36], bankruptcy risk of firms applying for a bank credit [38], finding indications for a surgery treatment [31] and classification of Siberian forests [2]. A special attention has been paid to application of the rough set approach in clinical practice, to support some diagnostic and managerial decisions in hospital emergency rooms. This application required extension of the rough set approach to handle incomplete data. The results were implemented as a "decision making core" of a clinical decision support system developed on a mobile platform [27]. The system, called MET (*Mobile Emergency Triage*), supports triage of pediatric patients with various acute conditions. It underwent a clinical trial in the Children's Hospital of Eastern Ontario in Ottawa [50].

Since the first formulation of DRSA, we have proposed many extensions of the approach that make it a useful tool for many specific decision problems. In this survey, we characterize the basic DRSA approach and its main extensions (for complementary surveys see [18, 19, 20, 45]).

The chapter is organized as follows. In the next section, we introduce the concept of knowledge discovery from preference-ordered data. Then, we present the basic Dominance-based Rough Set Approach (DRSA) and in the following sections we review its main extensions. In the last section some conclusions are given and current research directions are outlined.

1. KNOWLEDGE DISCOVERY FROM PREFERENCE ORDERED DATA

The data set in which classification patterns are searched for is called the *learning sample*. The learning of patterns from this sample should take into account available *prior knowledge* that may include the following items (see [40]):

(i) Domains of attributes, i.e. sets of values that an attribute may take while being meaningful to the user.

(ii) A division of attributes into condition and decision attributes, which restricts the range of patterns to functional relations between condition and decision attributes.

(iii) A preference order in the domains of some attributes and a semantic correlation between pairs of these attributes, requiring the patterns to observe the dominance principle.

In fact, item (i) is usually taken into account in knowledge discovery. With this prior knowledge only, one can discover patterns called *association rules* which show strong relationships between values of some attributes, without fixing which attributes will be on the condition and which ones on the decision side in all rules.

If item (i) is combined with item (ii) in the prior knowledge, then one can consider a partition of the learning sample into decision classes defined by decision attributes. The patterns to be discovered have then the form of *decision trees* or *decision rules* representing functional relations between condition and decision attributes. These patterns are typically discovered by machine learning and data mining methods [28]. As there is a direct correspondence between a decision tree and rules, we will concentrate our attention on decision rules only.

As item (iii) is crucial for decision support, let us explain it in more detail. 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. The domains of these attributes are composed of three values: *bad*, *medium* and *good*. This information constitutes item (i) of prior knowledge. Item (ii) is also available because, clearly, *Math* and *Ph* are condition attributes while *GA* is a decision attribute. The preference order of the attribute values is obvious: *good* is better than *medium* and *bad*, and *medium* is better than *bad*. It is known, moreover, that both *Math* and *Ph* are semantically correlated with *GA*. This is, precisely, item (iii) of the prior knowledge.

Attributes with preference-ordered domains are called *criteria* because they involve an evaluation. We will use the name of *regular attributes* for those attributes whose domains are not preference-ordered. *Semantic correlation between two criteria* (condition and decision) means that an improvement on one criterion should not worsen the evaluation on the second criterion, while other attributes and criteria are unchanged. In our example, an improvement of a pupil's score in *Math* or *Ph*, with other attribute values unchanged, should not worsen the pupil's general achievement (*GA*), but rather improve it. In general, semantic correlation between condition criteria and decision criteria requires that an object x dominating object y on all condition criteria (i.e. x having evaluations at least as good as y on all condition criteria) should also dominate y on all decision criteria (i.e. x should have evaluations at least as good as y on all decision criteria). This principle is called the *dominance principle* (or Pareto principle) and it is the only objective principle that is widely agreed upon in the multiple criteria comparisons of objects.

Let us consider two questions:

- What classification patterns can be drawn from the pupils' data set?
- How does item (iii) influences the classification patterns?

The answer to the first question is: "*if..., then...*" decision rules. Each decision rule is characterized by a *condition profile* and a *decision profile*, corresponding to vectors of threshold values of regular attributes and criteria in the condition and decision parts of the rule, respectively. The answer to the second question is that condition and decision profiles of a decision rule should observe the dominance principle 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 they both involve the same values of regular attributes and 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 learning sample, 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. 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 patterns.

If item (iii) were 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 inconsistency in knowledge discovery [26, 30, 32, 34, 35, 39]. 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 granules are bounded sets. Decision rules induced from rough approximation of a classification are also built up of such granules. While taking into account prior knowledge of type (i) and (ii), the rough approximation and the inherent rule induction ignore, however, prior knowledge of type (iii). In consequence, the resulting decision rules may be inconsistent with the dominance principle.

The authors have proposed an extension of the granular computing paradigm that enables us to take into account prior knowledge of type (iii), in addition to either (i) only [23], or (i) and (ii) together [8, 15, 40]. 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 of type (iii) 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 $C \cap D = \varnothing$. Also, let $X_C = \prod_{q=1}^{|C|} X_q$ and $X_D = \prod_{q=1}^{|D|} X_q$ 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 evaluation of objects on attributes from set $C = \{1,\ldots,|C|\}$ and from set $D = \{1,\ldots,|D|\}$, 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 $\boldsymbol{Cl} = \{Cl_t, t = 1,...,n\}$. Each $x \in U$ belongs to one and only one class $Cl_t \in \boldsymbol{Cl}$.

The above definitions take into account prior knowledge of type (i) and (ii) only. 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 patterns 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 prior knowledge includes item (iii) in addition to (i) and (ii), 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, $x_q \succeq_q y_q$ 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, x *P-dominates* y) in the condition attribute space X_P (denoted by xD_Py) if $x_q \succeq_q y_q$ for all $q \in P$. Assuming, without loss of generality, that the domains of the criteria are numerical (i.e. $X_q \subseteq \boldsymbol{R}$ for any $q \in C$) and that they are ordered so that the preference increases with the value, we can say that xD_Py is equivalent to $x_q \geq y_q$ for all $q \in P$, $P \subseteq C$. Observe that for each $x \in X_P$, xD_Px, i.e. P-dominance is reflexive. On the other hand, the analogous definition holds in the decision attribute space X_R (denoted by xD_Ry), where $R \subseteq D$.

The dominance relations xD_Py and xD_Ry ($P \subseteq C$ *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 $D_P^+(x) = \{y \in U\colon yD_Px\}$.

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 $D_P^-(x) = \{y \in U: xD_P y\}$.

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:

$Cl_R^{\geq x} = \{y \in U: yD_R x\}$, $Cl_R^{\leq x} = \{y \in U: xD_R y\}$.

$Cl_{t_q} = \{x \in X_D: x_q = t_q\}$ 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 patterns 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 $x \in X_P$.

In both cases above, the functions are sets of decision rules.

2. THE DOMINANCE-BASED ROUGH SET APPROACH (DRSA)

2.1. Granular computing with dominance cones

When discovering classification patterns, 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, $\boldsymbol{Cl} = \{Cl_t, t=1,...,n\}$. Each $x \in U$ belongs to one and only one class, $Cl_t \in \boldsymbol{Cl}$. 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,...,n. Notice that for t = 2,...,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 [8, 11, 14, 18, 19, 20, 45]).

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

If, given a set of criteria $P \subseteq C$, the inclusion of $x \in U$ to $Cl_t^{\geq}$, where $t = 2,\ldots,n$, 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 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, we can say:

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

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

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

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

$$\underline{P}(Cl_t^{\geq}) \subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq})$$
$$\underline{P}(Cl_t^{\leq}) \subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq}).$$

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

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

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

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

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

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 model [21].

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 $l\times 100\%$ of all objects $y\in U$ dominating x with respect to P also belong to $Cl_t^{\geq}$, i.e.

$$\frac{card\left(D_P^+(x)\cap Cl_t^{\geq}\right)}{card\left(D_P^+(x)\right)}\geq l$$

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 $(1-l)\times 100\%$ 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 $l\times 100\%$ of all the objects $y\in U$ dominated by x with respect to P also belong to $Cl_t^{\leq}$, i.e.

$$\frac{card\left(D_P^-(x)\cap Cl_t^{\leq}\right)}{card\left(D_P^-(x)\right)}\geq l$$

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,...,n) or with respect to the downward union $Cl_t^{\leq}$ (t = 1,...,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:

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

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

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 $\overline{P}^l\left(Cl_t^{\geq}\right)$ and $\overline{P}^l\left(Cl_t^{\leq}\right)$, respectively, by complementation of $\underline{P}^l\left(Cl_{t-1}^{\leq}\right)$ and $\underline{P}^l\left(Cl_{t+1}^{\geq}\right)$ with respect to U as follows:

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

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

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

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

where $t = 1,...,n$. The *variable consistency* model of the Dominance-based Rough Set Approach 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 $t = 2,...,n$,

$$\underline{P}^l\left(Cl_t^{\geq}\right) \subseteq \underline{P}^{l'}\left(Cl_t^{\geq}\right) \text{ and } \overline{P}^{l'}\left(Cl_t^{\geq}\right) \subseteq \overline{P}^l\left(Cl_t^{\geq}\right)$$

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 the objects *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 data sample set. Since the objects which are *P*-correctly classified are those that do not belong to any *P*-boundary of unions $Cl_t^{\geq}$ and $Cl_t^{\leq}$, $t = 1,...,n$, the quality of approximation of classification ***Cl*** by set of criteria *P*, can be written as

$$\gamma_P(\boldsymbol{Cl}) = \frac{\left|\left(U - \left(\bigcup_{t\in\{1,...,n\}} Bn_P\left(Cl_t^{\leq}\right)\right) \cup \left(\bigcup_{t\in\{1,...,n\}} Bn_P\left(Cl_t^{\geq}\right)\right)\right)\right|}{|U|} = \frac{\left|\left(U - \left(\bigcup_{t\in\{1,...,n\}} Bn_P\left(Cl_t^{\geq}\right)\right)\right)\right|}{|U|}$$

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

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

- *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 attributes and criteria conveying the relevant knowledge contained in the learning sample. This knowledge is relevant for the explanation of patterns in a given decision table but not necessarily for prediction.

It has been shown in [8, 12] 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 interdependence of the considered attributes and criteria, and can help to choose the best reduct.

2.2. Induction of decision rules

The dominance-based rough approximations of upward and downward unions of classes can serve to induce a generalized description of the objects contained in the decision table in terms of "*if..., then...*" decision rules. For a given upward or downward union of classes, $Cl_t^{\geq}$ or $Cl_s^{\leq}$, the decision rules induced under a hypothesis that objects belonging to $\underline{P}(Cl_t^{\geq})$ or $\underline{P}(Cl_s^{\leq})$ are positive and all the others 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 $\overline{P}(Cl_s^{\leq}) \cap \overline{P}(Cl_t^{\geq})$ 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:

$$if\ x_{q1} \succeq_{q1} r_{q1}\ and\ x_{q2} \succeq_{q2} r_{q2}\ and\ \ldots\ x_{qp} \succeq_{qp} r_{qp},\ then\ x \in Cl_t^{\geq},$$

where for each $w_q, z_q \in X_q$, "$w_q \succeq_q z_q$" means "w_q is <u>at least</u> 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:

$$if\ x_{q1} \succeq_{q1} r_{q1}\ and\ x_{q2} \succeq_{q2} r_{q2}\ and\ \ldots\ x_{qp} \succeq_{qp} r_{qp},\ then\ x \text{ possibly belongs to } Cl_t^{\geq}$$

3) *Certain* $D_{\leq}$*-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 $w_q,z_q\in X_q$, "$w_q\preceq_q z_q$" 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:

$$if\ x_{q1}\preceq_{q1}r_{q1}\ and\ x_{q2}\preceq_{q2}r_{q2}\ and\ ...\ x_{qp}\preceq_{qp}r_{qp},\ then\ x \text{ possibly belongs to } Cl_t^{\leq}$$

5) *Approximate* $D_{\geq\leq}$*-decision rules*. These represent simultaneously lower and upper profile descriptions for objects belonging to $Cl_s\cup Cl_{s+1}\cup\ldots\cup Cl_t$ without the possibility of discerning the actual class:

$$if\ x_{q1}\succeq_{q1}r_{q1}\ and...\ x_{qk}\succeq_{qk}r_{qk}\ and\ x_{qk+1}\preceq_{qk+1}r_{qk+1}\ and\ ...\ x_{qp}\preceq_{qp}r_{qp},$$
$$then\ x\in Cl_s\cup Cl_{s+1}\cup\ldots\cup Cl_t.$$

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".

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.

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 data 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 to the Variable-Consistency Dominance-based Rough Set Approach model mentioned above.

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 $\rho_1, \rho_2,\ldots,\rho_m$, having right hand sides $x\in Cl_{t1}^{\geq}$, $x\in Cl_{t2}^{\geq}$,…, $x\in Cl_{tm}^{\geq}$, then x is assigned to class Cl_t, where $t = \max\{t1,t2,\ldots,tm\}$. In the case of no matching, we can conclude that x belongs to 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 $\rho_1, \rho_2,\ldots,\rho_m$, having right hand sides $x\in Cl_{t1}^{\leq}$, $x\in Cl_{t2}^{\leq}$,…, $x\in Cl_{tm}^{\leq}$, then x is assigned to class Cl_t, where $t = \min\{t1,t2,\ldots,tm\}$. 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 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 α 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 traditional granular computing, the condition attribute space X_C need not be discretized.

Some procedures for rule induction from rough approximations have been proposed in [22, 26, 49].

In [3], a new methodology for the induction of monotonic decision trees from dominance-based rough approximations of preference-ordered decision classes has been proposed.

2.3. An illustrative example

To illustrate the application of the DRSA to multiple criteria classification, we will use a part of some data provided by a Greek industrial bank ETEVA which finances industrial and commercial firms in Greece [6, 48]. 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 sorting decision is represented by decision attribute d making a trichotomic partition of the 39 firms:

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

The partition is denoted by $\boldsymbol{Cl} = \{Cl_A, Cl_U, Cl_{NA}\}$ 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 (↑ means *preference increasing with value* and ↓ means *preference decreasing with value*):

- A_1 = earnings before interests and taxes/total assets, ↑
- A_2 = net income/net worth, ↑
- A_3 = total liabilities/total assets, ↓
- A_4 = total liabilities/cash flow, ↓
- A_5 = interest expenses/sales, ↓
- A_6 = general and administrative expense/sales, ↓
- A_7 = managers' work experience, ↑ (very low = 1, low = 2, medium = 3, high = 4, very high = 5)

- A_8 = firm's market niche/position, ↑ (bad = 1, rather bad = 2, medium = 3, good = 4, very good = 5)
- A_9 = technical structure-facilities, ↑ (bad = 1, rather bad = 2, medium = 3, good = 4, very good = 5)
- A_{10} = organization-personnel, ↑ (bad = 1, rather bad = 2, medium = 3, good = 4, very good = 5)
- A_{11} = special competitive advantage of firms, ↑ (low = 1, medium = 2, high = 3, very high = 4)
- A_{12} = market flexibility, ↑ (very low = 1, low = 2, medium = 3, high = 4, very high = 5)

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

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

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

We will answer these questions using the DRSA. The first result from this approach is a discovery that the financial data matrix is *consistent* for the complete set of criteria C. Therefore, the C-lower and C-upper approximations of $Cl_{NA}^{\leq}$, $Cl_{U}^{\leq}$ and $Cl_{U}^{\geq}$, $Cl_{A}^{\geq}$ are the same. In other words, the quality of approximation of all upward and downward unions of classes, 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:

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

$RED_{Cl}^{3}=\{A_3, A_4, A_6, A_7\}$, $RED_{Cl}^{4}=\{A_4, A_5, A_6, A_7\}$,

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

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

$RED_{Cl}^{9}=\{A_2, A_5, A_7, A_9\}$, $RED_{Cl}^{10}=\{A_4, A_5, A_7, A_9\}$,

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

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

$RED_{Cl}^{15}=\{A_4, A_5, A_6, A_{12}\}$, $RED_{Cl}^{16}=\{A_1, A_3, A_5, A_6, A_9, A_{12}\}$,

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

All 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 ***Cl*** is empty ($CORE_{Cl} = \varnothing$) which means that no criterion is indispensable for the approximation. Moreover, all the criteria are exchangeable and no criterion is redundant.

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

Table 3

Financial data matrix

Firm	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	d
F1	16.4	14.5	59.82	2.5	7.5	5.2	5	3	5	4	2	4	A
F2	35.8	67.0	64.92	1.7	2.1	4.5	5	4	5	5	4	5	A
F3	20.6	61.75	75.71	3.6	3.6	8.0	5	3	5	5	3	5	A
F4	11.5	17.1	57.1	3.8	4.2	3.7	5	2	5	4	3	4	A
F5	22.4	25.1	49.8	2.1	5.0	7.9	5	3	5	5	3	5	A
F6	23.9	34.5	48.9	1.7	2.5	8.0	5	3	4	4	3	4	A
F7	29.9	44.0	57.8	1.8	1.7	2.5	5	4	4	5	3	5	A
F8	8.7	5.4	27.4	3.3	4.5	4.5	5	2	4	4	1	4	A
F9	25.7	29.7	46.8	1.7	4.6	3.7	4	2	4	3	1	3	A
F10	21.2	24.6	64.8	3.7	3.6	8.0	4	2	4	4	1	4	A
F11	18.32	31.6	69.3	4.4	2.8	3.0	4	3	4	4	3	4	A
F12	20.7	19.3	19.7	0.7	2.2	4.0	4	2	4	4	1	3	A
F13	9.9	3.5	53.1	4.5	8.5	5.3	4	2	4	4	1	4	A
F14	10.4	9.3	80.9	9.4	1.4	4.1	4	2	4	4	3	3	A
F15	17.7	19.8	52.8	3.2	7.9	6.1	4	4	4	4	2	5	A
F16	14.8	15.9	27.94	1.3	5.4	1.8	4	2	4	3	2	3	A
F17	16.0	14.7	53.5	3.9	6.8	3.8	4	4	4	4	2	4	A
F18	11.7	10.01	42.1	3.9	12.2	4.3	5	2	4	2	1	3	A
F19	11.0	4.2	60.8	5.8	6.2	4.8	4	2	4	4	2	4	A
F20	15.5	8.5	56.2	6.5	5.5	1.8	4	2	4	4	2	4	A
F21	13.2	9.1	74.1	11.21	6.4	5.0	2	2	4	4	2	3	U
F22	9.1	4.1	44.8	4.2	3.3	10.4	3	4	4	4	3	4	U
F23	12.9	1.9	65.02	6.9	14.01	7.5	4	3	3	2	1	2	U
F24	5.9	-27.7	77.4	-32.2	16.6	12.7	3	2	4	4	2	3	U
F25	16.9	12.4	60.1	5.2	5.6	5.6	3	2	4	4	2	3	U
F26	16.7	13.1	73.5	7.1	11.9	4.1	2	2	4	4	2	3	U
F27	14.6	9.7	59.5	5.8	6.7	5.6	2	2	4	4	2	4	U
F28	5.1	4.9	28.9	4.3	2.5	46.0	2	2	3	3	1	2	U
F29	24.4	22.3	32.8	1.4	3.3	5.0	2	3	4	4	2	3	U
F30	29.7	8.6	41.8	1.6	5.2	6.4	2	3	4	4	2	3	U
F31	7.3	-64.5	67.5	-2.2	30.1	8.7	3	3	4	4	2	3	NA
F32	23.7	31.9	63.6	3.5	12.1	10.2	3	2	3	4	1	3	NA
F33	18.9	13.5	74.5	10.0	12.0	8.4	3	3	3	4	3	4	NA
F34	13.9	3.3	78.7	25.5	14.7	10.1	2	2	3	4	3	4	NA
F35	-13.3	-31.1	63.0	-10.0	21.2	23.1	2	1	4	3	1	2	NA
F36	6.2	-3.2	46.1	5.1	4.8	10.5	2	1	3	3	2	3	NA
F37	4.8	-3.3	71.9	34.6	8.6	11.6	2	2	4	4	2	3	NA
F38	0.1	-9.6	42.5	-20.0	12.9	12.4	1	1	4	3	1	3	NA
F39	13.6	9.1	76.0	11.4	17.1	10.3	1	1	2	1	1	2	NA

The 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* $x\in Cl_{NA}^{\leq}$, (4)
2. *if* $f(x,A_2)\leq 3.3$ *and* $f(x,A_7)\leq 2$, *then* $x\in Cl_{NA}^{\leq}$, (5)
3. *if* $f(x,A_3)\geq 63.6$ *and* $f(x,A_7)\leq 3$ *and* $f(x,A_9)\leq 3$, *then* $x\in Cl_{NA}^{\leq}$, (4)
4. *if* $f(x,A_2)\leq 12.4$ *and* $f(x,A_6)\geq 5.6$, *then* $x\in Cl_{U}^{\leq}$, (14)
5. *if* $f(x,A_7)\leq 3$, *then* $x\in Cl_{U}^{\leq}$, (18)
6. *if* $f(x,A_2)\geq 3.5$ *and* $f(x,A_5)\leq 8.5$, *then* $x\in Cl_{U}^{\geq}$, (26)
7. *if* $f(x,A_7)\geq 4$, *then* $x\in Cl_{U}^{\geq}$, (21)
8. *if* $f(x,A_1)\geq 8.7$ *and* $f(x,A_9)\geq 4$, *then* $x\in Cl_{U}^{\geq}$, (27)
9. *if* $f(x,A_2)\geq 3.5$ *and* $f(x,A_7)\geq 4$, *then* $x\in Cl_{A}^{\geq}$, (20)

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

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

The well-known machine discovery methods cannot deal with multiple criteria classification because they do not consider preference orders in the domains of attributes and among the classes. There are multiple criteria decision analysis methods for such classification. However, they are not discovering classification patterns from data. They simply apply a preference model, like the utility function in scoring methods, to a set of objects to be classified. In this sense, they are not knowledge discovery methods at all.

Comparing the DRSA to the standard rough set approach, we can notice the following differences between the two approaches. The standard 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 DRSA and multiple criteria 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 standard rough set approach and the DRSA.

There are three notable advantages of the DRSA over the standard 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 standard 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 matrix without any preprocessing 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 DRSA use relations from $\{\leq,=,\geq\}$, while those resulting from the standard rough set approach only use =. The DRSA syntax is more understandable to practitioners. The minimal sets of DRSA decision rules are smaller than the minimal sets which result from the standard rough set approach.

3. THE DOMINANCE-BASED ROUGH SET APPROACH FOR MULTIPLE CRITERIA CHOICE AND RANKING

One of the very first extensions of the DRSA concerned preference-ordered data representing pairwise comparisons (i.e. binary relations) between objects on both, condition and decision attributes [7, 8, 11]. 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 patterns (i.e. 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, from 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 preferential information of a decision maker in a multiple criteria 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 of criteria C. The best choice (or the preference ranking) in set A is semantically correlated with the criteria from set C. The preferential information concerning the multiple criteria 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 from C and a comprehensive preference relation. One such example is a weak preference relation called the *outranking relation*. By using the DRSA 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.

3.1. The pairwise comparison table as preferential information and as a learning sample

A set of reference objects represent a decision problem and a decision maker can express the preferences by pairwise comparisons. In the following, xSy denotes the presence, while xS^cy 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 then y, i.e. xS^cy,
3) the two objects are incomparable at the present stage.

An $m\times(n+1)$ pairwise comparison table, denoted by $\boldsymbol{S}_{PCT}$, is then created on the basis of this information. The first n columns correspond to the criteria from set C. The last, i.e. the $(n+1)$-th, column represents the comprehensive binary preference relation S or S^c. The m rows are pairs from B. For each pair in $\boldsymbol{S}_{PCT}$, 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 $\boldsymbol{S}_{PCT}$.

We will define $\boldsymbol{S}_{PCT}$ 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 $\boldsymbol{S}_{PCT} = \langle B, C\cup\{d\}, T_C\cup T_d, f\rangle$, where $B\subseteq A\times A$ is a non-empty *set of exemplary pairwise comparisons of reference objects*, $T_C = \bigcup_{g_i\in C} T_i$, d is a decision corresponding to the comprehensive pairwise comparison (comprehensive preference relation), and $f{:}B\times(C\cup\{d\})\rightarrow T_C\cup T_d$ is a total function such that $f[(x,y),q]\in T_i$ for every $(x,y)\in A\times A$ and for each $g_i\in C$, and $f[(x,y),q]\in T_d$ 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 $\boldsymbol{S}_{PCT}$ can be seen as decision table, since the set of considered criteria C and the decision d are distinguished.

We assume that the exemplary pairwise comparisons made by the decision maker can be represented in terms of *graded preference relations* (for example “very large preference”, “large preference”, “strict preference”, “strong preference” and “very strong preference”), denoted by P_q^h : For each $q\in C$ 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

- $x P_i^h y$, $h>0$, means that object x is preferred to object y by degree h with respect to criterion g_i,
- $x P_i^h y$, $h<0$, means that object x is not preferred to object y by degree h with respect to criterion g_i,
- $x P_i^0 y$ 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 $(x,y)\in A\times A$,

$$[x P_i^h y, h>0] \Rightarrow [y P_i^k x, k\le 0], [x P_i^h y, h<0] \Rightarrow [y P_i^k x, k\ge 0].$$

The set of binary relations T_d may be defined in a similar way, but $x P_d^h y$ means that object x is comprehensively preferred to object y by degree h. 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^cy 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.

3.2. Rough approximation of the outranking and non-outranking 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 $C^N \cap C^O = \varnothing$. 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 $P^N \cap P^O = \varnothing$.

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.

3.2.1. Multigraded dominance

Let $P = P^N$ and $P^O = \varnothing$. Given $P \subseteq C$ ($P \neq \varnothing$), $(x,y),(w,z) \in A \times A$, 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. $h_i \geq k_i$, where $h_i,k_i \in H_i$, $x\, P_i^{hi}\, y$ and $w\, P_i^{ki}\, z$, 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 $\{(w,z) \in A \times A\colon (w,z)D_P(x,y)\}$,
- A set of pairs of objects dominated by (x,y), called the *P-dominated set*, denoted by $D_P^-(x,y)$ and defined as $\{(w,z) \in A \times A\colon (x,y)D_P(w,z)\}$.

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 [8] 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

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

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 $RED_{\boldsymbol{S}_{PCT}}$). Note that $\boldsymbol{S}_{PCT}$ can have more than one reduct. The intersection of all B-reducts is called the *core* (denoted by $CORE_{\boldsymbol{S}_{PCT}}$).

It is also possible to use the Variable Consistency Model on $\boldsymbol{S}_{PCT}$ [41] but being aware that some of the pairs in the positive or negative dominance sets belong to the opposite relation but at least $l\times100\%$ of pairs belong to the correct one. Then the definition of the lower approximations of S and S^c boils down to:

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

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

3.2.2. 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 $P^N = \varnothing$, 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 $P^N\neq\varnothing$ and $P^O\neq\varnothing$, 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.

3.3. Induction of decision rules from rough approximations of outranking and non-outranking relations

Using the rough approximations of S and S^c defined in 3.2.1 and 3.2.2, it is possible to induce a generalized description of the preferential information contained in a given $\boldsymbol{S}_{PCT}$ in terms of suitable decision rules. The syntax of these rules is based on 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 $(x,y)\in A\times A$, $g_i\in C$ 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. *Certain* $\mathrm{D}_{\geq}$*-decision rules*:

$$\textit{if } x P_{i1}^{\geq h(i1)} y \textit{ and}\ldots x P_{ie}^{\geq h(ie)} y \textit{ and } g_{ie+1}(x)\geq r_{ie+1} \textit{ and } g_{ie+1}(y)\leq s_{ie+1} \textit{ and}\ldots g_{ip}(x)\geq r_{ip} \textit{ and } g_{ip}(y)\leq s_{ip}, \textit{ then } xSy,$$

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

2. *Certain* $\mathrm{D}_{\leq}$*-decision rules*:

$$\textit{if } x P_{i1}^{\leq h(i1)} y \textit{ and}\ldots x P_{ie}^{\leq h(ie)} y \textit{ and } g_{ie+1}(x)\leq r_{ie+1} \textit{ and } g_{ie+1}(y)\geq s_{ie+1} \textit{ and}\ldots g_{ip}(x)\leq r_{ip} \textit{ and } g_{ip}(y)\geq s_{ip}, \textit{ then } xS^cy,$$

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

3. *Approximate* $\mathrm{D}_{\geq\leq}$*-decision rules*:

$$\textit{if } x P_{i1}^{\geq h(i1)} y \textit{ and}\ldots x P_{ie}^{\geq h(ie)} y \textit{ and } x P_{ie+1}^{\leq h(ie+1)} y\ldots x P_{if}^{\leq h(if)} y \textit{ and } g_{if+1}(x)\geq r_{if+1} \textit{ and } g_{if+1}(y)\leq s_{if+1} \textit{ and}\ldots g_{ig}(x)\geq r_{ig} \textit{ and } g_{ig}(y)\leq s_{ig} \textit{ and } g_{ig+1}(x)\leq r_{ig+1} \textit{ and } g_{ig+1}(y)\geq s_{ig+1} \textit{ and}\ldots g_{ip}(x)\leq r_{ip} \textit{ and } g_{ip}(y)\geq s_{ip}, \textit{ then } xSy \textit{ or } xS^cy,$$

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

3.4. Use of decision rules for decision support

The decision rules induced from a given $\boldsymbol{S}_{PCT}$ 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 $\boldsymbol{S}_{PCT}$ represents a preference model from 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^cv, i.e. *true* outranking (denoted by uS^Tv)
- uS^cv and *not* uSv, i.e. *false* outranking (denoted by uS^Fv)
- uSv and uS^cv, i.e. *contradictory* outranking (denoted by uS^Kv)
- *not* uSv and *not* uS^cv, i.e. *unknown* outranking (denoted by uS^Uv)

The four above situations, which together constitute the so-called *four-valued outranking* [24], 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

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

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

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

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

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 [24].

3.5. 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: $V_1 = V_2 = V_3 = \{$sufficient, medium, good$\}$. The decision attribute (d) indicates the profitability of warehouses, expressed by the *Return On Equity* (*ROE*) ratio (in %). Table 2 presents a decision table which represents this situation.

Table 2

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:

- $x P_i^0 y$ (and $y P_i^0 x$), meaning that x is *indifferent* to y with respect to A_i, if $f(x,A_i) = f(y,A_i)$,
- $x P_i^1 y$ (and $y P_i^{-1} x$), meaning that x is *preferred* to y with respect to A_i, if $f(x,A_i) =$ good and $f(y,A_i) =$ medium, or if $f(x,A_i) =$ medium and $f(y,A_i) =$ sufficient,
- $x P_i^2 y$ (and $y P_i^{-2} x$), meaning that x is *strongly preferred* to y with respect to A_i, if $f(x,A_i) =$ good and $f(y,A_i) =$ 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^cy).

The pairwise comparisons of the reference objects result in $\boldsymbol{S}_{PCT}$. The rough set analysis of the $\boldsymbol{S}_{PCT}$ 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, $RED_{\boldsymbol{S}_{PCT}}$ = $CORE_{\boldsymbol{S}_{PCT}}$ = $\{A_1,A_2,A_3\}$. 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:

$\underline{C}(S)$ = {(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)}

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 $x\,P_1^{\geq 1}\,y$ *and* $x\,P_2^{\geq 1}\,y$, *then* xSy; ((1,6),(3,6),(7,6))

if $x\,P_2^{\geq 1}\,y$ *and* $x\,P_3^{\geq 0}\,y$, *then* xSy; ((1,2),(1,6),(1,8),(3,2),(3,6),(3,8),(7,2),(7,6),(7,8))

if $x\,P_2^{\geq 0}\,y$ *and* $x\,P_3^{\geq 1}\,y$, *then* xSy; ((1,4),(1,5),(3,4),(3,5),(7,4),(7,5))

if $x\,P_1^{\leq -1}\,y$ *and* $x\,P_2^{\leq -1}\,y$, *then* xS^cy; ((6,1),(6,3),(6,7))

if $x\,P_2^{\leq 0}\,y$ *and* $x\,P_3^{\leq -1}\,y$, *then* xS^cy; ((4,1),(4,3),(4,7),(5,1),(5,3),(5,7))

if $x\,P_1^{\leq 0}\,y$ *and* $x\,P_2^{\leq -1}\,y$ *and* $x\,P_3^{\leq 0}\,y$, *then* xS^cy; ((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 $x\,P_2^{\leq 0}\,y$ *and* $x\,P_2^{\geq 0}\,y$ *and* $x\,P_3^{\leq 0}\,y$ *and* $x\,P_3^{\geq 0}\,y$, *then* xSy *or* xS^cy;

((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 $x\,P_2^{\leq -1}\,y$ *and* $x\,P_3^{\geq 1}\,y$, *then* xSy *or* xS^cy; ((2,4),(2,5),(6,4),(6,5),(8,4),(8,5))

if $x\,P_2^{\geq 1}\,y$ *and* $x\,P_3^{\leq -1}\,y$, *then* xSy *or* xS^cy; ((4,2),(4,6),(4,8),(5,2),(5,6),(5,8))

if $x\,P_1^{\geq 1}\,y$ *and* $x\,P_2^{\leq 0}\,y$ *and* $x\,P_3^{\leq 0}\,y$, *then* xSy *or* xS^cy; ((1,3),(2,3),(2,6),(7,3),(8,3),(8,6)),

if $x\,P_1^{\geq 1}\,y$ *and* $x\,P_2^{\leq -1}\,y$, *then* xSy *or* xS^cy; ((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 3. The DRSA 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 3, as follows:

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

Table 3

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

Summary

We briefly presented the contribution of the DRSA to multiple 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 multiple criteria decision problem in terms of exemplary decisions.
- The rough set analysis of preferential 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 preferential 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 DRSA, while classical methods consider only quantitative ordered evaluations (with rare exceptions).
- No prior discretization of the quantitative domains of criteria is necessary.

Rough approximations of a comprehensive preference relation can be defined using other types of dominance than the Pareto dominance used in this section. In [43], the Lorenz dominance has been used for rough approximations, permitting induction of more robust decision rules, i.e. certain decision rules supported by consistent pairs of objects characterized by equitable distributions of intensities of preference on considered criteria

4. DRSA FOR DECISION UNDER RISK

In [13], we opened a new avenue for applications of the rough set concept. This avenue leads to the classical problem of *decision under risk*. To adapt the DRSA to this problem, we substituted the dominance relation by *stochastic dominance relation* defined on a set of objects meaning acts. We considered the case of traditional additive probability distribution over a set of states of the world, however, the model is rich enough to handle non-additive probability distributions and even qualitative ordinal distributions. The adapted DRSA gives a representation of DM's preferences under risk in terms of "*if..., then...*" decision rules induced from rough approximation of preference ordered classification of acts described in terms of outcomes in uncertain states of the world. The preference ordered classification constitutes, in this case, preferential information acquired from the DM.

4.1. DRSA based on stochastic dominance

To apply DRSA to decision under risk, we consider the following basic elements:

- a set $S = \{s_1, s_2, \ldots, s_s\}$ of states of the world, or simply *states*, which are supposed to be mutually exclusive and collectively exhaustive,
- an a priori probability distribution P over the states of the world: more precisely, the probabilities of states $s_1, s_2, \ldots, s_s$ are $p_1, p_2, \ldots, p_s$, respectively ($p_1 + p_2 + \ldots + p_s = 1, p_i \geq 0, i = 1, \ldots s$),

- a set $A = \{A_1, A_2, \ldots, A_m\}$ of *acts*,
- a set $X = \{x_1, x_2, \ldots, x_r\}$ of consequences or outcomes that, for the sake of simplicity, are supposed to be expressed in monetary terms, thus $X \subseteq \boldsymbol{R}$,
- a function g: $A \times S \rightarrow X$ assigning to each act-state pair $(A_i, s_j) \in A \times S$ a consequence $x_h \in X$,
- a set of classes $\boldsymbol{Cl} = \{Cl_1, Cl_2, \ldots, Cl_n\}$, such that $Cl_1 \cup Cl_2 \cup \ldots \cup Cl_n = A$, $Cl_p \cap Cl_q = \varnothing$ for each $p,q \in \{1,2\ldots,n\}$ with $p \neq q$; the classes of $\boldsymbol{Cl}$ are preference ordered according to an increasing order of their indices, in the sense that for each $A_i,A_j \in A$, if $A_i \in Cl_p$ and $A_j \in Cl_q$ with $p>q$, then A_i is preferred to A_j,
- a function e: $A \rightarrow \boldsymbol{Cl}$ assigning each act $A_i \in A$ to a class $Cl_t \in \boldsymbol{Cl}$.

In this context, two different types of dominance relations can be considered:

1) (classical) *dominance*: given $A_i,A_j \in A$, A_i dominates A_j iff for each possible state of the world act A_i gives an outcome at least as good as act A_j; more formally, $g(A_i,s_k) \geq g(A_j,s_k)$, for each $s_k \in S$,
2) *stochastic dominance*: given $A_i,A_j \in A$, A_i stochastically dominates A_j iff for each outcome $x \in X$, act A_i gives an outcome at least as good as x with a probability at least as large as act A_j.

Case 1) corresponds to a model in which the utility is state dependent while case 2) corresponds to a model of decision under risk proposed by. We consider the second case.

On the basis of an a priori probability distribution P, we can assign to each subset of states of the world $W \subseteq S$ ($W \neq \varnothing$) the probability $P(W)$ that one of the states in W is verified, i.e. $P(W) = \sum_{i:s_i \in W} p_i$ and then to build up the set $\boldsymbol{\Pi}$ of all the possible values $P(W)$, i.e.

$$\boldsymbol{\Pi} = \{\pi \in [0,1]: \pi = P(W),\ W \subseteq S\}$$

We define the following function z: $A \times S \rightarrow \boldsymbol{\Pi}$, assigning to each act-state pair $(A_i,s_j) \in A \times S$ a probability $\pi \in \boldsymbol{\Pi}$, as follows:

$$z(A_i,s_j) = \sum_{r:g(A_i,s_r) \geq g(A_i,s_j)} p_r$$

Therefore, $z(A_i,s_j)$ represents the probability of obtaining by act A_i an outcome whose value is at least $g(A_i,s_j)$.

On the basis of function $z(A_i, s_j)$, we can define the function ρ: $A \times \boldsymbol{\Pi} \rightarrow X$ as follows:

$$\rho(A_i,\pi) = \min_{j:z(A_i,s_j) \geq \pi} g(A_i,s_j)$$

Thus, $\rho(A_i,\pi) = x$ means that by act A_i one can gain *at least* x with a probability greater than or equal to π.

Using function $z(A_i, s_j)$, we can also define function ρ': $A\times\boldsymbol{\Pi}\rightarrow X$ as follows:

$$\rho'(A_i,\pi) = \max_{j:z(A_i,s_j)\le\pi} g(A_i,s_j)$$

$\rho'(A_i,\pi) = x$ means that by act A_i one can gain *at most* x with a probability smaller than or equal to π.

If the elements of $\boldsymbol{\Pi}$, $0 = \pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(w)} = 1$ (w = card($\boldsymbol{\Pi}$)), are reordered in such a way that $\pi_{(1)}\le\pi_{(2)}\le \ldots \le\pi_{(w)}$, then we have $\rho(A_i,\pi_{(j)})=\rho'(A_i,1-\pi_{(j-1)})$.

Therefore, $\rho(A_i,\pi_{(j)})\le x$ is equivalent to $\rho'(A_i,1-\pi_{(j-1)})\ge x$, $A_i\in A$, $\pi_{(j)}\in\boldsymbol{\Pi}$, $x\in X$.

Given $A_i,A_j\in A$, A_i stochastically dominates A_j if and only if $\rho(A_i,\pi)\ge\rho(A_j,\pi)$ for each $\pi\in\boldsymbol{\Pi}$. This is equivalent to say: given $A_i,A_j\in A$, A_i stochastically dominates A_j if and only if $\rho'(A_i,\pi)\le\rho'(A_j,\pi)$ for each $\pi\in\boldsymbol{\Pi}$.

We can apply DRSA in this context considering the following correspondence:

- the universe U is the set of acts A,
- the set of condition attributes (criteria) C is the set $\boldsymbol{\Pi}$,
- the domain V_π of each criterion $\pi\in\boldsymbol{\Pi}$ is the set X,
- the single decision attribute d specifies classification of acts from A into classes from $\boldsymbol{Cl}$,
- the information function f is a function $\boldsymbol{f}$ such that for all $A_i\in$A and $\pi\in\boldsymbol{\Pi}$, $\boldsymbol{f}(A_i,\pi) = \rho(A_i,\pi)$ and $\boldsymbol{f}(A_i,d) = e(A_i)$,
- the dominance relation on U is the stochastic dominance relation on A.

The aim of DRSA to decision under risk is to explain the preferences of the DM, represented by his/her assignments of the acts from A to the classes of $\boldsymbol{Cl}$, in terms of decision rules involving stochastic dominance on partial profiles corresponding to outcomes x for some probabilities π.

4.2. An illustrative example

The following example illustrates the approach. Let us consider

- set $S = \{s_1, s_2, s_3\}$ of states of the world,
- a priori probability distribution P over the states of the world defined as follows: $p_1 = 0.25$, $p_2 = 0.35$, $p_3 = 0.40$,
- set $A = \{A_1, A_2, A_3, A_4, A_5, A_6\}$ of acts,
- set $X = \{0, 10, 15, 20, 30\}$ of consequences,
- set of classes $\boldsymbol{Cl} = \{Cl_1, Cl_2, Cl_3\}$, where Cl_1 is the set of *bad* acts, Cl_2 is the set of *medium* acts, Cl_3 is the set of *good* acts,
- function g:$A\times S\rightarrow X$ assigning to each act-state pair $(A_i,s_j)\in A\times S$ a consequence $x_h\in X$, and a function e: $A\rightarrow \boldsymbol{Cl}$ assigning each act $A_i\in$A to a class $Cl_t\in\boldsymbol{Cl}$, as presented in Table 4.

Table 4

Acts, consequences and assignment to classes from ***Cl***

	p_j	A_1	A_2	A_3	A_4	A_5	A_6
s_1	0.25	30	0	15	0	20	10
s_2	0.35	10	20	0	15	10	20
s_3	0.40	10	20	20	20	20	20
d		*good*	*medium*	*medium*	*bad*	*medium*	*good*

DRSA is applied on Table 5 including the values of function $\rho(A_i,\pi)$. Let us explain what mean the entries in Table 5. If we consider the column of act, say A_3, we see that by act A_3,

- the value 20 in the row corresponding to 0.25 means that the outcome is at least 20 with a probability of at least 0.25,
- the value 15 in the row corresponding to 0.65 means that the outcome is at least 15 with a probability of at least 0.65,
- the value 0 in the row corresponding to 0.75 means that the outcome is at least 0 with a probability of at least 0.75.

Table 5

Acts, values of function $\rho(A_i,\pi)$ and assignment to classes from ***Cl***

	A_1	A_2	A_3	A_4	A_5	A_6
0.25	30	20	20	20	20	20
0.35	10	20	20	20	20	20
0.40	10	20	20	20	20	20
0.60	10	20	15	15	20	20
0.65	10	20	15	15	20	20
0.75	10	20	0	15	10	20
1	10	0	0	0	10	10
d	*good*	*medium*	*medium*	*bad*	*medium*	*good*

If we consider the row corresponding to 0.65, then

- the value 10 relative to A_1, means that by act A_1 the outcome is at least 10 with a probability of at least 0.65,
- the value 20 relative to A_2, means that by act A_2 the outcome is at least 20 with a probability of at least 0.65, and so on.

Applying DRSA, we approximate the following *upward* and *downward unions* of classes:

- $Cl_2^{\geq} = Cl_2 \cup Cl_3$, i.e. the set of the acts at least *medium*,
- $Cl_3^{\geq} = Cl_3$, i.e. the set of the acts (at least) *good*,

– $Cl_1^{\le} = Cl_1$, i.e. the set of the acts (at most) *bad*,

– $Cl_2^{\le} = Cl_1 \cup Cl_2$, i.e. the set of the acts at most *medium*.

The first result of the DRSA is a discovery that the data table (Table 5) is not consistent. Indeed, Table 5 shows that act A_4 stochastically dominates act A_3, however act A_3 is assigned to a better class (*medium*) than act A_4 (*bad*). Therefore, act A_3 cannot be assigned without doubt to the class of at least *medium* acts as well as act A_4 cannot be assigned without doubt to the class of (at most) *bad* acts. In consequence, lower approximation and upper approximation of $Cl_2^{\ge}$, $Cl_3^{\ge}$ and $Cl_1^{\le}$, $Cl_2^{\le}$ are equal, respectively, to

– $\underline{C}\left(Cl_2^{\ge}\right) = \{A_1,A_2,A_5,A_6\} = Cl_2^{\ge} - \{A_3\}$,
$\overline{C}\left(Cl_2^{\ge}\right) = \{A_1,A_2,A_3,A_4,A_5,A_6\} = Cl_2^{\ge} \cup \{A_4\}$,

– $\underline{C}\left(Cl_3^{\ge}\right) = \{A_1,A_6\} = Cl_3^{\ge}$, $\overline{C}\left(Cl_3^{\ge}\right) = \{A_1,A_6\} = Cl_3^{\ge}$,

– $\underline{C}\left(Cl_1^{\le}\right) = \varnothing = Cl_1^{\le} - \{A_4\}$, $\overline{C}\left(Cl_1^{\le}\right) = \{A_3,A_4\} = Cl_1^{\le} \cup \{A_3\}$,

– $\underline{C}\left(Cl_2^{\le}\right) = \{A_2,A_3,A_4,A_5\} = Cl_2^{\le}$, $\overline{C}\left(Cl_2^{\le}\right) = \{A_2,A_3,A_4,A_5\} = Cl_2^{\le}$.

Since there are two inconsistent acts on a total of six acts (A_3,A_4), then the quality of approximation (quality of classification) is equal to 4/6.

The second discovery is one reduct of condition attributes (criteria) ensuring the same quality of classification as the whole set $\boldsymbol{\Pi}$ of probabilities: $RED_{Cl} = \{0.25, 0.75, 1\}$. This means that we can explain the preferences of the DM using the probabilities in RED_{Cl} only. RED_{Cl} is also the core because no probability value can be removed from RED_{Cl} without deteriorating the quality of classification.

The third discovery gives sets of decision rules describing the DM's preferences. Below, we are presenting one of minimal sets of decision rules covering all the acts [*within brackets there is a verbal interpretation of the corresponding decision rule*] (within parentheses there are acts supporting the corresponding rule):

1) *if* $\rho(A_i,0.25) \ge 30$, *then* $A_i \in Cl_3^{\ge}$,
[*if the probability of gaining at least 30 is at least 0.25, then act A_i is (at least) good*] (A_1),

2) *if* $\rho(A_i,0.75) \ge 20$ *and* $\rho(A_i,1) \ge 10$, *then* $A_i \in Cl_3^{\ge}$,
[*if the probability of gaining at least 20 is at least 0.75 and the probability of gaining at least 10 is (at least) 1 (i.e. for sure the gain is at least 10), then act A_i is (at least) good*] (A_6),

3) *if* $\rho(A_i,1) \ge 10$, *then* $A_i \in Cl_2^{\ge}$,
[*if the probability of gaining at least 10 is (at least) 1 (i.e. for sure the gain is at least 10), then act A_i is at least medium*] (A_1, A_5, A_6),

4) *if* $\rho(A_i,0.75)\geq 20$, *then* $A_i \in Cl_2^{\geq}$,

[*if the probability of gaining at least 20 is at least 0.75, then act* A_i *is at least medium*] (A_2, A_6),

5) *if* $\rho(A_i,0.25)\leq 20$ (*i.e.* $\rho'(A_i,1)\geq 20$) *and* $\rho(A_i,0.75)\leq 15$ (*i.e.* $\rho'(A_i,0.35)\geq 15$), *then* $A_i \in Cl_2^{\leq}$,

[*if the probability of gaining at most 20 is (at least) 1 (i.e. for sure the gain is at most 20) and the probability of gaining at most 15 is at least 0.35, then act* A_i *is at most medium*] (A_3, A_4, A_5),

6) *if* $\rho(A_i,1)\leq 0$ *(i.e.* $\rho'(A_i,0.25)\geq 0$*), then* $A_i \in Cl_1 \cup Cl_2$,

[*if the probability of gaining at most 0 is at least 0.25, then act* A_i *is at most medium*] (A_2, A_3, A_4),

7) *if* $\rho(A_i,1)\geq 0$ *and* $\rho(A_i,1)\leq 0$ (i.e. $\rho(A_i,1)=0$) *and* $\rho(A_i,0.75)\leq 15$ (*i.e.* $\rho'(A_i,0.35)\geq 10$), *then* $A_i \in Cl_1 \cup Cl_2$,

[*if the probability of gaining at least 0 is 1 (i.e. for sure the gain is at least 0) and the probability of gaining at most 15 is at least 0.35, then act* A_i *is bad or medium, with no possibility of assigning* A_i *to only one of the two classes because of ambiguous knowledge*] (A_3, A_4).

Minimal sets of decision rules represent the most concise and non-redundant knowledge contained in Table 4 (and, consequently, in Table 5). The above minimal set of 7 decision rules uses 3 attributes (probabilities 0.25, 0.75 and 1) and 11 elementary conditions, i.e. 26% of descriptors from the original data table (Table 5). For larger sets of exemplary acts, the representation in terms of decision rules is even more synthetic (the percentage of descriptors from the original data table is smaller).

Let us observe that we considered an additive probability distribution, however, an extension to non-additive probability, and even to a qualitative ordinal probability, is straightforward. If the elements of set $\boldsymbol{\Pi}$ are numerous (like in real applications), a subset $\boldsymbol{\Pi'} \subset \boldsymbol{\Pi}$ of the most significant probability values (e.g. 0, 0.1, 0.2, …, 0.9, 1) can be considered.

5. COMPARISON OF DRSA WITH OTHER DECISION SUPORT PARADIGMS

DRSA aims to give an effective answer to the central problem of any decision-aiding methodology concerning multiple criteria and/or multiple attribute classification, that is the aggregation of the multiple criteria and attributes into a single preference model. In this section, we propose to compare different para-

digms used to solve this central problem by different theories (see Table 6). In [17, 42], this comparison was made at the level of axiomatic foundations, which has no precedence in the theoretical research concerning multiple criteria classification. The axiomatic approach is interesting for at least three reasons:
- it exhibits differences between preference models and methods,
- it permits to interpret methods conceived for one model in terms of another model,
- knowing the basic axioms, one can pass from one method to another with different preference models.

Table 6

Different paradigms of aggregation and preference representation

Theory (paradigm)	**Main preoccupation (axiomatic basis)**	**The aggregation result evidences**
Social Choice Theory	Voting system or aggregation of rankings	Final ranking
Decision Theory	Definition of preference structures	Relation in A
Measurement Theory	Cancellation property	Function, like in conjoint measurement
Measure Theory, Fuzzy Sets	Capacity or fuzzy measure	Weights or interaction between criteria, like in Choquet integral or Sugeno integral
Machine Learning, Logical Analysis of Data, Rough Sets	Boolean or pseudo-Boolean function, decision rules or decision trees	Knowledge, like in knowledge discovery or data mining

Moreover, in [17, 42], we have considered aggregation of ordinal criteria that has been studied much less than that of cardinal criteria. Among several aggregation models, a particular interest has been paid recently for an integral proposed by Sugeno, able to deal with ordinal data; it has been considered the most general ordinal aggregation function of the max-min average type. It appears, however, that this function has some unpleasant limitations: the most important is the so-called commensurability, i.e. the evaluations with respect to each considered criterion should be defined on the same scale. Comparison of the Sugeno integral with the decision rule model at the axiomatic level permits to show other limitation of the former.

Below, we present the main results concerning the comparison of axiomatic foundations of the decision rule model and two traditional models: utility function and outranking relation.

5.1. Axiomatic foundations of multiple criteria classification problems and associated preference models

In this point we consider a finite or denumerable *product space* $X = \prod_{i=1}^{n} X_i$, where X_i is an evaluation scale of *criterion* $i = 1,\ldots,n$. With appropriate topological conditions we can also work with infinite non-denumerable space, but in this paper, for the sake of simplicity, we will skip this possibility.

The following result is a representation theorem for the multiple criteria classification problem, stating equivalence between a very simple cancellation property, a general utility function, a very general outranking relation and a set of decision rules. Let us mention that equivalence of the considered cancellation property and the utility function was already noted by Goldstein (1991), within the conjoint measurement approach, for the special case of three decision classes.

Theorem 1 [17]. The following four propositions are equivalent:

1) (*cancellation property*) for each $i = 1,\ldots,n$, for each $x_i,y_i \in X_i$ and $a_{-i},b_{-i} \in X_{-i}$, and for each $r,s \in \{1,\ldots,m\}$:

$$\{(x_i a_{-i}) \in Cl_r \text{ and } (y_i b_{-i}) \in Cl_s\} \Rightarrow \{(y_i a_{-i}) \in Cl_r^{\geq} \text{ or } (x_i b_{-i}) \in Cl_s^{\geq}\}$$

2) (*utility function*) there exist:
- functions g_i: $X_i \rightarrow \boldsymbol{R}$ for each $i = 1,\ldots,n$, called criteria,
- function f: $\boldsymbol{R}^n \rightarrow \boldsymbol{R}$, non-decreasing in each argument, called utility function,
- m-1 ordered thresholds z_t, $t = 2,\ldots,m$, satisfying

$$z_2 < z_3 < \ldots. < z_m$$

such that for each $x \in X$ and each $t = 2,\ldots,m$

$$f[g_1(x_1), g_2(x_2),\ldots, g_n(x_n)] \geq z_t \Leftrightarrow x \in Cl_t^{\geq}$$

3) (*outranking function and relation*) there exist
- functions g_i: $X_i \rightarrow \boldsymbol{R}$, $i = 1,\ldots,n$, called criteria,
- function s: $\boldsymbol{R}^{2n} \rightarrow \boldsymbol{R}$, non-decreasing in each odd argument and non-increasing in each even argument, called outranking function,
- m-1 reference profiles p^t, $t = 2,\ldots,m$, satisfying

$$g_i(p^2) \leq g_i(p^3) \leq \ldots \leq g_i(p^m), \text{ for } i = 1,\ldots,n$$

such that for each $x \in X$ and each $t = 2,\ldots,m$

$$s[g_1(x_1), g_1(p^t), g_2(x_2), g_2(p^t),\ldots, g_n(x_n), g_n(p^t)] \geq 0 \Leftrightarrow x \in Cl_t^{\geq}$$

(N.B. $s[g_1(x_1), g_1(p^t), g_2(x_2), g_2(p^t),\ldots, g_n(x_n), g_n(p^t)] \geq 0 \Leftrightarrow x\,S\,p^t$,
where S is a binary outranking relation),

4) ("*at least*" *decision rules*) there exist:

- functions g_i: $X_i \rightarrow \boldsymbol{R}$ for each $i = 1,\ldots,n$, called criteria,
- a set of "at least" decision rules whose syntax is

$$\textit{if } g_{i1}(x_{i1}) \geq r_{i1} \textit{ and } g_{i2}(x_{i2}) \geq r_{i2} \textit{ and } \ldots \textit{ and } g_{ih}(x_{ih}) \geq r_{ih}, \textit{ then } x \in Cl_t^{\geq},$$

$$\text{with } \{i1,i2,\ldots,ih\} \subseteq \{1,\ldots,n\},\ t = 2,\ldots,m,$$

such that for each $y \in Cl_t$, $t = 2,\ldots,m$, there is at least one rule implying $y \in Cl_t^{\geq}$ and there is no rule implying $y \in Cl_r^{\geq}$, with $r>t$.

Let us remark that the above representation theorem for multiple criteria classification problem starts with a very weak axiomatic condition called cancellation property. Indeed, this property does not require existence of criterion functions g_i, $i = 1,\ldots,n$, or a dominance relation D on X in order to characterize the three preference models. Instead, the meaning of the above cancellation property is the following. Let us consider the binary large preference relation $\succeq_i$ defined on X_i, $i = 1,\ldots,n$, as follows: for each for each $x_i, y_i \in X_i$, for each for each $a_{-i} \in X_{-i}$ and for each $Cl_r \in \boldsymbol{Cl}$:

$$x_i \succeq_i y_i \Leftrightarrow [(y_i a_{-i}) \in Cl_r \Rightarrow (y_i a_{-i}) \in Cl_r^{\geq}]$$

Cancellation property ensures that the binary large preference relation $\succeq_i$ on X_i is a complete preorder, that is strongly complete (for each $x_i, y_i \in X_i$, $x_i \succeq_i y_i$ or $y_i \succeq_i x_i$) and transitive. Consequently, there exists a function g_i: $X_i \rightarrow \boldsymbol{R}$ such that for each $x_i, y_i \in X_i$

$$x_i \succeq_i y_i \Leftrightarrow g_i(x_i) \geq g_i(y_i)$$

On the basis of relations $\succeq_i$, $i = 1,...,n$, one can also define a dominance relation D on X as follows: for each $x,y \in X$

$$xDy \Leftrightarrow x_i \succeq_i y_i \text{ for all } i = 1,...,n$$

This is of course equivalent to

$$xDy \Leftrightarrow g_i(x_i) \geq g_i(y_i) \text{ for all } i = 1,...,n$$

Cancellation property 1) of Theorem 1 permits to state the following *condition of coherence between dominance relation D and classification* ***Cl***, for each $x,y \in X$

$$xDy \Rightarrow x \in Cl_r \text{ and } y \in Cl_s, \text{ with } r \geq s$$

For any subset of criteria $P \subseteq \{1,\ldots,n\}$ and for each pair $x,y \in X$ one can also define a dominance relation D_P on X:

$$xD_P y \Leftrightarrow x_i \succeq_i y_i \text{ for all } i \in P$$

which is equivalent to

$$xD_Py \Leftrightarrow g_i(x_i) \geq g_i(y_i) \text{ for all } i\in P$$

Dominance relations D_P, $P\subseteq\{1,\ldots,n\}$, are used in the condition part of decision rules. Being an intersection of complete preorders, binary relations D_P are partial preorders, i.e. they are reflexive and transitive.

Observe, moreover, that Theorem 1 regards a representation of classification ***Cl*** in terms of "lower bounds". Theorem 1 can be reformulated in terms of "upper bounds" in such a way that:

– condition of proposition 2) is expressed as

$$f[g_1(x_1), g_2(x_2),\ldots, g_n(x_n)] \leq w_t \Leftrightarrow x\in Cl_t^{\leq},$$

where w_t, $t = 1,\ldots,m$-1, are m-1 suitably ordered thresholds,

– condition of proposition 3) is expressed as

$$s[g_1(x_1), g_1(q^t), g_2(x_2), g_2(q^t),\ldots, g_n(x_n), g_n(q^t)] < 0 \Leftrightarrow x\in Cl_t^{\leq}$$

where q^t, $t = 1,\ldots,m$-1, are m-1 reference profiles q^t, such that q^{t+1} dominates q^t (i.e. q^{t+1} is at least as good as q^t with respect to each criterion $i = 1,\ldots,n$, and there is at least one criterion $j\in\{1,\ldots,n\}$ for which q^{t+1} is strictly preferred to q^t), $t = 1,\ldots,m$-2,

– condition of proposition 4) considers a set of decision rules whose syntax is

$$\textit{if } g_{i1}(x_{i1})\leq r_{i1} \textit{ and } g_{i2}(x_{i2})\leq r_{i2} \textit{ and } \ldots \textit{ and } g_{ih}(x_{ih})\leq r_{ih}, \textit{ then } x\in Cl_r^{\leq}$$

with $\{i1,i2,\ldots,ih\}\subseteq\{1,\ldots,n\}$. These decision rules are called "*at most*" decision rules. The classification of $x\in X$ with "at most" decision rules is done according to the following procedure:

- $x\in Cl_t$ if and only if there exists a rule matching x that assigns x to $Cl_t^{\leq}$, and there exists no rule matching x that assigns x to $Cl_s^{\leq}$, where $s<t$;
- $x\in Cl_m$ if and only if there exists no rule matching x.

6.2. Conjoint measurement for multiple criteria classification problems with inconsistencies

The conjoint measurement model presented in point 6.1 cannot represent the inconsistency with the dominance principle considered within the DRSA. In this point we present a more general model of conjoint measurement that permits representation of this inconsistency. This model is based on the concepts of dominance-based rough approximation of upward and downward unions of classes $Cl_t^{\geq}$ and $Cl_t^{\leq}$.

The following concepts will be useful: for each $x\in X$, the *lower class* and the *upper class* of x, denoted by $r_*(x)$ and $r^*(x)$, respectively, are defined as follows

$$r_*(x) = \max\{s\in\{1,..,m\}: x\in \underline{C}\left(Cl_s^{\geq}\right)\}$$

$$r^*(x) = \min\{s \in \{1,..,m\} : x \in \underline{C}(Cl_s^{\leq})\}$$

where $\underline{C}(Cl_s^{\geq})$ and $\underline{C}(Cl_s^{\leq})$ are the lower approximations of $Cl_s^{\geq}$ and $Cl_s^{\leq}$, respectively, with respect to set of criteria $C = \{g_1, g_2, \ldots, g_n\}$.

Theorem 2 [17]. For each set of binary relations $\succeq_i$, $i = 1,\ldots,n$, being complete preorders, and for each classification ***Cl*** there exist

- functions g_i: $X_i \rightarrow \boldsymbol{R}$, such that $x_i \succeq_i y_i \Leftrightarrow g_i(x_i) \geq g_i(y_i)$, $i = 1,\ldots,n$,
- functions $f^{\geq}$: $\boldsymbol{R}^n \rightarrow \boldsymbol{R}$ and $f^{\leq}$: $\boldsymbol{R}^n \rightarrow \boldsymbol{R}$, non-decreasing in each argument, such that
- $f^{\geq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] \leq f^{\leq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)]$
- m-1 ordered thresholds z_t, t=2,…,m,

$$z_2 < z_3 < \ldots < z_m$$

such that for each object $x \in X$, functions $f^{\geq}$ and $f^{\leq}$ assign x to a lower and an upper class, respectively:

$$f^{\geq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] \geq z_t \Leftrightarrow x \in \underline{C}(Cl_t^{\geq})$$

$$f^{\leq}[g_1(x_1), g_2(x_2), \ldots, g_n(x_n)] < z_t \Leftrightarrow x \in \underline{C}(Cl_{t-1}^{\leq})$$

where $t = 2,\ldots,m$, $C = \{g_1, g_2, \ldots, g_n\}$

Inconsistency with the dominance principle can also be represented in terms of a set of "*at least*" and "*at most*" decision rules considered together. More formally, a set of "*at least*" and "*at most*" decision rules does not contradict the classification ***Cl*** if for each $x \in Cl_t$ there exists no "*at least*" decision rule for which $x \in Cl_s^{\geq}$, with $s>t$, and there exists no "*at most*" decision rule for which $x \in Cl_s^{\leq}$, with $s<t$. A set of decision rules is complete if for each $x \in \underline{C}(Cl_t^{\geq})$ there exists a decision rule for which $x \in Cl_s^{\geq}$, with $s \geq t$, and for each $x \in \underline{C}(Cl_t^{\leq})$ there exists a decision rule for which $x \in Cl_s^{\leq}$, with $s \leq t$. A set of decision rules represents the classification ***Cl*** if it does not contradict ***Cl*** and it is complete.

Theorem 3 [17]. For each set of binary relations $\succeq_i$, $i = 1,\ldots,n$, being complete preorders, and for each classification ***Cl***, there exists a set of decision rules representing the classification ***Cl***.

The advantage of the DRSA with respect to competitive methodologies is the possibility of handling partially inconsistent data that are often encountered in preferential information, due to hesitation of decision makers, unstable character of their preferences, imprecise or incomplete information and the like. Therefore, we proposed a general model of conjoint measurement that, using the basic concepts of DRSA (lower and upper approximations), is able to represent these inconsistencies by a specific utility function. We showed that these inconsistencies can also be represented in a meaningful way by "*if..., then...*" decision rules induced from rough approximations.

As DRSA to multiple-criteria classification problems and the underlying decision rules exploit only the ordinal properties of the scales of criteria, they are appropriate for aggregation of ordinal criteria. This challenging problem of multiple-criteria decision making has been solved until now by using some 'max-min' aggregation functions, with the most general one – the fuzzy integral proposed by Sugeno. In [17, 42], we have shown that the decision rule model following from DRSA has advantages over the integral of Sugeno, in particular, it can represent some (even consistent) preferences that the Sugeno integral cannot.

The characterization of the decision rule preference model given in this section shows clearly its extraordinary capacity of criteria aggregation in multiple criteria classification problems. The decision rule preference model, apart from its capacity of representation, fulfils the postulate of transparency and interpretability of preference models in decision aiding. The characterization shows that the decision rule preference model is a strong alternative to functional and relational preference models to which it is formally equivalent. Recently, similar benefits of the decision rule model have been proved with respect to multiple criteria choice and ranking problems [16].

CONCLUSIONS AND PROMISING AREAS

We presented a knowledge discovery paradigm for multiple attribute and multiple criteria 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 in searching for patterns in the data that facilitate an understanding of the decision maker's preferences and that permit to recommend a decision concordant with these preferences. An original component of this paradigm is taking into account prior knowledge about preference semantics in patterns 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 inconsistent with the dominance principle. These deficien-

cies are repaired in DRSA based on the concept of rough approximations consistent with the dominance principle. DRSA permits, moreover, to apply rough set approach to some new fields, like multiple criteria decision making and decision under uncertainty. Many extensions proposed for DRSA make of this approach
a useful tool for practical applications. Let us mention below the most important ones:

- DRSA with missing values of attributes and criteria [9],
- fuzzy set extensions of DRSA [4, 10, 19, 46],
- DRSA for hierarchical decision making [1],
- induction of association rules from preference-ordered data sets [23].

DRSA gives, moreover, a methodology for building a preference model of a decision maker in terms of decision rules. The decision rules have a special syntax involving partial evaluation profiles and dominance relation on these profiles. The clarity of the rule representation of preferences permits to see the limits of other traditional aggregation functions: utility function and outranking relation. We proposed an axiomatic characterization of these aggregation functions in terms of conjoint measurement and in terms of a set of decision rules. The axioms of the "cancellation property" type are the weakest possible. In comparison to other studies on characterization of aggregation functions, our axioms do not require any preliminary assumption about the scales of criteria. A side-result of these investigations is the corollary that the decision rule aggregation (preference model) is the most general among the known aggregation functions.

The application of DRSA to analysis of data representing a preferential information supplies, moreover, some useful elements of knowledge about the decision situation; these are: the relevance of attributes and/or criteria, information about their interaction (from quality of approximation and its analysis using fuzzy measures theory), minimal subsets of attributes or criteria (reducts) conveying the relevant knowledge contained in the exemplary decisions, the set of the non-reducible attributes or criteria (core). Moreover, DRSA permits to handle heterogeneous information: qualitative and quantitative, preference-ordered or not, crisp and fuzzy, ordinal and cardinal, partially missing and inconsistent. Finally, the proposed methodology is based on elementary concepts and mathematical tools (sets and set operations, binary relations), without recourse to any algebraic or analytical structures; the main idea is very natural and even objective, in a certain sense, like the dominance relation is.

Due to the above features, DRSA contributes in a very promising way to many different areas, like:

- **knowledge discovery and data mining**, where without DRSA the preference order in data is ignored,
- **multiple criteria decision analysis**, for which DRSA is offering a natural, general and intelligible way of modeling DM's preferences in terms of "*if…, then…*" decision rules,
- **decision under risk**, where DRSA handles non-additive probability distributions and even qualitative ordinal distributions over possible states of the world, and offers a decision rule representation of DM's preferences,
- **approximate reasoning** based on fuzzy-rough modus ponens and gradual rules induced from fuzzy rough approximations,
- **fuzzy-rough control** involving gradual rules.

The DRSA 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 [16, 17, 20, 42] 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.

An interesting research problem concerns measuring attractiveness of decision rules taking into account three application perspectives: (i) knowledge representation, (ii) prediction of new decisions and (iii) interventions based on discovered rules in some other universe (see [44]). In order to choose attractiveness measures concordant with the above perspectives we analyzed semantics of particular measures which led us to a conclusion that the best suited measures for the above applications are: (i) support and certainty, (ii) a Bayesian confirmation measure [25], and (iii) two measures related to efficiency of intervention [5], respectively. These five measures induce a partial order in the set of rules giving a starting point for an interactive browsing procedure. For building a strategy of intervention, we proposed rules discovered using the DRSA – the "at least" type rules indicate opportunities for improving assignment of objects, and the "at most" type rules indicate threats for deteriorating assignment of objects.

REFERENCES

1. Dembczyński K., Greco S., Słowiński R.: Methodology of rough-set-based classification and sorting with hierarchical structure of attributesand criteria. “Control & Cybernetics” 2002, 31, pp. 891-920.
2. Flinkman M., Michalowski W., Nilsson S., Słowiński R., Susmaga R., Wilk Sz.: Use of rough sets analysis to classify Siberian forest ecosystem according to net primary production of phytomass. “INFOR” 2000, 38, No. 3, pp. 145-161.
3. Giove S. Greco S. Matarazzo B. Słowiński R.: Variable consistency monotonic decision trees. In: Rough Sets and Current Trends in Computing. Eds. J.J. Alpigini, J.F. Peters, A. Skowron, N. Zhong. LNAI 2475, Springer-Verlag, Berlin 2002, pp. 247-254.
4. Greco, S., Inuiguchi, M., Słowiński, R.: A new proposal for fuzzy rough approximations and gradual decision rule representation. In: Transactions in Rough Sets II, LNCS 3135, Springer-Verlag, Berlin 2004, pp. 332-355.
5. Greco S., Matarazzo B., Pappalardo N., Słowiński R.: Measuring expected effects of interventions based on decision rules. “Journal of Experimental and Applied Artificial Intelligence” 2005, 17, No.1-2, pp. 103-118.
6. Greco S., Matarazzo B., Słowiński R.: A new rough set approach to evaluation of bankruptcy risk. In: Operational Tools in the Management of Financial Risk. Ed. C. Zopounidis. Kluwer, Dordrecht 1998, pp. 121-136.
7. Greco S., Matarazzo, B., Słowiński, R.: Rough approximation of a preference relation by dominance relations. “European J. Operational Research” 1999, 117, pp. 63-83.
8. Greco S., Matarazzo B., Słowiński R.: The use of rough sets and fuzzy sets in MCDM. Chapter 14. In: Advances in Multiple Criteria Decision Making. Eds. Gal T., Stewart T., Hanne T. Kluwer, Dordrecht 1999, pp. 14.1-14.59.
9. Greco S., Matarazzo B., Słowiński R.: Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems. In: Decision Making: Recent Developments and Worldwide Applications. Eds. Zanakis S.H., Doukidis G., Zopounidis C. Kluwer, Dordrecht 2000, pp. 295-316.
10. Greco S., Matarazzo B., Słowiński R.: Fuzzy extension of the rough set approach to multicriteria and multiattribute sorting. In: Preferences and Decisions under Incomplete Knowledge. Eds. Fodor J., De Baets B., Perny P. Physica-Verlag, Heidelberg, 2000, pp.131-151
11. Greco S., Matarazzo B., Słowiński R.: Rough sets theory for multicriteria decision analysis. “European J. Operational Research” 2001, 129, pp. 1-47.
12. Greco S., Matarazzo B., Słowiński R.: Assessment of a value of information using rough sets and fuzzy measures. In: Fuzzy Sets and their Applications. Eds. Chojcan J., Łęski J. Silesian University of Technology Press, Gliwice 2001, pp. 185-193.

13. Greco S., Matarazzo B., Słowiński R.: Rough set approach to decisions under risk. In: Rough Sets and Current Trends in Computing. Eds. Ziarko W., Yao Y. LNAI 2005, Springer-Verlag, Berlin 2001, pp. 160-169.
14. Greco S., Matarazzo B., Słowiński R.: Rough sets methodology for sorting problems in presence of multiple attributes and criteria. "European J. Operational Research" 2002, 138, pp 247-259.
15. Greco S., Matarazzo B., Słowiński R.: Multicriteria classification. In: Handbook of Data Mining and Knowledge Discovery. Eds. Kloesgen W., Zytkow J. Oxford University Press, New York 2002, chapter 16.1.9, pp. 318-328.
16. Greco S, Matarazzo B, Słowiński R: Preference representation by means of conjoint measurement & decision rule model. In: Aiding Decisions with Multiple Criteria–Essays in Honor of Bernard Roy. Eds. Bouyssou D., Jacquet-Lagrèze E., Perny P., Słowiński R., Vanderpooten D., Vincke P. Kluwer, Dordrecht, 2002, pp. 263-313
17. Greco S., Matarazzo B., Słowiński R.: 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" 2004, 158, pp. 271-292.
18. Greco S., Matarazzo B., Słowiński R.: Dominance-Based Rough Set Approach to Knowledge Discovery (I) – General Perspective. Chapter 20. In: Zhong N., Liu J.: Intelligent Technologies for Information Analysis. Springer-Verlag, Berlin 2004, pp. 513-552.
19. Greco S., Matarazzo B., Słowiński R.: Dominance-Based Rough Set Approach to Knowledge Discovery (II) – Extensions and Applications. Chapter 21. In: Zhong N., Liu J. Intelligent Technologies for Information Analysis. Springer-Verlag, Berlin 2004, pp. 553-612.
20. Greco, S., Matarazzo, B., Słowiński, R.: Decision rule approach. Chapter 13 [in]: J.Figueira, S.Greco and M.Ehrgott (eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer-Verlag, New York, 2005, pp. 507-562
21. Greco S., Matarazzo B., Słowiński R., Stefanowski J.: Variable consistency model of dominance-based rough set approach. In: Ziarko W., Yao Y.: Rough Sets and Current Trends in Computing. LNAI 2005, Springer-Verlag, Berlin 2001, pp. 170-181
22. Greco S., Matarazzo B., Słowiński R., Stefanowski J.: An algorithm for induction of decision rules consistent with dominance principle. In: Rough Sets & Current Trends in Computing. Eds. Ziarko W., Yao Y. LNAI 2005, Springer-Verlag, Berlin 2001, pp. 304-313.
23. Greco S., Matarazzo B., Słowiński R., Stefanowski J.: Mining association rules in preference-ordered data. In: Foundations of Intelligent Systems. Eds. Hacid M.-S., Ras Z.W., Zighed D.A., Kodratoff Y. LNAI 2366, Springer-Verlag, Berlin 2002, pp. 442-450.
24. Greco S., Matarazzo B., Słowiński R., Tsoukias A.: Exploitation of a rough approximation of the outranking relation in multicriteria choice and ranking. [In]: Trends in Multicriteria Decision Making. Eds. Stewart T.J., van den Honert R.C. LNEMS 465, Springer-Verlag, Berlin 1998, pp. 45-60.

25. Greco S., Pawlak Z., Słowiński R.: Can Bayesian confirmation measures be useful for rough set decision rules? "Engineering Applications of Artificial Intelligence" 2004, 17, No 4, pp. 345-361.
26. Grzymala-Busse J.W.: A new version of the rule induction system LERS. "Fundamenta Informaticae" 1997, 31, pp. 27-39.
27. Michalowski W., Rubin S., Słowiński R., Wilk Sz.: Mobile clinical support system for pediatric emergencies. "Journal of Decision Support Systems" 2003, 36, pp. 161-176.
28. Michalski R.S., Bratko I. Kubat M.: Machine Learning and Data Mining – Methods and Applications. Eds. Wiley, New York 1998.
29. Pawlak Z.: Rough sets. Int. J. Information & Computer Sciences, 1982, 11, pp. 341-356.
30. Pawlak Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer, Dordrecht 1991.
31. Pawlak Z., Grzymala-Busse J.W., Słowiński R., Ziarko W.: Rough sets. „Communications of the ACM" 1995, 38, pp. 89-95.
32. Pawlak Z., Słowiński K., Słowiński R.: Rough classification of patients after highly selective vagotomy for duodenal ulcer. "International Journal of Man-Machine Studies" 1986, 24, pp. 413-433.
33. Pawlak Z., Słowiński R.: Rough set approach to multi-attribute decision analysis. "European J. Operational Research" 1994, 72, pp. 443-459.
34. Polkowski L.: Rough Sets: Mathematical Foundations. Physica-Verlag, Heidelberg 2002.
35. Polkowski L., Skowron A.: Calculi of granules based on rough set theory: approximate distributed synthesis and granular semantics for computing with words. In: New Directions in Rough sets, Data Mining and Soft-Granular Computing. Eds. Zhong N., Skowron A., Ohsuga S. LNAI 1711, Springer-Verlag, Berlin 1999, pp. 20-28.
36. Rossi L., Słowiński R., Susmaga R.: Rough set approach to evaluation of stormwater pollution. "International Journal of Environment and Pollution" 1999, 12, No 2/3, pp. 232-250.
37. Roy B.: Multicriteria Methodology for Decision Aiding. Kluwer, Dordrecht 1996.
38. Słowiński R.: Rough set learning of preferential attitude in multi-criteria decision making. In: Methodologies for Intelligent Systems. Eds. Komorowski J., Ras Z.W. LNAI 689, Springer-Verlag, Berlin 1993, pp. 642-651.
39. Słowiński R.: Rough set approach to decision analysis. "AI Expert Magazine" 1995, 10, No 3, pp. 18-25.
40. Słowiński R., Greco S., Matarazzo B.: Rough set analysis of preference-ordered data. In: Rough Sets and Current Trends in Computing. Eds. Alpigini J.J., Peters J.F., Skowron A., Zhong N. LNAI 2475, Springer-Verlag, Berlin 2002, pp. 44-59.
41. Słowiński R., Greco S., Matarazzo B.: Mining decision-rule preference model from rough approximation of preference relation. In: Proc. 26th IEEE Annual Int. Conf. on Computer Software & Applications. Oxford 2002, pp. 1129-1134.

42. Słowiński R., Greco S., Matarazzo B.: Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle. “Control and Cybernetics” 2002, 31, pp. 1005-1035.
43. Słowiński R., Greco S.: Inducing robust decision rules from rough approximations of a preference relation. In: Artificial Intelligence and Soft Computing. Eds. Rutkowski L., Siekmann J., Tadeusiewicz R., Zadeh L.A. LNAI 3070, Springer-Verlag, Berlin 2004, pp. 118-132.
44. Słowiński R., Greco S.: Measuring attractiveness of rules from the viewpoint of knowledge representation, prediction and efficiency of intervention. Keynote lecture. In: Proc. 3rd Atlantic Web Intelligence Conference, AWIC’05, Łódź, June 6-9, 2005.
45. Słowiński R., Greco S., Matarazzo B.: Rough set based decision support. Chapter 16. In: Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques. Eds. Burke E.K., Kendall G. Springer-Verlag, New York 2005, pp. 475-527.
46. Słowiński R., Stefanowski J.: Rough set reasoning about uncertain data. “Fundamenta Informaticae” 1996, 27, pp. 229-243.
47. Słowiński R., Vanderpooten D.: A generalised definition of rough approximations. “IEEE Transactions on Data and Knowledge Engineering” 2000, 12, pp. 331-336.
48. Słowiński R., Zopounidis C.: Application of the rough set approach to evaluation of bankruptcy risk. “Intelligent Systems in Accounting, Finance & Management” 1995, 4, pp. 27-41.
49. Stefanowski J., Vanderpooten D.: Induction of decision rules in classification and discovery-oriented perspectives. “International J. Intelligent Systems” 2001, 16, No 1, pp. 13-28.
50. Wilk Sz., Słowiński R., Michalowski W., Greco S.: Supporting triage of children with abdominal pain in the emergency room. “European J. Operational Research” 2005, 160, No 3, pp. 696-709.