



ELSEVIER

European Journal of Operational Research 129 (2001) 1–47

EUROPEAN
JOURNAL
OF OPERATIONAL
RESEARCH

www.elsevier.com/locate/dsw

Invited Review

Rough sets theory for multicriteria decision analysis

Salvatore Greco ^a, Benedetto Matarazzo ^a, Roman Slowinski ^{b,*}

^a *Faculty of Economics, University of Catania, Corso Italia 55, 95129 Catania, Italy*

^b *Institute of Computing Science, Poznan University of Technology, ul. Piotrowo 3a, 60-965 Poznan, Poland*

Received 1 July 1999; accepted 22 December 1999

Abstract

The original rough set approach proved to be very useful in dealing with inconsistency problems following from information granulation. It operates on a data table composed of a set U of objects (actions) described by a set Q of attributes. Its basic notions are: indiscernibility relation on U , lower and upper approximation of either a subset or a partition of U , dependence and reduction of attributes from Q , and decision rules derived from lower approximations and boundaries of subsets identified with decision classes. The original rough set idea is failing, however, when preference-orders of attribute domains (criteria) are to be taken into account. Precisely, it cannot handle inconsistencies following from violation of the dominance principle. This inconsistency is characteristic for preferential information used in multicriteria decision analysis (MCDA) problems, like sorting, choice or ranking. In order to deal with this kind of inconsistency a number of methodological changes to the original rough sets theory is necessary. The main change is the substitution of the indiscernibility relation by a dominance relation, which permits approximation of ordered sets in multicriteria sorting. To approximate preference relations in multicriteria choice and ranking problems, another change is necessary: substitution of the data table by a pairwise comparison table, where each row corresponds to a pair of objects described by binary relations on particular criteria. In all those MCDA problems, the new rough set approach ends with a set of decision rules playing the role of a comprehensive preference model. It is more general than the classical functional or relational model and it is more understandable for the users because of its natural syntax. In order to workout a recommendation in one of the MCDA problems, we propose exploitation procedures of the set of decision rules. Finally, some other recently obtained results are given: rough approximations by means of similarity relations, rough set handling of missing data, comparison of the rough set model with Sugeno and Choquet integrals, and results on equivalence of a decision rule preference model and a conjoint measurement model which is neither additive nor transitive. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Multicriteria decision analysis; Rough sets; Classification; Sorting; Choice; Ranking; Decision rules; Conjoint measurement

* Corresponding author. Tel.: +48-61-6652-375; fax: +48-61-8771-525.
E-mail address: slowinsk@sol.put.poznan.pl (R. Slowinski).

1. Introduction

The rough sets theory introduced by Pawlak (1982, 1991) has often proved to be an excellent mathematical tool for the analysis of a vague description of objects (called actions in decision problems). The adjective vague, referring to the quality of information, means inconsistency or ambiguity which follows from information granulation. The rough sets philosophy is based on the assumption that with every object of the universe there is associated a certain amount of information (data, knowledge), expressed by means of some attributes used for object description. Objects having the same description are indiscernible (similar) with respect to the available information. The *indiscernibility relation* thus generated constitutes a mathematical basis of the rough sets theory; it induces a partition of the universe into blocks of indiscernible objects, called elementary sets, that can be used to build knowledge about a real or abstract world. The use of the indiscernibility relation results in information granulation.

Any subset X of the universe may be expressed in terms of these blocks either precisely (as a union of elementary sets) or approximately only. In the latter case, the subset X may be characterized by two ordinary sets, called *lower* and *upper approximations*. A rough set is defined by means of these two approximations, which coincide in the case of an ordinary set. The lower approximation of X is composed of all the elementary sets included in X (whose elements, therefore, certainly belong to X), while the upper approximation of X consists of all the elementary sets which have a non-empty intersection with X (whose elements, therefore, may belong to X). Obviously, the difference between the upper and lower approximation constitutes the boundary region of the rough set, whose elements cannot be characterized with certainty as belonging or not to X , using the available information. The information about objects from the boundary region is, therefore, inconsistent or ambiguous. The cardinality of the boundary region states, moreover, to what extent it is possible to express X in exact terms, on the basis of the available information. For this reason, this cardinality may be used as a measure of vagueness of the information about X .

The rough sets theory, dealing with representation and processing of vague information, presents a series of intersections and complements with respect to many other theories and mathematical techniques handling imperfect information, like probability theory, evidence theory of Dempster–Shafer, fuzzy sets theory, discriminant analysis and mereology (see Dubois and Prade, 1990, 1992; Krusinska et al., 1992; Pawlak, 1985a,b; Polkowski and Skowron, 1994; Skowron and Grzymala-Busse, 1994; Slowinski, 1995).

Some important characteristics of the rough set approach make of this a particularly interesting tool in a number of problems and concrete applications. With respect to the input information, it is possible to deal with both quantitative and qualitative data, and inconsistencies need not to be removed prior to the analysis. With reference to the output information, it is possible to acquire a posteriori information regarding the relevance of particular attributes and their subsets to the quality of approximation considered in the problem at hand, without any additional inter-attribute preference information. Moreover, the final result in the form of “if..., then...” decision rules, using the most relevant attributes, is easy to interpret.

Several attempts have already been made to use the rough sets theory to decision support (Pawlak and Slowinski, 1994; Slowinski, 1993b). The original rough set approach is not able, however, to deal with preference-ordered attribute domains and decision classes. Solving this problem was crucial for application of the rough set approach to multicriteria decision analysis (MCDA). Why this application seems so important? The answer is connected with the nature of the input preferential information available in MCDA and of the output of the analysis. As to the input, the rough set approach requires a set of examples which is also convenient for acquisition of preferential information from decision makers (DMs). Very often in MCDA, this information has to be given in terms of preference model

parameters, like importance weights, substitution ratios and various thresholds. Giving such information requires a great cognitive effort of the DM. It is generally acknowledged that people prefer to make exemplary decisions than to explain them in terms of specific parameters. For this reason, the idea of inferring preference models from exemplary decisions provided by the DM is very attractive. Furthermore, the exemplary decisions may be inconsistent because of limited discriminatory power of criteria and because of hesitation of the DM (see, e.g., Roy, 1989). These inconsistencies cannot be considered as a simple error or noise. They can convey important information that should be taken into account in the construction of the DMs preference model. The rough set approach is intended to deal with inconsistency and this is another argument for its application to MCDA. Finally, the output of the analysis, i.e. the model of preferences in terms of decision rules seems very convenient for decision support because it is transparent and speaks the same language as the DM.

Let us explain shortly why the original rough set approach is not able to deal with inconsistencies coming from consideration of criteria, i.e. attributes with preference-ordered domains (scales), like product quality, market share, debt ratio. Consider, for example, two firms, *A* and *B*, evaluated for assessment of bankruptcy risk by a set of criteria including the “debt ratio” (total debt/total assets). If firm *A* has a low value while firm *B* has a high value of the debt ratio, and evaluations of these firms on other attributes are equal, then, from bankruptcy risk point of view, firm *A* dominates firm *B*. Suppose, however, that firm *A* has been assigned by a DM to a class of higher risk than firm *B*. This is obviously inconsistent with the dominance principle. Within the original rough set approach, the two firms will be considered as just discernible and no inconsistency will be stated.

For this reason, Greco et al. (1995, 1997a, 1998, 1999c,l) have proposed an extension of the rough sets theory that is able to deal with inconsistencies typical to exemplary decisions in MCDA problems. This innovation is mainly based on substitution of the indiscernibility relation by a dominance relation in the rough approximation of decision classes. An important consequence of this fact is a possibility of inferring from exemplary decisions the preference model in terms of decision rules being logical statements of the type “if..., then...” The separation of certain and doubtful knowledge about the DM’s preferences is done by distinction of different kinds of decision rules, depending whether they are induced from lower approximations of decision classes or from the boundaries of these classes composed of inconsistent examples that do not observe the dominance principle. Such preference model is more general than the classical functional models considered within Multi-Attribute Utility Theory (MAUT) or relational models considered, for example, in outranking methods.

The paper is organized as follows. In Section 2, a general view of the rough set approach is given. In Section 3, two extensions of the classical rough set approach based on generalizations of the basic concept of indiscernibility are presented: the first is the similarity relation being only reflexive and not necessarily symmetric and transitive; the second is a specific indiscernibility relation handling missing values in objects’ description – it is transitive but neither reflexive nor symmetric. In Section 4, we introduce a distinction between classification and sorting problems. The sorting problem involves preference-orders on domains of considered attributes (criteria) and among decision classes. To deal with multicriteria sorting problems rough set approximation based on dominance is proposed in this section. Furthermore, in order to handle missing values in multicriteria sorting problems a specific dominance relation is proposed. In Section 5, choice and ranking problems are considered. They are based on pairwise comparisons of objects, so the rough set approach concerns in this case approximation of a preference binary relation by specific dominance relations. These dominance relations can be multigraded, when the preferences with respect to considered criteria are cardinal, or without any degree of preference, when the preferences with respect to criteria are ordinal. Section 6 presents some results about equivalence between preference models of conjoint measurement and preference models expressed in terms of decision rules induced from rough approximations. Section 7 groups conclusions.

2. A general view of rough sets

2.1. Data table and indiscernibility relation

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

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

Therefore, each object x of U is described by a vector (string) $\text{Des}_Q(x) = [f(x, q_1), f(x, q_2), \dots, f(x, q_m)]$, called *description* of x in terms of the evaluations of the attributes from Q ; it represents the available information about x .

To every (non-empty) subset of attributes P is associated an *indiscernibility relation* on U , denoted by I_P :

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

If $(x, y) \in I_P$, it is said that the objects x and y are P -indiscernible. Clearly, the indiscernibility relation thus defined is an equivalence relation (reflexive, symmetric and transitive). The family of all the equivalence classes of the relation I_P is denoted by U/I_P and the equivalence class containing an element $x \in U$ by $I_P(x)$. The equivalence classes of the relation I_P are called P -*elementary sets*. If $P = Q$, the Q -elementary sets are called *atoms*.

2.2. Approximations

Let S be a data table, X a non-empty subset of U and $\emptyset \neq P \subseteq Q$. The P -*lower approximation* and the P -*upper approximation* of X in S are defined, respectively, by:

$$\underline{P}(X) = \{x \in U : I_P(x) \subseteq X\},$$

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

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

- The P -*boundary* of X in S , denoted by $Bn_P(X)$, is: $Bn_P(X) = \overline{P}(X) - \underline{P}(X)$.
- The following relation holds: $\underline{P}(X) \subseteq X \subseteq \overline{P}(X)$.

Therefore, if an object x belongs to $\underline{P}(X)$, it is certainly also an element of X , while if x belongs to $\overline{P}(X)$, it may belong to the set X . $Bn_P(X)$ constitutes the “doubtful region” of X : nothing can be said with certainty about the belonging of its elements to the set X .

The following relation, called *complementarity property*, is satisfied: $\underline{P}(X) = U - \overline{P}(U - X)$.

If the P -boundary of X is empty, $Bn_P(X) = \emptyset$, then the set X is an ordinary (exact) set with respect to P , that is, it may be expressed as the union of a certain number of P -elementary sets; otherwise, if $Bn_P(X) \neq \emptyset$, the set X is an approximate (rough) set with respect to P and may be characterized by means

of the approximations $\underline{P}(X)$ and $\overline{P}(X)$. The family of all the sets $X \subseteq U$ having the same P -lower and P -upper approximations is called a *rough set*.

The following ratio defines an *accuracy* of the approximation of X , $X \neq \emptyset$, by means of the attributes from P :

$$\alpha_P(X) = \frac{|\underline{P}(X)|}{|\overline{P}(X)|},$$

where $|Y|$ indicates the cardinality of a (finite) set Y . Obviously, $0 \leq \alpha_P(X) \leq 1$; if $\alpha_P(X) = 1$, X is an ordinary (exact) set with respect to P ; if $\alpha_P(X) < 1$, X is a *rough* (vague) set with respect to P .

Another ratio defines a *quality* of the approximation of X by means of the attributes from P :

$$\gamma_P(X) = \frac{|\underline{P}(X)|}{|X|}.$$

The quality $\gamma_P(X)$ represents the relative frequency of the objects correctly classified by means of the attributes from P . Moreover, $0 \leq \alpha_P(X) \leq \gamma_P(X) \leq 1$, and $\gamma_P(X) = 0$ iff $\alpha_P(X) = 0$, while $\gamma_P(X) = 1$ iff $\alpha_P(X) = 1$.

The definition of approximations of a subset $X \subseteq U$ can be extended to a classification, i.e. a partition $Y = \{Y_1, \dots, Y_n\}$ of U . Subsets $Y_i, i = 1, \dots, n$, are disjunctive classes of Y . By P -lower (P -upper) approximation of Y in S we mean sets $\underline{P}Y = \{\underline{P}Y_1, \dots, \underline{P}Y_n\}$ and $\overline{P}Y = \{\overline{P}Y_1, \dots, \overline{P}Y_n\}$, respectively. The coefficient

$$\gamma_P(Y) = \frac{\sum_{i=1}^n |\underline{P}Y_i|}{|U|}$$

is called *quality of the approximation of classification* Y by set of attributes P , or in short, *quality of classification*. It expresses the ratio of all P -correctly classified objects to all objects in the system.

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

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

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

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

$$\underline{P}(X) = \{x \in U : \mu_X^P(X) = 1\}, \quad \overline{P}(X) = \{x \in U : \mu_X^P(X) > 0\},$$

$$Bn_P(X) = \{x \in U : 0 < \mu_X^P(X) < 1\}, \quad \underline{P}(U - X) = \{x \in U : \mu_X^P(X) = 0\}.$$

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

2.3. Dependence and reduction of attributes

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

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

Another issue of great practical importance is that of “superfluous” data in a data table. Superfluous data can be eliminated, in fact, without deteriorating the information contained in the original table.

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

The set P is *independent* (orthogonal) if all its attributes are indispensable. The subset P' of P is a *reduct* of P (denotation $\text{Red}(P)$) if P' is independent and $I_{P'} = I_P$.

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

More than one Y -reduct (or reduct) of P may exist in a data table. The set containing all the indispensable attributes of P is known as the Y -core. Formally

$$\text{Core}_Y(P) = \bigcap \text{Red}_Y(P).$$

Obviously, since the Y -core is the intersection of all the Y -reducts of P , it is included in every Y -reduct of P . It is the most important subset of attributes of Q , because none of its elements can be removed without deteriorating the quality of classification.

The calculation of all the reducts is fairly complex (see Bazan et al., 1994; Kryszkiewicz and Rybinski, 1996; Skowron and Rauszer, 1992; Susmaga, 1998). Nevertheless, in many practical applications it is not necessary to calculate all the reducts, but only some of them. For example, in Slowinski et al. (1988), the following heuristic procedure has been used to obtain the “most satisfactory” reduct. Starting from single attributes, the one with the greatest quality of classification is chosen; then to the chosen attribute, another attribute is appended that gives the greatest increase to the quality of classification for the pair of attributes; then yet another attribute is appended to the pair giving the greatest increase to the quality of classification for the triple, and so on, until the maximal quality is reached by a subset of attributes. At the end of this procedure, it should be verified if the obtained subset is minimal, i.e. if elimination of any attribute from this subset keeps the quality unchanged. Then, for further analysis, it is often sufficient to take into consideration a reduced data table, where the set Q of attributes is confined to the “most satisfactory” reduct.

2.4. Decision table and decision rules

If in a data table the attributes of set Q are divided into *condition* attributes (set $C \neq \emptyset$) and *decision* attributes (set $D \neq \emptyset$), $C \cup D = Q$ and $C \cap D = \emptyset$, such a table is called a *decision table*. The decision attributes induce a partition of U deduced from the indiscernibility relation I_D in a way that is independent of the condition attributes. D -elementary sets are called *decision classes*. There is a tendency to reduce the

set C while keeping all important relationships between C and D , in order to make decisions on the basis of a smaller amount of information. When the set of condition attributes is replaced by one of its reducts, the quality of approximation of the classification induced by the decision attributes is not deteriorating.

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

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

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

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

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

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

Generation of decision rules from decision tables is a complex task and a number of procedures have been proposed to solve it (see, for example, Grzymala-Busse, 1992, 1997; Mienko et al., 1996a,b; Skowron, 1993; Skowron and Polkowski, 1997; Slowinski and Stefanowski, 1992; Stefanowski, 1998; Ziarko and Shan, 1994; Slowinski et al., 2000). The existing induction algorithms use one of the following strategies:

- (a) generation of a minimal set of rules covering all objects from a decision table;
- (b) generation of an exhaustive set of rules consisting of all possible rules for a decision table;
- (c) generation of a set of ‘strong’ decision rules, even partly discriminant, covering relatively many objects each but not necessarily all objects from the decision table.

2.5. Fuzzy measures and rough sets

From a formal point of view, the quality of classification satisfies the properties of set functions called *fuzzy measures*. As observed by Grabisch (1997), fuzzy measures constitute a useful tool for modeling the importance of coalitions. In Greco et al. (1998a), *fuzzy measures* have been used to assess a relative value of

information supplied by each attribute and to analyze the interactions among attributes, basing on the quality of classification calculated from the rough set approach. Let us explain this point in more detail.

Let $N = \{1, 2, \dots, n\}$ be a finite set, whose elements could be players in a game, criteria in a multicriteria decision problem, attributes in a data table, etc., and let $P(N)$ denote the power set of N , i.e. the set of all subsets of N . A fuzzy measure on N is a set function $\mu : P(N) \rightarrow [0, 1]$ satisfying the following axioms:

1. $\mu(\emptyset) = 0, \mu(N) = 1$,
2. $A \subseteq B$ implies $\mu(A) \leq \mu(B)$ for all $A, B \in P(N)$.

In the following, the first axiom is relaxed by considering the condition $\mu(N) \leq 1$ instead of $\mu(N) = 1$.

Within game theory, the function $\mu(A)$ is called characteristic function and represents the payoff obtained by the coalition $A \subseteq N$ in a cooperative game (Shapley, 1953; Banzhaf, 1965); in a multicriteria decision problem, $\mu(A)$ can be interpreted as the conjoint importance of the criteria from $A \subseteq N$ (Grabisch, 1994, 1996).

Some indices have been introduced in game theory as specific solutions of cooperative games. The most important are the Shapley value and the Banzhaf value. The Shapley value (Shapley, 1953) for every $i \in N$ is defined by

$$\phi_S(i) = \sum_{K \subseteq N - \{i\}} \frac{(n - |K| - 1)!|K|!}{n!} [\mu(K \cup \{i\}) - \mu(K)].$$

The Banzhaf value (Banzhaf, 1965) for every $i \in N$ is defined by

$$\phi_B(i) = \frac{1}{2^{n-1}} \sum_{K \subseteq N - \{i\}} [\mu(K \cup \{i\}) - \mu(K)].$$

The Shapley value and the Banzhaf value can be interpreted as specific kinds of weighted average contribution of element i alone to all coalitions. Let us remind that in the case of $\phi_S(i)$ the value of $\mu(N)$ is shared among the elements of N , i.e. $\sum_{i=1}^n \phi_S(i) = 1$, while an analogous property does not hold for $\phi_B(i)$.

The Shapley and Banzhaf values have also been proposed to represent the average importance of particular criteria within multicriteria decision analysis, when for the conjoint importance of criteria fuzzy measures are used (Murofushi, 1992). In addition to the indices concerning particular criteria, other indices have been proposed to measure the interaction between pairs of criteria. Interaction indices have been suggested by Murofushi and Soneda (1993) and Roubens (1996) with respect to Shapley value and Banzhaf value, respectively.

The Murofushi–Soneda interaction index for elements $i, j \in N$ is defined by

$$I_{MS}(i, j) = \sum_{K \subseteq N - \{i, j\}} \frac{(n - |K| - 2)!|K|!}{(n - 1)!} [\mu(K \cup \{i, j\}) - \mu(K \cup \{i\}) - \mu(K \cup \{j\}) + \mu(K)].$$

The Roubens interaction index for elements $i, j \in X$ is defined by

$$I_R(i, j) = \frac{1}{2^{n-2}} \sum_{K \subseteq N - \{i, j\}} [\mu(K \cup \{i, j\}) - \mu(K \cup \{i\}) - \mu(K \cup \{j\}) + \mu(K)].$$

The interaction indices $I_{MS}(i, j)$ and $I_R(i, j)$ can be interpreted as specific kinds of average added values resulting from putting i and j together in each possible coalition. The following cases can happen:

- $I_{MS}(i, j) > 0$ ($I_R(i, j) > 0$): i and j are complementary,
- $I_{MS}(i, j) < 0$ ($I_R(i, j) < 0$): i and j are substitutive,
- $I_{MS}(i, j) = 0$ ($I_R(i, j) = 0$): i and j are independent.

The definition of interaction indices can be extended from non-ordered pairs $i, j \in N$ to any subset $A \subseteq N, A \neq \emptyset$. Extensions of interaction indices in this sense have been proposed by Grabisch (1996) and Roubens (1996), with respect to Shapley index and Banzhaf index, respectively.

The Shapley interaction index of elements from $A \subseteq N$ is defined by

$$I_S(A) = \sum_{K \subseteq N-A} \frac{(n - |K| - |A|)!|K|!}{(n - |A| + 1)!} \sum_{L \subseteq A} (-1)^{|A|-|L|} \mu(L \cup K).$$

The Banzhaf interaction index of elements from $A \subseteq N$ is defined by

$$I_B(A) = \frac{1}{2^{n-|A|}} \sum_{K \subseteq N-A} \sum_{L \subseteq A} (-1)^{|A|-|L|} \mu(K \cup L).$$

In addition to the interaction indices, another concept useful for the interpretation of the fuzzy measures is the *Möbius representation* of μ , i.e. the set function $m : P(N) \rightarrow R$ defined by

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \mu(B)$$

for any $A \subseteq N$. Within Dempster–Shafer theory of evidence (Shafer, 1976), $m(A)$ is interpreted as basic probability assignment.

The relations between fuzzy measures μ , interaction indices I_S and I_B and Möbius representation m have been extensively studied in Grabisch (1997), Dennemberg and Grabisch (1996), Roubens (1996) and Grabisch and Roubens (1997).

Interaction indices I_S and I_B and Möbius representation m can be used within rough set analysis to study the relative value of the information supplied by different attributes (Greco et al., 1998a). Considering the quality of classification as a fuzzy measure, we conclude that:

1. the Shapley value $\phi_S(i)$ and the Banzhaf value $\phi_B(i)$ can be interpreted as measures of the contribution of attribute $i = 1, \dots, n$ to the quality of approximation of the considered classification,
2. the Murofushi–Soneda interaction index $I_{MS}(i, j)$ and Roubens interaction index $I_R(i, j)$ can be interpreted as the average conjoint contribution of the non-ordered pair of attributes $i, j = 1, \dots, n, i \neq j$, to the quality of classification when adjoined to all sets $K \subset C$ such that $K \cap \{i, j\} = \emptyset$,
3. the Shapley interaction index $I_S(A)$ and the Banzhaf interaction index $I_B(A)$ can be interpreted as the average conjoint contribution of the subset of attributes $A \subset C$ to the quality of classification when adjoined to all sets $B \subset C$ such that $B \cap A = \emptyset$,
4. the Möbius representation $m(A)$ of μ can be interpreted as the conjoint contribution of the subset of attributes $A \subseteq C$ to the quality of classification.

All of these indices are useful to study the informational dependence among the considered attributes and to choose the best reduct.

2.6. An example

Let us consider an example inspired by the example of *evaluation in a high school* proposed by Grabisch (1994). The director of the school wants to assign students to two classes: bad and good. To fix classification rules the director is asked to give some examples. The examples concern six students described by means of four attributes (see Table 1):

- A_1 , level in Mathematics,
- A_2 , level in Physics,

Table 1
Data table with examples of classification

Student	A_1	A_2	A_3	A_4
1	good	good	bad	good
2	medium	bad	bad	bad
3	medium	bad	bad	good
4	bad	bad	bad	bad
5	medium	good	good	bad
6	good	bad	good	good

- A_3 , level in Literature,
- A_4 , global evaluation (decision class).

The components of the data table S are:

$$\begin{aligned}
 U &= \{1, 2, 3, 4, 5, 6\}, \\
 Q &= \{A_1, A_2, A_3, A_4\}, \\
 V_1 &= \{\text{bad}, \text{medium}, \text{good}\}, \\
 V_2 &= \{\text{good}, \text{bad}\}, \\
 V_3 &= \{\text{good}, \text{bad}\}, \\
 V_4 &= \{\text{good}, \text{bad}\},
 \end{aligned}$$

the information function $f(x, q)$, taking values $f(1, A_1) = \text{good}$, $f(1, A_2) = \text{good}$, and so on.

Observe that each student has a different description in terms of the attributes A_1, A_2, A_3 and A_4 , so they can be distinguished (discerned) by means of the information supplied by the four attributes. Formally, the indiscernibility relation based on all four attributes is $I_Q = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$ and, therefore, there is no two distinct students x and y such that $(x, y) \in I_Q$. However, students 2 and 3 are indiscernible with respect to the attributes from $P = \{A_1, A_2, A_3\}$, since they have the same values on the three attributes. Formally, the indiscernibility relation based on P is $I_P = \{(1, 1), (2, 2), (2, 3), (3, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$. Similarly, students 2, 3, 4 are indiscernible with respect to the attributes from $P' = \{A_2, A_3\}$, and so on, considering all the possible subsets of attributes from Q .

Each $P \subseteq Q$ determines a partition $U|I_P$ that groups in the corresponding equivalence classes the objects having the same description in terms of the attributes from P , e.g., for $P' = \{A_2, A_3\}$, $U|I_{P'} = \{\{1\}, \{2, 3, 4\}, \{5\}, \{6\}\}$, and thus, $\{1\}, \{2, 3, 4\}, \{5\}, \{6\}$ are the P' -elementary sets.

Suppose that, using the set of attributes $P = \{A_1, A_2, A_3\}$, we wish to approximate the set X of students having a global evaluation “good”, i.e. $X = \{1, 3, 6\}$. Since $U|I_P = \{\{1\}, \{2, 3\}, \{4\}, \{5\}, \{6\}\}$, the result is

$$\underline{P}(X) = \{1, 6\}, \quad \overline{P}(X) = \{1, 2, 3, 6\}, \quad Bn_P(X) = \{2, 3\}.$$

The answer to the question whether it is possible to describe X by means of the information supplied by the attributes from P is not univocal. Observe that the P -boundary $Bn_P(X)$ is not empty: students 2 and 3, belonging to the P -boundary have the same description in terms of attributes considered, however, student 2 is globally bad while student 3 is good. Nevertheless, the P -lower approximation of X , $\underline{P}(X)$, is also not empty and consists of students 1 and 6, whose descriptions are different from those of all the students not belonging to X . Summing up, in intuitive terms, it may be said that, on the basis of the information supplied by the attributes from P :

- students 1 and 6, from the P -lower approximation of X , certainly belong to the set of “good” students,
- students 1, 2, 3 and 6, from the P -upper approximation of X , could belong to the set of “good” students,

- students 2 and 3, from the P -boundary of X , represent cases of uncertain membership to the set of “good” students.

Using the same set of attributes $P = \{A_1, A_2, A_3\}$, the approximation of the set Y of students having global evaluation “bad”, i.e. $Y = \{2, 4, 5\}$, gives the following result:

$$\underline{P}(Y) = \{4, 5\}, \quad \overline{P}(Y) = \{2, 3, 4, 5\}, \quad Bn_P(Y) = \{2, 3\}.$$

Let us consider now the following subsets of Q : $P = \{A_1, A_2, A_3\}$, $R = \{A_1, A_2\}$, $T = \{A_1, A_3\}$, $W = \{A_2, A_3\}$. It is easy to observe that $I_R = I_P$, $I_T = I_P$, while $I_W \neq I_P$. This means that R and T are reducts of P , while W is not. In other words, R and T are minimal subsets of P that induce the same partition of the elements of U as the set of attributes P . It can also be observed that in the core of P , defined by $R \cap T$, there is attribute A_1 , which is thus indispensable for the approximation of the class of “good” students (and also for the class of “bad” students), while other attributes from R and T may be mutually exchanged.

If in the set of attributes Q , condition attributes $C = \{A_1, A_2, A_3\}$ and decision attribute $D = \{A_4\}$ were distinguished, the data table could be seen as a decision table. In order to explain the evaluations of the decision attribute in terms of the evaluations of the condition attributes, one can represent the data table as a set of decision rules. Such a representation of Table 1 gives the following rules:

1. if $f(x, A_1) = \text{good}$ and $f(x, A_2) = \text{good}$ and $f(x, A_3) = \text{bad}$, then $f(x, A_4) = \text{good}$ (or, in linguistic terms, “if the level in Mathematics is good and the level in Physics is good and the level in Literature is bad, then the students is good”),
2. if $f(x, A_1) = \text{medium}$ and $f(x, A_2) = \text{bad}$ and $f(x, A_3) = \text{bad}$, then $f(x, A_4) = \text{bad}$,
3. if $f(x, A_1) = \text{medium}$ and $f(x, A_2) = \text{bad}$ and $f(x, A_3) = \text{bad}$, then $f(x, A_4) = \text{good}$,
4. if $f(x, A_1) = \text{bad}$ and $f(x, A_2) = \text{bad}$ and $f(x, A_3) = \text{bad}$, then $f(x, A_4) = \text{bad}$,
5. if $f(x, A_1) = \text{medium}$ and $f(x, A_2) = \text{good}$ and $f(x, A_3) = \text{good}$, then $f(x, A_4) = \text{bad}$,
6. if $f(x, A_1) = \text{good}$ and $f(x, A_2) = \text{bad}$ and $f(x, A_3) = \text{good}$, then $f(x, A_4) = \text{good}$.

The above set of rules may then be reduced by induction, obtaining a more concise representation of the decision table (within parentheses there are the objects supporting the corresponding rule):

- (1') if $f(x, A_1) = \text{good}$, then $f(x, A_4) = \text{good}$, (1, 6)
- (2') if $f(x, A_1) = \text{bad}$, then $f(x, A_4) = \text{bad}$, (4)
- (3') if $f(x, A_1) = \text{medium}$ and $f(x, A_2) = \text{good}$, then $f(x, A_4) = \text{bad}$, (5)
- (4') if $f(x, A_1) = \text{medium}$ and $f(x, A_2) = \text{bad}$, then $f(x, A_4) = \text{good or bad}$. (2, 3).

Observe that rules (1'), (2') and (3') have a univocal consequence and therefore these are exact rules, while rule (4') does not have a univocal consequence and for this reason it is an approximate rule.

Finally, the quality of approximation, interaction indices I_S and I_B and Möbius representation of all subsets of attributes in C were calculated. Their values are presented in Table 2. The results presented in Table 2 can be interpreted as follows:

Table 2
Quality of approximation, Möbius representation and interaction indices

Attributes	Quality	Möbius	Shapley	Banzhaf
$\{A_1\}$	0.5	0.5	0.44	0.5
$\{A_2\}$	0	0	0.11	0.17
$\{A_3\}$	0	0	0.11	0.17
$\{A_1, A_2\}$	0.67	0.17	−0.17	−0.17
$\{A_1, A_3\}$	0.67	0.17	−0.17	−0.17
$\{A_2, A_3\}$	0.5	0.5	0.17	0.17
$\{A_1, A_2, A_3\}$	0.67	−0.67	−0.67	−0.67

1. The second column shows the quality of approximation for the considered subset of attributes.
2. The third column presents the Möbius representation and gives a measure of the conjoint contribution of the corresponding subset of attributes to the quality of classification. The negative value corresponding to $\{A_1, A_2, A_3\}$ should be read as a measure of the information redundancy in conjoint contribution of the three attributes.
3. The fourth column shows the Shapley interaction index: more precisely, the first three values are the Shapley values and can be interpreted as measures of importance of the corresponding attributes in the rough approximation. One can notice a relatively great importance of A_1 in comparison to A_2 and A_3 . Furthermore, A_2 and A_3 are complementary, while A_1 and A_2 , as well as A_1 and A_3 , are substitutive. Finally, there is redundancy between A_1 , A_2 and A_3 , as pointed out by the negative value of the corresponding interaction index.
4. The fifth column presents the Banzhaf interaction index that has an interpretation analogous to the Shapley interaction index.

3. Generalization of the indiscernibility relation

As mentioned above, the classical definitions of lower and upper approximations are based on the use of the binary indiscernibility relation which is an equivalence relation. In this case, the sets to be approximated and the relation used for this approximation are both ordinary (*crisp*).

Generalizations consisting in approximation of *fuzzy* sets with a *fuzzy* indiscernibility relation have been considered by Dubois and Prade (1990, 1992), Slowinski (1993a), Slowinski and Stefanowski (1989, 1994, 1996), Yao (1996, 1998) and Inuiguchi and Tanino (2000). Nevertheless, this approach is still based on the use of the indiscernibility relation.

Further generalizations replacing the indiscernibility relation by a weaker binary *similarity* relation have considerably extended the capacity of the rough set approach, since in the least demanding case the similarity relation requires reflexivity only, relaxing the assumptions of symmetry and transitivity (see Slowinski and Vanderpooten, 1997, 2000).

3.1. Similarity

The indiscernibility implies an impossibility to distinguish two objects of U having the *same* description in terms of the attributes from Q . This relation induces equivalence classes on U , which constitute the basic granules of knowledge. In reality, due to the imprecision of data describing the objects, small differences are often not considered significant for the purpose of discrimination. This situation may be formally modeled by considering similarity or tolerance relations (see, e.g., Nieminen, 1988; Lin, 1989; Marcus, 1994; Polkowski et al., 1995; Skowron and Stepaniuk, 1995; Slowinski and Vanderpooten, 1997, 2000; Yao and Wong, 1995).

In general, the similarity relations R do not generate partitions on U ; the information regarding similarity may be represented using *similarity classes* for each object $x \in U$. Precisely, the similarity class of x , denoted by $R(x)$, consists of the set of objects which are similar to x :

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

It is obvious that an object y may be similar to both x and z , while z is not similar to x , i.e. $y \in R(x)$ and $y \in R(z)$, but $z \notin R(x)$, $x, y, z \in U$. The similarity relation is of course reflexive (each object is similar to itself). Slowinski and Vanderpooten (1997, 2000) have proposed a *similarity* relation which is only *reflexive*.

The abandon of the transitivity requirement is easily justifiable, remembering – for example – Luce’s paradox of the cups of tea (Luce, 1956). As for the symmetry, one should notice that yRx , which means “ y is similar to x ”, is directional; there is a *subject* y and a *referent* x , and in general this is not equivalent to the proposition “ x is similar to y ”, as maintained by Tversky (1977). This is quite evident when the similarity relation is defined in terms of a percentage difference between numerical evaluations of the objects, calculated with respect to the evaluation of the referent object. Therefore, the symmetry of the similarity relation should not be imposed and then it makes sense to consider the inverse relation of R , denoted by R^{-1} , where $xR^{-1}y$ means again “ y is similar to x ”; $R^{-1}(x)$, $x \in U$, is the class of referent objects to which x is similar:

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

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

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

The objects which are neither positive nor negative are said to be *ambiguous*.

A more general definition of lower and upper approximation may thus be offered (see Slowinski and Vanderpooten, 2000). Let $X \subseteq U$ and R a reflexive binary relation defined on U ; the lower approximation of X , denoted by $\underline{R}(X)$, and the upper approximation of X , denoted by $\overline{R}(X)$, are defined, respectively, as:

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

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

It may be demonstrated that the key property: $\underline{R}(X) \subseteq X \subseteq \overline{R}(X)$, still holds and that

$$\underline{R}(X) = U - \overline{R}(U - X) \quad (\text{complementarity property})$$

and

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

Moreover, the definitions proposed are the only ones that correctly characterize the set of positive objects (lower approximation) and the set of positive or ambiguous objects (upper approximation) when a similarity relation is reflexive, but not necessarily symmetric nor transitive.

Using similarity relation one is able to induce decision rules from a decision table. The syntax of a rule is the following:

if $f(x, q_1)$ is similar to r_{q_1} and $f(x, q_2)$ is similar to r_{q_2} and $\dots f(x, q_p)$ is similar to r_{q_p} , then x belongs to Y_{j1} or Y_{j2} or $\dots Y_{jk}$,

where $\{q_1, q_2, \dots, q_p\} \subseteq C$, $(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$ and $Y_{j1}, Y_{j2}, \dots, Y_{jk}$ are some classes of the considered classification (D -elementary sets). If $k = 1$ then the rule is *exact* or *certain*, otherwise it is *approximate* or *uncertain*. Procedures for generation of decision rules adapt the scheme described in Section 2.4. One such procedure has been proposed by Krawiec et al. (1998) – it involves a similarity relation that is learned from data. Let us add that, recently, Greco et al. (1998g, 2000) proposed a fuzzy extension of the similarity, i.e. rough approximation of fuzzy sets by means of fuzzy similarity relations.

3.2. Missing values

The classical rough set approach based on the use of indiscernibility relations requires the data table to be complete, i.e. without *missing values* on condition attributes describing the objects. In practice, however, the data tables are very often incomplete. To deal with these cases, we proposed in Greco et al. (1999e) and Greco et al. (1999f) an extension of the rough set methodology to the analysis of incomplete data tables. The extended indiscernibility relation between two objects is considered as a directional statement where a subject is compared to a referent object. We require that the referent object has no missing values. The extended rough set approach maintains all good characteristics of its original version. It also boils down to the original approach when there are no missing values. The rules induced from the rough approximations defined according to the extended relation verify a suitable property: they are robust in a sense that each rule is supported by at least one object with no missing value on the condition attributes represented in the rule.

For any two objects $x, y \in U$, we are considering a directional comparison of y to x ; object y is called subject and object x , referent. We say that subject y is *indiscernible* with referent x , with respect to condition attributes from $P \subseteq C$ (denotation yI_Px), if for every $q \in P$ the following conditions are met: (1) $f(x, q) \neq *$, (2) $f(x, q) = f(y, q)$ or $f(y, q) = *$, where $*$ denotes a missing value.

The above means that the referent object considered for indiscernibility with respect to P should have no missing values on attributes from set P . The binary relation I_P is not necessarily reflexive and also not necessarily symmetric. However, I_P is transitive.

For each $P \subseteq C$ let us define a set of objects having no missing values on attributes from P :

$$U_P = \{x \in U : f(x, q) \neq * \text{ for each } q \in P\}.$$

For each $x \in U$ and for each $P \subseteq C$ let $I_P(x) = \{y \in U : yI_Px\}$ denote the class of objects indiscernible with x . Given $X \subseteq U$ and $P \subseteq C$, we define lower and upper approximation of X with respect to P as follows:

$$\underline{I}_P(X) = \{x \in U_P : I_P(x) \subseteq X\}, \quad (1)$$

$$\bar{I}_P(X) = \{x \in U_P : I_P(x) \cap X \neq \emptyset\}. \quad (2)$$

Let $X_P = X \cap U_P$. The rough approximation defined as above satisfies the following properties:

- (*Rough inclusion*). For each $X \in U$ and for each $P \subseteq C$: $\underline{I}_P(X) \subseteq X_P \subseteq \bar{I}_P(X)$.
- (*Complementarity*). For each $X \in U$ and for each $P \subseteq C$: $\underline{I}_P(X) = U_P - \bar{I}_P(U - X)$.

Let us observe that a very useful property of lower approximation within classical rough sets theory is that if an object $x \in U$ belongs to the lower approximation of X with respect to $P \subseteq C$, then x belongs also to the lower approximation of X with respect to $R \subseteq C$ when $P \subseteq R$ (this is a kind of monotonicity property). However, definition (1) does not satisfy this property of lower approximation, because it is possible that $f(x, q) \neq *$ for all $q \in P$ but $f(x, q) = *$ for some $q \in R - P$. This is quite problematic for some key concepts of the rough sets theory, like quality of approximation, reduct and core.

Therefore, another definition of lower approximation should be considered to restore the concepts of quality of approximation, reduct and core in the case of missing values. Given $X \subseteq U$ and $P \subseteq C$, this definition is the following:

$$\underline{I}_P^*(X) = \bigcup_{R \subseteq P} \underline{I}_R(X). \quad (3)$$

$\underline{I}_P^*(X)$ is called *cumulative P-lower approximation* of X because it includes all the objects belonging to all R -lower approximations of X , where $R \subseteq P$.

It can be shown that another type of indiscernibility relation, denoted by I_p^* , permits a direct definition of the cumulative P -lower approximation in a usual way. For each $x, y \in U$ and for each $P \subseteq C$, yI_p^*x means that $f(x, q) = f(y, q)$ or $f(x, q) = *$ and/or $f(y, q) = *$, for every $q \in P$. Let $I_p^*(x) = \{y \in U : yI_p^*x\}$ for each $x \in U$ and for each $P \subseteq C$. I_p^* is reflexive and symmetric but not transitive (Kryszkiewicz, 1998). Greco et al. (1999e) proved that definition (3) is equivalent to the following definition: $\underline{I}_p^*(X) = \{x \in U_p^* : I_p^*(x) \subseteq X\}$, where $U_p^* = \{x \in U : f(x, q) \neq * \text{ for at least one } q \in P\}$.

Using I_p^* we can give definition of the P -upper approximation of X , complementary to $\underline{I}_p^*(X)$

$$\overline{I}_p^*(X) = \{x \in U_p^* : I_p^*(x) \cap X \neq \emptyset\}. \quad (4)$$

For each $X \subseteq U$, let $X_p^* = X \cap U_p^*$. Let us remark that $x \in U_p^*$ if and only if there exists $R \neq \emptyset$ such that $R \subseteq P$ and $x \in U_R$.

Rough approximation $\underline{I}_p^*(X)$ and $\overline{I}_p^*(X)$ satisfies the following properties:

- (*Rough inclusion*). For each $X \subseteq U$ and for each $P \subseteq C$: $\underline{I}_p^*(X) \subseteq X_p^* \subseteq \overline{I}_p^*(X)$;
- (*Complementarity*). For each $X \subseteq U$ and for each $P \subseteq C$: $\underline{I}_p^*(X) = U_p^* - \overline{I}_p^*(U - X)$.
- (*Monotonicity of the accuracy of approximation*). For each $X \subseteq U$ and for each $P, R \subseteq C$, such that $P \subseteq R$, the following inclusion holds: $\underline{I}_p^*(X) \subseteq \underline{I}_R^*(X)$. Furthermore, if $U_p^* = U_R^*$, the following inclusion is also true: $\overline{I}_p^*(X) \supseteq \overline{I}_R^*(X)$.

Due to the property of monotonicity, when augmenting a set of attributes P , we get a lower approximation of X that is at least of the same cardinality. Thus, we can restore for the case of missing values the following key concepts of the rough sets theory: accuracy and quality of approximation, reduct and core. These concepts have the same definitions as those given in Sections 2.2 and 2.3 but they use rough approximation $\underline{I}_p^*(X)$ and $\overline{I}_p^*(X)$.

3.2.1. An example

The illustrative example presented in this section is intended to explain the concepts introduced in Section 3.2. The director of the school must give a global evaluation to some students. This evaluation should be based on the level in Mathematics, Physics and Literature. However, not all the students have passed all three exams and, therefore, there are some missing values. The director gave the examples of evaluation as shown in Table 3.

The following lower and upper approximations can be calculated from Table 3:

$$\underline{I}_C(\text{good}) = \{6\}, \quad \underline{I}_C(\text{bad}) = \{1\}, \quad \overline{I}_C(\text{good}) = \{6\}, \quad \overline{I}_C(\text{bad}) = \{1\},$$

$$\underline{I}_C^*(\text{good}) = \{2, 6\}, \quad \underline{I}_C^*(\text{bad}) = \{1, 5\}, \quad \overline{I}_C^*(\text{good}) = \{2, 3, 4, 6\}, \quad \overline{I}_C^*(\text{bad}) = \{1, 3, 4, 5\}.$$

The quality of approximation of the partition of U using attributes from C is equal to 0.67. There are two reducts: $\text{Red}_1 = \{\text{Mathematics, Physics}\}$, $\text{Red}_2 = \{\text{Mathematics, Literature}\}$. The intersection of Red_1 and Red_2 constitutes the core, i.e. $\text{Core} = \{\text{Mathematics}\}$.

Table 3
Student evaluations with missing values

Student	Mathematics	Physics	Literature	Global evaluation
1	medium	bad	bad	bad
2	good	medium	*	good
3	medium	*	medium	bad
4	*	medium	medium	good
5	*	good	bad	bad
6	good	medium	bad	good

The following minimal exact rules can be induced from Table 3 (within parentheses there are objects supporting the corresponding decision rule):

1. “if Mathematics is good and Physics is medium, then the student is good” (students 2, 4, 6)
2. “if Mathematics is medium and Literature is bad, then the student is bad” (students 1, 5)

It is also possible to induce the following minimal approximate rule from Table 3:

3. “if Mathematics is medium and Literature is medium, then the student is bad or good” (students 3, 4)

We claim that decision rules induced from an incomplete data table according to our approach are *robust* in the following sense: among objects supporting a given decision rule there is at least one object matching exactly all elementary conditions of the rule. This is a distinctive feature in comparison with the approach proposed by Kryszkiewicz (1998). Her approach is based on the concept of a possible “completion” of an incomplete data table $S = \langle U, Q, V, f \rangle$. The completion is understood as a complete data table $S' = \langle U, Q, V, f' \rangle$ obtained by substitution of each missing value $*$ by some possible value from the domain of the corresponding attribute.

According to Kryszkiewicz’s approach, a decision rule “if $f(x, q_1) = r_{q_1}$ and $f(x, q_2) = r_{q_2}$ and $\dots f(x, q_p) = r_{q_p}$, then x is assigned to class Cl_i ” is certain (exact) if for any possible completion of data table S the implication expressed by the rule is true. A certain decision rule is called optimal (minimal) if no elementary condition can be eliminated from the condition part of the rule.

Let us observe that some of the decision rules obtained using Kryszkiewicz’s approach may not be robust, i.e. certain and founded on a real object present in the data table. Consider for instance the following optimal and certain decision rule induced using this approach from Table 3:

1. “if Mathematics is good and Physics is good, then student is bad” (student 5).

In Table 3 there is no object having a description matching exactly the condition part of this rule. In other words, there is no student characterized by a good level in Mathematics and by a good level in Physics. Therefore, rule (4) is not robust.

On the contrary, each of the rules generated using our approach is supported by at least one real object matching exactly the condition part of the corresponding decision rule and, therefore, they are robust. Precisely, rule (1) is founded on student 2 and 6, rule (2) on student 1, and rule (3) on student 3.

4. Multiattribute and multicriteria classification and sorting problems

As mentioned above, a decision table contains the information relative to a set of objects, described by a certain number of attributes. The traditional rough set analysis of such a table consists in approximating the classifications induced by decision attributes by means of the classifications induced by condition attributes. The two kinds of classifications are built independently, i.e. they are not deduced one from the other.

The aim of the decision analysis is to answer two basic questions. The first question is to explain decisions in terms of the circumstances in which they were made. The second is to give a recommendation how to make a decision under specific circumstances. Recommendation is mainly based on decision rules induced from a decision table. In this sense, the rough set approach is similar to the inductive learning approach (Michalski et al., 1998), however, the former one is going far beyond the latter because in the rough set approach, the recommendation task is preceded by the explanation which gives pertinent information useful for decision support (reducts, core, quality of approximation, relevance of attributes).

According to Roy (1985, 1993), it is possible to distinguish the following three, most frequent decision problems: classification, choice and ranking.

In general, decisions are based on some characteristics of objects (actions). For example, when buying a car, the decisions can be based on such characteristics as price, maximum speed, fuel consumption, color, country of production, etc. We refer to these characteristics calling them attributes. Let us observe that,

depending on interpretation given to the attributes by the DM, some of them may have ordinal properties expressing preference scales, while others may not. The former attributes are called criteria, while the latter ones keep the name of attributes. In the above example, price, maximum speed and fuel consumption are criteria, because, for instance, a low price is better than a high price; most probably, color and country of production are not criteria but simple attributes because, for instance, red is not better than green. However, one can imagine that also those attributes could become criteria, because a DM could consider, for example, red better than green.

Moreover, decisions may be ordinal, because of expressing a preference, or may not be ordinal. For example, classification of cars for a catalogue does not impose any preference order among the classes (sport cars, family cars, utility cars, etc.), however, choice of the best car, or ranking of a set of cars from the best to the worst surely impose a preference order. Let us also observe that, depending on interpretation given to the classification by the DM, the classes may express a preference, so also classification may be ordinal. For instance, the DM could be interested in classification of cars in three categories: acceptable, hardly acceptable, non-acceptable. This type of classification is called sorting.

In the case of any multicriteria and/or multiattribute decision problem, no recommendation can be elaborated before the DM provides some preferential information suitable to the preference model assumed.

There are two major models used until now in multicriteria decision analysis: functional and relational ones. The functional model has been extensively used within the framework of multiattribute utility theory (Keeney and Raiffa, 1976). The relational model has its most widely known representation in the form of an outranking relation (Roy, 1991) and a fuzzy relation (Fodor and Roubens, 1994). These models require specific preferential information more or less explicitly related with their parameters. For example, in the deterministic case, the DM is often asked for pairwise comparisons of objects, from which we can assess the substitution rates in the functional model or importance weights in the relational model (see Fishburn, 1967; Jacquet-Lagrèze and Siskos, 1982; Mousseau, 1993). This kind of preferential information seems to be close to the natural reasoning of the DM. He/she is typically more confident exercising his/her comparisons than explaining them. The representation of this information by functional or relational models seems, however, less natural. According to Slovic (1975), people make decisions by searching for *rules* that provide good justification of their choices. So, after getting the preferential information in terms of exemplary comparisons, it would be natural to build the preference model in terms of “*if... then...*” rules. Then, these rules can be applied to a set of objects (potential actions) in order to obtain specific preference relations. From the exploitation of these relations, a suitable recommendation can be obtained to support the DM in decision problem at hand.

The induction of rules from examples is a typical approach of artificial intelligence. It is concordant with the principle of posterior rationality by March (1988) and with aggregation–disaggregation logic by Jacquet-Lagrèze (1981). The rules explain the preferential attitude of the DM and enable his/her understanding of the reasons of his/her preferences. The recognition of the rules by the DM (Langley and Simon, 1998) justifies their use for decision support. So, the preference model in the form of rules derived from examples, fulfils both explanation and recommendation tasks mentioned above.

In Sections 4.2–5.4, we are presenting the main extensions of the rough set approach, resulting in a new methodology of modeling and exploitation of preferences in terms of decision rules. The rules are induced from the preferential information given by the DM in the form of examples of decisions. More precisely, for A being a finite set of objects (real or fictitious actions, potential or not) considered in a multicriteria problem, the examples of decisions are confined to a subset of objects $B \subseteq A$, relatively well known to the DM, called *reference objects*. Depending on the type of the multicriteria problem, the examples concern either assignment of reference objects to decision classes (sorting problem) or pairwise comparisons of reference objects (choice and ranking problems).

4.1. Problems of multiattribute classification

Up to now, the rough set approach to decision analysis has been limited to problems of multiattribute classification concerning the assignment of a set of objects described by a set of attributes (not criteria) to one of pre-defined categories (Pawlak and Slowinski, 1994). Rough set analysis is naturally adapted to this type of problems because the set of classification examples may be represented directly in the decision table and it is possible to extract all the essential knowledge contained in the table using indiscernibility or similarity relations.

The rough sets theory has been successfully applied to a number of real classification problems in different fields, such as medicine, pharmacology, engineering, credit management, market research, financial analysis, environmental economics, linguistics, databases and other important sectors. The interesting results have encouraged experts in various disciplines to study the rough sets theory and its applications. For a collection of studies on the application of the rough set approach to real-world problems (see Slowinski, 1992; Slowinski et al., 1988; Polkowski and Skowron, 1998). A brief but thorough review of the most important applications has been made by Pawlak (1997).

4.2. Problems of multicriteria sorting

As pointed out by Greco et al. (1996, 1998b,e, 1999c), the original rough set approach cannot extract all the essential knowledge contained in the decision table of multicriteria sorting problems, i.e. problems of assigning a set of objects described by a set of criteria to one of pre-defined and preference-ordered categories. Notwithstanding, in many real problems it is important to consider the *ordinal properties* of the considered criteria. For example, in bankruptcy risk evaluation, if the debt ratio (total debt/total activity) of company A has a modest value, while the same ratio of company B has a significant value, then, within the rough set approach, the two firms are just discernible, but no preference is given to one of them with respect to the attribute “debt ratio”. In reality, from the point of view of the bankruptcy risk evaluation, it would be reasonable to consider firm A better than firm B, and not simply different (discernible). Let us observe that the rough set approach based on the use of indiscernibility or similarity relations is not able to capture a specific kind of inconsistency which may occur when in the decision table there is at least one criterion. For instance, in the bankruptcy risk evaluation, which is a sorting problem, if firm A is better than B with respect to all the considered criteria (e.g. debt ratio, return on equity, etc.) but firm A is assigned to a class of higher risk than B, then there is an inconsistency which cannot be captured by the original rough set approach, because these firms are discernible. In order to detect this inconsistency, the rough approximation should handle the ordinal properties of criteria. This can be made by replacing the indiscernibility or similarity relation by the *dominance relation*, which is a very natural concept within multicriteria decision making.

On the basis of these considerations, Greco et al. (1998b,e, 1999g,h) have proposed a new rough set approach to multicriteria sorting problems, which is described in the following parts of the paper. Let us also mention that it is sometimes reasonable to consider both criteria and regular attributes (without preference-ordered domains) in a sorting problem. In this particular case, the rough approximation is based on the use of a binary relation which aggregates dominance (with respect to considered criteria) and indiscernibility (with respect to considered attributes), as proposed by Greco et al. (1998e). A more general binary relation that aggregates dominance, indiscernibility and similarity was considered by Greco et al. (1999d). Let us also mention that a fuzzy set extension of the rough approximation by dominance has been presented by Greco et al. (1999a). Moreover, a fuzzy set extension of the rough approximation based on dominance and similarity together has been described recently by Greco et al. (1999k).

4.2.1. Rough approximation by means of dominance relations

Let S_q be an *outranking* relation Roy (1985) on U with reference to criterion $q \in C$, such that xS_qy means “ x is at least as good as y with respect to criterion q ”. Suppose that S_q is a complete preorder, i.e. a strongly complete and transitive binary relation. Moreover, let $\mathbf{CI} = \{CI_t, t \in T\}$, $T = \{1, \dots, n\}$, be a set of classes of U , such that each $x \in U$ belongs to one and only one class $CI_t \in \mathbf{CI}$. We assume that for all $r, s \in T$, such that $r > s$, each element of CI_r is preferred (strictly or weakly) to each element of CI_s . More formally, if S is a comprehensive outranking relation on U , i.e. xSy means: “ x is at least as good as y ” for any $x, y \in U$, then it is supposed that

$$[x \in CI_r, y \in CI_s, r > s] \Rightarrow [xSy \text{ and not } ySx].$$

Let us also consider the following *upward* and *downward unions of classes*, respectively,

$$CI_t^{\geq} = \bigcup_{s \geq t} CI_s, \quad CI_t^{\leq} = \bigcup_{s \leq t} CI_s.$$

Observe that $CI_1^{\geq} = CI_n^{\leq} = U$, $CI_n^{\geq} = CI_1$ and $CI_1^{\leq} = CI_1$.

It is said that x *dominates* y with respect to $P \subseteq C$ (denotation $x D_P y$) if $x S_q y$ for all $q \in P$. Since the intersection of complete preorders is a partial preorder and S_q is a complete preorder for each $q \in P$, and $D_P = \bigcap_{q \in P} S_q$, then the dominance relation D_P is a partial preorder. Given $P \subseteq C$ and $x \in U$, let

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

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

represent, so-called, *P-dominating set* and *P-dominated set* with respect to x , respectively. In the dominance-based rough set approach, the sets to be approximated are upward and downward unions of classes and the items (granules of knowledge) used for this approximation are dominating and dominated sets.

The *P-lower* and the *P-upper approximation* of CI_t^{\geq} , $t \in T$, with respect to $P \subseteq C$ (denotation $\underline{P}(CI_t^{\geq})$ and $\overline{P}(CI_t^{\geq})$, respectively), are defined as:

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

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

Analogously, the *P-lower* and the *P-upper approximation* of CI_t^{\leq} , $t \in T$, with respect to $P \subseteq C$ (denotation $\underline{P}(CI_t^{\leq})$ and $\overline{P}(CI_t^{\leq})$, respectively) are defined as:

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

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

The *P-lower* and *P-upper approximations* defined as above satisfy the following properties for all $t \in T$ and for any $P \subseteq C$:

$$\underline{P}(CI_t^{\geq}) \subseteq CI_t^{\geq} \subseteq \overline{P}(CI_t^{\geq}), \quad \underline{P}(CI_t^{\leq}) \subseteq CI_t^{\leq} \subseteq \overline{P}(CI_t^{\leq}).$$

Furthermore, the following specific complementarity properties hold:

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

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

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

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

We define the *accuracy of approximation* of Cl_t^{\geq} and Cl_t^{\leq} for all $t \in T$ and for any $P \subseteq C$, respectively, as

$$\alpha_P(Cl_t^{\geq}) = \frac{|\underline{P}(Cl_t^{\geq})|}{|\overline{P}(Cl_t^{\geq})|}, \quad \alpha_P(Cl_t^{\leq}) = \frac{|\underline{P}(Cl_t^{\leq})|}{|\overline{P}(Cl_t^{\leq})|}.$$

The ratio

$$\gamma_P(\mathbf{CI}) = \frac{|U - ((\bigcup_{t \in T} Bn_P(Cl_t^{\geq})) \cup (\bigcup_{t \in T} Bn_P(Cl_t^{\leq})))|}{|U|}$$

defines the *quality of approximation of the partition CI* by means of the set of criteria P , or, briefly, *quality of sorting*. This ratio expresses the relation between all the P -correctly classified objects and all the objects in the table.

Every minimal subset $P \subseteq C$ such that $\gamma_P(\mathbf{CI}) = \gamma_C(\mathbf{CI})$ is called a *reduct* of C with respect to \mathbf{CI} and is denoted by $RED_{\mathbf{CI}}(P)$. Again, a data table may have more than one reduct. The intersection of all the reducts is known as the core, denoted by $CORE_{\mathbf{CI}}$.

4.2.2. Decision rules

On the basis of the approximations obtained by means of the dominance relations, it is possible to induce a generalized description of the preferential information contained in the decision table, in terms of decision rules (Slowinski et al., 2000).

The following three types of decision rules can be considered:

1. D_{\geq} -decision rules, which have the following form:

if $f(x, q_1) \geq r_{q_1}$ and $f(x, q_2) \geq r_{q_2}$ and ... $f(x, q_p) \geq r_{q_p}$, then $x \in Cl_t^{\geq}$,

where $P = \{q_1, q_2, \dots, q_p\} \subseteq C$, $(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$ and $t \in T$; these rules are supported only by objects from the P -lower approximations of the upward unions of classes Cl_t^{\geq} .

2. D_{\leq} -decision rules, which have the following form:

if $f(x, q_1) \leq r_{q_1}$ and $f(x, q_2) \leq r_{q_2}$ and ... $f(x, q_p) \leq r_{q_p}$, then $x \in Cl_t^{\leq}$,

where $P = \{q_1, q_2, \dots, q_p\} \subseteq C$, $(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$ and $t \in T$; these rules are supported only by objects from the P -lower approximations of the downward unions of classes Cl_t^{\leq} .

3. $D_{\geq \leq}$ -decision rules, which have the following form:

if $f(x, q_1) \geq r_{q_1}$ and $f(x, q_2) \geq r_{q_2}$ and ... $f(x, q_k) \geq r_{q_k}$ and $f(x, q_{k+1}) \leq r_{q_{k+1}}$ and ... $f(x, q_p) \leq r_{q_p}$, then $x \in Cl_t \cup Cl_{t+1} \cup \dots \cup Cl_s$,

where $O' = \{q_1, q_2, \dots, q_k\} \subseteq C$, $O'' = \{q_{k+1}, q_{k+2}, \dots, q_p\} \subseteq C$, $P = O' \cup O''$, O' and O'' not necessarily disjoint, $(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$, $s, t \in T$ such that $t < s$; these rules are supported only by objects from the P -boundaries of the unions of classes Cl_t^{\leq} and Cl_s^{\geq} .

Let us observe that the set of decision rules induced from the approximations defined using dominance relations gives, in general, a more synthetic representation of knowledge contained in the decision table than the set of rules induced from classical approximations defined using indiscernibility relations. The minimal sets of rules thus obtained have a smaller number of rules and use a smaller number of conditions. They also do not require any discretization of numerical scales of criteria prior to the rule induction. Moreover, the application of these rules to new objects gives better results, in general. This is due to the more general syntax of the rules (“ \geq ” and “ \leq ” are used instead of “ $=$ ”).

4.2.3. An example

Let us apply now the rough approximation by dominance relation to the decision table from Table 1. Within this approach we approximate the class Cl_1^{\leq} of “(at most) bad” students and the class Cl_2^{\geq} of “(at least) good” students. Since only two classes are considered, we have $Cl_1^{\leq} = Cl_1$ and $Cl_2^{\geq} = Cl_2$. As previously, $C = \{A_1, A_2, A_3\}$ and $D = \{A_4\}$. In this case, however, A_1, A_2 and A_3 are criteria and the classes are preference-ordered. This means that

- with respect to A_1 , “good” is better than “medium” and “medium” is better than “bad”,
- with respect to A_2 , “good” is better than “bad”,
- with respect to A_3 , “good” is better than “bad”,
- with respect to A_4 , “good” is better than “bad”.

The C -lower approximations, the C -upper approximations and the C -boundaries of classes Cl_1^{\leq} and Cl_2^{\geq} are equal, respectively, to:

$$\underline{C}(Cl_1^{\leq}) = \{4\}, \quad \overline{C}(Cl_1^{\leq}) = \{2, 3, 4, 5\}, \quad Bn_C(Cl_1^{\leq}) = \{2, 3, 5\},$$

$$\underline{C}(Cl_2^{\geq}) = \{1, 6\}, \quad \overline{C}(Cl_2^{\geq}) = \{1, 2, 3, 5, 6\}, \quad Bn_C(Cl_2^{\geq}) = \{2, 3, 5\}.$$

Therefore, the accuracy of the approximation is 0.25 for Cl_1^{\leq} and 0.4 for Cl_2^{\geq} , while the quality of sorting is equal to 0.5. There is only one reduct which is also the core, i.e. $Red_{Cl}(C) = Core_{Cl}(C) = \{A_1\}$.

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

1. if $f(x, A_1) \geq \text{good}$, then $x \in Cl_2^{\geq}$ (1, 6)
2. if $f(x, A_1) \leq \text{bad}$, then $x \in Cl_1^{\leq}$ (4)
3. if $f(x, A_1) \geq \text{medium}$ and $f(x, A_1) \leq \text{medium}$ (i.e. $f(x, A_1)$ is medium), then $x \in Cl_1 \cup Cl_2$ (2, 3, 5).

Let us notice that student 5 dominates student 3, i.e. student 5 is at least as good as student 3 with respect to all the three criteria, however, 5 has a comprehensive evaluation worse than 3. Therefore, this can be interpreted as an inconsistency revealed by the approximation based on dominance, that cannot be captured by the approximation based on indiscernibility.

Moreover, let us remark that the decision rules induced from approximations defined using dominance relations give a more synthetic representation of knowledge contained in the decision table. The minimal set of decision rules obtained from the dominance approach has a smaller number of stronger rules (3 against 4) and uses a smaller number of conditions (3 against 6). Furthermore, let us observe that some rules obtained from the original rough set approach make problems with their interpretation. For example, rule 3') obtained by the original rough set approach says that “if the level in Mathematics is medium and the level in Physics is good, then the student is bad”. One would expect that a student with lower marks, e.g. a student with the same level in Mathematics but with a medium level in Physics, should still be bad. Surprisingly, student 3 has these characteristics and, nevertheless, he/she is qualified as good one.

4.2.4. Another example: a comparison of rough sets with the Sugeno integral

Let us suppose that the director of the school was not satisfied with the obtained results. Therefore, after some interactions with an analyst, he made few modifications of the evaluation procedure. In consequence, the scales of the evaluation in Mathematics, Physics and Literature, as well as the global evaluation scale have been composed of three following grades: “bad”, “medium”, “good”. Moreover, a new set of examples presented in Table 4 has been considered.

According to the decision attribute, the students are divided into three preference-ordered classes: $Cl_1 = \{\text{bad}\}$, $Cl_2 = \{\text{medium}\}$, $Cl_3 = \{\text{good}\}$. Thus, the following unions of classes were approximated:

- $Cl_1^{\leq} = Cl_1$, i.e. the class of (at most) bad students,
- $Cl_2^{\leq} = Cl_1 \cup Cl_2$, i.e. the class of at most medium students,
- $Cl_2^{\geq} = Cl_2 \cup Cl_3$, i.e. the class of at least medium students,
- $Cl_3^{\geq} = Cl_3$, i.e. the class of (at least) good students.

When considering all the criteria, the unions of classes are perfectly approximated, i.e. for each union the lower and the upper approximation are the same and, therefore, coincide with the union:

$$\begin{aligned}\underline{C}(Cl_1^{\leq}) &= \overline{C}(Cl_1^{\leq}) = Cl_1^{\leq}, \\ \underline{C}(Cl_2^{\leq}) &= \overline{C}(Cl_2^{\leq}) = Cl_2^{\leq}, \\ \underline{C}(Cl_2^{\geq}) &= \overline{C}(Cl_2^{\geq}) = Cl_2^{\geq}, \\ \underline{C}(Cl_3^{\geq}) &= \overline{C}(Cl_3^{\geq}) = Cl_3^{\geq}.\end{aligned}$$

There is only one reduct which is also the core: $RED_{Cl}(C) = CORE_{Cl} = C = \{A_1, A_2, A_3\}$. Since the reduct is composed of all the criteria, each one is indispensable for precise explanation of the sorting.

The following set of minimal D_{\geq} – decision rules has been induced from Table 4:

1. “if Mathematics \geq medium and Physics \geq medium and Literature \geq medium, then student \geq medium”,
2. “if Mathematics \geq good and Physics \geq medium, then student \geq medium”,
3. “if Mathematics \geq medium and Physics \geq good, then student \geq medium”,
4. “if Mathematics \geq good and Physics \geq medium and Literature \geq medium, then student \geq good”,
5. “if Mathematics \geq medium and Physics \geq good and Literature \geq medium, then student \geq good”,
6. all uncovered students are bad.

Table 4
Decision table with examples of sorting

Student	Mathematics	Physics	Literature	Decision made by DM
1	medium	medium	bad	bad
2	good	medium	bad	medium
3	bad	good	bad	bad
4	medium	good	bad	medium
5	good	good	bad	medium
6	bad	bad	medium	bad
7	good	bad	medium	bad
8	medium	medium	medium	medium
9	good	medium	medium	good
10	medium	good	medium	good
11	good	bad	good	bad
12	medium	medium	good	medium
13	good	medium	good	good
14	bad	good	good	bad
15	medium	good	good	good

These rules permit to classify the students with all possible evaluations on the three considered criteria, as shown in Table 5 (within parentheses there are indicated the rules matching the student evaluation). If more than one D_{\geq} -decision rule matches a student evaluation, then the sorting decision corresponds to the highest suggested class.

Let us observe that the students represented in Table 5 could also be assigned to the same classes by the following set of D_{\leq} -decision rules:

- (1') “if Mathematics \leq bad, then student \leq bad”,
- (2') “if Physics \leq bad, then student \leq bad”,
- (3') “if Mathematics \leq medium and Physics \leq medium and Literature \leq bad, then student \leq bad”,
- (4') “if Literature \leq bad, then student \leq medium”,
- (5') “if Mathematics \leq medium and Physics \leq medium, then student \leq medium”,
- (6') all uncovered students are good.

It is interesting to note that the 27 decisions made by the rules and presented in Table 5 cannot be represented by the most general max–min aggregation operator permitting ordinal aggregation, i.e. the fuzzy integral proposed by Sugeno (1974). To apply the Sugeno integral a common ordinal scale must be assumed for criteria and for a fuzzy measure defined on the set of criteria. Let $\mathbf{V} = V_1 \times V_2 \times \dots \times V_n$ denote the evaluation space of the criteria. Each $x \in \mathbf{V}$ is called a profile. In our example, $\mathbf{V} = \{\text{bad, medium, good}\}^3$ and each of the 27 cases of evaluation from Table 5 corresponds to a profile. The scale value of $x \in \mathbf{V}$ on criterion c_i is denoted by $c_i(x)$, while the scale value of a subset of criteria $J = \{c_{j1}, c_{j2}, \dots, c_{jk}\}$ is denoted by $\mu(J)$. For each $x \in \mathbf{V}$, the criteria are ordered according to increasing values of $c_i(x)$

Table 5
Expression of preferences on all 27 cases

Student	Mathematics	Physics	Literature	Decision made by D_{\geq} -decision rules
1	bad	bad	bad	bad (#6)
2	medium	bad	bad	bad (#6)
3	good	bad	bad	bad (#6)
4	bad	medium	bad	bad (#6)
5	medium	medium	bad	bad (#6)
6	good	medium	bad	medium (#2)
7	bad	good	bad	bad (#6)
8	medium	good	bad	medium (#3)
9	good	good	bad	medium (#2,3)
10	bad	bad	medium	bad (#6)
11	medium	bad	medium	bad (#6)
12	good	bad	medium	bad (#6)
13	bad	medium	medium	bad (#6)
14	medium	medium	medium	medium (#1)
15	good	medium	medium	good (#1,2,4)
16	bad	good	medium	bad (#6)
17	medium	good	medium	good (#1,3,5)
18	good	good	medium	good (#1,2,3,4,5)
19	bad	bad	good	bad (#6)
20	medium	bad	good	bad (#6)
21	good	bad	good	bad (#6)
22	bad	medium	good	bad (#6)
23	medium	medium	good	medium (#1)
24	good	medium	good	good (#1,2,4)
25	bad	good	good	bad (#6)
26	medium	good	good	good (#1,3,5)
27	good	good	good	good (#1,2,3,4,5)

$c_{(1)}, c_{(2)}, \dots, c_{(n)}$, such that $c_{(1)}(x) \leq c_{(2)}(x) \leq \dots \leq c_{(n)}(x)$.

The Sugeno integral is defined as follows:

$$f[c_1(x), c_2(x), \dots, c_n(x)] = \max_{i=1, \dots, n} \{ \min \{ c_{(i)}(x), \mu(J_{(i)}) \} \},$$

where $J_{(i)} = \{c_{(i)}, \dots, c_{(n)}\}$.

Alternatively, the Sugeno integral can be presented as follows:

$$f[c_1(x), c_2(x), \dots, c_n(x)] = \max_{J \subseteq \{c_1, \dots, c_n\}} \{ \min \{ c_{j_1}(x), \dots, c_{j_k}(x), \mu(J) \} \}, \text{ where } J = \{c_{j_1}, \dots, c_{j_k}\}.$$

Greco et al. (1999j) have shown that the Sugeno integral can be represented in terms of single graded decision rules having the following syntax:

“if $c_{j_1}(x) \geq r$ and $c_{j_2}(x) \geq r$ and $\dots c_{j_h}(x) \geq r$, then $x \in Cl_r^{\geq}$ ”

and satisfying the following conditions:

- given the rule: “if $c_{j_1}(x) \geq r$ and $c_{j_2}(x) \geq r$ and $\dots c_{j_h}(x) \geq r$, then $x \in Cl_r^{\geq}$ ”, the following rules are also true for each $s < r$: “if $c_{j_1}(x) \geq s$ and $c_{j_2}(x) \geq s$ and $\dots c_{j_h}(x) \geq s$, then $x \in Cl_s^{\geq}$ ”;
- given the rule: “if $c_{j_1}(x) \geq r$ and $c_{j_2}(x) \geq r$ and $\dots c_{j_h}(x) \geq r$, then $x \in Cl_r^{\geq}$ ”, the following rules are also true for each $\{c_{i_1}, c_{i_2}, \dots, c_{i_k}\} \supseteq \{c_{j_1}, c_{j_2}, \dots, c_{j_h}\}$: “if $c_{i_1}(x) \geq r$ and $c_{i_2}(x) \geq r$ and $\dots c_{i_k}(x) \geq r$, then $x \in Cl_r^{\geq}$ ” (due to monotonicity of the fuzzy measure).

Why the 27 decisions made by the rules and presented in Table 5 cannot be represented by the Sugeno integral? This can be understood intuitively from the previous result: in fact, many of the rules applied for the sorting of the 27 cases are not single-graded, i.e. they use more than one grade of the evaluation scale. The answer can also be more direct: consider the D_{\geq} -decision rule (2): “if Mathematics \geq good and Physics \geq medium, then student \geq medium”, and check all possibilities permitting the Sugeno integral the same sorting as rule (2), without misclassification:

case 1: $\mu(\{\text{Mathematics, Physics}\}) = \text{medium}$,

case 2: $\mu(\{\text{Mathematics}\}) = \text{good}$,

case 3: $\mu(\{\text{Physics}\}) = \text{medium}$.

case 1: corresponds to the rule: “if Mathematics \geq medium and Physics \geq medium, then student \geq medium”, but it has the condition part weaker than rule (2);

case 2: corresponds to the rule: “if Mathematics \geq good, then student \geq medium”, but it has the condition part weaker than rule (2);

case 3: corresponds to the rule: “if Physics \geq medium, then student \geq medium”, but it has the condition part weaker than rule (2).

Therefore, there is no possibility to represent the sorting made by the rules in Table 5 by the Sugeno integral.

4.3. Multi-criteria sorting problem with missing values

Alike the original approach, the rough set approach based on dominance relations requires the data table to be complete. An extension of the rough set approach based on dominance to the analysis of incomplete data tables has been proposed in Greco et al. (1999e,f). It is assumed in this extension that the dominance relation between two objects is a directional statement where a subject is compared to a referent object having no missing values. The extended approach maintains all good characteristics of the dominance-based rough set approach and boils down to the latter when there are no missing values. The rules

induced from the rough approximations defined according to the extended relation are robust, i.e. each rule is supported by at least one object with no missing value on the criteria represented in the condition part of the rule.

Keeping in mind that a comparison of subject y to referent x is directional for any two objects $x, y \in U$, we say that subject y *dominates* referent x with respect to criteria from $P \subseteq C$ (denoted by $yD_P^+ x$) if for every criterion $q \in P$ the following conditions are met: (1) $f(x, q) \neq *$, (2) $f(y, q) \geq f(x, q)$ or $f(y, q) = *$.

We also say that subject y is *dominated* by referent x with respect to criteria from $P \subseteq C$ (denoted by $x D_P^- y$) if for every criterion $q \in P$ the following conditions are met: (1) $f(x, q) \neq *$, (2) $f(x, q) \leq f(y, q)$ or $f(y, q) = *$.

The above definition means that the referent object considered for dominance D_P^+ and D_P^- should have no missing values on criteria from set P .

The binary relations D_P^+ and D_P^- are not necessarily reflexive. However, D_P^+ and D_P^- are transitive.

For each $P \subseteq C$ we restore the definition of set U_P from Section 3.2. Given $P \subseteq C$ and $x \in U$, the “granules of knowledge” used for approximation are:

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

For any $P \subseteq C$ we say that $x \in U$ belongs to Cl_t^\geq *without any ambiguity* if $x \in Cl_t^\geq$ and for all the objects $y \in U$ dominating x with respect to P , we have $y \in Cl_t^\geq$, i.e. $D_P^+(x) \subseteq Cl_t^\geq$. Furthermore, we say that $x \in U$ *could belong* to Cl_t^\geq if there would exist at least one object $y \in Cl_t^\geq$ dominated by x with respect to P , i.e. $y \in D_P^-(x)$.

Thus, with respect to $P \subseteq C$, the set of all objects belonging to Cl_t^\geq without any ambiguity constitutes the *P-lower approximation* of Cl_t^\geq , denoted by $\underline{P}(Cl_t^\geq)$, and the set of all objects that could belong to Cl_t^\geq constitutes the *P-upper approximation* of Cl_t^\geq , denoted by $\overline{P}(Cl_t^\geq)$ for $t = 1, \dots, n$:

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

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

Analogously, one can define *P-lower approximation* and *P-upper approximation* of Cl_t^\leq for $t = 1, \dots, n$:

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

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

Let $(Cl_t^\geq)_P = Cl_t^\geq \cap U_P$ and $(Cl_t^\leq)_P = Cl_t^\leq \cap U_P$, $t = 1, \dots, n$. The rough approximation defined as above satisfies the following properties:

- (*Rough inclusion*). For each Cl_t^\geq and Cl_t^\leq , $t = 1, \dots, n$, and for each $P \subseteq C$:

$$\underline{P}(Cl_t^\geq) \subseteq (Cl_t^\geq)_P \subseteq \overline{P}(Cl_t^\geq), \quad \underline{P}(Cl_t^\leq) \subseteq (Cl_t^\leq)_P \subseteq \overline{P}(Cl_t^\leq);$$

- (*Complementarity*). For each Cl_t^\geq , $t = 2, \dots, n$, and Cl_t^\leq , $t = 1, \dots, n-1$, and for each $P \subseteq C$:

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

To preserve the monotonicity property of the lower approximation (see Section 3.2) it is necessary to use another definition of the approximation for a given Cl_t^\geq and Cl_t^\leq , $t = 1, \dots, n$, and for each $P \subseteq C$:

$$\underline{P}(Cl_t^\geq)^* = \bigcup_{R \subseteq P} \underline{R}(Cl_t^\geq), \quad (5)$$

$$\underline{P}(Cl_t^\leq)^* = \bigcup_{R \subseteq P} \underline{R}(Cl_t^\leq). \quad (6)$$

$\underline{P}(Cl_t^{\geq})^*$ and $\underline{P}(Cl_t^{\leq})^*$ will be called *cumulative P-lower approximations* of unions Cl_t^{\geq} and Cl_t^{\leq} , respectively, $t = 1, \dots, n$, because they include all the objects belonging to all R -lower approximations of Cl_t^{\geq} and Cl_t^{\leq} , where $R \subseteq P$.

It can be shown that another type of dominance relation, denoted by D_p^* , permits a direct definition of the cumulative P -lower approximations in a classical way. For each $x, y \in U$ and for each $P \subseteq Q$, yD_p^*x means that $f(y, q) \geq f(x, q)$ or $f(x, q) = *$ and/or $f(y, q) = *$, for every $q \in P$.

Now, given $P \subseteq C$ and $x \in U$, the “granules of knowledge” used for approximation are:

- a set of objects dominating x , called *P-dominating set*, $D_p^{+*}(x) = \{y \in U : yD_p^*x\}$,
- a set of objects dominated by x , called *P-dominated set*, $D_p^{-*}(x) = \{y \in U : xD_p^*y\}$.

D_p^* is reflexive but not transitive. Greco et al. (1999e,f) proved that definitions (5) and (6) are equivalent to the following definitions (for U_p^* see the definition in Section 3.2):

$$\underline{P}(Cl_t^{\geq})^* = \{x \in U_p^* : D_p^{+*}(x) \subseteq Cl_t^{\geq}\},$$

$$\underline{P}(Cl_t^{\leq})^* = \{x \in U_p^* : D_p^{-*}(x) \subseteq Cl_t^{\leq}\}.$$

Using D_p^* we can give definitions of the P -upper approximations of Cl_t^{\geq} and Cl_t^{\leq} , complementary to $\underline{P}(Cl_t^{\geq})^*$ and $\underline{P}(Cl_t^{\leq})^*$, respectively:

$$\overline{P}(Cl_t^{\geq})^* = \{x \in U_p^* : D_p^{-*}(x) \cap Cl_t^{\geq} \neq \emptyset\},$$

$$\overline{P}(Cl_t^{\leq})^* = \{x \in U_p^* : D_p^{+*}(x) \cap Cl_t^{\leq} \neq \emptyset\}.$$

For each $Cl_t^{\geq} \subseteq U$ and $Cl_t^{\leq} \subseteq U$, let $(Cl_t^{\geq})^* = Cl_t^{\geq} \cap U_p^*$ and $(Cl_t^{\leq})^* = Cl_t^{\leq} \cap U_p^*$. Let us remark that $x \in U_p^*$ if and only if there exists $R \neq \emptyset$ such that $R \subseteq P$ and $x \in U_R$.

Rough approximations $\underline{P}(Cl_t^{\geq})^*$, $\underline{P}(Cl_t^{\leq})^*$, $\overline{P}(Cl_t^{\geq})^*$ and $\overline{P}(Cl_t^{\leq})^*$ satisfy the following properties:

- (*Rough inclusion*). For each Cl_t^{\geq} and Cl_t^{\leq} , $t = 1, \dots, n$, and for each $P \subseteq C$:

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

- (*Complementarity*). For each Cl_t^{\geq} , $t = 2, \dots, n$, and Cl_t^{\leq} , $t = 1, \dots, n-1$, and for each $P \subseteq C$:

$$\underline{P}(Cl_t^{\geq})^* = U_p^* - \overline{P}(Cl_{t-1}^{\leq})^*, \quad \underline{P}(Cl_t^{\leq})^* = U_p^* - \overline{P}(Cl_{t+1}^{\geq})^*.$$

- (*Monotonicity of the accuracy of approximation*). For each Cl_t^{\geq} and Cl_t^{\leq} , $t = 1, \dots, n$, and for each $P, R \subseteq C$, such that $P \subseteq R$, the following inclusions hold:

$$\underline{P}(Cl_t^{\geq})^* \subseteq \underline{R}(Cl_t^{\geq})^*, \quad \underline{P}(Cl_t^{\leq})^* \subseteq \underline{R}(Cl_t^{\leq})^*.$$

Furthermore, if $U_p^* = U_R^*$, the following inclusions are also true

$$\overline{P}(Cl_t^{\geq})^* \supseteq \overline{R}(Cl_t^{\geq})^*, \quad \overline{P}(Cl_t^{\leq})^* \supseteq \overline{R}(Cl_t^{\leq})^*.$$

Due to the property of monotonicity, when augmenting a set of attributes P , we get lower approximations of Cl_t^{\geq} and Cl_t^{\leq} , $t = 1, \dots, n$, that are at least of the same cardinality. Thus, we can restore for the case of missing values the key concepts of the rough sets theory: accuracy and quality of approximation, reduct and core.

4.3.1. Example

Let us consider the example presented in Section 3.2.1 in the context of the multi-criteria sorting, where attributes are criteria with preference-ordered scales and the decision classes are also preference-ordered. We shall apply the dominance-based rough set approach to the same Table 3.

Using the extended rough set approach presented in Section 4.3, we will approximate the downward and the upward unions of classes, i.e. the class of students “at most bad” (Cl_1^{\leq}) and the class of students “at least good” (Cl_2^{\geq}). Since only two classes are considered, these unions coincide with the class of “bad” students (Cl_1) and with the class of “good” students (Cl_2), respectively.

The C -lower approximations, the C -upper approximations and the C -boundaries of the classes of “good” and “bad” students are equal, respectively, to:

$$\begin{aligned}\underline{C}(\text{good}) &= \emptyset, & \underline{C}(\text{bad}) &= \{1\}, & \overline{C}(\text{good}) &= \{6\}, & \overline{C}(\text{bad}) &= \{1\}, \\ \underline{C}(\text{good})^* &= \emptyset, & \underline{C}(\text{bad})^* &= \{1\}, & \overline{C}(\text{good})^* &= \{2, 3, 4, 5, 6\}, & \overline{C}(\text{bad})^* &= \{1, 2, 3, 4, 5, 6\}.\end{aligned}$$

Let us remark that students 2 and 6 belong to C -lower approximation of the class of “good” students when this approximation is calculated using the indiscernibility relation (see $\underline{L}_C^*(\text{good}) = \{2, 6\}$ in Section 3.2.1), however, they belong to the boundary of “good” students when this approximation is calculated using the dominance relation. This is true because there is no “bad” student indiscernible with students 2 and 6. Observe, however, that although student 5 has a comprehensive evaluation worse than students 2 and 6 (“bad” vs. “good”), he/she dominates students 2 and 6 with respect to the three criteria. Precisely, student 2 is dominated by student 5 because 2 has a worse level in Physics (medium vs. good), and on the levels in Mathematics and Literature either student 2 or student 5 has a missing value. For this reason, the assignments of students 2 and 5 are inconsistent and thus they both belong to the C -boundary of the “(at least) good” class constructed using the dominance relation. The inconsistency between student 2 and student 5 cannot be detected using the classical rough set approach based on indiscernibility because these students are discernible with respect to C . Similar explanation holds for students 5 and 6.

The quality of sorting using criteria from C is equal to 0.17. There is only one reduct which is also the core; it is composed of one criterion: {Physics}.

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

1. “if Physics \leq bad, then student is (at most) bad” (students 1,3)
2. “if Physics \geq medium and Physics \leq good, then student is bad or good” (not enough information to assign the student to one class only) (students 2,3,4,5,6).

5. Multicriteria choice and ranking problems

As pointed out above, the use of rough sets in the past has been limited to problems of multiattribute classification only. In Section 4.2, we presented an extension of the rough set approach to the multicriteria sorting problem. In the case of multicriteria choice and ranking problems we need further extensions, because the decision table in its original form does not allow the representation of preference binary relations between objects.

To handle binary relations within the rough set approach, Greco et al. (1995, 1996, 1998c) proposed to operate on, so-called, pairwise comparison table (PCT), i.e. with respect to a choice or ranking problem, a decision table whose rows represent pairs of objects for which multicriteria evaluations and a comprehensive preference relation are known.

The use of an indiscernibility relation on the PCT makes problems with interpretation of the approximations of the preference relation and of the decision rules derived from these approximations.

Indiscernibility permits handling *inconsistency* which occurs when two pairs of objects have preferences of the same strength on considered criteria, however, the comprehensive preference relations established for these pairs are not the same. When dealing with criteria, there may occur also another type of inconsistency connected with violation of the dominance principle: on a given set of criteria, one pair of objects is characterized by some preferences and another pair has all preferences at least of the same strength, however, for the first pair we have a comprehensive preference and for the other – inverse comprehensive preference. The indiscernibility relation is not able to handle this type of inconsistency. For this reason, another way of defining the approximations and decision rules has been proposed, which is based on the use of *graded dominance relations*.

5.1. The pairwise comparison table

Let C be the set of criteria used for evaluation of objects from A . For any criterion $q \in C$, let T_q 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 q , such that for any $(x, y) \in A \times A$ exactly one binary relation $\tau \in T_q$ is verified. More precisely, given the domain V_q of $q \in C$, if $v'_q, v''_q \in V_q$ are the respective evaluations of $x, y \in A$ by means of q and $(x, y) \in \tau$, with $\tau \in T_q$, then for each $w, z \in A$ having the same evaluations v'_q, v''_q by means of q , $(w, z) \in \tau$. For interesting applications it should be $\text{card}(T_q) \geq 2$, for each $q \in C$. Furthermore, let T_d be a set of binary relations defined on set A (comprehensive pairwise comparisons) such that at most one binary relation $\tau \in T_d$ is verified for any $(x, y) \in A \times A$.

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

We assume that the exemplary pairwise comparisons provided by the DM can be represented in terms of graded preference relations (for example “very weak preference”, “weak preference”, “strict preference”, “strong preference”, “very strong preference”) P_q^h : for each $q \in C$ and for any $(x, y) \in A \times A$,

$$T_q = \{P_q^h, h \in H_q\},$$

where H_q is a particular subset of the relative integers and $xP_q^h y$, $h > 0$, means that object x is preferred to object y by degree h with respect to the criterion q , $xP_q^h y$, $h < 0$, means that object x is not preferred to object y by degree h with respect to the criterion q , $xP_q^0 y$ means that object x is similar (asymmetrically indifferent) to y with respect to the criterion q .

Let us remark that P_q^0 is the same similarity relation as presented in Section 3.1 in very general terms, i.e. without any specific reference to preference modeling. Within the preference context, the similarity relation, even if not symmetric, resembles indifference relation. Thus, in this case, we call this similarity relation “asymmetric indifference”.

Notice that, for each $q \in C$ and for any $(x, y) \in A \times A$, $[xP_q^h y, h > 0] \Rightarrow [yP_q^k x, k \leq 0]$ and $[xP_q^h y, h < 0] \Rightarrow [yP_q^k x, k \geq 0]$.

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

Technically, the modeling of the binary relation P_q^h , i.e. the assessment of h , can be organized as follows:

- first, it is observed that for any $q \in C$ there exists a function $c_q : A \rightarrow \mathbf{R}$ which is increasing with respect to the preferences on q ,
- then, it is possible to define a function $k_q : \mathbf{R}^2 \rightarrow \mathbf{R}$ which measures the *strength of the preference* (positive or negative) of x over y (e.g. $k_q[c_q(x), c_q(y)] = c_q(x) - c_q(y)$); it should satisfy the following properties for any $x, y, z \in A$:

$$c_q(x) > c_q(y) \iff k_q[c_q(x), c_q(z)] > k_q[c_q(y), c_q(z)],$$

$$c_q(x) > c_q(y) \iff k_q[c_q(z), c_q(x)] < k_q[c_q(z), c_q(y)],$$

$$c_q(x) = c_q(y) \iff k_q[c_q(x), c_q(y)] = 0,$$

- next, the domain of k_q is divided into intervals, using a suitable set of thresholds Δ_q , for each $q \in C$; the intervals are numbered in such a way that $k_q[c_q(x), c_q(y)] = 0$ belongs to interval no. 0,
- the value of h in relation $xP_q^h y$ is then equal to the number of interval including $k_q[c_q(x), c_q(y)]$, for any $(x, y) \in A \times A$.

We are considering a PCT where the set T_d is composed of two binary relations defined on A :

1. x outranks y (denoted by $xS y$ or $(x, y) \in S$), where $(x, y) \in \mathbf{B}$,
2. x does not outrank y (denoted by $xS^c y$ or $(x, y) \in S^c$), where $(x, y) \in \mathbf{B}$,

and $S \cup S^c = \mathbf{B}$, where “ x outranks y ” means “ x is at least as good as y ” Roy (1985); observe that the binary relation S is reflexive, but neither necessarily transitive nor complete (Roy, 1991; Bouyssou, 1996).

5.2. Multigraded dominance

In Greco et al. (1996), we proposed a rough set approach to analysis of S_{PCT} using single-graded dominance relations, assuming common degrees of preference for all the criteria. While this permits a simple calculation of the approximations and of the resulting decision rules, it is lacking in precision. A dominance relation allowing a different degree of preference for each considered criterion gives a far more accurate picture of the preferential information contained in the pairwise comparison table S_{PCT} .

More formally, given $(x, y), (w, z) \in A \times A$, (x, y) is said to dominate (w, z) , taking into account the criteria from $\emptyset \neq P \subseteq C$ (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 $q \in P$. Precisely, “at least as strongly” means “by at least the same degree”, i.e. $h_q \geq k_q$, where $h_q, k_q \in H_q$, $xP_q^{h_q} y$ and $wP_q^{k_q} z$, for each $q \in P$. Let $D_{\{q\}}$ be the dominance relation confined to the single criterion $q \in P$. The binary relation $D_{\{q\}}$ is reflexive $((x, y)D_{\{q\}}(x, y))$, for any $(x, y) \in A \times A$, transitive $((x, y)D_{\{q\}}(w, z), (w, z)D_{\{q\}}(u, v) \implies (x, y)D_{\{q\}}(u, v))$, for any $(x, y), (w, z), (u, v) \in A \times A$, and complete $((x, y)D_{\{q\}}(w, z) \text{ and/or } (w, z)D_{\{q\}}(x, y))$, for any $(x, y), (w, z) \in A \times A$. Therefore, $D_{\{q\}}$ is a complete preorder. Since the intersection of complete preorders is a partial preorder and $D_P = \bigcap_{q \in P} D_{\{q\}}$, $P \subseteq C$, then the dominance relation D_P is a partial preorder.

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

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

Given $P \subseteq C$ and $(x, y) \in A \times A$, let us introduce the positive dominance (denoted by $D_P^+(x, y)$) and the negative dominance (denoted by $D_P^-(x, y)$)

$$D_P^+(x, y) = \{(w, z) \in A \times A : (w, z)D_P(x, y)\},$$

$$D_P^-(x, y) = \{(w, z) \in A \times A : (x, y)D_P(w, z)\}.$$

Using the dominance relations D_P , it is possible to define P -lower and P -upper approximations of the outranking relation S with respect to $P \subseteq C$, respectively, as:

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

$$\overline{P}(S) = \bigcup_{(x, y) \in S} D_P^+(x, y) = \{(x, y) \in \mathbf{B} : D_P^-(x, y) \cap S \neq \emptyset\}.$$

Analogously, it is possible to define the approximations of S^c :

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

$$\overline{P}(S^c) = \bigcup_{(x, y) \in S^c} D_P^-(x, y) = \{(x, y) \in \mathbf{B} : D_P^+(x, y) \cap S^c \neq \emptyset\}.$$

It may be proved 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:

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

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

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

Of course, $Bn_P(S) = Bn_P(S^c)$.

The concepts of accuracy, quality of approximation, reducts and core can be extended to the approximation of the outranking relation by multigraded dominance relations. The *accuracy of approximation* of S and S^c by $P \subseteq C$ is defined, respectively, by the ratios:

$$\alpha_P(S) = \frac{|\underline{P}(S)|}{|\overline{P}(S)|}, \quad \alpha_P(S^c) = \frac{|\underline{P}(S^c)|}{|\overline{P}(S^c)|}.$$

The coefficient

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

defines the quality of approximation of S and S^c by $P \subseteq C$. It expresses the ratio of all pairs of objects $(x, y) \in \mathbf{B}$ correctly assigned to S and S^c by the set P of criteria, to the number of all the pairs of objects contained in \mathbf{B} . Each minimal subset $P' \subseteq P$ such that $\gamma_{P'} = \gamma_P$ is called a *reduct* of P (denoted by $\text{RED}_S(P)$). Let us remark that \mathbf{S}_{PCT} can have more than one reduct. The intersection of all reducts is called the *core* (denoted by $\text{CORE}_S(P)$).

Using the approximations defined above, it is then possible to induce a generalized description of the preferential information contained in a given \mathbf{S}_{PCT} , in terms of suitable decision rules. The syntax of these rules is based on the concept of upward cumulated preferences (denoted by $P_q^{\geq h}$) and downward cumulated preferences (denoted by $P_q^{\leq h}$), having the following interpretation:

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

Exact definitions of the cumulated preferences, for each $(x, y) \in A \times A$, $q \in C$ and $h \in H_q$, are the following:

- $xP_q^{\geq h}y$ if $xP_q^k y$, where $k \in H_q$ and $k \geq h$,
- $xP_q^{\leq h}y$ if $xP_q^k y$, where $k \in H_q$ and $k \leq h$.

Using the above concepts, three types of decision rules can be obtained:

1. D_{\geq} -decision rules, being statements of the type:

if $xP_{q_1}^{\geq h(q_1)}y$ and $xP_{q_2}^{\geq h(q_2)}y$ and $\dots xP_{q_p}^{\geq h(q_p)}y$, then $xS^c y$,

where $P = \{q_1, q_2, \dots, q_p\} \subseteq C$ and $(h(q_1), h(q_2), \dots, h(q_p)) \in H_{q_1} \times H_{q_2} \times \dots \times H_{q_p}$; these rules are supported by pairs of objects from the P -lower approximation of S only;

2. D_{\leq} -decision rules, being statements of the type:

if $xP_{q_1}^{\leq h(q_1)}y$ and $xP_{q_2}^{\leq h(q_2)}y$ and $\dots xP_{q_p}^{\leq h(q_p)}y$, then $xS^c y$,

where $P = \{q_1, q_2, \dots, q_p\} \subseteq C$ and $(h(q_1), h(q_2), \dots, h(q_p)) \in H_{q_1} \times H_{q_2} \times \dots \times H_{q_p}$; these rules are supported by pairs of objects from the P -lower approximation of S^c only;

3. $D_{\geq \leq}$ -decision rules, being statements of the type:

if $xP_{q_1}^{\geq h(q_1)}y$ and $xP_{q_2}^{\geq h(q_2)}y$ and $\dots xP_{q_k}^{\geq h(q_k)}y$ and $xP_{q_{k+1}}^{\leq h(q_{k+1})}y$ and $\dots xP_{q_p}^{\leq h(q_p)}y$, then $xS y$ or $xS^c y$,

where $O' = \{q_1, q_2, \dots, q_k\} \subseteq C$, $O'' = \{q_{k+1}, q_{k+2}, \dots, q_p\} \subseteq C$, $P = O' \cup O''$, O' and O'' not necessarily disjoint, $(h(q_1), h(q_2), \dots, h(q_p)) \in H_{q_1} \times H_{q_2} \times \dots \times H_{q_p}$; these rules are supported by objects from the P -boundary of S and S^c only.

The decision rules, inferred from the approximation of S and S^c on A , are then applied to a set M of objects in order to obtain a recommendation for the decision problem at hand. After the application of the decision rules to each pair of objects $(u, v) \in M \times M$, one of the following four situations may occur:

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

The four above situations, which together constitute the so-called four-valued outranking (see Tsoukias and Vincke, 1995, 1997), 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 can be obtained upon a suitable exploitation of the presence and the absence of outranking S and S^c on M . A possible exploitation procedure consists in calculating a specific score, called Net Flow Score, for each object $x \in M$:

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

where

$$\begin{aligned} S^{++}(x) &= |\{y \in M : \text{there is at least one decision rule which affirms } xS y\}|, \\ S^{+-}(x) &= |\{y \in M : \text{there is at least one decision rule which affirms } yS x\}|, \\ S^{-+}(x) &= |\{y \in M : \text{there is at least one decision rule which affirms } yS^c x\}|, \\ S^{--}(x) &= |\{y \in M : \text{there is at least one decision rule which affirms } xS^c y\}|. \end{aligned}$$

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 $S_{nf}(x^*) = \max S_{nf}(x)$.

The exploitation procedure described above has been recently characterized with reference to a number of desirable properties (Greco et al., 1997, 1998), however, a thorough axiomatic analysis of this and other

exploitation procedures used to obtain a recommendation in choice and ranking problems has been carried out by Greco et al. (1997b).

Let us remark that a fuzzy extension of the multigraded approximation of relations S and S^c has been proposed by Greco et al. (1998f, 1999c).

5.3. Dominance without degrees of preference

The degree of graded preference considered in Section 5.1 is defined on a quantitative scale of the strength of preference k_q , $q \in C$. However, in many real world problems, the existence of such a quantitative scale is rather questionable. Roy (1999) distinguishes the following cases:

- preferences expressed on an ordinal scale: this is the case where the difference between two evaluations has no clear meaning;
- preferences expressed on a quantitative scale: this is the case where the scale is defined with reference to a unit clearly identified, such that it is meaningful to consider an origin (zero) of the scale and ratios between evaluations (ratio scale);
- preferences expressed on a numerical non-quantitative scale: this is an intermediate case between the previous two; there are two well-known particular cases:
 - interval scale, where it is meaningful to compare ratios between differences of pairs of evaluations,
 - scale for which a complete preorder can be defined on all possible pairs of evaluations.

The strength of preference k_q and, therefore, the graded preference considered in Section 5.1, is meaningful when the scale is quantitative or numerical non-quantitative. If the information about k_q is non-available, then it is possible to define a rough approximation of S and S^c using a specific dominance between pairs of objects from $A \times A$, defined on an ordinal scale represented by evaluations $c_q(x)$ on criterion q , for $x \in A$ (Greco et al., 1999c). Let us explain this latter case in detail.

Let C^O be the set of criteria expressing preferences on an ordinal scale, and C^N the set of criteria expressing preferences on a quantitative scale or a numerical non-quantitative scale, such that $C^O \cup C^N = C$ and $C^O \cap C^N = \emptyset$. Moreover, for each $P \subseteq C$, we denote by P^O the subset of P composed of criteria expressing preferences on an ordinal scale, i.e. $P^O = P \cap C^O$, and P^N the subset of P composed of criteria expressing preferences on a quantitative scale or a numerical non-quantitative 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 $P^O \cap P^N = \emptyset$.

If $P = P^N$ and $P^O = \emptyset$, then the definition of dominance is the same as in the case of multigraded dominance (Section 5.2). If $P = P^O$ and $P^N = \emptyset$, then, given $(x, y), (w, z) \in A \times A$, the pair (x, y) is said to dominate the pair (w, z) , with respect to criteria from P , if for each $q \in P$, $c_q(x) \geq c_q(w)$ and $c_q(z) \geq c_q(y)$. Let $D_{\{q\}}$ be the dominance relation confined to the single criterion $q \in P^O$. The binary relation $D_{\{q\}}$ is reflexive $((x, y)D_{\{q\}}(x, y))$, for any $(x, y) \in A \times A$, transitive $((x, y)D_{\{q\}}(w, z), (w, z)D_{\{q\}}(u, v) \text{ imply } (x, y)D_{\{q\}}(u, v))$, for any $(x, y), (w, z), (u, v) \in A \times A$, but non-complete (it is possible that *not* $(x, y)D_{\{q\}}(w, z)$ and *not* $(w, z)D_{\{q\}}(x, y)$ for some $(x, y), (w, z) \in A \times A$). Therefore, $D_{\{q\}}$ is a partial preorder. Since the intersection of partial preorders is also a partial preorder and $D_P = \bigcap_{q \in P} D_{\{q\}}$, $P = P^O$, then the dominance relation D_P is also a partial preorder.

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

Using the approximations of S and S^c based on the dominance relation defined above, it is possible to induce a generalized description of the available preferential information, in terms of decision rules. The decision rules are of the same type as the rules already introduced in the previous point, however, the conditions on criteria from C^O are expressed directly in terms of evaluations belonging to domains of these criteria. Let $C_q = \{c_q(x), x \in A\}$ denote the domain of ordinal criterion $q \in C^O$. The decision rules have in this case the following syntax:

1. D_{\geq} -decision rule, being a statement of the type:

$$\text{if } xP_{q1}^{\geq h(q1)}y \text{ and } \dots xP_{qe}^{\geq h(qe)}y \text{ and } c_{qe+1}(x) \geq r_{qe+1} \text{ and } c_{qe+1}(y) \leq s_{qe+1} \text{ and } \dots c_{qp}(x) \geq r_{qp} \text{ and } c_{qp}(y) \leq s_{qp}, \text{ then } xSy,$$

where $P = \{q1, \dots, qp\} \subseteq C$, $P^N = \{q1, \dots, qe\}$, $P^O = \{qe + 1, \dots, qp\}$, $(h(q1), \dots, h(qe)) \in H_{q1} \times \dots \times H_{qe}$ and $(r_{qe+1}, \dots, r_{qp}), (s_{qe+1}, \dots, s_{qp}) \in C_{qe+1} \times \dots \times C_{qp}$; these rules are supported by pairs of objects from the P -lower approximation of S only;

2. D_{\leq} -decision rule, being a statement of the type:

$$\text{if } xP_{q1}^{\leq h(q1)}y \text{ and } \dots xP_{qp}^{\leq h(qp)}y \text{ and } c_{qe+1}(x) \leq r_{qe+1} \text{ and } c_{qe+1}(y) \geq s_{qe+1} \text{ and } \dots c_{qp}(x) \leq r_{qp} \text{ and } c_{qp}(y) \geq s_{qp}, \text{ then } xS^cy,$$

where $P = \{q1, \dots, qp\} \subseteq C$, $P^N = \{q1, \dots, qe\}$, $P^O = \{qe + 1, \dots, qp\}$, $(h(q1), \dots, h(qe)) \in H_{q1} \times \dots \times H_{qe}$ and $(r_{qe+1}, \dots, r_{qp}), (s_{qe+1}, \dots, s_{qp}) \in C_{qe+1} \times \dots \times C_{qp}$; these rules are supported by pairs of objects from the P -lower approximation of S^c only;

3. $D_{\geq \leq}$ -decision rule, being a statement of the type:

$$\text{if } xP_{q1}^{\geq h(q1)}y \text{ and } \dots xP_{qe}^{\geq h(qe)}y \text{ and } xP_{qe+1}^{\leq h(qe+1)}y \dots xP_{qf}^{\leq h(qf)}y \text{ and } c_{qf+1}(x) \geq r_{qf+1} \text{ and } c_{qf+1}(y) \leq s_{qf+1} \text{ and } \dots c_{qg}(x) \geq r_{qg} \text{ and } c_{qg}(y) \leq s_{qg} \text{ and } c_{qg+1}(x) \leq r_{qg+1} \text{ and } c_{qg+1}(y) \geq s_{qg+1} \text{ and } \dots c_{qp}(x) \leq r_{qp} \text{ and } c_{qp}(y) \geq s_{qp}, \text{ then } xSy \text{ or } xS^cy,$$

where $O' = \{q1, \dots, qe\} \subseteq C$, $O'' = \{qe + 1, \dots, qf\} \subseteq C$, $P^N = O' \cup O''$, O' and O'' not necessarily disjoint, $P^O = \{qf + 1, \dots, qp\}$, $(h(q1), \dots, h(qf)) \in H_{q1} \times \dots \times H_{qf}$ and $(r_{qf+1}, \dots, r_{qp}), (s_{qf+1}, \dots, s_{qp}) \in C_{qf+1} \times \dots \times C_{qp}$; these rules are supported by pairs of objects from the P -boundary of S and S^c only.

5.4. An example

Let us consider the example (Evaluation in a High School) proposed by Grabisch (1994). The students are evaluated according to the level in Mathematics, Physics and Literature. Marks are given on a scale from 0 to 20. Three students presented in Table 6 are considered.

As the high school is “scientifically” oriented, the DM (director of the school) considers Mathematics and Physics as equally important, and more important than Literature. For this reason, he comprehensively

Table 6
Students' evaluation table

Student	Mathematics	Physics	Literature	(Comprehensive) Choquet evaluation
<i>a</i>	18	16	10	13.9
<i>b</i>	10	12	18	13.6
<i>c</i>	14	15	15	14.9

prefers student a over b . Moreover, he also comprehensively prefers student c over a , because c is as good in scientific subjects as in Literature while a is excellent in Mathematics and Physics but too bad in Literature. To represent this comprehensive evaluation, he tries to use the weighted sum model giving an equal weight to Mathematics and Physics, greater than the weight of Literature. He discovers, however, that this model will never represent his preference for c over a , because the simple sum of all marks of a and c gives the same score for them, making them indifferent; in order to obtain a better score for c , the director should give a higher weight to Literature than to Mathematics and Physics. This contradicts his original feeling about greater importance of scientific subjects than the Literature.

To solve this impossibility, Grabisch proposes to use the model of Choquet integral. The DM should give the following weights for calculation of Choquet integral:

- $\mu(\text{Mathematics}) = \mu(\text{Physics}) = 0.45$,
- $\mu(\text{Literature}) = 0.3$,
- $\mu(\text{Mathematics, Physics}) = 0.5 < \mu(\text{Mathematics}) + \mu(\text{Physics}) = 0.9$,
- $\mu(\text{Mathematics, Literature}) = 0.9 > \mu(\text{Mathematics}) + \mu(\text{Literature}) = 0.75$,
- $\mu(\text{Physics, Literature}) = 0.9 > \mu(\text{Physics}) + \mu(\text{Literature}) = 0.75$,
- $\mu(\text{Mathematics, Physics, Literature}) = 1$.

To apply the Choquet integral, for each student x the criteria are to be arranged in an order $c_{(1)}, c_{(2)}, \dots, c_{(n)}$, such that $c_{(1)}(x) \leq c_{(2)}(x) \leq \dots \leq c_{(n)}(x)$. The Choquet integral is then defined as follows:

$$C(c_1(x_1), c_2(x_2), \dots, c_n(x_n)) = \sum_{i=1}^n (c_{(i)}(x) - c_{(i-1)}(x)) \mu(J_{(i)}),$$

where $J_{(i)} = \{c_{(i)}, \dots, c_{(n)}\}$ and $c_{(0)}(x) = 0$ for each x .

In this case the Choquet integral assigns to each student the score presented in the last column of Table 6. Let us remark that this comprehensive evaluation satisfies the DM's preferences with respect to the three students and it respects his original feeling about greater importance of scientific subjects than the Literature. Moreover, the set of weights μ represents positive interaction (between Mathematics and Literature, and between Physics and Literature) and negative interaction (between Mathematics and Physics) among criteria.

Now, let us use for the same problem the rough set approach. Information from Table 6 and DM's preference-order of the students (c preferred to a preferred to b) are represented in terms of pairwise evaluations in Table 7. Note that “ x preferred to y ” means $xS^c y$ and $yS^c x$.

The Hasse diagram in Fig. 1 shows the partial preorder induced by dominance relation $D_{\{\text{Maths}\}}$ on all pairs of students with respect to the level in Mathematics.

Table 7
Pairwise comparison table

Pair of students	Mathematics	Physics	Literature	(Comprehensive) Outranking relation
(a, a)	18,18	16,16	10,10	S
(a, b)	18,10	16,12	10,18	S
(a, c)	18,14	16,15	10,15	S^c
(b, a)	10,18	12,16	18,10	S^c
(b, b)	10,10	12,12	18,18	S
(b, c)	10,14	12,15	18,15	S^c
(c, a)	14,18	15,16	15,10	S
(c, b)	14,10	15,12	15,18	S
(c, c)	14,14	15,15	15,15	S

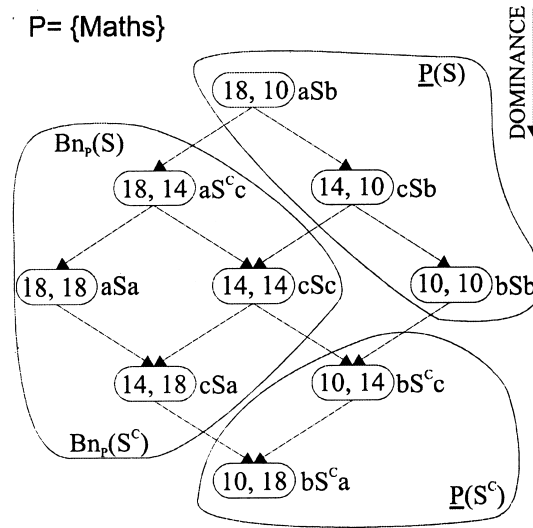


Fig. 1. Hasse diagram of the dominance relation between pairs of students with respect to {Mathematics}, and approximations of S and S^c (for the case: “ c preferred to a preferred to b ”).

The relation

$$(x, y) D_{\{\text{Maths}\}} (w, z)$$

is true if and only if the level in Mathematics of x is at least equal to the level in Mathematics of w and the level in Mathematics of y is at most equal to the level in Mathematics of z . Intuitively, the relation $D_{\{\text{Maths}\}}$ means that, with respect to Mathematics, the interval determined by evaluations of x and y is including the interval determined by evaluations of w and z , i.e. $[c_{\text{Maths}}(y), c_{\text{Maths}}(x)] \supseteq [c_{\text{Maths}}(z), c_{\text{Maths}}(w)]$. Let us remark that analogous partial preorders can be induced using dominance relations on Physics ($D_{\{\text{Physics}\}}$) and on Literature ($D_{\{\text{Lit.}\}}$).

Fig. 1 shows, moreover, lower approximations and the boundary region of S and S^c with respect to Mathematics. A pair of students (x, y) belongs to the lower approximation of S if xSy and there is no other pair of students (w, z) such that $[c_{\text{Maths}}(y), c_{\text{Maths}}(x)] \subseteq [c_{\text{Maths}}(z), c_{\text{Maths}}(w)]$, while wS^cz . Otherwise, the outranking (i.e. preference order) of pairs (x, y) and (w, z) is inconsistent with respect to the evaluation on Mathematics, so these pairs belong to the boundary region of S and S^c .

Analogously, a pair of students (x, y) belongs to the lower approximation of S^c if xS^cy and there is no other pair of students (w, z) such that $[c_{\text{Maths}}(z), c_{\text{Maths}}(w)] \subseteq [c_{\text{Maths}}(y), c_{\text{Maths}}(x)]$, while wSz . Otherwise, the outranking of the pairs (x, y) and (w, z) is inconsistent and they belong to the boundary region of S and S^c .

One can see in Fig. 1 that the boundary region of S and S^c is composed of four pairs of students $((a, a), (a, c), (c, a), (c, c))$. This means that it is not possible to approximate the outranking relation (i.e. preference-order) on all three students using the level in Mathematics only. In other words, the preference-order of students cannot be explained using Mathematics alone.

The Hasse diagram representing the dominance relation $D_{\{\text{Physics}\}}$ orders the pairs of students in the same way as $D_{\{\text{Maths}\}}$. Therefore, the lower approximations of S and S^c as well as their boundary region are the same as before. So, the information brought by the level in Physics is of the same quality as the one brought by the level in Mathematics.

As the pairs of students are ordered in the same way on the Hasse diagrams with respect to Mathematics and Physics, the conjoint consideration of Mathematics and Physics does not contribute to a better explanation of the preference-order of students.

Fig. 2 presents the Hasse diagram built on the basis of evaluations on Literature. Let us remark that there is one maximal element, the pair (b, a) . This means that for each pair (x, y) of students we have $(b, a)D_{\{Lit.\}}(x, y)$. However, $bS^c a$. As each pair (x', y') of students, for which $x'Sy'$, is dominated by (b, a) , the lower approximation of S with respect to the level in Literature is empty. The Hasse diagram has also a minimal element, the pair (a, b) , for which aSb . Therefore, also the lower approximation of S^c with respect to the level in Literature is empty. Consequently, all the pairs of students are in the boundary. In other words, the preference-order of students not only cannot be explained using Literature alone but does not give any explanation of this order.

The dominance relation $D_{\{Maths, Lit.\}}$ aggregating Mathematics and Literature puts all the pairs on the same level of the Hasse diagram, because there are no two pairs of students (x, y) and (w, z) such that $[c_{Maths}(y), c_{Maths}(x)] \supseteq [c_{Maths}(z), c_{Maths}(w)]$ and $[c_{Lit.}(y), c_{Lit.}(x)] \supseteq [c_{Lit.}(z), c_{Lit.}(w)]$. This permits to approximate perfectly the outranking relation S and its converse S^c , i.e. each pair of objects is assigned to the lower approximation of S or to the lower approximation of S^c and the boundary between S and S^c is empty. In other words, information brought by Mathematics and Literature permits to explain completely the given preference-order of students. The same result can be obtained for joint consideration of Physics and Literature. In terms of the rough sets theory, this means that there are two reducts for this problem

$$RED_1 = \{\text{Mathematics, Literature}\},$$

$$RED_2 = \{\text{Physics, Literature}\}$$

and, consequently, the core is

$$CORE = RED_1 \cap RED_2 = \{\text{Literature}\}.$$

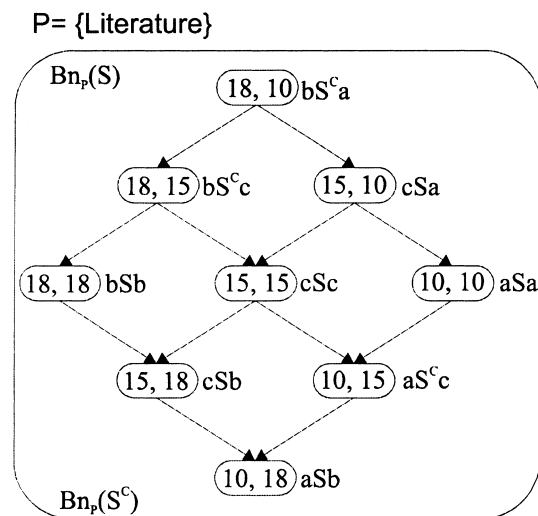


Fig. 2. Hasse diagram of the dominance relation between pairs of students with respect to $\{\text{Literature}\}$, and approximations of S and S^c (for the case: “ c preferred to a preferred to b ”).

According to the original preferential information given by the DM, Literature is not the most important criterion and, indeed, alone it is not able to explain anything about the preference-order of students. However, being the core, it is indispensable for explanation of preferences when considered together with other criteria.

The importance of and the interaction between the considered criteria can be calculated from the quality of P -approximation considered as fuzzy measure (see Section 2.5). Table 8 presents the quality of P -approximation, the Möbius index and the Shapley index with respect to considered subset P of criteria.

Results in Table 8 could be commented as follows. From the Möbius index we can remark the negative interaction (redundancy) of Mathematics and Physics ($-5/9$) and the positive interaction (synergy) of Mathematics and Literature ($4/9$), and Physics and Literature ($4/9$). The Shapley index confirms that Mathematics and Physics, ($19/54$) are more important criteria than Literature ($16/54$). It shows also a negative interaction of Mathematics and Physics ($-14/18$) and a positive interaction of Mathematics and Literature ($4/18$), and Physics and Literature ($4/18$).

It is worth noting that the quality of rough approximation and, in consequence, importance and interaction indices, are calculated from data, while the interactive weights of the Choquet integral have been given by the DM.

The following minimal set of decision rules can be induced from the considered examples (within parentheses there are pairs of students supporting the corresponding decision rules):

“if the level of x in Mathematics ≥ 10 and the level of y in Mathematics ≤ 10 , then xSy ”
 $((a, b), (c, b), (b, b))$,

“if the level of x in Mathematics ≤ 10 and the level of y in Mathematics ≥ 14 , then $xS^c y$ ” $((b, a), (b, c))$,

“if the level of x in Mathematics ≥ 14 and the level of y in Mathematics ≤ 18 and the level of x in Literature ≥ 15 and the level of y in Literature ≤ 15 , then xSy ” $((c, a), (c, c))$,

“if the level of x in Mathematics ≥ 14 and the level of y in Mathematics ≤ 18 and the level of x in Literature ≥ 10 and the level of y in Literature ≤ 10 , then xSy ” $((c, a), (a, a))$,

“if the level of x in Mathematics ≤ 18 and the level of y in Mathematics ≥ 14 and the level of x in Literature ≤ 10 and the level of y in Literature ≥ 15 , then $xS^c y$ ” $((a, c))$.

Now, let us suppose that there is a committee composed of three members (I, II, III) giving three different rankings of students, presented in Table 9.

Table 8
Importance and interaction indices^a

Subset P of criteria	Quality of rough approximation	Möbius index	Shapley index
M	5/9	5/9	19/54
Ph	5/9	5/9	19/54
L	0	0	16/54
M+Ph	5/9	$-5/9$	$-14/18$
M+L	1	4/9	4/18
Ph+L	1	4/9	4/18
M+Ph+L	1	$-4/9$	$-4/9$

Table 9
Rankings of students given by members of the committee

Student	Member I	Member II	Member III
a	2°	3°	1°
b	3°	1°	2°
c	1°	2°	3°

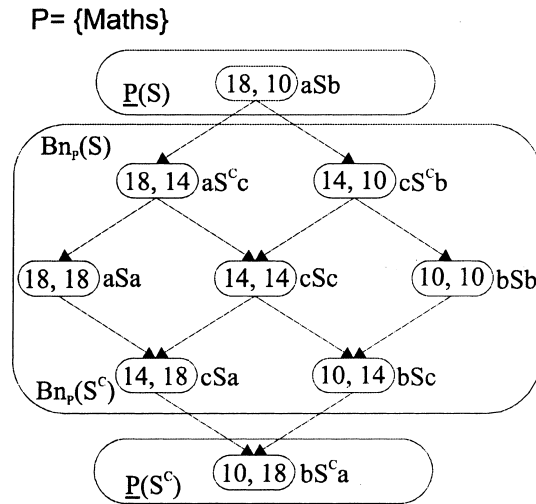


Fig. 3. Hasse diagram of the dominance relation between pairs of students with respect to {Mathematics}, and approximations of S and S^c (for the case: “ c preferred to a preferred to b preferred to c ”).

Let us assume that the committee takes final decision according to majority rule. Therefore, c is comprehensively preferred to a (because I and II vote in favor of c and III in favor of b), a is preferred to b (because I and III vote in favor of a and II in favor of b) and b is preferred to c (because II and III vote in favor of b and I in favor of c). There is a cycle in the preference-order of the commission: it is the very well-known Condorcet paradox.

Also in this case it is possible to represent lower and upper approximations of S and S^c on the basis of the Hasse diagrams used before. Fig. 3 shows the Hasse diagram with respect to Mathematics. The lower approximations of S and S^c are composed of only one pair of students each, (a, b) and (b, a) , respectively. This means that using information brought by Mathematics it is possible to explain the preference relative to two pairs of students only: (a, b) and (b, a) . Physics behaves similarly to Mathematics. Fig. 4 shows the Hasse diagram with respect to Literature and the lower approximations of S and S^c that are empty. Therefore, Literature alone is again not able to explain anything about the preference-order of students. Also in this case, information brought by Mathematics and Literature together permits to explain completely the preference-order of students. The same result can be obtained for joint consideration of Physics and Literature. In terms of the rough sets theory this means that there are two reducts for this problem

$$\text{RED}_1 = \{\text{Mathematics, Literature}\},$$

$$\text{RED}_2 = \{\text{Physics, Literature}\}$$

and, consequently, the core is

$$\text{CORE} = \text{RED}_1 \cap \text{RED}_2 = \{\text{Literature}\}.$$

As before, the importance of and the interaction between the considered criteria can be calculated from the quality of P -approximation considered as fuzzy measure. Table 10 presents the quality of P -approximation, the Möbius index and the Shapley index with respect to considered subset P of criteria.

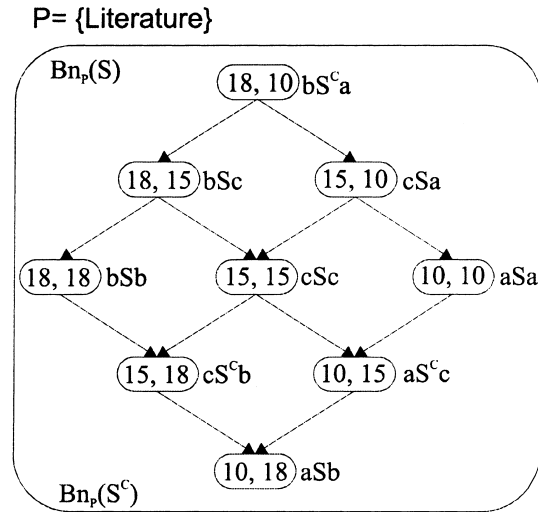


Fig. 4. Hasse diagram of the dominance relation between pairs of students with respect to $\{\text{Literature}\}$, and approximations of S and S^c (for the case: “ c preferred to a preferred to b preferred to c ”).

Table 10
Importance and interaction indices

Subset P of criteria	Quality of rough approximation	Möbius index	Shapley index
M	2/9	2/9	13/54
Ph	2/9	2/9	13/54
L	0	0	28/54
M+Ph	2/9	-2/9	-11/18
M+L	1	7/9	7/18
Ph+L	1	7/9	7/18
M+Ph+L	1	-7/9	-7/9

Let us observe that in this case the importance of Literature measured by the Shapley index increases and Literature becomes the most important criterion. Also the synergy of Mathematics and Literature or Physics and Literature increases.

Finally, the following minimal set of decision rules can explain the decision policy of the commission (between parentheses there are pairs of students supporting the corresponding decision rules):

- “if the level of x in Mathematics ≥ 18 and the level of y in Mathematics ≤ 10 , then xSy ” $((a, b))$,
- “if the level of x in Mathematics ≤ 10 and the level of y in Mathematics ≥ 18 , then $xS^c y$ ” $((b, a))$,
- “if the level of x in Mathematics ≥ 10 and the level of y in Mathematics ≤ 14 and the level of x in Literature ≥ 18 and the level of y in Literature ≤ 18 , then xSy ” $((b, c), (b, b))$,
- “if the level of x in Mathematics ≥ 14 and the level of y in Mathematics ≤ 18 and the level of x in Literature ≥ 15 and the level of y in Literature ≤ 15 , then xSy ” $((c, a), (c, c))$,
- “if the level of x in Mathematics ≥ 14 and the level of y in Mathematics ≤ 18 and the level of x in Literature ≥ 10 and the level of y in Literature ≤ 10 , then xSy ” $((c, a), (a, a))$,
- “if the level of x in Mathematics ≤ 14 and the level of y in Mathematics ≥ 10 and the level of x in Literature ≤ 15 and the level of y in Literature ≥ 18 , then $xS^c y$ ” $((c, b))$,
- “if the level of x in Mathematics ≤ 18 and the level of y in Mathematics ≥ 14 and the level of x in Literature ≤ 10 and the level of y in Literature ≥ 15 , then $xS^c y$ ” $((a, c))$.

The example shows that the Choquet integral is not able to represent the case of cyclic preferences, while the rough set approach is able.

6. Formal equivalence of decision rule preference models and conjoint measurement models

Traditionally, preferences are modeled using a value function $u(\cdot)$. In a multicriteria context, each object a is generally seen as a vector $c(a) = (c_1(a), c_2(a), \dots, c_m(a))$ of evaluations with reference to the m criteria $c_1(a), c_2(a), \dots, c_m(a)$. Greco et al. (1999j) characterized recently such a function for multicriteria sorting. They proved that a simple cancellation property permits to induce a preference order on the domain of each criterion from the order of considered classes $Cl_t, t = 1, \dots, n$. This is equivalent to the following model of sorting

$$u[c_1(a), c_2(a), \dots, c_m(a)] \geq z_t \iff a \in Cl_t^{\geq},$$

where $u(\cdot)$ is increasing in each argument and $n - 1$ ordered thresholds $z_t, t = 2, \dots, n$, satisfy the condition $z_2 \leq z_3 \leq \dots \leq z_n$. The authors proved, moreover, that such a model is equivalent to a sorting produced by a set of D_{\geq} -decision rules having the syntax defined in Section 4.2.2.

The above model of sorting can also be written as

$$u[c_1(a), c_2(a), \dots, c_m(a)] \leq w_t \iff a \in Cl_t^{\leq},$$

where $u(\cdot)$ is the same as above and $n - 1$ ordered thresholds $w_t, t = 1, \dots, n - 1$, satisfy the condition $w_1 \leq w_2 \leq \dots \leq w_{n-1}$. This model is equivalent to a sorting produced by a set of D_{\leq} -decision rules having the syntax defined in Section 4.2.2.

The authors proved, finally, that in presence of some inconsistent examples of sorting, i.e. if at least one C -boundary is nonempty, this model can be generalized as follows: there are two value functions, $u^{\geq}(\cdot)$ and $u^{\leq}(\cdot)$, increasing in each argument, which assign a lower and an upper value to each object a , respectively, i.e.

$$u^{\geq}[c_1(a), c_2(a), \dots, c_m(a)] \leq u^{\leq}[c_1(a), c_2(a), \dots, c_m(a)].$$

Moreover, there are $n - 1$ ordered thresholds $z_t, t = 2, \dots, n$, and $n - 1$ ordered thresholds $w_t, t = 1, \dots, n - 1$, satisfying

$$w_1 \leq z_2 \leq w_2 \leq \dots \leq z_{n-1} \leq w_{n-1} \leq z_n$$

such that for each object a

$$u^{\geq}[c_1(a), c_2(a), \dots, c_m(a)] \geq z_t \iff a \in \underline{C}(Cl_t^{\geq}),$$

$$u^{\leq}[c_1(a), c_2(a), \dots, c_m(a)] \leq w_t \iff a \in \underline{C}(Cl_t^{\leq}).$$

If there is no inconsistent example of sorting, the above model boils down to the previous ones.

When used for multicriteria ranking or choice problems, value function $u(\cdot)$ is requested to satisfy the property that object a is at least as good as object b , i.e. aSb , iff $u(a) \geq u(b)$. This implies that the relation S is complete (for each couple of objects a, b , aSb and/or bSa) and transitive (for each triple of objects a, b, c , aSb and bSc imply aSc). It is often assumed that the value function is additive (see, e.g., Keeney and Raiffa, 1976, and Krantz et al., 1978; Wakker, 1989, for an axiomatic characterization), i.e.,

$$u(a) = \sum_{q=1}^m u_q[c_q(a)],$$

where $u_q, q = 1, \dots, m$, are non-decreasing functions.

The additive and transitive model represented by the additive value function is inappropriate in many situations, because:

- the indifference (the symmetric part of S) may not be transitive,
- S may not be complete, that is some objects may be incomparable,
- the compensation between evaluations of conflicting criteria is far more complex than the capacity of representation by the additive value function.

To take these limitations into account, a variety of extensions have been proposed (e.g. Tversky, 1969, Fishburn, 1991). Bouyssou and Pirlot (1996) have proposed a model generalizing the previous ones and creating an axiomatic basis to many multicriteria decision methods proposed in the literature (see, e.g. Roy, 1993; Vincke, 1992). This model drops additivity, transitivity and completeness properties, and may be written as

$$(i) \quad aSb \text{ iff } F[\Psi_q(c_q(a), c_q(b)), q=1, \dots, m] \geq 0,$$

where $\Psi_q : \mathbf{R}^2 \rightarrow \mathbf{R}$ is a non-decreasing function in its first argument and a non-increasing function in its second argument, for $q = 1, \dots, m$, and $F : \mathbf{R}^m \rightarrow \mathbf{R}$ is a non-decreasing function in every one of its arguments. Observe that the values assumed by the function Ψ_q may be interpreted as a measure of the strength of preference of a over b with respect to criterion $q, q = 1, \dots, m$. Thus, Ψ_q plays the same role as the function k_q in the definition of the PCT.

Recently, Greco et al. (1998h, 1999b) have proposed some more general models of conjoint measurement. The first model can be written as (Greco et al., 1998h):

$$(ii) \quad aSb \text{ iff } G[\Psi_q(c_q(a), c_q(b)), q=1, \dots, k, c_q(a), c_q(b), q=k+1, \dots, m] \geq 0,$$

where the indices of the considered criteria are reordered such that $\{1, \dots, k\}$ is the set of criteria for which the preference is expressed on a quantitative or a numerical non-quantitative scale, and $\{k+1, \dots, m\}$ is the set of criteria for which the preference is expressed on an ordinal scale; Ψ_q is defined as above for $q = 1, \dots, k$, and $G : \mathbf{R}^{k+2(m-k)} \rightarrow \mathbf{R}$ is a non-decreasing function in its first k arguments, non-decreasing in each $(k+\text{'odd'})$ argument ('odd' = $1, 3, \dots, 2(m-k) - 1$) and non-increasing in each $(k+\text{'even'})$ argument ('even' = $2, 4, \dots, 2(m-k)$). We proved that model (ii) is based on the same axioms as the model (i), with the exception of an axiom which introduces a total preorder in the set of pairs $(c_q(a), c_q(b))$, for each $q \in \{1, \dots, m\}$. More precisely, this axiom is accepted only for $q = 1, \dots, k$, i.e. for the set of criteria with a quantitative or a numerical non-quantitative preference scale.

Moreover, Greco et al. (1999b) have proposed a model of conjoint measurement to represent some inconsistencies in the preferences. This model is based on the concepts of C -lower and C -upper approximation of S and S^c , and of C -boundary of S and S^c , where $C = \{1, \dots, m\}$. This model can be written as

$$(iiia) \quad (a, b) \in \underline{C}(S) \text{ iff } G[\Psi_q(c_q(a), c_q(b)), q=1, \dots, k, c_q(a), c_q(b), q=k+1, \dots, m] \geq t_2,$$

$$(iiib) \quad (a, b) \in \underline{C}(S^c) \text{ iff } G[\Psi_q(c_q(a), c_q(b)), q=1, \dots, k, c_q(a), c_q(b), q=k+1, \dots, m] \leq t_1,$$

$$(iiic) \quad (a, b) \in Bn_C(S) \text{ (or, equivalently, } (a, b) \in Bn_C(S^c))$$

$$\text{iff } t_1 < G[\Psi_q(c_q(a), c_q(b)), q=1, \dots, k, c_q(a), c_q(b), q=k+1, \dots, m] < t_2,$$

where Ψ_q and G are defined as above and $t_1, t_2 \in \mathbf{R}$ such that $t_1 < t_2$. With respect to the model (iiia)–(iiic), we proved that it is always possible to obtain that representation, i.e. S should not satisfy any specific axiom.

It is interesting to compare the above models of conjoint measurement with the *decision rule preference models* resulting from the rough set approach (Sections 5.2 and 5.3). The following results have been proved:

1. the outranking relation S may be represented by means of the non-additive, non-transitive and non-complete model (i) *if and only if* it may be represented by means of a set of D_{\geq} -decision rules having a syntax defined in Section 5.2 (Greco et al., 1998d),
2. the outranking relation S may be represented by means of the non-additive, non-transitive and non-complete model (ii) *if and only if* it may be represented by means of a set of D_{\geq} -decision rules having a syntax defined in Section 5.3 (Greco et al., 1998h).

Greco et al. (1999c) have also pointed out the clear equivalence between the representation of S and S^c obtained using the rough set approach proposed in Section 5 and the model of conjoint measurement (iiia)–(iiic). Furthermore, they observed that the rough set representation presented in Section 5.2 can be viewed as a particular case of the representation proposed in Section 5.3, when the set of criteria with an ordinal preference scale is empty.

7. Conclusions

In this paper, we made a synthesis of the contribution of the extended rough sets theory to MCDA. Classical use of the rough set approach, and more generally, of machine learning, data mining and knowledge discovery, is confined to problems of multiattribute classification, i.e. problems where neither the values of attributes describing the objects, nor the classes to which the objects are assigned, are preference-ordered. On the other hand, MCDA deals with problems where descriptions (evaluations) of objects (actions) by means of attributes (criteria), as well as decisions in sorting, choice and ranking problems, are preference-ordered. The extension of the rough set approach to problems in which preference-order properties are important is possible upon two main methodological contributions extensively discussed in this paper:

1. approximation by dominance relations, which allows to deal with preference-order properties of criteria,
2. analysis of pairwise comparison table, which allows to handle preference relations for choice and ranking problems.

Let us point out the main advantages of the extended rough set approach to MCDA in comparison with classical approaches:

- preferential information necessary to deal with a multicriteria decision problem is asked to the DM 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 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),
- the preference model induced from the preferential information is expressed in a natural and comprehensible language of “*if ... , then ...*” decision rules,
- heterogeneous information (qualitative and quantitative, preference-ordered or not, crisp and fuzzy evaluations, and ordinal and cardinal scales of preferences, missing values) can be processed within the extended rough set approach, while classical MCDA methods consider only quantitative ordered evaluations with rare exceptions,
- the decision rule preference model resulting from the rough set approach is more general than all existing models of conjoint measurement due to its capacity of handling inconsistent preferences (a new model of

conjoint measurement is formally equivalent to the decision rule preference model handling inconsistencies),

- 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, in a certain sense, even objective: dominance relation.

Let us conclude with a metaphor. Rough sets theory and multicriteria decision theory were like small worlds speaking their own languages. The language of rough sets did not include the words like criterion, choice, ranking, sorting. On the other hand, the language of multicriteria decision theory did not use words like approximation, reduct, core, decision rule. The results presented in this paper permitted rough sets theory and MCDA to communicate in a new language in which these words coexist and are semantically related. This communication permitted to create an added value for both theories. The rough sets theory entered the world of decision problems in which preference-orders are considered. Multicriteria decision analysis was equipped with new preference models composed of decision rules. Moreover, even if the concept of inconsistency was known in MCDA, there were only few tools to deal with it. The use of the decision rule model and the capacity of handling inconsistent preferential information opened a fascinating research field to MCDA.

Acknowledgements

The research of the first two authors has been supported by the Italian Ministry of University and Scientific Research (MURST). The third author wishes to acknowledge financial support from State Committee for Scientific Research (KBN), grant no. 8T11F00619.

References

- Bazan, J., Skowron, A., Synak, P., 1994. Dynamic reducts as a tool for extracting laws from decision tables. In: Zemankowa, M., Ras, Z. (Eds.), *Methodologies for Intelligent Systems*. LNAI, vol. 869. Springer, Berlin, pp. 346–355.
- Banzhaf, J.F., 1965. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review* 19, 317–343.
- Bouyssou, D., 1996. Outranking Relations: Do they have special properties? *Journal of Multi-Criteria Decision Analysis* 5 (2), 99–111.
- Bouyssou, D., Pirlot, M., 1996. A general framework for the aggregation of semiorders. Technical Report, ESSEC, Cergy-Pontoise.
- Dennenberg, D., Grabisch, M., 1996. Shapley value and interaction index. Working paper.
- Dubois, D., Prade, H., 1990. Rough fuzzy sets and fuzzy rough sets. *Int. J. of General Systems* 17, 191–200.
- Dubois, D., Prade, H., 1992. Putting rough sets and fuzzy sets together. In: Slowinski, R. (Ed.), *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*. Kluwer, Dordrecht, pp. 203–233.
- Fishburn, P.C., 1967. Methods for estimating additive utilities. *Management Science* 13, 435–453.
- Fishburn, P.C., 1991. Nontransitive additive conjoint measurement. *Journal of Mathematical Psychology* 35, 1–40.
- Fodor, J., Roubens, M., 1994. Fuzzy preference modelling and multicriteria decision support. Kluwer, Dordrecht.
- Grabisch, M., 1994. Fuzzy integral in multicriteria decision making. *Fuzzy Sets and Systems* 89, 279–298.
- Grabisch, M., 1996. The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research* 89, 445–456.
- Grabisch, M., 1997. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems* 89, 445–456.
- Grabisch, M., Roubens, M., 1997. Equivalent representations of a set function with application to decision making. In: *Proceedings of the FUZZ-IEEE '97 Conference*. Barcelona.
- Greco, S., Matarazzo, B., Slowinski, R., 1995. Rough set approach to multi-attribute choice and ranking problems. ICS Research Report 38/95, Warsaw University of Technology, Warsaw 1995. In: Fandel, G., Gal, T. (Eds.), *Multiple Criteria Decision Making, Proceedings of the 12th International Conference*. Hagen, Springer, Berlin, 1997, pp. 318–329.
- Greco, S., Matarazzo, B., Slowinski, R., 1996. Rough approximation of a preference relation by dominance relations, ICS Research Report 16/96, Warsaw University of Technology, Warsaw, 1996, *European Journal of Operational Research* 117, 1999, 63–83.
- Greco, S., Matarazzo, B., Slowinski, R., 1997a. Rough approximation of a preferential information. Working Paper, Poznan University of Technology.

- Greco, S., Matarazzo, B., Slowinski, R., 1997b. Exploitation procedures for rough set analysis of multicriteria decision problems. Working Paper, Poznan University of Technology.
- Greco, S., Matarazzo, B., Slowinski, R., 1998a. Fuzzy measures technique for rough set analysis. In: *Proceedings of the Sixth European Congress on Intelligent Techniques and Soft Computing*, vol. 1. Aachen, pp. 99–103.
- Greco, S., Matarazzo, B., Slowinski, R., 1998b. A new rough set approach to evaluation of bankruptcy risk. In: Zopounidis, C. (Ed.), *Operational Tools in the Management of Financial Risks*. Kluwer, Dordrecht, pp. 121–136.
- Greco, S., Matarazzo, B., Slowinski, R., 1998c. Rough approximation of a preference relation in a pairwise comparison table. In: Polkowski, L., Skowron, A. (Eds.), *Rough Sets in Data Mining and Knowledge Discovery*. Physica-Verlag, Heidelberg, pp. 13–36.
- Greco, S., Matarazzo, B., Slowinski, R., 1998d. Modellizzazione delle preferenze per mezzo di regole di decisione. In: *Atti del Ventiduesimo Convegno A.M.A.S.E.S.*, Bozzi Editore, Genova, pp. 233–247.
- Greco, S., Matarazzo, B., Slowinski, R., 1998e. A new rough set approach to multicriteria and multiattribute classification. In: Polkowski, L., Skowron, A. (Eds.), *Rough sets and Current Trends in Computing*, RSTCTC'98, Springer, pp. 60–67.
- Greco, S., Matarazzo, B., Slowinski, R., 1998f. Rough approximation of a fuzzy preference relation. Working Paper, Poznan University of Technology.
- Greco, S., Matarazzo, B., Slowinski, R., 1998g. Fuzzy similarity relation as a basis for rough approximation. In: Polkowski, L., Skowron, A. (Eds.), *Rough sets and Current Trends in Computing*, RSTCTC '98, Springer, pp. 283–289.
- Greco, S., Matarazzo, B., Slowinski, R., 1998h. A conjoint measurement model to represent preference on ordinal scales. Working paper, Poznan University of Technology.
- Greco, S., Matarazzo, B., Slowinski, R., 1999a. Fuzzy dominance as basis for rough approximations. In: *Proceedings of the Fourth Meeting of the EURO WG on Fuzzy Sets and Second International Conference on Soft and Intelligent Computing*, EUROFUSE-SIC'99, Budapest, pp. 273–278.
- Greco, S., Matarazzo, B., Slowinski, R., 1999b. Misurazione congiunta e incoerenze nelle preferenze. In: *Atti del ventitreesimo convegno A.M.A.S.E.S.*, Rende-Cosenza, pp. 255–269.
- Greco, S., Matarazzo, B., Slowinski, R., 1999c. The use of rough sets and fuzzy sets in MCDM. Chapter 14. In: Gal, T., Stewart, T., Hanne, T. (Eds.), *Advances in Multiple Criteria Decision Making*. Kluwer, Dordrecht, pp. 14.1–14.59.
- Greco, S., Matarazzo, B., Slowinski, R., 1999d. On joint use of indiscernibility, similarity and dominance in rough approximation of decision classes. In: Despotis, D.K., Zopounidis, C. (Eds.), *Proceedings of the Fifth International Conference of the Decision Sciences Institute*. Athens, pp. 1380–1382.
- Greco, S., Matarazzo, B., Slowinski, R., 1999e. Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems. In: Zhong, N., Skowron, A., Ohsuga, S. (Eds.), *New Directions in Rough Sets, Data Mining and Granular-Soft Computing*, RSFDGrC'99, *Lecture Notes in Artificial Intelligence*, vol. 1711. Springer, Berlin, pp. 146–157.
- Greco, S., Matarazzo, B., Slowinski, R., 1999f. Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems. In: Zanakakis, S.H., Doukidis, G., Zopounidis, C. (Eds.), *Recent Developments and Applications in Decision Making*. Kluwer Academic Publishers, Dordrecht, to appear.
- Greco, S., Matarazzo, B., Slowinski, R., 1999g. Multicriteria classification by dominance-based rough set approach. Chapter C5.1.9. In: Kloesgen, W., Zytkow, J. (Eds.), *Handbook of Data Mining and Knowledge Discovery*, Oxford University Press, New York, to appear.
- Greco, S., Matarazzo, B., Slowinski, R., 1999h. Dominance-based rough set approach to rating analysis. *Fuzzy Economic Review*, to appear.
- Greco, S., Matarazzo, B., Slowinski, R., 1999j. Conjoint measurement and rough set approach for multicriteria sorting problems in presence of ordinal criteria. In: Colomi, A., Parruccini, M., Roy, B. (Eds.), *Selected Papers from 49th and 50th Meeting of the EURO Working Group on MCDA*, EUR-Report, Ispra-Paris, to appear.
- Greco, S., Matarazzo, B., Slowinski, R., 1999k. A fuzzy extension of the rough set approach to multicriteria and multiattribute sorting. In: Fodor, J., De Baets, B., Perny, P. (Eds.), *Preferences and Decisions under Incomplete Information*. Physica, Heidelberg, to appear.
- Greco, S., Matarazzo, B., Slowinski, R., 2000. Rough set processing of vague information using fuzzy similarity relations. In: Calude, C.S., Paun, G. (Eds.), *Finite Versus Infinite – Contributions to an Eternal Dilemma*. Springer, London, pp. 149–173.
- Greco, S., Matarazzo, B., Slowinski, R., Tsoukias, A., 1997. Exploitation of a rough approximation of the outranking relation. *Cahier du LAMSADE no. 152*, Université de Paris Dauphine, Paris.
- Greco, S., Matarazzo, B., Slowinski, R., Tsoukias, A., 1998. Exploitation of a rough approximation of the outranking relation in multicriteria choice and ranking. In: Stewart, T.J., van den Honert, R.C. (Eds.), *Trends in Multicriteria Decision Making*. LNEMS, vol. 465. Springer, Berlin, pp. 450–460.
- Greco, S., Matarazzo, B., Slowinski, R., Zanakakis, S., 1999l. Rough set analysis of information tables with missing values. In: Despotis, D.K., Zopounidis, C. (Eds.), *Proceedings of the Fifth International Conference of the Decision Sciences Institute*. Athens, pp. 1359–1362.
- Grzymala-Busse, J.W., 1992. LERS - a system for learning from examples based on rough sets. In: Slowinski, R. (Ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*. Kluwer, Dordrecht, pp. 3–18.

- Grzymala-Busse, J.W., 1997. A new version of the rule induction system LERS. *Fundamenta Informaticae* 31, 27–39.
- Inuiguchi, M., Tanino, T., 2000. Fuzzy rough sets based on certainty qualifications. In: *Proceedings of the 4th AFSS Symp.*, Tsukuba, pp. 433–438.
- Jacquet-Lagrèze, E., 1981. *Systèmes de décision et acteurs multiples - Contribution à une théorie de l'action pour les sciences des organisations*, Thèse d'Etat, Université de Paris-Dauphine, Paris.
- Jacquet-Lagrèze, E., Siskos, J., 1982. Assessing a set of additive utility functions for multicriteria decision-making, the UTA method. *European Journal of Operational Research* 10, 151–164.
- Keeney, R.L., Raiffa, H., 1976. *Decision with Multiple Objectives - Preferences and value Tradeoffs*. Wiley, New York.
- Krantz, D.M., Luce, R.D., Suppes, P., Tversky, A., 1978. *Foundations of Measurements I*. Academic Press, New York.
- Krawiec, K., Slowinski, R., Vanderpooten, D., 1998. Learning of decision rules from similarity based rough approximations. In: Polkowski, L., Skowron, A. (Eds.), *Rough Sets in Knowledge Discovery*, vol. 2. Physica, Heidelberg, pp. 37–54.
- Kryszkiewicz, M., 1998. Properties of incomplete information systems in the framework of rough sets. In: Polkowski, L., Skowron, A. (Eds.), *Rough Sets in Knowledge Discovery*, vol. 1. Physica, Heidelberg, pp. 422–450.
- Krusinska, E., Slowinski, R., Stefanowski, J., 1992. Discriminant versus rough set approach to vague data analysis. *Applied Stochastic Models and Data Analysis* 8, 43–56.
- Kryszkiewicz, M., Rybinski, H., 1996. Computation of reducts of composed information systems. *Fundamenta Informaticae* 27, 183–195.
- Langley, P., Simon, H.A., 1998. Fielded applications of machine learning. In: Michalski, R.S., Bratko, I., Kubat, M. (Eds.), *Machine Learning and Data Mining*. Wiley, New York, pp. 113–129.
- Lin, T., 1989. Neighborhood systems and approximation in database and knowledge base systems. In: *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems*.
- Luce, R.D., 1956. Semi-orders and a theory of utility discrimination. *Econometrica* 24, 178–191.
- March, J.G., 1988. Bounded rationality, ambiguity, and the engineering of choice. In: Bell, D.E., Raiffa, H., Tversky, A. (Eds.), *Decision Making, Descriptive, Normative and Prescriptive Interactions*. Cambridge University Press, New York, pp. 33–58.
- Marcus, S., 1994. Tolerance rough sets, Cech topologies, learning processes. *Bull. of the Polish Academy of Sciences Technical Sciences* 42 (3), 471–487.
- Michalski, R.S., Bratko, I., Kubat, M. (Eds.), 1998. *Machine Learning and Data Mining – Methods and Applications*. Wiley, New York.
- Mienko, R., Slowinski, R., Stefanowski, J., Susmaga, R., 1996a. Rough Family – software implementation of rough set based data analysis and rule discovery techniques. In: Tsumoto, S., et al. (Eds.), *Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets and Machine Discovery*. Tokyo University Press, Tokyo, pp. 437–440.
- Mienko, R., Stefanowski, J., Toumi, K., Vanderpooten, D., 1996b. Discovery-oriented induction of decision rules. *Cahier du LAMSADE no. 141*, Université de Paris Dauphine, Paris.
- Mousseau, V., 1993. *Problèmes liés à l'évaluation de l'importance en aide multicritère à la décision: Réflexions théoriques et expérimentations*, Thèse de doctorat, Université de Paris-Dauphine, Paris.
- Murofushi, T., 1992. A technique for reading fuzzy measures (i): the Shapley value with respect to a fuzzy measure. In: *Proceedings of the Second Fuzzy Workshop*. Nagaoka, Japan, October, pp. 39–48, in Japanese.
- Murofushi, T., Soneda, S., 1993. Techniques for reading fuzzy measures (iii): interaction index. In: *Proceedings of the Ninth Fuzzy Systems Symposium*. Sapporo, Japan, May 1993, pp. 693–696, in Japanese.
- Nieminen, J., 1988. Rough tolerance equality. *Fundamenta Informaticae* 11 (3), 289–296.
- Pawlak, Z., 1982. Rough sets. *International Journal of Information & Computer Sciences* 11, 341–356.
- Pawlak, Z., 1985a. Rough probability. *Bull. Polish Acad. Scis., Technical Sci.* 33, 9–10.
- Pawlak, Z., 1985b. Rough sets and fuzzy sets. *Fuzzy Sets and Systems* 17, 99–102.
- Pawlak, Z., 1991. *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer, Dordrecht.
- Pawlak, Z., 1997. Rough set approach to knowledge-based decision support. *European Journal of Operational Research* 99, 48–57.
- Pawlak, Z., Slowinski, R., 1994. Rough set approach to multi-attribute decision analysis. *European Journal of Operational Research* 72, 443–459.
- Polkowski, L., Skowron, A., 1994. Rough mereology. In: *Proceedings Symposium on Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence*, vol. 869. Springer, Berlin, pp. 85–94.
- Polkowski, L., Skowron, A. (Eds.), 1998. *Rough Sets in Data Mining and Knowledge Discovery*. Physica, Heidelberg.
- Polkowski, L., Skowron, A., Zytkow, J., 1995. Rough foundations for rough sets. In: Lin, T.Y., Wildberger, A. (Eds.), *Soft Computing: Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery*. Simulation Councils, San Diego, CA, pp. 142–149.
- Roubens, M., 1996. Interaction between criteria through the use of fuzzy measures, Report 96.007, Institut de Mathématique, Université de Liège, Liège.
- Roy, B., 1985. *Méthodologie Multicritère d'Aide à la Décision*. Economica, Paris.
- Roy, B., 1989. Main sources of inaccurate determination, uncertainty and imprecision in decision models. *Mathematical and Computer Modelling* 12 (10/11), 1245–1254.

- Roy, B., 1991. The outranking approach and the foundation of ELECTRE methods. *Theory and Decision* 31, 49–73.
- Roy, B., 1993. Decision science or decision aid science?. *European Journal of Operational Research* 66, 184–203.
- Roy, B., 1999. Decision-aiding today: what should we expect? Chapter 1. In: Gal, T., Stewart, T., Hanne, T. (Eds.), *Advances in Multiple Criteria Decision Making*. Kluwer Academic Publishers, Dordrecht, pp. 1.1–1.35.
- Roy, B., 1993. Bouyssou, D.: Aide Multicritère à la Décision: Méthodes et Cas. Economica, Paris.
- Shafer, G., 1976. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton.
- Shapley, L.S., 1953. A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (Eds.), *Contributions to the Theory of Games II*. Princeton University Press, Princeton, pp. 307–317.
- Skowron, A., 1993. Boolean reasoning for decision rules generation. In: Komorowski, J., Ras, Z.W. (Eds.), *Methodologies for Intelligent Systems. Lecture Notes in Artificial Intelligence*, vol. 689. Springer, Berlin, pp. 295–305.
- Skowron, A., Grzymala-Busse, J.W., 1994. From the rough set theory to the evidence theory. In: Fedrizzi, M., Kacprzyk, J., Yager, R.R. (Eds.), *Advances in the Dempster-Shafer Theory of Evidence*. Wiley, New York, pp. 193–236.
- Skowron, A., Polkowski, L., 1997. Decision algorithms: a survey of rough set-theoretic methods. *Fundamenta Informaticae* 27 (3/4), 345–358.
- Skowron, A., Rauszer, C., 1992. The discernibility matrices and functions in information systems. In: Slowinski, R. (Ed.), *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Set Theory*. Kluwer Academic Publishers, Dordrecht, pp. 331–362.
- Skowron, A., Stepaniuk, J., 1995. Generalized approximation spaces. In: Lin, T.Y., Wildberger, A. (Eds.), *Soft Computing: Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery*. Simulation Councils, San Diego, CA, pp. 18–21.
- Slovic, P., 1975. Choice between equally-valued alternatives. *Journal of Experimental Psychology: Human Perception Performance* 1, 280–287.
- Slowinski, R. (Ed.), 1992. *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*. Kluwer Academic Publishers, Dordrecht.
- Slowinski, R., 1993a. A generalization of the indiscernibility relation for rough set analysis of quantitative information. *Rivista di matematica per le scienze economiche e sociali* 15, 65–78.
- Slowinski, R., 1993b. Rough set learning of preferential attitude in multi-criteria decision making. In: Komorowski, J., Ras, Z.W. (Eds.), *Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence*, vol. 689. Springer, Berlin, pp. 642–651.
- Slowinski, R., 1995. Rough set processing of fuzzy information. In: Lin, T.Y., Wildberger, A. (Eds.), *Soft Computing: Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery*. Simulation Councils, San Diego, CA, pp. 142–145.
- Slowinski, K., Slowinski, R., Stefanowski, J., 1988. Rough set approach to analysis of data from peritoneal lavage in acute pancreatitis. *Medical Informatics* 13, 143–159.
- Slowinski, R., Stefanowski, J., 1989. Rough classification in incomplete information systems. *Mathematical Computer Modelling* 12 (10/11), 1347–1357.
- Slowinski, R., Stefanowski, J., 1992. RoughDAS and RoughClass software implementations of the rough sets approach. In: Slowinski, R. (Ed.), *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory*. Kluwer Academic Publishers, Dordrecht, pp. 445–456.
- Slowinski, R., Stefanowski, J., 1994. Handling various types of uncertainty in the rough set approach. In: Ziarko, W.P. (Ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery*. Springer, London, pp. 366–376.
- Slowinski, R., Stefanowski, J., 1996. Rough set reasoning about uncertain data. *Fundamenta Informaticae* 27, 229–243.
- Slowinski, R., Stefanowski, J., Greco, S., Matarazzo, B., 2000. Rough sets processing of inconsistent information in decision analysis. *Control and Cybernetics* 29, 379–404.
- Slowinski, R., Vanderpooten, D., 1997. Similarity relation as a basis for rough approximations, ICS Research Report 53/95, Warsaw University of Technology, Warsaw, 1995. In: Wang, P.P. (Ed.), *Advances in Machine Intelligence & Soft-Computing*, vol. IV. Duke University Press, Durham, NC, pp. 17–33.
- Slowinski, R., Vanderpooten, D., 2000. A generalized definition of rough approximations based on similarity. *IEEE Transactions on Data and Knowledge Engineering* 12 (2), 331–336.
- Stefanowski, J., 1998. On rough set based approaches to induction of decision rules. In: Skowron, A., Polkowski, L. (Eds.), *Rough Sets in Data Mining and Knowledge Discovery*, vol. 1. Physica, Heidelberg, pp. 500–529.
- Susmaga, R., 1998. Experiments in incremental computation of reducts. In: Skowron, A., Polkowski, L. (Eds.), *Rough Sets in Data Mining and Knowledge Discovery*, vol. 1. Physica, Heidelberg, pp. 530–553.
- Sugeno, M., 1974. Theory of fuzzy integrals and its applications. Doctoral Thesis, Tokyo Institute of Technology.
- Tsoukias, A., Vincke, Ph., 1995. A new axiomatic foundation of the partial comparability theory. *Theory and Decision* 39, 79–114.
- Tsoukias, A., Vincke, Ph., 1997. Extended preference structures in MCDA. In: Climaco, J. (Ed.), *Multicriteria Analysis*. Springer, Berlin, pp. 37–50.
- Tversky, A., 1969. Intransitivity of preferences. *Psychological Review* 76, 31–48.

- Tversky, A., 1977. Features of similarity. *Psychological Review* 84 (4), 327–352.
- Vincke, Ph., 1992. *Multicriteria Decision-Aid*. Wiley, New York.
- Wakker, P.P., 1989. *Additive representations of preferences. A new foundation of decision analysis*. Kluwer Academic Publishers, Dordrecht.
- Yao, Y., 1996. Combination of rough sets and fuzzy sets based on α -level sets. In: Lin, T.Y., Cercone, N. (Eds.), *Rough Sets and Data Mining*. Kluwer Academic Publishers, Boston, pp. 301–321.
- Yao, Y., 1998. A comparative study of fuzzy sets and rough sets. *Information Sciences* 109, 227–242.
- Yao, Y., Wong, S., 1995. Generalization of rough sets using relationships between attribute values. In: *Proceedings of the Second Annual Joint Conference on Information Sciences*. Wrightsville Beach, NC, pp. 30–33.
- Ziarko, W., Shan, N., 1994. An incremental learning algorithm for constructing decision rules. In: Ziarko, W.P. (Ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery*. Springer, London, pp. 326–334.