Control and Cybernetics

vol. **31** (2002) No. 4

Methodology of rough-set-based classification and sorting with hierarchical structure of attributes and criteria

by

Krzysztof Dembczyński[1], **Salvatore Greco**[2] **and Roman Słowiński**[1]

[1] Institute of Computing Science, Poznań University of Technology, Piotrowo 3a, 60-965 Poznań, Poland

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

Abstract: We consider a hierarchical classification problem involving sets of attributes and criteria. The problem of classification concerns an assignment of a set of objects to pre-defined classes. The classification to preference-ordered classes is called sorting. The objects are described by two sorts of attributes: criteria and regular attributes, depending on whether the attribute domain is preference-ordered or not. The hierarchical classification and sorting is made in finite number of steps due to hierarchical structure of regular attributes and criteria in the form of a tree. We propose a methodology based on the decision rule preference model. The model is constructed by inductive learning from examples of hierarchical decisions made by the Decision Maker on a reference set of objects. To deal with inconsistencies appearing in decision examples we adapt the rough set approach to the hierarchical classification and sorting problems. Due to inconsistency and their propagation from the bottom to the top of the hierarchy, the description of an object on a particular attribute may be not a simple value but either a subset of a regular attribute domain or an interval on a criterion scale. An example illustrates the methodology presented.

Keywords: multicriteria decision problems, classification, sorting, hierarchical structure, rough sets, decision rule preference model.

1. Introduction

The aim of 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 as to how to make a good decision under specific circumstances. According to Roy (1985), it is possible to distinguish the following, most frequent decision problems: *choice, ranking* and *classification* or *sorting*. The choice problem consists in selection of the best object from a given set (often called actions in decision problems). The ranking problem consists in ordering the objects from the best to the worst, with respect to the Decision Maker's (DM's) preferences. *Classification* concerns an assignment of the objects to pre-defined classes. In the case where the classes are preference-ordered it is called the *sorting* problem.

In general, decisions are based on some characteristics of objects. For example, when buying a car, the decisions can be based on such characteristics as price, maximum speed, fuel consumption, colour, country of production, etc. These characteristics are called *attributes* or, more precisely, *condition attributes*. Let us observe that, depending on the 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 are called *regular attributes* or, briefly, *attributes*. In the above example, price, maximum speed and fuel consumption may be criteria because, for instance, low price is better than high price; most probably, colour and country of production are not criteria but regular attributes because, for instance, red is not better than green. However, one can imagine that also those attributes could become criteria.

Moreover, decisions may be also ordinal. For example, a standard 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 the interpretation assigned 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 into three categories: acceptable, hardly acceptable, non-acceptable. This type of classification is called sorting.

In many real life situations, the process of decision making is decomposable into subproblems; this decomposition may either follow from a, natural, hierarchical structure of the evaluation, or from a need of simplification of a complex decision problem. These situations are called *Hierarchical Decision Problems* (HDP), in particular, *Hierarchical Classification/Sorting Problems* (HCSP). They are considered in our paper. The hierarchical structure of HCSP has the form of a *tree* whose *nodes* are attributes and/or criteria describing objects. In the *root* of the tree there is an overall evaluation making assignment of objects to classes, then in the intermediate nodes are subattributes and/or subcriteria, called *hierarchical* and, finally, in the *leafs* there are attributes and/or criteria that do not branch further – they are called *flat* attributes or criteria. The hierarchy of attributes and criteria seems to be a natural and intuitive concept. For example, when considering car sorting, the criterion of fuel consumption may be considered as a hierarchical one consisting of following sub-criteria: fuel consumption in urban drive, in highway drive, at 60 km/h, and at 120 km/h.

Consideration of hierarchical decision problems is supported by psychological arguments, as pointed out by White, Wilson and Wilson (1969): "the use of hierarchical ordering must be as old as human thought, conscious and unconscious". Psychologists have proven that the human brain is limited to about seven (seven, plus or minus two) items in both short term memory capacity and its discrimination ability (Miller, 1956). Humans have learned how to deal with complexity by hierarchical decomposition. Surprisingly, the hierarchical decomposition of decision problems has gained little attention in scientific decision aiding.

The best-known exception is the controversial Analytic Hierarchy Process (Saaty, 1980, Belton and Gear, 1983, Barzilei, 1997). Two other methods for handling a hierarchical structure of attributes and criteria are V.I.S.A. (Belton, 1999) and MACBETH (Bana a Costa and Vansnick, 1999). It is well known that a multicriteria decision problem has no solution unless a preference model is defined. Traditionally, in multicriteria decision analysis (MCDA) two major preference models: functional and relational have been considered. For example, Analytic Hierarchy Process, V.I.S.A. and MACBETH use utility function as a preference model. The main difficulty with application of MCDA methods based on functional and relational preference models lies in acquisition of the DM's preferential information. Very often, this information has to be given in terms of pairwise comparisons over all objects or in terms of such parameters like importance weights, substitution rates and various thresholds. It is generally acknowledged, however, that people prefer to make exemplary decisions rather than to explain them in terms of the preference model adopted by an analyst. Our approach to HCSP is based on another type of preference model, which is a set of logical statements i.e. "*if...*, *then...*" *decision rules*, characterised by Greco, Matarazzo and Słowiński (1998a, 1998b, 2001a). The decision rules are induced from decision examples given by the DM. In the case of classification and sorting the decision examples concern assignment of some *reference objects* to decision classes. The reference objects are those objects in a set which are relatively well-know to the DM so that she/he is able to make decision with them. For the reasons mentioned above, the idea of inferring preference models from exemplary decisions provided by the DM is very attractive.

Very often, in a set of decision examples on reference objects, there may appear some inconsistency corresponding to the situation where two reference objects having the same description are assigned to different classes. To deal with such an inconsistency, the *rough set approach* has been proposed by Pawlak in the early 1980s (1982, 1991). The rough set theory is based on the assumption that objects analyzed may be considered only in the perspective of available information about them. This leads to the conclusion that knowledge has granular structure. Due to granularity of knowledge some objects of interest cannot be discerned or appear as identical or similar. The *indiscernibility relation* constitutes a mathematical basis of the rough set theory; it induces a partition of the universe into blocks of indiscernible objects, called *elementary sets* or *granules* that can be used to build knowledge about a real or abstract world. A set of objects (or class), which cannot be precisely described by elementary sets (is not a union of some elementary sets) is called *rough* (*approximate*) – otherwise it is referred to as a *crisp* (*exact*) set (class). A rough set is described by two ordinary sets called the *lower* and the *upper approximation*; the lower approximation consists of all elementary sets which surely and totally belong to the described set of object, while the upper approximation contains, additionally the elementary sets which partially belong to the described set of objects. Obviously, the difference between the upper and the lower approximation constitutes the *boundary region* of the set, whose elements cannot be characterized with certainty as belonging or not to the described set of objects, using the available information.

Rough set analysis is naturally adapted to problems of multiattribute classification because it is possible to extract all the essential knowledge contained in the set of examples using indiscernibility relation (Pawlak and Słowiński, 1994). However, as pointed out by Greco, Matarazzo and Słowiński (1996), the original rough set approach is insufficient for multicriteria sorting problems. Consider, for example two firms, A and B, where the firm A is characterized by better economical parameters, but it is assigned by the DM to a class of higher bankruptcy risk than the firm B. This is obviously inconsistent with the *dominance principle* that requires that an object having a better (in general, not worse) evaluation on considered criteria cannot be assigned to a worse class. Within the original rough set approach, called the *Classic Rough Set Approach* (CRSA), the two firms will be considered as just discernible and no inconsistency will be stated.

Greco, Matarazzo and Słowiński (1996, 1998b, 1999, 2002a, 2002c) have proposed an extension of the rough set theory called *Dominance-based Rough Set Approach* (DRSA) that is able to deal with inconsistencies typical to exemplary decisions in MCDA problems. In difference to CRSA, the indiscernibility relation is substituted by the *dominance relation* that identifies the *dominating* and the *dominated sets* as granules of knowledge. The dominating set of object x contains objects that are not worse then x on all considered criteria, while the dominated set of objects x contains objects that are not better than x on all considered criteria. DRSA prepares, moreover, a conceptual ground for inducing rules having syntax concordant with the dominance principle.

As Greco, Matarazzo and Słowiński (2001b) have proved, the decision rule preference model resulting from the rough set approach is more general than all the existing models of conjoint measurement due to its capacity of handling inconsistent preferences. Moreover, it is more understandable for the users because of its natural syntax. The decision rules explain a decision policy of the DM and may be used for classification or sorting of new objects.

In this paper, a methodology of hierarchical classification and sorting is presented. In this case, the decision examples specify a classification or sorting made by the DM at each node of the tree. To deal with inconsistencies in the set of decision examples we propose to adapt the rough set approaches. In HSCP, the main difficulty consists in *propagation* of inconsistencies along the tree, i.e. taking into account at each node of the tree the inconsistent information coming from lower level nodes. In the proposed methodology, the inconsistencies are propagated from the bottom to the top of the tree in the form of *subsets* of possible attribute values. In the case of hierarchical criteria, these subsets are *intervals* of possible criterion values. Subsets of possible values may also appear in leafs of the tree. To deal with multiple values of attributes for object description, we adequately adapt the rough set approaches: CRSA and DRSA. The rough set approaches based on generalized definitions of indiscernibility relation and dominance relation will be called *Multi-Valued* CRSA (MV-CRSA) and *Interval-Valued* DRSA (IV-DRSA), respectively. The sets of decision rules are induced from rough approximations at each node of the tree. The classification or sorting of new objects proceeds from the bottom to the top of the hierarchy, where the final decision is made.

The paper is organized as follows. In Section 2, a brief reminder of the decision rule methodology for non-hierarchical sorting and classification is given. In particular, data representation, decision rule preference model and formal description of CRSA and DRSA are presented. Section 3 contains description of the proposed methodology for classification and sorting with hierarchical structure of attributes and criteria. There we introduce the tree representation of the hierarchy, the extended rough set approaches: MV-CRSA and IV-DRSA, propagation of inconsistencies along the tree and application of decision rules to hierarchical classification or sorting of new objects. In Section 4 an illustrative example is presented. Section 5 contains conclusions.

2. Definitions and preliminaries

2.1. Data representation

Data are often presented as a table, where columns are labelled by *regular attributes* and/or *criteria*, rows by *reference objects*, and entries of the table are *attribute/criteria values.* Formally, a *decision table* is the 4-tuple $S = \langle U, A, V, f \rangle$, where U is a finite set of reference objects, A is a finite set of attributes and/or criteria, $V = \bigcup_{a \in A} V_a$, where V_a is the domain of the attribute/criterion a, and $f : U \times A \to V$ is an *information function* such that $f(x, a) \in V_a$ for every $(x, a) \in U \times A$. The set A is divided into *condition attributes* (set $C \neq \emptyset$) and *decision attributes* (set $D \neq \emptyset$), such that $C \cup D = A$ and $C \cap D = \emptyset$. Moreover, the set C is composed of subsets W and Q including regular attributes and criteria, respectively ($Q \cup W = C$; $Q \cap W = \emptyset$). The set D is a singleton ($D = \{d\}$), where d is a regular attribute or a criterion.

It is assumed that the domain of a criterion $q \in Q$ is completely pre-ordered by an *outranking relation* $\succeq_q$; $x \succeq_q y$ means that x is at least so good as y with respect to criterion q. The symmetric part of $\succeq_q$ is an *indifference relation* $\sim_q$, and the asymmetric part of $\succeq_q$ is a *strict preference relation* $\succ_q$ (Roy, 1985). In the following we are considering criteria having a numerical domain, i.e. if q is a criterion, then $V_q \subseteq \Re$ ($\Re$ denotes a set of real numbers), and belonging to one of the two following types: *gain criteria* and *cost criteria*. For the gain criterion: $x \succeq_q y \Rightarrow f(x,q) \geq f(y,q)$; for the cost criterion: $x \succeq_q y \Rightarrow f(x,q) \leq f(y,q)$, where $q \in Q$, $x, y \in U$.

Decision attribute d, whose domain $V_d = \{v_d^t, t \in T\}$, $T = \{1, \dots, n\}$ involves a partition $Cl(d) = \{Cl_t, t \in T\}$ of U into a finite number of classes $Cl_t = \{x \in U : f(x,d) = v_d^t\}$. Partition Cl is called *classification*. Each object $x \in U$ is assigned to one and only one class $Cl_t \in Cl(d)$. If decision attribute d is a criterion (it is assumed, without loss of generality, that a decision criterion is always of the gain type), then the classes from $Cl(d)$ are preference-ordered according to increasing order of class indices, i.e. for all $r, s \in T$, such that $r > s$, the objects from Cl_r are strictly preferred to the objects from Cl_s. In this case, classification $Cl(d)$ is called *sorting*. In the problem of sorting we will consider the *upward* and the *downward unions* of classes defined, respectively, as:

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

The statement $x \in Cl_t^{\geq}$ means "x belongs to *at least* class Cl_t", while $x \in Cl_t^{\leq}$ means "x belongs to *at most* class Cl_t". Observe that $Cl_1^{\geq} = Cl_n^{\leq} = U$, $Cl_n^{\geq} = Cl_n$ and $Cl_1^{\leq} = Cl_1$. Furthermore, for $t = 2, \dots, n$, it is: $Cl_{t-1}^{\leq} = U - Cl_t^{\geq}$ and $Cl_t^{\geq} = U - Cl_{t-1}^{\leq}$.

If the set C of condition attributes contains regular attributes only, then the decision problem represented by the table is a *multiattribute classification problem*, independently of whether d is criterion or attribute; if decision attribute d is a criterion and set C contains at least one criterion, then the corresponding decision problem is a *multicriteria sorting problem*.

2.2. Decision rules

Greco, Matarazzo and Słowiński (1998a, 1998b) have proposed and characterized the *decision rule preference model* for multicriteria sorting. According to Slovic (1975), people make decisions by searching for *rules* that provide good justification of their choices. Thus, it is natural to build the preference model in terms of "*if..., then...*" decision rules. The decision rules are induced from decision examples given by DM where the decision examples concern assignment of some reference objects to decision classes. Then, these rules can be applied to a set of new objects (potential actions) in order to obtain recommendations.

The induction of rules from examples is a typical approach of artificial intelligence. It is concordant with the principle of posterior rationality of March

(1988) and with aggregation-disaggregation logic of Jacquet-Lagrèze (1981). The rules explain the preferential attitude of the DM and enable her/his understanding of the reasons of her/his 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 that are the principal aims of decision analysis.

More formally, *decision rule* is a logical expression in the form:

if L, *then* K,

where L is an *antecedent* (*condition part*) and K is a *consequent* (*decision part*), meaning that an object satisfying L will be classified to class or classes described by K.

The condition part L is a conjunction of *elementary conditions* (*selectors*) w:

$$L = w_1 \wedge w_2 \wedge \ldots \wedge w_{l(r)},$$

where $l(r)$ is the number of elementary conditions, called *rule length*. The selector w_i is defined as $(f(x, a_i) \propto v_{a_i})$, $v_{a_i} \in V_{a_i}$, where $f(x, a_i)$ denotes a value of attribute a_i for object x; $\propto$ denotes relation operator $=, \geq, \leq$, etc. and term v_{a_i} denotes a value from V_{a_i} (Michalski, 1983, Stefanowski, 2001).

The decision part K is, in general, a disjunction of elementary decisions

$$x \in Cl_r \cup Cl_s \cup \ldots \cup Cl_t, \qquad r, s, t \in T.$$

An object $x \in U$ *supports* a decision rule r if it satisfies 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.

Procedures for generation of decision rules from a decision table use an inductive learning principle. When inducing the decision rules with consequent K, examples concordant with K are called *positive* and all the others *negative*. A decision rule is *discriminant* if it is consistent, i.e. distinguishes positive examples from the negative ones. By *minimal* decision rule we understand such implication that there is no other implication with an antecedent of at least the same generality, and a consequent of at least the same particularity. It is also interesting to look for *partly discriminant* rules. These are rules that, besides positive examples, cover also a limited number of negative ones.

It is assumed that a set of rules has to *reclassify* all examples from decision table. Such set of rules is called complete. If a decision table contains some inconsistent examples, then the induced set of rules should reflect such situation. This is possible while using the rough set approach.

2.3. Rough sets

2.3.1. Indiscernibility relation and dominance relation

In the case of multiattribute classification the CRSA is used. The relation that identifies the granules of knowledge is defined as follows.

For a given decision table S, where $C = W$, the *indiscernibility relation* I_a is a binary relation defined on U with respect to attribute $a \in C$, such that: $xI_ay \Leftrightarrow f(x,a) = f(y,a)$, $x, y \in U$, what means that objects x and y are indiscernible on attribute a. Objects x and y are indiscernible on subset $B \subseteq C$, when xI_ay, $\forall a \in B$, what is denoted by xI_By.

The indiscernibility relation thus defined is an equivalence relation (reflexive, symmetric and transitive). The family of all the equivalence classes of the relation I_B is denoted by $U|I_B$. The equivalence class identified by the relation I_B is called *B-elementary set*. The B-elementary set containing $x \in U$ is defined as: $I_B(x) = \{y \in U : xI_By\}$.

The DRSA extends the CRSA by substituting the indiscernibility relation by a *dominance relation*.

Given decision table S, where decision d is a criterion and $Q \neq \emptyset$, it is said that x *dominates* y with respect to $B \subseteq C$, denoted by xD_By, if $x \succeq_q y$, $\forall q \in Q \cap B$ and xI_ay, $\forall a \in W \cap B$. For each $B \subseteq C$, the dominance relation D_B is reflexive and transitive, i.e. it is a partial preorder.

Given $B \subseteq C$ and U, the granules of knowledge induced by dominance relation D_B are:

- a set of objects dominating x, $D_B^+(x) = \{y \in U : yD_Bx\}$,
- a set of objects dominated by x, $D_B^-(x) = \{y \in U : xD_By\}$,

called *B-dominating* set and *B-dominated* set with respect to $x \in U$, respectively. The granules will be used for rough approximation.

2.3.2. Lower and upper (rough) approximation

The *lower* and the *upper approximation* of a set are two principal concepts of the rough set theory used to description of sets.

In CRSA we are interested in approximation of classes. For given decision table S, where $C = W$, the B-lower and the B-upper approximation of class Cl_t, $t \in T$, with respect to $B \subseteq C$, are defined, respectively, as:

$$\underline{B}(Cl_t) = \{x \in U : I_B(x) \subseteq Cl_t\}, \quad \overline{B}(Cl_t) = \bigcup_{x \in Cl_t} I_B(x).$$

The *B-boundary* of Cl_t in S, denoted by $BN_B(Cl_t)$, is:

$$BN_B(Cl_t) = \overline{B}(Cl_t) - \underline{B}(Cl_t).$$

The lower approximation of Cl_t is composed of all B-elementary sets (granules of knowledge) included in Cl_t (whose elements, therefore, certainly belong to Cl_t), while the upper approximation of Cl_t consists of all the B-elementary sets that have a non-empty intersection with Cl_t (whose elements, therefore, possibly belong to Cl_t). The B-boundary of Cl_t constitutes the "doubtful region" of Cl_t: nothing can be said with certainty about the membership of its elements with respect to the class Cl_t. The objects from the boundary are inconsistent in the sense that they have been assigned to different classes while having the same description on considered condition attributes; such objects are also called *inconsistent examples.*

If the B-boundary of Cl_t is empty ($BN_B(Cl_t) = \emptyset$), then the class Cl_t is an crisp (exact) set with respect to B; otherwise, the class Cl_t is a rough (approximate) set with respect to B.

In the DRSA, the sets to be approximated are upward and downward unions of classes and the items (granules of knowledge) used for this approximation are B-dominating and B-dominated sets.

For given decision table S, where decision d is a criterion and $Q \neq \emptyset$, the B-lower and the B-upper approximation of $Cl_t^{\geq}$, $t \in T$, with respect to $B \subseteq C$, are defined, respectively, as:

$$\underline{B}(Cl_t^{\geq}) = \{x \in U : D_B^+(x) \subseteq Cl_t^{\geq}\}, \quad \overline{B}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_B^+(x).$$

Analogously, the B-lower and the B-upper approximation of $Cl_t^{\leq}$, $t \in T$, with respect to $B \subseteq C$, are defined, respectively, as:

$$\underline{B}(Cl_t^{\leq}) = \{x \in U : D_B^-(x) \subseteq Cl_t^{\leq}\}, \quad \overline{B}(Cl_t^{\leq}) = \bigcup_{x \in Cl_t^{\leq}} D_B^-(x).$$

The B-boundaries of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined as:

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

The interpretation of the above defined approximations is similar as in the case of CRSA. The objects from the boundary are inconsistent with the dominance principle in the sense that an object from a worse class dominates another object from a better class on the considered criteria; such objects are also called *inconsistent examples.* Formal properties of rough approximations for CRSA and DRSA may be found in (Pawlak, 1991, Greco, Matarazzo and Słowiński, 2002a).

For CRSA and DRSA we can consider the *quality of approximation of the partition* $Cl(d)$ by the set of attributes and/or criteria $B \subseteq C$. The definition is as follows:

$$\gamma(B,d) = \frac{|\bigcup_{Cl_t \in Cl(d)} \underline{B}(Cl_t)|}{|U|},$$

where $\underline{B}(Cl_t)$ is the lower approximation of class Cl_t; in CRSA, $\underline{B}(Cl_t)$ is defined above, and in DRSA, $\underline{B}(Cl_t) = \{x \in Cl_t : D_B^+(x) \subseteq Cl_t^{\geq} \wedge D_B^-(x) \subseteq Cl_t^{\leq}\}$. The quality expresses the ratio of all the B-correctly classified or sorted objects to all objects in the table.

Each minimal subset $B \subseteq C$ such that $\gamma(B, d) = \gamma(C, d)$ is called a *reduct* of $Cl(d)$ and denoted by $RED_{Cl(d)}$. Let us remark that a decision table can have more than one reduct. The intersection of all reducts is called the *core* and denoted by $CORE_{Cl(d)}$.

In CRSA, the notion of a generalized decision is used. The *B-generalized decision of* $x \in U$ is a function $\delta_B(x) : U \to 2^{V_d}$, defined as $\delta_B(x) = \{v \in V_d : \exists y \in U, yI_Bx \wedge f(y, d) = v\}$ where $x \in U$ and $B \subseteq C$. If an object is not inconsistent with any other object, then $|\delta_C(x)| = 1$.

A similar concept may be defined for DRSA. The *B-generalized decision* of object $x \in U$ with respect to the dominance relation is an interval $\delta_B^{\succ}(x) = [l(x, d), u(x, d)]$, where:

$$l(x, d) = \min\{v \in V_d : \exists y \in U, yD_Bx \wedge f(y, d) = v\},$$
$$u(x, d) = \max\{v \in V_d : \exists y \in U, xD_By \wedge f(y, d) = v\},$$

and $x, y \in U$, $B \subseteq C$.

In other words, $l(x, d)$ is the lowest decision class of objects dominating x, and $u(x, d)$ is the highest decision class of objects dominated by x. Remark that if an object does not cause inconsistency in decision table, then $l(x, d) = u(x, d)$.

2.3.3. Decision rules induced from rough approximations of decision classes

If the input decision table contains inconsistent examples, the lower and upper approximations of particular decision classes or unions of decision classes are computed. The decision rules are generated from these approximations and boundary regions. As a consequence, three basic kinds of rules are distinguished:

- *certain* rules induced from lower approximations of classes or unions of classes,
- *possible* rules induced from upper approximations of classes or unions of classes,
- *approximate* rules induced from boundary regions of classes or unions of classes.

Let us notice that certain rules are discriminant, while possible rules are partly discriminant. Moreover, the certain and possible rules indicate a unique decision to be made while approximate rules lead to a few possible decisions.

There are distinguished three different types of decision rule selectors:

1) $f(x, a) = r$, where $a \in W$, $r \in V_a$,
2) $f(x, q) \geq r$, where $q \in Q$, $r \in V_q$,
3) $f(x, q) \leq r$, where $q \in Q$, $r \in V_q$

In the case of CRSA, when inducing the certain and possible rules describing a class Cl_t, the positive examples are all examples belonging, respectively, to $\underline{C}(Cl_t)$ and $\overline{C}(Cl_t)$; the negative examples are all examples not belonging to corresponding approximations of Cl_t.

The approximate rules are induced from boundary regions composed of inconsistent examples. Let Y denote the set of examples having the same value of generalized decision function $\delta_C(x) = \Theta$, where $|\Theta| > 1$. For each Y identified with a set of positive examples, approximate decision rules are generated. Such decision rules have the consequent concordant with generalized decision function of objects belonging to Y. The negative examples are all examples not belonging to Y (Stefanowski, 2001).

In CRSA, the selectors of decision rules are the form 1). The decision part for these rules may take three possible forms:

- $x \in Cl_t$, for certain rules,
- x could belong to Cl_t, for possible rules,
- $x \in Cl_r \cup Cl_s \cup \ldots \cup Cl_t$, $r, s, t \in T$, for approximate rules.

For example, in classification of cars for a catalogue, the following decision rule could be induced: *if* $(f(x, speed) = high) \wedge (f(x, colour) = black)$, *then* $x \in sports\,car$.

In the case of DRSA, the decision rules are generated from approximations of upward and downward unions of classes. For a given upward or downward union of classes, $Cl_t^{\geq}$ or $Cl_t^{\leq}$, the decision rules induced from examples belonging to $\overline{C}(Cl_t^{\geq})$ or $\overline{C}(Cl_s^{\leq})$ suggest a possible assignment to "class Cl_t or better" or to "class Cl_s or worse", respectively. On the other hand, the decision rules induced from objects belonging to the intersection $\overline{C}(Cl_s^{\leq}) \cap \overline{C}(Cl_t^{\geq})$ are suggesting an assignment to some classes between Cl_s and Cl_t $(s < t)$.

The form of decision rules in DRSA is more general. We are distinguishing the following types of decision rules:

- *certain 'at least'* $D_{\geq}$-decision rules, where selectors are in the form 1), and 2), when q is a gain criterion, or 3), when q is a cost criterion, and the consequent is in the form: $x \in Cl_t^{\geq}$, where $t \in T$.
- *possible 'at least'* $D_{\geq}$-decision rules, where selectors are in the form 1), and 2), when q is a gain criterion, or 3), when q is a cost criterion, and the consequent is in the form: x could belong to $x \in Cl_t^{\geq}$, where $t \in T$.
- *certain 'at most'* $D_{\leq}$-decision rules, where selectors are in the form 1), and 3), when q is a gain criterion, or 2), when q is a cost criterion, and the consequent is in the form: $x \in Cl_t^{\leq}$, where $t \in T$.
- *possible 'at most'* $D_{\leq}$-decision rules, where selectors are in the form 1), and 3), when q is a gain criterion, or 2), when q is a cost criterion, and the consequent is in the form: x could belong to $Cl_t^{\leq}$, where $t \in T$.
- *approximate* $D_{\geq\leq}$-decision rules, where selectors are in all the forms given above, and the decision part in the form: $x \in Cl_s \cup Cl_{s+1} \cup \ldots \cup Cl_t$, where $s, t \in T$, and $s < t$.

For example, in sorting of cars to preference-ordered classes: acceptable, hardly acceptable, non-acceptable, the following decision rule could be induced (price and fuel consumption are cost criteria):

if $(f(x, price) \leq 9000) \wedge (f(x, fuel\,consumption) \leq 10\ lit./100\ km)$, *then* $x \in hardly\ acceptable^{\leq}$.

Let us observe that decision rules may be generated not only on the basis of all condition attributes/criteria but also with respect to a reduct or to any other subset of attributes/criteria.

Generation of decision rules from decision tables is a complex task. A number of procedures based on the rough set theory have been proposed to solve it (for example see: Grzymala-Busse, 1992, 1997, Skowron, 1993, Skowron and Polkowski, 1997, Słowiński and Stefanowski, 1992, Stefanowski, 1998; with respect to decision rule within DRSA see Greco et al., 2002b). The existing induction algorithms use one of the following strategies:

- generation of a *minimal* set of rules,
- generation of an *exhaustive* set of rules,
- generation of a set of *satisfactory* set of rules.

The first category of algorithms is focused on describing input objects using the minimum number of necessary rules covering all objects from a decision table, while the second group try to generate all decision rules in the simplest form. The third category of algorithms gives as a result the set of decision rules, which satisfy the a priori user's requirements, for example, the user can prefer to get decision rule characterized by a specified length of a rule (Stefanowski, 1998).

3. Methodology of hierarchical classification and sorting problem (HSCP)

As mentioned in the introduction, decision problems, in particular, classification and sorting problems, may have a hierarchical structure. In this case objects are described by regular attributes and/or criteria, which are organized in a hierarchical structure. In consequence, the problem analysis proceeds with respect to the hierarchy, such that decisions performed on subproblems (subattributes/subcriteria) influence the final decision.

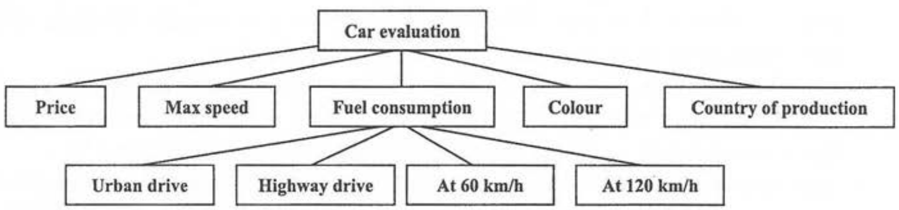

Figure 1. The hierarchy of attributes and criteria for a car sorting problem

An illustrative structure of a hierarchical sorting problem is shown in Fig. 1. The cars are sorted to three classes: acceptable, hardly acceptable, non-acceptable, on the basis of five characteristics, where one of them – fuel consumption – is further made more detailed.

Let us remind that the hierarchy can be represented by a *tree*, where *subtrees* represent subproblems and each *node* of the tree, different from the *leaf*, branches into succesive nodes called *direct descendents*. The *root* of the tree refers to an overall decision, each subnode refers to *hierarchical* attribute or criterion (subdecision or intermediate decision) and leafs are *flat* regular attributes and criteria. Notice that if a root of subtree is a hierarchical attribute, then the analysis of the relevant subproblem proceeds according to multiattribute classification; otherwise if a root is a hierarchical criterion, then the analysis proceeds according to multicriteria sorting.

In HSCP, the main difficulty consists in *propagation* of inconsistencies along the tree, i.e. taking into account at each node of the tree the inconsistent information coming from lower level nodes. In the proposed methodology, the inconsistencies are propagated from the bottom to the top of the tree in the form of *subsets* of possible attribute values. To deal with multiple values of attributes for object description, we propose the extended rough set approaches: MV-CRSA and IV-DRSA. Moreover, thanks to above adaptation we can deal with imprecise evaluations of objects on flat attributes/criteria.

The preference model corresponding to HSCP has the form of sets of decision rules for each subproblem. The particular sets of decision rules are generated from rough approximations at each node of the hierarchy. The application of the model consists in progressive classification or sorting of new objects at each node of the attribute and/or criteria hierarchy.

3.1. Data representation

The *hierarchical decision table* is presented as a tree T composed of *subtables*. At the node N_k of T, there is a subtable S_k containing decision attribute or criterion corresponding to the node N_k and condition attributes and/or criteria being direct descendants of N_k.

More formally, the subtable is the 4-tuple $S_k = \langle U, A_k, V_k, f\rangle$, where U is a finite set of objects; $A_k = C_k \cup \{d_k\}$ is a finite set of attributes and/or criteria, such that $C_k = W_k \cup Q_k$ denotes a set of condition attributes and criteria, respectively, corresponding to subnodes being direct descendents of N_k, and d_k is a decision attribute or criterion corresponding to N_k; $V_k = \bigcup_{a\in A_k} V_a$, V_a is a domain of attribute a; and $f : U \times C \to 2^{V_a}$ is an *information function* such that $f(x, w) \subseteq V_w$, for each regular attribute $w \in W_k$, and $f(x, q) \in [l(x, q), u(x, q)] \subseteq V_q$, for each criterion $q \in Q_k$, where $x \in U$. For each S_k a partition of U is considered with respect to decision attribute or criterion d_k, i.e. classification or sorting $Cl(d_k)$. We assume that in S_k each object is assigned to only one class, therefore, the information function for d_k is defined as $f : U \times \{d_k\} \to V_{d_k}$ and $f(x, d_k) \in V_{d_k}$ for each $x, y \in U$.

3.2. Rough set approach for attribute subset values and interval order

3.2.1. Indiscernibility relation for subsets of attribute values and MV-CRSA

Two objects $x, y \in U$, described by the subsets of attribute values will be considered as indiscernible on attribute $a \in W_k$ if and only if $f(x,a) \cap f(y,a) \neq \emptyset$. The corresponding *indiscernibility relation* is denoted by $I_a^{\cap}$. Two objects $x, y \in U$ are indiscernible with respect to the subset of attributes $B \subseteq W_k$, if and only if $xI_a^{\cap}y$ is true for each $a \in B$. The *B-indiscernibility relation* is denoted by $I_b^{\cap}$.

The above defined indiscernibility relation is reflexive, symmetric, but not transitive. This relation satisfies requirements of a tolerance relation. Comparable definitions were given, for example, by Słowiński (1992), Orłowska (1998), Stepaniuk (2000).

All definitions concerning CRSA, given in point 2.3, may be easily generalized for the above relation. For example, the lower and the upper approximations of a class Cl_t, $t \in T$, are defined, respectively, by:

$$\underline{B}(Cl_t) = \{x \in U : I_B^{\cap}(x) \subseteq Cl_t\}, \quad \overline{B}(Cl_t) = \bigcup_{x \in Cl_t} I_B^{\cap}(x)$$

where $I_B^{\cap}(x) = \{y \in U : xI_B^{\cap}y\}$ is a B-elementary set and $B \subseteq C_k$.

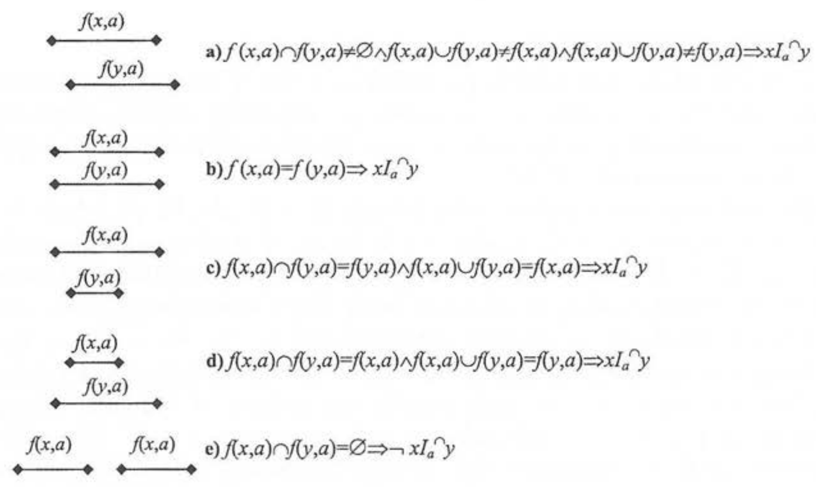

Figure 2. All possibilities of intersection of description of two objects on attribute a. Two objects x and y are indiscernible when their description has a common part.

3.2.2. Interval order and IV-DRSA

In order to enable imprecise description of objects on criteria, it is required to introduce the definition of interval order and dominance relation based on it.

The *interval order* is a binary relation R on set U if and only if there exist two functions $g : U \rightarrow \Re$ (where $\Re$ is a set of real numbers) and $h : U \rightarrow \Re^+$ such that:

$$xRy \Leftrightarrow g(x) + h(x) \geq g(y), \text{ and } x, y \in U.$$

Let us observe that:

$$xRy \wedge yRx \Leftrightarrow -h(x) \leq g(x) - g(y) \leq h(y)$$

and

$$xRy \wedge \neg yRx \Leftrightarrow g(x) - g(y) \geq h(y)$$

and if R is an outranking relation $\succeq$, then the first above situation corresponds to indifference $\sim$, and the second to preference $\succ$.

Let us assume, that there exist two functions $l : U \rightarrow \Re$ and $u : U \rightarrow \Re$ defined as follows: $l(x) = g(x)$ and $u(x) = g(x)+h(x)$, such that $xRy \Leftrightarrow u(x) \geq l(y)$. If R is an outranking relation $\succ$, then $x \sim y \Leftrightarrow u(x) \geq l(y) \wedge u(y) \geq l(x)$ and $x \succ y \Leftrightarrow l(x) > u(y)$.

Let us observe that an interval order is a strongly complete and *Ferrers transitive* binary relation. Ferrers transitivity could be illustrated as follows. Let us consider four objects, $x, y, w, z \in U$, such that $x \succeq y$ and $w \succeq z$; $x \succeq y$ means (1) $u(x) \geq l(y)$ and $w \succeq z$ means (2) $u(w) \geq l(z)$. Ferrers transitivity says that if $x \succeq y$ and $w \succeq z$, then at least one between $x \succeq z$ and $w \succeq y$ is verified. In fact if $x \succeq z$ is not verified, we have (3) $l(z) > u(x)$. From (1), (2) and (3) we obtain: (4) $u(w) \geq l(z) > u(x) \geq l(y)$. From (4) we obtain: $u(w) \geq l(y)$, i.e. $w \succeq y$.

The above definitions lead us to the following dominance relation defined on the basis of indiscernibility relation for subsets of attribute values and outranking relation $\succeq_q$ being interval order.

Let us assume, without loss of generality, that each criterion $q \in Q_k$ is a gain type criterion. For each we consider an outranking relation $\succeq_q$ on a set U on the basis of values $l(x,q)$ and $u(x,q)$ such that $\succeq_q = \succ_q \cup \sim_q$, thus: $x \succeq_q y \Leftrightarrow u(x,q) \geq l(y,q)$. The *dominance relation* with respect to $B \subseteq C_k$, considering objects description by both attributes and criteria, is defined as: $xD_B^{\langle\rangle}y \Leftrightarrow x \succeq_q y, \forall q \in Q_k \cap B \;\wedge\; xI_a^{\cap}y, \forall a \in W_k \cap B$.

The dominance relation $D_B^{\langle\rangle}$ is reflexive but it is not Ferrers transitive even if the considered attributes are all criteria. The following is a counterexample proving this point.

Let us consider four objects, $x, y, w, z \in U$, and two criteria $q_1, q_2 \in Q_k$ such that $x \succeq_{q_1} y$, $x \succeq_{q_2} y$, $w \succeq_{q_1} z$ and $w \succeq_{q_2} z$ and therefore $xD_B^{\langle\rangle}y$ and $wD_B^{\langle\rangle}z$, where $B = \{q_1, q_2\}$. Let us suppose that (1) $u(w,q_1) \geq l(z,q_1) > u(x,q_1) \geq l(y,q_1)$ and (2) $u(x,q_2) \geq l(y,q_2) > u(w,q_2) \geq l(z,q_2)$. Let us remark that (1) and (2) are concordant with: (a) $x \succeq_{q_2} y$ (in fact $u(x,q_1) \geq l(y,q_1)$), (b) $x \succeq_{q_2} y$ (in fact $u(x,q_2) \geq l(y,q_2)$), (c) $w \succeq_{q_1} z$ (in fact $u(w,q_1) \geq l(z,q_1)$), (d) $w \succeq_{q_2} z$ (in fact $u(w,q_2) \geq l(z,q_2)$). Remark that (a) and (b) gives $xD_B^{()}y$ while (c) and (d) $wD_B^{()}z$. However, from $xD_B^{()}y$ and $wD_B^{()}z$ neither $xD_B^{()}z$ nor $wD_B^{()}y$ is derived. In fact, $xD_B^{()}z$ cannot hold because $l(z,q_1) > u(x,q_1)$ while $wD_B^{()}y$ cannot hold because $l(y,q_2) > u(w,q_2)$.

Further definitions given in point 2.3 for DRSA may be easily generalized for the above dominance relation. Below we present as examples the definitions of the granules of knowledge and the approximations of class unions.

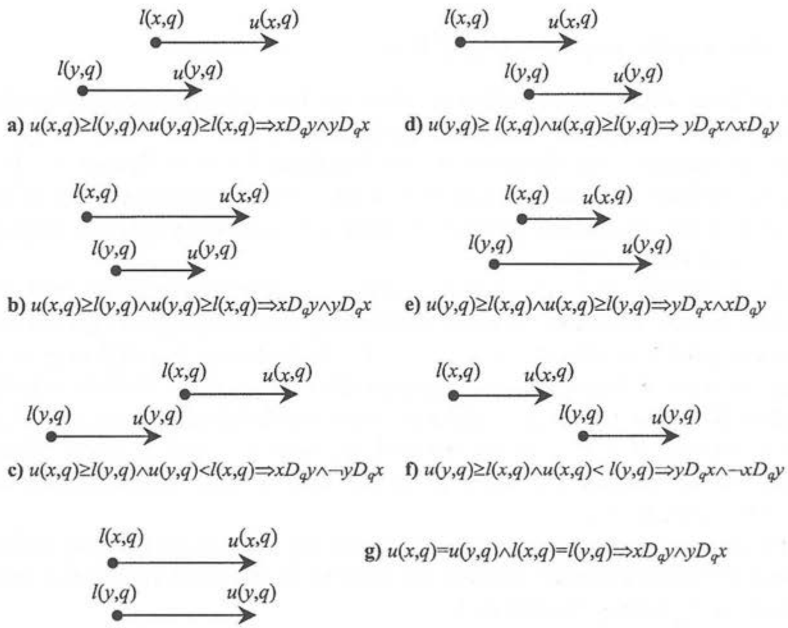

Figure 3. All possible relations between two object described using intervals on criterion q. Object x dominates object y when the upper bound of x is greater than the lower bound of y.

The B-dominated and the B-dominating sets are defined, respectively, as follows:

$$D_B^{()^+}(x) = \{y \in U : yD_B^{()}x\}, \;\; D_B^{()^-}(x) = \{y \in U : xD_B^{()}y\}, \text{ where } B \subseteq C_k.$$

The lower and the upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, for $t \in T$, are defined as:

$$\underline{B}(Cl_t^{\geq}) = \{x \in U : D_B^{\langle\rangle^+}(x) \subseteq Cl_t^{\geq}\}, \quad \overline{B}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_B^{\langle\rangle^+}(x),$$

$$\underline{B}(Cl_t^{\leq}) = \{x \in U : D_B^{\langle\rangle^-}(x) \subseteq Cl_t^{\leq}\}, \quad \overline{B}(Cl_t^{\leq}) = \bigcup_{x \in Cl_t^{\leq}} D_B^{\langle\rangle^-}(x).$$

3.2.3. Decision rules

The form of decision rules taking into account the imprecise description of objects, i.e. MV-CRSA and IV-DRSA rules, is changed in comparison with the form known from CRSA and DRSA. Moreover, the negative examples are defined differently.

The selectors for attributes are in the form:

1) $f(x,a) \cap v \neq \emptyset$, where $a \in W_k$ and $v \subseteq V_a$ is a value subset of attribute a.

The selectors for criteria are in the following forms:

2) $u(x,q) \geq r$, where $q \in Q_k$ is of gain type, $r \in V_q$, and $u(x,q)$ is a upper boundary of interval value of object x on q,

3) $l(x,q) \leq r$, where $q \in Q_k$ is of gain type, $r \in V_q$, and $l(x,q)$ is a lower boundary of interval value of object x on q,

4) $l(x,q) \leq r$, where $q \in Q_k$ is of cost type, $r \in V_q$, and $l(x,q)$ is a lower boundary of interval value of object x on q,

5) $u(x,q) \geq r$, where $q \in Q_k$ is of cost type, $r \in V_q$, and $u(x,q)$ is an upper boundary of interval value of object x on q.

The decision parts of rules keep the same form as in the case of traditional classification and sorting problems. Below, we list the types of IV-DRSA rules only, because they are the only ones that change:

- *certain 'at least'* $D_{\geq}$-decision rules, where the selectors are in the form: 1), 2) and 4), and the decision is in the form: $x \in Cl_t^{\geq}$,
- *possible 'at least'* $D_{\geq}$-decision rules, where the selectors are in the form: 1), 2) and 4), and the decision is in the form: x could belong to $x \in Cl_t^{\geq}$,
- *certain 'at most'* $D_{\leq}$-decision rules, where the selectors are in the form: 1), 3) and 5), and the decision in the form: $x \in Cl_t^{\leq}$,
- *possible 'at most'* $D_{\leq}$-decision rules, where the selectors are in the form: 1), 3) and 5), and the decision is in the form: x could belong to $x \in Cl_t^{\leq}$,
- *approximate* $D_{\geq\leq}$-decision rules, where selectors are in all the forms given above, and the decision is in the form: $x \in Cl_s \cup Cl_{s+1} \cup \ldots \cup Cl_t$, where $s, t \in T$, and $s < t$.

Because the generalized indiscernibility and dominance relations are not transitive, while generating any decision rule with a consequent K, the negative examples with respect to K are those examples that are not concordant with K and either are not indiscernible with the positive examples by an indiscernibility relation or are not indifferent with the positive examples in the sense of mutual dominance. This is explained below. Assume that there are three objects

$w, y, z \in U$ described on gain criterion q only, such that $u(w,q) = 7$, $l(w,q) = 6$, $u(y,q) = 8$, $l(y,q) = 4$, $u(z,q) = 5$ and $l(z,q) = 2$, and w, y belong to class Cl_2, while z belongs to class Cl_1. Lower approximation of $Cl_2^{\geq}$ with respect to $B = \{q\}$ contains only w, because $D_B^{\langle\rangle^+}(w) = \{w, y\}$ and $D_B^{\langle\rangle^+}(y) = \{w, y, z\}$. If y were considered as negative example, no certain rule for $Cl_2^{\geq}$ (supported by positive examples from lower approximation of $Cl_2^{\geq}$) would be induced. The following rule:

$$\textit{if } (u(x,q) \geq 6), \textit{ then } x \in Cl_2^{\geq}$$

covers $w \in Cl_2^{\geq}$, because $u(w,q) = 7 \geq 6$, but it covers also $y \in Cl_2^{\geq}$ because $u(y,q) = 8 \geq 6$. We accept, however, the above rule, even if it covers object y which is not included in the lower approximation of $Cl_2^{\geq}$, because it is concordant with the consequent of the rule.

3.3. Propagation of inconsistencies and application of decision rules

Let us comment on the propagation of inconsistencies along the tree. Inconsistencies are propagated from the bottom to the top of the hierarchy. Let S_k denote a decision subtable corresponding to node N_k of T and S_l a subtable of N_l being a direct descendent of N_k. If object $x \in U$ is inconsistent with other objects in S_l, then in the decision subtable S_k, the value of x on $d_l' \in C_k$ is the following:

- $\delta_{C_l}(x)$, if d_l is an attribute,
- $\delta_{C_l}^{\succeq}(x)$, if d_l is a criterion

and d_l corresponds to the decision attribute or criterion in S_l.

When new objects are submitted to hierarchical classification or sorting, the sets of decision rules are to be used progressively, starting from the lowest level of hierarchy. The decision from each node is propagated upward the hierarchy.

Let us comment on the application of decision rules to an object in a particular node using a car classification problem as an example. There are two possible cases. First, the node may correspond to an attribute and then the rules are in the form specified in MV-CRSA. In this case, if an object is matched by rules suggesting the same class Cl_t, then the object is assigned to Cl_t without any doubt. Nevertheless, the following doubtful situations can occur:

- object x matches one or more approximate decision rules, e.g.:

$$\textit{if } [\text{conditions}], \textit{ then } x \in \textit{sport cars} \cup \textit{family cars},$$

$$\textit{if } [\text{conditions}], \textit{ then } x \in \textit{utility cars} \cup \textit{family cars},$$

- object x matches certain or possible decision rules suggesting different classes, e.g.:

$$\textit{if } [\text{conditions}], \textit{ then } x \in \textit{utility cars},$$

$$\textit{if } [\text{conditions}], \textit{ then } x \in \textit{family cars}.$$

In all the above situations, the object is assigned imprecisely to all classes pointed by the matching decision rules. When an object does not match the condition part of any decision rule, then it is reasonable to conclude that it belongs to all classes.

Secondly, if a node corresponds to a criterion, then the decision rules are in the form specified in IV-DRSA. When applying $D_{\geq}$-decision rules to object x, it is possible that x either matches condition part of at least one decision rule or does not match condition part of any decision rule. In the case of at least one matching, it is reasonable to conclude that x belongs to class Cl_t, being the lowest class of the upward union $Cl_t^{\geq}$ resulting from intersection of all consequents of rules matching x. In the case of no matching, it is concluded that x belongs to Cl_1, i.e. to the worst class.

Analogously, when applying $D_{\leq}$-decision rules to object x, it is concluded that x belongs either to class Cl_t, being the highest class of the downward union $Cl_t^{\leq}$ resulting from intersection of all consequents of rules matching x, or to class Cl_n, i.e. to the best class, when x does not match any rule.

However, four other situations may occur:

(i) object x matches one or more approximate rules, e.g.:

if [conditions], *then* $x \in$ *non-acceptable* $\cup$ *hardly acceptable*

(ii) object x matches rules with intersecting downward and upward unions of classes, e.g.:

if [conditions], *then* $x \in$ *hardly acceptable*$^{\geq}$

if [conditions], *then* $x \in$ *acceptable*$^{\leq}$

(iii) object x matches rules with disjoint downward and upward unions of , e.g.:

if [conditions], *then* $x \in$ *acceptable*$^{\geq}$

if [conditions], *then* $x \in$ *hardly acceptable*$^{\leq}$

(iv) object x matches one or more approximate rules and certain rules with intersecting downward and upward unions of classes, e.g.:

if [conditions], *then* $x \in$ *non-acceptable* $\cup$ *hardly acceptable*

if [conditions], *then* $x \in$ *acceptable*$^{\geq}$.

The above situations correspond to: ambiguous [(i) and (iv)], incomplete [(ii)] and controversial [(iii) and (iv)] knowledge. In all the above situations object x is sorted imprecisely to an interval of classes. In the first situation object x is assigned to the interval of classes between the worst and the best class suggested by the rules matching x (from non-acceptable to hardly acceptable). In the second and in the third situation, object x is assigned to the interval of classes between the worst class of the matching at least rules and by the best class of the matching at most rules (i.e. in both situations to classes from hardly acceptable to acceptable). In the fourth situation, object x is assigned to the interval of classes which are extreme in the classes suggested by all approximate rules and the worst class from intersection of at least rules and/or the best class from the intersection of at most rules matching x.

4. Illustrative example

The illustrative example presented in this section will serve to explain the concepts introduced in previous point. Let us consider a problem of student qualification to the upper level of study. We are considering students of the Computer Science faculty choosing their specialization after third year of master course. In the considered problem the DM is looking for good students interested in studying Intelligent Decision Support Systems. The students are sorted into three classes: desirable, acceptable, and non-acceptable. Since during the first stage of study (bachelor studies) there are not many lectures linked directly with decision support systems, the decision is taken on the basis of all credits received by students. In the following example we consider six students described by means of the following two regular attributes and five criteria of gain type (see Table 1):

- Additional project
- Training
- examination in Statistics
- examination in Computer Networks
- project in Computer Networks
- examination in Databases
- project in Databases.

Assume that attributes "Additional project" and "Training" have the same domain composed of the following values: Artificial Intelligence (AI), Statistics (Stat), Databases (DB), and Programming (Progr). The domain of all the the above criteria is composed of the following evaluations: bad, sufficient and good. Of course, "good" is better than "sufficient" and "sufficient" is better than "bad".

The first step of the proposed methodology is to develop a hierarchy of the above evaluation. It seems natural that evaluations from examination and project from "Computer Networks" and "Databases", may be grouped into two independently subproblems. Similarly, it may be done with two other attributes: "Additional project" (AddProj) and "Training". On the basis of these two attributes we can judge the area of interest of a student. The structure of decomposed problem is presented in Fig. 4. The hierarchical criteria "Computer Network skills" (CN skills) and "Database skills" (DB skills) have the same domain as the above criteria, i.e. bad, sufficient and good. The hierarchical attribute "Area of Interest" (A. of I.) is described by following values: "Decision Support" (DS), "Knowledge Discovery" (KD) and "Other" (OT).

Table 1. Evaluations of students during bachelor studies

Student	Statistics (Examination)	Computer Networks (Examination)	Computer Networks (Project)	Databases (Examination)	Databases (Project)	Additional Project	Training
1	Sufficient	Bad-Sufficient	Good	Good	Good	Progr	AI
2	Good	Sufficient	Good	Good	Sufficient	Progr	AI
3	Good	Good	Sufficient	Sufficient	Good	DB	DB
4	Sufficient	Sufficient	Sufficient-Good	Sufficient	Sufficient	Progr	Stat
5	Good	Good	Good	Sufficient	Bad	AI	DB, Stat
6	Bad	Sufficient	Bad	Bad	Sufficient	AI	Stat

The next step of proposed methodology is the analysis of each subproblem. The analysis consists of following phases: evaluation of objects according to attributes and/or criteria characteristic for a subproblem, inconsistency analysis using MV-CRSA or IV-DRSA and induction of decision rules based on approximations of decision classes. After this the inconsistencies are propagated upward the hierarchy (if the root of hierarchy is not reached, of course).

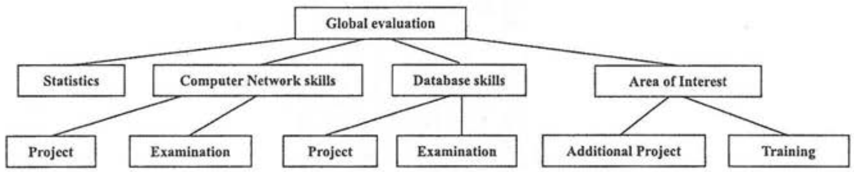

Figure 4. The hierarchy of student qualification problem

Consider for example the "Area of Interest" subproblem where the decision is in the form of hierarchical attribute, i.e. it is a multiattribute classification problem. The subattributes are: "Additional project" and "Training". The evaluation of DM is given in Table 2. For this subproblem the analysis consists in a multiattribute classification. Students are classified into three classes:

- students 1, 4 and 6 belong to class DS (Decision Support),
- students 2 and 5 belong to class KD (Knowledge Discovery), and
- student 3 belongs to class OT (Other).

Table 2. The subtree of the Area of Interest – the multiattribute classification problem

Student	Additional project	Training	Area of Interest
1	*Progr*	*AI*	*DS*
2	*Progr*	*AI*	*KD*
3	DB	DB	OT
4	Progr	Stat	DS
5	*AI*	*DB, Stat*	*KD*
6	*AI*	*Stat*	*DS*

Observe that students 1 and 2 have the same description but belong to different classes. A very interesting situation occurs in the case of student 5, with training having concerned both: "Databases" and "Statistics". According to the generalized indiscernibility relation student 5 could not be discerned from student 6. The inconsistent examples are marked out by italics in the table. Assuming that $B = \{Additional\ project,\ Training\}$, the approximation of classes are as follows:

$$\underline{B}(DS) = \{4\},\ \overline{B}(DS) = \{1,2,4,5,6\},\ BN_B(DS) = \{1,2,5,6\}$$
$$\underline{B}(KD) = \emptyset,\ \overline{B}(KD) = \{1,2,5,6\},\ BN_B(KD) = \{1,2,5,6\}$$
$$\underline{B}(OT) = \{3\},\ \overline{B}(OT) = \{3\},\ BN_B(OT) = \emptyset.$$

The following decision rules (certain and approximate) are induced:

$$\textit{if } (f(x, \textit{AddProj}) \cap \{\textit{Progr}\} \neq \emptyset) \wedge (f(x, \textit{Training}) \cap \{\textit{Stat}\} \neq \emptyset),$$
$$\textit{then } x \in DS \qquad (4)$$
$$\textit{if } (f(x, \textit{AddProj}) \cap \{DB\} \neq \emptyset),\ \textit{then } x \in OT \qquad (3)$$
$$\textit{if } (f(x, \textit{Training}) \cap \{AI\} \neq \emptyset),\ \textit{then } x \in DS \cup KD \qquad (1,2)$$
$$\textit{if } (f(x, \textit{AddProj}) \cap \{AI\} \neq \emptyset),\ \textit{then } x \in DS \cup KD. \qquad (5,6)$$

The inconsistencies between students 1 and 2, and students 5 and 6, are propagated and presented according to generalized decision function in the form of subsets of values on the hierarchical attribute "Area of Interest", i.e. on the upper level, as it is shown in Table 4.

The subsequent analysis concerns the "Computer Network skills" subproblem. It is described by two criteria: "Project" and "Examination". This problem is analysed according to the multicriteria sorting problem. The students are sorted into three classes: bad, sufficient and good. The evaluations given by DM are shown in Table 3.

Table 3. The subtree of the Computer Network skills – the multicriteria sorting problem

Student	Examination	Project	Computer Network skills
1	*Bad-Sufficient*	*Good*	*Good*
2	*Sufficient*	*Good*	*Good*
3	Good	Sufficient	Sufficient
4	*Sufficient*	*Sufficient-Good*	*Sufficient*
5	Good	Good	Good
6	Sufficient	Bad	Bad

The students 1, 2 and 5 belong to the class "good". The students 3 and 4 belong to the class "sufficient", and student 6 belongs to the class "bad". Observe that considering the interval order student 4 is not worse than students 1 and 2, i.e. student 4 dominates students 1 and 2. Nevertheless, student 4 is assigned to a worse class ("sufficient") than students 1 and 2 ("good"). Therefore, students 1,2 and 4 are inconsistent with respect to dominance principle. Italics in the table mark out the inconsistent examples. The approximations of all downward and upward unions of classes (for $B = \{Examination, Project\}$) are mentioned below:

$$\underline{B}(sufficient^{\geq}) = \{1,2,3,4,5\},\ \overline{B}(sufficient^{\geq}) = \{1,2,3,4,5\},$$
$$BN_B(sufficient^{\geq}) = \emptyset,$$
$$\underline{B}(good^{\geq}) = \{5\},\ \overline{B}(good^{\geq}) = \{1,2,4,5\},\ BN_B(good^{\geq}) = \{1,2,4\},$$
$$\underline{B}(sufficient^{\leq}) = \{3,6\},\ \overline{B}(sufficient^{\leq}) = \{1,2,3,4,6\},$$
$$BN_B(sufficient^{\leq}) = \{1,2,4\},$$
$$\underline{B}(bad^{\leq}) = \{6\},\ \overline{B}(bad^{\leq}) = \{6\},\ BN_B(bad^{\leq}) = \emptyset.$$

For this subproblem following decision rules are generated (certain and approximate rules):

$$\textit{if } (u(x, Project) \geq sufficient),\ \textit{then } x \in sufficient^{\geq} \qquad (1,2,3,4,5)$$
$$\textit{if } (u(x, Examination) \geq good),\ \textit{then } x \in sufficient^{\geq} \qquad (3,5)$$
$$\textit{if } (u(x, Examination) \geq good) \wedge (u(x, Project) \geq good),$$
$$\textit{then } x \in good^{\geq} \qquad (5)$$
$$\textit{if } (l(x, Project) \leq sufficient),\ \textit{then } x \in sufficient^{\leq} \qquad (3,4,6)$$
$$\textit{if } (l(x, Project) \leq bad),\ \textit{then } x \in bad^{\geq} \qquad (6)$$
$$\textit{if } (l(x, Exam.) \leq sufficient) \wedge (u(x, Project) \geq good),$$
$$\textit{then } x \in sufficient \cup good. \qquad (1,2,4)$$

On the upper level, for the hierarchical criterion "Computer Network skills", objects take the value according to the generalized decision interval, as shown in Table 4.

We do not present the analysis of "Database skills" subproblem because it is similar to the above. The final evaluations may be found in Table 4 – notice that decision table for this subproblem must be consistent.

The final problem is presented in Table 4. This decision table is composed of consistent examples. The evaluations modified on the basis of inconsistencies in the lower level are in italics.

Table 4. Global evaluation – multicriteria sorting problem

Student	Statistics (Examination)	Computer Network skills (CN skills)	Database skills (DB skills)	Area of Interest (A. of I.)	Global evaluation
1	Sufficient	Sufficient-Good	Good	DS, KD	Acceptable
2	Good	Sufficient-Good	Good	DS, KD	Desirable
3	Good	Sufficient	Sufficient	OT	Non-Acceptable
4	Sufficient	Sufficient-Good	Sufficient	DS	Acceptable
5	Good	Good	Sufficient	DS, KD	Acceptable
6	Bad	Bad	Bad	DS, KD	Non-acceptable

The induced decision rules (certain and approximate) for final decision are presented below:

$$\textit{if } (u(x, DB\ skills) \geq good),\ \textit{then } x \in acceptable^{\geq} \quad (1,2)$$
$$\textit{if } (u(x, CN\ skills) \geq sufficient) \wedge (u(x, A.\ of\ I.) \cap \{DS\} \neq \emptyset),$$
$$\textit{then } x \in acceptable^{\geq} \quad (1,2,4,5)$$
$$\textit{if } (u(x, Stat) \geq good) \wedge (u(x, DB\ skills) \geq good),$$
$$\textit{then } x \in desirable^{\geq} \quad (2)$$
$$\textit{if } (u(x, Stat) \leq sufficient),\ \textit{then } x \in acceptable^{\leq} \quad (1,2,6)$$
$$\textit{if } (u(x, DB\ skills) \leq sufficient),\ \textit{then } x \in acceptable^{\leq} \quad (3,4,5,6)$$
$$\textit{if } (u(x, A.\ of\ I.) \cap \{OT\} \neq \emptyset),\ \textit{then } x \in non\text{-}Acceptable^{\leq} \quad (3)$$
$$\textit{if } (u(x, CN\ skills) \leq bad),\ \textit{then } x \in non\text{-}Acceptable^{\leq}. \quad (6)$$

Let us observe that the analysis of this problem, without partitioning into smaller subproblems, is not easy. The cognitive effort to evaluate students on the basis of these seven attributes and criteria (in real case much and much more) is quite severe.

Finally, let us show the application of the model. Assume that there are two new students evaluated during bachelor studies as shown in Table 5.

Following rules cover the first student:

- subproblem of "Area of Interest":
  $\textit{if } (f(x, Training) \cap \{AI\} \neq \emptyset),\ \textit{then } x \in DS \cup KD;$
- subproblem of "Computer Network skills":
  $\textit{if } (u(x, Project) \geq sufficient),\ \textit{then } x \in sufficient^{\geq}$
  $\textit{if } (l(x, Project) \leq sufficient),\ \textit{then } x \in sufficient^{\leq}.$

In the case of student 2 the following rules are satisfied:

- subproblem of "Area of Interest":
  $\textit{if } (f(x, AddProj) \cap \{DB\} \neq \emptyset),\ \textit{then } x \in OT;$
- subproblem of "Computer Network skills":
  $\textit{if } (u(x, Examination) \geq good),\ \textit{then } x \in sufficient^{\geq}$
  $\textit{if } (l(x, Project) \leq bad),\ \textit{then } x \in bad^{\leq}.$

We assume that the students are assigned to the class "good" on the "Database skills" criterion. Remark that on "Computer Network skills" subproblem the student 2 is described by controversial information. The evaluations of students on the upper level are presented in Table 6.

On the final level the following rules cover student 1:

$\textit{if } (u(x, DB\ skills) \geq good),\ \textit{then } x \in acceptable^{\geq}$
$\textit{if } (u(x, CN\ skills) \geq sufficient) \wedge (f(x, A.\ of\ I.) \cap \{DS\} \neq \emptyset)\ \textit{then } x \in acceptable^{\geq}$
$\textit{if } (u(x, Statistics) \geq good) \wedge (u(DB\ skills, x) \geq good)\ \textit{then } x \in desirable^{\geq}.$

Table 5. New students evaluations during bachelor studies

Student	Statistics (only Examination)	Computer Networks (Examination)	Computer Networks (Project)	Databases (Examination)	Databases (Project)	Additional Project	Training
1	Good	Sufficient	Sufficient	Good	Good	Progr	AI
2	Sufficient	Good	Bad	Sufficient	Good	DB	Progr

Table 6. Global evaluations of new students

Student	Statistics (Examination)	Computer Network skills (CN skills)	Database skills (DB skills)	Area of Interest (A. of I.)	Global evaluation
1	Good	Sufficient	Good	DS, KD	?
2	Sufficient	Bad-Sufficient	Good	OT	?

Student 2 is matched by the following rules:

$if\ (u(x, DB\ skills) \geq good),\ then \in acceptable^{\geq}$

$if\ (u(x, Statistics) \leq sufficient)\ then\ x \in acceptable^{\geq}$

$if\ (u(x, CN\ skills) \leq bad)\ then\ x \in non\text{-}Acceptable^{\leq}$

$if\ (f(x, A.\ of\ I.) \cap \{OT\} \neq \emptyset)\ then\ x \in non\text{-}Acceptable^{\leq}$

Finally, student 1 is evaluated as "desirable", but student 2 is classified between "non-acceptable" and "acceptable", because of good evaluation on criterion "Database skills".

5. Conclusions

In this paper, we presented an extension of a methodology of classification and sorting previously proposed by Greco, Matarazzo and Słowiński. The extension consists in taking into account the Hierarchical Decision Problems i.e. problems containing smaller and smaller subproblems. The use of decision rule preference model resulting from the rough set approach seems to be very convenient and meaningful. Such preference model is more general than the classical functional models considered within the multiattribute utility theory or relational models considered, for example, in outranking methods. The proposed methodology consists of four steps. In the first step the problem is structured by means a hierarchy of attribute and criteria. In the second step, preference information in the form of a set of examples of classification or sorting is given. The third step concerns induction of decision rule preference model handling inconsistencies. The last step is the application of the model to classification or sorting of new objects. The main problem of Hierarchical Decision Problems regards inconsistencies from subproblems in making the intermediate and final decisions. The proposed methodology gives theoretically sound answer to this problem. A didactic example illustrates the very relevant potentiality of this approach with respect to real world applications.

Acknowledgement

The first and the third author wish to acknowledge financial support the State Committee for Scientific Research, KBN research grant no. 8T11F 006 19, and from the Foundation for Polish Science, subsidy no. 11/2001. The research of the second author has been supported by Italian Ministry of Education, University and Scientific Research (MIUR).

References

BANA E COSTA, C.A. and VANSNICK, J-C. (1999) The MACBETH approach: Basic ideas, software, and an application. In: Meskens, N., Roubens, M., eds., *Advances in Decision Analysis.* Kluwer Academic Publishers, Book Series: Mathematical Modelling: Theory and Applications, **4**, 131–157.

BARZILAI, J. (1997) Deriving Weights from Pairwise Comparison Matrices. *Journal of the Operational Society,* **48**, 12, 1226–1232.

BELTON, V. and GEAR, A.E. (1983) On the Shortcoming Saaty's method of Analytic Hierarchies. *Omega,* **11**, 228–230.

BELTON, V. (1999) Multi-criteria problem structuring and analysis in a value theory framework. Chapter 12 in: Gal, T., Stewart T., and Hanne, T., eds., *Advances in Multicriteria Decision Making.* Kluwer Academic Publishers, Dordrecht, 12.1–12.32.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (1996) Rough approximation of a preference relation by dominance relations, ICS Research Report 16/96, Warsaw University of Technology, Warsaw, and (1999) in: *European Journal of Operational Research* **117**, 63–83.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (1998a) A new rough set approach to evaluation of bankruptcy risk. In: Zopounidis, C., ed., *Operational Tools in the Management of Financial Risks.* Kluwer, Dordrecht, 121–136.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (1998b) A new rough set approach to multicriteria and multiattribute classification. In: Polkowski, L. and Skowron, A., eds., *Rough sets and Current Trends in Computing (RSTCTC'98).* Springer-Verlag, 60–67.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (1999) The use of rough sets and fuzzy sets in MCDM. Chapter 14 in: Gal, T., Stewart T., and Hanne, T., eds., *Advances in Multicriteria Decision Making.* Kluwer Academic Publishers, Dordrecht, 14.1–14.59.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2001a) Rough sets theory for multicriteria decision analysis. *EJOR,* **129**, 1, 1–47.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2001b) Conjoint measurement and rough set approach for multicriteria sorting problem in presence of ordinal criteria. In: Coloni, A., Paruccini, M., Roy, B., eds., *AMCDA – Aide Multicritère à la Décision (Multiple Criteria Decision Aiding).* EUR Report, Joint Research Centre, The European Commission. Ispra, 114–141.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2002a) Rough sets methodology for sorting problems in presence of multiple attributes and criteria. *EJOR,* **138**, 247–259.

GRECO, S., MATARAZZO, B. and SŁOWIŃSKI, R. (2002b) An Algorithm for Induction of Decision Rules Consistent with the Dominance Principle. In: Wojciech Ziarko, Y. Y. Yao, eds., *Rough Sets and Current Trends in Computing.* Lecture Notes in Computer Science, 2005, Springer, Berlin, 304–313.

GRECO, S., MATARAZZO B., SŁOWIŃSKI R. (2002c) Multicriteria classification. Chapter 16.1.9. In: Kloesgen, W. and Żytkow, J., eds., *Handbook of Data Mining and Knowledge Discovery*, Oxford University Press, New York, 318–328.

GRZYMALA-BUSSE, J.W. (1992) LERS – a system for learning from examples based on rough sets. In: Słowiński, R., ed., *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory.* Kluwer, Dordrecht, 3–18.

GRZYMALA-BUSSE, J.W. (1997) A new version of the rule induction system LERS. *Fundamenta Informaticae*, **31**, 27–39.

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.

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, 113–129.

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 33–58.

MICHALSKI, R.S. (1983) A theory and methodology of inductive learning. In: Michalski, R.S., Carbonell, J.G. and Mitchell, T.M., eds., *Machine Learning: An Artificial Intelligence Approach*, Morgan Kaufman, 83–134.

MILLER, G.A. (1956) The Magical Number Seven, Plus or Minus Two: Some limits on Our Capacity for Information Processing. *Psychological Review*, **63**, 81–97.

ORŁOWSKA, E. (1998) Introduction: What You Always Wanted to Know about Rough Sets. In. Orłowska, E., ed., *Incomplete Information, Rough Set Analysis*, Physica-Verlag, Heidelberg, New York, 1–20.

PAWLAK, Z. (1982) *Rough sets. International Journal of Information & Computer Sciences*, **11**, 341–356.

PAWLAK, Z. (1991) *Rough Sets. Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers, Dordrecht.

PAWLAK, Z. and SŁOWIŃSKI, R. (1994) Rough set approach to multi-attribute decision analysis. *EJOR*, **72**, 443–459.

ROY, B. (1985) *Méthodologie Multicritère d'Aide à la Décision.* Economica, Paris.

SAATY, T.L. (1980) *The Analytic Hierarchy Process.* McGraw-Hill, New York.

SKOWRON, A. (1993) Boolean reasoning for decision rules generation, In: Komorowski, J. and Ras, Z.W., eds., *Methodologies for Intelligent Systems. Lecture Notes in Artificial Intelligence*, **689**, Springer-Verlag, Berlin, 295–305.

SKOWRON, A. and POLKOWSKI, L. (1997) Decision algorithms: a survey of rough set-theoretic methods. *Fundamenta Informaticae*, **27** (3/4), 345–358.

SLOVIC, P. (1975) Choice between equally-valued alternatives. *Journal of Experimental Psychology: Human Perception Performance*, **1**, 280–287.

SŁOWIŃSKI, R. (1992) 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.

SŁOWIŃSKI, R. and STEFANOWSKI, J. (1992) RoughDAS and RoughClass software implementations of the rough sets approach". In: Słowiński, R., ed., *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory.* Kluwer Academic Publishers, Dordrecht, 445–456.

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

STEFANOWSKI, J. (2001) *Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy.* Wydawnictwo Politechniki Poznańskiej, Poznań, 361.

STEPANIUK, J. (2000) Knowledge Discovery by Application of Rough Set Models. In: Polkowski, L., Tsumoto S., Lin, T.Y., eds., *Rough Set Methods and Application*, Physica Verlag, Heidelberg, 137–231.

WHITE, L.L., WILSON, A.G. and WILSON, D. (eds.), (1969) *Hierarchical Structures.* Elsevier, New York.