Dedicated to
Professor Jakub Gutenbaum
on his 70th birthday

Control and Cybernetics

vol. **29** (2000) No. 1

Rough set based processing of inconsistent information in decision analysis

by

Roman Słowiński*, Jerzy Stefanowski*
Salvatore Greco, Benedetto Matarazzo****

* Institute of Computing Science
Poznań University of Technology
60-965 Poznań, Poland
e-mail: slowinsk@put.poznan.pl, Jerzy.Stefanowski@cs.put.poznan.pl
** Faculty of Economics, University of Catania
95129 Catania, Italy

Abstract: Inconsistent information is one of main difficulties in the explanation and recommendation tasks of decision analysis. We distinguish two kinds of such information inconsistencies: the first is related to indiscernibility of objects described by attributes defined in nominal or ordinal scales, and the other follows from violation of the dominance principle among attributes defined on preference ordered ordinal or cardinal scales, i.e. among criteria. In this paper we discuss how these two kinds of inconsistencies are handled by a new approach based on the rough sets theory. Combination of this theory with inductive learning techniques leads to generation of decision rules from rough approximations of decision classes. Particular attention is paid to numerical attribute scales and preference-ordered scales of criteria, and their influence on the syntax of induced decision rules.

Keywords: decision analysis, rule induction, rough sets, multicriteria decision analysis, classification.

1. Introduction

Various approaches to scientific decision analysis intend to clarify those elements of a decision situation which are not evident to the agents involved (decision makers, stake holders, experts) and which may influence their attitude towards decision. In other words, the elements revealed by scientific decision analysis either explain the situation, or recommend - or simply favor, a behavior that will increase the consistency between the possibilities offered by the situation, and the objectives and value systems of the agents (see Roy, 1985).

The main difficulty of the explanation and recommendation tasks is connected with the usually *inconsistent information* about the decision situation. One cannot expect to get any perfect explanation nor recommendation from inconsistent data, even if very sophisticated methods are used. However, one can expect to learn what are the *certain* conclusions and the *possible* conclusions that can be drawn from available information.

In this paper, we consider the case where the available input information is a record of experience, a list of observations, or a set of decision examples. Such information is often presented in the form of a *data table* where each row corresponds to a single case, observation or decision example, called *object*, and columns correspond to *attributes* characterizing objects. Drawing conclusions on the basis of this information naturally leads to representation in a form of *decision rules*. The decision rules are logical expressions of the form: **if** *condition-part* **then** *decision*, where condition-part means a conjunction of elementary conditions being some tests on values of attributes, and decision means disjunction of possible decision classes.

The rules are constructed using the *inductive learning principle* (Mitchell, 1997). Although other representations of knowledge are available, like (discriminant) functions, binary relations, or decision trees (see e.g. Weiss and Kulikowski, 1991, Michalski et al., 1998, or Mitchell, 1997), it is claimed that decision rules are more natural and readable for the users. Moreover, the rules "speak" the language of examples given in the input information, they generalize them by reducing all redundant pieces of information and they are acknowledged by real facts from the input information.

There are two general views on inductive learning: *descriptive* and *prescriptive*. The first view, also called explanatory, is connected with discovering useful dependencies in the data that could help in better understanding of circumstances in which decisions were made. The second aims at improving decision making for future cases. Although induction of rules from examples is a typical approach of artificial intelligence (in particular machine learning), it is concordant with the principle of posterior rationality of March (1988) and with the aggregation-disaggregation logic of Jacquet-Lagreze (1982). The rules explain the preferential attitude of the agents and enable an understanding of the reasons of his/her preferences. As pointed out by Langley and Simon (1998) the recognition of the rules by the agent justifies their use for decision support. So, the preference model in the form of rules derived from examples fulfils both the explanation and recommendation functions with respect to decision analysis.

The inconsistencies inherent to input information cannot be considered as noise or error. In decision context, they reflect the impression of the model used for description, the hesitations of agents, and the unstable character of agent's preferences. Thus, the inconsistencies should not be neglected nor amalgamated with certain information, but rather separated in order to get exact (certain) and approximate (possible) conclusions (rules).

The rough sets theory (RST) proposed by Z.Pawlak (1982,1991) is particularly useful in dealing with inconsistency of input information. It clearly separates certain and possible information by building lower and upper approximations of each decision class (category). The difference between upper and lower approximation constitutes a boundary that groups doubtful information. As a consequence, decision rules are induced either from lower approximations, upper approximations, or from boundaries of decision classes, and thus are categorized into exact and approximate ones.

Up to now, the rough set approach to decision analysis was focused on *multiattribute classification*, where objects were described by *attributes* defined on *nominal* or *ordinal scales*. There are several successful case studies of this approach (see e.g. Pawlak, 1991, 1995, Słowiński, 1996, or Komorowski et al., 1999). The difficulties arise, however, when attributes are defined on *numerical* scales. In this case, direct analysis of such data may lead to inducing rules of poor quality, so *discretization* techniques are often applied in a pre-processing phase (Chmielewski and Grzymala, 1995). Discretization converts numerical scales into ordinal scales represented by ordered sub-intervals. On the other hand new approaches have been proposed recently by Słowiński and Vanderpooten (1995), Krawiec et al. (1998), Skowron and Stepaniuk (1996), Greco et al. (1999), Grzymala and Stefanowski (1999), to directly handle numerical data.

The rough set approach can handle inconsistency manifested by *indiscernibility* of at least two objects having the same description by attributes but assigned to different decision classes.

The original rough set approach is not able, however, to discover inconsistencies coming from consideration of *criteria*, i.e. attributes with preference-ordered scales. The scales of criteria may be cardinal or ordinal (Shoemaker, 1982), depending whether the strength of preference is meaningful for the scale or not. The difference between scales of attributes and scales of criteria exists in consideration of *ordered preferences* in the latter case. Product quality, market share, debt ratio are examples of criteria. Regular attributes, e.g. symptoms, colors, textural features, traditionally considered in the rough set methodology, are different from criteria because their domains are not preference-ordered (see, e.g. Greco et al., 1996, 1999a). Therefore, the classical rough sets theory cannot be applied to *multicriteria sorting problems*, i.e. problems of assigning a set of objects described by a set of criteria to one of pre-defined and ordered classes.

Consider, for example, two firms, A and B, evaluated for 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 large value of the debt ratio, and evaluations of these firms on other attributes are equal, then, from the bankruptcy risk point of view, firm A *dominates* over firm B. Suppose, however, that firm A has been assigned to a class of higher risk than firm B. This is obviously inconsistent with the dominance principle. Within the original RST, the two firms will be considered as just discernible and no inconsistency will be stated.

For this reason, Greco, Matarazzo and Słowiński (1996) have proposed an extension of the RST that is able to deal with this kind of inconsistency typical to exemplary decisions (in input information) in Multi Criteria Decision Analysis (MCDA) problems. This innovation is mainly based on substitution of the *indiscernibility relation* by a *dominance relation* in the rough approximation of decision classes. It is also possible to infer from exemplary decisions the preference model in terms of the generalized decision rules.

The aim of this paper is to survey the induction of decision rules from rough approximations of decision classes while handling two kinds of inconsistencies in the input information. The first kind corresponds to the typical case of multi-attribute classification, involving indiscernibility relation. Particular attention will be paid to presentation of new approaches for handling numerical attributes. The second kind corresponds to multicriteria sorting problems involving dominance relation. We will briefly review a new generalization of the rough sets theory to handle this problem. Regarding the survey character of this paper, we will introduce an illustrative example for comparing approaches discussed.

2. Decision rules for multiattribute classification problems

2.1. The general idea of the rough sets theory

Information about objects is provided in the form of a *data table*. Rows of the table refer to objects (actions), whereas columns refer to different attributes considered. Each entry of this table contains the value of an attribute of a given object. Formally, the data table (also called information system or information table) is defined as a pair $S = (U, A)$ where U is a finite set of *objects* and A is a finite set of *attributes*. With every attribute $a \in A$, a set of its values V_a is associated. Each attribute a determines an information function f_a: $U \to V_a$ such that for any $a \in A$, and $x \in U$, $f_a(x) \in V_a$.

In practice, we are mostly interested in analyzing a special case of data table called *decision table*. It is a data table $(U, A \cup D)$, where D is a set of *decision attributes*. We consider a simple case where D is a singleton $\{d\}$; moreover in any case D can be always transformed to $\{d\}$. The elements of A are called *condition* attributes. Let us assume that the cardinality of the set V_d of values of the decision attribute d is equal a finite number k. The decision attribute determines the partition of the set of all objects U into k disjoint classes $X_1, X_2, \ldots, X_k$, called *decision classes*.

Solving *multiattribute classification* problems usually involves looking for dependencies between attributes, in particular between the condition attributes and the decision. This leads to discovering decision rules in the input table that could be used for two aims: either to *explain past decisions* made for learning examples or to *make recommendations* for future cases, i.e. supporting assignment of a new object to one of existing decision classes.

The rough sets theory, Pawlak (1982, 1991), is founded on the assumption that having some information about considered objects one can establish *relations* between these objects. The basic observation is that objects characterized by the same description are *indiscernible* (indistinguishable) due to limited information about them. The indiscernibility relation generated in this way is the mathematical basis of the original Pawlak's concept of the RST.

Formally, the *indiscernibility relation* is associated with every non-empty subset of attributes $P \subseteq A$ and $\forall x, y \in U$ is defined as

$$xI_Py \Leftrightarrow \ f_a(x) = f_a(y) \ \ \forall a \in P$$

The indiscernibility relation defined in this way is an *equivalence* relation (reflexive, symmetric and transitive) and generates the partition of objects from U. The family of all equivalence classes of relation I_P is denoted by U/I_P. These classes are called *P-elementary sets.* An elementary equivalence class (i.e. a single block of the partition U/I_P) containing element x is denoted $I_P(x)$.

The indiscernibility relation is not the only possible relation between objects. For instance, quite often due to imprecise description of objects by attributes, small differences between objects' description are not considered important for their discrimination. This situation may be formally modeled by *similarity* or *tolerance relations* (see e.g. Polkowski and Skowron, 1995, Skowron, 1995, Stepaniuk, 1996, Krawiec et al., 1998, Greco et al., 1998b, Marcus, 1994, Słowiński and Vanderpooten, 1998).

As the similarity relations R do not generate partitions on U, similarity classes $R(x)$ are defined for each object $x \in U$, instead of equivalence classes. Formally, the similarity class of x consists of objects y similar to x:

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

This relation has an interesting property, i.e. it may be only reflexive, relaxing symmetry and transitivity (see discussion in Słowiński and Vanderpooten, 1995). Although at a first glance a non symmetric similarity relation may appear odd we have several intuitive examples where such situation may occur. We always say that a child is similar to a parent or that a copy of a painting is similar to the original, without claiming the inverse. Therefore, symmetry is not imposed and it makes sense to consider the inverse relation

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

The $R^{-1}(x)$ is the class of objects y to which x is similar.

Then, considering the concept of ambiguous and non-ambiguous objects with respect to available information, we come naturally to defining *rough approximation* of a subset of objects $X \subset U$. Precisely, the rough approximation of X is characterized by its *lower and upper approximations* defined respectively as:

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

The set $Bn_R(X) = \underline{R}(X) - \bar{R}(X)$ is called the *R-boundary* of X. The lower approximation $\underline{R}X$ is a maximal set including objects that can be certainly classified as elements of X while the upper approximation $\bar{R}X$ is a minimal set of objects which can be possibly classified in X, taking into account information available. The set $Bn_R(X)$ reflects information ambiguity in describing the set X, i.e. it contains inconsistent (ambiguous) objects.

In the case where the similarity relation boils down to the indiscernibility relation defined on a set of attributes P, $R^{-1}(x) = R(x) = I_P(x)$ for any $x \in U$. Thus, *lower* and *upper approximations* are defined as:

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

The definition of approximations of a subset $X \subset U$ can be extended to a classification, i.e. a partition $\mathcal{Y}$ of U. Subsets Y_i, $i = 1, ..., n$, are disjunctive classes of $\mathcal{Y}$. By P-lower (P-upper) approximation of $\mathcal{Y}$ we mean sets $\underline{P}(\mathcal{Y}) = \{\underline{P}(Y_1), \underline{P}(Y_2), ..., \underline{P}(Y_n)\}$ and $\bar{P}(\mathcal{Y}) = \{\bar{P}(Y_1), \bar{P}(Y_2), ...\bar{P}(Y_n)\}$ respectively. The coefficient

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

is called quality of approximation of classification $\mathcal{Y}$ 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.

Another issue of great practical importance is that of "superfluous" data in an information table. Superfluous data can be eliminated, in fact, without deteriorating the information contained in the original table. For this elimination the concept of the so called reduct is of crucial importance.

A reduct of P is defined with respect to an approximation of a partition $\mathcal{Y}$ of U. It is then called $\mathcal{Y}$-reduct of P (notation $Red_{\mathcal{Y}}(P)$) and specifies a minimal subset P' of P which keeps the quality of classification unchanged, i.e. $\gamma_{P'}(\mathcal{Y}) = \gamma_P(\mathcal{Y})$. In other words, the attributes that do not belong to $\mathcal{Y}$-reduct of P are superfluous with respect to the classification $\mathcal{Y}$ of objects from U. More than one $\mathcal{Y}$-reduct (or reduct) of P may exist in an information table. The set containing all the indispensable attributes of P is known as the $\mathcal{Y}$-core. Obviously, since the $\mathcal{Y}$-core is the intersection of all the $\mathcal{Y}$-reducts of P, it is included in every $\mathcal{Y}$-reduct of P. It is the most important subset of attributes, because none of its elements can be removed without deteriorating the quality of classification.

2.2. Decision rules

Let us consider a decision table. The decision attribute d induces a partition of U deduced from the indiscernibility relation. There is a tendency to examine functional dependency between the partition induced by d and partitions induced by condition attributes from C. A decision table may also be seen as a set of learning examples which enable generation of *decision rules*. If the decision table is consistent, rules are induced from decision classes. Otherwise decision rules are generated from approximations of decision classes.

This special way of treating inconsistencies in the input data is the main point where the concept of the rough sets theory is used in the rule induction phase. The step of induction follows the inductive principle which is a common aspect with machine learning algorithms. As a consequence of using the approximations, induced decision rules are categorized into *certain (exact)* and *approximate (possible)* ones, depending on the used lower and upper approximations (or boundaries), respectively.

Decision rules are represented in the following form:

$$\wedge(c\#v) \rightarrow \vee(d\#w)$$

where $c \in C$ is condition attribute, v is a value of attribute c, and w is a value of decision d, $\#$ means one of the operators $=, \leq, \geq, <, >, \in, SIM$. If two consecutive elementary conditions $(c \geq v_1)$ and $(c \leq v_2)$, where $(v_1 < v_2)$ concern the same condition attribute c then we get the following new condition $(c \in [v_1, v_2])$. SIM is a '*similar to*' operator resulting from similarity measures used to define the relation of similarity.

We will call $s = \wedge(c\#v)$ and $t = \vee(d\#w)$ *condition* and *decision* part of a rule, respectively. If the decision part contains one element $(d = w)$ only, then the rule is exact, otherwise it is approximate. The exact decision rules, indicating assignment to class X_w, are induced under assumption that objects belonging to the lower approximation of decision class X_w are positive, while all the others are negative. The approximate rules, with decision part $t = (d = w) \vee (d = v)$, are induced under assumption that objects belonging to common part of subboundaries of classes X_w and X_v only are positive while all the others are negative.

It is said that an object $x \in A$ supports a rule $s \rightarrow t$ (or a rule covers object x) if its description satisfies both expressions s and t. Let $[s]$ denote set of such objects. A decision rule $s \rightarrow t$ is *exact* if $[s]$ is a subset of the lower approximation of the decision class indicated by t. The decision rule should have a *non-redundant* condition part, i.e. no other rule can be constructed from a proper subset of elementary conditions occurring in the given rule.

Induction of decision rules from decision tables is a complex task and a number of various algorithms have been already proposed (for some reviews see e.g. Grzymala, 1992, Skowron, 1995, Bazan, 1998, Grzymala et al., 1996, Stefanowski, 1998, Komorowski et al., 1999, Stefanowski and Vanderpooten, 2000). In fact, there is no unique "rough set approach" to rule induction as elements of rough sets can be used on different stages of the process of induction and data processing. Nowadays, the most often used approaches and software systems are:

- System LERS (Learning from Examples based on Rough Sets) - Grzymala (1992), which itself has four different options of rule induction; the most popular of them seems to be LEM2 algorithm.
- Systems based on a discernibility matrix and boolean reasoning techniques (Skowron, 1993, 1995). They were extended by several additional strategies connected with, e.g., approximation of reducts, looking for dynamic reducts, boundary region thinning, data filtration and tolerance relation, see Bazan (1998), Skowron (1995), Skowron and Polkowski (1997), Nguyen (1998a), Komorowski et al. (1999). Their implementations form a computational kernel of the system *Rosetta* (Komorowski et al., 1997).
- Systems RoughDAS, Profit and ROSE (Słowiński and Stefanowski, 1998, Predki and Wilk, 1999) which offer several rule induction options that are further described in this paper.
- Systems DataLogic and KDD-R that use the probabilistic extension of the original rough set model called *variable precision rough sets* model and are oriented towards data mining applications, Ziarko (1993, 1995).

Other learning algorithms inspired by data mining techniques are also known (see e.g. Kryszkiewicz, 1998a, Lin, 1996, Stefanowski and Vanderpooten, 1994, 2000). One should also remember about interesting proposals of the *probabilistic rough classifier* developed by Lenarcik and Piasta (1997).

If we notice that these algorithms aim at inducing the *rule descriptions of decision classes*, we can distinguish three main groups of existing approaches:

(1) algorithms inducing the minimum set of rules,
(2) algorithms inducing the exhaustive set of rules,
(3) algorithms inducing the satisfactory set of rules.

The first category of algorithms is focused on covering input objects using the minimum number of necessary rules while the second group tries to generate all possible decision rules in the simplest form. The third category of algorithms gives as a result the set of decision rules which satisfy user's requirements given a priori. The user may prefer to get strong decision rules (i.e. rules supported by a relatively large number of input objects), having good discriminatory ability, with emphasis on the syntax of the condition part (e.g. using some specific attributes or elementary conditions). The differences between these approaches will be further illustrated on a simple example.

It is worth noticing that the problem of finding a minimum set of rules (i.e. category 1) covering all examples can be seen as the problem of minimal cover which is NP-hard (see Pawlak, 1991, Stefanowski and Vanderpooten, 1994). Therefore, heuristic approximation algorithms are usually applied. It was shown, moreover, that category (2) of algorithms refers to the problem of searching for reducts of minimal length, also NP-hard (see Skowron and Rauszer, 1992). Finally, for category (3) the computation cost although still high, can be reduced in practice if one uses strong requirements for the rule support/strength and some pruning techniques that reject early the unnecessary candidates for rules (see discussions in Kryszkiewicz, 1998b, Stefanowski and Vanderpooten, 1994, 2000).

Two general perspectives of rule induction are considered: either explanation of existing decision situation using the rules or creation of classification systems. The first aim is also connected with the knowledge discovery perspective, Fayyad et al. (1996).

It must be noticed, however, that the two perspectives of rule induction are perceived as different. One of the basic distinctions consists in different evaluation criteria. In classification-oriented induction, rules are parts of a classification system; hence, evaluation refers to a complete set of rules. The evaluation criterion is usually unique and defined as classification (predictive) accuracy (see Weiss and Kulikowski, 1991). In explanation or discovery-oriented induction, each rule is evaluated individually and independently as a possible representation of an interesting pattern. The evaluation criteria are multiple and considering them together is not easy. Moreover, the definition of criteria depends on the application problem and the user's requirements. More details on this topic are given in Stefanowski (1998a), Stefanowski and Vanderpooten (2000).

2.3. Handling numerical attributes

The original rough set approach and rule induction techniques seem to be insufficient when applied directly to data sets containing *numerical* attributes, i.e. attributes with real number or integer domains. Rules induced directly from numerical attributes are of poor quality (very short, weak, and numerous).

The typical approach to solve this problem is to use *discretization techniques* which convert numerical attributes into discrete, ordinal ones. During discretization a number of cut-points are determined, dividing the attribute domain into consecutive subintervals. Many discretization methods (see reviews in Chmielewski and Grzymala, 1995, Dougherty et al., 1995, Nguyen, 1998a,b, Susmaga, 1997) can be applied as a *preprocessing step* before rule induction. In general, no one discretization method is optimal for all situations.

Additional aspect of employing these methods is that they determine discretization independently from the RST analysis. It is possible that obtained discretized subintervals may be more or less arbitrary and not lead to acceptable results. Therefore, some newly proposed extensions to the RST enable to analyse directly numerical attributes without any pre-discretization.

The generalization of the rough sets theory based on similarity relation, Słowiński and Vanderpooten (1995, 1999) is one of the solution. It is particularly useful if numerical attributes are affected by imprecise measurement, random fluctuation of some parameters, etc. In Krawiec et al.(1998) authors presented a quite efficient automatic procedure for inferring similarity relation from data, and then an algorithm for generating certain and robust decision rules.

Another group of extended approaches offers a new version of rough set based rule induction algorithms which do not require preliminary discretization of numerical attributes. One of these approaches, called MODLEM (Stefanowski, 1998c, Grzymala and Stefanowski, 1999), is a modified version of the LEM2 algorithm. The LEM2 (Grzymala, 1992) is a popular RST based rule induction algorithm for getting the minimum set of decision rules. Let us briefly comment the idea of modifications. Numerical attributes are handled by the learning algorithm MODLEM at the moment when elementary conditions of a rule are created. In the original version of LEM2 elementary conditions are represented as pairs $(c = v)$ where c is an attribute and v is its value. In MODLEM conditions are represented in the form of either $(c < v)$, $(c \geq v)$ or $(c = [v_1, v_2))$ (resulting from an intersection of two conditions $(c < v_2)$, $(c \geq v_1)$, $v_1 < v_2$, for the same attribute). The candidates for the cutpoint v are locally scanned for the range of each numerical attribute c taking into account unique values with their decision class assignment. The best cutpoint among all tested ones is chosen to be further compared against other attributes. The best condition for all compared attributes is chosen for adding to the condition part of the rule. As the evaluation measure indicating the best condition, typical entropy measures or Laplace accuracy (Clark and Boswell, 1991) are used. An experimental study performed in Grzymala and Stefanowski (1999) showed that the MODLEM algorithm used as a classifier performs better than the original version of LEM2 and produces classification accuracy comparable with such machine learning techniques as C4.5. Similar motivations were the basis of modifications introduced in the Explore algorithm which induces satisfactory set of rules (Stefanowski, 1998b). Some of the above techniques for handling numerical attributes are available within the newly offered ROSE (Rough Set data Explorer) system (Predki et al., 1999).

2.4. Illustrative example

Let us illustrate these concepts with a small example. We assume that 15 objects, considered as examples of classification, are described by 3 attributes a_1, a_2, a_3. Objects are classified into three decision classes according to the value of a decision attribute d. The data table analyzed is presented in Table 1.

First, let us analyze the original data table (objects described by original numerical values of attributes) using the 'classical' rough set approach based on the indiscernibility relation. All objects are discernible and there are no

No.	a_1	a_2	a_3	d
1	8.0	14	9.5	3
2	4.0	7	3.0	2
3	2.5	3	5.0	2
4	1.5	5	2.0	1
5	7.5	15	8.5	3
6	0.5	1	2.5	1
7	6.5	6.5	4.5	1
8	5.0	8	7.0	2
9	7.5	9	10.5	3
10	3.0	6	6.5	2
11	5.5	10	8.5	2
12	2.5	4	6.0	1
13	6.5	13	8.0	2
14	3.5	10	5.5	1
15	6.0	11	7.5	3

Table 1. Data table

inconsistencies. This means that lower approximations are equal upper approximations. As a result of applying LEM2 algorithm to this data table we can induce 14 exact rules - presented below:

rule 1. $(a_1{=}1.5) \rightarrow (d{=}1)$ $\{4\}$
rule 2. $(a_1{=}0.50) \rightarrow (d{=}1)$ $\{6\}$
rule 3. $(a_1{=}3.5) \rightarrow (d{=}1)$ $\{14\}$
rule 4. $(a_2{=}6.5) \rightarrow (d{=}1)$ $\{7\}$
rule 5. $(a_2{=}4.0) \rightarrow (d{=}1)$ $\{12\}$
rule 6. $(a_1{=}4.0) \rightarrow (d{=}2)$ $\{2\}$
rule 7. $(a_1{=}5.0) \rightarrow (d{=}2)$ $\{8\}$
rule 8. $(a_1{=}3.0) \rightarrow (d{=}2)$ $\{10\}$
rule 9. $(a_1{=}5.0) \rightarrow (d{=}2)$ $\{11\}$
rule 10. $(a_2{=}3.0) \rightarrow (d{=}2)$ $\{3\}$
rule 11. $(a_2{=}13) \rightarrow (d{=}2)$ $\{13\}$
rule 12. $(a_1{=}7.5) \rightarrow (d{=}3)$ $\{5,9\}$
rule 13. $(a_1{=}8.0) \rightarrow (d{=}3)$ $\{1\}$
rule 14. $(a_1{=}6.0) \rightarrow (d{=}3)$ $\{15\}$

The rules are presented in the following form: first the syntax of the rules, then the identifiers of objects covered by the rule.

One can easily notice that the quality of results obtained from non-discretized data is very poor. Induced decision rules are numerous, very specific and nearly all of them are supported by one learning example. Therefore, this input data should either be discretized before the rough set analysis or one of the approaches

No.	a_1	a_2	a_3	d
1	5	4	5	3
2	3	2	2	2
3	2	1	2	2
4	1	1	1	1
5	5	4	4	3
6	1	1	1	1
7	4	2	2	1
8	3	2	4	2
9	5	3	5	3
10	2	2	4	2
11	3	3	4	2
12	2	1	3	1
13	4	3	4	2
14	2	3	3	1
15	4	3	4	3

Table 2. Coded data table

specialized in direct handling of numerical data in rule induction applied.

For instance, let us consider the following proposal of discretization: for the attribute a_1, code 1 corresponds to ≤ 2.25, code 2 corresponds to interval $(2.25, 3.5]$, code 3 to $(3.5, 5.5]$, code 4 to $(5.5, 6.5]$ and code 5 corresponds to > 6.5. The attribute a_2 is discretized as: code 1 corresponds to ≤ 5, code 2 corresponds to interval $(5, 8]$, code 3 to $(8, 13]$, code 4 to > 13. Finally, for the attribute a_3 code 1 corresponds to ≤ 2.5, code 2 corresponds to interval $(2.5, 5]$, code 3 to $(5, 6]$, code 4 to $(6, 8.5]$ and code 5 corresponds to > 8.5.

If we use this discretization we can transform Table 1 into Table 2. Then, rough sets theory indicates two inconsistent examples, no. 13 and no. 15, having the same description in values of attributes $(a_1 = 4, a_2 = 3, a_3 = 4)$ and assigned to different decision classes.

The elementary sets (for all attributes) are the following:

$$
\begin{aligned}
I_P(1) &= \{1\},\\
I_P(2) &= \{2\},\\
I_P(3) &= \{3\},\\
I_P(4) &= I_P(6) = \{4,6\},\\
I_P(5) &= \{5\},\\
I_P(7) &= \{7\},\\
I_P(8) &= \{8\},\\
I_P(9) &= \{9\},\\
I_P(10) &= \{10\},\\
I_P(11) &= \{11\},\\
I_P(12) &= \{12\},\\
I_P(13) &= I_P(15) = \{13,15\},\\
I_P(14) &= \{14\}.
\end{aligned}
$$

Using them we can obtain rough approximations of each decision class. The class ($d = 1$) is exactly described (lower and upper approximations are the same and contain 5 objects). The class ($d = 2$) is roughly described, lower approximation consists of 5 objects, while upper approximation of 7 objects. Similary for class ($d = 3$); lower approximation is built of 3 objects while the upper one of 5 objects. The rough approximations are presented below:

- decision class 1:
  - lower approximation {4,6,7,12,14}
  - upper approximation {4,6,7,12,14}
- decision class 2:
  - lower approximation {2,3,8,10,11}
  - upper approximation {2,3,8,10,11,13,15}
- decision class 3:
  - lower approximation {1,5,9}
  - upper approximation {1,5,9,13,15}.

The boundary of decision classes 2 and 3 is composed of inconsistent objects {13,15}. The quality of classification of objects according to attribute d is

$$\sum_{i=1}^{3} |\underline{C}(\text{class i})|/|U|= (3+5+5)/15=0.867.$$

So, lower approximations of each decision class and boundary region of class 2 and 3 will be the basis of rule induction phase. The LEM2 algorithm produced 8 decision rules, including one approximate rule (no. 8):

Exact rules:

rule 1. $(a_1{=}1) \rightarrow (d{=}1)$ {4,6}
rule 2. $(a_3{=}3) \rightarrow (d{=}1)$ {12,14}
rule 3. $(a_1{=}4) \wedge (a_3{=}2) \rightarrow (d{=}1)$ {7}
rule 4. $(a_1{=}3) \rightarrow (d{=}2)$ {2,8,11}
rule 5. $(a_1{=}2) \wedge (a_3{=}2) \rightarrow (d{=}2)$ {3}
rule 6. $(a_1{=}2) \wedge (a_2{=}2) \rightarrow (d{=}2)$ {10}
rule 7. $(a_1{=}5) \rightarrow (d{=}3)$ {1,5,9}

Approximate rule:

rule 8. $(a_1{=}4) \wedge (a_2{=}3) \rightarrow (d{=}2) \vee (d{=}3)$ {13,15}

This result seems to be more readable than the previous result obtained for non-discretized data and also better supported by learning examples.

The result presents a minimal set of rules covering all examples. We can compare it to the set of all rules consisting of 16 rules (induced by means of *Explore* algorithm, Stefanowski and Vanderpooten, 1994, 2000). Below we give only these rules which do not appear in the minimal set:

Additional exact rules:

rule 9. $(a_3{=}1) \rightarrow (d{=}1)$ {4,6}
rule 10. $(a_1{=}2) \wedge (a_2{=}3) \rightarrow (d{=}1)$ {14}
rule 11. $(a_1{=}4) \wedge (a_2{=}2) \rightarrow (d{=}1)$ {7}
rule 12. $(a_1{=}2) \wedge (a_3{=}4) \rightarrow (d{=}2)$ {10}
rule 13. $(a_2{=}2) \wedge (a_3{=}4) \rightarrow (d{=}2)$ {8,10}
rule 14. $(a_2{=}1) \wedge (a_3{=}2) \rightarrow (d{=}2)$ {3}
rule 15. $(a_2{=}4) \rightarrow (d{=}3)$ {1,5}
rule 16. $(a_3{=}5) \rightarrow (d{=}3)$ {1,9}

One can notice that some of these rules give additional information about input data. For instance, rule no. 9 shows that examples 4 and 6 can be described in another way than with the rule no. 1. The same refers to using rule no. 11 instead of rule no. 3.

Let us consider now the new approaches that handle directly numerical attributes. First, we will use the generalization of the rough sets theory based on similarity relation. In general, various definitions of similarity measures can be applied (see reviews in Słowiński and Vanderpooten, 1995). Here we use one of the definitions that models similarity by means of ϵ-tolerance intervals, see Krawiec et al. (1998). More formally, if $f_q(x)$ is a value of attribute c for object x and $P \subseteq A$ is a subset of considered attributes then the relation of similarity is defined as:

$$xRy \Longleftrightarrow f_c(x) \in [f_c(y) - \epsilon_c^-(y), f_c(y) + \epsilon_c^+(y)] \ \forall_{c \in P}$$

The values $\epsilon_c^-(y)$ and $\epsilon_c^+(y)$ are assessed by an automatic procedure based on the analysis of examples from the input data table. For more details on this procedure see Krawiec et al. (1998).

Applying this procedure to our Table 1, we obtained a similarity relation presented in Table 3. It should be read as follows: in columns corresponding to

No.	a_1	a_2	a_3	d
1	8.0 [6.5,8.0]	14 [10,15]	9.5 [7.25,10.5]	3
2	4.0 [2.75,6.25]	7 [4.5,9.5]	3.0 [2.25,5.75]	2
3	2.5 [0.5,2.75]	3 [1,4.5]	5.0 [2.25,5.75]	2
4	1.5 [0.5,2.75]	5 [1,7.5]	2.0 [2,5.75]	1
5	7.5 [6.5,8]	15 [10,15]	8.5 [6.75,10]	3
6	0.5 [0.5,2.75]	1 [1,4.5]	2.5 [2,5.75]	1
7	6.5 [5.75,7]	6.5 [4.5,9.5]	4.5 [2.25,5.75]	1
8	5.0 [2.75,6.25]	8 [4.5,9.5]	7.0 [5.75,9]	2
9	7.5 [6.5,8]	9 [7.5,10]	10.5 [7.25,10.5]	3
10	3.0 [2,4.5]	6 [4.5,7.5]	6.5 [5.75,9]	2
11	5.5 [2.75,6.25]	10 [8.5,10.5]	8.5 [6.75,10]	2
12	2.5 [2,2.75]	4 [1,7.5]	6.0 [2.25,6.75]	1
13	6.5 [6.25,7.5]	13 [10,14.5]	8.0 [6.75,10]	2
14	3.5 [2.75,4.5]	10 [9.5,14.5]	5.5 [2.25,6.75]	1
15	6.0 [5.25,6.5]	11 [10,14.5]	7.5 [6.75,10]	3

Table 3. Similarity relation

an attribute there are values of the attribute for each object together with the ϵ intervals of similarity with respect to these values.

Having these intervals we can check that, for instance, object 13 is similar to object 1, since: $f_{a1}(13) = 6.5 \in [6.5, 8.0], f_{a2}(13) = 13 \in [10, 15], f_{a3}(13) = 8.0 \in [7.25, 10.5]$. However, the inverse relation '1 similar to 13' does not hold as, for instance, $f_{a1}(1) = 8.0 \notin [6.25, 7.5]$. This illustrates that similarity relation is non-symmetric. Using the ϵ-tolerance intervals calculated for all objects from table 1 we can calculate the following classes of the similarity relation:

$$\begin{array}{ll}
R^{-1}(1) = \{1,5\} & R(1) = \{1,5,13\}, \\
R^{-1}(2) = \{2\} & R(2) = \{2\}, \\
R^{-1}(3) = \{3,4,12\} & R(3) = \{3,6\}, \\
R^{-1}(4) = \{4\} & R(4) = \{3,4,6\}, \\
R^{-1}(5) = \{1,5\} & R(5) = \{1,5,13,15\}, \\
R^{-1}(6) = \{3,4,6\} & R(6) = \{3,6\}, \\
R^{-1}(7) = \{7\} & R(7) = \{7\}, \\
R^{-1}(8) = \{8\} & R(8) = \{8,10\}, \\
R^{-1}(9) = \{9\} & R(a9) = \{9\}, \\
R^{-1}(10) = \{8,10\} & R(10) = \{10\}, \\
R^{-1}(11) = \{11,15\} & R(11) = \{11\}, \\
R^{-1}(12) = \{12\} & R(12) = \{3,12\}, \\
R^{-1}(13) = \{1,5,13,15\} & R(13) = \{13,15\}, \\
R^{-1}(14) = \{14\} & R(14) = \{14\}, \\
R^{-1}(15) = \{13,15\} & R(15) = \{11,13,15\}
\end{array}$$

Using these similarity classes we can define rough approximations of decision classes:

- decision class 1:
  - lower approximation {4,7,12,14}
  - upper approximation {3,4,6,7,12,14}
- decision class 2:
  - lower approximation {2,8,10}
  - upper approximation {2,3,6,8,10,11,13,15}
- decision class 3:
  - lower approximation {1,5,9}
  - upper approximation {1,5,9,11,13,15}.

Then, a modification of the LEM2 approach can be used to induce exact rules from the lower approximations. Some objects from boundary regions remain uncovered. The set of rules is the following:

rule 1. $(a_1 \in [6.5,\ 8]) \rightarrow (d{=}3)$ {1,5,9}
rule 2. $(a_1 \in [2.75,\ 6.25]) \wedge (a_2 \in [4.5,\ 9.5]) \rightarrow (d{=}2)$ {2,8,10}
rule 3. $(a_1 \in [0.5,\ 2.75]) \wedge (a_2 \in [1,\ 7.5]) \rightarrow (d{=}1)$ {4,6,12}
rule 4. $(a_2 \in [8.5,\ 10.5]) \wedge (a_3 \in [2.25,\ 5.75]) \rightarrow (d{=}1)$ {14}
rule 5. $(a_1 \in [5.75, 7]) \wedge (a_2 \in [4.5,\ 7.5]) \rightarrow (d{=}1)$ {7}

One can notice that these rules are more general than the previous ones.

Let us apply now the next extended approach, i.e. the new algorithm MODLEM (Stefanowski, 1998c), to original data table without any preliminary discretization. The set of seven decision rules was obtained. There is no inconsistency and all rules are exact.

rule 1. $(a_3 \leq 6.25) \wedge (a_2 \geq 3.50) \wedge (a_1 \leq 3.75) \rightarrow (d{=}1)\ \{4,12,14\}$
rule 2. $(a_3 \leq 4.75) \wedge (a_2 \leq 6.75) \rightarrow (d{=}1)\ \{4,6,7\}$
rule 3. $(a_1 \in [2.0,5.75)) \wedge (a_3 \leq 5.25) \rightarrow (d{=}2)\ \{2,3\}$
rule 4. $(a_3 \geq 5.75) \wedge (a_1 \in [2.75, 5.75)) \rightarrow (d{=}2)\ \{8,10,11\}$
rule 5. $(a_2 \geq 10.5) \wedge (a_1 \in [6.25, 7.0)) \rightarrow (d{=}2)\ \{13\}$
rule 6. $(a_1 \geq 5.75) \wedge (a_2 \in [7.75, 12.0)) \rightarrow (d{=}3)\ \{9, 15\}$
rule 7. $(a_1 \geq 7.0) \rightarrow (d{=}3)\ \{1, 5, 9\}$

Let us stress that this result has been obtained without any prior discretization, the algorithm itself created necessary elementary conditions on numerical attributes. The representation of these conditions is more general, expressive and readable than in the classical approach. Moreover, we induced smaller number of stronger decision rules. The rules are supported by a greatex number of learning examples. Therefore, they are more justified both to explain the decision situation and to give better recommendations.

3. Decision rules for multicriteria sorting problems

Sorting of objects (actions), which consists in their assignment to some predefined and ordered classes, is one of the most frequent multicriteria decision problems (Roy, 1985). Objects are described by a set of criteria and decision classes are preference ordered.

As pointed out by Greco, Matarazzo and Słowiński (1996, 1998c) the original rough set approach does not take into account the preference-ordered scales of both condition and decision attributes and thus cannot handle inconsistencies manifested by violation of the dominance principle. Therefore, it cannot use properly all the essential information contained in the decision table of a multicriteria sorting problem. Greco, Matarazzo and Słowiński (1996) have proposed a new generalization of rough sets theory that is able to deal with ordinal preference information. It is mainly based on the idea of replacing the indiscernibility or similarity relation by the dominance relation, which is a very natural concept within multicriteria decision making. Let us shortly present basic concepts of this generalization. More information can be found in Greco et al. (1996, 1998a, c, 1999a).

Let S_q be an outranking relation (Roy, 1985) on a set U with reference to a 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, that is - a strongly complete and transitive binary relation.

Moreover, let $\mathbf{Cl} = \{Cl_t, t \in T\}$, $T = 1, ..., n$, denote classification on U, such that each $x \in U$ belongs to one and only one class $Cl_t \in \mathbf{Cl}$. As decision categories in multicriteria sorting are ordered, we assume that for all $r, s \in T$, such that $s > r$, each element of Cl_s is preferred to each element of Cl_r. More formally, if S is a comprehensive outranking relation on U, i.e. xSy means: for all $x, y \in U$: "x is at least as good as y", then it is supposed that $[x \in Cl_r, y \in Cl_s, r > s] \Rightarrow [xSy$ and not $ySx]$.

The ordinal classification of objects leads us to considering the following upward and downward cumulated classes (unions), respectively,

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

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

The meaning of cumulated classes is the following: $x \in Cl_t^{\geq}$ means that element x belongs to "at least decision class Cl_t" while $x \in Cl_t^{\leq}$ means that x belongs to "at most decision class Cl_t". Notice that $Cl_1^{\geq} = Cl_n^{\leq} = U$, $Cl_n^{\geq} = Cl_n$ and $Cl_1^{\leq} = Cl_1$.

For example, let us come back to the bankruptcy risk evaluation mentioned in the Introduction. The firms could be classified into three categories of risk: 1-unacceptable, 2-uncertain, 3-acceptable. Of course, category 3 is more preferred than category 2, and category 2 is more preferred than category 1. The definition of cumulated classes is the following: $Cl_1^{\leq}$ means "unacceptable" firms, $Cl_2^{\leq}$ means "at most uncertain", i.e. "uncertain or unacceptable" firms, $Cl_2^{\geq}$ means "at least uncertain" firms, i.e. "uncertain or acceptable", $Cl_3^{\geq}$ means "acceptable" firms.

We say that x dominates over y with respect to a set of criteria $P \subseteq C$ (notation xD_Py) if xS_qy for each $q \in P$.

Let us consider the example from Table 1. We can treat this classification problem as the multicriteria sorting problem, if we assume that a_1, a_2, a_3 are criteria and the decision classes are preference-ordered. For simplicity, we will assume that for all the criteria the direction of preference is increasing, i.e. a high value is more preferred than a small one. Additionally, decision classes 1,2,3 are also preference-ordered according to increasing class number. The cumulated classes are the following:

- $Cl_3^{\geq} = \{1, 5, 9, 15\}$
- $Cl_2^{\geq} = \{1, 2, 3, 5, 8, 9, 10, 11, 13, 15\}$
- $Cl_2^{\leq} = \{2, 3, 4, 6, 7, 8, 10, 11, 12, 13, 14\}$
- $Cl_1^{\leq} = \{4, 6, 7, 12, 14\}$.

One can notice that for instance object no. 1 dominates over object no. 15 as $(f_{a_2}(15) = 11)$ and $(f_{a_3}(1) = 9.5) > (f_{a_3}(15) = 7.5)$. In a similar way we can check that object no. 1 does not dominate over object no. 5 and vice versa. There is no dominance between them because object no. 1 is better than object no. 5 on criteria a_1 and a_3, but worse on criterion a_2. All these objects, no. 1, 5 and 15, belong to the same decision class Cl_3. By checking if the dominance between objects on the considered criteria is consistent with the class assignment, we can discover the following "troublesome" cases:

- object no. 15 (from the best class Cl_3) is dominated by object no. 13 (from the medium class Cl_2),

- objects no. 12 and 14 (from the worst class Cl_1) dominate over object no. 3 (from the medium class Cl_2).

In both cases, the dominance principle is violated since objects having better evaluation on all criteria are assigned to a decision class that is worse than for dominated objects. It is reasonable, of course, to expect the inverse assignment. These cases show inconsistencies of the second type, connected with the dominance principle.

Given a subset of criteria $P \subseteq C$ and an object $x \in U$, let us define two sets characterizing the dominance relation, called: P - *dominating set* and P - *dominated set*, respectively:

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

The P-dominating set groups all objects that dominate over the considered object x, i.e. which are at least as good as this object, according to preference on criteria from P. On the other hand, P-dominated set consists of all objects which are dominated by x. We can define the P-lower and the P-upper approximation of $Cl_t^{\geq}$ with respect to $P \subseteq C$ (notation $\underline{P}Cl_t^{\geq}$ and $\bar{P}Cl_t^{\geq}$, respectively), as:

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

Analogously, we define the P-lower and P-upper approximations of $Cl_t^{\leq}$:

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

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

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

Let us illustrate these concepts using our didactic example. We focus our interest on $Cl_3^{\geq}$. It consists of objects 1, 5, 9, 15.

P-dominating sets for these objects are the following:

$$\begin{aligned} D_P^+(1) &= \{1\} \\ D_P^+(5) &= \{5\} \\ D_P^+(9) &= \{9\} \\ D_P^+(15) &= \{1, 5, 13, 15\}. \end{aligned}$$

One can notice that inside the P-dominating set $D_P^+(15)$ there is also object no. 13 from another union $Cl_2^{\leq}$. Its P-dominating and P-dominated sets are $D_P^+(13) = \{1,5,13\}$ and $D_P^-(13) = \{2,3,4,6,7,8,10,12,13,14,15\}$. Therefore, we discovered that objects no. 13 and 15 axe inconsistent in the sense of the dominance principle.

The upper and lower approximations of the union $Cl_3^{\geq}$ are given below:

$$\begin{aligned} \underline{P}(Cl_3^{\geq}) &= \{1,5,9\} \\ \bar{P}(Cl_3^{\geq}) &= \{1,5,9,13,15\} \end{aligned}$$

The boundary region $Bn_P(Cl_3^{\geq}) = \{13,15\}$.

In the similar way we can calculate upper and lower approximations of union $Cl_2^{\leq}$, i.e.

$$\begin{aligned} \underline{P}Cl_2^{\leq} &= \{2,3,4,6,7,8,10,11,12,14\} \\ \bar{P}Cl_2^{\leq} &= \{2,3,4,6,7,8,10,11,12,13,14,15\} \end{aligned}$$

The boundary region $Bn_P(Cl_2^{\leq}) = \{13,15\}$.

The obtained approximations clearly separate certain part of information about decision assignments to both unions $Cl_3^{\geq}$ and $Cl_2^{\leq}$ from doubtful part that comprises inconsistent examples 13 and 15. Thus, the generalized rough set approach enables to discover the inconsistency referring to dominance principle.

Proceeding in an analogous way, one can discover another group of inconsistencies, i.e. objects no. 3,12 and 14. They belong to the boundary region of unions $Cl_3^{\geq}$ and $Cl_2^{\leq}$.

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.

The following three types of decision rules are considered:

1. decision rules of the type $D_{\geq}$, which have the following form:
   $(q_1 \geq r_{q_1}) \wedge (q_2 \geq r_{q_2}) \wedge \ldots (q_p \geq r_{qp}) \rightarrow (d = Cl_t^{\geq})$,
   where $P = \{q_1, q_2, \ldots, q_p\} \subseteq C$, and $r_{q_1}, r_{q_2}, \ldots, r_{q_p}$ are values from the domains of respective criteria and t is the index of a cumulated decision class $Cl_t^{\geq}$; these rules are supported only by objects from the P-lower approximations of $Cl_t^{\geq}$;
2. decision rules of the type $D_{\leq}$, which have the following form:
   $(q_1 \leq r'_{q_1}) \wedge (q_2 \leq r'_{q_2}) \wedge \ldots (q_p \leq r'_{qp}) \rightarrow (d = Cl_t^{\leq})$,
   where $P = \{q_1, q_2, \ldots, q_p\} \subseteq C$, and $r'_{q_1}, r'_{q_2}, \ldots, r'_{q_p}$ are values from the domains of respective criteria and t is the index of a cumulated class $Cl_t^{\leq}$; these rules are supported only by objects from the P-lower approximations of $Cl_t^{\leq}$;
3. decision rules of the type $D_{\leq\geq}$, which have the following form:
   $(q_1 \geq r_{q_1}) \wedge (q_2 \geq r_{q_2}) \wedge \ldots \wedge (q_k \geq r_{q_k}) \wedge$

$(q_{k+1} \leq r_{q_{k+1}}) \wedge (q_{k+2} \leq r_{q_{k+2}}) \wedge \ldots \wedge (q_p \leq r_{q_p}) \rightarrow (d = Cl_s \cup Cl_{s+1} \cup \ldots \cup Cl_t)$,
where $O' = \{q_1, q_2, \ldots, q_k\} \subseteq C$ and $O'' = \{q_{k+1}, q_{k+2}, \ldots, q_p\} \subseteq C$ are not necessarily disjoint, $r_{q_1}, r_{q_2}, \ldots, r_{q_p}$ are values from the domains of respective criteria, and $s, t \in T$ such that $t > s$ are the indices of cumulated classes; these rules are supported only by objects from the P-boundaries of $Cl_t^{\geq}$ and $Cl_s^{\leq}$.

Two first types of rules correspond to certain part of knowledge extracted from input data table, while the third type is the approximate one as it is built of the doubtful part.

The rule induction technique, called DOMLEM, generalizes the MODLEM algorithm in order to observe the dominance principle (for details of this modification see Greco et al.,1999b). DOMLEM and its archetype MODLEM are characterized by a polynomial computational complexity.

We give below the list of rules induced from Table 1 by the DOMLEM algorithm.

Exact rules:

rule 1. $(a_1 \geq 7.5) \rightarrow (d = Cl_3^{\geq})$ {1,5,9}
rule 2. $(a_3 \geq 6.5) \rightarrow (d = Cl_2^{\geq})$ {1,5,8,9,10,11,13,15}
rule 3. $(a_1 \geq 4) \wedge (a_2 \geq 7.0) \rightarrow (d = Cl_2^{\geq})$ {1,2,5,8,9,11,13,15 }
rule 4. $(a_1 \leq 5.5) \rightarrow (d = Cl_2^{\leq})$ {2,3,4,6,8,10,11,12,14}
rule 5. $(a_2 \leq 6.5) \rightarrow (d = Cl_2^{\leq})$ {3,4,6,7,10,12}
rule 6. $(a_3 \leq 4.5) \wedge (a_2 \leq 6.5) \rightarrow (d = Cl_1^{\leq})$ {4,6,7}

Approximate rules

rule 7. $(a_2 \geq 11) \wedge (a_3 \leq 8.0) \rightarrow (d = Cl_3$ or $Cl_2)$ {15,13,}
rule 8. $(a_3 \geq 5.0) \wedge (a_2 \leq 4.0) \rightarrow (d = Cl_1$ or $Cl_2)$ {3,12}
rule 9. $(a_1 \leq 3.5) \wedge (a_2 \leq 10.0) \rightarrow (d = Cl_1$ or $Cl_2)$ {14}

If the above eight rules were applied to the objects from the data table then objects no. 1,2, 4 -11 would be reclassified exactly to their classes, while objects no. 3,12 and 14 would be classified approximately to classes Cl_1 or Cl_2 and objects no. 13 and 15 to classes Cl_2 or Cl_3.

Let us observe that apart from the semantic difference, the set of decision rules induced from the rough 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 simple indiscernibility relations. The minimal sets of rules thus obtained have a smaller number of rules and use a smaller number of elementary conditions.

4. Conclusions

Handling inconsistency of information is of major importance for decision analysis. We claim that inconsistency should not be considered as a noise or error in data - a proper understanding of its semantics can help in drawing certain (exact) and possible (approximate) conclusions from available information. As to the semantics of inconsistency, we distinguish two kinds: one inconsistency is related to indiscernibility of objects described by regular attributes, and the other inconsistency follows from violation of the dominance principle among objects described by criteria, i.e. attributes with preference-ordered scales. The various versions of the rough sets theory can deal with both kinds of inconsistency. Rough sets theory has been combined with rule induction techniques in order to get exact and approximate decision rules.

We surveyed these combinations, starting from the simplest case where elementary conditions of the rules are of the form *attribute=single value*. It is typical for indiscernibility-based rough sets theory and nominal scales. Then, we considered an extended syntax in the form of *attribute#value*, where # means one of the operators: $=, \leq, \geq, <, >$. It is typical for indiscernibility or similarity-based rough sets theory and ordinal or numerical attribute scales. The above considerations were limited to handling only the first kind of inconsistency.

Finally, we considered criteria and preference-ordered decision classes where the second kind of inconsistency has to be taken into account. The rules induced from dominance-based rough approximations of decision classes have the extended syntax of elementary conditions mentioned above; moreover, their decision parts indicate assignment to "*at least*" or "*at most*", given decision class.

Current research directions concentrate on consideration of regular attributes and criteria in a joint model which handles both kinds of inconsistencies within the rough sets theory and rule induction techniques, Greco et al. (1998a). Yet, another challenging perspectives is connected with taking into account missing values in the data table, Greco et al. (1999d). Other studies try to introduce grades into the rough sets approach using fuzzy extensions, see Greco et al. (1998b, 1999c).

Acknowledgments
The two first authors' work on this paper was supported by DS research grant no. 43-323/49. The research of the second two authors was supported by the Italian Ministry of University and Scientific Research (MURST).

References

BAZAN, J.G. (1998) A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables. In: L. Polkowski, A. Skowron, eds., *Rough Sets in Data Mining and Knowledge Discovery*, vol. 1, Physica-Verlag, 321–365.

CHMIELEWSKI, M.R. and GRZYMALA-BUSSE, J.W. (1995) Global discretization of continuous attributes as preprocessing for machine learning. In: T. Y. Lin and A. M. Wilderberger, eds., *Soft Computing: Rough Sets, Fuzzy*

Logic, Neural Networks, Uncertainty Management, Knowledge Discovery. Simulation Councils, Inc., 294–297.

CLARK, P. and BOSWELL, R. (1991) Rule induction with CN2: some recent improvements. In: Kodratoff Y., ed., *Proc. of 5th European Working Session on Learning - ESWL 91* Porto, Portugal. Springer Verlag, 151–163.

FAYYAD, U.M., PIATETSKY-SHAPIRO, G., SMYTH, P., UTHURUSAMY, R., eds. (1996) *Advances in Knowledge Discovery and Data Mining.* AAAI/MIT Press, Cambridge Mass.

DOUGHERTY, J., KOHAVI, R., SAHAMI, M. (1995) Supervised and unsupervised discretizations of continuous features. *Proc. 12th Int. Conf. on Machine Learning*, 194–202.

GRECO, S., MATARAZZO, B., SŁOWIŃSKI, R. (1996) Rough approximation of a preference relation by dominance relations. ICS Research Report, 16/96 Warsaw University of Technology.

GRECO, S., MATARAZZO, B., SŁOWIŃSKI, R. (1998A) A new rough set approach to multicriteria and multiattribute classification. In: L. Polkowski, A. Skowron, eds., *Proc. of the First Int. Conf. on Rough Sets and Current Trends in Computing.* Springer Verlag LNCS 1424, Berlin, 60–67.

GRECO, S., MATARAZZO, B., SŁOWIŃSKI, R. (1998B) Fuzzy similarity relation as a basis for rough approximations. In: L. Polkowski, A. Skowron, eds., *Proc. of the First Int. Conf. on Rough Sets and Current Trends in Computing.* Springer LNAI 1424, 283-289.

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

GRECO, S., MATARAZZO, B., SŁOWIŃSKI, R. (1999A) The use of rough sets and fuzzy sets in MCDM. In: T. Gal, T. Hanne, T. Stewart, eds., *Multicriteria Decision Making. Advances in MCD Models, Algorithms, theory and applications*, 14.1-14.59.

GRECO, S., MATARAZZO, B., SŁOWIŃSKI, R., STEFANOWSKI, J. (1999B) Induction of decision rules for solving multicriteria sorting problems. Working Paper 1999 (presented at EURO XVI Brussels 1998).

GRECO, S., MATARAZZO, B., SŁOWIŃSKI, R. (1999C) Fuzzy dominance as basis for rough approximations. In: *Proc. 4th Meeting of the EURO WG on Fuzzy Sets and 2nd Internat. Conf. on Soft and Intelligent Computing*, EUROFUSE-SIC'99, Budapest, May 25-28, 273-278.

GRECO, S., MATARAZZO, B., SŁOWIŃSKI, R. (1999D) Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems. In: A. Skowron and N. Zhong, eds., *Proc. 7th Workshop New Directions in Rough Sets, Data Mining and Granular-Soft Computing* (Yamaguchi, Japan). LNAI 1711, Springer-Verlag, Berlin, 146-157.

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

GRZYMALA-BUSSE, J.W., STEFANOWSKI, J., ZIARKO, W. (1996) Rough sets: Facts versus misconceptions. *Informatica*, **20**, 4, 455-464.

GRZYMALA-BUSSE, J.W., STEFANOWSKI, J. (1999) Two Approaches to Numerical Attribute Discretization for Rule Induction. In: *Proceedings of the 5th Int. Conference of the Decision Sciences Institute*, Athens, Greece, July 4-7, 1377–1379 .

OHRN, A., KOMOROWSKI, J., SKOWRON, A., SYNAK, P. (1997) Rosetta - Part I: System Overview, Technical report. Dept. of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

KOMOROWSKI, J., PAWLAK, Z., POLKOWSKI, L., SKOWRON, A. (1999) Rough Sets: tutorial. In: S.K. Pal, A. Skowron, eds., *Rough Fuzzy Hybridization. A new trend in decision making.* Springer Verlag, Singapore, 3–98.

KRAWIEC, K., SŁOWIŃSKI, R., VANDERPOOTEN, D. (1998) Learning of decision rules from similarity based rough approximations. In: A. Skowron, L. Polkowski, eds., *Rough Sets in Knowledge Discovery* vol. II. Physica Verlag, Heidelberg, 37-54.

KRYSZKIEWICZ, M. (1998A) Strong Rules in Large Databases. In: *Proceedings of Int. Conf. on IPMU 98 - Information Processing and Management of Uncertainty in Knowledge-based Systems*, Paris, France 6-10 July 1998, Vol. 2, 1520-1527.

KRYSZKIEWICZ, M., RYBINSKI, H. (1998B) Knowledge Discovery from Large Databases using Rough Set. In: *Proceedings of 6th European Congress on Intelligent Techniques and Soft Computing* EUFIT ?98, Aachen, Germany, 7-10 September 1998, Vol. 1, 85-89

JACQUET-LAGREZE, E., SISKOS, J. (1982) Assesing a set of additive utility functions for multicriteria decision making, the UTA method. *European Journal of Operational Research*, **10**, 151-164.

LANGLEY, P., SIMON, A. (1998) Fielded applications of machine learning. In: R.S. Michalski, I. Bratko, M. Kubat, eds., *Machine Learning and Data Mining.* Wiley, New York, 113-129.

LENARCIK, A., PIASTA, Z. (1997) Probabilistic rough classifiers with mixture of discrete and continuous attributes. In: T.Y. Lin, N. Cecrone, eds., *Rough Sets and Data Mining.* Kluwer Academic Publishers, 373-384.

LIN, T.Y. (1996) Rough set theory in very large databases. In: *Proceedings of CESA'96*, vol. 2, Lille, July, 936-941.

MARCH, J.G., (1988) Bounded rationality, ambiguity and the engineering of choice. In: D.E. Bell, H. Raiffa, A. Tversky, eds., *Decision Making. Descriptive, Normative and Presciptive Interactions.* Cambridge University Press, New York, 33–58.

MARCUS, S. (1994) Tolerance rough sets. Cech topologies, learning process.

Bull. of the Polish Academy of Sciences, Tech. Sciences, **42**, 3, 471–487.

MICHALSKI, R.S., BRATKO, I., KUBAT, M., eds. (1998) *Machine Learning and Data Mining.* John Wiley & Sons.

MITCHELL, T. (1997) *Machine Learning.* Mac-Graw Hill, Boston.

NGUYEN, HUNG SON (1998A) Discretization problem for rough set methods. In: L. Polkowski, A. Skowron, eds., *Proc. of the First Int. Conf. on Rough Sets and Current Trends in Computing.* Springer Verlag, LNAI 1424, 545–552.

NGUYEN, HUNG SON, HOA NGUYEN (1998B) Discretization Methods in Data Mining. In: L. Polkowski, A. Skowron, eds., *Rough Sets in Data Mining and Knowledge Discovery*, vol. 1. Physica-Verlag, 451 – 482.

PAWLAK, Z. (1982) Rough sets. *Int. J. Computer and Information Sci.*, 11, 1982, 341–356.

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

PAWLAK, Z., GRZYMALA-BUSSE, J., SŁOWIŃSKI, R., ZIARKO, W. (1995) Rough sets. *Communications of the ACM*, vol. **38**, 11, 89-95.

POLKOWSKI, L., SKOWRON, A., ZYTKOW, J. (1995) Tolerance based rough sets. In: T.Y. Lin, A.M. Wildberg, eds., *Soft Computing: Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery.* Simulation Councils, Inc., 55–58.

PREDKI, B., WILK, SZ. (1999) ROSE - software implementation of the rough set theory. Proc. of the 11th Int. Symp. on Methodologies for Intelligent Systems, ISMIS-99, Warsaw, June 1999. LNAI Springer, 605-608.

ROY, B. (1985) *Multicriteria Methodology for Decision Aiding.* Kluwer Academic Publishers, Dordrecht, 1996 (French edition 1985).

SHOEMAKER, P.J.H. (1982) The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature*, **20**, 529–562.

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

SKOWRON, A. (1995) Extracting Laws from Decision Tables. *Int. Journal Computational Intelligence*, **11**, 2, 371-388.

SKOWRON, A., POLKOWSKI, L. (1997) Synthesis of decision systems from data tables. In: T.Y. Lin, N. Cecrone, eds., *Rough Sets and Data Mining.* Kluwer Academic Publishers, 289-299.

SKOWRON, A., RAUSZER, C. (1992) The discernibility matrices and functions in information systems. In: R. Słowiński, eds., *Intelligent Decision Support. Handbook of Applications and Advances of Rough Set Theory.* Kluwer Academic Publishers, Dordrecht, 331–362 .

SKOWRON, A., STEPANIUK, J. (1996) Tolerance approximation spaces. *Fundamenta Informaticae*, **27**, 2-3, 245-253.

SŁOWIŃSKI, R. (1995) Rough set approach to decision analysis. *AI Expert*, March, 19-25.

SŁOWIŃSKI, R., STEFANOWSKI, J. (1998) Rough Family - software implementation of the rough set theory. In: A. Skowron, L. Polkowski, eds., *Rough Sets in Knowledge Discovery*. Vol. 2. Physica Verlag, Heidelberg, 581-587.

SŁOWIŃSKI, R., VANDERPOOTEN, D. (1997) Similarity relation as a basis for rough approximations. In: P. Wang, ed., *Advances in Machine Intelligence and Soft Computing*, vol. IV. Duke University Press, 17-33.

SŁOWIŃSKI, R., VANDERPOOTEN, D. (1999) A generalized definition of rough approximations based on similarity. *IEEE Transactions on Data and Knowledge Engineering* (to appear).

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

STEFANOWSKI, J. (1998B) Handling continuous attributes in discovery of strong decision rules. In: L. Polkowski, A. Skowron, eds., *Proc. of the First Int. Conf. on Rough Sets and Current Trends in Computing*. Springer Verlag, LNAI 1424, 394-402.

STEFANOWSKI, J. (1998C) Rough set based rule induction techniques for classification problems. In: *Proc. 6th European Congress on Intelligent Techniques and Soft Computing*, vol. 1, Aachen Sept. 7-10, 109-113.

STEFANOWSKI, J., VANDERPOOTEN, D. (1994) A general two stage approach to rule induction from examples. In: W. Ziarko, ed., *Rough Sets, Fuzzy Sets and Knowledge Discovery*. Springer Verlag, 317-325.

STEFANOWSKI, J., VANDERPOOTEN, D. (2000) Induction of decision rules in classification and discovery-oriented perspectives. *The Journal of Intelligent Information Systems* (to appear).

SUSMAGA, R. (1997) Analyzing discretization of continuous attributes given a monotonic discrimination function. *Journal Intelligent Data Analysis*, 1, 3 (Elsevier on line journal http://www-east.elsevie.com/ida).

WEISS, S.M., KULIKOWSKI, C.A. (1991) *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural nets, Machine Learning and Expret Systems*. Morgan Kaufmann Publishers.

ZIARKO, W. (1993) Variable precision rough sets model. *Journal of Computer and Systems Sciences*, 46, 1, 1993, 39-59.

ZIARKO, W. and SHAN, N. (1995) KDD-R: A comprehensive system for knowledge discovery in databases using rough sets. In: T.Y. Lin, A.M. Wildberg, eds., *Soft Computing. Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery*. Simulation Council Inc., San Diego, 93-96.