Robust Ordinal Regression for Dominance-based Rough Set Approach Under Uncertainty

Roman Słowiński[1,2], Miłosz Kadziński[1], and Salvatore Greco[3,4]

[1] Institute of Computing Science, Poznań University of Technology, Poznań, Poland
[2] Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
[3] Department of Business and Economics, University of Catania, Catania, Italy
[4] Portsmouth Business School, University of Portsmouth, United Kingdom

Abstract. We consider decision under uncertainty where preference information provided by a Decision Maker (DM) is a classification of some reference acts, relatively well-known to the DM, described by outcomes to be gained with given probabilities. We structure the classification data using a variant of the Dominance-based Rough Set Approach. Then, we induce from this data all possible minimal-cover sets of rules which correspond to all instances of the preference model compatible with the input preference information. We apply these instances on a set of unseen acts, and draw robust conclusions about their quality using the Robust Ordinal Regression paradigm. Specifically, for each act we derive the necessary and possible assignments specifying the range of classes to which the act is assigned by all or at least one compatible set of rules, respectively, as well as class acceptability indices. The whole approach is illustrated by a didactic example.

1 Introduction

Decision under uncertainty is a classical topic of decision theory (see [3] for a review). The main approaches to modeling this decision are based on expected utility theory. Since many experiments showed systematic violation of expected utility hypotheses, alternative models weakening some original axioms have been proposed. In particular, Greco, Matarazzo and Słowiński [5] used an approach based on stochastic dominance, which is the weakest assumption possible. They applied for this Dominance-based Rough Set Approach (DRSA) [4], which adapts the rough set concept introduced by Pawlak [8] to handle ambiguity with respect to dominance.

The main difficulty in recommending a decision under uncertainty for some Decision Maker (DM) relies on such an aggregation of outcomes predicted for the considered acts with given probabilities, that an aggregated model, called preference model, indicates the act that best respects preferences of this DM. The preference model used in [5] has the form of a set of "if..., then..." decision rules. The input preference information to this model is a set of classification examples concerning some acts relatively well known to the DM, called reference acts. The decision rules are induced from these classification examples, however, for the reason of possible ambiguity in the set of examples caused by DM's violation of the stochastic dominance principle, the classification data are structured first using DRSA. DRSA permits distinguishing non-ambiguous from ambiguous classification examples, which corresponds to the lower approximation and to the boundary of the classification, respectively. In consequence, decision rules induced from data structured in this way are certain or ambiguous, respectively.

The most popular rule induction strategy consists in looking for a minimal set of rules covering the classification examples [1]. This strategy is called minimal-cover (MC) strategy. In practice, it is performed by a greedy heuristic of sequential covering type, giving an approximately-MC set of rules (later, for simplicity, we drop the prefix "approximately" from MC). However, representation of the DM's preferences with a MC set of rules is not unique, because there are in general many MC sets of rules for a given rough approximation of classification data, that reproduce equally well the provided preference information. Choosing among them is either arbitrary or requires involvement of the DM, which is not easy for most of them. This inconvenience has been already noticed in Multiple Criteria Decision Aiding (MCDA) and resulted in proposing Robust Ordinal Regression (ROR) ([6]; for a survey see [2]; for application of ROR to DRSA see [7]), which takes into account all preference model instances compatible with the DM's preference information.

The aim of this paper is to adapt ROR, involving MC sets of rules as compatible instances of the preference model, to decision under uncertainty formulated as a multi-attribute classification problem. In order to generate all MC sets of rules, first an exhaustive set of rules is induced, and then the MC sets are constructed by solving a series of Integer Linear Programming (ILP) problems.

Note that although compatible MC sets of rules reproduce the classification examples provided by the DM, the assignment of non-reference acts that result from application of any of these MC sets of rules can vary significantly. We investigate the diversity of the recommendations suggested by different MC sets of rules by producing two types of assignment, necessary and possible, for each act from the considered set of acts. Since all compatible MC sets of rules are known, we are also able to compute class acceptability indices defined as the share of compatible MC sets of rules assigning an act to a single class or a set of contiguous classes.

The paper is organized as follows. Section 2 recalls basics of DRSA for decision under uncertainty. Then, we present algorithms for generating all compatible rules and all compatible MC sets of rules in Sections 3 and 4, respectively. In Section 5, we show how to get the possible and the necessary assignments. To illustrate the whole approach, throughout the paper we refer to a didactic example of decision under uncertainty. Section 6 contains conclusions.

2 DRSA for decision under uncertainty

Following [5], we formulate the decision under uncertainty as a multi-attribute classification problem. For this, we consider the following basic elements:

- a finite set S=$\{s_1,\ s_2,\ \ldots,\ s_u\}$ of states of the world, or simply *states*, which are mutually exclusive and exhaustive,
- an *a priori probability distribution* P over S: more precisely; the probabilities of states $s_1,\ s_2,\ \ldots,\ s_u$ are given by $p_1,\ p_2,\ \ldots,\ p_u$, respectively, $p_1 + p_2 + \ldots + p_u = 1$, $p_i \geq 0$, $i = 1, \ldots, u$,
- a set A=$\{a_1,\ a_2,\ \ldots,\ a_m\}$ of all considered *acts*, and a set $A^R \subset A$ of *reference acts* for which the DM expressed her/his preferences,
- a set X=$\{x_1,\ x_2,\ \ldots,\ x_r\}$ of *outcomes* that, for the sake of simplicity, we consider expressed in monetary terms ($X \subseteq \Re$),
- a function $g : A \times S \to X$ assigning to each pair act-state $(a_i,s_j) \in A \times S$ an outcome $x_k \in X$,
- a set of *quality classes* $\boldsymbol{Cl}$=$\{Cl_1,\ Cl_2,\ \ldots,\ Cl_n\}$, such that $Cl_1 \cup Cl_2 \cup \ldots \cup Cl_n = A$, $Cl_r \cap Cl_q = \emptyset$ for each $r,q \in \{1,\ldots,n\}$ with $r \neq q$; the classes from $\boldsymbol{Cl}$ are preference-ordered according to the increasing order of their indices,
- a function e: $A \to \boldsymbol{Cl}$ assigning each act $a_i \in A$ to a quality class $Cl_j \in \boldsymbol{Cl}$.

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

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

Let us define the following functions z: $A \times S \to \Pi$ and z': $A \times S \to \Pi$ assigning to each act-state pair $(a_i,s_j) \in A \times S$ a probability $\pi \in \Pi$, as follows:

$$z(a_i, s_j) = \sum_{r:g(a_i,s_r) \geq g(a_i,s_j)} p_r \quad \text{and} \quad z'(a_i, s_j) = \sum_{r:g(a_i,s_r) \leq g(a_i,s_j)} p_r.$$

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

On the basis of functions $z(a_i,s_j)$ and $z'(a_i,s_j)$, we can define functions, respectively, ρ: $A \times \Pi \to X$ and ρ': $A \times \Pi \to X$ as follows:

$$\rho(a_i, \pi) = \max_{j:z(a_i,s_j) \geq \pi} \{g\,(a_i, s_j)\} \quad \text{and} \quad \rho'(a_i, \pi) = \min_{j:z'(a_i,s_j) \geq \pi} \{g\,(a_i, s_j)\}\,.$$

Thus, $\rho(a_i,\pi) = x$ ($\rho'(a_i,\pi) = x$) means that the outcome got by act a_i is greater (smaller) than or equal to x with a probability *at least* π. As observed in [5], information given by $\rho(a_i,\pi)$ and $\rho'(a_i,\pi)$ is related such that for all $a_i \in A$ and $\pi_{(j-1)}, \pi_{(j)} \in \Pi$:

$$\rho(a_i, \pi_{(j)}) = \rho'(a_i, 1 - \pi_{(j-1)}). \quad (1)$$

The above considerations lead us to formulation of the decision under uncertainty in terms of a multi-attribute classification problem, where the set of classified objects is the set of acts A, the set of condition attributes describing the acts is the set Π, $\boldsymbol{cl}$ denotes the decision attribute assigning the acts from A to classes from $\boldsymbol{Cl}$, the set X is a value set of condition attributes, the set $\boldsymbol{Cl}$ is the value set of the decision attribute, and $\boldsymbol{f}$ is an information function, such that $\boldsymbol{f}(a_i,\ \pi) = \rho(a_i,\ \pi)$ and $\boldsymbol{f}(a_i,\ \boldsymbol{cl})$=$e(a_i)$. Let us observe that due to the above stated equivalence, one can consider alternatively information function $\boldsymbol{f}'(a_i,\ \pi) = \rho'(a_i,\ \pi)$.

The classification examples concerning the reference acts A^R constitute the DM's preference information considered in the context of decision under uncertainty. Formally, they are presented as an *information table* whose rows correspond to reference acts belonging to set A^R, and columns correspond to condition attributes from set Π and to the decision attribute $\boldsymbol{cl}$. The entries of the information table are values $\rho(a_i,\ \pi)$ of the information function $\boldsymbol{f}$, as well as class assignments of the acts.

These classification data are structured using DRSA. In DRSA, we are approximating the upward $Cl_t^{\geq} = \bigcup\limits_{s \geq t} Cl_s$ and downward $Cl_t^{\leq} = \bigcup\limits_{s \leq t} Cl_s$, unions of classes, $t = 1, \ldots, n$, using *dominance cones* defined in the condition attribute space for any subset of condition attributes $\Theta \subseteq \Pi$. The fact that act a_p stochastically dominates act a_q with respect to $\Theta \subseteq \Pi$ (i.e., $\rho(a_p, \pi) \geq \rho(a_q, \pi)$ for each $\pi \in \Theta$) is denoted by $xD_\Theta y$. Given $\Theta \subseteq \Pi$ and $a_i \in A^R$, the cones of dominance are:

- a set of acts dominating a_i: $D_\Theta^+(a_i) = \{a_j \in A^R : a_j D_\Theta a_i\}$,
- a set of objects dominated by a_i: $D_\Theta^-(a_i) = \{a_j \in A^R : a_i D_\Theta a_j\}$.

With respect to $\Theta \subseteq \Pi$, the set of all acts belonging to $Cl_t^{\geq}$ ($Cl_t^{\leq}$) without any ambiguity constitutes the Θ-*lower approximation* of $Cl_t^{\geq}$ ($Cl_t^{\leq}$), denoted by $\underline{\Theta}(Cl_t^{\geq})$ ($\underline{\Theta}(Cl_t^{\leq})$), and the set of all acts that could belong to $Cl_t^{\geq}$ ($Cl_t^{\leq}$) constitutes the Θ-*upper approximation* of $Cl_t^{\geq}$ ($Cl_t^{\leq}$), denoted by $\overline{\Theta}(Cl_t^{\geq})$ ($\overline{\Theta}(Cl_t^{\leq})$), i.e., for $t = 1, \ldots, n$:

$$\underline{\Theta}(Cl_t^{\geq}) = \{a \in A^R : D_\Theta^+(a) \subseteq Cl_t^{\geq}\} \text{ and } \underline{\Theta}(Cl_t^{\leq}) = \{a \in A^R : D_\Theta^-(a) \subseteq Cl_t^{\leq}\},$$
$$\overline{\Theta}(Cl_t^{\geq}) = \{a \in A^R : D_\Theta^-(a) \cap Cl_t^{\geq} \neq \emptyset\} \text{ and } \overline{\Theta}(Cl_t^{\leq}) = \{a \in A^R : D_\Theta^+(a) \cap Cl_t^{\leq} \neq \emptyset\}.$$

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

$$Bn_\Theta(Cl_t^{\geq}) = \overline{\Theta}(Cl_t^{\geq}) - \underline{\Theta}(Cl_t^{\geq}) \text{ and } Bn_\Theta(Cl_t^{\leq}) = \overline{\Theta}(Cl_t^{\leq}) - \underline{\Theta}(Cl_t^{\leq}).$$

For every $t \in \{1, \ldots, n\}$ and for every $\Theta \subseteq \Pi$, we define the *quality of approximation of classification* $\boldsymbol{Cl}$ by set of attributes Θ, or in short, *quality of classification*:

$$\gamma_\Theta(\boldsymbol{Cl}) = \frac{card\left(A - \bigcup_{t=1}^{n-1} Bn_\Theta(Cl_t^{\leq})\right)}{card(A)} = \frac{card\left(A - \bigcup_{t=2}^{n} Bn_\Theta(Cl_t^{\geq})\right)}{card(A)}.$$

Each minimal (with respect to inclusion) subset $\Theta \subseteq \Pi$ for which $\gamma_\Theta(\boldsymbol{Cl}) = \gamma_\Pi(\boldsymbol{Cl})$, is called a *reduct* of $\boldsymbol{Cl}$ and denoted by $RED_{\boldsymbol{Cl}}$. The intersection of all reducts is called the *core* and denoted by $CORE_{\boldsymbol{Cl}}$.

Illustrative study: part 1. The following example illustrates the approach. Let us consider:

- a set $S = \{s_1, s_2, s_3\}$ of states of the world,
- an *a priori* probability distribution P over S defined as follows: $p_1 = 0.20$, $p_2 = 0.35$, $p_3 = 0.45$,
- a set $A = \{a_1, \ldots, a_{12}\}$ of acts, and a set of reference acts $A^R = \{a_1, \ldots, a_6\}$;
- a set $X = \{0, 10, 15, 20, 25, 30\}$ of possible consequences,
- a set of classes $\boldsymbol{Cl} = \{Cl_1, Cl_2, Cl_3\}$, where Cl_1 is the set of bad acts, Cl_2 is the set of medium acts, and Cl_3 is the set of good acts,
- a function $g : A \times S \to X$ assigning to each act-state pair $(a_i, s_j) \in A \times S$ a consequence $x_h \in X$ and a function $e : A^R \to \boldsymbol{Cl}$ assigning each reference act $a_i \in A^R$ to a class $Cl_j \in \boldsymbol{Cl}$, presented in Table 1a).

Table 1. a) Acts, consequences, and assignment to class from $\boldsymbol{Cl}$ by the DM. b) Acts, values of function $\rho(a_i, \pi)$ and assignment to class from $\boldsymbol{Cl}$.

Part a)

	s_1	s_2	s_3	$\boldsymbol{cl}$
p_j	0.25	0.35	0.40	
a_1	30	15	10	medium (2)
a_2	10	20	30	good (3)
a_3	15	0	20	bad (1)
a_4	0	15	20	bad (1)
a_5	20	10	30	medium (2)
a_6	0	10	30	good (3)
a_7	30	10	10	-
a_8	10	10	20	-
a_9	30	10	30	-
a_{10}	0	15	25	-
a_{11}	30	10	25	-
a_{12}	0	20	15	-

Part b)

$\rho(a_i, \pi)$	0.20	0.35	0.45	0.55	0.65	0.80	1.00	$\boldsymbol{cl}$
a_1	30	15	15	15	10	10	10	2
a_2	30	30	30	20	20	20	10	3
a_3	20	20	20	15	15	0	0	1
a_4	20	20	20	15	15	15	0	1
a_5	30	30	30	20	20	10	10	3
a_6	30	30	30	10	10	10	0	2
a_7	30	10	10	10	10	10	10	-
a_8	20	20	20	10	10	10	10	-
a_9	30	30	30	30	30	10	10	-
a_{10}	25	25	25	15	15	15	0	-
a_{11}	30	25	25	25	10	10	10	-
a_{12}	20	20	15	15	15	15	0	-

Table 1b) shows values of function $\rho(a_i, \pi)$. This information table is consistent. Indeed, there is no single reference act that would stochastically dominate some other reference act assigned to a better class. In consequence, lower and upper approximations of upward and downward unions of classes are equal to:

$$\underline{\Pi}(Cl_3^{\geq}) = \overline{\Pi}(Cl_3^{\geq}) = \{a_2, a_5\},\ \underline{\Pi}(Cl_2^{\geq}) = \overline{\Pi}(Cl_2^{\geq}) = \{a_1, a_6, a_2, a_5\},$$
$$\underline{\Pi}(Cl_1^{\leq}) = \overline{\Pi}(Cl_1^{\leq}) = \{a_3, a_4\},\ \underline{\Pi}(Cl_2^{\leq}) = \overline{\Pi}(Cl_2^{\leq}) = \{a_1, a_6, a_3, a_4\}.$$

The quality of approximation $\gamma(\boldsymbol{Cl}) = 1$. Moreover, there are four reducts of condition attributes (criteria) ensuring the same quality of sorting as the whole set Π of probabilities: $RED^1_{\boldsymbol{Cl}} = \{0.45, 1.00\}$, $RED^2_{\boldsymbol{Cl}} = \{0.35, 1.00\}$, $RED^3_{\boldsymbol{Cl}} = \{0.20, 0.65\}$, and $RED^4_{\boldsymbol{Cl}} = \{0.20, 0.55\}$. This means that we can explain the preferences of the DM using the probabilities from each $RED_{\boldsymbol{Cl}}$ only. The core is empty, i.e., each probability value can be removed individually from Π without deteriorating the quality of classification.

3 Generating all compatible minimal rules

The dominance-based rough approximations of upward and downward unions of classes serve to induce a generalized description of acts contained in the information table in terms of "*if ..., then ...*" decision rules. In what follows, we focus on certain decision rules. For a given upward or downward union of classes, $Cl_t^{\geq}$ or $Cl_s^{\leq}$, the decision rules induced under a hypothesis that reference acts belonging to $\underline{\Pi}(Cl_t^{\geq})$ or $\underline{\Pi}(Cl_s^{\leq})$ are *positive* and all the other reference acts *negative*, suggest a *certain* assignment to "at least class Cl_t" or to "at most class Cl_s", respectively. The syntax of decision rules obtained from DRSA is the following:

– $\text{D}_{\geq}$-*decision rules*:
if $\rho(a,\pi_{h1}) \geq x_{h1}$ *and, ..., and* $\rho(a,\pi_{hz}) \geq x_{hz}$, *then* $a \in Cl_r^{\geq}$
(i.e. "if by act a the outcome is at least x_{h1} with probability at least π_{h1}, and, ..., and the outcome is at least x_{hz} with probability at least π_{hz}, then $a \in Cl_r^{\geq}$"), where $\pi_{h1}, \ldots, \pi_{hz} \in \Pi$, $x_{h1}, \ldots, x_{hz} \in X$, and $r \in \{2, \ldots, n\}$;
– $\text{D}_{\leq}$-*decision rules*:
if $\rho'(a,p_{h1}) \leq x_{h1}$ *and, ..., and* $\rho'(a,p_{hz}) \leq x_{hz}$, *then* $a \in Cl_r^{\leq}$
(i.e. "if by act a the outcome is at most x_{h1} with probability at least π_{h1}, and, ..., and the outcome is at most x_{hz} with probability at least π_{hz}, then $a \in Cl_r^{\leq}$"), where $\pi_{h1}, \ldots, \pi_{hz} \in \Pi$, $x_{h1}, \ldots, x_{hz} \in X$, and $r \in \{1, \ldots, n-1\}$.

In the following, we discuss an algorithm which generates all certain decision rules for $\underline{\Pi}(Cl_t^{\geq})$, $t = 2, \ldots, n$. The algorithm for $\underline{\Pi}(Cl_t^{\leq})$, $t = 1, \ldots, n-1$, can be formulated analogously.

In the first phase, we generate a set of elementary conditions C_1 to be used in the construction of decision rules. If there exists an act $a_i \in \underline{\Pi}(Cl_t^{\geq})$, such that $\rho(a_i, \pi_{h1}) = x_{h1}$ and $\rho(a_i, \pi_{h2}) = x_{h2}$ and ... $\rho(a_i, \pi_{hz}) = x_{hz}$, then a_i is called basis of the rule. Each $D_{\geq}$-decision rule having a basis is called robust because it is "built" on an existing reference act. Although algorithms which construct robust rules require less computational effort than algorithms which construct non-robust rules, they usually generate a greater number of less general rules. Thus, in this paper, we rather focus on "mix of conditions" rules, which are possibly founded by multiple reference acts. For this purpose, C_1 needs to be composed of conditions in form $\rho(a,\pi_h) \geq x_h$, such that there exists $a_i \in \underline{\Pi}(Cl_t^{\geq}) : \rho(a_i, \pi_h) = x_h$.

In the second phase, we generate a set of conjunctions of elementary conditions which cover at least one reference act in $\underline{\Pi}(Cl_t^{\geq})$. It is an iterative process in which conjunctions of size $k+1$ are constructed from conjunctions of size k (i.e., first, a set C_2 of conjunctions of size 2 is constructed by combining elementary conditions from C_1; then, a set C_3 of conjunctions of size 3 is built from conjunctions of size 2, etc.). This procedure is repeated as long as it is possible to obtain conjunctions of a particular size. Precisely, each conjunction of size $k+1$ is obtained by merging a pair of conjunctions of size k which contain the same $k-1$ conditions, thus, differing by just a single elementary condition.

These differentiating conditions need to concern different criteria. At each stage, we neglect conjunctions of size k with negative support equal to 0, since they already contain all conditions necessary to discriminate reference acts in $\underline{\Pi}(Cl_t^{\geq})$ and $A^R \setminus \underline{\Pi}(Cl_t^{\geq})$. Moreover, the set C_{k+1} of conjunctions of size $k+1$ contains only these conjunctions whose positive support is greater than 0.

After generating all possible conjunctions of elementary conditions covering at least one reference act in $\underline{\Pi}(Cl_t^{\geq})$, we eliminate conjunctions covering any negative example in $A^R \setminus \underline{\Pi}(Cl_t^{\geq})$. Subsequently, we remove conjunctions of conditions which are not minimal, i.e. such that there exists some other conjunction:

- using a subset of elementary conditions or/and weaker elementary conditions,
- requiring in all elementary conditions the same cumulated outcome with less probability; for example, when considering two rules, r1 $\equiv$ if $\rho(a_i, 0.55) \geq 20$ then $a \in Cl_3^{\geq}$, and r2 $\equiv$ if $\rho(a_i, 0.8) \geq 20$ then $a \in Cl_3^{\geq}$, r1 is minimal among them, because it requires a cumulated outcome to be at least 20, but with less probability, 0.55 against 0.8. This allows removing the rules which are not minimal in the specific context of the DRSA analysis using stochastic dominance.

Thus filtered, the remained conjunctions are used to construct the rules with a decision part: $a \in Cl_t^{\geq}$. Let us denote by $\mathcal{R}_{all}^{\underline{\Pi}(Cl_t^{\geq})}$ ($\mathcal{R}_{all}^{\underline{\Pi}(Cl_t^{\leq})}$) the set of all compatible minimal rules induced from $\underline{\Pi}(Cl_t^{\geq})$ ($\underline{\Pi}(Cl_t^{\leq})$).

Table 2. All compatible certain minimal rules

Symbol	Rule	Support
$r^1_{\geq 2}$	if $\rho(a_i, 1.00) \geq 10$ then $a \in Cl_2^{\geq}$	$\{a_1, a_2, a_5\}$
$r^2_{\geq 2}$	if $\rho(a_i, 0.55) \geq 20$ then $a \in Cl_2^{\geq}$	$\{a_2, a_5\}$
$r^3_{\geq 2}$	if $\rho(a_i, 0.20) \geq 30$ then $a \in Cl_2^{\geq}$	$\{a_1.a_2, a_5, a_6\}$
$r^1_{\geq 3}$	if $\rho(a_i, 0.55) \geq 20$ then $a \in Cl_3^{\geq}$	$\{a_2, a_5\}$
$r^2_{\geq 3}$	if $\rho(a_i, 0.35) \geq 30$ and $\rho(a_i, 1.00) \geq 10$ then $a \in Cl_3^{\geq}$	$\{a_2, a_5\}$
$r^1_{\leq 1}$	if $\rho(a_i, 0.80) \leq 0$ ($\rho'(a_i, 0.35) \leq 0$) then $a \in Cl_1^{\leq}$	$\{a_3\}$
$r^2_{\leq 1}$	if $\rho(a_i, 0.20) \leq 20$ ($\rho'(a_i, 1.0) \leq 20$) then $a \in Cl_1^{\leq}$	$\{a_3, a_4\}$
$r^3_{\leq 1}$	if $\rho(a_i, 0.45) \leq 20$ ($\rho'(a_i, 0.65) \leq 20$) and $\rho(a_i, 1.0) \leq 0$ ($\rho'(a_i, 0.20) \leq 0$) then $a \in Cl_1^{\leq}$	$\{a_3, a_4\}$
$r^1_{\leq 2}$	if $\rho(a_i, 1.00) \leq 0$ ($\rho'(a_i, 0.20) \leq 0$) then $a \in Cl_2^{\leq}$	$\{a_3, a_4, a_6\}$
$r^2_{\leq 2}$	if $\rho(a_i, 0.65) \leq 15$ ($\rho'(a_i, 0.45) \leq 15$) then $a \in Cl_2^{\leq}$	$\{a_1, a_3, a_4, a_6\}$
$r^3_{\leq 2}$	if $\rho(a_i, 0.45) \leq 20$ ($\rho'(a_i, 0.65) \leq 20$) then $a \in Cl_2^{\leq}$	$\{a_1, a_3, a_4\}$

Illustrative study: part 2. A set of all minimal decision rules describing the DM's preferences is provided in Table 2. Minimal decision rules constitute the most concise and non-redundant representation of knowledge contained in Tables 1a) and b). There are 11 certain rules overall (5 and 6 rules for the upward and downward class unions, respectively). When it comes to the number of elementary conditions, there are 9 rules with just a single condition and 2 rules with two ones.

4 Generating all compatible minimal-cover sets of rules

A set of certain decision rules is minimal-cover if and only if it is complete, i.e., it is able to cover all reference acts and non-redundant, i.e., exclusion of any rule from this set makes it non-complete. Finding a minimal set of rules covering the reference acts in $\underline{\Pi}(Cl_t^{\geq})$ ($\underline{\Pi}(Cl_t^{\leq})$) is analogous to solving the minimum set cover problem. This classical problem in combinatorics and computer science can be solved as an Integer Linear Programming (ILP) problem (without loss of generality, we focus on $\underline{\Pi}(Cl_t^{\geq})$):

$$Minimize: \ f_k = \sum_{r_k \in \mathcal{R}_{all}^{\underline{\Pi}(Cl_t^{\geq})}} v_k, \tag{2}$$

s.t.
$$\left.\begin{array}{l} \sum_{r_k \in \mathcal{R}_{all}^{\underline{\Pi}(Cl_t^{\geq})} \text{ covering } a_i} v_k \geq 1, \ \text{for } a_i \in A^R, \\ v_k \in \{0,1\}, \ \text{for } r_k \in \mathcal{R}_{all}^{\underline{\Pi}(Cl_t^{\geq})}. \end{array}\right\}$$

Thus, v_k is a binary variable associated with each rule $r_k \in \mathcal{R}_{all}^{\underline{\Pi}(Cl_t^{\geq})}$. If $v_k = 1$, r_k is used in the set of rules covering all reference acts in $\underline{\Pi}(Cl_t^{\geq})$. The optimal solution of the above ILP indicates one of the MC sets of rules:

$$\mathcal{R}_k^{\underline{\Pi}(Cl_t^{\geq})} = \{r_k \in \mathcal{R}_{all}^{\underline{\Pi}(Cl_t^{\geq})}, \text{ such that } v_k^* = 1\},$$

where f_k^* is the optimal value of f_k and v_k^* are the values of the binary variables at the corresponding optimum found. Other sets can be identified by adding constraints that forbid finding again the solutions which have been already identified in the previously conducted optimizations:

$$\sum_{r_k \in \mathcal{R}_k^{\underline{\Pi}(Cl_t^{\geq})}} v_k \leq f_k^* - 1.$$

Let us denote by $\mathcal{R}_{mrc}^{\underline{\Pi}(Cl_t^{\geq})}$ ($\mathcal{R}_{mrc}^{\underline{\Pi}(Cl_t^{\leq})}$) all MC sets of rules for $\underline{\Pi}(Cl_t^{\geq})$ ($\underline{\Pi}(Cl_t^{\leq})$). All compatible minimal sets of rules $\mathcal{R}^{A^R}$ are formed by the following product:

$$\mathcal{R}^{A^R} = \mathcal{R}_{mrc}^{\underline{\Pi}(Cl_2^{\geq})} \times \ldots \times \mathcal{R}_{mrc}^{\underline{\Pi}(Cl_n^{\geq})} \times \mathcal{R}_{mrc}^{\underline{\Pi}(Cl_1^{\leq})} \times \ldots \times \mathcal{R}_{mrc}^{\underline{\Pi}(Cl_{n-1}^{\leq})}. \tag{3}$$

When computing each MC rule set in $\mathcal{R}^{A^R}$ according to (3), we should eliminate decision rules from $\mathcal{R}_{mrc}^{\underline{\Pi}(Cl_t^{\geq})}$ or $\mathcal{R}_{mrc}^{\underline{\Pi}(Cl_t^{\leq})}$ with a consequent having at least the same strength (i.e., rules assigning objects to the same union or sub-union of classes) as some other rules from, respectively $\mathcal{R}_{mrc}^{\underline{\Pi}(Cl_h^{\geq})}$, $h > t$, or $\mathcal{R}_{mrc}^{\underline{\Pi}(Cl_h^{\leq})}$, $h < t$.

Illustrative study: part 3. All minimal-cover sets of rules for the lower approximation of each class union are presented in Table 3. In particular, there are two MC sets of rules for reference acts in $\underline{\Pi}(Cl_1^{\leq})$, $\underline{\Pi}(Cl_2^{\leq})$, and $\underline{\Pi}(Cl_3^{\geq})$, and a unique way of covering all reference acts in $\underline{\Pi}(Cl_2^{\geq})$. Combination of these minimal rule covers leads to 8 minimal sets of minimal rules $\mathcal{R}^{A^R}$ which reproduce the preference information provided by the DM.

Table 3. All minimal-cover sets of rules for the lower approximations of class unions

	Minimal-cover sets of rules		Minimal-cover sets of rules
$\mathcal{R}_{mrc}^{\underline{C}(Cl_2^{\geq})}$	$\{r_{\geq 2}^3\}$	$\mathcal{R}_{mrc}^{\underline{C}(Cl_1^{\leq})}$	$\{r_{\leq 1}^2\}$, $\{r_{\leq 1}^3\}$
$\mathcal{R}_{mrc}^{\underline{C}(Cl_3^{\geq})}$	$\{r_{\geq 3}^1\}$, $\{r_{\geq 3}^2\}$	$\mathcal{R}_{mrc}^{\underline{C}(Cl_2^{\leq})}$	$\{r_{\leq 2}^2\}$, $\{r_{\leq 2}^1, r_{\leq 2}^3\}$

5 Class acceptability indices

Each set of rules covering classification examples constitutes a preference model of the DM and can be used to classify new (non-reference) acts $A \setminus A^R$. We apply the following sorting method. Let us denote by $l^{\mathcal{R}}$ ($u^{\mathcal{R}}$) the lowest (highest) class of the intersection of suggested unions of all D$_{\geq}$- (D$_{\leq}$-) decision rules in $\mathcal{R}$ covering a. If $l^{\mathcal{R}}$ and/or $u^{\mathcal{R}}$ are undefined or $l^{\mathcal{R}} \leq u^{\mathcal{R}}$, then a sorting procedure driven by a compatible set of rules $\mathcal{R}$ assigns an act $a \in A$ to an interval of classes $[Cl_{L^{\mathcal{R}}(a)}, Cl_{R^{\mathcal{R}}(a)}]$ such that

- $L^{\mathcal{R}}(a) = l^{\mathcal{R}}$, if $l^{\mathcal{R}}$ is defined, and $L^{\mathcal{R}}(a) = 1$, otherwise,
- $R^{\mathcal{R}}(a) = u^{\mathcal{R}}$, if $u^{\mathcal{R}}$ is defined, and $R^{\mathcal{R}}(a) = n$, otherwise.

In case of inconsistency (i.e., if $u^{\mathcal{R}} < l^{\mathcal{R}}$), a is left without recommendation (i.e., the procedure indicates an empty set $\emptyset$ of classes).

Since all compatible sets of rules are known, for each range of contiguous classes $[Cl_{h_L}, Cl_{h_L+1}, \ldots, Cl_{h_R}]$, with $1 \leq h_L \leq h_R \leq n$, we can define *class range acceptability index* $CAI(a, [h_L, h_R])$ as the share of compatible sets of rules $\mathcal{R} \in \mathcal{R}^{A^R}$ that assign alternative a precisely to the range of classes $[Cl_{h_L}, Cl_{h_L+1}, \ldots, Cl_{h_R}]$. We can also compute the share of $\mathcal{R} \in \mathcal{R}^{A^R}$ for which Cl_h is within $[Cl_{L^{\mathcal{R}}(a)}, \ldots, Cl_{R^{\mathcal{R}}(a)}]$, i.e. the share of sets of rules that either precisely or imprecisely assign a to Cl_h. Let us call such a share the *cumulative class acceptability index* $CuCAI(a, h)$.

Note that if $CuCAI(a, h) > 0$, a is *possibly* assigned to Cl_h (let us denote it by $h \in C_P(a)$), because there exists at least one compatible set of rules assigning a to Cl_h. If $CuCAI(a, h) = 1$, a is *necessarily* assigned to Cl_h, because all compatible sets of rules assign a to Cl_h (then $h \in C_N(a)$).

Illustrative study: part 4. Among the non-reference acts $A^R = \{a_7, \dots, a_{12}\}$, there are three ones $\{a_7, a_9, a_{12}\}$ which are possibly and necessarily assigned to just a single class (see Table 4). The remaining three acts are possibly assigned to two consecutive classes (i.e., $Cl_1 - Cl_2$ or $Cl_2 - Cl_3$). Note, however, that a_8 and a_{10} are necessarily assigned to, respectively, Cl_1 and a class interval $Cl_1 - Cl_2$. For a_{11} there is no agreement with respect to recommendation between all compatible sets of rules; moreover, some models indicate $\emptyset$.

When it comes to class acceptability indices, the 5 acts that are necessarily assigned to some class or class interval have the respective $CuCAI$ 100%. In general, for acts that are possibly assigned to at least two consecutive classes, we can analyze CAIs and $CuCAI$s to indicate a recommendation suggested by most of compatible MC sets of rules. While for a_{11} the shares of compatible sets of rules indicating Cl_2 or Cl_3 (either precisely or imprecisely) are the same, for a_8 all compatible sets of rules suggest Cl_1, even though half of them indicate Cl_2 as well.

Table 4. Class acceptability indices (CAIs), cumulative class acceptability indices (CUCAIs), and possible C_P and necessary C_N assignments

	CAIs						CUCAIs				Assignments	
Act	$1-1$	$1-2$	$2-2$	$2-3$	$3-3$	$\emptyset$	1	2	3	$\emptyset$	C_P	C_N
a_7	–	–	100.0	–	–	–	–	100.0	–	–	2	2
a_8	50.0	50.0	–	–	–	–	100.0	50.0	–	–	$1-2$	1
a_9	–	–	–	–	100.0	–	–	–	100.0	–	3	3
a_{10}	–	100.0	–	–	–	–	100.0	100.0	–	–	$1-2$	$1-2$
a_{11}	–	–	25.0	25.0	25.0	25.0	–	50.0	50.0	25.0	$2-3$	
a_{12}	100.0	–	–	–	–	–	100.0	–	–	–	1	1

6 Conclusions

We integrated Robust Ordinal Regression into Dominance-based Rough Set Approach to modeling the decision under uncertainty. The whole approach is addressing decision situations where preference information provided by a Decision Maker (DM) is a set of classification examples concerning some reference acts relatively well-known to the DM. All considered acts are described by outcomes to be gained with given probabilities. The method proceeds as follows. First, we structure the classification data using the concept of dominance-based rough sets, discerning consistent from inconsistent classification examples. Then, we induce from this data all possible minimal-cover sets of rules which correspond to all instances of the preference model compatible with the input preference information. Precisely, the rules are induced from lower approximations of the unions of ordered quality classes of the reference acts, i.e., from consistent classification examples. A minimal-cover set of rules is covering all consistent classification examples using a minimal number of "if..., then..." decision rules, chosen from among all possible rules. Finally, we apply these multiple instances of the compatible preference model on a set of unseen acts, and we draw robust conclusions about their quality using the Robust Ordinal Regression paradigm. Specifically, for each act we derive the necessary and possible assignments, specifying the range of classes to which the act is assigned by all or at least one minimal-cover set of rules, respectively. We also provide class acceptability indices informing for each particular act about the distribution of assignment decisions by all minimal-cover sets of rules.

The whole approach can be extended in several ways, in particular, by:

– considering non-additive probability instead of additive one,
– accounting for consequences distributed over time,
– selecting a single representative minimal-cover set of rules among all compatible ones.

Acknowledgment

The first two authors wish to acknowledge financial support from the Polish National Science Center (PNSC), grant no. DEC-2011/01/B/ST6/07318.

References

1. J. Błaszczyński, R. Słowiński, and M. Szelag. Sequential covering rule induction algorithm for variable consistency rough set approaches. *Information Sciences*, 181(5):987–1002, 2011.
2. S. Corrente, S. Greco, M. Kadziński, and R. Słowiński. Robust ordinal regression in preference learning and ranking. *Machine Learning*, 93(2-3):381–422, 2013.
3. P.C. Fishburn. *Nonlinear Preferences and Utility Theory.* The John Hopkins University Press, 1988.
4. S. Greco, B. Matarazzo, and R. Słowiński. Rough Sets Theory for Multicriteria Decision Analysis. *European Journal of Operational Research*, 129:1–47, 2001.
5. S. Greco, B. Matarazzo, and R. Słowiński. Dominance-based Rough Set Approach to decision under uncertainty and time preference. *Annals of Operations Research*, 176:41–75, 2010.
6. S. Greco, V. Mousseau, and R. Słowiński. Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research*, 191(2):415–435, 2008.
7. M. Kadziński, S. Greco, and R. Słowiński. Robust Ordinal Regression for Dominance-based Rough Set Approach to Multiple Criteria Sorting. *Information Sciences*, submitted for publication, 2014.
8. Z. Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers, 1991.