Bayesian Decision Theory for Dominance-Based Rough Set Approach

Salvatore Greco[1], Roman Słowiński[2], and Yiyu Yao[3]

[1] Faculty of Economics, University of Catania,
Corso Italia, 55, 95129 Catania, Italy
salgreco@unict.it
[2] Institute of Computing Science, Poznań University of Technology,
60-965 Poznań, and Institute for Systems Research,
Polish Academy of Sciences, 01-447 Warsaw, Poland
roman.slowinski@cs.put.poznan.pl
[3] Department of Computer Science, University of Regina,
S4S0A2 Regina, Saskatchewan, Canada
yyao@cs.uregina.ca

Abstract. Dominance-based Rough Set Approach (DRSA) has been proposed to generalize classical rough set approach when consideration of monotonicity between degrees of membership to considered concepts has to be taken into account. This is typical for data describing various phenomena, e.g., "the larger the mass and the smaller the distance, the larger the gravity", or "the more a tomato is red, the more it is ripe". These monotonicity relationships are fundamental in rough set approach to multiple criteria decision analysis. In this paper, we propose a Bayesian decision procedure for DRSA. Our approach permits to take into account costs of misclassification in fixing parameters of the Variable Consistency DRSA (VC-DRSA), being a probabilistic model of DRSA.

Keywords: Bayesian Decision Theory, Dominance, Rough Set Theory, Variable Consistency, Cost of Misclassification.

1 Introduction

Rough set theory has been proposed by Pawlak in the early 80s [5,6] as a tool for reasoning about data in terms of granules of knowledge. While the original rough set idea is very useful for classification support, it is not handling a background knowledge about monotonic relationship between evaluation of objects on condition attributes and their evaluation on decision attributes. Such a knowledge is typical for data describing various phenomena and for data describing multiple criteria decision problems. E.g., "the larger the mass and the smaller the distance, the larger the gravity", "the more a tomato is red, the more it is ripe" or "the better the school marks of a pupil, the better his overall classification". The monotonic relationships within multiple criteria decision problems follow from preferential ordering of value sets of attributes (scales of criteria), as well as preferential ordering of decision classes. In order to handle these monotonic

J.T. Yao et al. (Eds.): RSKT 2007, LNAI 4481, pp. 134–141, 2007.

relationships between conditions and decisions, Greco, Matarazzo and Słowiński [2,3,7] proposed to substitute the indiscernibility relation for a dominance relation. Dominance-based Rough Set Approach (DRSA) permits approximation of ordered sets. When dealing with preferences, monotonicity is expressed through the following relationship: "the better is an object with respect to (w.r.t.) considered points of view (criteria), the more it is appreciated". The definitions of rough approximations originally introduced in DRSA are based on a strict application of the dominance principle. However, when defining non-ambiguous objects, it is reasonable to accept a limited proportion of negative examples, particularly for large data tables. Such extended version of DRSA is called Variable Consistency DRSA model (VC-DRSA) [4] being a probabilistic model of DRSA. The focus of this paper is on extending the Bayesian decision theoretic framework [1], already introduced in case of classical rough set approach [8], to the VC-DRSA model. The paper is organized as follows. In the next section, the general principle of DRSA are recalled, together with a presentation of VC-DRSA. In the third section, a Bayesian decision procedure for DRSA is presented. The last sections contains conclusions.

2 Dominance-Based Rough Set Approach

In data analysis, information about objects can be represented in the form of an information table. The rows of the table are labelled by objects, whereas columns are labelled by attributes and entries of the table are attribute-values. Formally, by an information table we understand the 4-tuple $S =< U, Q, V, f >$, where U is a finite set of objects, Q is a finite set of attributes, $V = \bigcup_{q \in Q} V_q$, where V_q is a value set of the attribute q, and $f : U \times Q \rightarrow V$ is a total function such that $f(x, q) \rightarrow V_q$ for every $q \in Q$, $x \in U$, called an information function [6]. The set Q is, in general, divided into set C of condition attributes and set D of decision attributes. Assuming that all condition attributes $q \in C$ are criteria, let $\succeq_q$ be a weak preference relation on U w.r.t. criterion q such that $x \succeq_q y$ means "x is at least as good as y w.r.t. criterion q". We suppose that $\succeq_q$ is a complete preorder, i.e. a strongly complete and transitive binary relation, defined on U on the basis of evaluations $f(\cdot, q)$. Without loss of generality, we can assume that for all $x, y \in U$, $x \succeq_q y$ iff $f(x, q) \geq f(y, q)$. Furthermore, let us assume that the set of decision attributes D (possibly a singleton $\{d\}$) makes a partition of U into a finite number of decision classes $Cl = \{Cl_t, t \in T\}$, $T = \{1, ..., n\}$, such that each $x \in U$ belongs to one and only one class $Cl_t \in Cl$. We suppose that the classes are preference-ordered, i.e. for all $r, s \in T$, such that $r > s$, the objects from Cl_r are preferred to the objects from Cl_s. More formally, if $\succeq$ is a comprehensive weak preference relation on U, i.e. if for all $x, y \in U$, $x \succeq y$ means "x is at least as good as y", we suppose:

$$[x \in Cl_r, y \in Cl_s, r > s] \Rightarrow [x \succeq y \text{ and not } y \succeq x].$$

The above assumptions are typical for consideration of a multiple-criteria sorting problem. The sets to be approximated are called *upward union* and *downward union* of classes, respectively:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, \quad Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s, \; t = 1, ..., n.$$

The statement $x \in Cl_t^{\geq}$ means "x belongs to at least class Cl_t", while $x \in Cl_t^{\leq}$ means "x belongs to at most class Cl_t". Let us remark that $Cl_1^{\geq} = Cl_n^{\leq} = U$, $Cl_n^{\geq}{=}Cl_n$ and $Cl_1^{\leq}{=}Cl_1$. Furthermore, for t=2,...,n, we have:

$$Cl_{t-1}^{\leq} = U - Cl_t^{\geq} \text{ and } Cl_t^{\geq} = U - Cl_{t-1}^{\leq}.$$

The key idea of rough sets is approximation of one knowledge by another knowledge. In classical rough set approach (CRSA) [6], the knowledge approximated is a partition of U into classes generated by a set of decision attributes; the knowledge used for approximation is a partition of U into elementary sets of objects that are indiscernible with respect to a set of condition attributes. The elementary sets are seen as "granules of knowledge". In DRSA [2,3,7], where condition attributes are criteria and classes are preference-ordered, the knowledge approximated is a collection of upward and downward unions of classes and the "granules of knowledge" are sets of objects defined using a dominance relation, instead of an indiscernibility relation used in CRSA. This is the main difference between CRSA and DRSA. In the following, in order to gain some more flexibility, we use the *variable consistency* DRSA model [4] which has its counterpart within the CRSA in the variable precision rough set approach [9,10]. Let us define now the dominance relation. We say that "x dominates y w.r.t. $P \subseteq C$, denoted by xD_Py, if $x \succeq_q y$ for all $q \in P$.

Given a set of criteria $P \subseteq C$ and $x \in U$, the "granules of knowledge" used for approximation in DRSA are:

- a set of objects dominating x, called *P-dominating set*,

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

- a set of objects dominated by x, called *P-dominated set*,

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

For any $P \subseteq C$ we say that $x \in U$ belongs to $Cl_t^{\geq}$ with no ambiguity at consistency level $l \in (0, 1]$, if $x \in Cl_t^{\geq}$ and at least $l \times 100\%$ of all objects $y \in U$ dominating x w.r.t. P also belong to $Cl_t^{\geq}$, i.e.

$$\frac{|D_P^+(x) \cap Cl_t^{\geq}|}{|D_P^+(x)|} \geq l \qquad (i)$$

where, for any set A, $|A|$ denotes its cardinality.
In this case, we say that x is a non-ambiguous object at consistency level l w.r.t. the upward union $Cl_t^{\geq}$ $(t = 2, ..., n)$. Otherwise, we say that x is an ambiguous object at consistency level l w.r.t. the upward union $Cl_t^{\geq}$ $(t = 2, ..., n)$.

Let us remark that $\frac{|D_P^+(x)\cap Cl_t^{\geq}|}{|D_P^+(x)|}$ can be interpreted as an estimation of the probability $P(y \in Cl_t^{\geq}|yD_Px)$ in the data table and thus (i) can be rewritten as

$$P(y \in Cl_t^{\geq}|yD_Px) \geq l.$$

The level l is called consistency level because it controls the degree of consistency between objects qualified as belonging to $Cl_t^{\geq}$ without any ambiguity. In other words, if $l < 1$, then no more than $(1-l)\times 100\%$ of all objects $y \in U$ dominating x w.r.t. P do not belong to $Cl_t^{\geq}$ and thus contradict the inclusion of x in $Cl_t^{\geq}$. Analogously, for any $P \subseteq C$ we say that $x \in U$ belongs to $Cl_t^{\leq}$ with no ambiguity at consistency level $l \in (0,1]$, if $x \in Cl_t^{\leq}$ and at least $l \times 100\%$ of all objects $y \in U$ dominated by x w.r.t. P also belong to $Cl_t^{\leq}$, i.e.

$$\frac{|D_P^-(x)\cap Cl_t^{\leq}|}{|D_P^-(x)|} \geq l. \qquad (ii)$$

In this case, we say that x is a non-ambiguous object at consistency level l w.r.t. the downward union $Cl_t^{\leq}$ $(t = 1,...,n-1)$. Otherwise, we say that x is an ambiguous object at consistency level l w.r.t. the downward union $Cl_t^{\leq}$ $(t = 1,...,n-1)$. Let us remark that $\frac{|D_P^-(x)\cap Cl_t^{\leq}|}{|D_P^-(x)|}$ can be interpreted as an estimation of the probability $P(y \in Cl_t^{\leq}|xD_Py)$ in the data table and thus (ii) can be rewritten as

$$P(y \in Cl_t^{\leq}|xD_Py) \geq l.$$

The concept of non-ambiguous objects at some consistency level l leads naturally to the definition of P-*lower approximations* of the unions of classes $Cl_t^{\geq}$ and $Cl_t^{\leq}$, denoted by $\underline{P}^l(Cl_t^{\geq})$ and $\underline{P}^l(Cl_t^{\leq})$, respectively:

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

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

P-*lower approximations* of the unions of classes $Cl_t^{\geq}$ and $Cl_t^{\leq}$ can also be formulated in terms of conditional probabilities as follows:

$$\underline{P}^l(Cl_t^{\geq}) = \left\{x \in Cl_t^{\geq} : P(y \in Cl_t^{\geq}|yD_Px) \geq l\right\},$$

$$\underline{P}^l(Cl_t^{\leq}) = \left\{x \in Cl_t^{\leq} : P(y \in Cl_t^{\leq}|xD_Py) \geq l\right\}.$$

Given $P \subseteq C$ and consistency level $l \in (0,1]$, we can define the P-*upper approximations* of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, denoted by $\overline{P}^l(Cl_t^{\geq})$ and $\overline{P}^l(Cl_t^{\leq})$, by complementarity of $\underline{P}^l(Cl_t^{\geq})$ and $\underline{P}^l(Cl_t^{\leq})$ w.r.t. U:

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

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

$\overline{P}^{l}(Cl_t^{\geq})$ can be interpreted as the set of all the objects belonging to $Cl_t^{\geq}$, possibly ambiguous at consistency level l. Analogously, $\overline{P}^{l}(Cl_t^{\leq})$ can be interpreted as the set of all the objects belonging to $Cl_t^{\leq}$, possibly ambiguous at consistency level l. The P-boundaries (P-doubtful regions) of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined as:

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

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

The variable consistency model of the dominance-based rough set approach provides some degree of flexibility in assigning objects to lower and upper approximations of the unions of decision classes. It can easily be shown that for $0 < l' < l \leq 1$,

$$\underline{P}^{l}(Cl_t^{\geq}) \subseteq \underline{P}^{l'}(Cl_t^{\geq}), \quad \overline{P}^{l}(Cl_t^{\geq}) \supseteq \overline{P}^{l'}(Cl_t^{\geq}), \quad t = 2, ..., n,$$

$$\underline{P}^{l}(Cl_t^{\leq}) \subseteq \underline{P}^{l'}(Cl_t^{\leq}), \quad \overline{P}^{l}(Cl_t^{\leq}) \supseteq \overline{P}^{l'}(Cl_t^{\leq}), \quad t = 1, ..., n-1.$$

The dominance-based rough approximations of upward and downward unions of classes can serve to induce a generalized description of objects contained in the information table in terms of "if..., then..." decision rules. The following two basic types of variable-consistency decision rules can be induced from lower approximations of upward and downward unions of classes:

1. $D_{\geq}$-decision rules with the following syntax:
   "if $f(x, q1) \geq r_{q1}$ and $f(x, q2) \geq r_{q2}$ and ... $f(x, qp) \geq r_{qp}$, then $x \in Cl_t^{\geq}$" in α% of cases, where $t = 2, ..., n$, $P = \{q1, ..., qp\} \subseteq C$, $(r_{q1}, ..., r_{qp}) \in V_{q1} \times V_{q2} \times ... \times V_{qp}$.
2. $D_{\leq}$-decision rules with the following syntax:
   "if $f(x, q1) \leq r_{q1}$ and $f(x, q2) \leq r_{q2}$ and ... $f(x, qp) \leq r_{qp}$, then $x \in Cl_t^{\leq}$" in α% of cases, where $t = 1, ..., n-1$, $P = \{q1, ..., qp\} \subseteq C$, $(r_{q1}, ..., r_{qp}) \in V_{q1} \times V_{q2} \times ... \times V_{qp}$.

3 The Bayesian Decision Procedure for DRSA

Let $P(y \in Cl_t^{\geq}|yD_Px)$ be the probability of an object $y \in U$ to belong to $Cl_t^{\geq}$ given yD_Px, that is the probability that y belongs to a class of at least level t, given that y dominates x w.r.t. set of criteria $P \subseteq C$. Analogously, let $P(y \in Cl_t^{\leq}|xD_Py)$ be the probability of an object $y \in U$ to belong to $Cl_t^{\leq}$ given xD_Py, that is the probability that x belongs to a class of at most level t, given that y is dominated by x w.r.t. set of criteria $P \subseteq C$. One can also consider probabilities $P(y \in Cl_{t-1}^{\leq}|yD_Px)$ and $P(y \in Cl_{t+1}^{\geq}|xD_Py)$. Obviously, we have that $P(y \in Cl_{t-1}^{\leq}|yD_Px) = 1 - P(y \in Cl_t^{\geq}|yD_Px)$, $t = 2, ..., n$, and $P(y \in Cl_{t+1}^{\geq}|xD_Py) = 1 - P(y \in Cl_t^{\leq}|xD_Py)$, $t = 1, ..., n-1$.

Let $\lambda(z \in Cl_t^{\geq}|z \in Cl_t^{\geq})$ denote the loss for assigning an object $z \in U$ to $Cl_t^{\geq}$ when this is true, i.e. when condition $z \in Cl_t^{\geq}$ holds, $t = 2, ..., n$. Analogously,

- $\lambda(z \in Cl_t^{\geq}|z \in Cl_{t-1}^{\leq})$ denotes the loss for assigning an object $z \in U$ to $Cl_t^{\geq}$ when this is false, i.e. when condition $z \in Cl_{t-1}^{\leq}$ holds, $t = 2, ..., n$,
- $\lambda(z \in Cl_t^{\leq}|z \in Cl_t^{\leq})$ denotes the loss for assigning an object $z \in U$ to $Cl_t^{\leq}$ when this is true, i.e. when condition $z \in Cl_t^{\leq}$ holds, $t = 1, ..., n-1$,
- $\lambda(z \in Cl_t^{\leq}|z \in Cl_{t+1}^{\geq})$ denotes the loss for assigning an object $z \in U$ to $Cl_t^{\leq}$ when this is false, i.e. when condition $z \in Cl_{t+1}^{\geq}$ holds, $t = 1, ..., n-1$.

In the following, we suppose for simplicity that the above losses are independent from object z.

Given an object $y \in U$, such that yD_Px, the expected losses $R(y \in Cl_t^{\geq}|yD_Px)$ and $R(y \in Cl_{t-1}^{\leq}|yD_Px)$ associated with assigning y to $Cl_t^{\geq}$ and $Cl_{t-1}^{\leq}$, $t = 2, ..., n$, respectively, can be expressed as:

$$R(y \in Cl_t^{\geq}|yD_Px) = \lambda(y \in Cl_t^{\geq}|y \in Cl_t^{\geq})P(y \in Cl_t^{\geq}|yD_Px)+$$

$$\lambda(y \in Cl_t^{\geq}|y \in Cl_{t-1}^{\leq})P(y \in Cl_{t-1}^{\leq}|yD_Px),$$

$$R(y \in Cl_{t-1}^{\leq}|yD_Px) = \lambda(y \in Cl_{t-1}^{\leq}|y \in Cl_t^{\geq})P(y \in Cl_t^{\geq}|yD_Px)+$$

$$\lambda(y \in Cl_{t-1}^{\leq}|y \in Cl_{t-1}^{\leq})P(y \in Cl_{t-1}^{\leq}|yD_Px).$$

By applying the Bayesian decision procedure, we obtain the following minimum-risk decision rules:

- assign y to $Cl_t^{\geq}$ if $R(y \in Cl_t^{\geq}|yD_Px) \geq R(y \in Cl_{t-1}^{\leq}|yD_Px)$,
- assign y to $Cl_{t-1}^{\leq}$ if $R(y \in Cl_t^{\geq}|yD_Px) < R(y \in Cl_{t-1}^{\leq}|yD_Px)$.

It is quite natural to assume that

$$\lambda(z \in Cl_t^{\geq}|z \in Cl_t^{\geq}) < \lambda(z \in Cl_{t-1}^{\leq}|z \in Cl_t^{\geq}) \text{ and}$$

$$\lambda(z \in Cl_{t-1}^{\leq}|z \in Cl_{t-1}^{\leq}) < \lambda(z \in Cl_t^{\geq}|z \in Cl_{t-1}^{\leq}).$$

That is, the loss of classifying an object belonging to $Cl_t^{\geq}$ into the correct class $Cl_t^{\geq}$ is smaller than the loss of classifying it into the incorrect class $Cl_{t-1}^{\leq}$; whereas the loss of classifying an object not belonging to $Cl_t^{\geq}$ into the class $Cl_t^{\geq}$ is greater than the loss of classifying it into the class $Cl_{t-1}^{\leq}$. With this loss function and the fact that $P(y \in Cl_t^{\geq}|yD_Px) + P(y \in Cl_{t-1}^{\leq}|yD_Px) = 1$, the above decision rules can be expressed as:

- assign y to $Cl_t^{\geq}$ if $P(y \in Cl_t^{\geq}|yD_Px) \geq \alpha_t$,
- assign y to $Cl_{t-1}^{\leq}$ if $P(y \in Cl_t^{\geq}|yD_Px) < \alpha_t$,

$$\text{where} \quad \alpha_t = \frac{\lambda(z \in Cl_t^{\geq}|z \in Cl_{t-1}^{\leq}) - \lambda(z \in Cl_{t-1}^{\leq}|z \in Cl_{t-1}^{\leq})}{\Lambda^{\geq}}, \text{ and}$$

$$\Lambda^{\geq} = \lambda(z \in Cl_t^{\geq}|z \in Cl_{t-1}^{\leq}) + \lambda(z \in Cl_{t-1}^{\leq}|z \in Cl_t^{\geq}) -$$
$$\lambda(z \in Cl_{t-1}^{\leq}|z \in Cl_{t-1}^{\leq}) - \lambda(z \in Cl_t^{\geq}|z \in Cl_t^{\geq}).$$

Given an object $y \in U$, such that xD_Py, the expected losses $R(y \in Cl_t^{\leq}|xD_Py)$ and $R(y \in Cl_{t+1}^{\geq}|xD_Py)$ associated with assigning y to $Cl_t^{\leq}$ and $Cl_{t+1}^{\geq}$, respectively, can be expressed as:

$$R(y \in Cl_t^{\leq}|xD_Py) = \lambda(y \in Cl_t^{\leq}|y \in Cl_t^{\leq})P(y \in Cl_t^{\leq}|xD_Py) +$$
$$\lambda(y \in Cl_t^{\leq}|y \in Cl_{t+1}^{\geq})P(y \in Cl_{t+1}^{\geq}|xD_Py),$$
$$R(y \in Cl_{t+1}^{\geq}|xD_Py) = \lambda(y \in Cl_{t+1}^{\geq}|y \in Cl_t^{\leq})P(y \in Cl_t^{\leq}|xD_Py) +$$
$$\lambda(y \in Cl_{t+1}^{\geq}|y \in Cl_{t+1}^{\geq})P(y \in Cl_{t+1}^{\geq}|xD_Py).$$

By applying the Bayesian decision procedure, we obtain the following minimum-risk decision rules:

- assign y to $Cl_t^{\leq}$ if $R(y \in Cl_t^{\leq}|xD_Py) \geq R(y \in Cl_{t+1}^{\geq}|xD_Py)$,
- assign y to $Cl_{t+1}^{\geq}$ if $R(y \in Cl_t^{\leq}|xD_Py) < R(y \in Cl_{t+1}^{\geq}|xD_Py)$.

It is quite natural to assume that

$$\lambda(z \in Cl_t^{\leq}|z \in Cl_t^{\leq}) < \lambda(z \in Cl_{t+1}^{\geq}|z \in Cl_t^{\leq}) \quad \text{and}$$
$$\lambda(z \in Cl_{t+1}^{\geq}|z \in Cl_{t+1}^{\geq}) < \lambda(z \in Cl_t^{\leq}|z \in Cl_{t+1}^{\geq}).$$

That is, the loss of classifying an object belonging to $Cl_t^{\leq}$ into the correct class $Cl_t^{\leq}$ is smaller than the loss of classifying it into the incorrect class $Cl_{t+1}^{\geq}$; whereas the loss of classifying an object not belonging to $Cl_t^{\leq}$ into the class $Cl_t^{\leq}$ is greater than the loss of classifying it into the class $Cl_{t+1}^{\geq}$. With this loss function and the fact that $P(y \in Cl_t^{\leq}|xD_Py) + P(y \in Cl_{t+1}^{\geq}|xD_Py) = 1$, the above decision rules can be expressed as:

- assign y to $Cl_t^{\leq}$ if $P(y \in Cl_t^{\leq}|xD_Py) \geq \beta_t$,
- assign y to $Cl_{t+1}^{\leq}$ if $P(y \in Cl_t^{\leq}|xD_Py) < \beta_t$,

$$\text{where} \quad \beta_t = \frac{\lambda(z \in Cl_{t+1}^{\geq}|z \in Cl_{t+1}^{\geq}) - \lambda(z \in Cl_{t+1}^{\geq}|z \in Cl_t^{\leq})}{\Lambda^{\leq}}, \text{ and}$$

$$\Lambda^{\leq} = \lambda(z \in Cl_t^{\leq}|z \in Cl_{t+1}^{\geq}) + \lambda(z \in Cl_{t+1}^{\geq}|z \in Cl_t^{\leq}) -$$
$$\lambda(z \in Cl_{t+1}^{\geq}|z \in Cl_{t+1}^{\geq}) - \lambda(z \in Cl_t^{\leq}|z \in Cl_t^{\leq}).$$

Using the values of parameters α_t and β_t obtained using the Bayesian decision procedure, we can redefine the *P-lower approximations* of the unions of classes $Cl_t^{\geq}$ and $Cl_t^{\leq}$, denoted by $\underline{P}^{\alpha_t}(Cl_t^{\geq})$ and $\underline{P}^{\beta_t}(Cl_t^{\leq})$, as follows:

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

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

4 Conclusions

In this paper, we proposed a Bayesian decision procedure for DRSA that permits to take into account costs of misclassification in fixing parameters of the probabilistic model of DRSA, i.e. VC-DRSA. Future research will focus on investigation of the formal properties of the proposed model and on comparison of its performance with competitive models in data analysis.

Acknowledgement. The second author wishes to acknowledge financial support from the Polish Ministry of Science and Higher Education (3T11F 02127).

References

1. Duda, R. O., Hart., P. E.: Pattern Classification and Scene Analysis, Wiley (1973).
2. Greco, S., Matarazzo, B., Słowiński, R.: Rough set theory for multicriteria decision analysis, European Journal of Operational Research, **129** (2001) 1-47.
3. Greco, S., Matarazzo, B., Słowiński R.: Decision rule approach. In: J. Figueira, S. Greco, M. Erghott (Eds.), Multiple Criteria Decision Analysis: State of the Art Surveys, Springer (2005) 507-563.
4. Greco, S., Matarazzo, B., Słowiński, R., Stefanowski, J.: Variable consistency model of dominance-based rough sets approach. In: W. Ziarko, Y. Yao (Eds.), Rough Sets and Current Trends in Computing, LNAI **2005**. Springer, Berlin (2001) 170-181.
5. Pawlak, Z.: Rough Sets, International Journal of Computer and Information Sciences **11** (1982) 341-356.
6. Pawlak, Z.: Rough Sets, Kluwer (1991).
7. Słowiński, R., Greco, S., Matarazzo, B.: Rough set based decision support. In: E. Burke, G. Kendall (Eds.), Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques, Springer (2005) 475-527.
8. Yao, Y.Y., Wong, S.K.M.: A decision theoretic framwork for approximating concepts, International Journal of Man-machine Studies **37** (1992) 793-809.
9. Ziarko W.: Variable precision rough sets model, Journal of Computer and Systems Sciences **46** (1993) 39-59.
10. Ziarko W.: Rough sets as a methodology for data mining. In: L. Polkowski, A. Skowron (Eds.), Rough Sets in Data Mining and Knowledge Discovery, **1**. Physica (1998) 554-576.