Optimized Generalized Decision in Dominance-Based Rough Set Approach

Krzysztof Dembczyński[1], Salvatore Greco[2], and Wojciech Kotłowski[1], and Roman Słowiński[1,3]

[1] Institute of Computing Science, Poznań University of Technology, 60-965 Poznań, Poland
{kdembczynski,wkotlowski,rslowinski}@cs.put.poznan.pl
[2] Faculty of Economics, University of Catania, 95129 Catania, Italy
salgreco@unict.it
[3] Institute for Systems Research, Polish Academy of Sciences, 01-447 Warsaw, Poland

Abstract. Dominance-based Rough Set Approach (DRSA) has been proposed to deal with multi-criteria classification problems, where data may be inconsistent with respect to the dominance principle. However, in real-life datasets, in the presence of noise, the notions of lower and upper approximations handling inconsistencies were found to be excessively restrictive which led to the proposal of the variable consistency variant of the theory. In this paper, we deal with a new approach based on DRSA, whose main idea is based on the error corrections. A new definition of the rough set concept known as generalized decision is introduced, *the optimized generalized decision*. We show its connections with statistical inference and dominance-based rough set theory.

1 Introduction

The multicriteria classification problem [10] consists in assignment of objects to pre-defined *decision classes* $Cl_t, t \in T = \{1, \ldots, n\}$. It is assumed that the classes are preference-ordered according to an increasing order of class indices, i.e. for all $r, s \in T$, such that $r > s$, the objects from Cl_r are strictly preferred to the objects from Cl_s. The objects are evaluated on a set of *condition criteria* (i.e., attributes with preference-ordered domains). It is assumed that a better evaluation of an object on a criterion, with other evaluations being fixed, should not worsen its assignment to a decision class. The problem of multicriteria classification can also be seen as a data analysis problem, under assumption of *monotone relationship* between the decision attribute and particular condition attributes, i.e. that the expected decision value increases (or decreases) with increasing (or decreasing) values on condition attributes. This definition is valid only in the probabilistic sense, so it may happen that there exists in the dataset X an object x_i not worse than another object x_k on all condition attributes, however x_i is assigned to a worse class than x_k; such a situation violates the monotone nature of data, so we shall call objects x_i and x_k *inconsistent with respect to dominance principle*.

J.T. Yao et al. (Eds.): RSKT 2007, LNAI 4481, pp. 118–125, 2007.

Rough set theory [9] has been adapted to deal with this kind of inconsistency and the resulting methodology has been called *Dominance-based Rough Set Approach* (DRSA) [5,6]. In DRSA, the classical indiscernibility relation has been replaced by a dominance relation. Using the rough set approach to the analysis of multicriteria classification data, we obtain lower and the upper (rough) approximations of unions of decision classes. The difference between upper and lower approximations shows inconsistent objects with respect to the dominance principle. Another, equivalent picture of this problem can be expressed in terms of the *generalized decision* concept [3,4].

Unfortunately, it can happen, that due to the random nature of data and due to the presence of noise, we loose too much information, thus making the DRSA inference model not accurate. In this paper, a new approach is proposed, based on combinatorial optimization for dealing with inconsistency, which can be viewed as a slightly different way of introducing variable precision in the DRSA. The new approach is strictly based on the generalized decision concept. It is an invasive method (contrary to DRSA), which reassigns the objects to different classes when they are traced to be inconsistent. We show, that this approach has statistical foundations and is strictly connected with the standard dominance-based rough set theory.

We assume that we are given a set $X = \{x_1, \ldots, x_\ell\}$, consisting of ℓ objects, with their decision values (class assignments) $Y = \{y_1, \ldots, y_\ell\}$, where each $y_i \in T$. Each object is described by a set of m condition attributes $Q = \{q_1, \ldots, q_m\}$ and by $\mathrm{dom} q_i$ we mean the set of values of attribute q_i. By the *attribute space* we mean the set $V = \mathrm{dom} q_1 \times \ldots \times \mathrm{dom} q_m$. Moreover, we denote the evaluation of object x_i on attribute q_j by $q_j(x_i)$. Later on we will abuse the notation a little bit, identifying each object x with its evaluations on all the condition attributes, $x \equiv (q_1(x), \ldots q_m(x))$. By a *class* $Cl_t \subset X$, we mean the set of objects, such that $y_i = t$, i.e. $Cl_t = \{x_i \in X \colon y_i = t, 1 \leq i \leq \ell\}$.

The article is organized in the following way. Section 2 describes main elements of DRSA. Section 3 presents an algorithmic background for new approach. The concept of optimized generalized decision is introduced in Section 4. In Section 5 the connection with statistical inference is shown. The paper ends with conclusions. Proofs of the theorems are omitted due to the space limit.

2 Dominance-Based Rough Set Approach

Within DRSA [5,6], we define the *dominance* relation D as a binary relation on X in the following way: for any $x_i, x_k \in X$ we say that x_i *dominates* x_k, $x_i D x_k$, if on every condition attribute, x_i has evaluation not worse than x_k, $q_j(x_i) \geq q_j(x_k)$, for all $1 \leq j \leq m$. The dominance relation D is a partial preorder on X, i.e. it is reflexive and transitive. The *dominance principle* can be expressed as follows. For all $x_i, x_j \in X$ it holds:

$$x_i D x_j \Longrightarrow y_i \geq y_j \tag{1}$$

The rough approximations concern granules resulting from information carried out by the decisions. The *decision granules* can be expressed by unions of decision classes: for all $t \in T$

$$Cl_t^{\geq} = \{x_i \in X : y_i \geq t\}, \qquad Cl_t^{\leq} = \{x_i \in X : y_i \leq t\}. \tag{2}$$

The *condition granules* are dominating and dominated sets defined as:

$$D^+(x) = \{x_i \in X : x_i D x\}, \qquad D^-(x) = \{x_i \in X : x D x_i\}. \tag{3}$$

Lower rough approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, $t \in T$, are defined as follows:

$$\underline{Cl_t^{\geq}} = \{x_i \in X : D^+(x_i) \subseteq Cl_t^{\geq}\}, \qquad \underline{Cl_t^{\leq}} = \{x_i \in X : D^-(x_i) \subseteq Cl_t^{\leq}\}. \tag{4}$$

Upper rough approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, $t \in T$, are defined as follows:

$$\overline{Cl_t^{\geq}} = \{x_i \in X : D^-(x_i) \cap Cl_t^{\geq} \neq \emptyset\}, \quad \overline{Cl_t^{\leq}} = \{x_i \in X : D^+(x_i) \cap Cl_t^{\leq} \neq \emptyset\}. \tag{5}$$

In the rest of this section we focus our attention on the *generalized decision* [3]. Consider the following definition of generalized decision $\delta_i = [l_i, u_i]$ for object $x_i \in X$, where:

$$l_i = \min\{y_j : x_j D x_i, x_j \in X\}, \tag{6}$$

$$u_i = \max\{y_j : x_i D x_j, x_j \in X\}. \tag{7}$$

In other words, the generalized decision reflects an interval of decision classes to which an object may belong due to the inconsistencies with the dominance principle caused by this object. Obviously, $l_i \leq y_i \leq u_i$ for every $x_i \in X$ and if $l_i = u_i$, then object x_i is consistent with respect to the dominance principle with every other object $x_k \in X$.

Let us remark that the dominance-based rough approximations may be expressed using generalized decision:

$$\begin{array}{ll} \underline{Cl_t^{\geq}} = \{x_i \in X : l_i \geq t\} & \overline{Cl_t^{\geq}} = \{x_i \in X : u_i \geq t\} \\ \underline{Cl_t^{\leq}} = \{x_i \in X : u_i \leq t\} & \overline{Cl_t^{\leq}} = \{x_i \in X : l_i \leq t\}. \end{array} \tag{8}$$

It is also possible to obtain generalized decision using the rough approximation:

$$l_i = \max\left\{t : x_i \in \underline{Cl_t^{\geq}}\right\} = \min\left\{t : x_i \in \overline{Cl_t^{\leq}}\right\} \tag{9}$$

$$u_i = \min\left\{t : x_i \in \underline{Cl_t^{\leq}}\right\} = \max\left\{t : x_i \in \overline{Cl_t^{\geq}}\right\} \tag{10}$$

Those two descriptions are fully equivalent. For the purpose of this text we will look at the concept of generalized decision from a different point of view. Let us define the following relation: the decision range $\alpha = [l^\alpha, u^\alpha]$ is *more* *informative* than $\beta = [l^\beta, u^\beta]$ if $\alpha \subseteq \beta$. We show now that the generalized decision concept (thus also DRSA rough approximations) is in fact the unique optimal non-invasive approach that holds the maximum possible amount of information which can be obtained from given data:

Theorem 1. *The generalized decisions $\delta_i = [l_i, u_i]$, for $x_i \in X$, are most informative ranges from any set of decisions ranges of the form $\alpha_i = [l_i^\alpha, u_i^\alpha]$ that have the following properties:*

1. *The sets $\{(x_i, l_i^\alpha) : x_i \in X\}$ and $\{(x_i, u_i^\alpha) : x_i \in X\}$, composed of objects with, respectively, decisions l_i^α and u_i^α assigned instead of y_i are consistent with the dominance principle.*
2. *For each $x_i \in X$ it holds $l_i^\alpha \leq y_i \leq u_i^\alpha$.*

3 Minimal Reassignment

A new proposal of the definitions of lower and upper approximations of unions of classes is based on the concept of *minimal reassignment*. At first, we define the reassignment of an object $x_i \in X$ as changing its decision value y_i. Moreover, by minimal reassignment we mean reassigning the smallest possible number of objects to make the set X consistent (with respect to the dominance principle). One can see, that such a reassignment of objects corresponds to indicating and correcting possible errors in the dataset, i.e. it is an invasive approach.

We denote the minimal number of reassigned objects from X by R. To compute R, one can formulate a linear programming problem. Such problems were already considered in [2] (in the context of binary and multi-class classification) and also in [1] (in the context of boolean regression). Here we formulate a similar problem, but with a different aim.

For each object $x_i \in X$ we introduce $n-1$ binary variables d_{it}, $t \in \{2, \ldots, n\}$, having the following interpretation: $d_{it} = 1$ iff object $x_i \in Cl_t^\geq$ (note that always $d_{i1} = 1$, since $Cl_1^\geq = X$). Such interpretation implies the following conditions:

$$\text{if } t' \geq t \text{ then } d_{it'} \leq d_{it} \tag{11}$$

for all $i \in \{1, \ldots, \ell\}$ (otherwise it would be possible that there exists object x_i belonging to $Cl_{t'}^\geq$, but not belonging to $Cl_t^\geq$, where $t' > t$). Moreover, we give a new value of decision y_i' to object x_i according to the rule: $y_i' = 1 + \sum_{t=2}^n d_{it}$ (the highest t such that x_i belongs to $Cl_t^\geq$). So, for each object $x_i \in U$ the cost function of the problem can be formulated as $R_i = (1 - d_{i,y_i}) + d_{i,y_i+1}$. Indeed, the value of decision for x_i changes iff $R_i = 1$ [4].

The following conditions must be satisfied for X to be consistent according to (1):

$$d_{it} \geq d_{jt} \qquad \forall i,j\colon\ x_i D x_j \quad 2 \leq t \leq n \tag{12}$$

Finally, we can formulate the problem in terms of integer linear programming:

$$\text{minimize} \quad R = \sum_{i=1}^{\ell} R_i = \sum_{i=1}^{\ell} ((1 - d_{i,y_i}) + d_{i,y_i+1}) \tag{13}$$

$$\begin{aligned} \text{subject to} \quad & d_{it'} \leq d_{it} && 1 \leq i \leq \ell, \quad 2 \leq t < t' \leq n \\ & d_{it} \geq d_{jt} && 1 \leq i, j \leq \ell, \quad x_i D x_j, \quad 2 \leq t \leq n \\ & d_{it} \in \{0, 1\} && 1 \leq i \leq \ell, \quad 2 \leq t \leq n \end{aligned}$$

The matrix of constraints in this case is totally unimodular [2,8], because it contains in each row either two values 1 and -1 or one value 1, and the right hand sides of the constraints are integer. Thus, we can relax the integer condition reformulating it as $0 \leq d_{it} \leq 1$, and get a linear programming problem. In [2], the authors give also a way for further reduction of the problem size. Here, we prove a more general result using the language of DRSA.

Theorem 2. *There always exists an optimal solution of (13), $y_i^* = 1 + \sum_{t=2}^{n} d_{it}^*$, for which the following condition holds: $l_i \leq y_i^* \leq u_i$, $1 \leq i \leq \ell$.*

Theorem 2 enables a strong reduction of the number of variables. For each object x_i, variables d_{it} can be set to 1 for $t \leq l_i$, and to 0 for $t > u_i$, since there exists an optimal solution to (13) with such values of the variables. In particular, if an object x_i is consistent (i.e. $l_i = u_i$), the class assignment for this object remains the same.

4 Construction of the Optimized Generalized Decisions

The reassignment cannot be directly applied to the objects from X, since the optimal solution may not be unique. Indeed, in some cases one can find different subsets of X, for which the change of decision values leads to the same value of cost function R. It would mean that the reassignment of class labels for some inconsistent objects depends on the algorithm used, which is definitely undesirable. To avoid that problem, we must investigate the properties of the set of optimal feasible solutions of the problem (13).

Let us remark the set of all feasible solutions to the problem (13) by F, where by solution f we mean a vector of new decision values assigned to objects from X, i.e. $f = (f_1, \ldots, f_\ell)$, where f_i is the decision value assigned by solution f to object x_i. We also denote the set of optimal feasible solutions by OF. Obviously, $OF \subset F$ and $OF \neq \emptyset$, since there exist feasible solutions, e.g. $f = (1, \ldots, 1)$.

Assume that we have two optimal feasible solutions $f = (f_1, \ldots, f_\ell)$ and $g = (g_1, \ldots, g_\ell)$. We define "min" and "max" operators on F as $\min\{f, g\} = (\min\{f_1, g_1\}, \ldots, \min\{f_\ell, g_\ell\})$ and $\max\{f, g\} = (\max\{f_1, g_1\}, \ldots, \max\{f_\ell, g_\ell\})$. The question arises, whether if $f, g \in OF$ then $\min\{f, g\}$ and $\max\{f, g\}$ also belong to OF? The following lemma gives the answer:

Lemma 1. *Assume $f, g \in OF$. Then $\min\{f, g\}, \max\{f, g\} \in OF$.*

Having the lemma, we can start to investigate the properties of the order in OF. We define a binary relation $\succeq$ on OF as follows:

$$\forall_{f,g \in OF} \quad (f \succeq g \Leftrightarrow \forall_{1 \leq i \leq \ell} \, f_i \geq g_i) \tag{14}$$

It can be easily verified that it is a partial order relation. We now state the following theorem:

Theorem 3. *There exist the greatest and the smallest element in the ordered set* $(OF, \succeq)$

Theorem 3 provides the way to define for all $x_i \in X$ *the optimized generalized decisions* $\delta_i^* = [l_i^*, u_i^*]$ as follows:

$$l_i^* = y_{*i} = \min\{f_i : f \in OF\} \tag{15}$$

$$u_i^* = y_i^* = \max\{f_i : f \in OF\} \tag{16}$$

Of course, both l^* and u^* are consistent with respect to dominance principle (since they belong to OF). The definitions are more resistant to noisy data, since they appear as solutions with minimal number of reassigned objects. It can be shown that using the classical generalized decision, for any consistent set X we can add two "nasty" objects to X (one, which dominates every object in X, but has the lowest possible class, and another which is dominated by every object in X, but has the highest possible class) to make the generalized decisions completely noninformative, i.e. for every object $x_i \in X$, l_i equals to the lowest possible class and u_i equals to the highest possible class. If we use the optimized generalized decisions to this problem, two "nasty" objects will be relabeled (properly recognized as errors) and nothing else will change.

Optimized generalized decision is a direct consequence of the non-uniqueness of optimal solution to the minimal reassignment problem (so also to the problems considered in [2,1]). Also note that using (15) and (16) and reversing the transformation with 8 we end up with new definitions of *optimized lower and upper approximations.*

The problem which is still not solved is how to find the smallest and the greatest solutions in an efficient way. We propose to do this as follows: we modify the objective function of (13) by introducing the additional term:

$$R' = \epsilon \sum_{i=1}^{\ell} \sum_{t=l_i}^{u_i} d_{it} \tag{17}$$

and when we seek the greatest solution we subtract R' from the original objective function, while when we seek the smallest solution we add R' to the original objective function, so we solve two linear programs with the following objective functions:

$$R_{\pm} = \sum_{i=1}^{\ell} R_i \pm R' = R \pm R' \tag{18}$$

To prove, that by minimizing the new objective function we indeed find what we require, we define $I = \sum_{i=1}^{\ell} (u_i - l_i)$. The following theorem holds:

Theorem 4. *When minimizing objective functions (18) one finds the smallest and the greatest solution provided $\epsilon < I^{-1}$.*

Note that the solutions to the modified problem are unique.

5 Statistical Base of Minimal Reassignment

In this section we introduce a statistical justification for the described approach. We consider here only the binary (two-class) problem, however this approach can be extended to the multi-class case. We state the following assumptions: each pair $(x_i, y_i) \in X \times Y$ is a realization of random vector $(\mathcal{X}, \mathcal{Y})$, independent and identically distributed (i.i.d.) [7,12]. Moreover, we assume that the statistical model is of the form $\mathcal{Y} = b(\mathcal{X}) \oplus \epsilon$, where $b(\cdot)$ is some function, such that $b(x) \in \{0, 1\}$ for all $x \in V$ and $b(x)$ is isotonic (monotone and decreasing) for all $x \in V$. We observe y which is the composition ($\oplus$ is binary addition) of $b(x)$ and some variable ϵ which is the random noise. If $\epsilon(x) = 1$, then we say, that the decision value was *misclassified*, while if $\epsilon(x) = 0$, than we say that the decision value was correct. We assume that $\Pr(\epsilon = 1) < \frac{1}{2} \equiv p$ and it is independent of x, so each object is misclassified with the same probability p.

We now use the maximum likelihood estimate (MLE). We do not know the real decision values $b(x_i) \equiv b_i$ for all $x_i \in X$ and we treat them as parameters. We fix all x_i and treat only y_i as random. Finally, considering $B = \{b_1, \ldots, b_\ell\}$ and denoting by ϵ_i the value of variable ϵ for object x_i, the MLE is as follows:

$$L(B;Y) = \Pr(Y|B) = \prod_{i=1}^{\ell} \Pr(y_i|b_i) = \prod_{i=1}^{\ell} p^{\epsilon_i}(1-p)^{1-\epsilon_i} \tag{19}$$

Taking minus logarithm of (19) (the negative log-likelihood) we equivalently minimize:

$$-\ln L(B;Y) = -\sum_{i=1}^{\ell} (\epsilon_i \ln p - (1-\epsilon_i)\ln(1-p)) = \ln \frac{1-p}{p} \sum_{i=1}^{\ell} \epsilon_i + \ell \ln(1-p) \tag{20}$$

We see, that for any fixed value of p, the negative log-likelihood reaches its minimum when the sum $\sum_{i=1}^{\ell} \epsilon_i$ is minimal. Thus, for any p, to maximize the likelihood, we must minimize the number of misclassifications. This is equivalent to finding values $b_i, 1 \leq i \leq \ell$, which are monotone, i.e. consistent with the dominance principle, and such that the number of misclassifications $\sum_{i=1}^{\ell} \epsilon_i = \sum_{i=1}^{\ell} |b_i - y_i|$ is minimal. Precisely, this is the two-class problem of minimal reassignment.

Finally, we should notice that for each $x \in X$, $b(x)$ is the most probable value of y (decision) for given x, since $p < \frac{1}{2}$. Therefore, we estimate the decision values, that would be assigned to the object by the *optimal Bayes classifier* [7], i.e. the classifier which has the smallest expected error.

6 Conclusions

We propose a new extension of the Dominance-based Rough Set Approach (DRSA), which involves a combinatorial optimization problem concerning minimal reassignment of objects. As it is strongly related to the standard DRSA, we describe our approach in terms of the generalized decision concept. By reassigning the minimal number of objects we end up with a non-univocal optimal solution. However, by considering the whole set of optimal solution, we can optimize the generalized decision, so as to make it more robust in the presence of noisy data.

On the other hand, reassigning the objects to different classes in view of making the dataset consistent, has a statistical justification. Under assumption of common misclassification probability for all of the objects, it is nothing else than a maximum likelihood estimate of the optimal Bayes classifier.

References

1. Boros, E., Hammer, P. L., Hooker, J. N.: Boolean regression. Annals of Operations Research, **58** (1995) 201–226
2. Chandrasekaran, R., Ryu, Y. U., Jacob, V., Hong, S.: Isotonic separation. INFORMS J. Comput., **17** (2005) 462–474
3. Dembczyński, K., Greco, S., Słowinski, R.: Second-order Rough Approximations in Multi-criteria Classification with Imprecise Evaluations and Assignments, LNAI, **3641** (2005) 54–63
4. Dembczyński, K., Greco, S., Kotłowski, W., Słowiński, R.: Quality of Rough Approximation in Multi-Criteria Classification Problems. LNCS **4259** (2006) 318-327.
5. Greco S., Matarazzo, B.: Słowiński, R.: Rough approximation of a preference relation by dominance relations. European Journal of Operational Research, **117** (1999) 63–83
6. Greco S., Matarazzo, B., Słowiński, R.: Rough sets theory for multicriteria decision analysis. European Journal of Operational Research, **129** (2001) 1–47
7. Hastie, T., Tibshirani, R., Friedman J.: The Elements of Statistical Learning. Springer, Berlin 2003.
8. Papadimitriou, C. H., Steiglitz, K.: Combinatorial Optimization. Dover Publications, New York (1998).
9. Pawlak, Z.: Rough sets. International Journal of Information & Computer Sciences, **11** (1982) 341–356
10. Roy, B., Decision science or decision aid science?. European Journal of Operational Research, **66** (1993) 184–203.
11. Słowiński, R., Greco S., Matarazzo, B.: Rough Set Based Decision Support. Chapter 16 [in]: Burke, E. K., Kendall, G. (eds.): Search Methodologies: Introductory Tutorials on Optimization and Decision Support Techniques. Springer-Verlag, New York (2005) 475–527.
12. Vapnik, V.: Statistical Learning Theory. Wiley-Interscience, New York (1998)