Quality of Rough Approximation in Multi-Criteria Classification Problems

Krzysztof Dembczyński[1], Salvatore Greco[2], 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. In this paper, we consider different measures of the quality of approximation, which is the value indicating how much inconsistent the decision table is. We begin with the classical definition, based on the relative number of inconsistent objects. Since this measure appears to be too restrictive in some cases, a new approach based on the concept of generalized decision is proposed. Finally, motivated by emerging problems in the presence of noisy data, the third measure based on the object reassignment is introduced. Properties of these measures are analysed in light of rough set theory.

1 Introduction

The multi-criteria classification problem consists in assignment of objects from a set A 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 there exists a semantic correlation between evaluation of objects on criteria and their assignment to decision classes, i.e. a better evaluation of an object on a criterion with other evaluations being fixed should not worsen its assignment to a decision class.

In order to support multi-criteria classification, one must construct a preference model of the Decision Maker (DM). The construction of the preference model requires some *preference information* from the DM. One possible way is to induce the preference model from a set of exemplary decisions (assignments of objects to decision classes) made on a set of selected objects called *reference objects*. The reference objects are those relatively well-known to the DM who is able to assign them to pre-defined classes. In other words, the preference information comes from observation of DM's acts (comprehensive decisions). It is concordant with a paradigm of artificial intelligence and, in particular, of inductive learning. Moreover, the induced model can be represented in intelligible way, for example as a set of decision rules.

The reference objects and their evaluations and assignments are often presented in a *decision table* $S = \langle U, C, D \rangle$, where $U \subseteq A$ is a finite, non-empty set of reference objects, C is a set of condition criteria, and D is a set of decision criteria that contain information on assignment of objects to decision classes. D is often a singleton ($D = \{d\}$), where d is shortly called *decision*. C and D are disjoint, finite and non-empty sets that jointly constitute a set of all criteria Q. It is assumed, without loss of generality, that the domain of each criterion $q \in Q$, denoted by V_q, is numerically coded with an increasing order of preference. The domains of criteria may correspond to cardinal or ordinal scales, however, we are exploiting the ordinal information (the weakest) only, whatever is the scale. The domain of decision d is a finite set ($T = \{1, \dots, n\}$) due to a finite number of decision classes. Evaluations and assignments of objects on any criterion ($q \in Q$) are defined by an *information function* $f(x, q)$, $f : U \times Q \to V$, where $V = \bigcup_{q \in Q} V_q$.

There is, however, a problem with inconsistency often present in the set of decision examples. Two decision examples are inconsistent with respect to, so-called, *dominance principle*, if there exists an object not worse than another object on all considered criteria, however, it has been assigned to a worse decision class than the other. To deal with these inconsistencies, it has been proposed to construct the preference model in the form of a set of decision rules, after adapting rough set theory [7–9] to preference ordered data. Such an adaptation has been made by Greco, Matarazzo and Słowiński [4–6]; it consists in substituting the classical indiscernibility relation by a dominance relation, which permits taking into account the preference order in domains (scales) of criteria. The extended rough set approach is called Dominance-based Rough Set Approach (DRSA) - a complete overview of this methodology is presented in [10].

Using the rough set approach to the analysis of preference information, we obtain the 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. The rough approximations are then used in induction of decision rules representing, respectively, certain and possible patterns of DM's preferences. The preference model in the form of decision rules explains a decision policy of the DM and permits to classify new objects in line of the DM's preferences.

The ratio of the cardinality of all consistent objects to the cardinality of all reference objects is called quality of approximation. This ratio is very restrictive, because in the extreme case, if there existed one object having better evaluations on condition criteria than all the other objects from U and if it was assigned to the worst class being a singleton, this ratio would decrease to 0. In the paper, we consider two other measures of the quality of approximation. The first, based on the generalized decision, is more resistant to local inconsistencies, but still in the extreme case described above, its value would decrease to 0. The second, motivated by emerging problems in the presence of noisy data, is free of this disadvantage and is resistant to local inconsistencies. Its definition is based on the concept of object reassignment. All these measures are monotonically non-decreasing with the number of condition criteria considered.

The article is organized in the following way. Section 2 describes main elements of Dominance-based Rough Set Approach. Section 3 describes the classical ratio of quality of approximation and the ratio based on generalized decision. In Section 4, the third measure and its properties are presented. The last section concludes the paper.

2 Dominance-based Rough Set Approach

Within DRSA, the notions of *weak preference* (or *outranking*) relation $\succeq_q$ and P-dominance relation D_P are defined as follows. For any $x, y \in U$ and $q \in Q$, $x \succeq_q y$ means that x is at least as good as (*is weakly preferred to*) y with respect to criterion q. With respect to assumptions taken in the previous section, it is $x \succeq_q y \Leftrightarrow f(x,q) \geq f(y,q)$. Moreover, taking into account more than one criterion, we say that x dominates y with respect to $P \subseteq Q$ (shortly x P-dominates y), if $x \succeq_q y$ for all $q \in P$. The weak preference relation $\succeq_q$ is supposed to be a complete pre-order and, therefore, the P-dominance relation D_P, being the intersection of complete pre-orders $\succeq_q$, $q \in P$, is a partial pre-order in the set of reference objects. The dominance principle can be expressed as follows, for $x, y \in U$, and $P \subseteq C$:

$$xD_P y \Rightarrow xD_{\{d\}}y, \text{ i.e.,} \qquad (\forall_{q \in P}\ f(x,q) \geq f(y,q)) \Rightarrow f(x,d) \geq f(y,d). \tag{1}$$

The rough approximations concern granules resulting from information carried out by the decision criterion. The approximation is made using granules resulting from information carried out by condition criteria. These granules are called *decision* and *condition* granules, respectively. The decision granules can be expressed by unions of decision classes:

$$Cl_t^{\geq} = \{y \in U : f(y,d) \geq t\} \tag{2}$$

$$Cl_t^{\leq} = \{y \in U : f(y,d) \leq t\}. \tag{3}$$

The condition granules are P-dominating and P-dominated sets defined, respectively, as:

$$D_P^{+}(x) = \{y \in U : yD_P x\} \tag{4}$$

$$D_P^{-}(x) = \{y \in U : xD_P y\}. \tag{5}$$

Let us remark that both decision and condition granules are cones in decision and condition spaces, respectively. The origin of a decision cone is a class index $t \in T$, while the origin of a condition cone is an object $x \in U$. The dominating cones are open towards increasing preferences, and the dominated cones are open towards decreasing preferences.

P-lower dominance-based rough approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined for $P \subseteq C$ and $t \in T$, respectively, as follows:

$$\underline{P}(Cl_t^{\geq}) = \{x \in U : D_P^+(x) \subseteq Cl_t^{\geq}\}, \tag{6}$$

$$\underline{P}(Cl_t^{\leq}) = \{x \in U : D_P^-(x) \subseteq Cl_t^{\leq}\}. \tag{7}$$

P-upper dominance-based rough approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined for $P \subseteq C$ and $t \in T$, respectively, as follows:

$$\overline{P}(Cl_t^{\geq}) = \{x \in U : D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\}, \tag{8}$$

$$\overline{P}(Cl_t^{\leq}) = \{x \in U : D_P^+(x) \cap Cl_t^{\leq} \neq \emptyset\}, \tag{9}$$

Consider the following definition of P-generalized decision for object $x \in U$:

$$\delta_P(x) = \langle l_P(x), u_P(x) \rangle, \text{ where,} \tag{10}$$

$$l_P(x) = \min\{f(y, d) : yD_Px, y \in U\}, \tag{11}$$

$$u_P(x) = \max\{f(y, d) : xD_Py, y \in U\}. \tag{12}$$

In other words, the P-generalized decision reflects an interval of decision classes to which an object may belong due to inconsistencies with the dominance principle caused by this object. $l_P(x)$ is the lowest decision class, to which belong an object P-dominating x; $u_P(x)$ is the highest decision class, to which belong an object P-dominated by x. Obviously, $l_P(x) \leq u_P(x)$ for every $P \subseteq C$, $x \in U$ and if $l_P(x) = u_P(x)$, then object x is consistent with respect to the dominance principle in the decision table.

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

$$\underline{P}(Cl_t^{\geq}) = \{x \in U : l_P(x) \geq t\}, \tag{13}$$

$$\overline{P}(Cl_t^{\geq}) = \{x \in U : u_P(x) \geq t\}, \tag{14}$$

$$\underline{P}(Cl_t^{\leq}) = \{x \in U : u_P(x) \leq t\}, \tag{15}$$

$$\overline{P}(Cl_t^{\leq}) = \{x \in U : l_P(x) \leq t\}. \tag{16}$$

The lower and the upper rough approximations are then used in induction of decision rules representing, respectively, certain and possible patterns of DM's preferences. These rules are used in classification of new objects. In general, a new object is covered by several rules indicating rough approximations of upward and downward unions of decision classes. Intersection of the outputs of the rules gives an interval of decision classes to which an object is assigned. In many cases the object is assigned to only one class resulting from the intersection of the matching rules.

3 Quality of approximation

Let us begin with very restrictive definition of the quality of approximation. The quality of approximation is defined as a ratio of the number of objects from the decision table that are consistent with respect to the dominance principle, to the number of all objects from this decision table. A set of consistent objects can be defined in the following way, for any $P \subseteq C$:

$$\{x \in U : u_P(x) = l_P(x)\}. \tag{17}$$

The same may be expressed, equivalently, by:

$$\begin{aligned} &\bigcup\nolimits_{t \in T}\{x \in U : D_P^-(x) \subseteq Cl_t^{\leq} \wedge D_P^+(x) \subseteq Cl_t^{\geq}\} = \\ &= U - \left(\bigcup\nolimits_{t \in T} Bn_P^{\geq t}\right) = U - \left(\bigcup\nolimits_{t \in T} Bn_P^{\leq t}\right), \end{aligned}$$

where $Bn_P^{\geq t} = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq})$, and $Bn_P^{\leq t} = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq})$, are, so-called, *boundary regions.*

The *quality of approximation* can be defined as:

$$\gamma(P) = \frac{card\left(\{x \in U : u_P(x) = l_P(x)\}\right)}{card\,(U)}. \tag{18}$$

This definition is very restrictive, because in the extreme case, if there existed one object dominating all the other objects from U while being assigned to the lowest possible class, and if the lowest possible class was a singleton including this object, $\gamma(P)$ would decrease to 0, even if the other objects from U were perfectly consistent. It is not true, however, that $\gamma(P)$ does not count the relative number of objects which can be captured by deterministic rules (i.e., induced from the lower approximations of unions of decision classes), what was pointed by Düntsch and Gediga in [3]. This is in fact, the relative number of objects that are covered by these rules in the following way. When deterministic rules induced from lower approximations of upward and downward unions of decision classes are applied to an object, then the object is assigned by these rules to an interval of decision classes to which it may belong. For a consistent object this interval boils down to a single class. The relative number of these objects is just shown by $\gamma(P)$.

It is easy to show that, for any $P \subseteq R \subseteq C$, there holds:

$$\gamma(P) \leq \gamma(R). \tag{19}$$

In other words, $\gamma(P)$ possesses a monotonicity property well-known in rough set theory.

An improved ratio of the quality of approximation can be based on P-generalized decision. The *quality of approximation based on P-generalized decision* is defined as:

$$\eta(P) = 1 - \frac{\sum_{x \in U}\left(u_P(x) - l_P(x)\right)}{(n-1) \cdot card(U)}, \tag{20}$$

where n is the number of decision classes, and it is assumed that the domain of decision criterion is numbercoded and class indices are consecutive.

It is easy to see that $\eta(P) \in [0,1]$. The ratio expresses an average relative width of P-generalized decisions of reference objects. It is resistant to local inconsistencies, i.e. inconsistencies appearing between objects with similar evaluations and assignments. In fact, this ratio is equivalent to the formulation given by Düntsch and Gediga [3], however, differently motivated.

Theorem 1. *$\eta(P)$ is equivalent to the quality of approximation*

$$\gamma_{OO}(P) = \frac{\sum_{t=2}^{n} card(\underline{P}(Cl_t^{\geq})) + \sum_{t=1}^{n-1} card(\underline{P}(Cl_t^{\leq}))}{\sum_{t=2}^{n} card(Cl_t^{\geq}) + \sum_{t=1}^{n-1} card(Cl_t^{\leq})}, \tag{21}$$

defined in [3].

Proof. Taking into account that $U = Cl_t^{\geq} + Cl_{t-1}^{\leq}$, $t = 2, \ldots, n$, $\gamma_{OO}(P)$ may be expressed as follows:

$$\gamma_{OO}(P) = \frac{\sum_{t=2}^{n} (card(\underline{P}(Cl_t^{\geq})) + card(\underline{P}(Cl_{t-1}^{\leq})))}{(n-1) \cdot card(U)}. \tag{22}$$

Further, we have:

$$\gamma_{OO}(P) = \frac{\sum_{t=2}^{n} (card(\{x \in U : l_P(x) \geq t\}) + card(\{x \in U : u_P(x) \leq t-1\}))}{(n-1) \cdot card(U)}$$
$$= \frac{\sum_{x \in U} (l_P(x) - 1 + n - u_P(x))}{(n-1) \cdot card(U)} = \frac{\sum_{x \in U} ((n-1) - (u_P(x) - l_P(x)))}{(n-1) \cdot card(U)}$$
$$= \frac{(n-1) \cdot card(U) - \sum_{x \in U} (u_P(x) - l_P(x))}{(n-1) \cdot card(U)} = 1 - \frac{\sum_{x \in U} (u_P(x) - l_P(x))}{(n-1) \cdot card(U)} \quad \square$$

An interesting interpretation of (22) is that this ratio is also the average of the quality of approximations for $n-1$ binary classification problems for consecutive unions of decision classes ($Cl_1^{\leq}$ against $Cl_2^{\geq}$, $Cl_2^{\leq}$ against $Cl_3^{\geq}$, ..., $Cl_{n-1}^{\leq}$ against $Cl_n^{\geq}$).

It is easy to see that for any $P \subseteq R \subseteq C$, there holds:

$$\eta(P) \leq \eta(R).$$

4 Quality of Approximation Based on Reassignment of Objects

The measures of approximation described above were based on the notions of lower and upper approximations of the unions of classes. The common idea behind these definitions was the fact that a decision interval for a given object $x \in U$ is calculated taking into account all the other objects from U, dominating or being dominated by x. The problem is that it is enough to introduce one more object dominating x, with the class assignment lower than x (alternatively, being dominated by x, with higher class assignment) to enlarge the decision interval, thus lowering the measures of approximation.

The key idea of the new measure is the following. The *quality of approximation based on reassignment of objects* is the minimal number of objects in U that must be reassigned to make the reference objects from U consistent, i.e. satisfying the dominance principle (1). Formally, it is defined as:

$$\zeta(P) = \frac{m - L}{m} \tag{23}$$

where L is the minimal number of objects from U that have to be reassigned consistently and $m = card(U)$. It is easy to see that $\zeta(P) \in [0, 1]$, but one can give tighter lower bound: $\zeta(P) \geq \frac{m_{\max}}{m}$, where $m_{\max}$ is the number of objects belonging to the largest class. Notice that $\zeta(P) = 1$ iff set U is consistent for $P \subseteq C$.

To compute L one can formulate a linear programming problem. Similar problem was considered in [1] in the context of specific binary classification that has much in common with multi-criteria classification. The method presented in [1] is called *isotonic separation*. Here we formulate more general problem, used for different goal (measuring the quality of approximation), however the idea behind the algorithm for finding the optimal solution remains similar.

To formulate the problem in a linear form, for each object x_i, $i \in \{1, \ldots, m\}$, we introduce $n - 1$ binary variables d_{it}, $t \in \{1, \ldots, n\}$, with the following interpretation: $d_{it} = 1$ iff object $x_i \in Cl_t^{\geq}$. Such interpretation implies the following conditions:

$$\text{if } t' > t \text{ then } d_{it'} \leq d_{it} \tag{24}$$

for all $i \in \{1, \ldots m\}$ (otherwise it would be possible that there exists object x_i belonging to the $Cl_{t'}^{\geq}$, but not belonging to $Cl_t^{\geq}$, where $t' > t$). Moreover, we give a new value of decision f_i^* to object x_i according to the rule: $f_i^* = \max_{d_{it}=1}\{t\}$ (the highest t, for which we know that x_i belongs to $Cl_t^{\geq}$).

Then, for each object $x_i \in U$ with the initial class assignment $f_i = f(x_i, d)$, the cost function can be formulated as below:

$$L(x_i) = (1 - d_{i,f_i}) + d_{i,f_i+1} \tag{25}$$

Indeed, for $t = f_i + 1$, $d_{it} = 1$ means wrong assignment (to the class higher than f_i). For $t = f_i$, $d_{it} = 0$ means also wrong assignment, to the class lower than f_i. Moreover, according to (24), only one of those conditions can appear at the same time and one of those conditions is necessary for x_i to be wrongly assigned. Thus the value of decision for x_i changes iff $L(x_i) = 1$.

According to (1), the following conditions must be satisfied for U to be consistent:

$$d_{it} \geq d_{jt} \qquad \forall i, j\colon\ x_i D_P x_j \quad 1 \leq t \leq n \tag{26}$$

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

$$\begin{array}{ll} \text{minimize} & L = \sum_{i=1}^{m} L(x_i) = \sum_{i=1}^{m} (1 - d_{i,f_i} + d_{i,f_i+1})) \\ \text{subject to} & d_{it'} \le d_{it} \qquad 1 \le i \le m, \quad 1 \le t < t' \le n \\ & d_{it} \ge d_{jt} \qquad 1 \le i, j \le m, \quad x_i D_P x_j, \quad 1 \le t \le n \\ & d_{it} \in \{0, 1\} \qquad 1 \le i \le m, \quad 1 \le t \le n \end{array} \tag{27}$$

The matrix of constraints in this case is totally unimodular, 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:

$$0 \le d_{it} \le 1 \qquad 1 \le i \le m, \quad 1 \le t \le n \tag{28}$$

and get a linear programming problem. This property was previously applied in isotonic separation method for two class problems [1]. In this paper, 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 (27),* $f_i^* = \max_{d_{it}=1}\{t\}$, *for which the following condition holds:* $l_P(x_i) \le f_i^* \le u_P(x_i)$, $1 \le i \le m$.

Proof. First, notice that all the constraints in (27) are equivalent to introducing a new (optimal) class assignment variable $f_i^* = \max_{d_{it}=1}\{t\}$ and constraints $f_i^* \ge f_j^*$ for all x_i, x_j such that $x_i D_P x_j$.

Now, assume we have an optimal solution f_i^*, $i \in \{1, \ldots, m\}$. Assume also, that for some $I \subseteq \{1, \ldots, m\}$, $f_i^* < l_P(x_i)$, $i \in I$, and for some $J \subseteq \{1, \ldots, m\}$, $f_i^* > u_P(x_i)$, $j \in J$, holds. The solution can be modified to obtain new solution $f_i^{**} = l_P(x_i)$ for $i \in I$, $f_i^{**} = u_P(x_i)$ for $i \in J$ and $f_i^{**} = f_i^*$, $i \notin I \cup J$, which will not have higher cost than f^*. We will prove that the new solution f^{**} is also feasible (i.e. satisfies all the constraints), therefore, being optimal solution of the problem (27).

Thus, we must prove that for each $x_i, x_j \in U$, the following condition holds:

$$x_i D_P x_j \Rightarrow f_i^{**} \ge f_j^{**} \tag{29}$$

The proof consist of three parts. First, we consider object x_i, where $i \in I$. Then, we take into account $i \in J$. Finally, we check the consistency for $i \notin I \cup J$.

First, notice that for all $i \in I$, $f_i^{**} > f_i^*$, and for all $i \in J$, $f_i^{**} < f_i^*$.

Consider $i \in I$. Then, (29) holds for all $j \in \{1, \ldots, m\}$, since if $j \in I$, then $f_i^{**} = l_P(x_i)$, $f_j^{**} = l_P(x_j)$, and according to the definition of $l_P(x)$ it holds that $l_P(x_i) \ge l_P(x_j)$ for $x_i D_P x_j$. If $j \notin I$, then $f_i^{**} > f_i^* \ge f_j^* \ge f_j^{**}$.

Now, consider $i \in J$. Then, (29) holds for all $j \in \{1, \ldots, m\}$, since $f_i^{**} = u_P(x_i)$, $f_j^{**} \le u_P(x_j)$, and according to the definition of $u_P(x)$, it holds that $u_P(x_i) \ge u_P(x_j)$ for $x_i D_P x_j$, so $f_i^{**} = u_P(x_i) \ge u_P(x_j) \ge f_j^{**}$.

Finally, consider $i \notin I \cup J$. Then, (29) holds for all $j \in \{1, \ldots, m\}$, since if $j \in I$, then $f_i^{**} \ge l_P(x_i) \ge l_P(x_j) = f_j^{**}$. If $j \notin I$, then $f_i^{**} = f_i^* \ge f_j^* = f_j^{**}$. Thus, we proved the theorem. □

Table 1. Example of decision table; q_1, q_2 are criteria, d is decision criterion.

U	q_1	q_2	d	U	q_1	q_2	d
x_1	23	48	4	x_7	16	10	1
x_2	44	48	4	x_8	20	30	2
x_3	45	44	2	x_9	6	14	1
x_4	26	28	3	x_{10}	9	16	1
x_5	30	26	3	x_{11}	5	9	2
x_6	24	33	3	x_{12}	15	11	1

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 $t \leq l_P(x_i)$, and to 0 for $t > u_P(x_i)$, since there exists an optimal solution with such values of the variables. In particular, if an object x is consistent (i.e. $l_P(x) = u_P(x)$), the class assignment for this object remains the same.

The introduced ratio of the quality of approximation $\zeta(P)$ satisfies also the monotonicity property, as stated by the following theorem.

Theorem 3. *For any $P \subseteq R \subseteq C$, it holds:*

$$\zeta(P) \leq \zeta(R)$$

Proof. It results from the fact that for any $P \subseteq R \subseteq C$ and any $x, y \in U$, $xD_Ry \Rightarrow xD_Py$. Thus, any constraint in the optimization problem (27) for set R must also appear in the optimization problem for set P, so the feasible region (set of solutions satisfying all the constraints) for R includes the feasible region for P. Thus, the minimum of L for R cannot be greater than the minimum of L for P. $\square$

Finally, we should notice that the measure $\zeta(P)$ is more robust to the noise than $\gamma(P)$ and $\eta(P)$. Randomly changing an assignment of an object in the decision table will not change $\zeta(P)$ by more than $\frac{1}{m}$.

In Table 1, there is an example of decision table. If we consider set $U_1 = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ with classes $\{Cl_2, Cl_3, Cl_4\}$ then we have $\gamma(P) = \frac{1}{3}$, $\eta(P) = \frac{2}{3}$, $\zeta(P) = \frac{5}{6}$. However, for the set $U_2 = \{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}$ and classes $\{Cl_1, Cl_2\}$ we have $\gamma(P) = \frac{1}{6}$, $\eta(P) = \frac{1}{6}$, but $\zeta(P) = \frac{5}{6}$. Taking into account the whole decision table $U = U_1 \cup U_2$, we obtain $\gamma(P) = \frac{1}{4}$, $\eta(P) = \frac{3}{4}$, $\zeta(P) = \frac{5}{6}$.

5 Conclusions

The paper discusses different measures of the quality of approximation in the multi-criteria classification problem. There seems to be no one best way of calculating such a coefficient from the dataset. However, each measure can be characterized by showing its advantages and drawbacks. The classical measure is simple and intuitively clear, however, for real-life data it might be too restrictive in use.

The second one, based on the generalized decision concept, measures the width of decision ranges, thus allowing some local inconsistencies with small decrease of quality of approximation. However, both may boil down to 0 only because of one object being maximally inconsistent with the rest of reference objects. The third measure based on the objects reassignment, is more robust to noise, unfortunately the coefficient cannot be given explicitly, but has to be found in result of solving an optimization problem. All the proposed measures satisfy the monotonicity property typical for rough set theory.

Acknowledgements. The first, the third and the fourth author wish to acknowledge financial support from the Ministry of Education and Science (grant no. 3T11F 02127). The research of the second author has been supported by Italian Ministry of Education, University and Scientific Research (MIUR).

References

1. Chandrasekaran, R., Ryu, Y. U., Jacob, V., Hong, S.: Isotonic separation. INFORMS J. Comput. **17** (2005) 462–474
2. Dembczyński, K., Greco, S., Słowiński, R.: Second-order Rough Approximations in Multi-criteria Classification with Imprecise Evaluations and Assignments, LNAI, **3641** (2005) 54–63
3. Düntsch, I., Gediga, G.: Approximation quality for sorting rules. Computational Statistics & Data Analysis, **40** 3 (2002) 499–526
4. Greco S., Matarazzo, B., Słowiński, R.: Rough approximation of a preference relation by dominance relations. European Journal of Operational Research, **117** (1999) 63–83
5. Greco S., Matarazzo, B., Słowiński, R.: Rough sets theory for multicriteria decision analysis. European Journal of Operational Research, **129** 1 (2001) 1–47
6. Greco S., Matarazzo, B., Słowiński, R.: Rough sets methodology for sorting problems in presence of multiple attributes and criteria. European Journal of Operational Research, **238** (2002) 247–259
7. Pawlak, Z.: Rough sets. International Journal of Information & Computer Sciences, **11** (1982) 341–356
8. Pawlak, Z., Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
9. Polkowski, L.: Rough Sets: Mathematical Foundations. Physica-Verlag, Heidelberg (2002)
10. Słowiński, R., Greco S., Matarazzo, B.: Rough Set Based Decision Support. Chapter 16 [in]: Burke, E., Kendall, G. (eds.): Introductory Tutorials on Optimization, Search and Decision Support Methodologies. Springer-Verlag, Boston (2005)