Additive Preference Model with Piecewise Linear Components Resulting from Dominance-based Rough Set Approximations

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

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

Abstract. Dominance-based Rough Set Approach (DRSA) has been proposed for multi-criteria classification problems in order to handle inconsistencies in the input information with respect to the dominance principle. The end result of DRSA is a decision rule model of Decision Maker preferences. In this paper, we consider an additive function model resulting from dominance-based rough approximations. The presented approach is similar to UTA and UTADIS methods. However, we define a goal function of the optimization problem in a similar way as it is done in Support Vector Machines (SVM). The problem may also be defined as the one of searching for linear value functions in a transformed feature space obtained by exhaustive binarization of criteria.

1 Introduction

The rough set approach has often proved to be an interesting tool for solving a classification problem that consists in an assignment of objects from set A, described by *condition attributes*, to pre-defined *decision classes* Cl_t, where $t \in T$ and T is a finite set of numerically coded labels. In order to solve the problem (i.e., to classify all objects from A), a decision rule model is induced from a set of reference (training) objects $U \subset A$. The rough set analysis starts with computing lower and upper rough approximations of decision classes. The lower approximation of a decision class contains objects (from U) *certainly* belonging to the decision class without any inconsistency. The upper approximation of a decision class contains objects *possibly* belonging to the decision class that may cause inconsistencies. In the simplest case, the inconsistency is defined as a situation where two objects described by the same values of condition attributes (it is said that these objects are indiscernible) are assigned to different classes. In the next step, decision rules are induced from lower and upper rough approximations. These rules represent, respectively, certain and possible patterns explaining relationships between conditions and decisions. The model in the form of decision rules permits to classify all objects from A.

Greco. Matarazzo and Słowiński [5, 6, 13] have introduced a rough set approach (called Dominance-based Rough Set Approach — DRSA) for solving

the problem of multi-criteria classification. In this problem, it is additionally assumed that the domains of attributes (scales) are preference-ordered. The decision classes are also preference-ordered according to an increasing order of class labels, 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 condition attributes are often referred to as *condition criteria*. DRSA extends the classical approach by substituting the indiscernibility relation by a dominance relation, which permits taking into account the preference order. The inconsistency is defined in view of a dominance principle that requires that any object x, having not worse evaluations than any object y on the considered set of criteria, cannot be assigned to a worse class than y. Moreover, unlike in the classical rough set approach, there is no need in DRSA to make discretization of numerical attributes.

The preference model is a necessary component of a decision support system for multi-criteria classification. Construction of preference model requires some *preference information* from the Decision Maker (DM). Classically, these are substitution rates among criteria, importance weights, comparisons of lotteries, etc.. Acquisition of this preference information from the DM is not easy. In this situation, the preference model induced from decision examples provided by the DM has clear advantages over the classical approaches. DRSA, but also UTA [8] and UTADIS [9, 15], follows the paradigm of inductive learning (in Multi-Criteria Decision Analysis referred to as a preference-disaggregation approach). It is very often underlined by Słowiński, Greco and Matarazzo (see, for example [13]) that a decision rule model has another advantage over other models, i.e. it is intelligible and speaks the language of the DM. However, in the case of many numerical criteria and decision classes, the set of decision rules may be huge and may loose its intelligibility. In such situations, an additive model composed of marginal value (utility) functions, like in UTA and UTADIS, may be helpful. The marginal value functions are usually presented graphically to the DM in order to support her/his intuition.

In the following, we present an extension of DRSA, where after computing rough approximations, additive value functions are constructed instead of a set of decision rules. The additive value function is composed of piecewise linear marginal value functions. Its construction is proceeded by solving a problem of mathematical programming similar to that formulated in UTA and UTADIS. The main difference is that we define a goal function of the optimization problem in a similar way as it is done in Support Vector Machines (SVM) [14]. However, the obtained additive value functions, for lower and upper rough approximations of decision classes, may not cover accurately all objects belonging to corresponding rough approximations. It is caused by a limited capacity of an additive model based on piecewise linear functions to represent preferences as proved in [7, 12]. The problem may be also defined as the one of searching linear value functions in a transformed feature space obtained by exhaustive binarization of criteria.

The paper is organized as follows. In Section 2, DRSA involving piecewise linear marginal value functions is presented. Section 3 contains first experimental results of the methodology. The last section concludes the paper.

Table 1. Decision table: q_1 and q_2 indicate criteria, d class label. The last two columns present range of generalized decision function; objects x_2 and x_3 are inconsistent.

U	q_1	q_2	$d(x)$	$l(x)$	$u(x)$
x_1	0.25	0.3	-1	-1	-1
x_2	0.5	0.65	1	-1	1
x_3	0.75	0.7	-1	-1	1
x_4	1	0.6	1	1	1

2 Piecewise Linear Marginal Value Functions and Dominance-based Rough Set Approach

Assume, we have a set of objects A described by n criteria. We assign to each object x a vector $\mathbf{x} = (q_1(x), \ldots, q_n(x))$, where i-th coordinate $q_i(x)$ is a value (evaluation) of object x on criterion q_i, $i = 1, \ldots, n$. For simplicity, it is assumed that domains of criteria are numerically coded with an increasing order of preference. The objective of multi-criteria classification problem is to build a preference model, according to which a class label $d(x)$ from a finite set T is assigned to every object from A. Here, for simplicity, it is assumed that $T = \{-1, 1\}$. It corresponds to that the objects from Cl_1 are strictly preferred to the objects from Cl_{-1}. We assume that the DM provides a preference information concerning a set of *reference objects* $U \subset A$, assigning to each object $x \in U$ a label $d(x) \in T$. Reference objects described by criteria and class labels are often presented in the *decision table*. An example of the decision table is presented in Table 1.

The criteria aggregation model (preference model) is assumed to be additive value function:

$$\Phi(x) = \sum_{i=1}^{n} w_i \phi_i(q_i(x)) \tag{1}$$

where $\phi_i(q_i(x))$, $i = 1, \ldots, n$, are non-decreasing marginal value functions, normalized between 0 and 1, w_i is a weight of $\phi_i(q_i(x))$. A similar aggregation model with was used in [8] within UTA method (for ranking problems) and UTADIS [9, 15] (for multi-criteria classification problems), where marginal value functions were assumed to be piecewise linear. The use of this aggregation model for classification requires existence of threshold ϕ_0, such that $d(x) = 1$ if $\Phi(x) \geq \phi_0$ and $d(x) = -1$ otherwise (so $d(x) = \mathrm{sgn}(\Phi(x) - \phi_0)$). The error, which is the sum of differences $|\Phi(x) - \phi_0|$ of misclassified objects is minimized.

Assume however, that objects can be *inconsistent*. By inconsistency we mean violation of the *dominance principle*, requiring that any object x, having not worse evaluations than any object y on the considered set of criteria, cannot be assigned to a worse class than y. If such inconsistencies occur, the UTA method is not able to find any additive value function compatible with this information, whatever the complexity of the marginal functions (number of breakpoints) is, since none monotonic function can model this information. Within DRSA, such inconsistencies can be handled by using concepts of lower and upper approxi-

mations of classes. It was shown [3] that it corresponds to generalized decision function δ for an object $x \in U$:

$$\delta(x) = \langle l(x), u(x) \rangle \tag{2}$$

where

$$l(x) = \min\{d(x) : yDx, y \in U\} \quad u(x) = \max\{d(x) : xDy, y \in U\} \tag{3}$$

where D is a dominance relation defined as $xDy \Leftrightarrow \forall_{i \in \{1,\ldots,n\}} q_i(x) \geq q_i(y)$. In other words, given the preference information, for object x there is determined a range of decision classes to which x may belong. This range results from taking into account inconsistencies caused by x. Remark that without inconsistencies, for all $x \in U$, $l(x) = u(x)$. Moreover, if we assign to each $x \in U$ a class index $l(x)$ (instead of $d(x)$), the decision table becomes consistent (similarly if we assign a class index $u(x)$ for all $x \in U$). Thus one can deal with inconsistent set U, by considering two consistent sets with two different labelings. The values of generelized decision function are also presented in Table 1. In terms of further classification, the response of such model is a range of classes, to which an object may belong.

For the two consistent decision tables it is possible to derive compatible value functions $\Phi^l(x)$, $\Phi^u(x)$ respectively, and corresponding marginal value functions $\phi_i^l(q_i(x))$ and $\phi_i^u(q_i(x))$ We assume that both $\phi_i^l(q_i(x))$ and $\phi_i^u(q_i(x))$ have piecewise linear form:

$$\phi_i(q_i(x)) = \sum_{r=1}^{k-1} c_i^r + \frac{c_i^k}{h_i^k - h_i^{k-1}}(q_i(x) - h_i^{k-1}), \quad \text{for } h_i^{k-1} \leq q_i(x) \leq h_i^k \tag{4}$$

where h_i^k is the location of the k-th brakepoint on the i-th criterion ($k = 1, \ldots, \kappa_i$, where κ_i is a number of brakepoints on i-th criterion), and c_i^k is an increment of marginal value function between brakepoints h_i^{k-1} and h_i^k, i.e. $\phi_i(h_i^k) - \phi_i(h_i^{k+1})$. Equation (4) states that function ϕ_i evaluated at $q_i(x)$ equals to the sum of increments on all intervals on the left of $q_i(x)$ and linearly approximated increment in the interval where $q_i(x)$ is located. The example is shown on Figure 1.

In the simplest form, the corresponding optimization problem can be formulated for lower bound of generalized decision ($\Phi_i^l(x)$) as follows:

$$\min\colon \sum_{j=1}^{m} \sigma_j^l \tag{5}$$

subject to constraints:

$$\Phi^l(x_j) \leq \phi_0^l + \sigma_j^l \qquad \forall x_j\colon l(x_j) = -1 \tag{6}$$

$$\Phi^l(x_j) \geq \phi_0^l - \sigma_j^l \qquad \forall x_j\colon l(x_j) = 1 \tag{7}$$

$$\sigma_j^l \geq 0 \qquad \forall x_j \tag{8}$$

$$\phi_i^l(z_i^*) = 1 \qquad \forall i \in \{1, \ldots, n\} \tag{9}$$

$$\phi_i^l(z_{i*}) = 0 \qquad \forall i \in \{1, \ldots, n\} \tag{10}$$

$$c_i^k \geq 0 \quad \forall i \in \{1, \ldots, n\}, k \in \{1, \ldots, \kappa_i\} \tag{11}$$

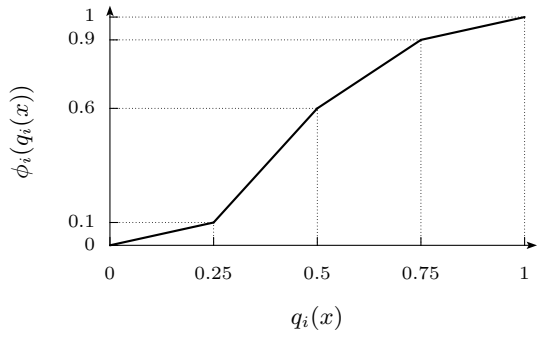

Fig. 1. Piecewise linear marginal value function ϕ_i defined by Equation 4. The increments are: $c_i^1 = 0.1$, $c_i^2 = 0.5$, $c_i^3 = 0.3$, $c_i^4 = 0.1$.

where m is the number of reference objects, σ_j are possible errors, z_i^* and z_{i*} are the highest and the lowest value on i-th criterion. Constraints (6) and (7) ensure correct separation, (9) and (10) control scaling and (11) preserves monotonicity. Analogous problem can be formulated for upper bound of generalized decision (function $\Phi^u(x)$). It is worth noting, that the method does not assure all errors become zero as the complexity of $\phi_i^l(q_i(x))$ and $\phi_i^u(q_i(x))$ grows, however, it avoids errors caused by inconsistencies. If all σ_i^l and σ_i^u become 0, the obtained model is concordant with DRSA in the sense that all objects belonging to lower or upper approximations will be reassigned by the obtained functions to these approximations.

It is worth introducing some measure of complexity of marginal functions and minimize it, to avoid building complex models. Notice, that as the function $\phi_i(q_i(x))$ is multiplied in (1) by weight w_i, we can introduce new coefficients $w_i^k = c_i^k w_i$ in order to keep the problem linear. Now control of the complexity is done by minimizing a regularization term:

$$||\mathbf{w}||^2 = \sum_{i=1}^{n} \sum_{k=1}^{\kappa_i} (w_i^k)^2 \tag{12}$$

instead of controlling the scale of functions in (9) and (10). Minimizing of term may lead to rescaling the utility function, so that all values of $\phi_i(q_i(x))$ will decrease down almost to zero. To avoid that, constraints are modified introducing the unit margin around threshold, in which no object may appear without error. Thus we rewrite equations (6) and (7) as:

$$(\Phi^l(x_j) - \phi_0^l) l(x_j) \geq 1 - \sigma_i^l \tag{13}$$

The objective of the optimization is now:

$$\min\colon ||\mathbf{w}^l||^2 + C \sum_{j=1}^{m} \sigma_j^l \tag{14}$$

where C is the complexity constant. Analogous reasoning may be proceeded for $\Phi^u(x)$. Such problem resembles maximal margin classifier and Support Vector Machines [14]. We will try to bring it even more similar.

Let us first modify $\Phi(x)$ to be $\Phi(x) = \sum_{i=1}^{n} w_i \phi_i(q_i(x)) - \phi_0$. Now, we decompose each function $\phi_i(q_i(x))$ into κ_i functions $\phi_{i,k}(q_i(x))$ in the following way:

$$\phi_i(q_i(x)) = \sum_{k=1}^{\kappa_i} c_k^i \phi_{i,k}(q_i(x)). \tag{15}$$

An example of such decomposition is shown on Figure 2. One can treat the family of functions $\{\phi_{1,1}(q_1(x)), \ldots, \phi_{n,\kappa_n}(q_n(x))\}$ as a transformation of the space of criteria. Namely, there is a map $\mathcal{T}\colon A \to \mathbb{R}^s$ where $s = \sum_{i=1}^{n} \kappa_i$, such that $\mathcal{T}(\mathbf{x}) = (\phi_{1,1}(q_1(x)), \ldots, \phi_{n,\kappa_n}(q_n(x)))$. By substituting $w_i^k = w_i c_i^k$ and denoting $\mathbf{w} = (w_1^1, \ldots, w_n^{\kappa_n})$, the function (1) becomes:

$$\Phi(x) = \langle \mathbf{w}, \mathcal{T}(\mathbf{x}) \rangle - \phi_0 \tag{16}$$

where $\langle \cdot, \cdot \rangle$ is a canonical dot product in $\mathbb{R}^s$.

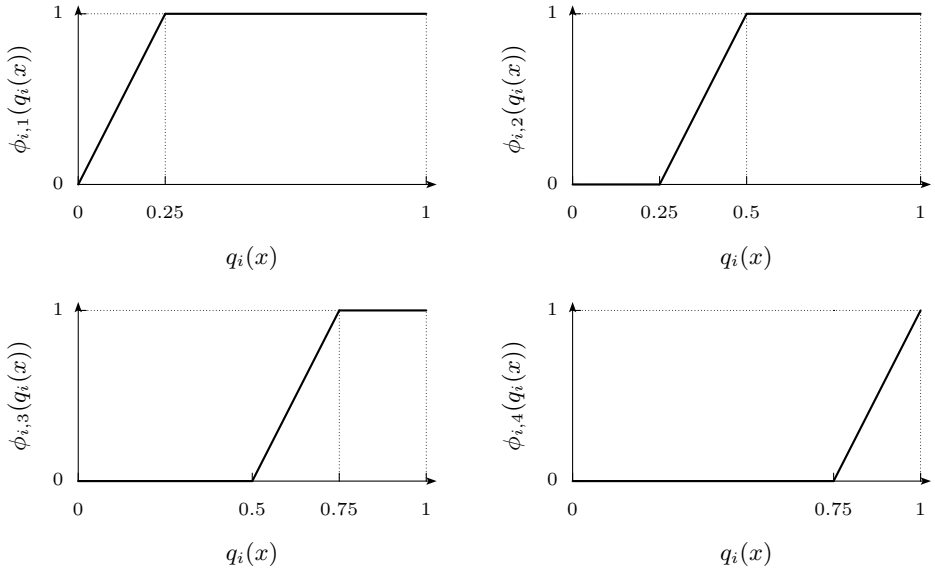

Fig. 2. Functions obtained by decomposing $\phi_i(q_i(x))$. Notice that $\phi_i(q_i(x) = 0.1\phi_{i,1}(q_i(x)) + 0.5\phi_{i,2}(q_i(x)) + 0.3\phi_{i,3}(q_i(x)) + 0.1\phi_{i,4}(q_i(x))$.

Finally, after reformulating the problem we obtain:

$$\min\colon ||\mathbf{w}^l|| + C\sum_{j=1}^{m}\sigma_j^l \tag{17}$$

subject to constraints:

$$(\langle \mathbf{w}^l, \mathcal{T}^l(\mathbf{x})\rangle - \phi_0^l)l(x) \geq 1 - \sigma_i^l \qquad \forall x_i \in U \tag{18}$$

$$\sigma_i^l \geq 0 \qquad \forall i \in \{1,\ldots,m\} \tag{19}$$

$$w_i^k \geq 0 \qquad \forall i \in \{1,\ldots,n\}, k \in \{1,\ldots,\kappa_i\} \tag{20}$$

Notice that the regularization term is just squared Euclidean norm of the weight vector in new feature space.

Motivated by the above result, we introduce a *kernel function* $k\colon A\times A \to \mathbb{R}$, defined as:

$$k(x,y) = \sum_{i=1}^{n} k_i(x,y) \tag{21}$$

where

$$k_i(x,y) = \sum_{j=1}^{\kappa_i} \phi_{i,j}(q_i(x))\phi_{i,j}(q_i(y)). \tag{22}$$

Notice, that $k(x,y) = \langle \mathcal{T}(\mathbf{x}), \mathcal{T}(\mathbf{y})\rangle$. Assume that we have a brakepoint in each evaluation point of all objects $x \in U$, the set of brakepoints for i-th criterion is $\{q_i(x_1),\ldots,q_i(x_m)\}$. Then the computing of the marginal kernel function $k_i(x,y)$ boils down to:

$$k_i(x,y) = \min\{rank_i(x), rank_i(y)\} \tag{23}$$

where $rank_i(x)$ is a position (in ranking) of value $q_i(x)$ on i-th criterion.

Thus, the problem may be formulated in a dual form. The most essential advantage of such approach is reduction in number of variables, irrespective to the number of brakepoints of marginal functions. As the complexity of marginal functions increases, the optimization problem remains the same and only the computation of the kernel function becomes harder. However there is a problem, how to ensure monotonicity of the resulting utility function. In the dual formulation the information about each criterion is lost, thus not all the weights may be non-negative.

Let us remark that the transformed criteria space obtained by image of mapping $\mathcal{T}(A)$ (i.e. $(\phi_{1,1}(q_1(x)),\ldots,\phi_{n,\kappa_n}(q_n(x))), x \in A$) may be also seen as a result of binarization of criteria. This type of binarization should be called an exhaustive one by analogy to other approaches well-known in rough set theory or in logical analysis of data (see for example, [1]).

The exhaustive binarization is proceeded by choosing cut points in each evaluation point of all objects $x \in U$. More precisely, the binarization of the i-th criterion is accomplished in a straightforward way by associating with each value

Table 2. Decision table from Table 1 with binarized criteria

U	$q_{1\,0.25}$	$q_{1\,0.5}$	$q_{1\,0.75}$	$q_{1\,1}$	$q_{2\,0.3}$	$q_{2\,0.6}$	$q_{2\,0.65}$	$q_{2\,0.7}$	d
x_1	1	0	0	0	1	0	0	0	-1
x_2	1	1	0	0	1	1	1	0	1
x_3	1	1	1	0	1	1	1	1	-1
x_4	1	1	1	1	1	1	0	0	1

v on this criterion, for which there exists an object x, such that $q_i(x) = v$, a boolean attribute q_{i_v} such that:

$$q_{i_v}(x) = \begin{cases} 1 \text{ if } q_i(x) \geq v \\ 0 \text{ if } q_i(x) < v \end{cases}. \tag{24}$$

Table 2 shows the exhaustive binarization of criteria from Table 1.

Moreover, let us remark that the binarized decision table contains almost the same information as dominance matrix introduced in [2]. The dominance matrix DM is defined as follows:

$$DM = \{dm(x,y) : x, y \in U\}, \text{ where } dm(x,y) = \{q_i \in Q : q_i(x) \geq q_i(y)\}. \tag{25}$$

where $Q = \{q_i, i = 1, \ldots, n\}$. Dominance Matrix DM is usually implemented as 3-dimensional binary cube $\mathbf{C}$ defined as $c_{jki} = 1$, if $q_i \in dm(x_j, x_k)$, and $c_{jki} = 0$ otherwise, where $j, k = 1, ..., m$ and $i = 1, \ldots, n$. Such a structure is very useful in a procedure of generating exhaustive set of decision rules [2], because all computations may be quickly proceeded as bitwise operations. It is a counterpart of a discernibility matrix [11] well-known in classical rough set approach. It is easy to see that the following occurs:

$$c_{jki} = 1 \Leftrightarrow q_{i_{q_i(x_k)}}(x_j) = 1, \quad x_j, x_k \in U.$$

3 Experimental results

We performed a computational experiment on Wisconsin breast cancer (BCW) data obtained from the UCI Machine Learning Repository [10]. This problem was chosen since it is known to have monotonic relationship between values on condition attributes and decision labels. Thus, all attributes can be interpreted as criteria, enabling DRSA. BCW consist of 699 instances described by 9 integer-valued attributes, from which 16 instances have missing values. Each instance is assigned to one of two classes (malignant and benign).

Several approaches have been compared with the methodology presented in the previous section that will be referred to as Piecewise Linear DRSA (PL-DRSA). These are k-Nearest Neighbours, linear Support Vector Machines, Logistic Regression, J48 Decision Trees and Naive Bayes. WEKA [4] software was used for the experiment. For all algorithms a criteria selection was performed, by

Table 3. Experimental results for Wisconsin breast cancer data.

Algorithm	Number of criteria	loo estimate
k-NN (k = 1)	6	96.8%
linear SVM	6	97.2%
J48	4	97.7%
Logistic Regr.	6	97.2%
Naive Bayes	6	97.1%
PL-DRSA	6	97.4%

using backward elimination. The number of criteria left and the leaving-one-out (loo) accuracy estimate are shown in Table 3.

Apparently, all the accuracy estimates are similar. PL-DRSA was conducted with setting breakpoints on each evaluation point of all objects on each criterion. Two models were created, one for lower bound of the decision range, second for the upper bound. The classification rule was the following: if $\Phi^l(x) - \Phi^u(x) \geq 0$ then assign x to class Cl_1 otherwise assign x to Cl_{-1}. The achieved accuracy was one of the best, competitive to other methods. However, marginal value functions constructed in PL-DRSA may be presented graphically and easily interpreted. Moreover, PL-DRSA shows the inconsistent data both in learning and classification stage.

4 Conclusions

Within DRSA framework, the decision rule model were always considered for multicriteria decision analysis. We presented an alternative method, related to additive aggregation model, similar to the one used in the UTA method. The described approach has several advantages. First, it is flexible and allows various shapes of separating function to be obtained. Marginal value functions may also be presented graphically and interpreted by DM. Finally, PL-DRSA can control the complexity of the additive function by fixing the number of breakpoints and minimizing the slopes in each breakpoint. DRSA plays important role in handling inconsistencies, which affect the data. Ignoring them may cause errors and, therefore generate wrong decision model. The method can be also interpreted in terms of criteria transformation (in a specific case, also in terms of binarization) and Support Vector Machines.

Acknowledgements. The authors wish to acknowledge financial support from the Ministry of Education and Science (grant no. 3T11F 02127).

References

1. Boros, E., Hammer, P. L., Ibaraki, T., Kogan, A., Mayoraz, E., Muchnik, I.: An Implementation of Logical Analysis of Data. IEEE Trans. on Knowledge and Data Engineering **12** (2000), 292–306

2. Dembczyński, K., Pindur, R., Susmaga R.: Generation of Exhaustive Set of Rules within Dominance-based Rough Set Approach. Electr. Notes Theor. Comput. Sci, 82 **4** (2003)
3. Dembczyński, K., Greco, S., Słowiński, R.: Second-order Rough Approximations in Multi-criteria Classification with Imprecise Evaluations and Assignments. LNAI **3641** (2005) 54–63
4. Frank, E., Witten, I. H.: Data Mining: Practical machine learning tools and techniques, 2nd Edition. Morgan Kaufmann, San Francisco, (2005)
5. Greco S., Matarazzo, B. and 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. and Słowiński, R.: Rough sets theory for multicriteria decision analysis. European Journal of Operational Research, **129** (2001) 1–47
7. Greco S., Matarazzo, B. and Słowiński, R.: Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. European Journal of Operational Research, **158**, (2004) 271–292
8. Jacquet-Lagréze, E., Siskos, Y.: Assessing a set of additive utility functions for multicriteria decision making: The UTA method. European Journal of Operational Research, **10** (1982) 151–164
9. Jacquet-Lagréze, E.: An application of the UTA discriminant model for the evaluation of R&D projects, In Pardalos, P.M., Siskos, Y., Zopounidis, C., (eds.): Advances in multicriteria analysis. Kluwer Academic Publishers, Dordrecht (1995) 203–211
10. Newman, D.J., Hettich, S., Blake, C.L., Merz, C.J.: UCI Repository of machine learning databases, `http://www.ics.uci.edu/~mlearn/MLRepository.html`. Irvine, CA: University of California, Department of Information and Computer Science (1998)
11. Skowron A., Rauszer C.: The discernibility matrices and functions in information systems. In Slowinski R. (ed.): Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory. Kluwer Academic Publishers, Dordrecht (1992) 331–362
12. Słowiński, R., Greco, S., Matarazzo, B.: Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle. Control & Cybernetics **31** (2002) 1005–1035
13. 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)
14. Vapnik, V.: The Nature of Statistical Learning Theory, New York, Springer-Verlag (1995)
15. Zopounidis, C., Doumpos, M.: PREFDIS: a multicriteria decision support system for sorting decision problems, Computers & Operations Research **27** (2000) 779–797