Incremental rule induction for multicriteria and multiattribute classification

Jerzy Stefanowski and Marcin Żurawski

Institute of Computing Science, Poznań University of Technology, 60-965 Poznań, Poland. E-mails *jerzy.stefanowski@cs.put.poznan.pl, mzurawski@mikrob.pl*

Abstract. Incremental induction of decision rules within the dominance-based rough set approach to the multicriteria and multiattribute classification is discussed. We introduce an algorithm, called Glance, that incrementally induces a set of all rules from examples. We experimentally evaluate it and compare with two other algorithms, previously proposed for this kind of classification problem.

1 Introduction

Induction of decision rules from provided learning examples is one of the main problems considered in machine learning and knowledge discovery. Several algorithms for rule induction have already been proposed for various types of *attributes* (nominal, numerical, etc.) describing examples, see e.g. reviews in [4,7]. However, in some problems one can meet yet another type of attributes with *preference order* in their domains. For example, when considering buying a car, an attribute like "fuel consumption" has a clear preference ordered scale - the less, the better. The attributes with preference ordered domains are called *criteria*. Let us notice that this kind of semantic information is often present in data related to cost, gain or quality of objects like in economical data; however it is neglected by typical knowledge discovery tools. On the other hand, consideration of criteria is the key aspect of Multiple-Criteria Decision Analysis (MCDA) [1,6].

In this paper we consider one of the major MCDA problems called *multiple-criteria classification* (or sorting) *problem*. It concerns an assignment of objects evaluated by a set of criteria into pre-defined, and also preference-ordered, decision classes. When solving such a problem, any reasonable regularities to be discovered from data have to take into account a *dominance principle*. This principle requires that an object x having better, or at least the same, evaluations on a set of criteria than object y cannot be assigned to the worse class than object y. Some data may contain objects violating this principle. This consitutes a kind of *inconsistency* that should be handled in a proper way from MCDA point of view. Therefore, Greco, Matarazzo and Slowinski [1] have proposed an extension of the rough sets theory [5] that is able to deal with this kind of inconsistency. It is called Dominance-based Rough Set Approach (DRSA). Decision rules, which are induced from examples within DRSA framework, have a special syntax, that requires a new type of algorithms for their induction.

Greco *et all* introduced an algorithm DOMLEM [3], which allows to induce a minimal set of DRSA rules covering all examples in the input data. On the other hand, sometimes it is also interesting to induce other sets of the rule, i.e. a set of all rules which can be induced from the given data or a subset of all rules which satisfy user predefined requirements concerning, e.g. minimal number of supporting objects or maximal length of the condition part. Two algorithms for inducing such sets of rules have already been discussed in [7]. However, these algorithms require all examples to be read into memory before induction. Experiments have also showed that they are efficient rather on smaller data sets. Hence, these algorithms are not suitable for handling large amounts of examples or for situations where only the part of data is available at the begining and other parts are provided later. Thus, we are interested in an incremental processing of examples while inducing rules.

The aim of this paper is to present a new algorithm for inducing dominance-based decision rules for the multicriteria and multiattribute classification problem. It is called "Glance" and induces all rules in an incremental way. Moreover, we experimentaly evaluate it on several data sets and compare its efficiency with other, previously proposed, algorithms.

2 Dominance based decision rules

Selected concepts of DRSA are presented, for more details see [1]. Let us assume that learning examples are represented in *decision table* $DT = (U, A \cup \{d\})$, where U is a set of examples (objects), A is a set of *condition attributes* describing examples, V_a is a domain of a. Let $f(x, a)$ denote the value of attribute $a \in A$ taken by object $x \in U$. The domains of attributes may be preference–ordered or not - in the first case the attributes are called *criteria* while in the latter case we call them *regular attributes*. Preference scale of each criterion a induces a *complete preorder* (i.e. a strongly complete and transitive binary relation) $\succeq$ in set U. For objects x and y, $\succeq$ means that object x is "*at least as good*" as y with respect to criterion a. The asymmetric part of $\succeq$ is denoted by $\succ$. An object x is preferred to object y on criterion a if $a(x) \succ a(y)$ and x is indiscernible to y if $a(x) \sim a(y)$.

$d \notin A$ is a decision attribute that partitions a set of examples into k decision classes $\{Cl_t : t = 1, \ldots, k\}$. These classes are preference ordered, i.e. for all r, s such that $s > r$, objects from Cl_s are prefered to objects from Cl_r. In multicriteria classfication problems it is typical to consider *upward union* and *downward union* of classes instead of single decision classes. The upward and downward unions are defined, respectively, as: $Cl_g^{\geq} = \bigcup_{h \geq g} Cl_h$, $Cl_g^{\leq} = \bigcup_{f \leq g} Cl_f$, $g = 1, \ldots k$. The statement $x \in Cl_g^{\geq}$ means "x belongs to at least class Cl_g", while $x \in Cl_g^{\leq}$ means "x belongs to at most class Cl_g.

The rough set theory [5] requires determining relation between objects. For each criterion the *dominance relation* is defined; Let $P^{\succeq}$ be a subset of criteria from A; the object x *dominates* object y with respect to subset $P^{\succeq}$

iff $a(x) \succeq a(y)$ for all $a \in P^{\succeq}$. For each regular attribute from $P^{=} \subset A$ there exists *indiscernible relation*, which is an equivalence relation as in the original rough sets theory. For any subset $P \subseteq A$ two sets of objects defined: $R_P^+(x)$ and $R_P^-(x)$. Given object $x \in U$, $R_P^+(x)$ is the set of all objects $y \in U$ which dominates x with respect to $P^{\succeq}$ (criteria of P) and are indiscernible with x with respect to $P^{=}$ (attributes of P). Analogously, $R_P^-(x)$ is the set of all objects $y \in U$ which are dominated by x with respect to $P^{\succeq}$ and are indiscernible with x with respects to $P^{=}$. If the decision table contains *inconsistent* objects, the sets $R_P^+(x)$ and $R_P^-(x)$ can be used to define lower and upper approximations of upward unions $Cl_g^{\geq}$ and downward unions $Cl_g^{\leq}$, respectively. For instance, the lower approximation of $Cl_g^{\geq}, g \leq k$ is the set $\{ x \in U : R_P^+(x) \subseteq Cl_q^{\geq} \}$, i.e. it contains all objects belonging to $Cl_q^{\geq}$ without ambiguity with respect to their descriptions.

Examples belonging to rough approximations of upward and downward unions of classes are used for induction of "$if \ldots then \ldots$" decision rules. The two kinds of these rules are distinguished:
$D_{\succeq}$–decision rules with the following syntax: *if* $(f(x, a_1) \succeq r_{a_1}) \wedge \ldots \wedge (f(x, a_m) \succeq r_{a_m}) \wedge f(x, a_{m+1}) \sim r_{a_{m+1}}) \wedge \ldots \wedge (f(x, a_p) \sim r_{a_p})$ *then* $x \in Cl_g^{\geq}$,
$D_{\preceq}$–decision rules with the following syntax: *if* $(f(x, a_1) \preceq r_{a_1}) \wedge \ldots \wedge (f(x, a_m) \preceq r_{a_m}) \wedge f(x, a_{m+1}) \sim r_{a_{m+1}}) \wedge \ldots \wedge (f(x, a_p) \sim r_{a_p})$ *then* $x \in Cl_g^{\leq}$,
where $P = \{a_1, \ldots, a_p\} \subseteq A, P^{\succeq} = \{a_1, \ldots, a_m\}, P^{=} = \{a_{m+1}, \ldots, a_p\}$, $(r_1, \ldots, r_p) \in V_{a1} \times V_{a2} \times V_{ap}$ and $g = 1, \ldots k$.

The $D_{\succeq}$–rule says: "if an evaluation of object x on criteria a_i is at least as good as a threshold value r_i $(i = 1, \ldots, m)$ and on attribute a_j object x is indiscernible with value r_j $(i = m+1, \ldots, p)$ then object x belongs to a least class Cl_g". Similarly the $D_{\preceq}$–decision rule means that an object is evaluated as "at most good as a value" and belongs to at most a given class.

Each dominance based decision rule has to be *minimal.* Since a decision rule is an implication, by a minimal decision rule we understand such an implication that there is no other implication with an antecedent of at least the same weakness (in other words, rule using a subset of elementary conditions or/and weaker elementary conditions) and a consequent of at least the same strength (in other words, rule assigning objects to the same union or sub-union of classes).

Consider a $D_{\succeq}$–decision rule *if* $(f(x, a_1) \succeq r_{a_1}) \wedge \ldots \wedge (f(x, a_m) \succeq r_{a_m}) \wedge f(x, a_{m+1}) \sim r_{a_{m+1}}) \wedge \ldots \wedge (f(x, a_p) \sim r_{a_p})$ *then* $x \in Cl_g^{\geq}$. If there exists an object y such that $f(y, a_1) \sim r_{a_1} \wedge \ldots \wedge (f(y, a_m) \sim r_{a_m} \wedge \ldots \wedge f(y, a_p) \sim r_{a_p}$, then y is called basis of the rule. Each $D_{\succeq}$–rule having a basis is called *robust* because it is "founded" on an object existing in the data table. Analogous definition of robust decision rules holds for the other types of rules.

We say that an object supports a decision rule if it matches both condition and decision parts of the rule. On the other hand, an object is covered by a decision rule if it matches the condition part of the rule. The rule can be described by a parameter *strength*, which is the number of supporting objects.

The set of induced rules is *complete*, if is covers all objects from the data table in such a way that objects belonging to lower approximations of unions are re-assigned to their original class while inconsistent objects are assigned to cluster of classes referring to their inconsistency [1]. The set of rules is called *minimal* if it is a set of minimal rules that is complete and non-redundant, i.e. exclusion of any rule from it makes it incomplete.

3 "Glance" algorithm

The minimal set of rules can be induced from examples of multicriteria and multiattribute classification problem by means of the *DOMLEM* algorithm [3,7]. Such a set of rules is usually sufficient for aims of *predicting classification* of new (or testing) objects. However, if the aim of the induction process is *descriptive*, i.e. discovered rules should help in explaining or better understanding of circumstances in which decisions were made, another kind of rule sets are also useful [7]. One possibility is to create all decision rules, which could be generated from the given decision table, while another option is to discover the satisfactory subset of all rules, which satisfy user's requirements e.g. sufficiently high strength and confidence, contain limitted number of elementary conditions. The algorithms for inducing such sets of dominance based rules were discussed in [7,8]. Let us notice however, that their computational costs are higher than for DOMLEM; in particular it is true in case of looking for all rules, which is a problem characterized by an exponential complexity with respect to a number of attributes. Moreover, these algorithms require all examples to be read into memory before induction – what is a serious drawback if knowledge is to be discovered from real databases. Let us also remind that some learning problems are characterized by an incremental processing of information, i.e. descriptions of examples are available sequentially in steps. To overcome all of these problems we have decided to consider *incremental learning*, i.e. the learning algorithm has to learn rules from provided examples and then refine knowledge representation when new examples become available. The paradigm of incremental learning has been already considered in the field of machine learning or knowledge discovery from databases, however we have to skip the review of existing approaches due to a limited size of the paper.

We introduce an algorithm called *Glance*, which is incremental and stores in memory only rules and not descriptions of learning examples. In its basic version, it induces the set of all rules from examples of multiattribute and multicriteria classification problems. These rules are not necessarily based on some particular objects, i.e. they are non-robust. The general scheme of the algorithm for the case of inducing certain rules is presented below.

Procedure GLANCE
(**input** U - set of examples; P - set of criteria and regular attributes;
Cl - set of unions of decision classes; **output R** set of decision rules);

begin
    for each union $u \in Cl$ **do**
    begin
        Let r be a rule with empty condition part; {it covers each example}
        $R_u \leftarrow r$; {add to the set of rules for union u}
    end{for}
    for each example $x \in U$ **do**
        for each u such that $f(x,d) \notin u$ **do** {d is a decision attribute}
            for $r \in R_u$ **do**
                if (r covers x) **then begin** {r does not discriminate properly}
                $R_u \leftarrow R_u \setminus r$;
                for each $a \in P$ **do**
                    begin { specialize rules }
                        Let s be a condition on attr. a excluding x;
                        $r_{new} \leftarrow r \cup s$;
                        if (r_{new} is minimal) **then** $R_u \leftarrow R_u \cup r_{new}$;
                    end; {for}
                end; {if }
    $\mathbf{R} \leftarrow \bigcup_u R_u$;
end{procedure}

The main idea of *Glance* is concordant with inductive extension principle, according to which an algorithm starts from the most general description and then update it (specialize) so that it is consistent with available training data (also considering approximations). The most general description are rules with empty condition part. When the new object, e.g. $x \in Cl_g$ is available, it is determined for which unions u this object should be treated as a negative example (these unions include all $Cl_f^{\leq}$ such that $f < g$ for downward cumulated unions and $Cl_h^{\geq}$ such that $h > g$ for upward cumulated unions). If any rule $r \in R_u$ covers x, then it has to be updated as it does not discriminate all negative examples properly. It is removed from R_u and specialized by adding on each attribute or criterion an elementary condition that its not satisfied by description of x. For criteria elementary conditions are in forms $(f(x,a) \succ v_a)$ or $(f(x,a) \prec v_a)$, depending on the direction of preference and unions; where v_a is the value of criterion a for object x. For regular attributes, conditions are in form $(f(x,a) \notin \{v_1, \ldots, v_l\}$ and v_a is added to this list. At the end of the induction process, these conditions in all rules are transformed to representation using operators $(\preceq, \succeq, \sim)$ as in the syntax of dominance based rules.

The computational complexity of basic version of *Glance* is exponential with respect to a number of attributes. The user can define the maximal accepted number of elementary conditions in the syntax of the rule. An extended version of *Glance*allows to induce a satisfactory set of rules with other pre-defined constraint expressing minimum accepted (relative) strength of the rule. However it requires to maintain in memory additional information about examples, for details see [8].

4 Experiments

The main aim of an experimental evaluation of the *Glance* algorithm is check how the computational time and the number of rules are changing with increasing number of objects and attributes/criteria. Moreover, we want to compare its performance with previously known algorithms *DomExplore* and *AllRules* [8]. Both of these algorithms are dedicated for multicrieria sorting problems, but they work in an non-incremental way. *DomExplore* induces the satisfactory set of *non-robust* rules and it is inspired by data mining algorithms, which search multiattribute (only) data for all "strong" rules with strength and confidence not less than predefined thresholds [7]. The *AllRules* algorithm has been developed by Greco and Stefanowski and it induces the set of *all roboust* rules, which are based on some particular (non-dominated) objects belonging to approximations of upward or downward unions of classes. It is strongly tied to dominance based rough sets approach and uses its specific properties to reduce descriptions of objects being basis for roboust rules.

Table 1. The computation time (in seconds) for compared algorithms while changing the number of objects

Algorithm	No. of attrib.	Number of objects 500	1000	1500	2000	3000	4000	5000	6000
DomExplore	3	0.10	0.33	0.72	1.15	2.69	4.9	8.3	14.99
	6	3.68	21.4	73.7	178.4	704.9	–	–	–
AllRules	3	0.10	0.16	1.41	3.13	7.69	18.4	34.3	53.01
	6	1.43	5.5	13.1	27.2	108.1	281.9	502.9	712.7
Glance	3	0.05	0.05	0.05	0.06	0.06	0.06	0.1	0.12
	6	1.26	3.13	5.5	7.86	15.38	21.2	26.31	38.7

The experiments were performed on family of artificial data sets. They were randomly generated according to chosen probability distribution (for details see [8]) and differ by the number of objects. Moreover for each series of data sets we could change the number of attributes and the proportion of criteria to regular attributes. Firstly, all three algorithms were compared on the same data sets with respect to time of computation. In two series of data, with number of attributes 3 (2 criteria and 1 regular attribute) and 6 (2 criteria and 4 regular attributes), we systematically changed the number of objects from 500 till 6000. The results are presented in Table 1; symbol "–" means that the algorithm exceeded the accepted resources. Let us remark, that the number of rules induced by algorithms may be different as *Allrules* generates robust rules. Then, in Table 2 we present the change of the number of rules induced by *Glance* algorithm while incrementally increasing the number of objects (with 6 attributes).

Table 2. The number of rules induced by Glance while processing the different number of objects

	Number of objects										
	500	1000	1500	2000	2500	3000	3500	4000	4500	5000	6000
no. rules	1024	1580	1961	2153	2271	2495	2629	2705	2791	2962	3072

Table 3. The computational time (in seconds) of algorithms while changing the number of regular attributes and criteria in data containing 100 objects

	No. of criteria and regular attributes					
Algorithm	3	6	9	12	15	18
DomExplore	0	0.11	0.77	9	41.14	284.79
AllRules	0	0.06	0.82	7.58	68.66	624.83
Glance	0	0.22	44.9	439.41	–	–

Furthermore, we examine the influence of changing the number of attributes for the time of computation (with fixed number of objects equal to 100), see Table 3. For *Allrules* the results are growing with the number of attributes. For *DomExplore* the time also grows exponentially with the number of attributes, but respective results are around 10 times longer than *AllRules*. However, for *Glance* time exceeded the accepted limit when number of attributes was greater than 12.

5 Discussion of results and final remarks

We introduced a new algorithm *Glance*, which induces the set of all rules from provided examples. Unlike the previous algorithms, the new one works in an incremental way. It stores only rules in memory but not processed examples. The user can also define the constraint for maximal number of elementary condition to be used in a rule. Moreover, the algorithm can be extended to allow the user specifying constraints on minimal (relative) strength of a rule - for details see [8].

The algorithm *Glance* has been experimentally evaluated and compared with algorithms *AllRules* and *DomExplore*. The results given in Table 1 show that the computational time of *Glance* increases "linearly" with the number of objects. Furthermore it is always the best, while *AllRules* is the second and *DomExplore* is the worst. On the other hand, additional experiments indicate that if the number of attributes/ criteria is higher than 9, the winner is *AllRules*. Probably it is due to the fact that for all three algorithms the number of rules increases with the number of attributes, and *Glance* needs to check all generated rules when a new example is being processed. This aspect could be somehow reduced by using in *Glance* the minimal rule length constrain. Analysing the number of rules from Table 2, one can notice that the highest increase occurs at the begining of learning, then next parts of examples ($\geq$ 3500) cause smaller changes. Additional results presented in [8] indicate that both *Glance* and *AllRules* "prefer" data characterized by criteria and are less efficient on regular attributes.

The above observations obtained on artificial data, have been also confirmed in other experiments on some real data coming from UCI ML Repository - we skip them due to limited size of this paper, for details see [8].

In future research, we also want to examine classification accuracy of the *Glance* algorithm. However, predicting classification of a new object is a different problem than in only attribute case, because rules indicate unions not single classes. So, new strategies should be developed for determining final decision when the new objects matches conditions of many rules.

References

1. Greco S., Matarazzo B., Slowinski R., (1999) The use of rough sets and fuzzy sets in MCDM. In: Gal T., Stewart T., Hanne T. (Eds) Advances in Multiple Criteria Decision Making, Kluwer, chapter 14, pp. 14.1-14.59.
2. Greco S., Matarazzo B., Slowinski R. (2001), Rough sets theory for multicriteria decision analysis. European Journal of Operational Research, **129** (1), 1–47.
3. Greco S., Matarazzo B., Slowinski R., Stefanowski J. (2000), An algorithm for induction decision rules consistent with the dominance principle. In: Ziarko W., Yao Y. (eds), Proc. 2nd Int. Conference on Rough Sets and Current Trends in Computing, Banff, October 16- 19, 2000, 266-275.
4. Grzymala-Busse J.W. (1992) LERS - a system for learning from examples based on rough sets. In: Slowinski R. (Ed.), Intelligent Decision Support, Kluwer Academic Publishers, 3–18.
5. Pawlak Z. (1991), Rough sets. Theoretical aspects of reasoning about data. Kluwer Academic Publishers.
6. Roy B. (1995), Multicriteria Methodology for Decision Aiding. Kluwer Academic Publishers (Orignal French edition 1985).
7. Stefanowski J. (2001), Algorithims of rule induction for knowledge discovery. (In Polish), Habilitation Thesis published as Series Rozprawy no. 361, Poznan Univeristy of Technology Press, Poznan.
8. Zurawski M. (2001), Algorithms of induction decision rules for multricriteria decision problems. M.Sc. Thesis, Poznań University of Technology. Supervisor Stefanowski J., 110 pages.