Measuring Attractiveness of Rules from the Viewpoint of Knowledge Representation, Prediction and Efficiency of Intervention

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

[1] Institute of Computing Science, Poznań University of Technology,
60-965 Poznań, Poland
[2] Institute for Systems Research, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland
Roman.Slowinski@cs.put.poznan.pl
[3] Faculty of Economics, University of Catania,
Corso Italia 55, 95129 Catania, Italy
salgreco@unict.it

Abstract. Rules mined from a data set represent knowledge patterns relating premises and decisions in '*if* ..., *then* ...' statements. Premise is a conjunction of elementary conditions relative to independent variables and decision is a conclusion relative to dependent variables. Given a set of rules, it is interesting to rank them with respect to some attractiveness measures. In this paper, we are considering rule attractiveness measures related to three semantics: knowledge representation, prediction and efficiency of intervention based on a rule. Analysis of existing measures leads us to a conclusion that the best suited measures for the above semantics are: support and certainty, a Bayesian confirmation measure, and two measures related to efficiency of intervention, respectively. These five measures induce a partial order in the set of rules. For building a strategy of intervention, we propose rules discovered using the Dominance-based Rough Set Approach – the "at least" type rules indicate opportunities for improving assignment of objects, and the "at most" type rules indicate threats for deteriorating assignment of objects.

Keywords: Knowledge discovery, Rules, Attractiveness measures, Efficiency of intervention, Dominance-based Rough Set Approach.

1 Introduction

Knowledge patterns discovered from data are usually represented in a form of '*if* ..., *then* ...' rules, being consequence relations between premise built of independent variables and decision expressed in terms of dependent variables. In data mining and knowledge discovery such rules are induced from data sets concerning a finite set of objects described by a finite set of condition and decision attributes, corresponding to dependent and independent variables, respectively. The rules mined from data may be either decision rules or association rules,

P.S. Szczepaniak et al. (Eds.): AWIC 2005, LNAI 3528, pp. 11–22, 2005.

depending if the division into condition and decision attributes has been fixed or not. Association rules and decision rules have a double utility:

- they **represent knowledge** about relationships between dependent and independent variables existing in data,
- they can be used for **prospective decisions**.

The use of rules for prospective decisions can be understood, however, in two ways:

- **matching up the rules to new objects** with given values of independent variables, in view of **predicting** possible values of dependent variables,
- **building a strategy of intervention** based on discovered rules, in view of transforming a universe in a desired way.

For example, rules mined from data concerning medical diagnosis are useful to represent relationships between symptoms and diseases. Moreover, from one side, the rules can be used to diagnose new patients, assuming that a patient with particular symptoms will probably be sick of a disease suggested by a rule showing a strong relationship between the disease and these symptoms. From the other side, such rules can be seen as general laws and can be considered for application in course of an intervention which consists in modifying some symptoms strongly related with a disease, in order to get out from this disease.

While the first kind of prospective use of rules is rather usual, building a strategy of intervention is relatively new.

Problems related to mining rules from data in view of knowledge representation and building a strategy of intervention can be encountered in many fields, like medical practice, market basket analysis, customer satisfaction and risk analysis. In all practical applications, it is crucial to know how good the rules are for both knowledge representation and efficient intervention. “How good” is a question about attractiveness measures of discovered rules. A review of literature on this subject shows that there is no single measure which would be the best for applications in all possible perspectives (see e.g. [1], [6], [7], [12]).

We claim that the adequacy of interestingness measures to different application perspectives of discovered rules is dependent on **semantics** of these measures. In this paper, we will distinguish three main semantics and for each of them we propose some adequate measures:

- knowledge representation semantics, characterized by the strength and by the certainty degree of discovered rules,
- prediction semantics, underlining the strength of support that a premise gives to a conclusion of a particular rule, known as confirmation degree,
- efficiency of intervention semantics, referring to efficiency of an action based on a rule discovered in one universe and performed in another universe.

The differences between these semantics make impossible any compensatory aggregation of the corresponding measures for ranking the discovered rules according to a comprehensive value. Thus, we postulate to use them all in view of establishing a partial order in the set of discovered rules. While this leaves some rules incomparable, it permits to identify a set of most attractive rules with respect to preferred application perspective.

Considerations of the present article are valid for both association rules and for decision rules; however, for the sake of brevity, we speak about decision rules only.

The paper is organized as follows. In the preliminaries, we introduce some notation and basic definitions concerning rules. Then, we characterize attractiveness measures corresponding to the three semantics mentioned above and, finally, we give an interpretation of the intervention based on monotonic rules coming from Dominance-based Rough Set Approach (DRSA).

2 Preliminaries

Discovering rules from data is the domain of inductive reasoning. Contrary to deductive reasoning, where axioms expressing some universal truths constitute a starting point of reasoning, inductive reasoning uses data about a sample of larger reality to start inference.

Let $S = (U, A)$ be a *data table*, where U and A are finite, non-empty sets called the *universe* and the set of *attributes*, respectively. If in the set A two disjoint subsets of attributes, called *condition* and *decision attributes*, are distinguished, then the system is called a *decision table* and is denoted by $S = (U, C, D)$, where C and D are sets of condition and decision attributes, respectively. With every subset of attributes, one can associate a formal language of logical formulas $\boldsymbol{L}$ defined in a standard way and called the *decision language*. Formulas for a subset $B \subseteq A$ are build up from attribute-value pairs (a, v), where $a \in B$ and $v \in V_a$ (set V_a is a domain of a), by means of logical connectives $\wedge$ (*and*), $\vee$ (*or*), $\neg$ (*not*). We assume that the set of all formula sets in $\boldsymbol{L}$ is partitioned into two classes, called *condition* and *decision formulas*, respectively.

A *decision rule* induced from S and expressed in $\boldsymbol{L}$ is presented as $\Phi \to \Psi$, and read as "*if* Φ*, then* Ψ", where Φ and Ψ are condition and decision formulas in $\boldsymbol{L}$, called *premise* and *decision*, respectively. A decision rule $\Phi \to \Psi$ is also seen as a binary relation between premise and decision, called *consequence relation* (see a critical discussion about interpretation of decision rules as logical implications in [6]).

Let $||\Phi||_S$ denote the set of all objects from universe U, having property Φ in S. If $\Phi \to \Psi$ is a decision rule, then $supp_S(\Phi, \Psi) = card(||\Phi \wedge \Psi||_S)$ is the *support* of the decision rule and

$$str_S(\Phi, \Psi) = \frac{supp_S(\Phi, \Psi)}{card(U)} \tag{1}$$

is the *strength* of the decision rule.

With every decision rule $\Phi \to \Psi$ we associate a *certainty* factor, called also *confidence*,

$$cer_S(\Phi, \Psi) = \frac{supp_S(\Phi, \Psi)}{card(||\Phi||_S)}, \tag{2}$$

and a coverage factor

$$cov_S(\Phi, \Psi) = \frac{supp_S(\Phi, \Psi)}{card(||\Psi||_S)}. \tag{3}$$

Certainty and coverage factors refer to Bayes' theorem:

$$cer_S(\Phi, \Psi) = Pr(\Psi|\Phi) = \frac{Pr(\Psi \wedge \Phi)}{Pr(\Phi)}, \; cov_S(\Phi, \Psi) = Pr(\Phi|\Psi) = \frac{Pr(\Phi \wedge \Psi)}{Pr(\Psi)}$$

Taking into account that given decision table S, the probability (frequency) is calculated as:

$$Pr(\Phi) = \frac{card(||\Phi||_S)}{card(U)}, \; Pr(\Psi) = \frac{card(||\Psi||_S)}{card(U)}, \; Pr(\Phi \wedge \Psi) = \frac{card(||\Phi \wedge \Psi||_S)}{card(U)}$$

one can observe the following relationship between certainty and coverage factors, without referring to prior and posterior probability:

$$cer_S(\Phi, \Psi) = \frac{cov_S(\Phi, \Psi) card(||\Psi||_S)}{card(||\Phi||_S)} \tag{4}$$

Indeed, what is certainty factor for rule $\Phi \to \Psi$ is a coverage factor for inverse rule $\Psi \to \Phi$, and vice versa. This result underlines a directional character of the statement '*if* Φ, *then* Ψ'.

If $cer_S(\Phi, \Psi) = 1$, then the decision rule $\Phi \to \Psi$ is *certain*, otherwise the decision rule is *uncertain*. A set of decision rules supported in total by the universe U creates a *decision algorithm* in S.

3 Attractiveness Measures with Different Semantics

3.1 Knowledge Representation Semantics

Decision rules $\Phi \to \Psi$ induced from some universe U represent knowledge about this universe in terms of laws relating some properties Φ with properties Ψ. These laws are naturally characterized by a number of cases from U supporting them, and by a probability of obtaining a particular decision Ψ considering a condition Φ. These correspond precisely to the *strength* form one side, and to the *certainty* or *coverage* factor from the other side. With respect to the latter side, we saw in the previous section that due to (4), in order to characterize the truth of the relationship between Φ and Ψ, it is enough to use one of these factors only; moreover, for the directional character of the statement '*if* Φ, *then* Ψ', it is natural to choose the certainty factor.

In consequence, we propose to use **strength** $str_S(\Phi, \Psi)$ and **certainty** $cer_S(\Phi, \Psi)$ as attractiveness measures of rules, adequate to the semantics of knowledge representation.

For example, in a data table with medical information on a sample of patients, we can consider as condition attributes a set of symptoms $C = \{c_1, \ldots, c_n\}$, and as decision attributes, a set of diseases $D = \{d_1, \ldots, d_m\}$. In the decision table so obtained we can induce decision rules of the type: "*if* symptoms $c_{i1}, c_{i2}, \ldots, c_{ih}$ appear, *then* there is disease d_j", with $c_{i1}, c_{i2}, \ldots, c_{ih} \in C$ and $d_j \in D$. Such a rule has interpretation of a law characterized as follows (the % is calculated from a hypothetical data table):

- the patients having symptoms $c_{i1}, c_{i2}, \ldots, c_{ih}$ and disease d_j constitute 15% of all the patients in the sample, i.e. 15% is the *strength* of the rule,
- 91% of the patients having symptoms $c_{i1}, c_{i2}, \ldots, c_{ih}$ have also disease d_j, i.e. 91% is the *certainty factor* of the rule.

It is worth noting that strength $str_S(\Phi, \Psi)$ and certainty $cer_S(\Phi, \Psi)$ are more general than a large variety of statistical interestingness measures, like entropy gain, gini, laplace, lift, conviction, chi-squared value and the measure proposed by Piatetsky-Shapiro. Bayardo and Agrawal [1] demonstrated that, for given data table S, the set of Pareto-optimal rules with respect to strength and certainty includes all rules that are best according to any of the above measures.

3.2 Prediction Semantics

The use of rule $\Phi \to \Psi$ for prediction is based on reasoning by analogy: an object having property Φ will have property Ψ. The truth value of this analogy has the semantics of a degree to which a piece of evidence Φ supports the hypothesis Ψ. As shown in [6], this corresponds to a Bayesian confirmation measure (see e.g. [3] and [8] for surveys). While the confirmation measure is certainly related to the strength of relationship between Φ and Ψ, its meaning is different from a simple statistics of co-occurrence of properties Φ and Ψ in universe U, as shown by the following example borrowed from Popper [9].

Consider a possible result of rolling a die: 1,2,3,4,5,6. We can built a decision table, presented in Table 1, where the fact that the result is even or odd is the condition attribute, while the result itself is the decision attribute.

Table 1. Decision Table

Condition attribute (result odd or even)	Decision attribute (result of rolling the die)
odd	1
even	2
odd	3
even	4
odd	5
even	6

Now, consider the case Ψ = "the result is 6" and the case $\neg\Psi$ = "the result is not 6". Let us also take into account the information Φ = "the result is an even number (i.e. 2 or 4 or 6)". Therefore, we can consider the following two decision rules:

- $\Phi \rightarrow \Psi$ = "if the result is even, then the result is 6", with certainty $cer_S(\Phi,\Psi) = 1/3$,
- $\Phi \rightarrow \neg\Psi$ = "if the result is even, then the result is not 6", with certainty $cer_S(\Phi,\neg\Psi) = 2/3$.

Remark that rule $\Phi \rightarrow \Psi$ has a smaller certainty than rule $\Phi \rightarrow \neg\Psi$. However, the probability that the result is 6 is 1/6, while the probability that the result is different from 6 is 5/6. Thus, the information Φ raises the probability of Ψ from 1/6 to 1/3, and decreases the probability of $\neg\Psi$ from 5/6 to 2/3. In conclusion, we can say that Φ confirms Ψ and disconfirms $\neg\Psi$, independently of the fact that the certainty of $\Phi \rightarrow \Psi$ is smaller than the certainty of $\Phi \rightarrow \neg\Psi$.

From this simple example, one can see that certainty and confirmation are two completely different concepts.

Bayesian confirmation measure, denoted by $c(\Phi,\Psi)$, exhibits the impact of evidence Φ on hypothesis Ψ by comparing probability $Pr(\Psi|\Phi)$ with probability $Pr(\Psi)$ as follows:

$$c(\Phi,\Psi)\begin{cases} > 0 \text{ if } Pr(\Psi|\Phi) > Pr(\Psi) \\ = 0 \text{ if } Pr(\Psi|\Phi) = Pr(\Psi) \\ < 0 \text{ if } Pr(\Psi|\Phi) < Pr(\Psi) \end{cases} \tag{5}$$

In data mining, the probability Pr of Ψ is substituted by the relative frequency Fr in the considered data table S, i.e.

$$Fr_S(\Psi) = \frac{card(||\Phi||)}{card(U)}.$$

Analogously, given Φ and Ψ, $Pr(\Psi|\Phi)$ is substituted by the certainty factor $cer_S(\Phi,\Psi)$ of the decision rule $\Phi \rightarrow \Psi$, therefore, a measure of confirmation of property Ψ by property Φ can be rewritten as:

$$c(\Phi,\Psi)\begin{cases} > 0 \text{ if } cer_S(\Phi,\Psi) > Fr_S(\Psi) \\ = 0 \text{ if } cer_S(\Phi,\Psi) = Fr_S(\Psi) \\ < 0 \text{ if } cer_S(\Phi,\Psi) < Fr_S(\Psi) \end{cases} \tag{6}$$

(6) can be interpreted as follows:

- $c(\Phi,\Psi) > 0$ means that property Ψ is satisfied more frequently when Φ is satisfied (then, this frequency is $cer_S(\Phi,\Psi)$), rather than generically in the whole decision table (where this frequency is $Fr_S(\Psi)$),
- $c(\Phi,\Psi) = 0$ means that property Ψ is satisfied with the same frequency when Φ is satisfied and generically in the whole decision table,
- $c(\Phi,\Psi) < 0$ means that property Ψ is satisfied less frequently when Φ is satisfied, rather than generically in the whole decision table.

In other words, the confirmation measure for rule $\Phi \rightarrow \Psi$ is the credibility of the following proposition: **Ψ is satisfied more frequently when Φ is satisfied rather than when Φ is not satisfied**.

Apart from property (5), many authors have considered other properties of confirmation measures (see [3] for a survey). Among the desirable properties there is a kind of symmetry called *hypothesis symmetry* [2]:

$$c(\Phi, \Psi) = -c(\Phi, \neg\Psi) \tag{7}$$

Greco, Pawlak and Słowiński [6] have formulated yet another desirable property for confirmation measures of rules mined from data tables – this property is called *monotonicity*. It underlines an important difference existing between rules considered as consequence relations and rules considered as logical (material) implications.

Using the denotation: $a = supp_S(\Phi, \Psi),\ b = supp_S(\neg\Phi, \Psi),\ c = supp_S(\Phi, \neg\Psi)$, $d = supp_S(\neg\Phi, \neg\Psi)$, the monotonicity property says that $c(\Phi, \Psi) = F(a, b, c, d)$, where F is a function non-decreasing with respect to a and d and non-increasing with respect to b and c.

While monotonicity of the confirmation measure with respect to a and c makes no doubt, the monotonicity with respect to b and d needs a comment. Remembering that $c(\Phi, \Psi)$ is the credibility of the proposition: Ψ is satisfied more frequently when Φ is satisfied rather than when Φ is not satisfied, we can state the following. An evidence in which Φ is not satisfied and Ψ is satisfied (objects $||\neg\Phi \wedge \Psi||$) increases the frequency of Ψ in situations where Φ is not satisfied, so it should decrease the value of $c(\Phi, \Psi)$. Analogously, an evidence in which both Φ and Ψ are not satisfied (objects$||\neg\Phi \wedge \neg\Psi||$) decreases the frequency of Ψ in situations where Φ is not satisfied, so it should increase the value of $c(\Phi, \Psi)$.

In [6], six confirmation measures well known from the literature have been analyzed from the viewpoint of the desirable monotonicity property. It has been proved that only three of them satisfy this property. Moreover, among these three confirmation measures, only two satisfy the hypothesis symmetry (7); these are:

$$\begin{aligned} l(\Phi, \Psi) &= \log\left[\frac{cer_S(\Psi, \Phi)}{cer_S(\neg\Psi, \Phi)}\right] = \log\left[\frac{a/(a+b)}{c/(c+d)}\right] \text{ and} \\ f(\Phi, \Psi) &= \frac{cer_S(\Psi, \Phi) - cer_S(\neg\Psi, \Phi)}{cer_S(\Psi, \Phi) + cer_S(\neg\Psi, \Phi)} = \frac{ad - bc}{ad + bc + 2ac}. \end{aligned} \tag{8}$$

As proved by Fitelson [3], these measures are ordinally equivalent, i.e. for all rules $\Phi \to \Psi$ and $\Phi' \to \Psi'$, $l(\Phi, \Psi) \geq l(\Phi', \Psi')$ if and only if $f(\Phi, \Psi) \geq f(\Phi', \Psi')$. Thus, it is sufficient to use one of them, e.g. $f(\Phi, \Psi)$.

In consequence, we propose to use **confirmation measure** $f(\Phi, \Psi)$ as attractiveness measure of rules, adequate to the semantics of reasoning by analogy for prediction.

3.3 Efficiency of Intervention Semantics

The attractiveness measures considered above can be interpreted as characteristics of the universe U where the rules come from, and do not measure the future effects of a possible intervention based on these rules. In [5], we considered expected effects of an intervention which is a three-stage process:

1. Mining rules in universe U.
2. Modification (manipulation) of universe U', based on a rule mined from U, with the aim of getting a desired result.
3. Transition from universe U' to universe U'' due to the modification made in stage 2.

For example, let us suppose a medical rule has been induced from universe U: $r \equiv$ '*if absence of symptom Φ, then no disease Ψ*' whose certainty is 90% (i.e. in 90% of cases where symptom Φ is absent there is no disease Ψ). On the basis of r, an intervention may be undertaken in universe U' consisting in eliminating symptom Φ to get out from disease Ψ in universe U''. This intervention is based on a hypothesis of *homogeneity* of universes U and U'. This homogeneity means that r is valid also in U' in the sense that one can expect that 90% of sick patients with symptom Φ will get out from the sickness due to the intervention.

In another application concerning customer satisfaction analysis, the universe is a set of customers and the intervention is a strategy (promotion campaign) modifying perception of a product so as to increase customer satisfaction.

Measures of efficiency of intervention depend not only on characteristics of rules in universe U, but also on characteristics of universe U' where the intervention takes place.

Let $S = (U, A)$, $S' = (U', A)$ and $S'' = (U'', A)$ denote three data tables referring to universes U, U' and U'', respectively.

In [5], the following reasoning has been applied to measure the effect of an intervention based on rule $\Phi \to \Psi$: if we modify property $\neg\Phi$ to property Φ in the set $||\neg\Phi \wedge \neg\Psi||_{S'}$, we may reasonably expect that $cer_S(\Phi, \Psi) \times supp_{S'}(\neg\Phi, \neg\Psi)$ objects from set $||\neg\Phi \wedge \neg\Psi||_{S'}$ in universe U' will enter decision class Ψ in universe U''. In consequence, the expected ***relative increment*** of objects from U' entering decision class Ψ in universe U'' is:

$$incr_{SS'}(\Psi) = cer_S(\Phi, \Psi) \times \frac{card(||\neg\Phi \wedge \neg\Psi||_{S'})}{card(U')} \tag{9}$$

The relative increment (9) can be rewritten as:

$$\begin{aligned} incr_{SS'}(\Psi) &= cer_S(\Phi, \Psi) \times \tfrac{card(||\neg\Phi \wedge \neg\Psi||_{S'})}{card(||\neg\Psi||_{S'})} \times \tfrac{card(||\neg\Psi||_{S'})}{card(U')} = \\ &= cer_S(\Phi, \Psi) \times cer_{S'}(\neg\Psi, \neg\Phi) \times \tfrac{card(||\neg\Psi||_{S'})}{card(U')} \end{aligned} \tag{10}$$

where $cer_{S'}(\neg\psi, \neg\phi)$is a certainty factor of the contrapositive rule $s \equiv \neg\Psi \to \neg\Phi$ in U'. Taking into account that $card(||\neg\Psi||_{S'})/card(U')$ is a fraction of all objects having not property Ψ in universe U', the remaining part of (10) is just expressing the **efficiency of the intervention**:

$$eff_{SS'}(\Phi, \Psi) = cer_S(\Phi, \Psi) \times cer_{S'}(\neg\Psi, \neg\Phi). \tag{11}$$

Assuming that the condition formula Φ is composed of n elementary conditions $\Phi_1 \wedge \Phi_2 \wedge \ldots \wedge \Phi_n$, we consider rule $r \equiv \Phi_1 \wedge \Phi_2 \wedge \ldots \wedge \Phi_n \to \Psi$, with certainty $cer_S(\Phi, \Psi)$. Using this rule, one can perform a multi-attribute intervention which consists in modification of attributes with indices from each subset $P \subseteq N = \{1, \ldots, n\}$ on all objects from U' having none of properties Φ_i, $i \in P$, while having all properties Φ_j, $j \notin P$, and having not property Ψ. In this case, the **relative increment** (10) takes the form:

$$incr_{SS'}(\Psi) = \tag{12}$$
$$= \sum_{\emptyset \subset P \subset N} \left[cer_S(\Phi, \Psi) \times cer_{S'} \left(\neg\Psi, \bigwedge_{i \in P} \neg\Phi_i \wedge \bigwedge_{j \notin P} \Phi_j \right) \right] \times \frac{card(||\neg\Psi||_{S'})}{card(U')}$$

where $cer_{S'} \left(\neg\Psi, \bigwedge_{i \in P} \neg\Phi_i \wedge \bigwedge_{j \notin P} \Phi_j \right)$ is a certainty factor of the contrapositive rule $s_P \equiv \neg\Psi \rightarrow \bigwedge_{i \in P} \neg\Phi_i \wedge \bigwedge_{j \notin P} \Phi_j$ in U', for $P \subseteq N$. From (12) it follows that the **efficiency of the multi-attribute intervention** is equal to:

$$\mathit{eff}_{SS'}(\Phi, \Psi) = cer_S(\Phi, \Psi) \times \sum_{\emptyset \subset P \subset N} cer_{S'} \left(\neg\Psi, \bigwedge_{i \in P} \neg\Phi_i \wedge \bigwedge_{j \notin P} \Phi_j \right). \tag{13}$$

Using calculations analogous to calculation of the Shapley value in terms of the Möbius transform of the Choquet capacity, one can assess a contribution of each particular elementary condition Φ_i, $i \in N$, in the efficiency of the whole intervention [5].

Remark that relative increment $incr_{SS'}(\Psi)$ and efficiency of intervention $\mathit{eff}_{SS'}(\Phi, \Psi)$ have a meaning analogical to knowledge representation measures, i.e. strength $str_S(\Phi, \Psi)$ and certainty factor $cer_S(\Phi, \Psi)$, respectively; they refer, however, to intervention in another universe than that of the knowledge representation.

3.4 Partial Order of Rules with Respect to the Five Measures of Attractiveness

A set of rules can be partially ordered using the five attractive measures proposed in this section. These are:

- rule strength $str_S(\Phi, \Psi)$ (1),
- certainty factor $cer_S(\Phi, \Psi)$ (2),
- confirmation measure $f(\Phi, \Psi)$ (8),
- relative increment due to intervention $incr_{SS'}(\Psi)$ (12),
- efficiency of intervention $\mathit{eff}_{SS'}(\Phi, \Psi)$ (13).

Such a partial ranking supports an interactive search in which the user can browse the best rule according to preferences related to a specific application: representation, prediction or intervention. Remark that it also makes sense to use these measures in a lexicographic procedure, ordering first the rules with respect to the most important measure, then, ordering a subset of best rules using the second-most important measure, and so on.

4 Interpretation of the Intervention Based on Monotonic Rules

Let us complete our considerations by interpretation of the intervention based on monotonic rules coming from the Dominance-based Rough Set Approach (DRSA) [4], [10].

Considering decision table $S = (U, C, D)$, where C is a finite set of attributes with preference-ordered domains X_q ($q \in C$), and D is a finite set of decision attributes partitioning U into a finite set of preference-ordered decision classes $Cl_1, Cl_2, \ldots, Cl_k$ (the higher the index the better the class), DRSA permits to mine two kinds of decision rules:

- **"at least" rules**
  if $x_{q1} \succeq_{q1} r_{q1}$ *and* $x_{q2} \succeq_{q2} r_{q2}$ *and* $\ldots$ $x_{qp} \succeq_{qp} r_{qp}$, *then* $x \in Cl_t^{\geq}$,
  where for each $w_q, z_q \in X_q$, "$w_q \succeq_q z_q$" means "w_q is at least as good as z_q", and $x \in Cl_t^{\geq}$ means "x belongs to class Cl_t or better",
- **"at most" rules**
  if $x_{q1} \preceq_{q1} r_{q1}$ *and* $x_{q2} \preceq_{q2} r_{q2}$ *and* $\ldots$ $x_{qp} \preceq_{qp} r_{qp}$, *then* $x \in Cl_t^{\leq}$,
  where for each $w_q, z_q \in X_q$, "$w_q \preceq_q z_q$" means "w_q is at most as good as z_q", and $x \in Cl_t^{\leq}$ means "x belongs to class Cl_t or worse".

The rules "at least" indicate **opportunities for improving** the assignment of object x to class Cl_t or better, if it was not assigned as high and its evaluation on $q1, q2, \ldots, qp$ would grow to $r_{q1}, r_{q2}, \ldots, r_{qp}$ or better.

The rules "at most" indicate **threats for deteriorating** the assignment of object x to class Cl_t or worse, if it was not assigned as low and its evaluation on $q1, q2, \ldots, qp$ would drop to $r_{q1}, r_{q2}, \ldots, r_{qp}$ or worse.

In the context of these two kinds of rules, an **intervention** means either an action of **taking the opportunity** of improving the assignment of a subset of objects, or an action of **protecting against threats** of deteriorating the assignment of a subset of objects.

For example, consider the following "at least" rule mined from a hypothetical data set of customer satisfaction questionnaires:

$$\text{‘}\textit{if}\,(q1 \geq 5) \wedge (q5 \geq 4), \textit{then}\ \text{Satisfaction} \succeq \text{High’}$$

Suppose that an intervention based on this rule is characterized by $incr_{SS'}(\text{High}) = 77\%$; this means that increasing $q1$ above 4 and increasing $q5$ above 3 will result in improvement of customer satisfaction from Medium or Low to High for 77% of customers with Medium or Low satisfaction.

Now, consider the following "at most" rule:

$$\text{‘}\textit{if}\,(q2 \leq 4) \wedge (q4 \leq 4) \wedge (q6 \leq 4), \textit{then}\ \text{Satisfaction} \preceq \text{Medium’}$$

In this case, $incr_{SS'}(\text{Medium}) = 89\%$ means that dropping $q2, q4$ and $q6$ below 5 will result in deterioration of customer satisfaction from High to Medium or Low for 89% of customers with High satisfaction.

In practical applications, the choice of rules used for intervention can also be supported by some additional measures, like:

– length of the rule chosen for intervention (the shorter the better),
– cost of intervention on attributes present in the rule,
– priority of intervention on some types of attributes, like short-term attributes or attributes on which competing firms perform better.

Remark that intervention based on rules shows some similarity with an interesting concept of **action rules** considered by Tsay and Raś [11], however, action rules are pairs of rules representing two scenarios for assignment of an object: one desired and another unsatisfactory, and the action consists in passing from the undesired scenario to desired one, by changing values of so-called flexible attributes. Action rules are characterized by support and confidence only.

5 Conclusions

In this paper, we considered attractiveness measures of rules mined from data, taking into account three application perspectives: knowledge representation, prediction of new classifications and interventions based on discovered rules in some other universe. In order to choose attractiveness measures concordant with the above perspectives we analyzed semantics of particular measures which lead us to a conclusion that the best suited measures for the above applications are: support and certainty, a Bayesian confirmation measure, and two measures related to efficiency of intervention, respectively. These five measures induce a partial order in the set of rules, giving a starting point for an interactive browsing procedure. For building a strategy of intervention, we proposed rules discovered using the Dominance-based Rough Set Approach – the "at least" type rules indicate opportunities for improving assignment of objects, and the "at most" type rules indicate threats for deteriorating assignment of objects.

Acknowledgements

The first author wishes to acknowledge financial support from the State Committee for Scientific Research (KBN). The research of the second author was supported by the Italian Ministry for Education, University and Research (MIUR).

References

1. Bayardo, R.J., Agrawal, R.: Mining the most interesting rules. [In]: Proc. of the Fifth ACM SIGKDD Intl Conf. on Knowledge Discovery and Data Mining, (1999) 145–154
2. Eells, E., Fitelson, B.: Symmetries and asymmetries in evidential support. Philosophical Studies, **107** (2) (2002) 129–142

3. Fitelson, B.: Studies in Bayesian Confirmation Theory. Ph.D. thesis, University of Wisconsin – Madison (2001)
4. Greco, S., Matarazzo, B., Słowiński, R.: Dominance-Based Rough Set Approach to Knowledge Discovery – (I) General Perspective; (II) Extensions and Applications. Chapters 20 and 21 [in]: N.Zhong and J.Liu, Intelligent Technologies for Information Analysis. Springer-Verlag, New York, (2004) 513–612
5. Greco, S., Matarazzo, B., Pappalardo, N., Słowiński, R.: Measuring expected effects of interventions based on decision rules. Journal of Experimental and Applied Artificial Intelligence, **17** (1-2) (2005) 103–118
6. Greco, S., Pawlak, Z., Słowiński, R.: Can Bayesian confirmation measures be useful for rough set decision rules? Engineering Applications of Artificial Intelligence, **17** (4) (2004) 345–361
7. Hilderman, R.J., Hamilton, H.J.: Knowledge Discovery and Measures of Interest. Kluwer Academic Publishers, Boston (2001)
8. Kyburg, H.: Recent work in inductive logic. [In]: K.G. Lucey and T.R. Machan (eds.), Recent Work in Philosophy. Rowman and Allanheld, Totowa, N.J., (1983) 89–150
9. Popper, K. R.: The Logic of Scientific Discovery. Hutchinson, London (1959)
10. Słowiński, R., Greco, S., Matarazzo, B.: Rough Set Based Decision Support. Chapter 15 [in]: E. Burke and G. Kendall (eds.), Introductory Tutorials on Optimization, Search and Decision Support Methodologies, Springer-Verlag, New York (2005)
11. Tsay, L.-S., Raś, Z.: Action rule discovery: system DEAR2, method and experiments. Journal of Experimental and Applied Artificial Intelligence **17** (1-2) (2005) 119–128
12. Yao, Y.Y., Zhong, N.: An analysis of quantitative measures associated with rules. [In]: Proceedings of the Third Pacific-Asia Conference on Knowledge Discovery and Data Mining, LNAI **1574**, Springer-Verlag, Berlin, (1999) 479–488