Control and Cybernetics

vol. **38** (2009) No. 1

Analysis of monotonicity properties of some rule interestingness measures*

by

Salvatore Greco[1], **Roman Słowiński**[2,3] **and Izabela Szczęch**[2]

[1] Faculty of Economics, University of Catania
Corso Italia, 55, 95129 Catania, Italy

[2] Institute of Computing Science, Poznań University of Technology
60-965 Poznań, Poland

[3] Systems Research Institute, Polish Academy of Sciences
01-447 Warsaw, Poland

e-mail: salgreco@mbox.unicit.it, {Roman.Slowinski,
Izabela.Szczech}@cs.put.poznan.pl

Abstract: One of the crucial problems in the field of knowledge discovery is development of good interestingness measures for evaluation of the discovered patterns. In this paper, we consider quantitative, objective interestingness measures for "*if..., then...*" association rules. We focus on three popular interestingness measures, namely *rule interest function* of Piatetsky-Shapiro, *gain measure* of Fukuda et al., and *dependency factor* used by Pawlak. We verify whether they satisfy the valuable property M of monotonic dependency on the number of objects satisfying or not the premise or the conclusion of a rule, and property of hypothesis symmetry (HS). Moreover, analytically and through experiments we show an interesting relationship between those measures and two other commonly used measures of rule support and anti-support.

Keywords: association rules, Piatetsky-Shapiro's rule interest function, gain measure, dependency factor, support, anti-support, Pareto-optimal border.

1. Introduction

In data mining and knowledge discovery, the discovered knowledge patterns are often expressed in the form of "*if..., then...*" rules. They are consequence relations representing correlation, association, causation etc. between independent and dependent attributes. If the division into independent and dependent

*Submitted: June 2008; Accepted: October 2008.

attributes has been fixed, the rules mined from data are regarded as decision rules, otherwise as association rules.

It has been recognized early on in the knowledge discovery literature that the number of rules discovered in databases can be quite large and can easily overwhelm the human capabilities to understand them and to find useful results. This is due to the fact that many rules are either irrelevant or obvious, and do not provide new knowledge (Morzy and Zakrzewicz, 2003). To address the problem of evaluation of attractiveness of the mined rules, various quantitative measures of interestingness have been defined and studied (e.g. support, confidence, anti-support, gain, rule interest function, lift) (Bramer, 2007). They allow for reducing the number of rules that need to be considered by ranking them and filtering out the useless ones. Each of the interestingness measures has been introduced to reflect different characteristics of rules.

Generally, interestingness measures can be divided into objective and subjective. The first group can be established through statistical arguments, derived from data to determine whether a rule is interesting or not. For example, rules that cover only very few transactions, and can therefore capture spurious relationships in data, are discarded by objective measures. On the other hand, the group of subjective measures regards a rule as uninteresting unless it reveals unexpected information about the data or provides knowledge that can lead to profitable actions (Tan, Steinbach and Kumar, 2006). Thus, for subjective evaluation criteria rare cases in the data are often interesting and rules that cover them are of high value. All in all, objective measure can be seen as those that depend on the structure of the rules and the underlying data used in the discovery process, whereas the subjective measures depend on the class of users who examine the rule (Silberschatz and Tuzhilin, 1996).

Let us also stress that there is no general interestingness evaluation approach that will work for any real-life problem. The choice of an interestingness measure for a certain application is a non-trivial task that should be closely related to the domain of a particular problem and should take advantage of available domain knowledge.

Since the literature is now a rich resource of interestingness measures, naturally, there arises a need of studying and analyzing relationships between various measures. Such studies could show similarities and differences in the behavior of the measures (e.g. whether the measures rank the rules in the same way) and are a useful tool helping to choose a proper measure for the particular use.

While choosing interestingness measure(s) for a certain application, the users also often take into consideration properties (features) of measures, which reflect the user's expectations toward the behavior of the measures in particular situations. For example, one may demand that the measure used increase its value for a given rule (or at least does not decrease) when the number of objects in the dataset that support this rule increases. Thus, verification whether particular interestingness measures satisfy some valuable features is another valid problem from both theoretical and practical points of view. Such analysis would widen our understanding of measures and of their applicability, and could also unveil some relationships between different measures.

In this paper, we focus on three well-known objective measures: rule interest function proposed by Piatetsky-Shapiro (1991), gain measure of Fukuda et al. (1996) and dependency factor, considered by Pawlak (2004) and Popper (1959). We investigate whether they possess a useful feature called property M introduced by Greco, Pawlak and Słowiński (2004), and hypothesis symmetry (HS) advocated by Eells and Fitelson (2002) and Fitelson (2001). Moreover, on the basis of satisfying the property M, we draw some conclusions about very particular relationship between rule interest function and gain measure, and two other simple but meaningful measures of rule support and anti-support.

In order to achieve the above objectives, the rest of the paper is organized as follows. In Section 2, there are preliminaries on rules and their quantitative description. In Section 3, we verify analytically whether rule interest function, gain measure and dependency factor have the analyzed property M. In Section 4, we investigate the relationship between the first two measures and the Pareto-optimal border with respect to support and anti-support. Illustration of the results on a real life dataset is presented to support the theoretical considerations with experimental results. Next, in Section 5, we analyze if rule interest function, gain measure and dependency factor satisfy the hypothesis symmetry. The paper ends with conclusions.

2. Preliminaries

The discovery of knowledge from data is done by induction. It is a process of creating patterns which are true in the world of the analyzed data. In this paper we consider discovering knowledge represented in the form of the rules. The starting point for such rule induction (mining) is a sample of a larger reality often represented in the form of a data table.

Formally, a data table is a pair $S = (U, A)$, where U is a nonempty finite set of objects, called *universe*, and A is a nonempty finite *set of attributes*. For every attribute $a \in A$, let us denote by V_a the domain of a, and $a(x)$ will stand for the value of attribute a for an object $x \in U$. A *rule* induced from a data table S is denoted by $\phi \rightarrow \psi$ (read as "*if* ϕ, *then* ψ"), where ϕ and ψ are built up from elementary conditions using logical operator $\wedge$ (and). The *elementary conditions* of a rule are defined as $(a(x)\ rel\ v)$ where *rel* is a relational operator from the set $\{=, <, \leq, \geq, >\}$ and v is a constant belonging to V_a. The antecedent ϕ of a rule is also referred to as *premise* or *condition*, whereas the consequent ψ of a rule is often called *conclusion*, *decision* or *hypothesis*. Generally, a rule can be seen as a consequence relation (see critical discussion in Greco, Pawlak and Słowiński, 2004, about interpretation of rules as logical implications) between premise and conclusion. The attributes that appear in elementary conditions of the premise (respectively, conclusion) are called *condition attributes* (*respectively, decision attributes*). Obviously, within one rule, the sets of condition and decision attributes must be disjoint. The rules induced from data may be either *decision* or *association rules*, depending on whether the division of A into condition and decision categories of attributes has been fixed or not.

2.1. Support and anti-support measures of rules

One of the most popular measures used to identify frequently occurring association rules in sets of items from information table S is *support* (Agrawal, Imielinski and Swami, 1993). The *support of condition* ϕ (analogously, ψ), denoted as $sup(\phi)$ (analogously, $sup(\psi)$), is equal to the number of objects in U having property ϕ (analogously, property ψ). The *support of rule* $\phi \to \psi$ (also simply referred to as support), denoted as $sup(\phi \to \psi)$, is the number of objects in U having property ϕ and ψ. Thus, it corresponds to statistical significance (Hilderman and Hamilton, 2001). The domain of the measure of support can cover any natural number. The greater the value of support for a given rule, the more desirable the rule is, thus, support is a gain-type criterion.

Anti-support of a rule $\phi \to \psi$ (also simply referred to as anti-support), denoted as $anti-sup(\phi \to \psi)$, is equal to the number of objects in U having the property ϕ but not having the property ψ. Thus, anti-support is the number of counterexamples, i.e. objects for which the premise ϕ evaluates as true but which fall into a class different than ψ. Note that anti-support can also be regarded as $sup(\phi \to \neg\psi)$. Similarly to support, the anti-support measure can obtain any natural value. However, its optimal value is 0, because it reflects the situation in which a rule has no counterexamples at all. Any value greater than zero means that the considered rule is not certain, i.e. there are some counterexamples for that rule. The less counterexamples we observe in the dataset, the better, and therefore anti-support is considered a cost-type criterion.

Some authors define support and anti-support as relative values with respect to the number of all objects in the dataset U (Tan, Steinbach, and Kumar, 2006). Then, the rule support (respectively, anti-support) can be interpreted as the percentage of objects satisfying both the premise and conclusion (respectively, counterexamples) of the rule, in the dataset. In this paper we will consider the former definition of support and anti-support, however, using the latter would not influence the generality of the conducted analysis and the obtained results.

2.2. Piatetsky-Shapiro's rule interest function, gain and dependency factor

The $rule\ interest\ function$, RI, introduced by Piatetsky-Shapiro (1991) is used to quantify the correlation between the premise and conclusion. It is given by the following formula:

$$RI(\phi \to \psi) = sup(\phi \to \psi) - \frac{sup(\psi)sup(\phi)}{|U|}. \tag{1}$$

For rule $\phi \rightarrow \psi$, when $RI = 0$, then ϕ and ψ are statistically independent and thus, such rule should be considered as uninteresting. When $RI > 0$ ($RI < 0$), then there is a positive (negative) correlation between ϕ and ψ (Hilderman and Hamiltonet, 2001). Obviously, it is a gain-type criterion, as greater values of RI reflect stronger trend toward desirable positive correlation.

The *gain* function of Fukuda et al. (1996) is defined in the following manner:

$$gain(\phi \rightarrow \psi) = sup(\phi \rightarrow \psi) - \Theta sup(\phi) \tag{2}$$

where Θ is a fraction constant between 0 and 1. Note that, for a fixed value of $\Theta = sup(\psi)/|U|$, the gain measure becomes identical to the above rule interest function RI. Moreover, if Θ is zero then *gain* boils down to calculation of the support of the rule, and when Θ is equal to 1, *gain* will take negative values unless all objects satisfying ϕ also satisfy ψ (in that case *gain* will be 0). Thus, *gain* can take any integer value depending on what value Θ is set at. For a fixed Θ, greater values of *gain* are more desirable, thus it is a gain-type criterion.

The $dependency\ factor$ used by Pawlak (2004) and also considered earlier by Popper (1959), is defined in the following manner:

$$\eta(\phi \rightarrow \psi) = \frac{\dfrac{sup(\phi \rightarrow \psi)}{sup(\phi)} - \dfrac{sup(\psi)}{|U|}}{\dfrac{sup(\phi \rightarrow \psi)}{sup(\phi)} + \dfrac{sup(\psi)}{|U|}}. \tag{3}$$

The dependency factor expresses the degree of dependency, and can be seen as a counterpart of correlation coefficient used in statistics. When ϕ and ψ are independent of each other, then $\eta(\phi \rightarrow \psi) = 0$. If $-1 < \eta(\phi \rightarrow \psi) < 0$, then ϕ and ψ are negatively dependent, and if $0 < \eta(\phi \rightarrow \psi) < 1$, then ϕ and ψ are positively dependent on each other. The dependency factor is a gain-type criterion.

2.3. Property of monotonicity M

Greco, Pawlak and Słowiński (2004) considered a group of interestingness measures, called Bayesian confirmation measures, from the viewpoint of their usefulness for measuring interestingness of decision rules. In general, Bayesian confirmation measures say in what degree a piece of evidence in premise confirms a hypothesis in the conclusion of a rule. Greco, Pawlak and Słowiński (2004) claim that confirmation measures should enjoy a valuable property M describing monotonic dependency on the number of objects satisfying or not the premise or the conclusion of the rule. Though the property was introduced in the perspective of confirmation measures, its definition is wide enough to cover any interestingness measures and we are strongly convinced that it is a desirable property for any measure.

The *property M* was formally defined in Greco, Pawlak and Słowiński (2004) as follows:
An interestingness measure

$$F = [sup(\phi \to \psi), sup(\neg\phi \to \psi), sup(\phi \to \neg\psi), sup(\neg\phi \to \neg\psi)] \quad (4)$$

being a gain-type criterion, has the property M if and only if it is a function

- non-decreasing with respect to $sup(\phi \to \psi)$,
- non-increasing with respect to $sup(\neg\phi \to \psi)$,
- non-increasing with respect to $sup(\phi \to \neg\psi)$, and
- non-decreasing with respect to $sup(\neg\phi \to \neg\psi)$.

The property M with respect to $sup(\phi \to \psi)$ (or, analogously, with respect to $sup(\neg\phi \to \neg\psi)$) means that any evidence in which ϕ and ψ (or, analogously, neither ϕ nor ψ) hold together increases (or at least does not decrease) the credibility of the rule $\phi \to \psi$. On the other hand, the property M with respect to $sup(\neg\phi \to \psi)$ (or, analogously, with respect to $sup(\phi \to \neg\psi)$) means that any evidence in which ϕ does not hold and ψ holds (or, analogously, ϕ holds and ψ does not hold) decreases (or at least does not increase) the credibility of the rule $\phi \to \psi$.

Let us use the following example, considered by Hempel (1945), to show the interpretation of the property. Consider a rule $\phi \to \psi$:

if x is a raven then x is black.

In this case ϕ stands for being a raven and ψ stands for being black. If an interestingness measure $F(\phi \to \psi)$ (being a gain-type criterion) possesses the property M then:

- the more black ravens there are in the dataset, the more credible is the rule, and thus $F(\phi \to \psi)$ obtains greater (or at least not smaller) values,
- $F(\phi \to \psi)$ also obtains greater (or at least not smaller) values when the number of non-black non-ravens increases,
- the more black non-ravens appear in the dataset, the less credible becomes the rule and thus, $F(\phi \to \psi)$ obtains smaller (or at least not greater) values,
- $F(\phi \to \psi)$ also obtains smaller (or at least not greater) values when the number of non-black ravens in the dataset increases.

Property M makes use of elementary parameters of the considered dataset (numbers of objects satisfying some properties) and therefore is an easy and intuitive criterion helping to choose an appropriate interestingness measure for a certain application.

2.4. Property of Hypothesis Symmetry (HS)

Eells and Fitelson (2002) have analyzed some confirmation measures from the viewpoint of four properties of symmetry, introduced by Carnap (1962). Again, we believe that these properties should be considered for any interestingness measure, and not be limited to the group of Bayesian confirmation measures.

Considering an interestingness measure $c(\phi \to \psi)$, the considered symmetries were defined as follows:

- evidence symmetry (ES): $c(\phi \to \psi) = -c(\neg\phi \to \psi)$
- commutativity symmetry (CS): $c(\phi \to \psi) = c(\psi \to \phi)$
- hypothesis symmetry (HS): $c(\phi \to \psi) = -c(\phi \to \neg\psi)$
- total symmetry (TS): $c(\phi \to \psi) = c(\neg\phi \to \neg\psi)$.

It has been concluded in Eells and Fitelson (2002) that, in fact, only (HS) is a desirable property, while (ES), (CS) and (TS) are not. The meaning behind the hypothesis symmetry is that the significance of the premise with respect to the conclusion of a rule should be of the same strength, but of the opposite sign, as the significance of the premise with respect to a negated conclusion.

The arguments for (HS) can be presented by an exemplary situation of randomly drawing a card from a standard deck (Earman, 1992; Greco, Pawlak and Słowiński, 2004). Let the premise ϕ of a rule stand for that *the drawn card is the seven of spades*, and let ψ be the hypothesis that *the card is black*. It is clear that the premise confirms the hypothesis in 100%. Moreover, obviously, the evidence that the card is the seven of spades (ϕ) is negatively conclusive (completely disconfirms) for the hypothesis that the card is not black ($\neg\psi$).

2.5. Support–anti-support Pareto-optimal border

Let us denote by $\preceq_{s\neg a}$ a partial preorder given by the dominance relation on a set X of rules in terms of two interestingness measures: support and anti-support, i.e. given a set of rules X and two rules $r_1, r_2 \in X, r_1 \prec_{s\neg a} r_2$ if and only if

$$sup(r_1) \leq sup(r_2) \ \wedge \ anti - sup(r_1) \geq anti - sup(r_2). \tag{5}$$

Recall that a partial preorder on a set X is any binary relation R on X that is reflexive (i.e. for all $x \in X$, xRx) and transitive. In simple words, if the semantics of xRy is "x is at most as good as y", then a complete preorder permits to order the elements of X from the best to the worst, with possible ex-aequo (i.e. cases of $x, y \in X$ such that xRy and yRx) and with possible incomparability (i.e. cases of $x, y \in X$ such that *not* xRy and *not* yRx).

The partial preorder $\preceq_{s\neg a}$ can be decomposed into its asymmetric part $\prec_{s\neg a}$ and its symmetric part $\sim_{s\neg a}$ in the following manner: given a set of rules X and two rules $r_1, r_2 \in X, r_1 \prec_{s\neg a} r_2$ if and only if

$$\begin{array}{l} sup(r_1) \leq sup(r_2) \ \wedge \ anti - sup(r_1) > anti - sup(r_2), \text{ or} \\ sup(r_1) < sup(r_2) \ \wedge \ anti - sup(r_1) \geq anti - sup(r_2) \end{array} \tag{6}$$

moreover, $r_1 \sim_{s\neg a} r_2$ if and only if

$$sup(r_1) = sup(r_2) \ \wedge \ anti - sup(r_1) = anti - sup(r_2). \tag{7}$$

If for a rule $r \in X$ there does not exist any rule $r' \in X$, such that $r \prec_{s\neg a} r'$ then r is said to be non–dominated (i.e. Pareto–optimal) with respect to support and anti-support. A set of all non-dominated rules forms a *Pareto–optimal border* of the set of rules in the evaluation space. A set of all non-dominated rules with respect to support and anti-support will be called a *support-anti-support Pareto-optimal border*. In other words, it is the set of rules such that there is no other rule having greater support and smaller anti-support.

The approach to evaluation of the set of rules in terms of two interestingness measures being rule support and anti-support was proposed and presented in detail in Brzezińska, Greco and Słowiński (2007), and later also considered in Słowiński et al. (2007). The idea of combining those two dimensions came as a result of looking for a set of rules that would include all rules optimal with respect to any measure with the desirable property M. It was proved by Brzezińska, Greco and Słowiński (2007) that the best rules according to any measure with M must reside in the set of rules non-dominated with respect to support and anti-support:

THEOREM 1 *When considering rules with the same conclusion, rules that are optimal with respect to any interestingness measure that has the property M must reside on the support–anti-support Pareto-optimal border.*

Thus, we can consider satisfying of the property of monotonicity M by a measure as a sufficient condition for stating that rules optimal with respect to this measure will be found on the support–anti-support Pareto-optimal border. It is a valuable result as it unveils relationships between different interestingness measures. Among the practical applications of the above result, one can mention potential efficiency gains as rules optimal with respect to measures with the property M can be found in the support–anti-support Pareto-optimal set instead of searching through the set of all rules. Moreover, rule evaluation can be narrowed down to mining only the support–anti-support Pareto-optimal set instead of conducting rule evaluation separately with respect to many measures with property M, as we are sure that rules optimal according to any of them, are in that Pareto set.

Fig. 1 presents a general outlook of the support–anti-support evaluation space. Since anti-support is a cost-type criterion (the smaller its value the better), the shape of the support–anti-support Pareto-optimal border resembles a curve concave up.

Another valuable and practical feature of the support–anti-support Pareto-optimal border is that it contains the set of non-dominated rules with respect to another evaluation space based on support and confidence. The confidence of a rule (Agrawal, Imielinski and Swami, 1993), denoted as $conf(\phi \to \psi)$ is a popular interestingness measure defined as:

$$conf(\phi \to \psi) = \frac{sup(\phi \to \psi)}{sup(\phi)}. \tag{8}$$

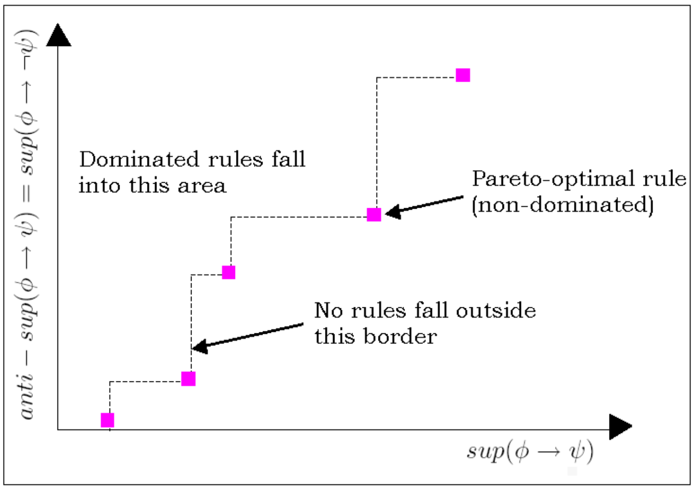

Figure 1. Support–anti-support Pareto-optimal border

The support–confidence evaluation space was proposed by Bayardo and Agrawal (1999), who proved that for rules with the same conclusion, rules that are optimal with respect to many interestingness measures such as Laplace (Clark and Boswell, 1991; Webb, 1995), lift (IBM, 1996) (also known as *interest*, Brin et al., 1997, or *strength*, Dhar and Tuzhilin, 1993), conviction (Brin et al., 1997), and other will reside on the support–confidence Pareto-optimal border. This evaluation space was also considered in Słowiński Brzezińska and Greco (2006). Thorough analysis conducted in Brzezinska, Greco and Słowiński (2007) showed that the support–confidence Pareto-optimal border has the advantage of presenting a smaller number of rules (more precisely a not greater number of rules) than the support–anti-support Pareto-optimal border. However, its disadvantage is that it does not present the rules optimizing any attractiveness measure satisfying the property M. In fact, all the rules which are present on the support–anti-support Pareto-optimal border and not present on the support–confidence Pareto-optimal border maximize an attractiveness measure which is not monotone with respect to support.

3. Analysis of property M

For the simplicity of presentation, the following notation shall be used throughout the next sections: $a = sup(\phi \to \psi)$, $b = sup(\neg\phi \to \psi)$, $c = sup(\phi \to \neg\psi)$, $d = sup(\neg\phi \to \neg\psi)$, $a + c = sup(\phi)$, $a + b = sup(\psi)$, $b + d = sup(\neg\phi)$, $c + d = sup(\neg\psi)$, $a + b + c + d = |U|$. We also assume that set U is not empty, so that at least one of a, b, c or d is strictly positive. Moreover, we also assume that any value in the denominator of any ratio is different from zero.

In order to prove that a measure has the property M we need to show that it is non-decreasing with respect to a and d, and non-increasing with respect to b and c.

THEOREM 2 *Measure RI has the property M.*

Proof. Let us observe that measure RI can be rewritten as:

$$RI(\phi \to \psi) = a - \frac{(a+b)(a+c)}{a+b+c+d}. \tag{9}$$

After some simple algebraic transformation, we obtain

$$RI(\phi \to \psi) = \frac{ad-bc}{a+b+c+d}. \tag{10}$$

Taking into account equation (10), to prove the monotonicity of RI with respect to a we have to show that if a increases by $\Delta > 0$, then RI does not decrease, i.e.

$$\frac{(a+\Delta)d-bc}{a+b+c+d+\Delta} - \frac{ad-bc}{a+b+c+d} \geq 0. \tag{11}$$

After few simple algebraic passages, and remembering that a, b, c, d and Δ are non-negative, we get

$$\begin{aligned}&\frac{(a+\Delta)d-bc}{a+b+c+d+\Delta} - \frac{ad-bc}{a+b+c+d} = \\ &= \frac{b(b+c+d)\Delta + bc\Delta}{(a+b+c+d)(a+b+c+d+\Delta)} > 0 \geq 0\end{aligned} \tag{12}$$

so that we can conclude that RI is non-decreasing (more precisely, strictly increasing) with respect to a. Analogous proof holds for the monotonicity of RI with respect to d.

Now, to prove the monotonicity of RI (10) with respect to b we have to show that an increase of b by $\Delta > 0$, will not result in an increase of RI, i.e.

$$\frac{ad-(b+\Delta)c}{a+b+c+d+\Delta} - \frac{ad-bc}{a+b+c+d} \leq 0. \tag{13}$$

Through simple algebraic transformations we get that:

$$\begin{aligned}&\frac{ad-(b+\Delta)c}{a+b+c+d+\Delta} - \frac{ad-bc}{a+b+c+d} = \\ &= -\frac{c(a+c+d)\Delta + ad\Delta}{(a+b+c+d)(a+b+c+d+\Delta)} < 0 \leq 0.\end{aligned} \tag{14}$$

Since a, b, c, d and Δ are non-negative, we can conclude that RI is non-increasing (more precisely, strictly decreasing) with respect to b. Analogous proof holds for the monotonicity of RI with respect to c. ■

THEOREM 3 *The gain measure has the property M.*

Proof. Let us consider the gain measure expressed as follows:

$$gain(\phi \rightarrow \psi) = a - \Theta(a + c) \quad (15)$$

where Θ is a fractional constant between 0 and 1. As $gain(\phi \rightarrow \psi)$ does not depend on b nor d, it is clear that the change of b or d does not result in any change of $gain(\phi \rightarrow \psi)$. Thus, we only need to verify if :

- (i) the increase of a results in non-decrease of $gain(\phi \rightarrow \psi)$,
- (ii) the increase of c results in non-increase of $gain(\phi \rightarrow \psi)$.

Condition (i). Let us assume that $\Delta > 0$ is the value by which a increases. Condition (i) will be satisfied if and only if

$$gain(\phi \rightarrow \psi) = a - \Theta(a + c) \leq gain'(\phi \rightarrow \psi) = (a + \Delta) - \Theta(a + \Delta + c) \quad (16)$$

Let us observe that

$$\begin{aligned} & a - \Theta(a + c) \leq (a + \Delta) - \Theta(a + \Delta + c) \Leftrightarrow \\ & \Leftrightarrow a - a\Theta - c\Theta \leq a + \Delta - a\Theta - c\Theta - \Theta\Delta \Leftrightarrow \\ & \Leftrightarrow \Delta - \Theta\Delta \geq 0 \Leftrightarrow \Delta(1 - \Theta) \geq 0. \end{aligned} \quad (17)$$

The last inequality is always satisfied as $\Delta > 0$ and $(1 - \Theta) \geq 0$, because Θ is a fractional constant between 0 and 1. Thus, condition (i) is satisfied.

Condition (ii). Let us assume that $\Delta > 0$ is the value, by which c increases. Condition (ii) will be satisfied if and only if

$$gain(\phi \rightarrow \psi) = a - \Theta(a + c) \geq gain'(\phi \rightarrow \psi) = a - \Theta(a + \Delta + c). \quad (18)$$

Let us observe that

$$\begin{aligned} & a - \Theta(a + c) \geq a - \Theta(a + \Delta + c) \Leftrightarrow \\ & \Leftrightarrow a - a\Theta - c\Theta \geq a - a\Theta - c\Theta - \Theta\Delta \Leftrightarrow \\ & \Leftrightarrow 0 \geq -\Theta\Delta \Leftrightarrow \Delta\Theta \geq 0. \end{aligned} \quad (19)$$

The last inequality is always satisfied as $\Delta > 0$ and $\Theta \geq 0$. Thus, condition (ii) is satisfied. Since all four conditions are satisfied, the hypothesis that gain measure has the property M is true. ■

Having determined that both of the analyzed measures do satisfy the desired property M, we can draw conclusion that rules optimal according to them will be found on the support–anti-support Pareto-optimal border.

Now, let us prove by counterexample that the dependency factor $\eta(\phi \rightarrow \psi)$ does not have the property M.

THEOREM 4 *Dependency factor $\eta(\phi \rightarrow \psi)$ does not have the property M.*

Proof. Let us consider the dependency factor rewritten as follows:

$$\eta(\phi \rightarrow \psi) = \frac{\frac{a}{a+c} - \frac{a+b}{a+b+c+d}}{\frac{a}{a+c} + \frac{a+b}{a+b+c+d}}. \quad (20)$$

It will be shown by the following counterexample that $\eta(\phi \to \psi)$ does not satisfy the condition that the increase of a results in non-decrease of $\eta(\phi \to \psi)$, thus this measure does not have the property M. Let us consider case α, in which a=7, b=2, c=3, d=3, and case α', in which a increases to 8 and b, c, d remain unchanged. The dependency factor does not have the property M as such increase of a results in the decrease of the measure:

$$\eta(\phi \to \psi) = 0.0769 > 0.0756 = \eta'(\phi \to \psi). \tag{21}$$

■

4. Experimental illustration of the result

It was proved by Brzezińska, Greco and Słowiński (2007) that rules optimal with respect to any interestingness measure that has the property M will reside on the support–anti-support Pareto-optimal border. Since the above analysis shows that both *RI* and *gain* satisfy the property M, we can conclude that rules optimal with respect to them will be found in the set of rules non-dominated according to support and anti-support. Several computational experiments analyzing rules optimal with respect to *RI* and *gain* in the perspective of rule support and anti-support have been conducted in order to illustrate the theoretical results concerning their possession of the property M and thus their occurrence on the support–anti-support Pareto-optimal border.

Fig. 2 shows an exemplary diagram from those experiments. For a real life dataset containing information about technical state of buses, a set of all possible rules was generated. A set of 85 rules with the same conclusion was then isolated, and dominated and non-dominated rules with respect to support and anti-support were found. The support–anti-support Pareto-optimal border is indicated in Fig. 2 by circles connected by a line. Four points marked as r_1, r_2, r_3, r_4 form the Pareto-optimal border. Each of those points represents rules characterized by particular values of support and anti-support (i.e., r_1 represents rules with $sup(\phi \to \psi) = 50$ and $anti - sup(\phi \to \psi) = 4$, r_2 rules with $sup(\phi \to \psi) = 49$ and $anti - sup(\phi \to \psi) = 2$, r_3 rules with $sup(\phi \to \psi) = 48$ and $anti - sup(\phi \to \psi) = 1$, and r_4 rules with $sup(\phi \to \psi) = 45$ and $anti - sup(\phi \to \psi) = 0$). In the generated set of 85 rules, we have distinguished rules optimal according to *RI* (marked by r_3), and *gain* for different values of Θ. For $\Theta = 0.33$ the rules with maximal *gain* are marked as r_1; when $\Theta = 0.5$ these are the rules marked as r_2 or r_3; finally when $\Theta = 0.66$ these are the rules marked as r_3. The diagram shows that, indeed, rules optimal with respect to those measures lie on the support–anti-support Pareto-optimal border. It means that rules optimal with respect to *RI* or *gain* can be found more efficiently by looking for them in the support–anti-support Pareto-optimal set instead of searching the set of all rules. Moreover, if the user is not interested in knowing which particular rules are optimal according to *RI* or *gain*, we can narrow down the data mining process to searching only for the support–anti-support

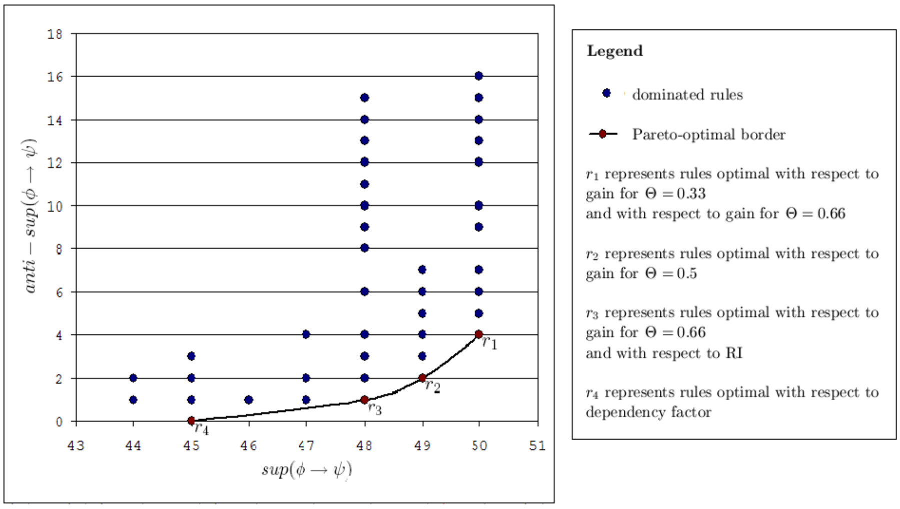

Figure 2. Pareto-optimal border with respect to rule support and anti-support includes rules optimal in *RI* and *gain*

Pareto-optimal set because we are sure that it contains *RI* and *gain*-optimal rules (though we do not know which ones they are).

During this experiment we have also calculated the optimal value of the dependency factor. This measure does not have the property M, so we could not conclude right away that rules optimal according to it will be on the support–anti-support Pareto-optimal border. However, since possession of the property M is only a sufficient condition for lying on that border, we cannot exclude a situation in which rules optimal with respect to the dependency factor will be found on the support–anti-support Pareto-optimal border. For this dataset we have such a case. Rules marked as r_4 are optimal according to dependency factor and they also belong to the set of non-dominated rules with respect to support and anti-support. Thus, r_4 can also be regarded as a counterexample proving that possession of property M is not a necessary condition for lying on the support–anti-support Pareto-optimal border.

5. Analysis of hypothesis symmetry (HS)

The verification of the property of hypothesis symmetry was done for all three considered measures separately, by checking if their values for rules $\phi \to \psi$ and $\phi \to \neg\psi$ are the same but of opposite sign.

THEOREM 5 *Measure RI has the property of hypothesis symmetry.*

Proof. Let us consider RI expressed as follows:

$$RI(\phi \to \psi) = a - \frac{(a+c)(a+b)}{a+b+c+d}. \tag{22}$$

For a negated conclusion RI is defined as:

$$RI(\phi \to \neg\psi) = c - \frac{(a+c)(c+d)}{a+b+c+d}. \tag{23}$$

The hypothesis symmetry will be satisfied by RI if and only if:

$$a - \frac{(a+c)(a+b)}{a+b+c+d} = -[c - \frac{(a+c)(c+d)}{a+b+c+d}]. \tag{24}$$

Through simple mathematical transformation we obtain that:

$$a - \frac{(a+c)(a+b)}{a+b+c+d} = \frac{ad-bc}{a+b+c+d} \tag{25}$$

and

$$-c + \frac{(a+c)(c+d)}{a+b+c+d} = \frac{ad-bc}{a+b+c+d} \tag{26}$$

and thus, we can conclude that RI has the property of hypothesis symmetry. ■

THEOREM 6 *The gain measure has the property of hypothesis symmetry if and only if* $\Theta = 1/2$.

Proof. Let us consider *gain* expressed as follows:

$$gain(\phi \to \psi) = a - \Theta(a+c). \tag{27}$$

For a negated conclusion *gain* is defined as:

$$gain(\phi \to \neg\psi) = c - \Theta(a+c). \tag{28}$$

The hypothesis symmetry will be satisfied by *gain* if and only if:

$$a - \Theta(a+c) = -[c - \Theta(a+c)]. \tag{29}$$

Through simple mathematical transformation we obtain that the above equality is satisfied only when

$$a + c = 2\Theta(a+c) \tag{30}$$

that is, when $\Theta = 1/2$. ■

THEOREM 7 *The dependency factor* η *does not have the property of hypothesis symmetry.*

Proof. Let us consider dependency factor expressed as follows:

$$\eta(\phi \to \psi) = \frac{\frac{a}{a+c} - \frac{a+b}{a+b+c+d}}{\frac{a}{a+c} + \frac{a+b}{a+b+c+d}}. \tag{31}$$

For a negated conclusion it is defined as:

$$\eta(\phi \to \neg\psi) = \frac{\frac{c}{a+c} - \frac{c+d}{a+b+c+d}}{\frac{c}{a+c} + \frac{c+d}{a+b+c+d}}. \tag{32}$$

To prove that the dependency factor does not satisfy the hypothesis symmetry let us use the following counterexample. Let us consider a situation in which $a = b = c = 10$ and $d = 20$. We can easily verify that

$$\eta(\phi \to \psi) = 0.11 \neq 0.09 = \eta(\phi \to \neg\psi). \tag{33}$$

■

6. Conclusions

As an active research area in data mining, rule evaluation has been considered by many authors from different perspectives. This paper concentrated on measuring the relevance and utility of induced rules according to three popular interestingness measures: rule interest function of Piatetsky-Shapiro, gain measure of Fukuda et al., and dependency factor of Pawlak.

A theoretical analysis has been conducted verifying which of those measures satisfy valuable properties M and hypothesis symmetry (HS). It has been proved that the rule interest function and gain measure are characterized by both of those properties, while the dependency factor does not satisfy any of them. Such analysis of properties of interestingness measures was carried out in order to widen our knowledge and understanding of those measures, and of their applicability.

Since measures *RI* and *gain* satisfy property M, they can be regarded as functions non-decreasing with respect to $sup(\phi \to \psi)$ and $sup(\neg\phi \to \neg\psi)$, and non-increasing with respect to $sup(\neg\phi \to \psi)$ and $sup(\phi \to \neg\psi)$. Moreover, the possession of the property M unveils an interesting relationship between rule interest function and *gain* on one hand, and two other interestingness measures: rule support and anti-support, on the other hand. It has been shown that rules maximizing rule interest function or *gain* will surely be found on the rule support–anti-support Pareto-optimal border (when considering rules with the same conclusion). Thus, one can concentrate on mining the set of non-dominated rules with respect to support and anti-support and be sure to obtain in that set all rules that are optimal with respect to any measure with the property M, which includes rule interest function and gain measure. These results have also been illustrated on an exemplary dataset, containing information about technical state of buses.

The results obtained are useful for practical applications because they show which interestingness measures are relevant for meaningful rule evaluation. By using the measures which enjoy the desirable properties one can avoid analyzing unimportant rules.

References

AGRAWAL, R., IMIELINSKI, T. and SWAMI, A. (1993) Mining Associations Between Sets of Items in Massive Databases. In: P. Buneman and S. Jajodia, eds., *Proceedings of the 1993 ACM-SIGMOD Int'l Conf. on Management of Data*, Washington, D.C., USA. ACM Press, 207–216.

BAYARDO, R.J. and AGRAWAL, R. (1999) Mining the Most Interesting Rules. *Proceedings of Fifth ACM-SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, San Diego, CA, USA. ACM Press, New York, NY, 145–154.

BRAMER, M. (2007) *Principles of Data Mining.* Springer-Verlag, New York.

BRIN, S., MOTWANI, R., ULLMAN, J. and TSUR, S. (1997) Dynamic itemset counting and implication rules for market basket data. *Proceedings of the 1997 ACM-SIGMOD Int'l Conf. on the Management of Data.* ACM Press, New York, NY, 255–264.

BRZEZIŃSKA, I., GRECO, S. and SŁOWIŃSKI, R. (2007) Mining Pareto-optimal rules with respect to support and anti-support. *Engineering Applications of Artificial Intelligence*, **20** (5), 587–600.

CARNAP, R. (1962) *Logical Foundations of Probability, 2nd ed.* University of Chicago Press, Chicago.

CLARK, P. and BOSWELL, P. (1991) Rule Induction with CN2: Some Recent Improvements. *Proceedings of the European Working Session on Machine Learning.* **LNCS 482**, Springer, London, 151–163.

DHAR, V. and TUZHILIN, A. (1993) Abstract-driven pattern discovery in databases. *IEEE Transactions on Knowledge and Data Engineering* **5** (6), 926-938.

EARMAN, J. (1992) *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory.* MIT Press, Cambridge, MA.

EELLS, E and FITELSON, B. (2002) Symmetries and assymmetries in evidential support. *Philosophical Studies*, **107** (2), 129–142.

FITELSON, B. (2001) *Studies in Bayesian confirmation theory.* Ph.D. Thesis, University of Wisconsin, Madison.

FUKUDA, T., MORIMOTO, Y., MORISHITA, S. and TOKUYAMA, T. (1996) Data Mining using Two-Dimensional Optimized Association Rules: Schemes, Algorithms, and Visualization. *Proceedings of the 1996 ACM SIGMOD Int'l Conference on Management of Data.* Montreal, Canada. ACM Press, New York, 13–23.

GRECO, S., PAWLAK, Z. and SŁOWIŃSKI, R. (2004) Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence* **17**, 345–361.

HEMPEL, C.G. (1945) Studies in the logic of confirmation (I). *Mind* **54**, 1–26.

HILDERMAN, R. and HAMILTON, H. (2001) *Knowledge Discovery and Measures of Interest.* Kluwer Academic Publishers, Boston.

INTERNATIONAL BUSINESS MACHINES (1996) *IBM Intelligent Miner User's Guide* Version 1, Release 1.

MORZY, T. and ZAKRZEWICZ, M. (2003) Data mining. In: J. Błażewicz, W. Kubiak, T. Morzy, M.E. Rusinkiewicz, eds., *Handbook on Data Management in Information Systems.* Springer-Verlag, 487–565.

PAWLAK, Z. (2004) Some issues on Rough Sets. *Transactions on Rough Sets I.* **LNCS 3100**, 1–58.

PIATETSKY-SHAPIRO, G. (1991) Discovery, analysis and presentation of strong rules. *Knowledge Discovery in Databases.* AAAI/MIT Press, **2**, 29–248.

POPPER, K.R. (1959) *The Logic of Scientific Discovery.* Hutchinson, London.

SILBERSCHATZ, A. and TUZHILIN, A. (1996) What Makes Patterns Interesting in Knowledge Discovery Systems. *IEEE Transactions on Knowledge and Data Engineering* **8** (6), 970–974.

SŁOWIŃSKI, R., BRZEZIŃSKA, I. and GRECO, S. (2006) Application of Bayesian confirmation measures for mining rules from support-confidence Pareto-optimal set. Invited paper in: L. Rutkowski, R. Tadeusiewicz, L.A. Zadeh, J. Zurada, eds., *Artificial Intelligence and Soft Computing.* **LNAI 4029**, Springer-Verlag, Berlin, 1018–1026.

SŁOWIŃSKI, R., SZCZĘCH, I., URBANOWICZ, M. and GRECO, S. (2007) Mining association rules with respect to support and snti-support - experimental results. In: M. Kryszkiewicz, J.F. Peters, H. Rybiński, A. Skowron, eds., *Rough Sets and Intelligent Systems Paradigms.* **LNAI 4585**, Springer-Verlag, Berlin, 534–542.

TAN, P-N., STEINBACH, M. and KUMAR, V. (2006) *Introduction to Data Mining.* Pearson Education, Inc., USA.

WEBB, G.I. (1995) OPUS: An efficient admissible algorithm for unordered search. *Journal of Artificial Intelligence Research*, **3**, 431–465.