Mining Pareto-Optimal Rules with Respect to Support and Confirmation or Support and Anti-Support

Izabela Brzezinska[1], Salvatore Greco[2], Roman Slowinski[1,3]

[1]Institute of Computing Science, Poznan University of Technology,
Piotrowo 2, 60-965 Poznan, Poland
e-mail: {izabela.brzezinska, roman.slowinski}@cs.put.poznan.pl
[2] Faculty of Economics, University of Catania, Corso Italia, 55, 95129 Catania, Italy
e-mail: salgreco@unict.it
[3] Institute for Systems Research, Polish Academy of Sciences, 01-447 Warsaw, Poland

Abstract: In knowledge discovery and data mining many measures of interestingness have been proposed in order to measure the relevance and utility of the discovered patterns. Among these measures, an important role is played by Bayesian confirmation measures, which express in what degree a premise confirms a conclusion. In this paper, we are considering knowledge patterns in a form of "*if…, then…*" rules with a fixed conclusion. We investigate a monotone link between Bayesian confirmation measures, and classic dimensions being rule support and confidence. In particular, we formulate and prove conditions for monotone dependence of two confirmation measures enjoying some desirable properties on rule support and confidence. As the confidence measure is unable to identify and eliminate non-interesting rules, for which a premise does not confirm a conclusion, we propose to substitute the confidence for one of the considered confirmation measures in mining the Pareto-optimal rules. We also provide general conclusions for the monotone link between any confirmation measure enjoying the desirable properties and rule support and confidence. Finally, we propose to mine rules maximizing rule support and minimizing rule anti-support, which is the number of examples which satisfy the premise of the rule but not its conclusion (called counter-examples of the considered rule). We prove that in this way we are able to mine all the rules maximizing any confirmation measure enjoying the desirable properties. We also prove that this Pareto-optimal set includes all the rules from the previously considered Pareto-optimal borders.

Keywords: Knowledge discovery; Association rules; Decision rules; Bayesian confirmation measures; Monotonicity in rule support, confidence and confirmation; Pareto-optimal rules

1. Introduction

Discovering knowledge from data is the domain of inductive reasoning. Knowledge patterns induced from data are usually expressed in a form of "*if…, then…*" rules. They are consequence relations representing correlation, association, causation between independent and dependent attributes. If the division into independent and dependent attributes has been fixed, the rules mined from data are regarded as decision rules, otherwise as association rules. Typically, the number of rules generated from massive datasets is quite large, but only a few of them are likely to be useful for the domain expert analysing the data. Therefore, in order to measure the relevance and utility of the discovered patterns, quantitative measures, also known as attractiveness or interestingness measures (metrics), have been proposed and studied. There is a number of widely known interestingness measures such as support and confidence [1], gain [11], and conviction [3]. An important place is taken, moreover, by some Bayesian confirmation measures. In general, Bayesian confirmation measures quantify the degree to which a piece of evidence built of the independent attributes provides "evidence for or against" or "support for or against" the hypothesis built of the dependent attributes [10]. Confirmation measures enjoy different properties related to specific understanding of symmetry [9] and monotonicity [13]. Taking these properties into account, we focus our attention in this study on two specific Bayesian confirmation measures: measure f (denotation used in [10] and other studies), and measure s (proposed in [7]) (see definitions of measures f and s in Section 2.3.1).

Bayardo and Agrawal have proved in [2] that for a class of rules with fixed conclusion, the set of non-dominated, Pareto-optimal rules with respect to both rule support and confidence (i.e. the upper support-confidence Pareto-optimal border) includes optimal rules according to several different interestingness measures, such as gain, Laplace [8], lift [18], conviction, and unnamed measure proposed by Piatetsky-Shapiro [24]. This practically useful result allows to identify the most interesting rules according to several interestingness measures by solving an optimised rule mining problem with respect to rule support and confidence only.

As shown in [13], the semantics of the scale of confidence is not as meaningful as that of confirmation measures. Moreover, it has been analytically shown in [4] that there exists a monotone link between some confirmation measures on one side, and confidence and support, on the other side. In consequence, we propose in this paper, three alternative approaches to mining interesting rules. The first one consists in searching for a Pareto-optimal border with respect to rule support and confirmation measure *f*, the second concentrates on searching for a Pareto-optimal border with respect to rule support and confirmation measure *s*, and the last one proposes to search for a Pareto-optimal border with respect to rule support and rule anti-support which is the number of examples which satisfy the premise of the rule but not its conclusion (called counter-examples of the considered rule). Of course, in the last approach the rule anti-support is minimized. We prove that it allows to mine all rules maximizing any confirmation measure enjoying some desirable properties. We also prove that the support-anti-support Pareto-optimal set includes all the rules from the previously considered Pareto-optimal borders.

The paper is organized as follows. In the next section, there are preliminaries on rules and their quantitative description. In section 3, we investigate the idea and the advantages of mining rules constituting Pareto-optimal border with respect to support and confirmation measure *f*. Section 4 concentrates on the proposal of mining Pareto-optimal rules with respect to support and confirmation measure *s*. In section 5, we generalize the approach from sections 3 and 4 to a broader class of confirmation measures. In section 6 we consider Pareto-optimal set of rules with respect to support and anti-support. The paper ends with conclusions.

2. Preliminaries

Since discovering rules from data is the domain of inductive reasoning, its starting point is a sample of larger reality often given in a 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* such that $a : U \rightarrow V_a$ for every $a \in A$. The set V_a is a domain of a. Let us associate a formal language L of logical formulas with every subset of attributes. Formulas for a subset $B \subseteq A$ are built up from attribute-value pairs (a,v), where $a \in B$ and $v \in V_a$, using logical connectives $\neg$ (not), $\wedge$ (and), $\vee$ (or). A *rule* induced from S and expressed in L is denoted by $\phi \rightarrow \psi$ (read as "*if* ϕ, *then* ψ"). It consists of antecedent ϕ and consequent ψ, being formulas expressed in L, called *premise* and *conclusion*, respectively, and therefore it can be seen as a consequence relation (see critical discussion about interpretation of rules as logical implications in [13]) between premise and conclusion. The rules mined from data may be either *decision* rules or *association* rules, depending on whether the division of A into condition and decision attributes has been fixed or not.

2.1. Monotonicity

Let x be an element of a set of rules X and let $g(x)$ be a real function associated with this set, such that $g: X \rightarrow \mathbf{R}$. Assuming an ordering relation $\succ$ in X, function g is said to be monotone (resp. anti-monotone) in x, if for any $x,y \in X$, relation $x \succ y$ implies that $g(x) \geq g(y)$ (resp. $g(x) \leq g(y)$).

2.2. Support and confidence measures of rules

With every rule induced from data table S two coefficients called *support* and *confidence* can be associated. The *support* of condition ϕ, denoted as $sup(\phi)$, is equal to the number of objects in U having property ϕ. The support of rule $\phi \rightarrow \psi$, denoted as $sup(\phi \rightarrow \psi)$, is equal to the number of objects in U having both property ϕ and ψ; for those objects, both conditions ϕ and ψ evaluate to true.

The *confidence* of a rule (also called *certainty*), denoted as $conf(\phi \rightarrow \psi)$, is defined as follows:

$$conf(\phi \rightarrow \psi) = \frac{sup(\phi \rightarrow \psi)}{sup(\phi)}.$$

Obviously, when considering rule $\phi \rightarrow \psi$, it is reasonable to assume that the set of objects having property ϕ is not empty, i.e. $sup(\phi) \neq \varnothing$.

Under the "closed world assumption" adopted in inductive reasoning, and because U is a finite set, it is legitimate to express probabilities $\Pr(\phi)$ and $\Pr(\psi)$ in terms of frequencies $sup(\phi)/|U|$ and $sup(\psi)/|U|$, respectively. In consequence, the confidence measure $conf(\phi \rightarrow \psi)$ can be regarded as conditional probability $\Pr(\psi|\phi) = \Pr(\phi \wedge \psi)/\Pr(\phi)$ with which conclusion ψ evaluates to true, given that premise ϕ evaluates to true.

2.3. Bayesian confirmation measures

Bayesian confirmation measures constitute an important group of interestingness measures. In general, they say in what degree a piece of evidence in premise confirms a hypothesis in the conclusion. According to Fitelson [10], measures of confirmation quantify the degree to which a premise ϕ provides "support for or against" a conclusion ψ. In this context, a confirmation measure denoted by $c(\phi, \psi)$ is required to satisfy the following definition:

$$c(\phi, \psi) = \begin{cases} > 0 & if \ \Pr(\psi \mid \phi) > \Pr(\psi), \\ = 0 & if \ \Pr(\psi \mid \phi) = \Pr(\psi), \\ < 0 & if \ \Pr(\psi \mid \phi) < \Pr(\psi). \end{cases}$$

Many authors have considered desirable properties of confirmation measures. Eells and Fitelson have analysed in [9] a set of best-known confirmation measures from the viewpoint of the following four properties of symmetry introduced by Carnap in [6]:

- evidence symmetry (ES): $c(\phi, \psi) = -c(\neg\phi, \psi)$
- commutativity symmetry (CS): $c(\phi, \psi) = c(\psi, \phi)$
- hypothesis symmetry (HS): $c(\phi, \psi) = -c(\phi, \neg\psi)$

- total symmetry (TS): $c(\phi,\psi)=c(\neg\phi,\neg\psi)$.

It has been concluded in [9] that, in fact, only (HS) is a desirable property, while (ES), (CS) and (TS) are not.

Greco, Pawlak and Slowinski have considered in [13] Bayesian confirmation measures from the viewpoint of their usefulness for measuring interestingness of decision rules. In this context, given decision rule $\phi\rightarrow\psi$, confirmation measure $c(\phi,\psi)$ should give the credibility of the proposition: ψ is satisfied more frequently when ϕ is satisfied rather than when ϕ is not satisfied. According to [13], in order to satisfy fully this requirement, the confirmation measure should enjoy a property, called property of monotonicity (M), defined as follows:

(M) $c(\phi\rightarrow\psi)=$
$F[sup(\phi\rightarrow\psi), sup(\neg\phi\rightarrow\psi),$
$sup(\phi\rightarrow\neg\psi), sup(\neg\phi\rightarrow\neg\psi)]$

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

The property of monotonicity (M) of $c(\phi,\psi)$ with respect to $sup(\phi\rightarrow\psi)$ (or, analogously, with respect to $sup(\neg\phi\rightarrow\neg\psi)$) 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 $\phi\rightarrow\psi$. On the other hand, the property of monotonicity of $c(\phi,\psi)$ with respect to $sup(\neg\phi\rightarrow\psi)$ (or, analogously, with respect to $sup(\phi\rightarrow\neg\psi)$) 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 $\phi\rightarrow\psi$.

2.3.1. Bayesian confirmation measures *f* and *s*

Among the best-known and widely studied confirmation measures, there are confirmation measures denoted by *f* and *s*, defined as follows:

$$f(\phi\rightarrow\psi)=\frac{\Pr(\phi\mid\psi)-\Pr(\phi\mid\neg\psi)}{\Pr(\phi\mid\psi)+\Pr(\phi\mid\neg\psi)},$$

$$s(\phi\rightarrow\psi)=\Pr(\psi\mid\phi)-\Pr(\psi\mid\neg\phi).$$

Taking into account that conditional probability $\Pr(\circ\mid *)=conf(\circ\rightarrow *)$, confirmation measures *f* and *s* can be re-written as:

$$f(\phi\rightarrow\psi)=\frac{conf(\psi\rightarrow\phi)-conf(\neg\psi\rightarrow\phi)}{conf(\psi\rightarrow\phi)+conf(\neg\psi\rightarrow\phi)},$$

$$s(\phi\rightarrow\psi)=conf(\phi\rightarrow\psi)-conf(\neg\phi\rightarrow\psi).$$

Among authors advocating for confirmation measure *f*, there are Good [12], Heckerman [14], Pearl [23] and Fitelson [10]. Measure *s* has been proposed by Christensen [7] and Joyce [19]. It is worth noting that confirmation measure *f* is monotone (and therefore gives the same ranking) with respect to the Bayes factor originally proposed by Jeffrey [16] and reconsidered as an interestingness measure by Kamber and Shingal [17] having the following formulation:

$$k(\phi\rightarrow\psi)=\frac{conf(\psi\rightarrow\phi)}{conf(\neg\psi\rightarrow\phi)}.$$

Confirmation measures *f* and *s* play an important role in the whole group of confirmation measures for having the desirable property of monotonicity (M), which was verified in [13].

Moreover, as shown in [9], confirmation measure *f* enjoys the most useful version of the property of symmetry (it satisfies hypothesis symmetry (HS), and does not satisfy other symmetry properties (ES), (CS) and (TS)). The reason in favor of (HS) is that the significance of ϕ with respect to ψ should be of the same strength, but of opposite sign, as the significance of ϕ with respect to $\neg\psi$. The arguments against (ES), (CS) and (TS) can be found in [9].

As to confirmation measure *s*, it has three properties of symmetry (HS), (ES) and (TS), while not (CS). It is attractive, however, for another reason: it reflects the rule support better than confirmation measure *f*. This can be seen from the following hypothetical example: assume two different universes, composed of 110 objects each – they will correspond to case a) and b). In case a), $sup(\phi)$=10, $sup(\neg\phi)$=100, $sup(\psi)$=10, $sup(\neg\psi)$=100, $sup(\phi\rightarrow\psi)$=4, $sup(\neg\phi\rightarrow\psi)$=6, and $sup(\phi\rightarrow\neg\psi)$=6. In case b), $sup(\phi)$=100, $sup(\neg\phi)$=10, $sup(\psi)$=100, $sup(\neg\psi)$=10, $sup(\phi\rightarrow\psi)$=95, $sup(\neg\phi\rightarrow\psi)$=5, and $sup(\phi\rightarrow\neg\psi)$=5. According to the above formulae, in case:

a) s_a = 34/100 = 0.34 and f_a = 34/46 = 0.739,
while in case
b) s_b = 45/100 =0.45 and f_b = 45/145 = 0.31.

As the support of rule $\phi\rightarrow\psi$ in case a) is smaller than that of case b), it is desirable that the confirmation measures reflect this fact, however, only $s_b > s_a$, while $f_b < f_a$.

2.4. Partial preorder on rules in terms of two interestingness measures

Let us denote by $\preceq_{AB}$ a partial preorder on rules in terms of any two different interestingness measures *A* and *B*. 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 (i.e. for all $x,y,z\in X$, *xRy* and *yRz* imply *xRz*). 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 $\preceq_{AB}$ can be decomposed into its asymmetric part $\prec_{AB}$ and its symmetric part $\sim_{AB}$ in the following manner:

given two rules r_1 and r_2, $r_1 \prec_{AB} r_2$ if and only if

$A(r_1) \le A(r_2) \wedge B(r_1) < B(r_2)$, or

$A(r_1) < A(r_2) \wedge B(r_1) \le B(r_2)$,

moreover $r_1 \sim_{AB} r_2$ if and only if

$A(r_1) = A(r_2) \wedge B(r_1) = B(r_2)$.

2.5. Implication of a complete preorder $\preceq_t$ by partial preorder $\preceq_{AB}$

Application of some measures that quantify the interestingness of a rule induced from an information table S creates a complete preorder, denoted as $\preceq_t$, on set of those rules. Recall that a *complete preorder* on a set X is any binary relation R on X that is strongly complete, (i.e. for all $x,y \in X$, xRy or yRx) 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 but without any incomparability. In particular, measures such as gain, Laplace, lift, conviction, measure proposed by Piatetsky-Shapiro, or confirmation measures f and s result in such a complete preorder on the set of rules, ordering them according to their interestingness value.

A complete preorder $\preceq_t$ *is implied by* a partial preorder $\preceq_{AB}$ if:

$r_1 \prec_{AB} r_2 \Rightarrow r_1 \prec_t r_2$, *and*

$r_1 \sim_{AB} r_2 \Rightarrow r_1 \sim_t r_2$.

It has been proved by Bayardo and Agrawal in [2] that if a complete preorder $\preceq_t$ is implied by a particular support-confidence partial preorder $\preceq_{sc}$, then the optimal rules with respect to $\preceq_t$ can be found in the set of non-dominated rules with respect to rule support and confidence. Thus, having proved that a complete preorder defined for a new interestingness measure is implied by $\preceq_{sc}$, one can concentrate on discovering non-dominated rules with respect to rule support and confidence. Moreover, Bayardo and Agrawal have shown in [2] that the following conditions are sufficient for proving that a complete preorder $\preceq_t$ defined over a rule value function $g(r)$ is implied by partial preorder $\preceq_{AB}$:

- $g(r)$ is monotone in A over rules with the same value of B, and
- $g(r)$ is monotone in B over rules with the same value of A.

2.6. Advantages of confirmation measures over confidence in the context of APRIORI approach.

Support and confidence are the two measures most frequently used in association rule extraction algorithms based on the selection of frequent itemsets. Having obtained from a user minimum rule support and confidence thresholds, the basic version of the well known APRIORI algorithm [1] proceeds in a two step framework:

- find frequent itemsets (i.e. sets of items which occur more frequently than the minimum support threshold),
- generate rules from frequent itemsets and filter out those that do not exceed the minimum confidence threshold.

The efficiency of this algorithm lies in the fact that for an itemset which is not frequent, none of its supersets can be frequent. Therefore, the APRIORI algorithm starts its analysis from itemsets of k=1 size and generates itemsets of size k+1 only from frequent itemsets of size k.

The second step of the algorithm generates association rules from frequent itemsets and discards those that are beyond the minimum confidence threshold. However, rules within the confidence framework might not be necessarily interesting for the experts due to a weakness of the confidence scale. This disadvantage of confidence can be easily seen when its scale is compared with the scale of any confirmation measure, from the view point of semantics.

Confidence can obtain values between 0 and 1 (where 1 is regarded as the best) whereas confirmation measures take values between –1 and 1 (again 1 being the most desirable). The utility of scale of confirmation measures outranks the utility of confidence's scale. The confidence measure has no means to show that the rule is useless when its premise disconfirms the conclusion. Such situation is expressed by a negative value of any confirmation measure, and thus useless rules can be filtered out simply by observing the confirmation measure's sign.

The difference of semantics and utility of $conf(\phi\to\psi)$ on one hand, and $f(\phi\to\psi)$ or $s(\phi\to\psi)$on the other hand, can be shown on the following example. Consider the possible result of rolling a die: 1,2,3,4,5,6, and let the conclusion ψ="the result is divisible by 2". Given two different premises: ϕ_1="the result is a number from a set {1,2,3}", ϕ_2="the result is a number from a set {2,3,4}", we get, respectively: $conf(\phi_1\to\psi)=1/3$, $f(\phi_1\to\psi)=-1/3$, $s(\phi_1\to\psi)=-1/3$, $conf(\phi_2\to\psi)=2/3$, $f(\phi_2\to\psi)=1/3$, $s(\phi_2\to\psi)=1/3$. This example, of course, acknowledges the monotone link between confirmation measure f or s and confidence. However, it also clearly shows that the values of confirmation measures have a more useful interpretation than confidence. In particular, in the case of rule $\phi_1\to\psi$, the premise actually disconfirms the conclusion as it reduces the probability of conclusion ψ from $1/2=sup(\psi)$ to $1/3=conf(\phi_1\to\psi)$. This fact is expressed by a negative value of confirmation measure f and s , but cannot be concluded by observing only the value of confidence.

The difference of semantics and utility of confidence and the two confirmation measures can also be seen when we consider two different conclusions and only one premise Let ϕ="the result is divisible by 2", while ψ_1="the result is 6" and ψ_2="the result is *not* 6". Then, $conf(\phi\to\psi_1)=1/3$, $f(\phi\to\psi_1)=3/7$, $s(\phi\to\psi_1)=1/3$ and

$conf(\phi\rightarrow\psi_2)$=2/3, $f(\phi\rightarrow\psi_2)$=–3/7, $s(\phi\rightarrow\psi_1)$=-1/3. In this example, rule $\phi\rightarrow\psi_2$ has greater confidence than rule $\phi\rightarrow\psi_1$, however, rule $\phi\rightarrow\psi_2$ is less interesting than rule $\phi\rightarrow\psi_1$ because premise ϕ reduces the probability of conclusion ψ_2 from 5/6=$sup(\psi_2)$ to 2/3=$conf(\phi\rightarrow\psi_2)$, while it augments the probability of conclusion ψ_1 from 1/6=$sup(\psi_1)$ to 1/3=$conf(\phi\rightarrow\psi_1)$. In consequence, premise ϕ disconfirms conclusion ψ_2, which is expressed by a negative value of $f(\phi\rightarrow\psi_2)$=–3/7 and $s(\phi\rightarrow\psi_2)$=–1/3, and it confirms conclusion ψ_1, which is expressed by a positive value of $f(\phi\rightarrow\psi_1)$=3/7 and $s(\phi\rightarrow\psi_2)$=1/3.

It is therefore clear, that even high values of confidence can be misleading. Thus, in order to reduce the number of induced rules, it is not enough to filter out rules that do not satisfy an assumed confidence threshold, like it is done in a basic version of the APRIORI algorithm. We find it valuable to substitute the confidence threshold by confirmation measure f or confirmation measure s threshold. Then, the second phase of the APRIORI algorithm, would be modified to generate, from the frequent itemsets, rules for which the value of confirmation measure f or s is larger than the threshold set by a user. It is obvious that only thresholds larger than 0 are reasonable.

3. Mining the Pareto-optimal border with respect to confirmation measure *f* and rule support

As proved in [2], mining the support-confidence border identifies optimal rules according to several different interestingness metrics. This is a practically useful result that assures that rules maximizing many popular measures shall be found by solving an optimised rule mining problem with respect to rule support and confidence only. It has been shown in [4] that even confirmation measure f is among those interestingness metrics and, therefore, all rules maximizing confirmation measure f can be found on the Pareto-optimal support-confidence border (concerning rules with a fixed conclusion). For the sake of completeness we give here the two fundamental results formally stating these properties.

Theorem 1. [4] Confirmation measure f is independent of rule support, and, therefore, monotone in rule support, when the value of confidence is held fixed.

Proof. Let us consider the confirmation measure f transformed such that, for given U and ψ, it only depends on confidence of rule $\phi\rightarrow\psi$ and support of ψ:

$$f(\phi\rightarrow\psi)=\frac{|U|conf(\phi\rightarrow\psi)-sup(\psi)}{(|U|-2sup(\psi))conf(\phi\rightarrow\psi)+sup(\psi)}.$$

As we consider rules with a fixed conclusion ψ, the values of $|U|$ and $sup(\psi)$ are constant. Thus, for a fixed confidence, we have a constant value of the confirmation measure f, no matter what the rule support is. Hence, confirmation measure f is monotone in rule support when the confidence is held constant. □

Theorem 2. [4] Confirmation measure f is increasing in confidence, and, therefore, monotone in confidence.

Proof. Again, let us consider confirmation measure f given in the same form as above:

$$f(\phi\rightarrow\psi)=\frac{|U|conf(\phi\rightarrow\psi)-sup(\psi)}{(|U|-2sup(\psi))conf(\phi\rightarrow\psi)+sup(\psi)}$$

For the clarity of the presentation, let us express the above formula as a function of confidence, still regarding $|U|$ and $sup(\psi)$ as constant values greater than 0:

$$y=\frac{kx-m}{nx+m},$$

where $y=f(\phi\rightarrow\psi)$, $x=conf(\phi\rightarrow\psi)$, $k=|U|$, $m=sup(\psi)$, $n=|U|-2sup(\psi)$.

It is easy to observe that:

- $k=|U|>0$, and
- $0<m\le|U|$.

In order to verify the monotonicity of f in confidence, let us differentiate y with respect to x. We obtain:

$$\frac{\partial y}{\partial x}=\frac{m(k+n)}{(nx+m)^2}.$$

As $m>0$, and $k+n=|U|+|U|-2sup(\psi)=2|U|-2sup(\psi)>0$ for $|U|\ge sup(\psi)$, the whole derivative is always not smaller than 0. Therefore, confirmation measure f is monotone in confidence. □

However, as the utility of confirmation measure f outranks the utility of confidence, it has been found interesting to propose a new Pareto-optimal border – with respect to rule support and confirmation measure f. It is valuable to combine those two measures in the border, as for a fixed value of confidence, confirmation measure f is independent of rule support, and rules that have high values of confirmation measure f are often characterized by small values of the rule support.

For the completeness of the research, an analysis of the monotonicity of confidence in rule support for a fixed value of confirmation f, as well as in confirmation f for a fixed value of support was also performed.

Corollary 1. Confidence is independent of rule support, and, therefore, monotone in rule support, when the value of confirmation measure f is held fixed.

Proof. Let us consider Bayesian confirmation measure f:

$$f(\phi\rightarrow\psi)=\frac{conf(\psi\rightarrow\phi)-conf(\neg\psi\rightarrow\phi)}{conf(\psi\rightarrow\phi)+conf(\neg\psi\rightarrow\phi)}$$

Through simple mathematical transformations, the above definition can be converted to outline how confidence depends on confirmation measure f:

$$conf(\phi\rightarrow\psi)=\frac{f(\phi\rightarrow\psi)sup(\psi)+sup(\psi)}{|U|-f(\phi\rightarrow\psi)(|U|-2sup(\psi))}.$$

As we consider a set of rules with a fixed conclusion ψ, the values of $|U|$ and $sup(\psi)$ are constant. Thus, for a fixed confirmation f, we have a constant value of confidence, no matter what the rule support is. Hence, we can conclude that confidence is monotone in rule support when the confirmation measure f is held constant. □

Corollary 2. Confidence is increasing in confirmation measure f, and therefore monotone in confirmation measure f.

Proof. Since confirmation measure f is strictly monotone (i.e. increasing) in confidence (see proof of *Theorem 2*), then it is obvious that confidence is also strictly monotone in confirmation measure f. □

The above results are in fact a verification of the two sufficient conditions proving that the complete preorder introduced by confidence is implied by the partial preorder of rule support and confirmation measure f. Thus,
Theorem 1, *Theorem 2* and *Corollary 1*, *Corollary 2* show that the set of rules located on the support-confidence Pareto-optimal border is exactly the same as the set of rules located on the support-confirmation-f Pareto-optimal border.

Moreover, it is clear that having three functions A, B and C such that A is strictly monotone in B and B is strictly monotone in C, A is also strictly monotone in C. Thus, it is straightforward to observe that other interestingness measures that are monotone in confidence, must also be monotone in confirmation measure f, due to the monotone link between confidence and confirmation measure f. Hence, all the interestingness measures that were found on the support-confidence Pareto-optimal border shall also reside on the Pareto-optimal border with respect to rule support and confirmation measure f. Then, for rules with a fixed conclusion, mining the newly proposed Pareto-optimal border will identify rules optimal according to confidence, conviction, lift, Laplace, Piatetsky-Shapiro, gain, etc.

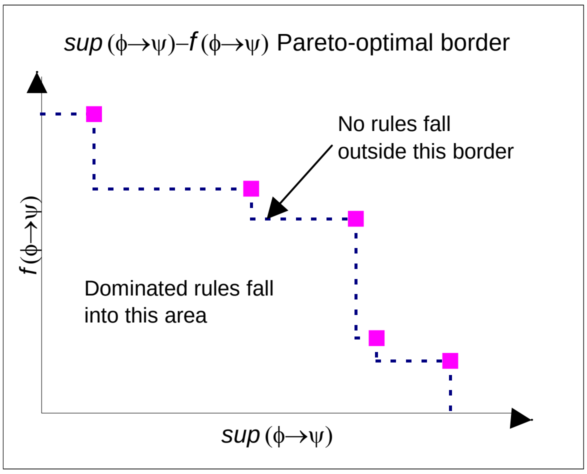

Fig.1 Pareto-optimal border with respect to rule support and confirmation measure f

However, any non-dominated rule with a negative value of confirmation measure f must be discarded from further analysis as its premise only disconfirms the conclusion. In particular, if the highest value of confirmation measure f in the Pareto set is negative, then the whole set should be excluded as it does not contain any interesting rules. In this paper, similarly to [2], a separate Pareto-optimal border is considered for each set of rules with the same conclusion. A final set of rules representing patterns discovered from the whole dataset shall be a union of all the non-negative-in-f rules from all the Pareto-optimal borders (all possible conclusions) with respect to rule support and confirmation measure f.

Several computational experiments analysing rules in confirmation measure f and rule support have been conducted in order to illustrate the theoretical results concerning the support-confirmation f Pareto-optimal border. Below, on Fig.2 is an exemplary diagram from that experiment. For a real life dataset containing information about technical state of buses a set of all possible rules was generated. A set of 78 rules with the same conclusion was then isolated and rules non-dominated with respect to support and confirmation measure f were found. Determining which of those 78 rules are optimal according to such measures as confidence, lift, Laplace, Piatetsky-Shapiro etc., has shown that, as was earlier theoretically proved, they all reside on the Pareto-optimal support-confirmation f border.

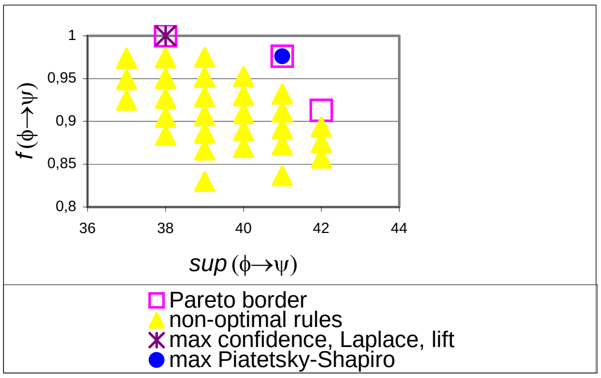

Fig.2 Pareto-optimal border with respect to rule support and confirmation measure f includes rules being optimal in many other measures

It can be easily observed that the Pareto-optimal set of rules (marked in Fig.2 by squares) includes rules maximizing such interestingness measures as confidence, Laplace, lift (marked in Fig.2 by asterisk), Piatetsky-Shapiro (marked in Fig.2 by a cross).

Summarizing, due to the fact that the utility of scale of confirmation measure f outranks the scale of confidence, we are strongly in favour of mining the Pareto-optimal border with respect to rule support and confirmation f and not rule support and confidence as it was proposed in [2]. Those two Pareto sets can, in fact, be regarded as monotone transformations of each other. Hence, substitution of confidence by confirmation measure *f* does not induce any losses to the set of other interestingness metrics for which optimal rules reside on the Pareto-optimal border. As semantics of confirmation *f* are more useful, it is straightforward to see the reasons for such substitution.

4. Mining the Pareto-optimal border with respect to confirmation measure *s* and rule support

From the group of Bayesian confirmation measures, an important role is also played by confirmation measure *s*. Similarly to confirmation measure *f*, it also has the desirable property of monotonicity (M). Its monotone link (within a set of rules with the same conclusion) with confidence and rule support came into the scope of our analysis.

For the clarity of further presentation, let us use the following notation:

$a = sup(\phi \to \psi)$,

$b = sup(\neg\phi \to \psi)$,

$c = sup(\phi \to \neg\psi)$,

$d = sup(\neg\phi \to \neg\psi)$.

Throughout the following analysis, we shall assume that *a*, *b*, *c* and *d* are positive numbers.

The confirmation measure *s* is then defined as:

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

The analysis considers only a set of rules with the same conclusion, thus the values of $|U| = a+b+c+d$ and $sup(\psi) = a+b$ are constant.

First, the monotonicity of confirmation measure *s* in confidence for a fixed value of support has been considered.

Theorem 3. When the rule support value is held fixed, then confirmation measure *s* is increasing with respect to confidence (i.e. confirmation measure *s* is monotone in confidence).

Proof. Let us consider the confirmation measure *s*:

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

For the hypothesis, $sup(\phi \to \psi) = a$ is supposed to be constant. Therefore, it is clear that $conf(\phi \to \psi) = \frac{a}{a+c}$ can only increase with the decrease of *c*. Hence, let us consider $c' = c - \Delta,$ where $\Delta > 0$. Now, operating on c' the only way to guarantee that $|U|$ and $sup(\psi)$ still remain constant is to increase *d* such that $d' = d + \Delta$. The values of *a* and *b* cannot change: $a' = a$ and $b' = b$. Now, the new value of confirmation measure *s* takes the following form:

$$s'(\phi \to \psi) = \frac{a'}{a'+c'} - \frac{b'}{b'+d'} = \frac{a}{a+c-\Delta} - \frac{b}{b+d+\Delta}.$$

Since $\Delta > 0$, it is clear that $s'(\phi \to \psi) > s(\phi \to \psi)$. This means that for a fixed value of rule support, increasing confidence results in an increase of the value of confirmation measure *s* and therefore confirmation measure *s* is monotone with respect to confidence.□

Moreover, the monotonicity of confirmation measure *s* in rule support for a fixed value of confidence has been analysed.

Theorem 4. When the confidence value is held fixed, then:

a) confirmation measure *s* is increasing in rule support (i.e. strictly monotone) if and only if *s*>0,
b) confirmation measure *s* is constant in rule support (i.e. monotone) if and only if *s*=0,
c) confirmation measure *s* is decreasing in rule support (i.e. strictly anti-monotone) if and only if *s*<0.

Proof. We present the proof of part a) only, as the other points are analogous. Let us consider the confirmation measure *s*:

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

Let us consider an increase of $sup(\phi \to \psi) = a$ expressed in the form of $a' = a + \Delta,$ where $\Delta > 0$.

Since $conf(\phi \to \psi) = \frac{a}{a+c}$ is to be constant, thus *c* should change into $c' = c + \varepsilon$ in such a way that:

$$conf(\phi \to \psi) = \frac{a}{a+c} = \frac{a'}{a'+c'} = \frac{a+\Delta}{a+\Delta+c+\varepsilon}.$$

Simple mathematical transformation lead to the conclusion that:

$$\frac{a}{a+c} = \frac{a+\Delta}{a+\Delta+c+\varepsilon} \Leftrightarrow \frac{\Delta}{\Delta+\varepsilon} = \frac{a}{a+c} \tag{1}$$

Let us observe that (1) implies that if $c = 0$ then $\varepsilon = 0$ and moreover if $c > 0$ then $\varepsilon > 0$. Since $|U|$ and $sup(\psi)$ must be kept constant, *b* and *d* need to decrease in such a way that $b' = b - \Delta$ and $d' = d - \varepsilon$. In this situation, the new confirmation measure *s* will be:

$$s'(\phi \to \psi) = \frac{a'}{a'+c'} - \frac{b'}{b'+d'} = \frac{a+\Delta}{a+\Delta+c+\varepsilon} - \frac{b-\Delta}{b-\Delta+d-\varepsilon}$$

Remembering that $conf(\phi \to \psi) = \frac{a}{a+c}$ is constant, let us observe that:

$$s'(\phi \to \psi) > s(\phi \to \psi) \Leftrightarrow \frac{b}{b+d} > \frac{b-\Delta}{b-\Delta+d-\varepsilon} \Leftrightarrow$$
$$d\Delta > b\varepsilon \Leftrightarrow d\Delta + b\Delta > b\varepsilon + b\Delta \Leftrightarrow$$
$$\frac{\Delta}{\Delta+\varepsilon} > \frac{b}{b+d} \tag{2}$$

Considering (1) and (2) it can be concluded that:

$$s'(\phi \to \psi) > s(\phi \to \psi) \Leftrightarrow \frac{a}{a+c} > \frac{b}{b+d} \Leftrightarrow s(\phi \to \psi) > 0.$$

This proves that, for a fixed value of confidence, confirmation measure *s* is increasing with respect to rule support if and only if $s(\phi \to \psi) > 0$ and therefore in its positive range confirmation measure *s* is strictly monotone in rule support. □

As rules with negative values of confirmation measure *s* should always be discarded from consideration, the result from *Theorem 4* states the monotone relationship just in the interesting subset of rules.

Confirmation measure *s* was not found in [9] as a satisfying measure with respect to the *property of symmetry*. It was due to the fact that though it does satisfy the desirable property of *hypothesis symmetry* (HS), it also has the undesirable properties of *evidence symmetry* (ES) and *total symmetry* (TS). However, it was proved in [13] that, similarly to *f*, confirmation measure *s* has the desirable property of monotonicity (M).

Since confirmation measure *s* has the property of monotonicity (M), we propose to generate interesting rules by searching for rules maximizing confirmation measure *s* and support, i.e. substituting the confidence in the support-confidence Pareto-optimal border with the confirmation measure *s* and obtaining in this way a support-confirmation-*s* Pareto-optimal border. This approach differs from the idea of finding the Pareto-optimal border according to rule support and confirmation measure *f*, because support-confirmation-*f* Pareto-optimal border contains the same rules as the support-confidence Pareto-optimal border, while in general support-confirmation-*s* Pareto-optimal border contains a subset of the support-confidence Pareto-optimal border as stated in the following theorem.

Theorem 5. If a rule resides on the support-confirmation-*s* Pareto-optimal border (in case of positive value of confirmation measure *s*), then it resides also on the support-confidence Pareto-optimal border, while one can have rules being on the support-confidence Pareto-optimal border which are not on the support-confirmation-*s* Pareto-optimal border.

Proof. Let us consider a rule *r* $\phi\to\psi$ residing on the support-confirmation-*s* Pareto-optimal border and let us suppose that confirmation measure *s* has a positive value. This means that for any other rule *r'* $\phi'\to\psi$ we have that:

$sup(\phi'\to\psi)>sup(\phi\to\psi)\Rightarrow s(\phi'\to\psi)<s(\phi\to\psi)$, (*i*)

On the basis of monotonicity of confirmation measure *s* with respect to support and confidence in case of positive value of *s*, we have that $sup(\phi'\to\psi)>sup(\phi\to\psi)$ and $s(\phi'\to\psi)<s(\phi\to\psi)$ implies that $conf(\phi'\to\psi)<conf(\phi\to\psi)$.

This means that (*i*) implies that for any other rule *r'*

$sup(\phi'\to\psi)>sup(\phi\to\psi)\Rightarrow conf(\phi'\to\psi)<conf(\phi\to\psi)$.

This means that rule *r* residing on the support-confirmation-*s* Pareto-optimal border is also on the support-confidence Pareto-optimal border because one cannot have any other rule *r'* such that

$sup(\phi'\to\psi)>sup(\phi\to\psi)$ and $conf(\phi'\to\psi)\geq conf(\phi\to\psi)$.

Now, we prove with a counter-example that there can be rules being on the support-confidence Pareto-optimal border which are not on the support-confirmation-*s* Pareto-optimal border. Let us consider rules *r* and *r'* residing on the support-confidence Pareto-optimal border such that for rule *r* we have support $sup(\phi\to\psi)=200$ and confidence $conf(\phi\to\psi)=0.667$, while for rule *r'* we have support $sup(\phi'\to\psi)=150$ and confidence $conf(\phi'\to\neg\psi)=0.68$. We have that $s(\phi\to\psi)=0.167$ which is greater than $s(\phi'\to\psi)=0.142$. Thus, rule *r'* is not on the support-confirmation-*s* Pareto-optimal border because it is dominated with respect to support- confirmation-*s* by rule *r* having a greater support and a greater confirmation measure *s*. □

Theorem 5 states that some rules from the support-confidence Pareto-optimal border may not be present on the support-confirmation-*s* Pareto-optimal border. *Theorem 5* can be easily generalized by substituting confirmation measure *s* for any interestingness measure monotone with respect to support and confidence. The following theorem states formally this point.

Theorem 6. Given an interestingness measure *i*, which is monotone with respect to support and confidence, if a rule resides on the support-interestingness-*i* Pareto-optimal border, then it also resides on the support-confidence Pareto-optimal border, while the opposite assertion is not necessarily true.

Proof. Analogous to Theorem 5.

5. Optimal rules with respect to any confirmation measure having the property of monotonicity (M)

The investigation of monotone link with confidence and rule support has also been extended to a more general class of all the confirmation measures that have the property of monotonicity (M). For a set of rules with a fixed conclusion a general analysis has been conducted verifying under what conditions a confirmation measure with the property of monotonicity (M):

- is monotone in confidence when the value of rule support is kept unchanged,
- is monotone in rule support when the value of confidence is held fixed.

Again, for the simplicity of presentation, let us use the following notation:

$a = sup(\phi \to \psi)$,

$b = sup(\neg\phi \to \psi)$,

$c = sup(\phi \to \neg\psi)$,

$d = sup(\neg\phi \to \neg\psi)$.

Let us consider a Bayesian confirmation measure $F(a,b,c,d)$ having the property of monotonicity (M). The analysis concerns only a set of rules with the same conclusion, thus the values of $|U| = a+b+c+d$ and $sup(\psi) = a+b$ are constant.

One can observe that *a*, *b*, *c*, and *d* can be transformed in the following way:

$a = sup(\phi \to \psi)$,

$b = sup(\psi) - sup(\phi \to \psi)$,

$c = \dfrac{1}{conf(\phi \to \psi)} sup(\phi \to \psi) - sup(\phi \to \psi)$,

$d = |U| - sup(\psi) - \dfrac{1}{conf(\phi \to \psi)} sup(\phi \to \psi) +$

$+ sup(\phi \to \psi)$.

Then, a Bayesian confirmation measure *F* can be expressed as:

$F(a,b,c,d) = F(sup(\phi \to \psi),$

$sup(\psi) - sup(\phi \to \psi),$

$\dfrac{1}{conf(\phi \to \psi)} sup(\phi \to \psi) - sup(\phi \to \psi),$

$|U| - sup(\psi) - \dfrac{1}{conf(\phi \to \psi)} sup(\phi \to \psi) + sup(\phi \to \psi)).$

Theorem 7. When the value of rule support is held fixed, then the confirmation measure *F*(*a*, *b*, *c*, *d*) is monotone in confidence.

Proof. Confidence determines the value of Bayesian confirmation measure *F*(*a*, *b*, *c*, *d*) through variables *c* and *d*. We have that variable *c* is non-increasing in confidence. In fact,

$$c = \frac{1}{conf(\phi \to \psi)} sup(\phi \to \psi) - sup(\phi \to \psi)$$

and $sup(\phi \to \psi)$ is non-negative. Since for the property of monotonicity (M), *F* is non-increasing with respect to variable *c*, we get that *F* is non-decreasing with respect to the value of confidence in variable *c*.

We have also that variable *d* is non-decreasing in confidence. In fact,

$$d = |U| - sup(\psi) - \frac{1}{conf(\phi \to \psi)} sup(\phi \to \psi) + sup(\phi \to \psi)$$

and $sup(\phi \to \psi)$ is non-negative. Since for the property of monotonicity (M), *F* is non-decreasing with respect to variable *d*, we get that *F* is non-decreasing with respect to the value of confidence in variable *d*. □

Let us remark that we can say nothing in general about monotonicity with respect to support of a confirmation measure *F*(*a*, *b*, *c*, *d*) satisfying the property of monotonicity (M). In fact confirmation measure *F*(*a*, *b*, *c*, *d*) is clearly non-decreasing with respect to the value of support in variable *a* and *b*, however confirmation measure *F*(*a*, *b*, *c*, *d*) is non-increasing with respect to the value of support in variable *c* and *d*. The latter point merits some explanations. We have that

$$c = \frac{1}{conf(\phi \to \psi)} sup(\phi \to \psi) - sup(\phi \to \psi) = sup(\phi \to \psi)\left(\frac{1}{conf(\phi \to \psi)} - 1\right)$$

and $\left(\dfrac{1}{conf(\phi \to \psi)} - 1\right)$ is non-negative. Since for the property of monotonicity (M), *F* is non-increasing with respect to variable *c*, we get that *F* is non-increasing with respect to the value of support in variable *c*.

Analogously, we have that

$$d = |U| - sup(\psi) - \frac{1}{conf(\phi \to \psi)} sup(\phi \to \psi) + sup(\phi \to \psi) = |U| - sup(\psi) - sup(\phi \to \psi)\left(\frac{1}{conf(\phi \to \psi)} - 1\right)$$

and $\left(\dfrac{1}{conf(\phi \to \psi)} - 1\right)$ is non-negative. Since for the property of monotonicity (M), *F* is non-decreasing with respect to variable *d*, we get that *F* is non-increasing with respect to the value of support in variable *d*.

In order to find a condition for monotonicity of confirmation measure *F*(*a*, *b*, *c*, *d*) with respect to support $sup(\phi \to \psi)$, in the following theorem we suppose that confirmation measure *F*(*a*, *b*, *c*, *d*) is differentiable with respect to all its variables *a*, *b*, *c* and *d*.

Theorem 8. When the value of confidence is held fixed, then the confirmation measure *F*(*a*, *b*, *c*, *d*) admitting derivative with respect to all its variables *a*, *b*, *c* and *d*, is monotone in rule support if:

$$\frac{\partial F}{\partial c} = \frac{\partial F}{\partial d} = 0 \quad \text{or} \quad \frac{\dfrac{\partial F}{\partial a} - \dfrac{\partial F}{\partial b}}{\dfrac{\partial F}{\partial d} - \dfrac{\partial F}{\partial c}} \geq \frac{1}{conf(\phi \to \psi)} - 1.$$

Proof. Let us assume that $conf(\phi \to \psi) = \text{const}$. Let us differentiate *F*(*a*, *b*, *c*, *d*) with respect to $sup(\phi \to \psi)$ Then, we obtain:

$$\frac{\partial F}{\partial sup(\phi \to \psi)} = \frac{\partial F}{\partial a} - \frac{\partial F}{\partial b} + \left(\frac{\partial F}{\partial c} - \frac{\partial F}{\partial d}\right)\left(\frac{1}{conf(\phi \to \psi)} - 1\right) \qquad (4)$$

Since *F* is supposed to satisfy the property of monotonicity (M), it must be non-increasing with respect to *b*, *c* and non-decreasing with respect to *a*, *d*, such that $\dfrac{\partial F}{\partial b} \leq 0$, $\dfrac{\partial F}{\partial c} \leq 0$ and $\dfrac{\partial F}{\partial a} \geq 0$, $\dfrac{\partial F}{\partial d} \geq 0$.

Hence, if $\dfrac{\partial F}{\partial c} = \dfrac{\partial F}{\partial d}$, then $\dfrac{\partial F}{\partial sup(\phi \to \psi)} \geq 0$.

It is clear, that due to the property of monotonicity (M) of *F*, $\dfrac{\partial F}{\partial c} = \dfrac{\partial F}{\partial d}$ if and only if $\dfrac{\partial F}{\partial c} = \dfrac{\partial F}{\partial d} = 0$.

Let us observe, moreover, that if $\frac{\partial F}{\partial c} \neq \frac{\partial F}{\partial d}$, then $\frac{\partial F}{\partial sup(\phi \to \psi)} \geq 0 \Leftrightarrow \frac{\frac{\partial F}{\partial a} - \frac{\partial F}{\partial b}}{\frac{\partial F}{\partial d} - \frac{\partial F}{\partial c}} \geq \frac{1}{conf(\phi \to \psi)} - 1$. □

Theorem 7 states that for a set of rules with the same conclusion, any Bayesian confirmation measure satisfying the property of monotonicity (M) is always non-decreasing with respect to confidence when the value of rule support is kept fixed. Moreover, due to *Theorem 8*, all those confirmation measures that are independent of $sup(\phi \to \neg\psi)$ and $sup(\neg\phi \to \neg\psi)$ are always found monotone in rule support when the value of confidence remains unchanged. However, for a constant value of confidence, Bayesian confirmation measures which do depend on the value of $sup(\phi \to \neg\psi)$ and $sup(\neg\phi \to \neg\psi)$ are also non-decreasing with respect to rule support if and only if they satisfy the following condition:

$$\frac{\frac{\partial F}{\partial a} - \frac{\partial F}{\partial b}}{\frac{\partial F}{\partial d} - \frac{\partial F}{\partial c}} \geq \frac{1}{conf(\phi \to \psi)} - 1 .$$

The general analysis in *Theorem 7* and *Theorem 8* outlines an easy method of verification whether there exists a monotone link between any Bayesian confirmation measure with the property of monotonicity (M), and confidence and rule support, respectively. Confirmation measures that positively undergo such ascertainment are, in our opinion, good candidates for substituting the confidence dimension in the Pareto-optimal border with respect to rule support and confidence proposed by Bayardo and Agrawal [2]. Due to the monotonicity of a confirmation measure in rule support and confidence, the Pareto-optimal border with respect to rule support and a confirmation measure includes rules optimal according to all the metrics that were found on the support-confidence Pareto-optimal border.

The scale of confirmation measures is more useful than that of confidence, therefore, we propose searching for the non-dominated set of rules with respect to rule support and a confirmation measure. In particular, we find confirmation measures *f* and *s* valuable and useful for such application.

6. Beyond support-confidence and support-confirmation Pareto-optimal borders

We believe that the considered Pareto-optimal borders include interesting rules, however, some critical remarks may help to better understand its limitations.

When inducing rules from data we are interested in a set of rules that characterise a given concept (conclusion ψ), rather than in one rule being the best with respect to one or two interestingness measures. In this sense, a Pareto-optimal border is merely a starting point for identification of the interesting set of rules. Precisely, all rules lying in the area delimited by minimum acceptable support, minimum acceptable confidence or confirmation measure, and the Pareto-optimal border with respect to those two dimensions, can be considered interesting. Remark that this area may include dominated rules, however, from the view point from a good representation of a concept ψ, these dominated rules may be found better than some non-dominated rules from outside this area.

The following example clarifies this point. Let us consider induction of rules with a given conclusion ψ from a universe U such that $sup(\psi)$=100 and $|U|$=300. Consider three rules r_1, r_2 and r_3, such that

$sup(r_1)$=80 and $conf(r_1)$=0.8,
$sup(r_2)$=75 and $conf(r_2)$=0.75 and
$sup(r_3)$=100 and $conf(r_3)$=0.4.

We have, therefore, that $r_2 \preceq_{sc} r_1$ and thus r_2 is not on the support-confidence Pareto-optimal border that contains rules r_1 and r_3. Suppose, however, that the user sets the minimum acceptable support to 50, and the minimum acceptable confidence to 0.5. Then, the set of interesting rules with respect to these two dimensions will be composed of rules r_1 and r_2. Remark that even though rule r_3 is non-dominated, it has been found less interesting than the dominated rule r_2.

Another critical issue is related to *Theorem 8*. In fact, it says that a rule maximizing a confirmation measure satisfying the property (M) is on the support-confidence Pareto-optimal border only if a specific condition is satisfied. In other words this means that, in general, not all rules maximizing a confirmation measure satisfying the property (M) are on the support-confidence Pareto-optimal border. Thus, on the basis of the observation that a confirmation measure is more meaningful than confidence, one can think about mining all rules that maximize confirmation measures satisfying the property (M) without taking into account the rule confidence. Let us consider induction of rules with a given conclusion ψ from a universe U such that $sup(\psi)$ and $|U|$ can be considered fixed. Again, for the simplicity of presentation, let us use the following notation:

$a = sup(\phi \to \psi)$,
$b = sup(\neg\phi \to \psi)$,
$c = sup(\phi \to \neg\psi)$,
$d = sup(\neg\phi \to \neg\psi)$.

One can observe that a, b, c, and d can be transformed in the following way:

$a = sup(\phi \to \psi)$,
$b = sup(\psi) - sup(\phi \to \psi)$,
$c = sup(\phi \to \neg\psi)$,
$d = |U| \text{ - } sup(\psi) - sup(\phi \to \neg\psi))$.

Then, a Bayesian confirmation measure F can be expressed as:

$$F(a,b,c,d) = F(sup(\phi \to \psi),$$
$$sup(\psi) - sup(\phi \to \psi),$$
$$sup(\phi \to \neg\psi),$$
$$|U| \text{ - } sup(\psi) - sup(\phi \to \neg\psi)).$$

Let us call $sup(\phi\to\neg\psi)$ the anti-support of rule $\phi\to\psi$. It represents the number of counter-examples to the rule $\phi\to\psi$. For example, if ϕ="x is a raven" and ψ="x is black", then $\phi\to\psi$ is the rule "if x is a raven, then x is black" and the anti-support $sup(\phi\to\neg\psi)$ is the number of non-black ravens.

Theorem 9. When the value of rule support is held fixed, then the confirmation measure $F(a, b, c, d)$ is anti-monotone (non-increasing) in rule anti-support.

Proof. Function F depends on the anti-support $sup(\phi\to\neg\psi)$ through variables c and d. Observe that c is increasing while d is decreasing with respect to $sup(\phi\to\neg\psi)$. Remembering that for the property of monotonicity (M) F is non-increasing in c and non-decreasing in d, we get the thesis. □

Theorem 10. When the value of rule anti-support is held fixed, then the confirmation measure $F(a, b, c, d)$ is monotone (non-decreasing) in rule support.

Proof. Function F depends on the support $sup(\phi\to\psi)$ through variables a and b. Observe that a is increasing with respect to $sup(\phi\to\psi)$ while b is decreasing. Remembering that for the property of monotonicity (M) F is non-decreasing in a and non-increasing in b, we get the thesis.□

Theorem 9 and *Theorem 10* say that F is monotone (non-decreasing) with respect to rule support $sup(\phi\to\psi)$ and anti-monotone (non-increasing) with respect to rule anti-support $sup(\phi\to\neg\psi)$. Therefore, the best rule according to any of these monotone confirmation measures must reside on the support-anti-support Pareto-optimal border being the set of rules such that there is no other rule having greater support and smaller anti-support.. This result is interesting from the viewpoint of searching for some efficient algorithms to mine decision rules optimal in the sense of maximizing the support and minimizing the anti-support. Of course, the first critical remark to Pareto-optimal borders also concern the support-anti-support Pareto-optimal border.

Moreover, this result has an interesting relation with the so called Nicod's Principle. Nicod's Principle [22] says that an evidence confirms an implication "A implies B" if and only if it satisfies both the antecedent and the consequent of the implication; it disconfirms the implication if and only if it satisfies the antecedent, but not the consequent of the implication. Thus, according to the Nicod's principle, an evidence is neutral, or irrelevant, with respect to the implication if it does not satisfy the antecedent.

To illustrate this point Hempel introduced an example which became very well known in the specialized literature. The implication, denoted by (I1), is "if x is a raven, then x is black" or, in everyday language, "All ravens are black". Consider that with respect to the considered implication there are four possible evidences:

(a) a black raven,
(b) a non-black raven (for example, a white raven),
(c) a black non-raven (for example, a black shoe),
(d) a non-black non-raven (for example, a white shoe).

From Nicod's Principle, (a) is a positive instance of the implication, and so (a) confirms the implication "All ravens are black". (b) is a negative instance, and so (b) disconfirms the implication "All ravens are black". (c) and (d) do not satisfy the antecedent of the implication "All ravens are black", (i.e., neither (c) nor (d) is a raven), and so they are non-instances and irrelevant to I1.

With the aim of discussing the Nicod's Principle, Hempel introduced the Equivalence Condition which says "Whatever confirms (disconfirms) one of two equivalent sentences, also confirms (disconfirms) the other" [15].

It seems that the truth of this condition is quite uncontestable. As Hempel claimed, the Equivalence Condition is "a necessary condition" and "fulfillment of this condition makes the confirmation of a hypothesis independent of the way in which it is formulated" [15].

Even if the Nicod's Principle seems so natural and the Equivalence Condition so necessary in the theory of confirmation, when we put the two together, some problems arise.

To illustrate, let us come back to the implication "All ravens are black". Now, we can imagine another implication, denoted in the following by (I2), "All non-black things are not ravens" which is logically equivalent. Using the same logic from Nicod's Principle, evidence (d) will confirm (I2), (b) will disconfirm (I2) while (a) and (c) becomes irrelevant to (I2). Now, we know that (I1) is logically equivalent to (I2). From the Equivalence Condition, whatever confirms (I1) should confirm (I2) as well. However, we have that (a) confirms (I1) but not (I2) and (d) confirms (I2) but not (I1). In this case, we have that the application of Nicod's creates a violation of the Equivalence condition. In this case, as Hempel says, "This means that Nicod's Principle makes confirmation depend not only on the content of the hypothesis, but also on its formulation" [15].

Instead of the Nicod's Principle, Hempel proposes the Positive Instances Principle. Hempel thinks that since (a) confirms (I1) and (d) confirms (I2), and (I1) is logically equivalent to (I2), then both (a) and (d) confirm (I1) and (I2). Moreover, (a) and (d) are not the only evidence that can confirm (I1) and (I2). Now, let us consider the following hypothesis (I3): "Anything which is or is not a raven is either no raven or black". Anything that is not a raven or black can confirm (I3). (I3) is logically equivalent to (I1) and (I2). Therefore, according to the Positive Instances Principle, anything that is not a raven or a black raven confirms (I1) (and (I2)) as well. In other words, (a), (c), and (d) would confirm (I1), while only (b) disconfirms (I1). Greco, Pawlak and Slowinski in [13] further generalized the Hempel criterion by means of the already mentioned monotonicity principle M which can be rephrased as (a) and (d) support the implication, while (b) and (c) are against it. Their arguments were the following. Given a probability Pr, an evidence ϕ confirms a hypothesis ψ, if $\Pr(\psi|\phi)>\Pr(\psi|\neg\phi)$. Translating probability in terms of confidence, one can say that an evidence ϕ confirms hypothesis ψ, if $conf(\psi|\phi)>conf(\psi|\neg\phi)$. Greco, Pawlak and Slowinski [13] proved that it is possible to pass from one situation in which evidence ϕ does not confirm hypothesis ψ, i.e. $conf(\psi|\phi)<conf(\psi|\neg\phi)$, to a situation in which evidence ϕ confirms hypothesis ψ, i.e. $conf(\psi|\phi)>conf(\psi|\neg\phi)$, when $sup(\phi\rightarrow\psi)$ or $sup(\neg\phi\rightarrow\neg\psi)$ increases, or $sup(\neg\phi\rightarrow\psi)$ or $sup(\phi\rightarrow\neg\psi)$ decreases. Thus, it is reasonable to expect that a confirmation measure F is monotone with respect to $sup(\phi\rightarrow\psi)$ and $sup(\neg\phi\rightarrow\neg\psi)$, corresponding to (a) and (d), and anti-monotone with respect to $sup(\phi\rightarrow\neg\psi)$ and $sup(\neg\phi\rightarrow\psi)$, corresponding to (b) and (c).

Within the context of this discussion, *Theorem 9* and *Theorem 10* permit to reconcile the monotonicity (M) of Greco, Pawlak and Slowinski with the Nicod's Principle. In fact, when the conclusion of the implication is fixed any monotone confirmation measure is non-decreasing with respect to positive instance (support) and non-increasing with respect to negative instance (anti-support).

Another interesting observation with respect to the Pareto-optimal border of support-anti-support is that it contains the support-confidence Pareto-optimal border. The following Theorem states formally this point.

Theorem 11. If a rule resides on the support-confidence Pareto-optimal border, then it resides also on the support-anti-support Pareto-optimal border, while one can have rules being on the support-anti-support Pareto-optimal border which are not on the support-confidence Pareto-optimal border.

Proof. Let us consider a rule r $\phi\rightarrow\psi$ residing on the support-confidence Pareto-optimal border. This means that for any other rule r' $\phi'\rightarrow\psi$ we have that:

$$sup(\phi'\rightarrow\psi)>sup(\phi\rightarrow\psi)\Rightarrow conf(\phi'\rightarrow\psi)<conf(\phi\rightarrow\psi), \quad (i)$$

Observe that

$$conf(\phi'\rightarrow\psi)<conf(\phi\rightarrow\psi) \Leftrightarrow \frac{sup(\phi'\rightarrow\psi)}{sup(\phi'\rightarrow\psi)+sup(\phi'\rightarrow\neg\psi)} < \frac{sup(\phi\rightarrow\psi)}{sup(\phi\rightarrow\psi)+sup(\phi\rightarrow\neg\psi)}$$

and since we are supposing

$sup(\phi'\rightarrow\psi)>sup(\phi\rightarrow\psi)$, we get that

$sup(\phi'\rightarrow\neg\psi)>sup(\phi\rightarrow\neg\psi)$.

This means that (i) implies that for any other rule r'

$$sup(\phi'\rightarrow\psi)>sup(\phi\rightarrow\psi)\Rightarrow sup(\phi'\rightarrow\neg\psi)>sup(\phi\rightarrow\neg\psi).$$

This means that rule r residing on the support-confidence Pareto-optimal border is also on the support-anti-support Pareto-optimal border because one cannot have any other rule r' such that

$$sup(\phi'\rightarrow\psi)>sup(\phi\rightarrow\psi) \text{ and } sup(\phi'\rightarrow\neg\psi)>sup(\phi\rightarrow\neg\psi).$$

Now, we prove with a counter-example that there can be one decision rule being on the support-anti-support Pareto-optimal border which is not on the support-confidence Pareto-optimal border. Let us consider two rules r and r' residing on the support-anti-support Pareto-optimal border such that for rule r we have support $sup(\phi\rightarrow\psi)=200$ and anti-support $sup(\phi\rightarrow\neg\psi)=100$, while for rule r' we have support $sup(\phi'\rightarrow\psi)=150$ and anti-support $sup(\phi'\rightarrow\neg\psi)=99$. We have that $conf(\phi\rightarrow\psi)=0.667$ which is greater than $conf(\phi'\rightarrow\psi)=0.602$. Thus, rule r' is not on the support-confidence Pareto-optimal border because it is dominated in the sense of support-confidence by rule r having a larger support and a larger confidence. □

Thus, 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, support-confidence Pareto-optimal border has the disadvantage that it does not present all the rules maximizing a confirmation measure satisfying the property (M). In fact, all the rules present on the support-anti-support Pareto-optimal border and not present on the support-confidence Pareto-optimal border maximize some confirmation measure which is not monotone with respect to support because it does not satisfy the condition of above *Theorem 8*. The support-anti-support Pareto-optimal border has also another advantage. It only depends on the search of frequent and "infrequent" itemsets, independently of the confidence. Indeed, a rule $\phi\rightarrow\psi$ lying on the support-anti-support Pareto-optimal border is "frequent enough" with respect to the pattern $\phi\wedge\psi$ (this frequency corresponds to $sup(\phi\rightarrow\psi)$) and "infrequent enough" with respect to the pattern $\phi\wedge\neg\psi$ (this frequency corresponds to $sup(\phi\rightarrow\neg\psi)$). From an algorithmic viewpoint this should be particularly useful because of the closure property of support and anti-support. In fact,

a) if an itemset is frequent, then all its subsets are also frequent,

b) if an itemset is infrequent, then all its supersets are also infrequent.

Property a) means that support is downward closed, i.e. if an itemset has a required support, then all its subsets also have it. Whereas, property b) means that anti-support is upward closed, i.e. if an itemset has not a required support, then neither of its subsets has it. The conjoint use of properties a) and b) permits to develop an efficient algorithm for finding rules on the support-anti-support Pareto-optimal border. We plan to deal with this issue in the future.

7. Conclusions

Bayardo and Agrawal have proved in [2] that complete preorders of many interestingness measures such as gain, Laplace, lift, conviction, the one proposed by Piatetsky-Shapiro, etc. are implied by the rule support-confidence partial order. This result is practically very useful because it ensures that the rules maximizing any of the above measures are included in the set of Pareto-optimal rules with respect to both rule support and confidence.

In this paper, for a class of rules with the same conclusion, we have analysed the monotonicity of two Bayesian confirmation measures: f and s in rule support when the value of confidence is held fixed, and in confidence when the value of rule support remained unchanged. Those particular measures came into the scope of our interest for their valuable properties and the meaningful semantics of the scale of Bayesian confirmation measures in general.

The analysis has also been extended to a more general class of all the confirmation measures that have the property of monotonicity (M). As the result, precise conditions in which such confirmation measures are monotone in rule support and confidence were presented.

The overall results show that it is reasonable to propose a new approach in which we search for a Pareto-optimal border with respect to rule support and confirmation measure f, and not rule support and confidence as it was suggested in [2]. Due to the monotone link between confirmation measure f and confidence, the new set of non-dominated rules includes the same rules that reside on the support-confidence Pareto-optimal border. Thus, without any loss, we only present the same Pareto-optimal border in a more meaningful way because the semantic utility of confirmation f is higher than that of confidence.

Moreover, we believe that discovering rules optimal with respect to support and confirmation measure s would bring valuable results. This is because, unlike confirmation measure f, measure s is dependent on rule support when the value of confidence is held fixed. The Pareto-optimal border thus obtained is contained in the support-confidence Pareto-optimal border.

We also proposed to search for a Pareto-optimal border with respect to rule support and rule anti-support. Of course, the rule anti-support is minimized. We proved that through the support-anti-support Pareto-optimal border we are able to mine all the rules maximizing any confirmation measure that has the property of monotonicity (M). We also proved that the support-anti-support Pareto-optimal border contains the support-confirmation Pareto-optimal border, but we argue that the disadvantage of having a greater number of rules on the Pareto-optimal border is counter-balanced by the fact that in this way we have also rules maximizing confirmation measures which are not monotone with respect to support and confidence, even if they satisfy the property of monotonicity (M).

Consequently, our future research will concentrate on adapting the APRIORI algorithm [1], based on the frequent itemsets, for mining most interesting rules with respect to rule support and confirmation measure s or with respect to rule support and anti-support. In the latter case, we are especially confident that a very efficient algorithm can be developed using closure properties of support and anti-support.

Acknowledgements

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

References

[1] Agrawal, R., Imielinski, T., Swami, A.., 1993. Mining associations between sets of items in massive databases, in: Proc. of the 1993 ACM-SIGMOD Int'l Conf. on Management of Data, 207-216.

[2] Bayardo, R.J., Agrawal, R., 1999. Mining the most interesting rules, in: Proc. of the Fifth ACM-SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, 145-154.

[3] Brin, S., Motwani, R., Ullman, J. and Tsur, S., 1997. Dynamic itemset counting and implication rules for market basket data, in: Proc. of the 1997 ACM-SIGMOD Int'l Conf. on the Management of Data, 255-264.

[4] Brzezinska, I., Slowinski, R., 2005. Monotonicity of a Bayesian confirmation measure in rule support and confidence, in: T. Burczyński, W. Cholewa, W. Moczulski (Eds.), Recent Developments in Artificial Intelligence Methods, AI-METH Series, Gliwice, 39-42.

[5] Brzezinska, I., Greco, S., Slowinski, R., 2005. Investigation of monotone link between confirmation measures and rule support and confidence. Research Report RA-025/05, Institute of Computing Science, Poznan University of Technology, Poznan.

[6] Carnap, R., 1962. Logical Foundations of Probability, second ed, University of Chicago Press, Chicago.

[7] Christensen, D., 1999. Measuring confirmation. Journal of Philosophy XCVI, 437-461.

[8] Clark, P., Boswell, P., 1991. Rule induction with CN2: some recent improvements, in: Machine Learning: Proc. of the Fifth European Conference, 151-163.

[9] Eells, E., Fitelson, B., 2002. Symmetries and asymmetries in evidential support. Philosophical Studies, 107 (2): 129-142.

[10] Fitelson, B., 2001. Studies in Bayesian Confirmation Theory. Ph.D. Thesis, University of Wisconsin, Madison.

[11] Fukuda, T., Morimoto, Y., Morishita, S., Tokuyama, T., 1996. Data mining using two-dimensional optimized association rules: scheme, algorithms, and visualization, in: Proc. of the 1996 ACM-SIGMOD Int'l Conf. on the Management of Data, 13-23.

[12] Good, I.J., 1984. The best explicatum for weight of evidence. Journal of Statistical Computation and Simulation, 19: 294-299.

[13] Greco, S., Pawlak, Z., Slowinski, R., 2004. Can Bayesian confirmation measures be useful for rough set decision rules? Engineering Applications of Artificial Intelligence, 17: 345-361.

[14] Heckerman, D, 1988. An axiomatic framework for belief updates, in: Kanal, L., Lemmer, J. (Eds), Unconfidence in Artificial Intelligence 2. Elsevier Science Publishers, New York, 11-22.

[15] Hempel, C.G., 1945. Studies in the logic of confirmation (I). Mind 54, 1-26.

[16] Jeffreys, H., 1935. Some tests of significance treated by theory of probability, in: Proceedings of the Cambridge Philosophical Society, 31: 203-222.

[17] Kamber, M., Shingal, R., 1996. Evaluating the interestingness of characteristic rules, in: Proccedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, pp. 263–266.

[18] International Business Machines, 1996. IBM Intelligent Miner User's Guide, Version 1, Release 1.

[19] Joyce, J., 1999. The Foundations of Causal Decision Theory. Cambridge University Press, Cambridge.

[20] Morimoto, Y., Fukuda, T., Matsuzawa, H., Tokuyama, T. and Yoda, K., 1998. Algorithms for mining association rules for binary segmentation of huge categorical databases, in: Proc. of the 24th Very Large Data Bases Conf., 380-391.

[21] Morishita, S., 1998. On classification and regression, in: Proc. of the First Int'l Conf. on Discovery Science – Lecture Notes in Artificial Intelligence, 1532: 40-57.

[22] Nicod, J., 1923. Le probleme de la logique de l'induction. Alcan, Paris.

[23] Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman, San Francisco.

[24] Piatetsky-Shapiro, G., 1991. Discovery, analysis, and presentation of strong rules. Chapter 12,in: Knowledge Discovery in Databases, AAAI/MIT Press.

[25] Slowinski, R. and Greco, S., 2005. Measuring attractiveness of rules from the viewpoint of knowledge representation, prediction and efficiency of intervention, in: Szczepaniak, P., Kacprzyk, J., Niewiadomski, A. (Eds.), Advances in Web Intelligence. LNAI, vol 3528, Springer-Verlag, Berlin, 1-12.

[26] Webb, G.I., 1995. OPUS: An efficient admissible algorithm for unordered search. Journal of Artificial Intelligence Research, 3: 431-465.