Application of Bayesian confirmation measures for mining rules from support-confidence Pareto-optimal set

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

[1] Institute of Computing Science Poznan University of Technology, 60-965 Poznan, Poland

[2] Institute for Systems Research, Polish Academy of Sciences, 01-447 Warsaw, Poland
{Izabela.Brzezinska, Roman.Slowinski}@cs.put.poznan.pl

[3] Faculty of Economics, University of Catania, Corso Italia, 55, 95129 Catania, Italy
salgreco@unict.it

Abstract. We investigate a monotone link between Bayesian confirmation measures and rule support and confidence. In particular, we prove that two confirmation measures enjoying some desirable properties are monotonically dependent on at least one of the classic dimensions being 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. We also provide general conclusions for the monotone link between any confirmation measure enjoying some desirable properties and rule support and confidence.

1 Introduction

Knowledge patterns discovered from data are usually expressed in a form of "*if..., then...*" rules. They are consequence relations representing mutual relationship, association, causation, etc. between independent and dependent attributes. Typically, the number of rules generated from massive datasets is very large, and only a small portion of them is likely to be useful. 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. Measures such as confidence and support, gain [10], conviction [3], etc. have been introduced to capture different characteristics of rules. Among widely studied interestingness measures, there is, moreover, a group of Bayesian confirmation measures, which 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 [9]. An important role is played by a confirmation measure denoted in [9] and other studies by$f$, and by a confirmation measure s proposed by [6]. Both of them have a valuable property of monotonicity (M) introduced in [12].

Bayardo and Agrawal [2] have proved that for a class of rules with fixed conclusion, the upper support-confidence Pareto border (i.e. the set of non-dominated, Pareto-optimal rules with respect to both rule support and confidence) includes optimal rules according to several different interestingness measures, such as gain, Laplace [7], lift [13], conviction [3], an unnamed measure proposed by Piatetsky-Shapiro [16]. This practically useful result allows to identify, the most interesting rules according to several interestingness measures by solving an optimized rule mining problem with respect to rule support and confidence only.

As shown in [12], 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 exist 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, two 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.

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 approaches from sections 3 and 4 to a broader class of confirmation measures. The paper ends with conclusions.

2 Preliminaries

Discovering rules from data is a domain of inductive reasoning. To start inference it uses information about a sample of larger reality. This sample is often given in a form of an information table, containing objects of interest characterized by a finite set of attributes. Let us consider information table $S = (U, A)$, where $U$and A are finite, non-empty sets called *universe* and *set of attributes*, respectively. One can associate a formal language L of logical formulas with every subset of attributes. Conditions for a subset $B \subseteq A$ are built up from attribute-value pairs (a,v), where $a \in B$ and $v \in V_a$ (set V_a is a domain of attribute a), using logical connectives $\neg$ (not), $\wedge$ (and), $\vee$ (or). A *decision rule* induced from S and expressed in L is denoted by $\phi \rightarrow \psi$ (read as "*if* ϕ, *then* ψ") and consists of condition and decision formulas in L, called premise and conclusion, respectively.

In this paper, similarly to [2], we only consider all minimal rules with the same conclusion, which can be induced from a dataset. Let us remind that a rule is minimal if, for a given conclusion, there is no other rule with weaker conditions.

2.1 Monotonicity of a function in its argument

For x belonging to a set ordered by the relation $\succ$ and for the values of g belonging to a set ordered by the relation $\leq$, a function $g(x)$ is understood to be *monotone* (resp. *anti-monotone*) in x, if $x_1 \prec x_2$ implies that $g(x_1) \leq g(x_2)$ (resp. $g(x_1) \geq g(x_2)$).

2.2 Support and confidence measures of rules

With every rule induced from information table S, measures 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 premise ϕ and conclusion ψ evaluate to true.

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

$$conf(\phi \rightarrow \psi) = \frac{sup\,(\phi \rightarrow \psi)}{sup\,(\phi)},\ \ sup(\phi) > 0 \tag{1}$$

Note, that it can be regarded as a conditional probability $Pr(\psi|\phi)$ with which conclusion ψ evaluates to true, given that premise ϕ evaluates to true, however, expressed in terms of frequencies.

2.3 Bayesian confirmation measures f and s

In general, confirmation measures quantify the strength of confirmation that premise ϕ gives to conclusion ψ. All confirmation measures take (desired) positive values in situations where the conclusion of the rule is verified more often when its premise is verified, rather than when its premise is not verified. For the confirmation measures a desired property of monotonicity (M) was proposed in [12]. This monotonicity property says that, given an information system S, a confirmation measure 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 $sup(\phi \rightarrow \neg\psi)$. Among confirmation measures that have property (M) there is confirmation measure f [9] and confirmation measure s [6].

The confirmation measures f and s are defined as follows:

$$f(\phi \rightarrow \psi) = \frac{Pr(\phi|\psi) - Pr(\phi|\neg\psi)}{Pr(\phi|\psi) + Pr(\phi|\neg\psi)}, \tag{2}$$

$$s(\phi \rightarrow \psi) = Pr(\psi|\phi) - Pr(\psi|\neg\phi). \tag{3}$$

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

$$f(\phi \to \psi) = \frac{\mathit{conf}(\psi \to \phi) - \mathit{conf}(\neg\psi \to \phi)}{\mathit{conf}(\psi \to \phi) + \mathit{conf}(\neg\psi \to \phi)}, \tag{4}$$

$$s(\phi \to \psi) = \mathit{conf}(\phi \to \psi) - \mathit{conf}(\neg\phi \to \psi). \tag{5}$$

2.4 Partial order on rules in terms of two interestingness measures

Let us denote by $\preceq_{AB}$ a partial order on rules in terms of any two different interestingness measures A and B. The partial order $\preceq_{AB}$ can be decomposed into its asymmetric part $\prec_{AB}$ and 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

$$\begin{array}{l} A(r_1) \leq A(r_2) \wedge B(r_1) < B(r_2), \text{or} \\ A(r_1) < A(r_2) \wedge B(r_1) \leq B(r_2); \end{array} \tag{6}$$

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

$$A(r_1) = A(r_2) \wedge B(r_1) = B(r_2). \tag{7}$$

2.5 Implication of a total order $\preceq_t$ by partial order $\preceq_{AB}$

Application of some measures that quantify the interestingness of a rule induced from an information table S creates a total order, denoted as $\preceq_t$, on those rules. In particular, measures such as gain, Laplace, lift, conviction, one proposed by Piatetsky-Shapiro, or confirmation measures f and s result in such a total order on the set of rules with a fixed conclusion, ordering them according to their interestingness value.

A total order $\preceq_t$ *is implied by* partial order $\preceq_{AB}$ if:

$$\begin{array}{l} r_1 \preceq_{AB} r_2 \Rightarrow r_1 \preceq_t r_2, \quad \text{and} \\ r_1 \sim_{AB} r_2 \Rightarrow r_1 \sim_t r_2. \end{array} \tag{8}$$

It has been proved by Bayardo and Agrawal in [2] that if a total order $\preceq_t$ is implied by support-confidence partial order $\preceq_{sc}$, 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, when one proves that a total order defined over 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 total order $\preceq_t$ defined over a rule value function $g(r)$ is implied by partial order $\preceq_{AB}$:

$g(r)$ 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.

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

Due to the semantic importance and utility of confirmation measure f, a verification of the monotonicity of confirmation measure f in rule support and confidence has been conducted in [4]. It has been proved that rules maximizing confirmation measure f can be found on the Pareto-optimal support-confidence border. However, the utility of confirmation measure f outranks the utility of confidence. 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, thus useless rules can be filtered out simply by observing the confirmation measure's sign. Therefore, we find it interesting to propose a new Pareto-optimal border – with respect to rule support and confirmation measure f.

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 rule support, has been performed. The following theorems have been proved in [5].

Theorem 1. *Confidence is monotone in confirmation measure f.*

Theorem 2. *Confidence is independent of rule support, and therefore monotone in rule support, when the value of confirmation measure f is held fixed.*

It follows from the above results that rules optimal in confidence lie on the Pareto-optimal border with respect to rule support and confirmation measure f. Even more, the Pareto-optimal border with respect to support and onfirmation measure f is identical with the Pareto-optimal border with respect to support and confidence.

Consequently, 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. Thus, 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. We find it valuable to combine those two measures in the border, as 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.

A computation experiment showing rules in confirmation measure f and rule support has been conducted. A real life dataset containing information about technical state of buses was analyzed. The set consisted of 76 objects described by 8 criteria and divided into 3 decision classes. For one of those classes a set of all rules was generated. The values of confirmation measure f and rule support for those rules were placed on Fig.1. It can be easily observed that the Pareto-optimal set of rules (marked in Fig.1 by squares) includes rules maximizing such interestingness measures as confidence, Laplace, lift (marked in Fig.1 by asterisk), Piatetsky-Shapiro (marked in Fig.1 by a cross).

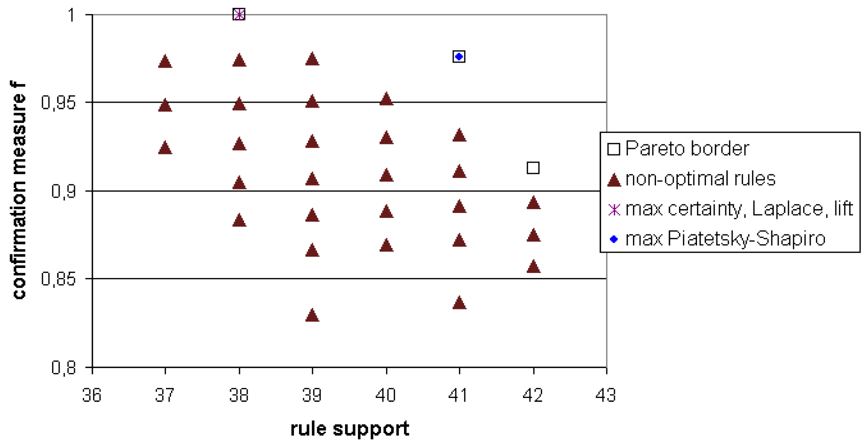

Fig. 1. Pareto-optimal border with respect to rule support and confirmation measure f includes rules being optimal in many other measures (technical state of buses dataset)

For rules with a fixed conclusion, mining the set of non-dominated rules with respect to rule support and confirmation measure f will identify rules optimal according to such interestingness measures as confidence, conviction, lift, Laplace, Piatetsky-Shapiro, gain, etc. However, if those non-dominated rules are characterized by a negative value of confirmation measure f, then they must be discarded because in those rules the premise just disconfirms the conclusion. A final set of rules representing "‘the best"’ 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 support and confirmation measure f.

4 Rules optimal with respect to confirmation measure s

The second confirmation measure that came into the scope of our interest was confirmation measure s. Similarly to confirmation measure f, it also has the desirable property of monotonicity (M). On the contrary to confirmation measure f, however, it is dependent on both rule support and confidence. The monotonicity of confirmation measure s in confidence for a fixed value of support, as well as in rule support for a fixed value of confidence, has been analyzed. The following theorems have been proved in [5].

Theorem 3. *When the rule support value is held fixed, then confirmation measure s is monotone in confidence.*

Theorem 4. *When the confidence value is held fixed, then:*
– *confirmation measure s is monotone in rule support if and only if $s \geq 0$,*

– *confirmation measure s is anti-monotone in rule support if and only if $s < 0$.*

As rules with negative values of confirmation measure s are discarded from consideration, the result from *Theorem 4* states the monotone relationship just in the interesting subset of rules. 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, the support-confirmation-s Pareto-optimal border can differ from the support-confidence Pareto-optimal border. Moreover, as measure f, unlikely to s, is a satisfying confirmation measure with respect to the *property of symmetry* verified in [8], mining the Pareto-optimal border with respect to rule support and confirmation measure f still remains a good alternative idea.

5 Rules optimal with respect to any confirmation measure having the property of monotonicity (M)

A general analysis of the monotonicity of any confirmation measure that enjoys the property of monotonicity (M) has also been conducted.

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 $F(a, b, c, d)$ being differentiable and having the property of monotonicity (M). The following theorems have been proved in [5].

Theorem 5. *When the value of rule support is held fixed, then the confirmation measure $F(a, b, c, d)$ is monotone in confidence.*

Theorem 6. *When the value of confidence is held fixed, then the confirmation measure $F(a, b, c, d)$ is monotone in rule support if:*

$$\frac{\partial F}{\partial c} = \frac{\partial F}{\partial d} = 0 \ \ or \ \ \frac{\frac{\partial F}{\partial a} - \frac{\partial F}{\partial b}}{\frac{\partial F}{\partial d} - \frac{\partial F}{\partial c}} \geq \frac{1}{conf(\phi \to \psi)} - 1. \qquad (9)$$

It is worth noting, that, due to *Theorem 6*, all those confirmation measures that are independent of $sup(\phi \to \neg\psi)$ and $sup(\neg\phi \to \neg\psi)$ are found monotone in rule support when the value of confidence is kept unchanged.

Theorem 5 and *Theorem 6* outline an easy method of verification of existence of the monotone link between any confirmation measure with the property of monotonicity (M), and rule support and confidence. Confirmation measures that positively undergo such verification 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 in [2]. Thanks to the monotonicity of a confirmation measure in rule support and confidence, a monotone link of that confirmation measure with other interestingness measures such as lift, gain, Laplace, etc. is assured. Therefore, the Pareto-optimal border with respect to rule support and a confirmation measure includes rules optimal according to the same metrics as the support-confidence Pareto-optimal border. Due to the fact that the scale of confirmation measures is more useful than that of confidence, we propose searching for the non-dominated set of rules with respect to rule support and a confirmation measure with the propery of monotonicity (M). We find confirmation measure f particularly valuable for its property of monotonicity (M) and for being a satisfying measure with respect to the property of symmetry, and confirmation measure s for its property of monotonicity (M) and its simplicity.

6 Conclusions

Bayardo and Agrawal have opted in [2] for an approach to mining interesting rules based on extracting a Pareto-optimal border with respect to rule support and confidence. We have analyzed and described the monotone link between the confirmation measures f and s, and rule support and confidence. This analysis has also been extended to a more general class of all the confirmation measures that have the property of monotonicity (M). The 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 a confirmation measure, in particular, we are in favor of confirmation measure f or s. Consequently, our future research will concentrate on adapting the "APRIORI" algorithm [1], based on the frequent itemsets, for mining most interesting association rules with respect to rule support and either confirmation measure f or s.

Acknowledgements

The research of the first two authors has been supported by the Polish Ministry of Education and Science. The third author wishes to acknowledge financial support from the Italian Ministry of Education, University and Scientific Research (MIUR).

References

1. Agrawal, R., Imielinski, T. and Swami, A.: 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 (1993).

2. Bayardo, R.J., Agrawal, R.: Mining the most interesting rules. [In]: *Proc. of the Fifth ACM-SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, 145-154 (1999).
3. Brin, S., Motwani, R., Ullman, J. and Tsur, S.: 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 (1997).
4. Brzezinska, I., Slowinski, R.: Monotonicity of a Bayesian confirmation measure in rule support and confidence. [In]: T. Burczynski, W. Cholewa, W. Moczulski (Eds.) *Recent Developments in Artificial Intelligence Methods*, AI-METH Series, Gliwice, 39-42 (2005).
5. Brzezinska, I., Greco, S., Slowinski, R.: *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, (2005).
6. Christensen, D.: Measuring confirmation. *Journal of Philosophy XCVI*, 437-461 (1999).
7. Clark, P. and Boswell, P.: Rule induction with CN2: Some recent improvements. [In]: *Machine Learning: Proc. of the Fifth European Conference*, 151-163 (1991).
8. Eells, E., Fitelson, B.: Symmetries and asymmetries in evidential support. *Philosophical Studies*, 107 (2): 129-142 (2002).
9. Fitelson, B.: *Studies in Bayesian Confirmation Theory.* Ph.D. Thesis, University of Wisconsin, Madison (2001).
10. Fukuda, T., Morimoto, Y., Morishita, S. and Tokuyama, T.: 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 (1996).
11. Good, I.J.: The best explicatum for weight of evidence. *Journal of Statistical Computation and Simulation*, 19: 294-299 (1984).
12. Greco, S., Pawlak, Z. and Slowinski, R.: Can Bayesian confirmation measures be useful for rough set decision rules? *Engineering Applications of Artificial Intelligence*, 17: 345-361 (2004).
13. International Business Machines: *IBM Intelligent Miner User's Guide*, Version 1, Release 1 (1996).
14. Morimoto, Y., Fukuda, T., Matsuzawa, H., Tokuyama, T. and Yoda, K.: Algorithms for mining association rules for binary segmentation of huge categorical databases. [In]: *Proc. of the 24^{th} Very Large Data Bases Conf.*, 380-391 (1998).
15. Morishita, S.: On classification and regression. [In]: *Proc. of the First Int'l Conf. on Discovery Science – Lecture Notes in Artificial Intelligence*, 1532: 40-57 (1998).
16. Piatetsky-Shapiro, G.: Discovery, Analysis, and Presentation of Strong Rules. Chapter 12 [in]: *Knowledge Discovery in Databases*, AAAI/MIT Press (1991).